226–228 *Nucleic Acids Research, 1998, Vol. 26, No. 1*

aCHEdb: the database system for ESTHER, the α/β fold family of proteins and the Cholinesterase gene server

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

Différenciation Cellulaire et Croissance and [1]Unité Informatique, INRA-ENSAM, 2 place Viala, 34060 Montpellier, France and [2]Department of Structural Biology, Weizmann Institute of Science, 76100 Rehovot, Israel

Received October 1, 1997; Accepted October 3, 1997

ABSTRACT

Acetylcholinesterase belongs to a family of proteins, the α/β hydrolase fold family, whose constituents evolutionarily diverged from a common ancestor and share a similar structure of a central β sheet surrounded by α helices. These proteins fulfil a wide range of physiological functions (hydrolases, adhesion molecules, hormone precursors) [Krejci,E., Duval,N., Chatonnet,A., Vincens,P. and Massoulié,J. (1991) *Proc. Natl. Acad. Sci. USA*, 88, 6647–6651]. ESTHER (for esterases, α/β hydrolase enzymes and relatives) is a database aimed at collecting in one information system, sequence data together with biological annotations and experimental biochemical results related to the structure–function analysis of the enzymes of the family. The major upgrade of the database comes from the use of a new database management system: aCHEdb which uses the ACeDB progam designed by Richard Durbin and Jean Thierry-Mieg. It can be found at http://www.ensam.inra.fr/cholinesterase

THE α/β FOLD FAMILY OF PROTEINS

When a large number of protein structures was known it became clear that the proteins could be classified in four clearly separated classes depending on the fact that they contained only α helices or only β strands or α helix and β strand segments segregated along the polypeptide chain or α helices and β strands mixed or approximately alternating (1). Each class contains families of proteins which derive from a common ancestor. The α/β fold hydrolase family belongs to the last class. The alternating α helices and β strands are arranged so that a single β sheet contains most of the β strands surrounded by the α helices (2–4). Previous updates of our database contained descriptions of the signatures that are used to recognize by primary sequence homology if a sequence belongs to the α/β fold hydrolase family before any structural analysis is available (5,6). The members of the family which are homologous can be subdivided in three subfamilies (C, L and H) depending on different conserved motifs. Some members of the family do not show sequence homology with cholinesterases but their three-dimensional structures can be superimposed and probably diverged from a common ancestor (subfamily X). Since the last release (5), a new group of sequences (cutinases) for which the structure is known has been included. The structure of cutinases is interesting as it shows so far the smallest number of β strands while being clearly superimposable with structures of the other subfamilies (7,8).

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, EMBL, PIR, Swiss-Prot, etc.) and references.

The database contains now a total of 263 entries. New sequences are added regularly. There are 28 entries of cholinesterases; 19 contain a complete mature protein sequence. There are 164 entries in the C-family, 57 in the L-family, 27 in the H-family and 15 in the X-family. The genome sequencing programs allowed the characterization of a lot of sequences from model organisms. One can now find in the database 36 genes from *Caenorhabditis elegans*, 19 from *Drosophila melanogaster*, 13 from mouse and 23 from human.

NEW DEVELOPMENTS

The whole database has been ported to the ACeDB format. ACeDB is the database developed by Jean Thierry-Mieg (CRBM-CNRS Montpellier) and Richard Durbin (Sanger Center Cambridge) (9) integrating all data from the *C.elegans*: genetic map, physical map and sequence. It is also a database management system for a large number of genomic databases (10). The system shows great versatility, is very powerful and able to manipulate a large number of objects and answer complex questions. More than 80 specialized databases exist under this system. Information, tutorials, models and discussion groups exist and are easily accessible through the web at http://probe.

*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

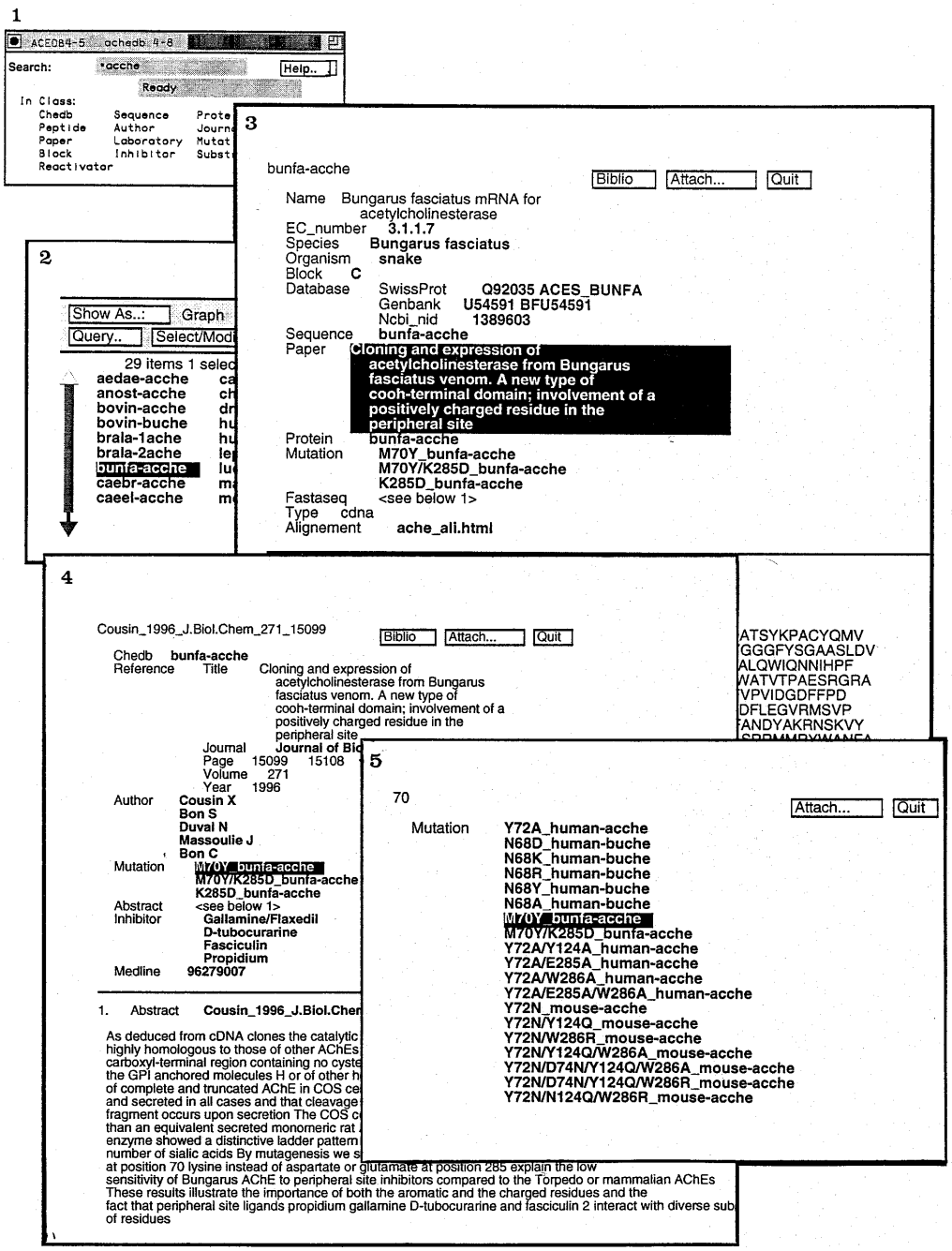

Figure 1. Screen dump of a typical search in ESTHER. Five windows have been successively opened. Window 1 is the window which opens automatically at the start of the program. It shows some of the main classes. Searching the AChEs in the Chedb Class, gives on the Main Keyset window 2 the list of entries corresponding to the search (entering the truncated string *acche in the search window means looking for all entries names ending with acche). Selecting the bunfa-acche prompts window 3 which contains all information on *Bungarus fasciatus* acetylcholinesterase. Selecting the Paper field, prompts window 4 with the reference. A fifth window shows all mutations studied in cholinesterases involving an amino-acid change at the position 70. 70 refers to the numbering of Torpedo acetylcholinesterase sequence and corresponds to residue 72 in human and mouse AChEs and 68 in human BChE.

nalusda.gov:8000/acedocs/index.html . Now the aCHEdb database is the source information of the ESTHER server.

The ACeDB interface appears as multi-windows. Figure 1 shows a screen dump of a typical consultation of aCHEdb. The first window shows the principal classes available (sequence, paper, author, inhibitor, etc.). The second one shows a list of objects in a selected class. In the case presented all acetylcholinesterase entries of aCHEdb are shown. Selecting bunfa-acche gives a window with all information on *Bungarus fasciatus* acetylcholinesterase. Selecting the paper opens a window on the reference related to this sequence. Selecting a mutation (M70Y_bunfa-acche)

on this window gives the information on this particular amino acid change.

The database is accessible in different formats in our ftp server: ftp://ftp.toulouse.inra.fr/pub/esther . The repository is organised in sub-directories each with a readme file. Stand alone versions are available for Unix system or Macintosh. For Macintosh 28 Mbytes of RAM are necessary. It is available as self-extractable archives. The Unix version exists for different operating systems. Users need the binary file of the program corresponding to the local machine and the update files.

The advantages of these stand alone versions are that they do not rely on the network (for those who have a slow link to the web), they can be customized to new functions or data from the user. The drawback is that updates are to be recovered and installed from time to time or need to be totally reinstalled when a lot of corrections are necessary on the previous updates.

The system is also now available as an X-client-server in test. Its purpose is to provide a complete graphic acedb interface without loading all the data. The advantage is that distant user recovers and saves data from the server when using the database. The data recovered and saved is always accessible locally and only new data needs to be recovered from the server. The size of the database in the client system is then limited to the data really useful to the client. This version uses the latest version 4.5 of the ACeDB software. At this time it is available only for the Unix system.

FUTURE DEVELOPMENT

The ACeDB management system is in constant improvement and new displays and programs get interrelated to it. A Java interface to acedb called Jade is under active development. Jade is developed by Lincoln Stein, Jean Thierry-Mieg, Doug Bigwood, John Barnett and Sam Cartinhour. A test system can be seen at http://alpha.crbm.cnrs-mop.fr . Jade gives full access to the database through the web. This relies totally on the web, so it does not need the acedb program on the client machine. It is available to all operating systems supporting the latest versions of browsers which are Java capable. We hope to soon be able to use this system to get the advantages of the ACeDB system and keep at the same time the service of ESTHER on the web.

AVAILABILITY OF ESTHER AND CITATION

The W3 URL is : http://www.ensam.inra.fr/cholinesterase . The ftp server is ftp://ftp.toulouse.inra.fr in the directory /pub/esther

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 the current article. Use of ACeDB should be cited as per ref. 9.

COMMUNITY OUTREACH AND USER SUPPORT

Demonstration and discussions of the use of the database will be organized during workshops at the 6th internationnal meeting on cholinesterase in San Diego (March 20–24, 1998) and the 3rd international meeting on esterases reacting with organophosphorus compounds in Dubrovnick, Crotia (April 15–18, 1998). More information can be found on the ESTHER server. To get included in the cholinesterase mailing list an Email 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.

ACKNOWLEDGEMENTS

The new developments would not have been possible without the constant support of Jean Thierry-Mieg (CNRS-Montpellier). We thank Solange Prime and Daniel Kahn (INRA, Toulouse) for providing anonymous ftp server space. This work was supported by the Institut National de la Recherche Agronomique (INRA) and grants from the Association Française contre les Myopathies (AFM).

REFERENCES

1 Levitt,M. and Chothia,C. (1976) *Nature*, **261**, 552–558.
2 Shrag,J.D., Li,Y., Wu,S. and Cygler,M. (1991) *Nature*, **351**, 761–765.
3 Sussman,J.L., Harel,M., Frolow,F., Oefner,C., Goldman,A., Toker,L. and Silman,I. (1991) *Science*, **253**, 872–879.
4 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.
5 Cousin,X., Hotelier,T., Lievin,P., Toutant,J.-P. and Chatonnet,A. (1996) *Nucleic Acids Res.*, **24**, 132–136.
6 Cousin,X., Hotelier,T., Giles,K., Lievin,P., Toutant,J.-P. and Chatonnet,A. (1997) *Nucleic Acids Res.*, **25**, 143–146.
7 Martinez,C., De Geus,P., Lauwereys,M., Matthyssens,G. and Cambillau,C. (1992) *Nature*, **356**, 615–618.
8 Martinez,C., Nicolas,A., van Tilbeurgh,H., Egloff,M.P., Cudrey,C., Verger,R. and Cambillau,C. (1994) *Biochemistry*, **33**, 83–89.
9 Durbin,R. and Thierry-Mieg,J. (1991) *A C. elegans Database*. Documentation, code and data available from anonymous FTP servers at lirmm.lirmm.fr, cele.mrc-lmb.cam.ac.uk and ncbi.nlm.nih.gov
10 Matthews,D.E. and Sherman,B.K. (1996) ACEDB Genome Database Software FAQ, ftp://rtfm.mit.edu/pub/usenet/news.answers/acedb-faq , http://probe.nalusda.gov:8000/acedocs/acedbfaq.html