Biologia, Bratislava, **57**/3: 283—288, 2002

Evolutionary relationships of glycolytic $(\beta/\alpha)_8$-barrel enzymes present in completely sequenced genomes

Andrea Kováčová & Štefan Janeček*

Institute of Molecular Biology, Slovak Academy of Sciences, Dúbravská cesta 21, SK-84251 Bratislava, Slovakia; tel.: ++ 421 2 5930 7420, fax: ++ 421 2 5930 7416, e-mail: umikstef@savba.sk

Kováčová, A. & Janeček, Š., Evolutionary relationships of glycolytic $(\beta/\alpha)_8$-barrel enzymes present in completely sequenced genomes. Biologia, Bratislava, **57:** 283—288, 2002; ISSN 0006-3088. (Biologia). ISSN 1335-6399 (Biologia. Section Cellular and Molecular Biology).

Fifty completely sequenced genomes of bacterial, archaeal and eukaryotic organisms were searched for the genes coding for the four glycolytic enzymes adopting the structure of a parallel $(\beta/\alpha)_8$-barrel fold. Since the enzyme fructose-1,6-bisphosphate aldolase forms two distinct classes, the search was focused on the three rest $(\beta/\alpha)_8$-barrels. The genes encoding the triosephosphate isomerase, enolase and pyruvate kinase were found in 43 genomes. The corresponding amino acid sequences were collected and aligned, and the three respective evolutionary trees were constructed and discussed. The results obtained indicate that: (i) the glycolytic pathway may be in some organisms incomplete or some other, alternative and/or unique enzymes (reaction steps) should be taken into account in order to complete the pathway; and (ii) the three $(\beta/\alpha)_8$-barrel glycolytic enzymes, as the enzymes belonging to the same biochemical pathway, are sequentially independent proteins which moreover may have their own evolutionary history as reflected by their evolutionary trees.

Key words: TIM-barrel, glycolysis, complete genome, sequence analysis, evolution.

Abbreviations: ENOL, enolase; FALD, fructose-1,6-bisphosphate aldolase; PK, pyruvate kinase; TIM, triosephosphate isomerase.

Introduction

The glycolysis, as the biochemical pathway, is very suitable for studying the protein evolution. It is a central metabolic pathway, virtually ubiquitous, so that it enables to compare the enzymes isolated from a wide spectrum of organisms (Fothergill-Gilmore, 1986). Three-dimensional structures of all 10 enzymes catalysing the individual steps of glycolysis have already been determined (Muirhead & Watson, 1992). The solved structures together with the fact that the glycolytic enzymes belong probably to the most highly conserved enzymes known, i.e. evolving slowly (Fothergill-Gilmore, 1986), make the evolutionarily oriented studies justified. These efforts are strengthened by the wealth of sequence data available from about 50 sequenced complete genomes of all the three domains of life, Bacteria, Archaea and Eucarya (Woese & Fox, 1977; Woese et al., 1990). For the actual status on sequenced genomes follow, e.g., the link "Genomes" on the ENTREZ web-site (Schuler et al., 1996) at the URL: http://www.ncbi.nlm.nih.gov/.

* Corresponding author

Four of the ten glycolytic enzymes rank among the $(\beta/\alpha)_8$-barrel proteins, i.e. the proteins adopting the structure of a parallel β-sheet forming the inner barrel surrounded by the outside α-helical cylinder (for reviews, see JANEČEK & BATEMAN, 1996; PUJADAS & PALAU, 1999). Since this folding motif is perhaps the most frequently occurring motif in proteins, found approximately in every tenth enzyme whose structure was solved (FARBER & PETSKO, 1990; BRÄNDÉN, 1991; WOLF et al., 1999), as well as it has been deeply studied in recent years (GERLT & BABBITT, 2001; HÖCKER et al., 2001; WIERENGA, 2001), our attention was primarily focused on the $(\beta/\alpha)_8$-barrel glycolytic enzymes. These enzymes are: fructose-1,6-bisphosphate aldolase (FALD), triosephosphate isomerase (TIM), enolase (ENOL) and pyruvate kinase (PK). However, there are two classes of FALD which are sequentially and structurally dissimilar appearing thus to be evolutionary distinct (SYGUSCH et al., 1987). Although both forms of the class I and class II FALDs have been recognised as $(\beta/\alpha)_8$-barrels (SYGUSCH et al., 1987; BLOM et al., 1996), the fact, that they need not be necessarily present in each organism, does not make it possible to take this $(\beta/\alpha)_8$-barrel glycolytic enzyme into the present evolutionary comparison.

This study therefore brings the comparable evolutionary trees for the three $(\beta/\alpha)_8$-barrel glycolytic enzymes, TIM, ENOL and PK, constructed for the equivalent sets of bacterial, archaeal and eukaryotic organisms whose genomes were already completely sequenced.

Material and methods

All amino acid sequences were retrieved from the complete sequenced genomes from GenBank (BENSON et al., 2000) on the ENTREZ system (SCHULER et al., 1996). The SWISSPROT sequence database (BAIROCH & APWEILER, 2000) was also used. FALD was not taken into the analysis due to the existence of two different classes that are not present in each organism simultaneously. In order to make the results comparable focus was on the organisms from which the sequences of the three rest $(\beta/\alpha)_8$-barrel glycolytic enzymes (TIM, ENOL and PK) were available. Thus from the 50 completely sequenced genomes only 43 contained the sequences of TIM, ENOL and PK (Tab. 1).

The sequences were aligned using the program CLUSTAL W (THOMPSON et al., 1994) and the computer-produced alignments were slightly manually tuned where applicable. The final alignments served for calculation by the neighbor-joining method (SAITOU & NEI, 1987) of the evolutionary trees, one for each $(\beta/\alpha)_8$-barrel enzyme. The Phylip format tree output was applied using the bootstrapping procedure (FELSENSTEIN, 1985); the number of bootstrap trials used was 1000. The trees were drawn with the program TreeView (PAGE, 1996).

Results and discussion

The search for the $(\beta/\alpha)_8$-barrel enzymes involved in glycolysis revealed that this important metabolic pathway may be in some organisms incomplete since a few of the sequenced

Table 1. The list of completely sequenced genomes containing the genes coding for TIM, ENOL and PK.

Domain of life	Organism
Bacteria	*Bacillus halodurans*
	Bacillus subtilis
	Borrelia burgdorferi
	Buchnera sp. APS
	Campylobacter jejuni
	Caulobacter crescentus
	Chlamydia muridarum
	Chlamydia trachomatis
	Chlamydophila pneumoniae AR39
	Chlamydophila pneumoniae CWL029
	Chlamydophila pneumoniae J138
	Deinococcus radiodurans R1
	Escherichia coli K12
	Escherichia coli O157
	Lactococcus lactis
	Mesorhizobium loti
	Mycobacterium leprae
	Mycobacterium tuberculosis CDC1551
	Mycobacterium tuberculosis H37Rv
	Mycoplasma genitalium
	Mycoplasma pneumoniae
	Mycoplasma pulmonis
	Neisseria meningitidis A Z2491
	Neisseria meningitidis B MC58
	Pseudomonas aeruginosa PA01
	Staphylococcus aureus Mu50
	Staphylococcus aureus N315
	Streptococcus pyogenes M1 GAS
	Synechocystis sp. PCC6803
	Thermotoga maritima
	Ureaplasma urealyticum
	Vibrio cholerae
	Xylella fastidiosa 9a5c
Archaea	*Methanococcus jannaschii*
	Pyrococcus abyssi
	Pyrococcus horikoshii
	Sulfolobus solfataricus
	Thermoplasma volcanium
Eucarya	*Arabidopsis thaliana*
	Ceanorhabditis elegans
	Homo sapiens
	Mus musculus
	Saccharomyces cerevisiae

genomes lack the respective genes. Thus the genomes of seven microorganisms seem to miss one of the three $(\beta/\alpha)_8$-barrel enzymes, mostly PK: *Aquifex aeolicus* (PK), *Archaeoglobus fulgidus* (PK), *Halobacterium* sp. (ENOL), *Helicobacter pylori* (PK), *Methanothermobacter thermoautotrophicus* (PK), *Plasmodium falciparum* (PK) and *Treponema pallidum* (PK). DANDEKAR et al. (1999) have proposed that the absence of PK may indicate the presence of yet another undetected enzyme which has displaced the "classical" form of PK. This fact is not too surprising because the other important metabolic pathway of the citric-acid cycle was found to be incomplete or absent in several completely sequenced genomes (HUYNEN et al., 1999).

Genomes of 43 different organisms belonging to all the three domains of life (Bacteria, Archaea and Eucarya) contain the genes (or ORFs coding for the putative proteins) of the three $(\beta/\alpha)_8$-barrel enzymes, TIM, ENOL and PK (as of June 2001). Five organisms belong to the domain Archaea, five further ones to Eucarya and the rest 33 ones to Bacteria (Tab. 1). The fact that the organisms listed in Table 1 possess in their genomes the genes encoding the TIM, ENOL and PK does not necessarily mean that these organisms contain the full set of glycolytic enzymes, i.e. they are able to perform the complete glycolysis. The comparative study focused on all the enzymes from the complete glycolytic pathway is under way and the results will be published elsewhere.

The amino acid sequences of all 43 TIMs, ENOLs and PKs were aligned (the alignments are not shown) in order to calculate the evolutionary trees (Fig. 1) based on the alignments. Table 2 indicates simplified trends of evolutionary relatedness among $(\beta/\alpha)_8$-barrel glycolytic enzymes derived from bacterial (*Bacillus subtilis*), archaeal (*Methanococcus jannaschii*) and eukaryotic (*Saccharomyces cerevisiae*) origins. In general, it seems that the degree of conservation decreases in the order ENOLs (average sequence identity 55.8%), PKs (36.9%) and TIMs (29.3%). The substantially higher sequence identity between the *Bacillus* and yeast TIMs (41.3% in comparison to values lower than 25%) is of interest (Tab. 2).

The evolutionary relationships among 43 organisms in the frame of all the three $(\beta/\alpha)_8$-barrel glycolytic enzymes can be inferred from the trees shown individually for TIM, ENOL and PK (Fig. 1). First of all, it seems evident that each glycolytic enzyme may have its own evolutionary history since the relatedness among the individual living systems is different for each enzyme. The clustering does not reflect the known taxonomy deduced from the comparison of small-subunit rRNA sequences (PACE, 1997). The eukaryotes (human, mouse, nematode, plant and yeast) and archaeons (two *Pyrococcus*, *Methanococcus*, *Sulfolobus* and *Thermoplasma*) are the best conserved groups in all the three trees (Fig. 1), however, they are not comparably related to each other. Remarkably the two groups are on adjacent branches in the ENOL tree (Fig. 1). All the archaeons used in this study (Tab. 1) belong to extremophilic microorganisms producing mainly thermostable proteins and enzymes (LÉVÊQUE et al., 2000). The extremely thermophilic bacterium *Thermotoga maritima* is found in one larger cluster with the archaeons in the ENOL and PK trees. In the TIM tree (Fig. 1) *Thermotoga* is placed close to chlamydiae. The other bacterium clustered next to archaebacteria in all the evolutionary trees is *Campylobacter jejuni*, the fact that is especially evident in the TIM and ENOL trees. The rest of bacteria, in fact, respect the well-established bacterial taxonomic groups, however, the observed evolutionary relationships are not uniformly conserved for the three $(\beta/\alpha)_8$-barrel glycolytic enzymes. This observation can be supported by the three-dimensional structures of TIM, ENOL and PK that, although being characterised by a common $(\beta/\alpha)_8$-barrel domain, have been classified into different structural groups (FARBER & PETSKO, 1990).

In conclusion, our results are in agreement with the remarks by COPLEY & BORK (2000)

Table 2. The values in % of degrees of sequence identity and similarity (in parenthesis) between the pairs of triosephosphate isomerases, enolases and pyruvate kinases representing the bacterial, archaeal and eukaryotic domain of life.

Triosephosphate isomerase	*B. subtilis*	*M. jannaschii*
M. jannaschii	24.2 (40.6)	
S. cerevisiae	41.3 (59.5)	22.4 (39.3)
Enolase	***B. subtilis***	***M. jannaschii***
M. jannaschii	57.5 (72.6)	
S. cerevisiae	52.1 (69.5)	57.7 (71.4)
Pyruvate kinase	***B. subtilis***	***M. jannaschii***
M. jannaschii	35.6 (57.5)	
S. cerevisiae	41.0 (57.2)	34.0 (52.4)

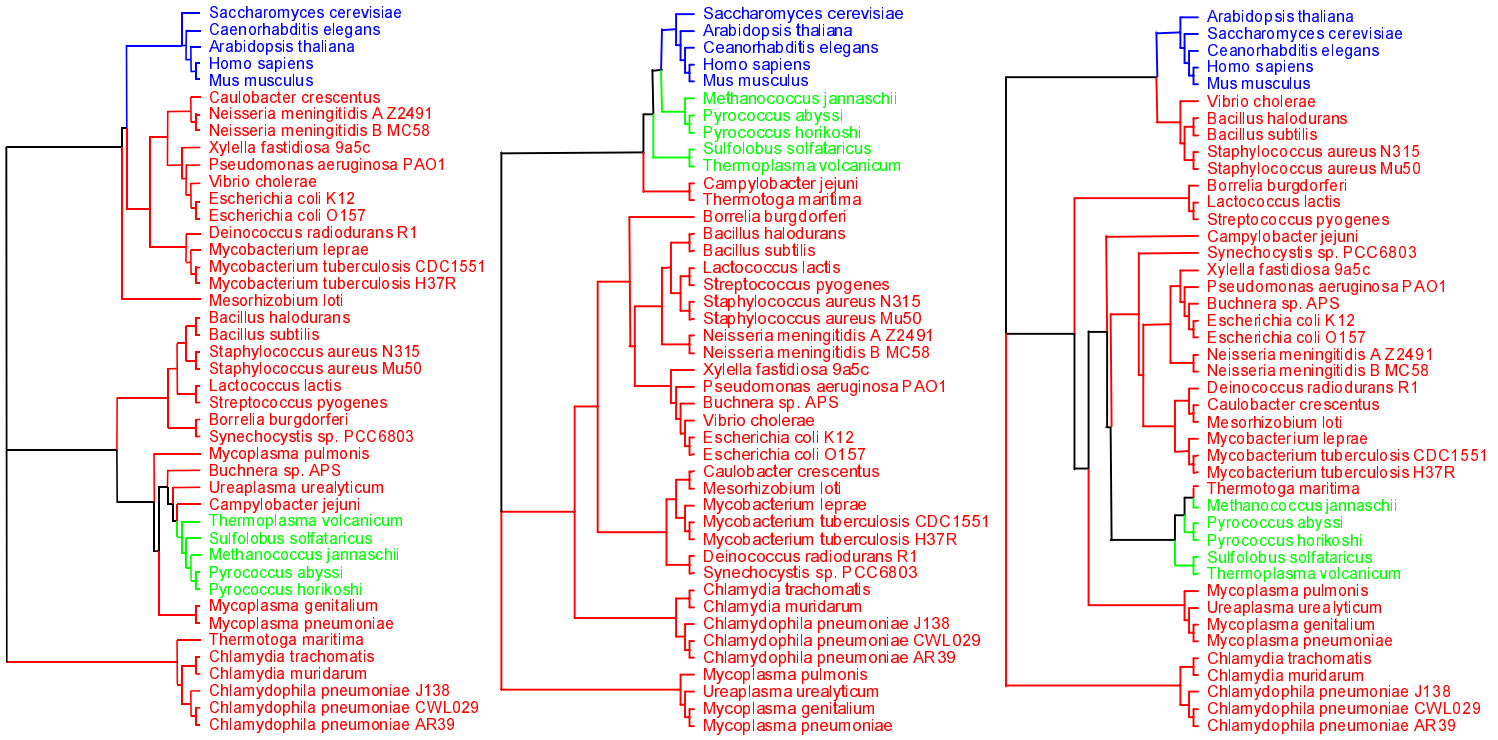

Fig. 1. Evolutionary trees of the three $(\beta/\alpha)_8$-barrel glycolytic enzymes. The trees are based on the alignment of entire amino acid sequences of triosephosphate isomerases, enolases and pyruvate kinases present if 43 completely sequenced genomes. Colour code: bacteria – red, archaea – green, and eukarya – blue.

who have found no strong evidence on evolution of biochemical pathways in the terms that enzymes within a pathway are likely to be homologous. The TIM, ENOL and PK as the enzymes and the glycolysis as the pathway seem to be very good examples of such an evolutionary behaviour (Fig. 1). With regard to the conservative glycine and proline residues flanking in loops the strand $\beta 2$ of many $(\beta/\alpha)_8$-barrel enzymes, the three glycolytic enzymes were also revealed to be structurally different (Janeček, 1996): ENOL contains both Gly and Pro residues, PK has the Gly whereas TIM possesses the Pro. Finally, concerning the apparent incompleteness of glycolysis in several genomes, it should be pointed out that it is not the feature characteristic of glycolysis only. Beside the incompleteness of the citric-acid cycle mentioned above (Huynen et al., 1999) some archaeons were found to miss any glycoside hydrolases indicating their life with no sugars (Coutinho & Henrissat, 1999). Thus where enzymes of genomes expected in the "classical" biochemical steps are missing, some gene and functional modifications and alternative enzymes have to be taken into account (Cordwell, 1999). To shed more light on this phenomenon, the evolutionary relationships of the three $(\beta/\alpha)_8$-barrel enzymes of glycolysis were described for 43 completely sequenced genomes and the analysis of the entire glycolytic pathway is under way.

Acknowledgements

This work was supported in part by the VEGA grant No. 2/2057/22.

References

Bairoch, A. & Apweiler, R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. **28:** 45–48.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Rapp, B. A. & Wheeler, D. L. 2000. GenBank. Nucleic Acids Res. **28:** 15–18.

Blom, N. S., Tétreault, S., Coulombe, R. & Sygusch, J. 1996. Novel active site in *Escherichia coli* fructose-1,6-bisphosphate aldolase. Nature Struct. Biol. **3:** 856–862.

Brändén, C. I. 1991. The TIM-barrel – the most frequently occurring folding motif in proteins. Curr. Opin. Struct. Biol. **1:** 978–983.

Copley, R. R. & Bork, P. 2000. Homology among $(\beta\alpha)_8$-barrels: implications for the evolution of metabolic pathways. J. Mol. Biol. **303:** 627–640.

Cordwell, S. J. 1999. Microbial genomes and "missing" enzymes: redefining biochemical pathways. Arch. Microbiol. **172:** 269–279.

Coutinho, P. M. & Henrissat, B. 1999. Life with no sugars? J. Mol. Microbiol. Biotechnol. **1:** 307–308.

Dandekar, T., Schuster, S., Snel., B., Huynen, M. & Bork, P. 1999. Pathway alignment: application to the comparative analysis of glycolytic enzymes. Biochem. J. **343:** 115–124.

Farber, G. K. & Petsko, G. A. 1990. The evolution of α/β barrel enzymes. Trends Biochem. Sci. **15:** 228–234.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:** 783–791.

Fothergill-Gilmore, L. A. 1986. The evolution of glycolytic pathway. Trends Biochem. Sci. **11:** 47–51.

Gerlt, J. A. & Babbitt, P. C. 2001. Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. Annu. Rev. Biochem. **70:** 209–246.

Höcker, B., Jürgens, C., Wilmanns, M. & Sterner, R. 2001. Stability, catalytic versatility and evolution of the $(\beta\alpha)_8$-barrel fold. Curr. Opin. Biotechnol. **12:** 376–381.

Huynen, M. A., Dandekar, T. & Bork, P. 1999. Variation and evolution of the citric-acid cycle: a genomic perspective. Trends Microbiol. **7:** 281–291.

Janeček, Š. 1996. Invariant glycines and prolines flanking in loops the strands $\beta 2$ of various $(\alpha/\beta)_8$-barrel enzymes: a hidden homology? Protein Sci. **5:** 1136–1143.

Janeček, Š. & Bateman, A. 1996. The parallel $(\alpha/\beta)_8$-barrel: perhaps the most universal and the most puzzling protein folding motif. Biologia, Bratislava **51:** 613–628.

Lévêque, E., Janeček, Š., Haye, B. & Belarbi, A. 2000. Thermophilic archaeal amylolytic enzymes. Enzyme Microb. Technol. **26:** 3–14.

Muirhead, H. & Watson, H. 1992. Glycolytic enzymes: from hexose to pyruvate. Curr. Opin. Struct. Biol. **2:** 870–876.

Pace, N. R. 1997. A molecular view of microbial diversity and the biosphere. Science **276:** 734–740.

Page, R. D. 1996. TreeView: an application to display phylogenetic trees on personal computers. Comput. Applic. Biosci. **12:** 357–358.

Pujadas, G. & Palau, J. 1999. TIM-barrel fold: structural, functional and evolutionary characteristics in natural and designed molecules. Biologia, Bratislava **54:** 231–254.

Saitou, N. & Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

Schuler, G. D., Epstein, J. A., Ohkawa, H. & Kans, J. A. 1996. Entrez: molecular biology database and retrieval system. Methods Enzymol. **266:** 141–162.

Sygusch, J., Beaudry, D. & Allaire, M. 1987. Molecular architecture of rabbit skeletal muscle aldolase at 2.7-Å resolution. Proc. Natl. Acad. Sci. USA **84:** 7846–7850.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment trough sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Woese, C. R. & Fox, G. E. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. USA **74:** 5088–5090.

Woese, C. R., Kandler, O. & Wheelis, M. L. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc. Natl. Acad. Sci. USA **87:** 4576–4579.

Wolf, Y. I., Brenner, S. E., Bash, P. A. & Koonin, E. V. 1999. Distribution of protein folds in the three superkingdoms of life. Genome Res. **9:** 17–26.

Wierenga, R. K. 2001. The TIM-barrel fold: a versatile framework for efficient enzymes. FEBS Lett. **492:** 193–198.

Received September 3, 2001
Accepted February 12, 2002