

Molecular evolution of enolase*

Michał Piast[✉], Irena Kustrzeba-Wójcicka, Małgorzata Matusiewicz and Teresa Banaś

Department of Medical Biochemistry, Wrocław Medical University, Wrocław, Poland;

[✉]e-mail: piast@bioch.am.wroc.pl

Received: 13 October, 2004; revised: 24 January, 2005; accepted: 07 February, 2005

available on-line: 15 May, 2005

Enolase (EC 4.2.1.11) is an enzyme of the glycolytic pathway catalyzing the dehydration reaction of 2-phosphoglycerate. In vertebrates the enzyme exists in three isoforms: α , β and γ . The amino-acid and nucleotide sequences deposited in the GenBank and SwissProt databases were subjected to analysis using the following bioinformatic programs: ClustalX, GeneDoc, MEGA2 and S.I.F.T. (sort intolerant from tolerant). Phylogenetic trees of enolases created with the use of the MEGA2 program show evolutionary relationships and functional diversity of the three isoforms of enolase in vertebrates. On the basis of calculations and the phylogenetic trees it can be concluded that vertebrate enolase has evolved according to the “birth and death” model of evolution. An analysis of amino acid sequences of enolases: non-neuronal (NNE), neuron specific (NSE) and muscle specific (MSE) using the S.I.F.T. program indicated non-uniform number of possible substitutions. Tolerated substitutions occur most frequently in α -enolase, while the lowest number of substitutions has accumulated in γ -enolase, which may suggest that it is the most recently evolved isoenzyme of enolase in vertebrates.

Keywords: glycolysis, enolase, vertebrates evolution, gene duplication, concerted evolution, “birth and death” evolution

Enolase (EC 4.2.1.11) is an enzyme catalyzing the reversible conversion of 2-phosphoglycerate into phosphoenolpyruvate in the presence of Mg^{2+} ions. Our team has been investigating glycolytic enzymes, particularly enolase, for many years. Enolase in crystalline form was obtained for the first time from human muscles by Baranowski and coworkers in 1968 in the Department of Medical Biochemistry, Wrocław Medical University (Baranowski *et al.*, 1968; Baranowski & Wolna, 1975). Our research has led to the extraction of enolase from various sources (Pietkiewicz *et al.*, 1983; Kustrzeba-Wójcicka & Golczak, 2000). For some of them nucleotide and amino acid sequences have been derived (Nowak *et al.*, 1981). The enzyme can be found from archaeobacteria to mammals, and is a member of a large superfamily comprising among others carboxyphosphoenolpyruvate synthase (Babbit *et al.*, 1996; Pegg & Babbit, 1999).

In vertebrates have been identified three tissue-specific isoenzymes: non-neuronal enolase (NNE, α), neuron-specific enolase (NSE, γ) and muscle-specific enolase (MSE, β) (Rider & Taylor, 1975). Biochemical research of enolases has practical

applications in cancer and allergy diagnostics (Nitter-Marszalska *et al.*, 1998).

Phylogenetic analysis of enolases and related gene sequences points to early gene duplication. In vertebrates two duplication events took place, leading to the formation of three enolase isoenzymes (Tracy & Hedges, 2000).

The “driving force” of evolution are changes of genes such as: nucleotide substitutions, insertions, deletions and conversions; all of these changes lead to mutations of gene or DNA sequences. If the amino acid remains unchanged the substitution is called synonymous or silent. A nonsynonymous substitution leads to a change of the encoded amino acid. The vast majority of synonymous substitutions are in the third nucleotide in a codon, rarely in the first position (Nei & Kumar, 2000).

How did the genes encoding enolase evolve? The theory of concerted evolution assumes that individual members of a gene family evolve as a group. Concerted evolution leads to a high sequence identity of many copies of a gene within a species and not between species. Intraspecific homogenization, characteristic for this model of evolution, is a result

*Appendix – available at: www.actabp.pl

Abbreviations: MSE (β), muscle-specific enolase; NNE (α), non-neuronal enolase; NSE (γ) neuron-specific enolase.

of recombinant repair of DNA, non-uniform crossing-over or genetic conversion.

It has to be taken into account that the similarity of sequences in a given species does not have to be caused by active homogenization but may be an effect of gene amplification or transposition. If there is no sequence homogenization among members of a gene family, it should not be expected that they would build a monophyletic group. Some members of such a family might even link up on a phylogenetic tree with sequences of other species. Multigene families may "escape" from concerted evolution and evolve in a different way. These genes do not necessarily evolve in concerted fashion and may undergo "birth and death" evolution (Liao, 1999).

In the "birth and death" model of evolution a repeated duplication of a gene is responsible for the formation of a new gene. According to this model some genes remain in the genome for a long period, while others are removed or become nonfunctional (Nei *et al.*, 1997).

The goal of this work was to investigate the possible pathway of enolase evolution and the changes in the structure of vertebrate isoenzymes of enolase. For this purpose phylogenetic trees of available enolase protein sequences and of the coding fragments of nucleotide sequences were generated. The obtained comparisons of amino-acid sequences were subjected to analysis in order to distinguish more and less conserved regions in the enolase molecule. Finally, tolerated and non-tolerated substitutions in mammalian isoenzymes of enolase were identified.

MATERIALS AND METHODS

Amino-acid and nucleotide sequences of enolases from numerous species representing most taxonomic groups were the subject of analysis. A detailed list of the sequences together with their accession numbers can be found in the Appendix. The sequences were labeled by appropriate species names. The amino-acid and nucleotide sequences used for the analysis come from the GenBank and SwissProt databases.

The amino-acid and nucleotide sequences written in the FASTA format were aligned using CLUSTAL_X v1.83 program (Thompson *et al.*, 1997). Alignment of the sequences was conducted according to the default program configuration. The obtained comparisons were saved as *.msf files (accepted by the GeneDoc program). Some of the alignments, especially those of nucleotide sequences, contained poorly aligned regions. Edition of alignments using GeneDoc tools (Nicholas & Nicholas, 1997)¹ was

conducted to exclude areas of poor reliability and to remove unnecessary gaps.

The MEGA2 program (Kumar *et al.*, 2001) was used to construct phylogenetic trees of the studied amino-acid and nucleotide sequences. The trees were constructed using the neighbor-joining method (Saitou & Nei, 1987). In the case of the amino acid tree a Poisson correction was applied. The nucleotide sequence tree was computed using p-distance and complete deletion options. Bootstrap probability (2000 replications) was introduced to assess the statistical significance of groups in phylogenetic trees. The program was also employed for calculation of the P_S (synonymous nucleotide differences to synonymous sites) to P_N (non-synonymous nucleotide differences to non-synonymous sites) ratio for the isoforms of enolase from the selected species of vertebrates. The calculations were conducted using the Nei-Gojobori method.

S.I.F.T. tools (sort intolerant from tolerant) were applied to classify as tolerated or deleterious the substitutions occurring in enolases. S.I.F.T. analyzes amino-acid sequences in a more precise way than so called matrices, for example BLOSUM62, calculating substitutions. S.I.F.T. checks comparisons of amino-acid sequences in the *.msf format. The program conducts numerous comparisons of loaded amino-acid sequences and subsequently calculates the probability of all possible substitutions for each amino-acid residue (Ng & Henikoff, 2001).

RESULTS

Comparisons of amino-acid sequences

The Appendix* presents a comparison of amino-acid sequences of the studied enolases. The sequences are between 366 (*Archaeoglobus fulgidus*) and 456 (*Mycoplasma pneumoniae*) amino acid residues in length. Cysteine is a relatively rarely occurring amino acid and is entirely absent in enolases from *Candida albicans* and *A. pernix*. A total absence of tryptophan is characteristic for enolase from *Methanococcus jannaschii*.

Enolases have a two-domain structure. The N-terminal domain extends from amino acid 1 to approximately 134, further lies the longer C-terminal domain, adopting a β structure (residues approximately from 143 to 434). Between these domains there is a short fragment with random structure.

Pairways similarities of amino-acid sequences of analyzed enolases comprise a broad range from 22% (*Pseudomonas putida* – *A. fulgidus*) to 99% (*Mus* sp. γ – *Rattus norvegicus* γ).

¹Nicholas KB, Nicholas Jr HB (1997) GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by the authors.

Table 1. Percent similarity between enolase sequences.

	1	2	3	4	5	6	7	8	9	10
1	-									
2	22	-								
3	23	50	-							
4	24	49	84	-						
5	24	47	75	78	-					
6	25	47	73	76	85	-				
7	26	47	73	74	81	80	-			
8	25	49	75	78	83	83	89	-		
9	25	49	75	77	83	83	89	99	-	
10	25	45	72	73	77	82	82	84	84	-

1, *Pseudomonas putida* (almond acid racemase); 2, *Archaeoglobus fulgidus*; 3, *Aeropyrum alternate*; 4, *Saccharomyces cerevisiae*; 5, *Drosophila melanogaster*; 6, *H. gammarus*; 7, *Gallus gallus*; 8, *Mus sp. γ*; 9, *Rattus norvegicus γ*; 10, *Homo sapiens α*.

The similarity of selected sequences is presented in Table 1.

The N- and C-terminal regions exhibit various degrees of conservation in particular taxonomic groups. In fungal sequences the similarity in both domains is around 57%. In vertebrates the N-terminal region is more variable than the C-terminal domain, the conservation of which is 68%. In the case of amino-acid sequences derived from enolases from plants the situation is similar. The evolutionary stability of the C-terminal region is higher (70%) than of the N-terminal domain (65%).

In both these domains there exist regions which are much less susceptible to changes i.e., showing high conservation. In the N-terminal domain a fragment with the sequence Asp-Ser-Arg-Gly-Asn-Pro-Thr-Val-Glu (approx. from residue 15 to 23), which can be found in this or slightly altered form in most of the studied enolases, is one such region. Among the exceptions are enolases from *A. fulgidus*, *Drosophila pseudoobscura*, *Drosophila subobscura*, *Chlamydomonas reinhardtii* and *P. putida*. Another highly stable region in the N-terminal domain — Pro-Ser-Gly-Ala-Ser-Thr-Gly (residues approx. from 36 to 43) — is absent only from enolases from *C. reinhardtii* and *P. putida*. Exclusively in enolases from Viridiplantae, *Pseudomonas falciparum* and *Toxoplasma gondii* the Glu-Trp-Gly-Trp-Cys-Lys insert is present.

The N-terminal domain ends with highly non-variable elements Ile-Asp-Gly-Thr and Gly-Ala-Asn-Ala-Ile-Leu-Gly-Val-Ser-Leu-Ala-Val, which are absent only in the enolase from *P. putida*. The C-terminal region starts with the weakly variable fragment Leu-Pro-Val-Pro (residues from approx. 146 to 149), absent only from *P. putida*. The C-terminal domain, (approx. 290 amino-acid residues), is relatively rich in conservative elements, such as Gly-Asp-Glu-Gly-Gly-Phe-Ala-Pro, Val/Leu-Ser-His-Arg, and Gly-Gln-Ile-Lys-Thr, divided by highly variable fragments.

Evolution of enolase

Figure 1 presents a phylogenetic tree of enolase amino-acid sequences. The tree was constructed using the neighbor-joining (NJ) method because of the considerable number of sequences (77) and their substantial length. The phylogenetic tree besides the evolutionary relationships, reveals the evident differentiation of particular variants of vertebrate enolases (α , β , γ) due to their different functions. The par-

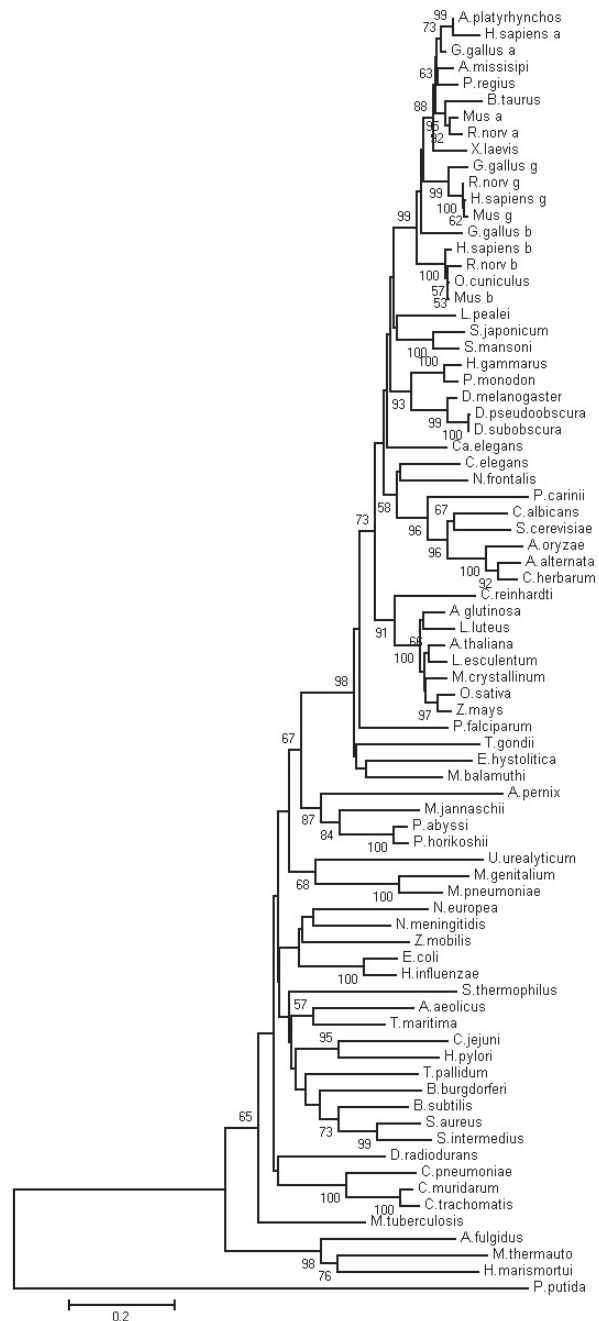


Figure 1. Phylogenetic tree of amino-acid sequences of enolases.

Neighbor-joining method (NJ) using Poisson correction. Numbers indicate bootstrap values (>50%).

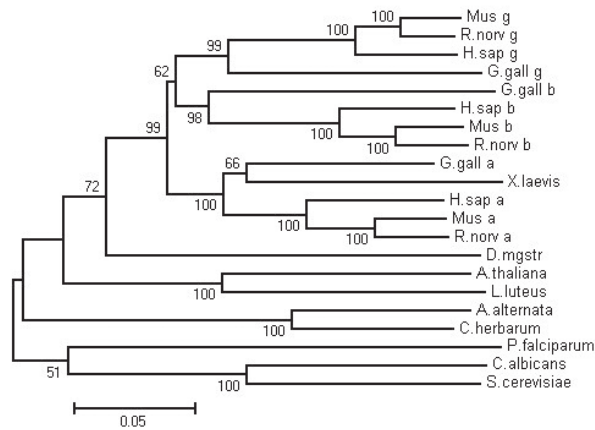


Figure 2. Phylogenetic tree of nucleotide sequences of selected enolases.

Neighbor joining (NJ) method using p -distance and complete deletion. Numbers indicate bootstrap values (>50%).

ticular isoforms of vertebrate enolases form monophyletic branches in the tree.

The most ancestral amino-acid structure, from which the further branches of the tree are derived, was found in *P. putida*. Figure 1 demonstrates the tendency of particular classes of enolases to cluster into separate tree branches. This is not governed by a strict rule, however. Enolases from *Ureaplasma urealyticum*, *M. genitalium* and *M. pneumoniae* do not assemble with other Eubacterial but with Archaeobacterial enolases. Vertebrate enolases form a big cluster, characterized by greater intraspecific than interspe-

cies similarity, manifested as separate monophyletic branches for three different enolase isoforms: non-neuronal (NNE, α), muscle-specific (MSE, β), and neuron-specific (NSE, γ).

Figure 2 presents a phylogenetic tree of selected nucleotide sequences of enolases. This tree seems to confirm the general tendencies in enolase evolution, with several exceptions. It has to be noted that the gene of *P. falciparum* — a representative of Protista — assembles on the tree with fungal genes; there is no such situation on the amino-acid sequences tree. The second difference concerns the relationships between the isoforms of vertebrate enolases. According to the generated tree of nucleotide sequences, the gene encoding NSE might be evolutionarily the most recent. However, one must bear in mind that the topology derived from nucleotide sequences is not very reliable because of the high level of saturation. Because of that only first and second codon positions were used to compute this tree.

Prediction of possible amino-acid substitutions (S.I.F.T.)

Figures 3 to 5 show variable substitution sites in three mammalian enolase isoenzymes. α -Enolase, which collected 33 possible substitution sites, nine of which are in the N-terminal domain and 24 in the C-terminal one, is the most variable (Fig. 3). The distribution of substitutions is irregular, in the C-terminal domain the lowest number of substitutions is possible in the flanking region. In mammalian β -enolases the number of tolerated substitution

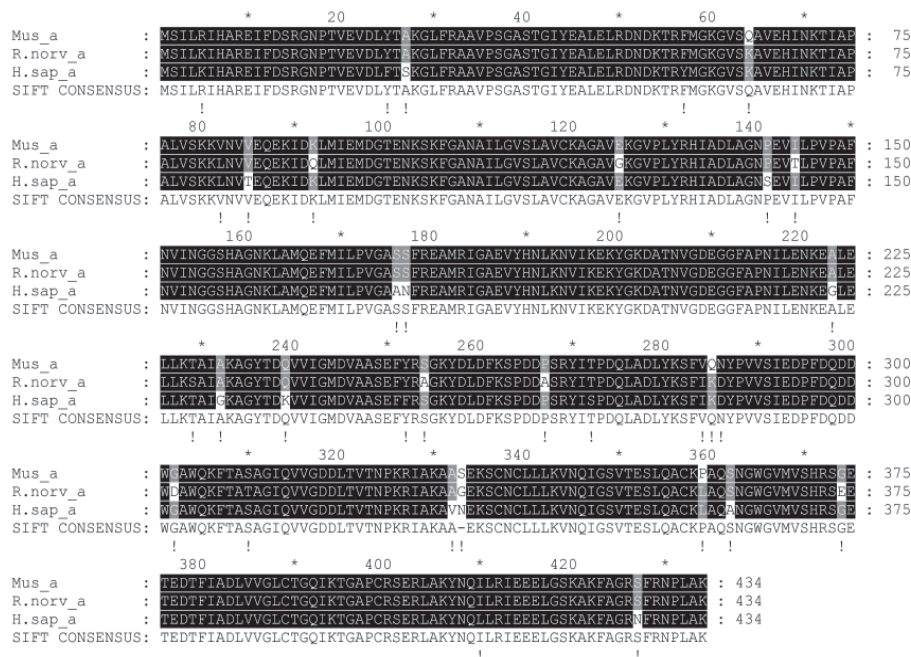


Figure 3. Tolerated amino acid substitutions in mammalian α -enolases.

The exclamation marks indicate the sites of tolerated substitutions.

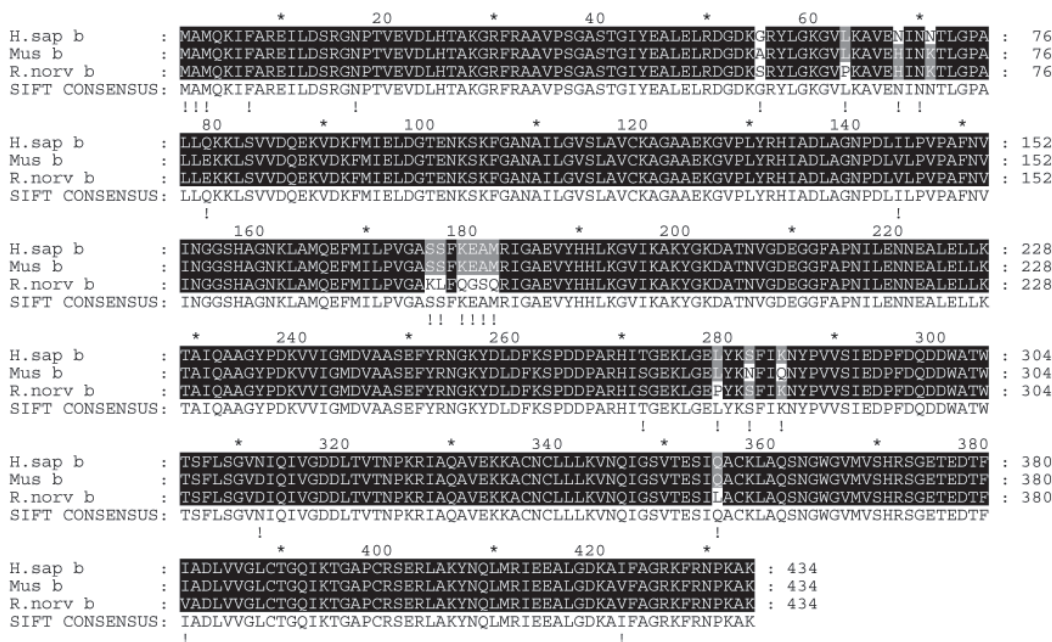


Figure 4. Tolerated amino-acid substitutions in mammalian β -enolases.

The exclamation marks indicate the sites of tolerated substitutions.

sites is lower — 25 (10 in the N-terminal and 15 in the C-terminal region). They are the most abundant in the 176–183 area (Fig. 4). γ -Enolases accumulated the lowest number of substitutions among the three classes of mammalian enolases. In their molecules only 11 sites of tolerated substitutions are present, six of which lie in the N-terminal domain (Fig. 5).

DISCUSSION

Amino-acid sequences

Enolase, an enzyme of the glycolytic pathway, catalyzes the conversion of 2-phosphoglycerate into phosphoenolpyruvate. It is present in all organisms

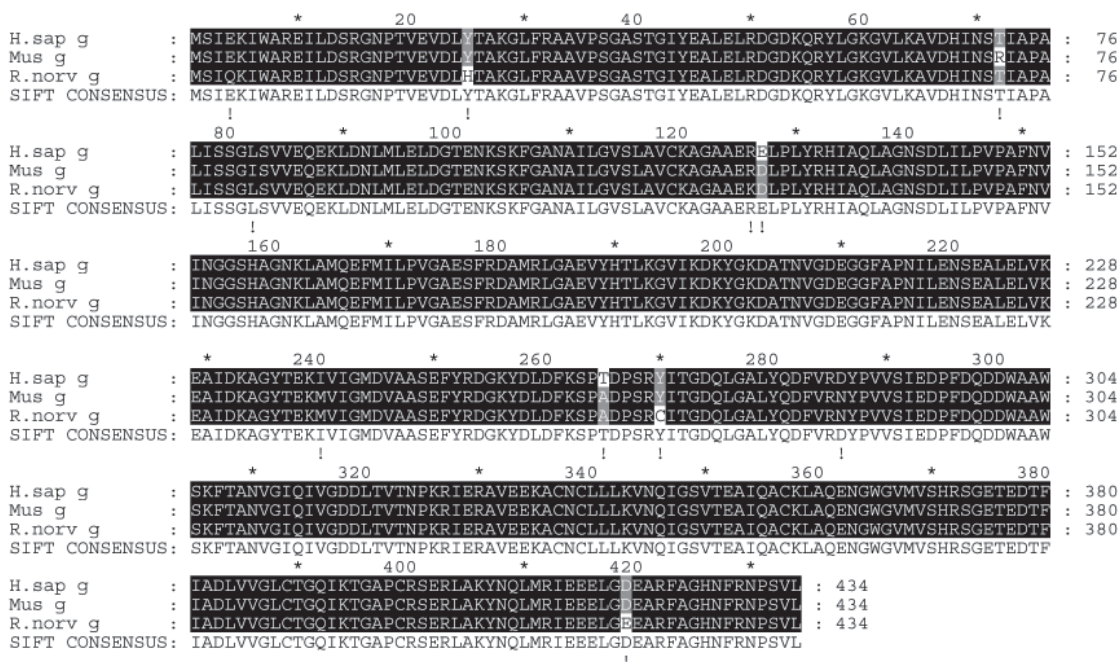


Figure 5. Tolerated amino-acid substitutions in mammalian γ -enolases.

The exclamation marks indicate the sites of tolerated substitutions.

and due to the significance of the process in which it participates (glycolysis is an important step in the process of ATP production) it has not been a subject of profound changes. The conducted comparison of amino-acid sequences, similarly as in earlier works (Van Der Straeten *et al.*, 1991; Stamm & Young, 1997; Hannaert *et al.*, 2000) demonstrated a high degree of amino acid sequence conservation in enolases. The sequence identity is high even for evolutionarily distant species, for example for *A. fulgidus* – *H. sapiens* α sequences it is 45%. The differences between enolase sequences from vertebrates and fungi are very small.

The amino-acids critical for enzyme function are situated in especially well conserved fragments of the molecule. The active site histidine (His-157 in vertebrates) is absent only in *P. putida* (mandelate racemase). Equally stable are also the ligand binding sites. Aspartic acid residues (Asp-244 and Asp-317 in vertebrates) and glutamic acid (Glu-292 in vertebrates) are present in all analyzed enolases with the exception of racemase from *P. putida*, in which there is no equivalent to Asp-244 and in position 317 asparagine is present.

Evolution of enolase

Gene duplications are probably the main source of new genes. Duplicated genes can undergo differentiation and gain new functions, can become inactive and remain in the genome as pseudogenes, and also can preserve their primary role (Liao, 1999). Duplications played an essential role in the case of enolase evolution. Two successive duplications of a single enolase gene resulted in the formation of three paralog sequences. Earlier research (Rider & Taylor, 1975) suggested that this event took place 200–300 million years ago, but according to recent work (Tracy & Hedges, 2000) these three subtypes probably arose 450 million years or earlier. The sequenced enolase isoenzymes are mostly mammalian and avian, although the presence of isoforms in other phyla of vertebrates has been demonstrated (Segil *et al.*, 1988).

The model of evolution can be derived from the number of synonymous and non-synonymous differences per nucleotide in the sequences of the analyzed genes. If we assume that gene conversion or interlocus recombination has taken place then we can expect that the number of synonymous differences within a species will be slightly higher than the number of nonsynonymous differences, regardless of the occurrence of purifying selection (Piontkivska *et al.*, 2002; Rooney *et al.*, 2002).

Purifying selection determines the frequency of particular codons encoding the same amino acid.

Codons that are not recognized by abundant isoacceptor tRNA are removed by purifying selection, because their participation in protein synthesis can be considered insignificant. This is true in the case of genes with high expression. In genes with an average level of expression the selection pressure is low and various codons specific for a given amino-acid participate in protein synthesis (Nei & Kumar, 2000). In the case of the “birth and death” model more synonymous differences are expected because of the divergence of genes due to silent nucleotide substitutions (Rooney *et al.*, 2002).

Which of the models reflects the evolution of enolase better? Calculations were conducted for the whole gene coding sequences of the three vertebrate enolase isoenzymes. The number of synonymous differences per nucleotide (P_s) is in all enolase isoforms higher than the number of nonsynonymous differences (P_n). Although P_s/P_n ratio is 1.95 it seems more probable that enolase genes evolved through “birth and death” evolution. Such a conclusion can be drawn from phylogenetic trees of amino-acid and nucleotide sequences, as well as from the fact of gene duplication event occurring in the past. From phylogenetic trees it is evident that particular isoenzymes form monophyletic branches not within the phyla but enzyme types, and hence on both trees separate branches for NSE, NNE and MSE enolases are present.

Tolerated substitutions in mammalian enolase

The studied enolase isoforms exhibit higher interspecies (orthologic variants) than intraspecies (paralogic variants) similarity. Analysis of the amino-acid sequences with the S.I.F.T. tool indicates non-uniform distribution of the tolerated substitutions. In the studied sequences of non-neuronal (NNE) and muscle-specific (MSE) enolase most of the tolerated substitutions are localized in the areas of the C-terminal domain. A uniform distribution of tolerated substitutions was found for γ -enolase (NSE). A molecule of NSE contains less positions in which the substitutions are possible – it suggests higher selective pressure of this isoform, but also that it is relatively new in comparison to other types of mammalian enolase. The active sites (His-157) and ligand binding sites (Asp-244, Glu-292, Asp-317) remain unchanged. Thiol groups may play a critical role in glycolytic enzymes reactivity and function (Banaś *et al.*, 1988). It is worth mentioning that the α and β isoforms contain five cysteine residues (119, 337, 339, 357, 389 and 399) and γ -enolase contains all those mentioned above plus one additional Cys-270. It might have important implications on structure and function of this form.

REFERENCES

- Babbit PC, Hasson MS, Wedekind JE, Palmer DR, Barrett WC, Reed GH, Rayment I, Ringe D, Kenyon GI, Gerlt JA (1996) The enolase superfamily: a general strategy for enzyme-catalyzed abstraction of the alpha-protons of carboxylic acids. *Biochemistry* **35**: 16489–16501.
- Banaś T, Gontero B, Drews VL Johnson SL, Marcus F, Kemp RG (1988) Reactivity of the thiol groups of *Escherichia coli* phosphofructo-1-kinase. *Biochim Biophys Acta* **957**: 178–184.
- Baranowski T, Wolna E (1975) Enolase from human muscle. *Methods Enzymol* **42**: 335–338.
- Baranowski T, Wolna E, Morawiecki A (1968) Purification and properties of crystalline 2-phospho-D-glycerate hydro-lyase from human muscle. *Eur J Biochem* **5**: 119–123.
- Hannaert V, Brinkmann H, Nowitzki U, Lee JA, Albert MA, Sensen CW, Gaasterland T, Muller M, Michaels P, Martin W (2000) Enolase from *Trypanosoma brucei* from the amitochondriate protist *Mastigamoeba balamuthi* and from the chloroplast and cytosol of *Euglena gracilis*: pieces in the evolutionary puzzle of the eukaryotic glycolytic pathway. *Mol Biol Evol* **17**: 989–1000.
- Kumar SK, Tamura I, Jakobsen B, Nei M (2001) MEGA2: Molecular evolutionary analysis software. *Bioinformatics* **17**: 1244–1245.
- Kustrzeba-Wójcicka I, Golczak M (2000) Enolase from *Candida albicans* – purification and characterization. *Comp Biochem Physiol* **126 B**: 109–120.
- Liao D (1999) Concerted evolution: molecular mechanism and biological implications. *Am J Hum Genet* **64**: 24–30.
- Nei M, Kumar S (2000) *Molecular Evolution and Phylogenetics*, pp 10–13, Oxford University Press Oxford.
- Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc Natl Acad Sci USA* **94**: 7799–7806.
- Ng P, Henikoff S (2001) Predicting deleterious amino acid substitutions. *Genome Res* **11**: 863–874.
- Nitter-Marszalska M, Kustrzeba-Wójcicka I, Pisarczyk-Bo-gacka E, Medrala W, Patkowski J (1998) Skin test with enolase in the diagnosis of fungal allergy. *Allergy* **53**: 30.
- Nowak K, Wolny M, Banaś T (1981) The complete amino acid sequence of human muscle glyceraldehyde 3-phosphate dehydrogenase. *FEBS Lett* **134**: 143–146.
- Pegg SC, Babbit PC (1999) Shotgun: getting more from sequence similarity searches. *Bioinformatics* **15**: 729–740.
- Pietkiewicz J, Kustrzeba-Wójcicka I, Wolna E (1983) Purification and properties of enolase from carp (*Cyprinus carpio*). Comparison with enolases from mammals' muscles and yeast. *Comp Biochem Physiol* **75B**: 693–698.
- Piontkivska H, Rooney A, Nei M (2002) Purifying selection and birth-and-death evolution in the histone H4 gene family. *Mol Biol Evol* **19**: 689–697.
- Rider CC, Taylor CB (1975) Enolase isoenzymes. II. Hybridization studies developmental and phylogenetic aspects. *Biochim Biophys Acta* **405**: 175–187.
- Rooney A, Piontkivska H, Nei M (2002) Molecular evolution of the nontandemly repeated genes of the histone H3 multigene family. *Mol Biol Evol* **19**: 68–75.
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425.
- Segil N, Shrutkowski A, Dworkin MB, Dworkin-Rastl E (1988) Enolase isoenzymes in adult and developing *Xenopus laevis* and characterization of a cloned enolase sequence. *Biochem J* **251**: 31–39.
- Stamm LV, Young NR (1997) Nucleotide sequence of the *Treponema pallidum* eno gene. *DNA Seq* **7**: 261–265.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**: 4876–4882.
- Tracy MR, Hedges SB (2000) Evolutionary history of the enolase gene family. *Gene* **259**: 129–138.
- Van Der Straeten D, Rodrigues-Pousada RA, Goodman HM, Van Montagu M (1991) Plant enolase: gene structure expression and evolution. *Plant Cell* **3**: 719–735.