Eur. J. Biochem. 224, 519–524 (1994)

Sequence similarities and evolutionary relationships of microbial, plant and animal α-amylases

Štefan JANEČEK

Institute of Ecobiology, Slovak Academy of Sciences, Bratislava, Slovakia

(Received March 22/June 17, 1994) – EJB 94 0392/3

Amino acid sequence comparison of 37 α-amylases from microbial, plant and animal sources was performed to identify their mutual sequence similarities in addition to the five already described conserved regions. These sequence regions were examined from structure/function and evolutionary perspectives. An unrooted evolutionary tree of α-amylases was constructed on a subset of 55 residues from the alignment of sequence similarities along with conserved regions. The most important new information extracted from the tree was as follows: (a) the close evolutionary relationship of *Alteromonas haloplanctis* α-amylase (thermolabile enzyme from an antarctic psychrotroph) with the already known group of homologous α-amylases from streptomycetes, *Thermomonospora curvata*, insects and mammals, and (b) the remarkable 40.1% identity between starch-saccharifying *Bacillus subtilis* α-amylase and the enzyme from the ruminal bacterium *Butyrivibrio fibrisolvens*, an α-amylase with an unusually large polypeptide chain (943 residues in the mature enzyme). Due to a very high degree of similarity, the whole amino acid sequences of three groups of α-amylases, namely (a) fungi and yeasts, (b) plants, and (c) *A. haloplanctis*, streptomycetes, *T. curvata*, insects and mammals, were aligned independently and their unrooted distance trees were calculated using these alignments. Possible rooting of the trees was also discussed. Based on the knowledge of the location of the five disulfide bonds in the structure of pig pancreatic α-amylase, the possible disulfide bridges were established for each of these groups of homologous α-amylases.

Although all α-amylases catalyze the hydrolysis of α-1,4-D-glucosidic bonds, their amino acid sequences have high variability [1]. Generally, five highly conserved sequence regions [2, 3] are described. Despite the very low sequence identity (10% or less) [2], the α-amylases have been grouped in one family of similar sequences along with other enzymes involved in starch metabolism [4]. As indicated from prediction studies [1, 5] and confirmed by the crystal structures of four α-amylases (pig pancreas [6], *Aspergillus niger* [7], *Aspergillus oryzae* [8] and barley [9]), the α-amylases contain a $(\beta/\alpha)_8$-barrel domain [10].

More than 50 complete amino acid sequences of α-amylases from different microbial, plant and animal sources are currently known [11, 12], the vast majority being deduced from the DNA sequences. The total number of available α-amylase sequences is much higher due to the presence of multiple α-amylase genes or isozymes in one organism. Although there is extensive information on α-amylases sequences from different sources, an overview of their diversity is lacking. Several pairs or groups of α-amylases are known to be homologous, such as liquefying *Bacillus* α-amylases [13], the α-amylases from *A. oryzae* and *Saccharomycopsis fibuligera* [14], *Aeromonas hydrophila* and *Xanthomonas campestris* α-amylases [15], the *Streptomyces* α-amylases along with *Thermomonospora curvata*, *Drosophila melanogaster* and mammalian α-amylases [16, 17], and the α-amylases from plants [18]. For all these examples, sequence analysis remains to be performed.

Therefore, the main goal of the present study was to summarize the findings resulting from the analysis of amino acid sequences of available microbial, plant and animal α-amylases. To further elucidate the evolutionary aspects of α-amylases, unrooted distance trees were constructed and discussed.

Correspondence to Š. Janeček, Institute of Ecobiology, Slovak Academy of Sciences, Štefánikova 3, SK-81434 Bratislava, Slovakia
Fax: +42 7 334967.

Enzymes. α-Amylase (EC 3.2.1.1); cyclodextrin glycosyltransferase (EC 2.4.1.19).

MATERIALS AND METHODS

Amino acid sequences of α-amylases (Table 1) were extracted from the Swiss-Prot protein and GenBank DNA sequence data banks [11, 12].

Conserved positions and semi-conservative substitutions from the author's previous alignment of eight α-amylases [3] and other alignments [1, 15, 16, 19] were used in searching for sequence similarities in the present set of α-amylases.

The identified similarities, along with the isolated five conserved sequence regions of each α-amylase [2, 3] were aligned to give sub-sets of 55 amino acid residues. An unrooted distance tree was calculated for 30 out of 37 α-amylases based on this alignment by the neighbour-joining method [20] implemented in the program CLUSTAL V [21]. The seven α-amylase sequences, three of which (*A. niger* and *Aspergillus shirousami*, and *Streptomyces griseus*) were in the sequence sub-sets identical to *A. oryzae* and *S. limosus* enzymes, respectively, and four of which (mouse, rat and human pancreas, and human saliva) were different from their pancreatic (pig pancreas) and salivary (mouse saliva) coun-

Table 1. α-Amylases included in the present study. All Swiss-Prot database numbers start with P.

Abbreviation	Organism (specification)	SwissProt or GenBank accession number
Aerhy	*Aeromonas hydrophila*	P22630
Altha	*Alteromonas haloplanctis*	P29957
Bacam	*Bacillus amyloliquefaciens*	P00692
Bacli	*Bacillus licheniformis*	P06278
Bacme	*Bacillus megaterium*	P20845
Bacst	*Bacillus stearothermophilus*	P06279
Bacsu	*Bacillus subtilis*	P00691
Butfi	*Butyrivibrio fibrosolvens*	P30269
Dicth	*Dictyoglomus thermophilum* (amy B)	P14898
Escco	*Escherichia coli*	P26612
Micsp	*Micrococcus* sp.	X55799
Salty	*Salmonella typhimurium*	P26613
Strgr	*Streptomyces griseus*	P30270
Strhy	*Streptomyces hygroscopicus*	P08486
Strli	*Streptomyces limosus*	P09794
Strth	*Streptomyces thermoviolaceus*	P27350
Strvi	*Streptomyces violaceus*	P22998
Thecu	*Thermomonospora curvata*	P29750
Xanca	*Xanthomonas campestris*	M85252
Aspni	*Aspergillus niger*	X52755
Aspor	*Aspergillus oryzae*	P10529
Aspsh	*Aspergillus shirousami*	P30292
Sacfi	*Saccharomycopsis fibuligera*	P21567
Schoc	*Schwanniomyces occidentalis*	P19269
Barle	*Hordeum vulgare* (barley isozyme A)	P00693
Maize	*Zea mays* (maize)	L25805
Ricea	*Oryza sativa* (rice isozyme 1B)	P17654
Vigmu	*Vigna mungo* (rice bean, black gram)	P17859
Wheat	*Triticum aestivum* (wheat amy 3)	P08117
Anoga	*Anopheles gambiae*	L04753
Drome	*Drosophila melanogaster*	P08144
Muspa	*Mus musculus* (mouse pancreas)	P00688
Mussa	*Mus musculus* (mouse saliva)	P00687
Pigpa	*Sus scrofa* (pig pancreas)	P00690
Ratpa	*Rattus norvegicus* (rat pancreas)	P00689
Humpa	*Homo sapiens* (human pancreas)	P04746
Humsa	*Homo sapiens* (human saliva)	P04745

terparts in only one residue, were not used in the construction of this tree.

Whole amino acid sequences of three groups of α-amylases with high degree of mutual sequence similarities that formed compact clusters in the evolutionary tree were aligned independently. These alignments were made using the program CLUSTAL V [21]. The groups were (a) fungi and yeasts, (b) plants, and (c) streptomycetes, insects and mammals along with *Alteromonas haloplanctis* and *T. curvata*. Based on these alignments, the unrooted distance trees were constructed by the neighbour-joining method [20]. The cysteine residues involved in probable disulfide bridges of the third group of α-amylases represented by streptomycetes and animals were extracted from the amino acid sequence alignment of this group by using the known cystines of pig pancreatic α-amylase [22] as the template.

RESULTS AND DISCUSSION

Sequence similarities

α-Amylases contain five conserved sequence regions, four of which are well established [23] and usually noted in new α-amylase sequences. The fifth conserved region has been demonstrated only recently [3]. It comprises the sequence region around Asp167 (pig pancreatic α-amylase numbering) involved in the binding of Ca^{2+} [6]. For example, the regions in the amino acid sequence of pig pancreatic α-amylase are as follows: 1−95 DAVINH 63 LLDLA 23 GFRLDASKH 31 EVID 58 FVDNHD 301−496; the numbers indicate the size of the sequence between the different regions and the sizes of the start and end of the sequence. The present study adds four sequence similarities (Fig. 1) to these conserved regions. These regions were identified along the sites where certain single amino acid residues are strongly conserved or where so-called semi-conservative substitutions (hydrophobic or hydrophilic residues) are present. Taking into account the refined three-dimensional structure of the $(\beta/\alpha)_8$-barrel of pig pancreatic α-amylase [6], the locations of short regions found in this study to be similar in all α-amylases are as follows: region I (FEW, pig pancreatic α-amylase in Fig. 1), loop 1; region II (GFGGVQVSP), strand β2; region III (SYND), the longest loop 3 (domain B); region IV (GFTRVMSSY), strand β8. Both the loops (loop 1, loop 3) are the loops joining the C-terminus of a β-strand to the N-terminus of the adjacent helix, i.e. the loops in β-loop-α units. For comparison, the locations of previously described conserved regions are also in β-strands and loops, i.e. none of the α-helices of the $(\beta/\alpha)_8$-barrel is involved in the conserved regions. This reflects that the sequence requirements of the inner β-barrel are more stringent than the requirement of the outside helical cylinder. This is a feature characteristic of all known $(\beta/\alpha)_8$-barrel starch hydrolases and related enzymes [24].

There are two main differences between the previously described conserved regions [1−3] and the sequence simi-

		I		II		III		IV	
Aerhy	*1-6*	FNW	*15*	GYKQVLISP	*108*	DWSD	*177*	GSPLVYSDH	*332-443*
Altha	*1-8*	FEW	*16*	GYAAVQVSP	*89*	DYGN	*167*	GYPKVMSSY	*306-453*
Bacam	*1-8*	FEW	*22*	GITAVWIPP	*119*	DWDE	*191*	GYPQVFYGD	*366-483*
Bacli	*1-10*	FEW	*22*	GITAVWIPP	*119*	DWDE	*189*	GYPQVFYGD	*366-483*
Bacme	*1-18*	YVN	*37*	GIWMMPVNP	*83*	SWGQ	*185*	GNPYIYYGE	*349-493*
Bacst	*1-11*	FEW	*22*	GITALWLPP	*119*	DWDE	*190*	GYPCVFYGD	*368-515*
Bacsu	*1-14*	WNW	*15*	GYTAIQTSP	*87*	NWSD	*165*	GSTPLFFSR	*307-619*
Butfi	*1-120*	FCW	*15*	GYTAVQTSP	*95*	GYTD	*198*	GTPLFFSRP	*454-943*
Dicth	*1-139*	FID	*36*	GINTIWISP	*78*	DYET	*160*	AIPIIYNGQ	*439-562*
Escco	*1-9*	FHW	*22*	GINMVWLPP	*141*	GWND	*172*	GVPSVFYPD	*370-495*
Micsp	*1-353*	LTD	*43*	GVNTIWISP	*90*	RFSD	*180*	GQPVIYYGE	*692-1104*
Salty	*1-9*	FHW	*22*	GINMVWLPP	*141*	GWND	*172*	GVPSVFYPD	*370-494*
Strgr	*1-12*	FEW	*16*	GYGYVQVSP	*89*	DYGN	*163*	GSPDVHSGY	*306-538*
Strhy	*1-12*	FER	*16*	GYGYVEVSP	*87*	DYTN	*159*	GSPNVYSGY	*300-448*
Strli	*1-12*	FEW	*16*	GYGYVQVSP	*89*	DYGN	*163*	GSPDVHSGY	*306-538*
Strth	*1-11*	FEW	*16*	GYGYVQVSP	*89*	DYQD	*163*	GAPDINSGY	*305-431*
Strvi	*1-12*	FEW	*16*	GYGYVQVSP	*89*	NYQD	*163*	GSPDVHSGY	*306-541*
Thecu	*1-12*	FQW	*16*	GFGAVQVSP	*98*	NWND	*166*	GTPKVMSSY	*318-572*
Xanca	*1-6*	FNW	*15*	GYRKVLVAP	*108*	NYND	*177*	GVPMVYTDN	*332-440*
Aspni	*1-14*	LTD	*38*	GFTAIWITP	*89*	NYED	*165*	GIPIIYAGQ	*332-477*
Aspor	*1-14*	LTD	*38*	GFTAIWITP	*89*	NYED	*165*	GIPIIYAGQ	*332-478*
Aspsh	*1-14*	LTD	*38*	GFTAIWITP	*89*	NYED	*165*	GIPIIYAGQ	*332-478*
Sacfi	*1-15*	VTD	*38*	GFTAIWISP	*89*	NYDD	*165*	GIPVIYYGQ	*333-468*
Schoc	*1-25*	VTD	*38*	GFTAIWISP	*89*	NYND	*165*	GIPIIYYGQ	*343-487*
Barle	*1-7*	FNW	*23*	GVTHVWLPP	*87*	KYSD	*184*	GIPCIFYDH	*327-414*
Maize	*1-7*	FNW	*23*	GATHVWLPP	*87*	QYSN	*181*	GTPCIFYDH	*324-412*
Ricea	*1-6*	FNW	*23*	GITHVWLPP	*86*	PYGD	*184*	GNPCIFYDH	*325-403*
Vigmu	*1-6*	FNW	*22*	GITHVWLPP	*87*	AYSD	*181*	GTPSIFYDH	*322-398*
Wheat	*1-7*	FNW	*23*	GATHVWLPP	*86*	KYSN	*166*	GIPCIFYDH	*308-389*
Anoga	*1-16*	FEW	*16*	GYGGVQLSP	*101*	DWGN	*176*	GQLRIMSSF	*335-489*
Drome	*1-16*	FEW	*16*	GYAGVQVSP	*94*	NYND	*179*	GTPRVMSSF	*331-476*
Muspa	*1-16*	FEW	*16*	GFGGVQVSP	*102*	NYND	*180*	GFTRVMSSY	*340-493*
Mussa	*1-16*	FEW	*16*	GFAGVQVSP	*105*	NYQD	*180*	GFTRVMSSY	*343-496*
Pigpa	*1-16*	FEW	*16*	GFGGVQVSP	*105*	SYND	*180*	GFTRVMSSY	*343-496*
Ratpa	*1-16*	FEW	*16*	GFGGVQVSP	*102*	NYND	*180*	GFTRVMSSY	*340-493*
Humpa	*1-16*	FEW	*16*	GFGGVQVSP	*105*	NYND	*180*	GFTRVMSSY	*343-496*
Humsa	*1-16*	FEW	*16*	GFGGVQVSP	*105*	NYND	*180*	GFTRVMSSY	*343-496*
				* *					
Cons		f w		G v sP		y d		g p v	

Fig. 1. Sequence similarities in α-amylases. The abbreviations of enzyme sources are given in Table 1. α-Amylases are numbered from the N-terminus of the mature (when known) enzymes. The numbers represent the number of amino acid residues between two regions and the sizes of the peptide chain preceding the first and ending the last region, respectively. Invariable amino acid residues are indicated (*). A residue is written in the consensus (*Cons*) sequence if it is present in more than half of the enzymes.

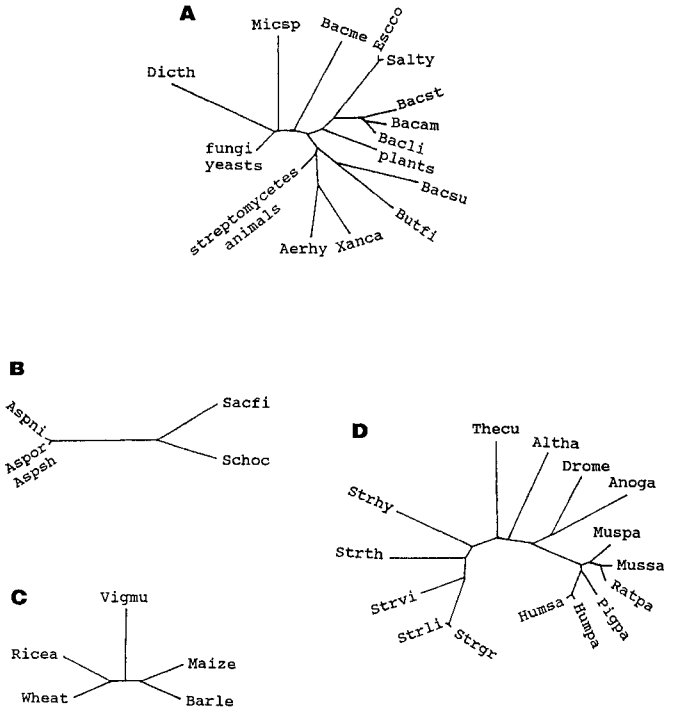

Fig. 2. Unrooted evolutionary trees of α-amylases. The abbreviations of enzyme sources are given in Table 1. The branch lengths are proportional to the divergency of the amino acid sequences of α-amylases, the sum of the lengths of the branches linking any α-amylases being a measure of the evolutionary distance between them. All different α-amylases are presented in A. The three groups of mutually homologous enzymes are depicted in detail in B (fungi and yeasts), C (plants) and D (streptomycetes and animals along with *A. haloplanctis* and *T. curvata* α-amylases).

larities shown in Fig. 1. The first difference is the higher degree of mutual similarity among the conserved regions than in the regions in Fig. 1. The second and more important difference is the presence of amino acid residues involved in the active site and binding of Ca^{2+} and Cl^- in the structure of pig pancreatic α-amylase [6]: 25% of the content of conserved regions versus 4% for sequence similarities. As for the potential catalytic residues alone (the invariant residues Asp197, Glu233 and Asp300; pig pancreatic α-amylase numbering), all are in the most conserved regions.

Nevertheless, the similarity is obvious throughout the whole set of α-amylases (Fig. 1). Considering the similarity of the amino acid sequences of α-amylases (10% or less), there are no doubts of the importance of the regions shown in Fig. 1. The impact of these regions on the function of various α-amylases is difficult to define but, as will be shown below, from the evolutionary point of view their importance is comparable with the importance of conserved regions.

Evolution of α-amylases

Evolutionary relationships of α-amylases are shown in Fig. 2. The unrooted distance tree (Fig. 2A) was based on the alignment of sequence similarities (Fig. 1) along with the five conserved regions. There were three compact clusters containing (a) fungal and yeast enzymes, (b) plant enzymes, and (c) the enzymes from streptomycetes, insects and mammals along with *T. curvata* and *A. haloplanctis* α-amylases. Since each one of these three groups comprises homologous α-amylases, their entire amino acid sequences were aligned independently (the alignments are not shown) and used for construction of unrooted distance trees (Fig. 2B–D). Since the trees (Fig. 2A and Fig. 2B–D) are based on different alignments (Fig. 2A, sequence regions; Fig. 2B–D, entire sequences), the branches linking the enzymes in these trees are not mutually fully compatible. To place the three trees (Fig. 2B–D) into the tree shown in Fig. 2A, the branches would have to be shortened to approximately 50% of their lengths.

The tree in Fig. 2A reflects several facts already reported, such as the high degree of sequential similarity between the α-amylases of *Bacillus amyloliquefaciens*, *Bacillus licheniformis* and *Bacillus stearothermophilus* [13] (liquefying α-amylases), between the *A. hydrophila* and *Xanthomonas campestris* [15] enzymes (next to the cluster of streptomycetes and animal α-amylases), and between those from *Escherichia coli* and *Salmonella typhimurium* [25] (cytoplasmic α-amylases). Although the three above-mentioned groups of α-amylases (fungi and yeasts, plants, and streptomycetes and animals) are already described as homologous [14, 16–18], their individual evolutionary trees (Fig. 2B–D) for most of the available sequences are presented in this study.

Two other important points are presented in this investigation: (a) the α-amylase from the antarctic psychrotroph *A. haloplanctis* can be added to the previously identified group representing streptomycetes, *T. curvata*, *D. melanogaster* and mammals [16, 17] (Fig. 2D), and (b) the grouping together of the α-amylases from *Butyrivibrio fibrisolvens* (with unusually large polypeptide chains, 943 amino acid residues of the mature enzyme) and *Bacillus subtilis* (Fig. 2A). Although their branches in the tree are relatively long, the amino acid sequence of *B. fibrisolvens* α-amylase has 40.1% identity (the ratio of identical amino acid residues calculated using the number of residues of the smaller enzyme) with the sequence of *B. subtilis* α-amylase. As for the α-amylase from *A. haloplanctis*, the amino acid sequence of this enzyme has 49.0% identity with the sequence of pig pancreatic α-amylase (Fig. 3). This bacterial α-amylase appears next to the insect (animal) α-amylases (Fig. 2D) represented by *Anopheles gambiae* and *D. melanogaster* enzymes, while the α-amylases from streptomycetes are more distantly related to those of mammals. The location of the enzymes from *S. griseus*, *S. limosus* and *S. violaceus* may reflect that the sequences of these enzymes comprise an additional putative raw-starch-binding C-terminal domain not present in the other α-amylases [26].

The evolutionary importance of the sequence similarities (Fig. 1) is further supported by the fact that these regions (except for region I) can be identified in the sequences of different starch hydrolases and related enzymes [24]. Recently, Jespersen et al. [24] have aligned the amino acid sequences of such a set of enzymes and constructed an evolutionary tree. This tree was based on the four well-accepted conserved regions. Interestingly, when a distance tree was calculated using the five conserved regions and three (Fig. 1, II–IV) of the four sequence similarities (data not shown), the resulting evolutionary conclusions are similar to those given for the tree reported by Jespersen et al. [24]. One could expect that if more variable regions are included in the construction of a distance tree, then the background noise will increase. This is not the case, however, for the sequence similarity regions of α-amylases (Fig. 1), indicating the comparable evolutionary importance of the four well-accepted con-

```
Pigpa   QYAPQTQSGRTDIVHLFEWRWVDIALEcERYLGPKGFGGVQVSPPNENVVVTNPS   55
Altha   --------TPTTFVHLFEWNWQDVAQEcEQYLGPKGYAAVQVSPPNEHIT----G   43
                *  ****** * * * *** ******   ********

Pigpa   RPWWERYQPVSYKLcTRSGNENEFRDMVTRcNNVGVRIYVDAVINHMCGSGAAAG  110
Altha   SQWWTRYQPVSYELQSRGGNRAQFIDMVNRcSAAGVDIYVDTLINHM---AAGSG   95
          ** ******** *  * **   * *** **   ** ****  ****     *  *

Pigpa   TGTTcGSYCNPGNREFPAVPYSAWDFNDGKcKTASGGIESYNDP-YQVRDcQLVG  164
Altha   TGTAGNSF---GNKSFPI--YSPQDFHES-cTINN---SDYGNDRYRVQNcELVG  141
        ***   *    **  **   **  **    *          *    * *  * ***

Pigpa   LLDLALEKDYVRSMIADYLNKLIDIGVAGFRLDASKHMWPGDIKAVLDKLHNLNT  219
Altha   LADLDTASNYVQNTIAAYINDLQAIGVKGFRFDASKHVAASDIQSLMAKVN----  191
        * **     **   ** * * *  *** *** *****    **     *

Pigpa   NWFPAGSRPFIFQEVIDLGGEAIKSGEYFSNGRVTEFKYGAKLGTVVRKWSGEKM  274
Altha   ------GSPVVFQEVIDQGGEAVGASEYLSTGLVTEFKYSTELGNTFRN---GSL  238
              *  ****** ****    ** * * ******   **   *

Pigpa   SYLKNWGEGWGFMPSDRALVFVDNHDNQRGHGAGGSSILTFWDAYRKLVAVGFML  329
Altha   AWLSNFGEGWGFMPSSSAVVFVDNHDNQRGHGGAGN-VITFEDGRLYDLANVFML  292
          * * **********  * **************  *     ** *      *  ***

Pigpa   AHPYGFTRVMSSYRWARNFVNGEDVNDWIGPPNNNGVIKEVTINADTTc-GNDWV  383
Altha   AYPYGYPKVMSSY----DFHGDTDAGGPNVPVHNNGNLE---------cFASNWK  334
        * ***   *****     *    *      *  ***              *   *

Pigpa   cEHRWREIRNMVWFRNVVDGE-PFANWWDNGSNQVAFGRGNRGFIVFNNDDWQLS  437
Altha   cEHRWSYIAGGVDFRNNTADNWAVTNWWDNTNNQISFGRGSSGHMAINKEDSTLT  389
        *****  *   * ***          *****  **  ****  *    *  *  *

Pigpa   STLQTGLPAGTYcDVISGDKVGN--ScTGIKVYVSSDGKAQFSISNSAEDPFIAI  490
Altha   ATVQTDMASGQYcNVLKGELSADAKScSGEVITVNSDGTINLNIGAWDA---MAI  441
          * **    * ** *  *       ** *    * ***     *         **

Pigpa   HAESKL------   496
Altha   HKNAKLNTSSAS   453
        *   **
```

Fig. 3. Amino acid sequence alignment of the α-amylases from pig pancreas (Pigpa) and *A. haloplanctis* (Altha). The sequences are numbered from the N-terminus of the mature enzymes. Identical amino acid residues (*) and gaps (−) are indicated. The ten cysteine residues involved in the five disulfide bridges of pig pancreatic α-amylase [22] and the eight equivalent cysteine residues of the enzyme from *A. haloplanctis* are shown in lower-case letters and are italicized.

served regions [1, 2], the fifth conserved region [3] and the sequence similarities (this study, Fig. 1).

The evolutionary relationships of fungal and yeast α-amylases are shown in Fig. 2B. Taxonomy is respected despite the very similar amino acid sequences, and fungi and yeasts form their own evolutionary groups. The fungal α-amylases are mutually more closely related than yeast enzymes. The position of these α-amylases in the whole tree (Fig. 2A) next to the enzymes with the longest branches (*Dictyoglomus thermophilum*, *Micrococcus* sp., *Bacillus megaterium*) is remarkable.

The plant α-amylases appear between the liquefying and saccharifying α-amylases of the genus *Bacillus* (Fig. 2A). These α-amylases form another group of closely related enzymes (Fig. 2C). They differ from the rest of the α-amylases especially with respect to the sequence in the third conserved region, e.g. AWRLDFARG in barley (low pI isozyme) compared with GFRLDASKH in pig pancreatic α-amylase.

As mentioned above, the α-amylases from *D. thermophilum*, *Micrococcus* sp. and *B. megaterium* are on the longest branches of the tree in Fig. 2A. This indicates that these enzymes least resemble the other enzymes. Indeed, several distinct differences from the rest of the α-amylases can be found also in the regions of sequence similarity shown in Fig. 1, such as threonine residues at the fourth position of region II for *D. thermophilum* and *Micrococcus* sp. α-amylases and a proline residue at the sixth position of this region for the *B. megaterium* enzyme. Moreover, *D. thermophilum* and *B. megaterium* enzymes possess an intermediary sequence in region I, i.e. Phe (Tyr) and Asp (Asn) at the first and the third positions, respectively, in comparison with FEW or LTD for most α-amylases (compare Fig. 1). Furthermore, the α-amylase from *Micrococcus* sp. has the largest polypeptide chain (1104 amino acid residues) of all α-amylases known to date [11, 12]. Despite its size, this α-amylase, in addition to the supposed $(\beta/\alpha)_8$-barrel domain, apparently does not contain parts recognized elsewhere, e.g. the C-terminal starch-binding domain [26]. *D. thermophilum* produces, in addition to the enzyme investigated in this study (α-amylase B), two other α-amylases designated A and C [27, 28]. The sequence of α-amylase A [27] is exceptional, since the conserved sequence regions [2, 3] are not apparent in this sequence. The position of *B. megaterium* α-amylase in the tree in Fig. 2A can be explained only by the lower degree of sequence similarity throughout the conserved regions (data not shown) and the sequence similarities (Fig. 1).

As for the rooting of the trees presented in Fig. 2, the midpoint method [29] could be used. The root in this method can be placed at the half-way mark along the longest reconstructed lineage between two taxa. Thus, the tree presenting the evolutionary relationships of all α-amylases in this study (Fig. 2A) could be rooted along the branch of *D. thermophilum* α-amylase. Correct rooting in this case should be warranted by the fact that this organism produces one α-amylase (α-amylase A) that is homologous with the *Pyrococcus furiosus* α-amylase [30], which is probably one of the most ancient α-amylases. The rest of the trees in Fig. 2 can be rooted as follows by analogy: (a) fungal α-amylases would be separated from the yeast enzymes in the tree in Fig. 2B; (b) placing the root along the branch of *Vigna mungo* α-amylase in the plant enzyme tree (Fig. 2C), two different taxonomic groups [*Leguminosae* (*V. mungo*) and *Gramineae* (the rest of plant α-amylases)] would result in this case, and (c) the root placed along the branch of *T. curvata* α-amylase in the tree in Fig. 2D would divide the enzymes produced by microorganisms and animals with the exception of *A. haloplanctis* α-amylase. It should be pointed out, however, that the *A. haloplanctis* enzyme is on a branch with length comparable to the branch length of *T. curvata* α-amylase.

One interesting finding can be extracted from the alignment of the amino acid sequences of α-amylases from streptomycetes, *T. curvata*, *A. haloplanctis*, insects and mammals (data not shown), i.e. the alignment of cysteine residues when compared with these residues of pig pancreatic α-amylase, which is known to contain five disulfide bridges [22]. All of these bridges are present probably only in the α-amylases of mammals. The results for all α-amylases of this group are summarized in Table 2. There is no evidence that the disulfide bonds are really formed, but the unambiguous alignment of the cysteine residues of pig pancreatic α-amylase (involved in the five disulfide bridges [22]) with the cysteine residues of the rest of the α-amylases (Fig. 3) support the possibility of the presence of structurally equivalent bonds. Analogical data for fungal and yeast α-amylases are already implemented in the Swiss-Prot protein sequence data bank [11]. As for the plant enzymes, only two invariant cysteine residues (Cys106 and Cys126; barley α-amylase numbering) can be found in their alignment (data not shown). The first cysteine residues could be equivalent to Cys115, the second residue could be equivalent to the Cys141 of pig pancreatic α-amylase, both cysteines of the animal enzyme being involved in two different disulfide bridges [22] (compare Table 2). The determination of possible locations of disulfide bonds in plant α-amylases should be based, however, on the knowledge of locations of these bonds in one of these enzymes since plant α-amylases form, like the enzymes from streptomycetes and animals, their own homologous group

Table 2. Possible disulfide bridges of α-amylases. Full names of the sources of α-amylases are given in Table 1. The locations of disulfide bridges in the sequence of pig pancreatic α-amylase were taken from [22].

α-Amylase	Pig pancreatic α-amylase disulfide bond positions				
	28–86	70–115	141–160	378–384	450–462
	possible positions of disulfide bonds in other α-amylase				
Altha	20–74	–	120–137	328–335	402–416
Strgr	24–78	–	124–140	323–330	–
Strhy	24–78	–	122–138	–	382–391
Strli	24–78	–	124–140	323–330	–
Strth	24–78	–	–	–	–
Strvi	24–78	–	124–140	323–330	–
Thecu	24–83	–	133–149	340–347	411–423
Anoga	28–88	–	142–156	362–368	434–446
Drome	28–84	–	135–149	358–364	430–442
Muspa	28–86	70–115	141–157	375–381	447–459
Mussa	28–86	70–115	141–160	378–384	450–462
Ratpa	28–86	70–115	141–157	375–381	447–459
Humpa	28–86	70–115	141–160	378–384	450–462
Humsa	28–86	70–115	141–160	378–384	450–462

(Fig. 2A and C). Such information will probably be available soon [9].

The last remark concerns the three α-amylases that were excluded from the studied set of α-amylases (Table 1). These α-amylases are the enzymes from *Bacillus circulans*, *Bacillus* sp. B1018 and *Clostridium thermosulfurogenes*, with the accession numbers in Swiss-Prot protein sequence database P08137, P17692 and P26827, respectively [11]. Based on the presence of some sequence features characteristic of a cyclodextrin glycosyltransferase, including the presence of the sequence Phe-Ala-Pro in cyclodextrin glycosyltransferases in the first conserved region [31] and the insertion of the glutamine residue in front of the invariant proline in cyclodextrin glycosyltransferases in the second sequence similarity [32], the possibility of an erroneous identification of these enzymes seems probable. There is no biochemical evidence as yet to support the re-classification of the three α-amylases. When taking into account that the functions of these two enzymes are related (the bond cleavage in cyclodextrin glycosyltransferases is followed by a transglycosylation step [33]), their biochemical re-evaluation is relevant.

The author would like to express his gratitude to Dr B. Svensson (Carlsberg Laboratory, Copenhagen) for her valuable comments on the construction of the evolutionary tree of starch hydrolases and related enzymes, and for kindly providing her manuscript prior to publication. The author also thanks two anonymous referees for their remarks on the manuscript.

REFERENCES

1. MacGregor, E. A. (1988) α-Amylase structure and activity, *J. Protein Chem. 7*, 399–415.
2. Nakajima, R., Imanaka, T. & Aiba, S. (1986) Comparison of amino acid sequences of eleven different α-amylases, *Appl. Microbiol. Biotechnol. 23*, 355–360.
3. 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.
4. Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities, *Biochem. J. 280*, 309–316.
5. Raimbaud, E., Buleon, A., Perez, S. & Henrissat, B. (1989) Hydrophobic cluster analysis of the primary sequences of α-amylases, *Int. J. Biol. Macromol. 11*, 217–225.
6. Qian, M. X., Haser, R. & Payan, F. (1993) Structure and molecular model refinement of pig pancreatic α-amylase at 2.1 Å resolution, *J. Mol. Biol. 231*, 785–799.
7. Brady, R. L., Brzozowski, A. M., Derewenda, Z. S., Dodson, E. J. & Dodson, G. G. (1991) Solution of the structure of *Aspergillus niger* acid α-amylase by combined molecular replacement and multiple isomorphous methods, *Acta Crystallogr. Sect. B 47*, 527–535.
8. Swift, H. J., Brady, L., Derewenda, Z. S., Dodson, E. J., Dodson, G. G., Turkenburg, J. P. & Wilkinson, A. J. (1991) Structure and molecular model refinement of *Aspergillus oryzae* (TAKA) α-amylase. An application of the simulated-annealing method, *Acta Crystallogr. Sect. B 47*, 535–544.
9. Kadziola, A., Abe, J., Svensson, B. & Haser, R. (1992) Crystal structure and function of barley malt α-amylase, Abstract from the Workshop on Cereal Polysaccharides, Le Croisic, France.
10. Janeček, Š. & Baláž, Š. (1993) Evolution of parallel β/α-barrel enzyme family lightened by structural data on starch-processing enzymes, *J. Protein Chem. 12*, 509–514.
11. SwissProt protein sequence database, release 27, October 1993.
12. GenBank DNA sequence database, release 80, October 1993.
13. Yuuki, T., Nomura, T., Tezuka, H., Tsuboi, A., Yamagata, H., Tsukagoshi, N. & Udaka, S. (1985) Complete nucleotide sequence of a gene coding for heat- and pH-stable α-amylase of *Bacillus licheniformis*. Comparison of the amino acid sequences of three bacterial liquefying α-amylases deduced from the DNA sequences, *J. Biochem. (Tokyo) 98*, 1147–1156.
14. Itoh, T., Yamashita, I. & Fukui, S. (1987) Nucleotide sequence of the α-amylase gene (*ALP1*) in the yeast *Saccharomycopsis fibuligera*, *FEBS Lett. 219*, 339–342.
15. Hu, N.-T., Hung, M.-N., Huang, A.-M., Tsai, H.-F., Yang, B.-Y., Chow, T.-Y. & Tseng, Y.-H. (1992) Molecular cloning, characterization and nucleotide sequence of the gene for secreted α-amylase from *Xanthomonas campestris* pv. *campestris*, *J. Gen. Microbiol. 138*, 1647–1655.
16. Long, C. M., Virolle, M.-J., Chang, S.-Y., Chang, S. & Bibb, M. J. (1987) α-Amylase gene of *Streptomyces limosus*. Nucleotide sequence, expression motifs, and amino acid sequence homology to mammalian and invertebrate α-amylases, *J. Bacteriol. 169*, 5745–5754.

17. Petříček, M., Tichý, P. & Kuncová, M. (1992) Characterization of the α-amylase-encoding gene from *Thermomonospora curvata*, *Gene (Amst.) 112*, 77–83.
18. Huang, N., Stebbins, G. L. & Rodriquez, R. L. (1992) Classification and evolution of α-amylase genes in plants, *Proc. Natl Acad. Sci. USA 89*, 7526–7530.
19. Rumbak, E., Rawlings, D. E., Lindsey, G. G. & Woods, D. R. (1991) Cloning, nucleotide sequence, and enzymatic characterization of an α-amylase from the ruminal bacterium *Butyrivibrio fibrisolvens* H17c, *J. Bacteriol. 173*, 4203–4211.
20. Saitou, N. & Nei, M. (1987) The neighbour-joining method: A new method for reconstructing phylogenetic trees, *Mol. Biol. Evol. 4*, 406–425.
21. Higgins, D. G., Bleasby, A. J. & Fuchs, R. (1992) CLUSTAL V. Improved software for multiple sequence alignment, *Comput. Appl. Biosci. 8*, 189–191.
22. Pasero, L., Mazzei-Pierron, Y., Abadie, B., Chicheportiche, Y. & Marchis-Mouren, G. (1986) Complete amino acid sequence and location of the five disulfide bridges in porcine pancreatic α-amylase, *Biochim. Biophys. Acta 869*, 147–157.
23. Vihinen, M. & Mäntsälä, P. (1989) Microbial amylolytic enzymes, *Crit. Rev. Biochem. Mol. Biol. 24*, 329–418.
24. Jespersen, H. M., MacGregor, E. A., Henrissat, B., Sierks, M. R. & 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.
25. Raha, M. A., Kawagishi, I., Mueller, V., Kihara, M. & Macnab, R. M. (1992) *Escherichia coli* produces a cytoplasmic α-amylase, *Amy A*, *J. Bacteriol. 174*, 6644–6652.
26. Svensson, B., Jespersen, H., Sierks, M. & MacGregor, E. A. (1989) Sequence homology between putative raw-starch binding domains from different starch-degrading enzymes. *Biochem. J. 264*, 309–311.
27. Fukusumi, S., Kamizono, A., Horinouchi, S. & Beppu, T. (1988) Cloning and nucleotide sequence of a heat-stable amylase gene from an anaerobic thermophile, *Dictyoglomus thermophilum*, *Eur. J. Biochem. 174*, 15–21.
28. Horinouchi, S., Fukusumi, S., Ohshima, T. & 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.
29. Stewart, C.-B. (1993) The powers and pitfalls of parsimony, *Nature 361*, 603–607.
30. Laderman, K. A., Asada, K., Uemori, T., Mukai, H., Taguchi, Y., Kato, I. & Anfinsen, C. B. (1993) α-Amylase from the hyperthermophilic archaebacterium *Pyrococcus furiosus*. Cloning and sequencing of the gene and expression in *Escherichia coli*, *J. Biol. Chem. 268*, 24402–24407.
31. MacGregor, E. A. & Svensson, B. (1989) A super-secondary structure predicted to be common to several α-1,4-D-glucan-cleaving enzymes, *Biochem. J. 259*, 145–152.
32. Janeček, Š. & Tóth, D. (1994) Amino acid sequence alignment of microbial, plant and animal α-amylases: Some evolutionary implications, *Biologia (Bratislava) 49*, in the press.
33. Svensson, B. & Søgaard, M. (1993) Mutational analysis of glycosylase function, *J. Biotechnol. 29*, 1–37.