Nucleic Acids Research, 1997, Vol. 25, No. 1 **143–146**

The α/β fold family of proteins database and the cholinesterase gene server ESTHER

Xavier Cousin, Thierry Hotelier[1], Kurt Giles[2], Philippe Lievin[1], Jean-Pierre Toutant and Arnaud Chatonnet*

Différenciation Cellulaire et Croissance and [1]Centre de Calcul, INRA-ENSAM, 2 Place Viala, 34060 Montpellier, France and [2]Department of Structural Biology, Weizmann Institute of Science, 76100 Rehovot, Israel

Received September 30, 1996; Accepted October 8, 1996

ABSTRACT

ESTHER (for esterases, α/β hydrolase enzyme and relatives) is a database of sequences phylogenetically related to cholinesterases. These sequences define a homogeneous group of enzymes (carboxylesterases, lipases and hormone-sensitive lipases) sharing a similar structure of a central β-sheet surrounded by α-helices. Among these proteins a wide range of functions can be found (hydrolases, adhesion molecules, hormone precursors). The purpose of ESTHER is to help comparison of structures and functions of members of the family. Since the last release, new features have been added to the server. A BLAST comparison tool allows sequence homology searches within the database sequences. New sections are available: kinetics and inhibitors of cholinesterases, fasciculin–acetylcholinesterase interaction and a gene structure review. The mutation analysis compilation has been improved with three-dimensional images. A mailing list has been created.

INTRODUCTION

Cholinesterases are phylogenetically related to proteins which have a similar α/β fold structure. It consists of a central, highly twisted, 8–11-stranded β-sheet in which most strands are parallel. This sheet is flanked on both sides by α-helices (2–5). The comparison of properties of these proteins is of great value in many fields of research. Thanks to accumulating crystallographic data, both the normal catalytic mechanism of ACh hydrolysis and the structure–activity relationships of toxic acetylcholinesterase (AChE) inhibitors can now be studied in detail (4–8). Results are relevant not only for antidotal therapy but also for therapeutic use of anticholinesterase agents in neurodegenerative disorders (9). The comparison is also of benefit to research on carboxylesterases involved in resistance of insects to insecticides (10). We have created a database of biological data related to this family of genes which derive from a common ancestor. These data are available on the ESTHER server at the universal resource locator (URL) http://www.ensam.inra.fr/cholinesterase/ . Methods used to build the database and the rationale for determining the limits of the family were described previously (1).

THE α/β FOLD FAMILY OF PROTEINS

The family can be subdivided in three subfamilies (C, L and H). Sequences in the C-family have retained the sequence signatures: [ED]-D-C-L-[YT]-[LIV]-[DNS]-[LIV]-[LIVFYW]-x-[PQR] and F-[GR]-G-x-x-x-x-[LIVM]-X-[LIV]-x-G-x-S-[STAG]-G where alternative possible amino acids are shown in brackets and x represents any amino acid. These two consensus sequences are designed ps00941 and ps00122 respectively in the prosite database. This consensus is also known as the Carboxylesterase_B pattern. Information on this group can be found on the ExPASy (http://expasy.hcuge.ch/) server with the prosite accession number PDOC00112. Sequences in the L-family have retained the sequence signature: [LIV]-x-[LIVFY]-[LIVST]-G-[HYWV]-S-x-G-[GSTAC] designed in prosite database as ps00120 and also known as the LIPASE_SER pattern. Information on this group can be found on the ExPASy server with the prosite accession number PDOC00110.

Hormone-sensitive lipases are phylogenetically related to the L-family but they do not have the full range of sequence signature. We kept these sequences in a separate group called the H-family. New prosite entries have been recently created for this group of sequences ps01173 LIPASE_GDXG_HIS and ps01174 LIPASE_GDXG_SER with the signatures [LIVMF]-[LIVMF]-x-[LIVMF]-H-G-G-[SAG]-[FY]-x-x-x-[STDN]-x-x-[ST]-H and [LIVM]-x-[LIVMF]-[SA]-G-D-S-[CA]-G-[GA]-x-L-[CA].

A tree illustrating the similarities of typical proteins within the families in the database is shown in Figure 1. In the C and L families, enzymes and non-enzyme proteins can be found. The figure has been made interactive, picking on dots will lead to tables of sequences.

RAW DATA

One entry corresponds to one gene in one species and contains a compilation of all data available for this gene: cDNA and protein sequences, gene structure, promoter region, links to appropriate accession numbers in other databases (NCBI-UID, GenBank,

* To whom correspondence should be addressed at: Différenciation cellulaire et Croissance, 2 Place Viala, 34060 Montpellier cedex 1, France. Tel: +33 467 61 28 14; Fax: +33 467 54 56 94; Email: chatonne@ensam.inra.fr

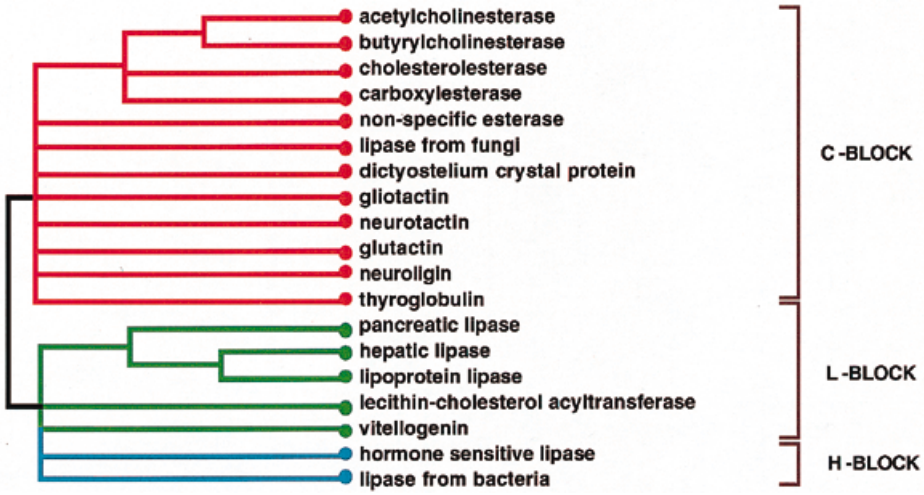

Figure 1. Schematic tree of α/β hydrolase enzymes and relatives. This tree gives a schematic representation of the relatedness of the subfamilies of proteins collected in ESTHER. The red, green and blue branches correspond respectively to the C-family, the L-family and H-family defined in text. On W3 browsers, picking the subfamilies or the dots of the tree will lead to the corresponding ESTHER tables and sequence alignments.

EMBL, PIR and Swiss-Prot) and references. The initial full table was getting too large to be easily downloaded. A new table containing only names of sequences with links to full data has been built.

The database contains now a total of 208 entries. New sequences are added regularly. There are 26 entries of cholinesterases; 19 contain a complete mature protein sequence. The entries can be subdivised in 141 sequences in the C-family, 53 in the L-family and 14 in the H-family. The genome sequencing programs allowed the characterization of a lot of sequences from model organisms. One can now find in the database 23 genes from *Caenorhabditis elegans*, 19 *Drosophila melanogaster*, 12 mouse and 21 human. These figures give only a small idea of the paralogous diversity acquired during evolution in this family as a large number of sequences has yet to be described in these organisms. In the C-family no sequence homologies was found with genome sequence of *Saccharomyces cerevisiae* nor with known complete genome of bacteria. This could mean a recent origin of the C-family.

NEW SECTIONS

Molecular biology

Molecular forms. Cholinesterases are well known for the large number of molecular forms that can be found in one organism. It arises from oligomerization and attachment to other proteins and/or glycolipids. This section (still in construction) is dedicated to the description of various molecular forms. A figure derived from Massoulié *et al*. (11) is available. Picking areas of the figure will lead to pages on mode of membrane attachment.

Gene structures. Figure 2 shows the comparison of gene structures in the cholinesterase group. The figure is associated with a Java applet, picking on individual exons leads directly to an alignment of the prototypic sequences with the sequence of the selected exon coloured.

Protein structures. For proteins in the family which has been crystallized so far, links to PDB files (Brookhaven Protein DataBase, http://www.pdb.bnl.gov) and to images are available in this section.

Phylogenetic analysis

Different alignments of groups of sequences from the database are available. Two pages allow a quick back and forth access from alignments to phylogenetic analysis. We used different subsets of the databases to recognize paralogous and orthologous genes within the database. The ALIGN page describes alignments and the PHYLO page contains links to dendrograms or phylogenetic trees deduced from these alignments. A small icon over each tree allows to see the alignment used for each analysis. A code is introduced to give the statistical significance of the branching of the tree given by bootstraping analysis. Alignments in nucleotides and proteins were built using CLUSTALW (12) or ALIGN (13). Parsimony methods on DNA or protein alignment or Maximum likelihood on DNA alignment were used according to the PHYLIP package (14). Trees were edited using the program TREETOOL (15). Some alignments were kindly provided by Gentry and Doctor (16).

BLAST server

BLAST performs fast database searching (17). Statistics for judging the significance of matches are also given by the program. BLAST searches can be performed with a sequence sent by a distant user against the full database or subsets: only cholinesterases, or the subset C-family, L-family or H-family. Options include choices of matrix, expected and cutoff scores. The output is given in hypertext and links are available to the entries in the ESTHER format.

Mutation analysis

This section presents a compilation of all mutations analyzed so far on cholinesterases. It corresponds to an updated version of the

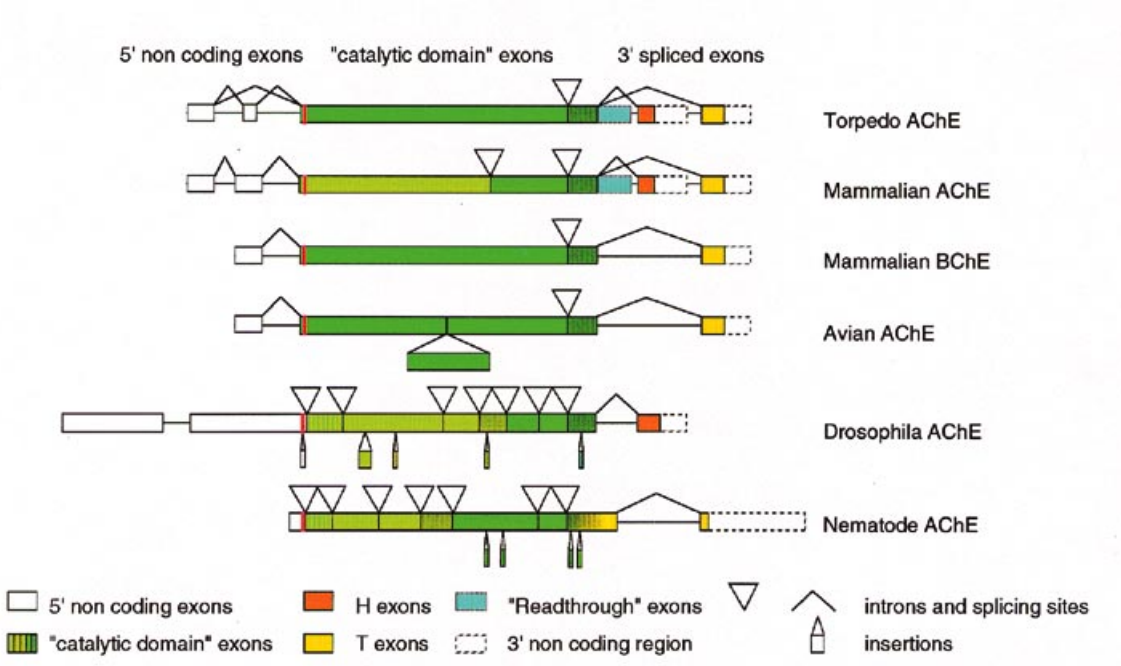

Figure 2. Comparison of gene structure in the cholinesterase subfamily of ESTHER. By convention the gene and protein of *Torpedo* AChE are taken as the prototype of the cholinesterase family (11). When insertions are necessary for keeping the alignment of homologous regions with the *Torpedo* sequence, it is indicated under the protein. This does not imply a genetic interpretation as deletions may have occured. Homologous exons have the same color in different species. On the server this figure is a clickable map: highlighted regions are links to alignments of proteins which show the homologies between cholinesterases. Introns are not to scale.

table from Shafferman *et al.* (18) and the hypertext alignment Mutalign (1). The advantages of the Mutalign representation is that it allows a rapid access to information on mutants created experimentally by various groups on different but related enzymes overpassing the scattering of the published information. The alignment allows a quick access and comparison of mutation data.

Effects of a point mutation on the properties of the enzyme depends on the position of the altered amino acid in the three-dimensional structure. Thus a rapid visual localization of a particular mutated amino acid is often necessary. A new feature of Mutalign is the possibility to recover interactive three-dimensional images of acetylcholinesterase showing the position of the particular mutated amino acid relative to the catalytic center. Selecting the highlighted symbol (!) placed over the alignment or on each line in the table will allow the recovery of a 32 kbites file in kinemage format (much smaller than the PDB files), which can be viewed and manipulated locally on the user's monitor. The *Torpedo marmorata* AChE structure was used to build these images. Only the backbone of the protein is shown coloured in gold. The side chains of the amino acids of the catalytic triad are coloured in cyan and those of the aromatic residues bording the gorge are in green. The side chains of the mutated amino acids are in red. There are 133 images stored this way corresponding to all mutations or multiple mutations visible on the structure. The entire image can be rotated in real time, parts of the display can be turned on or off, and any amino acid can be identified by picking it. MAGE is a display program to view and explore kinemages (19). Programs, full instructions and the proposed MIME standard are described and can be obtained from the Protein Science Gopher Space:

http://www.prosci.uci.edu/ kinemages/MageSoftware.html

Browsers correctly configured will open directly the kinemage file with the MAGE application when picking the highlighted (!). Possible extension of this feature of ESTHER will be files in RASMOL or Cn3d formats or Java scripts.

Kinetics and inhibition

This section is intended to be a repository of information useful in various domains of both theoretical and applied cholinesterase research.

On a fundamental point of view acetylcholinesterase is one of the most rapid enzymes with a speed close to that of a diffusion-controlled reaction (20). The position of the catalytic triad in a deep gorge has triggered a lot of work to elucidate the mechanism of action of the enzyme. Acetylcholinesterase possesses a peripheral binding site where some compounds can act as non-competitive inhibitors. These aspects have been studied using different classes of inhibitors, on wild-type or mutated enzymes and crystallography of the enzyme complexed with these inhibitors (6–8).

Inhibition of cholinesterases is also the focus of intense applied research. Many organophosphates and carbamates are used in agriculture for pest control. Resistance of insect population is a common phenomenon when these inhibitors are massively used. Both mutations in acetylcholinesterase (21) or gene amplifications in related esterases (10) are mainly responsible for this resistance. In medicine, anticholinesterase agents like tacrine are used in the treatment of Alzheimer's disease and can improve cognitive capacities of patients (9).

The diversity of uses and impacts of anticholinesterase chemicals explains the need for a comprehensive review. The compilation of substrates, inhibitors and reactivators of cholinesterases presented in this section of the database is intended to fullfil this need. Inhibitors are classified as natural inhibitors [e.g., onchidal (22), huperzine (23), fasciculin (26)], carbamates and organophosphates. Substrates include those of human butyrylcholinesterase used as myorelaxant (24) or bronchodilator prodrugs (25). Each compound has an entry with name, synonyms, chemical formula, a GIF image of structure and references.

Fasciculin and acetylcholinesterase inhibition

Fasciculins are toxins purified from the venom of snakes belonging to the genus *Dendroaspis* (family: *Elapidae*, the common name for *Dendroaspis* is mamba) (26).

The aim of this chapter is to present results which have been obtained in the characterization of fasciculins and mechanisms of action of these 'natural' acetylcholinesterase peripheral site ligands.

The resolution of the crystal structure of the complex fasciculin–AChE was realized by two differents groups using recombinant mouse (7) or *torpedo* (8) AChEs. Both studies confirmed the interaction of fasciculin with the peripheral site and indicated a very tight complex implicating hydrophobic interactions.

Mailing list

A mailing list has been created. In order to be included in the list a mail should be sent to: majordomo@wicc.weizmann.ac.il leaving the subject line blank, then writing 'subscribe cholinesterase' in the body of the message. Messages to the list should be sent to: cholinesterase@wicc.weizmann.ac.il. Withdrawing one's name from the list can be done by sending the message 'unsubscribe cholinesterase' to majordomo@wicc.weizmann.ac.il.

FUTURE DEVELOPMENT

The exponential growth of sequence data provided by genome sequencing projects or expressed sequence tags increases the need for annotated and specialized databases. This is the goal of ESTHER in the particular field of α/β fold family.

We wish to develop access to programs that enable users to manipulate sequences in the database together with their own sequences. Besides BLAST search programs, alignments and tree constructing programs will be proposed. We will use the new standards emerging in Java programs for production of three-dimensional images of protein structures. We wish to extend the relationship to other generalist sequence databases. The mailing list is a first step towards a more general forum of information on this α/β fold hydrolases family.

AVAILABILITY OF ESTHER AND CITATION

The world wide web (W3) URL is:
http://www.ensam.inra.fr/cholinesterase

When recovering sequences from this server do not acknowledge the server but the original author(s). If you extensively use ESTHER or documents from ESTHER, please cite this article.

ACKNOWLEDGEMENTS

We thank Jérome Gouzy, Daniel Kahn (PRODOM-I.N.R.A., Toulouse) and Meng Wong (University of Pennsylvania) for providing BLAST Perl scripts. We thank Yves Bourne, Palmer Taylor and Pascale Marchot (University of California, San Diego) for providing pictures of the mouse AChE–fasciculin complex. This work was supported by the Institut National de la Recherche Agronomique (I.N.R.A) and grants from the Association Française contre les Myopathies (A.F.M.).

REFERENCES

1 Cousin,X., Hotelier,T., Lievin,P., Toutant,J.-P. and Chatonnet,A. (1996) *Nucleic Acids Res.* **24**, 132–136.
2 Krejci,E., Duval,N., Chatonnet,A., Vincens,P. and Massoulié,J. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 6647–6651.
3 Shrag,J.D., Li,Y., Wu,S. and Cygler,M. (1991) *Nature* **351**, 761–765.
4 Sussman,J.L., Harel,M., Frolow,F., Oefner,C., Goldman,A., Toker,L. and Silman,I. (1991) *Science* **253**, 872–879.
5 Ollis,D.L., Cheah,E., Cygler,M., Dijkstra,B., Frolow,F., Franken,S.M., Harel,M., Remington,S.J., Silman,I., Shrag,J., Sussman,J.L., Verschueren,K.H.G. and Goldman,A. (1992) *Protein Eng.* **5,** 197–211.
6 Harel,M., Schalk,I., Ehret-Sabatier,L., Bouet,F., Goeldner,M., Hirth,C., Axelsen,P., Silman,I. and Sussman,J. (1993)*Proc. Natl. Acad. Sci. USA* **90** 9031.
7 Bourne,Y., Taylor,P. and Marchot,P. (1995) *Cell* **83**, 503–512.
8 Harel,M., Kleywegt,G.J., Ravelli,R.B.G., Silman,I. and Sussman,J. (1995) *Structure* **3**, 1355–1366.
9 Mesulam,M.-M. and Morán,A. (1987) *Ann. Neurol.* **22**, 223–228.
10 Smyth,K.-A., Russel,R.J. and Oakeshott,J.G. (1994) *Biochem. Genet.* **32**, 437–453.
11 Massoulié,J., Pezzementi,L., Bon,S., Krejci,E. and Vallette,F.-M. (1993) *Prog. Neurobiol.* **41**, 31–91.
12 Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) *Nucleic Acids Res.* **22**, 4670–4680.
13 Hein,J. (1990) *Methods Enzymol.* **183**, 626–645.
14 Felsenstein,J. (1985) *Evolution* **39**, 783–791.
15 Maciukenas,M. (1992) The Ribosomal Database Project. University of Illinois.
16 Gentry,M.K. and Doctor,B.P. (1995) In Quinn,D.M., Balasubramanian,A. S., Doctor,B.P. and Taylor,P. (eds) *Enzymes of the Cholinesterase Family.* Plenum Press, New York, pp. 493–505.
17 Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) *J. Mol. Biol.* **215**, 403–410.
18 Shafferman,A., Kronman,C. and Ordentlich,A (1995) In Quinn,D.M., Balasubramanian,A.S., Doctor,B.P. and Taylor,P. (eds) *Enzymes of the Cholinesterase Family.* Plenum Press, New York, pp. 480–488.
19 Richardson,D.C. and Richardson,J.S. (1992) *Protein Sci.* **1**, 3–9.
20 Quinn,D.M. (1987) *Chem. Rev.* **87**, 955–979.
21 Fournier,D. and Mutero,A. (1994) *Comp. Biochem. Physiol.* **108C**, 19–31.
22 Abramson,S.N., Radic,Z., Manker,D., Faulkner,D.J. and Taylor,P. (1989) *Mol. Pharm.* **36**, 349–354.
23 Ashani,Y., Grunwald,J., Kronman,C., Velan,B. and Shafferman,A. (1994) *Mol. Pharmacol.* **45**, 555–560.
24 Karlsson,E., Mbugua,D. and Rodriguez-Ithurralde,D. (1984) *J. Physiol. (Paris)* **79**, 232–240.
25 Lockridge,O. (1990) *Pharmacol. Ther.* **47**, 35–60.
26 Olsson,T. and Svensson,L.-Å. (1984) *Pharm. Res.* **1**, 19–23.