

ESTHER, the database of the α/β -hydrolase fold superfamily of proteins

Thierry Hotelier, Ludovic Renault¹, Xavier Cousin², Vincent Negre, Pascale Marchot¹ and Arnaud Chatonnet^{2,*}

UIC INRA-AgroM, Place Viala, 34060 Montpellier, France, ¹UMR-6560 CNRS Ingénierie des protéines, IFR Jean Roche, Université de la Méditerranée, Boulevard Pierre Dramard, 13916 Marseille Cedex 20, France and ²UMR-866 INRA-UMII-AgroM Différenciation Cellulaire et Croissance, Montpellier, France

Received September 15, 2003; Revised October 14, 2003; Accepted October 29, 2003

ABSTRACT

The α/β -hydrolase fold is characterized by a β -sheet core of five to eight strands connected by α -helices to form a $\alpha/\beta/\alpha$ sandwich. In most of the family members the β -strands are parallels, but some show an inversion in the order of the first strands, resulting in antiparallel orientation. The members of the superfamily diverged from a common ancestor into a number of hydrolytic enzymes with a wide range of substrate specificities, together with other proteins with no recognized catalytic activity. In the enzymes the catalytic triad residues are presented on loops, of which one, the nucleophile elbow, is the most conserved feature of the fold. Of the other proteins, which all lack from one to all of the catalytic residues, some may simply be 'inactive' enzymes while others are known to be involved in surface recognition functions. The ESTHER database (<http://bioweb.ensam.inra.fr/esther>) gathers and annotates all the published information related to gene and protein sequences of this superfamily, as well as biochemical, pharmacological and structural data, and connects them so as to provide the bases for studying structure–function relationships within the family. The most recent developments of the database, which include a section on human diseases related to members of the family, are described.

INTRODUCTION

The ESTHER database (<http://bioweb.ensam.inra.fr/esther>) was created 10 years ago when the α/β -hydrolase fold superfamily of proteins, as originally described by Ollis *et al.* (1) on the basis of crystalline structures, encompassed only five homologous proteins. Since then the number of protein structures has grown to 314 corresponding to 85 proteins, covering 33 classes. Recently, many predicted proteins conceptually deduced from genome sequences were assigned to the superfamily on the basis of sequence alignments. Currently the database contains more than 3500 gene loci

and their protein products, of which 365 are present in Swiss-Prot. The superfamily covers eight families in Pfam-A and is split into 50 families in ESTHER. New families are introduced. One family (the Ndr_family) was found independently by Shaw *et al.* (2) using a Bayesian computational algorithm. Two other families (abh_upf00227 and Duf_676), were detected using a sample sequence as a search template with 3D-PSSM (3).

The diversity of catalytic and non-catalytic functions fulfilled by members of the superfamily have been reviewed on several occasions (4–10). Previous updates of ESTHER presented the tools developed for comparing the structures and effects of mutations on the enzyme kinetics of members of the carboxylesterase and cholinesterase subfamilies (11–14). Since then these tools have been largely improved and expanded. The present update introduces selected recent developments related to the whole α/β -hydrolase fold superfamily of proteins.

OVERALL TABLE

The major generalist databases attribute the α/β -hydrolase fold proteins to different subfamilies, presumably because gathering processes based on short active site motifs and on Hidden Markov Models result in classifications that are not fully congruent with each other. However, sufficiently accurate correspondances allowed us to summarize relations between databases in a single table (Table 1, Supplementary Material). The fully linked table is available at <http://bioweb.ensam.inra.fr/ESTHER/general?what=synthese>. The table lists the Interpro Pfam Prints families or signatures, the ESTHER families, the ESTHER blocks [based on: Prosite PS00122-PS00941 for block C; PS01173-PS01174 for block H; PS00120 for block L; Block X all other entries as defined in reference (12)] and PDB accession codes for representative 3D structures. It is fully linked to original entries in the Interpro (15), Pfam (16), Prosite (17), Prints (18), Scop (19) and ESTHER databases.

ALIGNMENTS AND TREES

For each family, trees, pruned trees, subtrees and summarized coloured alignments and subalignments are available. These

*To whom correspondence should be addressed. Tel: +33 4 99 61 28 14; Fax: +33 4 67 54 56 94; Email: chatonne@ensam.inra.fr

are built automatically on a regular basis using the 'Rich Family Description' format developed by Corpet and collaborators for the Prodom database (20,21).

Human genetic diseases involving members of the α/β -hydrolase fold family

Mutations in 14 genes encoding proteins of the α/β -hydrolase fold have been found to be responsible for genetic diseases in man. Recently four genes were found to be mutated in neurological diseases: neuroligin 3 and 4 in autism (22), NDRG1 in motor and sensory neuropathy-Lom (23), maspardin in Mast syndrome (24). Mutations in three other genes appear to be associated with risk factors for three more diseases. Moreover, mutations in two more genes are related to increased sensitivity to xenobiotics. These diseases and their phenotypes are included in ESTHER as three classes (Disease, Risk_factor and Xenobiotic_sensitivity).

A page (<http://bioweb.ensam.inra.fr/ESTHER/disease.table>) summarizes this information, and links to genes, mutations and the OMIM database (25) are available. A distinct class (Symptom_of_disease) has been created which includes diseases where altered expression of a member of the α/β -hydrolase fold superfamily results from but is not the cause of the disease; a critical example in this class is Alzheimer's disease where inhibitors of cholinesterases are widely used for symptomatic treatment of the cognitive and neurological impairments associated with the disease.

COMPLEX QUERIES

For those users who are interested in programming access to the resource, ESTHER provides scriptable access to the underlying ACeDB database via the AcePerl module (stein.cshl.org/aceperl/) and the AQL query language. Precomputed data sets for common searches are available on the server and a page of examples is proposed.

Links to model organisms databases

Links to Wormbase (26), Flybase (27) and Ensembl (28) are provided when available for genes present in the genome of model organisms. For *Caenorhabditis elegans* genes a link to a subset of the genome browser of Wormbase is implemented automatically (e.g. http://bioweb.ensam.inra.fr/ESTHER/family?name=caeel-acche1&class=Gene_locus).

FUTURE DEVELOPMENTS

New tools are in preparation in particular for protein analysis (physicochemical parameters) based on primary sequence information.

For members of the superfamily with a known 3D (mostly crystalline) structure, ESTHER provides direct links to the structure files in the RCSB protein data bank. RCSB now provides fixed images of the 'assumed biological molecules' and links to these images will be provided in ESTHER. However, these images are overall views displayed with random orientation and are not always informative for ESTHER users, hence, as a further upgrade we will prepare our own overall views of the 'biological molecules' and close up views of the catalytic site with bound ligands.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We thank F. Corpet, J. Gouzy and D. Kahn (Laboratoire de Génétique Cellulaire, INRA, Toulouse) for providing us with tools from the Prodom database. This work was supported by the Génopôle Montpellier Languedoc-Roussillon (to A.C.) and the Programme Bioinformatique inter EPST 2002 (to A.C. and P.M.).

REFERENCES

- Ollis,D.L., Cheah,E., Cygler,M., Dijkstra,B., Frolow,F., Franken,S.M., Harel,M., Remington,S.J., Silman,L., Schrag,J. *et al.* (1992) The α/β hydrolase fold. *Protein Eng.*, **5**, 197–211.
- Shaw,E., McCue,L.A., Lawrence,C.E. and Dordick,J.S. (2002) Identification of a novel class in the α/β hydrolase fold superfamily: the *N*-myc differentiation-related proteins. *Proteins*, **47**, 163–168.
- Kelley,L.A., MacCallum,R.M. and Sternberg,M.J.E. (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.*, **299**, 501–522.
- Schrag,J.D. and Cygler,M. (1997) Lipases and α/β hydrolase fold. *Methods Enzymol.*, **284**, 85–107.
- Nardini,M. and Dijkstra,B.W. (1999) α/β hydrolase fold enzymes: the family keeps growing. *Curr. Opin. Struct. Biol.*, **9**, 732–737.
- Heikinheimo,P., Goldman,A., Jeffries,C. and Ollis,D.L. (1999) Of barn owls and bankers: a lush variety of α/β hydrolases. *Struct. Fold. Des.*, **7**, 141–146.
- Oakeshott,J.G., Claudianos,C., Russell,R.J. and Robin,G.C. (1999) Carboxyl/cholinesterases: a case study of the evolution of a successful multigene family. *Bioessays*, **21**, 1031–1042.
- Holmquist,M. (2000) α/β -hydrolase fold enzymes: structures, functions and mechanisms. *Curr. Prot. Pept. Sci.*, **1**, 209–235.
- Fischer,M. and Pleiss,J. (2003) The Lipase Engineering Database: a navigation and analysis tool for protein families *Nucleic Acids Res.*, **31**, 319–321.
- Soreq,H. and Seidman,S. (2001) Acetylcholinesterase: new roles for an old actor. *Nature Rev. Neurosci.*, **2**, 294–302.
- Cousin,X., Hotelier,T., Lievin,P., Toutant,J.P. and Chatonnet,A. (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.*, **24**, 132–136.
- Cousin,X., Hotelier,T., Giles,K., Lievin,P., Toutant,J.P. and Chatonnet,A. (1997) The α/β fold family of proteins database and the cholinesterase gene server ESTHER. *Nucleic Acids Res.*, **25**, 143–146.
- Cousin,X., Hotelier,T., Giles,K., Toutant,J.P. and Chatonnet,A. (1998) aChEDb: the database system for ESTHER, the α/β fold family of proteins and the Cholinesterase gene server. *Nucleic Acids Res.*, **26**, 226–228.
- Chatonnet,A., Cousin,X. and Robinson,A. (2001) Links between kinetic data and sequences in the α/β -hydrolases fold database. *Brief. Bioinform.*, **2**, 30–37.
- Mulder,N.J., Apweiler,R., Attwood,T.K., Bairoch,A., Barrell,D., Bateman,A., Binns,D., Biswas,M., Bradley,P., Bork,P. *et al.* (2003) The InterPro Database, 2003 brings increased coverage and new features. *Nucleic Acids Res.*, **31**, 315–318.
- Bateman,A., Birney,E., Cerruti,L., Durbin,R., Eddy,S.R., Griffiths-Jones,S., Howe,K.L., Marshall,M. and Sonnhammer,E.L. (2002) The Pfam protein families database. *Nucleic Acids Res.*, **30**, 276–280.
- Sigrist,C.J., Cerutti,L., Hulo,N., Gattiker,A., Falquet,L., Pagni,M., Bairoch,A. and Bucher,P. (2002) PROSITE: a documented database using patterns and profiles as motif descriptors. *Brief. Bioinform.*, **3**, 265–274.
- Attwood,T.K., Bradley,P., Flower,D.R., Gaulton,A., Maudling,N., Mitchell,A.L., Moulton,G., Nordle,A., Paine,K., Taylor,P. *et al.* (2003) PRINTS and its automatic supplement, prePRINTS. *Nucleic Acids Res.*, **31**, 400–402.

19. LoConte,L., Brenner,S.E., Hubbard,T.J.P., Chothia,C. and Murzin,A. (2002) SCOP database in 2002: refinements accommodate structural genomics. *Nucleic Acids Res.*, **30**, 264–267.
20. Corpet,F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.*, **16**, 10881–10890.
21. Corpet,F., Gouzy,J. and Kahn,D. (1999) Browsing protein families via the 'Rich Family Description' format. *Bioinformatics*, **15**, 1020–1027.
22. Jamain,S., Quach,H., Betancur,C., Rastam,M., Colineaux,C., Gillberg,I.C., Soderstrom,H., Giros,B., Leboyer,M., Gillberg,C. *et al.* (2003) Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nature Genet.*, **34**, 27–29.
23. Kalaydjieva,L., Gresham,D., Gooding,R., Heather,L., Baas,F., de Jonge,R., Blechschmidt,K., Angelicheva,D., Chandler,D., Worsley,P., *et al.* (2000) *N-myc* downstream-regulated gene 1 is mutated in hereditary motor and sensory neuropathy-Lom. *Am. J. Hum. Genet.*, **67**, 47–58.
24. Simpson,M.A., Cross,H., Proukakis,C., Pryde,A., Hershberger,R., Chatonnet,A., Patton,M.A. and Crosby,A.H. (2003) Maspardin is mutated in a complicated form of hereditary spastic paraplegia associated with dementia (Mast syndrome). *Am. J. Hum. Genet.*, **73**, 1146–1157.
25. Hamosh,A., Scott,A.F., Amberger,J., Bocchini,C., Valle,D. and McKusick,V.A. (2002) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, **30**, 52–55.
26. Harris,T.W., Lee,R., Schwarz,E., Bradnam,K., Lawson,D., Chen,W., Blasier,D., Kenny,E., Cunningham,F., Kishore,R. *et al.* (2003) WormBase: a cross-species database for comparative genomics. *Nucleic Acids Res.*, **31**, 133–137.
27. FlyBase Consortium (2003) The FlyBase database of the *Drosophila* genome projects and community literature. *Nucleic Acids Res.*, **31**, 172–175.
28. Clamp,M., Andrews,D., Barker,D., Bevan,P., Cameron,G., Chen,Y., Clark,L., Cox,T., Cuff,J. and Curwen,V. (2003) Ensembl 2002: accommodating comparative genomics. *Nucleic Acids Res.*, **31**, 38–42.