BIOCHEMICAL JOURNAL LETTERS

The tentative $(\alpha/\beta)_8$-barrel in the pathway of β-carotene biosynthesis: lycopene cyclase has an amino acid sequence similar to that of xylose isomerase

The family of $(\alpha/\beta)_8$-barrel proteins grows rapidly: from 17 members (all enzymes) in 1990 [1] to 46 different proteins (mostly enzymes) at present [2]. Many other sequenced proteins and enzymes whose three-dimensional structures are to be solved very probably adopt the $(\alpha/\beta)_8$-barrel fold because of their sequence similarity to the crystallographically determined barrels [3]. Due to the fact that arguments of more or less equivalent power have been marshalled to support both divergent and convergent ways of evolution of the $(\alpha/\beta)_8$-barrel structural family, the process of how this fold has arisen remains still unclear. It seems that the structural principles determining the packing of the sheet structure in these barrels, and indicating a convergent relation between their two classes [4], cannot be used as a strong evolutionary argument, since a simple genetic mechanism has been found [5] that may be responsible for switching the nature of the β-sheet arrangement in $(\alpha/\beta)_8$-barrel proteins. The other structural criteria, such as direction of the axis of the barrel ellipse, the length of the secondary-structure elements, the presence and location of some extra strands, helices and domains, and the eventual sequence similarity, that divide the whole $(\alpha/\beta)_8$-barrel family into several subfamilies in terms of its divergence [1,3,6], are very attractive, but still not unambiguous. Nor do the recently published 'hidden homologies' comprising the strands β2 and β5 of various $(\alpha/\beta)_8$-barrels [7,8] solve the basic problem of their evolution, since at this stage of our knowledge they are acceptable for divergent as well as convergent opinions.

Thus each observation of any new evolutionary connection in this protein family has been always a welcomed event. Many of the barrels in the family apparently form pairs or groups related by divergence due to sequence similarities, such as mandelate racemase, muconate lactonizing enzyme and chloromuconate cycloisomerase [9,10], α-amylase, cyclodextrin glycosyltransferase and oligo-1,6-glucosidase [11,12], and *N*-acetylneuraminate lyase and dihydrodipicolinate synthase [13,14]. Many of these enzymes were predicted to contain the $(\alpha/\beta)_8$-barrel fold before their structures were solved, e.g. chitinases [15]. It is not a very difficult task today, mainly if the sequence (found by, e.g., BLAST [16]) exhibits a high degree of sequence similarity to the known barrel. On the other hand, several known barrel proteins (e.g. β-amylase, pyruvate kinase, xylose isomerase, etc.) seem to be 'solitary' from the evolutionary point of view, since they have produced no results when used for similarity searches in the sequence databases [3]. This indicates that, apart from the eventuality of imperfection in the method used, these $(\alpha/\beta)_8$-barrels may be the products either of a very early recruitment from the barrel ancestor(s) or contain the stretches in their sequence that are not equivalent from the evolutionary point of view. If one searches for the unknown $(\alpha/\beta)_8$-barrel having only residual similarity to the known barrel(s), the selection of a polypeptide chain segment of this known $(\alpha/\beta)_8$-barrel protein that will be used as the query is of crucial importance. In this case, the idea that some structural parts of an $(\alpha/\beta)_8$-barrel fold (the so-called 'hidden homologies', which are described elsewhere [7,8,17]) evolve more slowly or are more resistant to the evolutionary pressure than the rest of the barrel structure could be very helpful.

For instance, the second β-strand of many different $(\alpha/\beta)_8$-barrel enzymes is flanked in loops by invariant glycine and proline residues [7]. In order to test the hypothesis that the hidden homologies are important from the point of view of the $(\alpha/\beta)_8$-barrel fold evolution, the segment comprising the strand β2 of *Streptomyces olivochromogenes* xylose isomerase (46-GAHGVTFHDDDLIP, β2-strand underlined [18]) was used as the query in a BLAST [16] search throughout the non-redundant database (GenBank coding sequence translations, Protein Data Bank, SwissProt, SwissProt update, and PIR; 203478 sequences). Remarkably, the output from the BLAST search (not shown) contained, except for many various xylose isomerases, two different proteins of unknown structure: lycopene cyclase from *Erwinia herbicola*, an enzyme implicated in the pathway of β-carotene biosynthesis, and a replicase of as many as 1924 amino acid residues from garlic latent virus. Despite the fact that the segment of replicase (692-GGHGIGFHRDD) exhibited a high degree of similarity to that of xylose isomerase, it was impossible to make a justified conclusion concerning the rest of replicase sequence (GenBank accession no. Z68502) in comparison with that of xylose isomerase (the alignment not shown).

As far as the lycopene cyclase [19] is concerned, the situation was entirely different. Just a few manual adjustments taking into account the $(\alpha/\beta)_8$-barrel elements of *S. olivochromogenes* xylose isomerase [18] was enough to obtain the reliable alignment using the program CLUSTAL V [20] (Figure 1). For comparison, the sequence of a bacterial xylose isomerase from *Escherichia coli* was added because it contains differences from the sequence of actinomycetal *S. olivochromogenes* xylose isomerase (reflected also in different length of their sequences). This can be seen especially in the parts where the similarity between either actinomycetal xylose isomerase and lycopene cyclase or bacterial xylose isomerase and lycopene cyclase is very high (Figure 1), i.e. the N-terminal end of strand β2 and helix α6, and strand β5 and the segment following the C-terminal end of strand β6 respectively.

As far as the degrees of identity and similarity between the two xylose isomerases and lycopene cyclase are concerned, these values are 16.1 % and 17.6 % (identities) and 40.4 % and 43.5 % (similarities) for *S. olivochromogenes* xylose isomerase and lycopene cyclase and for *E. coli* xylose isomerase and lycopene cyclase respectively. The values mean 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. Although the degree of identity is not too high, it should be taken into account that the segments of unambiguous similarity are positioned in equivalent

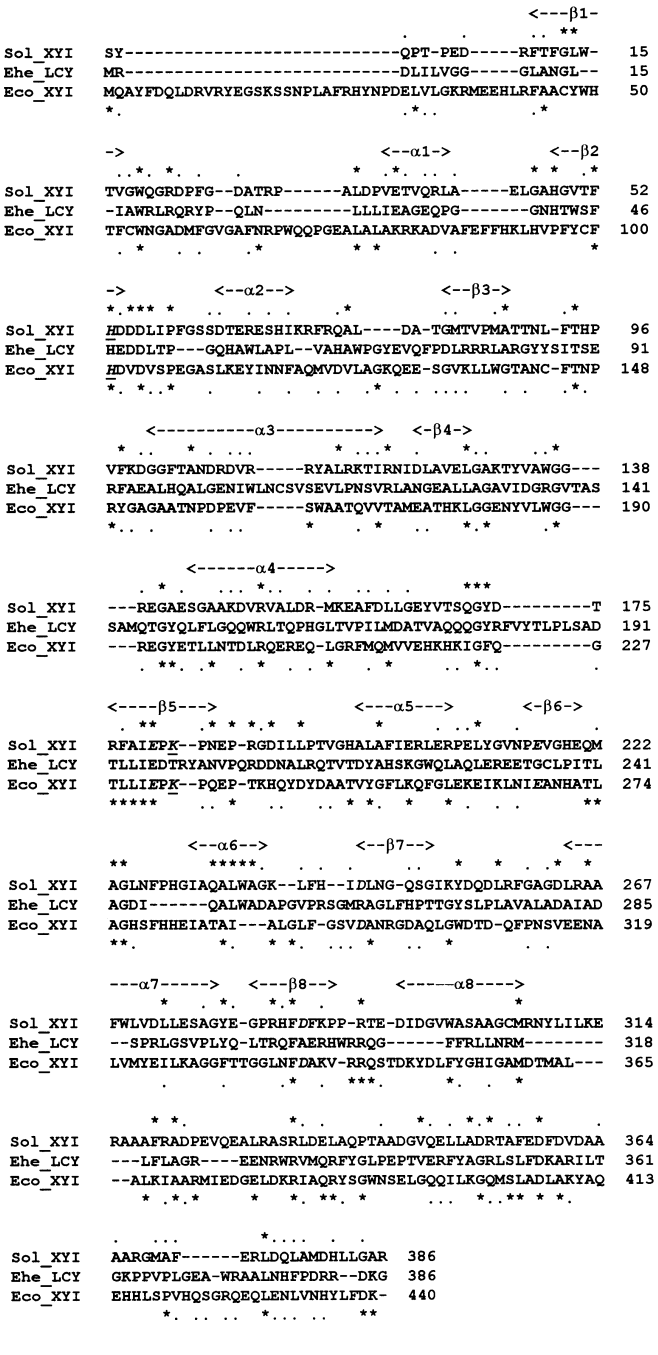

Figure 1 Sequence alignment of xylose isomerases and lycopene cyclase

Xylose isomerases from *Strep. olivochromogenes* (Sol-XYI; SwissProt accession number: P15587) and *E. coli* (Eco-XYI; P00944) and lycopene cyclase from *Erwinia herbicola* (Ehe-LCY; Q01331). The secondary-structure elements of the *S. olivochromogenes* xylose isomerase $(\alpha/\beta)_8$-barrel (taken from [18]) are indicated in the top lines of the alignment blocks. The asterisks and dots signify the identical amino acid residues and conservative substitutions respectively, for (i) Sol-XYI and Ehe-LCY over the alignment, and (ii) Eco-XYI and Ehe-LCY under the alignment. Gaps are indicated by dashes. The residues involved in the metal-ion coordination and catalysis of xylose isomerases (His-53, Glu-180, Lys-182, Glu-216, Asp-244, Asp-286 in the sequence of Sol-XYI) are italicized, and italicized and underlined, respectively.

distances in respective polypeptide chains. Moreover, the similarity is conserved also in the C-terminal parts of xylose isomerases and lycopene cyclase, i.e. behind the part corresponding with the $(\alpha/\beta)_8$-barrel of *S. olivochromogenes* xylose isomerase.

This raises a question of whether the presented observation concerning the lycopene cyclase from *E. herbicola* can be generalized for all lycopene cyclases or not. There are five other lycopene cyclase sequences available from the GenBank Sequence Database (release 95.0): *Agrobacterium aurantiacum* (accession number D58420), *Arabidopsis thaliana* (L40176), *Erwinia uredovora* (D90087), *Nicotiana tabacum* (X81787), and *Synechococcus* sp. (X74599). It should be pointed out that there is a significant difference in sequences between plant (*Arabidopsis* and *Nicotiana*) and bacterial (the rest) lycopene cyclases (the alignment not shown). It is apparent that proteobacterial (*Agrobacterium* and *Erwinia*) lycopene cyclases share the sequences of xylose isomerases as it is demonstrated in Figure 1. The difference between plant and bacterial lycopene cyclases is reflected also in the length of their sequences, i.e. the plant enzymes are about 100 amino acid residues longer. Interestingly, the cyanobacterial *Synechococcus* lycopene cyclase goes mostly with the plant enzymes. Despite the fact that the histidine equivalent to the catalytic His-53 of *S. olivochromogenes* xylose isomerase (His-47 in *E. herbicola* lycopene cyclase; cf. Figure 1) is not invariantly conserved in all available lycopene cyclase sequences, there are several equivalent segments (mainly the strands β2 and β4 with following C-terminal loops, and strand β5 including the invariant glutamate equivalent to the magnesium ligand in xylose isomerases) that indicate a relatedness between xylose isomerases and lycopene cyclases.

In conclusion, the importance of the present study is dual: (i) it predicts that lycopene cyclase very probably adopts the structure of an $(\alpha/\beta)_8$-barrel fold related to that of xylose isomerase; and (ii) it demonstrates that hidden homologies of $(\alpha/\beta)_8$-barrel proteins are able to catch a related sequence based on an oligopeptide as short as 14 amino acid residues.

This work was supported in part by the grant no. 2/3013/96 from the Slovak Grant Agency for Science. I am grateful to Greg Farber for fruitful discussions.

Štefan JANEČEK

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

1 Farber, G. K. and Petsko, G. A. (1990) Trends Biochem. Sci. **15**, 228–234
2 Janeček, Š. and Bateman, A. (1996) Biologia **51**, in the press
3 Reardon, D. and Farber, G. K. (1995) FASEB J. **9**, 497–503
4 Lesk, A. M., Brändén, C.-I. and Chothia, C. (1989) Proteins **5**, 139–148
5 Raine, A. R. C., Scrutton, N. S. and Mathews, F. S. (1994) Protein Sci. **3**, 1889–1892
6 Farber, G. K. (1993) Curr. Opin. Struct. Biol. **3**, 409–412
7 Janeček, Š. (1996) Protein Sci. **5**, 1136–1143
8 Janeček, Š. and Baláž, Š. (1995) Protein Eng. **8**, 809–813
9 Neidhart, D. J., Kenyon, G. L., Gerlt, J. A. and Petsko, G. A. (1990) Nature (London) **347**, 692–694
10 Hoier, H., Schlömann, M., Hammer, A., Glusker, J. P., Carrell, H. L., Goldman, A., Stezowski, J. J. and Heinemann, U. (1994) Acta Crystallogr. **D50**, 75–84
11 Klein, C. and Schulz, G. E. (1991) J. Mol. Biol. **217**, 737–750
12 Kizaki, H., Hata, Y., Watanabe, K., Katsube, Y. and Suzuki, Y. (1993) J. Biochem. (Tokyo) **113**, 646–649
13 Izard, T., Lawrence, M. C., Malby, R. L., Lilley, G. G. and Colman, P. M. (1994) Structure **2**, 361–369
14 Mirwaldt, C., Korndörfer, I. and Huber, R. (1995) J. Mol. Biol. **246**, 227–239
15 Coulson, A. F. W. (1994) FEBS Lett. **354**, 41–44
16 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) J. Mol. Biol. **215**, 403–410
17 Janeček, Š. (1995) Protein Sci. **4**, 1239–1242
18 Farber, G. K., Glasfeld, A., Tiraby, G., Ringe, D. and Petsko, G. A. (1989) Biochemistry **28**, 7289–7297
19 Hundle, B., Alberti, M., Nievelstein, V., Beyer, P., Kleinig, H., Armstrong, G. A., Burke, D. H. and Hearst, J. E. (1994) Mol. Gen. Genet. **254**, 406–416
20 Higgins, D. G., Bleasby, A. J. and Fuchs, R. (1992) Comput. Applic. Biosci. **8**, 189–191

Received 12 July 1996