*132–136* *Nucleic Acids Research, 1996, Vol. 24, No. 1* 

# A cholinesterase genes server (ESTHER): a database of cholinesterase-related sequences for multiple alignments, phylogenetic relationships, mutations and structural data retrieval

**Xavier Cousin[1,2], Thierry Hotelier[3], Philippe Liévin[2], Jean-Pierre Toutant[2] and Arnaud Chatonnet[2,*]**

[1]Unité des Venins, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris cedex 15, France, [2]Différenciation cellulaire et Croissance and [3]Centre de Calcul, INRA, 2 place Viala, 34060 Montpellier, France

Received August 31, 1995; Revised and Accepted October 16, 1995

## ABSTRACT

**We have built a database of sequences phylogenetically related to cholinesterases (ESTHER for esterases, α/β hydrolase enzymes and relatives). These sequences define a homogeneous group of enzymes (carboxylesterases, lipases and hormone-sensitive lipases) with some related proteins devoid of enzymatic activity. The purpose of ESTHER is to help comparison and alignment of any new sequence appearing in the field, to favour mutation analysis of structure–function relationships and to allow structural data recovery. ESTHER is a World Wide Web server with the URL http://www.montpellier.inra.fr:70/cholinesterase.**

## INTRODUCTION

A number of previous results (1–3) have suggested that cholinesterases belong to a large family of phylogenetically related esterases and lipases which have representatives in all higher eukaryotes, as well as in numerous microorganisms. Comparison of three-dimensional structures of *Geotrichum* lipase (4), *Torpedo* acetylcholinesterase (5) and human pancreatic lipase (6) has shown that they share a similar arrangement of eight β-sheets connected by α-helices (the α/β hydrolase fold) that is common to several other hydrolytic enzymes (2) differing widely in phylogenetic origin.

We have created a database of sequences which are phylogenetically related to cholinesterases (ESTHER, for esterases, α/β hydrolase enzymes and relatives). It includes carboxylesterases, lipases (from vertebrates and microorganisms) and hormone-sensitive lipases (HSL; 3), as well as some related proteins lacking enzymatic activity, such as neurotactin, glutactin, thyroglobulin, vitellogenins (see 1,3), neuroligin (7) and gliotactin (8). At present (October 95) 166 non-redundant entries are available in ESTHER.

A preliminary version of the database was presented (as a CHE-DB gopher) at the Fifth International Meeting on Cholinesterases, Madras, India (9).

## CONSTRUCTION OF ESTHER

Within the α/β hydrolase fold family only carboxylesterases and lipases are obviously phylogenetically related. Other α/β hydrolases, such as carboxypeptidase II from wheat, dienelactone hydrolase from *Pseudomonas* and haloalkane dehalogenase from *Xanthobacter*, which are representatives of different subgroups (2), do not show sequence homology and were not included in ESTHER. A PROSITE signature exists for lipases (PS00120) and for esterases (PS00122) (10). Using BLOCKSEARCH (11) it appears that lipases possess five blocks of conserved sequences, PS00120A–E (the lipase pattern), and that carboxylesterases possess the four blocks PS00122A–D (the carboxylesterase B pattern). Full sequences of blocks PS00120 and PS00122, as well as their positions on three-dimensional images of *Torpedo* AChE and horse triglyceride lipase are given in ESTHER (see section I of the home page; Fig. 1). As described by Hemilä *et al.* (3), another group of related enzymes (HSL) contains some vertebrate as well as some bacterial lipases. We used four kinds of sequence on BLAST searching (12) generalist databases. These sequences were: a carboxylesterase; a vertebrate lipase; a HSL; an example of lipases from microorganism which possess the five blocks of PS00120 but are shorter than vertebrate lipases. The servers used to recover sequences were: the NCBI server for GenBank (http://www.ncbi.nlm.nih.gov); the EBI server for EMBL (http://www.ebi.ac.uk/); Expasy for SWISS-PROT (http://expasy.hcuge.ch/). For sequences selected by this method which did not appear obviously related a BLOCKSEARCH (11) was performed in addition to the BLAST analysis (12). This allowed rejection of marginal false positives.

## ESTHER SEQUENCES

A complete table of ESTHER sequences can be displayed by selecting 'ESTHER table' in the menu (Fig. 2A). Clicking on a given ESTHER name in this table gives all related information available in the database (Fig. 2B). Each entry is related to one of the three main groups defined above, carboxylesterases, lipases and HSL (C, L and H families respectively; see 'BLOCK' in Fig. 2B).

* To whom correspondence should be addressed

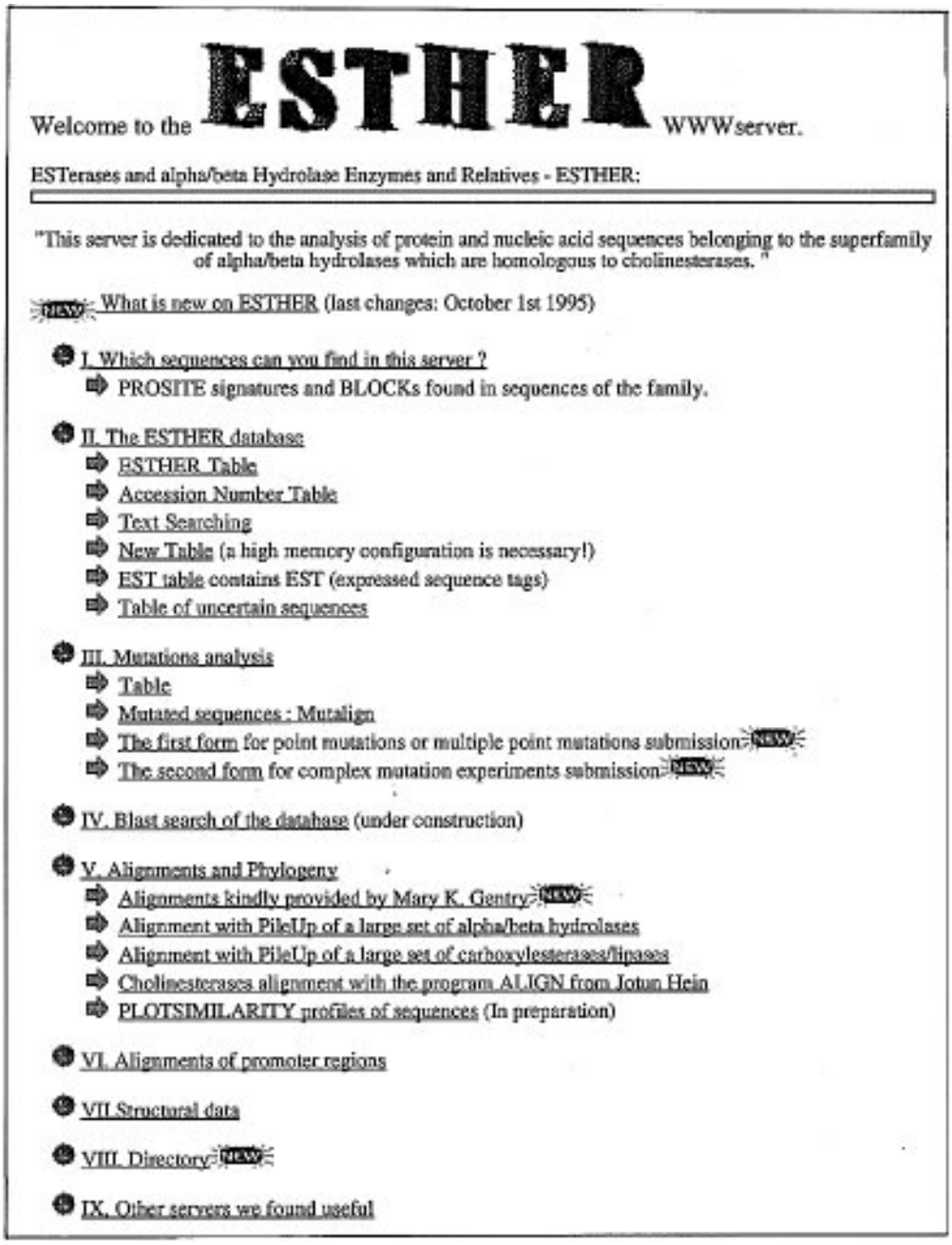


**Figure 1.** The home page of ESTHER. 'BLOCKS' gives sequences of PS00120 and PS00122 (lipase and carboxylesterase PROSITE signatures). The 'ESTHER table' opens the general table (see Fig. 2). Mutalign is illustrated in Figure 3. Underlined items are hypertext links to other parts of the database. See text for comments on all sections (I–IX).

Among carboxylesterases of type B (those carboxylesterases sensitive to organophosphate inhibition) are included invertebrate and vertebrate acetylcholinesterases, vertebrate butyrylcholinesterases and a number of esterases. Some non-enzymatic proteins, which are thus insensitive to organophosphate inhibition, are also included in this group (vertebrate thyroglobulins, *Drosophila* neurotactin, glutactin and gliotactin and rat neuroligin).

Among lipases are included pancreatic lipases (13), as well as triglyceride lipases (EC 3.1.1.3) and their tissue-specific isozymes in higher vertebrates (pancreatic, hepatic and gastric/lingual), lipoprotein lipases (EC 3.1.1.34), some prokaryotic lipases, lecithin-cholesterol acyltransferases (EC 2.3.1.43) and related non-enzymatic vitellogenins from flies.

The third subfamily (HSL) was defined by Hemilä *et al.* (3) by reference to a mammalian enzyme of this group. The sequence homology between HSL family members and the two other groups of ESTHER proteins is restricted to a 100 amino acid region upstream of the active serine (the HG-GESAG region; 3). Although this region contains only block D of PS00122 (equivalent to block B of PS00120), the corresponding sequences were included in

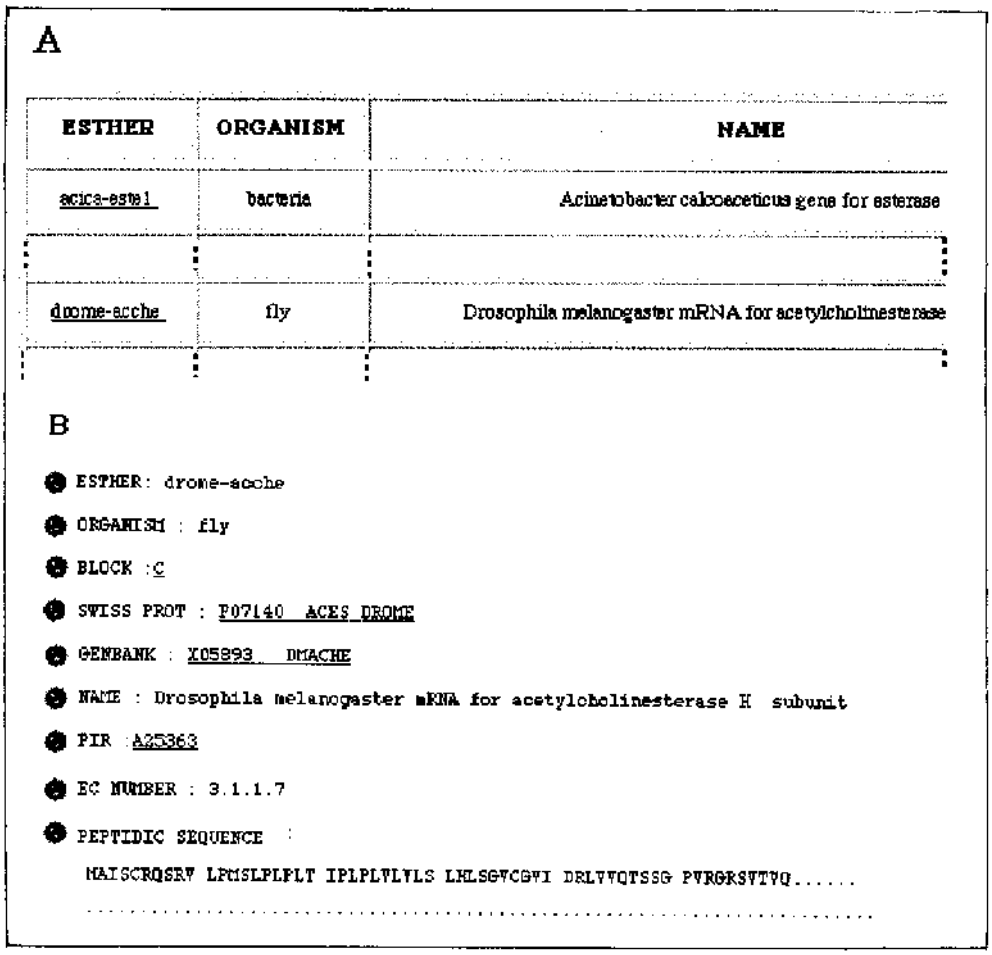


**Figure 2.** The ESTHER table (from home page section II). (**A**) Sequences are presented by alphabetical order of their ESTHER names. This name is built of 11 characters: five characters for the organism (three for genus and two for species, except when a five character name exists, such as human, chick, mouse), followed by a hyphen and five characters for the name of the protein. The correspondence between the abbreviation used in ESTHER and the trivial name is indicated. (**B**) As an example we have selected a particular entry (*Drosophila melanogaster* acetylcholinesterase, drome-acche) by clicking on its ESTHER name. Information available includes the subfamily (C, H or L in 'BLOCK') and the EC number, as well as its accession numbers in all other data banks where it appears. These accession numbers are hypertext links to the corresponding WWW servers and allow access to further details including reference(s) to the original work(s).

ESTHER under the 'H' family. The H family includes vertebrate HSL, some bacterial lipases and some invertebrate lipases.

At present 166 (non-redundant) sequences are included in ESTHER. To update our collection of sequences a BLAST search (similar to that used above) is made for each update of GenBank/EMBL or SWISS-PROT. A form for submission of new sequences is available in this section.

## LIMITS OF ESTHER COLLECTION OF SEQUENCES

A growing number of 'expressed sequence tags' (EST) from different organisms (human, *Caenorhabditis elegans* and *Arabidopsis thaliana*) have been released in generalist databases. The presence of all blocks or complete alignments as defined above cannot be tested on these partial sequences. In addition, these sequences have been sequenced only once and shifts in the reading frame are often necessary for translation and alignment. These EST were not included in ESTHER, but those presenting homologies with sequences already included in ESTHER are listed in a special table, 'EST of interest'. Other partial sequences were usually not included in ESTHER, except for a few when the method used for cloning allowed their formal identification as obvious orthologues (see for example BChE from *Canis* and *Ovis*; 14). A 'table of uncertain sequences' is also available in this directory. It contains sequences with low BLAST scores possessing less than two blocks of either PS00120 or PS00122.

## PRINCIPAL FEATURES OF ESTHER

The home page of Figure 1 summarizes what can be performed with ESTHER.

### ESTHER collection of sequences (home page section II)

In the 'ESTHER table' one can find the list of all sequences, the name of the protein and of the organism concerned (Fig. 2A). For each entry the sequence itself is available and the correspondence between accession numbers in different databases is indicated (Fig. 2B). These accession numbers are connected with their original databases and this allows access to the original reference(s).

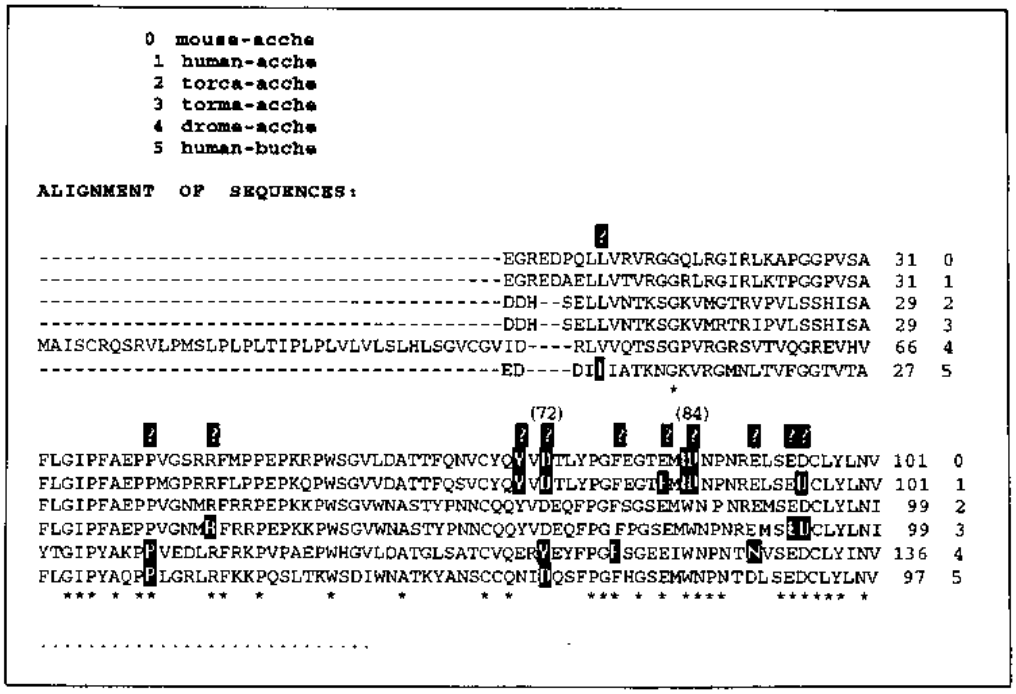


**Figure 3.** The N-terminal portion of Mutalign. Only the N-terminal portions of sequences are shown for clarity. 0–4 are mouse, human, *T.californica*, *T.marmorata* and *Drosophila* AChEs. 5 is human BChE. The numbering used on each line corresponds to that used in the original work. The conventional numbering is that of *Torpedo* AChE (see 26). Residues subjected to directed mutagenesis are shown on a dark field. In the region shown D72, considered as a component of the peripheral site, and W84, a critical side chain of the anionic site in AChEs, were particularly studied. Selecting an individual highlighted residue or question mark gives references to original works and a summary of the effects observed. Natural variants (see for example D72 in human BChE), as well as mutations induced by EMS treatment in *Drosophila*, are also listed. Multiple alignment was performed using the ALIGN program (19).

## Mutation analysis (section III)

This section presents a compilation of all mutations analysed so far in cholinesterases. An updated version of the table from Shafferman *et al.* (15) is available. Natural mutants or variants (*Drosophila* AChE and human BChE), results of site-directed mutagenesis and random generated mutants (by EMS treatment of *Drosophila* or *C.elegans*; see 16,17), as well as computer model analyses, are indicated as such. We have also built a hypertext alignment (Mutalign) of AChEs from mouse, human, *Torpedo californica*, *T.marmorata* and *Drosophila* and human BChE (Fig. 3). The positions of mutated residues are indicated by a question mark above the sequences and a highlighted residue in the sequence itself. Selecting highlighted residues or question marks leads to all references concerned with a summary of the effects observed. Natural mutants or variants (*Drosophila* AChE and human BChE) are also indicated. Figure 3 shows Mutalign for a short N-terminal portion of the sequences. Forms for submission of mutation experiments and references to be included in the table or in Mutalign are also available in this section.

## Alignment and phylogeny (section V)

A FASTA format text file of all sequences can be easily recovered for local comparison and alignment programs.

Under alignment/phylogeny we also present several examples of phylogenetic trees built from ESTHER sequences. Orthologues and paralogues were studied separately. For example, a phylogenetic tree of five paralogous genes (resulting from a genomic duplication) in the esterase family and the lipase family respectively in human and rat is presented. Also installed in the server are several other alignment/phylogenetic trees for orthologous genes (the same gene in different species). This is presented for all acetylcholinesterases, for all vertebrate cholinesterases (acetyl and butyrylcholinesterases) or for a mixed set of data (esterases, lipases and HSL).

An example of multiple alignments of cholinesterases, esterases, lipases and related proteins, built independently by M.K.Gentry and Dr B.P.Doctor (18), is also available in this section.

The programs used for our alignments were either ALIGN (19) or CLUSTALW (20) or the PILEUP program of the GCG package (21). To infer the corresponding phylogeny a protein parsimony program (PROTPARS) was used on amino acid alignments. In addition, an alignment of the corresponding nucleotides was obtained using the MAKEINF program. The third, less informative, base of each codon was not taken into account when the DNAPARS (DNA parsimony) or DNAML (DNA maximum likelihood) programs were used to produce phylogenetic trees. To test the statistical significance of the topology of trees (branching order and nodes) aligments were resampled using the bootstrap method (using the SEQBOOT program, one hundred resamplings). The PROTPARS, MAKEINF, DNAPARS, DNAML and SEQBOOT programs are parts of the PHYLIP package (22). Trees were edited using the program TREETOOL (23).

### Retrieval of structural data (section VII)

In this directory three-dimensional images of AChE and certain lipases are available. Structural data were recovered from the Brookhaven Protein Data Base (PDB, http://www.pdb.bnl.gov/). One can also recover PDB files with coordinates and use local programs to visualize the protein structure under certain conditions (for example AChE in the presence of inhibitors). Kinemages illustrations using interactive graphics can be displayed with the client MAGE program (24).

### Other servers (section IX)

Links to servers from other laboratories working in the field are available.

## FUTURE ADDITIONS AND DEVELOPMENT

ESTHER is still under construction. In particular, we are preparing a BLAST search of the database (home page section IV). It will soon be possible to open a BLAST page (built with the help of H.Recipon from NCBI) allowing a comparison of any personal sequence with those in the database using the BLAST program (12). The collection of sequences used for this comparison can be all the non-redundant ESTHER sequences, all cholinesterases, all carboxylesterases or all lipases. The search can be performed with nucleotide or amino acid sequences. In section V we will soon include results obtained with the PLOTSIMILARITY program of the GCG package (21). This will display a diagram of conservation along aligned sequences of esterases and/or lipases. The positions of the different blocks of PS00120 and PS00122 will be indicated.

We are also working on a new directory of promoter regions of cholinesterases (section VI). At present only alignments of 5′ regions of human and rabbit BChEs are available (25). Also in progress is a comparison of genomic structures in cholinesterases and esterases. Other programs will be proposed on-line to be used with ESTHER data and distant users data (PHYLIP and CLUSTALW). We hope also to present, in the near future, structural data concerning the electric fields in the catalytic gorge of cholinesterases.

A directory of ESTHER users (section VIII) can be completed by giving an individual mailing address on a specific form. A newsgroup and a diffusion list will be proposed for free discussion in the field. A form for direct email communication is also proposed for submission of new sequences, new mutants, comments and suggestions.

## AVAILABILITY OF ESTHER AND CITATION

The World Wide Web URL is
http://www.montpellier.inra.fr:70/cholinesterase.
When recovering sequences from this server do not acknowledge the server, but the original author(s). Original references can be found through the general table.

## ACKNOWLEDGEMENTS

We thank Avigdor Shafferman and his colleagues (Ness-Ziona, Israel) for communication of the mutagenenis compilation on cholinesterases, Mary K.Gentry and Bhupendra Doctor (Walter Reed Army Institute of Research, Washington, DC) for their multiple alignment and Kurt Giles (Weizmann Institute, Israel) for useful comments and corrections. Hervé Recipon (NCBI, Bethesda, MD) is acknowledged for very useful advice. This work was supported by grants from the Association Franaise contre les Myopathies and Institut National de la Recherche Agronomique.

## REFERENCES

1 Krejci,E., Duval,N., Chatonnet,A., Vincens,P. and Massoulié,J. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 6647–6651.
2 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 Engng*, **5**, 197–211.
3 Hemilä,H., Koivula,T.T. and Palva,I. (1994) *Biochim. Biophys. Acta*, **1210**, 249–253.
4 Shrag,J.D., Li,Y., Wu,S. and Cygler,M. (1991) *Nature* 351, 761–765.
5 Sussman,J.L., Harel,M., Frolow,F., Oefner,C., Goldman,A., Toker,L. and Silman,I. (1991) *Science*, **253**, 872–879.
6 Winkler,F.K., D'Arcy,A. and Hunziker,W. (1990) *Nature*, **343**, 771–774.
7 Ichtchenko,K., Hata,Y., Nguyen,T., Ullrich,B., Missler,M., Moomaw,C. and Südhof,T.C. (1995) *Cell*, **81**, 435–443.
8 Auld,V.J., Fetter,R.D., Broadie,K. and Goodman,C.S. (1995) *Cell.*,**81**, 757–767.
9 Cousin,X., Hotelier,T., Mazzoni,C., Arpagaus,M., Toutant,J.-P. and Chatonnet,A. (1995) In Quinn,D.M. *et al.* (eds), *Enzymes of the Cholinesterase Family.* Plenum Press, New York, NY, in press.
10 Bairoch,A. and Boeckmann,B. (1994) *Nucleic Acids Res.*, **22**, 3578–3580.
11 Henikoff,S. and Henikoff,J.G. (1994) *Genomics*, **19**, 97–107.
12 Altschul,S.F., Gish,W., Miller,W., Myers,E.W., Lipman,D.J. (1990) *J. Mol. Biol.*, **215**, 403–410.
13 Schrag,J.D., Winkler,F.K. and Cygler,M. (1992) *J. Biol. Chem.*, **267**, 4300–4303.
14 Arpagaus,M., Chatonnet,A., Masson,P., Newton,M., Vaughan,T.A., La Du,B.N. and Lockridge,O. (1991) *J. Biol. Chem.*, **266**, 6966–6974.
15 Shafferman,A., Kronman,C. and Ordentlich,A (1995) In Quinn,D.M. *et al.* (eds), *Enzymes of the Cholinesterase Family.* Plenum Press, New York, NY, in press.
16 Mutero,A., Bride,J.-M., Pravalorio,M. and Fournier,D. (1994) *Mol. Gen. Genet.*, **243**, 699–705.
17 Talesa,V., Culetto,E., Schirru,N., Bernardi,H., Fedon,Y., Toutant,J.-P. and Arpagaus,M. (1995) *FEBS Lett.*, **357**, 265–268.
18 Gentry,M.K. and Doctor,B.P. (1995) In Quinn,D.M. *et al.* (eds), *Enzymes of the Cholinesterase Family.* Plenum Press, New York, NY, in press.
19 Hein,J. (1990) *Methods Enzymol.*, **183**, 626–645.
20 Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) *Nucleic Acids Res.*, **22**, 4670–4680.
21 Genetics Computer Group (1991) *Program Manual for the GCG Package, Version 7.* Computer Graphics Group, Madison, WI.
22 Felsenstein,J. (1985) *Evolution*, **39**, 783–791.
23 Maciukenas,M. (1992) The Ribosomal Database Project, University of Illinois.
24 Richardson,D.C. and Richardson,J.S. (1992) *Protein Sci.*, **1**, 3–9.
25 Jbilo,O., Toutant,J.-P., Vatsis,K.P., Chatonnet,A. and Lockridge,O. (1994) *J. Biol. Chem.*, **269**, 20829–20837.
26 Massoulié,J., Pezzementi,L., Bon,S., Krejci,E. and Vallette,F.-M. (1993) *Prog. Neurobiol.*, **41**, 31–91.