

**Arnaud Chatonnet and  
Xavier Cousin**  
are at the Unité Différenciation  
Cellulaire et Croissance,  
Institut National de la  
Recherche Agronomique,  
Montpellier, France.

**Alan Robinson**  
is at the EMBL European  
Bioinformatics Institute,  
Hinxton, Cambridge, UK.

**Keywords:** *alpha/beta-  
hydrolase fold,  
acetylcholinesterase, kinetics,  
inhibitors*

Dr Arnaud Chatonnet,  
Département de Physiologie  
Animale, Institut National de la  
Recherche Agronomique,  
Place Viola,  
Montpellier Cedex 34060,  
France

Tel: +(33) 04 9961 26 77  
Fax: +(33) 04 6754 56 94  
E-mail: chatonnet@ensam.inra.fr

# Links between kinetic data and sequences in the alpha/beta-hydrolases fold database

Arnaud Chatonnet, Xavier Cousin and Alan Robinson

Date received (in revised form): 6th November 2000

## Abstract

While the number of sequenced genes is increasing dramatically, the number of different protein structural families is expected to be more limited. Changes in enzymatic activity or protein interactions can dramatically modify the role of homologous proteins in different organisms or mutants. However, experimental data associated with sequences or mutations stored in databases are often limited to a short description of the enzymatic pathway, molecular interaction or phenotype associated with the changes in amino acid sequence. In the alpha/beta-hydrolases fold database ESTHER, we are experimenting with links between experimental kinetic data and sequences, mutations and protein structures. This effort will lead to the integration of pharmacological data with genome-wide databases.

## INTRODUCTION

With the advent of rapid genome sequencing, it has become apparent that the number of genes estimated for single organisms had been overestimated previously. Furthermore, it came as a surprise that the variation in the number of genes between multicellular model organisms is much less than expected. Of course the prediction of genes from genomic sequences is still in its infancy and many different proteins can result from alternative splicing. But although the number of sequences in databases will grow exponentially owing to the investigation of new model organisms, parasite mutants and variants of all kinds, there will not be a similar increase in the number of new sequences that are not related to a previously known family. Similarly, structural proteomics is expected to provide new structural folds mainly for proteins that are difficult to crystallise, such as membrane bound proteins. A great part of the observed diversity of chemical reactions, cellular specialisations and physiological adaptations are due to subtle differences in proteins, masked by the overall homology of their sequences or structures.

Gene and protein interactions give rise to a combinatorial complexity that is only now being investigated on a larger scale. The first level is the genetic interactions as have been developed in the model organisms *Drosophila* and *Caenorhabditis*, but also in humans, farm animals and plants. These may be further studied mechanistically at the protein level. One can distinguish direct protein–protein interactions from more distant relationships, for example in the case of transcription factors and their target genes. The links can be even more distant in the case of hormonal neurotransmission regulation or in metabolic pathways. For enzyme functions, the critical factors are kinetic parameters as well as the concentration of substrates and regulators. Neither the enzyme concentration nor the level of gene expression necessarily affects the actual physiological condition.

Thus the challenge is to manage the flow of information coming from the functional analysis of isolated proteins and from the large-scale analysis of gene expression. Databases have to develop easy-to-use and reliable tools for these new types of functional annotation. Comprehensive databases and servers such

as PubMed Central or E-Biomed, which will give direct free access to experimental data, are an incentive to search for new ways of presenting biochemical data associated with gene sequences.

Since 1994, ESTHER, the alpha/beta hydrolase fold database,<sup>1-4</sup> has provided a server dedicated to presenting a wide variety of information related to this ever-growing family of proteins.<sup>5-8</sup> In this family, most proteins are enzymes for which a large number of natural and synthetic substrates or inhibitors are known. Most of the time, pharmacological data for these proteins are not available from sequence or gene databases.

In this paper we summarise some of the problems in the annotation of enzyme function in databases. Subsequently the paper aims to provide an overview of the first attempts made in the ESTHER database to present enzyme kinetics data linked to mutation analysis in cholinesterases. Finally, future directions and expected limits are discussed.

## DIFFICULTIES IN THE ANNOTATION OF ENZYME FUNCTION

### Large-scale analysis

Large-scale analysis of protein interactions or gene expression will not give all the clues necessary for deciphering the function of proteins of neurohumoral systems or enzymes. For example, the hypoglycaemic role of insulin *in vivo* would be difficult to deduce from the sequence homology of insulin with insulin-like growth factors. Furthermore, other knowledge such as the expression of the insulin gene in the pancreas, the protein's excretion in plasma and the ubiquitous interaction of insulin with receptors in tissues would also be of little help for the inference of the function of this hormone without physiological measurements.

In our case, for an enzyme such as acetylcholinesterase (AChE), protein interactions are important for the localisation of the enzyme but not for its

activity. Among these protein interactions, ColQ of the collagen tail anchors the enzyme in the basal lamina at the neuromuscular junction, the P subunit binds the enzyme to the plasma membrane of neurones in the brain, and the proteins of the post-translational machinery add a glycolipid to the C terminus of the enzyme anchor in blood cells.<sup>9</sup> All these interactions are important for positioning the enzyme outside the cells, but are not involved directly in its function of acetylcholine hydrolysis. On the contrary, no direct protein interaction is seen between acetylcholinesterase and the acetylcholine receptors or cholinacetyltransferase, which provide the base of the functional cholinergic system. The expression profile is also misleading as acetylcholinesterase is often found in the same neurons as cholinacetyltransferase, leading to confusion of cholinergic and cholinceptive neurons. Thus, how can a database present experimental data relevant for understanding the control of acetylcholine concentration by cholinesterases?

Proteins that do not participate in complex interactions give rise to other annotation problems including that the genes coding for these proteins are more prone to horizontal gene transfer between organisms.<sup>10</sup> Through such transfers, organisms not only may gain new functions, but also the newly acquired gene can evolve to participate in new pathways.

### Proteins in the alpha/beta-hydrolase fold superfamily with very homologous 3D structures perform a wide variety of functions

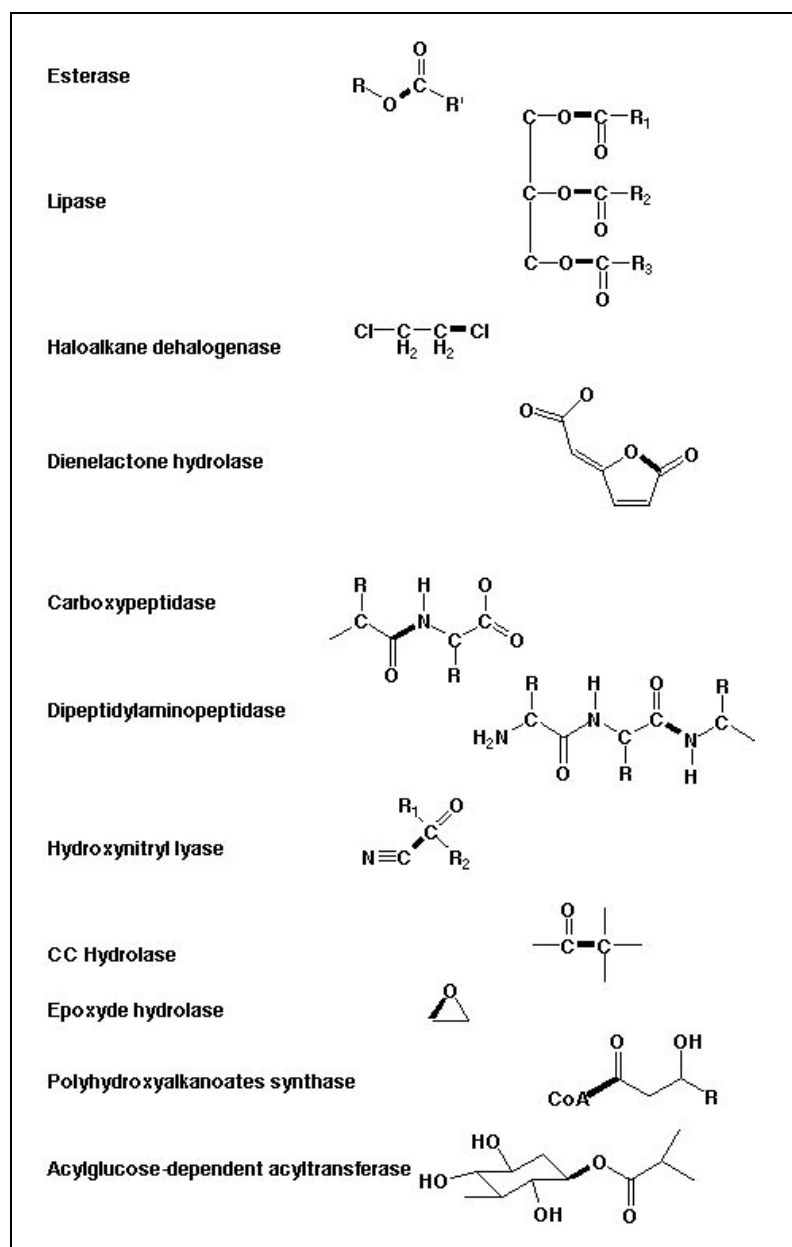
When the alpha/beta-hydrolase fold was first defined and described, it was evident that this structure is widely represented in nature and that it serves to fulfil a great variety of enzymatic reactions.<sup>5</sup> The panel of reactions in which it is observed to participate has since then increased (see

**long range interactions impede from annotating enzyme function precisely**

**enzyme function cannot always be inferred from sequence or structure**

Figure 1).<sup>6-8</sup> The enzyme reaction can be specific for a substrate found in the environment, eg xylane, cutine, lipids or proteins in the case of microorganism enzymes. The reaction can be less specific for extracellular or secreted enzymes in the case of eukaryotes, eg lipases or carboxylesterases. For those proteins that do not belong to a housekeeping

pathway, true orthology in other taxonomic orders will be rare and comparison of sequences between organisms will be less informative. Another aspect of this variety is that a reaction performed by an alpha/beta-hydrolase fold protein may also be performed in another organism by a structurally unrelated enzyme. This has been shown to be the case for haloacetate dehalogenase, a member of the alpha/beta-hydrolase fold superfamily.<sup>11</sup>



**Figure 1:** Alpha/beta-hydrolase fold enzyme subfamilies and their substrates. The attacked bond is shown as a thick line. Note that the final two enzymes participate in a synthesis reaction

### Proteins in the alpha/beta-hydrolase fold superfamily with very homologous sequences may have completely different functions

It has been shown that apparent homologues of acetylcholinesterase have cell adhesion properties, but completely lack enzymatic activity. Among these, glutactin neurotactin and neurologins have all the landmarks of the alpha/beta-hydrolase fold, but not the activity. A model of the structure of neurotactin even shows what resembles an active site gorge, but without the active site amino acid necessary for esterase activity. There is not a clear motif for adhesion and another one for enzymatic activity; some may even have both properties. Without experimental testing, the only way to identify those proteins without enzymatic activity is the absence of key residues such as the active serine in a binding site motif. Many proteins have been wrongly classified as esterases in the annotated genomes of *C. elegans* and *Drosophila*, since the annotation of such genes requires expertise that cannot be expected necessarily from the large-scale sequencing groups.

We try to bring some solutions to these difficulties with a specialised database of alpha/beta-hydrolases fold named ESTHER. The restricted scope of the database allows the curators to focus on documenting the information from experimental data.

new tools to compare effects of mutations

## FUNCTIONAL (PHARMACOLOGICAL) DATA IN ESTHER

When a new sequence is identified that is very similar to an esterase or an alpha/beta-hydrolase fold, and that the aforementioned pitfalls have been dealt with, it is very useful to know what previously well-documented homologous proteins do. A few very valuable general databases provide some information on enzymatic reactions, pathways and ligands in a comprehensive way, eg the ENZYME database,<sup>12</sup> the KEGG database,<sup>13</sup> the LIGAND database,<sup>14</sup> the BRENDA server<sup>15</sup> the EMP/MPW databases<sup>16</sup> and the Protein Mutant Database.<sup>17</sup> The information found in these databases can be extended to a higher degree of precision and include more sophisticated links for structure/function analysis if a new database is dedicated to only a limited number of genes and proteins, as is done with the ESTHER database.

### Structure of the ESTHER database

The database management system used for ESTHER is a well-known system used for genomic data ACeDB.<sup>18</sup> The interface of the database with the Internet is the webserver Perlace.<sup>19</sup> A general presentation of the database has been made previously.<sup>4</sup>

Sequence and structure information in ESTHER is derived from the primary databases of GenBank, EBI and PDB, and direct links back to the original entries in these databases are available. Additionally, a collection of chemicals identified as substrates or inhibitors of this family of enzymes has been introduced. The mutations and pharmacological data are extracted from the literature, and may now be submitted directly through forms on the World-Wide Web.

### Linking mutation analysis to function

To date, at least 400 mutations have been studied in the cholinesterase family. Only

a database can provide an efficient way to compare and analyse the results of these studies. In ESTHER we use different tools to present the mutation analysis of cholinesterases and provide links to enzyme kinetics.

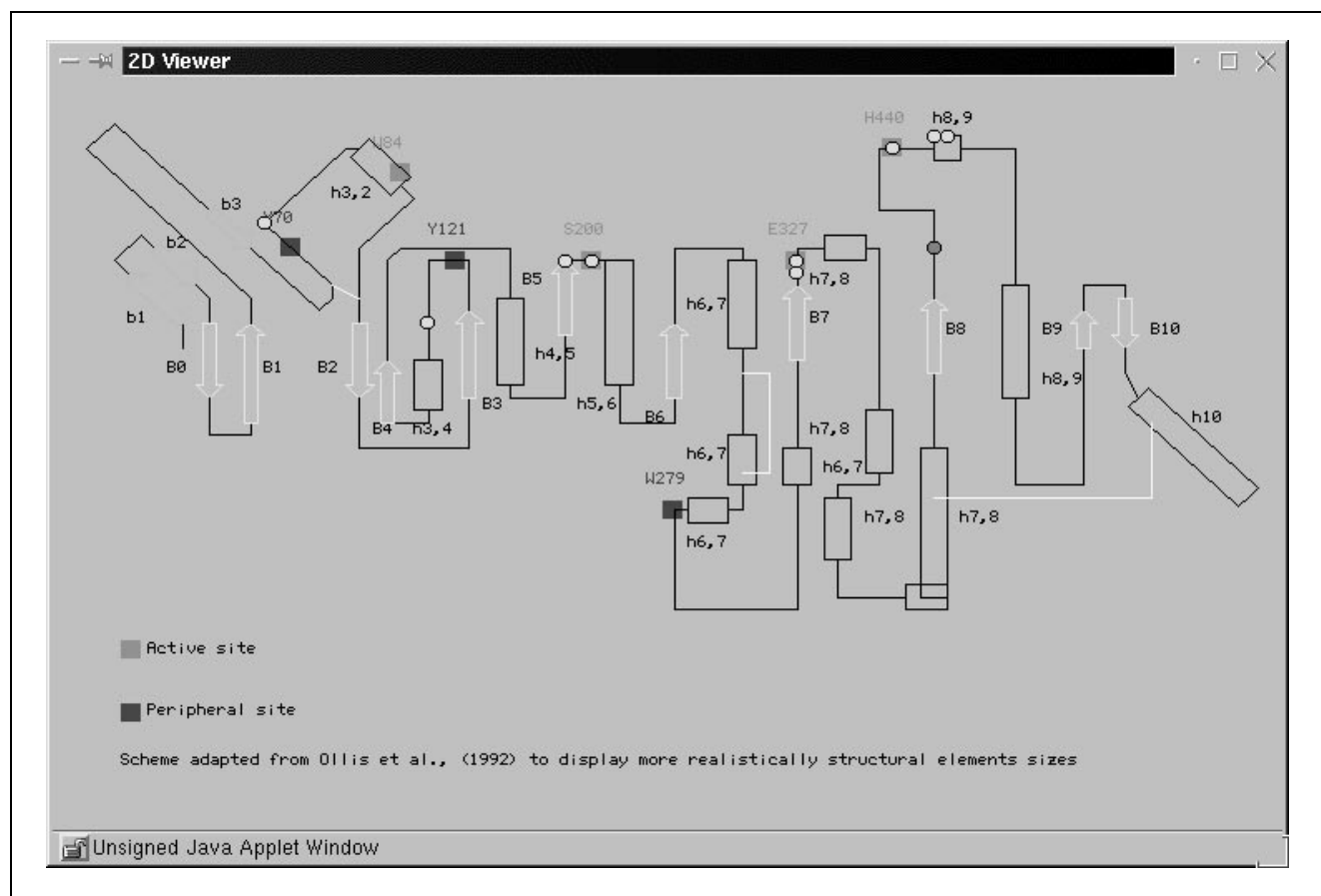
The first tool is Mutalign which provides a hypertext alignment of the cholinesterase sequences in which site-directed or natural mutations have been analysed. Links between any mutated amino acid positions and experimental data obtained for the mutation can be selected and browsed from this web page. Symbols over each column of the alignment lead to combined information for mutations in all sequences at this similar position. This tool has been described previously,<sup>4</sup> but now the mutation pages also include kinetic data obtained on the mutated enzymes (see below).

The second tool is a mutation viewer written in Java that was developed originally to visualise mutations in the p53 gene and protein.<sup>20</sup> It allows a keyword search of mutation summaries and interactive displays in 1D, 2D and 3D of the AChE protein.

When the applet starts, four windows appear, of which three are representations of the AChE protein:

- A 1D view of the AChE protein is presented at two scales using a 'semantic lens'. The peptide sequence of the prototypic *Torpedo californica* AChE is readable in the lower part of the window.
- A 2D representation of the AChE protein with localisation of structural elements (see Figure 2).
- A 3D structure of *Torpedo* AChE as a wire frame representation that can be manipulated interactively with the mouse (rotation, zoom).

Included with the browser, a text area allows dialogue with the applet. For example, keywords listed in Table 1 may



**Figure 2:** A 2D representation of the AChE protein with localisation of structural elements. The keyword 'Catalysis' has been entered as a search term and all amino acids mutated during the study of the catalysis of acetylcholine appear as green dots. One dot corresponding to residue 425 (highlighted in red) has been selected. This selection causes the display of the description of the mutant in the Text Viewer (see Figure 3)

**Table 1:** Keywords used for Mutation Viewer queries. Each of the terms correspond to a particular catalytic subsite or to groups of amino acids involved in substrate or inhibitor specificities. Choosing one of these terms allows one to see relative positions of these groups of amino acids in the secondary or tertiary structure

Natural_mutation	Catalysis
Peripheral_Anionic_Site	Succinylcholine_hydrolysis
Oxyanion_hole	Succinylthiocholine_binding
Acyl_pocket	Substrate_inhibition
Choline_binding_site	Dibucaine_inhibition
Salt_bridge	Phosphorylation
Omega_loop	OP-resistance
Electrostatic	Aging
Back_door_hypothesis	OP-specificity
N-glycosylation	Fasciculin_inhibition
Minimal_C-term	Acylation
Glypiation	Signal_transduction
Folding	Phosphotriesterase_activity
Biosynthesis	

be entered here to perform a query. The last window presents the results of such a query (see Figure 2).

When search terms are entered into the query entry box in the browser window, the 'hits' will be highlighted on the different representations of the AChE. For example when the term 'Catalysis' is chosen, then the window with the secondary structure of AChE will appear as shown in Figure 3. The circles represent the position of amino acids that were mutated in order to study catalysis of AChE and they will appear with a similar colour in the primary and tertiary structure windows. When a single amino acid is selected, it is highlighted in red in

Query	Accession	Method	Species	Enzyme	Position	Effect	Reference
'Catalysis':	99385	sdm	H4329_human-acche	425	Catalysis	impairs secretion but not activity	Shafferian_1992_J.Biol.Chem 267
'Catalysis':	99384	sdm	H4250_torca-acche	425	Catalysis	only 50% loss of activity	Gibney_1990_Proc.Natl.Acad.Sci.USA 87_75
'Catalysis':	99381	sdm	H4250/H4400_torca-acche	425	Catalysis	No_activity_H440_of_the_catalytic_triad	Gibney_1990_Proc.Natl.Acad.Sci.U

**Figure 3:** A text window describing the mutation selected on the secondary structure (see Figure 2). Column 1 shows the query performed. Column 3 indicates that these mutations were made by site-directed mutagenesis (sdm). Column 4 describes the mutations using a numbering according to the species and the mutated enzyme. Column 5 is the torpedo\_number, since AChE in different species shows some small variation in size, *Torpedo* AChE has been chosen as a typical AChE. Column 7 shows the effects of mutations and Column 8 is the bibliographic reference for each mutation

all the displays, and details of the mutation and its effect are displayed in the text window (see Figure 3).

We hope in the future to link the applet to the rest of the database and to be able to navigate from the viewer to the kinetic data pages.

### Substrates and inhibitors

A collection of data on 300 inhibitors and 70 substrates is available from the database. For the moment, most of these are compounds specific to the cholinesterase subset of the database.

Kinetic parameters obtained experimentally for 40 inhibitors and 36 substrates on 12 enzymes and 146 of their mutants have been collected. Shown in Figure 4 is a sample of the data collected for Tacrine, an inhibitor of acetylcholinesterase used for the treatment of Alzheimer's disease. The page includes the chemical structure and nomenclature, the kinetic parameters with summaries of experimental conditions, and references which are hypertext links to the details about the paper in which the experiment has been reported. Other links are available to pages about the mutations or the enzymes on which the inhibitor has been tested. When the mutation Y337A of mouse-acche (mouse acetylcholinesterase) is selected on this page, it gives the page presented in Figure 5 with the kinetic parameters determined for this particular mutation. A dictionary of terms is necessary to relate the sequence data to biological results. Objects associated with the sequence gene are obvious: mutations and amino acid changes. Parameters include Km, Kss,

Kcat, Kapp, Ki, Ksi, Kon, Koff and Ic50. This nomenclature allows easy construction of queries using the query language available in ACeDB. A few terms have not yet been introduced as classes, but exist in text format and should in the future be indexed, eg temperature, ionic strength and enzyme concentration. Yet the vocabulary available already allows rational comparison of kinetic parameters for similar mutations in different organisms and multiple mutations in a unique sequence.

### FUTURE DIRECTIONS

As with the general databases (eg GenBank, EBI, SWISS-PROT and PDB), the bottleneck for the development of the ESTHER database is the accurate entry of primary data and supplementary annotation. Facilitating the participation of individual scientists and laboratories in the submission of formatted data is of utmost priority. Data already in computer-readable formats in laboratories should be incorporated directly in the database, providing the appropriate conversion tools are available. Hopefully, it will soon be as common to submit such data as it is to submit sequence and structure data currently.

Ideally the end-users and their needs will direct the future development. Designs of new inhibitors, such as insecticides or therapeutic agents, could benefit from the development of the ESTHER database. Protein engineering of cholinesterases and related enzymes is already an active field with the design of biosensors with immobilised cholinesterases or decontaminating

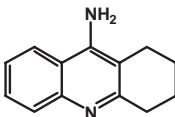
kinetic parameters in ESTHER

developing tools to enrich the database

**Figure 4:** A web page corresponding to the Tacrine entry in the ESTHER database. Tacrine is an inhibitor used in the treatment of Alzheimer's disease. Cells in the table are hypertext links to other objects in the database, eg papers, mutations, sequences. The table is truncated at the bottom and does not show all the kinetic parameters available from this page

# ESTHER

| Tables | BLAST | Mutations | Kinetics | Structures | Bibliography | Align/Phylo | AceEsther |  
 ↑ Home ⇒ Comments/Submission

Molecule	Tacrine					
Type	Acridine					
Other name	THA; tetrahydroamino-acridine; Cognex					
MW	Formula	CAS Number	Other database			
198.27	C13H14N2	321-64-2				
Chemical nomenclature	tetrahydroamino-acridine; 1,2,3,4,-tetrahydro-9 aminoacridine					
Perlace	More information on Tacrine....					
Comment	Treatment of Alzheimer Disease, trademark is cognex					
Structure						
Kinetic parameters	WT table					
*human-acche						
Enzyme	Ki	Ksi	Ic50	Substrate	Exp.conditions	Ref
D74N	4600 & 700 nM	65000 & 15000 nM	-	ACh	50mM Na P pH8 T22C	Ariel_1998
E202A	3300 & 150 nM	16000 & 3000 nM		ACh	50mM Na P pH8	Ariel_1998
.....						
*mouse-acche						
Enzyme	Ki	Ksi	Ic50	Substrate	Exp.conditions	Ref
WT	40 & 10 nM	130 & 50 nM	-	ACh	100mM NaP pH7.0 T22C	Radic_1993
Y337A	1.5 & 0.5 nM	5 & 1 nM	-	ACh	100mM NaP pH7.0 T22C	Radic_1993
Y337F	150 & 30 nM	680 & 180 nM	-	ACh	100mM NaP pH7.0 T22C	Radic_1993
.....						
*mouse-buche						
Enzyme	Ki	Ksi	Ic50	Substrate	Exp.conditions	Ref
WT	17 & 4 nM	47 & 22 nM	-	ACh	100mM NaP pH7.0 T22C	Radic_1993
.....						

**Figure 5:** A mutation selected in the previous table of Tacrine (see Figure 4) gives this new window which shows kinetic parameters tested on the mouse acetylcholinesterase for which the tyrosine 337 has been mutated to alanine. Links are available to substrate, inhibitors, references or torpedo\_number, which is the numbering reference of the aligned cholinesterases

# ESTHER

| Tables | BLAST | Mutations | Kinetics | Structures | Bibliography | Align/Phylo | AceEsther |  
 ↑ Home ⇒ Comments/Submission

**Mutation :**      **mouse -ac che\_Y337A**

This mutation corresponds to following torpedo number :      330

**Subst rate**

Molecu le	Ki	Ksi	Exp.conditions	Reference
Acety lthio ch oline	110 & 16 uM	52000 & 16000/min	100mM NaP pH7 T22C	Radic_1993

**Inhibitor**

Molecu le	Ki	Ksi	Ic50	Exp.conditions	Reference
9-Aminoac ridine	2.3 & 0.7 nM	5.9 & 1.0 nM	-	100mM NaP pH7.0 T22C	Radic_1993
Acridine	0.87 & 0.15 uM	3.4 & 0.8 uM	-	100mM NaP pH7.0 T22C	Radic_1993
Chlo rproma zine	17 & 7 uM	91 & 38 uM	-	100mM NaP pH7.0 T22C	Radic_1993
Edro phon ium	5.6 & 1.3 uM	26 & 14 uM	-	100mM NaP pH7.0 T22C	Radic_1993
Etho propa zine	0.041 & 0.003 uM	0.13 & 0.002 uM	-	100mM NaP pH7.0 T22C	Radic_1993
WT-fasciculin2	-	9.9 pM	-	100mM NaP pH7.0 T22C	Radic_1994
WT-fasciculin2	-	4.2 & 1.7 pM	-	100mM NaP pH7.0 T22C	Radic_1994
N-Methyl-ac ridin ium	0.049 & 0.010 uM	0.13 & 0.03 uM	-	100mM NaP pH7.0 T22C	Radic_1993
Prom a zine	37 & 11 uM	150 & 27 uM	-	100mM NaP pH7.0 T22C	Radic_1993
Prom eth a zine	29 & 11 uM	130 & 29 uM	-	100mM NaP pH7.0 T22C	Radic_1993
Tacrine	1.5 & 0.5 nM	5 & 1 nM	-	100mM NaP pH7.0 T22C	Radic_1993

devices for biological treatment of wastewater and hazardous waste. Improved lipases for food and household products is a very promising field that necessitated a database of its own.<sup>21</sup>

If such specialised databases correspond to a need, then practical actions will be necessary. Potentially, a thousand different types of protein fold implies a thousand individual databases for their characterisation. These will not all be of the same size, but those like the alpha/beta-hydrolases database will require much more expertise and will need to be split into as many databases as there are functions or pathways fulfilled by proteins with this particular fold. The ESTHER database would become the merging of 10 to 20 individual subdatabases curated by specialists of each subfamily. The interconnection and the definition of standards for some thousands of databases with expert annotation may be the challenge of the years to come.

### References

1. Alpha/beta-hydrolases fold database ESTHER: <http://www.ensam.inra.fr/cholinesterase/>
2. Cousin, X., Hotelier, T., Liévin, P. *et al.* (1996), 'A cholinesterase genes server (ESTHER): a database of cholinesterase-related sequences for multiple alignments, phylogenetic relationships, mutations and structural data retrieval', *Nucleic Acids Res.*, Vol. 24, pp. 132–136.
3. Cousin, X., Hotelier, T., Giles, K. *et al.* (1997), 'The alpha/beta fold family of proteins database and the cholinesterase gene server ESTHER', *Nucleic Acids Res.*, Vol. 25, pp. 143–146.
4. Cousin, X., Hotelier, T., Giles, K. *et al.* (1998), 'aCHEdb: the database system for ESTHER, the alpha/beta fold family of proteins and the cholinesterase gene server', *Nucleic Acids Res.*, Vol. 26(1), pp. 226–228.
5. Ollis, D. L., Cheah, E., Cygler, M. *et al.* (1992), 'The alpha/beta hydrolase fold', *Protein Eng.*, Vol. 5(3), pp. 197–211.
6. Schrag, J. D. and Cygler, M. (1997), 'Lipases and alpha/beta hydrolase fold', *Methods Enzymol.*, Vol. 284, pp. 85–107.
7. Nardini, M. and Dijkstra, B. W. (1999), 'Alpha/beta hydrolase fold enzymes: The family keeps growing', *Curr. Opin. Struct. Biol.*, Vol. 9(6), pp. 732–737.
8. Heikinheimo, P., Goldman, A., Jeffries, C. and Ollis, D. L. (1999), 'Of barn owls and bankers: A lush variety of alpha/beta hydrolases', *Structure Fold Des.*, Vol. 7(6), pp. 141–146.
9. Massoulié, J., Anselmet, A., Bon, S. *et al.* (1998), 'Acetylcholinesterase: C-terminal domains, molecular forms and functional localization', *J. Physiol. (Paris)*, Vol. 92(3–4), pp. 183–190.
10. Jain, R., Rivera, M. C. and Lake, J. A. (1999), 'Horizontal gene transfer among genomes: The complexity hypothesis', *Proc. Natl Acad. Sci. USA*, Vol. 96 (7), pp. 3801–3806.
11. Galperin, M. Y., Walker, D. R. and Koonin, E. V. (1998), 'Analogous enzymes: Independent inventions in enzyme evolution', *Genome Res.*, Vol. 8, pp. 779–790.
12. Bairoch A. (2000), The ENZYME data bank in 2000, *Nucleic Acids Res.*, Vol. 28, 304–305; <http://www.expasy.ch/enzyme>
13. Kanehisa, M. and Goto, S. (2000), 'KEGG: Kyoto encyclopedia of genes and genomes', *Nucleic Acids Res.*, Vol. 28(1), pp. 27–30; <http://www.genome.ad.jp/kegg/kegg2.html>
14. Goto, S., Nishioka, T. and Kanehisa, M. (2000), 'LIGAND: Chemical database of enzyme reactions', *Nucleic Acids Res.*, Vol. 28(1), pp. 380–382; [http://www.genome.ad.jp/dbget-bin/www\\_bfind?ligand](http://www.genome.ad.jp/dbget-bin/www_bfind?ligand)
15. <http://www.brenda.uni-koeln.de/>
16. <http://wit.mcs.anl.gov/EMP/>
17. <http://pmd.ddbj.nig.ac.jp/>
18. Durbin, R. and Thierry-Mieg, J. (1991), 'A *C. elegans* database.' Documentation, code and data available from anonymous FTP servers at [lirmm.lirmm.fr](http://lirmm.lirmm.fr), [cele.mrc-lmb.cam.ac.uk](http://cele.mrc-lmb.cam.ac.uk) and [ncbi.nlm.nih.gov](http://ncbi.nlm.nih.gov)
19. Stein, L. D. (1999), 'Internet access to the *C. elegans* genome', *Trends Genet.*, Vol. 15(10), pp. 425–427.
20. Hainaut, P., Hernandez, T., Robinson, A. *et al.* (1998), 'IARC Database of p53 gene mutations in human tumors and cell lines: updated compilation, revised formats and new visualisation tools', *Nucleic Acids Res.*, Vol. 26(1), pp. 205–213.
21. Pleiss, J., Fischer, M., Peiker, M. *et al.* (2000), 'Lipase engineering database: Understanding and exploiting sequence–function relationships', *J. Mol. Catal. B: Enzym.*, Vol. 10, pp. 491–508.