

An evolutionary perspective on the first disulfide bond in members of the cholinesterase-carboxylesterase (COesterase) family: Possible outcomes for cholinesterase expression in prokaryotes

Arnaud Chatonnet, Xavier Brazzolotto, Thierry Hotelier, Nicolas Lenfant, P. Marchot, Yves Bourne

▶ **To cite this version:**

Arnaud Chatonnet, Xavier Brazzolotto, Thierry Hotelier, Nicolas Lenfant, P. Marchot, et al.. An evolutionary perspective on the first disulfide bond in members of the cholinesterase-carboxylesterase (COesterase) family: Possible outcomes for cholinesterase expression in prokaryotes. Chemico-Biological Interactions, 2019, 308 (308), pp.179-184. 10.1016/j.cbi.2019.05.016 . hal-02269496

HAL Id: hal-02269496
https://hal.science/hal-02269496

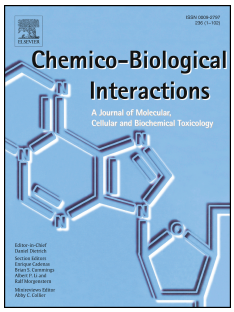
Submitted on 22 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Version of Record: https://www.sciencedirect.com/science/article/pii/S0009279719302388
Manuscript_e91de739ad8aa2f761aa90083965de80

Accepted Manuscript



An evolutionary perspective on the first disulfide bond in members of the cholinesterase-carboxylesterase (COesterase) family: Possible outcomes for cholinesterase expression in prokaryotes

Arnaud Chatonnet, Xavier Brazzolotto, Thierry Hotelier, Nicolas Lenfant, Pascale Marchot, Yves Bourne

PII: S0009-2797(19)30238-8

DOI: https://doi.org/10.1016/j.cbi.2019.05.016

Reference: CBI 8650

To appear in: *Chemico-Biological Interactions*

Received Date: 11 February 2019

Revised Date: 16 April 2019

Accepted Date: 13 May 2019

Please cite this article as: A. Chatonnet, X. Brazzolotto, T. Hotelier, N. Lenfant, P. Marchot, Y. Bourne, An evolutionary perspective on the first disulfide bond in members of the cholinesterase-carboxylesterase (COesterase) family: Possible outcomes for cholinesterase expression in prokaryotes, *Chemico-Biological Interactions* (2019), doi: https://doi.org/10.1016/j.cbi.2019.05.016.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2019 published by Elsevier. This manuscript is made available under the Elsevier user license
https://www.elsevier.com/open-access/userlicense/1.0/

ACCEPTED MANUSCRIPT

An evolutionary perspective on the first disulfide bond in members of the cholinesterase-carboxylesterase (COesterase) family: Possible outcomes for cholinesterase expression in prokaryotes

Arnaud Chatonnet[1]*, Xavier Brazzolotto[2], Thierry Hotelier[1], Nicolas Lenfant[1,3$], Pascale Marchot[3], Yves Bourne[3]

1) Institut National de la Recherche Agronomique (INRA) / Université Montpellier, Dynamique Musculaire et Métabolisme, Montpellier France.

2) Institut de Recherche Biomédicale des Armées (IRBA), Département de Toxicologie et Risques Chimiques, Brétigny-sur-Orge, France.

3) Centre National de la Recherche Scientifique (CNRS) / Aix-Marseille Univ, Architecture et Fonction des Macromolécules Biologiques, Marseille, France.

[$]Current address: Institut National de la Santé et de la Recherche Médicale (INSERM) / Aix-Marseille Univ, Génétique Médicale et Génomique Fonctionnelle, Marseille, France.

Abstract

Within the alpha/beta hydrolase fold superfamily of proteins, the COesterase group (carboxylesterase type B, block C, cholinesterases…) diverged from the other groups through simultaneous integration of an N-terminal, first disulfide bond and a significant increase in the protein mean size. This first disulfide bond ties a large Cys loop, which in the cholinesterases is named the omega loop and forms the upper part of the active center gorge, essential for the high catalytic activity of these enzymes. In some non-catalytic members of the family, the loop may be necessary for heterologous partner recognition. Reshuffling of this protein portion occurred at the time of emergence of the fungi/metazoan lineage. Homologous proteins with this first disulfide bond are absent in plants but they are found in a limited number of bacterial genomes. In prokaryotes, the genes coding for such homologous proteins may have been acquired by horizontal transfer. However, the cysteines of the first disulfide bond are often lost in bacteria. Natural expression in bacteria of CO-esterases comprising this disulfide bond may have required compensatory mutations or expression of new chaperones. This disulfide bond may also challenge expression of the eukaryote-

specific cholinesterases in prokaryotic cells. Yet recently, catalytically active human cholinesterase variants with enhanced thermostability were successfully expressed in *E. coli*. The key was the use of a peptidic sequence optimized through the Protein Repair One Stop Shop process, an automated structure- and sequence-based algorithm for expression of properly folded, soluble and stable eukaryotic proteins. Surprisingly however, the crystal structures of the optimized cholinesterase variants expressed in bacteria revealed co-existing formed and unformed states of the first disulfide bond. Whether the bond never formed, or whether it properly formed then broke during the production/analysis process, cannot be inferred from the structural data. Yet, these features suggest that the recently acquired first disulfide bond is difficult to maintain in *E. coli*-expressed cholinesterases. To explore the fate of the first disulfide bond throughout the cholinesterase relatives, we reanalyzed the crystal structures of representative COesterases members from natural prokaryotic or eukaryotic sources or produced as recombinant proteins in *E. coli*. We found that in most cases this bond is absent.

Introduction

Overproduction and purification of a recombinant protein from *Escherichia coli* can be faster, cheaper, and easier than from other organisms. The engineering of bacterial strains and vectors along with improvement of culture, extraction and solubilization protocols have substantially increased the number of eukaryotic proteins successfully produced in bacteria [1-3]. However, it is still often necessary to modify the sequence, such as optimizing it by substituting codons more frequently used in bacteria. Recently a new bioinformatics tool was developed with the aim of designing specific mutations that can markedly improve production of eukaryotic proteins in bacteria [4]. Yet, expression in *E. coli* of proteins with tridimensional folds only found in eukaryotes can be challenging, because bacteria lack specific pathways necessary for proper folding and post-translational modifications of these particular proteins. Disulfide bond formation is a post-translational modification that requires particular machineries, and several laboratory cell strains have been engineered to fulfill this task [3].

The alpha/beta hydrolase fold superfamily of proteins encompasses hydrolytic enzymes of widely differing substrate specificities. The common core comprises a central alpha/beta sheet, made of eight parallel or antiparallel beta strands connected by six alpha helices [5-7]. The proteins are believed to have diverged from a common ancestor in retaining little obvious sequence similarity. The superfamily can be split into subfamilies according to

structurally unrelated additional domains that can simply decorate the common core or bring a new function [8]. For example, members of the S9 protease family acquired a bulky beta propeller domain that covers the active center cavity of the alpha/beta hydrolase subunit and controls the enzyme specificity. The COesterase family is one of these subfamilies that are well represented in eukaryotic genomes. It is named Carboxylesterase_B IPR002018 in the InterPro database [9] and COesterase PF00135 in the PFAM database [10], and it corresponds to Block C in the ESTHER database [11, 12]. This COesterase family emerged from the shuffling of a genomic sequence coding for a N-terminal sequence with a gene coding for a core alpha/beta hydrolase fold scaffold (figure 1). This extra N-terminal sequence increased the length of the polypeptide from ca. 300 to ca. 500 amino acid residues and accordingly the size of the alpha/beta hydrolase subunit. In most cases the additional sequence includes two cysteines separated by up to 65 residues yet spatially close to each other, and whose disulfide bonding ties a 'Cys loop' that serves as an important regulator of the various functions of the subfamilies. The second cysteine of the pair belongs to a conserved SEDCLYLN motif which is a signature of the COesterase family (Prosite PS00941: [ED]-D-C-L-[YT]-[LIV]-[DNS]-[LIV]-[LIVFYW]-x-[PQR]). The shuffling is such an ancient event that the members of the family now fold as single domain subunits. For example, in enzymes of the family one can hardly separate the N-terminus extension from the core and still retain enzymatic activity. As well, the making of active chimera implies a high level of homology between the two parental enzymes [13, 14]. Attempts to identify an evolutionary origin of the shuffled N-terminal part failed so far, as no clear homology outside the family was found. The Pfam and Interpro databases sometimes annotate the second cysteine of the first disulfide bond as belonging to a motif of the Ubiquitin-conjugating enzyme family (Prosite PS50127: [FYWLSP]-H-[PC]-[NHL]-[LIV]-x(3,4)-G-x-[LIVP]-C-[LIV]-x(1,2)-[LIVR], where the cysteine is an active center residue, not involved in disulfide bonding). For example, the cholinesterase-like protein from the King cobra, *Ophiophagus hannah* (Uniprot: V8P7P1), is annotated that way, and this annotation is propagated in few other cholinesterase entries. However, the serendipity of this annotation suggests that it is a false positive association.

The acquisition of disulfide bonds is correlated with the organism complexity [15, 16], and the shuffling of the first disulfide bond of COesterase members probably occurred after the split of the Archeplastidae lineage from other eukaryotic lineages. The COesterase family is not represented in plants or red algae (figure 2) [17]. This 'first bond' signature is found in a limited number of bacterial genomes and could have been acquired by horizontal transfer from eukaryotes [18]. In particular it is absent in *E. coli*. The family was very successfully populated in the

fungi-metazoa lineage (>30,000 representatives in InterPro and >10,000 in ESTHER, which excludes some redundancy). No such proteins are found in *Saccharomyces cerevisiae* but they are highly represented in pathogenic and multicellular fungi, where protein enlargement permitted the emergence of a mobile lid essential for the lipid-water interfacial function of the extracellular lipases (see below). Not all protozoans produce enzymes of this family. For example, members of the family are present in *Monosiga brevicolis*, *Trichomonas vaginalis* and *Cryptosporidum hominis* but they are absent in *Paramecium tetraurelia* and *Plasmodium falciparum*. Proteins of the family are well represented in early multicellular organisms such as *Dictyostelium discoideum* and *Trichoplax adhaerens* [19]. In triploblastic animals the family is even more populated. Numerous carboxylesterases are present in arthropods, nematodes and chordates where the genes are often organized in clusters. Among these, carboxylesterases hydrolyze both endogenous substrates and substrates or xenobiotics found in the environment. In insects, some of these enzymes evolved to hydrolyze or scavenge organophosphates used as insecticides, a process that led to the selection of resistant strains [20]. Numerous lipases of triploblastic animals belong to the family and also possess the lid necessary for their lipolytic function. In the cholesterol esterase lipases the Cys loop is a mobile surface lid whose opening upon association/interaction with the lipid-water interface exposes the active center cavity and makes it fully accessible to the lipid substrate. In the cholinesterases the corresponding 'omega loop' (so-called because shaped as the capital Greek letter omega, 'Ω') contributes to form the active center gorge, essential for the high catalytic activity, and the peripheral anionic site, where allosteric regulators of catalysis bind [21, 22].

Expression of animal COesterases in bacteria has been a challenge in particular for cholinesterases, which had to be produced from eukaryotic cell lines. Recently the Protein Repair One Stop Shop (PROSS) process, an automated structure- and sequence-based algorithm for expression of properly folded, soluble eukaryotic proteins with an enhanced stability [23], was applied to human acetylcholinesterase (AChE) and butyrylcholinesterase (BChE) and the modified, yet functional enzyme variants were successfully expressed in *E. coli*, purified and crystallized [4, 24]. However, despite the remarkable retention of the genuine tridimensional structures by the variants, the first disulfide bond was found to be not always formed, a feature that prompted us to analyze comparatively the structurally related proteins expressed in bacteria.

Materials and methods

Sequences were retrieved from GenBank (http://www.ncbi.nlm.nih.gov/), UniProt (http://www.uniprot.org/), the current assembly of the genomes using the BLAST servers dedicated to species from DOE Joint Genome Institute (http://genome.jgi-psf.org/) or ENSEMBL (http://www.ensembl.org/) or NCBI (http://www.ncbi.nlm.nih.gov/sutils/genom_tree.cgi). The sequences used along with individual references and links to GenBank and UniProt can be retrieved from ESTHER. Alignments were performed with ClustalW. Crystal structures and their electron density maps were visualized with the graphics program Coot [25]. Figure 3 was generated with PyMOL (The PyMOL Molecular Graphics System, Version 2.2.0, Schrödinger, LLC).

Results and discussion

At present 383 crystal structures of members of the CO-esterase subfamily, corresponding to 33 distinct proteins, are available from the RCSB Protein Data Bank. The significant redundancy results from structures of a same protein with distinct mutations or ligands (e.g., there are 113 structures of *Torpedo californica* AChE). Out of the 33 distinct proteins, there are seven AChEs and seven neuroligins from various animal species, and three and four bacterial and fungi carboxylesterases, respectively. We selected those 10 proteins whose crystal structure was determined after recombinant protein expression in *E. coli*. Ten other coordinate files of the same enzymes generated from eukaryotic expression systems were also analyzed. The characteristics of the analyzed proteins and structures are summarized in Table 1. Compared to the disparity observed in the formation of the first disulfide bonds, the other two bonds were found to display higher stability as reflected in their continuous electron density patterns.

Of the bacterial proteins, the one from *Geobacillus stearothermophilus* (PDB code, 2OGS) has no cysteine at all [26]. The other two, from *Bacillus subtilis* and *Clostridium hathewayi* (1QE3, 5A2G), have cysteines well conserved in other COesterases, yet their structures show that these cysteines are too far apart (over 4 Å) to form a disulfide bond [27, 28]. All four fungal proteins purified from the organism of origin or expressed in another yeast strain have the first disulfide bond well formed as seen in their crystal structures (PDB: 1CRL, 1THG, 1UKC, 4BE4) [29-33] None of the fungal esterases or lipases produced in *E. coli* have been crystallized or analyzed for disulfide bond formation.

Both human AChE and BChE expressed from mammalian or insect cell lines and crystallized have their first disulfide bond fully formed (4PQE, 4AQD, 1P0I) [34, 35]. However, their variants generated through the

PROSS process and expressed in a bacterial strain optimized for disulfide bond formation [2] show possible wobble in the connectivity of the two cysteines, otherwise well positioned relative to each other, and in the disulfide bonding state (5HQ3, 6EMI) [4, 24]. Indeed, the crystal structure of the optimized BChE variant revealed co-existence of formed and unformed states of the first disulfide bond [4]. A similar situation was found upon reexamination of the structure of the AChE variant expressed the same way. In both cases the electron density map of each subunit in the dimer present in the asymmetric unit clearly reflects a significant proportion of molecules whose first disulfide bond is not formed. Whether the unformed bond involves two cysteines (i.e., the bond never formed) or two half-cystines (i.e., the bond properly formed, then broke at some point of protein production or x-ray data collection) cannot be inferred from the structural data. A recent review on the occurrence of allosteric disulfides in proteins and how they are cleaved to alter protein structure and function points out a non-favorable conformation of each of the three disulfide bridges in AChE and close homologs, thus ruling out the occurrence of allosteric disulfides in these proteins [36].

However, chemical and structural damages encountered by crystalline *T. californica* AChE subjected to extreme doses and duration of synchrotron radiation (i.e., much greater than those used for 'regular' data collection) were reported previously [37, 38]. In particular these studies focused on the fate of the three disulfide bonds of AChE. Throughout the nine temporal states analyzed (frames A to I), the second disulfide bond, which ties the base of a short loop located at the surface of the subunit (loop 254-265 in *T. californica* AChE), encountered critical damage as soon as frame B and was identified as being the most susceptible in *T. californica*, while the third bond, which stabilizes the C-terminal third of the subunit, began to elongate from frame C but did not break [37, 38]. In contrast, the first disulfide bond was found to be substantially more stable over the same period of time with no alteration in the inter-sulfur distance [37, 38]. Our examination of the electron density maps around the first disulfide bond pointed to partial loss from frame E only, indicating that this first bond was more resistant to radiation damage than the second bond. In *T. californica* the most susceptible second disulfide bond displays the largest solvent accessible surface area, while the first disulfide bond is buried inside the protein core, although one can find two small internal cavities on each side of this bond (as seen in 1EA5).

Hence in the structures of the human AChE and BChE variants expressed in bacteria, presence of a fully intact second disulfide bond rules out breaking of the first bond because of radiation damage during x-ray data collection. Despite the substantial proportion of proteins lacking the first bond, the unaltered hydrolytic activities of

the two variants, relative to their representative ATCh and BTCh substrates, point to an overall unaltered alpha/beta hydrolase fold and a fully functional active center. This suggests that the enhanced stability of the variants generated through the PROSS procedure by introducing a set of substitutions (including a slightly greater proline-to-glycine ratio) may compensate the partial weakness of the first disulfide bond [4, 23].

In the Vertebrates, the closest relatives to the cholinesterases are the neuroligins, of which the available structures, all solved from soluble forms expressed from eukaryotic cell lines, display a fully formed first disulfide bond [39]. One step further are the carboxylesterases, where a properly formed first disulfide bond is found not only in human CES1 produced in eukaryotic cell lines (4AB1, 1MX1) [40, 41], but also in pig liver esterase (Uniprot: a9gvw6) produced in *E. coli* (5FV4) [42]. Human bile salt activated lipase (cholesterol esterase), which has only the first two disulfide bonds, was produced both in *E. coli* and from eukaryotic cells. In the first case, the protein was found in inclusion bodies and refolded in vitro, and the crystal structure revealed that the first disulfide bond was only partly formed (1F6W) [43]. In the second case, the protein displayed full disulfide bonding (6H0T) [44].

For insect proteins, the COesterases members have been classified in three main phylogenetic classes broadly distinguished by dietary/detoxification, hormone/semiochemical processing, and neuro/developmental functions [45]. The classes are further subdivided in clades. The proteins that have been crystalized belong to different classes and clades. Insecticide-metabolizing carboxylesterase LcαE7 from the blowfly *Lucilia cuprina* and insecticide-sequestering carboxylesterase Cqestβ2 from *Culex quinquefasciatus* both fall in clade B (alpha esterases generally microsomal), esterase 6 from *Drosophila melanogaster* falls in clade E (secreted esterases), the juvenile hormone esterase of *Manduca sexta* belongs to clade G (lepidopteran-type JHEs), insect AChEs belong to clade J. Although sequence similarity networks analysis including sequences from more divergent groups pointed to variation in clustering in some clades, it did not change the repartition of the sequences analyzed here.

In both Cqestβ2 and LcαE7 the cysteines present in the N-terminal part of the sequences are too distant from each other to form a disulfide bond (>10 Å) [46, 47]. Also, in Cqestβ2 the consensus SEDCLYL sequence motif is replaced by the homologous sequence CEDLYN that shifts the Cys residue four positions backward. These two enzymes probably evolved from the original COesterase structure so that the folding of the loop was thermodynamically determined without the need of a disulfide bond to rigidify it. Whether this evolution was accompanied with an increase of expression of these enzymes for better hydrolysis or scavenging of organophosphates, or even a better affinity for insecticides, remains to be documented.

The *D. melanogaster* esterase 6 produced in *E. coli* crystallized with half of the molecules being apparently devoid of the first disulfide bond (5THM) [48]. The *M. sexta* juvenile hormone esterase and AChEs from both *D. melanogaster* and *Anopheles gambiae* were all crystalized after production from eukaryotic cells and all three structures display a properly formed first bond [49-51].

Conclusion

All these observations suggest that the first disulfide bond can properly form in *E. coli*, but that depending on the nature and sequence of the enzyme, the use (or not) of a fused chaperone protein, and the bacterial strain selected for production, the capacity of the produced polypeptide to correctly fold varies.

Acknowledgment

This work was supported by BioMedef NBC-5-C-4210 grant to XB.

Table 1 Summary of the characteristics of the analyzed proteins and structure. Note that in *L. cuprina* insecticide-metabolizing carboxylesterase (LcaE7) and *C. quinquefasciatus* insecticide-sequestering carboxylesterase (Cqestb2) four cysteines are found close to the expected spatial positions of the two cysteines of the first disulfide bond in other analyzed sequences.

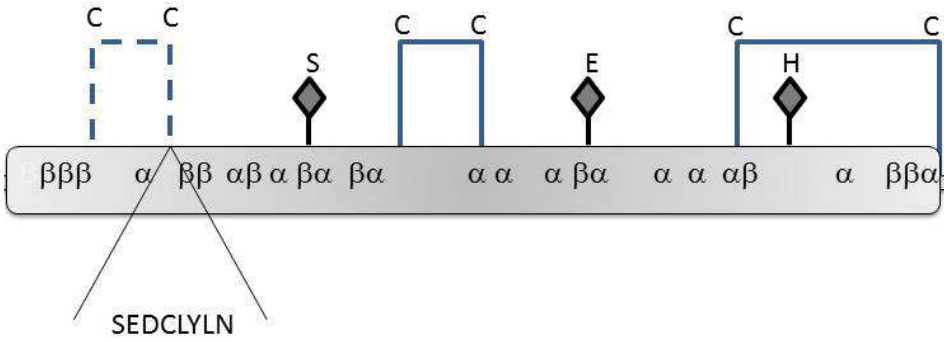


Fig 1

Disulfide bond pattern of a canonical COesterase sequence, based on the AChE structure. The cysteine positions are indicated with C letters. The first disulfide bond is displayed as a dashed line and the other two bonds as solid lines. The first bond is the one described in this study, the second bond is the most conserved within the alpha/beta hydrolase fold superfamily of proteins, and the third bond is the most facultative throughout the superfamily. The secondary structure elements are indicated with α for alpha helices and β for beta strands. The position of the conserved SEDCLYLN sequence motif that surrounds the second cysteine involved in the first disulfide bond is highlighted. The positions of the active center residues serine (S), glutamate (E) and histidine (H) are indicated with grey lozenges.

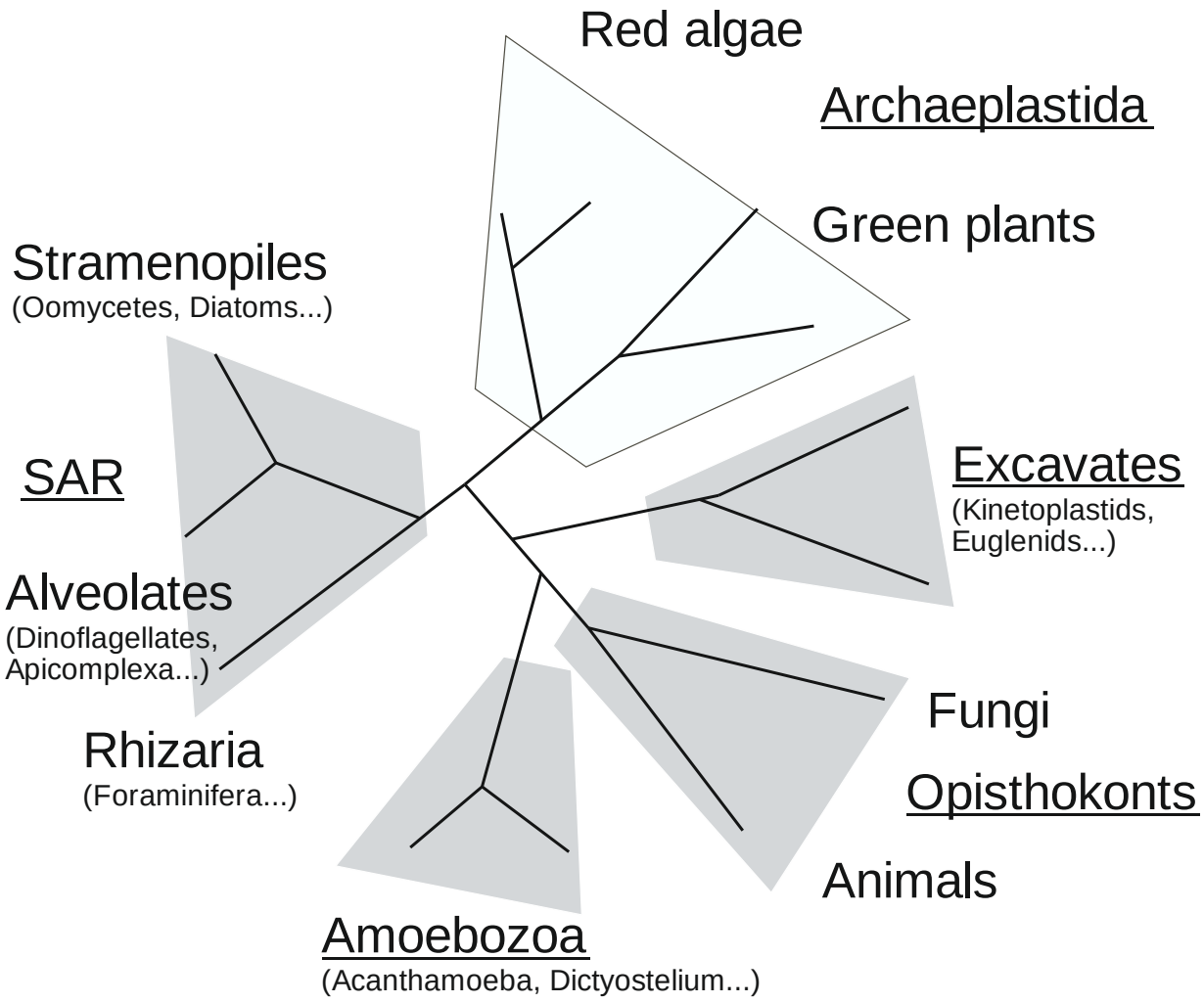


Fig 2

Simplified phylogeny of eukaryotes according to Keeling et al. [17]. Phyla where COesterases are found are represented by grey polygons. In the Archaeplastida phylum, which includes green plants and red algae, no COesterases are found.

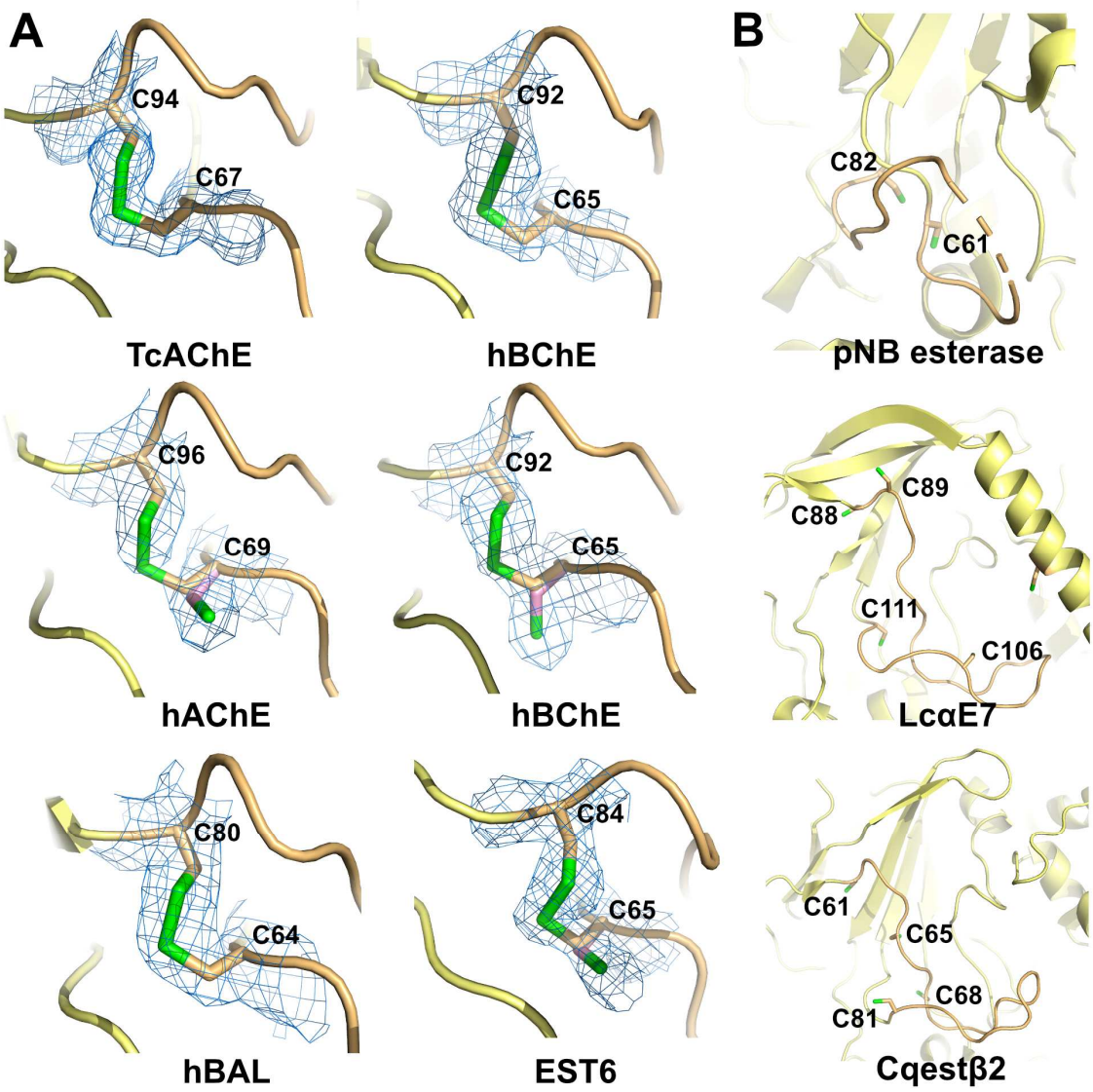


Fig 3

Structural comparison of the first disulfide bond topology in COesterases. (A) Close-up views of the crystal structures of six representative COesterases isolated from the genuine source (TcAChE, accession code 1EA5) or produced recombinantly in *E. coli* (pNB esterase, 1QE3; hAChE, 5HQ3; hBChE, 6EMI; LcαE7, 5IVH; hBAL, 1F6W; Cqestβ2, 5W1U) or from eukaryotic cells (hBChE, 1P0I; EST6, 5THM) showing the first disulfide bond (green sulfur atoms) along with the 2fo-fc electron density maps contoured at 1.0 σ. The Cys loop of variable length that is tightened by the first disulfide bond is in gold. For structures hAChE/5HQ3, hBChE/6EMI and EST6/5THM the alternate formed and unformed states of the bond are overlaid. (B) Close-up views of the crystal structures of

three representative insect COesterases devoid of a disulfide bond at the base of the long loop. (Abbreviations are: TcAChE, *T. californica* AChE; hBChE/hAChE, human BChE/AChE; hBAL, human bile salt activated lipase; EST6, *D. melanogaster* esterase 6; pNB esterase, *B. subtilis* para-nitrobenzyl esterase; LcaE7, *L. cuprina* insecticide-metabolizing carboxylesterase; Cqestb2, *C. quinquefasciatus* insecticide-sequestering carboxylesterase.)

References

[1] G. Ren, N. Ke, M. Berkmen, Use of the SHuffle Strains in Production of Proteins, Curr. Protoc. Protein Sci. 85 (2016) 5 26 21-25 26 21.
[2] J. Lobstein, C.A. Emrich, C. Jeans, M. Faulkner, P. Riggs, M. Berkmen, SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm, Microb. Cell Fact. 11 (2012) 56.
[3] N. Ke, M. Berkmen, Production of Disulfide-Bonded Proteins in Escherichia coli, Curr. Protoc. Mol. Biol. 108 (2014) 16 11B 11-21.
[4] A. Goldenzweig, M. Goldsmith, S.E. Hill, O. Gertman, P. Laurino, Y. Ashani, O. Dym, T. Unger, S. Albeck, J. Prilusky, R.L. Lieberman, A. Aharoni, I. Silman, J.L. Sussman, D.S. Tawfik, S.J. Fleishman, Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability, Mol. Cell. 63 (2016) 337-346.
[5] D.L. Ollis, E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S.M. Franken, M. Harel, S.J. Remington, I. Silman, J. Schrag, et al., The alpha/beta hydrolase fold, Protein Eng. 5 (1992) 197-211.
[6] M. Nardini, B.W. Dijkstra, Alpha/beta hydrolase fold enzymes: the family keeps growing, Curr. Opin. Struct. Biol. 9 (1999) 732-737.
[7] M. Holmquist, Alpha/Beta-hydrolase fold enzymes: structures, functions and mechanisms, Curr. Protein Pept. Sci. 1 (2000) 209-235.
[8] P. Marchot, A. Chatonnet, Enzymatic activity and protein interactions in alpha/beta hydrolase fold proteins: moonlighting versus promiscuity, Protein Pept. Lett. 19 (2012) 132-143.
[9] R.D. Finn, T.K. Attwood, P.C. Babbitt, A. Bateman, P. Bork, A.J. Bridge, H.Y. Chang, Z. Dosztanyi, S. El-Gebali, M. Fraser, J. Gough, D. Haft, G.L. Holliday, H. Huang, X. Huang, I. Letunic, R. Lopez, S. Lu, A. Marchler-Bauer, H. Mi, J. Mistry, D.A. Natale, M. Necci, G. Nuka, C.A. Orengo, Y. Park, S. Pesseat, D. Piovesan, S.C. Potter, N.D. Rawlings, N. Redaschi, L. Richardson, C. Rivoire, A. Sangrador-Vegas, C. Sigrist, I. Sillitoe, B. Smithers, S. Squizzato, G. Sutton, N. Thanki, P.D. Thomas, S.C. Tosatto, C.H. Wu, I. Xenarios, L.S. Yeh, S.Y. Young, A.L. Mitchell, InterPro in 2017-beyond protein family and domain annotations, Nucleic Acids Res. 45 (2017) D190-D199.
[10] R.D. Finn, P. Coggill, R.Y. Eberhardt, S.R. Eddy, J. Mistry, A.L. Mitchell, S.C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G.A. Salazar, J. Tate, A. Bateman, The Pfam protein families database: towards a more sustainable future, Nucleic Acids Res. 44 (2016) D279-285.
[11] N. Lenfant, T. Hotelier, E. Velluet, Y. Bourne, P. Marchot, A. Chatonnet, ESTHER, the database of the alpha/beta-hydrolase fold superfamily of proteins: tools to explore diversity of functions, Nucleic Acids Res. 41 (2013) D423-D429.
[12] N. Lenfant, Y. Bourne, P. Marchot, A. Chatonnet, Relationships of human alpha/beta hydrolase fold proteins and other organophosphate-interacting proteins, Chem. Biol. Interact. 259 (2016) 343-351.
[13] Z. Radic, N.A. Pickering, D.C. Vellom, S. Camp, P. Taylor, Three distinct domains in the cholinesterase molecule confer selectivity for acetyl- and butyrylcholinesterase inhibitors, Biochemistry 32 (1993) 12074-12084.
[14] D.C. Vellom, Z. Radic, Y. Li, N.A. Pickering, S. Camp, P. Taylor, Amino acid residues controlling acetylcholinesterase and butyrylcholinesterase specificity, Biochemistry 32 (1993) 12-17.
[15] J.W. Wong, S.Y. Ho, P.J. Hogg, Disulfide bond acquisition through eukaryotic protein evolution, Mol Biol Evol 28 (2011) 327-334.
[16] R.R. Thangudu, M. Manoharan, N. Srinivasan, F. Cadet, R. Sowdhamini, B. Offmann, Analysis on conservation of disulphide bonds and their structural features in homologous protein domain families, BMC Struct Biol 8 (2008) 55.
[17] P.J. Keeling, G. Burger, D.G. Durnford, B.F. Lang, R.W. Lee, R.E. Pearlman, A.J. Roger, M.W. Gray,The tree of eukaryotes, Trends Ecol Evol 20 (2005) 670-676.

[18] L. Pezzementi, A. Chatonnet, Evolution of cholinesterases in the animal kingdom, Chem. Biol. Interact. 187 (2010) 27-33.
[19] A. Chatonnet, X. Cousin, T. Hotelier, Evolution of the carboxylesterase_B (COesterase) family: the closest relatives of cholinesterases among alpha/beta hydrolase fold family. Hypothesis of horizontal transfer from eucaryotes to bacteria, N.C. Inestrosa, E.O. Campos (Eds.), Cholinesterases in the second millennium: biomolecular and pathological aspects,, FONDAP- Biomedicina, Pontificia Universidad Catolica de Chile, Santiago, Chile (2004) 75-80.
[20] T. Hotelier, V. Negre, P. Marchot, A. Chatonnet, Insecticide resistance through mutations in cholinesterases or carboxylesterases: data mining in the ESTHER database, J. Pestic. Sci. 35 (2010) 315-320.
[21] O. Lockridge, S. Adkins, B.N. La Du, Location of disulfide bonds within the sequence of human serum cholinesterase, J Biol Chem 262 (1987) 12945-12952.
[22] K. MacPhee-Quigley, T.S. Vedvick, P. Taylor, S.S. Taylor, Profile of the disulfide bonds in acetylcholinesterase, J Biol Chem 261 (1986) 13565-13570.
[23] A. Goldenzweig, M. Goldsmith, S.E. Hill, O. Gertman, P. Laurino, Y. Ashani, O. Dym, T. Unger, S. Albeck, J. Prilusky, R.L. Lieberman, A. Aharoni, I. Silman, J.L. Sussman, D.S. Tawfik, S.J. Fleishman, Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability, Mol. Cell 70 (2018) 380.
[24] X. Brazzolotto, A. Igert, V. Guillon, G. Santoni, F. Nachon, Bacterial Expression of Human Butyrylcholinesterase as a Tool for Nerve Agent Bioscavengers Development, Molecules 22 (2017) 1828.
[25] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, Acta Crystallogr D Biol Crystallogr 66 (2010) 486-501.
[26] P. Liu, H.E. Ewis, P.C. Tai, C.D. Lu, I.T. Weber, Crystal structure of the Geobacillus stearothermophilus carboxylesterase Est55 and its activation of prodrug CPT-11, J. Mol. Biol. 367 (2007) 212-223.
[27] B. Spiller, A. Gershenson, F.H. Arnold, R.C. Stevens, A structural view of evolutionary divergence, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 12305-12310.
[28] V. Perz, A. Baumschlager, K. Bleymaier, S. Zitzenbacher, A. Hromic, G. Steinkellner, A. Pairitsch, A. Lyskowski, K. Gruber, C. Sinkel, U. Kuper, D. Ribitsch, G.M. Guebitz, Hydrolysis of synthetic polyesters by Clostridium botulinum esterases, Biotechnol. Bioeng. 113 (2016) 1024-1034.
[29] J. Gutierrez-Fernandez, M.E. Vaquero, A. Prieto, J. Barriuso, M.J. Martinez, J.A. Hermoso, Crystal structures of Ophiostoma piceae sterol esterase: Structural insights into activation mechanism and product release, J. Struct. Biol. 187 (2014) 215-222.
[30] Y. Bourne, A.A. Hasper, H. Chahinian, M. Juin, L.H. De Graaff, P. Marchot, Aspergillus niger protein EstA defines a new class of fungal esterases within the alpha/beta hydrolase fold superfamily of proteins, Structure 12 (2004) 677-687.
[31] J.D. Schrag, M. Cygler, 1.8 Å refined structure of the lipase from Geotrichum candidum, J. Mol. Biol. 230 (1993) 575-591.
[32] P. Grochulski, Y. Li, J.D. Schrag, F. Bouthillier, P. Smith, D. Harrison, B. Rubin, M. Cygler, Insights into interfacial activation from an open structure of Candida rugosa lipase, J. Biol. Chem. 268 (1993) 12843-12847.
[33] J. Gutierrez-Fernandez, M.E. Vaquero, A. Prieto, J. Barriuso, M.J. Martinez, J.A. Hermoso, Crystal structures of Ophiostoma piceae sterol esterase: structural insights into activation mechanism and product release, J Struct Biol 187 (2014) 215-222.
[34] Y. Nicolet, O. Lockridge, P. Masson, J.C. Fontecilla-Camps, F. Nachon, Crystal structure of human butyrylcholinesterase and of its complexes with substrate and products, J. Biol. Chem. 278 (2003) 41141-41147.

[35] X. Brazzolotto, M. Wandhammer, C. Ronco, M. Trovaslet, L. Jean, O. Lockridge, P.Y. Renard, F. Nachon, Human butyrylcholinesterase produced in insect cells: huprine-based affinity purification and crystal structure, FEBS. J. 279 (2012) 2905-2916.
[36] J. Chiu, P.J. Hogg, Allosteric disulfides: Sophisticated molecular structures enabling flexible protein regulation, J Biol Chem 294 (2019) 2949-2960.
[37] M. Weik, R.B. Ravelli, G. Kryger, S. McSweeney, M.L. Raves, M. Harel, P. Gros, I. Silman, J. Kroon, J.L. Sussman, Specific chemical and structural damage to proteins produced by synchrotron radiation, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 623-628.
[38] M. Weik, J. Berges, M.L. Raves, P. Gros, S. McSweeney, I. Silman, J.L. Sussman, C. Houee-Levin, R.B. Ravelli, Evidence for the formation of disulfide radicals in protein crystals upon X-ray irradiation, J. Synchrotron Radiat. 9 (2002) 342-346.
[39] I.P. Fabrichny, P. Leone, G. Sulzenbacher, D. Comoletti, M.T. Miller, P. Taylor, Y. Bourne, P. Marchot, Structural analysis of the synaptic protein neuroligin and its beta-neurexin complex: determinants for folding and cell adhesion, Neuron 56 (2007) 979-991.
[40] S. Bencharit, C.L. Morton, J.L. Hyatt, P. Kuhn, M.K. Danks, P.M. Potter, M.R. Redinbo, Crystal structure of human carboxylesterase 1 complexed with the Alzheimer's drug tacrine: from binding promiscuity to selective inhibition, Chem. Biol. 10 (2003) 341-349.
[41] H.M. Greenblatt, T.C. Otto, M.G. Kirkpatrick, E. Kovaleva, S. Brown, G. Buchman, D.M. Cerasoli, J.L. Sussman, Structure of recombinant human carboxylesterase 1 isolated from whole cabbage looper larvae, Acta Crystallogr. Sec.t F Struct. Biol. Cryst. Commun. 68 (2012) 269-272.
[42] A. Hummel, E. Brusehaber, D. Bottcher, H. Trauthwein, K. Doderer, U.T. Bornscheuer, Isoenzymes of pig-liver esterase reveal striking differences in enantioselectivities, Angew. Chem. Int. Ed. Engl. 46 (2007) 8492-8494.
[43] S. Terzyan, C.S. Wang, D. Downs, B. Hunter, X.C. Zhang, Crystal structure of the catalytic domain of human bile salt activated lipase, Protein Sci. 9 (2000) 1783-1790.
[44] C. Touvrey, C. Courageux, V. Guillon, R. Terreux, F. Nachon, X. Brazzolotto, X-ray structures of human bile-salt activated lipase conjugated to nerve agents surrogates, Toxicology 411 (2018) 15-23.
[45] J.G. Oakeshott, R.M. Johnson, M.R. Berenbaum, H. Ranson, A.S. Cristino, C. Claudianos, Metabolic enzymes associated with xenobiotic and chemosensory responses in Nasonia vitripennis, Insect Mol Biol 19 Suppl 1 (2010) 147-163.
[46] C.J. Jackson, J.W. Liu, P.D. Carr, F. Younus, C. Coppin, T. Meirelles, M. Lethier, G. Pandey, D.L. Ollis, R.J. Russell, M. Weik, J.G. Oakeshott, Structure and function of an insect alpha-carboxylesterase (alphaEsterase7) associated with insecticide resistance, Proc. Natl. Acad. Sci. U. S. A. 18 (2013) 10177-10182.
[47] D.H. Hopkins, N.J. Fraser, P.D. Mabbitt, P.D. Carr, J.G. Oakeshott, C.J. Jackson, The structure of an insecticide sequestering carboxylesterase from the disease vector Culex quinquefasciatus: what makes an enzyme a good insecticide sponge?, Biochemistry 56 (2017) 5512-5525.
[48] F. Younus, N.J. Fraser, C.W. Coppin, J.W. Liu, G.J. Correy, T. Chertemps, G. Pandey, M. Maibeche, C.J. Jackson, J.G. Oakeshott, Molecular basis for the behavioral effects of the odorant degrading enzyme Esterase 6 in Drosophila, Sci. Rep. 7 (2017) 46188.
[49] M. Wogulis, C.E. Wheelock, S.G. Kamita, A.C. Hinton, P.A. Whetstone, B.D. Hammock, D.K. Wilson, Structural studies of a potent insect maturation inhibitor bound to the juvenile hormone esterase of Manduca sexta, Biochemistry 45 (2006) 4045-4057.
[50] M. Harel, G. Kryger, T.L. Rosenberry, W.D. Mallender, T. Lewis, R.J. Fletcher, J.M. Guss, I. Silman, J.L. Sussman, Three-dimensional structures of Drosophila melanogaster acetylcholinesterase and of its complexes with two potent inhibitors, Protein Sci 9 (2000) 1063-1072.

[51] J. Cheung, A. Mahmood, R. Kalathur, L. Liu, P.R. Carlier, Structure of the G119S Mutant Acetylcholinesterase of the Malaria Vector Anopheles gambiae Reveals Basis of Insecticide Resistance, Structure 26 (2018) 130-136.

ACCEPTED MANUSCRIPT

Protein	Species of origin	PDB code / resolution	Expression strain	Compartment	Cys positions for first bond	First disulfide bond	Cys to cys distance	Reference	Comments
Bacteria Carboxylesterases									
bacsu-pnbae	*Bacillus subtilis*	1QE3 / 1.50 Å	*E. coli* BL21	Total lysis, soluble fraction	61, 81	100% no	4.1 Å	Spiller et al. 1989 [27]	-
geost-est50	*Geobacillus stearothermophilus*	2OGS / 2.00 Å	*E. coli* Top 10	Total lysis, soluble fraction	no Cys	100% no		Liu et al. 2007 [26]	Thermostable enzyme
9clot-r5t6k9	*Clostridium hathewayi*	5A2G / 1.90 Å	*E. coli* BL21	Not reported	64, 89	100% no	4.1 - 4.3 Å	Perz et al. 2015 [28]	-
Fungi Carboxylesterases-lipases									
9pezi-q2tfw1	*Ophiostoma piceae*	4BE4 / 2.60 Å	*Komagataella pastoris*	Secreted	72, 108	yes		Gutierrez-Fernandez et al. 2014 [29]	-
aspni-EstA	*Aspergillus niger*	1UKC / 2.10 Å	*Aspergillus niger*	Secreted	85, 101	yes		Bourne et al. 2004 [30]	-
geoca-2lipa	*Geotrichum candidum*	1THG / 1.80 Å	*Geotrichum candidum*	Secreted	61, 105	yes		Schrag and Cygler 1993 [31]	-
canru-1lipa	*Candida rugosa*	1CRL / 2.06 Å	*Diutina rugosa*	Secreted	60, 97	yes		Grochulski et al. 1993 [32]	-
Mammalian Cholinesterases									
human-AChE	*Homo sapiens*	4PQE / 2.90 Å 5HQ3 / 2.60 Å	HEK *E. coli* SHuffle	Secreted Total lysis, soluble fraction	69, 96	yes 100/90% no (subunit A/B)		- Goldenzweig et al. 2016 [4]	- PROSS, 51 substitutions from WT sequence
human-BChE	*Homo sapiens*	1P0I / 2.00 Å 4AQD / 2.50 Å 6EMI / 2.47 Å	CHO *D. melanogaster* S2 *E. coli* SHuffle	Secreted Secreted Total lysis, soluble fraction	65, 92	yes yes 50/90% no (subunit A/B)		Nicolet et al. 2003 [33] Brazzolotto et al. 2012 [34] Brazzolotto et al. 2017 [24]	- - PROSS, 47 substitutions from WT sequence
Mammalian Neuroligin									
human-NLGNX4	*Homo sapiens*	2WQZ / 3.90 Å	HEK	Secreted		yes		Fabrichny et al. 2007 [38]	-
Mammalian Carboxylesterases									
human-CES1	*Homo sapiens*	4AB1 / 2.12 Å 1MX1 / 2.40 Å	*Baculovirus* *Baculovirus*	Secreted Secreted	87, 116 1087, 1116	yes yes		Greenblatt et al. 2012 [40] Bencharit et al. 2003 [39]	- -
pig-a9gyw6	*Sus scrofa*	5FV4 / 2.40 Å	*E. coli* BL21	Not reported	70, 99	yes		Hummel et al. 2007 [41]	-
Mammalian Bile salt activated lipase (Cholesterol esterase)									
human-CEL	*Homo sapiens*	1F6W / 2.30 Å	*E. coli* BL21	Inclusion bodies / refolding	64, 80	90% no		Terzyan et al. 2000 [42]	-
Insect Carboxylesterases Juvenile hormone esterase									
luccu-E3aest7 (LcαE7)	*Lucilia cuprina*	5IVH / 1.70 Å	*E. coli* BL21	Total lysis, soluble fraction	88-89, 106, 111	100% no	> 10 Å	Jackson et al. 2013 [45]	No substitutions
drome-este6	*Drosophila melanogaster*	5THM / 2.15 Å	*E. coli* Origami	Total lysis, soluble fraction	65, 84	50% no		Younus et al. 2017 [47]	16 substitutions to increase thermostability
culqu-1estb (Cqestβ1)	*Culex quinquefasciatus*	5W1U / 2.50 Å	*E. coli* BL21	Total lysis, soluble fraction	61, 65, 68, 81	100% no	> 10 Å	Hopkins et al. 2017 [46]	CEDSLYN instead of SEDCLYN
manse-jhest	*Manduca sexta*	2FJ0 / 2.70 Å	*Baculovirus*	Secreted	92, 112	yes		Wogulis et al. 2006 [48]	-

ACCEPTED MANUSCRIPT

The first C-C bond in cholinesterases and carboxylesterases is eukaryote-specific.

Formation of this bond may challenge bacterial expression of mammal cholinesterases.

We explored the bond density maps in structures of eukaryotic or prokaryotic enzymes.

Formed / unformed states of the bond coexist in cholinesterases expressed in bacteria.

Only the unformed state of the bond is observed in bacterial carboxylesterases.