Volume 316, number 1, 23–26 FEBS 12001 January 1993

Sequence similarities in $(\alpha/\beta)_8$-barrel enzymes revealed by conserved regions of α-amylase

Štefan Janeček

Department of Biochemical Technology, Faculty of Chemical Technology, Slovak Technical University, Radlinského 9, CS-81237 Bratislava, Czech and Slovak Federal Republic

Received 22 October 1992

The parallel $(\alpha/\beta)_8$-barrel is a frequently occurring protein-folding motif. Although the arrangement of secondary structural elements along the barrel is very similar in different $(\alpha/\beta)_8$-barrel enzymes, there is a very low mutual amino acid sequence homology among the enzymes, contributing in part to the hazy view of their evolution. Here an approach to identifying at least the rough of evolutionarily conserved $(\alpha/\beta)_8$-barrel sequence is presented. Based on the idea that highly conserved sequence regions of a particular enzyme should be more or less conserved in the sequences of the other evolutionary related enzymes, five sequence similarities of ten different $(\alpha/\beta)_8$-barrel enzymes were revealed, using the five conserved regions of the amino acid sequence of the α-amylase $(\alpha/\beta)_8$-barrel as the templates.

α-Amylase; $(\alpha/\beta)_8$-Barrel enzyme; Conserved region; Sequence similarity; Evolutionary relationship

1. INTRODUCTION

The daunting number of known protein amino acid sequences (over 17,000) in comparison with about a 40-times lower number of known protein tertiary structures [1], is fortunately weakened by the existence of families of proteins that have similar folding patterns [2]. One of them, the parallel 8-folded $(\alpha/\beta)_8$-barrel, first recognized in chicken muscle triosephosphate isomerase [3], has been observed in the structure of roughly every tenth enzyme solved to date [4]. Almost all $(\alpha/\beta)_8$-barrel structures have 8 consecutive parallel β-strands surrounded by 8 α-helices [5]. The motif enables the enzyme to catalyze quite different biochemical reactions [6] and the $(\alpha/\beta)_8$-barrel enzyme family covers all types of enzyme nomenclature except for ligases.

Two evolutionary histories have been explored for $(\alpha/\beta)_8$-barrel enzymes [5]. The first one is based on a proposal by Gilbert [7] that proteins evolved by different combinations of exons coding for small functional or structural units, and has been documented by a comparison of the intron–exon arrangement in triosephosphate isomerase from different species [8]. In the second case Farber and Petsko [4] have elucidated the reasons favoring the more probable divergent evolution of $(\alpha/\beta)_8$-barrel enzymes from a common ancestor in contrast with convergent evolution. Despite these efforts the challenging question of how the $(\alpha/\beta)_8$-barrel fold has arisen remains unanswered.

Due to very low mutual sequence homology [4,5], a comparison of all the $(\alpha/\beta)_8$-barrel sequences is not yet available. If there is any evolutionary relationship among them, however, at least the stretches of the $(\alpha/\beta)_8$-barrel's ancestors or descendant amino acid sequences should be identified by using the conserved regions of a particular $(\alpha/\beta)_8$-barrel enzyme as structural templates where some similarities have been protected or selected throughout the enzyme family. This report reports 5 such sequence similarities of 10 different $(\alpha/\beta)_8$-barrel enzymes. They were revealed in the parts of their $(\alpha/\beta)_8$-barrels that correspond with five conserved regions of the $(\alpha/\beta)_8$-barrel of α-amylase.

2. MATERIALS AND METHODS

α-Amylase (EC 3.2.1.1) from porcine pancreas (AMYL) was used as the template $(\alpha/\beta)_8$-barrel enzyme in this study. Its amino acid sequence, as well as the arrangement of α-helices and β-sheets along the $(\alpha/\beta)_8$-barrel superstructure, were taken from the literature [9,10].

Five conserved amino acid regions ([1], Š. Janeček, submitted for publication) were found to form approximately the same parts of the $(\alpha/\beta)_8$-barrels of different α-amylases. Their locations in the AMYL $(\alpha/\beta)_8$-barrel were then used in the search for sequence similarities among the amino acid sequences of the other 10 $(\alpha/\beta)_8$-barrel enzymes.

The enzymes were chosen based on the proposed organization of the whole $(\alpha/\beta)_8$-barrel enzyme family into 4 different structural subfamilies [4]; one representative from each family was considered here. Needed structural details, such as amino acid sequences and composition of secondary structure in individual $(\alpha/\beta)_8$-barrels, were taken from the literature, and the enzymes were abbreviated as follows: GO, spinach glycolate oxidase [12]; RBCO, ribulose-1,5-bisphosphate carboxylase/oxygenase from *Rhodospirillum rubrum* [13,14]; MR, mande-

Correspondence address: Š. Janeček, Department of Biochemical Technology, Faculty of Chemical Technology, Slovak Technical University, Radlinského 9, CS-81237 Bratislava, Czech and Slovak Federal Republic.

Table I

Five conserved amino acid sequence regions of $(\alpha/\beta)_8$-barrel enzymes

Region	Enzyme[a]	Numbering	Sequence[b]	Secondary structure[c]
I	AMYL	96–101	D A V I N H	β3, loop 3
	GO	124–129	R F F Q L Y	β3, loop 3
	RBCO	227–232	L F S A N I	β3, loop 3
	MR	193–198	M V D Y N Q	β3, loop 3
	MLE	183–188	H V V A I K	β3
	CGT	135–140	D F A P N H	β3, loop 3
	XYI	87–92	M A T T N L	β3, loop 3
	PRAI	309–314	G V F R N H	β3, loop 3
	IGPS	111–116	I L C K D F	β3, loop 3
	TIM	60–65	T V G A Q N	β3, loop 3
	FALD	102–107	I I L G I K	turn[d], β3
II	AMYL	165–169	L L D L A	loop 3
	GO	131–135	Y K D R N	loop 3, α3
	RBCO	233–237	T A D D P	loop 3, α3
	MR	199–203	S L D V P	loop 3, α3
	MLE	191–195	L G D S A	loop 3, α3
	CGT	197–201	L A D F N	loop 3
	XYI	98–102	F K D G G	loop 3, α3
	PRAI	316–320	I A D V V	α3
	IGPS	117–121	I I D P Y	loop 3, α3
	TIM	83–87	I K D V G	α3
	FALD	127–131	L D D L A	loop 3, α3
III	AMYL	193–201	G F R L D A S K H	β4, loop 4
	GO	153–161	A L T V D T P R L	β4, loop 4
	RBCO	259–267	A L L V D G Y V A	β4, loop 4
	MR	219–227	W I E E P T L Q H	turn, β4, loop 4
	MLE	210–218	A I R A C Q V L G	β4, loop 4, α4
	CGT	225–233	G I R V D A V K H	β4, loop 4
	XYI	123–131	D L A V E L A R K	β4, loop 4
	PRAI	330–338	A V Q L H G N E E	β4, loop 4
	IGPS	135–143	L L M L S V L D D	β4, loop 4, α4
	TIM	90–98	W V I L G H S E R	β4, loop 4
	FALD	148–156	R C V L K I G K N	β4, loop 4
IV	AMYL	233–236	E V I D	loop 5
	GO	230–233	K G V I	β5, loop 5
	RBCO	287–290	H R A G	β5, loop 5
	MR	247–250	E N W L	loop 5
	MLE	235–238	S G Q V	β5, loop 5
	CGT	257–260	E W F L	β5, loop 5
	XYI	180–183	E P K P	β5
	PRAI	363–366	E T L P	loop 5
	IGPS	163–166	E V S N	β5, loop 5
	TIM	129–132	E T L E	β5, loop 5
	FALD	189–192	E V L P	β5, loop 5
V	AMYL	295–300	F V D N H D	β7, loop 7
	GO	283–288	F L D G G V	β7, loop 7
	RBCO	366–371	I I S G G M	β7, loop 7
	MR	293–298	P M S S H L	β7, loop 7
	MLE	285–290	V L R T A Q	β7, loop 7
	CGT	323–328	F I D N H D	loop 7
	XYI	249–254	S G I K Y D	β7, loop 7
	PRAI	403–408	L L A G G L	β7, loop 7
	IGPS	212–217	I S E S G I	β7, loop 7
	TIM	206–211	I I Y G G S	β7, loop 7
	FALD	269–274	F L S G G Q	β7, loop 7

[a] Full names of the enzymes and their sources are given in section 2.
[b] Amino acids are shown in the one-letter code.
[c] Elements of secondary structure in loops are not specified.
[d] Turn means a stretch preceding a β-strand.

late racemase from *Pseudomonas putida* [15,16]; MLE, muconate lactonizing enzyme from *Pseudomonas putida* [17,18]; CGT, cyclodextrin glycosyltransferase from *Bacillus circulans* [19]; XYI, xylose isomerase from *Streptomyces olivochromogenes* [20]; PRAI, phosphoribosylanthranilate isomerase from *Escherichia coli* [21]; IGPS, indoleglycerolphosphate synthase from *Escherichia coli* [21]; TIM, yeast triosephosphate isomerase [22,23]; FALD, fructose-1,6-bisphosphate aldolase from *Drosophila melanogaster* ([24,25], K. Piontek, personal communication).

3. RESULTS AND DISCUSSION

This is the first report giving several sequence stretches of different $(\alpha/\beta)_8$-barrel enzymes that might be conserved during evolution. The stretches were deduced from the assumption that the highly conserved regions in the amino acid sequence of one $(\alpha/\beta)_8$-barrel enzyme should be more or less conserved in the other enzymes with this folding motif because they are evolutionary related.

The results from the search for sequence similarities are shown in Table I. Although the similarities do not exhibit a strict homology, there are mostly semiconservative substitutions of amino acid residues among the compared enzymes. The fact that these regions with a high degree of similarity are localized in nearly the same locations of the individual $(\alpha/\beta)_8$-barrels (last column in Table I) is of special importance. Moreover, the regions containing only 30 amino acid residues, i.e. less than 10% of the average length of an $(\alpha/\beta)_8$-barrel enzyme, include about 50% of all the residues involved in the catalytic functions of the enzymes studied here (Table II).

Recently, based on the common property of several $(\alpha/\beta)_8$-barrel enzymes (binding of phosphate), Wilmanns et al. [26] found some local sequence similarity for GO, RBCO, PRAI, IGPS, TIM, α-subunit of tryptophan synthase (TS_a) and flavocytochrome b_2 (FCB2). It should be pointed out that a part of that similarity [26] fell down in the fifth conserved region proposed here, where the alignments of amino acid sequences from both studies were almost identical.

Table II

Amino acid residues involved in catalytic function in $(\alpha/\beta)_8$-barrel enzymes found in five sequence similar regions

Enzyme[a]	Catalytic residues[b]	In five conserved regions
AMYL	D197[c], D300	D197 (III)[d], D300 (V)
GO	Y24, Y129, H254, R257	Y129 (I)
RBCO	K166, K158, K329, S368	S368 (V)
MR	K166, H297	H297 (V)
MLE[e]	–	–
CGT	D229, E257, D328	D229 (III), E257 (IV), D328 (V)
XYI	E180, H219, D286	E180 (V)
PRAI	C260, Q332, D379	Q332 (III)
IGPS	S60, K114, E163, N184, R186	K114 (I), E163 (IV)
TIM	H95, E165	H95 (III)
FALD[f]	K107, K146, R148, K229	K107 (I), R148 (III)

[a] Full names of the enzymes and their sources are given in section 2.
[b] Residues involved in catalysis were extracted from literature given also in section 2.
[c] Amino acid type (using the one-letter code) and the position in the sequence of the particular enzyme are shown.
[d] The number of conserved regions (in parentheses).
[e] Residues have not been determined.
[f] Residues have been proposed only as the active site residues.

Although the comparison of all the $(\alpha/\beta)_8$-barrel enzymes sequences is not yet available [4], several groups, such as AMYL and CGT [27], MR and MLE [28], GO and FCB2 [29], and PRAI, IGPS and TS_a [26], have been shown to have sufficient sequence and, eventually, structural identity. Therefore, these groups of $(\alpha/\beta)_8$-barrel enzymes are considered to have evolved by divergent evolution, either from the same ancestor or, in part, from each other. Despite the lack in overall homology in the $(\alpha/\beta)_8$-barrel enzyme family [4,5], the existing structural similarities almost certainly arise from the stringent requirements of the $(\alpha/\beta)_8$-barrel fold [30].

From the results of this study, however, it might be concluded that, independent of the type of evolution (divergent or convergent), the rough amino acid sequence of the ancestor or the descendant $(\alpha/\beta)_8$-barrel or, at least, a substantial part of it, can be designed simply by putting together the highly conserved regions of the particular $(\alpha/\beta)_8$-barrel enzymes. The possibility of constructing an evolutionarily conserved amino acid sequence which will adopt the parallel 8-folded $(\alpha/\beta)_8$-barrel might be useful not only in the elucidation of evolutionary relationships in the $(\alpha/\beta)_8$-barrel enzyme family, but also in the prediction of further enzymes with this folding motif prior to their X-ray crystallographic studies.

Acknowledgements: The author is grateful to Dr. Š. Baláž for general support, and Dr. K. Piontek for providing the secondary structural elements of FALD.

REFERENCES

[1] Baron, M., Norman, D.G. and Campbell, I.D. (1991) Trends Biochem. Sci. 16, 13–17.
[2] Finkelstein, A.V. and Ptitsyn, O.B. (1987) Prog. Biophys. Mol. Biol. 50, 171–190.
[3] Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C., Pogson, C.I., Wilson, I.A., Corran, P.H., Furth, A.J., Milman, J.D., Offord, R.E., Priddle, J.D. and Waley, S.G. (1975) Nature 255, 609–614.
[4] Farber, G.K. and Petsko, G.A. (1990) Trends Biochem. Sci. 15, 228–234.
[5] Brändén, C.-I. (1991) Curr. Opin. Struct. Biol. 1, 978–983.
[6] Scheerlinck, J.-P., Lasters, I., Claessens, M., De Maeyer, M., Pio, F., Delhaise, P. and Wodak, S.J. (1992) Proteins 12, 299–313.
[7] Gilbert, W. (1978) Nature 271, 501.
[8] Straus, D. and Gilbert, W. (1985) Mol. Cell. Biol. 5, 3497–3506.
[9] Pasero, L., Mazzéi-Pierron, Y., Abadie, B., Chicheportiche, Y. and Marchis-Mouren, G. (1986) Biochim. Biophys. Acta 869, 147–157.
[10] Buisson, G., Duée, E., Haser, R. and Payan, F. (1987) EMBO J. 6, 3909–3916.

[11] Nakajima, R., Imanaka, T. and Aiba, S. (1986) Appl. Microbiol. Biotechnol. 23, 355–360.
[12] Lindqvist, Y. (1989) J. Mol. Biol. 209, 151–166.
[13] Schneider, G., Lindqvist, Y. and Lundqvist, T. (1990) J. Mol. Biol. 211, 989–1008.
[14] Brändén, C.-I., Lindqvist, Y. and Schneider, G. (1991) Acta Crystalogr. B47, 824–835.
[15] Ransom, S.G., Gerlt, J.A., Powers, V.M. and Kenyon, G.L. (1988) Biochemistry 27, 540–545.
[16] Neidhart, D.J., Howell, P.L., Petsko, G.A., Powers, V.M., Li, R., Kenyon, G.L. and Gerlt, J.A. (1991) Biochemistry 30, 9264–9273.
[17] Aldrich, T.L., Frantz, B., Gill, J.F., Kilbane, J.J. and Chakrabarty, A.M. (1987) Gene 52, 185–195.
[18] Goldman, A., Ollis, D.L. and Steitz, T.A. (1987) J. Mol. Biol. 194, 143–153.
[19] Klein, C. and Schulz, G.E. (1991) J. Mol. Biol. 217, 737–750.
[20] Farber, G.K., Glasfeld, A., Tiraby, G., Ringe, D. and Petsko, G.A. (1989) Biochemistry 28, 7289–7297.
[21] Wilmanns, M., Priestle, J.P., Niermann, T. and Jansonius, J.N. (1992) J. Mol. Biol. 223, 477–507.
[22] Lolis, E., Alber, T., Davenport, R.C., Rose, D., Hartman, F.C. and Petsko, G.A. (1990) Biochemistry 29, 6609–6618.
[23] Lolis, E. and Petsko, G.A. (1990) Biochemistry 29, 6619–6625.
[24] Malek, A.A., Hy, M., Honegger, A., Rose, K. and Brenner-Holzach, O. (1988) Arch. Biochem. Biophys. 266, 10–31.
[25] Hester, G., Brenner-Holzach, O., Rossi, F.A., Struck-Donatz, M., Winterhalter, K.H., Smit, J.D.G. and Piontek, K. (1991) FEBS Lett. 292, 237–242.
26] Wilmanns, M., Hyde, C.C., Davies, D.R., Kirschner, K. and Jansonius, J.N. (1991) Biochemistry 30, 9161–9169.
[27] MacGregor, E.A. and Svensson, B. (1989) Biochem. J. 259, 145–152.
[28] Neidhart, D.J., Kenyon, G.L., Gerlt, J.A. and Petsko, G.A. (1990) Nature 347, 692–694.
[29] Lindqvist, Y., Brändén, C.-I., Mathews, F.S. and Lederer, F. (1991) J. Biol. Chem. 266, 3198–3207.
[30] Chothia, C. (1988) Nature 333, 598–599.