Comparative analysis of human and mammalias genes of protein ortologs of cytochrome P450 2E1

D. R. Duplij, O. V. Maksymchuk, V. P. Duplij[1], N. A. Chaschin

Institute of Molecular Biology and Genetics
150 Academician Zabolotny Street, Kiev, Ukraine, 03143

[1]Institute of Cell Biology and Genetic Engineering
148 Academician Zabolotny Street, Kiev, Ukraine, 03680

duplijv@icbge.org.ua; duplijd@gmail.com

Aim. *To carry out a comparison analysis of nucleotide sequences for the cytochrome P450 2E1 protein ortologs genes and re- establish the connection of gene evolution with nucleotide con- tent.* ***Methods***. *In silico: BLAST, ClustalW, MEGA4, PHP.* ***Results***. *A general phylogeny of CYP2E1 genes is described. The most affinity is found for human translated sequences with Pan troglo- dytes. The transition C > T is the most often, it occurs in introns by 2.6 times more than in exons. The correlation of keto-amino skew of CYP2E1 genes and evolution age of species is stated.* ***Conclusions***. *The analysis carried out in the paper allows us to assume that a common ancestor of the CYP2E1 protein isoform lived before the divergence between rodent and Primates orders, i.e. 70 million years ago. The single nucleotide substitution is accumulated in int- rons during evolution.*

Keywords: cytochrome P450 2E1, CYP2E1, transition, phylogeny, nucleotide composition skew.

Introduction. The multigen family of cytochrome P450 is one of the most intensively studied issues, because it plays an important role in metabolism of endo- and exogenous substratum. The ethanol inducible isoform of P450 2E1 cytochrome (CYP2E1) is of a special interest [1].

The induction of *Cyp2E1* expression leads to an increase in the level of oxygen radicals generated by CYP2E1, which are able to initiate the NADPH-depended lipid peroxidation. Because of this the cell acid-base balance is hereupon violated, and the oxidative stress develops [2]. The interaction of the genetic and environmental factors results in the tissue-specific changes of synthesis and activity of

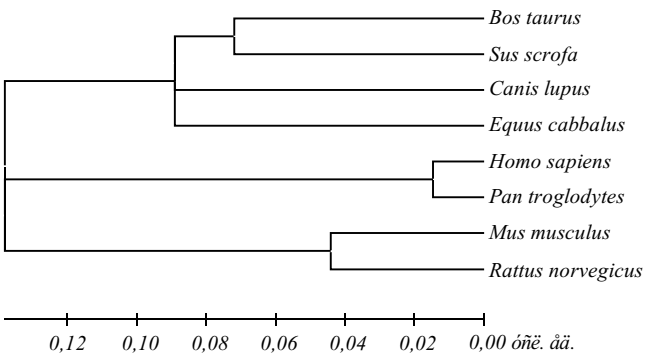

Fig. 1. Phylogenetic tree of the P450 2E1 cytochrome genes for human and seven mammals. Numerals mark evolutional distances in the reference units.

Cyp2E1, which brings in a certain contribution to the pathogeny of liver, pancreas diseases and other. [3, 4, 5].

There are some data about the impact of exogenous factors on expression of *Cyp2E1* [6], while the influence of genetic factors on expression of enzyme has not been studied enough. From the evolutional viewpoint, the cytochrome P450 is a unique protein [7], which survived from the primitive aerobes-prokaryotes to the human. This allows us to believe that most modern genes of cytochromes had a general predecessor, existed about 2 billions of years ago [8].

An opinion exists, that the primary evolutional function of cytochromes P450 in a cell was connected with its participation in the plastic and power metabolisms. The cytochrome P450 got the main specialization as xenobiotic biotransformator about 800 million years ago, which coincides with time of animals appearance [8].

In this paper, a comparative analysis of nucleotide sequences of mammalian and human genes of the Cyp2E1 orthology proteins is made for the restoration of evolutional history of the gene and mutational events.

Materials and methods The genomic sequences *CYP2E1* of human and seven mammals (*Pan troglodytes* XM_508139.2, *Bos taurus* NM_174530.2, *Canis lupus* NM_001003339.1, *Equus caballus* NM_001111303.1, *Mus musculus* NM_021282.2, *Rattus norvegicus* NM_031543.1, *Sus scrofa* NM_214421 [9]) were used in the investigation.

The initial alignment of nucleotide sequences was realized by BLASTN [10] and multiple alignment was executed by ClustalW [11]. The phylogenetic analysis was carried out by the MEGA4 packet [12]. The analysis of single nucleotide substitutions was fulfilled by using PHP computer language [13]. The value of keto-amino skew (K_{skew}) was calculated using the formula K_{skew}= (NG+NT -NA -NC)/L, where NA, NG, NT, NC are absolute quantities of corresponding nucleotides in a fragment of length (L = NA + NG + NT + NC) [14]. The arithmetical mean values of K_{skew} for exons and introns sequences were calculated, and the standard error of sample was found using the formula STD_ERR=RMSD/vN, where N is the size of a sample [15].

Results and discussions The structure of the *Cyp2E1* genes coincides for the majority of studied mammals. It consists of nine exons and eight introns. A unique variant of mRNA *Cyp2E1* is described for all species, in which the translated sequence of gene begins in the first exon and terminates in the last one. In spite of identical number of exons, the length of *Cyp2E1* differs for different species due to the varied introns length.

Thus, Homo sapiens has the longest sixth intron (due to the repetitive GGG sequence of 576 b.p.), and Mus musculus has the longest second intron. The structural likeness of the intron parts of *CYP2E* from Rattus norvegicus and Mus musculus is of special interest. They have around 86% similarity between the sequences of second and sixth introns, and 93% – between the fourth introns.

To reconstruct the evolutional history of the P450 2E1 cytochrome gene the multiple alignments of the translated sequences were executed and a phylogenetic tree was built (Fig.1).

It follows from Fig.1, that the evolution of *CYP2E1* gene undergoes a few divergences; in particular, there are independent ways of rodents, primates and other mammalian genes development. It is possible that the analyzed *CYP2E1* genes had a common ancestor before the division of rodents and primates that is about 70 millions years ago.

The pairwise alignments of the *CYP2E1* exonic and intronic sequences between each species and human were executed and frequencies of single nucleotide

Total frequencies of mutational events in gene CYP2E1

Name of species compared with *Homo sapiens*	Frequencies of single nucleotide substitutions				
	Transitions	Transversions	Total	Insertions	Deletions
Exons					
Pan troglodytes	0,014	0,011	0,025	0,067	0,003
Sus scrofa	0,111	0,092	0,203	0,035	0,000
Bos taurus	0,107	0,087	0,194	0,108	0,000
Canis upus	0,119	0,098	0,217	0,045	0,004
Equus cabbalus	0,110	0,081	0,192	0,045	0,035
Mus musculus	0,132	0,095	0,227	0,054	0,001
Rattus norvegicus	0,128	0,088	0,217	0,000	0,017
Introns					
Pan troglodytes	0,025	0,024	0,049	0,081	0,001
Sus scrofa	0,211	0,235	0,446	0,078	0,029
Bos taurus	0,200	0,196	0,396	0,017	0,143
Canis lupus	0,198	0,213	0,411	0,010	0,185
Equus cabbalus	0,208	0,223	0,431	0,037	0,048
Mus musculus	0,236	0,265	0,501	0,015	0,074
Rattus norvegicus	0,198	0,187	0,386	0,013	0,142

substitutions were analyzed in homologous sites to establish the nucleotide composition changes of the examined genes during their evolution.

Twelve possible types of single nucleotide substitutions were estimated. It should be noted that the substantial distinctions in frequencies of individual types of replacements in exonic and inronic parts of genes were found out. In particular, the A - G transitions were observed more frequent, than the T- C ones, and the G-A transitions were observed more frequent than the C-T ones. The final results are presented in Table. The substitutions are found to be more frequent in introns, than in exons. At the same time the transitions in exons are by 23% more frequent, than transversions.

Rather small predominance (about 5%). of transversions was observed in introns due to the transition G>C, which, possibly, related to the species features of CpG rich areas, where *Cyp2E1* is located. This fact is of special interest and requires a more thorough research.

The fixed distinctions between nucleotide ratios in different functional portions of genes were shown in our previous work [14]. Indeed, the high values of the keto-amine skew of nucleotide composition (K_{skew}) were registered in introns. This reflects predominance of the most frequent types of substituted nucleotides during the spontaneous mutagenesis. In order to check changes of K_{skew} value in evolution, we calculated the value K_{skew} for introns and exons of *Cyp2E1* in each species and compared it with early obtained data [14]. The changes in the K_{skew} value depending on the evolutional affinity of nucleotide sequences is presented in Fig.2.

Fig.2A shows that the introns K_{skew} values at evolution remote from *Homo sapiens* are negative due to the predominance of A and C. During evolution the K_{skew} value increases from -0,041±0,021 (*Mus*

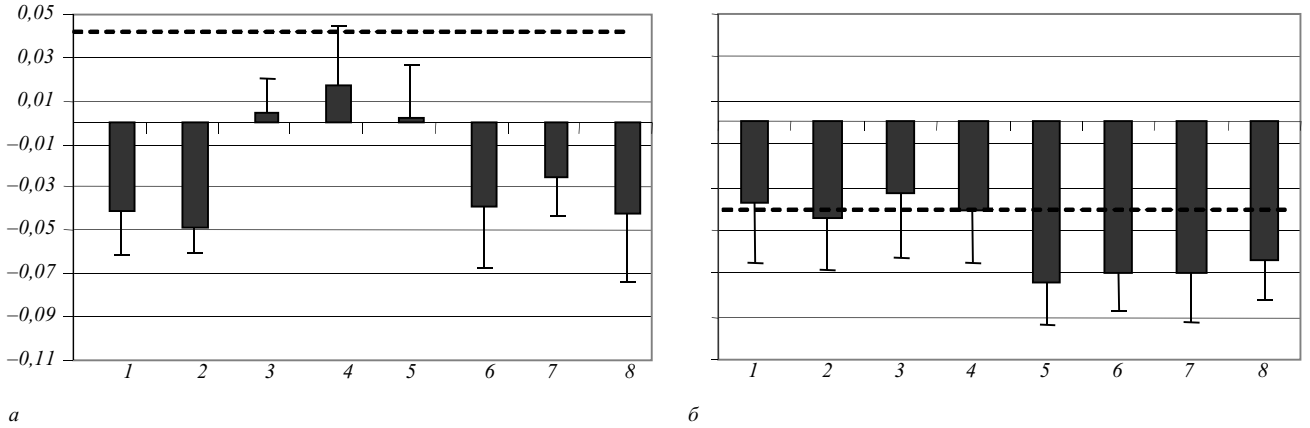

Fig. 2. Values of K_{skew} in introns (*a*) and exons (*b*) of the *CYP2E1* genes for eight mammals. The biological species are located on abscises in order of their affinity to the *CYP2E1* gene: *1 – Mus musculus*; *2 – Rattus norvegicus*; *3 – Pan troglodytes*; *4 – Homo sapiens*; *5 – Equus caballus*; *6 – Canis lupus*; *7 – Sus scrofa*; *8 – Bos taurus*. The proper averages of K_{skew} are indicated on ordinate for introns (*A*) and exons (*B*). The dotted line marks the average K_{skew} value for 10839 human genes [14].

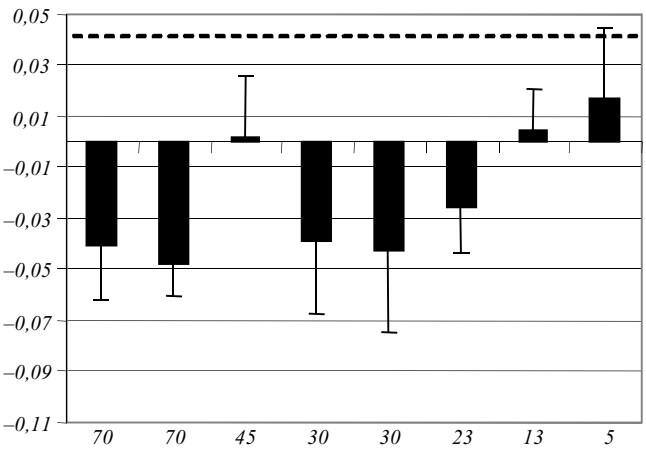

Fig. 3. Values K_s *CYP2E1* introns of eight mammalian. The divergention time (MY) of species is denoted at abscissa axis [16]: 70-*Mus musculus*, *Rattus norvegicus*; 45- *Equus caballus*; 30 -*Bos taurus*; *Canis lupus*; 23- *Sus scrofa*; 13-*Pan troglodytes*; 5-*Homo sapiens*. The corresponding mean K_s values with standard errors are depicted at ordinate axis.

musculus) and -0,042±0,032 (*Bos taurus*) to 0,004±0,016 (*Pan troglodytes*), and for *Homo sapiens* it arrives 0,017±0,027. On the contrary, all the exons K_{skew} values are negative (Fig.2B) and similar for eight mammalian species. This fact can be explained by the accumulation of thymine and guanine in the loci of the least selection pressure.

Thus, the K_{skew} value can be of particular interest for further research, because it can serve as an additional criterion for the estimation of nucleotide sequence evolutional age. The presented numerical analysis of the gene sequences of P450 2E1 cythochrome orthologous proteins allowed us to describe general regularities of the *Cyp2E1* phylogeny. We have shown that the frequencies of single nucleotide substitutions in the *Cyp2E1* introns are by 2.62 times higher in comparison with the exons .The most frequent mutational events are transitions which come to 48% of all single nucleotide substitutions.

Д. Р. Дуплий, О. В. Максимчук, В. П. Дуплий, Н. А. Чащин

Сравнительный анализ генов ортологичных белков цитохрома Р450 2Е1 человека и млекопитающих

Резюме

Цель. *Провести сравнительный анализ нуклеотидных последовательностей генов ортологичных белков цитохрома Р450 2Е1 и установить связь эволюции гена с нуклеотидным составом.* ***Методы***. *In silico: BLAST, ClustalW, MEGA4, PHP.* ***Результаты***. *Изучены особенности общей филогении генов CYP2E1. Наибольшее родство транслируемой последовательности гена CYP2E1 Homo sapiens обнаружено с Pan troglodytes. Выявлено, что транзиция С Т встречается в интронах в 2,6 раза чаще, чем в экзонах. Установлена связь между кето-аминовой асимметрией генов CYP2E1 и эволюционным возрастом вида.* ***Выводы***. *На примере CYP2E1 показано, что в течение эволюции интронный состав генов изменяется в сторону увеличения количества гуанина и тимина. Таким образом, величину кето-аминовой асимметрии можно использовать как дополнительный критерий эволюционного анализа.*

Ключевые слова: цитохром Р450 2Е1, CYP2E1, транзиции, филогения, асимметрии нуклеотидного состава.

Д. Р. Дуплій, О. В. Максимчук, В. П. Дуплій, М. О. Чащин

Порівняльний аналіз генів ортологічних білків цитохрому Р450 2Е1 людини і ссавців

Резюме

***Мета**. Провести порівняльний аналіз нуклеотидних послідовностей генів ортологічних білків цитохрому Р450 2Е1 і встановити зв'язок еволюції гена з нуклеотидним складом. **Методи**. In silico: BLAST, ClustalW, MEGA4, PHP. **Результати**. Описано особливості загальної філогенії генів CYP2E1. Найбільшу спорідненість трансльованої послідовності гена CYP2E1 Homo sapiens виявлено з Pan troglodytes. Визначено, що транзиція C T зустрічається в інтронах у 2,6 разу частіше, ніж в экзонах. Встановлено зв'язок між кето-аміновою асиметрією генів CYP2E1 і еволюційним віком виду. **Висновки**. На прикладі CYP2E1 показано, що впродовж еволюції нуклеотидний склад інтронів змінюється у напрямку збільшення кількості гуаніну і тиміну. Таким чином, величину кето-аміновой асиметрії можна використовувати як додатковий критерій еволюційного аналізу.*

Ключові слова: цитохром Р450 2Е1, CYP2E1, транзиції, філогенія, асиметрії нуклеотидного складу.

REFERENCES

1. *Chen Q., Galleano M., Cederbaum A. I.* Cytotoxicity and apoptosis produced by arachidonic acid in HepG2 cell overexpressing human cytochrome P4502E1 // J. Biol. Chem.–1997.–**272**, N 23.–P. 14532–14541.
2. *Wu D., Cederbaum A. I.* Oxidative stress mediated toxicity exerted by ethanol-inducible CYP2E1 // Toxicol. Appl. Pharmacol.–2005.–**207**, N 2 (suppl.).–P. 70–76.
3. *Botto F., Seree E., el Khyari S., de Sousa G., Massacrier A., Placidi M., Cau P., Pellet W., Rahmani R., Barra Y.* Tissue specific expression and methylation of the human *CYP2E1* gene // Biochem. Pharmacol.–1994.–**48**, N 6.–P. 1095–1103.
4. *Danko I. M., Chaschin N. A.* Association of CYP2E1 gene polymorphism with predisposition to cancer development // Exp. Oncol.–2005.–**27**, N 4.–P. 248–256.
5. *Sidorik L, Kyyamova R., Bobyk V., Kapustjan L., Rozhko O., Vigontina O., Ryabenko D., Danko I., Maksymchuk O., Kovalenko L., Chaschin N.* Molecular chaperon, HSP60, and cytochrome P450 2E1 co-expression in dilated cardiomyopathy // Cell Biol. Int.– 2005.–**29**, N 1.–P. 51–55.
6. *Maksymchuk O. V., Bobyk V. I. , Sydoryk L. L., Chashchyn M. O.* Influence of long-term combined gamma-radiation and ethanol on cytochrome P450 2E1 expression in the mice liver // Ukr. Biokhim. Zhur.–2008.–**80**, N 5.–P. 105–111.
7. *Danko I. M., Odynets K. A., Kitam V. O., Chaschin N. A.* Computer modeling of cytochrome P450 2E1 three-dimensional structure // Ukr. Biokhim. Zhur.–2006.–**78**, N 2.–P. 154–162.
8. *Nelso D. R. Strobel H. W.* Evolution of cytochrome P450 proteins // Mol. Biol. Evol.–1987.–**4**, N 6.–P. 572–593.
9. Web-resource: ftp://ftp.ncbi.nih.gov/genbank/genomes/
10. *Altschul S. F., Gish W., Miller W., Myers E. W., Lipman D. J.* Basic local alignment search tool. // J. Mol. Biol.–1990.–**215,** N3.–P. 403–410.
11. *Larkin M. A, Blackshields G., Brown N. P., Chenna R., McGettigan P. A., McWilliam H., Valentin F., Wallace I. M., Wilm A., Lopez R., Thompson J. D., Gibson T. J., Higgins D. G.* Clustal W and Clustal X version 2.0 // Bioinformatics.–2007.–**23**, N 21.–P. 2947–2948.
12. *Koterov D. V., Kostarev A. F.* PHP 5 v podlinnike.–Sankt-Petersburg: BHV, 2006.–1120 p.
13. *Tamura K., Dudley J., Nei M., Kumar S.* MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 // Mol. Biol. Evol.–2007.–**24**, N 8.–P. 1596–1599.
14. *Duplij D. R., Kalashnikov V. V., Chaschin N. A., Tolstorukov M. Y.* Comparative analysis of base-pair composition bias in exons and introns of human genes // Biopolym. cell.–2008.–**24**, N 5.–P. 1–9.

UDC 577.3
Received 06.04.10