ELSEVIER

International Journal of Biological Macromolecules
21 (1997) 277–280

INTERNATIONAL JOURNAL OF
Biological
Macromolecules
STRUCTURE, FUNCTION AND INTERACTIONS

Short communication

Sequence similarity between xylose isomerase and replicase: Another TIM-barrel in the replicase structure?

Štefan Janeček *

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

Received 16 April 1997; accepted 22 April 1997

Abstract

The BLAST search using the strand $\beta 2$ (46_GAHGVTFHDDDLIP) of the $(\alpha/\beta)_8$-barrel of xylose isomerase from *Streptomyces olivochromogenes* resulted in retrieving the sequentially similar segment of replicase from garlic latent virus (692_GGHGIGFHRDD). The detailed analysis of the entire amino acid sequences of both xylose isomerase and replicase suggested that the polypeptide segment 644–1046 of replicase (the entire length of this enzyme is 1924 residues) could share the structure of xylose isomerase (20.7% identity using the entire sequence of xylose isomerase). The relatedness of replicase and xylose isomerase is supported by the fact that the sequence similarity can be observed along the whole sequence of xylose isomerase (386 amino acid residues). The sequence of replicase exhibits moreover the similarity with that of lycopene cyclase, an enzyme implicated in the β-carotene biosynthesis, that was previously found to share similarity with xylose isomerase. Thus the relevant segment of replicase is predicted to adopt an $(\alpha/\beta)_8$-barrel topology similar to that of xylose isomerase. © 1997 Elsevier Science B.V.

Keywords: $(\alpha/\beta)_8$-barrel enzyme; Strand $\beta 2$; Motif search

One of the most universal protein folding motif is the $(\alpha/\beta)_8$- or TIM-barrel fold composed of eight parallel β-strands forming the inner β-barrel that is surrounded by eight α-helices forming the cylinder outside the barrel. This protein family comprises more than 40 different enzyme specificities and three proteins with no known enzymatic function [1]. Although the question of evolutionary relationships (either convergent or divergent) of this broad protein family are still not strictly understood, there are more arguments for divergent evolution from a common ancestor [2–6].

* Tel.: +421 7 392990; fax: +421 7 393824; e-mail: sjanecek@ue.savba.sk

PII S0141-8130(97)00067-6

One of the partial solutions of the evolutionary problems in this protein folding family has been offered recently: The so-called hidden homologies of $(\alpha/\beta)_8$-barrel proteins [4,6]. If we accept a view that there can be a family of proteins that have diverged so far that no detectable homologies are left [7], then the postulate that the homologies in such a family have to be unsatisfactory from the structural point of view and are in fact hidden, should also be acceptable [8]. The recently described hidden homology covering the strand $\beta 2$ flanked in loops by the invariant glycines and prolines, respectively, of nearly half of the present-day $(\alpha/\beta)_8$-barrel enzymes [6], has indicated that some parts of the $(\alpha/\beta)_8$-barrel structure may have evolved more slowly than the rest of the structure or are more resistant to the evolutionary pressure. However, it should be pointed out that the hidden homologies can be explained also by the structural point of view alone, i.e. they may be due to structural constraints of the $(\alpha/\beta)_8$-barrel fold or for folding of the fold [1].

Therefore it is necessary to test the hypothesis whether the hidden homologies are important from the evolutionary point of view. One way of doing this is to use the strand $\beta 2$ of a known $(\alpha/\beta)_8$-barrel as the query in a BLAST [9] search. And indeed, the strand $\beta 2$ of the $(\alpha/\beta)_8$-barrel of a small sugar isomerase, xylose isomerase, was able to catch two proteins of unknown three-dimensional structure that could be of eventual evolutionary relatedness with xylose isomerase: Lycopene cyclase and replicase. While the analysis of the lycopene cyclase sequence was reported recently [10] and was considered to be easier due to the comparable sequence lengths of xylose isomerase and lycopene cyclase, the analysis concerning the replicase sequence is presented here.

The BLAST search was carried out using the $\beta 2$-strand segment of the xylose isomerase from *Streptomyces olivochromogenes* [11–13] (46_GA*HGVTFH*DDDLIP; strand $\beta 2$ in italics) throughout the non-redundant database (GenBank coding sequence translations, Protein Data Bank, SwissProt, SwissProt update and PIR; 207851 sequences). The four other structures of related xylose isomerases are available: Those from *Actinoplanes missouriensis* [14], *Arthrobacter* strain B3278 [15], *Streptomyces rubiginosus* [16] and *Streptomyces albus* [17].

As shown in Fig. 1, starting with the residue 644, the replicase from garlic latent virus [18] goes well with xylose isomerase. For comparison, the sequence of lycopene cyclase from *Erwinia herbicola*, an enzyme implicated in the β-carotene biosynthesis [19] that was previously identified as similar to the xylose isomerase [10], is added. The alignment indicates a possibility that the replicase (1924 residues) could be a multidomain protein containing an $(\alpha/\beta)_8$-barrel fold which is similar to that of xylose isomerase and formed approximately by the residues 644–980.

As far as the degrees of sequence identity and similarity between xylose isomerase and replicase are concerned, these values are given in Table 1. It is worth mentioning that Farber et al. [11] have found that there is different conservation of amino acid residues in the first and second halves of the $(\alpha/\beta)_8$-barrel of xylose isomerases. Interestingly, the value of identical amino acid residues between xylose isomerase and replicase is about 5% higher when taking only the first half of the xylose isomerase barrel (Table 1).

Moreover, the loop connecting the strand $\beta 2$ to helix $\alpha 2$ (54–64) which allows access to the active site of xylose isomerase [11], belongs to one of the most similar sequence features between xylose isomerase and replicase. It begins just after the catalytic His53 of xylose isomerase that has unambiguous equivalents in both replicase and lycopene cyclase (Fig. 1). Furthermore, this part of xylose isomerase sequence was, in fact, able to catch both sequences of structurally unknown enzymes, lycopene cyclase and replicase, using the BLAST tool. Of the two additional highly conserved sequence regions of xylose isomerases described by Farber et al. [11]: 136_WGGREG and 179_IEPKP (see Fig. 1), the former region seems to be absent either in replicase or in lycopene cyclase and is thus proposed to play its functional role (making contacts in dimer [11]) only in xylose isomerases. The later region contains two functionally important residues, Glu180 (metal ion co-ordination) and Lys182 (catalysis) [11], from which only the glutamate could have its equiva-

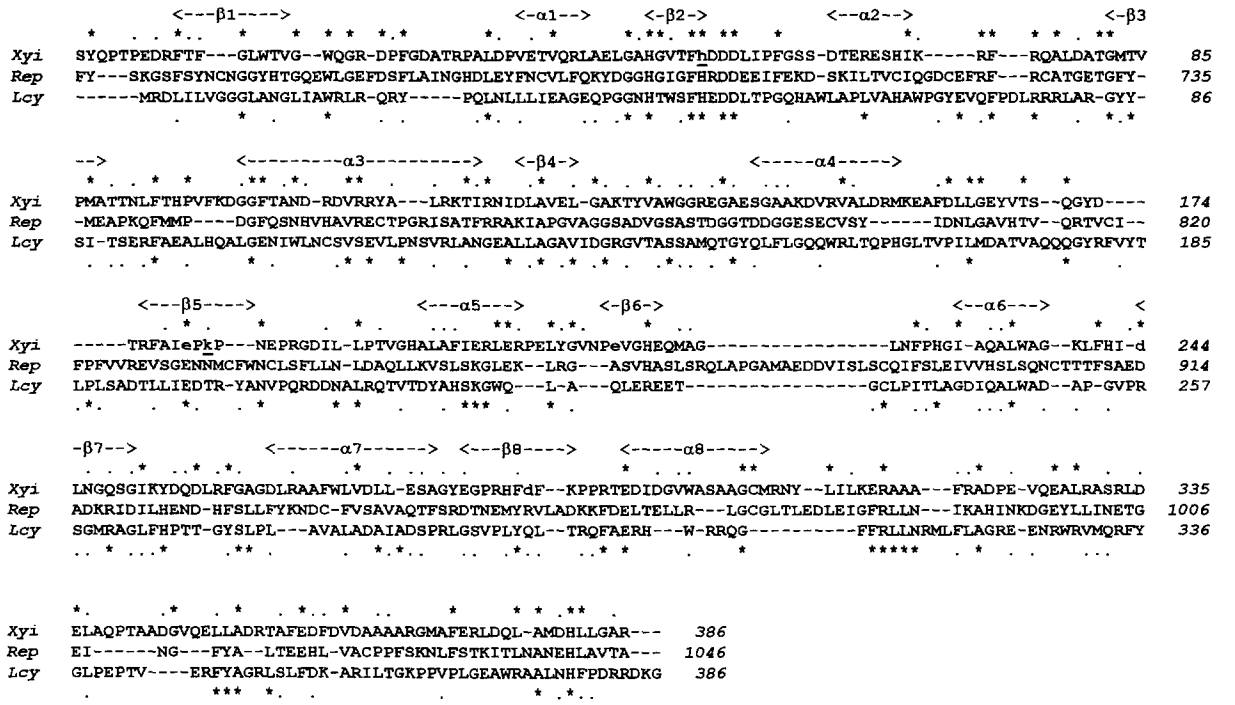

Fig. 1. Sequence alignment (made by the CLUSTAL V [21] and manually tuned where applicable) of replicase from garlic latent virus (*Rep*; GenBank: Z68502), xylose isomerase from *Streptomyces olivochromogenes* (*Xyi*; SwissProt: P15587) and lycopene cyclase from *Erwinia herbicola* (*Lcy*; SwissProt: Q01331). The secondary structure elements of the xylose isomerase $(\alpha/\beta)_8$-barrel (taken from [11]) are indicated in the top lines of the alignment blocks. The asterisks and dots signify the identical amino acid residues and conservative substitutions, respectively. Gaps are indicated by dashes. The residues involved in the metal ion co-ordination of xylose isomerase (Glu180, Glu216, Asp244, Asp286) are shown in lower case letters and those involved in the catalysis (His53, Lys182) are moreover underlined.

lent in the replicase (as well as in lycopene cyclase). However, the second important histidine, His219 (that was also considered to play a role in catalysis [11]), has very probably its equivalent histidine in replicase, as supported by the alignment with corresponding part of *Escherichia coli* xylose isomerase [20] (the alignment not shown).

As far as the similarity (identity) behind the catalytic $(\alpha/\beta)_8$-barrel of xylose isomerase is concerned, it is slightly lower in comparison with the similarity (identity) involving the entire sequence of xylose isomerase (Table 1). However, the identical pentapeptide stretches of replicase and lycopene cyclase in this C-terminal part (984_FRRLN in the replicase) correspond to an additional α-helix of xylose isomerase located just behind the $(\alpha/\beta)_8$-barrel and starting the second C-terminal loop domain [11].

Table 1
Sequence identity and similarity (%) between xylose isomerase and replicase[a]

	Entire length (386)	$(\alpha/\beta)_8$-Barrel (306)	C-terminus (80)	First half of the barrel (157)
Identity	20.7	21.2	18.8	25.5
Similarity	42.2	43.1	38.8	45.9

[a] The values represent the ratio of identical amino acid residues (for identity) or both identical residues and conservative substitutions (for similarity) calculated using the number of residues of the smaller enzyme (given in parentheses).

To summarise, it can be concluded that (1) the replicase from garlic latent virus could be a multidomain protein with an $(\alpha/\beta)_8$-barrel domain similar to that of xylose isomerase formed approximately by the residues 644–980. (2) The strand $\beta 2$ of xylose isomerase's $(\alpha/\beta)_8$-barrel could be of higher evolutionary importance than the rest of the structure since by using the BLAST tool [9] it was able to catch related sequences of lycopene cyclase [10] and replicase (this study) based on an oligopeptide as short as 14 amino acid residues.

Acknowledgements

This work was supported in part by the VEGA grant No. 2/3013/96.

References

[1] Janeček Š, Bateman A. Biologia (Bratislava) 1996;51:613.
[2] Wilmanns M, Hyde CC, Davies DR, Kirschner K, Jansonius JN. Biochemistry 1991;30:9161.
[3] Raine ARC, Scrutton NS, Mathews FS. Protein Sci 1994;3:1889.
[4] Janeček Š, Baláž S. Protein Eng 1995;8:809.
[5] Reardon D, Farber GK. FASEB J 1995;9:497.
[6] Janeček Š. Protein Sci 1996;5:1136.
[7] Lesk AM, Boswell DR. BioEssays 1992;14:407.
[8] Janeček Š. Protein Sci 1995;4:1239.
[9] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. J Mol Biol 1990;215:403.
[10] Janeček Š. Biochem J 1996;319:1005.
[11] Farber GK, Glasfeld A, Tiraby G, Ringe D, Petsko GA. Biochemistry 1989;28:7289.
[12] Tiraby G, Bejar S, Drocourt D, Reynes JP, Sicard PJ, Farber GK, Glasfeld A, Ringe D, Petsko GA. In: Hershberger CL, Queener SW, Hegeman G, (Eds.), Genetics and Molecular Biology of Industrial Microorganisms. Washington: American Society for Microbiology, 1989:119.
[13] Lavie A, Allen KN, Petsko GA, Ringe D. Biochemistry 1994;33:5469.
[14] Rey F, Jenkins J, Janin J, Lasters I, Alard P, Claessens M, Matthyssens G, Wodak S. Proteins: Struct Funct Genet 1988;4:165.
[15] Henrick K, Collyer CA, Blow DM. J Mol Biol 1989;208:129.
[16] Carrell HL, Glusker JP, Burger V, Manfre F, Tritsch D, Biellmann J-F. Proc Natl Acad Sci USA 1989;86:4440.
[17] Dauter Z, Terry H, Witzel H, Wilson KS. Acta Crystallogr 1990;B46:833.
[18] Jin Nam C, Sang Ik S, Jong Tae S, Jong Seob L, Yang Do C. Unpublished GenBank Accession No. Z68502.
[19] Hundle B, Alberti M, Nievelstein V, Beyer P, Kleinig H, Armstrong GA, Burke DH, Hearst JE. Mol Gen Genet 1994;245:406.
[20] Schellenberg GD, Sarthy A, Larson AE, Backer MP, Crabb JW, Lindstrom M, Hall BD, Furlong CE. J Biol Chem 1983;255:6826.
[21] Higgins DG, Bleasby AJ, Fuchs R. Comput Applic Biosci 1992;8:189.