FEBS Letters 377 (1995) 6–8 FEBS 16380

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

Štefan Janeček*

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

Received 2 October 1995

Abstract **A short conserved sequence equivalent to the fifth conserved sequence region of α-amylases (173_LPDLD, *Aspergillus oryzae* α-amylase) comprising the calcium-ligand aspartate, Asp-175, was identified in the amino acid sequences of several members of the family of $(\alpha/\beta)_8$-barrel glycosyl hydrolases. Despite the fact that the aspartate is not invariantly conserved, the stretch can be easily recognised in all sequences to be positioned 26–28 amino acid residues in front of the well-known catalytic aspartate (Asp-206, *A. oryzae* α-amylase) located in the β4-strand of the barrel. The identification of this region revealed remarkable similarities between some α-amylases (those from *Bacillus megaterium*, *Bacillus subtilis* and *Dictyoglomus thermophilum*) on the one hand and several different enzyme specificities (such as oligo-1,6-glucosidase, amylomaltase and neopullulanase, respectively) on the other hand. The most interesting example was offered by *B. subtilis* α-amylase and potato amylomaltase with the regions LYDWN and LYDWK, respectively. These observations support the idea that all members of the family of glycosyl hydrolases adopting the structure of the α-amylase-type $(\alpha/\beta)_8$-barrel are mutually closely related and the strict evolutionary borders separating the individual enzyme specificities can be hardly defined.**

Key words: α-Amylase; Glycosyl hydrolase; Conserved sequence region; Evolutionary relatedness

1. Introduction

At present, there is an enormous increase of determined sequences of various proteins the structures of which are to be solved. This terrible fact is fortunately weakened by the existence of the limited set of folding patterns adopting by most of proteins [1–3]. One of the patterns, the parallel eight-folded $(\alpha/\beta)_8$-barrel, is formed by the inner β-barrel sheet consisting of 8 parallel β-strands surrounded by 8 α-helices [4]. This motif is adopted also by the family of glycosyl hydrolases and related enzymes grouped around the α-amylase as suggested by the sequence-oriented and prediction studies [5,6] in combination with determined three-dimensional structures of several α-amylases, cyclodextrin glycosyltransferases and one oligo-1,6-glucosidase [7–9].

The glycosyl hydrolases of the α-amylase family exhibit several sequence similarities [9,10]. Four of them, those at or around the strands β3, β4, β5 and β7 of the $(\alpha/\beta)_8$-barrel, are well-known as conserved regions important from both functional and evolutionary points of view [6]. Recently, the fifth conserved sequence region has been pointed out in the sequences of α-amylases [11] to be localised outside the catalytic $(\alpha/\beta)_8$-barrel in domain B comprising the very long third loop of the barrel. The region contains the Asp-175 (*Aspergillus oryzae* α-amylase (TAA) numbering) mostly involved in the binding of Ca^{2+} (e.g. [12]). It is worth mentioning that not only the sequence of this region is conserved. In α-amylases, this stretch is sequentially positioned predominantly 26–28 amino acid residues in front of the well-recognised catalytic aspartate (Asp-206 in TAA) located in the β4-strand of the $(\alpha/\beta)_8$-barrel domain.

The main goal of the present work was to trace (if possible) the fifth conserved sequence region in the sequences of the other members of continuously expanding family of α-amylase-type $(\alpha/\beta)_8$-barrel glycosyl hydrolases and related enzymes and to evaluate its evolutionary importance in relation to the enzyme specificities brought about the respective amino acid sequences.

2. Materials and methods

Amino acid sequences of the $(\alpha/\beta)_8$-barrel glycosyl hydrolases from the α-amylase family (Table 1) were extracted from the SwissProt Protein or GenBank DNA Data Banks.

The sequences of all these enzymes were searched for the fifth conserved sequence region pointed out firstly in the sequences of α-amylases [11]. The region was expected to be positioned about 26–28 amino acid residues in front of the well-known catalytic aspartate (Asp-206 in TAA) located in the β4-strand of the $(\alpha/\beta)_8$-barrel domain.

To demonstrate the evolutionary importance of the fifth conserved sequence region, the parts of the amino acid sequences comprising their eventual domains B (from the third β-strand to the fourth β-strand) of the α-amylases from *Bacillus subtilis* and *Butyrivibrio fibrisolvens* (40% identity [10]) and of the functionally distantly related potato amylomaltase (its sequence exhibits substantial variability also in the four well-known conserved regions [8]), were aligned using the program CLUSTAL V [13].

3. Results and discussion

The fifth conserved sequence region of the enzymes studied here is shown in Table 1. It is remarkable that not only the sequence but also the position of this stretch in the whole amino acid sequence of a glycosyl hydrolase from the studied set is perfectly conserved (26–28 amino acid residues in front of the catalytic aspartate (Asp-206 in TAA) from the fourth β-strand). This fact warrants the possibility to identify the regions cor-

*Corresponding author. Fax: (42) (7) 493824.
E-mail: sjanecek@ue.savba.sk

Abbreviation: TAA, TAKA amylase A (α-amylase from *Aspergillus oryzae*).

SSDI 0014-5793(95)01309-1

```
            strand β3                                                                   strand β4
         *  ****   ****     *               *    *     **   *  *  ** *  * **  * *   **      ****** * *
Bfi  200 GVAVIVDILPNHTTPSTGSIAKALMEAAGGSDALYHATGKIGGGYTDRLELTYYSMGGLPDVDTENTGFQQYFYEFLKDCVYLGADGFRIDTA 292
Bsu   91 GIKVIVDAVINHTTSDYAAISNEVKSIPN----WTHGNTQIKN-WSDRWDVTQNSLLGLYDWNTQNTQVQSYLKRFLDRALNDGADGFRFDAA 178
Pot  298 VGYHSADVWANKKQFLLNRKGFPLIVSGVPPDA-FSETGQL---WGSP----------LYDWKAMEKDGFSWWVRRIQRATDL-FDEFRIDHF 375
               *   *                            *    *             ****        *   *   **     * ** *
```

Fig. 1. Sequence comparison of closely and distantly related homologous $(\alpha/\beta)_8$-barrel glycosyl hydrolases. The closely related enzymes: Bfi, *B. fibrisolvens* α-amylase; Bsu, *B. subtilis* α-amylase. Distantly related enzyme to both of them: Pot, amylomaltase from potato. The alignment comprises approximately the parts of their amino acid sequences from the third β-strand to the fourth β-strand of the catalytic $(\alpha/\beta)_8$-barrels (domains B). The asterisks signify the identical amino acid residues for (i) the two α-amylases over the alignment, and (ii) the *B. subtilis* α-amylase and potato amylomaltase under the alignment. Gaps are indicated by dashes. The fifth conserved sequence region in all the three enzymes is in bold print. For illustration, two of the four well-accepted conserved sequence regions (strands β3 and β4) are italicised.

rectly in spite of the observed Asp→Lys substitution in a few cases. The fifth conserved sequence region can be unambiguously traced also in the sequences of pullulanase from *Klebsiella pneumoniae* (632_CSDSA) [14] or amylopullulanase from *Clostridium thermohydro-sulfuricum* (566_WADFI) [15]. On the other hand, this region can be hardly traced in the sequences of, e.g. glycogen branching and debranching enzymes [6].

The most important point of the present study is to consider the sequence of the fifth conserved sequence region of the enzymes listed in Table 1 in relation to the enzyme specificity brought about the relevant amino acid sequence. Thus, for instance, this stretch of α-amylases from *Bacillus megaterium* (174_MPDLN) and *Dictyoglomus thermophilum* amyC (181_MPDLN) is similar to the regions of oligo-1,6-glucosidase, α-glucosidase, dextran glucosidase and trehalose-6-phosphate hydrolase (mostly QPDLN). The regions of *D. thermophilum* amyB α-amylase (276_MPKIN) on the one hand and the regions of *Bacillus sphaericus* cyclomaltodextrinase and *Bacillus* sp. neopullulanase (both enzymes 294_MPKLN) on the other hand, can manifest the other clear example of the similarity of these regions independently of the enzyme specificities. Perhaps, the most interesting case can be demonstrated by the stretches of *B. subtilis* α-amylase (144_LYDWN) and the amylomaltase from potato (342_LYDWK). It is worth mentioning that this partial similarity was not reflected in the evolutionary tree of these enzymes (Š. Janeček, unpubl. results). On the other hand, this fact unambiguously indicates that all these $(\alpha/\beta)_8$-barrel glycosyl hydrolases from the α-amylase family are mutually closely related since the individual enzymes harbouring different enzyme specificities that are also in the frame of one homologous family distantly related (i.e. *B. subtilis* α-amylase and potato amylomaltase) contain common relic sequence features. To make this proposal more convincing, the sequence alignment was made (Fig. 1) that comprises the part of the amino acid sequences from the β3-strand to the β4-strand of the catalytic $(\alpha/\beta)_8$-barrel domain containing the fifth conserved sequence region of the above-mentioned *B. subtilis* α-amylase and potato amylomaltase along with the *B. fibrisolvens* α-amylase. The enzyme from the ruminal bacterium has about 40% identity with the saccharifying α-amylase from *B. subtilis*, both enzymes being also clustered together in the evolutionary tree of α-amylases [10]. In spite of this fact, the fifth conserved sequence region of microbial *B. subtilis* α-amylase is more similar to the fifth conserved sequence region of potato amylomaltase, i.e. of the enzyme that is less similar in the rest part of the amino acid sequence (cf. Fig. 1), moreover is not an α-amylase, moreover is produced by a plant.

It should be pointed out, however, the presented similarities are not in contradiction with the evolutionary trees published previously [6,9] that have divided the members of this enzyme family according to their substrate specificities. These observations only demonstrate that all these enzymes very probably evolved from a common ancestor. Paradoxically, its relic sequence features can be found more convincingly in the enzymes that are more distantly related (*B. subtilis* α-amylase and amylomaltase from potato) than in the closely related enzymes (*B. subtilis* and *B. fibrisolvens* α-amylases; cf. Fig. 1).

Table 1
The fifth conserved sequence region of glycosyl hydrolases from the α-amylase family included in the present study

Enzyme	EC[a]	Source	Conserved region[b]		Ac. No.[c]
α-Amylase	3.2.1.1	*Aspergillus oryzae*	173_LPDLD	(28)	P10529
		Bacillus subtilis	144_LYDWN	(27)	P00691
		Bacillus megaterium	174_MPDLN	(27)	P20845
		Butyrivibrio fibrisolvens	258_LPDVD	(27)	P30269
		Dictyoglomus thermophilum amyB	276_MPKIN	(28)	P14898
		Dictyoglomus thermophilum amyC	181_MPDLN	(27)	P14899
Oligo-1,6-glucosidase	3.2.1.10	*Bacillus cereus*	167_QPDLN	(27)	P21332
α-Glucosidase	3.2.1.20	*Aedes aegypti* (yellowfever mosquito)	169_QPDLN	(27)	P13080
Cyclomaltodextrinase	3.2.1.54	*Bacillus sphaericus*	294_MPKLN	(28)	X62576
Dextran glucosidase	3.2.1.70	*Streptococcus mutans*	162_QPDLN	(27)	M30944
Trehalose-6-phosphate hydrolase	3.2.1.93	*Bacillus subtilis*	170_QADLN	(27)	P39795
Neopullulanase	3.2.1.0	*Bacillus* sp. strain KSM-1876	294_MPKLN	(28)	M74130
Amylomaltase	2.4.1.25	*Solanum tuberosum* (potato)	342_LYDWK	(26)	X68664

[a] The enzymes are ordered according to their EC numbers.
[b] The numbers in parenthesis specify the length of the polypeptide-chain segment (given by the number of amino acid residues) that separates the conserved sequence region from the catalytic aspartate (Asp-206 in TAA) located in the β4-strand of $(\alpha/\beta)_8$-barrel domain.
[c] The accession numbers from SwissProt Protein (starting with P) or GenBank DNA Sequence Databases.

In conclusion, it is suggested that the fifth conserved sequence region (173_LPDLD in TAA) recognised firstly in the sequences of α-amylases [11] and now demonstrated in the sequences of the other known enzymes from this family (Table 1) should be taken into account in studying both structure-function (Asp-175 in TAA is the calcium ion ligand [12]) and evolutionary relationships (this work) of these enzymes. Although the characteristic sequence differences that allow to discriminate functionally and structurally closely related enzymes, such as α-amylases and cyclodextrin glycosyltransferases, can be revealed (e.g. [16]), the strict evolutionary borders that would separate the individual substrate specificities of this family can be hardly defined.

Acknowledgements: The advance of this work was promoted in part by the financial support from the Slovak Literary Fund that is gratefully acknowledged.

References

[1] Finkelstein, A.V. and Ptitsyn, O.B. (1987) Prog. Biophys. Mol. Biol. 50, 171–190.
[2] Chothia, C. (1992) Nature 357, 543–544.
[3] Orengo, C.A., Jones, D.T. and Thornton, J.M. (1994) Nature 372, 692–694.
[4] Richardson, J.C. (1985) Methods Enzymol. 115, 341–358.
[5] Henrissat, B. (1991) Biochem. J. 280, 309–316.
[6] Jespersen, H.M., MacGregor, E.A., Henrissat, B., Sierks, M.R. and Svensson, B. (1993) J. Protein Chem. 12, 791–805.
[7] MacGregor, E.A. (1993) Starch 7, 232–237.
[8] Svensson, B. (1994) Plant Mol. Biol. 25, 141–157.
[9] Janeček, Š. (1994) FEBS Lett. 353, 119–123.
[10] Janeček, Š. (1994) Eur. J. Biochem. 224, 519–524.
[11] Janeček, Š. (1992) Biochem. J. 288, 1069–1070.
[12] Boel, E., Brady, L., Brzozowski, A.M., Derewenda, Z., Dodson, G.G., Jensen, V.J., Petersen, S.B., Swift, H., Thim. L. and Woldike, H.F. (1990) Biochemistry 29, 6244–6249.
[13] Higgins, D.G., Bleasby, A.J. and Fuchs, R. (1992) Comput. Applic. Biosci. 8, 189–191.
[14] Mathupala, S.P., Lowe, S.E., Podkovyrov, S.M. and Zeikus, J.G. (1993) J. Biol. Chem. 268, 16332–16344.
[15] Kornacker, M.G. and Pugsley, A.P. (1990) Mol. Microbiol. 4, 73–85.
[16] Janeček, Š., MacGregor, E.A. and Svensson, B. (1995) Biochem. J. 305, 685–686.