Structure and Function Studies of Insect Carboxylesterases

A thesis submitted for the degree of Doctor of Philosophy of The Australian National University

Australian National University

Nicholas James Fraser

June 2017

Declaration

I declare that this thesis is the result of my own work and that the work contained herein is my own except where explicitly stated otherwise in the text before each chapter.

This work has been carried out by myself in the laboratory of Associate Professor Colin Jackson. To the best of my knowledge, the work presented in this thesis has not been previously submitted, in part or whole, to any university or institution for any degree, diploma, or other qualification.

Nicholas J. Fraser

June 2017

Acknowledgements

This thesis would have not been possible without the great support of many people over the years and I would like to thank all the people who contributed in some way to the work described in this thesis. First and foremost, I would like to express my sincere gratitude to my advisor, Associate Professor Colin Jackson for allowing me to conduct research in his laboratory. I very much appreciate all the opportunities and projects he has provided me and his patience, guidance and motivation during my studies.

Besides my primary advisor, I would like to thank the rest of my PhD panel members especially Dr. John Oakeshott, for all the insightful comments and invaluable suggestions over the years. I am profoundly grateful to Professor John White also for the opportunity to conduct research overseas.

My sincere thanks to Dr. Peter Mabbitt for all the constructive comments and guidance during the last years of my doctorate studies. I am also thankful for all the collaborators on my projects helping the work and papers come to fruition.

I am grateful for the financial assistance from the Australian Government and the Research School of Chemistry, without the funding, this thesis would not be possible. I am also grateful for Dr. Colin Scott and members at CSIRO Black Mountain (Matt & Nigel) for getting my laboratory skills during honours to a high standard for the start of my doctorate studies.

I am deeply thankful to everyone of the Jackson group past and present for important friendships and stimulating discussions over the many years. I would also like to express my gratitude to Jo for the assistance and feedback during the write-up.

And last but not least, my deep and sincere gratitude goes to my family and friends back home. My parents for their continuous encouragement and support over the years. My older brother Sam, for enduring my presence during my time in Canberra and to my younger brother Ben, for keeping me in check.

Abstract

Insect carboxylesterases (CBEs) have great importance in insect biology and play a crucial role in the ever-increasing occurrence of insecticide resistance. This resistance to insecticides (e.g. organophosphates) leads to downstream effects on agriculture and to human health (through reduced crop production and the spread of insect-borne diseases). Structural and functional studies of insect CBEs are vital for combating insecticide resistance, finding new potential targets and in providing insight into the evolution of new protein function. The work described in this thesis contributes to the overall understanding of the structure, function and evolution of insect CBEs.

Chapter 1 is the introduction into the structure, function and evolution of insect CBEs. In particular, the classification of CBEs is discussed, as well, the current understanding of the role CBEs play in insect species. Chapter 2 describes the evolution of new oligomeric structure in the CBE αE7 from *Lucilia cuprina* and shows that higher order thermostable oligomers can be selected for during evolution. This work affirms the importance of oligomers in evolution to maintain or increase protein stability and demonstrates structure/activity trade-offs that are observed throughout enzyme evolution.

Chapter 3 describes structural studies into the CBE esterase-6 (EST6) from *Drosophila melanogaster* - a CBE that has shown to be important for the reproductive success of *Drosophila* species. The enzyme has a unique active site entry for insect CBEs, which results in a narrow and shortened active site. Docking simulations in combination with kinetic analyses show that the enzyme is a probable odorant-degrading enzyme. It also indicates that EST6 does not directly interact with the sex pheromone 11-*cis* vaccenyl acetate (cVA), which contradicts the previously held belief that it is the main substrate of the enzyme.

The evolution of CBEs involved in insecticide resistance in several species is discussed in Chapter 4. α-esterase orthologs from different dipteran species were analysed to discern important factors for the evolution of qualitative insecticide resistance and the signifiance of evolutionary contingency. A large variation in binding and turnover for the organophosphate compound was observed with the introduction of the Gly137Asp mutation into the orthologs. Given the similarity in the predicted structures of the orthologs, it suggests that second and third shell mutations are important in mediating the catalytic effects conferred by the Gly137Asp mutation in the orthologs.

Chapter 5 discusses the development and testing of new inhibitors to treat the symptoms of Alzheimer's disease, through molecular docking simulations. Galanthamine derivatives showed no significant binding to AChE, however, the tested marinoquinoline derivatives displayed a large variation in affinity for AChE. Chapter 6 summarises the conclusions that these studies have allowed us to make regarding the structures, functions and evolution of insect CBEs. The chapter finishes with a discussion of future work that could be undertaken to extend these findings and further develop our understanding of this important enzyme family.

Contents

Abbreviations

4-NPB – 4-nitrophenyl butyrate

AAA – Aplidopsamine A

A. aegypti - *Aedes aegypti*

Ach – acetylcholine

AChE – acetylcholinesterase

AD – Alzheimer's disease

A. gambiae – *Anopheles gambiae*

Asp – aspartate

B. dorsalis - *Bactrocera dorsalis*

BChE – butyrylcholinesterase

CBE – carboxylesterase

C. capitata - *Ceratitis capitate*

CNS – central nervous system

C. hominivorax – *Cochliomyia hominivorax*

C. pipiens – *Culex pipiens*

C. quinquefasciatus - *Culex quinquefasciatus*

Cys – cysteine

C. stygia – *Calliphora stygia*

cVA – 11-*cis* vaccenyl acetate

D. melanogaster – *Drosophila melanogaster*

DEUP – Diethylumbelliferyl phosphate

DMSO – dimethyl sulfoxide

EC - Enzyme Commission

E. coli – *Escherichia coli*

Est – esterase

EST6 – esterase-6

EST7 – esterase-7

GABA - *gamma*-aminobutyric acid

GAL - (-)-galanthamine

Glu – glutamate

Gly – glycine

GST – glutathione S-transferase

HEPES - 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

H. irritans – Haematobia irritans

His – histidine

IC_{50} – half maximal inhibitory concentration

IUBMB – International Union of Biochemistry and Molecular Biology

L. cuprina – Lucilia cuprina

Leu - leucine

M. domestica – Musca domestica

MQ – marinoquinoline

MW – molecular weight

N. vitripennis – Nasonia vitripennis

OP – organophosphate

P450 – cytochrome P450 monoxygenase

PDB – Protein Databank

P. falciparum – Plasmodium falciparum

Phe – phenylalanine

RMSD – root-mean-square deviation

S. calcitrans - Stomoxys calcitrans

SEC – size exclusion chromatography

Ser - serine

Trp – tryptophan

TT_{50} – transition temperature 50

Tyr – tyrosine

PAGE – polyacrylamide gel electrophoresis

WT - wildtype

Chapter 1. Introduction

1.1. Introduction

1.1.1. Overview

This section will introduce carboxylesterases (CBEs), a large family of enzymes that have been of major interest in humans (*e.g.* carboxylesterase 1 (1)) and insects (*Lc*αE7 (2)), where the enzymes have been shown to play important roles in xenobiotic metabolism. CBEs will be introduced in terms of their classification, followed by a discussion of the structures of CBEs with a particular focus on insect CBEs.

1.1.2. Esterases

Esterases are a class of hydrolase enzymes that are ubiquitous in nature and catalyse the cleavage of ester bonds by the addition of water **(Figure 1.1)** (3–6). Esterases are important for the metabolism of several classes of compounds and have evolved independently in multiple superfamilies (7–10). Esterases perform crucial functions ranging from hormone metabolism, pheromone processing and neurotransmission in a wide variety of organisms (5, 11–14). They have received attention in recent years for their role in insecticide resistance, metabolism of pharmaceuticals and for use as biocatalysts (11, 15–17). Esterases that hydrolyse carboxyl esters (carboxylesterases or CBEs) will be discussed in this thesis as most esterases belong to the carboxyl/cholinesterase family (14)**.**
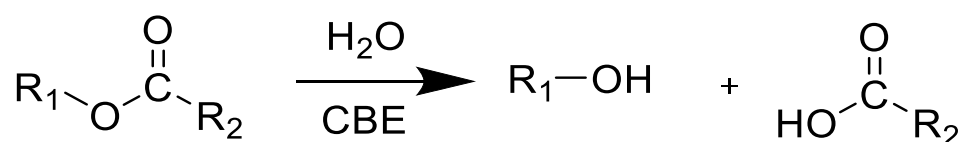
Figure 1.1. Basic reaction of a CBE: the breakage of the ester bond leads to an alcohol and a carboxylic acid as products. Figure adapted from Wheelock et al (18).

1.1.3. Classification of Carboxylesterases

There has been a significant effort to functionally define and classify eukaryotic CBEs. One of the main ways to classify enzymes is through activity, however, due to overlapping substrate ranges, this has proven difficult with CBEs (14, 19). Historically, interactions with inhibitors have been used to subdivide esterases, with the initial classification by Aldridge defining esterases into three classes (20–22). This system was widely used due to the ease of classification and lack of other classification systems (18). The three classes were A, B and C which subdivided esterases that hydrolyse organophosphates (Est-A), those that are irreversibly inhibited by organophosphates (Est-B) and esterases that do not

interact with organophosphates (Est-C) (18, 20, 23, 24). A problem with this classification was that it was considered to be too broad as a large majority of insect esterases would be categorised as Est-B (18, 19).

With more refined studies, an alternative scheme was proposed to address the difficulties in classifying esterases (20). Esterases were grouped into four classes: arylesterases, acetylesterases, carboxylesterases and cholinesterases (14, 20). The classes were defined by the sensitivity to three different types of inhibitors and several ester substrates following native PAGE **(Table 1.1)** (25–27).

Table 1.1. Classification of esterases based on inhibition and substrate specificity (27).

Esterases	Classification
Arylesterases	Inhibited only by sulfhydryl inhibitors and prefers esters with aromatic alcohol groups.
Carboxylesterases	Inhibited only by organophosphates and a preference for aliphatic esters.
Cholinesterases	Inhibited by organophosphates and eserine sulphates. Prefers charged substrates (choline esters).
Acetylesterases	No sensitivity to the 3 inhibitors and a preference for aliphatic esters with acetyl acid or aromatic alcohol groups.

A further classification has been used to group eukaryotic CBEs, which has been adopted in *Drosophila* and *Culex* species. This classifies esterases based on the preferential hydrolysis of two artificial carboxylic ester substrates: α-naphthyl acetate (Est-α) or β-naphthyl acetate (Est-β) and on electrophoretic mobility (28–31). A superscript number is used after the α or β to designate alleles for each variant as the increasing numbers of esterases are described in the species (28, 31, 32). The α/β nomenclature is still used for studying enzyme clusters in insects despite the limitations in a classification that only uses two substrates, such as having no broad biological distinctions and not being an informative predictor of enzyme function (19, 32, 33). In particular species (*e.g.* aphids) a different nomenclature is used to classify esterases based only on their different mobility on electrophoretic gels (34, 35). Electrophoretic mobility has limitations for classification as 20-30% of the variability in mobility is hidden, resulting in variants not being identified (36).

The International Union of Biochemistry and Molecular Biology (IUBMB) has also attempted to classify esterases on reaction type, however, the overlapping substrate range has made classification by substrate specificity difficult. The IUBMB gives a four series of numbers to classify enzymes (23, 37). EC 3 designates the subgroup of hydrolases, 3.1 are for hydrolases that act on an ester bond which also include lipases and exonucleases. EC 3.1.1 are for carboxyl ester hydrolases which includes acetylcholinesterase **(Table 1.2)** (23, 37). EC 3.1.1.1 is designated for esterases with a preference for aliphatic esters that are inhibited by paraoxon (14). This designation is also known as the CBEs or Est-B. In some cases, the overlapping substrate specificity means esterases in EC 3.1.1.1 fit into EC 3.1.1.2 as well (esterases that preferably act on aromatic compounds and hydrolyse triester organophosphates) but the inhibition by paraoxon distinguishes the two groups (14).

Table 1.2. EC 3.1.1 classification for hydrolases that work on ester bonds. Only the first eight in the 3.1.1.X designation are shown (23, 37).

EC Number	Classification	Substrates	Products
3.1.1.1	Carboxylesterase	Carboxylic Ester	Alcohol, Carboxylate
3.1.1.2	Arylesterase	Phenyl acetate	Phenol, Acetate
3.1.1.3	Triacylglycerol lipase	Triacylglycerol	Diacylglycerol, Carboxylate
3.1.1.4	Phospholipase A2	Phosphatidylcholine	Acylglycerophosphocholine, Carboxylate
3.1.1.5	Lysophospholipase	Lysophosphatidylcholine	Glycerophosphocholine , Carboxylate
3.1.1.6	Acetylesterase	Acetic Ester	Acetate, Alcohol
3.1.1.7	Acetylcholinesterase	Acetylcholine	Acetate, Choline
3.1.1.8	Cholinesterase	Acylcholine	Choline, Carboxylate

More recently, phylogenetics has been used to split insect CBE families into 14 major clades (annotated as A-N) and into three major groups for classification (32, 38). This classification was based on 35 *Drosophila melanogaster* sequences, 51 *Anopheles gambiae* sequences and 35 other insect sequences. Similarity of the gene, physiological characteristics, catalytic competence and cellular localisation is used to differentiate the CBEs into three main functional groups: dietary detoxification (mostly intracellular active

enzymes, enzymes with broad substrate specificity and those that are involved in insecticide resistance, A-C), hormone and pheromone degradation (secreted and catalytic enzymes, D-G) and neurodevelopmental functions (mostly non-catalytic, secreted, membrane-associated enzymes, H-N) (32, 38). An update to the classification was released in 2010 with the genome of *Nasonia vitripennis* sequenced (39). The updated clades were based on the genomes of *N. vitripennis*, *D. melanogaster*, *A. gambiae*, *Apis mellifera* and *Tribolium castaneum* with other previously characterized CBE sequences included **(Table 1.3)** (39). The assignment of the clades is similar to the previous assignment, however, clade B and C are combined to reflect a closer relationship between higher and lower dipteran CBEs and a new clade comprising of coleopteran CBEs of unknown function. Many authors have used the phylogenetic classification criteria recently for classification as it has the ability to functionally distinguish esterases, however, a better understanding of the physiological function of the esterases by genomic and biochemical studies is required.

Table 1.3. Clade classification of invertebrate esterases by Oakeshott et al. (2010) (39) based on the initial clades of Oakeshott et al. (2005) (38) (A) Dietary/detoxification function class (red) (B) Hormone/semiochemical processing class (blue) (C) Neuro-developmental, cell adhesion functions class (brown).

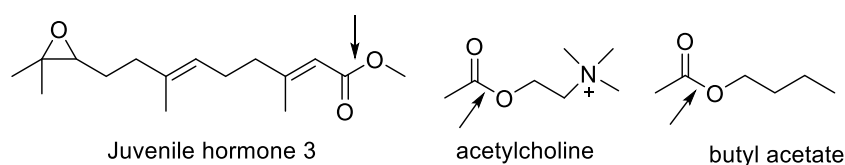| Clade	Classified	Examples
A	Hymenopteran xenobiotic metabolising esterases	-
B	α-esterases type enzymes, generally microsomal	*Lc*αE7
C	Unknown function	-
D	Integument esterases	-
E	Secreted β-esterases	Esterase-6, E4
F	Dipteran type juvenile hormone esterases	Juvenile hormone esterase
G	Lepidoteran type juvenile hormone esterases	Juvenile hormone esterase
H	Glutactin enzymes	Glutactin
I	Unknown function	-
J	Acetylcholinesterases	Acetylcholinesterase
K	Gliotactins	Gliotactin
L	Neuroligins	Neuroligin
M	Neurotactins	Neurotactin

Overall, there is no universal standard used in the classification of esterases, and multiple systems have been adopted in attempt to find a single method to classify esterases.

Subdividing esterases by substrate specificity, interactions with inhibitors and electrophoretic mobility all have issues for classification. However, the phylogenetic criteria is the favoured method currently, especially with the increase in sequencing information. Future biochemical, functional and structural studies are important to help establish the understanding of members of the 14 clades in the future. It has been suggested that future classification could involve the use of selective inhibitors or substrates, which would help to elucidate isozyme abundance (18).

1.1.4. Carboxylesterases

The CBEs are a multigene family that extensively radiated in higher eukaryotes (mammals and insects) (5, 12, 13). Despite their abundance in eukaryotes, CBE genes are comparatively rare in bacterial genomes, which suggests that the gene family evolved from ancient bacteria but has not expanded in prokaryotes (40, 41). The CBEs in higher eukaryotes show a large diversity in function that is highlighted in their broad substrate spectrum and complex expression profiles **(Figure 1.2)**. In some cases, there is strong substrate selectivity and specialised function (acetylcholinesterase) or less specific function and poor substrate selectivity (*Lc*αE7) and in rare cases, proteins from this fold can be non-catalytic (neuroligin) (32).

Natural Ester Substrates

Juvenile hormone 3   acetylcholine   butyl acetate

Artificial Ester Substrates

α-napthyl acetate   β-napthyl acetate   4-nitrophenyl butyrate   4-methylumbelliferyl acetate
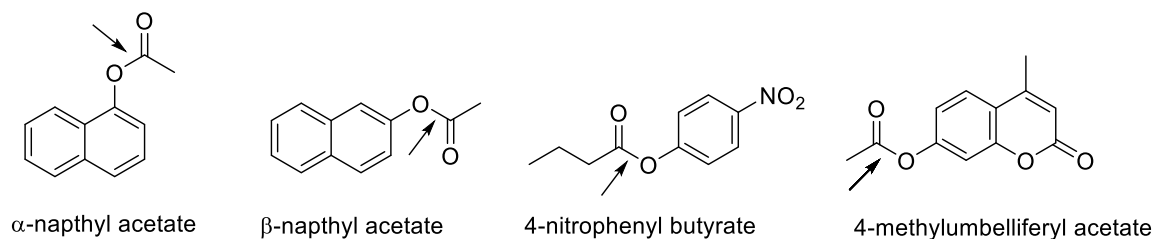
Drugs containing ester bonds

procaine   cocaine   ritalin
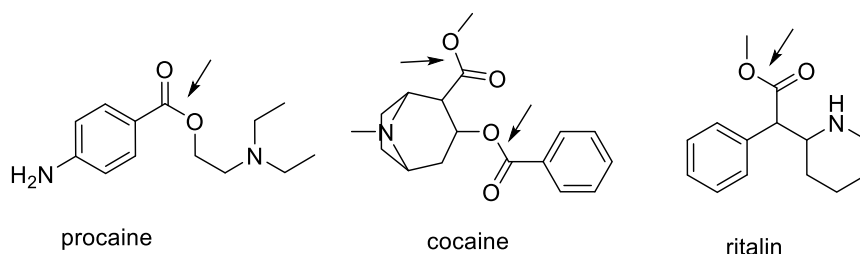
Figure 1.2. List of substrates (natural or artificial) hydrolysed by CBEs. The bonds that are subject to hydrolysis are indicated by the arrow. Figure adapted from Montella et al (14).

1.1.5. α/β hydrolase fold

Members of the CBE family do not share a high degree of sequence identity with an average sequence identity of 29% (42). Despite the low average sequence identity, the conservation of the tertiary structure is high. CBEs belong to the α/β hydrolase fold protein superfamily with a conserved arrangement of residues in the catalytic site (43–45). The conserved structure and arrangement of the active site suggest that CBEs originated from a common ancestor (46). This canonical fold is found in many functionally diverse enzymes that hydrolyse a wide range of substrates and is one of the most common protein folds. This superfamily includes proteases, esterases, lipases, dehalogenases, peroxidases, phosphotriesterases and epoxide hydrolases (4, 45, 46). The structure is well conserved with at least 20% amino acid identity between the most distantly related family members and 73 conserved residues identified (47). The ESTHER database collects published protein sequences and gene information of proteins with the α/β hydrolase fold and the list of CBEs is growing with 6157 of the 46165 sequences being CBEs as of May 2016 (45).

The core of the α/β hydrolase fold contains eight β-sheets in a parallel orientation (except β2), which are connected by six surrounding α-helices **(Figure 1.3)** (4). Loops and helices are spread among the β-strands in the primary sequence. The core fold is important for maintaining the position of the catalytic triad found in the active site (4). The protein fold allows for a large variation in domain size depending on the loop sizes and other secondary structure (25 - 65 kDa) (48).
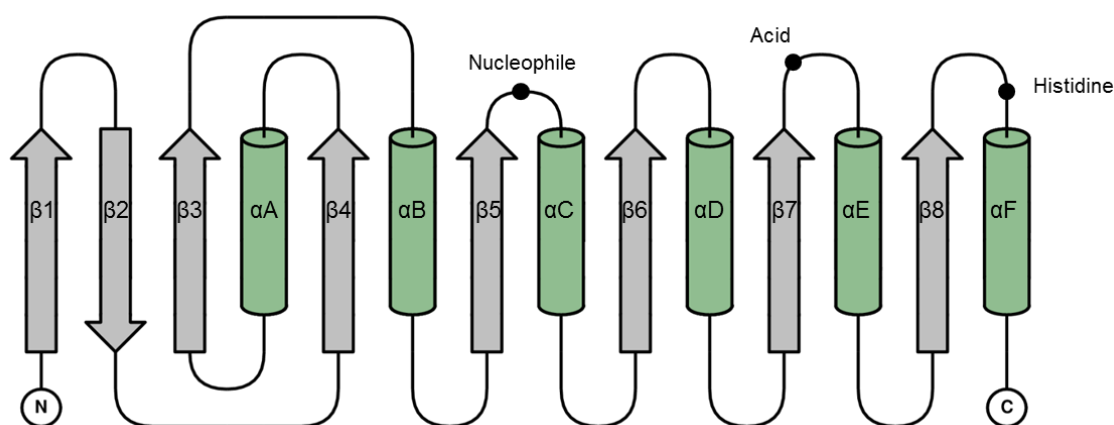
Figure 1.3. Secondary structure diagram of the 'canonical' α/β hydrolase fold (4). α-helices are represented by green cylinders and β-strands by gray arrows. The location of the catalytic triad is indicated by black dots (47).

The active site is formed by highly conserved loops and contains a nucleophilic residue (serine, cysteine or aspartate), a histidine residue and an acidic residue (glutamate or aspartate) (3, 4). The common triad is Ser, Glu, His (eukaryotic CBEs) but can also be Ser, Asp, His (minority of eukaryotic CBEs) and Cys, Asp, His (bacterial arylesterases) **(Figure 1.4)** (42). The nucleophilic residue is optimally positioned on a highly conserved turn known as the nucleophilic elbow (Gly-X-Ser-X-Gly) (4). Other conserved regions include the oxyanion hole, which stabilises the carbonyl group of the substrate, the P1 subsite (leaving group pocket) and the P2 subsite (acyl pocket) (43). All catalytic residues are required for esterase activity but there are examples of CBEs that are non-catalytic (49). The inactive members have large insertions with the catalytic residues relative positions still maintained (49). This highlights the large variation found within the α/β hydrolase fold and shows the great capacity of the fold for adaptation (50–52).

Figure 1.4. The canonical α/β hydrolase active site makeup. The oxyanion hole is shown and the catalytic triad of Glu, His and Ser (18).

Three structures of insect carboxyl/cholinesterases had been reported at the time of writing this thesis: acetylcholinesterase (AChE) from the species *D. melanogaster* (53), juvenile hormone esterase from *Manduca sexta* (54) and αE7 from *Lucilia cuprina* (55). The structure of AChE has been used in the past for modelling insect esterases to give struct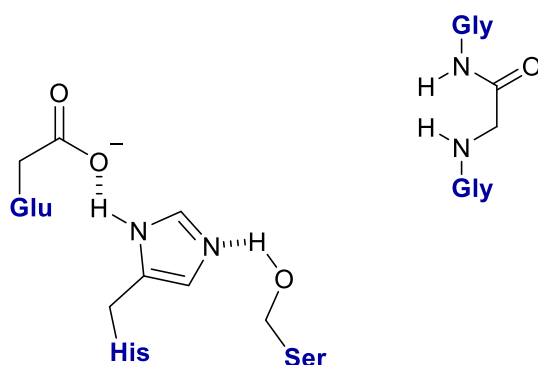ure-function insight (56). The active site of the solved insect CBEs varies significantly in the two P1 and P2 subsites (55, 57–59). Other noticeable differences are the entry to the catalytic site, alternative binding sites involved in allostery, and the number of disulfide bonds (54, 55, 57–59). Little is known about the aspects of the structure affecting catalytic processes in CBEs, as small sequence changes can affect catalytic machinery. To improve the understanding of the functional diversity found within CBEs, further structural data are required.

1.1.6. Regulation of CBEs

As well as the lack of structural information about insect CBEs, there is little known about the gene regulation of insect CBEs particularly in response to insecticides. Insecticide resistance can be achieved by the amplification of CBE genes in the genome of resistant insects during all life stages (60). The amplification level is usually variable with minimal control and exposure to an insecticide leads to a higher copy number (60). More recently it has been suggested cis-regulation is important to control gene amplification (60, 61). Resistance can also be achieved by CBE genes being 'switched on' in resistant species by methylation (34, 62). This form of transcriptional regulation is able to reduce fitness costs in the absence of insecticide selection by stopping the overproduction of enzyme by demethylation (34, 62). The methylation control is evidence for an epigenetic feature playing a role in insecticide resistance as the resistance mechanism is inherited to future generations and contributes to phenotype variability (34, 62, 63). Recent evidence is

suggesting histone modifications can affect gene expression of CBEs and that epigenetics plays a major regulatory role in the development of insecticide resistance and adaptation (63). More work is required, however, to identify the relationship with insect CBEs, genetic evolution, and epistasis.

1.1.7. Reaction Mechanism

The proposed mechanism of CBEs proceeds by a two-step reaction catalysed by the catalytic triad *via* an acyl-enzyme intermediate to generate an acid and an alcohol as products (64). The hydrolysis is base-mediated, requiring water, and is highly conserved among CBEs (64). The currently accepted mechanism is based on work by Satoh, Hosokawa, Sogorb, Vilanova, and Redinbo **(Figure 1.5)** (1, 65–68). The first step of the proposed reaction mechanism involves covalent binding between the catalytic nucleophilic residue and the carbonyl carbon of the substrate, resulting in liberation of the alcohol (1, 65–68). Nucleophilic attack on the carbonyl carbon forms the relatively stable acyl-enzyme intermediate. The nucleophilicity of the serine is increased and the reaction stabilised by the catalytic histidine (1, 65–68). The histidine itself is stabilised and oriented by the carboxyl group of the acidic residue (1, 65–68). In the second step, the acyl-enzyme tetrahedral intermediate that is covalently attached is attacked by a nucleophilic water molecule to regenerate the free enzyme and to liberate the acid portion of the substrate (37, 69). The water molecule in this step is activated by the histidine and the acid residue of the triad.
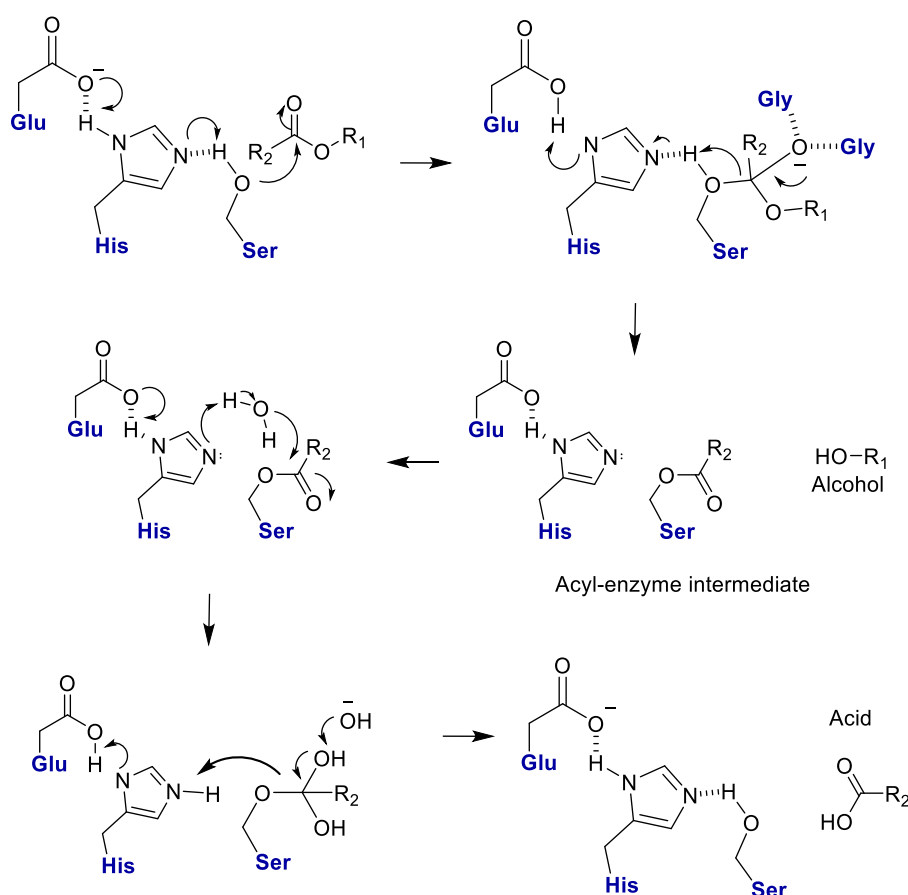
Figure 1.5. The mechanism of ester hydrolysis by CBEs. The essential catalytic amino acids are shown (Ser, His and Glu) and the mechanism depicts the hydrolysis of a generic ester. Figure adapted from Montella et al (14).

The glycine residues in the oxyanion hole are important in stabilising and lowering the energy of the tetrahedral transition state during both nucleophilic substitution reactions (1, 65–68). The tetrahedral transition state involves a charge being moved to the carbonyl oxygen and the nitrogens of the oxyanion hole help to stabilise the transition state by hydrogen bonding interactions (1, 65–68). This reaction mechanism is similar to those of serine proteases and is a form of convergent evolution as the structures are unrelated (32). Recently, it has been proposed that the mechanism involves four steps instead of two, based on computational studies (70).

The hydrolysis is energetically favourable and in some cases can be diffusion limited (57, 71–73). The low energy barrier also allows for this reaction to be easily reversible and in non-aqueous environments, esterification reactions are possible (74, 75). In this reaction, the alcohol molecule is in competition with the water molecules in the active site during the attack of the acyl-enzyme intermediate. This results in an ester being obtained instead of a carboxylic acid (70).

1.1.8. Insect Carboxylesterases

Insect esterases have been well studied in insect biochemical research for over 50 years and fall under the carboxyl/cholinesterase gene family (32). There is little sequence information regarding acetyl or arylesterases from invertebrates, with most research focusing on CBEs (76). CBEs are rich in the genomes of insects and have been investigated for their diverse functions in insecticide resistance (xenobiotic defence), lipid metabolism, reproduction and insect development (3, 43, 57, 77). Certain insect pheromones or hormones have been identified as being metabolized by CBEs, which has led to significant research on these enzymes (32). Insecticide detoxification of organophosphates, carbamates and pyrethroids has also been a major research area involving CBEs (32).

In *D. melanogaster* and other insect species, there appears to be between 30-76 different choline/carboxylesterase genes and a majority are catalytically active (38, 39, 43, 78–80). Most of the CBEs identified have no known physiological role (43, 44). The two main clusters of focus have been separated as the α and β clusters. The large cluster in *D. melanogaster* is known as the α-cluster with over 10 genes and has roles in xenobiotic metabolism (81). The second cluster is known as the β-cluster, which is much smaller and appears to be rapidly evolving (42). This cluster has been well studied for models of microevolution and contains two genes in *D. melanogaster* (42).

An important characteristic observed in insect species is the large number of duplications and tolerance to mutations in CBEs, which has resulted in the functional diversification of this family and a significant expansion of CBEs (43, 44, 50, 77, 79). With the functional diversity found, the sequence diversity is similarly distant with as little as 20% amino acid identity between members of the CBEs (43). It has suggested CBEs are rapidly evolving, as the enzymes are commonly associated with xenobiotic detoxification mechanisms in insect species (38, 39). Gene redundancy (from overlapping substrate ranges) and mutation tolerance are possible reasons for the rapid evolution of CBEs (43, 44, 50, 77, 79).

1.1.9. Vertebrate Carboxylesterases

CBEs in vertebrate species have been studied for their roles in the biotransformation of compounds and fall under the carboxyl/cholinesterase gene family (66). The CBEs are responsible for the pharmacokinetic behaviour of several drugs and interact with compounds such as heroin, cocaine and the anticancer drug CPT-11 (5, 64, 82). Similar to insect CBEs, structures of vertebrate CBEs are also lacking in the PDB. The major

structural representation of vertebrate CBEs is liver carboxylesterase 1, which has been solved in humans and rabbits (83, 84). Both human cholinesterase protein structures (acetylcholinesterase and butyrylcholinesterase) have also been solved and are well studied for their association with Alzheimer's disease and interactions with OP nerve agents (85–88). This is further discussed in chapter five.

1.2. Focus of the Thesis

This thesis will encompass five chapters that will cover α-esterases and β-esterases from a variety of species. The first three chapters include three separate projects on insect CBEs from dipteran species to understand oligomerization, evolution, structure, and function in CBEs. Chapter five discusses the development of inhibitors for vertebrate acetylcholinesterase.

There is a current lack of understanding regarding the structures of CBEs from different clades and structure-function relationships within the superfamily. Some questions that are unanswered include: do CBEs from different clades vary significantly in structure? Do possible small changes in structure affect the evolution of insect resistance? In this work, the structure of esterase-6 from Clade E will attempted to be solved to further understand the structures of CBEs and the variation within the 14 clades. This can also provide evidence for the function of esterase-6, which has been under investigation for several decades. To further learn about structure-function relationships in insect CBEs, the activity of dipteran CBEs will be characterised with an organophosphate compound in combination with homology model analysis to expand on the results. This will also help to understand if historical contingency playing a role in the evolution of insecticide resistance involving CBEs.

As well as structural analysis, the effect of quaternary structure on stability in insect CBEs is not understood in detail. By studying the formation of new oligomeric species from a directed evolution experiment on an insect CBE, the question if quaternary structure directly provides stability to proteins can be answered. Lastly, to further develop drugs to target the carboxyl/cholinesterase superfamily, two different types of potential inhibitors will be tested and analysed structurally against the enzyme, acetylcholinesterase. It is unknown how the two natural and modified inhibitors will interact with the protein, therefore assays and structural analysis is important to know how effective the new potential inhibitors are.

Chapter 2. Evolution of oligomeric structure in response to heat stress

2.1. Declaration of Author Contribution

The research article in this chapter was peer-reviewed and published as a regular original research article. I contributed to >50% of the work included in this paper: I wrote a major portion of the text and designed and carried out the majority of the experimental work. James M. Murphy was involved in the experimental work involving SAXS (Figure 2F), and Matthew A. Perugini contributed towards the AUC data (Figure 2E). Jian-Wei Liu was involved in the initial (previously published) directed evolution experiment along with Chris W. Coppin and John G. Oakeshott. The initial observation of the difference in oligomeric structure between the wild-type and variant proteins involved the authors Colin J. Jackson, Mathilde Lethier and Martin Weik. The contributions of the authors Jian-Wei Liu, Peter D. Mabbitt, Galen J. Correy, John G. Oakeshott and Colin J. Jackson were towards the writing of the paper.

2.2. Introduction

2.2.1. Oligomers

A large number of proteins consist of more than one polypeptide chain in their native state and the arrangement of these folded chains forms the quaternary structure **(Figure 2.1)**. Oligomeric proteins (two or more associating polypeptide chains) are vital and highly abundant in nature, comprising a significant percentage of cellular proteins (89–93). Oligomerization represents a fundamental strategy for generating complex protein structure, functional complexity and regulation in enzymes (94). This has led to protein-protein interactions and assemblies being the target of structural genomic initiatives to understand oligomers (94).
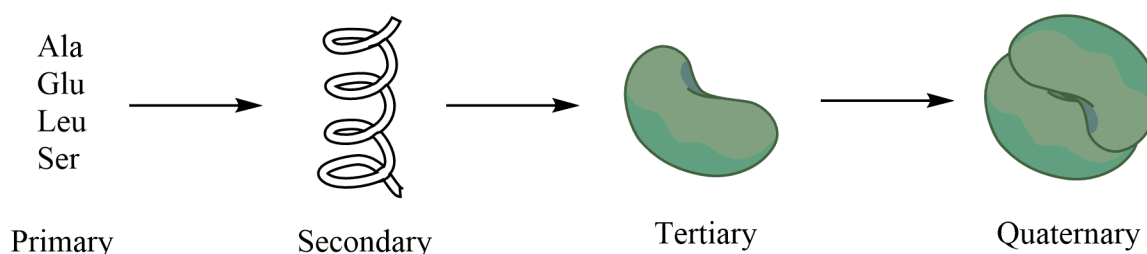
Figure 2.1. The four levels of protein structure. Primary (amino acid sequence), secondary (local sub-substructure), tertiary (three-dimensional structure of a monomeric protein), quaternary (three-dimensional structure of a multi-subunit protein) (95).

A majority of oligomers are composed of identical protein subunits known as homo-oligomers, based on a snapshot of protein crystal structures from the Protein Databank (PDB) and from *Escherichia coli* annotation (92, 96–98). In contrast, heteromeric complexes are composed of multiple different protein subunits and structures for these complexes are not well represented in the PDB (99). Higher-order oligomers are less frequently found and odd-numbered stoichiometries make up a small number of all oligomeric proteins (92, 100). Several explanations have been proposed to account for the large amount of homo-oligomeric proteins including self-attraction, symmetry, stability, foldability and evolutionary optimization arguments (101–103).

2.2.2. Why study oligomerization?

Oligomeric proteins are ubiquitous within the cell and are essential in most biological processes (104–106). Given this, it is important to understand how oligomeric protein assembly occurs, and factors affecting quaternary structure organisation (104–106). The study of oligomeric proteins can provide information about the early protein environment,

protein evolution and the general characteristics of an oligomeric protein (107). It is important to understand the structural and evolutionary principles of protein-protein interactions as many unfavourable interactions can result in disease (108, 109). With a greater understanding, the ability to predict and engineer oligomerization is possible. This has the potential for the development of new drugs to target complex assembly or induce disassembly (110–113).

2.2.3. Oligomer Evolution

Although oligomeric proteins are central to biology, the general mechanisms of oligomerization have not been fully understood and only anecdotal knowledge exists on the factors affecting oligomer evolution and assembly (107, 114). Investigations into the formation of new oligomeric species are difficult as the generation of new oligomeric states can expose hydrophobic surfaces, which in some cases, results in aggregation and solubilisation issues (107). However, it has been shown that soluble oligomerization can occur to form stable functional proteins and to protect proteins from aggregation (102, 115, 116). Currently, no work has focussed on the directed evolution of a protein to a new oligomeric state (117). There are several proposed mechanisms for the evolution of new oligomeric species and it is likely that oligomerization in different proteins have evolved by alternative routes **(Figure 2.2)** (107, 118). Some proposed mechanisms of oligomer evolution are (1) amino acid substitutions, (2) domain swapping, (3) insertions and deletions, (4) gene duplication and (5) gene fusion.
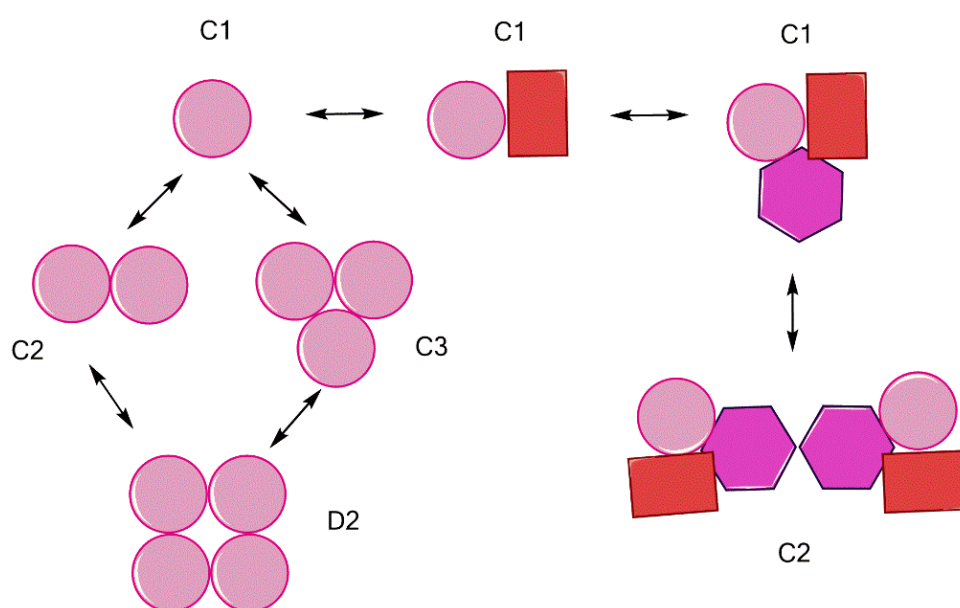
Figure 2.2 An example of the multiple pathways for the evolution of new oligomers. Changes in oligomers can be through point mutations or by the gain of new subunits. Symmetry point groups are given for each oligomer. Figure adapted from Marsh et al (99).

(1) Amino acid substitutions. Point mutations introduced on the protein surface or interface can cause the association of oligomers (119–121). The mutations can either be located directly in an interface or located away from the interface exerting an indirect effect **(Figure 2.3)** (119–121). The replacement of surface residues to more hydrophobic and/or larger residues can favour the formation of oligomers (121, 122). It is difficult to distinguish an interface from the surface based solely on amino acid composition, and as few as two surface substitutions can generate a new oligomeric interface (121, 122). This shows that protein surfaces are similar to interfaces based on amino acid composition alone (122–125).
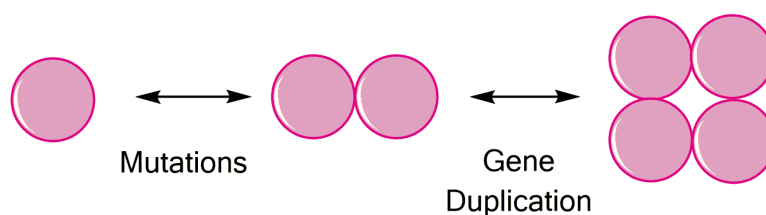
Figure 2.3. The introduction of surface mutations can give rise to new oligomer structure and gene duplication results in further changes in the oligomerization state. Figure adapted from Hashimoto et al (126).

(2) Domain Swapping. Oligomers can occur naturally or artificially through the exchange of small or large regions **(Figure 2.4)** (127). In some cases, entire domains of monomeric subunits can be exchanged (127). This mechanism is known as three-dimensional domain

18

swapping where tertiary structure exchanges with another copy of itself (128). 24% of homo-oligomers have been identified as having domain swapped structures (129).
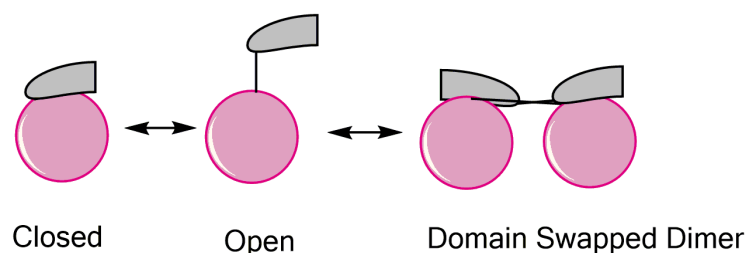
Figure 2.4. An example of three-dimensional domain swapping with two monomeric proteins exchanging secondary structure. The opening of the loop in the monomer allows for the formation of the dimer. Figure adapted from Gotte et al (130).

(3) Insertions and deletions. In a single step, the insertion of a short loop can introduce a new interface (118, 131–133). This is an important mechanism to control oligomeric states in protein evolution (126). The removal of insertions has been shown to result in a complete or partial loss of oligomer stability (118, 131, 132).

(4) Gene duplication. Protein complex evolution can occur through the gain of subunits by gene duplication events **(Figure 2.3)** (115, 134–137). If a gene encoding an oligomeric protein is duplicated, at first the two identical genes will still form the same oligomer (99). However, as one copy of the gene retains its original function, the other gene copy will be under relaxed evolutionary constraints, allowing it to develop a new specific function. Eventually, this mechanism will lead to two dissimilar genes and the formation of oligomeric paralogs (115, 132, 137–139).

(5) Gene fusion. Oligomerization can arise by gene fusion when two separate genes become fused into one open reading frame **(Figure 2.5)** (140). For example, a gene encoding an oligomerization domain can fuse with a previously monomeric protein, forming a new oligomeric protein (107).
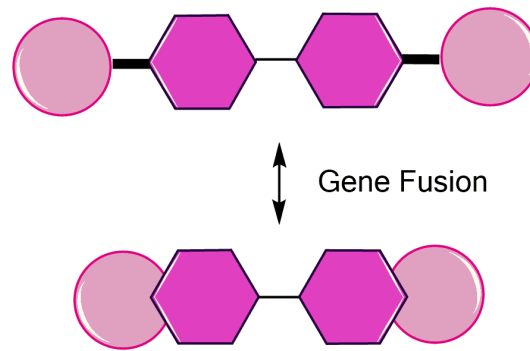
Figure 2.5. An example of gene fusion where a new covalent association between two separate subunits of the same complex is formed. Figure adapted from Marsh et al (99).

2.2.4. Features of oligomerization

As oligomeric proteins are important in numerous pathways including gene expression, ion channels, and cell-cell adhesion processes, the question of how oligomeric proteins evolved is a fundamental question which is still being explored (98, 107, 141). There are several advantages oligomerization offers compared to a monomeric protein that can be selected for during evolution. Currently, there is no single theory to explain how a specific oligomer has emerged and it has been suggested that oligomeric structure could occur by chance (114).

2.2.5. General characteristics of oligomerization.

(**1**) Complex scaffolds can result in the better support for function or functional gain (142–144). For example, a new interface can be directly involved in protein function (142–144). This is commonly found in heteromeric proteins where distinct functions are brought together (92).

(**2**) Allostery introduces an additional level of control and is commonly associated with oligomeric proteins. Oligomerization can provide sites for cofactors to interact with, resulting in substrate-induced cooperativity (**Figure 2.6**) (145, 146).
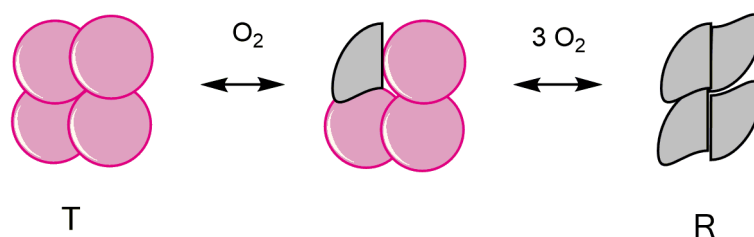
Figure 2.6. Allosteric control between two states (T, R) associated with oligomerization is observed in haemoglobin induced by oxygen binding (145).

(3) Oligomerization can result in a protein being encoded using less genetic material. This represents a form of coding economy, as larger proteins can be formed without increasing the genome size (147).

(4) The increase in oligomerization state results in larger proteins that are typically more resistant to degradation and denaturation (92, 116, 125, 148). This is due to a reduced surface-to-volume ratio in oligomeric proteins than that of free monomers. Oligomerization can increase stability in this manner and also offers protection against denaturation (92, 116, 125, 148).

2.2.6. Oligomer Equilibrium and Regulation

While new oligomers can evolve, the formation of a new oligomer can be induced by ligands, posttranslational modifications or by changes in physiological conditions (pH, ionic strength, temperature and protein concentration) (98, 149). The strength and duration of an association between subunits can vary significantly (107). Oligomeric proteins can be found primarily in one higher oligomeric state or can weakly associate and fluctuate between several states (107). This can serve a regulatory mechanism by changing oligomeric states in response to a stimulus (98, 149–151).

2.3. Research Article

2.3.1. Preface

There are several suggested mechanisms of how new oligomers can form and the functional or stability advantages that arise from forming a new oligomeric species. However, direct evidence for this is missing in the literature. Previous evidence comes from the comparison of ancestry proteins to modern day proteins and few studies have directly validated stability with newly formed oligomeric species. It has also been suggested that oligomerisation can arise during evolution neutrally and offer no benefits. Investigations into evolutionary experiments involving oligomeric proteins can provide the necessary evidence to refute this claim. In the carboxylesterase αE7 from the Australian sheep blowfly (*L. cuprina*), a directed evolution experiment for increase thermal stability resulted in a variant with increased amounts of unknown oligomeric species compared to the wild type protein (55). The following article discusses the characterisation of the identified oligomeric species and the nature of the stabilisation of αE7 in the directed evolution experiment.

2.3.2. Published Research Article: Evolution of Protein Quaternary Structure in Response to Selective Pressure for Increased Thermostability

Supplementary Information is found in Appendix A
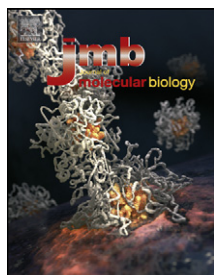
Evolution of Protein Quaternary Structure in Response to Selective Pressure for Increased Thermostability

Nicholas J. Fraser[1], **Jian-Wei Liu**[2], **Peter D. Mabbitt**[1], **Galen J. Correy**[1],
Chris W. Coppin[2], **Mathilde Lethier**[3], **Matthew A. Perugini**[4], **James M. Murphy**[5,6],
John G. Oakeshott[2], **Martin Weik**[3] and **Colin J. Jackson**[1]

1 - Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia
2 - Commonwealth Scientific and Industrial Research Organization, Land and Water Flagship, Canberra, ACT 0200, Australia
3 - Institut de Biologie Structurale, Commissariat a l'Energie Atomique, F-38027 Grenoble, France
4 - Department of Biochemistry, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, VIC 3086, Australia
5 - Cell Signalling and Cell Death Division, The Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, VIC 3052, Australia
6 - Department of Medical Biology, University of Melbourne, Melbourne, VIC 3050, Australia

Correspondence to Colin J. Jackson: cjackson@rsc.anu.edu.au
http://dx.doi.org/10.1016/j.jmb.2016.03.014
Edited by S. A. Teichmann

Abstract

Oligomerization has been suggested to be an important mechanism for increasing or maintaining the thermostability of proteins. Although it is evident that protein–protein contacts can result in substantial stabilization in many extant proteins, evidence for evolutionary selection for oligomerization is largely indirect and little is understood of the early steps in the evolution of oligomers. A laboratory-directed evolution experiment that selected for increased thermostability in the αE7 carboxylesterase from the Australian sheep blowfly, *Lucilia cuprina*, resulted in a thermostable variant, LcαE7-4a, that displayed increased levels of dimeric and tetrameric quaternary structure. A trade-off between activity and thermostability was made during the evolution of thermostability, with the higher-order oligomeric species displaying the greatest thermostability and lowest catalytic activity. Analysis of monomeric and dimeric LcαE7-4a crystal structures revealed that only one of the oligomerization-inducing mutations was located at a potential protein–protein interface. This work demonstrates that by imposing a selective pressure demanding greater thermostability, mutations can lead to increased oligomerization and stabilization, providing support for the hypothesis that oligomerization is a viable evolutionary strategy for protein stabilization.

Introduction

Oligomerization is one of the most fundamental biophysical interactions in protein chemistry. Bioinformatics and pull-down experiments have revealed that a large number of proteins exist as homooligomers consisting of two or more identical chains [1–4]. Similarly, many proteins function in the cell as hetero-oligomers composed of non-identical chains [5–9]. Oligomerization has been shown to play important roles in the genetic economy [10], functional gain [11–13], structural stability [14,15], allosteric regulation [16–19], and protection from degradation [17]. Given the biological importance of oligomerization, there is substantial interest in understanding how the evolution of new oligomeric species occurs [20,21], engineering new oligomeric structure [22,23], and developing drugs targeted at complex assembly and disassembly [24,25].

The cellular milieu is a precarious environment for the evolution of oligomeric proteins as the physical forces that drive beneficial protein association are the same as those that drive deleterious aggregation [26,27]. It is well established that changes in protein sequence through non-synonymous point mutations, insertions, and deletions can shift the balance of oligomeric states [21,28–30]. Both rational mutagenesis and directed evolution have been exploited to advance our understanding of how oligomers form [31–34] and to design new hetero-oligomers [35,36].

Protein flexibility, shape, and symmetry have all been identified as being important for the formation of new oligomeric structure [37,38], since symmetrical and complementary interfaces may form stronger interactions than heterologous surfaces [39–41]. Despite a growing understanding of the nature of oligomers and protein complexes, it is difficult to predict how a new and beneficial protein–protein interface will develop, let alone design one *de novo* [23,42,43]. Moreover, despite intense study, *de novo* evolution of oligomeric structure has not been observed directly, to the best of our knowledge.

The fundamental importance of stability in protein evolution and the stabilizing contribution of oligomerization (*via* reduction of the surface-to-volume ratio of the complex compared to a free monomer) are well established [44,45,2]. Protein–protein interactions can provide stabilizing polar or hydrophobic contacts (although interfaces in homodimers predominantly involve polar interactions), leading to tighter molecular packing and offering protection from denaturation [2,46]. Indirect evidence for the stabilizing effects of protein–protein interactions comes from thermophilic Archaea and bacteria, where oligomerization has been suggested to be one of the contributing factors to the high thermostability of proteins in these organisms [47,48]. Other examples of oligomerization leading to increased stability come from studies in which the disruption of protein–protein interfaces leads to decreased thermostability or enzymatic activity [14,49] or the engineered formation of oligomers leads to increased stability [50,51]. However, despite several lines of evidence that oligomerization increases protein stability, most of the evidence for an evolutionary role for oligomerization had been indirect until the recent work of Perica et al., who used ancestral protein reconstruction to identify specific mutations directly involved in the process [20]. Notably, many of the mutations were found to be remote from the protein–protein interface.

LcαE7 is a carboxylesterase involved in organophosphate insecticide resistance in the sheep blowfly, *Lucilia cuprina* [52]. We have previously reported an experiment in which we performed directed evolution of LcαE7 in order to stabilize the protein for crystallization. However, there was little analysis of the evolution of thermostability, beyond the observation that after four rounds of evolution a variant (D83A, M364L, I419F, A472T, I505T, K530E, and D554G) was obtained that readily crystallized and displayed enhanced thermostability [53]. In this work, we have focused on the process by which the thermostable variant evolved and the nature of the stabilization. Here, we demonstrate that stabilization of the protein occurred *via* two related routes: stabilization of the monomeric protein through improved side-chain packing, and through the enrichment of more stable dimeric and tetrameric species. Oligomerization as a route to thermostability has been inferred through analysis

of extant proteins and through ancestral reconstruction [20], but seldom has *de novo* evolution of oligomerization been observed. These results allow us to understand the first steps in this process, which will benefit future engineering efforts.

Results

Directed evolution of LcαE7 for increased thermal stability

As reported previously, the wild-type (WT) LcαE7 protein was unstable both during and after purification [53]. Thus, the original rationale behind the design of this directed evolution experiment was to extend the half-life of LcαE7 at temperatures that it might experience during expression and purification to facilitate crystallization. To achieve this, we designed a medium throughput screen in which LcαE7 was heterologously expressed in *Escherichia coli*, replica plated onto filter paper, and assayed colorimetrically for activity using 2-napthyl acetate and fast-red dye. By randomly mutating the LcαE7 gene in an expression vector and incubating the replica-plated *E. coli* transformed with the mutant library on filter paper for up to 1 h at elevated temperatures, we were able to iteratively select more thermostable variants and progressively increase the thermostability of the protein. After one round of random mutagenesis, only marginal improvements in activity over that of the WT enzyme were evident. Accordingly, these variants were not sequenced, but potentially improved variants were pooled for further random mutagenesis (which includes some recombination [54]) ahead of three further rounds of evolution. By round 2, significant increases in thermostability

Table 1. Mutations incorporated into LcαE7 during laboratory-directed evolution for thermostability

	Round of evolution and mutation						
	2a	2b	2c	2d	3a	3b	4a
Residue							
Asp83							Ala
Ala285			Ser				
Met364							Leu
Ile419				Phe	Phe		Phe
Ala472		Thr	Thr	Thr	Thr	Thr	Thr
Phe478	Leu						
Ile505						Thr	Thr
Lys530		Glu				Glu	Glu
Asp554							Gly

Mutations were not recorded in the first round of mutagenesis. In round 2, four variants (2a, 2b, 2c, and 2d) were selected for DNA shuffling and mutagenesis resulting in round 3 variants. In round 3, two variants (3a and 3b) were selected for DNA shuffling and mutagenesis resulting in the round 4a variant.

were apparent, allowing specific mutants to be isolated and sequenced (Table 1). In round 2, four variants (2a, 2b, 2c, and 2d) were selected for DNA shuffling and mutagenesis resulting in round 3 variants. In round 3, two variants (3a and 3b) were selected for DNA shuffling and mutagenesis resulting in the round 4a variant. The most stable variant from the experiment was LcαE7-4a (Fig. 1a), which incorporates three mutations first observed in round 2 (I419F, A472T, K530E), one mutation from round 3 (I505T), and three mutations from round 4 (D83A, M364L, D554G). The mutation D83A was accidently missed in the previously reported structure of this protein [49]. Subsequent DNA sequencing and analysis of electron density maps (5IKX, 5CH3) of LcαE7-4a have confirmed that this mutation is present.

The LcαE7-4a variant was then characterized in detail. We found that its half-life (time at which 50% of original activity remains) at 40 °C was indeed much greater than that of the WT enzyme (>60 min versus 1.6 min), indicating that some of the mutations in LcαE7-4a stabilize the protein (Fig. 1a). Notably, whereas the activity of the WT protein

rapidly diminished to background levels (7.5% of the initial activity) with a half-life of 1.6 min, consistent with a single population of enzymes with similar thermostability, the activity of LcαE7-4a decayed with a half-life of 24 min, but extrapolates to 56% of its original activity when fit to a one-phase exponential decay curve. Thus, LcαE7-4a appears to be heterogeneous, consisting of one species with a half-life of 24 min, and other species that are essentially stable at 40 °C.

To investigate the composition of LcαE7-4a in greater detail, we performed size exclusion chromatography (SEC), with kinetic and SDS-PAGE analysis of each fraction (Fig. 1b). These analyses revealed that, other than minor contamination by a protein of approximately 70 kDa in fractions 12–18, LcαE7-4a comprised the majority of the protein in the sample. However, the protein quaternary structure was not homogeneous. In fact, there were four distinct species present in the sample: a high-molecular weight (HMW) species, which eluted near the void volume of the column, and three species relatively close in elution volume which primarily eluted in fractions 22, 24, and 26, respectively.

Characterization of the oligomeric species

The SEC was repeated several times, using a range of protein concentrations (5–20 mg/mL). LcαE7-WT was consistently observed to elute primarily as a monomer, with small amounts of an HMW species (~120 mL) and small shoulders (170 mL, 185 mL) on the primary monomer peak (~215 mL) (Fig. 2a–c). In contrast and at all concentrations, LcαE7-4a eluted with a significantly greater proportion of larger species eluting at ~170 mL and 185 mL. The peak UV absorbance of these fractions relative to that of the monomer in LcαE7-4a was approximately double those found for LcαE7-WT. Because the cytosol of the cell, where LcαE7-WT and LcαE7-4a were expressed and assayed during the laboratory evolutionary selection, is a viscous environment [55], we repeated these experiments in the presence of 10% glycerol. This revealed that in a more viscous solution, the tetrameric and dimeric HMW species comprised approximately half of the total LcαE7-4a protein, whereas for LcαE7-WT, the majority of the protein was monomeric (SI Fig. 1).

We next performed SEC–multi-angle laser light scattering (SEC–MALLS) analysis to obtain more accurate estimates of the absolute molecular weight of the various LcαE7 forms [56]. This analysis indicated that the three lower-molecular-weight species, which are composed exclusively of LcαE7 protein (Fig. 1b), are approximately 60 kDa, 120 kDa, and 260 kDa (Fig. 2d). These molecular masses correspond to the theoretical molecular masses of LcαE7 monomer (66.3 kDa), dimer

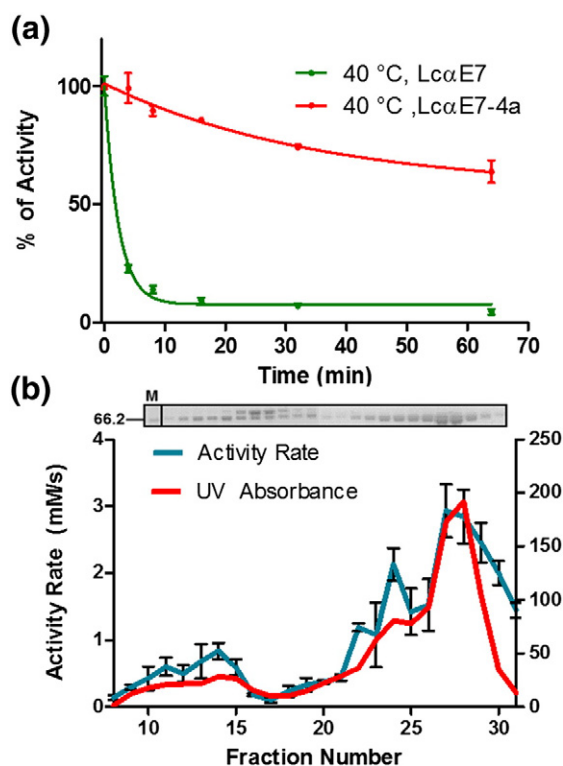**Fig. 1.** Directed evolution of LcαE7 for enhanced thermostability. (a) When incubated at 40 °C over a time-course, there was a one-phase exponential decay curve of activity in the WT protein, and of the LcαE7-4a protein (b) SEC detected the presence of LcαE7 in four main peaks, fraction 12, fraction 22, fraction 24, and fraction 28. A contaminating band was present, represented by a peak in fraction 14.

Fig. 2. Characterization of the oligomeric species observed in this work. (a and b) Comparison between size exclusion chromatographic spectra of WT LcαE7 (a) and LcαE7-4a (b) reveals that directed evolution has resulted in substantial increases in the amounts of dimeric and tetrameric species (c). Monomer (M), dimer (D), and tetramer (T) peaks are labeled. Each line in the graphs (a and b) corresponds to an independent purification. (d) SEC–MALLS analysis reveals the three species of LcαE7-4a correspond to monomer, dimer, and tetramer stoichiometry. (e) AUC sedimentation velocity analysis of LcαE7-4a at an initial concentration of 3.0 mg/mL. The continuous sedimentation coefficient [$c(s)$] distribution is plotted as a function of the standardized sedimentation coefficient. Analysis was performed using a resolution of 200 with a sedimentation coefficient range of 0–20 S and a P-value of 0.95. The nonlinear least squares fit shown resulted in an rmsd of 0.00585 and a runs test Z of 4.61. Inset: residuals from the $c(s)$ distribution best fit plotted as a function of radial distance from the axis of rotation. (f) Small angle X-ray scattering analysis of the HMW species reveals it to be a heterogeneous aggregate.

(132.6 kDa), and tetramer (265.2 kDa). We then performed analytical ultracentrifugation (AUC) sedimentation velocity experiments on the complex sample of LcαE7-4a (Fig. 2e). These results confirm that LcαE7-4a exists as a mixture of monomeric species and high-order oligomeric species, with the majority of the protein sedimenting as monomer

(66 kDa versus 66.3 kDa theoretical molecular mass) but a significant amount forming dimers (137 kDa versus 132.6 kDa theoretical molecular mass) and tetramers (230 kDa versus 265.2 kDa theoretical mass).

Whereas the stoichiometry of the ~260, ~136, and ~66 kDa species is readily apparent, the HMW
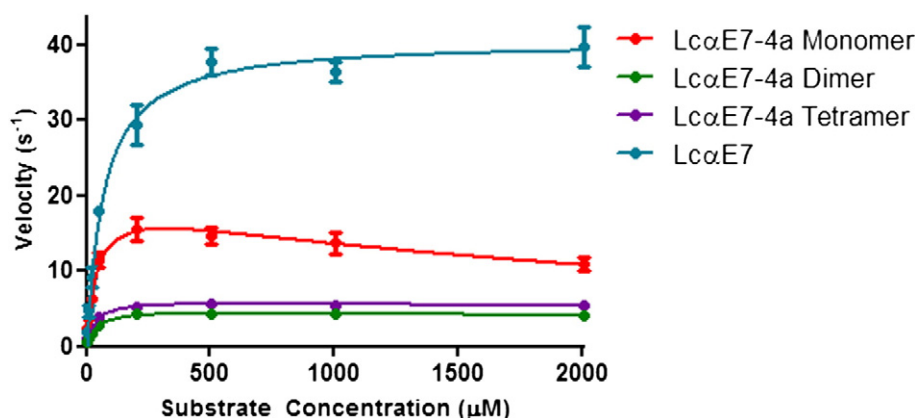
Fig. 3. Characterization of the oligomeric species. Rate versus concentration plots for the catalysis of 4-nitrophenyl butyrate hydrolysis by WT LcαE7 and LcαE7-4a. All curves were fit to the Michaelis–Menten equation $v = (k_{cat} \cdot [S])/(K_M + [S])$, with the exception of LcαE7-4a monomer, which was fitted to the equation modified for substrate inhibition $v = (k_{cat} \cdot [S])/((K_M + [S]) \cdot (1 + [S]/K_{si}))$.

Table 2. Kinetic parameters for the oligomers of LcαE7 with the model carboxylesterase substrate 4-nitrophenyl butyrate

	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})	K_i (mM)
LcαE7	68.6 ± 12.3	40.6 ± 2.2	6.0 × 10^5 ± 1.3 × 10^5	—
LcαE7-4 monomer	38.1 ± 8.3	14.5 ± 1.1	3.8 × 10^5 ± 1.6 × 10^5	2.6 ± 0.8
dimer	30.4 ± 4.8	4.5 ± 0.2	1.5 × 10^5 ± 0.2 × 10^5	—
LcαE7-4 tetramer	35.6 ± 4.9	5.7 ± 0.3	1.6 × 10^5 ± 0.3 × 10^5	—

Kinetic data for LcαE7, LcαE7-4 dimer, and LcαE7-4 tetramer were fit to a model for standard Michaelis–Menten kinetics ($v = (k_{cat} \cdot [S])/(K_M + [S])$), whereas kinetic data for LcαE7-4 monomer were fit to a Michaelis–Menten kinetic model incorporating substrate inhibition ($v = (k_{cat} \cdot [S])/((K_M + [S]) \cdot (1 + [S]/K_{si}))$).

species (~400 kDa) was more difficult to analyze. We performed small angle X-ray scattering on this sample in order to determine, at low resolution, its molecular structure (Fig. 2f). These results indicate that this species is very heterogeneous, consistent with HMW protein aggregates that likely formed during recombinant expression and purification. The observation that the HMW peak is present in the WT sample at similar abundance to the LcαE7-4a samples suggests that this species is not directly relevant to the evolutionary stabilization of the protein. Accordingly, it was not considered further in this study.

We then investigated whether the dimeric and tetrameric oligomeric species were representative of the protein in an equilibrium state, or were artifacts of the purification and concentration process. We obtained single 5-mL fractions from the center of the monomer and tetramer peaks from SEC and then performed SEC again after approximately 5 h at 4 °C (monomer) or 24 h at 4 °C (tetramer) (SI Fig. 2). We found that the tetramer had equilibrated into approximately equal distributions of monomer, dimer, and tetramer, while the monomer had also equilibrated to form dimeric and tetrameric species. Thus, the tendency of LcαE7-4a to self-associate is a genuine equilibrium property.

The effect of oligomerization on enzyme activity and thermostability

Kinetic analyses of each species from LcαE7-4a (Fig. 3 and Table 2) showed that the monomeric, dimeric, and tetrameric forms were all significantly less active than the WT enzyme with the model substrate 4-nitrophenyl butyrate. However, the dimeric and tetrameric species were kinetically distinct from the monomer. Interestingly, the monomeric LcαE7-4a species exhibited significant substrate inhibition (K_i), which was not apparent in either the dimeric or tetrameric species or the monomeric WT species. When the activity of the tetrameric and monomeric species was assayed at 46 °C over a time course (SI Fig. 3), the tetrameric species retained greater activity, after 1 h (the monomeric species was almost fully inactivated). Given that the evolutionary selection was used to identify mutants that could retain activity after a 1-h incubation at elevated temperature, this suggests that the greater thermostability of the higher-order species was selected for and made a significant contribution to the global enhancement of LcαE7 thermostability.

Circular dichroism (CD) spectroscopy was used to investigate the thermostability of the three LcαE7-4a species (Table 3; SI Fig. 4). None of these species underwent reversible folding, meaning that accurate estimates of the free energy of folding energy could not be obtained. Nevertheless, the thermal melting curves provide an appropriate measurement of the temperature at which the majority of secondary structure in these proteins was lost (unfolding). All three LcαE7-4a species were significantly more stable that the WT protein. As was the case with the catalytic activity of these species, the thermostability was different between the monomer, dimer, and tetramer. Given the propensity of the higher-order species to slowly re-equilibrate (SI Fig. 2), we were unable to obtain completely homogeneous samples, although those sampled were analyzed within 2 h of SEC (after buffer exchange). This was apparent in the monomer and dimer samples, where the use of a three-state model showed that separate unfolding transitions corresponding to the monomer (55.5 °C) and dimer (66.2 °C) were apparent (Table 3; SI Fig. 4). The tetrameric sample showed a single transition at ~65 °C. Thus, the higher-order species are significantly more thermostable, consistent with the results from the activity decay measurement (SI Fig. 3).

Table 3. Calculated transition temperature (TT$_{50}$) for the LcαE7 proteins as measured by temperature-ramp CD

Protein	TT$_{50}$ (°C)
LcαE7	48.8 ± 0.2
LcαE7-4a monomer	55.5 ± 0.5, 66.2 ± 0.6
dimer	56.3 ± 0.6, 61.7 ± 0.9
LcαE7-4a tetramer	65.2 ± 2.0

Ellipticity was recorded at 208 nm and fitted by non-linear regression to two-state model (LcαE7 monomer and LcαE7-4a tetramer) and a three-state model (LcαE7-4a monomer and LcαE7-4a tetramer. The best fitting model was chosen based on the sum of residuals. Raw data is plotted in SI Fig. 2.

We repeated these thermostability measurements using differential scanning fluorimetry (DSF) [57]. These results showed that, whereas a complex mixture (pre-SEC) of LcαE7-WT appeared to comprise a single dominant species that permitted the use of a two-state unfolding model, the complex mixture of LcαE7-4a consisted of two species, with one having a significantly higher melting temperature (SI Fig. 5). When monomeric and tetrameric fractions from SEC (0.3 mg/mL) were immediately analyzed by DSF (the dimeric fraction was found to comprise a mixture of both species), we could obtain pure samples with a single transition temperature. This confirmed that the higher-order tetramer fraction was more stable by ~ 5 °C. After concentration to 3 mg/mL, some re-equilibration occurred and both tetrameric and monomeric samples showed two transitions, one corresponding to monomer and the other corresponding to tetramer (SI Table 1; SI Fig. 6). Overall, the melting temperatures were slightly lower with DSF than with CD (although the results were qualitatively identical), which can be explained by local exposure of hydrophobic surface (enabling binding of the fluorescent dye) before complete loss of secondary structure.

Structural comparison between monomeric and dimeric LcαE7-4a

We have previously reported the structures of monomeric and dimeric LcαE7-4a [53]. Monomeric LcαE7-4a was crystallized in the $P2_1$ space group (5IKX), with two molecules in the asymmetric unit. However, there was no close association reminiscent of a dimer/tetramer interface between the two molecules, or between them and any of the other molecules in the crystal lattice *via* symmetry (Fig. 5). In confirmation of this, the PISA algorithm did not detect any potential association interfaces or any potential assemblies [58]. In contrast, the LcαE7-4a protein crystallized in the $C222_1$ space group (5CH3) had a monomer in the asymmetric unit, yet formed extensive, symmetrical, interactions with a neighboring molecule that were visually consistent with a typical oligomeric interface (Fig. 5a). Analysis using PISA confirmed that this structure is likely to be a dimer under physiological conditions: 2260 Å^2 of protein surface areas is buried within the dimer interface. A total of 16 hydrogen bonds and 8 salt bridges were observed between the two-monomer chains (Fig. 5). The monomeric ($P2_1$) and dimeric ($C222_1$) structures were very similar in terms of bother overall topology (C-α RMSD 0.339 Å) and the main chain conformation at the dimer interface. Thus, in the case of LcαE7-4a, dimerization does not require any large-scale rearrangement.

Interestingly, none of the mutations observed in LcαE7-4a are located at the dimer interface (Fig. 5a and b). However, three of the mutations are located on the protein surface near regions that are involved in the dimer interface and display some conformational disorder, as indicated by their atomic *B*-factors: D83A is located at the rear of the protein, M364L is located at a tight turn between two α-helices that overhang the dimer interface and active site entrance, D554G is located on an α-helix that is adjacent to one of the α-helices that comprise a significant part of the dimer interface. K530E is located at the rear of the protein in another comparatively disordered region.
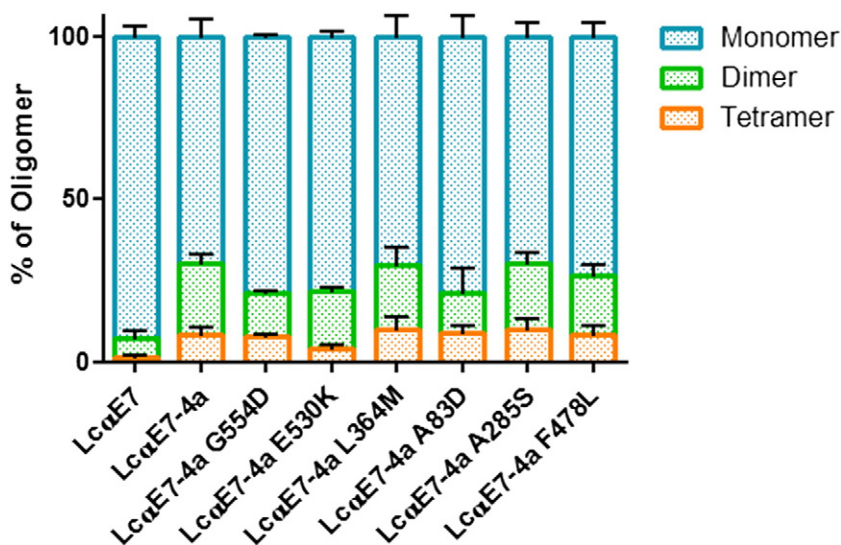
Fig. 4. Proportion of oligomeric species in the reverse mutations LcαE7-4a (A83D LcαE7-4a, L364M LcαE7-4a, E530K LcαE7-4a, G554D LcαE7-4a) and two mutations found in round 2 of the directed evolution (A285S LcαE7-4a and F478L LcαE7-4a) as estimated by peak heights obtained by SEC. Standard errors are shown for each species.

The effect of amino acid substitutions on oligomerization and thermal stability of LcαE7-4a

To better understand the effects of the four surface mutations (D83A, M364L, K530E, and D554G) present in the thermostable LcαE7-4a protein, they were individually reverted to the WT state. Purified proteins were subject to SEC in order to investigate the formation of oligomeric species (Fig. 4). The A83D and L364M reversion mutations had no discernible effect on the oligomeric equilibrium. The G554D reversion mutation resulted in a loss of dimer, which suggests that it may stabilize the adjacent α-helix that participates in the dimer interface, but had no effect on the proportion of tetrameric species. This observation is consistent with recent work that has shown that mutations can alter the oligomeric equilibrium of proteins through long-range effects on mobile regions (protein dynamics) [20]. The E530K mutation had no effect on the amount of dimer that was formed, but resulted in a significant reduction in the abundance of tetramer. In the X-crystal structure of LcαE7-4a, E530 is located in a putative tetramerization interface (Fig. 5c).

A salt bridge was observed between E530 and K104 from the adjacent chain; this salt bridge would be absent from the WT protein and may act to stabilize the LcαE7-4a tetramer. Relative to other amino acids, a large proportion of lysine residues are present on protein surface of LcαE7. Lysine residues are less frequently found at interfaces of oligomeric proteins [59], and removal of lysine residues has been shown to increase the probability of protein crystallization [60], which is a controlled form of aggregation. Thus, it is likely that the K530E mutation is at least partially responsible for formation of the tetrameric species and the increased propensity of LcαE7-4a to crystallize in the C222$_1$ space group.

The effects of the surface mutations on the thermostability of LcαE7-4a were measured by CD (Table 4). The four surface mutations in LcαE7-4a were individually returned to the WT state. The charge reversal mutation, E530K, was the only mutation to significantly affect the thermostability of the LcαE7-4a monomer, while the other three surface mutations A83D, L364M, and G554D had no significant effect on the thermostability of LcαE7-4a. This suggests that the
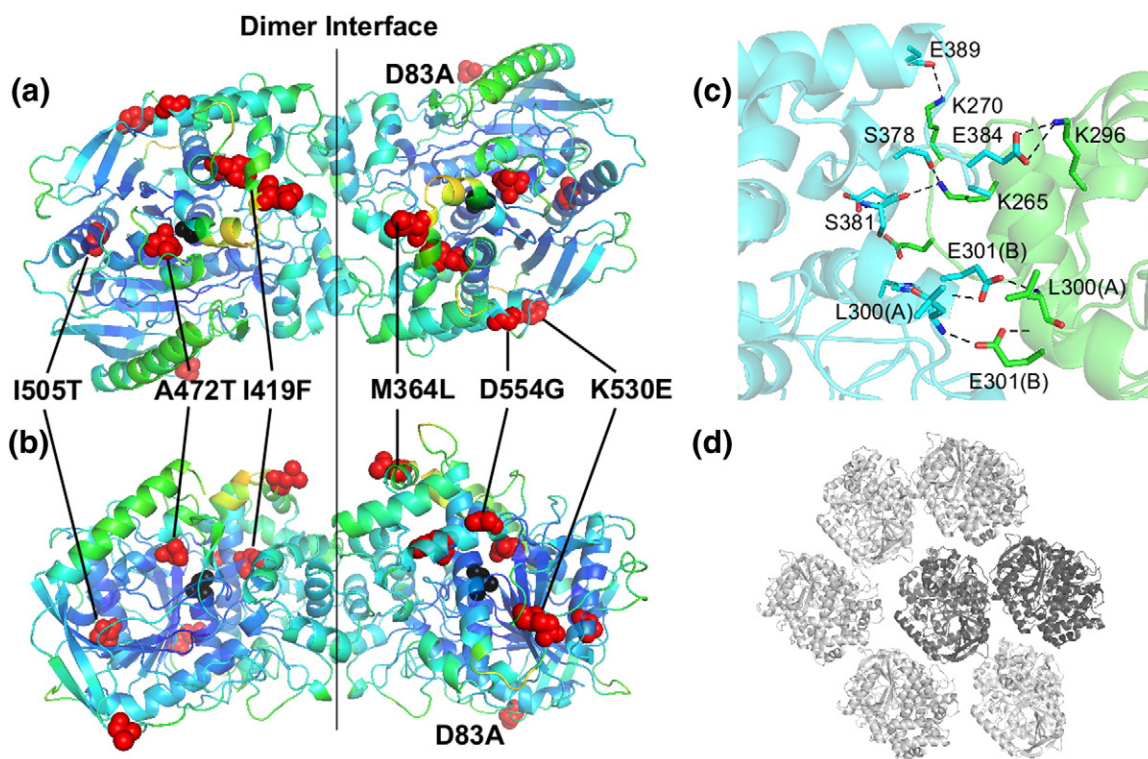
Fig. 5. (a and b) The structure of dimeric LcαE7-4a (5CH3), showing the location of the dimer interface and the location of the mutations. The surface mutations D83A, M364L, K530E, and D554G are shown on the right, while the three internal mutations I505T, A472T, and I419F are shown on the left. The catalytic serine (S218) is shown in black to indicate the location of the active site. B is rotated 90° relative to A. (c) A number of salt bridges and hydrogen bonds are formed at the dimer interface. Monomer A is shown in cyan and monomer B in green. (d) The comparatively loose crystal packing interactions of the P2$_1$ crystal form of LcαE7-4a (5IKX). The two molecules in the asymmetric unit are colored dark gray; symmetry mates, light gray. Other symmetry-related molecules are omitted for clarity.

Table 4. Calculated transition temperature (TT$_{50}$) for the LcαE7-4a proteins variants with reversed surface mutations for thermostability as measured by temperature-ramp CD

Protein	TT$_{50}$ (°C)
LcαE7-4a	55.5 ± 0.5
LcαE7-4a A83D	55.7 ± 0.4
LcαE7-4a L364M	56.5 ± 0.6
LcαE7-4a E530K	53.9 ± 0.2
LcαE7-4a G554D	56.3 ± 0.8

Ellipticity was recorded at 208 nm and fitted to non-linear regression to a three-state model (Eq. (1)).

K530E mutation along with at least one of the three internal mutations (I419F, T472A, and I505T) plays a role in the stabilization of monomeric LcαE7-4a protein, relative to LcαE7. We also investigated the A285S and F478L mutations that occurred in round 2 mutants 2a and 2c, but did not fixate during the evolution. In terms of both oligomerization (Fig. 4) and thermostability (SI Fig. 7), these mutations were selectively neutral, explaining their disappearance by round 4.

Stabilizing and oligomerization-inducing mutations are found in fly species closely related to *L. cuprina*

To determine if the amino acid substitutions we observed in this experiment also occur in orthologs of LcαE7, we performed a multiple sequence alignment of α-esterase 7 genes from four closely related species of flies: *Calliphora stygia* (New Zealand brown blowfly), *Cochliomyia hominivorax* (New World screw-worm fly), *Musca domestica* (house fly), and *Haematobia irritans* (horn fly) (SI Fig. 8). Interestingly, many of the mutations that fixated during this laboratory evolution experiment to produce LcαE7-4a are commonly occurring amino acid differences among these sequences. Three of them, I419F, T472A, and I505T, occur among the five species' sequences exactly as we find them in LcαE7-4a and variation is also found at three of the other sites, albeit not to exactly the same amino acid D83(E), M364(A/E/H), and D554(E) in those cases. Only K530 is strongly conserved across the five species. Thus, several of the mutations that we observed in our directed evolution experiment are also present in orthologous enzymes.

Discussion

In this work, we have evolved an enzyme with a greater propensity to adopt higher-order molecular structure. In practical terms, the evolution of higher-molecular-weight species could be advantageous to the possible application of LcαE7 as a biosca-venger to treat individuals poisoned by organophosphates, since it is known that proteins less than ~60 kDa are cleared through glomerular filtration [52,61]. The use of protein engineering or directed evolution to produce proteins with enhanced thermostability is perhaps one of the most common applications of these techniques: several high-profile studies have shed substantial light on the mechanisms by which thermostability can evolve [44,45,62–64]. The increase in the thermostability of monomeric LcαE7-4a over monomeric LcαE7 conforms to our expectations of how monomers can be stabilized and is likely due to improved packing of hydrophobic residues in the core of LcαE7, as has we have discussed previously [53]. However, the oligomeric states of these stabilized proteins are seldom characterized; enzymes are most often assayed either in crude lysate or after affinity purification. It is therefore possible that our observation that the increased thermostability of LcαE7-4a can be partially attributed to enrichment of higher-order oligomeric species could be a more common process than currently thought.

Our results exemplify a trade-off between activity and stability that has been demonstrated in a number of studies of other enzymes [44,45,65]. Specifically, we observe the monomeric species to be more active than the dimeric and tetrameric species. Furthermore, we find that the monomer displays substrate inhibition, whereas the other species do not. This suggests that the monomeric species is able to adopt conformations in which multiple substrate molecules can bind at one time, inhibiting the reaction, whereas these, presumably more open, conformations are not available to the dimeric and tetrameric species. This is consistent with the location of the dimer interface: directly at the entrance to (but not occluding) the active site. Thus, we see that oligomerization can have significant stabilizing effects on certain protein regions [37] that can indirectly affect activity [66].

Recent work has shown that oligomers are more tolerant to mutation and thus more evolvable, primarily owing to their stability [64]. Indeed, numerous studies have implicated oligomerization in protein thermostability, through observation of oligomers in thermophiles [47,48], or engineered disruption/formation of protein–protein interfaces [14,49]. Despite this, there have been few, if any, direct observations (to the best of our knowledge), of proteins spontaneously evolving thermostability through the gain in additional oligomeric structure in laboratory settings in response to a thermal challenge. Because of our inability to replicate the stepwise evolution of oligomers in the laboratory and characterize evolutionary intermediates, our understanding of the sequence changes and structural mechanisms that underlie protein self-association has been limited. Only recently has the use of ancestral

protein reconstruction allowed an evolutionary process of oligomerization to be recapitulated, yielding valuable new insights into the mechanism of oligomerization and highlighting the role of remote mutations and protein dynamics [20]. The results presented here show that the evolution of thermostability in LcαE7 has occurred in parallel with enrichment of higher-order oligomeric states that display greater thermostability than the monomeric species, that point mutations can lead to a gradual shift in the oligomeric equilibrium of states in a protein population, and that these mutations can alter the propensity for monomers to oligomerize either *via* the formation of new bonds, or indirectly *via* the stabilization of mobile regions of the protein.

Materials and Methods

Directed evolution

The directed evolution experiment was performed as described previously [53]. Selection of improved variants was carried out through library screening (approximately 100,000 colonies). Cells were spread onto Lysogeny Broth agar plates supplemented with 100 μg/mL ampicillin at a cell density which resulted in ~200 colonies per plate. Colonies were blotted onto Whatman grade 3 filter paper (GE Healthcare) and incubated for 1 h at 50 °C (increased to 55 °C, 60 °C, and 70 °C for rounds 2, 3, and 4). The residual esterase activity was assayed by spraying heat-treated colonies with substrate solution [0.8% wt/vol Fast Red, 10 mM β-naphthyl acetate, and 100 mM Tris (pH 7.0)]. The most active mutants after thermal challenge were identified as those colonies that produced the most intense red color. A secondary screen of the best variants was carried out, in which the best variants from the plate screen were picked and grown in 96-deep well plate formats. The overnight cultures were heat stressed for 1 h at the same temperatures used in the primary screen, and 25 μL overnight cultures were assayed using a Molecular Devices plate reader at 490 nm in the presence of 0.5 mM β-naphthyl acetate, 0.5 mM Fast Red dye, and 100 mM Tris (pH 7.0). The best 5–10 variants of each generation were carried forward to the next generation, and the final product LcαE7-4a and several variants from rounds 2 and 3 were sequenced at the Micromon Sequencing Facility, Melbourne, Australia.

Site-directed mutagenesis

Surface mutations of LcαE7-4a were individually reverted back to the WT state (A83D, L364M, E530K, or G554D) by the DNA-fragment assembly method described by Gibson and coworkers [67]. A285S and F478L were introduced into LcαE7-4a by the same assembly method. Pairs of primers were designed to introduce the required codon change in the LcαE7 gene located in a pETMCSIII vector. The resulting vectors were sequenced at the Biomolecular Resource Facility, Australian National University, Australia.

Protein expression and purification

LcαE7-WT and LcαE7-4a were expressed in *E. coli* BL21 (DE3) cells and grown in Lysogeny Broth supplemented with 100 μg/mL of ampicillin in the leaky expression vector pETMCSIII [68]. Cultures were grown at 25 °C for 18 h, without induction. The cells were resuspended in buffer A [10 mM imidazole, 50 mM Hepes (pH 7.5), 300 mM NaCl] and lysed by sonication using the Sonic Ruptor 400 (Omni International). Clarified lysate was applied to a 5 mL Ni-NTA column (Qiagen) for nickel-affinity chromatography and the protein was eluted with buffer A supplemented with 300 mM imidazole. Protein purified by affinity chromatography was tested for esterase activity with 4-nitrophenyl butyrate and analyzed by SDS-PAGE to identify pure fractions.

SEC–MALLS

Pure LcαE7 fractions were pooled and further purified by SEC using a Hiload 26/600 Superdex-200 column (GE Healthcare), equilibrated with buffer B [200 mM NaCl, 20 mM Hepes (pH 7.5)]. The effect of viscosity on LcαE7-WT and LcαE7-4a was investigated by SEC as described above, with the addition of 10% glycerol to the running buffer. The total protein loaded (estimated by total area under the curve) was 24.5, 12.1, 11.5, and 10.8 mg (highest to lowest) for the LcαE7-WT samples and 16.6, 15.3, 13.2, and 13 mg (highest to lowest) for LcαE7-4a. For SEC–MALLS, purified LcαE7-4a (3.0 mg/mL) was applied at a flow rate of 1 mL/min to a WTC-030S5-column (Wyatt Technology) connected to a DAWN8+ MALLS detector and tREX refractive index detector (Wyatt Technology). The column was equilibrated with buffer B. Data analysis was performed using ASTRA software (Wyatt Technology).

Small-angle X-ray scattering

Data were collected in 200 mM NaCl, 20 mM Hepes (pH 7.5), and 5% glycerol using an inline S200, 5/150GL (3.2 mL; GE Healthcare) column at a 0.2-mL/min flow rate at the Australian Synchrotron SAXS beam line. Data were collected serially in 2-s quanta. The radius of gyration was unable to be estimated for any of the eluted protein owing to scatter patterns synonymous with the presence of HMW aggregates. Data were collected using a Pilatus 1 M detector at a distance of 3.3 m allowing collection of data in the Q range: 0.00486–0.25 nm^{-1}.

AUC

Sedimentation velocity experiments were conducted in a Beckman model XL-A analytical ultracentrifuge at a temperature of 20 °C. LcαE7 and LcαE7-4a were solubilized in 20 mM Hepes, and 500 mM NaCl (pH 7.5) and were analyzed at an initial concentration of 3.0 mg/mL. The sample and reference solution were loaded into conventional double-sector quartz cells and mounted in a Beckman 4-hole An-60 Ti rotor. The sample of 380 μL and reference solution of 400 μL were centrifuged at a rotor speed of 40,000 rpm, and the data were collected at a single wavelength (295 nm) in continuous mode, using a step-size of 0.003 cm without averaging. The

partial specific volume (0.736 mL/g at 20 °C for both WT and variant proteins), buffer density (1.0202 g/mL), and buffer viscosity (1.0635 cp) were computed using the program SEDNTERP [69]. Sedimentation velocity data at multiple time points were fitted to a continuous sedimentation coefficient [$c(s)$] distribution and a continuous mass [$c(M)$] distribution model [70–72] using the program SEDFIT, which is available at www.analyticalultracentrifugation.com.

CD

Purified LcαE7 at 0.3 mg/mL was exchanged into 20 mM NaH_2PO_4 and 20 mM NaCl buffer (pH 7.5). CD data were collected with a Chirascan CD spectrometer (Applied Photophysics) in a 1-mm quartz cuvette. The protein was heated from 25 °C to 90 °C at a rate of 1 °C/min, while ellipticity was monitored at 208 nm. The thermally induced unfolding of LcαE7 was not reversible. CD data were fitted using the by non-linear regression to a two-state model (Eq. (1)) using GraphPad Prism 6.00 (GraphPad Software).

$$y_{obs} = \frac{y_F - y_U}{1 + \exp\left(\frac{TT_{50} - T}{c}\right)} \tag{1}$$

where y_{obs} is the observed elipticity, y_F nand y_U are the elipticity values observed for the native and unfolded states, respectively. TT_{50} is the temperature at which the population of unfolded protein is 50% and c is the slope. Curves were fit by non-linear regression using GraphPad Prism 6.00 (GraphPad Software, USA). A three-state model was used for curve fitting with the monomer and dimer data using the software CDpal [73].

Thermal shift fluorescence assay

The thermal shift of LcαE7 and LcαE7-4a was measured using the ViiA7 real-time PCR system (Life Technologies) in MicroAmp EnduraPlate Optical 384 well plates (Life Technologies) with a final volume of 20 μL per well. The plates were covered with optical seal and shaken before denaturation. Each sample was measured in duplicate using the SYPRO Orange dye to measure protein unfolding. Protein was diluted to concentrations between 0.3 mg/mL and 3 mg/mL in buffer B. The 5000× SYPRO Orange dye stock was diluted in buffer B to a final concentration of 20× in the plate. Thermal denaturation was measured by increasing the temperature from 20 °C to 90 °C at a rate of 0.017 °C/s, and the plate was measured with wavelengths of excitation at 470 nm and emission at 580 nm. The TT_{50} was calculated by nonlinear regression with the Boltzmann sigmoidal equation in Graphpad Prism 6.00 (GraphPad Software, USA). A three-state model was used for curve fitting with the LcαE7-4a data using the software CDpal [73].

Enzyme kinetic assays

Purified LcαE7-WT and LcαE7-4a were assayed against the ester substrate 4-nitrophenyl butyrate in buffer B [200 mM NaCl, 20 mM Hepes (pH 7.5)] at 25 °C. Formation of the product 4-nitrophenolate was monitored at 405 nm

using an Epoch microplate spectrophotometer (BioTeK Instruments). Velocities were obtained from the initial linear portion of the reaction progress curves, and product concentration was determined with the extinction coefficient calculated from a standard curve of 4-nitrophenol. Assays were done in triplicate, and the protein concentration was used in the range of 0.01 to 0.05 μM as final concentration. The Michaelis constant and the inhibition constant were determined by non-linear regression of the initial velocities to the Michaelis–Menten equation or substrate inhibition equation using GraphPad Prism 6.00 (GraphPad Software). For thermal stability assays, purified LcαE7-WT and LcαE7-4a were incubated at 40 °C, aliquots were removed at set time points, and specific activity was assayed as described above with 200 μM 4-nitrophenyl butyrate in triplicate. To measure the decay of LcαE7-4a monomer and tetramer activity over a time course, the proteins were incubated at 46 °C in a buffer consisting of 20 mM Hepes, 200 mM NaCl, and 10% glycerol. Aliquots removed at various time points and subsequently assayed with 200 μM 4-nitrophenyl butyrate in triplicate as described above. The data were fit by non-linear regression to a one-phase exponential decay curve using GraphPad Prism 6.00 (GraphPad Software, USA).

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2016.03.014.

molecular weight; SEC, size exclusion chromatography; SEC–MALLS, size exclusion chromatography–multi-angle laser light scattering; WT, wild-type.

References

[1] S. Kühner, V. van Noort, M.J. Betts, A. Leo-Macias, C. Batisse, M. Rode, et al., Proteome organization in a genome-reduced bacterium, Science 326 (2009) 1235–1240, http://dx.doi.org/10.1126/science.1176343.

[2] D.S. Goodsell, A.J. Olson, Structural symmetry and protein function, Annu. Rev. Biophys. Biomol. Struct. 29 (2000) 105–153, http://dx.doi.org/10.1146/annurev.biophys.29.1.105.

[3] E.D. Levy, J.B. Pereira-Leal, C. Chothia, S.A. Teichmann, 3D complex: a structural classification of protein complexes, PLoS Comput. Biol. 2 (2006) 1395–1406, http://dx.doi.org/10.1371/journal.pcbi.0020155.

[4] H. Nishi, K. Hashimoto, T. Madej, A.R. Panchenko, Evolutionary, physicochemical, and functional mechanisms of protein homooligomerization, Prog. Mol. Biol. Transl. Sci. 117 (2013) 3–24, http://dx.doi.org/10.1016/B978-0-12-386931-9.00001-5.

[5] P.C. Havugimana, G.T. Hart, T. Nepusz, H. Yang, A.L. Turinsky, Z. Li, et al., A census of human soluble protein complexes, Cell 150 (2012) 1068–1081, http://dx.doi.org/10.1016/j.cell.2012.08.011.

[6] K. Tarassov, V. Messier, C.R. Landry, S. Radinovic, M.M. Serna Molina, I. Shames, et al., An *in vivo* map of the yeast protein interactome, Science 320 (2008) 1465–1470, http://dx.doi.org/10.1126/science.1153878.

[7] C.V. Robinson, A. Sali, W. Baumeister, The molecular sociology of the cell, Nature 450 (2007) 973–982, http://dx.doi.org/10.1038/nature06523.

[8] N.J. Krogan, G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, et al., Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*, Nature 440 (2006) 637–643, http://dx.doi.org/10.1038/nature04670.

[9] C. Wan, B. Borgeson, S. Phanse, F. Tu, K. Drew, G. Clark, et al., Panorama of ancient metazoan macromolecular complexes, Nature 525 (2015) 339–344, http://dx.doi.org/10.1038/nature14877.

[10] F.H. Crick, J.D. Watson, Structure of small viruses, Nature 177 (1956) 473–475, http://dx.doi.org/10.1038/177473a0.

[11] R.C.J. Dobson, K. Valegård, J.A. Gerrard, The crystal structure of three site-directed mutants of *Escherichia coli* dihydrodipicolinate synthase: further evidence for a catalytic triad, J. Mol. Biol. 338 (2004) 329–339, http://dx.doi.org/10.1016/j.jmb.2004.02.060.

[12] M.A. Navia, P.M. Fitzgerald, B.M. McKeever, C.T. Leu, J.C. Heimbach, W.K. Herber, et al., Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1, Nature 337 (1989) 615–620, http://dx.doi.org/10.1038/337615a0.

[13] C.H. Heldin, Dimerization of cell surface receptors in signal transduction, Cell 80 (1995) 213–223, http://dx.doi.org/10.1016/0092-8674(95)90404-2.

[14] A.D. Malay, S.L. Procious, D.R. Tolan, The temperature dependence of activity and structure for the most prevalent mutant aldolase B associated with hereditary fructose intolerance, Arch. Biochem. Biophys. 408 (2002) 295–304, http://dx.doi.org/10.1016/S0003-9861(02)00546-5.

[15] E.J. Loveridge, R.J. Rodriguez, R.S. Swanwick, R.K. Allemann, Effect of dimerization on the stability and catalytic activity of dihydrofolate reductase from the hyperthermophile thermotoga maritima, Biochemistry 48 (2009) 5922–5933, http://dx.doi.org/10.1021/bi900411a.

[16] J. Monod, J. Wyman, J.P. Changeux, On the nature of allosteric transitions: a plausible model, J. Mol. Biol. 12 (1965) 88–118, http://dx.doi.org/10.1016/S0022-2836(65)80285-6.

[17] T. Perica, J.A. Marsh, F.L. Sousa, E. Natan, L.J. Colwell, S.E. Ahnert, et al., The emergence of protein complexes: quaternary structure, dynamics and allostery. Colworth Medal Lecture, Biochem. Soc. Trans. 40 (2012) 475–491, http://dx.doi.org/10.1042/BST20120056.

[18] K. Gunasekaran, B. Ma, R. Nussinov, Is allostery an intrinsic property of all dynamic proteins? Proteins Struct. Funct. Genet. 57 (2004) 433–443, http://dx.doi.org/10.1002/prot.20232.

[19] D.E. Koshland, G. Némethy, D. Filmer, Comparison of experimental binding data and theoretical models in proteins containing subunits, Biochemistry 5 (1966) 365–385, http://dx.doi.org/10.1021/bi00865a047.

[20] T. Perica, Y. Kondo, S.P. Tiwari, S.H. McLaughlin, K.R. Kemplen, X. Zhang, et al., Evolution of oligomeric state through allosteric pathways that mimic ligand binding, Science 346 (2014) 1254346, http://dx.doi.org/10.1126/science.1254346.

[21] J.A. Marsh, S.A. Teichmann, Structure, dynamics, assembly, and evolution of protein complexes, Annu. Rev. Biochem. 84 (2014) 551–575, http://dx.doi.org/10.1146/annurev-biochem-060614-034142.

[22] S.J. Fleishman, J.E. Corn, E.-M. Strauch, T.A. Whitehead, J. Karanicolas, D. Baker, Hotspot-Centric *de novo* Design of protein binders, J. Mol. Biol. 413 (2011) 1047–1062, http://dx.doi.org/10.1016/j.jmb.2011.09.001.

[23] J. Karanicolas, J.E. Corn, I. Chen, L.A. Joachimiak, O. Dym, S.H. Peck, et al., A *de novo* protein binding pair by computational design and directed evolution, Mol. Cell 42 (2011) 250–260, http://dx.doi.org/10.1016/j.molcel.2011.03.010.

[24] B. Kim, A. Eggel, S.S. Tarchevskaya, M. Vogel, H. Prinz, T.S. Jardetzky, Accelerated disassembly of IgE–receptor complexes by a disruptive macromolecular inhibitor, Nature 491 (2012) 613–617, http://dx.doi.org/10.1038/nature11546.

[25] H. Jubb, A.P. Higueruelo, A. Winter, T.L. Blundell, Structural biology and drug discovery for protein–protein interactions, Trends Pharmacol. Sci. 33 (2012) 241–248, http://dx.doi.org/10.1016/j.tips.2012.03.006.

[26] T. Laue, B. Demeler, A postreductionist framework for protein biochemistry, Nat. Chem. Biol. 7 (2011) 331–334, http://dx.doi.org/10.1038/nchembio.575.

[27] A. Gershenson, L.M. Gierasch, A. Pastore, S.E. Radford, Energy landscapes of functional proteins are inherently risky, Nat. Chem. Biol. 10 (2014) 884–891, http://dx.doi.org/10.1038/nchembio.1670.

[28] L.N. Kinch, N.V. Grishin, Evolution of protein structures and functions, Curr. Opin. Struct. Biol. 12 (2002) 400–408, http://dx.doi.org/10.1016/S0959-440X(02)00338-X.

[29] E. Akiva, Z. Itzhaki, H. Margalit, Built-in loops allow versatility in domain-domain interactions: lessons from self-interacting domains, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 13292–13297, http://dx.doi.org/10.1073/pnas.0801207105.

[30] K. Hashimoto, A.R. Panchenko, Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 20352–20357, http://dx.doi.org/10.1073/pnas.1012999107.

[31] M.C. Mossing, R.T. Sauer, Stable, monomeric variants of lambda cro obtained by insertion of a designed beta-hairpin sequence, Science 250 (1990) 1712–1715, http://dx.doi.org/10.1126/science.2148648.

[32] R.R. Dickason, D.P. Huston, Creation of a biologically active interleukin-5 monomer, Nature 379 (1996) 652–655, http://dx.doi.org/10.1038/379652a0.

[33] G. MacBeath, P. Kast, D. Hilvert, Probing enzyme quaternary structure by combinatorial mutagenesis and selection, Protein Sci. 7 (1998) 1757–1767, http://dx.doi.org/10.1002/pro.5560070810.

[34] G. MacBeath, P. Kast, D. Hilvert, Redesigning enzyme topology by directed evolution, Science 279 (1998) 1958–1961, http://dx.doi.org/10.1126/science.279.5358.1958.

[35] Y.-T. Lai, D. Cascio, T.O. Yeates, Structure of a 16-nm cage designed by using protein oligomers, Science 336 (2012) 1129-1129, http://dx.doi.org/10.1126/science.1219351.

[36] Y.-T. Lai, N.P. King, T.O. Yeates, Principles for designing ordered protein assemblies, Trends Cell Biol. 22 (2012) 653–661, http://dx.doi.org/10.1016/j.tcb.2012.08.004.

[37] J.A. Marsh, S.A. Teichmann, Protein flexibility facilitates quaternary structure assembly and evolution, PLoS Biol. 12 (2014), e1001870 http://dx.doi.org/10.1371/journal.pbio.1001870.

[38] J.A. Marsh, H.A. Rees, S.E. Ahnert, S.A. Teichmann, Structural and evolutionary versatility in protein complexes with uneven stoichiometry, Nat. Commun. 6 (2015) 6394, http://dx.doi.org/10.1038/ncomms7394.

[39] X. Zhang, T. Perica, S.A. Teichmann, Evolution of protein structures and interactions from the perspective of residue contact networks, Curr. Opin. Struct. Biol. 23 (2013) 954–963, http://dx.doi.org/10.1016/j.sbi.2013.07.004.

[40] T.L. Blundell, N. Srinivasan, Symmetry, stability, and dynamics of multidomain and multicomponent protein systems, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 14243–14248, http://dx.doi.org/10.1073/pnas.93.25.14243.

[41] I. André, C.E.M. Strauss, D.B. Kaplan, P. Bradley, D. Baker, Emergence of symmetry in homooligomeric biological assemblies, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 16148–16152, http://dx.doi.org/10.1073/pnas.0807576105.

[42] S.J. Fleishman, T.A. Whitehead, D.C. Ekiert, C. Dreyfus, J.E. Corn, E. Strauch, et al., Computational design of proteins targeting the conserved stem region of influenza hemagglutinin, Science 332 (2011) 816–821, http://dx.doi.org/10.1126/science.1202617.

[43] D. Grueninger, N. Treiber, M.O.P. Ziegler, J.W.A. Koetter, M.-S. Schulze, G.E. Schulz, Designed protein–protein association, Science 319 (2008) 206–209, http://dx.doi.org/10.1126/science.1150421.

[44] J.D. Bloom, S.T. Labthavikul, C.R. Otey, F.H. Arnold, Protein stability promotes evolvability, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 5869–5874, http://dx.doi.org/10.1073/pnas.0510098103.

[45] S. Bershtein, M. Segal, R. Bekerman, N. Tokuriki, D.S. Tawfik, Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein, Nature 444 (2006) 929–932, http://dx.doi.org/10.1038/nature05385.

[46] T.A. Larsen, A.J. Olson, D.S. Goodsell, Morphology of protein–protein interfaces, Structure 6 (1998) 421–427.

[47] H. Walden, G.S. Bell, R.J. Russell, B. Siebers, R. Hensel, G.L. Taylor, Tiny TIM: a small, tetrameric, hyperthermostable triosephosphate isomerase, J. Mol. Biol. 306 (2001) 745–757, http://dx.doi.org/10.1006/jmbi.2000.4433.

[48] M. Robinson-Rechavi, A. Alibés, A. Godzik, Contribution of electrostatic interactions, compactness and quaternary structure to protein thermostability: lessons from structural genomics of thermotoga maritima, J. Mol. Biol. 356 (2006) 547–557, http://dx.doi.org/10.1016/j.jmb.2005.11.065.

[49] R. Thoma, M. Hennig, R. Sterner, K. Kirschner, Structure and function of mutationally generated monomers of dimeric phosphoribosylanthranilate isomerase from Thermotoga maritima, Structure 8 (2000) 265–276, http://dx.doi.org/10.1016/S0969-2126(00)00106–4.

[50] B. Kuhlman, J.W. O'Neill, D.E. Kim, K.Y. Zhang, D. Baker, Conversion of monomeric protein L to an obligate dimer by computational protein design, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 10687–10691, http://dx.doi.org/10.1073/pnas.181354398.

[51] M. Kirsten Frank, F. Dyda, A. Dobrodumov, A.M. Gronenborn, Core mutations switch monomeric protein GB1 into an intertwined tetramer, Nat. Struct. Biol. 9 (2002) 877–885, http://dx.doi.org/10.1038/nsb854.

[52] R.D. Newcomb, P.M. Campbell, D.L. Ollis, E. Cheah, R.J. Russell, J.G. Oakeshott, A single amino acid substitution converts a carboxylesterase to an organophosphorus hydrolase and confers insecticide resistance on a blowfly, Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 7464–7468, http://dx.doi.org/10.1073/pnas.94.14.7464.

[53] C.J. Jackson, J.-W. Liu, P.D. Carr, F. Younus, C. Coppin, T. Meirelles, et al., Structure and function of an insect α-carboxylesterase (αesterase7) associated with insecticide resistance, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 10177–10182, http://dx.doi.org/10.1073/pnas.1304097110.

[54] R.H. Brakenhoff, J.G. Schoenmakers, N.H. Lubsen, Chimeric cDNA clones: a novel PCR artifact, Nucleic Acids Res. 19 (1991) 1949, http://dx.doi.org/10.1093/nar/19.8.1949.

[55] A.D. Keith, W. Snipes, Viscosity of cellular protoplasm, Science 183 (1974) 666–668, http://dx.doi.org/10.1126/science.183.4125.666.

[56] P.J. Wyatt, Multiangle light scattering: the basic tool for macromolecular characterization, Instrum. Sci. Technol. 25 (1997) 1–18, http://dx.doi.org/10.1080/10739149709351443.

[57] F.H. Niesen, H. Berglund, M. Vedadi, The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability, Nat. Protoc. 2 (2007) 2212–2221, http://dx.doi.org/10.1038/nprot.2007.321.

[58] E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state, J. Mol. Biol. 372 (2007) 774–797, http://dx.doi.org/10.1016/j.jmb.2007.05.022.

[59] L. Lo Conte, C. Chothia, J. Janin, The atomic structure of protein–protein recognition sites, J. Mol. Biol. 285 (1999) 2177–2198, http://dx.doi.org/10.1006/jmbi.1998.2439.

[60] Z.S. Derewenda, P.G. Vekilov, Entropy and surface engineering in protein crystallization, Acta Crystallogr. Sect. D: Biol. Crystallogr. 62 (2006) 116–124, http://dx.doi.org/10.1107/S0907444905035237.

[61] B. Meibohm, H. Zhou, Characterizing the impact of renal impairment on the clinical pharmacology of biologics, J. Clin. Pharmacol. 52 (2012) 54S–62S, http://dx.doi.org/10.1177/0091270011413894.

[62] C. Mitchinson, J.A. Wells, Protein engineering of disulfide bonds in subtilisin BPN', Biochemistry 28 (1989) 4807–4815, http://dx.doi.org/10.1021/bi00437a043.

[63] L. Giver, A. Gershenson, P.O. Freskgard, F.H. Arnold, Directed evolution of a thermostable esterase, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 12809–12813, http://dx.doi.org/10.1073/pnas.95.22.12809.

[64] S. Bershtein, W. Mu, E.I. Shakhnovich, Soluble oligomerization provides a beneficial fitness effect on destabilizing mutations, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 4857–4862, http://dx.doi.org/10.1073/pnas.1118157109.

[65] B.M. Beadle, B.K. Shoichet, Structural bases of stability-function tradeoffs in enzymes, J. Mol. Biol. 321 (2002) 285–296, http://dx.doi.org/10.1016/S0022-2836(02)00599-5.

[66] S.R. Devenish, J.A. Gerrard, The role of quaternary structure in (beta/alpha)(8)-barrel proteins: evolutionary happenstance or a higher level of structure–function relationships? Org. Biomol. Chem. 7 (2009) 833–839, http://dx.doi.org/10.1039/b818251p.

[67] D.G. Gibson, L. Young, R.-Y. Chuang, J.C. Venter, C.A. Hutchison, H.O. Smith, Enzymatic assembly of DNA molecules up to several hundred kilobases, Nat. Methods 6 (2009) 343–345, http://dx.doi.org/10.1038/nmeth.1318.

[68] C. Neylon, S.E. Brown, A.V. Kralicek, C.S. Miles, C.A. Love, N.E. Dixon, Interaction of the *Escherichia coli* replication terminator protein (Tus) with DNA: a model derived from DNA-binding studies of mutant proteins by surface plasmon resonance, Biochemistry 39 (2000) 11989–11999, http://dx.doi.org/10.1021/bi001174w.

[69] T.M. Laue, B.D. Shah, T.M. Ridgeway, S.L. Pelletier, Computer-aided interpretation of analytical sedimentation data for proteins, in: S.E. Harding, A.J. Rowe, J.C. Horton (Eds.), Anal. Ultracentrifugation Biochem. Polym. Sci. Royal Society of Chemistry, Cambridge, U.K. 1992, pp. 90–125.

[70] M.A. Perugini, P. Schuck, G.J. Howlett, Self-association of human apolipoprotein E3 and E4 in the presence and absence of phospholipid, J. Biol. Chem. 275 (2000) 36758–36765, http://dx.doi.org/10.1074/jbc.M005565200.

[71] P. Schuck, Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling, Biophys. J. 78 (2000) 1606–1619, http://dx.doi.org/10.1016/S0006-3495(00)76713-0.

[72] P. Schuck, M.A. Perugini, N.R. Gonzales, G.J. Howlett, D. Schubert, Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems, Biophys. J. 82 (2002) 1096–1111, http://dx.doi.org/10.1016/S0006-3495(02)75469-6.

[73] M. Niklasson, C. Andresen, S. Helander, M.G.L. Roth, A. Zimdahl Kahlin, M. Lindqvist Appell, et al., Robust and convenient analysis of protein thermal and chemical stability, Protein Sci. 24 (2015) 2055–2062, http://dx.doi.org/10.1002/pro.2809.

Chapter 3. Molecular basis for the behavioral effects of the odorant degrading enzyme Esterase 6 in *Drosophila*

3.1. Declaration of Author Contribution

The research article in this chapter was peer-reviewed and published as a regular original research article. The work shown was completed primarily by myself and Faisal Younus. The experimental work to obtain enzyme kinetics was performed by Faisal Younus, Christopher Coppin, Jian-Wei Liu and Gunjan Pandey. The covalent docking was performed by myself and Galen J. Correy. The electrophysiological experimental was done by Thomas Chertemps and Martine Maïbèche. The crystallization work was performed by myself. I solved and refined the structure with assistance from Colin J. Jackson. The writing of the draft was conducted by Faisal Younus, myself, Colin J. Jackson, John G. Oakeshott and Martine Maïbèche.

3.2. Introduction

3.2.1. β-Esterases

Clade E is one of the most studied clades of CBEs in *Drosophila* species (152). Enzymes in this clade have been collectively known as β-esterases based on the preferential hydrolysis of β-napthyl acetate in *Drosophila* (32). This cluster however, is able to breakdown α-napthyl acetate which makes substrate preference not necessarily indicative of the evolutionary history. Regardless, the name for the β-cluster is widely used and has been retained in the literature (32). The β-esterase cluster generally contains two to four members and it has been suggested to have evolved after a relatively ancient gene duplication event **(Table 3.1)** (42, 43, 56, 152–155). After the gene duplication event, the proteins have adopted distinct functions and expression patterns (42, 152, 153, 156–158). The β-cluster has been studied to investigate the molecular basis of functional evolution and for understanding allozymes (56, 152, 153).

Table 3.1. List of the β-esterase gene cluster in various *Drosophila* species, red represents the orthologs of EST6 from *D. melanogaster* and blue represents the EST7 equivalent (32).

Drosophila Species	Members of the β-Cluster
melanogaster, simulans, mauritiana, sechellia, yakuba, teissieri, erecta, orena	EST6, EST7
ananassae	EST6a, EST6b, EST6c, EST7
pseudoobscura, persimilis, miranda	EST5a, EST5b, EST5c
affinis	EST5b
wilistoni	EST6
mojavensis	EST1, EST2a, EST2b, EST2c
virilis	EST1, EST2d, EST2e, EST2f
grimshawi	EST2

3.2.2. EST6 and EST7

In several *Drosophila* species, the cluster is comprised of two tandemly arranged genes noted as *esterase6* and *esterase7* that encode the proteins esterase-6 (EST6) and esterase-7 (EST7) **(Table 3.1)** (153, 154, 159). Both proteins are soluble, secreted and glycosylated with distinct expression patterns (153, 154, 159). The first mention of this cluster of enzymes in the literature was in 1963 and since then over 40 research papers have focused

on the enzymes (32, 160). The study of EST6 and EST7 has established a well-understood model of microevolution and has had a large impact on *Drosophila* genetics research (32).

3.2.3. EST6

EST6 in *D. melanogaster* has acquired a novel function involved with reproduction biology (56). In other *Drosophila* species (*D. erecta*), EST6 has a functional role in the hemolymph with unknown function (56). However, EST6 in *D. melanogaster* is highly expressed in the sperm ejaculatory duct of the adult male fly (161). The enzyme is transferred during copulation from the male to the female in the seminal fluid during the first few minutes of mating. The transfer results in the triggering of egg-laying and remating behaviour being repressed in the female fly (162, 163). After transfer, EST6 is translocated to the hemolymph of the female, where traces of the enzyme survives for several days (164, 165). It is unknown how the male-EST6 in the female hemolymph affects the female-EST6, however, it has been shown to affect the female's subsequent reproductive behaviour in relation to EST6 activity (153, 165–168).

EST6 is important in the reproductive biology of *Drosophila* as differences in enzyme expression in the male can affect the reproductive success of their mates (168). Also while EST6 is expressed in the ejaculatory duct, it is also expressed in the antennae (27, 157, 161, 169, 170). The expression in the antennae has recently expanded EST6 to have a secondary role in the sensory physiological and behavioural dynamics of *D. melanogaster*. It has been proposed that EST6 acts as an odorant-degrading enzyme involved in 11-*cis* vaccenyl acetate (cVA) detection in the antennae (170, 171).

3.2.4. EST6 Substrate Specificity and 11-*cis* vaccenyl acetate

The mechanism of EST6 in the reproductive success of *Drosophila* is not well understood (32). EST6 shows activity against a broad range of substrates *in vitro* suggesting to be a non-specific CBE. However, the *in vivo* substrates are unknown **(Figure 3.1)** (168, 172). A proposed mechanism of how EST6 affects reproduction involves the pheromone cVA **(Figure 3.1)** (162, 173).
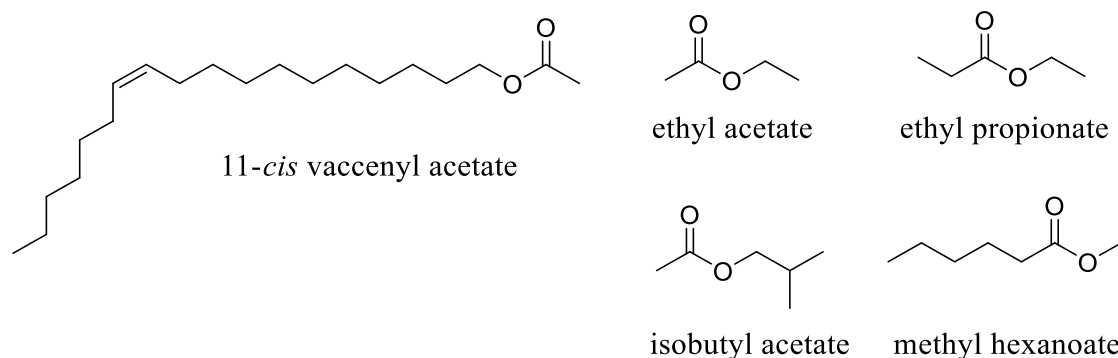
Figure 3.1. Structures of cVA, the sex aggregation hormone in *Drosophila* and short-chain esters (ethyl acetate, ethyl propionate, isobutyl acetate, methyl hexanoate) that are known odorants involved with *Drosophila* species (174).

cVA is an oxygenated semi-volatile hydrocarbon and is the major sex and aggregation pheromone in *D. melanogaster* species (175). cVA is transferred from the ejaculatory bulb of the male to the female during copulation. After transfer, cVA reduces the attractiveness of the recently mated female to males which suppresses male mating behaviour (173, 175–178). The mated female once copulated will rarely accept another male, resulting in cVA being an indicator of the mating status of the female (179). This reduces a potential loss of progeny for the female's first mate (170).

The evidence for EST6 inactivating cVA has been speculative (32). EST6 and cVA are both found in close proximity in the male reproductive system and both transferred to the female during copulation. Preliminary assays had led researchers to believe that *in vivo,* cVA is the substrate for EST6 and the mechanism of how EST6 affects reproduction (162, 173). However, these claims have since been refuted (162, 180–183). Recent work involving electrophysiological investigations on the *Drosophila* antennae indicate that EST6 has a role in the deactivation of cVA despite the location of EST6 in the antennae not being responsive to cVA (170). Further characterization of EST6 suggests the enzyme acts on various short chain fatty acid food esters (32, 174). This raises the question of the substrate range of the enzyme, suggesting EST6 has the ability to turnover a large range of esters with carbon length ranging from 3 to 20. The evidence of the *in vivo* substrates of EST6, the role the enzyme plays in the antennae and the mechanism by which the enzyme influences female reproductive behaviour remains unknown. Likewise, the interaction between EST6 and cVA is not well understood and still under discussion (170, 173, 174, 183).

3.2.5. Prediction of EST6 Structure

Studies have modelled the structure of EST6 using the enzyme acetylcholinesterase (56, 153). The structure of EST6 is predicted to contained three disulfide bonds and four N-linked glycosylation sites which have been suggested to help the longevity of the protein during transfer from the male to the female (56, 153). The loop regions, substrate binding regions and active site gorge which are critical for functional differences in CBEs are unknown in the structure based on the EST6 homology model with only the main α/β hydrolase fold identified with high accuracy.

3.3. Research Article

3.3.1. Preface

For the past 50 years, the β-cluster has been heavily studied, yet questions still remain on the structure/function relationships of enzymes in this cluster. The most studied member of this cluster, EST6 from *D. melanogaster* has been a major focus due to the important role the enzyme has in the reproductive success of *Drosophila*. EST6 has been modelled using the structure of AChE, but this protein has significant differences in substrate specificity and shares 30% sequence identity to EST6. Thus, only the highly conserved regions of EST6 have been modelled with reasonable certainty (56, 153). A structure of EST6 is important to understand the function of this protein and the role of this protein in *D. melanogaster* reproduction and in the antennae. The following article discusses the structure of EST6 in combination with enzyme kinetics and electrophysiological experiments to provide insight into the substrate range of the enzyme and predicted role of EST6 in *Drosophila.*

3.3.2. Published Research Article: Molecular basis for the behavioral effects of the odorant degrading enzyme Esterase 6 in *Drosophila*

Supplementary Information is found in Appendix B

SCIENTIFIC REP⚙RTS

Molecular basis for the behavioral effects of the odorant degrading enzyme Esterase 6 in *Drosophila*

Faisal Younus[1,2], Nicholas J. Fraser[2], Chris W. Coppin[1], Jian-Wei Liu[1], Galen J. Correy[2], Thomas Chertemps[3], Gunjan Pandey[1], Martine Maïbèche[3], Colin J. Jackson[2] & John G. Oakeshott[1]

Previous electrophysiological and behavioural studies implicate esterase 6 in the processing of the pheromone cis-vaccenyl acetate and various food odorants that affect aggregation and reproductive behaviours. Here we show esterase 6 has relatively high activity against many of the short-mid chain food esters, but negligible activity against cis-vaccenyl acetate. The crystal structure of esterase 6 confirms its substrate-binding site can accommodate many short-mid chain food esters but not cis-vaccenyl acetate. Immunohistochemical assays show esterase 6 is expressed in non-neuronal cells in the third antennal segment that could be accessory or epidermal cells surrounding numerous olfactory sensilla, including basiconics involved in food odorant detection. Esterase 6 is also produced in trichoid sensilla, but not in the same cell types as the cis-vaccenyl acetate binding protein LUSH. Our data support a model in which esterase 6 acts as a direct odorant degrading enzyme for many bioactive food esters, but not cis-vaccenyl acetate.

Insects' olfactory systems are both primary drivers of their interactions with the environment and an emerging model for studying the molecular basis of eukaryote signaling processes. They are also of enormous interest in applied entomology because they are the targets for various pest control strategies based on mating disruption[1]. Many aspects of insects' olfactory system have recently been elucidated but others, such as their odorant degrading enzymes (ODEs), are still poorly understood[2,3]. It is proposed that ODEs are vital in the maintenance of the ongoing sensitivity of the olfactory system to incoming signals through the rapid inactivation of the relevant pheromones and kairomones once they have activated their receptors[2,4]. However few of these have yet been characterized in any detail and fundamental questions remain about their modes of action. In particular there is ongoing debate, both about whether individual ODEs are specific for particular odorants or act generally against many[2], and about whether they act alone or in combination with odorant binding proteins (OBPs)[2,5]. OBPs have been strongly implicated in the transport of incoming odorants through the sensillar lymph to their corresponding receptors, but any subsequent role for them in the deactivation process remains controversial[2].

Most of the work to date on ODEs has been done on certain Lepidoptera that have antennae large enough for classical biochemical and physiological studies[4]. One of the best characterized is the antennal specific esterase Apo1SE from the giant silk moth *Antheraea polyphemus*, which is estimated to have a k_{cat} of $127\,s^{-1}$ for its natural E6Z11-16:acetate pheromone substrate[6], but little activity for other isomers of this compound or for several other volatile esters tested. Relatively high k_{cat} values for their putative pheromone ester substrates have also been reported for a few other lepidopteran antennal esterases, although in at least two of these cases their substrate ranges seem be to less specific[3,7,8], perhaps suggesting broad rather than specific ODE functions.

By far the best characterized ODE for the model insect *Drosophila melanogaster* is esterase 6 (EST6). This enzyme was originally reported to degrade the major volatile sex and aggregation pheromone cis-vaccenyl acetate (cVA)[9]. Subsequent electrophysiological comparisons of EST6 wildtype and null flies on comparable genetic backgrounds have confirmed a role for the enzyme in the dynamics of cVA processing[10]. A specific OBP, LUSH, has been identified for cVA in *D. melanogaster* but the latest genetic evidence suggests that the interaction of cVA with its receptor OR67d is independent of LUSH[11]. Notably, the distribution of EST6 in the third antennal

[1]CSIRO Land and Water, Black Mountain, Canberra, ACT, 2601, Australia. [2]Research School of Chemistry, Australian National University, Canberra, ACT, 2601, Australia. [3]Université Pierre et Marie Curie, Institut d'Ecologie et des Sciences de l'Environnement de Paris, 75252, Paris, France. Correspondence and requests for materials should be addressed to J.G.O. (email: john.oakeshott@csiro.au)

segment also includes sensilla that are responsive to other odorants but not to cVA[10], and further biochemical, electrophysiological and behavioral comparisons of the EST6 active and null strains indicate that the enzyme also acts on various short chain fatty acid food esters[12,13]. There is indeed some relationship between the level of EST6 activity for the different esters and the size of the electrophysiological effect[13], suggesting that EST6 does act as a general ODE with activities for several ester odorants.

As further evidence for pleiotropic effects of the enzyme, EST6 is also known to be expressed at high levels in the male ejaculatory duct, from where it is transferred to the female reproductive tract during mating[14]. It is then rapidly (within minutes) translocated to her hemolymph, where it remains for several days. Comparisons of females mated with null and wildtype EST6 males indicate it acts in the female to stimulate her egg-laying and delay her receptivity to re-mating[15,16]. Early claims that this effect was mediated by EST6 action on endogenous cVA[9] have since been refuted[17], but the substrate responsible for the effect nevertheless remains unknown.

EST6 is a member of the carboxyl/cholinesterase (CCE) family of proteins[18], which is represented by 30–110 different gene/enzyme systems encoding diverse functions in the insect genomes so far sequenced[19]. However, the juvenile hormone esterase from the moth *Manduca sexta* (*Ms*JHE)[20], an insecticide metabolizing carboxylesterase from the blowfly *Lucilia cuprina* (*Lc*αE7)[21] and acetylcholinesterase from *D. melanogaster* (*Dm*AChE)[22] are the only insect CCEs for which crystal structures have been determined, so relatively little is known of the structure-function relationships underlying their diverse functions. The structural features of EST6 have so far been inferred from the structure of the *D. melanogaster* AChE or its orthologue from the electric ray *Torpedo californica*, but the low sequence similarity between EST6 and AChE (27%) means that the fine structural features of the enzyme responsible for its substrate specificity have not yet been understood[23].

In this paper, we present a comprehensive analysis of the substrate range of semi-purified EST6, showing it has significant activity for a range of short chain fatty acid esters but negligible activity for long chain fatty acid esters. In particular, we find that EST6 is not active against cVA, either in the presence or absence of LUSH, but does degrade various volatiles emitted by rotting fruits and the yeasts therein on which the flies naturally live; these volatiles have recently been shown to be key regulators of *Drosophila* mating behavior[24]. We also present a crystal structure for the enzyme which, together with *in silico* docking studies, supports the kinetic data and shows that its active site can readily accommodate short chain fatty acid esters, including the yeast and fruit volatiles above, but not long chain fatty acid esters like cVA. A unique active site location and entry is identified, which appears to explain the enzyme's substrate preferences. Finally, we present data from immunohistochemical and behavioral assays with RNAi knock-down constructs that localize the expression of EST6 to a large proportion of non-neuronal cells surrounding the olfactory neurons of almost all the olfactory sensilla, but in different cells than those producing LUSH in the trichoid sensilla.

Results

Enzyme kinetics. Wildtype EST6 was tested for activity against 85 bioactive ester odorants and two model substrates; 4-nitrophenyl acetate (4 NPA) and 2-naphthyl acetate (2 NA). It showed detectable activity (generally, a specificity constant $k_{cat}/K_M^{Est} > 1.5 \times 10^4 M^{-1}.s^{-1}$) for 47 of the bioactive esters as well as the two model substrates (Fig. 1 and Supplementary Table S1). Specificity constants for most (42) of these 49 were above $1 \times 10^5 M^{-1}.s^{-1}$, although none exceeded $1.3 \times 10^6 M^{-1}.s^{-1}$, consistent with typical k_{cat}/K_M values for enzymatic reactions in secondary metabolism[25]. The highest activities were seen with esters containing longer (C > 6) or more complex (branched, unsaturated or cyclic) leaving groups and acetate or propionate acid moieties, although a combination of mid-length leaving groups and acid groups (butyl decanoate) was also a relatively good substrate in these assays. The 38 compounds for which little or no activity could be detected were mainly methyl or ethyl esters or those with more complex acidic groups. cVA, which has a very long leaving group, was not hydrolysed at significant rates.

Precise K_M values for most substrates could not be calculated because of low substrate solubility. However, estimates of K_M values could be obtained for some of the more soluble esters (4 NPA, 2 NA, benzyl acetate, phenyl acetate, phenethyl acetate) and were found to be in the range 121–880 μM under these assay conditions, which included 5% ethanol (Supplementary Table S1). Previous kinetic analyses of EST6 with 2NA[26] and 4 NPA (Supplementary Fig. S1) indicated the K_M values were ~5–20 fold lower in the absence of 5% ethanol. The K_M values that were obtained generally exceed the concentration of substrate in the reaction mixtures (200 μM), which means that the k_{cat}/K_M^{Est} calculated will be a reasonable approximation of the true k_{cat}/K_M value (in those cases where K_M is lower than 200 μM, the estimated value will underestimate the true k_{cat}/K_M for the assay conditions used – see Methods). Given the measured K_M values are typically >100 μM, the measured k_{cat}/K_M values therefore imply relatively high k_{cat} values (in some cases >1,000 s^{-1}). These results indicate that EST6 is a relatively "fast" enzyme (high k_{cat} values) that displays broad specificity, working moderately efficiently with a very wide range of natural esters. In comparison, the related enzyme acetylcholinesterase catalyses acetylcholine hydrolysis with very high efficiency but has an extraordinarily narrow substrate range, essentially catalysing a single substrate[18].

The assays with cVA were repeated in the presence of the cVA binding protein LUSH, which again indicated negligible activity, even in the presence of a great excess of EST6 (57 nM compared with the 3 nM used previously). The only other known pheromone among the compounds tested was the fatty acid ester methyl myristate, which is also a plant volatile and functions as an attractant to *D. melanogaster*[27]. EST6 also had relatively little activity with this compound (~1.5 × 10^4 M^{-1}.s^{-1}).

Apart from the two pheromones and two model substrates, all the esters tested for which EST6 was found to have significant activity are food odorants that are known to be bioactive against *D. melanogaster* in *in vivo* (behavioral) and/or *in vitro* (receptor binding) assays (Supplementary Table S2)[27–29]. Five of the major odorant receptors in this species that are known to have affinity for ester ligands (Or10a, Or22a, Or35a, Or67a and Or98a) all bind a variety of such esters, with substantially overlapping ranges[30,31]. Notably, many of the alcohol and aldehyde metabolites of these esters are also known ligands for various *D. melanogaster* odorant receptors[32].
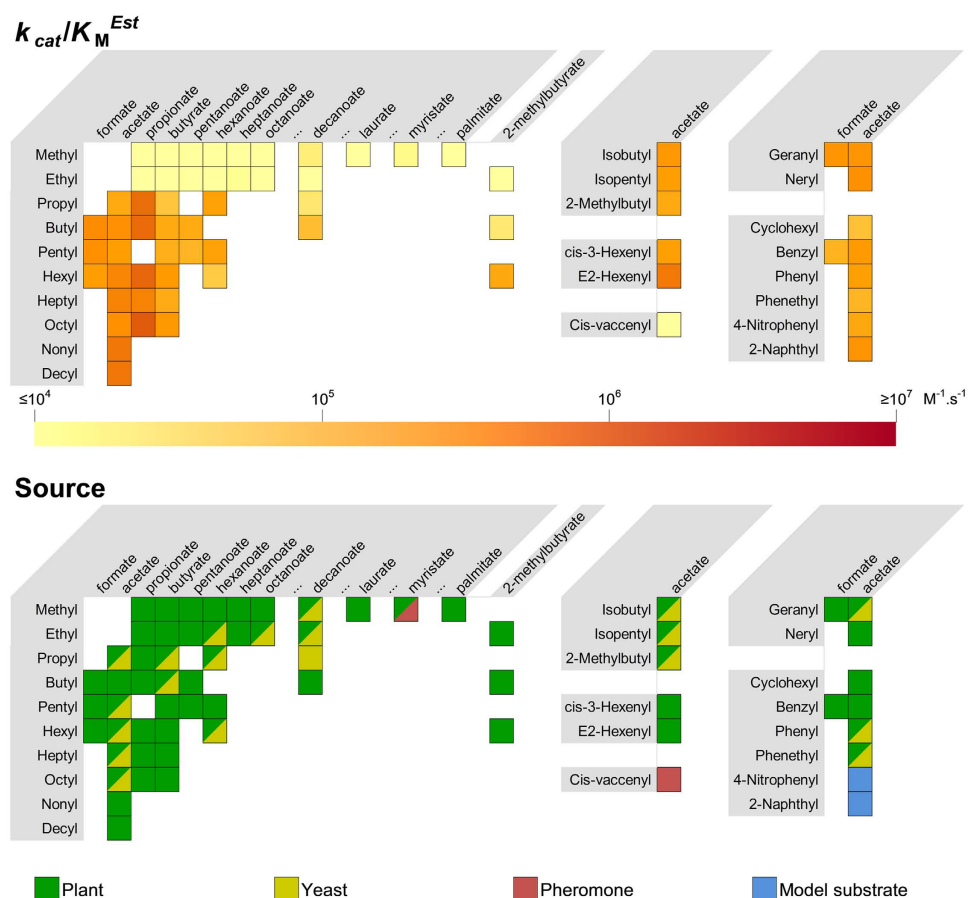
Figure 1. EST6 k_{cat}/K_M^{Est} and biological source of the most active substrates tested and other substrates of particular structural or physiological significance. Alcohol moieties are listed on the vertical and are grouped according to structural similarity. Acid moieties are listed on the horizontal. An ellipsis (…) demarcates a break in an otherwise incremental series. Data on the biological source of the substrates are taken from Supplementary Table S2. Activity results for all 87 compounds tested are given in Supplementary Table S1.

Structure determination of EST6. One of the main barriers to crystallizing EST6 was its very low soluble expression in *Escherichia coli*. To address this, we used the same approach as we did to solve the structures of the α-Esterase 7 carboxylesterase from *Lucilia cuprina*[21]. Briefly, we utilized directed evolution to screen libraries of EST6 variants lacking the N-terminal signal peptide[33] for enhanced activity (as a result of enhanced soluble expression) in *E. coli* (Supplementary Fig. S2). After six rounds of directed evolution, the EST6 variant with greatest soluble expression (EST6-1) contained 16 mutations; K15V, V145L, R208K, G229E, N237S, T247A, D290G, I292F, I335V, E383G, S400G, A416V, F450S, F456S, N485D, I511T (note that amino acids are numbered from the first residue of the mature EST6 protein as it would be processed in its native form within the fly[33] and omits the start methionine included to permit heterologous expression in *E. coli*). Four of these mutations have been found in EST6 from several *Drosophila* species (V145A, R208K, T247A and I292F)[23,34]. Importantly, the catalytic activity of EST6-1 was very similar to that of EST6-WT (Supplementary Fig. S3), suggesting that the 16 mutations principally affected folding, rather than function, consistent with their being located remote from the active site.

Using the Origami B strain of *E. coli*, a cell line that has been designed to enhance disulfide bond formation in the cytoplasm in prokaryotic systems[35], high levels of soluble EST6-1 were expressed (~20 mg. l^{-1}) (Supplementary Fig. S4). Expression of EST6-WT in *E. coli* Origami B cells resulted in substantially lower soluble expression (~0.5 mg. l^{-1}). Size exclusion chromatography showed EST6-1 eluted primarily as a monomer, although there was secondary peak present that indicted a small amount of dimer (Supplementary Fig. S4). Crystallization trials of the EST6-1 monomer fraction at two different concentrations did not yield crystals. We then performed surface lysine methylation, which has been shown to increase the propensity of proteins to crystallize[36], which yielded crystals in conditions of 0.2 M ammonium acetate, 0.1 M Tris pH 8.5 and 25% *w/v* PEG 3,350 that diffracted to 2.1 Å resolution.

The structure of EST6-1 contains 520 amino acids, 353 water molecules, 32 surface carboxylated lysines and one monomer per asymmetric unit. All but the first four N-terminal amino acids are present in reasonable electron density. EST6-1 adopts an α/β-hydrolase fold, including the conserved catalytic triad and oxyanion hole (Fig. 2a,b). The eight-stranded β-sheet (β1–8) surrounded by six α-helices (A-F), that comprises the canonical fold is present, along with the two antiparallel β-strands at the start and two antiparallel β-strands at the end of the structure that are found in the other three insect carboxylesterases whose structures have been solved[20–22]. The
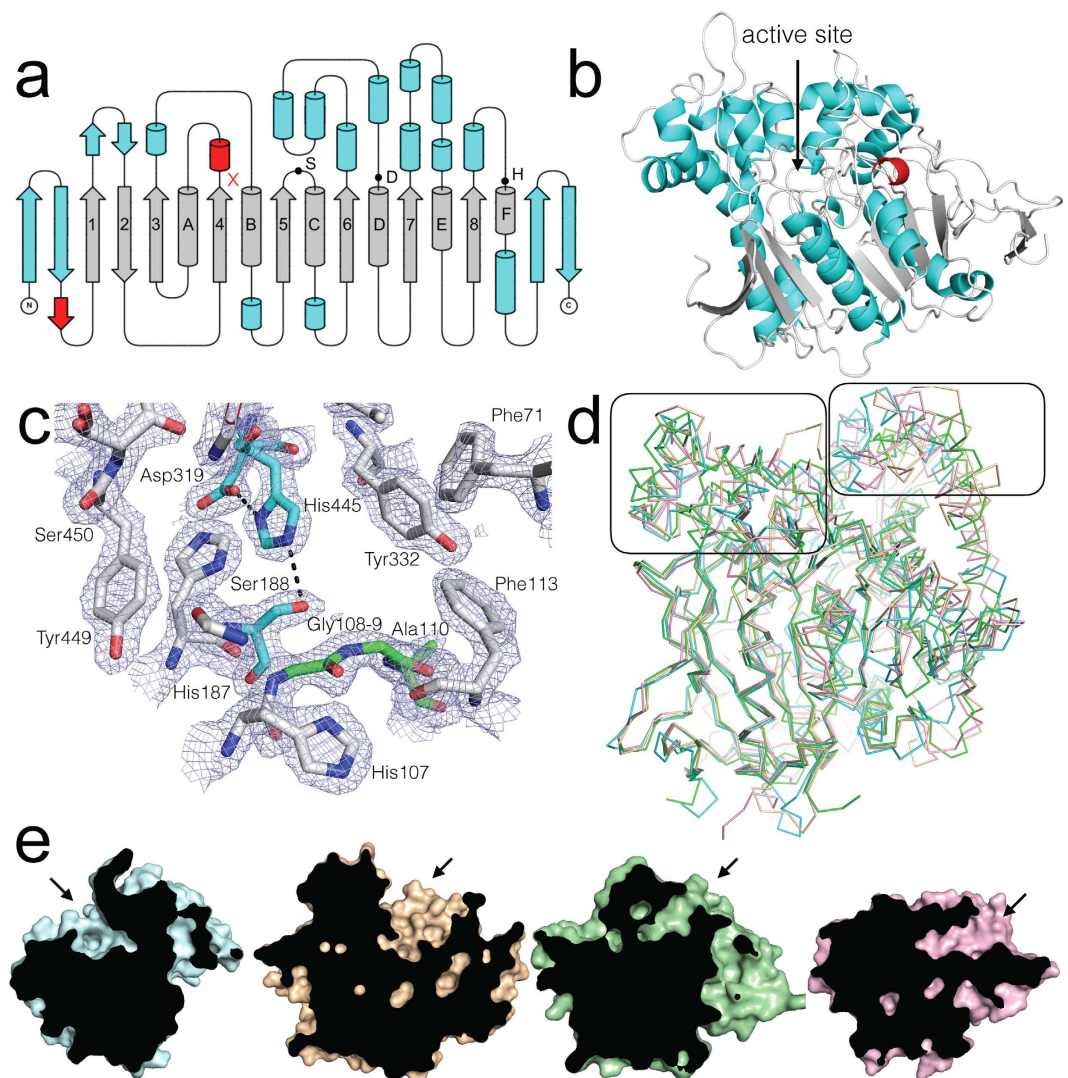
Figure 2. The structure of EST6 from *D. melanogaster*. (**a**) Topology representation of EST6 displaying the conserved α/β-hydrolase fold (grey), secondary structure found in the structurally similar proteins (blue) and unique secondary structure (red). S, D, H represent the Ser188, Asp319 and His445 residues that make up the catalytic triad. The oxyanion hole is located in the loop following sheet 4 (marked by a red x). (**b**) Cartoon diagram of EST6 with features shown in the topology model similarly coloured. The location of the active site is indicated. (**c**) The active site of EST6 with $2mF_o-dF_c$ electron density contoured at 1.5 σ. The active site serine and histidine from the catalytic triad are coloured cyan, the oxyanion hole (Gly108, Gly109, Ala110) is coloured green. (**d**) An overlay of EST6 (cyan), *Lc*αE7 (tan; 4FNM), *Dm*AChE (green; 1QO9) and *Ms*JHE (pink; 2FJ0). Conservation of the core β-sheet and conserved α-helices is apparent, but the structures diverge in the region that forms the active site entrance. These regions, either side of the active site, are boxed for clarity. (**e**) A superposition of EST6, LcαE7, DmAChE and MsJHE, with cut-aways through the middle of the active site. The location of the active site entrance difference between EST6 (cyan) and the other related insect carboxylesterases *Lc*αE7 (tan; 4FNM), *Dm*AChE (green; 1QO9) and *Ms*JHE (pink; 2FJ0).

entrance of the active site is formed by loops following β1, loops and two helices following β4, and loops following β8, including helix F that makes up part of the canonical α/β hydrolase fold. The active site itself is formed from the catalytic triad (Ser188, His445 and Asp319), oxyanion hole (Gly108, Gly109 and Ala189) and additional residues (Tyr322 Tyr449, Phe450, Asn455, Phe456 and Val457) from a helix after β7 and a loop following β5 (Fig. 2c). There are three intramolecular disulfide bonds present (65–84, 240–252, and 493–514) on surface loop regions. The first two disulfide bonds are also seen in the other two insect carboxylesterase structures containing disulfides (AChE, JHE), but the third disulfide is unique to EST6, which also has a shortened C-terminus relative to the other three carboxylesterases structures.

Comparison to known structures. Analysis of the ESTHER database[37], which comprehensively describes the α/β hydrolase fold across a wide range of organisms, reveals that EST6-1 falls into Block C, which also includes the other three known insect carboxylesterase structures. Amongst the insect carboxylesterases,

Protein	Active Site Volume (Å³)	Distance from surface to active site Serine (Å)
EST6 WT FoldX Model	408	15.1
EST6-1 Crystal Structure	935	15.1
*Lc*αE7 (4FNG)	2727	20.2
*Dm*AChE (1QO9)	782	17.2
*Ms*JHE (2FJ0)	1308	18.1
Lipase (1AQL)	3074	17.4

Table 1. Active site volume calculated using the CASTp server.

which Oakeshott *et al.*[18] have divided into 14 Clades, EST6-1 falls into Clade E, with *Lc*αE7 (PDB - 4FNM) in Clade B, *Ms*JHE (PDB - 2FJO) in Clade G and *Dm*AChE (PDB - 1QO9) in Clade J. Application of the SALAMI server[38] confirmed there were no structural homologues in PDB closer to EST6-1 than these three enzymes (Supplementary Table S3). The four clades are well separated from one another phylogenetically (26–29% amino acid identity) but all four structures superimpose well over the canonical fold (2.27, 2.09 and 2.41 Å C-α r.m.s.d. for the other three compared with EST6-1, respectively). In contrast, in the loop regions above the canonical β-sheet and α-helices, there is significant variance between the structures (Fig. 2d).

Closer inspection of the structures and alignment revealed that EST6-1 is missing the C-terminal helix present in *Dm*AChE, *Lc*αE7 and *Ms*JHE. Another feature of interest is the length and composition of the surface-exposed loop regions after strands β1, β6 and β8, which contribute to the active site entrance in the other three proteins. In the latter three, the opening of the active site is formed from helices after β6 and β7 and loops and helices after β1 and β8, but this region has closed over in EST6-1. Its active site entry is instead formed by loops and helices after β1, β4, and β8 on the opposite face of the protein (Fig. 2e). The result is a narrower and shorter active site entrance in EST6-1 in comparison to the open and accessible active site in *Lc*αE7 and the deep gorges leading to the catalytic triads in AChE and JHE.

A comparison of the four structures using the CASTp server[39] also revealed that the active site volume of EST6-1 was significantly less than in *Lc*αE7 (Table 1). The relative sizes of the active sites of *Dm*AChE, *Lc*αE7 and *Ms*JHE reflect their native substrate preferences: *Lc*αE7 natively hydrolyses a wide range of medium chain fatty acid methyl esters and has a large active volume (2727 Å³)[21], while AChE and JHE both have narrower substrate specificities, for the smaller acetylcholine and juvenile hormone molecules respectively, and have much smaller active site volumes, of 782 and 1308 Å³, respectively. The active site volume of EST6-1 is estimated to be 935 Å³, which is consistent with the observed preference of EST6 for smaller substrates than *Lc*αE7 (Fig. 1).

The substrate binding pocket.   Given that EST6-1 is ~97% identical to EST6-WT, and the mutations distinguishing them are all remote from the active site, it is highly likely that the structures will be essentially identical in this region. Nevertheless, for analysis of the substrate binding site, a model of EST6-WT was produced using the empirical structure of EST6-1 and the FoldX force field, which has been developed to allow accurate modeling of point mutations, among other things[40]. As noted above, the conserved catalytic triad of EST6 consists of Ser188, His445 and Asp319, while the backbone NH groups of Gly108, Gly109 and Ala189 create the oxyanion hole (Fig. 2c). His187 is adjacent to the catalytic serine and as with the other three structures its side chain extends into the active site; in the others it has been suggested to affect substrate specificity[26]. EST6 has an asymmetrical binding pocket with a very small, hydrophobic and buried sub-site consisting of Ala110, Trp221, Phe276, Tyr322, Phe397 and His445 that could accommodate the carboxyl group. Opposite this, there is a larger cavity (the putative alcohol leaving group site) that extends into the active site exit/entrance and is slightly less hydrophobic, consisting of Gln70, Phe71, Phe113, Gly114, Gln118, Asn119, Ile429, Tyr449, Phe450, Asn455, Phe456 and Val457 (Fig. 3a).

A representative range of potential substrates that EST6 was tested with were docked into the active site of EST6 using flexible docking with DOCKovelent[41], which is able to screen binding modes for substrates or inhibitors that form covalent bonds with the target enzyme (Fig. 3b, Supplementary Fig. S5 and Supplementary Table S4). The docking results are entirely consistent with the kinetic data, in so much as acylated enzyme intermediates for substrates that were hydrolyzed at significant rates were well accommodated by the substrate binding pocket, whereas no suitable binding poses (without steric clashes) could be obtained for the acylated enzyme intermediates that would result from reaction with compounds that were shown not to be substrates of EST6 (such as cVA). A clear trend is evident: the small sub-site can easily accommodate chains of 1–6 carbons, while the leaving group site has a preference for longer saturated chains, such as hexyl and octyl, over smaller chains, such as methyl and ethyl, but not as large as cVA (C18). This is also consistent with the high activity and complementary binding of geranyl and neryl acetate, with the short carboxyl side chains being accommodated in the small sub-site and the unsaturated leaving group being accommodated in the leaving group site. Likewise, those substrates with aromatic leaving groups and short carboxyl groups are also well accommodated (Supplementary Fig. S5). This analysis provides a molecular explanation for the observed substrate preference for typical food odorants with carboxyl groups of 0–6 carbons and leaving alcohol groups up to ~10 carbons, including branched and aromatic moieties. This structural analysis also strongly supports the kinetic analysis and the initially surprising observation that cVA does not appear to be a physiological substrate for EST6, in that it is clearly far too large for the EST6 substrate binding pocket.
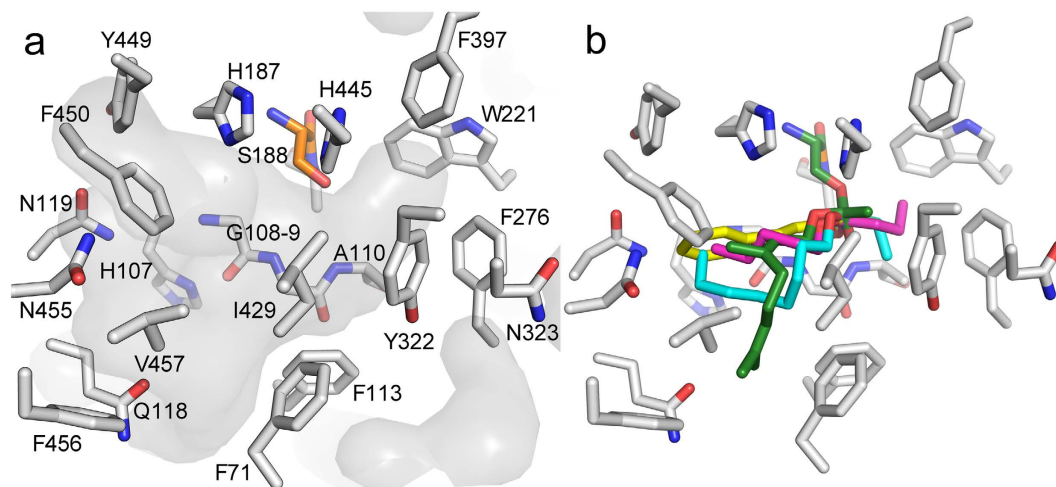
Figure 3. The substrate binding site of EST6. (a) The surface of the substrate binding site is shaded grey and the residues that comprise the small and large pockets are shown (grey) as is the catalytic serine (orange). The small site consists of Ala110, Trp221, Phe276, Tyr322, Phe397 and His445, and the large site consists of Gln70, Phe71, Phe113, Gly114, Gln118, Asn119, Ile429, Tyr449, Phe450, Asn455, Phe456, and Val457. **(b)** An overlay of representative acylated enzyme intermediates covalently docked into EST6: the efficiently hydrolyzed substrates pentyl butyrate (magenta), octyl propionate (cyan), geranyl acetate (green) and phenethyl acetate (yellow) all produce acylated intermediates that are accommodated by the substrate binding site.

Localization of EST6 in the antennae. *Est6* is known to be highly expressed in the antenna[8], in particular in the third antennal segment[10], but its expression in this tissue at the cellular level was unknown. Labelling of EST6 with anti-EST6 antibody and of Red Fluorescent Protein (RFP) under the control of the *Orco* promoter (*Orco* encodes the universal odorant co-receptor Orco) in transgenic adults showed EST6 immunoreactivity in numerous cells at the base of olfactory sensilla throughout the third antennal segment whereas, as expected[42], the *Orco* promoter directed expression in numerous olfactory receptor neurons (ORNs) and in cilia entering the sensillar lumen (Fig. 4). As was earlier suggested by Chertemps *et al.*[10], there was thus no co-localization of the two signals, showing that EST6 is not expressed in ORNs. Similarly, a complementary experiment showed no co-localization of EST6 and the neuron-specific expression of Green Fluorescent Protein (GFP) under the control of the *elav* promoter[43] (Supplementary Fig. S7). Given that EST6 is a secreted enzyme, this confirms that the enzyme surrounds the Orco+ dendrites within the sensillar lymph of various sensilla.

Co-labelling of EST6 and *lush* was then performed to investigate whether the location of EST6 in the sensillar lymph includes the T1 trichoid sensilla involved in cVA detection. LUSH is known to be expressed in all trichoid types[44]. Labelling of EST6 with anti-EST6 antibody and of RFP under the control of the *lush* promoter in transgenic adults found that both signals were closely associated but with no co-localization of the two. RFP was found at the base of trichoid sensilla in accessory cells (Supplementary Fig. S8) that could correspond to trichogen and tormogen cells[44], whereas EST6 was apparently produced by different support cells for the trichoid sensilla than the LUSH-producing cells[45], and possibly also by the epidermal cells surrounding the sensilla. To corroborate this result we also performed RNAi knock-down experiments. These results are also consistent with *Est6* is not being co-expressed with *lush* (Supplementary text).

Altogether, these data show that EST6 is produced by non-neuronal cells in the olfactory sensilla, most probably in a large population of accessory cells surrounding ORNs. It localization in the sensillar lymph is compatible with a function of a general ODE in the basiconic sensilla involved in the detection of almost all the substrates tested here[46]. Its function in the T1 trichoid sensilla is not yet clear but its effect on cVA processing in the absence of any direct hydrolytic activity for the compound may reflect a general scavenging role for other ester odorants which might otherwise impede the processing of cVA by its own, as yet unknown, ODE. It is possible that it also plays an equivalent broad scavenging role in some of the other sensilla where it is abundant, although its strong hydrolytic activity for many ester kairomones suggests it has a direct ODE function for several of them.

Discussion

Notwithstanding the genetic evidence that EST6 contributes to cVA processing *in vivo*[10], we find that the enzyme has negligible activity ($<1.5\,M^{-1}.s^{-1}$) for this substrate *in vitro*, with or without LUSH in the assay mix. Our results in fact confirm the only other direct measure of its *in vitro* activity, by Mane *et al.*[9]; their estimation of 55 picomoles of cVA per min per g of purified EST6, or $3.4\,M.\,min^{-1}.M^{-1}$, (in the absence of LUSH) is in the range that was too low to measure accurately in our assays. We concur with Vandermeer *et al.*[17] that activity in this range is most unlikely to be physiologically relevant. This indicates that the *in vivo* effects of EST6 on cVA processing seen by Chertemps *et al.*[10] must be indirect.

While we found that EST6 had low activity against cVA, it clearly has physiologically significant ($k_{cat}/K_M > 10^5\,M^{-1}.s^{-1}$)[25] activity with a wide range of esters with acyl chains up to six carbons in length and alcohol groups from mid length (3–10 carbon atoms), aliphatic moieties to branched, secondary, unsaturated,
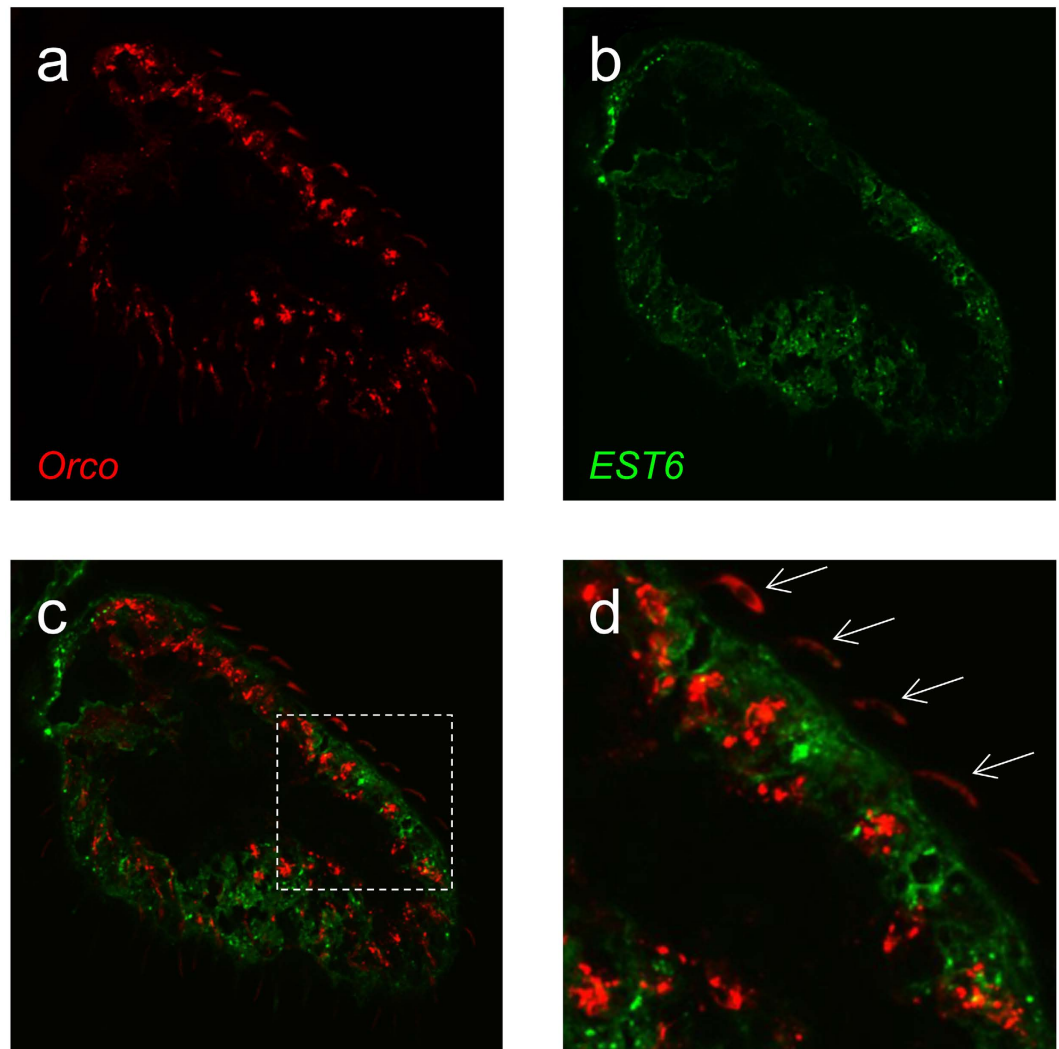
Figure 4. EST6 and *Orco* expression in the third antennal segment, longitudinal sections. (**a**) Membrane-tethered RFP expressed with the *Orco* promoter (*Orco^Gal4^* /*UAS-mCD8::RFP* transgenic flies). (**b**) EST6 protein localization in the same section. (**c**) Merge image of (**a**,**b**): Est-6 and *Orco* are not expressed in the same cells. (**d**) Higher magnifications of (**c**): EST6 protein surrounds the Orco+ dendrites. Arrows indicate the dendrites of *Orco* expressing ORNs. Western blots and immunohistochemistry showing the specificity of the anti-EST6 antibody are shown in Supplementary Fig. S6.

cyclic and aromatic groups. These substrates include many fruit and yeast volatiles that are known to be bioactive against *Drosophila*, consistent with the results of electrophysiological and behavioural comparisons of wildtype and EST null flies by Chertemps *et al.*[10], which show that the enzyme contributes to the processing of many such molecules *in vivo*. As such, our biochemical data support the proposition that EST6 is a general, rather than specific, odorant degrading enzyme (ODE), but with a substrate range tuned to various volatile esters with relatively short chain acyl groups that are commonly emitted by the food sources for the flies.

Significantly, the bioactivity of many of these better substrates for EST6 involves attraction behaviours[47]. For example, fruity smelling acetate esters such as isopentyl and pentyl acetate, which are produced by both plants and yeasts, are highly attractive to *Drosophila*[48], wherein they activate several fairly broadly tuned odorant receptors, such as Or43b, Or47a and Or85b[46,49]. Likewise, the phenolic yeast volatile phenethyl acetate elicits an attraction response from the fly[50] and activates its Or85d receptor[49]. Notably, some of these attraction behaviors also manifest as effects on reproductive traits; for example, citrus fruits emit many short-mid chain volatile acetates (e.g. propyl, hexyl, heptyl, nonyl, decyl, neryl and geranyl acetates[51]), which attract females to lay eggs[28].

It has been shown that several food odors, including ester substrates for EST6, can act synergistically with cVA in both aggregation and courtship bioassays[52–55]. Indeed, some evidence suggests that cVA only acts as an aggregation pheromone in the presence of attractive food odors[55]. It is suggested that the co-processing of pheromonal and kairomonal stimuli would help coordinate feeding and oviposition site selection with reproductive behaviors[53]. However, we cannot see how this synergism would explain the indirect effects of EST6 activity on cVA processing observed by Chertemps *et al.*[10]. One reason is that the experimental design of that previous study meant

that food odors would not have been present in the cVA atmospheres they tested. Furthermore, the co-processing of the signals from cVA and the food odors must occur downstream of their receptors, since they have different receptors and the signals from their receptors are transmitted to different glomeruli in the brain, but the effects of EST6 on EAG responses to cVA seen by Chertemps et al.[10] must occur prior to or at the time when the cVA interacts with its receptors. Other indirect effects of EST6 on cVA processing must therefore explain the data of Chertemps et al.[10]. For example, as noted above, EST6 may facilitate cVA processing simply by removing other potential substrates (or inhibitors) of the ODE that does degrade cVA. As noted, the latter ODE may be a lipase, and indeed, with an 18-carbon leaving group, cVA is more like a typical lipase substrate than an esterase substrate. Our localization studies would certainly allow for that, given the broad distribution of the enzyme through the sensillar lymph. Further work is needed to elucidate the molecular basis for the effects seen by Chertemps et al.[13].

Our biochemical studies also bear on the question of the molecular basis for the effects on female oviposition and remating behaviors due to the ejaculatory duct EST6 transferred from their mates[15,16]. This enzyme is known to be transferred from the female's reproductive tract to her haemolymph within minutes of mating[14], but its fate from there and its substrate in the female are unknown. Our results indicate that a wide variety of esters of terpene or aromatic alcohol groups and short-mid chain acids could be candidate substrates. Notably, some such compounds are precursors for various hormones and other key molecules in the fly[56,57]. Modern metabolomic technologies may be useful in identifying the *in vivo* substrate for the transferred EST6, particularly given the availability of the *Est6°* flies and wildtype revertants on the same genetic background[10].

EST6, in Clade E of the carboxylcholinesterase gene family, is not closely related in sequence (26–29% amino acid identity) to any of the three insect esterases for which structures have been solved previously (in Clades B, G and J). While it's overall structure is similar to the other three, we noted several significant differences in relation to its active site. Of particular note was the appearance of an active site entrance on the opposite face of the protein to that containing the active site entrance in the other three structures. Interestingly, the entrance in EST6 corresponds to the alternative 'back door' entrance that has been proposed for AChE[58]. Moreover, the corresponding surface of the catalytically inactive ligand-binding 'esterase' neuroligin is the site to which its ligand binds[59].

Transcriptomic analyses of sensory tissues in various insects have shown as many as half of the catalytically competent carboxyl/cholinesterases in some insects may be expressed at readily detectable levels in their sensory tissues[8,60]. The few for which there is any empirical support for ODE functions have spanned four major Clades (A, D, E and G)[18,61], suggesting that esterase ODEs may have evolved independently on several occasions. However, there is a concentration of putative esterase ODEs in the particular lineage within Clade E that contains EST6 (31% amino acid identity)[18]. This lineage contains esterases from at least four insect orders, including one of the best-understood ODE's at a physiological level, the Apo1PDE from the silkmoth *Antheraea polyphemus*. Apo1PDE is highly specific ODE for a particular sex pheromone substrate[62], whereas we find EST6 has both broad activity for many kairomones and an indirect effect on cVA processing whose mechanism we currently do not understand. Further work on this lineage could elucidate a range of biochemical, physiological and evolutionary phenomena concerning the function of esterases in insect antennae.

Methods

EST6 activity assays. The expression of wildtype EST6 and an inactive EST6 variant in the baculovirus system has been described previously[13]. These two enzymes were assayed here for activity against 85 ester odorants of potential ecological relevance[49,63,64] and two other model substrates (listed in Supplementary Table S1). All these esters were purchased in the highest available purity.

Eighty two of the esters were first subjected individually to gas chromatography-mass spectrometry (GC-MS, 7890 series, Agilent Technologies, USA) to determine their respective retention times. A J&W DB-WAX column (30 m × 0.25 mm × 0.25 μm, Agilent Technologies, USA) was used with He (2 ml . min^{-1}) as the carrier gas. The oven temperature was initially set at 50 °C for 2 mins and then subsequently increased over a gradient of 10 °C to 275 °C and held for 10 mins. The injector and detector temperature was set at 250 °C with a 10:1 split ratio.

Mixtures of up to 17 compounds with non-overlapping GC-MS retention times were then made in Tris HCl buffer pH 8.0 for a set of preliminary 'group assays'. Each group included pentyl acetate as a common ester substrate standard. All compounds had been dissolved in ethanol to give a 5% *v/v* final solvent concentration; preliminary assays on some of the more water-soluble esters showed that this ethanol concentration increased K_M by 5–20 fold (see below) but lower concentrations of ethanol were insufficient to solubilize some compounds and equivalent concentrations of other organic solvents tested were more disruptive to EST6 activity. Several reactions were set up at 25 °C in Tris-HCl buffer pH 8.0 with each ester in the mixture at a final concentration of 200 μM and the enzyme (added last) at 8.2 nM. Individual reactions were then stopped by the addition of 0.5 volumes of ice-cold hexane containing 200 μM heptanone as an internal non-ester standard at intervals from 5 to 65 mins. The concentrations of the various esters remaining were then determined by GC-MS as above. EST6 activity was calculated from the difference in substrate usage between the wildtype and null enzymes, but all values for the latter were essentially negligible.

Subsequently, 43 substrates from the group assay, including all the better substrates, were assayed individually in order to obtain estimates of k_{cat}/K_M using equation (1):

$$k_{cat}/K_M{}^{Est} = V_0/([E] . [S]) \tag{1}$$

where [E] and [S] are the starting enzyme and substrate concentrations respectively, and V_0 is the initial velocity of the reaction[65].

Aside from the single substrate, these assays were the same as those for the group assays, except that a lower enzyme concentration was used (0.1 to 3.6 nM). The appropriate enzyme concentration was inferred from the enzyme's activity towards each substrate in the group assay. Three other esters that were not included in the

group assays but with closely similar chain lengths and structures to some of the best substrates were also assayed individually in this way. Individual substrate assays with two model esters, 4 NPA and 2 NA, were also carried out using previously described 420 and 390 nm UV/vis protocols for monitoring substrate loss[8,21]. K_M estimates could be obtained from these data for a few substrates and a few were also obtained using the competitive inhibition method with 4 NPA as substrate as described in Younus *et al.*[8]. All the above assays were conducted in triplicate.

Assays with LUSH. Some assays were also conducted in the presence of the odorant binding protein LUSH. In preparation for this the *lush* coding region was synthesized by Invitrogen and cloned into the expression vector pETMCSI[66]. The LUSH protein was overexpressed in inclusion bodies of *E. coli* BL21 (DE3) star (Invitrogen) cells after overnight growth in Lysogeny Broth (LB) broth containing 100 mg.l^{-1} of ampicillin at 37 °C. The cells were harvested by centrifugation at 5,000 g for 5 min at 4 °C, the cells lysed by three passages through a French Press, and the inclusion bodies collected by centrifugation at 10,000 g for 20 min at 4 °C. The inclusion bodies were then solubilized and refolded following the method of Kruse *et al.*[67] using a cysteine-cystine redox reaction in the presence of 1% *v/v* ethanol. The only modifications to this method were that 8 M urea was used to solubilize the inclusion bodies and the soluble protein was dialyzed in 20 mM Tris pH 7.4, 50 mM NaCl. The soluble LUSH was further purified by using a Superdex 200 preparation size exclusion column (GE Healthcare, UK) and assayed for binding activity with the model ligand N-phenyl-1-naphthylamine (NPN) according to the method described by Katti *et al.*[68]. This involved titrating LUSH (1 μM) with increasing amounts of NPN to final concentrations ranging from 0.5 μM to 20 μM. A saturable NPN fluorescence change was recorded by a fluorometer and the dissociation constant was found to be 2.39 μM. Katti *et al.*[68] showed that LUSH does not display a saturable NPN fluorescence change if it is not fully refolded.

Assays to investigate the activity of EST6 towards cVA in the presence of LUSH were set up the same as those for the group assays except for changes to the substrate (150 μM) and enzyme (3 and 57 nM) concentrations, and the addition of LUSH (300 μM). Duplicate reaction mixtures were set up without LUSH as controls. Equivalent reactions using a better, mid-chain ester substrate, decyl acetate, were also set up as further controls.

Protein engineering and expression. Six generations of directed evolution were undertaken to improve the soluble expression of *E. coli*-expressed wildtype EST6. The method followed Jackson *et al.*[21], but in this case the coding region of *Est6* from the iso-1 $y^1cn^1bw^1sp^1$ reference strain (http://flybase.org/reports/FBsn0000272.html), omitting the 63 bp encoding the N-terminal signal peptide[33], was cloned into the expression vector pETMCSIII[66] between the *Nde*I and *Eco*RI sites in frame with the ATG start codon of the *Nde*I site. Adequate expression of *Est6* could be achieved by 'leaky expression' because of the presence of trace amounts of lactose in the LB media used. The error-prone PCR protocol used to construct the initial mutant library involved a reaction mixture comprising 100–200 ng of pETMCSIII-*Est6*, 1 μM primers pET3 and pET4 (5′CGACTCACTATAGGGAGACCACCAC3′ and 5′CCTTTCGGGCTTTGTTAGCAG3′), 1 × Taq DNA polymerase buffer, 5 mM MgCl$_2$, 0.1–0.4 mM MnCl$_2$, 0.5 mM dNTPs, 5U Taq DNA polymerase, and milliQ H$_2$O to a final volume of 50 μl. Thermocycling involved 30 cycles of 94 °C for 10 s, 45 °C for 10 s and 30 s at 72 °C. The *Nde*I- and *Eco*RI-digested PCR product was gel extracted, ligated back into pETMCSIII, and then used to transform competent BL21 (DE3) star cells. Transformed cells were plated onto LB plates containing 100 mg.l^{-1} ampicillin. After incubation at 30 °C overnight, the colonies were blotted onto 3 M filter papers and esterase activity was assayed by staining the filter paper with a solution consisting of 10 ml of 0.1% *w/v* Fast Red and 0.2 ml of 1% *w/v* 2 NA in 0.1 M Tris pH 7.0. Between 200–300 (approximately 1%) of the colonies generating the most intense red colour were then picked by hand and grown overnight in 500 μl of LB, 100 mg.l^{-1} ampicillin, in 96-well culture plates. 50 μl of each of these cultures was then added to the corresponding well of a 96-well assay plate that contained 250 μl of a reaction mixture consisting of 0.5 mM 2 NA, 0.5 mM Fast Red, and 0.1 M Tris pH 7.0. The reaction was monitored with a spectrophotometer at 490 nm, and the 10–20 colonies generating the highest activities were sequenced and used as parents for the next generation of mutation and selection. The protocols for generations 2 to 6 followed those above. The sixth generation mutant generating the highest activity in the spectrophotometric assay, denoted EST6-1, was used for crystallization.

EST6-1 Crystallization and Computational Analysis. EST6-1 was expressed in *E. coli* Origami B (DE3) pLysS Cells (Merck) grown in LB media with 100 μg.ml^{-1} ampicillin to an optical density of 0.6. The cells were induced with 700 μM IPTG and harvested after 18 hours at 25 °C. The cells were then lysed by sonication in 50 mM Hepes (pH 7.5), 300 mM NaCl, 10 mM imidazole (buffer A). The soluble lysate was separated by centrifugation at 23,000 g and filtered with a 0.45 μM filter before being loaded onto a 5 ml Ni-NTA column. The protein was eluted from the column with buffer A supplemented with 300 mM imidazole. Fractions were pooled after confirmation by SDS-PAGE and further purified by size exclusion chromatography in 20 mM Hepes (pH 7.5) and 50 mM NaCl (buffer B) using a Hiload 26/600 Superdex 200 pg column (GE Healthcare). The concentration of EST6-1 was determined at 280 nm with the Nanodrop 1000 (Thermo Scientific) using an extinction coefficient of 74,635 M^{-1} cm^{-1} estimated by the ProtParam server[69].

Surface lysine residues of purified EST6-1 (1 mg. ml^{-1}) were methylated following the protocol of Walter *et al.*[36], and the reaction was quenched with 1 M glycine, followed by concentration of methylated EST6-1 to 18.2 mg. ml^{-1} and dialysis into buffer B. Crystals of methylated EST6 were grown by the sitting drop diffusion technique with a reservoir solution containing 0.2 M ammonium acetate, 0.1 M Tris pH 8.5 and 25% *w/v* PEG 3,350. 35% *w/v* PEG 3,350 was used as a cyroprotectant during flash cooling of the crystals in nitrogen at 100 K. Diffraction data were collected at the MX2 beamline at the Australian Synchrotron, Victoria, Australia with a wavelength of 0.9655 Å. Data collection methods and statistics as well as details of the informatics methods used to solve the enzymes structure are given in Supplementary Table S5.

A model EST6-WT was built from the 97% identical structure of EST6-1 using FoldX[40]. Covalent docking was performed with DOCKovalent, a covalent version of DOCK3.7[41]. The program pre-generates a set of conformations for each ligand, covalently attaches the ligand to a receptor, and exhaustively samples ligand orientations around the covalent bond. Ligands are then ranked via a physics-based scoring function. Esters investigated in this work were represented as SMILES strings with the covalent attachment to the catalytic serine $O\gamma$ marked with a dummy atom. The esters were docked in the form of a tetrahedral intermediate, after nucleophilic addition of the serine $O\gamma$ to the carbonyl carbon and prior to departure of the alcohol, with the carbonyl oxygen bearing a negative charge. The generation of ligand conformations and preparation of the receptor (EST6-WT model) was carried out as described previously[41]. The catalytic histidine was represented in its doubly protonated form. The selected esters were covalently docked onto the Ser188 $O\gamma$ with a $O\gamma$-ligand bond length of 1.6 ± 0.1 Å sampled at 0.05 Å increments and with the $C\beta$-$O\gamma$-ligand and $O\gamma$-ligand-ligand bond angles set to $109.5 \pm 5°$ and sampled at 1° increments. The lowest energy pose for each ligand was selected for analysis. Protein structure images were produced with PyMol V 1.3 and a topology diagram was generated using TOPDRAW[70].

Immunohistochemistry.
Flies. *Orco*[Gal4] flies were generously provided by G. Galizia (University of Konstanz, Germany), *lush*[Gal4] flies (originally from R Benton, Université de Lausanne, Switzerland) from J-F. Ferveur (CSGA, Dijon, France) and *elav*[LexA], *LexAOP-mCD8::GFP*, *UAS-mCD8::RFP* and *Est6* null mutant flies from the Bloomington Stock Center (stocks 52676, 32203, 27392 and 4211 respectively). All flies were raised at 25 °C on standard yeast/cornmeal/agar medium in a 12-hr light/12-hr dark cycle, 50–60% relative humidity.

Generation of anti-EST6 antiserum. Preparation of denatured EST6 antigen and production of polyclonal antibody followed the methods of Han *et al.*[71]. Briefly, wildtype EST6 was overexpressed in inclusion bodies in *E. coli* using the expression vector pETMCS III as above. Cells were harvested and lysed and inclusion bodies collected as above. The latter were then dissolved in 6 M guanidine HCl in a buffer containing 20 mM phosphate pH 7.4 and the solubilized denatured proteins loaded onto a 5 ml Ni-NTA column. The EST6 was eluted from the column with a gradient of buffer containing 6 M guanidine HCl, 20 mM phosphate, 0.5 M imidazole, pH 7.4. Fractions containing EST6 were identified from the presence of a 59.7 kDa band on denaturing PAGE and then pooled and loaded onto a Superdex 200 preparative scale exclusion column (GE Healthcare) equilibrated with 6 M of guanidine HCl, 20 mM phosphate buffer, pH 7.4. EST6 fractions from this column were concentrated to 1 mg.ml^{-1} using an Amicon Ultra-15 centrifugal device (Millipore, US) and the guanidine HCl removed by dialysis in 20 mM phosphate buffer, pH 7.4. The purified denatured EST6 was used as antigen for polyclonal antibody production by IMVS Veterinary Services, South Australia. Four doses of 0.5 mg antigen were administered to a rabbit at 3 weekly intervals. The polyclonal antibodies were purified from antiserum using an IgG affinity column and the protein concentration was estimated at 3 mg.ml^{-1}.

The specificity of the antiserum was then tested by western blotting against extracts from heads of wildtype (*Canton S*) and *Est6°* null mutant flies. Mass homogenates of heads from each strain in 20 mM Tris-HCl buffer (pH 7.4) were briefly sonicated, centrifuged at 8,000 g for 10 min and the supernatants isolated. Twenty μg of protein from each homogenate were then separated by SDS-PAGE and blotted onto a polyvinylidene fluoride (PVDF) membrane. After blocking in Tris Buffered Saline-Tween 10% (TBST-10%) blocking reagent (Invitrogen), membranes were incubated overnight at 4 °C with the anti EST6 antibody (1:3,000), then incubated with rabbit-peroxidase-labelled antibody (1:10,000). Blots were then washed and incubated with chemiluminescent substrate (ECL Plus Western Detection Kit, GE Healthcare).

Localization of EST6 within antennae.
To localize EST6 in the antenna, we performed immunohistochemistry with the anti-EST6 antibody above on transgenic flies expressing RFP under the control of either the *Orco* or *lush* promoter or GFP under the control of the *elav* promoter. *Est6* null mutant flies were used as a control for the specific labelling of the antibody. Specifically, heads with antennae still attached from 5-day-old *Orco*[Gal4]/*UAS-mCD8::RFP*, *elav*[LexA]/*LexAOP-mCD8::GFP*, *lush*[Gal4]/*UAS-mCD8::RFP* or *Est6* null mutant males were fixed for 3 h in 4% paraformaldehyde with 0.2% Triton X-100, then washed for 1 h with PBS containing 0.2% Triton X-100 (PBST). The heads were then embedded in Tissue-Tek™ (CellPath) and cryosections (15 μm) were set in cell culture insert (Greiner Bio-one). After blocking with 3% normal goat serum and 1% BSA in PBST (1 h at room temperature), the anti-EST6 antibody was diluted from 1:3,000 to 1:750 (*v:v*) in the blocking solution (3% normal goat serum in PBST) and incubated overnight at room temperature. After a brief rinse in PBST, an anti-mouse conjugated Alexa-488 or Alexa-596 (Invitrogen) was applied at a concentration of 1:800 (*v:v*) in the blocking solution for 4 h at room temperature. Tissues were mounted in Slowfade reagent with DAPI (Invitrogen). Images were captured on a Leica SP5 confocal microscope and analysed using ImageJ 1.47 v (http://imagej.nih.gov/ij).

References
1. Carey, A. F. & Carlson, J. R. Insect olfaction from model systems to disease control. *Proc. Natl. Acad. Sci. USA* **108,** 12987–12995, doi: 10.1073/pnas.1103472108 (2011).
2. Leal, W. Odorant reception in insects: Roles of receptors, binding proteins, and degrading enzymes. *Annu. Rev. Entomol.* **58,** doi: 10.1146/annurev-ento-120811-153635 (2013).
3. Durand, N. *et al.* Degradation of pheromone and plant volatile components by a same odorant-degrading enzyme in the cotton leafworm, Spodoptera littoralis. *PLoS One* **6,** doi: 10.1371/journal.pone.0029147 (2011).
4. Vogt, R. G. Molecular Basis of Pheromone Detection in Insects. In *Comprehensive Molecular Insect Science* Vol. 3 (ed. Gilbert, L. I.) 753–803 (Elsevier, 2005).
5. Vogt, R. G., Callahan, F. E., Rogers, M. E. & Dickens, J. C. Odorant binding protein diversity and distribution among the insect orders, as indicated by LAP, an OBP-related protein of the true bug Lygus lineolaris (Hemiptera, Heteroptera). *Chem. Senses* **24,** 481–495, doi: 10.1093/chemse/24.5.481 (1999).

6. Vogt, R., Riddiford, L. & Prestwich, G. Kinetic properties of a sex pheromone-degrading enzyme: the sensillar esterase of Antheraea polyphemus. *Proc. Natl. Acad. Sci. USA* **82,** doi: 10.1073/pnas.82.24.8827 (1985).

7. He, P. *et al.* Functional characterization of an antennal esterase from the noctuid moth, Spodoptera exigua. *Arch. Insect Biochem. Physiol.* **86,** doi: 10.1002/arch.21164 (2014).

8. Younus, F. *et al.* Identification of candidate odorant degrading gene/enzyme systems in the antennal transcriptome of *Drosophila melanogaster*. *Insect Biochem. Mol. Biol.* **53,** doi: 10.1016/j.ibmb.2014.07.003 (2014).

9. Mane, S. D., Tompkins, L. & Richmond, R. C. Male esterase 6 catalyzes the synthesis of a sex pheromone in *Drosophila* melanogaster females. *Science* **222,** doi: 10.1126/science.222.4622.419 (1983).

10. Chertemps, T. *et al.* A carboxylesterase, Esterase-6, modulates sensory physiological and behavioral response dynamics to pheromone in *Drosophila*. *BMC Biol.* **10,** doi: 10.1186/1741-7007-10-56 (2012).

11. Gomez-Diaz, C., Reina, J., Cambillau, C. & Benton, R. Ligands for pheromone-sensing neurons are not conformationally activated odorant binding proteins. *PLoS Biol.* **11,** doi: 10.1371/journal.pbio.1001546 (2013).

12. Benton, R. Sensitivity and specificity in *Drosophila* pheromone perception. *Trends Neurosci.* **30,** 512–519, doi: 10.1016/j.tins.2007.07.004 (2007).

13. Chertemps, T. *et al.* An antennal carboxylesterase from *Drosophila* melanogaster, esterase 6, is a candidate odorant-degrading enzyme toward food odorants. *Front. Physiol.* **6,** doi: 10.3389/fphys.2015.00315 (2015).

14. Richmond, R., Gilbert, D., Sheehan, K., Gromko, M. & Butterworth, F. Esterase 6 and reproduction in *Drosophila* melanogaster. *Science* **207,** doi: 10.1126/science.6767273 (1980).

15. Gilbert, D. Ejaculate esterase 6 and initial sperm use by female *Drosophila* melanogaster. *J. Insect Physiol.* **27,** doi: 10.1016/0022-1910(81)90112-8 (1981).

16. Scott, D. Inhibition of female *Drosophila* melanogaster remating by a seminal fluid protein (Esterase 6). *Evolution* **40,** doi: 10.2307/2408766 (1986).

17. Vander Meer, R., Obin, M., Zawistowski, S., Sheehan, K. & Richmond, R. A reevaluation of the role of cis-vaccenyl acetate, cis-vaccenol and esterase 6 in the regulation of mated female sexual attractiveness in *Drosophila* melanogaster. *J. Insect Physiol.* **32,** doi: 10.1016/0022-1910(86)90109-5 (1986).

18. Oakeshott, J. G., Claudianos, C., Campbell, P. M., Newcomb, R. D. & Russell, R. J. Biochemical Genetics and Genomics of Insect Esterases. In *Comprehensive Molecular Insect Science* Vol. 5 (ed. Gilbert, L. I.) 309–381 (Elsevier, 2005).

19. Rane, R. V. *et al.* Are feeding preferences and insecticide resistance associated with the size of detoxifying enzyme families in insect herbivores? *Curr. Opin. Insect Sci.* **13,** 70–76, doi: 10.1016/j.cois.2015.12.001 (2016).

20. Wogulis, M. *et al.* Structural studies of a potent insect maturation inhibitor bound to the juvenile hormone esterase of Manduca sexta. *Biochemistry* **45,** 4045–4057, doi: 10.1021/bi0521644 (2006).

21. Jackson, C. J. *et al.* Structure and function of an insect α-carboxylesterase (αEsterase7) associated with insecticide resistance. *Proc. Natl. Acad. Sci. USA* **110,** 10177–10182, doi: 10.1073/pnas.1304097110 (2013).

22. Harel, M. *et al.* Three-dimensional structures of *Drosophila* melanogaster acetylcholinesterase and of its complexes with two potent inhibitors. *Protein Sci.* **9,** doi: 10.1110/ps.9.6.1063 (2000).

23. Oakeshott, J. G. *et al.* An episode of accelerated amino acid change in *Drosophila* esterase-6 associated with a change in physiological function. *Genetica* **110,** 231–244, doi: 10.1023/A:1012727814167 (2000).

24. Becher, P. *et al.* Yeast, not fruit volatiles mediate *Drosophila* melanogaster attraction, oviposition and development. *Funct. Ecol.* **26,** doi: 10.1111/j.1365-2435.2012.02006.x (2012).

25. Bar-Even, A. *et al.* The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **50,** 4402–4410, doi: 10.1021/bi2002289 (2011).

26. Myers, M. A., Healy, M. J. & Oakeshott, J. G. Effects of the residue adjacent to the reactive serine on the substrate interactions of *Drosophila* esterase 6. *Biochem. Genet.* **31,** 259–278, doi: 10.1007/bf00553170 (1993).

27. Dweck, H. *et al.* Pheromones mediating copulation and attraction in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **112,** doi: 10.1073/pnas.1504527112 (2015).

28. Dweck, H. *et al.* Olfactory preference for egg laying on citrus substrates in *Drosophila*. *Curr. Biol.* **23,** doi: 10.1016/j.cub.2013.10.047 (2013).

29. Galizia, C., Münch, D., Strauch, M., Nissler, A. & Ma, S. Integrating heterogeneous odor response data into a common response model: A DoOR to the complete olfactome. *Chem. Senses* **35,** doi: 10.1093/chemse/bjq042 (2010).

30. Mansourian, S. & Stensmyr, M. The chemical ecology of the fly. *Curr. Opin. Neurobiol.* **34,** doi: 10.1016/j.conb.2015.02.006 (2015).

31. Münch, D. & Galizia, C. DoOR 2.0 - Comprehensive mapping of *Drosophila* melanogaster odorant responses. *Sci. Rep.* **6,** doi: 10.1038/srep21841 (2016).

32. Mathew, D. *et al.* Functional diversity among sensory receptors in a *Drosophila* olfactory circuit. *Proc. Natl. Acad. Sci. USA* **110,** 43, doi: 10.1073/pnas.1306976110 (2013).

33. Cooke, P. H. & Oakeshott, J. G. Amino acid polymorphisms for esterase-6 in *Drosophila* melanogaster. *Proc. Natl. Acad. Sci. USA* **86,** 1426–1430 (1989).

34. Karotam, J., Delves, A. C. & Oakeshott, J. G. Conservation and change in structural and 5′ flanking sequences of esterase 6 in sibling *Drosophila* species. *Genetica* **88,** 11–28 (1993).

35. de Marco, A. Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli. *Microb. Cell Fact.* **8,** doi: 10.1186/1475-2859-8-26 (2009).

36. Walter, T. *et al.* Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* **14,** doi: 10.1016/j.str.2006.09.005 (2006).

37. Lenfant, N. *et al.* ESTHER, the database of the α/β-hydrolase fold superfamily of proteins: tools to explore diversity of functions. *Nucleic Acids Res.* **41,** doi: 10.1093/nar/gks1154 (2013).

38. Margraf, T., Schenk, G. & Torda, A. The SALAMI protein structure search server. *Nucleic Acids Res.* **37,** doi: 10.1093/nar/gkp431 (2009).

39. Binkowski, A. T., Naghibzadeh, S. & Liang, J. CASTp: Computed Atlas of Surface Topography of proteins. *Nucleic Acids Res.* **31,** 3352–3355, doi: 10.1093/nar/gkg512 (2003).

40. Schymkowitz, J. *et al.* The FoldX web server: an online force field. *Nucleic Acids Res.* **33,** doi: 10.1093/nar/gki387 (2005).

41. London, N. *et al.* Covalent docking of large libraries for the discovery of chemical probes. *Nat. Chem. Biol.* **10,** 1066–1072, doi: 10.1038/nchembio.1666 (2014).

42. Larsson, M. C. *et al.* Or83b encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction. *Neuron* **43,** 703–714, doi: 10.1016/j.neuron.2004.08.019 (2004).

43. Yao, K. M. & White, K. Neural specificity of elav expression: defining a *Drosophila* promoter for directing expression to the nervous system. *J. Neurochem.* **63,** 41–51, doi: 10.1046/j.1471-4159.1994.63010041.x (1994).

44. Shanbhag, S. R. *et al.* Expression mosaic of odorant-binding proteins in *Drosophila* olfactory organs. *Microsc. Res. Tech.* **55,** 297–306, doi: 10.1002/jemt.1179 (2001).

45. Kim, M. S., Repp, A. & Smith, D. P. LUSH odorant-binding protein mediates chemosensory responses to alcohols in *Drosophila* melanogaster. *Genetics* **150,** 711–721 (1998).

46. Hallem, E., Ho, M. & Carlson, J. The molecular basis of odor coding in the *Drosophila* antenna. *Cell* **117,** doi: 10.1016/j.cell.2004.05.012 (2004).

47. Knaden, M., Strutz, A., Ahsan, J., Sachse, S. & Hansson, B. S. Spatial representation of odorant valence in an insect brain. *Cell Rep*. **1,** 392–399, doi: 10.1016/j.celrep.2012.03.002 (2012).
48. Christiaens, J. *et al.* The fungal aroma gene ATF1 promotes dispersal of yeast cells through insect vectors. *Cell Rep*. **9,** doi: 10.1016/j.celrep.2014.09.009 (2014).
49. Hallem, E. & Carlson, J. Coding of odors by a receptor repertoire. *Cell* **125,** doi: 10.1016/j.cell.2006.01.050 (2006).
50. Stökl, J. *et al.* A deceptive pollination system targeting drosophilids through olfactory mimicry of yeast. *Curr. Biol*. **20,** doi: 10.1016/j.cub.2010.09.033 (2010).
51. González-Mas, M., Rambla, J., Alamar, M., Gutiérrez, A. & Granell, A. Comparative analysis of the volatile fraction of fruit juice from different citrus species. *PLoS One* **6,** doi: 10.1371/journal.pone.0022016 (2011).
52. Bartelt, R. J., Schaner, A. M. & Jackson, L. L. Cis-Vaccenyl acetate as an aggregation pheromone in *Drosophila* melanogaster. *J. Chem. Ecol*. **11,** doi: 10.1007/BF01012124 (1985).
53. Ejima, A. Pleiotropic actions of the male pheromone cis-vaccenyl acetate in *Drosophila* melanogaster. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol*. **201,** doi: 10.1007/s00359-015-1020-9 (2015).
54. Griffith, L. & Ejima, A. Courtship learning in *Drosophila* melanogaster: Diverse plasticity of a reproductive behavior. *Learn. Mem*. **16,** doi: 10.1101/lm.956309 (2009).
55. Grosjean, Y. *et al.* An olfactory receptor for food-derived odours promotes male courtship in *Drosophila*. *Nature* **478,** 236–240, doi: 10.1038/nature10428 (2011).
56. Schlief, M. & Wilson, R. Olfactory processing and behavior downstream from highly selective receptor neurons. *Nat. Neurosci*. **10,** doi: 10.1038/nn1881 (2007).
57. Strutz, A. *et al.* Decoding odor quality and intensity in the *Drosophila* brain. *Elife* **3,** doi: 10.7554/eLife.04147 (2014).
58. Gilson, M. *et al.* Open "back door" in a molecular dynamics simulation of acetylcholinesterase. *Science* **263,** 1276–1278, doi: 10.1126/science.8122110 (1994).
59. Biswas, S. *et al.* Bridging the synaptic gap: Neuroligins and neurexin I in Apis mellifera. *PLoS One* **3,** doi: 10.1371/journal.pone.0003542 (2008).
60. Jordan, M. *et al.* Expressed sequence tags and proteomics of antennae from the tortricid moth, Epiphyas postvittana. *Insect Mol. Biol*. **17,** doi: 10.1111/j.1365-2583.2008.00812.x (2008).
61. Durand, N. *et al.* A diversity of putative carboxylesterases are expressed in the antennae of the noctuid moth Spodoptera littoralis. *Insect Mol. Biol*. **19,** doi: 10.1111/j.1365-2583.2009.00939.x (2010).
62. Ishida, Y. & Leal, W. Rapid inactivation of a moth pheromone. *Proc. Natl. Acad. Sci. USA* **102,** doi: 10.1073/pnas.0505340102 (2005).
63. Silbering, A. F. *et al.* Complementary function and integrated wiring of the evolutionarily distinct *Drosophila* olfactory subsystems. *J. Neurosci*. **31,** doi: 10.1523/JNEUROSCI.2360-11.2011 (2011).
64. Stensmyr, M., Giordano, E., Balloi, A., Angioy, A.-M. & Hansson, B. Novel natural ligands for *Drosophila* olfactory receptor neurones. *J. Exp. Biol*. **206,** doi: 10.1242/jeb.00143 (2003).
65. Eisenthal, R., Danson, M. J. & Hough, D. W. Catalytic efficiency and kcat/KM: a useful comparator? *Trends Biotechnol*. **25,** 247–249, doi: 10.1016/j.tibtech.2007.03.010 (2007).
66. Love, C. A., Lilley, P. E. & Dixon, N. E. Stable high-copy-number bacteriophage lambda promoter vectors for overproduction of proteins in Escherichia coli. *Gene* **176,** 49–53 (1996).
67. Kruse, S. W., Zhao, R., Smith, D. P. & Jones, D. N. M. Structure of a specific alcohol-binding site defined by the odorant binding protein LUSH from *Drosophila* melanogaster. *Nat. Struct. Biol*. **10,** 694–700, doi: 10.1038/nsb960 (2003).
68. Katti, S., Lokhande, N., González, D., Cassill, A. & Renthal, R. Quantitative analysis of pheromone-binding protein specificity. *Insect Mol. Biol*. **22,** doi: 10.1111/j.1365-2583.2012.01167.x (2013).
69. Gasteiger, E. *et al.* Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook* (ed. J. M. Walker) 571–607 (Humana Press, 2005).
70. Bond, C. S. TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics* **19,** 311–312 (2003).
71. Han, Y. *et al.* Proteomic and molecular analyses of esterases associated with monocrotophos resistance in Helicoverpa armigera. *Pestic. Biochem. Physiol*. **104,** doi: 10.1016/j.pestbp.2012.09.005 (2012).

Acknowledgements

Author Contributions

F.Y., N.J.F., C.W.C., J.W.L., G.P., T.C., C.J.J. performed experiments; F.Y., N.J.F., C.J.J., C.W.C., J.W.L., G.J.C., M.M.C., G.P., J.G.O., T.H. designed research, F.Y., N.J.F., C.W.C., G.P., C.C., C.J.J., J.G.O. analyzed data; F.Y., N.J.F., C.J.J. and J.G.O. wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing Interests: The authors declare no competing financial interests.

How to cite this article: Younus, F. *et al.* Molecular basis for the behavioral effects of the odorant degrading enzyme Esterase 6 in *Drosophila*. *Sci. Rep.* **7**, 46188; doi: 10.1038/srep46188 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Chapter 4. Historical contingency in the evolution of organophosphate insecticide resistance

4.1. Declaration of Author Contribution

I conducted the majority of the research described in this chapter. Specifically, I completed the molecular cloning, protein purification, size exclusion chromatography and enzyme kinetic assays of the orthologs. I also performed the structural analysis and modelling. Dr. Peter D. Mabbitt from the Research School of Chemistry (RSC), The Australian National University (ANU) and Galen J. Correy at the RSC, ANU were involved in a large proportion of the enzyme assays. Davis H. Hopkins of the RSC, ANU cloned three of the orthologs.

4.2. Introduction

4.2.1. Insects

Insects are one of the most abundant group of animals on the planet, some of which are associated with detrimental effects on human health and the environment (184). Insects are also a major agricultural pest, which results in billions of dollars of damage to agriculture every year (14). There are a number of ways to deal with insects such as mechanical (traps, barriers) or biological controls (promoting beneficial insects to eat target pests), but most control programs rely on the use of insecticides to control the insect population (14). It has been estimated that 4 million tons of pesticides are used on world crops annually for pest control, with the market estimated to be in the order of $11 billion US (185, 186).

4.2.2. Organophosphate insecticides

Five major classes of insecticides have been widely used, including: organochlorines, carbamates, pyrethroids, organophosphates and neonicotinoids (**Figure 4.1**) (185, 187). All of these insecticides are neurotoxins that target the central nervous system of the insect and aim for one of the following targets: acetylcholinesterase, GABA receptors, sodium channels or nicotinic acetylcholine receptors (188). These insecticides, with the exception of organochlorines and neonicotinoids, are susceptible to cleavage as they contain ester bonds. This has led to resistance by detoxification and interaction with carboxylesterases (CBEs) (16, 68, 69, 185, 187–191). Two alternative insecticide types have been introduced in recent years including: biopesticides (*Bacillus thuringiensis*) and insect growth regulators (methoprene) (192–195).
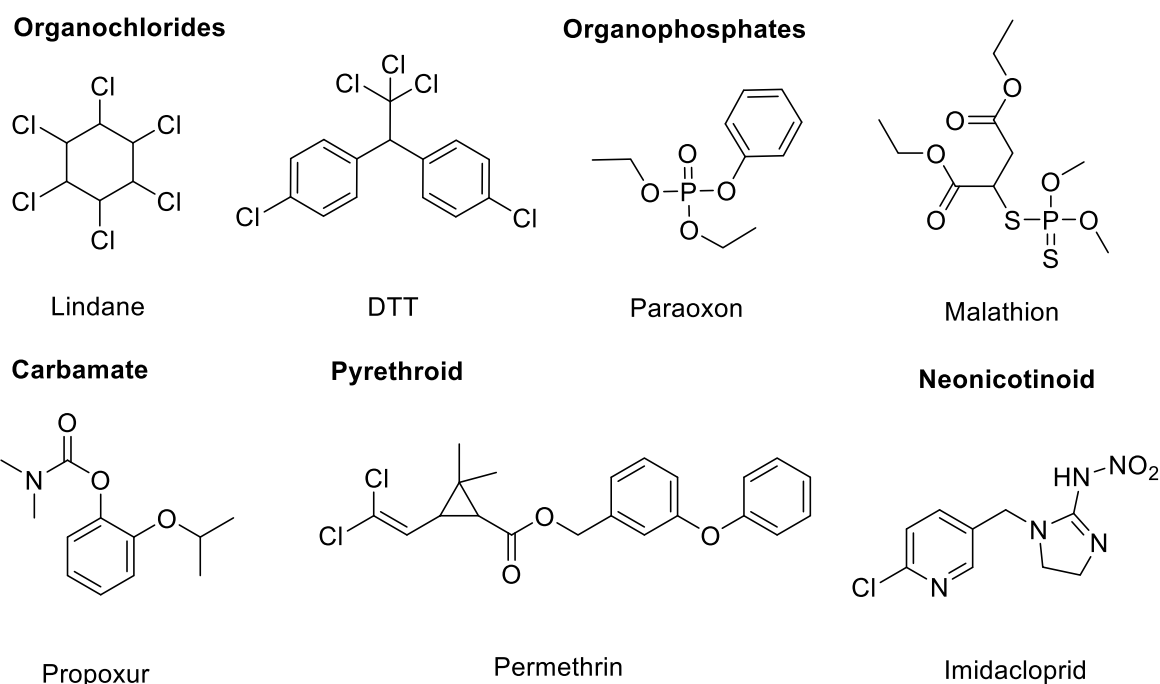
Figure 4.1. A selection of insecticides from the five major classes: organochlorides, organophosphates, carbamates, pyrethroids and neonicotinoids (185).

Organophosphate (OP) insecticides are covalent, irreversible inhibitors of acetylcholinesterase (AChE) and are used worldwide (17, 196, 197). They contain triesters of phosphoric acid with two short alkyl side chains (*o*-methyl or *o*-ethyl), a variable leaving group and a terminal oxygen **(Figure 4.2)** (17, 196–198). The leaving groups of OPs are electron withdrawing groups, which are displaced in the first step of the mechanism by the catalytic serine residue (17, 198). Over 30 commercial OP insecticides are in use and are often synthesised as "thion" analogs form, containing a sulphur at the phosphoryl oxygen position, which are activated to the active oxon form by P450 enzymes during metabolism (199). The oxon form of the OP is a much more effective inhibitor of AChE (200).
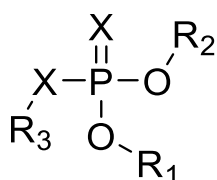
Figure 4.2. A generic OP insecticide. X is either an oxygen or sulfur and R_1 and R_2 are typically a methyl or ethyl group. R_3 is the variable leaving group (197).

OPs disrupt the function of cholinergic synapses in higher eukaryotes, which is the basis of their insecticidal effect (17, 197). The target, AChE is an essential enzyme that terminates

signal transduction at nerve synapses by hydrolysis of acetylcholine (Ach) (17, 201, 202). Insects are able to survive a high amount of AChE inhibition until a certain threshold, at which interminable signal transduction occurs at the cholinergic synapse leading to cholinergic nerve failure, paralysis, and death (72, 203, 204). Insect AChE is more susceptible to inhibition compared to human AChE leading to the popularity and widespread use of OPs (205–207).
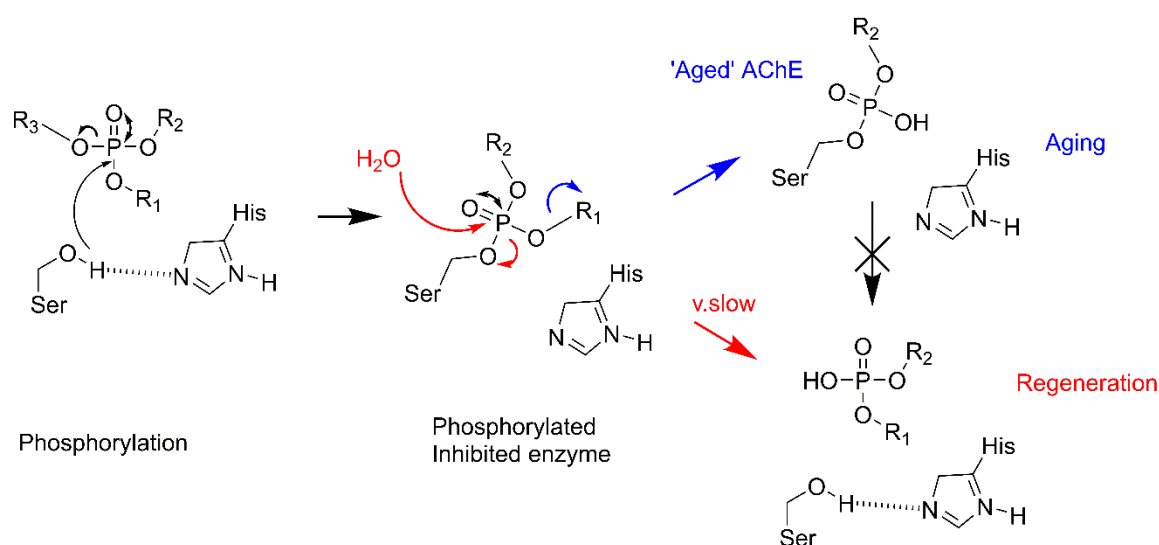
Figure 4.3. Mechanism of AChE inhibition by OPs. The enzyme can go either two pathways: aging to an irreversibly inhibited enzyme or can be slowly regenerated by the solvent to active protein. Figure adapted from Jackson et al (208).

The central phosphotriester moiety of OPs is important to their mode of action as inhibitors (208). In the initial steps of inhibition by OPs, the hydrolytic attack of the central phosphorous leads to the departure of the leaving group and the stable phosphor-Ser intermediate is formed **(Figure 4.3)** (32, 208). The first step of the reaction mechanism proceeds fast but the OP is a poor substrate for the second step and proceeds extremely slowly, rendering the enzyme essentially catalytically inactive (32, 208). The phosphor-Ser intermediate is tetrahedral, not planar like in the acyl-Ser intermediate. This causes steric constraint and prevents nucleophilic attack from an activated water molecule to regenerate the enzyme as in the acylated intermediate (209).

The enzyme can have two fates after the phosphor-Ser intermediate is formed – either (i) be slowly deactivated by non-activated water molecules, which is a very slow hydrolytic process or (ii) an aging reaction **(Figure 4.3)** (208). The aging reaction is from dealkylation of one of the side chains of the phosphor-Ser intermediate that leads to irreversible

inhibition of AChE (210, 211). Inhibition by OPs occurs to a certain degree with various CBEs due to the similarity in structure and catalytic mechanism (208, 212, 213). However, CBEs are resistant to the aging reaction due to subtle differences in structure (208, 212, 213).

4.2.3. Insecticide Resistance

Living organisms have evolved to adapt to xenobiotics and insecticide resistance is a well-documented example (32, 214). Insecticides are the preferred control method for agricultural and disease vector insect pests, yet the widespread use of insecticides worldwide creates a strong selection pressure (32, 214). This has led to a global increase of insecticide resistance due to the evolutionary adaptability of insects (215, 216). This resistance has made insecticide use ineffective and limited the available options for controlling arthropod populations, resulting in the re-emergence of diseases and a decrease in crop yields and profitability (217, 218). This resistance has caused disruption to arthropod control programs and is an ever growing and complicated global problem (14).

Insects exposed to these insecticides have evolved altered physiology and unique metabolic mechanisms in order to survive under the strong selection pressure imposed (32, 69, 187, 214, 219, 220). Currently, there are over 500 insect species that are reported to have resistance to at least one of the five classes of insecticides **(Table 4.1)** (193, 195, 221). Some of the most famous examples of insecticide resistance are from the house fly (222, 223), the cotton bollworm (224–226) and several mosquitoes species (227–229). Research worldwide is focussed on elucidating the mechanisms involved with the development of insecticide resistance, which is necessary for the creation of more effective strategies to hinder resistance, to control resistant species and to reduce the prevalence of disease and crop loss (230). To combat resistant arthropod populations, alternative techniques are being implemented such as using multiple insecticides on rotation at an increased concentration (217, 231).

Insecticide resistance in arthropod species has evolved by two main routes: by a decrease in sensitivity of the target site, or by upregulation of proteins able to sequester or break down the insecticide (32, 219). Metabolic resistance occurs through one or a combination of either quantitative (sequester) or qualitative (breakdown) and CBEs are most commonly responsible for this type of resistance **(Table 4.1)** (14, 19, 44, 187, 208, 215). CBEs are an abundant protein family in insects, with up to 76 CBE encoding genes identified in the

genomes of insect species (232). Two other major enzyme families are also involved in metabolic resistance: glutathione S-transferases (GST) and cytochrome P450 monoxygenases (P450s) **(Table 4.1)** (77, 187, 228). Metabolic resistance mediated by CBEs has rapidly evolved and has been shown to give resistance to three major insecticide classes (pyrethroids, carbamates, OPs) (69, 187, 208, 230).

Table 4.1. Summary of insecticide resistance to five types of insecticides (organophosphates, organochlorines, pyrethroids, carbamates and neonicotinoids) (32, 233).

Type of Resistance	Insecticide	Target Site (s)	Mechanism of Resistance
Desensitivity	OPs	Acetylcholinesterase	G117S
	Pyrethroids	Sodium Channels/GABA Receptors	L1014F
	Carbamates	Acetylcholinesterase	G117S
	Neonicotinoids	Nicotinic Acetylcholine Receptors	Y151S, R81T
	Organochlorines	Sodium Channels	L1014F
Metabolic Resistance	OPs	Acetylcholinesterase	CBEs
	Pyrethroids	Sodium Channels/GABA Receptors	P450s, GSTs, CBEs
	Carbamates	Acetylcholinesterase	CBEs
	Neonicotinoids	Nicotinic Acetylcholine Receptors	P450s, GSTs
	Organochlorines	Sodium Channels	-

4.2.4. Organophosphate Resistance

OP insecticides are commonly used worldwide, and have superseded organochlorine insecticides since they were mostly banned in 1972 (208). There have been over 150 separate cases of OP resistance reported with the earliest reports coming in 1963 (14, 234). Insects have three main resistance mechanisms to OP insecticides associated with CBEs (14, 32).

(1) Evolution of AChE (target) for lower sensitivity towards OPs.

(2) Upregulation of CBE genes that are able to bind and sequester the insecticide – potentially turning it over very slowly.

(3) Evolution of CBEs to catalyse hydrolytic detoxification of the OP insecticide.

4.2.5. Mechanisms of Resistance to Organophosphates

(1) Desensitivity of Acetylcholinesterase. Most insects have two forms of cholinesterase; one (AChE) is involved in neurotransmission and the target of insecticides. The other form (butylcholinesterase) is not discussed in further detail here (235–237). Amino acid substitutions in AChE cause a change in the structure that results in an enzyme that is less sensitive to inhibition by OPs (238–241). The amino acid position and the nature of resistance conferring mutations vary depending on the organism. For example, the Gly151Ser mutation is commonly found in the mosquitoes *Culex pipiens* and *Anopheles gambiae,* however, the Ser371Phe mutation is observed in the aphid, *Myzus persicae* (237, 238, 242–246). In both cases, resistance is achieved from a smaller residue to a bulkier one that constricts space in the active site gorge for the OP. This limits access for the OP to the catalytic residues that results in a reduction of AChE sensitivity (32). The substrate, Ach can still enter for normal cholinergic function, however, in some cases with reduced efficiency (247, 248). The mutation that appears depends on the particular insecticide, slight structural differences and genetic constraints (codon usage and variability). There are over 10 mutations that have been reported to be involved in resistance throughout multiple species (249–252).

(2) Quantitative changes involving carboxylesterases as a mechanism of metabolic resistance. This mechanism of resistance is from the overproduction of CBEs through gene amplification (multiple copies of the same gene being found on the same genome). This results in CBEs effectively acting as an OP sink, minimising the inhibition of AChE (28, 32, 187, 188). In some cases, the CBEs genes can be amplified up to 200 copies, which results in the overexpression of the protein (34, 253, 254). This mechanism is found in several orders including Hemiptera (aphids and whiteflies), Hymenoptera, Lepidoptera and lower dipteran species (*Culex* mosquitoes). The quantitative mechanism has been utilized in insect species for resistance against OPs, carbamates, and pyrethroids insecticides (32, 34, 35, 188, 253, 255–259). The sequestration mechanism is effective because of the similarity in structure and catalytic mechanism of CBEs to AChE, which results in CBEs having high affinity for OPs (260). Despite the negligible OP hydrolase activity, the

overabundance of the CBE is able to sequester the OP sufficiently to confer resistance to the organism (261–263).

(3) Qualitative changes in carboxylesterases by mutations for increased activity. Instead of a change in copy number, a favourable structural mutation in a CBE can provide resistance. A mutation can enhance OP hydrolysis resulting in an increased capability for insecticide metabolism (2, 208). This mechanism of resistance by catalytic improvement is a successful detoxification strategy because of the high affinity for OPs observed in CBEs (2, 223, 264). The turnover rates of the OP are typically slow for an enzyme but are sufficient for resistance (2, 223, 264). The most studied CBE for this strategy is the protein αE7 from the Australian sheep blowfly (*Lucilia cuprina*) (2, 264). A single point mutation was identified to give increased activity towards OP insecticides, however, the mutation resulted in a decrease in CBE activity (72, 223, 264–266).

4.2.6. Qualitative Resistance in Diptera
In higher Diptera, resistance to OPs is achieved by a structural change that reduces CBE activity for OP hydrolase activity (2). Resistance by this mechanism has been observed in multiple Diptera species including *Lucilia cuprina* (Australian sheep blowfly), *Lucilia sericata*, *Musca domestica* (Common housefly), *Cochliomyia hominivorax* (New World screw-worm fly), and one Hymenopteran species, *Anisopteromalus calandrae* (Howard Parastic Wasp) (2, 223, 267–269). In blowfly species, the CBE involved has been identified as αE7 with a single amino acid substitution found in the active site (Gly137Asp or Trp251Leu). The single mutation Gly137Asp results in broad-spectrum OP resistance in preference to diethyl OPs. The Trp251Leu substitution results in broad-spectrum resistance in preference to dimethyl OPs and malathion specific resistance (264, 270). This has been identified as an example of the ali-esterase mechanism, as there is a characteristic shift in substrate specificity from carboxylesters to OPs (2, 223, 265, 267).

As mentioned, the best characterized mutant CBE involved in the qualitative resistance mechanism is αE7 from the species *L. cuprina* (2, 72, 271, 272). The Gly137Asp mutant enzyme has a 55-fold increase in hydrolase activity for diethyl OPs and a 33 fold increase for dimethyl OPs compared to the wild-type enzyme (72). The turnover number is poor, with approximately three turnovers per hour per molecular enzyme (72, 264, 270). The poor catalytic efficiency (k_{cat}/K_M of 10^2 s^{-1}M^{-1}) is low for a typical naturally evolved enzyme (10^4- 10^7 s^{-1}M^{-1}), however, the system has been present for less than 70 years rather

than millions of years for evolution to take place (32). The Trp251Leu mutant enzyme provides a 130 fold increase in activity against malathion compared to the wild-type enzyme (264, 270). The selective pressure and protective benefit has resulted in alleles with the Gly137Asp substitution becoming dominant over the Trp251Leu mutant in *L. cuprina* populations (11, 273). The opposite trend (Trp251Leu dominant over Gly137Asp) has been observed with *M. domestica* populations in Turkey where malathion is commonly used (274). The two mutations are not found together in nature, each conferring resistance independently (69).

A full understanding of the mutations has been achieved with the structure of αE7 from *L. cuprina* (55). For the Gly137Asp mutation, the non-conservative replacement of Gly to Asp disrupts the oxyanion hole in the active site. This compromises function, resulting in a reduction of CBE activity (2, 55, 223). However, the introduction of Asp at position 137 increases OP turnover by Asp acting as a general base. This allows for the activation of a water molecule in a suitable position to attack the phosphor-Ser intermediate that is formed between the catalytic serine and the OP (2, 72). The hydrolysis of the intermediate regenerates the enzyme and is similar to the second step of native carboxylester hydrolysis activity (32). A new activation site is necessary for OP hydrolysis, as the intermediate has tetrahedral rather than planar trigonal geometry that is found with carboxylester substrates (233). A similar mechanism for an increase in OP hydrolase activity has been observed in human BChE during a directed evolution experiment. However, in this experiment, a histidine residue was found at this position instead of an aspartate residue (275).

The Trp251Leu mutation is found in the acyl pocket of the active site (P2 subsite). The presence of a bulky tryptophan residue at this position limits the hydrolysis of the phosphor-Ser bond (55, 264, 270). The change from Trp251 to a smaller residue (Leu) provides more space for the OP by opening up the acyl binding pocket. This reduces steric hindrance for the hydrolysis of the phosphor-Ser bond (55, 223, 264). There is also a loss of a stabilising interaction between the phosphorylated serine and the protein that leads to an increase in active site regeneration (208). This mutation has been found before the introduction of insecticides, suggesting it could be neutral in regards to native function (264, 268).

A further understanding for how the OP resistance is conferred by CBEs can be gained from the comparison between CBEs and AChE. The high active site complementary between

CBEs and AChE can result in CBEs having a tighter affinity to OPs than AChE (2, 55, 206, 208, 212). Furthermore, the resistance to the aging reaction and a high abundance results in CBEs acting as effective catalytic bioscavengers (55, 212, 272, 276, 277).

4.2.7. Evolutionary Contingency and Insecticide Resistance

There are a larger number of examples of the quantitative resistance mechanism in insects compared to the qualitative resistance mechanism (193, 195, 221). There are several proposed reasons for why certain insect species favour the quantitative mechanism over the qualitative mechanism of resistance. These include:

- Two or more base pairs changes from Gly to Asp results in a low chance of evolution to occur (278).
- Subtle differences in structure limits the ability for OP hydrolase activity to evolve (55).
- The loss of wild-type function with the introduced mutation leads to a loss of fitness (278, 279).
- The sequestration ability confers resistance to a wider range of insecticides, therefore, provides greater protection than the qualitative resistance mechanism (32, 278, 279).

Despite the large amount of effort in showing the critical role of CBEs in the evolution of insecticide resistance, the molecular principles regarding the evolution of new enzyme function have not been fully elucidated. Furthermore, there is no full understanding of what dictates the evolution of qualitative or quantitative mechanism in insect species (32, 279). The understanding of resistance could lead to the ability to engineer a protein for an increase in OP hydrolase activity to treat OP poisoning (72, 208). However, without structural and biochemical information on a large number of insect proteins, it is difficult to understand what factors dictate a protein's ability to hydrolyse OPs effectively.

Subtle differences between the CBEs could explain why some proteins cannot confer resistance by the qualitative mechanism. This could be caused by neutral drift, which is related to the concept of evolutionary contingency (280). The key evolutionary concept of historical contingency was first mentioned by Stephen Jay Gould, who describes historical contingency as the constriction of evolutionary paths by historical events that are often random (281–284). In other words, the initial state influences the evolutionary outcome. During the selection process, small genetic differences can cause major changes in the

evolutionary outcome (285). Genetic differences can be due to past selective pressures and to random changes in alleles, which is related to the sampling processes (286–289).

The fundamentals of evolution involve a balance between random and deterministic processes (284, 290–295). The deterministic processes of evolution will produce the most accessible form in the fewest possible steps, regardless of random events (296, 297). When non-deterministic processes (genetic drift) play a strong part in evolution, chance events determine the outcome of evolution (298). This results in the unpredictability when history is repeated, with different results despite identical environmental conditions (299). Non-deterministic processes are important for the generation of new beneficial mutations (299).

In some cases, beneficial mutations can be dependent on prior mutations that are neutral when isolated but must be present for beneficial mutations to take effect (284, 300–306). These mutations cannot be fixed by selection processes and must be acquired independently (298). Therefore, the chance of a mutation appearing in a population may be contingent on past mutations (282, 284, 307, 308). These historical contingencies or differences in initial conditions that are a universal property of dynamic systems, make evolution largely unpredictable (282, 309).

In most studies, the conditions that facilitate the evolution of new function are destroyed during the evolutionary processes (307). However, the recent example of OP resistance in insects presents a modern example where the past is able to be recovered (310–313). Understanding how random historical events shape evolutionary processes is a main goal of evolutionary biology and it is difficult to characterize other potential pathways from the past (284, 297, 305, 308). However, as insecticide resistance is a relatively recent event, direct insights into evolutionary contingency can be investigated.

There has been a suggestion of contingency previously with Diptera αE7 orthologs. The *Drosophila* αE7 protein has a lower affinity for OPs compared to *Lc*αE7, despite few changes in the active site residues (206, 272, 277). Development of prediction strategies is required to look at the active site and second or third shell residues. These residues are important for shaping the active site and influencing the correct orientation of Asp137 for OP hydrolysis (314–316). To investigate OP resistance in Diptera species, the role of evolutionary contingency in Diptera and why OP resistance is not widespread by the qualitative mechanism, multiple genes were ordered of *Lc*αE7 orthologs with and without

the Gly137Asp mutation. The orthologs would be characterized biochemically and structurally to analyse their ability to turnover OPs.

4.3. Preface

How enzymes evolve new functions has been of great interest and insecticide resistance is a valuable model system to study this. While a large number of insect pests have evolved resistance, there appears to be limited options for CBE-based metabolic resistance. Studying the evolution of resistance is important for designing new effective control strategies for insecticide resistance and the creation of new insecticides to make resistance difficult to evolve. A large amount of effort has been focussed on resistance at the genetic level with protein analysis been relatively unexplored. The difficulties of recombinant expression of eukaryotic proteins in *E. coli* has been a major hurdle in the characterization of resistance proteins. There are still a few questions that remain to be answered with resistance proteins including if there are limited routes for resistance by the qualitative resistance mechanism for CBEs and can the Asp137 residue hydrolyse OPs in close orthologs of *Lc*αE7? This could show that resistance by the qualitative resistance mechanism is a general characteristic or a unique feature in dipteran species. The following research also intends to understand if evolutionary contingency plays a role in the development of insecticide resistance involving CBEs. In this chapter, work has been focussed on the evolution of qualitative resistance in *Lc*αE7 orthologs. The Gly137Asp mutation in the protein *Lc*αE7 has been found to give resistance to the blowfly against OPs. Analysis of this mutation on OP hydrolase activity in *Lc*αE7 orthologs can establish if contingency plays a role in the evolution of insecticide resistance in Diptera species.

4.4. Materials and Methods

4.4.1. Cloning

Genes encoding αE7 orthologs bearing the Gly137Asp mutation were synthesized by Integrated DNA Technologies (IDT, USA). The genes were cloned by the DNA-fragment assembly method described by Gibson and coworkers into the pETMCSIII vector (317, 318). The mutations Asp137Gly or Gly137Asp were introduced into the vectors by Gibson assembly using the primers listed in Appendix F. The resulting vectors were sequenced at the Biomolecular Resource Facility, Australian National University, Australia. Plasmids encoding *L. cuprina* αE7 were described previously (315). Plasmids and sequence information was submitted to the Addgene vector database.

4.4.2. Protein Expression and Purification

The proteins were expressed in BL21 (DE3) *Escherichia coli* at 24 °C for 22 hours with autoinduction in lysogeny broth (LB). Cells were collected by centrifugation and resuspended in buffer A (50 mM HEPES pH 7.5, 300mM NaCl, 10 mM imidazole) and lysed by sonication. Cell debris was pelleted by centrifugation, and the soluble fraction was applied to a HisTrap-HP Ni-Sepharose column (GE Healthcare, USA). Protein was eluted in buffer B (50 mM HEPES pH 7.5, 300 mM NaCl, 300 mM imidazole). The major fraction containing the protein of interest was applied to a HiLoad 26/60 Superdex-200 size-exclusion column (GE Healthcare, USA), pre-equilibrated in buffer C (20 mM HEPES pH 7.5, 150 mM NaCl). The protein *Cc*αE7 was purified in the same buffer conditions but with the pH adjusted to 8.5. Protein elution was monitored at 280 nm, and fractions containing the monomeric protein were pooled for enzyme assays. Protein concentration was calculated from absorbance at 280 nm with the extinction coefficient. The extinction coefficient and molecular mass for each protein was calculated using the Protparam online tool (319). Molecular weights of the oligomeric species were estimated using a gel filtration HMW calibration kit (GE Healthcare, USA) following the manufacturer's instructions. Protein purity was assessed by SDS-PAGE using precast ExpressPlus 4 to 20% PAGE gels (GenScript) and visualized by staining with Coomassie brilliant blue (Sigma, USA.)

4.4.3. Carboxylesterase Activity

Hydrolysis of the ester substrate 4-nitrophenyl butyrate (4-NPB; Sigma-Aldrich, USA) was monitored at 405 nm in buffer C at 25 °C using an Epoch microplate spectrophotometer (BioTeK Instruments, USA). Buffer C was supplemented with 0.4 mg/mL bovine serum albumin in the assays (Sigma-Aldrich, USA) to increase the stability of the enzyme. 4-NPB

was serially diluted 1:2 from 1000 to 7.81 µM and the assays were measured in triplicate. Velocities were obtained from the initial linear portion of the reaction progress curves, and product concentration was determined with the extinction coefficient calculated from a standard curve of 4-nitrophenol (Sigma-Aldrich, USA). K_M and k_{cat} were determined by nonlinear regression to the Michaelis−Menten equation using GraphPad Prism 6.0 (GraphPad Software, USA).

4.4.4. Organophosphate Hydrolase Activity

The measurement of OP hydrolase activity of the proteins was followed by the protocol developed by Mabbitt et al. 2016 using the substrate diethylumbelliferyl phosphate (DEUP, Sigma-Aldrich, USA) (315). DEUP hydrolysis was monitored fluorometrically (excitation at 330 nm and emission at 450 nm) with a Cary Eclipse Fluorescence spectrophotometer (Agilent Technologies, USA). The assay was carried out at 25 °C and pH 7.5. Product formation was quantified using a standard curve of 4-methylumbelliferone (Sigma-Aldrich, USA) in the assay buffer C. The proposed scheme for hydrolysis of DEUP is presented in scheme 1.

$$\mathrm{E} + \mathrm{S} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \overset{k_3}{\longrightarrow} \mathrm{EPQ} \overset{k_5}{\longrightarrow} \mathrm{E} + \mathrm{P} + \mathrm{Q} \qquad \text{(scheme 1)}.$$
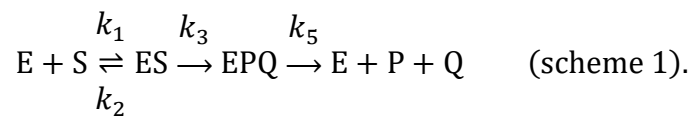
where E, S, P, Q are the enzyme, substrate, product and fluorescent umbelliferone leaving group, respectively. k_1 and k_2 are the rate constants for the formation of the Michaelis complex (ES), k_3 is the rate constant for the formation of the phosphorylated intermediate (EP), and k_5 is the rate constant for dephosphorylation.

The rate constant describing the pre-steady state burst was determined by solutions containing DEUP serially diluted 1:2 from 500 to 1.5 µM, and 0.75 µM αE7. The solution was mixed in a 1:1 ratio using a RX2000 rapid mixing stopped-flow unit fitted with a RX/DA pneumatic drive (Applied Photophysics, UK). The values of k_3 and K_M were determined from the concentration dependence of the pre-steady state burst (315). The apparent values of k_5 were determined from the linear portion of the fluorescence trace. Curve fitting was conducted using GraphPad Prism 6.0 (GraphPad Software, USA).

4.4.5. Structure Modelling

Models of the orthologs were generated using the I-TASSER server with the structure of *Lc*αE7 (PDB ID: 4FNG) and the unpublished structure of *Cq*B2 (55, 320). *Aa*A2 and *Ag*B2

models were generated with the structure of *Cq*B2 and the other ortholog models were generated with the structure of *Lc*αE7.

4.4.6. Native-PAGE

Peaks from the size exclusion chromatography were concentrated to 2 mg/mL and loaded onto NativePAGE Novex 4-16% Bis-Tris protein gels (Invitrogen) according to the manufacturer's instructions. Protein samples were visualized by staining with Coomassie brilliant blue (Sigma, USA).

4.5. Results

4.5.1. Expression and Purification

The proteins for analysis were selected on their similarity to *Lc*αE7 from a BLAST search **(Table 4.2)**. The protein, B2 from the species *Culex quinquefasciatus* (Southern house mosquito) was used as the end point in the search, as the introduction of the Gly137Asp mutation in the *Culex* protein resulted in no change in the rate of OP turnover (278). The structure of this protein was recently solved in Dr. Colin Jackson's research group (work unpublished). This structure can be used to understand the structural differences that affect OP kinetics. The pairwise identities of the orthologs from the 11 species range from 38% to 89% compared to *Lc*αE7 **(Table 4.3)**.

The amino acid residues for the orthologs corresponding to position Gly137 found in the *L. cuprina* CBE are shown in Table 4.4. The targeted position 137 was Ala in the enzyme from *Aedes aegypti*, appears as Asp in the *Musca domestica* and *Cochliomyia hominivorax* proteins and is found as Gly for the other proteins **(Table 4.4)**. The proteins share similar molecular weights and pIs with the exception of *Cc*αE7 with a pI approximately 1 unit above the other orthologs.

Table 4.2. Summary of the Diptera α-esterases selected for analysis. GenBank codes are listed for each protein.

Species	Common Name	Protein (GenBank)	Name
Lucilia cuprina	Australian sheep blowfly	AAB67728.1	*Lc*αE7
Calliphora stygia	Brown blowfly	AID61335	*Cs*αE7
Cochliomyia hominivorax	New World screw-worm fly	ACR56068.1	*Ch*αE7
Musca domestica	Common house fly	NP_001295905	*Md*αE7
Haematobia irritans	Horn fly	AF139082_1	*Hi*αE7
Stomoxys calcitrans	Biting house fly	XP_013116190	*Sc*αE7
Bactrocera dorsalis	Oriental fruit fly	AKN90083.1	*Bd*B1
Ceratitis capitata	Mediterranean fruit fly	NP_001266341.1	*Cc*αE7
Aedes aegypti	Yellow fever mosquito	XP_001654509.1	*Aa*A2
Anopheles gambiae	African malaria mosquito	XP_316296.3	*Ag*B2
Culex quinquefasciatus	Southern house mosquito	CAA83643.1	*Cq*B2

Table 4.3. Percentage identity (%) between the Diptera orthologs. The range of identity varies between 89% and 36%.

	*Lc*αE7	*Cs*αE7	*Ch*αE7	*Md*αE7	*Hi*αE7	*Sc*αE7	*Bd*B1	*Cc*αE7	*Aa*A2	*Ag*B2	*Cq*B2
*Lc*αE7	100	89	88	76	74	73	58	57	41	39	38
*Cs*αE7	89	100	86	74	71	70	56	55	39	39	36
*Ch*αE7	88	86	100	76	73	73	57	56	40	40	37
*Md*αE7	76	74	76	100	75	75	57	55	39	41	37
*Hi*αE7	74	71	73	75	100	77	56	53	39	39	36
*Sc*αE7	73	70	73	75	77	100	55	55	39	39	37
*Bd*B1	58	56	57	57	56	55	100	59	40	42	38
*Cc*αE7	57	55	56	55	53	55	59	100	38	39	36
*Aa*A2	41	39	40	39	39	39	40	38	100	49	51
*Ag*B2	39	39	40	41	39	39	42	39	49	100	63
*Cq*B2	38	36	37	37	36	37	38	36	51	63	100

Table 4.4. Summary of the general characteristics of the Diptera esterases analysed including the calculated molecular weight, theoretical pI, equivalent position number to 137 in *Lcα*E7, codon found at this site and the resistance found in the species mentioned.

Protein	MW (kDa)	Theoretical pI	Position Equivalent to 137 in *Lcα*E7	Codon at 137	Observed OP Resistance involving esterases in Diptera species
*Lcα*E7	65.4	6.27	137	GGT/GAT (Gly/Asp)	Gly137Asp, Trp251Leu (2)
*Csα*E7	65.6	6.13	137	GGT (Gly)	Unknown
*Chα*E7	65.2	6.56	137	GAC (Asp)	Gly137Asp, Trp251Leu (267)
*Mdα*E7	65.4	6.27	137	GAT (Asp)	Gly137Asp, Trp251Leu (223)
*Hiα*E7	65.7	6.35	137	GGT (Gly)	Upregulation (321)
*Scα*E7	65.5	6.69	137	GGT (Gly)	Unknown
*Bd*B1	65.3	6.09	145	GGC (Gly)	Upregulation (322)
*Ccα*E7	65.7	7.52	145	GGA (Gly)	Unknown
*Aa*A2	62.2	5.34	118	GCT (Ala)	Upregulation (323)
*Ag*B2	61.1	5.26	118	GGA (Gly)	Unknown
*Cq*B2	61.7	5.70	118	GGC (Gly)	Upregulation (262)

The genes of the dipteran esterases were ordered as WT or with Asp at the equivalent 137 position. Cloning by Gibson assembly introduced the genes into the pETMCSIII expression vector. The reverse mutation Asp137Gly or the Gly137Asp mutation were mutated into the Diptera esterases genes by Gibson assembly. The wild type and variant (Gly137Asp) CBEs were expressed and purified in similar conditions to the expression of *Lcα*E7 (324). There was a significant absorbance reading (>1000 mA) following affinity chromatography that indicates a high amount of soluble expression for the recombinant CBEs **(Figure 4.4)**. The exception was with the ortholog, *Hiα*E7, where after affinity chromatography a low absorbance was observed (<300 mA). This suggests the protein is largely insoluble or does not express and was subsequently not analysed in further detail. There was no change in soluble expression of *Hiα*E7 with variations in expression media and temperature.
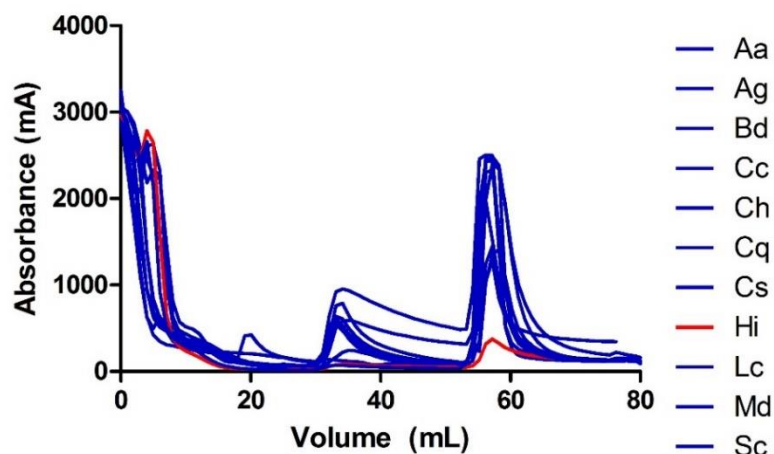
Figure 4.4. Chromatogram of the 11 orthologs after nickel affinity chromatography. Blue lines represent the 10 proteins that express in large amounts. The red line represents *Hi*αE7 that was found to not express in high amounts.

Following affinity chromatography, the proteins were subjected to size exclusion chromatography (SEC). This revealed a large variation in the amounts of soluble protein, aggregation and oligomeric states between the proteins **(Table 4.5) (Figure 4.5)**. The mosquito proteins (*Aa*A2, *Ag*B2, and *Cq*B2) expressed in high amounts compared to the blowfly orthologs **(Table 4.5)**. The three mosquito orthologs lack the amphipathic α-helix that is found in the other seven blowfly orthologs. The missing α-helix in the mosquito orthologs could explain the high amounts of expression and low aggregation associated with the three mosquito enzymes **(Figure 4.5)**.

Table 4.5. Summary of the average peak size, area and average concentration of the orthologs after SEC.

Protein	Average Monomer Peak Absorbance (mA)	Total Peak Area (ml*aMU)	Estimated Monomer Concentration of a single main peak fraction (mg/mL)
*Lc*αE7	180 ±50	4900 ± 900	1
*Cs*αE7	105 ± 10	5200 ± 900	0.5
*Ch*αE7	300 ± 50	8200 ± 1200	1.5
*Md*αE7	490 ± 40	11000 ±2000	2
*Sc*αE7	160 ± 30	5800 ± 2600	1
*Bd*B1	50 ± 30	4800 ± 1100	0.5
*Cc*αE7	40 ±10	3700 ± 900	0.3
*Aa*A2	880 ± 500	14000 ± 7000	3
*Ag*B2	1500 ± 200	24000 ± 3000	4
*Cq*B2	960 ± 100	17000 ± 3000	3

To analyse in further detail the oligomeric states of the proteins, the absorbance peaks were compared to protein standards subjected to SEC. It is surprising the differences in oligomeric state between the orthologs, given the close similarity of the proteins to each other. Proteins sharing a sequence identity level of 90% have a 93% likelihood of quaternary structure conservation (114). Furthermore, when proteins share 30-40% sequence identity, the likelihood of the proteins sharing similar quaternary structure is 70% (114). This indicates the orthologs have a high likelihood of sharing quaternary structure, however, this was not the case. Comparing the elution volume of the orthologs to protein standards and using previous work on the identified oligomeric peaks of *Lc*αE7, the different oligomeric species can be analysed **(Table 4.6)** (324).
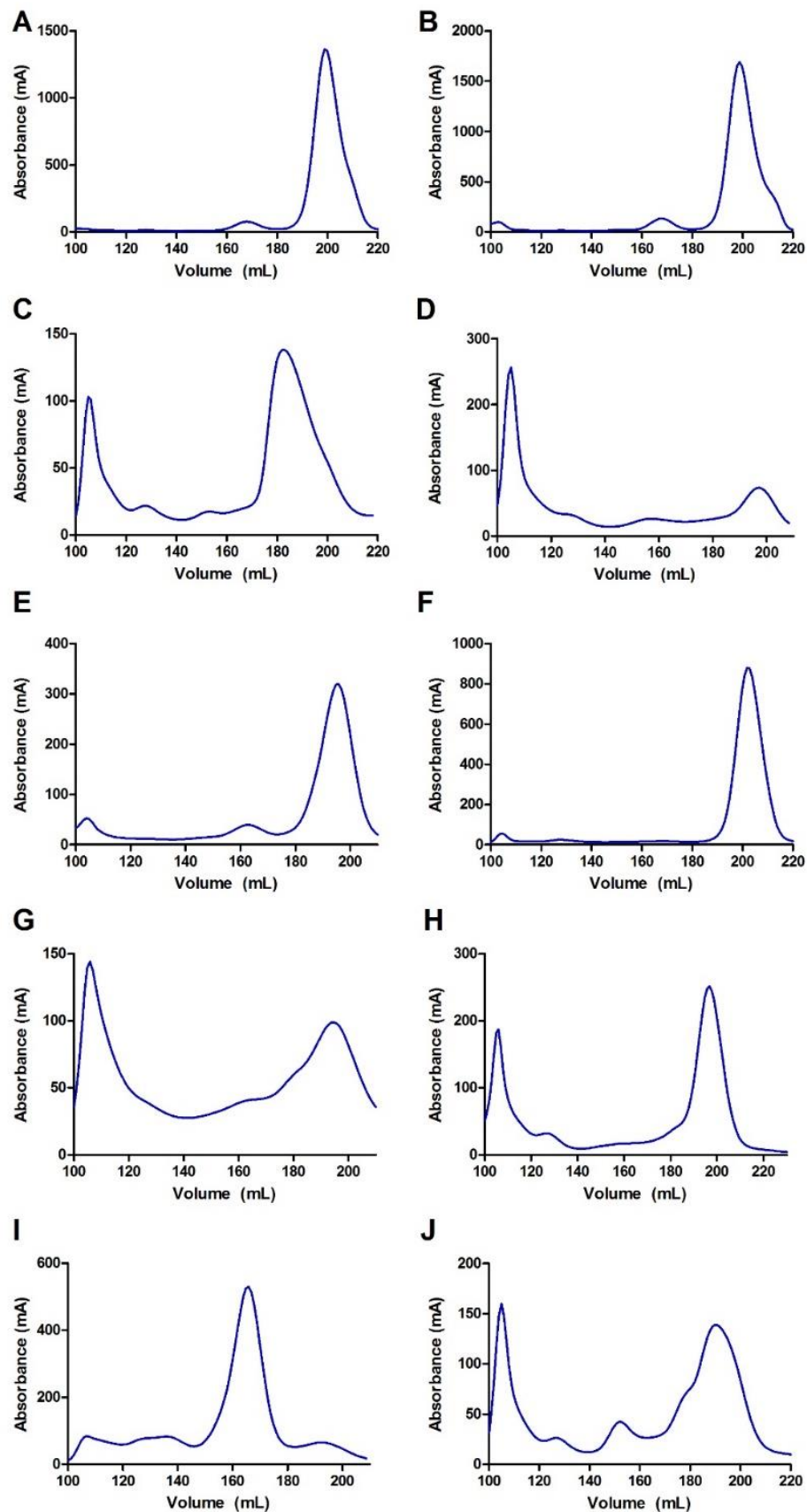
Figure 4.5. Chromatograms of the 10 purified orthologs after SEC (A) *Aa*A2 (B) *Ag*B2 (C) *Bd*B1 (D) *Cc*αE7 (E) *Ch*αE7 (F) *Cq*B2 (G) *Cs*αE7 (H) *Lc*αE7 (I) *Md*αE7 (J) *Sc*αE7.

Four of the orthologs show similar quaternary structure to *Lc*αE7 (*Cs*αE7, *Sc*αE7, *Cc*αE7, *Ch*αE7). The surprising difference is observed with *Md*αE7 **(Figure 4.6, Panel I)**, as the appearance of the main peak is at the similar position to the dimer peak in *Lc*αE7 **(Figure 4.6, Panel H)**. To investigate the peak at this position, the main fractions of *Md*αE7 and *Lc*αE7 were subjected to native-PAGE. This revealed that for *Lc*αE7 there is a band at 66 kDa, consistent of a monomeric species. However, for the main peak of *Md*αE7, there appears to be no band at this position. A major band for *Md*αE7 is present at 130 kDa suggesting *Md*αE7 appears as a dimer species **(Figure 4.6)**. This is consistent with the protein standards that estimated the main peak of *Md*αE7 to be at 135 kDa. A variant of *Md*αE7 (Lys557Pro) resulted in a large amount of aggregation and a shift in the oligomeric state of *Md*αE7. The variant displayed similar amounts of dimer and monomer species that indicates the oligomeric state equilibrium of *Md*αE7 is dynamic **(Figure 4.7)**. *Bd*B1 elutes in a similar position to *Md*αE7 and suggests the protein is mainly found as a dimer species similar to *Md*αE7. The orthologs from the mosquito species (*Aa*A2, *Ag*B2, and *Cq*B2) express in high amounts and are predominantly found as monomers **(Tables 4.5, 4.6)** **(Figure 4.5)**. After affinity chromatography and SEC, the proteins with the exception of *Cc*αE7 were >90% purity for further analysis by kinetic assays **(Figure 4.8)**.

Table 4.6. Summary of the predicted oligomeric states of the Diptera esterases estimated from protein standards.

Protein	Estimated Sizes of Peaks (kDa)	Monomer	Dimer	Tetramer
*Lc*αE7	**55**, 89, 176	**Yes**	Yes	Yes
*Cs*αE7	**59**, 91, 165	**Yes**	Yes	Yes
*Ch*αE7	**58**, 165	**Yes**	Possible	Possible
*Md*αE7	63, **145**	Yes	**Yes**	No
*Sc*αE7	**69**,100, 224	**Yes**	Yes	Yes
*Bd*B1	49, **88**	Yes	**Yes**	No
*Cc*αE7	**52**, 76, 180	**Yes**	Yes	Yes
*Aa*A2	**50**, 132	**Yes**	Possible	Possible
*Ag*B2	**47**, 132	**Yes**	Possible	Possible
*Cq*B2	**46**	**Yes**	No	No

*Bold text highlights the predominant oligomeric species for each protein
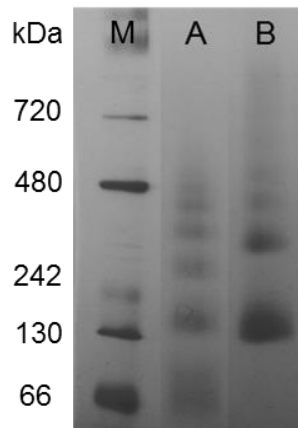
Figure 4.6. Native-PAGE gel of *Md*αE7 and *Lc*αE7 displaying the change in oligomeric state between the two proteins. There is an absence of a protein band at 66 kDa for *Md*αE7 (A) *Lc*αE7 (B) *Md*αE7 (M) Marker.
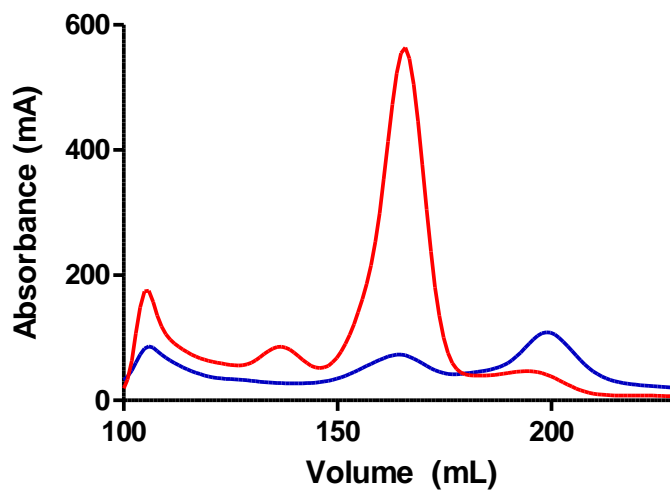
Figure 4.7. Size exclusion chromatogram displaying the difference in oligomeric species between *Md*αE7 WT (red) and *Md*αE7 Lys557Pro (blue).
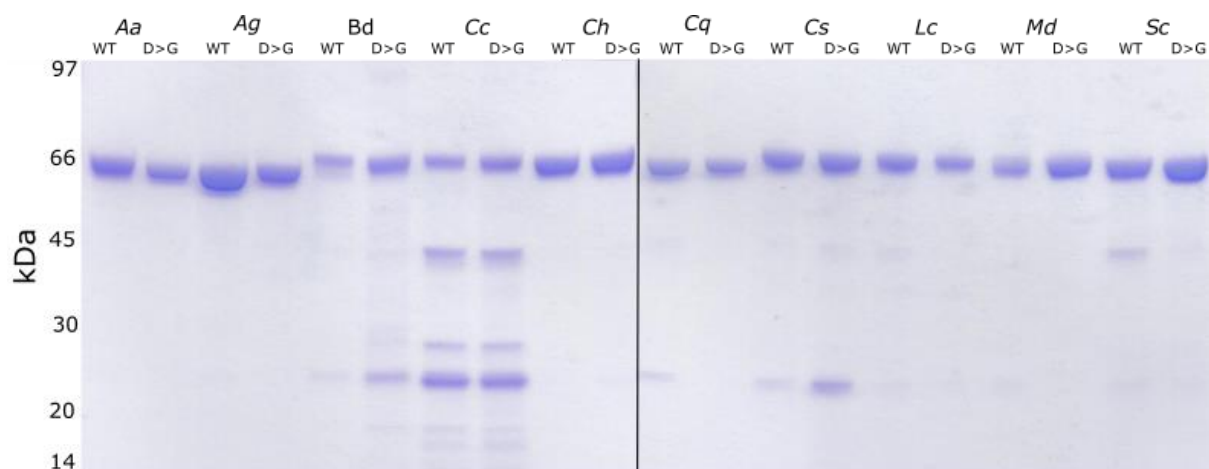
Figure 4.8. SDS-PAGE gel of the 10 orthologs (WT and Gly137Asp variants) after size exclusion chromatography that was further used for kinetic analysis.

4.5.2. Kinetic Assays – Esterase Activity

To understand a potential trade-off between CBE activity and OP hydrolase activity, the orthologs were first characterized using a model ester substrate 4-nitrophenyl butyrate (4-NPB) **(Figures 4.9, 4.10) (Table 4.7)**. The WT orthologs have a large range in k_{cat} (3.9 - 1317 s^{-1}), K_M (20 – 105 μM) and k_{cat}/K_M (3.7 x 10^4 - 2.2 x 10^7 $s^{-1}M^{-1}$). The wild-type enzyme from *A. aegypti* has the highest catalytic efficiency (k_{cat}/K_M) towards the ester substrate.
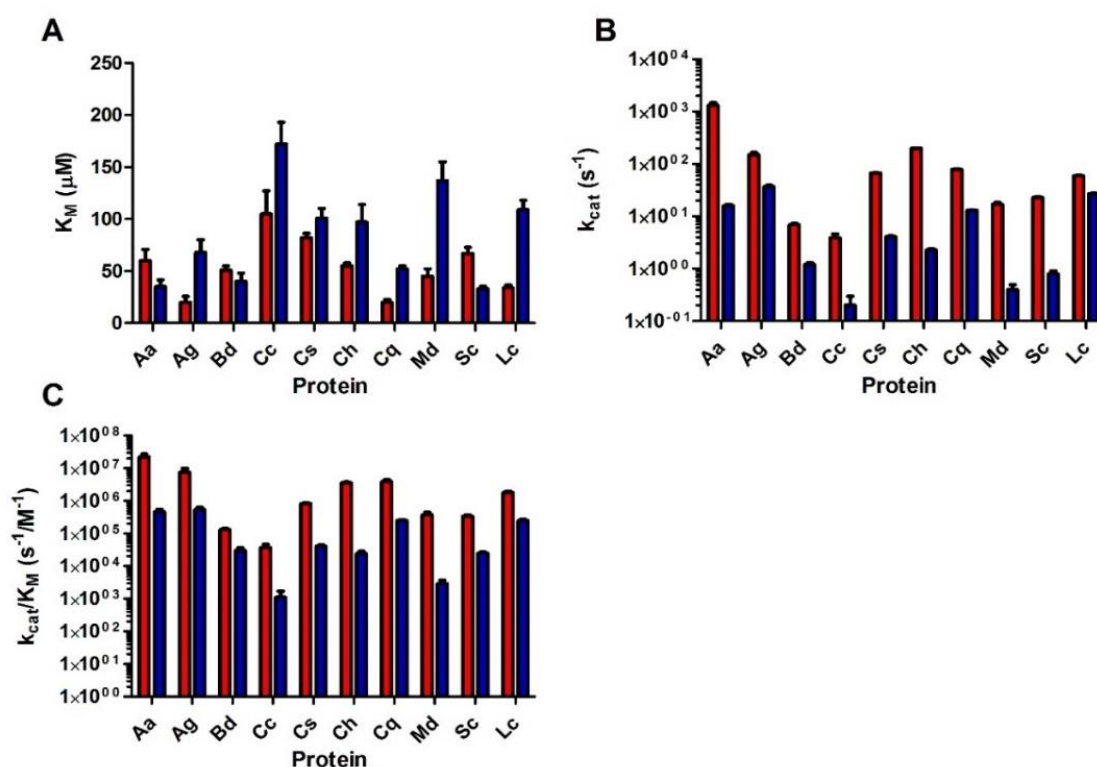
Figure 4.9. Kinetic data of the 10 orthologs for the ester substrate 4-NPB. Red represents WT of the ortholog for a Diptera species and blue represents the Gly137Asp variant for the species (A) K_M (B) k_{cat} (C) k_{cat}/K_M.

The Asp residue at position 137 decreases the enzymatic activity (k_{cat} and k_{cat}/K_M) against the ester substrate in all cases. This result has been previously reported in *Lc*αE7 and *Md*αE7 (72, 223, 264–266). The decrease in catalytic efficiency is a major example for the mutant ali-esterase theory and suggests there is a potential cost of native function with the Gly137Asp mutation in all of the orthologs (2, 72, 223, 264, 265, 268). In most cases, the K_M is increased for the Gly137Asp variant, however, the exception is with the *Sc*αE7 ortholog. There is a statistically significant decrease in K_M with Asp at position 137 for the *Sc*αE7 variant. This was an unexpected result, however, for the Gly137Asp variant of *Sc*αE7, the k_{cat} is significantly reduced, resulting in a lower catalytic efficiency compared

to WT enzyme. This result was found to be consistent with multiple batches of purified *Sc*αE7 and the Gly137Asp variant.

Table 4.7. Summary of the kinetic data for the orthologs (WT and Gly137Asp) against the ester substrate 4-NPB.

Species	Variant	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M (s^{-1}/M^{-1})
Aedes aegypti (*Aa*A2)	WT	1317 ± 150	60 ± 11	2.2 x 10^7 ± 4.8 x 10^6
	Ala118Asp	16 ± 1.1	35 ± 6.6	4.6 x 10^5 ± 9.2 x 10^4
Anopheles gambiae (*Ag*B2)	WT	149 ± 17	20 ± 6.0	7.5 x 10^6 ± 2.4 x 10^6
	Gly118Asp	37 ± 2.7	68 ± 12	5.4 x 10^5 ± 1.0 x 10^5
Bactrocera dorsalis (*Bd*B1)	WT	6.9 ± 0.4	51 ± 4.1	1.4 x 10^5 ± 1.3 x 10^4
	Gly145Asp	1.2 ± 0.1	40 ± 8.0	3.0 x 10^4 ± 6.5 x 10^3
Ceratitis capitata (*Cc*αE7)	WT	3.9 ± 0.7	105 ± 22	3.7 x 10^4 ± 1.0 x 10^4
	Gly145Asp	0.2 ± 0.1	172 ± 21	1.1 x 10^3 ± 6.0 x 10^2
Calliphora stygia (*Cs*αE7)	WT	67 ± 1.0	82 ± 4.3	8.2 x 10^5 ± 4.4 x 10^4
	Gly137Asp	4.1 ± 0.1	101 ± 9.3	4.1 x 10^4 ± 3.9 x 10^3
Cochliomyia hominivorax (*Ch*αE7)	WT	199 ± 2.7	55 ± 2.8	3.6 x 10^6 ± 1.9 x 10^5
	Gly137Asp	2.3 ± 0.1	97 ± 17	2.4 x 10^4 ± 4.3 x 10^3
Culex quinquefasciatus (*Cq*B2)	WT	78 ± 2.2	20 ± 2.7	3.9 x 10^6 ± 5.5 x 10^5
	Gly118Asp	13 ± 0.2	52 ± 2.8	2.5 x 10^5 ± 1.4 x 10^4
Musca domestica (*Md*αE7)	WT	17 ± 1.4	45 ± 7.2	3.8 x 10^5 ± 6.8 x 10^4
	Gly137Asp	0.4 ± 0.1	137 ± 18	2.9 x 10^3 ± 8.3 x 10^2
Stomoxys calcitrans (*Sc*αE7)	WT	23 ± 0.5	67 ± 5.8	3.4 x 10^5 ± 3.0 x 10^4
	Gly137Asp	0.8 ± 0.1	33 ± 2.5	2.5 x 10^4 ± 2.0 x 10^3
Lucilia cuprina (*Lc*αE7)	WT	60 ± 1.1	34 ± 2.6	1.8 x 10^6 ± 1.4 x 10^5
	Gly137Asp	27 ± 0.7	109 ± 9.3	2.5 x 10^5 ± 2.3 x 10^4
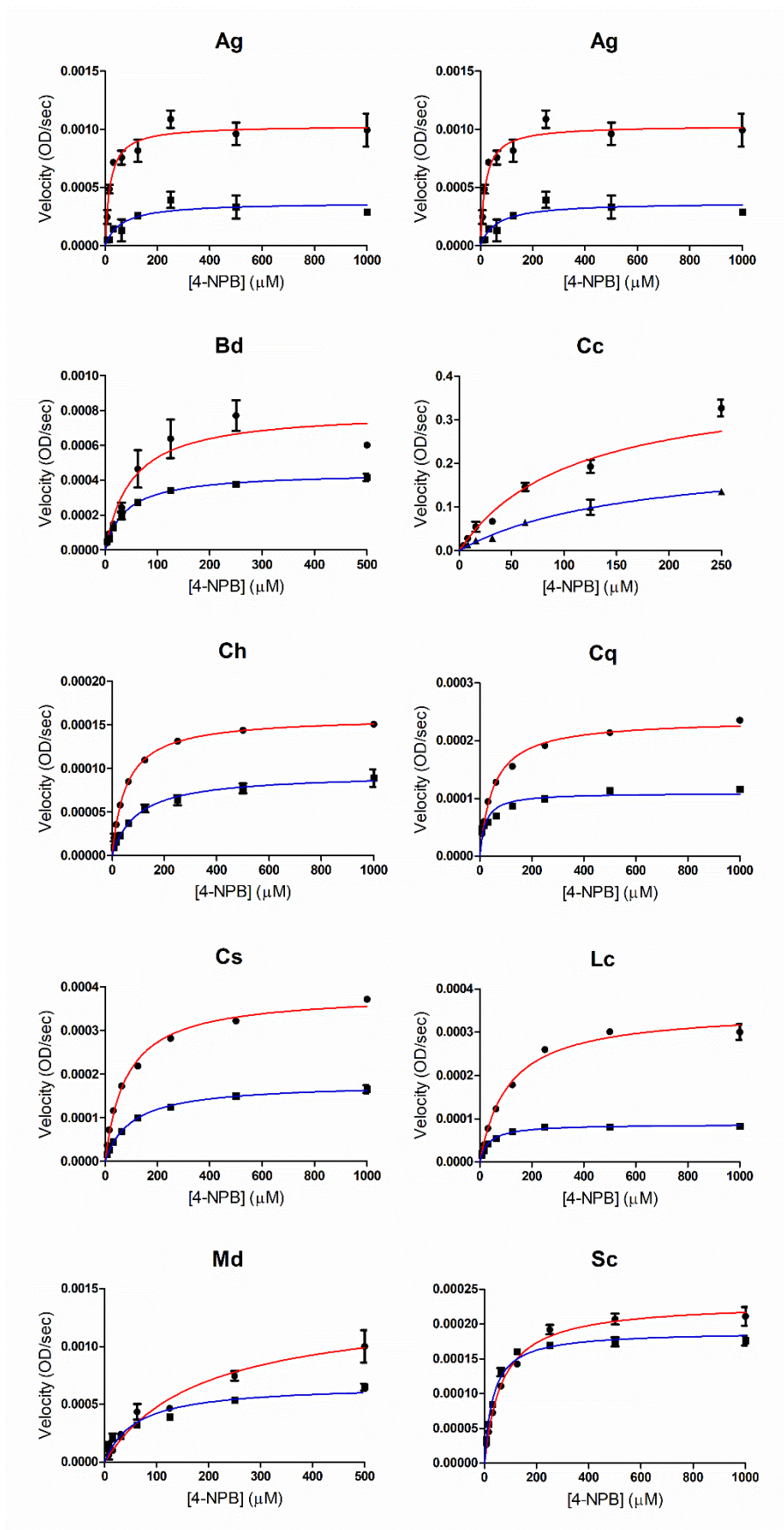
Figure 4.10. Michaelis-Menten kinetics graphs for the WT protein (red) and Gly137Asp variant (blue) of the 10 orthologs.

4.5.3. Kinetic Assays – Organophosphate Activity

The pre-steady and steady state kinetics of OP hydrolysis were monitored to understand the potential for OP hydrolase activity in the orthologs **(Figure 4.11) (Table 4.8)**. Overall, despite the closer similarity in sequence identity between the orthologs, there is a large variation in the OP hydrolase activity.

k_3 represents the rate constant for the formation of phosphorylated intermediate and in the orthologs, this rate is slow (<1). The introduction of the Gly137Asp mutation generally decreases k_3 in the orthologs with exception of *Cs*αE7. This indicates this variant has the potential for an increase in the formation of the intermediate. k_3 could not be calculated for four orthologs (*Ag*B2, *Bd*B1, *Cc*αE7, *Md*αE7) and therefore, the K_M for these species was estimated from the concentration dependence of k_5.

k_5 is the turnover rate or the rate of hydrolysis for the bound OP. This rate constant dictates the potential of the ortholog to be involved in the quantitative mechanism. For the majority of the orthologs, the introduction of the Gly137Asp mutation results in an increase of k_5. This suggests the mutation can potentially increase the turnover of the phosphor-Ser intermediate. In some cases (*Md*αE7 and *Sc*αE7), there is a decrease in the rate constant. The biggest change in k_5 is associated with *Lc*αE7, where the enzyme has a 28 fold increase in k_5 with the introduction of the Gly137Asp mutation. This indicates a high amount of OP turnover for this variant.

With the exception of *Cs*αE7 and *Lc*αE7, the Gly137Asp variants have a 1-2 fold increase or decrease in k_5. This shows the residue Asp137 has little effect on the turnover of the phosphor-Ser intermediate in the majority of the orthologs. This indicates the qualitative mechanism of resistance may not be selected for in some species by the orthologs tested. However, OPs are a strong selection pressure for insect species, therefore, any associated increase in OP hydrolysis could be selected for. The *Bd*B1 WT ortholog has high OP turnover and there is no statistical difference in rate with the Gly145Asp mutation. The rate of turnover for the WT protein is higher than any other WT ortholog and similar to *Lc*αE7 Gly137Asp. This indicates the expression of the protein could be sufficient for resistance and it is observed in the species (322).

The introduction of the Gly137Asp mutation in most cases, increases the K_M and shows the Asp residue interferes with OP binding. The largest increase in K_M is associated with *Cq*B2 (76 fold increase), however, there is no statistically change in K_M for the *Md*αE7, *Sc*αE7,

*Ch*αE7 and *Bd*B1 orthologs with the introduced mutation. There is a potential for inaccuracies with measuring the K_M as the differences in phosphorylation rate at different substrate concentrations are very small. This could explain the small change in K_M for the four orthologs.

Overall, the catalytic efficiency (k_5/K_M) shows a large variation with the introduction of Asp137. The protein *Lc*αE7 is the only ortholog to have a statistically significant increase in catalytic efficiency. The Gly137Asp mutation is associated with a decrease or no change in catalytic efficiency in the orthologs, with the large increase of K_M for the OP not equally compensated by k_5. The OP kinetic results shows despite the close similarity between the proteins, the orthologs have a large variation in OP kinetics with and without the Gly137Asp mutation. This shows OP turnover and binding is easily affected by differences in the protein sequence and structure.
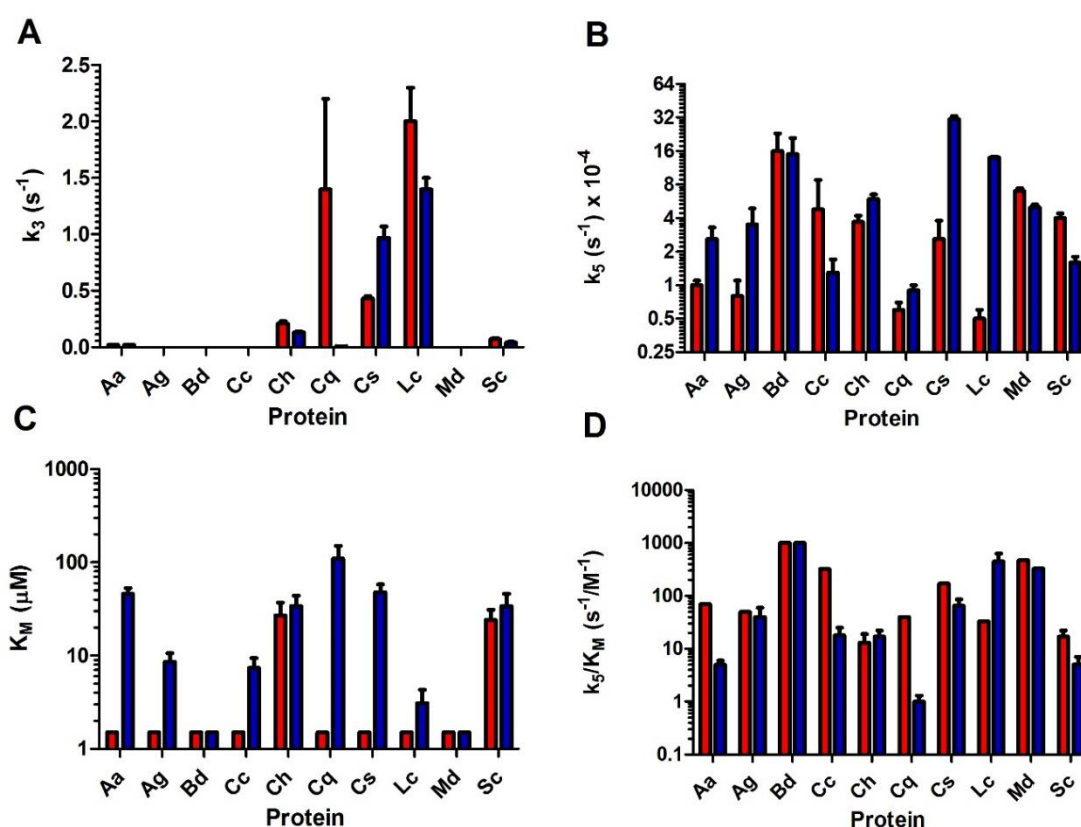
Figure 4.11. OP kinetic data for the 10 orthologs. WT protein of the ortholog is shown in red and Gly137Asp variant of the ortholog is in blue (A) k_3 (B) k_5 (C) K_M (D) k_{cat}/K_M.

Table 4.8. Summary of the kinetic data for the 20 Diptera proteins (WT and Gly137Asp) against the OP, DEUP.

Species	Variant	k_3 (s^{-1})	k_5 (s^{-1}) $\times 10^{-4}$	K_M (µM)	k_5/K_M (s^{-1}. M^{-1})
Aedes aegypti (*Aa*A2)	WT	0.02±0.001	1.0±0.1	<1.5	>70
	Ala118Asp	0.02±0.001	2.6±0.7	46±7	6±2
Anopheles gambiae (*Ag*B2)*	WT	-	0.8±0.3	<1.5	>50
	Gly118Asp	-	3.5±1.4	8.6±2	40±20
Bactrocera dorsalis (*Bd*B1)*	WT	-	16±7	<1.5	>1000
	Gly145Asp	-	15±6	<1.5	>1000
Ceratitis capitata (*Ccα*E7)*	WT	-	4.8±4	<1.5	>320
	Gly145Asp	-	1.3±0.4	7.4±2	18±7
Cochliomyia hominivorax (*Chα*E7)	WT	0.21±0.02	3.7±0.5	27±10	13±5
	Gly137Asp	0.13±0.01	5.9±0.6	34±10	17±5
Culex quinquefasciatus (*Cq*B2)	WT	1.4±0.8	0.6±0.1	<1.5	>40
	Gly118Asp	0.008 ±0.001	0.9±0.1	110±40	1 ±0.3
Calliphora stygia (*Csα*E7)	WT	0.43±0.02	2.6±1.2	<1.5	>170
	Gly137Asp	0.97±0.1	31±2	48±10	65±20
Lucilia cuprina (*Lcα*E7)	WT	2.0±0.3	0.5±0.1	<1.5	>33
	Gly137Asp	1.4±0.1	14±0.3	3.1±1.2	450±180
Musca domestica (*Mdα*E7)*	WT	-	7±0.4	<1.5	>470
	Gly137Asp	-	5±0.3	<1.5	>330
Stomoxys calcitrans (*Scα*E7)	WT	0.07±0.01	4.0±0.4	24±7	17±5
	Gly137Asp	0.04±0.01	1.6±0.2	34±12	5±2

*No pre-steady-state burst (*) was observed for *Anopheles gambiae* (*Ag*B2), *Bactrocera dorsalis* (*Bd*B1), *Ceratitis capitata* (*Ccα*E7) or *Musca domestica* (*Mdα*E7), the K_M for these species was estimated from the concentration dependence of k_5. Values are mean ± standard error for three replicates.

4.5.4. Structure alignment and analysis

To investigate if evolutionary contingency plays a role in the evolution of qualitative resistance in Diptera, the differences in sequence and structure was examined in further detail. Investigation of the sequence and structural differences between the proteins could suggest potential mutations that influence levels of activity.

First, the active site sequence changes were analysed including the second and third shell residues that are known to affect substrate binding **(Table 4.9)**. A large body of work has focused on OP hydrolase activity in the enzyme *Lc*αE7. The substrate-binding pocket of *Lc*αE7 is asymmetrical and is made up of the p1 subsite (Trp251, Met308, Phe309, Phe355, Tyr420, Phe421) and the p2 subsite (Tyr148, Phe354, Tyr457) **(Figure 4.12, Panel A)** (55). Second and third shell residues have been identified to have a significant effect on OP binding and the rate of OP turnover (315, 316).
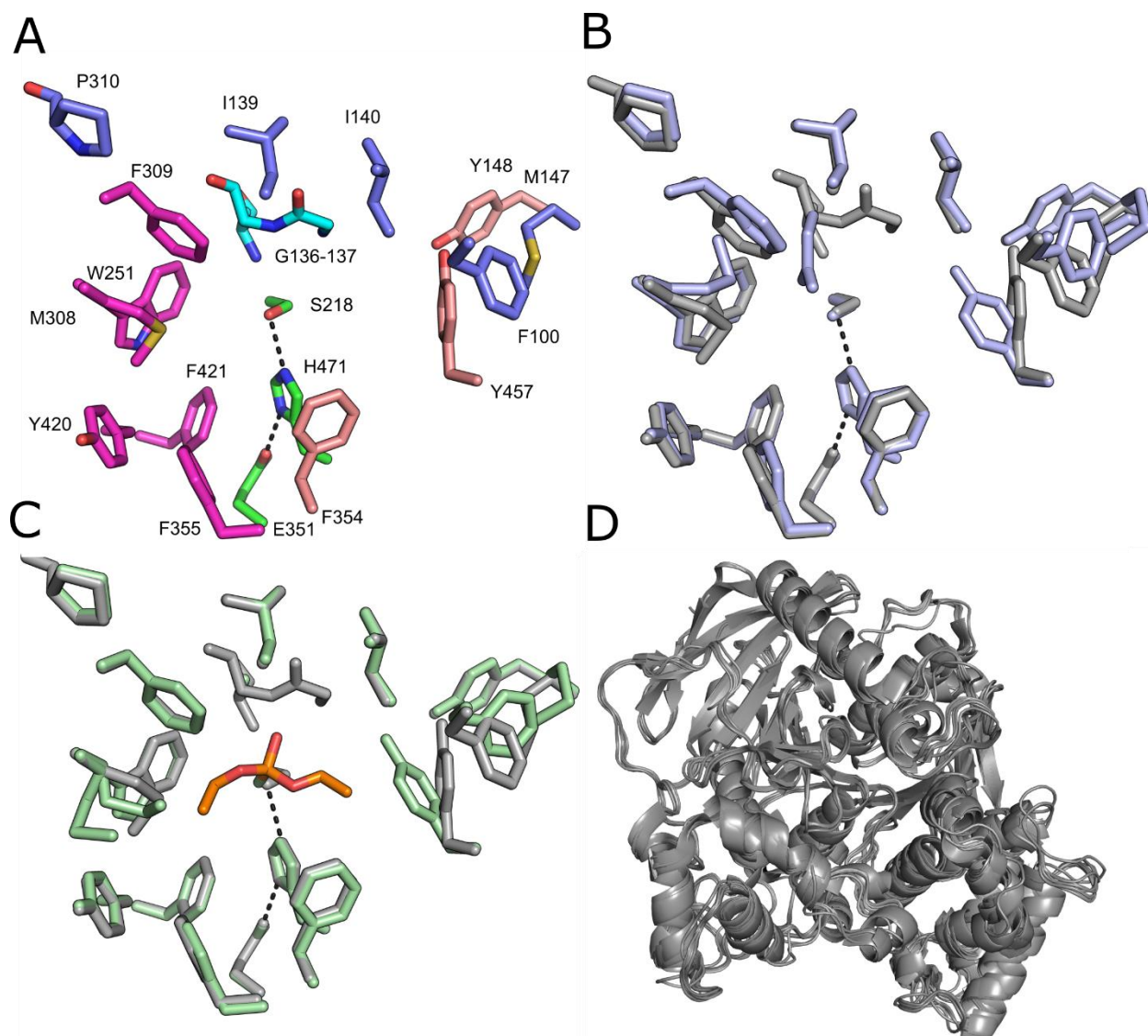
Figure 4.12. (A) Active site residues of WT *Lc*αE7 including the catalytic triad (green), oxyanion hole (cyan), p1 subsite (pink), p2 subsite (wheat) and the 2nd and 3rd shell residues (blue) **(B)** The structure of WT *Lc*αE7 (PDB ID: 4FNG) (grey) aligned against the structure of *Lc*αE7 Gly137Asp (blue) (PDB ID: 5C8V) **(C)** The structure of WT *Lc*αE7 aligned against the structure of WT *Lc*αE7 bound with diethyl hydrogen phosphate to the active site serine (PDB ID: 4FNM) **(D)** Cartoon structure of *Lc*αE7 aligned against the solved structure of *Cq*B2 and the 8 modelled orthologs.

Table 4.9. Analysis of the variability in residues making up the active site of the 10 proteins.

Esterase		*Lc*	*Cs*	*Ch*	*Md*	*Sc*	*Cq*	*Aa*	*Ag*	*Bd*	*Cc*
Catalytic		S218	S	S	S	S	S	S	S	S	S
Triad		E351	E	E	E	E	E	E	E	E	E
		H471	H	H	H	H	H	H	H	H	H
Oxyanion		G136	G	G	G	G	G	G	G	G	G
Hole		G137	G	G	G	G	G	A	G	G	G
		A219	A	A	A	A	A	A	A	A	A
p1 subsite		W251	W	W	W	W	W	W	W	W	W
(leaving		M308	M	M	M	M	F	M	L	M	S
group		F309	F	F	F	F	T	F	S	F	T
pocket)		F355	F	F	F	F	L	S	L	M	F
		Y420	Y	Y	Y	Y	V	L	I	L	Y
		F421	F	F	F	F	F	F	F	F	L
p2 subsite		Y148	Y	Y	F	Y	Y	Y	Y	Y	Y
(acyl binding		F354	L	L	L	L	L	F	I	V	I
pocket)		Y457	Y	Y	Y	F	Y	Y	Y	Y	Y
2nd and 3rd		F100	F	F	F	I	R	K	R	L	I
shell residues		I139	I	V	I	I	T	N	V	V	C
		I140	I	I	F	C	E	R	E	V	T
		M147	M	Y	W	Y	L	M	L	W	W
		P310	A	A	P	A	P	E	P	P	P

*Non-conservatives changes in residues from *L. cuprina* are highlighted in red. Conservatives changes in residues from *L. cuprina* are highlighted in blue.

The second shell residues Met308 and Phe309 in the *Lc*αE7 structure are dynamic in nature and are important for the correct orientation of Asp137 for hydrolysis of the phosphor-Ser intermediate (315). Asp137 is highly mobile and requires a conformational change from the apoenzyme state to a productive state (315). Met308 and Phe309 undergo a conformation change for the correct orientation of Asp137 to be adopted, to allow for turnover of the intermediate **(Figure 4.12, Panel C)**. As Asp137 frequently adopts non-productive conformations, it suggests small changes in residues and structure could have significant effects on the conformational organization of Asp137. At positions Met308 and Phe309, there are three orthologs with changes at these positions, *Cq*B2 (Phe, Thr), *Ag*B2 (Leu, Ser) and *Cc*αE7 (Ser, Thr). It is unknown how the changes effect the OP kinetics, but as the 308 and 309 positions are important for the correct orientation of Asp137 in *Lc*αE7 for OP turnover, it could explain the differences in OP hydrolase activity for the orthologs.

Other residues identified to affect OP kinetics in the *Lc*αE7 structure include Tyr457, Met147, Tyr148 and Phe100 (316). In the open state, these residues are in a dynamic

network that is lost when the phosphorylated intermediate state is formed. Replacement of Tyr457 or Met147 to Ala affects the dissociation constant of the OP-$Lc\alpha$E7 Michaelis complex, which increases the K_D (316).

For analysis of the sequence differences in more detail, modelling of the proteins was performed using the I-TASSER server (325). The I-TASSER server has a reputation of being the most accurate modelling server based on the **C**ritical **A**ssessment of Protein **S**tructure **P**rediction (CASP) (326). Currently, there are crystal structures of two orthologs, CqB2 (unpublished) and $Lc\alpha$E7 (PDB ID: 4FNG). The proteins share 38% sequence identity to each other (55). Despite the low sequence similarity between the proteins, they share a close similarity in structure when aligned (RMSD: 1.102). This suggests modelling of other Diptera orthologs will have some level of accuracy. The alignment of the 10 orthologs have significantly overlap (RMSD range from 0.275-1.012), however, differences are observed in the active site makeup **(Figures 4.12 Panel D, 4.13)**.

Comparing the binding affinity for the OP between the WT orthologs, the K_M is below 1.5 µM for 8 of the orthologs with the exception of $Sc\alpha$E7 (24 µM) and $Ch\alpha$E7 (27 µM). The active sites of the proteins are similar to $Lc\alpha$E7. However, for $Sc\alpha$E7, Phe is found at position 457 instead of Tyr **(Figure 4.13, Panel I)**. This site is important for the binding of the OP and the change to Phe457 is likely to be responsible for the poor binding that is also observed for the Gly137Asp variant (34 µM). In the modelled structure of $Sc\alpha$E7, Phe457 is positioned away from the active site and not in a position to interact with the bound OP. Tyr is found at the second shell position of 147 instead of Met for both proteins and could have another effect on binding. This residue in $Lc\alpha$E7 is involved in a dynamic network with Tyr457 for OP binding. Replacement to Tyr is likely to effect the network and resulting OP binding.

The Asp137 residue increases the K_M, as the bulky size of the Asp residue in the oxyanion hole results in a steric clash with the bound OP intermediate (316). The Gly137Asp variants of the orthologs show large variations in K_M, generally with an increase in K_M compared to the WT enzymes. This is found to be significant in the CqB2 Gly118Asp ortholog where the K_M is 110 µM compared to the K_M of <1.5 µM for the WT enzyme. For the binding of the OP in $Lc\alpha$E7, Phe309 buries into the structure for the productive conformation of Asp137. However, in the structure of CqB2, Met308 and Phe309 are replaced by Phe308 and Thr309. The Phe308 residue in CqB2 occupies the space between Met308 and Phe309

in *Lc*αE7 **(Figure 4.13, Panel F)**. Phe308 in the structure of *Cq*B2 has limited movement and would be in a steric clash with Asp137. This would result in a little change of OP turnover and very poor affinity for the OP.

For the other orthologs, there are multiple reasons for the large increase in K_M compared to *Lc*αE7 with the introduction of Asp137. Pro is found at position 310 in *Lc*αE7, however, is found as Ala in the orthologs *Cs*αE7, *Ch*αE7, and *Sc*αE7. This third shell residue is adjacent to the dynamic second shell residues Met308 and Phe309. From the structure, it could be predicted that this could increase the movement of Met308 and Phe309 and hinder OP binding **(Figure 4.13, Panels E, G, I)**. For the *Aa*A2 protein, the residue Arg at position 140 obstructs Asp137 in the modelled structure **(Figure 4.13, Panel A)**. This will steric clash with Asp137 and will likely cause a high K_M and poor turnover of the OP. The Glu residue at position 310 also interferes with the ability of Phe309 to bury itself for the productive conformation of Asp137.

The *Md*αE7 protein has little change in K_M with the introduction of the Asp137 and from the modelled structure it is difficult to provide explanations **(Figure 4.13, Panel H)**. The residue Phe140 could play a role, as well as the second shell Trp and Phe residues at positions 147 and 148 instead of Met and Trp. The position of Phe140 in the structure appears to shift the residue Ile139. The shift causes the Ile139 to a steric clash with Asp137 and it could explain the decrease in k_5. Arg461 in the modelled structure of *Md*αE7 is orientated towards Asp137 and could be another explanation for the low OP turnover.

*Cc*αE7 has a number of active site changes compared to *Lc*αE7 including the second shell positions 308 (Ser) and 309 (Thr). Though changes in these residues do not effect K_M in a large amount for the Gly145Asp variant, it appears to interfere with the ability to turnover OPs. This suggests Asp145 cannot reach an optimal conformation for hydrolysis in *Cc*αE7. The residue Val312 in the modelled structure also steric clashes with Asp145. *Ag*B2 has significant changes in the second and third shell residues including Asn at position 139 **(Figure 4.13, Panel B)**. This residue may interfere with Asp137 and could explain the small change in k_5 associated with the variant enzyme.

The ortholog *Cs*αE7 shares the highest amino acid identity to *Lc*αE7 (89%), as well as the fewest changes in the active site. Surprisingly, the Gly137Asp variant of *Cs*αE7 has a higher rate of OP turnover (k_5) (31 vs. 14 s⁻¹), but a significantly higher K_M (48 vs. 3 μM). This results in lower catalytic efficiency (65 vs. 450 s⁻¹. M⁻¹). In the structure it is difficult

to rationalize the differences in OP kinetics between the enzymes. However, the third shell residue Ala310 instead of Pro310 could affect the dynamic nature of residues Met308 and Phe309, which are important for positioning of Asp137 for OP hydrolysis.

In regards to the rate constant k_5 (OP turnover), there is a large variation across the WT orthologs. However, it is unknown what factors dictate and affect k_5 in the WT proteins. With the introduction of Asp137, there are some orthologs with little to no increase of k_5. The poor OP turnover for the Gly137Asp variant of $Sc\alpha$E7 could be from the presence of a cysteine residue at position 140 in the active site. This residue could hinder the correct positioning of Asp137 for hydrolysis. For $Ch\alpha$E7, Val at position 139 causes a steric clash with Asp137 and could explain a low k_5 for the Gly137Asp variant of this enzyme. BdB1 is unique for the orthologs where the enzyme has a high k_5 and low K_M in the WT enzyme and the Gly145Asp variant. It is difficult to explain why, however, two changes in the active site include Trp at second shell positions 147 and 460. These residues could be responsible for high affinity and turnover for the OP.
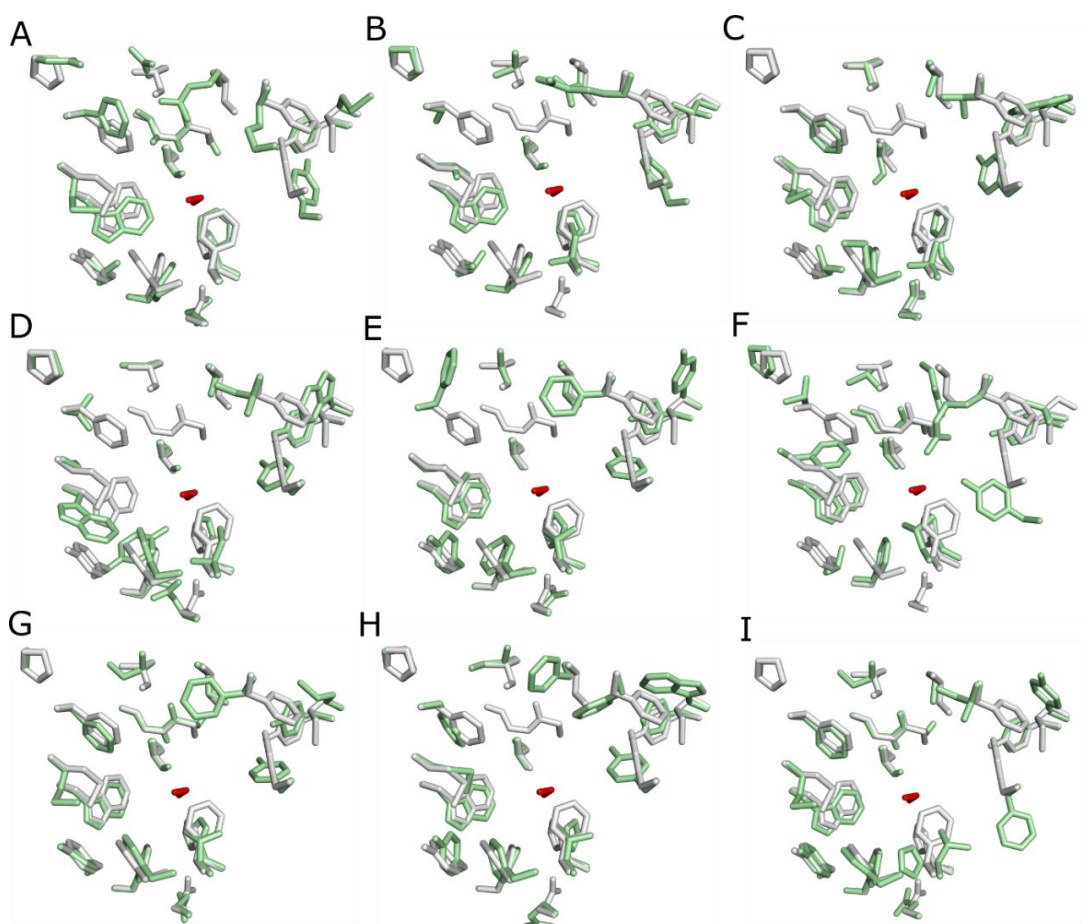
Figure 4.13. Active sites of the modelled orthologs (green) aligned with the structure of $Lc\alpha$E7 (grey). The active site serine is shown in red **(A)** AaA2 **(B)** AgB2 **(C)** BdB1 **(D)** $Cc\alpha$E7 **(E)** $Ch\alpha$E7 **(F)** CqB2 **(G)** $Cs\alpha$E7 **(H)** $Md\alpha$E7 **(I)** $Sc\alpha$E7.

4.6. Discussion

In this work, the effect of the point mutation at position 137 to Asp has been investigated in 10 dipteran CBEs. This work was performed to understand if the ability of the Asp137 residue to hydrolyse OPs is unique or a general characteristic common to close orthologs of $Lc\alpha$E7. The results confirm that OP hydrolase activity is increased in some dipteran orthologs with the mutation, albeit at the significant cost of OP affinity (K_M), which results in a poor catalytic efficacy. In some cases, there is no change in OP turnover with the introduced mutation. This supports the idea that the Gly137Asp mutation is limited in terms of being a viable route for resistance and there are limited options for evolving CBE-based metabolic resistance to OPs (268). Previous studies into the aspartate mutation in nonspecific CBEs from various insect species have supported the idea that the Gly137Asp mutation is a limited route for resistance (278, 279). In some cases, there is no increase in OP hydrolysis with the introduction of the mutation. When OP hydrolysis is increased, there is a total loss of native CBE function suggesting the mutation has a significant negative effect on fitness (278, 279).

Three of the proteins ($Lc\alpha$E7, $Ch\alpha$E7, $Md\alpha$E7) tested for OP activity are involved with insecticide resistance by the Asp137 mutation in wild blowfly species while three of the proteins tested (BdB1, AaA2, CqB2) are involved with resistance by upregulation without the Asp137 mutation present (2, 223, 262, 267, 322, 323). It is unknown if the other four proteins tested ($Cs\alpha$E7, $Sc\alpha$E7, $Cc\alpha$E7, AgB2) are involved in resistance. For the two mosquito proteins, CqB2 and AaA2, the introduced Asp137 mutation results in a significant increase in K_M and minor increases to k_5 compared to the three blowfly proteins ($Lc\alpha$E7, $Ch\alpha$E7, $Md\alpha$E7) where k_5 is increased significantly and changes to K_M are minor. Slight changes in the structure are likely to explain why the Asp137 cannot adopt the conformation for sufficient OP hydrolysis activity in the mosquito species. For the Asp137 residue to be favourable for resistance, it must not affect K_M in a large amount for the protein to be able to scavenge the OP efficiently. The residue must also result in sufficient OP hydrolysis activity otherwise resistance is achieved by other means (upregulation). This is likely to be the reasoning why the mosquito species achieve resistance by upregulation as they lack both features for qualitative resistance in the selected proteins. The three blowfly species that have resistance by the Asp137 mutation have a sufficient k_5 rate for OP hydrolysis activity to provide insecticide resistance to the strong selection pressure of OPs and a relatively low K_M with the mutation to effectively scavenge the OP.

The *Bd*B1 protein is involved in OP resistance by upregulation, the protein itself has a high turnover and low K_M, and therefore, the benefits the Asp137 mutation offers may not be necessary for the species to achieve resistance. The four other proteins tested (*Cs*αE7, *Sc*αE7, *Cc*αE7, *Ag*B2) may not be involved in resistance by the Asp137 mutation with similar reasoning to the two mosquito proteins. The minor increase in OP hydrolytic activity is matched with a significant increase in K_M and therefore not favourable for selection with the Asp137 mutation. The low K_M for the WT proteins and the reasonable turnover rate could suggest the proteins may be involved in resistance through upregulation and be able to scavenge the OP effectively. *Cc*αE7 and *Ag*B2 also require two nucleotide changes from Gly to Asp and therefore, it would be very unlikely for the Asp mutation to appear. The kinetic data from the 10 proteins suggests that the qualitative resistance mechanism through the Asp137 mutation may be a less common resistance mechanism in dipteran than previously thought.

Homology models of the orthologs show amino acid differences in the active site and second shell residues affect the position of Asp137. This prevents the residue sampling optimal geometries for hydrolysis of the phosphor-Ser intermediate. The active site and second shell residues between the orthologs have varied over time, before the selection pressure of OPs. This suggests random neutral drift has produced few esterases that can evolve resistance with a sufficient OP turnover, without losing native function and a large increase in K_M for the OP. This also suggests historical contingency plays a role in the protein structure, affecting the mechanism of insecticide resistance observed. This is supported by a laboratory mutagenesis experiment in the fruit fly species *D. melanogaster*. It was observed that resistance to OPs does not occur through the protein, *Dm*αE7 (72). This protein shares 74% sequence identity to *Lc*αE7 and it was suggested the enzyme's low affinity for OPs (lower than AChE) does not allow for resistance (72).

There are a number of additional reasons why the mutation Gly137Asp is not present in the orthologs in nature and the quantitative resistance mechanism is common in insect species. In two of the proteins (*Aa*A2, *Bd*B1), two nucleotide substitutions are required to convert Gly to Asp, which suggests a low chance for resistance by the quantitative mechanism to evolve. In other species, P450s or GSTs could be responsible for insecticide resistance, along with target site insensitivity. Another reason could be the loss of wild-type function and therefore, another resistance mechanism may be selected for (278, 279). The

quantitative mechanism is also able to sequester a wider range of insecticides than the qualitative mechanism of resistance and therefore, offering greater protection for insect species (32, 278, 279). Other factors contributing to the widespread of the quantitative mechanism of resistance in insects is from the limitations in the eukaryote genetic systems in introducing multiple mutations and recruiting genetic variation in a short amount of time (233). This limitation results in difficulties in modifying an enzyme's active site for a new function, few options for resistance, and convergent evolution (233).

It has been argued that the living world would be very different if evolution was replayed, however, it has also been argued there are limited evolutionary routes life can take (282, 309). This will result in convergent evolution and contingency only affecting the minor details. This is found in insecticide resistance where selection finds similar adaptations and suggests evolution is repeatable (193, 195, 221). However, contingency affects the mechanism of resistance (quantitative or qualitative), which results in the small happenstances of history leading to different evolutionary paths of resistance (299). Given this, it makes predicting resistance difficult and furthermore, complications in predicting the resistance mechanism based on the choice of insecticide (274, 327).

Overall, a detailed kinetic comparison of the wild-type and Gly137Asp variants of *α*E7 orthologs from 10 dipteran species has suggested contingency has played a role in the evolution of resistance to OPs. Comparison of the sequences and structures of the *L. cuprina α*E7 (*Lc*αE7) and several αE7 orthologs reveals that amino acid substitutions in the vicinity of residue 137 prevent the newly introduced Asp from sampling conformations necessary for the catalytic cycle effectively. However, the amino acid substitutions that restrict Asp137 do not substantially impact on the native CBE activity of the enzyme. Genetic drift before the introduction of insecticides predisposed blowflies to a mutational path that was not accessible to other insect species. This suggests that evolutionary innovation, even in the presence of strong selection pressure, is limited by the initial states or past events.

Chapter 5. Inhibitors of Acetylcholinesterase

5.1. Declaration of Author Contribution

The research article in this chapter was peer-reviewed and published as a regular original research article. The synthesis of the compounds was conducted by Joshua Buckler and Ehab Taher. The crystallography work was done by Anthony Willis and Paul Carr. I performed the molecular docking simulations of the compounds and kinetic analysis. The writing of the draft was conducted by myself, Colin J. Jackson, Martin Banwell and Joshua Buckler. The research in the chapter involving the marinoquinolines was mainly conducted by myself and members of the Dr. Martin Banwell's research group. The docking of the compounds and analysis were performed by myself and the synthesis of the compounds was managed by members of the Dr. Martin Banwell's research group.

5.2. Introduction

5.2.1. Acetylcholinesterase

Acetylcholinesterase (AChE) is a member of the carboxyl/cholinesterase gene family that is the target for OP insecticides which reduce pest invertebrate populations (17, 197). AChE catalyses the hydrolysis of the excitatory neurotransmitter, acetylcholine (Ach) (17, 197). In invertebrates, Ach is the prevalent excitatory neurotransmitter in the central nervous system (CNS), however, in the CNS of vertebrates, glutamate mediates most of the excitatory neurotransmission (328, 329). Instead of playing a strong role in the CNS, Ach is primarily found in the neuromuscular junctions and nerve-electro plaque junctions in vertebrates (328, 329). The hydrolysis of Ach by AChE is important to regulate neurotransmission and occurs in cholinergic brain synapses **(Figure 5.1)** (4, 330). The rapid degradation of Ach at the synaptic junction results in decreases in and termination of the signal transmission in nerve synapses (330). Hydrolysis occurs at an extremely fast catalytic rate (25000 molecules of Ach per second), approaching the diffusion controlled limit (202, 331).
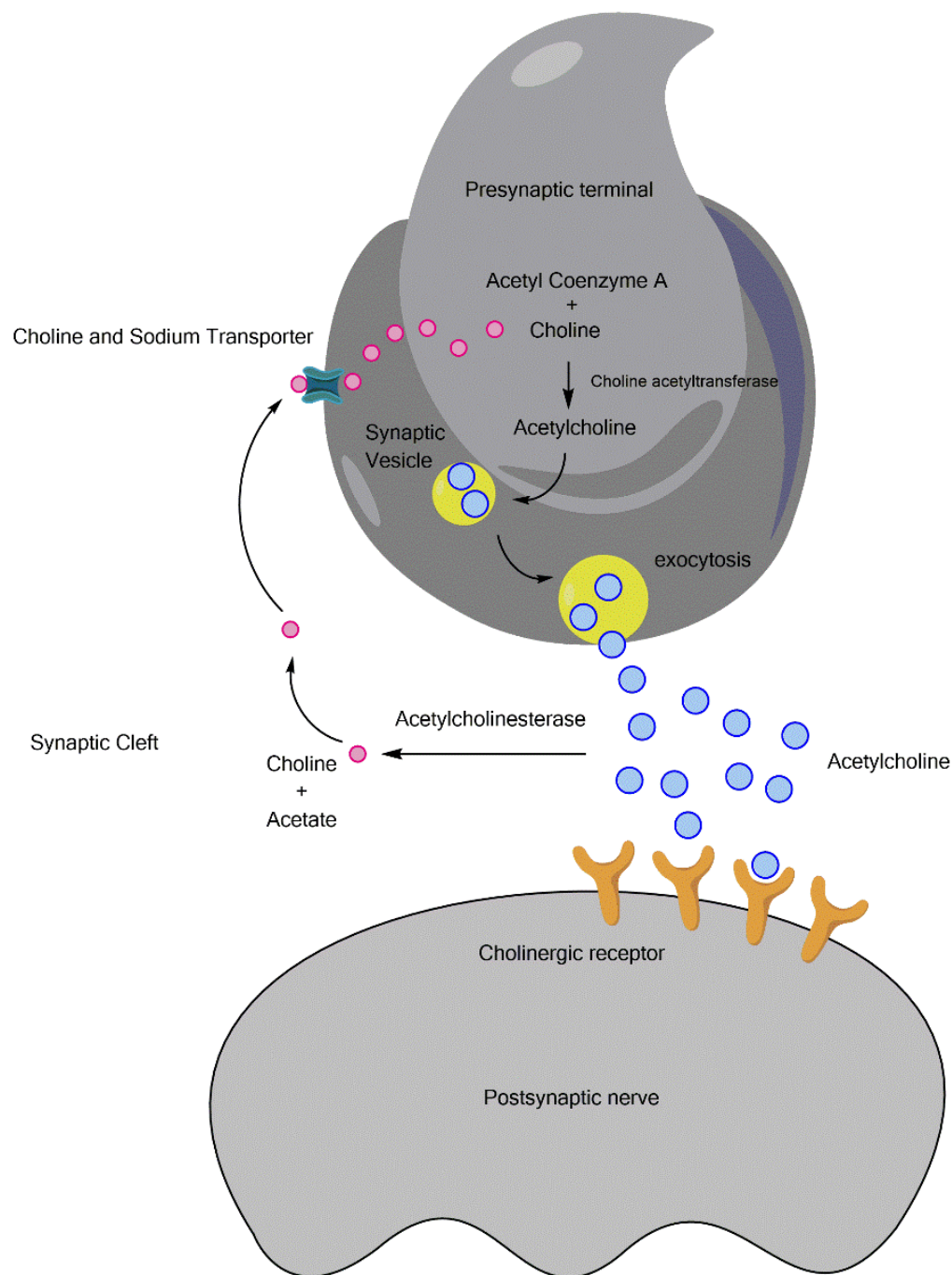
Figure 5.1. Transmission of the nerve impulse with the release of Ach from the presynaptic nerve terminal to the Ach receptors on the postsynaptic nerve terminal. After signalling, Ach is released from receptors and broken down by AChE to be recycled in a continuous process (330). Figured adapted from Lilienfeld (332).

The X-ray crystal structure of AChE was solved in 1991 and was important to understand the fast catalytic rate of the enzyme (53). It was thought due to the high rate of Ach turnover, the binding site of AChE would be found on the surface of the protein. However, the catalytic triad and binding site of AChE are found at the bottom of a deep narrow binding gorge (53). The enzyme contains five main sites, which includes the catalytic triad (Ser,

94

His, Glu) and the oxyanion hole **(Figure 5.2)** (333–336). The third site is the catalytic anionic site that is comprised of aromatic residues. The aromatic residues bind the positive quaternary choline moiety of Ach *via* cation-π interactions (333–336). The acyl binding site is also found in the binding site, this region determines selectivity for AChE (333–336). The last site of AChE is the peripheral anionic site (PAS), this site is found at the entrance of the active site gorge and allosterically modifies catalysis (333–336). The binding of inhibitors to this region reduces Ach's ability to enter the active site, leading to substrate inhibition (333, 337–339).
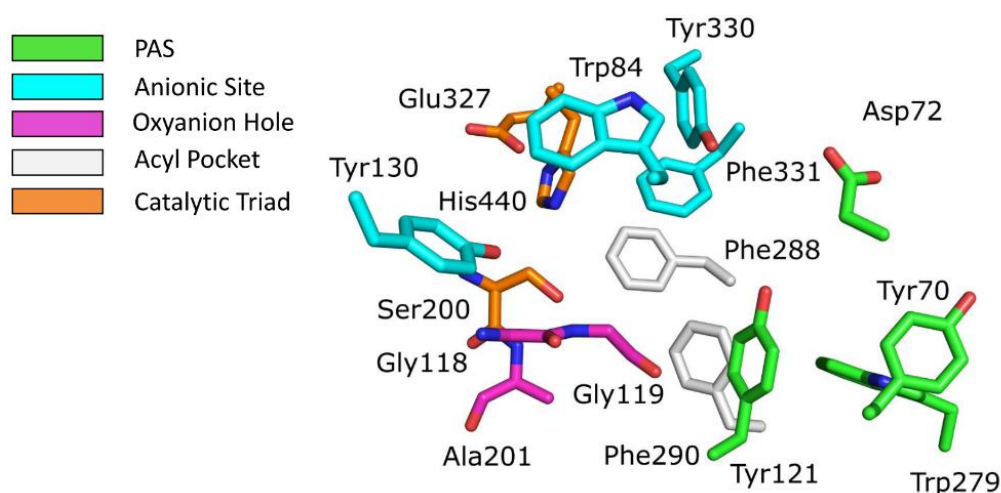
Figure 5.2. The active site make up of AChE including the five regions: the catalytic triad, the oxyanion hole, the peripheral anionic site, the acyl pocket and the anionic site (53, 333–336). Figure adapted from Bajda et al (340).

5.2.2. Alzheimer's disease

AChE is a drug target for treating the progressive neurodegenerative disorder Alzheimer's disease (AD). This is due to the key role the substrate Ach plays in memory, cognition and connectivity between neurons (341). AD affects over 20 million individuals worldwide and is set to increase, becoming a major public health issue (342, 343). Impairment of cholinergic function, decreased cholinergic neurons, β-amyloid plaque formation and the decreased level of Ach is strongly associated with AD **(Figure 5.3)** (344, 345).

Symptoms

Mood and Behaviour changes

Loss of memory

Disorientation in time and space

Poor judgement

Language problems

Observed

Impaired cholinergic function

Decreased cholingeric neurons

Beta-amyloid plague

Decreased level of Ach
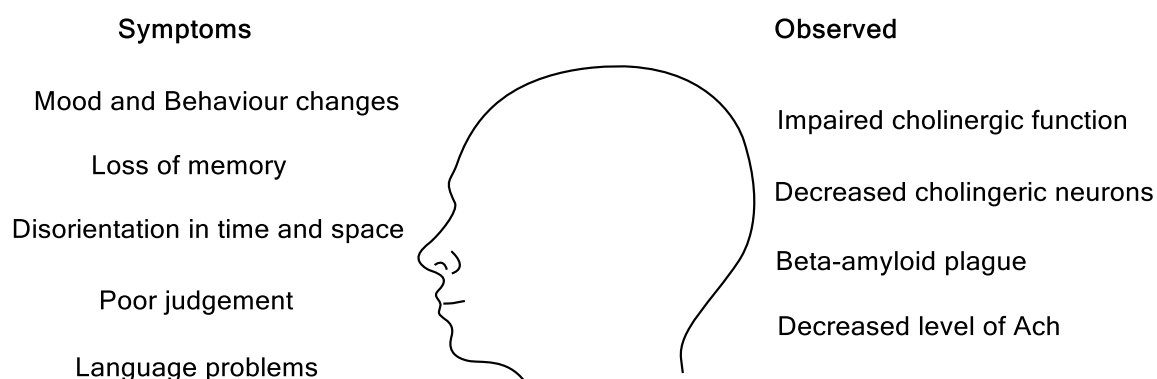
Figure 5.3. Symptoms and the physical observations in patients affected by AD (342).

There are two hypotheses for the pathology of AD, either from the decrease in Ach production, or by the increase of toxic β-amyloids aggregating in the brain (343, 346–348). It has been suggested that inhibitors of AChE can slow the progression of AD (343, 349). Inhibiting AChE results in an increase in both level and duration of the Ach in the synapse. This compensates for the reduction in cholinergic function, particularly at the synaptic terminals (343, 347–349). AChE also has a role in accelerating β-amyloid aggregation and deposition onto fibrils. This is associated with the PAS of the enzyme (350). Therefore, inhibition of the PAS and the AChE active site are desired to target both Ach hydrolysis and plaque formation (348, 349).

5.2.3. AChE Inhibitors to treat AD symptoms

Due to the association of AChE in AD, it has been a major therapeutic target for treating AD (344, 351). Though there is no cure for AD, reversible inhibitors of AChE treat the symptoms of AD **(Figure 5.3)** (352, 353). The discovery of AChE inhibitors to treat AD symptoms has resulted in some promising leads (354). Physostigimine was the first reversible inhibitor of AChE used for the treatment of AD but only had a small beneficial effect in AD patients with many side effects (355). For similar reasons, metrifonate, a non-active prodrug was discontinued from use before approval for significant muscle weakness side effects (354, 356). The compound is transformed to dichlorvos in vivo non-enzymatically and is an irreversible cholinesterase inhibitor (354, 356). Huperzine A is a promising new option in treating AD symptoms (354, 355). Huperzine A is an alkaloid that is a reversible and highly selective inhibitor of AChE (357). This compound is more potent than the three approved drugs and has the least amount of inhibition to BChE (355, 357, 358). It is currently used in China for treatment of memory disorders and AD symptoms (357, 358). Huperzine A is currently under trials in Europe and remains the next promising

inhibitor to treat symptoms of AD (355, 357, 358). Along with huperzine A, phenserine is a possible new drug candidate for targeting AChE (355). The compound is a pseudo-irreversible cholinesterase inhibitor with a strong preference for AChE (355, 359) Phase II trials of the compound had mixed results but there may still be promise for this compound to treat AD symptoms (359). Currently, three AChE inhibitors have been approved by the Food and Drug Administration in the U.S.A and the European Medicines Agency in Europe for use to treat AD symptoms to improve life quality (344, 360). The compounds are rivastigmine, donepezil, and galanthamine **(Figure 5.4)** (344, 360). All three compounds inhibit both AChE and BChE with a difference in potency and selectivity, however, all three demonstrate similar efficacy for treatment of AChE (361).

Galanthamine is a reversible AChE inhibitor that is further discussed in the next section (362, 363). The compound interacts with AChE, nicotinic Ach receptors and weakly inhibits BChE (362, 363). Donepezil is another reversible AChE inhibitor with a high specificity for AChE and the CNS, with limited activity in the peripheral tissue (363, 364). Donepezil was the second approved anti-Alzheimer's drug and interacts with the active site and the PAS of AChE (363, 364). The compound also affects serotonin concentrations, which can influence the mood of AD patients (365). Rivastigmine is the third approved AChE inhibitor drug to treat AD symptoms and is a pseudo-irreversible inhibitor (363, 366). The drug binds covalently to AChE and BChE forming a carbamylated complex (363, 366). After the covalent link is formed, the complex is slowly hydrolysed over several hours. Similar to donepezil, rivastigmine has CNS selectivity and is rapidly excreted without binding to plasma proteins with limited interaction in the peripheral tissue (367, 368).

Tacrine is a rapidly-reversible inhibitor of AChE that was previously used to treat AD, however, was discontinued due to the fatal long-term side effects including hepatotoxicity (369, 370). The compound is non-selective and inhibits both cholinesterases (AChE and BChE) with a strong preference for BChE (371). Tacrine also has several other interactions including overstimulation of the peripheral cholinergic system, inhibition of L-type calcium channels, monoamine oxidase inhibition, potassium channel blockade and inhibition of the muscarinic and nicotinic receptors (372, 373). Improving the selectivity for AChE is important for fewer side effects and it has been proposed that high amounts of BChE inhibition are responsible for the unwanted side effects (363, 374).

Recently it has been shown AChE isoforms react differently to AChE inhibitors and raises questions on isoform selectivity for future drug development (375). G_4 is a tetrameric isoform of AChE that is found in the mammalian brain and cholinergic synapses, it is also the relevant target for treating AD symptoms (375–377). On the contrast, the monomeric G_1 isoform of AChE is found primarily in the cytoplasm and is unlikely to affect synaptic physiology in large amounts (375–377). Selective inhibition of the G_4 isoform is suggested to prolong the action of Ach, necessary to treat AD symptoms (375). The currently approved drugs show a large variety in inhibition of the isoforms. Rivastigmine inhibits G_1 preferentially while donepezil varies selectivity between the G_4 and G_1 isoforms of AChE depending on the region but inhibits the G_4 isoform of AChE with high potency in the cortex (375, 378). Tacrine and galanthamine have no preference and bind the G_4 and G_1 isoforms equally well. Future testing and development of effective inhibitors would inhibit brain AChE (G_4) without any effect on the peripheral tissue AChE (G_1) (375, 378). The G_4 isoform of AChE emerges as a potential molecular target for development of AChE inhibitors to treat AD symptoms and another consideration in drug development (375, 379, 380).
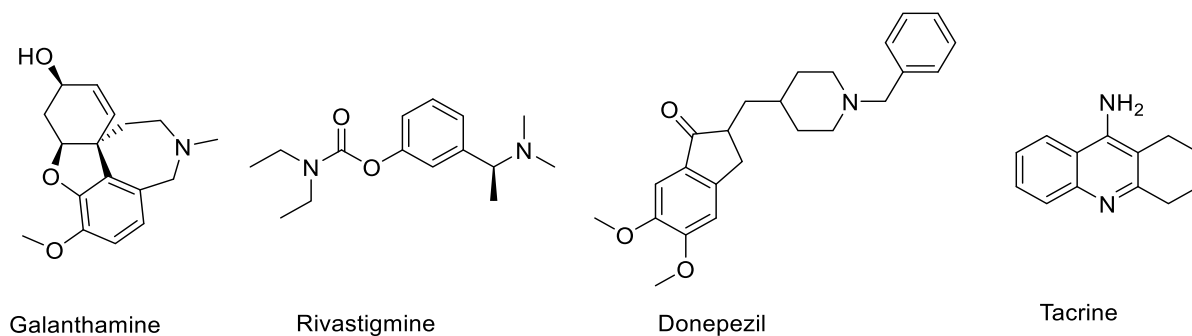
Figure 5.4. The four drugs that have been used for the treatment of AD, galanthamine, rivastigmine, donepezil and tacrine (currently not in use) (344).

5.3. Preface

In this chapter, two types of inhibitors that target AChE will be investigated. In the first section, derivatives based on the drug galanthamine have been designed and tested for inhibition against AChE. Section two describes a new potential set of drugs to treat AD identified as the marinoquinolines that are similar to tacrine in structure.

Section 1. The synthesis of certain derivatives and analogues of (−)- and (+)-Galanthamine and an assessment of their capacities to inhibit acetylcholine esterase

5.4. Introduction

5.4.1 Galanthamine

(-)-Galanthamine (GAL) is an alkaloid obtained from the bulbs and flowers of *Galanthus caucasicus* and related *Amaryllidaceae* family species (381). It was first used as a drug for the treatment of myasthenia gravis, motor disorders of the CNS and residual poliomyelitis paralysis syndromes (381, 382). GAL has been widely successful in treatments, as the compound is able to cross the blood-brain barrier (381). GAL in the 1980s was investigated for the treatment of mild to moderate forms of AD. Following the initial trials, the compound was approved for use in Europe, U.S.A and Asia by the 2000s as GAL exhibits a significant improvement in cognitive function for those affected by AD (347, 383, 384).

GAL is a selective and reversible AChE inhibitor (IC_{50} = 800 nM) that interacts with the anionic site and the aromatic gorge of AChE (385–387). GAL also allosterically affects nicotinic cholinergic receptors therefore, displaying a dual mechanism in treatment (388). This is beneficial for AD treatment as it can affect other neurotransmitter systems and can improve cognitive dysfunction, major depression and bipolar disorder (389).

Due to the clinical effects of GAL, total synthesis of the compound has been of great interest for pharmaceutical companies (390, 391). Large scale production by a synthetic route is economically beneficial and more efficient for worldwide distribution than extraction from the botanical source (390, 391). For these reasons, several total synthetic routes of GAL have been developed (392–394).

5.4.2. Interactions of GAL

A X-ray crystal structure of GAL bound to AChE has identified four key sites for chemical modification based on structure-activity studies: (a) the hydroxyl site, (b) the cyclohexene ring, (c) the tertiary amine site and (d) the methoxy site **(Figure 5.5)** (385, 395, 396).
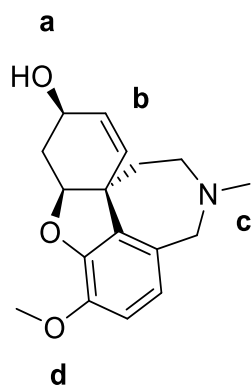
Figure 5.5. The four key sites of GAL that are known to interact with AChE. (a) the hydroxyl site, (b) the cyclohexene ring, (c) the tertiary amine site, and (d) the methoxy site (385, 395, 396).

In the structure, there are two classic hydrogen bonds. The first hydrogen bond is a strong interaction between the hydroxyl oxygen of the inhibitor and Glu199 **(Figure 5.6)** (385, 395, 396). The hydroxyl moiety also interacts with two additional water molecules that are conserved in structures of AChE. The second hydrogen bond is between Ser200 and the O-methyl group of the inhibitor. The O-methyl group of GAL is also involved in occupying the acyl-binding pocket by non-bonding s-p interactions with residues Phe288 and Phe290. A non-classical hydrogen bond interaction has been identified between the N-methyl group of GAL, Asp72 and with a conserved water molecule. The residue Asp72 commonly traps cationic ligands in the active site gorge and contributes to the tight binding of GAL to AChE (385, 395, 396). The final identified interaction between GAL and AChE involves the indole ring of Trp84 that stacks against the double bond of the cyclohexene ring with favourable π-π interactions (385, 395, 396).
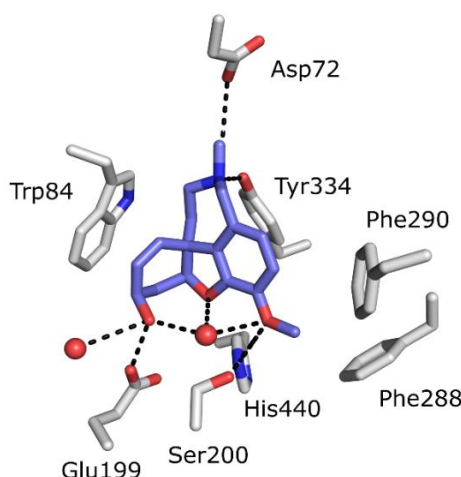
Figure 5.6. The X-ray crystal structure of GAL bound to AChE, showing the key interactions between GAL and AChE (385, 395, 396).

5.4.3. Galanthamine Derivatives

Pharmaceutical companies and research groups have focussed on the synthesis of GAL derivatives to improve the inhibition against AChE (341, 397, 398). Previous work has shown a large variation in IC_{50} depending on the site modified **(Figure 5.7) (Table 5.1)** (385). Certain modified sites of GAL (hydroxyl site) have a major effect on AChE inhibition that can either cause an increase or decrease in inhibition **(Table 5.1)** (385). The large variation in IC_{50} between the derivatives suggests certain interacting sites of GAL are more sensitive to change compared to other sites.
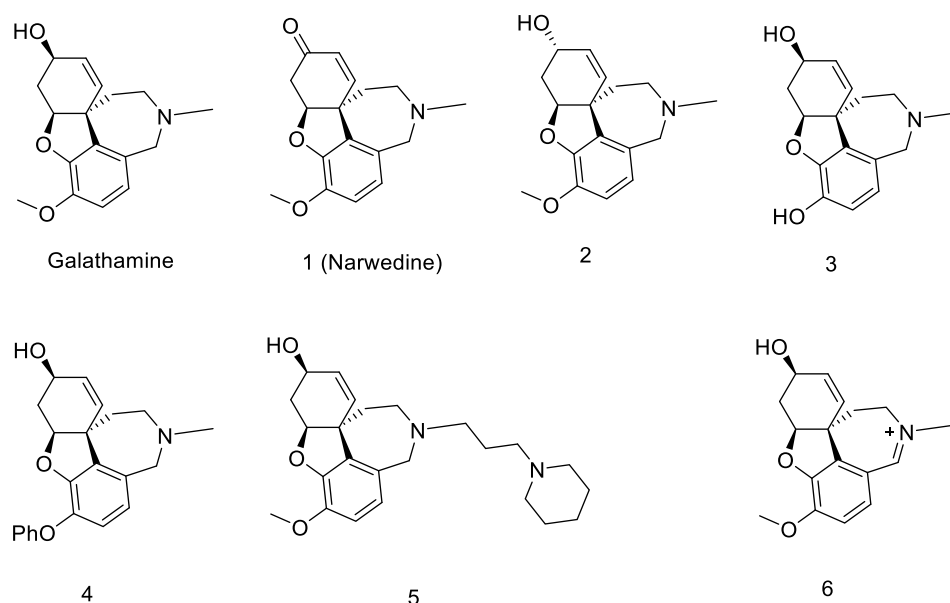
Figure 5.7. Selection of GAL derivatives that have been synthesised and tested against AChE (385)

Table 5.1. Calculated IC_{50} for the GAL derivatives and the sites of GAL with the affect on inhibition upon modification (385).

Compounds	IC_{50} (µM)	Site of Modification	Generally associated with:
Galanthamine	0.35	-	-
1 (Narwedine)	30	Hydroxyl group	Decreased inhibition (higher IC_{50})
2	30	Hydroxyl group	Decreased inhibition
3	0.03	Methoxy group	Increased inhibition (lower IC_{50})
4	3.15	Methoxy group	Decreased Inhibition
5	0.03	Tertiary amine Site	Increased inhibition
6	0.14	Tertiary amine Site	Increased inhibition

5.5. Research Article

5.5.1. Preface

Currently, there are three drugs (rivastigmine, donepezil, and GAL) used to treat symptoms of AD by inhibiting the enzyme, AChE. A major focus in synthetic chemistry is to improve the effectiveness of existing drugs by changing functional groups. Well-established synthetic routes and an increasing market for GAL has resulted in numerous GAL derivatives being synthesised to increase the inhibition towards AChE. The following research article describes the synthesis and testing of new GAL derivatives. The inhibition towards AChE was measured by enzymatic assays. Molecular docking simulations were also performed to understand and further support the inhibition results.

5.5.2. Published Research Article: The Synthesis of Certain Derivatives and Analogues of (−)- and (+)-Galanthamine and an Assessment of their Capacities to Inhibit Acetylcholine Esterase

The Synthesis of Certain Derivatives and Analogues of (−)- and (+)-Galanthamine and an Assessment of their Capacities to Inhibit Acetylcholine Esterase

Joshua N. Buckler, Ehab S. Taher, Nicolas J. Fraser, Anthony C. Willis, Paul D. Carr, Colin J. Jackson, and Martin G. Banwell*

Research School of Chemistry, Institute of Advanced Studies, The Australian National University, Canberra, Australian Capital Territory 2601, Australia

S *Supporting Information*
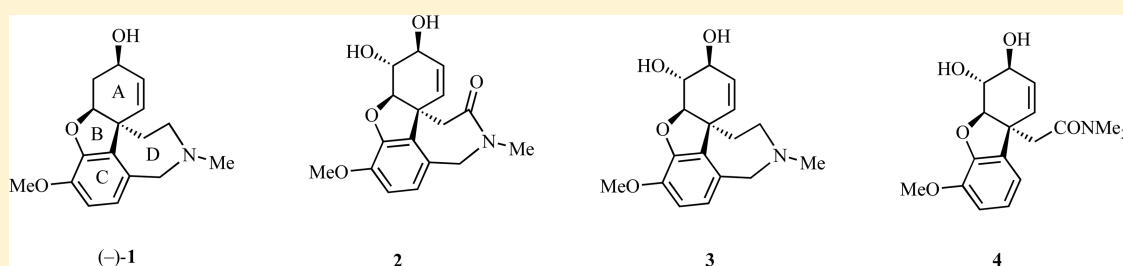
ABSTRACT: Syntheses of certain di- and mono-oxygenated derivatives (e.g., **2** and **3**, respectively) and analogues (e.g., **4**, a D-ring monoseco-analogue of **2**) of both the (−)- and (+)-enantiomeric forms of the alkaloid galanthamine [(−)-**1**] are reported. All have been assessed for their capacities to inhibit acetylcholine esterase but, in contrast to the predictions from docking studies, none bind strongly to this enzyme.

INTRODUCTION

The alkaloid (−)-galanthamine [(−)-**1**] (Figure 1) has been isolated from a range of plant sources and is currently used in
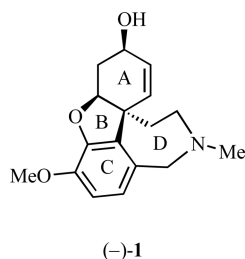
Figure 1. Structure of the alkaloid (−)-galanthamine.

the clinic for the symptomatic treatment of mild to moderate forms of Alzheimer's disease.[1] It exerts its beneficial effects by crossing the blood-brain barrier and then, in part at least, inhibiting acetylcholine esterase (AChE). It also acts as an allosteric modulator of the nicotinic acetylcholine receptor.[1,2] The non-natural enantiomer of compound **1**, namely *ent*-**1** or (+)-galanthamine, has also been shown to accumulate in brain tissue but does so through nonspecific binding.[3]

Currently, (−)-galanthamine is produced industrially by extraction from various plants sources, most notably the red spider lily (*Lycoris radia*), the wild daffodil (*Narcissus pseudonarcissus*), the summer snowflake (*Leucojum aestivum*), and the Caucasian snowdrop (*Galanthus woronowii*).[1] However,

both the increasing demands for the natural product and the erosion of habitat of at least some of the producing plants has prompted investigations into other methods for obtaining it or for identifying analogues with improved efficacy. As part of such efforts, galanthamine has been the subject of a significant number of total synthesis studies with the first of these being reported by Barton and Kirby[4] and involving mimicking of the proposed biogenesis. Substantial refinements of this process have been reported in the interim,[5] and one of these has formed the basis of a pilot-plant scale synthesis of the alkaloid[5a] although it is not clear if this contributes significantly to the commercial production of the alkaloid. Magnus and co-workers have described[6] a related and highly effective approach. Intramolecular Heck reactions have provided another means for assembling the tetracyclic framework of galanthamine,[7] including those accomplished in an enantioselective manner, while various chirons (corresponding to the A-ring) have been employed for the assembly in either enantiomeric form of the alkaloid.[8,9a] New routes to (−)-galanthamine continue to be reported,[1e] including approaches from our group.[9]

The identification and biological evaluation of analogues of galanthamine has been another focus of significant activity[10] that is now greatly assisted by data derived from high-resolution X-ray analysis of the alkaloid bound to the active site of acetylcholine esterase.[11] Biomimetic diversity-oriented syn-
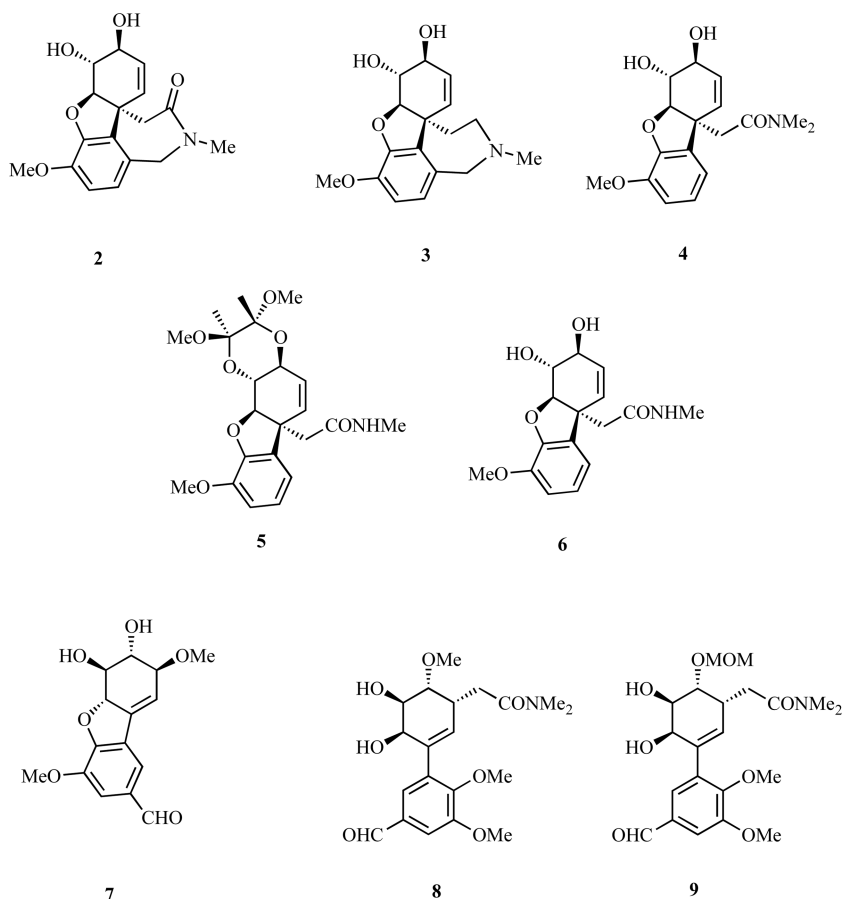
Figure 2. Galanthamine derivatives/analogues **2**−**9** targeted for synthesis.

thesis, sophisticated QSAR analyses, multicomponent-coupling chemistries, as well as more conventional studies have all revealed active compounds.[12] The screening of natural products for relevant activities remains an ongoing area of investigation and has also resulted in the isolation of new inhibitors of the title enzyme.[13]

As part of an ongoing program to establish concise routes to the tetracyclic framework of galanthamine,[1e,9] we now report chemoenzymatic syntheses of two oxygenated derivatives, **2** and **3**, as well as analogues **4**−**9** of the natural or (−)-form of the alkaloid (Figure 2).[14]

Syntheses of the enantiomers of the first five of these derivatives/analogues, namely compounds ent-**2** to ent-**6**, are also described. As detailed below, and contrary to the predictions arising from molecular docking studies, none of these is an effective inhibitor of acetylcholine esterase.

Our motivation for undertaking the studies described herein was that the introduction of additional functionalities within the galanthamine framework could increase water solubility and/or provide the means for conjugating the drug with various groups including peptide fragments[10f] or other entities (e.g. memantine)[10d] and thus allow for the development of a multi-targeted therapeutic approach.[10d] In principle, then, the attachment of such moieties could provide new inhibitors (or even just prodrug forms[10e] of galanthamine) that are superior to existing therapies. As revealed below, the nature of our synthetic strategy is such that additionally oxygenated forms of the galanthamine A-ring were likely to be the most readily accessible. Accordingly, and because this region of alkaloid has

not been "explored" previously, such derivatives became a major focus of the work detailed below.

■ RESULTS AND DISCUSSION

Chemical Synthesis Studies. The reaction sequence used to assemble ABC-ring substructures of (−)-galanthamine is shown in Scheme 1 and employs the readily available L-tartaric acid (**10**) as starting material. Thus, as specified in a recent publication,[15] compound **10** can be converted over 11 steps, involving reduction, Grignard addition, and ring-closing metatheses as key transformations, into the 1,2-diacetal annulated bromocyclohexene **11**. Coupling of compound **11** with the readily available[16] aryl boronic acid ester **12** proceeded under conventional conditions to give the anticipated arylated cyclohexene **13** (68%). Following protocols established during the course of our syntheses of the ribisins,[17] this last compound was engaged in an intramolecular Mitsunobu reaction using triphenylphosphine in conjunction with di-iso-propyl azodicarboxylate (DIAD) wherein the phenolic hydroxyl group served as the internal nucleophile and so affording the acid-sensitive isobenzofuran **14** (33%). As demonstrated through work in the enantiomeric series (see below), if appropriate account is taken of this acid-sensitivity, then the Mitsunobu reaction can be a high yielding one. Despite concerns about the potential for competing isomerization of compound **14** to its more conjugated (fully aromatic) counterpart, upon subjecting it to conditions previously employed for effecting the Eschenmoser−Claisen rearrangement of allylic alcohols,[18] amide **15** was produced in 84% yield. The structure of this compound followed not only from the derived NMR, IR, and mass spectral

104

Scheme 1. Synthesis of Key Acid 17 Embodying the ABC-Ring Substructure and Associated Quaternary Carbon of (−)-Galanthamine
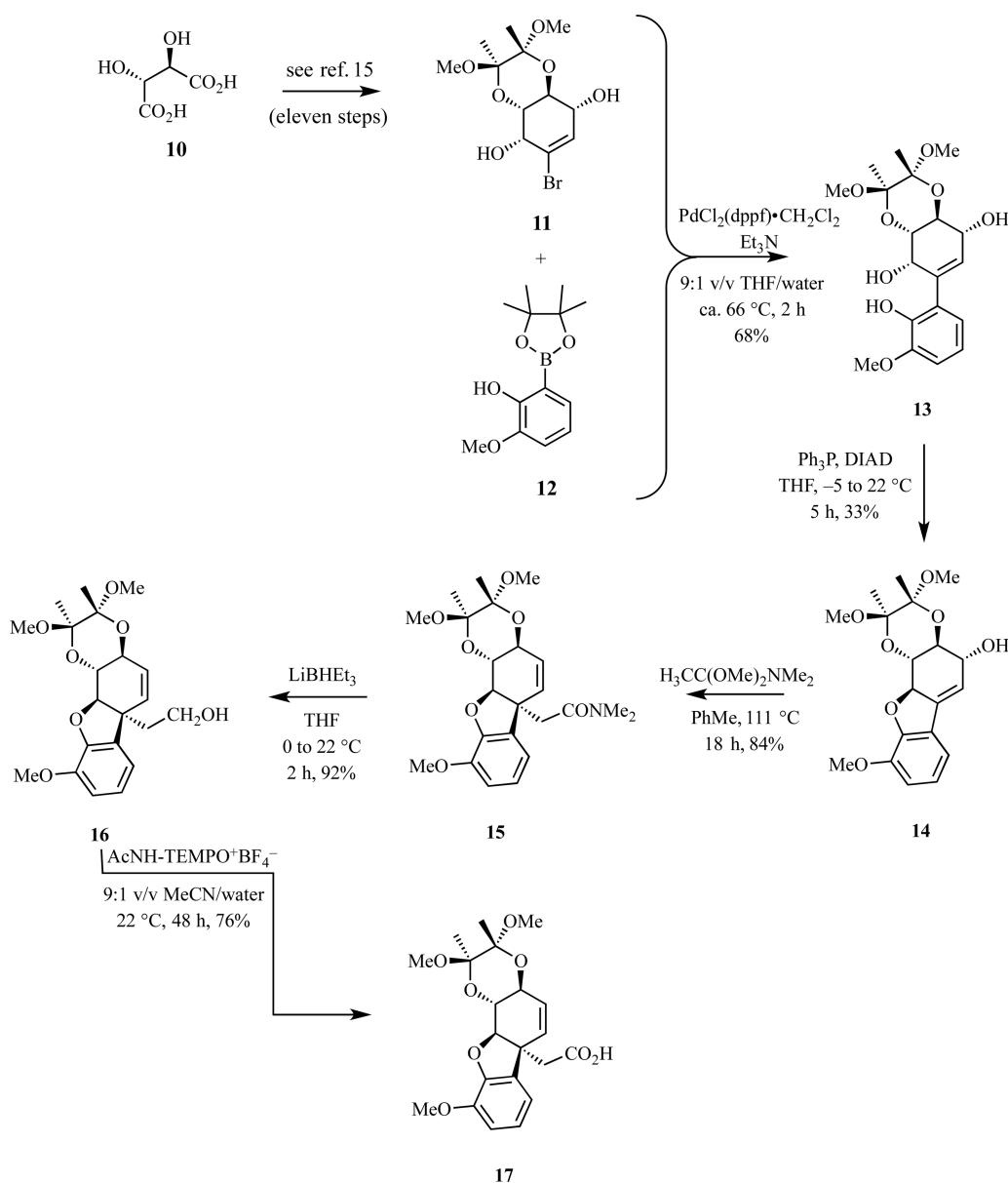
data but also from a single-crystal X-ray analysis of its enantiomer (see below). Compound **15** embodies both the targeted ABC-ring substructure of (−)-galanthamine and the associated quaternary carbon. In anticipation of installing the final D-ring of the title alkaloid, the amide residue within compound **15** was reduced using LiBHEt₃ and 1° alcohol **16** thereby obtained in 92% yield. Oxidation of compound **16** under conditions defined by Bobbitt and Bailey[19] then gave acid **17** (76%), which represents the key precursor to the targeted galanthamine derivatives/analogues **2**−**6**.

The straightforward manipulations of acid **17** leading to the targeted (−)-galanthamine derivatives/analogues **2**, **3**, **5**, and **6** are shown in Scheme 2. Thus, coupling of compound **17** and methylamine, using 1,1′-carbonyldiimidazole (CDI) for activation of the acid, afforded amide **5** (79%), and on exposure of the latter to modified Pictet−Spengler conditions[7c,20] involving paraformaldehyde in the presence of trifluoroacetic acid (TFA), tetracyclic lactam **2** (47%) was formed as a result of

concomitant cleavage of the Ley acetal moiety. Upon treatment with sodium bis(2-methoxyethoxy)aluminum dihydride, compound **2** was reduced to azepine **3** (44%), and hydrolysis of the 1,2-diacetal residue within compound **5** using aqueous trifluoroacetic acid (TFA) afforded diol **6** (66%). Compound **6** can be considered as a hybrid of the title alkaloid and the neurologically active natural product ribisin D.[17,21]

The synthesis of compound **4**, a monoseco derivative of 5R-hydroxy-(−)-galanthamine (**3**), simply involved (Scheme 3) acid-catalyzed hydrolysis of precursor **15** under conventional conditions. This reaction proceeded in 88% yield.

All of the spectral data obtained on targeted compounds **2**−**6** were in complete accordance with the assigned structures but final confirmation of that of the first (i.e., **2**) followed from a single-crystal X-ray analysis of its enantiomer (see below).

In an effort to establish a more meaningful SAR profile for the above-mentioned analogues, the enantiomerically related derivatives/analogues *ent*-**2** to *ent*-**6** (Figure 3) were sought.

105

Scheme 2. Conversion of Acid 17 into the (−)-Galanthamine Derivatives/Analogues 2, 3, 5, and 6
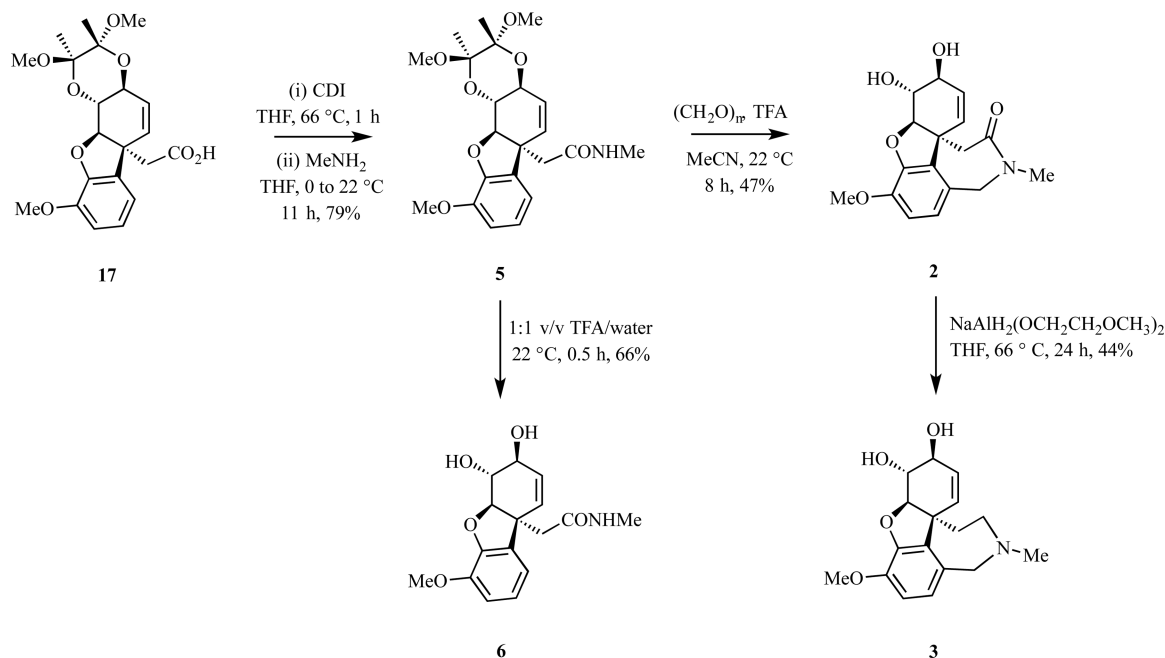
Scheme 3. Hydrolysis of Biacetal 15 Leading to Compound 4

enzymatically derived chiron that is readily available in the required enantiomeric form (but less so in the opposite one required to prepare the derivatives/analogues just described).

The synthesis of the enantiomeric series of compounds started, as shown in Scheme 4, with the enzymatically derived and enantiomerically pure *cis*-1,2-dihydrocatechol **18**[22] that was converted into the corresponding and well-known[23] acetonide **19** under standard conditions. Immediate treatment of the last compound with *m*-chloroperbenzoic acid (*m*-CPBA) then afforded, in a completely regio- and stereo-selective fashion, epoxide **20**[23] (93% from **18**). Reaction of compound **20** with a large excess of *p*-methoxybenzylalcohol (*p*-MBnOH) in the presence of boron trifluoride etherate (BF₃·Et₂O) resulted in selective nucleophilic opening of the epoxide ring at the allylic

Although the routes defined above could simply be adapted for this purpose by starting with D- rather than L-taratric acid, we were able to establish a shorter pathway by starting with an

ent-**2**   *ent*-**3**   *ent*-**4**

ent-**5**   *ent*-**6**
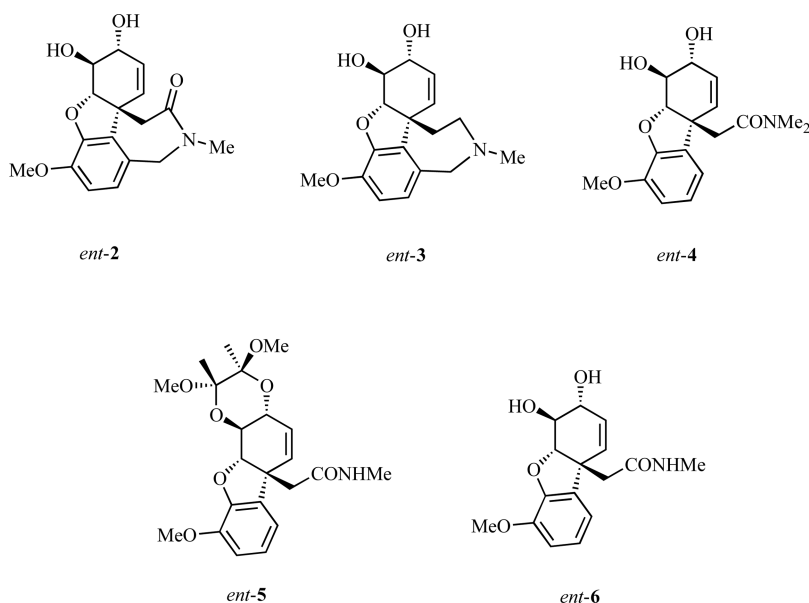
Figure 3. Enantiomeric series of derivatives/analogues.

106

Scheme 4. Synthesis of Bromoconduritol *ent*-11 from the Enzymatically-Derived and Homochiral *cis*-1,2-Dihydrocatechol
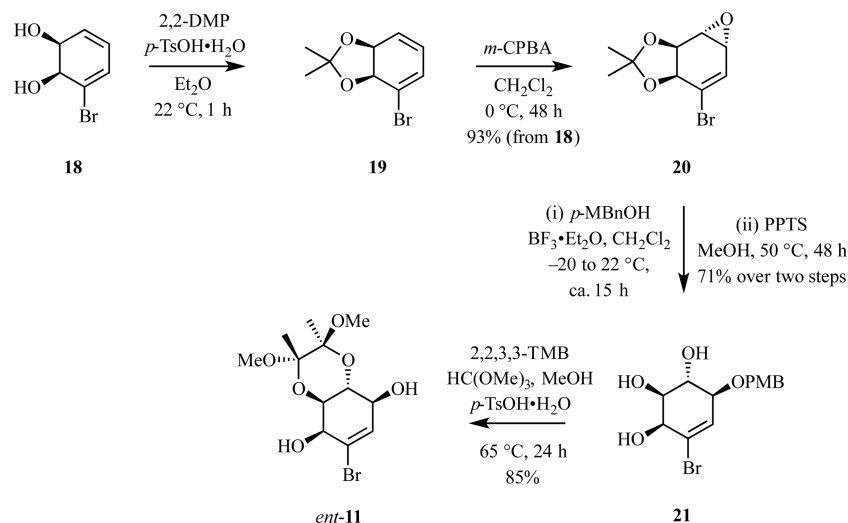
Scheme 5. Synthesis of Key Acid *ent*-17 from Bromoconduritol *ent*-11
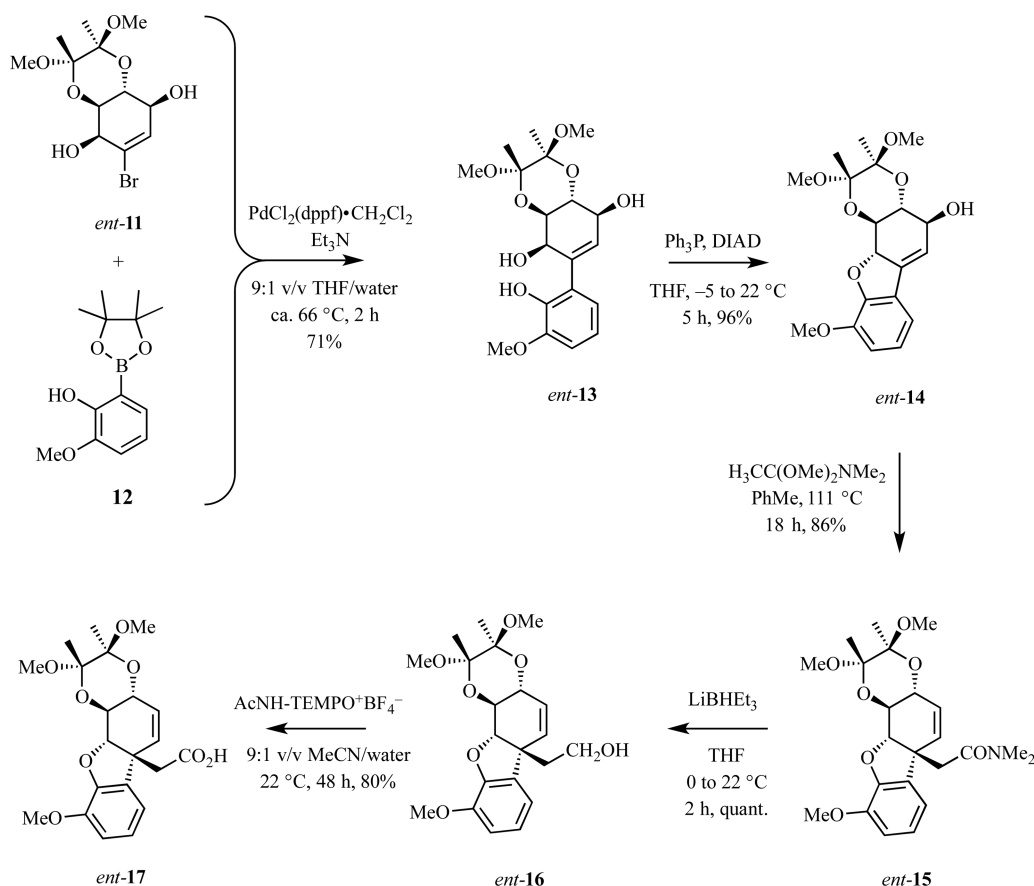
carbon,[24] but the protected bromoconduritol thus formed was not isolated. Rather, it was simply allowed to react with methanol in the presence of pyridinium tosylate (PPTS) and thereby affording triol **21** (71%). Treatment of a methanolic solution of this last compound with 2,2,3,3-tetramethoxybutane (2,2,3,3-TMB) in the presence of *p*-toluenesulfonic acid monohydrate (*p*-TsOH·H$_2$O) then resulted in the selective formation of the "Ley"-acetal[25] *ent*-**11** (85%) and so establishing an "enantiomeric overlap" with the synthetic sequence leading to the original sets of analogues.

During the conversion, **21** → *ent*-**11** acetal formation is presumed to take place prior to cleavage of the PMB-ether unit because the reverse order of events would lead to an intermediate conduritol embodying two adjacent *trans*-diol moieties and the formation of two isomeric Ley-type acetals would therefore be expected. Indeed, when the tetra-ol derived from hydrolysis of compound **21** was subjected to reaction with 2,2,3,3-TMB in the presence of *p*-TsOH·H$_2$O then an ~1:1 mixture of the two possible bis-acetals is formed.

107

Scheme 6. Synthesis of Aryl Boronate Ester 27
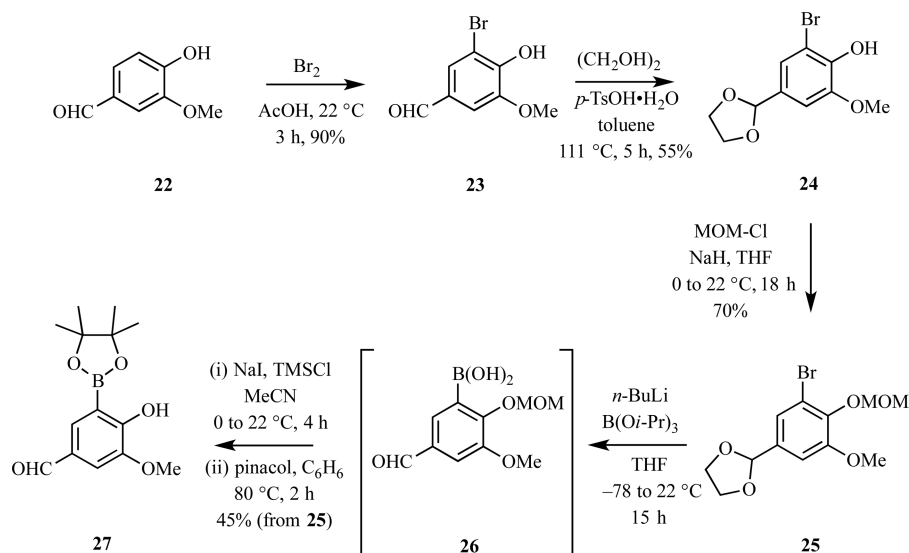
Scheme 7. Synthesis of Analogue 7
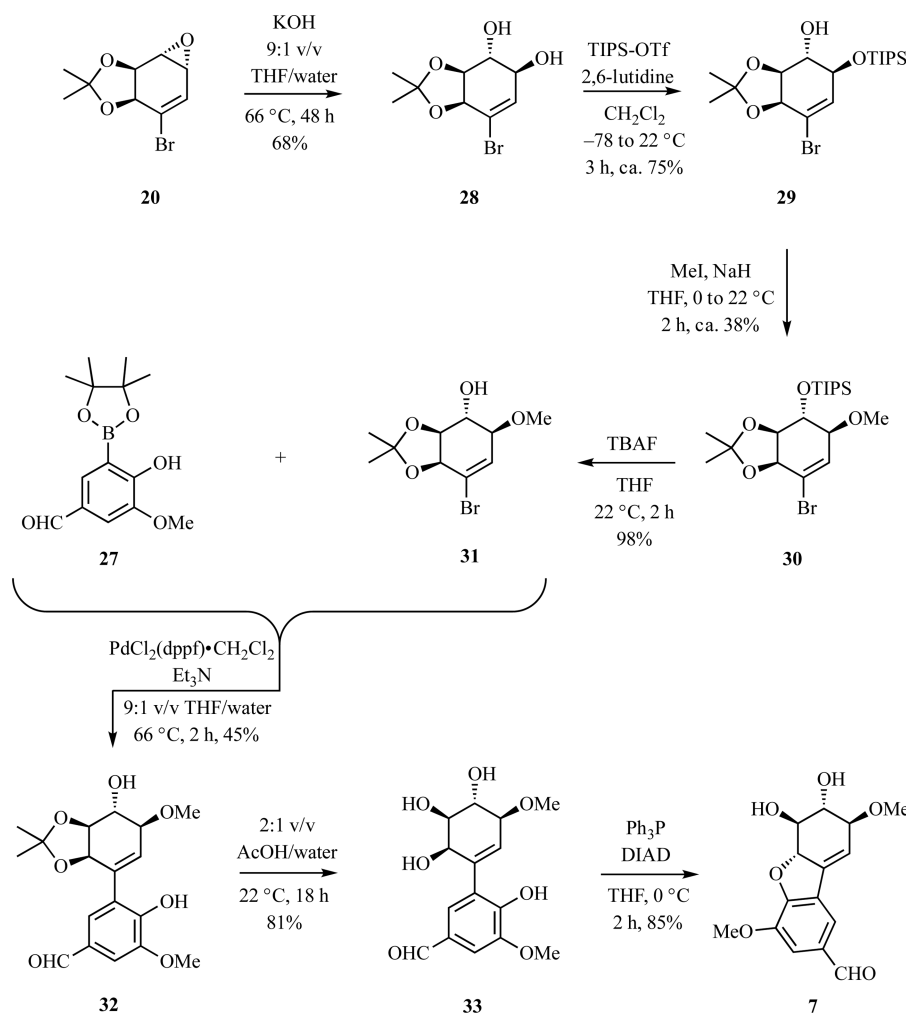
As shown in Scheme 5, and as was the case in the enantiomeric series, compound *ent*-**11** could be engaged in a Suzuki–Miyaura cross-coupling reaction[26] with arylboronic acid ester **12**, thus affording the anticipated product *ent*-**13** (71%) that participated in an intramolecular Mitsunobu reaction to give compound *ent*-**14** (96%). Allylic alcohol *ent*-**14** underwent an Eschenmoser–Claisen rearrangement reaction on thermolysis with the dimethyl acetal of *N,N*-dimethylacetamide in refluxing toluene and thereby affording *N,N*-dimethylamide *ent*-**15** (86%). The structure of this last

108

Scheme 8. Synthesis of Analogue 8
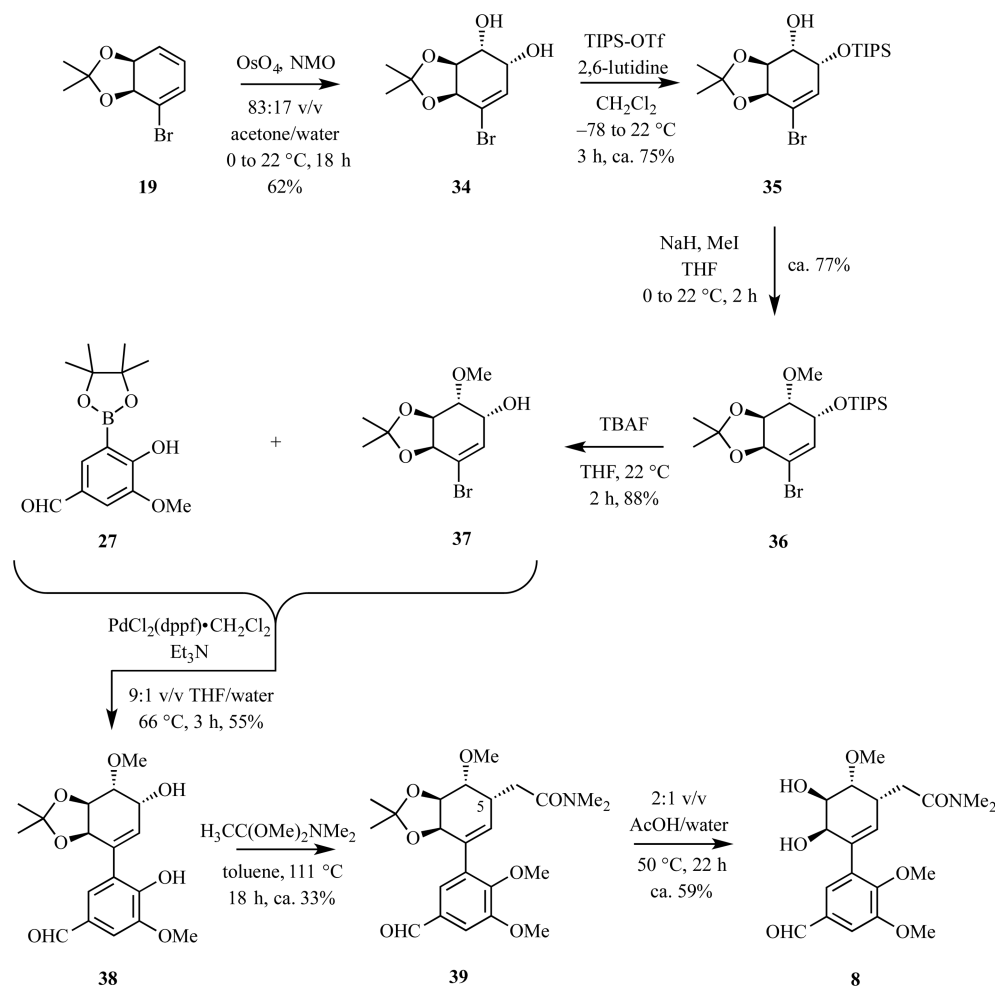
compound was confirmed by a single-crystal X-ray analysis, details of which are provided in the Experimental Section and the Supporting Information (SI). Reduction of compound *ent*-**15** with lithium triethylborohydride and oxidation of the resulting 1° alcohol *ent*-**16** (quant.) using the Bobbitt–Bailey protocol[19] then gave acid *ent*-**17** (80%).

Following precisely the same reaction sequences as detailed above in Schemes 2 and 3, compound *ent*-**17** was converted into target derivatives *ent*-**2** to *ent*-**6** (Figure 3) with the structure of the first of these being confirmed by single-crystal X-ray analysis. The monoseco analogue of compound *ent*-**2**, namely amide *ent*-**4**, was readily obtained in 77% yield by simply treating bis-acetal *ent*-**15** with aqueous trifluoroacetic acid (compare to Scheme 3).

As part of an effort to develop galanthamine derivatives containing additional functionality in the aromatic C-ring, especially ones capable of conjugation with motifs that might bind to the so-called peripheral and anionic binding site of acetylcholine esterase,[27] we sought to exploit the synthetic chemistry detailed above for this purpose. As such, the relevant arylboronate acid ester (representing a synthon for the C-ring) was required and the route used to obtain this is shown in Scheme 6. Thus, vanillin (**22**) was converted, under established conditions, into its bromo-derivative **23**[28] (90%), and the aldehydic residue within the latter then protected as the corresponding ethylene acetal using standard conditions and thus affording compound **24** in 55% yield. The readily derived

MOM-ether **25** (70%) of the last compound was subjected to a metalation/borylation protocol, and the intermediate boronic acid **26** obtained on hydrolytic work up was reacted with sodium iodide/trimethylsilyl chloride (to cleave the MOM ether) and then pinacol and thus affording the hitherto unreported ester **27** (45% from **25**). The structure of this last compound was confirmed by single-crystal X-ray analysis (see Experimental Section and SI for details).

In the opening stages of attempts to exploit ester **27** in the production of galanthamine analogues related to those described above, epoxide **19** (Scheme 7) was subjected to hydrolytic cleavage, thus affording previously reported[29] *trans*-diol **28** (68%). The allylic hydroxyl group of diol **28** was selectively protected through its reaction with tri-*iso*-propylsilyl trifluoromethanesulfonate (TIPS-OTf) in the presence of 2,6-lutidine and thus producing allylic ether **29** (88%).[30] The location of the TIPS group within this product was established using COSY experiments. Upon treating compound **29** with methyl iodide in the presence of sodium hydride, *bis*-ether **30** (49%) rather than the anticipated regioisomer was obtained. Because the illustrated locations of the ether residues within compound **30** (that thwart the application of the anticipated Eschenmoser–Claisen rearrangement) were not appreciated until an X-ray analysis was carried out on a derivative, this was carried forward by first treating it with TBAF and thus affording homoallylic alcohol **31** (98%). Compound **31** was engaged in a Suzuki–Miyaura cross-coupling reaction with ester **27** to give

Scheme 9. Synthesis of Analogue 9
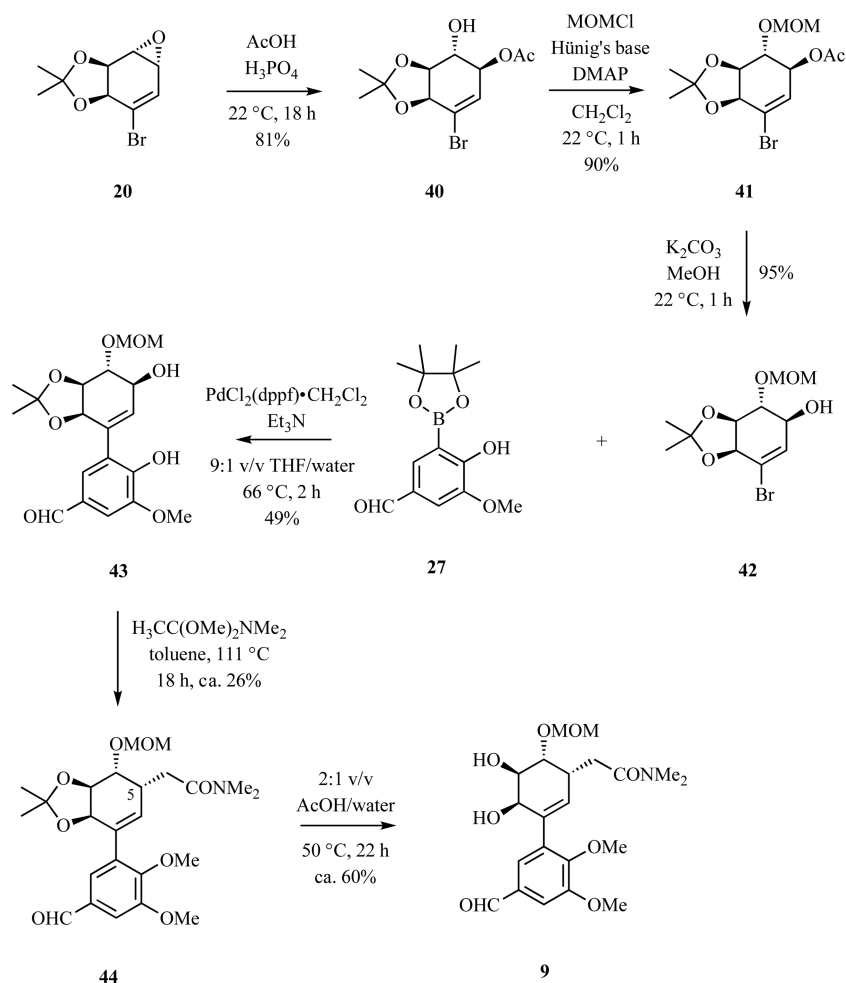
the arylated and crystalline cyclohexene **32** (45%), the illustrated structure of which was established by single-crystal X-ray analysis. Acid-catalyzed hydrolysis of the acetonide residue within the last compound then gave triol **33** (81%). Even though the position of the methoxy group within compound **33** precluded the application of the type of EC rearrangement used earlier, it was subjected to an intramolecular Mitsunobu reaction so as to produce a system, namely compound **7** (85%), that embodies the ABC-ring ensemble associated with (−)-galanthamine.

The synthesis of analogue **8** is shown in Scheme 8 and involved, as the initial step, the regio- and diastereo-selective *cis*-dihydroxylation of diene **19**. Selective silylation of the allylic hydroxyl group within the resulting and previously reported[17b,31] diol **34** (62%) was readily achieved using TIPS-OTf in the presence of 2,6-lutidine, and the product ether **35** (88%) was subjected to *O*-methylation using methyl iodide/base.

Treatment of product *bis*-ether **36** (92%) with TBAF resulted in cleavage of the *O*-TIPS bond and formation of alcohol **37** (88%) that could be cross-coupled with boronate ester **27** under Suzuki−Miyaura conditions and thus affording the anticipated product **38** (55%). Contrary to expectations, however, the allylic alcohol residue within compound **38** failed to engage in an Eschenmoser−Claisen rearrangement upon treatment with *N,N*-dimethylacetamide dimethyl acetal. Rather, the replacement of the associated allylic hydroxyl group with an *N,N*-dimethylacetamide residue took place to give, in a

stereoselective manner and after accompanying *O*-methylation of the phenolic hydroxyl group, amide **39** (40%). This was contaminated by small amounts (7%) of its chromatographically inseparable C5 epimer. "Thwarted E-C rearrangements" of this type have been reported previously[32] and in this instance the process may be driven by electron-rich arene residues facilitating ionization of the intermediate mixed acetal with the ensuing and extensively stabilized cation then undergoing nucleophilic capture by 1-methoxy-*N,N*-dimethylethen-1-amine that is itself generated through thermal cracking of the starting dimethyl acetal. The basic structure of product **39** follows from the observation that the diastereotopic methylene hydrogens of the acetamide side-chain both show vicinal couplings to the adjacent allylic hydrogen. The illustrated configuration of the C5 acetamide residue is assigned on the basis that this would be introduced preferentially during the course of the nucleophilic capture process mentioned above from that face of the intermediate cation opposite to the sterically demanding and nearby acetonide residue. Hydrolysis of this acetonide residue within compound **33** under standard conditions then gave analogue **8** in 69% yield.

The synthesis of final galanthamine analogue **9** followed a very similar route to that used in preparing congener **8**. The reaction sequence involved is shown in Scheme 9. Thus, reaction of epoxide **19** with acetic acid in the presence of a mineral acid gave the previously reported *trans*-diol monoacetate **40**[9a] (81%), the free hydroxyl of which was protected as

110

the corresponding MOM ether using standard protocols and thus affording compound **41**[9a] in 90% yield.

Cleavage of the acetate unit with the last compound could be effected under conventional conditions, and the resulting allylic alcohol **42**[9a] (95%) then cross-coupled with boronate ester **27** in the usual manner to give product **43** (49%). As was the case with congener **32** (Scheme 8), upon subjecting compound **43** to conditions often used to effect the EC rearrangement, this substrate also engaged in both an allylic substitution reaction and *O*-methylation of the phenolic hydroxyl of the precursor. As a consequence, a chromatographically inseparable and 3:1 mixture of amide **44** and its *β*-epimer (35% combined yield) was formed. The salient spectral features of compound **44** resembled those of congener **33**. Hydrolysis of the mixture of acetonide **44** and its C5-epimer under conventional conditions then gave, after chromatographic purification, diol **9** (69%) that, like the other galanthamine derivatives/analogues, was subjected to molecular docking studies and evaluation as a potential inhibitor of acetylcholine esterase. Details are presented in the following section.

Biological Evaluation and Molecular Docking Studies. The above-mentioned derivatives/analogues of (−)- and (+)-galanthamine were each evaluated for their ability to inhibit AChE using a modified method involving addition of DMSO so as to ensure dissolution of these otherwise rather insoluble compounds.[33] The inhibitory effect of DMSO itself on the AChE was taken into account by subtracting a control measurement for obtaining the IC_{50} values of the tested materials. A summary of the inhibition data thus obtained is shown in Table 1. These assays reveal that only one of the

Table 1. Outcomes of Evaluating Galanthamine Derivatives and Analogues as Inhibitors of AChE and Their Calculated Docking Binding Energies (BEs)

entry	compd	IC_{50} (μM)	docking BE (kcal)
1	**2**	>500	−8.9
2	*ent*-**2**	>500	−9.6
3	**3**	420 ± 57	−8.4
4	*ent*-**3**	>500	−9.5
5	**4**	>500	−6.0
6	*ent*-**4**	>500	−9.1
7	**5**	>500	
8	*ent*-**5**	>500	
9	**6**	>500	−6.3
10	*ent*-**6**	>500	−9.2
11	**7**	>500	−7.8
12	**8**	>500	
13	**9**	>500	
14	**1** (+ve control)	0.9 ± 0.2	−10.2

compounds, namely derivative **3**, showed a measurable IC_{50} value (of 420 μM) compared to the positive control (−)-galanthamine [(−)-**1**] which had an IC_{50} value of 0.9 μM. Clearly, then, and regardless of the absolute stereo-chemistries of the systems involved, none of the above-mentioned derivatives/analogues are strong inhibitors of AChE.

Analysis of these inhibition data was undertaken through molecular docking simulations and using the structure of (−)-galanthamine bound to human AChE.[34] Crystallographic studies have revealed that (−)-galanthamine binds into the active site of AChE with the tetrahydroazepine or D-ring assuming a boat-like conformation and the associated *N*-methyl

group in a pseudoequatorial orientation spanning the acyl- and choline-binding sites.[35] Docking simulations, using AUTODOCK, matched the solved structure (see Figure 4A) with the
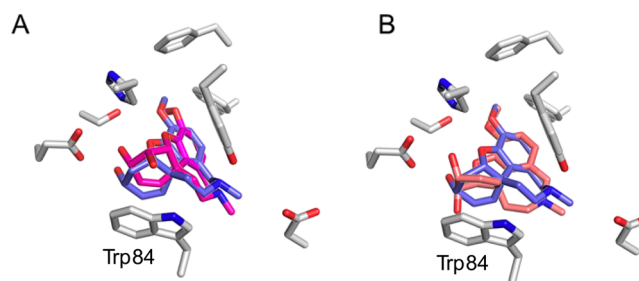
Figure 4. Overlap of (−)-galanthamine (**1**) (blue) and the docked derivatives **3** (**A**, purple) and *ent*-**3** (**B**, peach) in the active site of AChE.

key interactions between AChE and (−)-galanthamine being evident, thus suggesting that docking simulations of this type can provide the correct orientation of binding for the compounds. Surprisingly, with the exception of compounds **5** and *ent*-**5**, **8** and **9**, for which no bound structures could be obtained (in the case of the first two of these compounds, this may be a consequence of their significantly greater size), the derivatives all had significant docking binding energies, albeit weaker than (−)-galanthamine (see right-hand column, Table 1). This highlights some of the known limitations in the prediction of binding affinity by docking programs.[36−38] The docking studies do, however, reveal the likely structural basis for the reduced affinity of these analogues. For example, the configuration of compound **3** matches the orientation of (−)-galanthamine, except that the additional (C5) hydroxyl moiety is positioned toward tryptophan 84 and thus produces a distortion in the shape of the A-ring. This most likely results in destabilization of the $\pi-\pi$ stacking interaction between the indole ring of Trp84 and this ring (A)[39] with this loss of interaction impairing the compound's capacity to inhibit AChE.

The (+)-enantiomer, *ent*-**3**, of compound **3** also has a hydroxyl positioned toward Trp84, again disrupting the stabilizing cyclohexene−indole interactions, although on this occasion it is the same hydroxyl moiety that is present in galanthamine, potentially explaining, in part at least, why (+)-galanthamine derivatives are not potent inhibitors of AChE.

■ CONCLUSIONS

The synthetic chemistry studies detailed above have established that the ABC-ring system of galanthamine is readily obtained through the Suzuki−Miyaura cross-coupling of *o*-hydroxyarylboronates with conduritols incorporating a brominated double-bond and then engaging the products of such processes in an intramolecular Mitsunobu reaction. Furthermore, most of the polyhydroxylated tetrahydrodibenzofurans arising from such a reaction sequence engage in a thermally promoted Eschenmoser−Claisen-type rearrangement reaction upon treatment with *N*,*N*-dimethylacetamide dimethyl acetal in refluxing toluene. The angularly substituted tetrahydrodibenzofurans thus formed, which embody the quaternary carbon center associated with the title alkaloid and represent monoseco analogues of the same, can then be elaborated, using Pictet−Spenger chemistry, to give oxygenated derivatives of galanthamine, certain variants of which have recently been isolated

111

from Chinese medicinal plants.[40] Interestingly, these new natural products were also poor inhibitors of AChE.

The biological evaluation of the galanthamine derivatives and analogues obtained by the pathways described above reveals the finely tuned nature of the interactions of the parent alkaloid with the target enzyme AChE. In particular, structurally "modest" changes to the galanthamine framework, as embodied in the oxygenated derivatives **2**, *ent*-**2**, **3**, and *ent*-**3**, can completely disrupt binding such that the compounds are rendered inactive. These studies have also revealed that the computational prediction of the likely binding affinity of galanthamine analogues to AChE is fraught.

■ EXPERIMENTAL SECTION

General Protocols. Unless otherwise specified, proton (^{1}H) and carbon (^{13}C) NMR spectra were recorded at room temperature in base-filtered CDCl$_3$ on a spectrometer operating at 400 MHz for proton and 100 MHz for carbon nuclei. The signal due to residual CHCl$_3$ appearing at δ_H 7.26 and the central resonance of the CDCl$_3$ "triplet" appearing at δ_C 77.0 were used to reference ^{1}H and ^{13}C NMR spectra, respectively. ^{1}H NMR data are recorded as follows: chemical shift (δ) [multiplicity, coupling constant(s) J (Hz), relative integral] where multiplicity is defined as s = singlet; d = doublet; t = triplet; q = quartet; m = multiplet or combinations of the above. Infrared spectra (ν_{max}) were recorded on an FTIR spectrometer. Samples were analyzed as thin films or finely divided solids. Low-resolution ESI mass spectra were recorded on a single quadrupole mass spectrometer interfaced with a liquid chromatograph, whereas high-resolution measurements were conducted on a time-of-flight instrument. Low- and high-resolution EI mass spectra were recorded on a magnetic-sector machine. Melting points were measured on an automated melting point system and are uncorrected. Analytical thin layer chromatography (TLC) was performed on aluminum-backed 0.2 mm thick silica gel 60 F$_{254}$ plates. Eluted plates were visualized using a 254 nm UV lamp and/or by treatment with a suitable dip followed by heating. These dips included phosphomolybdic acid:ceric sulfate:sulfuric acid (concd):water (37.5 g: 7.5 g: 37.5 g: 720 mL), potassium permanganate:potassium carbonate:5% w/v aq. sodium hydroxide solution:water (3 g: 20 g: 5 mL: 300 mL)), *p*-anisaldehyde or vanillin:sulfuric acid (concd):ethanol (15 g: 2.5 mL: 250 mL). Flash chromatographic separations were carried out following protocols defined by Still et al.[41] with silica gel 60 (40−63 μm) as the stationary phase and using the AR- or HPLC-grade solvents indicated. The melting points of solids purified by such means were recorded directly (i.e., after they had crystallized from the concentrated chromatographic fractions). Starting materials, reagents, drying agents, and other inorganic salts were generally commercially available and used as supplied. THF, methanol, and CH$_2$Cl$_2$ were dried using a solvent purification system that is based upon a technology originally described by Grubbs et al.[42]

Specific Chemical Transformations (2R,3R,4aS,5S,8R,8aS)-6-(2-Hydroxy-3-methoxy-phenyl)-2,3-dimethoxy-2,3-dimethyl-2,3,4a,5,8,8a-hexahydrobenzo[b][1,4]dioxine-5,8-di-ol (13) and Enantiomer ent-13. A magnetically stirred solution of *bis*-acetal **11**[15] (805 mg, 2.37 mmol), ester **12** (720 mg, 2.88 mmol), and triethylamine (3 mL) in THF/water (20 mL of a 9:1 v/v mixture) was subjected to sonication under an atmosphere of nitrogen for 0.5 h. PdCl$_2$dppf·CH$_2$Cl$_2$ (140 mg, 0.191 mmol) was then added, and the ensuing mixture was heated under reflux for 2 h before being cooled and quenched with phosphate buffer (5 mL of a 1 M aqueous solution at pH 7). The mixture thus obtained was cooled to 0 °C, treated with methanol/30% aq hydrogen peroxide (10 mL of a 1:1 v/v mixture), and then allowed to warm to room temperature over 1 h. The resulting mixture was diluted with water (50 mL) and extracted with ethyl acetate (5 × 20 mL). The combined organic phases were washed with brine (2 × 20 mL) before being dried (Na$_2$SO$_4$), filtered, and concentrated under reduced pressure. The ensuing thick orange oil was triturated with diethyl ether (5 × 2 mL), and the resulting yellow

solid was subjected to flash column chromatography (silica, 4:15:1 v/v ethyl acetate/hexane/methanol →12:7:1 v/v ethyl acetate/hexane/methanol gradient elution). Concentration of the appropriate fractions (R_f = 0.3 in 10:9:1 v/v/v ethyl acetate/hexane/methanol) afforded phenol **13** (618 mg, 68%) as a white powder: mp 202−210 °C; $[\alpha]^{20}_D$ −93.8 (*c* 0.1, methanol); ^{1}H NMR (400 MHz, CD$_3$OD) δ 6.89 (dd, J = 7.5 and 1.8 Hz, 1H), 6.83−6.73 (complex m, 2H), 5.76 (d, J = 2.5 Hz, 1H), 4.60 (d, J = 3.7 Hz, 1H), 4.23 (dd, J = 8.0 and 2.5 Hz, 1H), 4.09 (dd, J = 11.0 and 8.0 Hz, 1H), 3.85 (s, 3H), 3.69 (dd, J = 11.0 and 3.7 Hz, 1H), 3.34 (s, 3H), 3.28 (s, 3H), 1.34 (s, 3H), 1.33 (s, 3H) (signals due to hydroxyl protons not observed); ^{13}C NMR (100 MHz, CD$_3$OD) δ 149.0, 145.0, 138.9, 132.9, 128.6, 123.6, 120.3, 112.2, 100.7, 100.3, 71.4, 70.7, 69.7, 69.0, 56.6, 48.3, 48.2, 18.1(5), 18.1(0); IR ν_{max} 3456, 3283, 2941, 1589, 1468, 1218, 1129, 1119, 914, 594 cm^{-1}; MS (ESI, +ve) *m/z* 788 ([(2 M + Na)$^+$, 80%], 405 [(M + Na)$^+$, 100]; HRMS (ESI, +ve) *m/z* (M + Na)$^+$ calcd for C$_{19}$H$_{26}$NaO$_8$ 405.1525, found 405.1526.

Compound *ent*-**13** was prepared in an analogous fashion from *ent*-**11** (2.04 g, 5.34 mmol) to give 1.60 g (71%) of product; $[\alpha]^{20}_D$ +86.4 (*c* 0.1, methanol). All of the other spectral data acquired on this material were identical with those reported above for compound **13**.

(2R,3R,4aS,5R,11aR,11bS)-2,3,10-Trimethoxy-2,3-dimethyl-2,3,4a,5,11a,11b-hexahydrobenzo[b][1,4]dioxino[2,3-g]-benzofuran-5-ol (14) and Enantiomer ent-14. A magnetically stirred solution of phenol **13** (605 mg, 1.58 mmol) and PPh$_3$ (456 mg, 1.74 mmol) in THF (70 mL) maintained at −5 °C was treated dropwise over 0.5 h with di-*iso*-propyl azodicarboxylate (364 μL, 1.74 mmol). The resulting solution was stirred at −5 °C for 4 h and then allowed to warm to room temperature over 1 h before being concentrated under reduced pressure. The orange oil thus obtained was subjected to two successive flash chromatographic separations (silica, hexane → 1:1 v/v ethyl acetate/hexane gradient elution), and concentration of appropriate fractions (R_f = 0.7, 2:1 v/v ethyl acetate/hexane) afforded allylic alcohol **14** (193 mg, 33%) as a white foam; $[\alpha]^{20}_D$ −157 (*c* 0.6, CHCl$_3$); ^{1}H NMR (400 MHz, CDCl$_3$) δ 6.99 (dd, J = 7.5 and 1.2 Hz, 1H), 6.89 (apparent t, J = 7.5 Hz, 1H), 6.82 (dd, J = 8.1 and 1.2 Hz, 1H), 5.81 (m, 1H), 5.14 (dt, J = 9.0 and 3.6 Hz, 1H), 4.67−4.61 (complex m, 1H), 4.04 (dd, J = 10.5 and 9.0 Hz, 1H), 3.93 (dd, J = 10.5 and 7.7 Hz, 1H), 3.88 (s, 3H), 3.34 (s, 6H), 2.34 (d, J = 4.8 Hz, 1H), 1.36(3) (s, 3H), 1.35(9) (s, 3H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 151.6, 145.4, 139.1, 125.7, 122.1, 116.2, 113.6, 113.5, 99.2, 98.9, 84.0, 74.1, 71.3, 70.1, 56.0, 48.4, 48.0, 17.7, 17.6; IR ν_{max} 3486, 2959, 2897, 1614, 1597, 1494, 1445, 1133, 1114, 1098, 1035, 1015, 786 cm^{-1}; MS (ESI, +ve) *m/z* 752 [(2 M + Na)$^+$, 25%], 387 [(M + Na)$^+$, 100%]; HRMS (ESI, +ve) *m/z* (M + Na)$^+$ calcd for C$_{19}$H$_{24}$NaO$_7$ 387.1420, found 387.1423.

Compound *ent*-**14** was prepared in an analogous fashion from *ent*-**13** (345 mg, 0.90 mmol) to give 314 mg (96%) of product; $[\alpha]^{20}_D$ +102 (*c* 1.0, CHCl$_3$). All the other spectral data acquired on this material were identical with those reported above for compound **14**.

N,N-Dimethyl-2-((2R,3R,4aS,6aS,11aR,11bR)-2,3,10-trimethoxy-2,3-dimethyl-2,3,11a,11b-tetrahydrobenzo[b][1,4]dioxino[2,3-g]-benzofuran-6a(4aH)-yl)acetamide (15) and Enantiomer ent-15. A magnetically stirred solution of allylic alcohol **14** (180 mg, 0.494 mmol) in toluene (20 mL) maintained at 22 °C was treated with *N,N*-dimethylacetamide dimethyl acetal (900 μL, 4.9 mmol), and the resulting solution was heated under reflux for 18 h. The cooled reaction mixture was concentrated under reduced pressure, and the residue thus obtained was triturated with diethyl ether. The resulting waxy solid was subjected to flash chromatography (silica, 20:79:1 v/v ethyl acetate/hexane/methanol → 60:39:1 v/v ethyl acetate/hexane/methanol gradient elution), and concentration of appropriate fractions (R_f = 0.2 in 2:1 v/v ethyl acetate/hexane) afforded amide **15** (178 mg 84%) as a white foam; $[\alpha]^{20}_D$ −89.9 (*c* 1.0, CHCl$_3$); ^{1}H NMR (400 MHz, CDCl$_3$) δ 6.83 (m, 1H), 6.75 (d, J = 7.8 Hz, 1H), 6.72 (d, J = 7.6 Hz, 1H), 6.02 (dd, J = 10.0 and 2.7 Hz, 1H), 5.74 (dd, J = 10.0 and 2.0 Hz, 1H), 5.04 (d, J = 9.6 Hz, 1H), 4.55 (dt, J = 9.6 and 2.0 Hz, 1H), 3.85 (s, 3H), 3.76 (t, J = 9.6 Hz, 1H), 3.30 (s, 3H), 3.16 (s, 3H), 2.89 (s, 3H), 2.78 (s, 3H), 2.71 (d, J = 15.8 Hz, 1H), 2.61 (d, J = 15.8 Hz, 1H), 1.34 (s, 3H), 1.28 (s, 3H); ^{13}C NMR (100 MHz, CDCl$_3$) δ

112

169.0, 145.6(4), 145.5(6), 134.2, 128.9, 126.9, 121.5, 115.3, 112.0, 99.8, 87.1, 70.5, 65.9, 56.1, 51.0, 48.1, 47.4, 42.5, 37.3, 35.3, 17.7(4), 17.7(0) (one signal obscured or overlapping); IR ν_{max} 2954, 2889, 1651, 1491, 1459, 1380, 1276, 1115, 1061, 1000, 954, 739 cm^{-1}; MS (ESI, +ve) m/z 456 [(M + Na)$^+$, 100%], 201 (35); HRMS (ESI, +ve) m/z (M + Na)$^+$ calcd for $C_{23}H_{31}NNaO_7$ 456.1998, found 456.2000.

Compound *ent*-**15** was prepared in an analogous fashion from *ent*-**14** (234 mg, 0.64 mmol) to give 238 mg (86%) of product as a white foam. A small sample was crystallized (diethyl ether/methanol/hexane) to give a white, crystalline solid; mp 160−165 °C (dec); $[\alpha]^{20}_D$ +87.1 (c 1.0, CHCl$_3$). All of the other spectral data acquired on this material were identical with those reported above for compound **15**.

2-((2R,3R,4aS,6aS,11aR,11bR)-2,3,10-Trimethoxy-2,3-dimethyl-2,3,11a,11b-tetrahydrobenzo[b][1,4]dioxino[2,3-g]benzofuran-6a-(4aH)-yl)ethan-1-ol (16) and Enantiomer ent-16. A magnetically stirred solution of amide **15** (146 mg, 0.337 mmol) in THF (17 mL) maintained at 0 °C was treated dropwise with lithium triethylborohydride (1.7 mL of a 1 M solution in THF, 1.7 mmol). The resulting mixture was warmed to room temperature over 2 h, recooled to 0 °C, quenched with methanol (2 mL), and then treated with silica (200 mg of flash chromatographic-grade material) before being subjected to flash chromatography (silica, 1:4 v/v ethyl acetate/hexane → 1:1 v/v ethyl acetate/hexane gradient elution). Concentration of relevant fractions (R_f = 0.6, 2:1 v/v ethyl acetate/hexane) afforded compound **16** (122 mg, 92%) as a white foam; $[\alpha]^{20}_D$ −131.9 (c 1.8, CHCl$_3$); ^{1}H NMR (400 MHz, CDCl$_3$) δ 6.84 (apparent t, J = 7.8 Hz, 1H), 6.74 (m, 1H), 6.69 (dd, J = 7.5 and 1.2 Hz, 1H), 5.91 (broad d, J = 10.0 Hz, 1H), 5.70 (broad d, J = 10.0 Hz, 1H), 4.89 (d, J = 9.6 Hz, 1H), 4.31 (ddd, J = 9.6, 2.7, and 1.5 Hz, 1H), 3.84 (s, 3H), 3.72 (t, J = 9.6 Hz, 1H), 3.67−3.57 (complex m, 2H), 3.28 (s, 3H), 3.17 (s, 3H), 1.96 (dt, J = 13.7 and 6.7 Hz, 1H), 1.85 (dt, J = 13.7 and 6.4 Hz, 1H), 1.40 (triplet, J = 5.2 Hz, 1H), 1.34 (s, 3H), 1.28 (s, 3H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 145.7, 145.4, 133.6, 130.1, 126.4, 121.8, 114.9, 112.0, 99.9, 99.7, 86.9, 70.8, 65.4, 59.3, 56.1, 51.7, 48.0, 47.5, 43.5, 17.7, 17.6; IR ν_{max} 3508, 2948, 2837, 1619, 1588, 1491, 1459, 1281, 1130, 1116, 1036, 753, 736 cm^{-1}; MS (ESI, +ve) m/z 415 [(M + Na)$^+$, 100%]; HRMS (ESI, +ve) m/z (M + Na)$^+$ calcd for $C_{21}H_{28}NaO_7$ 415.1733, found 415.1733.

Compound *ent*-**16** was prepared in an analogous fashion from *ent*-**15** (177 mg, 0.408 mmol) to give 160 mg (quantitative) of product; $[\alpha]^{20}_D$ +108 (c 0.9, CHCl$_3$). All of the other spectral data acquired on this material were identical with those reported above for compound **16**.

2-((2R,3R,4aS,6aS,11aR,11bR)-2,3,10-Trimethoxy-2,3-dimethyl-2,3,11a,11b-tetrahydrobenzo[b][1,4]dioxino[2,3-g]benzofuran-6a-(4aH)-yl)acetic Acid (17) and Enantiomer ent-17. A magnetically stirred solution of alcohol **16** (122 mg, 0.310 mmol) in acetonitrile/water (9:1 v/v, 2.3 mL) maintained at room temperature was treated in one portion with 4-(acetylamino)-2,2,6,6-tetramethyl-1-oxo-piperidinium tetrafluoroborate (280 mg, 0.93 mmol). The resulting dark-brown solution was stirred at 22 °C for 48 h (by which time it was a pale-yellow color) and then poured into water (3 mL) and extracted with diethyl ether (5 × 4 mL). The combined organic extracts were washed with HCl (2 × 1 mL of a 1 M aqueous solution) and brine (2 × 5 mL) before being dried (Na$_2$SO$_4$), filtered, and concentrated under reduced pressure. The residue thus obtained was subjected to flash chromatography (silica, CH$_2$Cl$_2$ → 1:19 v/v methanol/CH$_2$Cl$_2$ gradient elution) and concentration of appropriate fractions (R_f = 0.2 in 1:19 v/v methanol/CH$_2$Cl$_2$) afforded acid **17** (96 mg, 76%) as a clear, colorless gum; $[\alpha]^{20}_D$ −88.5 (c 0.6, CHCl$_3$); ^{1}H NMR (400 MHz, CDCl$_3$) δ 6.85 (m, 1H), 6.76 (d, J = 7.4 Hz, 1H), 6.74 (d, J = 7.4 Hz, 1H), 5.96 (dd, J = 10.0 and 2.6 Hz, 1H), 5.77 (d, J = 10.0 Hz, 1H), 5.17 (d, J = 9.7 Hz, 1H), 4.40 (d, J = 9.7 Hz, 1H), 3.85 (s, 3H), 3.75 (t, J = 9.7 Hz, 1H), 3.29 (s, 3H), 3.17 (s, 3H), 2.76 (d, J = 15.6 Hz, 1H), 2.63 (d, J = 15.6 Hz, 1H), 1.36 (s, 3H), 1.28 (s, 3H) (signal due to carboxylic acid group proton not observed); ^{13}C NMR (100 MHz, CDCl$_3$) δ 174.8, 145.6, 132.9, 128.1, 127.9, 121.9, 115.0, 112.3, 99.9, 99.8, 86.0, 70.5, 65.5, 56.1, 50.9, 48.0, 47.5, 44.0, 17.7 (signals due to two carbons obscured or overlapping); IR ν_{max} 2950, 1710,

1712, 1619, 1491, 1459, 1284, 1128, 1116, 1034, 960, 732 cm^{-1}; MS (ESI, +ve) m/z 429 [(M + Na)$^+$, 100%]; HRMS (ESI, +ve) m/z (M + Na)$^+$ calcd for $C_{21}H_{26}NaO_8$ 429.1525, found 429.1524.

Compound *ent*-**17** was prepared in an analogous fashion from *ent*-**16** (150 mg, 0.382 mmol) to give 115 mg (80%) of product; $[\alpha]^{20}_D$ +95.7 (c 1.2, CHCl$_3$). All of the other spectral data acquired on this material were identical with those reported above for compound **17**.

N-Methyl-2-((2R,3R,4aS,6aS,11aR,11bR)-2,3,10-trimethoxy-2,3-dimethyl-2,3,-11a,11b-tetrahydrobenzo[b][1,4]dioxino[2,3-g]-benzofuran-6a(4aH)-yl)acetamide (5) and Enantiomer ent-5. A magnetically stirred solution of acid **17** (93 mg, 0.23 mmol) in THF (12 mL) maintained at room temperature was treated with 1,1′-carbonyldiimidazole (49 mg, 0.30 mmol). The resulting solution was heated under reflux for 1 h before being cooled to room temperature and then placed in an ice bath at 0 °C. Methylamine (700 μL of a 2 M solution in THF, 1.4 mmol) was then added dropwise, and the ensuing solution was maintained at 0 °C for 3 h before being warmed to 22 °C and stirred at this temperature for another 8 h. After this time, the reaction mixture was diluted with ethyl acetate (40 mL) and washed with NH$_4$Cl (3 × 15 mL of a saturated aqueous solution). The combined aqueous phases were extracted with ethyl acetate (3 × 10 mL), and the combined organic phases were washed with brine (2 × 10 mL) before being dried (Na$_2$SO$_4$), filtered, and concentrated under reduced pressure. The residue thus obtained was subjected to flash chromatography (silica, 4:15:1 v/v ethyl acetate/hexane/methanol → 8:11:1 v/v ethyl acetate/hexane/methanol gradient elution) and concentration of appropriate fractions (R_f = 0.3 in 10:9:1 v/v ethyl acetate/hexane/methanol) gave amide **5** (76 mg, 79%) as a clear, colorless oil; $[\alpha]^{20}_D$ −102.1 (c 1.2, CHCl$_3$); ^{1}H NMR (400 MHz, CDCl$_3$) δ 6.81 (m, 1H), 6.72 (dd, J = 8.0 and 1.2 Hz, 1H), 6.68 (dd, J = 7.5 and 1.2 Hz, 1H), 6.09 (dd, J = 10.0 and 2.6 Hz, 1H), 5.72 (dd, J = 10.0 and 1.5 Hz, 1H), 5.36 (broad q, J = 4.8 Hz, 1H), 4.97 (d, J = 9.7 Hz, 1H), 4.37 (ddd, J = 9.7, 2.6, and 1.5 Hz, 1H), 3.81 (s, 3H), 3.71 (t, J = 9.7 Hz, 1H), 3.27 (s, 3H), 3.13 (s, 3H), 2.71 (d, J = 4.8 Hz, 3H), 2.51 (d, J = 14.0 Hz, 1H), 2.42 (d, J = 14.0 Hz, 1H), 1.32 (s, 3H), 1.26 (s, 3H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 169.5, 145.5, 133.4, 128.7, 126.9, 121.6, 115.2, 112.1, 99.8, 99.7, 86.9, 70.3, 65.5, 56.0, 51.2, 48.0, 47.4, 46.8, 26.2, 17.7, 17.6 (signal due to one carbon obscured or overlapping); IR ν_{max} 3317, 2952, 2921, 1646, 1548, 1491, 1457, 1377, 1197, 1114, 1034, 999, 955, 735 cm^{-1}; MS (ESI, +ve) m/z 442 [(M + Na)$^+$, 100%]; HRMS (ESI, +ve) m/z (M + Na)$^+$ calcd for $C_{22}H_{29}NNaO_7$ 442.1842, found 442.1842.

Compound *ent*-**5** was prepared in an analogous fashion from *ent*-**17** (115 mg, 0.28 mmol) to give 91 mg (79%) of product; $[\alpha]^{20}_D$ +89.1 (c 1.3, CHCl$_3$). All of the other spectral data acquired on this material were identical with those reported above for compound **5**.

2-((5aR,6R,7S,9aS)-6,7-Dihydroxy-4-methoxy-6,7-dihydrodibenzo[b,d]furan-9a(5aH)-yl)-N-methylacetamide (6) and Enantiomer ent-6. A round-bottomed flask charged with a magnetic stirrer bar and amide **5** (16.7 mg, 0.04 mmol) was treated sequentially with water (250 μL) and trifluoroacetic acid (250 μL) and the ensuing mixture stirred at 22 °C for 0.5 h; then, the volatiles were removed under reduced pressure. The residue thus obtained was subject to flash chromatography (silica, 5:4:1 v/v ethyl acetate/hexane/methanol) and concentration of appropriate fractions (R_f = 0.3 in 8:1:1 v/v ethyl acetate/hexane/methanol) afforded diol **6** (8.1 mg, 66%) as a white foam; $[\alpha]^{20}_D$ −11 (c 0.4, methanol); ^{1}H NMR [400 MHz, (CD$_3$)$_2$CO] δ 7.04 (broad s, 1H), 6.84−6.74 (complex m, 3H), 6.00 (dd, J = 10.1 and 2.3 Hz, 1H), 5.67 (dd, J = 10.1 and 1.5 Hz, 1H), 5.25 (d, J = 8.7 Hz, 1H), 4.75 (broad s, 1H), 4.22 (partially obscured and broad s, 1H), 4.18 (d, J = 8.7 Hz, 1H), 3.82 (s, 3H), 3.40 (t, J = 8.7 Hz, 1H), 2.63 (d, J = 4.6 Hz, 3H), 2.48 (d, J = 13.9 Hz, 1H), 2.41 (d, J = 13.9 Hz, 1H); ^{13}C NMR [100 MHz, (CD$_3$)$_2$CO] δ 170.7, 146.9, 146.2, 135.5, 131.6, 128.5, 122.1, 116.1, 113.3, 90.4, 75.8, 69.9, 56.4, 51.6, 46.8, 25.9; IR ν_{max} 3315, 2943, 1642, 1491, 1272, 1199, 1179, 1132, 1064, 945, 723 cm^{-1}; MS (ESI, +ve) m/z 634 [(2 M + Na)$^+$, 40], 328 [(M + Na)$^+$, 100]; HRMS (ESI, +ve) m/z (M + Na)$^+$ calcd for $C_{16}H_{19}NNaO_5$ 328.1161, found 328.1158.

Compound *ent*-**6** was prepared in an analogous fashion from *ent*-**5** (16.7 mg, 0.04 mmol) to give 11.8 mg (97%) of product; $[\alpha]^{20}_D$ +16.7

113

(4aR,5R,6S,8aS)-5,6-Dihydroxy-3-methoxy-11-methyl-4a,5,11,12-tetrahydro-6H-benzo[2,3]benzofuro[4,3-cd]azepin-10(9H)-one (2) and Enantiomer ent-2. A magnetically stirred solution of amide **5** (56 mg, 0.133 mmol) in acetonitrile (14 mL) was treated sequentially with paraformaldehyde (20 mg, 0.67 mmol) and trifluoroacetic acid (63 μL, 0.82 mmol). The resulting solution was stirred at room temperature for 8 h and then quenched with phosphate buffer (15 mL of a 1 M aqueous solution at pH 7) and extracted with CH_2Cl_2 (3 × 10 mL). The combined organic extracts were washed with brine (1 × 5 mL) and dried (Na_2SO_4) before being filtered and then concentrated under reduced pressure. The residue thus obtained was subjected to flash chromatography (silica, 4:15:1 v/v ethyl acetate/hexane/methanol → 8:1:1 v/v ethyl acetate/hexane/methanol gradient elution) and concentration of the appropriate fractions (R_f = 0.4 in 8:1:1 v/v ethyl acetate/hexane/methanol) afforded lactam **2** (19.7 mg, 47%) as a white, amorphous solid; $[\alpha]^{20}_D$ −139 (c 0.4, methanol). ^{1}H NMR (400 MHz, CDCl$_3$) δ 6.71 (apparent s, 2H), 5.97 (dd, J = 10.2 and 4.7 Hz, 1H), 5.58 (d, J = 10.2 Hz, 1H), 4.78−4.73 (complex m, 1H), 4.67−4.63 (complex m, 1H), 4.53 (d, J = 16.0 Hz, 1H), 4.32 (d, J = 16.0 Hz, 1H), 4.12 (m, 1H), 3.85 (s, 3H), 3.18 (d, J = 14.2 Hz, 1H), 3.07 (d, J = 14.2 Hz, 1H), 3.04 (s, 3H) (signals due to hydroxyl group protons not observed); ^{13}C NMR (100 MHz, CDCl$_3$) δ 171.8, 145.9, 144.8, 131.9, 128.7, 125.7, 125.1, 120.4, 111.9, 88.4, 68.5, 67.0, 56.1, 52.1, 43.2, 42.3, 35.9; IR ν_{max} 3357, 2924, 1623, 1508, 1438, 1280, 1101, 1070, 1031, 971, 880, 793 cm^{-1}; MS (ESI, +ve) m/z 657 [(2 M + Na)$^+$, 40%], 340 [(M + Na)$^+$, 100]; HRMS (ESI, +ve) m/z (M + Na)$^+$ calcd for C$_{17}$H$_{19}$NNaO$_5$ 340.1161, found 340.1159.

Compound ent-**2** was prepared in an analogous fashion from ent-**5** (20 mg, 0.048 mmol) to give 9.5 mg (63%) of product as a white, amorphous solid. A small sample was recrystallized (diethyl ether) to give a white, crystalline solid; mp 135−140 °C; $[\alpha]^{20}_D$ +127 (c 0.4, methanol). All of the other spectral data acquired on this material were identical with those reported above for compound **2**.

(4aR,5R,6S,8aS)-3-Methoxy-11-methyl-4a,5,9,10,11,12-hexahydro-6H-benzo[2,3]-benzofuro[4,3-cd]azepine-5,6-diol (3) and Enantiomer ent-3. A magnetically stirred solution of lactam **2** (9.5 mg, 0.03 mmol) in THF (5 mL) maintained at room temperature was treated with NaAlH$_2$(OCH$_2$CH$_2$OCH$_3$)$_2$ (60 μL of a 60% w/v solution in toluene, 0.184 mmol) and the ensuing mixture heated under reflux for 24 h after which time it was cooled to 0 °C (ice-bath), quenched with potassium sodium tartrate (2 mL of a saturated aqueous solution), diluted with water (10 mL) and then extracted with CHCl$_3$ (3 × 5 mL). The combined organic extracts were washed with brine (1 × 2 mL) before being dried (Na$_2$SO$_4$), filtered, and concentrated under reduced pressure. The residue thus obtained was subjected to flash chromatography (silica, 1:19 v/v NH$_3$ saturated methanol/CHCl$_3$ → 1:9 v/v NH$_3$ saturated methanol/CHCl$_3$ gradient elution) and concentration of appropriate fractions (R_f = 0.2 in 1:9 v/v NH$_3$ saturated methanol/CHCl$_3$) afforded amine **3** (4.0 mg, 44%) as a white, amorphous solid; $[\alpha]^{20}_D$ −86.6 (c 0.3, CDCl$_3$); ^{1}H NMR (400 MHz, CDCl$_3$) δ 6.66 (d, J = 8.2 Hz, 1H), 6.63 (d, J = 8.2 Hz, 1H), 6.23 (d, J = 10.2 Hz, 1H), 5.93 (dd, J = 10.2 and 3.8 Hz, 1H), 4.57 (d, J = 4.6 Hz, 1H), 4.30 (m, 1H), 4.19−4.09 (complex m, 2H), 3.84 (s, 3H), 3.63 (m, 1H), 3.27 (m, 1H), 2.98 (broad d, J = 14.8 Hz, 1H), 2.34 (s, 3H), 2.18 (m, 1H), 1.81 (dd, J = 13.4 and 2.2 Hz, 1H) (signals due to hydroxyl group protons not observed); ^{13}C NMR (100 MHz, CDCl$_3$) δ 145.4, 144.0, 132.3, 129.1, 128.1, 126.6, 122.4, 111.1, 90.9, 72.4, 67.6, 60.1, 55.9, 54.2, 49.8, 42.0, 36.0; IR ν_{max} 3345, 2921, 1626, 1596, 1507, 1439, 1282, 1041, 948, 793, 726 cm^{-1}; MS (ESI, +ve) m/z 629 [(2 M + Na)$^+$, 30%], 326 [(M + Na)$^+$, 100], 304 [(M + H)$^+$, 7]; HRMS (ESI, +ve) m/z (M + H)$^+$ calcd for C$_{17}$H$_{22}$NO$_4$ 304.1549, found 304.1544.

Compound ent-**3** was prepared in an analogous fashion from ent-**2** (9.5 mg, 0.030 mmol) to give 4.0 mg (44%) of product, $[\alpha]^{20}_D$ +60.5 (c 0.4, CDCl$_3$). All of the other spectral data acquired on this material were identical with those reported above for compound **3**.

2-((5aR,6R,7S,9aS)-6,7-Dihydroxy-4-methoxy-6,7-dihydrodibenzo[b,d]furan-9a-(5aH)-yl)-N,N-dimethylacetamide (4)

and Enantiomer ent-4. Amide **15** (15 mg, 0.035 mmol) was treated trifluoroacetic acid/water (200 μL of a 1:1 v/v mixture), and the resulting mixture was stirred at 22 °C for 0.5 h; then, the volatiles were removed under reduced pressure. The residue thus obtained was subjected to flash chromatography (silica, 1:19 v/v methanol/CH$_2$Cl$_2$), and concentration of the relevant fractions (R_f = 0.1) afforded diol **4** (9.8 mg, 88%) as a clear, colorless oil; $[\alpha]^{20}_D$ +5.7 (c 0.9, CH$_3$OH); ^{1}H NMR (400 MHz, CDCl$_3$) δ 6.86 (m, 1H), 6.78−6.73 (complex m, 2H), 5.98 (d, J = 9.9 Hz, 1H), 5.83 (d, J = 9.9 Hz, 1H), 5.01 (d, J = 7.5 Hz, 1H), 4.45 (broad s, 1H), 3.85 (s, 3H), 3.71 (m, 1H), 2.88 (s, 3H), 2.81 (s, 3H), 2.79 (m, 1H), 2.58 (d, J = 15.8 Hz, 1H) (signals due to hydroxyl protons not observed); ^{13}C NMR (100 MHz, CDCl$_3$) δ 169.7, 145.5, 145.3, 134.1, 129.4, 128.5, 122.0, 115.4, 111.9, 90.5, 74.4, 69.2, 56.1, 50.1, 42.7, 37.5, 35.6; IR ν_{max} 3389, 2927, 1616, 1489, 1457, 1402, 1271, 1093, 1058, 941, 749, 731 cm^{-1}; MS (ESI, +ve) m/z 661 [(2 M + Na)$^+$, 40%], 342 [(M + Na)$^+$, 100]; HRMS (ESI, +ve) m/z (M + Na)$^+$ calcd for C$_{17}$H$_{21}$NNaO$_5$ 342.1317, found 342.1311.

Compound ent-**4** was prepared in an analogous fashion from ent-**15** (23.7 mg, 0.055 mmol) to give 13.4 mg (77%) of product; $[\alpha]^{20}_D$ −8.2 (c 1.3, methanol). All of the other spectral data acquired on this material were identical with those reported above for compound **4**.

(1S,2S,3S,6S)-4-Bromo-6-((4-methoxybenzyl)oxy)cyclohex-4-ene-1,2,3-triol (21). Step i. BF$_3$·OEt$_2$ (100 μL of a 10% v/v solution in CH$_2$Cl$_2$) was added dropwise over 0.25 h to a magnetically stirred solution of epoxide **20**[23] (2.00 g, 8.09 mmol) and p-methoxybenzyl alcohol (23.8 g, 175 mmol) in dry CH$_2$Cl$_2$ (80 mL) maintained at −20 °C. The resulting solution was allowed to warm to −10 °C over 2 h after which time a second aliquot of BF$_3$·OEt$_2$ (150 μL of a 10% v/v solution in CH$_2$Cl$_2$) was added dropwise over 0.25 h. The reaction mixture thus formed was warmed to 22 °C over 12 h then quenched with phosphate buffer (3 mL of a 1 M aqueous solution at pH 7), and the solvent was then removed under reduced pressure. The residue thus obtained, which was comprised of a mixture of the desired PMB-ether and p-methoxybenzyl alcohol, was submitted directly to step ii as detailed immediately below.

Step ii. A magnetically stirred solution of the material obtained from step i in methanol (160 mL) was treated with pyridinium p-toluenesulfonate (2.03 g, 8.09 mmol) and the mixture so-formed was heated at 50 °C for 48 h; then, it was cooled to 22 °C and treated with NaHCO$_3$ (500 mg), and the solvent was removed under reduced pressure. The residue thus obtained was treated with ethyl acetate (200 mL) and then water (100 mL), and the separated aqueous layer was extracted with ethyl acetate (4 × 50 mL). The combined organic phases were washed with brine (1 × 30 mL) and then dried (Na$_2$SO$_4$), filtered, and concentrated under reduced pressure. The resulting yellow oil was subjected to flash column chromatography (silica, CH$_2$Cl$_2$ → 1:19 v/v methanol/CH$_2$Cl$_2$ gradient elution), and concentration of appropriate fractions (R_f = 0.2 in 1:19 v/v methanol/CH$_2$Cl$_2$) afforded triol **21** (1.98 g, 71% from **20**) as a white foam; $[\alpha]^{20}_D$ +132.1 (c 1.1, CHCl$_3$); ^{1}H NMR [400 MHz, (CD$_3$)$_2$CO] δ 7.32 (d, J = 8.8 Hz, 2H), 6.89 (d, J = 8.8 Hz, 2H), 6.14 (d, J = 2.4 Hz, 1H), 4.67 (s, 2H), 4.58 (d, J = 5.1 Hz, 1H), 4.27−4.20 (complex m, 2H), 4.03 (broad s, 1H), 3.86 (dd, J = 7.4 and 2.0 Hz, 1H), 3.82 (broad d, J = 9.7 Hz, 1H), 3.78 (s, 3H), 3.54 (m, 1H); ^{13}C NMR [100 MHz, (CD$_3$)$_2$CO] δ 160.2, 132.8, 131.8, 130.2, 124.8, 114.4, 80.8, 74.1, 72.3(2), 72.2(5), 71.8, 55.5; IR (KBr) ν_{max} 3437, 2918, 2848, 1732, 1449, 1368, 1241, 1072, 1026 cm^{-1}; MS (ESI, +ve) m/z 369 and 367 [(M + Na)$^+$, 100 and 98%]; HRMS (ESI, +ve) m/z (M + Na)$^+$ calcd for C$_{14}$H$_{17}^{79}$BrNaO$_5$ 367.0155, found 367.0157.

(2S,3S,4aR,5S,8S,8aR)-6-Bromo-2,3-dimethoxy-2,3-dimethyl-2,3,4a,5,8,8a-hexahydrobenzo[b][1,4]dioxine-5,8-diol (ent-11). A magnetically stirred solution of triol **21** (2.57 g, 7.44 mmol), 2,2,3,3-tetramethoxybutane (1.91 g, 10.7 mmol), and trimethyl orthoformate (3.40 mL, 31.1 mmol) in dry methanol (50 mL) was treated with p-TsOH·H$_2$O (73 mg, 5 mol %). The resulting mixture was heated under reflux for 24 h, cooled to 22 °C, and treated with NaHCO$_3$ (2.00 g), and the solvent was removed under reduced pressure. The residue thus obtained was treated with ethyl acetate (100 mL) and then NaHCO$_3$ (30 mL of a saturated aqueous solution). The separated

114

organic phase was washed with NaHCO$_3$ (1 × 30 mL of a saturated aqueous solution) and water (1 × 30 mL), and then the combined aqueous layers were extracted with ethyl acetate (3 × 30 mL). The combined organic phases were washed with brine (2 × 20 mL) and then dried (Na$_2$SO$_4$), filtered, and concentrated under reduced pressure. The resulting thick, orange oil was subjected to flash column chromatography (silica, 1:20 v/v methanol/CH$_2$Cl$_2$ elution), and concentration of appropriate fractions (R_f = 0.3) afforded bis-acetal *ent*-**11** (2.14 g, 85%) as a white foam; $[\alpha]^{20}_D$ +72.6 (*c* 1.1, CHCl$_3$) {lit.[15] (for **11**); $[\alpha]^{20}_D$ −76.5 (*c* 1.0, CHCl$_3$)}; ^{1}H NMR (400 MHz, CD$_3$OD) δ 6.05 (d, *J* = 2.5 Hz, 1H), 4.22 (d, *J* = 4.1 Hz, 1H), 4.06 (dd, *J* = 7.9 and 2.5 Hz, 1H), 3.87 (dd, *J* = 11.1 and 7.9 Hz, 1H), 3.63 (dd, *J* = 11.1 and 4.1 Hz, 1H), 3.30 (s, 3H), 3.25 (s, 3H), 1.32 (s, 3H), 1.29 (s, 3H) (signals due to hydroxyl group protons not observed); ^{13}C NMR (100 MHz, CD$_3$OD) δ 135.2, 124.0, 100.8, 100.2, 73.1, 71.8, 69.8, 69.2, 48.3, 48.2, 18.1, 18.0; IR (KBr) ν_{max} 3160, 2940, 1636, 1454, 1375, 1136, 1077, 1031, 980, 915 cm^{-1}; MS (ESI, +ve) *m/z* 363 and 361 [(M + Na)$^+$, 95 and 100%]; HRMS (ESI, +ve) (M+Na)$^+$ calcd for C$_{12}$H$_{19}$^{79}BrNaO$_6$ 361.0263, found 361.0263.

3-Bromo-4-hydroxy-5-methoxybenzaldehyde (23). A magnetically stirred solution of vanillin (**22**) (4.00 g, 26.3 mmol) in acetic acid (10 mL) was treated with molecular bromine (1.34 mL, 0.03 mol), and the ensuing mixture was stirred at 22 °C for 3 h during which time a precipitate appeared. The reaction mixture was quenched with water (30 mL), and the precipitate was filtered off, washed with water (1 × 50 mL) and then methanol (1 × 20 mL) before being dried under vacuum to afford compound **23**[28] (5.40 g, 90%) as a white, crystalline solid; mp 164 °C (lit.[28] mp 160−162 °C); ^{1}H NMR (CDCl$_3$, 400 MHz) δ 9.79 (s, 1H), 7.64 (d, *J* = 1.7 Hz, 1H), 7.36 (d, *J* = 1.7 Hz, 1H), 6.52 (s, 1H), 3.99 (s, 3H); ^{13}C NMR (CDCl$_3$, 100 MHz) δ 189.6, 148.9, 147.7, 130.1, 130.0, 108.2, 108.0, 56.6; IR ν_{max} 3305, 2980, 1674, 1590, 1463, 1290, 1157, 1047, 680 cm^{-1}; MS (ESI, + ev) *m/z* 255 and 253 [(M+Na)$^+$, 100 and 99%]; HRMS (ESI, +ve) (M+Na)$^+$ calcd for C$_8$H$_7$^{79}BrNaO$_3$ 252.9476, found 252.9479.

2-Bromo-4-(1,3-dioxolan-2-yl)-6-methoxyphenol (24). Compound **23** (5.00 g, 21.8 mmol), toluene (120 mL), *p*-TsOH·H$_2$O (39 mg, 0.21 mmol) and ethylene glycol (3.60 mL, 65.2 mmol) were placed in a round-bottom flask fitted with a Dean−Stark trap and condenser. The ensuing mixture was heated under reflux for 5 h before being cooled, quenched with NaHCO$_3$ (100 mL of a saturated solution), and extracted with ethyl acetate (1 × 100 mL). The organic phase was dried (Na$_2$SO$_4$), filtered, and concentrated under reduced pressure, and the light-yellow oil thus obtained was subjected flash chromatography (silica, 5:1 v/v hexane/ethyl acetate elution). Concentration of the relevant fractions (R_f = 0.4 in 1:9 v/v ethyl acetate/hexane) gave compound **24** (3.30 g, 55%) as a light-yellow oil. ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.24 (d, *J* = 1.5 Hz, 1H), 6.94 (d, *J* = 1.5 Hz, 1H), 5.97 (s, 1H), 5.71 (s, 1H), 4.14−4.11 (complex m, 2H), 4.04−3.98 (complex m, 2H), 3.92 (s, 3H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 147.2, 143.8, 130.7, 123.2, 108.0, 107.9, 102.9, 65.2, 56.4; IR ν_{max} 3358, 2963, 2887, 1684, 1603, 1589, 1503, 1466, 1426, 1280, 1181, 1090, 1044 cm^{-1}; MS (ESI, +ve) *m/z* 299 and 297 [(M+Na)$^+$, 92 and 100%] 277 and 275 [(M+H)$^+$, 45 and 40]; HRMS (ESI, +ve) (M+H)$^+$ calcd for C$_{10}$H$_{12}$^{79}BrO$_4$ 274.9919, found 274.9921.

2-(3-Bromo-5-methoxy-4-(methoxymethyl)phenyl)-1,3-dioxolane (25). A magnetically stirred mixture of phenol **24** (3.23 g, 11.79 mmol) in dry THF (30 mL) maintained at 0 °C was treated with NaH (564 mg of a 60% suspension in oil, 14.2 mmol). After 0.5 h, the reaction mixture was treated with chloromethyl methyl ether (980 μL, 12.9 mmol) and then stirred at 22 °C for 18 h before being quenched with water (100 mL; *CAUTION!* possibility of hydrogen gas evolution). The separated aqueous layer was extracted with diethyl ether (3 × 30 mL), and the combined organic phases were washed with brine (1 × 30 mL) and then dried (Na$_2$SO$_4$), filtered, and concentrated under reduced pressure. The residue thus obtained was subjected to flash chromatography (silica, 1:4 v/v ethyl acetate/hexane elution) and gave, after concentration of the relevant fractions (R_f = 0.5 in 1:3 v/v ethyl acetate/hexane), bromide **25** (2.65 g, 70%) as a clear, colorless oil; ^{1}H NMR (CDCl$_3$, 400 MHz) δ 7.28 (d, *J* = 1.9 Hz, 1H), 6.98 (d, *J* = 1.9 Hz, 1H), 5.74 (s, 1H), 5.17 (s, 2H), 4.13−4.01

(complex m, 4H), 3.86 (s, 3H), 3.64 (s, 3H); ^{13}C NMR (CDCl$_3$, 100 MHz) δ 153.3, 143.9, 135.1, 123.1, 117.6, 109.6, 102.6, 98.6, 65.3, 58.0, 56.1; IR ν_{max} 2941, 2891, 2839, 1696, 1570, 1484, 1463, 1416, 1384, 1274, 1157, 1081, 1043, 942, 854 cm^{-1}; MS (EI, 70 eV) *m/z* 320 and 318 (M$^{+\bullet}$, 99 and 100%), 289 and 287 (55 and 53), 239 (85), 166 (25); HRMS (EI, 70 eV) M$^{+\bullet}$ calcd for C$_{12}$H$_{15}$^{79}BrO$_5$ 318.0103, found 318.0104.

4-Hydroxy-3-methoxy-5-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)benzaldehyde (27). Step i. A magnetically stirred mixture of bromide **25** (1.30 g, 4.15 mmol) in dry THF (15 mL) maintained at −78 °C was treated with *n*-BuLi (3.1 mL of a 1.6 M solution in THF, 5.0 mmol). After 1 h, the reaction mixture was treated with tri-*iso*-propyl borate (1.9 mL, 8.3 mmol) and then stirred at 22 °C for 15 h before being quenched with HCl (10 mL of a 10% w/v aqueous solution). The separated aqueous layer was extracted with ethyl acetate (3 × 50 mL), and the combined organic phases were washed with brine (1 × 50 mL) before being dried (Na$_2$SO$_4$), filtered, and concentrated under reduced pressure. The residue thus obtained, and which is presumed to contain boronic acid **26**, was subjected directly to step ii of the reaction sequence.

Step ii. A magnetically stirred mixture of the product obtained from step i in dry acetonitrile (40 mL) maintained at 0 °C was treated with sodium iodide (0.62 g, 14.5 mmol) and chlorotrimethylsilane (530 μL, 14.5 mmol). The resulting solution was warmed to 22 °C over 4 h and then treated with Na$_2$S$_2$O$_3$ (20 mL of a saturated aqueous solution). The separated aqueous layer was extracted with ethyl acetate (3 × 20 mL), and the combined organic phases were washed with brine (1 × 20 mL) before being dried (Na$_2$SO$_4$), filtered, and concentrated under reduced pressure. The residue thus obtained was immediately subjected to step iii of the reaction sequence.

Step iii. A magnetically stirred suspension of the product obtained from step ii in benzene (30 mL) was treated with pinacol (990 mg, 8.40 mmol), and the solution thus obtained was heated under reflux for 2 h in an apparatus fitted with a Dean−Stark trap and a condenser. The cooled reaction mixture was treated with water (20 mL), and the separated aqueous layer was extracted with ethyl acetate (3 × 20 mL). The combined organic phases were then washed with brine (1 × 50 mL) before being dried (Na$_2$SO$_4$), filtered, and concentrated under reduced pressure. The residue thus obtained was subjected to flash chromatography (silica, 2:3 v/v diethyl ether/hexane elution) to afford, after concentration of the relevant fractions (R_f = 0.4), boronic ester **27** (520 mg, 45% from **25**) as a white, crystalline solid; mp 73 °C; ^{1}H NMR (400 MHz, CDCl$_3$) δ 9.82 (s, 1H), 8.33 (s, 1H), 7.72 (d, *J* = 1.9 Hz, 1H), 7.49 (d, *J* = 1.9 Hz, 1H), 3.93 (s, 3H), 1.38 (s, 12H); ^{13}C NMR (CDCl$_3$, 100 MHz) δ 190.9, 158.6, 148.3, 133.8, 129.6, 111.7, 85.1, 56.0, 24.8; IR ν_{max} 3407, 2992, 2931, 2830, 2797, 2714, 1683, 1620, 1587, 1467, 1388, 1372, 1298, 1256, 1140, 1056, 980, 846, 674 cm^{-1}; MS (EI, 70 eV) *m/z* 278 (M$^{+\bullet}$, 38%), 221 (100), 178 (51), 177 (30); HRMS (EI, 70 eV) M$^{+\bullet}$ calcd for C$_{14}$H$_{19}$BO$_5$ 278.1326, found 278.1326.

(3aS,4S,5S,7aS)-7-Bromo-2,2-dimethyl-5-((triiso-propylsilyl)oxy)-3a,4,5,7a-tetrahydrobenzo[d][1,3]dioxol-4-ol (29). Tri-*iso*-propylsilyl trifluoromethanesulfonate (1.95 mL, 7.3 mmol) was added dropwise to a magnetically stirred solution of compound **28**[29] (1.40 g, 5.3 mmol) and 2,6-lutidine (2.5 mL, 21.5 mmol) in CH$_2$Cl$_2$ (30 mL) maintained at −78 °C under a nitrogen atmosphere. The ensuing mixture was allowed to warm to 22 °C over 3 h and then treated with NH$_4$Cl (60 mL of a saturated aqueous solution). The separated aqueous phase was extracted with CH$_2$Cl$_2$ (1 × 20 mL), and the combined organic phases were dried (MgSO$_4$), filtered, and concentrated under reduced pressure. The resulting light yellow oil was subjected to flash chromatography (silica, 3:100 v/v ethyl acetate/hexane elution) and gave, after concentration of the appropriate fractions (R_f = 0.3 in 0.5:2.5:5.5 v/v/v ethyl acetate/CH$_2$Cl$_2$/hexane), an ∼6:1 mixture of compound **29** and its regioisomeric silyl ether (1.95 g, 88% combined yield) as a light-yellow oil; ^{1}H NMR (400 MHz, CDCl$_3$) δ 6.42 (s, 1H), 4.66 (d, *J* = 6.6 Hz, 1H), 4.17 (m, 1H), 4.11 (m, 1H), 3.55 (t, *J* = 8.7 Hz, 1H), 2.45 (s, 1H), 1.54 (s, 3H), 1.40 (s, 3H), 1.15−1.04 (complex m, 21H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 144.5, 110.2, 92.2, 79.3, 77.0, 74.5, 73.6, 28.0, 25.7, 18.0(1), 17.9(9),

115

12.4; IR ν_{max} 3469, 2943, 2893, 2866, 1635, 1463, 1382, 1248, 1070, 1019, 997, 882, 866, 828 cm^{-1}; MS (ESI, +ve) m/z 445 and 443 [(M +Na)$^+$, 100 and 97%]; HRMS (ESI, +ve) (M + Na)$^+$ calcd for C$_{18}$H$_{33}$^{79}BrNaO$_4$Si 443.1229, found 443.1232.

(((3aR,4R,5S,7aS)-7-Bromo-5-methoxy-2,2-dimethyl-3a,4,5,7a-tetrahydrobenzo[d][1,3]-dioxol-4-yl)oxy)tri-iso-propylsilane **(30).** Sodium hydride (342 mg of a 60% dispersion in mineral oil, 8.6 mmol) was added to a magnetically stirred solution of an ~6:1 mixture of compound **29** and its regioisomer (1.20 g, 2.9 mmol) and iodomethane (391 μL, 6.3 mmol) in dry THF (20 mL) maintained at 0 °C under a nitrogen atmosphere. Stirring was continued for 2 h at 22 °C, and then the reaction mixture was treated with ice-water (70 mL) (*CAUTION!* possibility of hydrogen evolution). The separated aqueous phase was extracted with ethyl acetate (1 × 30 mL), and the combined organic phases were then dried (MgSO$_4$), filtered, and concentrated under reduced pressure. The ensuing light-yellow oil was subjected to flash chromatography (silica, 1:50 v/v ethyl acetate/hexane elution) to give, after concentration of the appropriate fractions (R_f = 0.4 in 0.5:2.5:5.5 v/v/v ethyl acetate/CH$_2$Cl$_2$/ hexane), an ~5:1 mixture of compound **30** and its regioisomer (612 mg, 49% combined yield) as a yellowish oil; ^{1}H NMR (400 MHz, CDCl$_3$) δ (major regioisomer) 6.16 (m, 1H), 4.63 (m, 1H), 4.47 (m, 1H), 4.36 (m, 1H), 3.86 (m, 1H), 3.41 (s, 3H), 1.42 (s, 3H), 1.40 (s, 3H), 1.12−0.95 (complex m, 21H); ^{13}C NMR (100 MHz, CDCl$_3$) δ (major regioisomer) 129.4, 122.3, 110.0, 77.3, 77.2(3), 77.2(2), 69.9, 57.3, 27.5, 26.2, 18.0(2), 17.9(5), 12.5; IR ν_{max} 2939, 2892, 2866, 1645, 1463, 1381, 1234, 1163, 1078, 1039, 1011, 941, 882, 815, 680 cm^{-1}; MS (ESI, +ve) m/z 459 and 457 [(M+Na)$^+$, 98 and 96%], 355 (100), 347 and 345 (67 and 65); HRMS (ESI, +ve) (M+Na)$^+$ calcd for C$_{19}$H$_{35}$^{79}BrNaO$_4$Si 457.1386, found 457.1389.

(3aS,4R,5S,7aS)-7-Bromo-5-methoxy-2,2-dimethyl-3a,4,5,7a-tet-rahydrobenzo-[d][1,3]dioxol-4-ol **(31).** A magnetically stirred solution of an ~5:1 mixture of compound **30** and its regioisomer (600 mg, 1.4 mmol) in THF (10 mL) maintained at 22 °C under a nitrogen atmosphere was treated with tetra-*n*-butylammonium fluoride (2 mL of 1.0 M solution in THF, 2.0 mmol). After 2 h, the reaction mixture was concentrated under reduced pressure, and the residue so-formed was subjected to flash chromatography (silica, 1:2 v/v ethyl acetate/ hexane elution) to provide, after concentration of the appropriate fractions (R_f = 0.4 in 4:2.5:5.5 v/v/v ethyl acetate/CH$_2$Cl$_2$/hexane), compound **31** (377 mg, 98%) as a light-yellow oil; $[\alpha]^{20}_D$ −15.7 (*c* 2.4, CHCl$_3$); ^{1}H NMR (400 MHz, CDCl$_3$) δ 6.28 (s, 1H), 4.68 (d, J = 6.5 Hz, 1H), 4.15 (m, 1H), 3.66 (m, 2H), 3.48 (s, 3H), 2.64 (broad s, 1H), 1.55 (s, 3H), 1.42 (s, 3H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 132.2, 118.8, 111.0, 80.3, 77.5, 77.2, 72.5, 57.5, 28.1, 25.9; IR ν_{max} 3453, 2987, 2934, 2826, 1646, 1457, 1381, 1217, 1164, 1074, 869 cm^{-1}; MS (EI, 70 eV) m/z 280 and 278 (M$^{+\bullet}$, both 3%), 265 and 263 (both 35%), 101 (100); HRMS (EI, 70 eV) M$^{+\bullet}$ calcd for C$_{10}$H$_{15}$^{79}BrO$_4$ 278.0154, found 278.0148.

4-Hydroxy-3-((3aR,6S,7R,7aS)-7-hydroxy-6-methoxy-2,2-dimeth-yl-3a,6,7,7a-tetrahydrobenzo[d][1,3]dioxol-4-yl)-5-methoxybenzal-dehyde **(32).** A magnetically stirred solution of compound **31** (150 mg, 0.54 mmol), ester **27** (180 mg, 0.65 mmol), PdCl$_2$dppf·CH$_2$Cl$_2$ (31.5 mg, 0.04 mmol), and triethylamine (3 mL) in THF/water (18 mL of a 9:1 v/v mixture) was purged with nitrogen for 0.5 h and then heated under reflux for 2 h before being cooled, poured into water (50 mL), and extracted with ethyl acetate (3 × 30 mL). The combined organic phases were washed with brine (1 × 40 mL) and then dried (Na$_2$SO$_4$), filtered, and concentrated under reduced pressure. The ensuing light-yellow oil was subjected to flash chromatography (silica, 1:4 v/v ethyl acetate/hexane elution), and concentration of the relevant fractions (R_f = 0.5 in 1:3 v/v ethyl acetate/hexane) afforded phenol **32** (85 mg, 45%) as a white, crystalline solid; mp 129 °C, $[\alpha]^{20}_D$ +16.0 (*c* 0.2, CHCl$_3$); ^{1}H NMR (400 MHz, CDCl$_3$) δ 9.83 (s, 1H), 7.45 (d, J = 1.8 Hz, 1H), 7.39 (d, J = 1.8 Hz, 1H), 7.09 (s, 1H), 6.19 (d, J = 1.3 Hz, 1H), 5.17 (d, J = 6.4 Hz, 1H), 4.27 (dd, J = 8.9 and 6.4 Hz, 1H), 3.98 (s, 3H), 3.86 (m, 1H), 3.79 (m, 1H), 3.54 (s, 3H), 2.71 (broad s, 1H), 1.53 (s, 3H), 1.40 (s, 3H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 190.9, 149.2, 147.7, 132.3, 131.7, 129.2, 128.2, 125.1, 110.7, 108.4, 79.9, 77.7, 73.9, 72.9, 57.5, 56.4, 28.2, 26.0; IR ν_{max} 3400, 2986,

2935, 2830, 1681, 1588, 1488, 1455, 1432, 1373, 1301, 1254, 1217, 1148, 1067, 990, 863, 732 cm^{-1}; MS (EI, 70 eV) m/z 350 (M$^{+\bullet}$, 5%), 292 (53), 260 (55), 232 (100), 231 (72), 218 (53), 203 (39), 189 (33); HRMS (EI, 70 eV) M$^{+\bullet}$ calcd for C$_{18}$H$_{22}$O$_7$ 350.1366, found 350.1369.

(2'R,3'R,4'S,5'S)-2',3',4',6-Tetrahydroxy-5,5'-dimethoxy-2',3',4',5'-tetrahydro-[1,1'-biphenyl]-3-carbaldehyde **(33).** Compound **32** (60 mg, 0.17 mmol) was treated with acetic acid/water (10 mL of a 2:1 v/v mixture), and the resulting mixture was stirred at 22 °C for 18 h and then cooled and concentrated under reduced pressure. Subjection of the residue thus obtained to flash chromatography (silica, 1:8:1 v/v/v methanol/ethyl acetate/hexane elution) gave, after concentration of the appropriate fractions (R_f = 0.4), compound **33** (43 mg, 81%) as a white powder; mp 191 °C; $[\alpha]^{20}_D$ +7.5 (*c* 0.1, CHCl$_3$); ^{1}H NMR [400 MHz, (CD$_3$)$_2$CO] δ 9.83 (s, 1H), 7.45 (d, J = 1.9 Hz, 1H), 7.39 (d, J = 1.9 Hz, 1H), 6.03 (d, J = 2.5 Hz, 1H), 4.63 (d, J = 3.8 Hz, 1H), 3.97 (m, 1H), 3.92 (s, 3H), 3.84 (dd, J = 7.6 and 2.5 Hz, 1H), 3.62 (dd, J = 10.3 and 3.8 Hz, 1H), 3.49 (s, 3H), 2.83 (broad, s, 3H) (signal due to a hydroxyl group proton not observed); ^{13}C NMR [100 MHz, (CD$_3$)$_2$CO] δ 191.3, 151.3, 149.5, 137.4, 132.1, 129.9, 128.6, 128.4, 110.1, 82.9, 72.7, 71.8, 70.2, 57.6, 56.6; IR ν_{max} 3346, 2926, 2839, 2821, 1678, 1586, 1454, 1429, 1383, 1298, 1257, 1145, 1106, 1070, 943, 858, 696 cm^{-1}; MS (ESI, +ve) m/z 643 [(2M + Na)$^+$, 35%], 333 [(M + Na)$^+$, 100]; HRMS (ESI, +ve) (M+Na)$^+$ calcd for C$_{15}$H$_{18}$NaO$_7$ 333.0950, found 333.0952.

(5aS,6S,7S,8S)-6,7-Dihydroxy-4,8-dimethoxy-5a,6,7,8-tetrahydrodibenzo[b,d]-furan-2-carbaldehyde **(7).** A magnetically stirred solution of phenol **33** (40 mg, 0.13 mmol) in THF (12 mL) was treated with Ph$_3$P (40 mg, 0.15 mmol), cooled to 0 °C, and treated dropwise with a solution of di-*iso*-propyl azodicarboxylate (25 μL, 0.13 mmol) in THF (1 mL). The reaction mixture thus obtained was stirred at 0 °C for 2 h and then concentrated under reduced pressure, and the ensuing light-yellow oil was subjected to flash chromatography (silica, 3:97 v/v methanol/CH$_2$Cl$_2$ elution). Concentration of the relevant fractions (R_f = 0.4 in 0.5:9.5 v/v methanol/ CH$_2$Cl$_2$) afforded benzofuran **7** (32 mg, 85%) as a clear, light-yellow oil; $[\alpha]^{20}_D$ +92.0 (*c* 0.2, CHCl$_3$); ^{1}H NMR [400 MHz, (CD$_3$)$_2$CO] δ 9.87 (s, 1H), 7.69 (d, J = 1.5 Hz, 1H), 7.44 (d, J = 1.5 Hz, 1H), 6.19 (t, J = 3.6 Hz, 1H), 5.09 (m, 1H), 4.94 (s, 1H), 4.48 (broad s, 1H), 4.09 (m, 1H), 3.95 (s, 3H), 3.81−3.77 (complex m, 2H), 3.49 (s, 3H); ^{13}C NMR [100 MHz, (CD$_3$)$_2$CO] δ 191.1, 157.6, 146.9, 138.0, 133.1, 127.7, 118.2, 117.6, 114.0, 89.1, 83.8, 77.3, 75.6, 57.5, 56.6; IR ν_{max} 3339, 2926, 2892, 2853, 2823, 1686, 1604, 1590, 1437, 1337, 1313, 1185, 1119, 1092, 1069, 997, 920, 721, 694 cm^{-1}; MS (ESI, +ve) m/z 607 [(2M + Na)$^+$, 40%], 315 [(M + Na)$^+$, 100], 293 (23), 195 (30); HRMS (ESI, +ve) (M + Na)$^+$ calcd for C$_{15}$H$_{16}$NaO$_6$ 315.0845, found 315.0847.

(3aS,4S,5R,7aS)-7-Bromo-2,2-dimethyl-5-((tri-iso-propylsilyl)oxy)-3a,4,5,7a-tetrahydrobenzo[d][1,3]dioxol-4-ol **(35).** Tri-*iso*-propylsilyl trifluoromethanesulfonate (620 μL, 2.3 mmol) was added dropwise to a magnetically stirred solution of compound **34**[31] (500 mg, 1.9 mmol) and 2,6-lutidine (880 μL, 7.6 mmol) in CH$_2$Cl$_2$ (15 mL) maintained at −78 °C under a nitrogen atmosphere. The ensuing mixture was allowed to warm to 22 °C over 3 h and then treated with NH$_4$Cl (30 mL of a saturated aqueous solution). The separated aqueous phase was extracted with CH$_2$Cl$_2$ (1 × 10 mL), and the combined organic phases were dried (MgSO$_4$), filtered, and concentrated under reduced pressure. The resulting light-yellow oil was subjected to flash chromatography (silica, 3:100 v/v ethyl acetate/hexane elution) and gave, after concentration of the appropriate fractions (R_f = 0.3 in 0.5:2.5:5.5 v/v/v ethyl acetate/CH$_2$Cl$_2$/hexane), an ~6:1 mixture of ether **35** and its regioisomer (700 mg, 88% combined yield) as a light-yellow oil. ^{1}H NMR (400 MHz, CDCl$_3$) δ (major regioisomer) 6.00 (m, 1H), 4.62 (dd, J = 5.2 and 1.5 Hz, 1H), 4.49 (m, 2H), 4.24 (m, 1H), 2.67 (d, J = 1.5 Hz, 1H), 1.40 (s, 3H), 1.39 (s, 3H), 1.12−0.99 (complex m, 21H); ^{13}C NMR (100 MHz, CDCl$_3$) δ (major regioisomer) 130.8, 123.5, 110.0, 75.9, 75.7, 69.3, 68.3, 27.5, 26.2, 17.9(4), 17.9(3), 12.2; IR ν_{max} 3560, 2943, 2893, 2867, 1645, 1463, 1382, 1370, 1340, 1236, 1080, 1055, 882, 863, 682 cm^{-1}; MS (EI, 70 eV) m/z 423 and 421 [(M+ H)$^+$, 10 and 9%], 407 and 405 [(M −

116

CH$_3$•)$^+$, 7 and 6], 321 and 319 (100 and 97); HRMS (EI, 70 eV) (M − CH$_3$•)$^+$ calcd for C$_{17}$H$_{30}$^{79}BrO$_4$Si 405.1097, found 405.1096.

(((3aS,4S,5R,7aS)-7-Bromo-4-methoxy-2,2-dimethyl-3a,4,5,7a-tetrahydrobenzo[d][1,3]diox-ol-5-yl)oxy)tri-iso-propylsilane (36). Sodium hydride (172 mg of a 60% dispersion in mineral oil, 4.3 mmol) was added to a magnetically stirred solution of an ~6:1 mixture of compound **35** and its regioisomer (600 mg, 1.4 mmol) and iodomethane (267 μL, 4.3 mmol) in dry THF (15 mL) maintained at 0 °C under a nitrogen atmosphere. Stirring was continued for 2 h at 22 °C, and then the reaction mixture was treated with ice-water (30 mL) (*CAUTION!* possibility of hydrogen gas evolution). The separated aqueous phase was extracted with ethyl acetate (1 × 15 mL), and the combined organic phases were then dried (MgSO$_4$), filtered, and concentrated under reduced pressure. The ensuing light-yellow oil was subjected to flash chromatography (silica, 1:50 v/v ethyl acetate/hexane elution) to give, after concentration of the appropriate fractions (R_f = 0.4 in 0.5:2.5:5.5 v/v/v ethyl acetate/CH$_2$Cl$_2$/hexane), an ~5:1 mixture of compound **36** and its regioisomer (570 mg, 92% combined yield) as a clear, colorless oil; ^{1}H NMR (400 MHz, CDCl$_3$) δ (major regioisomer) 6.15 (d, J = 3.0 Hz, 1H), 4.62 (m, 1H), 4.56 (m, 1H), 4.44 (t, J = 5.4 Hz, 1H), 3.70 (m, 1H), 3.54 (s, 3H), 1.41 (s, 3H), 1.38 (s, 3H), 1.10−0.98 (complex m, 21H); ^{13}C NMR (100 MHz, CDCl$_3$) δ (major regioisomer) 133.1, 122.2, 109.9, 80.3, 76.9, 75.1, 68.2, 59.7, 27.5, 26.0, 17.9(8), 17.9(6), 12.2; IR ν_{max} 2941, 2887, 2865, 1649, 1460, 1383, 1335, 1241, 1197, 1138, 1121, 1079, 1040, 956, 880, 858, 680 cm^{-1}; MS (ESI, +ve) m/z 459 and 457 [(M + Na)$^+$, 83 and 81%], 437 and 435 (88 and 86), 205 and 203 (97 and 100); HRMS (ESI, +ve) (M + Na)$^+$ calcd for C$_{19}$H$_{35}$^{79}BrNaO$_4$Si 457.1386, found 457.1375.

(3aS,4R,5R,7aS)-7-Bromo-4-methoxy-2,2-dimethyl-3a,4,5,7a-tetrahydrobenzo[d][1,3]-dioxol-5-ol (37). A magnetically stirred solution of an ~5:1 mixture of compound **36** and its regioisomer (600 mg, 1.4 mmol) in THF (10 mL) maintained at 22 °C under a nitrogen atmosphere was treated with tetra-n-butylammonium fluoride (2 mL of 1.0 M solution in THF, 4.1 mmol). After 2 h, the reaction mixture was concentrated under reduced pressure, and the residue so-formed was subjected to flash chromatography (silica, 1:2 v/v ethyl acetate/hexane elution). Concentration of the relevant fractions (R_f = 0.4 in 4:2.5:5.5 v/v/v ethyl acetate/CH$_2$Cl$_2$/hexane) gave compound **37** (338 mg, 88%) as a white powder; mp 61 °C; [α]^{20}$_D$ +15.3 (c 0.6, CHCl$_3$); ^{1}H NMR (400 MHz, CDCl$_3$, 400 MHz) δ 6.15 (d, J = 3.0 Hz, 1H), 4.59 (m, 1H), 4.53 (t, J = 5.1 Hz, 1H), 4.33 (complex m, 1H), 3.76 (t, J = 4.4 Hz, 1H), 3.54 (s, 3H), 2.55 (broad s, 1H), 1.42 (s, 3H), 1.40 (s, 3H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 132.0, 123.2, 110.3, 78.7, 76.2, 73.9, 66.3, 59.2, 27.6, 26.2; IR ν_{max} 3453, 2987, 2935, 2900, 2831, 1646, 1457, 1381, 1372, 1231, 1109, 1077, 1041, 964, 868 cm^{-1}; MS (EI, 70 eV) (EI, 70 eV) 280 and 278 (M$^{+•}$, both 1%), 265 and 263 (74 and 76), 177 and 175 (13 and 15), 124 (28), 115 (100); HRMS M$^{+•}$ calcd for C$_{10}$H$_{15}$^{79}BrO$_4$ 278.0154, found 278.0153.

4-Hydroxy-3-((3aR,6R,7R,7aR)-6-hydroxy-7-methoxy-2,2-dimethyl-3a,6,7,7a-tetrahydrobenzo[d][1,3]dioxol-4-yl)-5-methoxybenzaldehyde (38). A magnetically stirred solution of compound **37** (116 mg, 0.42 mmol), boronate ester **27** (139 mg, 0.50 mmol), PdCl$_2$dppf·CH$_2$Cl$_2$ (25 mg, 0.03 mmol), and triethylamine (2 mL) in THF/water (15 mL of a 9:1 v/v mixture) was purged with nitrogen for 0.5 h and then heated under reflux for 3 h before being cooled, poured into water (40 mL), and extracted with ethyl acetate (3 × 20 mL). The combined organic phases were washed with brine (1 × 30 mL) and then dried (Na$_2$SO$_4$), filtered, and concentrated under reduced pressure. The ensuing light-yellow oil was subjected to flash chromatography (silica, 1:4 v/v ethyl acetate/hexane elution), and concentration of the relevant fractions (R_f = 0.5 in 1:3 v/v ethyl acetate/hexane) gave phenol **38** (80 mg, 55%) as a light-yellow oil; [α]^{20}$_D$ +7.0 (c 0.2, CHCl$_3$); ^{1}H NMR (400 MHz, CDCl$_3$) δ 9.82 (s, 1H), 7.39 (d, J = 1.5 Hz, 1H), 7.37 (d, J = 1.5 Hz, 1H), 6.10 (d, J = 3.3 Hz, 1H), 5.15 (d, J = 5.8 Hz, 1H), 4.64 (t, J = 5.5 Hz, 1H), 4.52 (m, 1H), 3.96 (s, 3H), 3.79 (t, J = 4.6 Hz, 1H), 3.59 (s, 3H), 2.65 (broad d, J = 8.3 Hz, 1H), 1.46 (s, 3H), 1.40 (s, 3H) (signal due to hydroxyl group proton not observed); ^{13}C NMR (100 MHz, CDCl$_3$) δ 190.8, 149.4, 147.9, 134.8, 132.0, 129.2, 128.3, 125.9, 109.8, 108.4, 79.7, 73.6,

73.3, 65.0, 59.1, 56.3, 27.6, 25.9; IR ν_{max} 3368, 2985, 2936, 1681, 1588, 1456, 1432, 1297, 1149, 1120, 1071, 913, 873 cm^{-1}; MS (EI, 70 eV) m/z 350 (M$^{+•}$, 4%), 274 (41), 115 (100); HRMS (EI, 70 eV) M$^{+•}$ calcd for C$_{18}$H$_{22}$O$_7$ 350.1366, found 350.1370.

2-((3aS,4R,5R,7aR)-7-(5-Formyl-2,3-dimethoxyphenyl)-4-methoxy-2,2-dimethyl-3a,4,5,7a-tetrahydrobenzo[d][1,3]dioxol-5-yl)-N,N-dimethylacetamide (39). A magnetically stirred solution of compound **38** (60 mg, 0.17 mmol) in toluene (10 mL) was treated with N,N-dimethylacetamide dimethyl acetal (126 μL, 0.87 mmol), and the ensuing mixture was heated under reflux for 18 h. The cooled reaction mixture was concentrated under reduced pressure, and the residue thus obtained was subjected to flash chromatography (silica, 4:1 v/v ethyl acetate/hexane elution). Concentration of the appropriate fractions (R_f = 0.3 in 5:1 v/v ethyl acetate/hexane) then gave an ~5:1 mixture of amide **39** and its β-epimer (30 mg, 40% combined yield) as a clear, colorless oil; ^{1}H NMR (400 MHz, CDCl$_3$) δ (major epimer) 9.87 (s, 1H), 7.39−7.36 (complex m, 2H), 6.01 (d, J = 4.6 Hz, 1H), 5.20 (d, J = 6.7 Hz, 1H), 4.41 (t, J = 6.7 Hz, 1H), 3.91 (s, 3H), 3.86 (s, 3H), 3.65−3.59 (complex m, 1H), 3.48 (s, 3H), 3.34 (m, 1H), 3.01 (s, 3H), 2.94 (s, 3H), 2.73 (dd, J = 15.9 and 5.1 Hz, 1H), 2.27 (dd, J = 15.9 and 9.4 Hz, 1H), 1.44 (s, 3H), 1.35 (s, 3H); ^{13}C NMR (100 MHz, CDCl$_3$) δ (major epimer) 191.2, 171.3, 153.3, 152.1, 134.6, 133.9, 133.1, 132.3, 127.5, 109.6, 109.0, 79.7, 74.4, 73.6, 60.8, 58.1, 56.0, 37.3, 35.5, 33.7, 33.4, 27.9, 25.7; IR ν_{max} 2982, 2932, 2857, 2831, 1691, 1646, 1578, 1458, 1421, 1384, 1240, 1142, 1034, 1002, 913, 863, 733 cm^{-1}; MS (ESI, +ve) m/z 456 [(M + Na)$^+$, 78%], 434 (100), 376 (21); HRMS (ESI, +ve) (M + Na)$^+$ calcd for C$_{23}$H$_{31}$NNaO$_7$ 456.1998, found 456.1996.

2-((3R,4R,5R,6R)-5′-Formyl-5,6-dihydroxy-2′,3′,4-trimethoxy-3,4,5,6-tetrahydro[1,1′-biphenyl]-3-yl)-N,N-dimethylacetamide (8). An ~5:1 mixture of compound **39** and its β-epimer (16 mg, 0.04 mmol) was treated with acetic acid/water (10 mL of a 2:1 v/v mixture), and the resulting solution was heated at 50 °C for 22 h and then cooled to 22 °C and concentrated under reduced pressure. Subjection of the residue thus obtained to flash chromatography (silica, 1:8:1 v/v methanol/ethyl acetate/hexane elution) gave, after concentration of the appropriate fractions (R_f = 0.4 in 1:9 v/v methanol/ethyl acetate), an ~6:1 mixture of compound **8** and its β-epimer (10 mg, 69% combined yield) as a light-yellow oil; ^{1}H NMR (400 MHz, CDCl$_3$) δ (major epimer) 9.87 (s, 1H), 7.39 (complex m, 2H), 5.99 (m, 1H), 4.71 (m, 1H), 4.01 (dd, J = 9.2 and 4.0 Hz, 1H), 3.92 (s, 3H), 3.87 (s, 3H), 3.80 (m, 1H), 3.50 (m, 1H), 3.46 (s, 3H), 3.02 (s, 3H), 2.95 (s, 3H), 2.85 (broad s, 1H), 2.74 (dd, J = 15.9 and 5.0 Hz, 1H), 2.20 (dd, J = 15.9 and 9.0 Hz, 1H) (signal due to a hydroxyl group proton not observed); ^{13}C NMR (CDCl$_3$, 100 MHz) δ (major epimer) 191.1, 171.4, 153.2, 151.7, 135.1, 134.9, 133.3, 132.4, 127.5, 109.8, 77.7, 69.0, 68.5, 61.0, 57.9, 56.0, 37.3, 35.6, 34.3, 32.4; IR ν_{max} 3395, 2957, 2935, 2828, 1688, 1627, 1463, 1420, 1387, 1260, 1128, 1105, 1089, 797 cm^{-1}; MS (ESI, +ve) m/z 787 [(2M + H)$^+$, 30%], 416 [(M + Na)$^+$, 100], 394 (92), 376 (20); HRMS (ESI, +ve) (M + Na)$^+$ calcd for C$_{20}$H$_{27}$NNaO$_7$ 416.1685, found 416.1685.

4-Hydroxy-3-((3aR,6S,7R,7aR)-6-hydroxy-7-(methoxymethoxy)-2,2-dimethyl-3a,6,7,7a-tetrahydrobenzo[d][1,3]dioxol-4-yl)-5-methoxybenzaldehyde (43). A magnetically stirred solution of compound **42**[9a] (100 mg, 0.33 mmol), boronate ester **27** (108 mg, 0.39 mmol), PdCl$_2$dppf·CH$_2$Cl$_2$ (19 mg, 0.02 mmol), and triethylamine (3 mL) in THF/water (18 mL of a 9:1 v/v mixture) was purged with nitrogen for 0.5 h and then heated under reflux for 2 h before being cooled, poured into water (50 mL), and extracted with ethyl acetate (3 × 30 mL). The combined organic phases were washed with brine (1 × 40 mL) and then dried (Na$_2$SO$_4$), filtered, and concentrated under reduced pressure. The ensuing light-yellow oil was subjected to flash chromatography (silica, 1:1 v/v ethyl acetate/hexane elution), and concentration of the relevant fractions (R_f = 0.3 in 1:3 v/v ethyl acetate/hexane) then gave phenol **43** (60 mg, 49%) as a light-yellow oil; [α]^{20}$_D$ +18.4 (c 0.9, CHCl$_3$); ^{1}H NMR (400 MHz, CDCl$_3$) δ 9.82 (s, 1H), 7.46 (d, J = 1.6 Hz, 1H), 7.39 (d, J = 1.6 Hz, 1H), 7.07 (s, 1H), 6.16 (d, J = 2.2 Hz, 1H), 5.21 (d, J = 6.2 Hz, 1H), 4.90 (d, J = 6.8 Hz, 1H), 4.86 (d, J = 6.8 Hz, 1H), 4.36 (dd, J = 8.4 and 6.2 Hz, 1H), 4.27 (broad d, J = 8.1 Hz, 1H), 4.08 (d, J = 3.3 Hz, 1H),

3.97 (s, 3H), 3.67 (m, 1H), 3.50 (s, 3H), 1.51 (s, 3H), 1.40 (s, 3H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 190.8, 149.2, 147.6, 133.6, 131.8, 129.2, 128.5, 125.0, 110.5, 108.1, 98.0, 83.2, 76.5, 74.0, 69.4, 56.4, 55.9, 28.2, 26.0; IR ν_{max} 3370, 2985, 2948, 2936, 1683, 1588, 1488, 1456, 1432, 1372, 1300, 1250, 1217, 1149, 1105, 1058, 1037, 996, 866, 732 cm^{-1}; MS (EI, +ve) m/z 380 (M$^{+\bullet}$, 2%), 362 (12), 260 (59), 259 (54), 231 (100), 218 (48); HRMS (EI, +ve) M$^{+\bullet}$ calcd for C$_{19}$H$_{24}$O$_8$ 380.1471, found 380.1478.

2-((3aS,4R,5R,7aR)-7-(5-Formyl-2,3-dimethoxyphenyl)-4-(methoxymethoxy)-2,2-dimethyl-3a,4,5,7a-tetrahydrobenzo[d][1,3]-dioxol-5-yl)-N,N-dimethylacetamide (44). A magnetically stirred solution of compound **43** (33 mg, 0.09 mmol) in toluene (10 mL) was treated with *N,N*-dimethylacetamide dimethyl acetal (64 μL, 0.44 mmol), and the ensuing mixture was heated under reflux for 18 h. The cooled reaction mixture was concentrated under reduced pressure, and subjection of the ensuing residue to flash chromatography (silica, 9.5:0.5 v/v ethyl acetate/hexane elution) gave, after concentration of the appropriate fractions (R_f = 0.4 in ethyl acetate), an ~3:1 mixture of amide **44** and its β-epimer (14 mg, 35% combined yield) as a clear, colorless oil; ^{1}H NMR (400 MHz, CDCl$_3$) δ (major epimer) 9.87 (s, 1H), 7.38 (m, 2H), 5.98 (d, J = 4.3 Hz, 1H), 5.22 (d, J = 6.0 Hz, 1H), 4.77 (m, 2H), 4.46 (t, J = 6.5 Hz, 1H), 4.04 (dd, J = 6.9 and 4.5 Hz, 1H), 3.91 (s, 3H), 3.87 (s, 3H), 3.43 (s, 3H), 3.30 (m, 1H), 3.01 (s, 3H), 2.94 (s, 3H), 2.75 (dd, J = 15.8 and 5.1 Hz, 1H), 2.33 (dd, J = 15.8 and 9.5 Hz, 1H), 1.44 (s, 3H), 1.35 (s, 3H); ^{13}C NMR (mixture of epimers) (100 MHz, CDCl$_3$) δ 191.3, 171.2, 153.3, 152.0, 134.6, 134.0, 133.0, 132.3, 129.3, 127.7, 113.8, 109.6, 109.2, 100.0, 96.5, 76.2, 75.0, 74.1, 73.8, 60.8, 56.0, 55.7, 37.2, 35.5, 34.2, 33.8, 28.0, 27.9, 26.0; IR ν_{max} 2982, 2933, 2843, 1692, 1647, 1579, 1455, 1384, 1243, 1145, 1070, 1037, 918, 863 cm^{-1}; MS (ESI, +ve) m/z 486 [(M + Na)$^+$, 100%], 464 (3), 60 (10); HRMS (ESI, +ve) (M + Na)$^+$ calcd for C$_{24}$H$_{33}$NNaO$_8$ 486.2104, found 486.2105.

2-((3R,4R,5R,6R)-5'-Formyl-5,6-dihydroxy-2',3'-dimethoxy-4-(methoxymethoxy)-3,4,5,6-tetrahydro[1,1'-biphenyl]-3-yl)-N,N-dimethylacetamide (9). Compound **44** (8 mg, 0.02 mmol) was treated with acetic acid/water (10 mL of a 2:1 v/v mixture), and the resulting solution was heated at 50 °C for 22 h and then cooled and concentrated under reduced pressure. Subjection of the residue thus obtained to flash chromatography (silica, 1:8:1 v/v methanol/ethyl acetate/hexane elution) gave, after concentration of the appropriate fractions (R_f = 0.4 in 1:9 v/v methanol/ethyl acetate), an ~6.5:1 mixture of compound **9** and its β-epimer (5 mg, 69%) as a light-yellow oil; ^{1}H NMR (400 MHz, CDCl$_3$) δ (major epimer) 9.87 (s, 1H), 7.38 (m, 2H), 5.97 (d, J = 4.4 Hz, 1H), 4.76−4.71 (complex m, 3H), 4.14 (dd, J = 9.1 and 5.7 Hz, 1H), 4.01 (m, 1H), 3.92 (s, 3H), 3.87 (s, 3H), 3.45 (s, 3H), 3.34 (broad s, 1H), 3.02 (s, 3H), 2.94 (s, 3H), 2.84 (dd, J = 15.9 and 5.2 Hz, 1H), 2.25 (dd, J = 15.9 and 8.9 Hz, 1H) (signal due to hydroxyl group protons not observed); ^{13}C NMR (100 MHz, CDCl$_3$) δ (major epimer) 191.2, 171.4, 153.2, 151.7, 135.2, 135.0, 133.3, 132.5, 127.4, 109.8, 97.5, 76.9, 69.2, 68.9, 61.0, 56.0(1), 55.9(6), 37.3, 35.8, 35.6, 33.2; IR ν_{max} 3384, 2919, 2848, 1688, 1630, 1579, 1463, 1419, 1387, 1329, 1292, 1245, 1147, 1130, 1037, 917, 862 cm^{-1}; MS (ESI, +ve) m/z 446 [(M + Na)$^+$, 68%], 424 (100); HRMS (ESI, +ve) (M + Na)$^+$ calcd for C$_{21}$H$_{29}$NNaO$_8$ 446.1791, found 446.1780.

Crystallographic Studies. *Crystallographic Data.* Compound *ent*-**2**. C$_{17}$H$_{19}$NO$_5$, M = 317.34, T = 150 K, orthorhombic, space group $P2_12_12_1$, Z = 4, a = 6.58612(5) Å, b = 9.28140(8) Å, c = 23.3001(2) Å; V = 1424.30(2) Å^3, D_x = 1.480 g cm^{-3}, 2794 unique data ($2\theta_{max}$ = 144.6°), R = 0.027 [for 2725 reflections with I > 2.0$\sigma(I)$]; Rw = 0.072 (all data), S = 1.01.

Compound *ent*-**15**. C$_{23}$H$_{31}$NO$_7$, M = 433.50, T = 150 K, monoclinic, space group $P2_1$, Z = 2, a = 10.14845(9) Å, b = 10.61199(7) Å, c = 10.80864(10) Å; β = 106.8796(9)°; V = 1113.89(2) Å^3, D_x = 1.292 g cm^{-3}, 3907 unique data ($2\theta_{max}$ = 144.8°), R = 0.026 [for 3843 reflections with I > 2.0$\sigma(I)$]; Rw = 0.067 (all data), S = 1.03.

Compound **27**. C$_{14}$H$_{19}$BO$_5$, M = 278.11, T = 150 K, orthorhombic, space group *Pbam*, Z = 8, a = 23.0535(6) Å, b = 18.2756(6) Å, c = 6.8183(2) Å; V = 2872.66(15) Å^3, D_x = 1.286 g cm^{-3}, 3098 unique

data ($2\theta_{max}$ = 145.2°), R = 0.084 [for 2982 reflections with I > 2.0$\sigma(I)$]; Rw = 0.191 (all data), S = 1.06.

Compound **32**. C$_{18}$H$_{22}$O$_7$, M = 359.38, T = 150 K, monoclinic, space group $P2_1$, Z = 4, a = 7.3153(1) Å, b = 31.4358(3) Å, c = 7.8245(1) Å; β = 94.8743(8)°; V = 1792.83(4) Å^3, D_x = 1.331 g cm^{-3}, 7018 unique data ($2\theta_{max}$ = 145°), R = 0.030 [for 6919 reflections with I > 2.0$\sigma(I)$]; Rw = 0.080 (all data), S = 0.99.

Structure Determination. Diffraction images for compounds *ent*-**2**, *ent*-**15**, **27**, and **32** were all measured on a diffractometer (Mo Kα, mirror monochromator, λ = 0.71073 Å or, for **32**, Cu Kα mirror monochromator, λ = 1.54184 Å) fitted with an area detector, and the data were extracted using the DENZO/Scalepack package.[43] The structure solutions for all four compounds were solved by direct methods (SIR92)[44] and then refined using the CRYSTALS program package.[45] Atomic coordinates, bond lengths and angles, and displacement parameters have been deposited at the Cambridge Crystallographic Data Centre (CCDC nos. 1517512, 1517513, 1517514 and 1517515). These data can be obtained free-of-charge via www.ccdc.cam.ac.uk/data_request/cif, by emailing data_request@ccdc.cam.ac.uk, or by contacting The Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, UK; fax: + 44 1223 336033.

AChE Inhibition Testing. The galanthamine derivatives/analogues described above were tested for inhibition against AChE as described by Sangnoi et al.[46] Thus, the compounds were dissolved in DMSO and made up to concentrations ranging from 5 mM to 3.05 μM by serial dilution with DMSO. Then, 2.5 μL of a solution of each compound was added to a solution of 5,5'-dithiobis(2-nitrobenzoic acid) (125 μL of a 1.5 mM aqueous solution), tris(hydroxymethyl)-aminomethane buffer (72.5 μL of a 15 mM solution at pH 8.0), and acetylthiocholine iodide (25 μL of a 150 μM aqueous solution) in water. Enzyme activity was followed after the addition of AChE (25 μL of 0.03 μM solution of *Electrophorus electricus*, Type V−S, EC 3.1.1.7) by measuring the absorption at 412 nm using a microplate spectrophotometer. Assays were repeated in triplicate, and the hydrolysis rate was calculated during the data from the absorptions observed over the first 2 min. Standard commercially available graphic software was used to calculate the tabulated IC$_{50}$ values.

Molecular Docking Studies. The three-dimensional coordinates of each compound were generated with the PRODRG server (http://davapc1.bioch.dundee.ac.uk/cgi-bin/prodrg).[47] Galanthamine and the above-mentioned derivatives/analogues were docked into the structure of human acetylcholine esterase (4EY6) using Autodock Vina v1.1.2 after removal of galanthamine from the active site gorge.[48]

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.joc.7b01062.

> Data derived from the single-crystal X-ray analyses of compounds *ent*-**2**, *ent*-**15**, **27**, and **32** and ^{1}H and ^{13}C NMR spectra of compounds **2**-**9**, *ent*-**11**, **13**-**17**, **21**, **23**, **24**, **25**, **27**, **29**−**33**, **35**−**39**, **43**, and **44** (PDF)
> CIFs of compounds *ent*-**2**, *ent*-**15**, **27**, and **32** (CCDC Nos. 1517512, 1517513, 1517514, and 1517515, respectively)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: Martin.Banwell@anu.edu.au.

ORCID ⓘ

Colin J. Jackson: 0000-0001-6150-3822

Martin G. Banwell: 0000-0002-0582-475X

Notes

The authors declare no competing financial interest.

118

ACKNOWLEDGMENTS

REFERENCES

(1) For useful points-of-entry into the literature on the history, chemistry, and pharmacology of galanthamine, see (a) Heinrich, M.; Teoh, H. L. *J. Ethnopharmacol.* **2004**, *92*, 147. (b) Marco-Contelles, J.; do Carmo Carreiras, M.; Rodríguez, C.; Villarroya, M.; García, A. G. *Chem. Rev.* **2006**, *106*, 116. (c) Marco, L.; do Carmo Carreiras, M. *Recent Pat. CNS Drug Discovery* **2006**, *1*, 105. (d) Heinrich, M. *Alkaloids* **2010**, *68*, 157. (e) Banwell, M. G.; Buckler, J. N.; Jackson, C. J.; Lan, P.; Ma, X.; Matoušová, E.; Nugent, J. *Strategies and Tactics in Organic Synthesis* **2015**, *11*, 29. (f) Rinner, U.; Dank, C.; Hudlicky, T. In *Targets in Heterocyclic Systems − Chemistry and Properties*; Attanasi, O. A., Merino, P., Spinelli, D., Eds., **2017**; Vol 20, pp 286−318;. (g) Fraser, M. D.; Davies, J. R. T.; Chang, X. *J. Alzheimer's Dis.* **2017**, *55*, 1321.

(2) For a useful point-of-entry into the literature on this topic, see: Gallagher, R.; Chebib, M.; Balle, T.; McLeod, M. D. *Aust. J. Chem.* **2015**, *68*, 1834.

(3) Kimura, H.; Kawai, T.; Hamashima, Y.; Kawashima, H.; Miura, K.; Nakaya, Y.; Hirasawa, M.; Arimitsu, K.; Kajimoto, T.; Ohmomo, Y.; Ono, M.; Node, M.; Saji, H. *Bioorg. Med. Chem.* **2014**, *22*, 285.

(4) Barton, D. H. R.; Kirby, G. W. *J. Chem. Soc.* **1962**, *806*, 1.

(5) For key examples, see: (a) Fröhlich, J.; Jordis, U. *Org. Process Res. Dev.* **1999**, *3*, 425. (b) Kodama, S.; Hamashima, Y.; Nishide, K.; Node, M. *Angew. Chem., Int. Ed.* **2004**, *43*, 2659.

(6) Magnus, P.; Sane, N.; Fauber, B. P.; Lynch, V. *J. Am. Chem. Soc.* **2009**, *131*, 16045.

(7) (a) Trost, B. M.; Tang, W.; Toste, F. D. *J. Am. Chem. Soc.* **2005**, *127*, 14785 and references cited therein. (b) Satcharoen, V.; McLean, N. J.; Kemp, S. C.; Camp, N. P.; Brown, R. C. D. *Org. Lett.* **2007**, *9*, 1867. (c) Chen, J. Q.; Xie, J.-H.; Bao, D.-H.; Liu, S.; Zhou, Q.-L. *Org. Lett.* **2012**, *14*, 2714. (d) Zang, Y.; Ojima, I. *J. Org. Chem.* **2013**, *78*, 4013. (e) Choi, J.; Kim, H.; Park, S.; Tae, J. *Synlett* **2013**, *24*, 379. For a rather different approach, see (f) Hu, X.-D.; Tu, Y. Q.; Zhang, E.; Gao, S.; Wang, S.; Wang, A.; Fan, C.-A.; Wang, M. *Org. Lett.* **2006**, *8*, 1823.

(8) See, for example: (a) Kato, T.; Tanimoto, H.; Yamada, H.; Chida, N. *Heterocycles* **2010**, *82*, 563. (b) Chida, N.; Sato, T. *Chem. Rec.* **2014**, *14*, 592. (c) Endoma-Arias, M. A. A.; Hudlicky, T. *Chem. - Eur. J.* **2016**, *22*, 14540.

(9) (a) Banwell, M. G.; Ma, X.; Karunaratne, O. P.; Willis, A. C. *Aust. J. Chem.* **2010**, *63*, 1437. (b) Nugent, J.; Matoušová, E.; Banwell, M. G. *Eur. J. Org. Chem.* **2015**, *2015*, 3771. (c) Nugent, J.; Banwell, M. G. *Eur. J. Org. Chem.* **2016**, *2016*, 5862.

(10) For some representative and seminal studies, see: (a) Han, S.-Y.; Mayer, S. C.; Schweiger, E. J.; Davis, B. M.; Joullié, M. M. *Bioorg. Med. Chem. Lett.* **1991**, *1*, 579. (b) Mary, A.; Renko, D. Z.; Guillou, C.; Thal, C. *Tetrahedron Lett.* **1997**, *38*, 5151. (c) Jia, P.; Sheng, R.; Zhang, J.; Fang, L.; He, Q.; Yang, B.; Hu, Y. *Eur. J. Med. Chem.* **2009**, *44*, 772. (d) Rosini, M.; Simoni, E.; Minarini, A.; Melchiorre, C. *Neurochem. Res.* **2014**, *39*, 1914. (e) Bhattacharya, S.; Maelicke, A.; Montag, D. *J. Alzheimer's Dis.* **2015**, *46*, 123. (f) Vezenkov, L. T.; Ilieva, L.; Danalev, D. L.; Bakalova, A.; Vassilev, D. N.; Danchev, N.; Nikolova, I. *Protein Pept. Lett.* **2015**, *22*, 913.

(11) For a useful discussion of such matters, see: Kitisripanya, N.; Saparpakorn, P.; Wolschann, P.; Hannongbua, S. *Nanomedicine* **2011**, *7*, 60 and references cited therein.

(12) See, for example: (a) Pelish, H. E.; Westwood, N. J.; Feng, Y.; Kirchhausen, T.; Shair, M. D. *J. Am. Chem. Soc.* **2001**, *123*, 6740. (b) Santra, S.; Andreana, P. R. *Angew. Chem., Int. Ed.* **2011**, *50*, 9418. (c) Yan, A.; Wang, K. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 3336.

(13) See, for example: (a) Yan, X.; Tang, J.; dos Santos Passos, C.; Nurisso, A.; Avello Simões-Pires, C.; Ji, M.; Lou, H.; Fan, P. *J. Agric. Food Chem.* **2015**, *63*, 10611. (b) Wang, M.; Sun, M.; Hai, H.; Lu, C. *J. Nat. Prod.* **2015**, *78*, 3067. (c) Zhang, J.-J.; Yang, X.-W.; Liu, X.; Ma, J.-Z.; Liao, Y.; Xu, G. *J. Nat. Prod.* **2015**, *78*, 3075. (d) Dong, L.-B.; Wu, X.-D.; Shi, X.; Zhang, Z.-J.; Yang, J.; Zhao, Q.-S. *Org. Lett.* **2016**, *18*, 4498.

(14) See, for example: Lan, P.; Jackson, C. J.; Banwell, M. G.; Willis, A. C. *J. Org. Chem.* **2014**, *79*, 6759.

(15) Buckler, J. N.; Schwartz, B. D.; Banwell, M. G. *Heterocycles* **2017**, *95*, 290.

(16) (a) Boebel, T. A.; Hartwig, J. F. *J. Am. Chem. Soc.* **2008**, *130*, 7534. (b) Sumida, Y.; Kato, T.; Hosoya, T. *Org. Lett.* **2013**, *15*, 2806.

(17) (a) Lan, P.; Banwell, M. G.; Ward, J. S.; Willis, A. C. *Org. Lett.* **2014**, *16*, 228. (b) Lan, P.; Banwell, M. G.; Willis, A. C. *J. Org. Chem.* **2014**, *79*, 2829.

(18) (a) Wick, A. E.; Felix, D.; Steen, K.; Eschenmoser, A. *Helv. Chim. Acta* **1964**, *47*, 2425. (b) Mulzer, J.; Bats, J. W.; List, B.; Opatz, T.; Trauner, D. *Synlett* **1997**, *1997*, 441. (c) Varin, M.; Barré, E.; Iorga, B.; Guillou, C. *Chem. - Eur. J.* **2008**, *14*, 6606. For the application of Claisen rearrangements in closely related settings, see: (d) Tanimoto, H.; Kato, T.; Chida, N. *Tetrahedron Lett.* **2007**, *48*, 6267. (e) Tanimoto, H.; Saito, R.; Chida, N. *Tetrahedron Lett.* **2008**, *49*, 358. (f) Ichiki, M.; Tanimoto, H.; Miwa, S.; Saito, R.; Sato, T.; Chida, N. *Chem. - Eur. J.* **2013**, *19*, 264. (g) Varghese, V.; Hudlicky, T. *Synlett* **2013**, *24*, 369.

(19) Qiu, J. C.; Pradhan, P. P.; Blanck, N. B.; Bobbitt, J. M.; Bailey, W. F. *Org. Lett.* **2012**, *14*, 350.

(20) For an example of the application of this process in the construction of the galanthamine D-ring, see: (a) Feng, Y.; Yu, Z.-X. *J. Org. Chem.* **2015**, *80*, 1952. For a useful review on the Pictet−Spengler reaction, see: (b) Maryanoff, B. E.; Zhang, H.-C.; Cohen, J. H.; Turchi, I. J.; Maryanoff, C. A. *Chem. Rev.* **2004**, *104*, 1431.

(21) For other examples of this sort of hybridization process, see: Banwell, M. G.; Hamel, E.; Hockless, D. C. R.; Verdier-Pinard, P.; Willis, A. C.; Wong, D. J. *Bioorg. Med. Chem.* **2006**, *14*, 4627.

(22) For reviews on methods for generating *cis*-1,2-dihydrocatechols by microbial dihydroxylation of the corresponding aromatics as well as the synthetic applications of these metabolites, see: (a) Hudlicky, T.; Gonzalez, D.; Gibson, D. T. *Aldrichimica Acta* **1999**, *32*, 35. (b) Johnson, R. A. *Org. React.* **2004**, *63*, 117. (c) Hudlicky, T.; Reed, J. W. *Synlett* **2009**, *2009*, 685. (d) Rinner, U. *Comprehensive Chirality*; Carreira, E. M., Yamamoto, H., Eds.; Elsevier: Amsterdam, 2012; Vol 2, p 240. (e) Lewis, S. E. *Chem. Commun.* **2014**, *50*, 2821. (f) Banwell, M. G.; Bolte, B.; Buckler, J. N.; Chang, E. L.; Lan, P.; Taher, E. S.; White, L. V.; Willis, A. C. *J. Proc. Royal Soc.* **2016**, *149*, 34.

(23) White, L. V.; Lan, P.; Schwartz, B. D.; Willis, A. C.; Banwell, M. G. *Aust. J. Chem.* **2015**, *68*, 1467.

(24) Banwell, M. G.; Haddad, N.; Hudlicky, T.; Nugent, T. C.; Mackay, M. F.; Richards, S. L. *J. Chem. Soc., Perkin Trans. 1* **1997**, 1779.

(25) Ley, S. V.; Baeschlin, D. K.; Dixon, D. J.; Foster, A. C.; Ince, S. J.; Priepke, H. W. M.; Reynolds, D. J. *Chem. Rev.* **2001**, *101*, 53.

(26) (a) Miyaura, N.; Suzuki, A. *Chem. Rev.* **1995**, *95*, 2457. (b) Han, F.-S. *Chem. Soc. Rev.* **2013**, *42*, 5270.

(27) For useful discussions of this topic, see: (a) Luttmann, E.; Linnemann, E.; Fels, G. *J. Mol. Model.* **2002**, *8*, 208. (b) Singh, M.; Kaur, M.; Kukreja, H.; Chugh, R.; Silakari, O.; Singh, D. *Eur. J. Med. Chem.* **2013**, *70*, 2013.

(28) Georgiev, D.; Saes, B. W. H.; Johnston, H. J.; Boys, S. K.; Healy, A.; Hulme, A. N. *Molecules* **2016**, *21*, 88.

(29) Hudlicky, T.; Luna, H.; Olivio, H. F.; Andersen, C.; Nugent, T.; Price, J. D. *J. Chem. Soc., Perkin Trans. 1* **1991**, 2907.

(30) For other examples of this type of selective *O*-silylation of allylic rather than homoallylic alcohols, see: Ma, X.; Banwell, M. G.; Willis, A. C. *J. Nat. Prod.* **2013**, *76*, 1514.

(31) (a) Hudlicky, T.; Price, J. D.; Rulin, F.; Tsunoda, T. *J. Am. Chem. Soc.* **1990**, *112*, 9439. (b) Findlay, A. D.; Banwell, M. G. *Org. Lett.* **2009**, *11*, 3160.

(32) Mukhanova, T. I.; Kukushkin, S.; Yu; Ivanov, P.; Yu; Alekseeva, L. M.; Granik, V. G. *Russ. Chem. Bull.* **2007**, *56*, 325.

119

(33) Ellman, G. L.; Courtney, K. D.; Andres, V.; Featherstone, R. M. *Biochem. Pharmacol.* **1961**, *7*, 88.

(34) Cheung, J.; Rudolph, M. J.; Burshteyn, F.; Cassidy, M. S.; Gary, E. N.; Love, J.; Franklin, M. C.; Height, J. J. *J. Med. Chem.* **2012**, *55*, 10282.

(35) Greenblatt, H. M.; Kryger, G.; Lewis, T.; Silman, I.; Sussman, J. L. *FEBS Lett.* **1999**, *463*, 321.

(36) Stark, J.; Powers, R. *J. Am. Chem. Soc.* **2008**, *130*, 535.

(37) Perola, E.; Walters, W. P.; Charifson, P. S. *Proteins: Struct., Funct., Genet.* **2004**, *56*, 235.

(38) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. *J. Med. Chem.* **2006**, *49*, 5912.

(39) Bartolucci, C.; Perola, E.; Pilger, C.; Fels, G.; Lamba, D. *Proteins: Struct., Funct., Genet.* **2001**, *42*, 182.

(40) Zhan, G.; Zhou, J.; Liu, R.; Liu, T.; Guo, G.; Wang, J.; Xiang, M.; Xue, Y.; Luo, Z.; Zhang, Y.; Yao, G. *J. Nat. Prod.* **2016**, *79*, 760.

(41) Still, W. C.; Kahn, M.; Mitra, A. *J. Org. Chem.* **1978**, *43*, 2923.

(42) Pangborn, A. B.; Giardello, M. A.; Grubbs, R. H.; Rosen, R. K.; Timmers, F. J. *Organometallics* **1996**, *15*, 1518.

(43) DENZO−SMN; Otwinowski, Z.; Minor, W. Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology, Vol. 276: Macromolecular Crystallography, Part A*; Carter, C. W., Jr., Sweet, R. M., Eds.; Academic Press: New York, 1997; pp 307−326.

(44) SIR92. Altomare, A.; Cascarano, G.; Giacovazzo, C.; Guagliardi, A.; Burla, M. C.; Polidori, G.; Camalli, M. *J. Appl. Crystallogr.* **1994**, *27*, 435.

(45) Betteridge, P. W.; Carruthers, J. R.; Cooper, R. I.; Prout, K.; Watkin, D. J. *J. Appl. Crystallogr.* **2003**, *36*, 1487.

(46) Sangnoi, Y.; Sakulkeo, O.; Yuenyongsawad, S.; Kanjana-opas, A.; Ingkaninan, K.; Plubrukarn, A.; Suwanborirux, K. *Mar. Drugs* **2008**, *6*, 578.

(47) Schuettelkopf, A. W.; van Aalten, D. M. F. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 1355.

(48) Trott, O.; Olson, A. A. J. *J. Comput. Chem.* **2010**, *31*, 455.

120

Section 2. Identification of new inhibitors - Marinoquinolines

5.6. Introduction

5.6.1. Marinoquinolines

AChE is a drug development target for treating symptoms of AD, as a deficit in Ach is a neurochemical characteristic of patients affected by AD (341, 343, 347, 348). AChE inhibitors prevent the hydrolysis of Ach and compensate for the Ach deficiency in the synapse (341, 343, 347, 348). Currently, three drugs have been approved for the treatment of AD (344, 360). The identification of novel natural products is of great importance for new drugs to combat diseases worldwide (399). Therefore, for treating AD, it is crucial to search for new effective compounds that inhibit AChE (399).

The marine gliding bacterium, *Rapidithrix thailandica,* has been studied for novel secondary metabolites. One compound isolated was marinoquinoline A (MQ-A), which contains a rare 3H-pyrrolo[2,3-c]quinolone system **(Figure 5.8)** (400). This molecule was shown to be a strong inhibitor of AChE (IC_{50} = 4.9 μM), however, it is unknown how the compound interacts with AChE (400–402). MQ-A has been proposed to interact with AChE by a π-π stacking interaction with Trp84 found in the active site. This is similar to the drug, tacrine that shows high structural similarity to MQ-A (333, 349).

Five other derivatives of MQ-A have also been isolated from the bacterium, *Ohtaekwangia kribbensis,* and have been labelled marinoquinolines B–F (MQ-B-F) **(Figure 5.8)** (403). The compounds have a similar structural motif but with a different 'R' group **(Figure 5.8)**. The five compounds have strong activity against the resistant malaria parasite, *Plasmodium falciparum* and also have cytotoxicity against cancer cell lines (403). The compounds, MQ-B-F, have not been tested against AChE, therefore, it is important to investigate if the molecules are effective inhibitors of AChE similar to MQ-A (403).
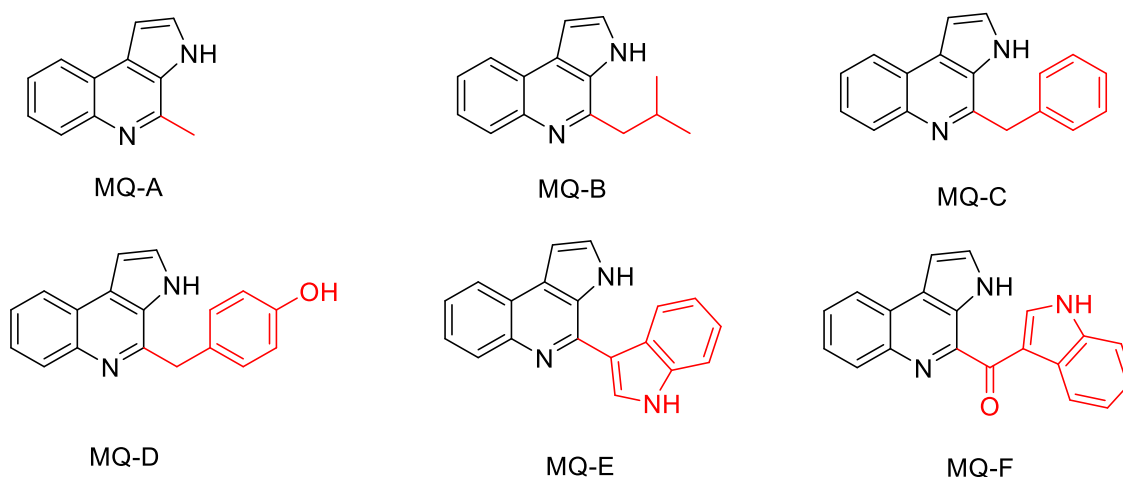
Figure 5.8. The identified marinoquinolines from the species *Rapidithrix thailandica* and *Ohtaekwangia kribbensis* (400, 403). Red shows the variation in the 'R' group between the compounds.

Three other compounds (MQ-F Isomer, MQ-H and lactame) have been synthesised by Martin Banwell's research group at the Research School of Chemistry, The Australian National University, Australia. The compounds were designed to investigate the binding of the marinoquinolines (MQ) to AChE **(Figure 5.9)**. Another compound of interest is Aplidiopsamine A (AAA). This compound is an antiplasmodial alkaloid that has been isolated from the temperate Australian ascidian (*Aplidiopsis confluata*). Similar to compounds MQ-B-F, AAA shows inhibition to resistant strains of *P. falciparum*, with minimal toxicity to human cells. However, the inhibition to AChE has not been characterized for AAA (404, 405).
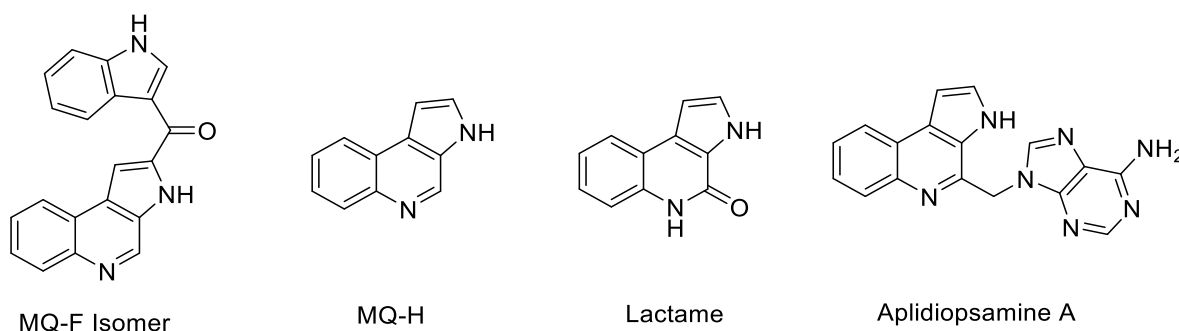
Figure 5.9. Synthesised MQs from Dr. Martin Banwell's research group to study interaction of MQs with AChE (404).

The aim of the project was to understand the interaction between the MQs and AChE. Currently, it is unknown if the identified MQ compounds (MQ-B-F) interact with AChE similar to MQ-A. Molecular docking simulations in combination with inhibition assays were performed to further elucidate the interaction of the compounds with AChE and for the potential development of compounds based on the results.

5.7. Materials and Methods

5.7.1. Synthesis of MQ derivatives

The total synthesis of the MQ compounds was done by members of Dr. Martin Banwell's research group (Australian National University, Research School of Chemistry, Australia)

5.7.2. AChE Inhibition Assay

The compounds were tested for inhibition against AChE as described previously by Sangnoi et al. (400, 406). Compounds were dissolved in DMSO and made up to concentrations ranging from 5 mM to 3.05 µM by half dilutions. 2.5 µl of each compound was added to a solution consisting of 125 µl 5,5'-dithiobis-(2-nitrobenzoic acid) (1.5 mM), 72.5 µl Tris buffer (15 mM, pH 8.0) and 25 µl acetylthiocholine iodide (150 µM). Enzyme activity was followed by the addition of 25 µl AChE at 0.03 µM (*Electrophorus electricus*, Type V-S, EC 3.1.1.7) measuring at 412 nm using the Epoch microplate spectrophotometer (BioTek). The assays were repeated in triplicates and the rate was calculated using the first two minutes. Prism v6.0 (GraphPad Software, USA.) was used to calculate the IC_{50}.

5.7.3. Molecular Docking Studies

The three-dimensional coordinates of each compound were generated with the PRODRG server (http://davapc1.bioch.dundee.ac. uk/cgi-bin/prodrg) (407). The MQs and AAA were docked into the structure of human AChE (**4BDT**) using autodock Vina v1.1.2 with tacrine removed from the active site (408).

5.8. Results and Discussion

5.8.1. Inhibition studies of MQ

After testing the 10 MQs compounds against AChE (**Figures 5.8, 5.9**), it was observed that six of the compounds inhibited the enzyme AChE with IC_{50}'s between 4.7 µM and 42.7 µM (**Table 5.2**). Three compounds (MQ-C, MQ-D and MQ-H) were not as effective inhibitors of AChE compared to MQ-A. There was no statistically significant difference in inhibition for MQ-B (4.7 µM) and MQ-E (5.8 µM) compared to MQ-A (5.5 µM). Four compounds (MQ-F, MQ-F isomer, lactame, AAA) did not show any inhibition towards AChE at the highest concentration tested.

The results from the inhibition assays indicate the MQ structures containing an extended 'R' group chain or a bulky 'R' group (MQ-C, MQ-D, MQ-F, MQ-F Isomer, AAA) are poor

inhibitors of AChE. This suggests there is little room in the structure of AChE for the 'R' group to occupy when the group is bulky or when the 'R' group carbon chain is extended (C0 to C1). MQ-E with a bulky 'R' group has similar inhibition to MQ-A, which contradicts the trend, however, the carbon chain making up the 'R' group is not increased in length (C0) as in MQ-C and MQ-D (C1 in length). The introduction of hydrophilic nature into the MQ structure (MQ-D) results in poor binding. This suggests the pocket in AChE where the 'R' group interacts with is generally hydrophobic in nature. The MQ-F isomer compound has the 'R' group extending from a different site in the molecule and has no detectable inhibition towards AChE. This indicates the 'R' group extends into a region of the AChE structure that is limited in space, therefore, the inhibitor cannot bind similar to the other structures. Removal of the 'R' group (MQ-H) removes a large proportion of inhibition towards AChE (8-fold) and shows the 'R' group interaction is important for inhibition.

Table 5.2. IC_{50} results from the 10 MQs tested against AChE along with the predicted docking binding energy from autodock Vina.

Compound	IC_{50} (µM)	Docking Binding Energy (kcal/mol) with 4BDT
MQ-A	5.5 ± 0.8	-8.8
MQ-B	4.7 ± 1.2	-9.6
MQ-C	8.1 ± 1.5	-10.5
MQ-D	31.8 ± 3.1	-11.0
MQ-E	5.8 ± 0.8	-12.1
MQ-F	$IC_{50} > 31.25$ (No inhibition observed at solubility limit)	-11.9
MQ-F Isomer	$IC_{50} > 125$ (No inhibition observed at solubility limit)	-10.9
MQ-H	42.7 ± 3.4	-8.3
lactame	$IC_{50} > 125$ (No inhibition observed at solubility limit)	-9.1
AAA	$IC_{50} > 72.5$ (No inhibition observed at solubility limit)	-10.2

5.8.2. Docking studies of MQ

Molecular docking simulations were performed with autodock Vina to elucidate the binding site of the MQ derivatives in the structure of AChE and for future direction in designing stronger binding inhibitors. Docking simulations of MQ-A were previously done in 2014 by Stoddard et al (349). This reported the compound, MQ-A, interacted with AChE

by π-π stacking interactions with Trp84 in the choline binding site (349). This residue is important in AChE to interact with the positive charge character of the Ach substrate (333).

Molecular docking simulations showed all 10 compounds binding into the active site of AChE. Nine compounds interact with Trp84 forming the π-π stacking interaction in all low-energy states **(Table 5.2)**. The three ring structure is stacked against the indole ring of Trp84 and sandwiched between Phe330, providing a strong interaction **(Figure 5.10)**. The MQ derivative, lactame, does not appear to stack with Trp84 or Phe330 in the docking studies.
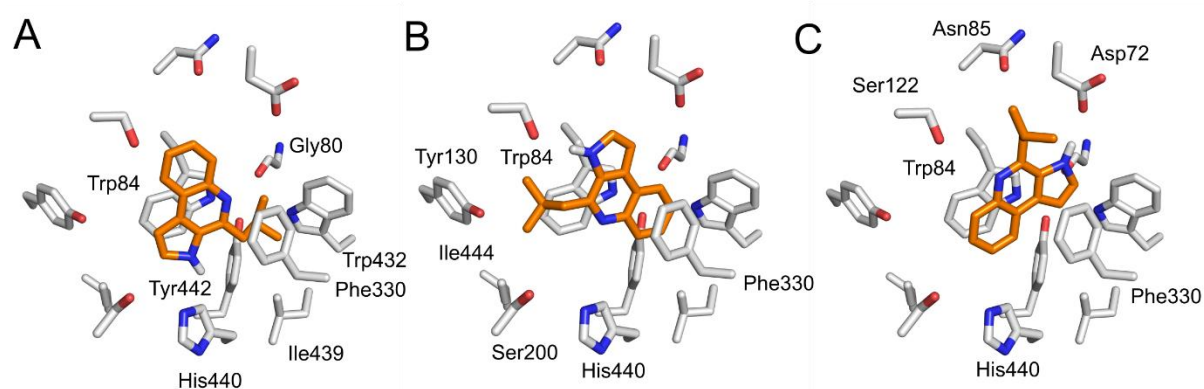
Figure 5.10. MQ-B docked into the structure of human acetylcholinesterase (4BDT) showing the main interaction between Trp84 and Phe330. The three different orientations of the 'R' group are shown in panels A,B and C.

Similar to tacrine, the nitrogen of the indole ring hydrogen bonds to the main chain carbonyl oxygen of His440 (333). Without water molecules in the docking simulations, it is unknown if a hydrogen bond is formed with the amino nitrogen of the MQs and the water molecule that is observed in the structure of tacrine bound to AChE.

From the molecular docking simulations, the 'R' group can orient into two pockets while retaining the His440 hydrogen bond. One pocket (A) is lined with Gly80, Trp432, Ile439 and Tyr442. The other potential binding pocket (B) is towards Tyr130, Ile444 and Ser200. A third pocket (C) that is commonly observed in the docking experiments as the lowest energy position is towards Asp72, Asn85 and Ser122. However, binding into this pocket, the hydrogen bond between His440 and MQ is lost **(Figure 5.10)**. It is difficult to rationalize the correct orientation of the 'R' group from the docking simulations.

Pocket C is unlikely to be the correct pocket for the 'R' group to interact with. In the tacrine structure, this side of the pocket is where a hydrogen bonding network is formed **(Figure 5.11)** (333). The 'R' group in this direction would disrupt this, however, the compound

could nullify the loss of the hydrogen bonding interaction with new interactions involving the 'R' group.
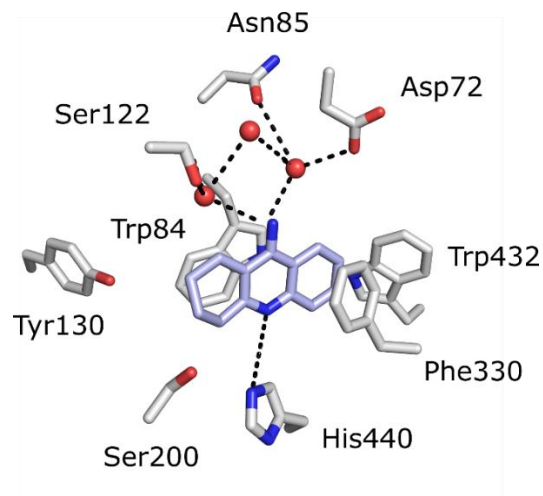
Figure 5.11. Tacrine co-crystallised into the structure of Pacific electric ray AChE (1ACJ) showing the main interaction between Trp84, His440 and tacrine (333).

Pocket A and B are similar in residue makeup but the increase in the hydrophobic nature of pocket A would match the results with compounds MQ-C and MQ-D. The comparison between the compounds suggests an increase in hydrophobic character produces a tighter binding inhibitor. Future experiments with soaking the compounds into AChE would be required to elucidate the binding site of the MQs in full detail.

The 'R' group orientating into pocket A could explain why MQ-F and AAA do not bind to AChE (at the concentrations measured). This is due to the narrow space of pocket A. The reduced flexibility with MQ-F and the bulkiness of AAA would prevent the correct orientation of the main ring for interaction with Trp84 and His440, which will result in poor binding.

This is further supported by the difference between MQ-E and MQ-F. MQ-E binds at similar affinity to MQ-A, but MQ-F does not bind to AChE at the concentrations measured, despite similar structures of the compounds. This suggests the increase in length in the MQ-F 'R' group compared to MQ-E (C0-C1) or the decrease in the hydrophobicity character does not allow for the 'R' group to favourably orientate into pocket A, resulting in no observed inhibition.

It is difficult to rationalize why MQ-E binds at similar affinity to MQ-A given the large 'R' group. It is possible the compound has additional interactions with residues in pocket A, however, this is not observed in the docking simulations. Lactame has a carbonyl in the

main ring structure and in the docking experiments, the compound binds into a similar site as the MQs. The carbonyl in this compound distorts the ring resulting in the disruption of the key π-π stacking interactions. The MQ-F Isomer compound with the 'R' group at an alternative position does not bind to AChE at the concentrations measured. The large 'R' group orientates into an unfavourable position in the structure, which will result in poor binding.

The values of the binding energies varied significantly and did not match the experimental data. Four compounds (MQ-F, MQ-F Isomer, lactame, AAA) that showed no inhibition in the assays, showed a significantly tighter binding energy than MQ-A in the molecular docking studies. This falsely indicates inhibition by the compounds and shows the false positives from the docking results (binding into pocket C). This highlights the importance of experimental data over molecular docking studies in designing derivatives.

MQ-A and tacrine are similar in structure and binding affinity that could suggest MQ-A can be used as a drug against AD. However, as the structures are similar in aromatic nature, this could cause issues with side effects. The hepatotoxicity of tacrine resulted in the discontinued use of the drug, which could be a similar issue for MQ-A (369, 370). MQ-A would need to be extensively tested for selectivity with BChE and other enzymes to identify if MQ-A behaves similarly to tacrine in being a non-selective cholinesterase inhibitor and having similar affinity for both isoforms of AChE. The compound may not be suitable for treatment if the selectivity is similar to tacrine.

The testing of the inhibitors designed to target human AChE for both projects involved AChE from *Electrophorus electricus*. The findings for the human-based inhibitors on *E. electricus* AChE can be applied to human AChE due to the similarity in structure. The RMSD of the two proteins is 0.63, highlighting the close similarity of the proteins, with no changes in the residues making up the active site and the PAS. It should be noted that as it was observed in chapter four, subtle differences in structure can have a major effect, therefore, the IC_{50} calculated for *E. electricus* AChE may not be the same for human AChE. There may be potential differences in inhibitor sensitivity between AChE's from different species, however, differences between the AChE structures are negligible between species, suggesting to not affect IC_{50}'s significantly (409).

Chapter 6. Conclusions and Future Work

6.1. Conclusions

In this thesis, four chapters of work are described that mainly focussed on dipteran carboxylesterases. Chapter 2 investigated the higher-order oligomeric species present in the thermostable protein *Lc*αE7-4a after a directed evolution experiment towards greater thermostability. This work highlights the importance of oligomeric species during evolution for maintaining protein stability. The chapter provides a significant case study for the investigation of stability associated with oligomerisation and advances our understanding of the evolution of new oligomers through point mutations. It also reinforces the nature of structure/activity/stability trade-offs, which have been well illustrated throughout enzyme evolution and that oligomerisation impacts evolution, and in this case, was not selectively neutral.

Chapter 3 describes the structure determination and kinetic assays of Esterase-6 from *D. melanogaster,* an enzyme important for reproductive success in *Drosophila* species. This section shows the importance of new insect carboxylesterase structures to understand protein function, as homology models of EST6 were not informative with regard to the enzyme's active site. The structure is the first of an odorant degrading enzyme which is able to provide a probable function of esterase-6. This chapter also shows the enzyme is not directly involved in the degradation of long chain pheromones, which had been strongly believed.

Chapter 4 is focussed on the characterization of several orthologs from Diptera to investigate the ability of the enzymes to evolve through the incorporation of a point mutation to catalyse the hydrolysis of organophosphates. This work shows a large variation in organophosphate binding and turnover, despite the similar sequence similarity and structure between the proteins. It suggests neutral drift plays an important role for the evolution of qualitative resistance in Diptera and emphasises that there are limited evolutionary routes from which qualitative resistance to organophosphates can evolve.

Chapter 5 describes the testing of inhibitors in targeting AChE. This chapter demonstrates the importance of docking simulations in combination with rational design for the development of new drugs. This chapter also describes the pitfalls with molecular docking simulations giving false positive results.

Overall, this body of work has answered fundamental questions on insect CBEs and also expanded the knowledge on this superfamily. EST6 gives a small insight into the structural diversity within insect CBEs, as this protein is completely unrelated to the functions of the other solved CBE proteins. This sheds light on the potential structural differences (alternative active site entries) between the 14 clades and the difficulties in predicting function and structure from homology models alone that have been used previously for insect CBEs.

The versatility of insect CBEs to break down multiple substrates has been highlighted in chapters two, three and four, especially with the enzyme, EST6. EST6 is able to hydrolyse multiple functionally diverse esters from small short-chain esters (methyl propionate) to larger complex esters (E2-hexenyl acetate). This strongly relates to animal CBEs, where the human liver CBE is important to hydrolyse and biotransform multiple functionally diverse ester substrates.

Chapter four highlights how malleable the active sites of CBEs are with small sequences changes affecting the catalytic machinery resulting in the ability to break down new substrates. A single point mutation (Gly137Asp) is able to convert the function of a CBE towards a functionally different substrate, completing changing the kinetics of the CBE from native function. This is observed throughout the multiple CBEs tested and in the CBE literature showcasing the ability of CBEs to adapt to new selection pressures. The CBE active site makeup and subtle structural differences between CBEs have also shown to be vital in this thesis (chapter four) for a new function to be favourable. The minor active site differences between closely related CBEs shown in chapter four add to the complications in predicting insecticide resistance and protein evolution.

Subtle structural differences in CBEs are further highlighted in chapter two and five, where a few mutations can change the oligomerisation state of an insect CBE or for new inhibitors to be effective. This shows the plasticity and tolerance of CBEs to mutations with the ever going ability to adapt to new roles and diverse functions that are found throughout past work on insects proteins.

The work has provided a significant contribution to fields of structural biology with the fourth insect CBE structure and answered key questions to the function of an important CBE involved in odorant degradation. Contributions have also been made to protein evolution with a greater understanding for the formation of oligomeric species and evidence

has been provided for the limitations for the evolution of new function involved with insecticide resistance. Additionally, this thesis also provides the framework to investigate new questions on the evolution of insecticide resistance and to further understand structural differences throughout the CBE family.

6.2. Future Work

6.2.1. Chapter 2.

Future experiments to investigate the evolution of quaternary structure in insect CBEs should focus on increasing the level of the higher order oligomeric species. The ratio between the oligomeric species is approximately 60:30:10 between the monomer, dimer and tetramer species. Directed evolution experiments towards thermostability could be continued to further amplify the relative amounts of the higher order species. This could be useful in a number of ways, as the oligomeric species are more stable, the tetramer/dimer species could have increase longevity *in vivo* if *Lc*αE7 is used as a bio-scavenger (55). This could also lead to future structural studies using crystallization to elucidate the interface of the tetramer and to confirm the dimer interface. Cross linking experiments could also be used to further support the results.

The ratio of dimer to monomer could also be increased using the protein *Md*αE7. This protein has been identified in this work (Chapter 4) to elute from SEC as a dimer species. Mapping a homology model of this protein to the proposed dimer interface of *Lc*αE7 generated from the PISA server could locate surface mutations located in the interface to result in *Lc*αE7 being essentially a dimer.

6.2.2. Chapter 3.

With the structure of EST6 being solved in this thesis, there are a number of additional pathways available to extend this project. There are numerous mutations that have been found in wild-type EST6 including the two identified allozymes 'Fast' and 'Slow' EST6 (410). These allozymes have been suggested to have major differences in stability/activity and response to inhibitors (410). Mapping the mutations found in the EST6 population to the structure and recombinant testing of the mutations could result in a further understanding of the natural mutations and allozymes. As the recombinant characterization

of EST6 has not been performed in great detail due to the low expression, this could be an opportunity to characterize the allozymes with high purity enzyme. Preliminarily work suggests that the 'fast' and 'slow' mutations are located at positions remote from the active site.

The structure of EST6 could be used to understand the other member of the β-cluster, EST7. This protein has been suggested to be a "dying" enzyme, with many non-catalytic alleles of this protein being observed in *D. melanogaster* species (152). By recombinant expression of the enzyme or generating homology models, further analysis could reveal possible reasons into why the enzyme is not under strong evolutionary selective pressure. The substrate range and stability could also give possible suggestions into the enzyme's function. This technique could also be adapted to other members of the β-cluster in other species to understand the microevolution that has been observed.

The contributions of the disulfide bonds in EST6 to protein folding have been shown to be essential for the recombinant soluble expression of EST6. The structure of EST6 shows a high degree of overlap with *Lc*αE7, yet in EST6, a disulfide bond is formed between Cys65 and Cys84 while at the analogous positions in *Lc*αE7 (Ser95 and Cys111), there is no disulfide bond. At position 95 in *Lc*αE7, a cysteine is replaced with a serine. To introduce additional stability to *Lc*αE7, a disulfide bond could possibly be introduced into *Lc*αE7 at this position. Likewise, the effect of the disulfide bonds on stability/activity of EST6 could be studied by mutation of one of these cysteine residues to serine.

6.2.3. Chapter 4.

To further understand the structural differences between the proteins, crystal structures could be solved of the orthologs. In combination with inhibitor soaking, this could further elucidate the differences between the proteins and could explain the observed changes with OP binding and activity between the orthologs. To also understand the differences in OP binding, different OP insecticides could be assayed against the orthologs. This could result in OP affinity differences between the orthologs that could give more insight into the proteins and understand if the orthologs can be involved in the qualitative resistance mechanism.

There is a large variation in the oligomeric state between the orthologs, despite the close sequence identity. Initial work has estimated the oligomeric state of the orthologs, but this could be investigated by detailed characterisation of the oligomeric species including

stability and activity data to coincide with the work in Chapter 2. With one of the orthologs (*Cs*αE7) there appears to be a large increase in OP turnover with the introduction of Gly137Asp mutation. Sequencing of *C. stygia* flies in the wild could investigate if the protein has the Gly137Asp mutation and could be another example of insecticide resistance.

6.2.4. Chapter 5.

No derivatives of tacrine have been observed with modifications at the 'R' group position of MQ. Generally, groups have focussed on extending from the amine group of tacrine to reach the PAS site of AChE for tighter binding and better inhibition (411, 412). Future work could focus on tacrine derivatives that extend out from the 'R' position of MQ-A. Similarly, tighter binding MQs could be designed with the extension from the nitrogen to reach the PAS of AChE.

6.2.5. The Future of Carboxylesterase Research.

Overall, there are still many questions this thesis has not answered, and there are more research areas involving CBEs to pursue. Currently, four of the fourteen insect CBE clades have been structurally identified. Since all four CBEs are structurally unique, it is vital to understand structural features that make up each member of a clade to find out the structural diversity in this superfamily. Furthermore, there are two clades with unknown function and should be studied in more detail. This is important to understand the functions of the unknown clades to know the variety of roles CBEs plays in insect species. Future genomic projects and sequencing information will also help to comprehend the fourteen clades and could further separate the CBEs to better characterize this superfamily. This could not only help to understand the evolution of CBEs, but can identify potential new targets for insecticides.

Juvenile hormone esterase is separated in the clades between Lepidopteran and Diptera. Studying the functional and structural differences between the two juvenile hormone esterases are important to know the differences in reproduction and evolution between the two orders and why they are separated in phylogenetic studies.

There is also no current understanding of what determines a range of substrates for a CBE. Further testing the substrate specificity for several CBEs could help to find out the full range of esters that can be turned over and structural features that determine a substrate

range for a CBE. This could play a role for biocatalysts projects involving transesterification or in developing CBEs for detoxifying certain compounds.

References

1. Redinbo MR, Bencharit S, Potter PM (2003) Human carboxylesterase 1: from drug metabolism to drug discovery. *Biochem Soc Trans* 31(3):620–624.

2. Newcomb RD, et al. (1997) A single amino acid substitution converts a carboxylesterase to an organophosphorus hydrolase and confers insecticide resistance on a blowfly. *Proc Natl Acad Sci U S A* 94(14):7464–7468.

3. Myers M, Richmond RC, Oakeshott JG (1988) On the origins of esterases. *Mol Biol Evol* 5(2):113–119.

4. Ollis DL, et al. (1992) The α/β hydrolase fold. *Protein Eng* 5(3):197–211.

5. Satoh T, Hosokawa M (2006) Structure, function and regulation of carboxylesterases. *Chem Biol Interact* 162(3):195–211.

6. Wheelock CE, et al. (2008) Applications of carboxylesterase activity in environmental monitoring and toxicity identification evaluations (TIEs). *Rev Environ Contam Toxicol* 195:117–178.

7. Orengo C, et al. (1997) CATH - a hierarchic classification of protein domain structures. *Structure* 17(8):1093–1109.

8. Zschunke F, et al. (1991) cDNA cloning and characterization of human monocyte/macrophage serine esterase-1. *Blood* 78(2):506–512.

9. Cygler M, Schrag JD, Ergan F (1992) Advances in structural understanding of lipases. *Biotechnol Genet Eng Rev* 10(1):143–184.

10. Benning MM, Kuo JM, Raushel FM, Holden HM (1995) Three-dimensional structure of the binuclear metal center of phosphotriesterase. *Biochemistry* 34(25):7973–7978.

11. Newcomb RD, Gleeson DM, Yong CG, Russell RJ, Oakeshott JG (2005) Multiple mutations and gene duplications conferring organophosphorus insecticide resistance have been selected at the Rop-1 locus of the sheep blowfly, *Lucilia cuprina*. *J Mol Evol* 60(2):207–220.

12. Jeon JH, et al. (2011) Novel metagenome-derived carboxylesterase that hydrolyzes β-lactam antibiotics. *Appl Environ Microbiol* 77(21):7830–7836.

13. Marshall SDG, Putterill JJ, Plummer KM, Newcomb RD (2003) The carboxylesterase gene family from *Arabidopsis thaliana*. *J Mol Evol* 57(5):487–500.

14. Montella IR, Schama R, Valle D (2012) The classification of esterases: An important gene family involved in insecticide resistance - a review. *Mem Inst Oswaldo Cruz* 107(4):437–449.

15. Stok JE, et al. (2004) Identification, expression, and purification of a pyrethroid-hydrolyzing carboxylesterase from mouse liver microsomes. *J Biol Chem* 279(28):29863–29869.

16. Gupta RC, Dettbarn WD (1993) Role of carboxylesterases in the prevention and potentiation of N-methylcarbamate toxicity. *Chem Biol Interact* 87(1–3):295–303.

17. Casida JE, Quistad GB (2004) Organophosphate toxicology: safety aspects of nonacetylcholinesterase secondary targets. *Chem Res Toxicol* 17(8):983–998.

18. Wheelock CE, Shan G, Ottea J (2005) Overview of carboxylesterases and their role in the metabolism of insecticides. *J Pestic Sci* 30(2):75–83.

19. Qiao C, Cui F, Yan S (2009) Structure, function and applications of carboxylesterases from insects for insecticide resistance. *Protein Pept Lett* 16(10):1181–1188.

20. Aldridge WN, Reiner E (1972) Enzyme inhibitors as substrates: interactions of esterases with esters of organophosphorus and carbamic acids. *Frontiers of Biology (Vol. 26)* (North-Holland, Amsterdam), p 328.

21. Aldridge WN (1953) Serum esterases. II. An enzyme hydrolysing diethyl p-nitrophenyl phosphate (E600) and its identity with the A-esterase of mammalian sera. *Biochem J* 53(1):117–124.

22. Aldridge WN, Davison AN (1953) The mechanism of inhibition of cholinesterases by organophosphorus compounds. *Biochem J* 55(5):763–766.

23. Reiner E (1993) Recommendations of the IUBMB nomenclature committee: comments concerning classification and nomenclature of esterases hydrolysing organophosphorus compounds. *Chem Biol Interact* 87(1–3):15–16.

24. Walker CH (1993) The second international meeting on esterases interacting with organophosphorus compounds. The classification of esterases which hydrolyse organophosphates: recent developments. *Chem Biol Interact* 87(1):17–24.

25. Holmes RS, Masters CJ (1967) The developmental multiplicity and isoenzyme status of cavian esterases. *Biochim Biophys Acta - Enzymol* 132(2):379–399.

26. Coates PM, Mestriner MA, Hopkinson DA (1975) A preliminary genetic interpretation of the esterase isozymes of human tissues. *Ann Hum Genet* 39(1):1–20.

27. Healy MJ, Dumancic MM, Oakeshott JG (1991) Biochemical and physiological studies of soluble esterases from *Drosophila melanogaster*. *Biochem Genet* 29(7–8):365–388.

28. Hemingway J, Karunaratne SHPP (1998) Mosquito carboxylesterases: A review of the molecular biology and biochemistry of a major insecticide resistance mechanism. *Med Vet Entomol* 12(1):1–12.

29. Georghiou GP, Pasteur N, Hawley MK (1980) Linkage relationships between organophosphate resistance and a highly active esterase-B in *Culex quinquefaciatus* from California. *J Econ Entomol* 73(2):301–305.

30. Raymond M, Callaghan A, Fort P, Pasteur N (1991) Worldwide migration of amplified insecticide resistance genes in mosquitoes. *Nature* 350(6314):151–153.

31. Vaughan A, Rodriguez M, Hemingway J (1995) The independent gene amplification of electrophoretically indistinguishable B esterases from the insecticide-resistant mosquito *Culex quinquefasciatus*. *Biochem J* 305(2):651–658.

32. Oakeshott JG, Claudianos C, Campbell PM, Newcomb RD, Russell RJ (2005) Biochemical genetics and genomics of insect esterases. *Comprehensive Molecular Insect Science*, eds Gilbert L, Iatrou K, Gill S (Elsevier, San Diego, CA), pp 309–381.

33. Zouros E, van Delden W (1982) Substrate-preference polymorphism at an esterase locus of *Drosophila mojavensis*. *Genetics* 100(2):307–314.

34. Field LM, Devonshire AL (1998) Evidence that the E4 and FE4 esterase genes responsible for insecticide resistance in the aphid *Myzus persicae* (Sulzer) are part of a gene family. *Biochem J* 330(1):169–173.

35. Field LM, Williamson MS, Moores GD, Devonshire AL (1993) Cloning and analysis of the esterase genes conferring insecticide resistance in the peach-potato aphid, *Myzus persicae* (Sulzer). *Biochem J* 294(2):569–574.

36.   Thorpe JP (1983) Enzyme variation, genetic distance and evolutionary divergence in relation to levels of taxonomic separation. *Protein Polymorphism: Adaptive and Taxonomic Significance*, eds Oxford GS, Rollinson D (Academic Press, London, England), pp 131–152.

37.   Testa B, Krämer SD (2007) The biochemistry of drug metabolism - an introduction part 3. Reactions of hydrolysis and their enzymes. *Chem Biodivers* 4(9):2031–2122.

38.   Claudianos C, et al. (2006) A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. *Insect Mol Biol* 15(5):615–636.

39.   Oakeshott JG, et al. (2010) Metabolic enzymes associated with xenobiotic and chemosensory responses in *Nasonia vitripennis*. *Insect Mol Biol* 19(Suppl. 1):147–163.

40.   Bornscheuer UT (2002) Microbial carboxyl esterases: classification, properties and application in biocatalysis. *FEMS Microbiol Rev* 26(1):73–81.

41.   Tatusov RL, Koonin E V, Lipman DJ (1997) A genomic perspective on protein families. *Science* 278(5338):631–637.

42.   Oakeshott JG, van Papenrecht EA, Boyce TM, Healy MJ, Russell RJ (1993) Evolutionary genetics of *Drosophila* esterases. *Genetica* 90(2–3):239–268.

43.   Oakeshott JG, Claudianos C, Russell RJ, Robin GC (1999) Carboxyl/cholinesterases: a case study of the evolution of a successful multigene family. *Bioessays* 21(12):1031–1042.

44.   Hotelier T, Nègre V, Marchot P, Chatonnet A (2010) Insecticide resistance through mutations in cholinesterases or carboxylesterases: data mining in the ESTHER database. *J Pestic Sci* 35(3):315–320.

45.   Hotelier T, et al. (2004) ESTHER, the database of the α/β-hydrolase fold superfamily of proteins. *Nucleic Acids Res* 32:D145–D147.

46.   Nardini M, Dijkstra BW (1999) α/β Hydrolase fold enzymes: the family keeps growing. *Curr Opin Struct Biol* 9(6):732–737.

47.   Cygler M, et al. (1993) Relationship between sequence conservation and three-dimensional structure in a large family of esterases, lipases, and related proteins. *Protein Sci* 2(3):366–382.

48.   Heikinheimo P, Goldman A, Jeffries C, Ollis DL (1999) Of barn owls and bankers: a lush variety of α/β hydrolases. *Structure* 7(6):R141–R146.

49.   Araç D, et al. (2007) Structures of neuroligin-1 and the neuroligin-1/neurexin-1β complex reveal specific protein-protein and protein-Ca 2+ interactions. *Neuron* 56(6):992–1003.

50.   Aharoni A, et al. (2005) The "evolvability" of promiscuous protein functions. *Nat Genet* 37(1):73–76.

51.   Khersonsky O, Roodveldt C, Tawfik DS (2006) Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr Opin Chem Biol* 10(5):498–508.

52.   Karve R, et al. (2010) Evolutionary lineages and functional diversification of plant hexokinases. *Mol Plant* 3(2):334–346.

53.   Sussman JL, et al. (1991) Atomic structure of acetylcholinesterase from *Torpedo californica*: a prototypic acetylcholine-binding protein. *Science* 253(5022):872–879.

54.   Wogulis M, et al. (2006) Structural studies of a potent insect maturation inhibitor bound to the juvenile hormone esterase of *Manduca sexta*. *Biochemistry* 45(13):4045–4057.

55.   Jackson CJ, et al. (2013) Structure and function of an insect α-carboxylesterase

(αEsterase7) associated with insecticide resistance. *Proc Natl Acad Sci U S A* 110(25):10177–82.

56. Oakeshott JG, et al. (2001) An episode of accelerated amino acid change in *Drosophila* esterase-6 associated with a change in physiological function. *Genetica* 110(3):231–244.

57. Taylor P, Radić Z (1994) The cholinesterases: from genes to proteins. *Annu Rev Pharmacol Toxicol* 34(1):281–320.

58. Estrada-Mondaca S, Fournier D (1998) Stabilization of recombinant *Drosophila* acetylcholinesterase. *Protein Expr Purif* 12(2):166–172.

59. Massoulié J, et al. (1999) The polymorphism of acetylcholinesterase: post-translational processing, quaternary associations and localization. *Chem Biol Interact* 119–120:29–42.

60. Callaghan A, Guillemaud T, Makate N, Raymond M (1998) Polymorphisms and fluctuations in copy number of amplified esterase genes in *Culex pipiens* mosquitoes. *Insect Mol Biol* 7(3):295.

61. Li T, et al. (2015) A G-protein-coupled receptor regulation pathway in cytochrome P450-mediated permethrin-resistance in mosquitoes, *Culex quinquefasciatus*. *Sci Rep* 5:17772.

62. Field LM, Blackman RL (2003) Insecticide resistance in the aphid *Myzus persicae* (Sulzer): chromosome location and epigenetic effects on esterase gene expression in clonal lineages. *Biol J Linn Soc* 79(1):107–113.

63. Oppold A-M, Müller R (2017) Epigenetics: a hidden target of insecticides. *Advances in Insect Physiology*, eds Jurenka R, Verlinden H (Academic Press), p 342. 1st Ed.

64. Redinbo MR, Potter PM (2005) The mammalian carboxylesterases: from drug targets to protein therapeutics. *Drug Discov Today* 10(5):313–325.

65. Satoh T, et al. (2002) Current progress on esterases: from molecular structure to function. *Drug Metab Dispos* 30(5):488–493.

66. Satoh T, Hosokawa M (1998) The mammalian carboxylesterases: from molecules to functions. *Biomed Res* 38(1):257–288.

67. Hosokawa M, et al. (1995) Interindividual variation in carboxylesterase levels in human liver microsomes. *Drug Metab Dispos* 23(10):1022–1027.

68. Sogorb MA, Vilanova E (2002) Enzymes involved in the detoxification of organophosphorus, carbamate and pyrethroid insecticides through hydrolysis. *Toxicol Lett* 128(1–3):215–228.

69. Russell RJ, et al. (2011) The evolution of new enzyme function: lessons from xenobiotic metabolizing bacteria versus insecticide-resistant insects. *Evol Appl* 4(2):225–248.

70. Aranda J, et al. (2014) The catalytic mechanism of carboxylesterases: a computational study. *Biochemistry* 53(36):5820–5829.

71. Zera AJ, Sanger T, Hanes J, Harshman L (2002) Purification and characterization of hemolymph juvenile hormone esterase from the cricket, *Gryllus assimilis*. *Arch Insect Biochem Physiol* 49(1):41–55.

72. Devonshire AL, et al. (2003) Kinetic efficiency of mutant carboxylesterases implicated in organophosphate insecticide resistance. *Pestic Biochem Physiol* 76(1):1–13.

73. Kamita SG, et al. (2003) Juvenile hormone (JH) esterase: why are you so JH specific? *Insect Biochem Mol Biol* 33(12):1261–1273.

74. Arpigny JL, Jaeger KE (1999) Bacterial lipolytic enzymes: classification and properties.

Biochem J 343(1):177–183.

75. Fojan P, Jonson PH, Petersen MTN, Petersen SB (2000) What distinguishes an esterase from a lipase: a novel structural approach. *Biochimie* 82(11):1033–1041.

76. Campbell PM, et al. (2003) Developmental expression and gene/enzyme identifications in the alpha esterasa gene cluster of *Drosophila melanogaster*. *Insect Mol Biol* 12(5):459–471.

77. Ranson H, et al. (2002) Evolution of supergene families associated with insecticide resistance. *Science* 298(5591):179–181.

78. Jochens H, et al. (2009) Converting an esterase into an epoxide hydrolase. *Angew Chemie - Int Ed* 48(19):3532–3535.

79. Strode C, et al. (2008) Genomic analysis of detoxification genes in the mosquito *Aedes aegypti*. *Insect Biochem Mol Biol* 38(1):113–123.

80. Ramsey JS, et al. (2010) Comparative analysis of detoxification enzymes in *Acyrthosiphon pisum* and *Myzus persicae*. *Insect Mol Biol* 19(Suppl. 2):155–164.

81. Robin C, Russell RJ, Medveczky KM, Oakeshott JG (1996) Duplication and divergence of the genes of the alpha-esterase cluster of *Drosophila melanogaster*. *J Mol Evol* 43(3):241–252.

82. Matsuzaki T, Yokokura T, Mutai M, Tsuruo T (1988) Inhibition of spontaneous and experimental metastasis by a new derivative of camptothecin, CPT-11, in mice. *Cancer Chemother Pharmacol* 21(4):308–312.

83. Bencharit S, et al. (2003) Crystal structure of human carboxylesterase 1 complexed with the Alzheimer's drug tacrine: from binding promiscuity to selective inhibition. *Chem Biol* 10(4):341–349.

84. Bencharit S, et al. (2002) Structural insights into CPT-11 activation by mammalian carboxylesterases. *Nat Struct Mol Biol* 9(5):337.

85. Nicolet Y, Lockridge O, Masson P, Fontecilla-Camps JC, Nachon F (2003) Crystal structure of human butyrylcholinesterase and of its complexes with substrate and products. *J Biol Chem* 278(42):41141–41147.

86. Kryger G, et al. (2000) Structures of recombinant native and E202Q mutant human acetylcholinesterase complexed with the snake-venom toxin fasciculin-II. *Acta Crystallogr Sect D Biol Crystallogr* 56(11):1385–1394.

87. Allgardsson A, et al. (2016) Structure of a prereaction complex between the nerve agent sarin, its biological target acetylcholinesterase, and the antidote HI-6. *Proc Natl Acad Sci* 113(20):5514–5519.

88. Haviv H, et al. (2005) Crystal packing mediates enantioselective ligand recognition at the peripheral site of acetylcholinesterase. *J Am Chem Soc* 127(31):11029–11036.

89. Giot L, et al. (2003) A protein interaction map of *Drosophila melanogaster*. *Science* 302(5651):1727–1736.

90. Ge H, et al. (2004) A map of the interactome network of the metazoan *C. elegans*. *Science* 303(5657):540–543.

91. Gavin A, et al. (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature* 440(7084):631–636.

92. Goodsell DS, Olson AJ (2000) Structural symmetry and protein function. *Annu Rev*

Biophys Biomol Struct 29(1):105–153.

93. Krogan NJ, et al. (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 440(7084):637–643.

94. Janin J, Bahadur RP, Chakrabarti P (2008) Protein-protein interaction and quaternary structure. *Q Rev Biophys* 41(2):133–180.

95. Nelson DL, Lehninger AL, Cox MM (2008) *Lehninger principles of biochemistry* (W.H. Freeman, New York, NY). 5th Ed.

96. Levy ED, Pereira-Leal JB, Chothia C, Teichmann SA (2006) 3D complex: a structural classification of protein complexes. *PLoS Comput Biol* 2(11):1395–1406.

97. Kühner S, et al. (2009) Proteome organization in a genome-reduced bacterium. *Science* 326(5957):1235–1240.

98. Nishi H, Hashimoto K, Madej T, Panchenko AR (2013) Evolutionary, physicochemical, and functional mechanisms of protein homooligomerization. *Progress in Molecular Biology and Translational Science* (Elsevier Inc, Netherlands), pp 3–24. 1st Ed.

99. Marsh JA, Teichmann SA (2015) Structure, dynamics, assembly, and evolution of protein complexes. *Annu Rev Biochem* 84(1):551–575.

100. Jones S, Thornton JM (1996) Principles of protein-protein interactions. *Proc Natl Acad Sci U S A* 93(1):13–20.

101. Wolynes PG (1996) Symmetry and the energy landscapes of biomolecules. *Proc Natl Acad Sci U S A* 93(25):14249–14255.

102. Lukatsky DB, Shakhnovich BE, Mintseris J, Shakhnovich EI (2007) Structural similarity enhances interaction propensity of proteins. *J Mol Biol* 365(5):1596–1606.

103. André I, Strauss CEM, Kaplan DB, Bradley P, Baker D (2008) Emergence of symmetry in homooligomeric biological assemblies. *Proc Natl Acad Sci U S A* 105(42):16148–16152.

104. Robinson C V, Sali A, Baumeister W (2007) The molecular sociology of the cell. *Nature* 450(7172):973–982.

105. Tarassov K, et al. (2008) An in vivo map of the yeast protein interactome. *Science* 320(5882):1465–1470.

106. Havugimana PC, et al. (2012) A census of human soluble protein complexes. *Cell* 150(5):1068–1081.

107. Ali MH, Imperiali B (2005) Protein oligomerization: how and why. *Bioorganic Med Chem* 13(17):5013–5020.

108. Alberts B (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* 92(3):291–294.

109. Eisenberg D, Marcotte EM, Xenarios I, Yeates TO (2000) Protein function in the post-genomic era. *Nature* 405(6788):823–826.

110. Lai YT, King NP, Yeates TO (2012) Principles for designing ordered protein assemblies. *Trends Cell Biol* 22(12):653–661.

111. Jubb H, Higueruelo AP, Winter A, Blundell TL (2012) Structural biology and drug discovery for protein-protein interactions. *Trends Pharmacol Sci* 33(5):241–248.

112. Kim B, et al. (2012) Accelerated disassembly of IgE–receptor complexes by a disruptive macromolecular inhibitor. *Nature* 491(7425):613–617.

113.   Wells JA, McClendon CL (2007) Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* 450(7172):1001–1009.

114.   Levy ED, Boeri Erba E, Robinson C V, Teichmann SA (2008) Assembly reflects evolution of protein complexes. *Nature* 453(7199):1262–1265.

115.   Pereira-Leal JB, Levy ED, Kamp C, Teichmann SA (2007) Evolution of protein complexes by duplication of homomeric interactions. *Genome Biol* 8(4):1–12.

116.   Bershtein S, Mu W, Shakhnovich EI (2012) Soluble oligomerization provides a beneficial fitness effect on destabilizing mutations. *Proc Natl Acad Sci U S A* 109(13):4857–4862.

117.   Sudha G, Nussinov R, Srinivasan N (2014) An overview of recent advances in structural bioinformatics of protein-protein interactions and a guide to their principles. *Prog Biophys Mol Biol* 116(2–3):141–150.

118.   Hashimoto K, Panchenko AR (2010) Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proc Natl Acad Sci U S A* 107(47):20352–20357.

119.   Perica T, Chothia C, Teichmann SA (2012) Evolution of oligomeric state through geometric coupling of protein interfaces. *Proc Natl Acad Sci U S A* 109(21):8127–8132.

120.   Malay AD, Allen KN, Tolan DR (2005) Structure of the thermolabile mutant aldolase B, A149P: Molecular basis of hereditary fructose intolerance. *J Mol Biol* 347(1):135–144.

121.   Nishi H, Ota M (2010) Amino acid substitutions at protein-protein interfaces that modulate the oligomeric state. *Proteins Struct Funct Bioinforma* 78(6):1563–1574.

122.   Levy ED (2010) A simple definition of structural regions in proteins and its use in analyzing interface evolution. *J Mol Biol* 403(4):660–670.

123.   Neuvirth H, Raz R, Schreiber G (2004) ProMate: A structure based prediction program to identify the location of protein-protein binding sites. *J Mol Biol* 338(1):181–199.

124.   de Vries SJ, Bonvin AMJJ (2008) How proteins get in touch: interface prediction in the study of biomolecular complexes. *Curr Protein Pept Sci* 9(4):394–406.

125.   Levy ED, De S, Teichmann SA (2012) Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. *Proc Natl Acad Sci U S A* 109(50):20461–20466.

126.   Hashimoto K, Nishi H, Bryant S, Panchenko AR (2011) Caught in self-interaction: evolutionary and functional mechanisms of protein homooligomerization. *Phys Biol* 8(3):35007.

127.   Bennett MJ, Schlunegger MP, Eisenberg D (1995) 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci* 4(12):2455–2468.

128.   Huang Y, Cao H, Liu Z (2012) Three-dimensional domain swapping in the protein structure space. *Proteins Struct Funct Bioinforma* 80(6):1610–1619.

129.   Mackinnon SS, Malevanets A, Wodak SJ (2013) Intertwined associations in structures of homooligomeric proteins. *Structure* 21(4):638–649.

130.   Gotte G, Libonati M (2014) Protein oligomerization. *Oligomerization of Chemical and Biological Compounds*, ed Claire Le (InTech), pp 239–278. 1st Ed.

131.   Akiva E, Itzhaki Z, Margalit H (2008) Built-in loops allow versatility in domain-domain interactions: lessons from self-interacting domains. *Proc Natl Acad Sci U S A* 105(36):13292–13297.

132. Hashimoto K, Madej T, Bryant SH, Panchenko AR (2010) Functional states of homooligomers: insights from the evolution of glycosyltransferases. *J Mol Biol* 399(1):196–206.

133. Nishi H, Koike R, Ota M (2011) Cover and spacer insertions: small nonhydrophobic accessories that assist protein oligomerization. *Proteins Struct Funct Bioinforma* 79(8):2372–2379.

134. Archibald JM, Logsdon JM, Doolittle WF (1999) Recurrent paralogy in the evolution of archaeal chaperonins. *Curr Biol* 9(18):1053–1056.

135. Archibald JM, Logsdon JM, Doolittle WF (2000) Origin and evolution of eukaryotic chaperonins: phylogenetic evidence for ancient duplications in CCT genes. *Mol Biol Evol* 17(10):1456–1466.

136. Ispolatov I, Yuryev A, Mazo I, Maslov S (2005) Binding properties and evolution of homodimers in protein-protein interaction networks. *Nucleic Acids Res* 33(11):3629–3635.

137. Reid AJ, Ranea JAG, Orengo CA (2010) Comparative evolutionary analysis of protein complexes in *E. coli* and yeast. *BMC Genomics* 11(79):1–16.

138. Dayhoff JE, Shoemaker BA, Bryant SH, Panchenko AR (2010) Evolution of protein binding modes in homooligomers. *J Mol Biol* 395(4):860–870.

139. Pereira-Leal JB, Teichmann SA (2005) Novel specificities emerge by stepwise duplication of functional modules. *Genome Res* 15(4):552–559.

140. Taylor WR, May ACW, Brown NP, Aszódi A (2001) Protein structure: geometry, topology and classification. *Reports Prog Phys* 64(4):517–590.

141. D'Alessio G (1999) The evolutionary transition from monomeric to oligomeric proteins: tools, the environment, hypotheses. *Prog Biophys Mol Biol* 72(3):271–298.

142. Mei G, Di Venere A, Rosato N, Finazzi-Agrò A (2005) The importance of being dimeric. *FEBS J* 272(1):16–27.

143. Dobson RCJ, Valegård K, Gerrard JA (2004) The crystal structure of three site-directed mutants of *Escherichia coli* dihydrodipicolinate synthase: further evidence for a catalytic triad. *J Mol Biol* 338(2):329–339.

144. Navia MA, et al. (1989) Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* 337(6208):615–620.

145. Monod J, Wyman J, Changeux J-PP (1965) On the nature of allosteric transitions: a plausible model. *J Mol Biol* 12(1):88–118.

146. Koshland DE, Némethy G, Filmer D (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* 5(1):365–385.

147. Crick FHC, Watson JD (1956) Structure of small viruses. *Nature* 177(4506):473–475.

148. Yang JR, Liao BY, Zhuang SM, Zhang JZ (2012) Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. *Proc Natl Acad Sci U S A* 109(14):E831–E840.

149. Nooren IMA, Thornton JM (2003) Structural characterisation and functional significance of transient protein-protein interactions. *J Mol Biol* 325(5):991–1018.

150. Nooren IMA, Thornton JM (2003) Diversity of protein-protein interactions. *EMBO J* 22(14):3486–3492.

151. Devenish SR, Gerrard JA (2009) The role of quaternary structure in $(\beta/\alpha)_8$-barrel proteins:

evolutionary happenstance or a higher level of structure-function relationships? *Org Biomol Chem* 7(5):833–9.

152. Robin C, Bardsley LMJ, Coppin C, Oakeshott JG (2009) Birth and death of genes and functions in the β-esterase cluster of *Drosophila*. *J Mol Evol* 69(1):10–21.

153. Oakeshott JG, Boyce TM, Russell RJ, Healy MJ (1995) Molecular insights into the evolution of an enzyme: esterase 6 in *Drosophila*. *Trends Ecol Evol* 10(3):103–110.

154. Collet C, et al. (1990) Molecular analysis of duplicated esterase genes in *Drosophila melanogaster*. *Mol Biol Evol* 7(1):9–28.

155. Karotam J, Delves AC, Oakeshott JG (1993) Conservation and change in structural and 5' flanking sequences of esterase 6 in sibling *Drosophila* species. *Genetica* 88(1):11–28.

156. Sergeev P V, et al. (1993) Regulation of tissue-specific expression of the esterase S gene in *Drosophila virilis*. *Nucleic Acids Res* 21(15):3545–3551.

157. Brady JP, Richmond RC, Oakeshott JG (1990) Cloning of the esterase-5 locus from *Drosophila pseudoobscura* and comparison with its homologue in *D. melanogaster*. *Mol Biol Evol* 7(6):525–546.

158. East P, Graham A, Whitington G (1990) Molecular isolation and preliminary characterisation of a duplicated esterase locus in *Drosophila buzzatii*. *Ecological and Evolutionary Genetics of Drosophila*, eds Barker JSF, Starmer WT, MacIntyre RJ (Springer US, Boston, MA), pp 389–406.

159. Dumancic MM, Oakeshott JG, Russell RJ, Healy MJ (1997) Characterization of the EstP protein in *Drosophila melanogaster* and its conservation in Drosophilids. *Biochem Genet* 35(7–8):251–271.

160. Wright TRF (1963) The genetics of an esterase in *Drosophila melanogaster*. *Genetics* 48(6):787–801.

161. Sheehan K, Richmond RC, Cochrane BJ (1979) Studies of esterase 6 in *Drosophila melanogaster*. III. The developmental pattern and tissue distribution. *Insect Biochem* 9(5):443–450.

162. Richmond RC, Gilbert DG, Sheehan KB, Gromko MH, Butterworth FM (1980) Esterase 6 and reproduction in *Drosophila melanogaster*. *Science* 207(4438):1483–1485.

163. Gilbert DG, Richmond RC (1982) Esterase 6 in *Drosophila melanogaster*: reproductive function of active and null males at low temperature. *Proc Natl Acad Sci U S A* 79(9):2962–2966.

164. Richmond RC, Senior A (1981) Esterase 6 of *Drosophila melanogaster* kinetics of transfer to females, decay in females and male recovery. *J Insect Physiol* 27(12):849–853.

165. Meikle DB, Sheehan KB, Phillis DM, Richmond RC (1990) Localization and longevity of seminal-fluid esterase 6 in mated female *Drosophila melanogaster*. *J Insect Physiol* 36(2):93–101.

166. Saad M, Game AY, Healy MJ, Oakeshott JG (1994) Associations of esterase-6 allozyme and activity variation with reproductive fitness in *Drosophila melanogaster*. *Genetica* 94(1):43–56.

167. Oakeshott JG, Saad M, Game AY, Healy MJ (1994) Causes and consequences of esterase-6 enzyme-activity variation in pre-adult *Drosophila melanogaster*. *Hered* 73:160–169.

168. Richmond RC, Nielsen KM, Brady JP, Snella EM (1990) Physiology, biochemistry and molecular biology of the Est-6 locus in Drosophila melanogaster. *Ecological and*

Evolutionary Genetics of Drosophila, eds Barker JSF, Starmer WT, McIntyre RJ (Springer US, New York, NY), pp 273–292.

169.   Anholt RRH, Williams TI, Carolina N (2010) The soluble proteome of the *Drosophila* antenna. *Chem Senses* 35(1):21–30.

170.   Chertemps T, et al. (2012) A carboxylesterase, esterase-6, modulates sensory physiological and behavioral response dynamics to pheromone in *Drosophila*. *BMC Biol* 10(56):1–12.

171.   Ha TS, Smith DP (2006) A pheromone receptor mediates 11-*cis*-vaccenyl acetate-induced responses in *Drosophila*. *J Neurosci* 26(34):8727–8733.

172.   White MM, Mane SD, Richmond RC (1988) Studies of esterase 6 in *Drosophila melanogaster*. XVIII. Biochemical differences between the slow and fast allozymes. *Mol Biol Evol* 5(1):41–62.

173.   Mane SD, Tompkins L, Richmond RC (1983) Male esterase 6 catalyzes the synthesis of a sex pheromone in *Drosophila melanogaster* females. *Science* 222(4622):419–421.

174.   Chertemps T, et al. (2015) An antennal carboxylesterase from *Drosophila melanogaster*, esterase 6, is a candidate odorant-degrading enzyme toward food odorants. *Front Physiol* 6.

175.   Butterworth FM (1969) Lipids of *Drosophila*: a newly detected lipid in the male. *Science* 163(3873):1356–1357.

176.   Guiraudie-Capraz G, Pho DB, Jallon J-M (2007) Role of the ejaculatory bulb in biosynthesis of the male pheromone *cis*-vaccenyl acetate in *Drosophila melanogaster*. *Integr Zool* 2(2):89–99.

177.   Jallon J-M (1984) A few chemical words exchanged by *Drosophila* during courtship and mating. *Behav Genet* 14(5):441–478.

178.   Brieger G, Butterworth FM (1970) *Drosophila melanogaster*: identity of male lipid in reproductive system. *Science* 167(3922):1262.

179.   Fernández MP, Kravitz EA (2013) Aggression and courtship in *Drosophila*: pheromonal communication and sex recognition. *J Comp Physiol A* 199(11):1065–1076.

180.   Gilbert DG (1981) Ejaculate esterase 6 and initial sperm use by female *Drosophila melanogaster*. *J Insect Physiol* 27(9):641–650.

181.   Scott D (1986) Inhibition of female *Drosophila melanogaster* remating by a seminal fluid protein (esterase 6). *Evolution* 40(5):1084–1091.

182.   Ronderos DS, Smith DP (2010) Activation of the T1 neuronal circuit is necessary and sufficient to induce sexually dimorphic mating behavior in *Drosophila melanogaster*. *J Neurosci* 30(7):2595–2599.

183.   Vander Meer RK, Obin MS, Zawistowski S, Sheehan KB, Richmond RC (1986) A reevaluation of the role of *cis*-vaccenyl acetate, *cis*-vaccenol and esterase 6 in the regulation of mated female sexual attractiveness in *Drosophila melanogaster*. *J Insect Physiol* 32(8):681–686.

184.   Ødegaard F (2000) How many species of arthropods? Erwin's estimate revised. *Biol J Linn Soc* 71(4):583–597.

185.   Nauen R (2007) Insecticide resistance in disease vectors of public health importance. *Pest Manag Sci* 63(7):628–633.

186. Pimentel D (1983) Effects of pesticides on the environment. *Proceedings of the 10th International Congress on Plant Protection*, pp 685–691.

187. Li X, Schuler MA, Berenbaum MR (2007) Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. *Annu Rev Entomol* 52(1):231–253.

188. Hemingway J, Hawkes NJ, McCarroll L, Ranson H (2004) The molecular basis of insecticide resistance in mosquitoes. *Insect Biochem Mol Biol* 34(7):653–665.

189. Hemingway J (2000) The molecular basis of two contrasting metabolic mechanisms of insecticide resistance. *Insect Biochem Mol Biol* 30(11):1009–1015.

190. Abernathy CO, Casida JE (1973) Pyrethroid insecticides: esterase cleavage in relation to selective toxicity. *Science* 179(4079):1235–1236.

191. Casida JE, Quistad GB (1998) Golden age of insecticide research: past, present, or future? *Annu Rev Entomol* 43(1):1–16.

192. Braga IA, et al. (2005) Effectiveness of methoprene, an insect growth regulator, against temephos-resistant *Aedes aegypti* populations from different Brazilian localities, under laboratory conditions. *J Med Entomol* 42(5):830–837.

193. Gould F, Anderson A, Reynolds A, Bumgarner L, Moar W (1995) Selection and genetic analysis of a *Heliothis virescens* (Lepidoptera: Noctuidae) strain with high levels of resistance to *Bacillus thuringiensis* toxins. *Insectic Resist Resist Manag* 88(6):1545–1559.

194. World Health Organization (2006) *Pesticides and their application: for the control of vectors and pests of public health importance* (World Health Organization, Geneva, Switzerland).

195. Sparks TC, Graves JB, Leonard BR (1993) Insecticide resistance and the tobacco budworm: past, present and future. *Rev Pestic Toxicol* 2:149–184.

196. Ecobichon DJ (2001) Pesticide use in developing countries. *Toxicology* 160(1):27–33.

197. Gupta RC (2005) *Toxicology of organophosphate & carbamate compounds* (Elsevier Academic Press, Burlington, MA). 1st Ed.

198. Krueger HR, O'Brien RD, Dauterman W (1960) Relationship between metabolism and differential toxicity in insects and mice of diazinon, dimethoate, parathion, and acethion. *J Econ Entomol* 53(1):25–31.

199. Sultatos LG, Murphy SD (1983) Kinetic analyses of the microsomal biotransformation of the phosphorothioate insecticides chlorpyrifos and parathion. *Toxicol Sci* 3(1):16–21.

200. Feyereisen R (1999) Insect P450 Enzymes. *Annu Rev Entomol* 44(1):507–533.

201. Dvir H, Silman I, Harel M, Rosenberry TL, Sussman JL (2010) Acetylcholinesterase: from 3D structure to function. *Chem Biol Interact* 187(1):10–22.

202. Quinn DM (1987) Acetylcholinesterase: enzyme structure, reaction dynamics, and virtual transition states. *Chem Rev* 87(5):955–979.

203. Smissaert HR, El Hamid FMA, Overmeer WPJ (1975) The minimum acetylcholinesterase (AChE) fraction compatible with life derived by aid of a simple model explaining the degree of dominance of resistance to inhibitors in AChE "mutants." *Biochem Pharmacol* 24(9):1043–1047.

204. Hoffmann F, Fournier D, Spierer P (1992) Minigene rescues acetylcholinesterase lethal mutations in *Drosophila melanogaster*. *J Mol Biol* 223(1):17–22.

205. Villatte F, Marcel V, Estrada-Mondaca S, Fournier D (1998) Engineering sensitive

acetylcholinesterase for detection of organophosphate and carbamate insecticides. *Biosens Bioelectron* 13(2):157–164.

206. Chen Z, Newcomb R, Forbes E, McKenzie J, Batterham P (2001) The acetylcholinesterase gene and organophosphorus resistance in the Australian sheep blowfly, *Lucilia cuprina*. *Insect Biochem Mol Biol* 31(8):805–816.

207. Boublik Y, et al. (2002) Acetylcholinesterase engineering for detection of insecticide residues. *Protein Eng* 15(1):43–50.

208. Jackson CJ, Oakeshott JG, Sanchez-Hernandez JC, Wheelock CE (2011) Carboxylesterases in the metabolism and toxicity of pesticides. *Anticholinesterase Pesticides: Metabolism, Neurotoxicity, and Epidemiology*, eds Satoh T, Gupta RC (John Wiley & Sons, Inc, Hoboken, NJ), pp 57–75. 1st Ed.

209. Millard CB, et al. (1999) Crystal structures of aged phosphonylated acetylcholinesterase: nerve agent reaction products at the atomic level. *Biochemistry* 38(22):7032–7039.

210. Nachon F, Asojo OA, Borgstahl GEO, Masson P, Lockridge O (2005) Role of water in aging of human butyrylcholinesterase inhibited by echothiophate: the crystal structure suggests two alternative mechanisms of aging. *Biochemistry* 44(4):1154–1162.

211. Shafferman A, et al. (1996) Aging of phosphylated human acetylcholinesterase: catalytic processes mediated by aromatic and polar residues of the active centre. *Biochem J* 318(3):833–840.

212. Fleming CD, et al. (2007) Crystal structures of human carboxylesterase 1 in covalent complexes with the chemical warfare agents soman and tabun. *Biochemistry* 46(17):5063–5071.

213. Maxwell DM, Brecht KM (2001) Carboxylesterase: specificity and spontaneous reactivation of an endogenous scavenger for organophosphorus compounds. *J Appl Toxicol* 21:S103–S107.

214. Hemingway J, Ranson H (2000) Insecticide resistance in insect vectors of human disease. *Annu Rev Entomol* 45(1):371–391.

215. Cui F, et al. (2015) Carboxylesterase-mediated insecticide resistance: quantitative increase induces broader metabolic resistance than qualitative change. *Pestic Biochem Physiol* 121:88–96.

216. Hardstone MC, Scott JG (2010) A review of the interactions between multiple insecticide resistance loci. *Pestic Biochem Physiol* 97(2):123–128.

217. Georghiou GP (1983) Management of resistance in arthropods. *Pest Resistance to Pesticides*, eds Georghiou GP, Saito T (Springer US, Boston, MA), pp 769–792.

218. World Health Organization (2013) *Test procedures for insecticide resistance monitoring in malaria vector mosquitoes* (World Health Organization, Geneva, Switzerland).

219. Bass C, et al. (2014) The evolution of insecticide resistance in the peach potato aphid, *Myzus persicae*. *Insect Biochem Mol Biol* 51(1):41–51.

220. Whalon ME, Mota-Sanchez D, Hollingworth R (2016) Arthropod Pesticide Resistance Database, Michigan State University. Available at: http://www.pesticideresistance.org/index.php [Accessed May 6 2016].

221. Zhang JL, Qiao C-L, Lan WS (2004) Detoxification of organophosphorus compounds by recombinant carboxylesterase from an insecticide-resistant mosquito and oxime-induced amplification of enzyme activity. *Environ Toxicol* 19(2):154–159.

222. Rinkevich FD, Hamm RL, Geden CJ, Scott JG (2007) Dynamics of insecticide resistance alleles in house fly populations from New York and Florida. *Insect Biochem Mol Biol* 37(6):550–558.

223. Claudianos C, Russell RJ, Oakeshott JG (1999) The same amino acid substitution in orthologous esterases confers organophosphate resistance on the house fly and a blowfly. *Insect Biochem Mol Biol* 29(8):675–686.

224. Li Y, et al. (2013) Organophosphate and pyrethroid hydrolase activities of mutant esterases from the cotton bollworm *Helicoverpa armigera*. *PLoS One* 8(10):e77685.

225. Wu S, et al. (2011) Overexpressed esterases in a fenvalerate resistant strain of the cotton bollworm, *Helicoverpa armigera*. *Insect Biochem Mol Biol* 41(1):14–21.

226. Tabashnik BE, et al. (2006) DNA screening reveals pink bollworm resistance to Bt cotton remains rare after a decade of exposure. *J Econ Entomol* 99(5):1525–1530.

227. Paton MG, Karunaratne SHPP, Giakoumaki E, Roberts N, Hemingway J (2000) Quantitative analysis of gene amplification in insecticide-resistant *Culex* mosquitoes. *Biochem J* 346(1):17–24.

228. Che-Mendoza A, Penilla RP, Rodríguez DA (2009) Insecticide resistance and glutathione S-transferases in mosquitoes: a review. *J Biotechnol* 8(8):1386–1397.

229. Casimiro S, Coleman M, Hemingway J, Sharp B (2006) Insecticide resistance in *Anopheles arabiensis* and *Anopheles gambiae* from Mozambique. *J Med Entomol* 43(2):276–282.

230. Liu N (2015) Insecticide resistance in mosquitoes: impact, mechanisms, and research directions. *Annu Rev Entomol* 60(1):537–559.

231. Wang K-Y, Liu T-X, Yu C-H, Jiang X-Y, Yi M-Q (2002) Resistance of *Aphis gossypii* (Homoptera: Aphididae) to fenvalerate and imidacloprid and activities of detoxification enzymes on cotton and cucumber. *J Econ Entomol* 95(2):407–413.

232. Yu Q-Y, Lu C, Li W, Xiang Z-H, Zhang Z (2009) Annotation and expression of carboxylesterases in the silkworm, *Bombyx mori*. *BMC Genomics* 10(1):553.

233. Oakeshott JG, et al. (2005) Comparing the organophosphorus and carbamate insecticide resistance mutations in cholin- and carboxyl-esterases. *Chem Biol Interact* 157–158:269–275.

234. Burns EC, Wilson BH (1963) Field resistance of horn flies to the organic phosphate insecticide Ronnel. *J Econ Entomol* 56(5):718.

235. Toda S, Komazaki S, Tomita T, Kono Y (2004) Two amino acid substitutions in acetylcholinesterase associated with pirimicarb and organophosphorous insecticide resistance in the cotton aphid, *Aphis gossypii* Glover (Homoptera: Aphididae). *Insect Mol Biol* 13(5):549–553.

236. Weill M, Berticat C, Raymond M, Chevillon C (2000) Quantitative polymerase chain reaction to estimate the number of amplified esterase genes in insecticide-resistant mosquitoes. *Anal Biochem* 285(2):267–270.

237. Weill M, et al. (2002) A novel acetylcholinesterase gene in mosquitoes codes for the insecticide target and is non-homologous to the ace gene in *Drosophila*. *Proc R Soc B Biol Sci* 269(1504):2007–2016.

238. Weill M, et al. (2004) The unique mutation in ace-1 giving high insecticide resistance is easily detectable in mosquito vectors. *Insect Mol Biol* 13(1):1–7.

239. Cassanelli S, Reyes M, Rault M, Manicardi GC, Sauphanor B (2006) Acetylcholinesterase mutation in an insecticide-resistant population of the codling moth *Cydia pomonella* (L.). *Insect Biochem Mol Biol* 36(8):642–653.

240. Liu N, Xu Q, Li T, He L, Zhang L (2009) Permethrin resistance and target site insensitivity in the mosquito *Culex quinquefasciatus* in Alabama. *J Med Entomol* 46(6):1424–1429.

241. Ramphul U, et al. (2009) Insecticide resistance and its association with target-site mutations in natural populations of *Anopheles gambiae* from eastern Uganda. *Trans R Soc Trop Med Hyg* 103(11):1121–1126.

242. Weill M, Lutfalla G, Mogensen K, Chandre F (2003) Insecticide resistance in mosquito vectors. *Nature* 423(6936):136.

243. Alout H, Berthomieu A, Hadjivassilis A, Weill M (2007) A new amino-acid substitution in acetylcholinesterase 1 confers insecticide resistance to *Culex pipiens* mosquitoes from Cyprus. *Insect Biochem Mol Biol* 37(1):41–47.

244. Wirth MC (1998) Isolation and characterization of two novel organophosphate resistance mechanisms in *Culex pipiens* from Cyprus. *J Am Mosq Control Assoc* 14(4):397–405.

245. Nabeshima T, Kozaki T, Tomita T, Kono Y (2003) An amino acid substitution on the second acetylcholinesterase in the pirimicarb-resistant strains of the peach potato aphid, *Myzus persicae*. *Biochem Biophys Res Commun* 307(1):15–22.

246. Moores GD, Devine GJ, Devonshire AL (1994) Insecticide resistance due to insensitive acetylcholinesterase in *Myzus persicae* and *Myzus nicotianae*. *Pestic Biochem Physiol* 49(2):413–418.

247. Fournier D, Mutero A (1994) Modification of acetylcholinesterase as a mechanism of resistance to insecticides. *Comp Biochem Physiol Part C Pharmacol* 108(1):19–31.

248. Devonshire AL, et al. (1998) The evolution of insecticide resistance in the peach-potato aphid, *Myzus persicae*. *Philos Trans R Soc B Biol Sci* 353(1376):1677–1684.

249. Mutero A, Pralavorio M, Bride JM, Fournier D (1994) Resistance-associated point mutations in insecticide-insensitive acetylcholinesterase. *Proc Natl Acad Sci U S A* 91(13):5922–5926.

250. Kozaki T, Shono T, Tomita T, Kono Y (2001) Fenitroxon insensitive acetylcholinesterases of the housefly, *Musca domestica* associated with point mutations. *Insect Biochem Mol Biol* 31(10):991–997.

251. Vontas JG, et al. (2001) Altered acetylcholinesterase confers organophosphate resistance in the olive fruit fly *Bactrocera oleae*. *Pestic Biochem Physiol* 71(2):124–132.

252. Walsh SB, et al. (2001) Identification and characterization of mutations in housefly (*Musca domestica*) acetylcholinesterase involved in insecticide resistance. *Biochem J* 359(1):175–181.

253. Devonshire AL, Field LM (1991) Gene amplification and insecticide resistance. *Annu Rev Entomol* 36(1):1–21.

254. Qiao CL, Raymond M (1995) The same esterase B1 haplotype is amplified in insecticide-resistant mosquitoes of the *Culex pipiens* complex from the Americas and China. *Hered* 74(4):339–345.

255. Field LM, Blackman RL, Tyler-Smith C, Devonshire AL (1999) Relationship between amount of esterase and gene copy number in insecticide-resistant *Myzus persicae* (Sulzer). *Biochem J* 339:737–742.

256. Field LM, Devonshire AL, Forde BG (1988) Molecular evidence that insecticide resistance in peach-potato aphids (*Myzus persicae* Sulz.) results from amplification of an esterase gene. *Biochem J* 251(1):309–312.

257. Raymond M, Callaghan A, Fort P, Pasteur N (1986) Amplification of an esterase gene is responsible for insecticide resistance in a California *Culex* mosquito. *Science* 233(4765):778–780.

258. Karunaratne SHPP, Vaughan A, Paton MG, Hemingway J (1998) Amplification of a serine esterase gene is involved in insecticide resistance in Sri Lankan *Culex tritaeniorhynchus*. *Insect Mol Biol* 7(4):307–315.

259. Vontas JG, Small GJ, Hemingway J (2000) Comparison of esterase gene amplification, gene expression and esterase activity in insecticide susceptible and resistant strains of the brown planthopper, *Nilaparvata lugens* (Stal). *Insect Mol Biol* 9(6):655–660.

260. Chanda SM, Mortensen SR, Moser VC, Padilla S (1997) Tissue-specific effects of chlorpyrifos on carboxylesterase and cholinesterase activity in adult rats: an in vitro and in vivo comparison. *Fundam Appl Toxicol* 38(2):148–157.

261. Karunaratne SHPP, Jayawardena KG, Hemingway J, Ketterman AJ (1993) Characterization of a B-type esterase involved in insecticide resistance from the mosquito *Culex quinquefasciatus*. *Biochem J* 294(2):575–579.

262. Ketterman AJ, Jayawardena KG, Hemingway J (1992) Purification and characterization of a carboxylesterase involved in insecticide resistance from the mosquito *Culex quinquefasciatus*. *Biochem J* 287(2):355–360.

263. Ketterman AJ, Karunaratne SHPP, Jayawardenal KGI, Hemingway J (1993) Qualitative differences between populations of *Culex quinquefasciatus* in both the esterases A2 and B2 which are involved in insecticide resistance. *Pest Biochem Physiol* 47(2):142–148.

264. Campbell PM, Newcomb RD, Russell RJ, Oakeshott JG (1998) Two different amino acid substitutions in the ali-esterase, E3, confer alternative types of organophosphorus insecticide resistance in the sheep blowfly, *Lucilia cuprina*. *Insect Biochem Mol Biol* 28(3):139–150.

265. Oppenoorth FJ, van Asperen K (1960) Allelic genes in the housefly producing modified enzymes that cause organophosphate resistance. *Science* 132(3422):298–299.

266. Heidari R, et al. (2004) Hydrolysis of organophosphorus insecticides by in vitro modified carboxylesterase E3 from *Lucilia cuprina*. *Insect Biochem Mol Biol* 34(4):353–363.

267. de Carvalho RA, Torres TT, de Azeredo-Espin AML (2006) A survey of mutations in the *Cochliomyia hominivorax* (Diptera: Calliphoridae) esterase E3 gene associated with organophosphate resistance and the molecular identification of mutant alleles. *Vet Parasitol* 140(3):344–351.

268. Hartley CJ, et al. (2006) Amplification of DNA from preserved specimens shows blowflies were preadapted for the rapid evolution of insecticide resistance. *Proc Natl Acad Sci U S A* 103(23):8757–8762.

269. Zhu YC, Dowdy AK, Baker JE (1999) Detection of single-base substitution in an esterase gene and its linkage to malathion resistance in the parasitoid Anisopteromalus calandrae (Hymenoptera: Pteromalidae). *Pestic Sci* 55(4):398–404.

270. Campbell PM, et al. (1998) Cross-resistance patterns among *Lucilia cuprina* (Diptera: Calliphoridae) resistant to organophosphorus insecticides. *J Econ Entomol* 91(2):367–375.

271. Heidari R, et al. (2005) Hydrolysis of pyrethroids by carboxylesterases from *Lucilia*

cuprina and *Drosophila melanogaster* with active sites modified by in vitro mutagenesis. *Insect Biochem Mol Biol* 35(6):597–609.

272. Parker AG, Russell RJ, Delves AC, Oakeshott JG (1991) Biochemistry and physiology of esterases in organophosphate-susceptible and resistant strains of the Australian sheep blowfly, *Lucilia cuprina*. *Pestic Biochem Physiol* 318(3):305–318.

273. Smyth KA, Boyce TM, Russell RJ, Oakeshott JG (2000) MCE activities and malathion resistances in field populations of the Australian sheep blowfly (*Lucilia cuprina*). *Hered* 84(1):63–72.

274. Gacar F, Taşkin V (2009) Partial base sequence analysis of MdαE7 gene and ali-esterase enzyme activities in field collected populations of house fly (*Musca domestica* L.) from Mediterranean and Aegean Regions of Turkey. *Pestic Biochem Physiol* 94(2):86–92.

275. Lockridge O, et al. (1997) A single amino acid substitution, Gly117His, confers phosphotriesterase (organophosphorus acid anhydride hydrolase) activity on human butyrylcholinesterase. *Biochemistry* 36(4):786–795.

276. McKenzie JA, O'Farrell K (1993) Modification of developmental instability and fitness: malathion-resistance in the Australian sheep blowfly, *Lucilia cuprina*. *Genetica* 89(1):67–76.

277. Parker A, Campbell P, Spackman M, Russell R, Oakeshott J (1996) Comparison of an esterase associated with organophosphate resistance in *Lucilia cuprina* with an orthologue not associated with resistance in *Drosophila melanogaster*. *Pestic Biochem Physiol* 55(2):85–99.

278. Cui F, Qu H, Cong J, Liu X-L, Qiao C-L (2007) Do mosquitoes acquire organophosphate resistance by functional changes in carboxylesterases? *FASEB J* 21(13):3584–3591.

279. Cui F, et al. (2011) Two single mutations commonly cause qualitative change of nonspecific carboxylesterases in insects. *Insect Biochem Mol Biol* 41(1):1–8.

280. Labbe P, Lenormand T, Raymond M (2005) On the worldwide spread of an insecticide resistance gene: a role for local selection. *J Evol Biol* 18(6):1471–84.

281. Gould SJ (1989) *Wonderful life: the burgess shale and the nature of history* (WW Norton, New York, NY). 1st Ed.

282. Gould SJ (2002) *The structure of evolutionary theory* (The Belknap Press of Harvard University Press, London, England). 1st Ed.

283. Travisano M, Mongold JA, Bennett AF, Lenski RE (1995) Experimental tests of the roles of adaptation, chance, and history in evolution. *Science* 267(5194):87.

284. Blount ZD, Borland CZ, Lenski RE (2008) Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc Natl Acad Sci U S A* 105(23):7899–7906.

285. Simões P, et al. (2008) How repeatable is adaptive evolution? The role of geographical origin and founder effects in laboratory adaptation. *Evolution* 62(8):1817–1829.

286. Bryant EH, McCommas SA, Combs LM (1986) The effect of an experimental bottleneck upon quantitative genetic variation in the housefly. *Genetics* 114(4):1191–1211.

287. Cheverud JM, et al. (1999) Epistasis and the evolution of additive genetic variance in populations that pass through a bottleneck. *Evolution* 53(4):1009–1018.

288. Naciri-Graven Y, Goudet J (2003) The additive genetic variance after bottlenecks is affected by the number of loci involved in epistatic interactions. *Evolution* 57(4):706–716.

289. Zhang X-S, Wang J, Hill WG (2004) Redistribution of gene frequency and changes of genetic variation following a bottleneck in population size. *Genetics* 167(3):1475–1492.

290. Mani GS, Clarke BC (1990) Mutational order: a major stochastic process in evolution. *Proc R Soc London B Biol Sci* 240(1297):29 LP-37.

291. Lenski RE, Rose MR, Simpson SC, Tadler SC (1991) Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am Nat* 138(6):1315–1341.

292. Cooper VS, Lenski RE (2000) The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature* 407(6805):736–739.

293. Cooper TF, Remold SK, Lenski RE, Schneider D (2008) Expression profiles reveal parallel evolution of epistatic interactions involving the CRP regulon in *Escherichia coli*. *PLoS Genet* 4(2):e35.

294. Weinreich DM, Watson RA, Chao L (2005) Perspective: sign epistasis and genetic costraint on evolutionary trajectories. *Evolution* 59(6):1165–1174.

295. Weinreich DM, Delaney NF, DePristo MA, Hartl DL (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312(5770):111–114.

296. Fisher RA (1999) *The genetical theory of natural selection: a complete variorum edition* ed Bennett H (Oxford University Press, Oxford, U.K.). 1st Ed.

297. Losos JB, Jackman TR, Larson A, de Queiroz K, Rodríguez-Schettino L (1998) Contingency and determinism in replicated adaptive radiations of island lizards. *Science* 279(5359):2115–2118.

298. Harms MJ, Thornton JW (2014) Historical contingency and its biophysical basis in glucocorticoid receptor evolution. *Nature* 512(7513):203–207.

299. Bedhomme S, Lafforgue G, Elena SF (2013) Genotypic but not phenotypic historical contingency revealed by viral experimental evolution. *BMC Evol Biol* 13(1).

300. Martin RE, et al. (2009) Chloroquine transport via the malaria parasite's chloroquine resistance transporter. *Science* 325(5948):1680–1682.

301. Field SF, Matz M V (2010) Retracing evolution of red fluorescence in GFP-Like proteins from faviina corals. *Mol Biol Evol* 27(2):225–233.

302. Bloom JD, Gong LI, Baltimore D (2010) Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science* 328(5983):1272–1275.

303. Lynch VJ, May G, Wagner GP (2011) Regulatory evolution through divergence of a phosphoswitch in the transcription factor CEBPB. *Nature* 480(7377):383–386.

304. Gong LI, Suchard MA, Bloom JD (2013) Stability-mediated epistasis constrains the evolution of an influenza protein. *Elife* 2:e00631.

305. Ortlund EA, Bridgham JT, Redinbo MR, Thornton JW (2007) Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* 317(5844):1544–1548.

306. Meyer JR, et al. (2012) Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science* 335(6067):428–432.

307. Bridgham JT, Ortlund EA, Thornton JW (2009) An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* 461(7263):515–519.

308. Beatty J (2006) Replaying life's tape. *J Philos* 103(7):336–362.

309. Vermeij GJ (2006) Historical contingency and the purported uniqueness of evolutionary innovations. *Proc Natl Acad Sci U S A* 103(6).

310. Gould SJ (1970) Dollo on Dollo's law: irreversibility and the status of evolutionary laws. *J Hist Biol* 3(2):189–212.

311. Bull JJ, Charnov EL (1985) On irreversible evolution. *Evolution* 39(5):1149–1155.

312. Wagner GP (1982) The logical structure of irreversible systems transformations: A theorem concerning Dollo's law and chaotic movement. *J Theor Biol* 96(3):337–346.

313. Lewontin RC (1966) Is nature probable or capricious? *Bioscience*:25–27.

314. Drawz SM, et al. (2009) The role of a second-shell residue in modifying substrate and inhibitor interactions in the SHV β-lactamase: a study of ambler position Asn276. *Biochemistry* 48(21):4557–4566.

315. Mabbitt PD, et al. (2016) Conformational disorganization within the active site of a recently evolved organophosphate hydrolase limits its catalytic efficiency. *Biochemistry* 55(9):1408–1417.

316. Correy GJ, et al. (2016) Mapping the accessible conformational landscape of an insect carboxylesterase using conformational ensemble analysis and kinetic crystallography. *Structure* 24(6):977–987.

317. Gibson DG, et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6(5):343–345.

318. Neylon C, et al. (2000) Interaction of the *Escherichia coli* replication terminator protein (Tus) with DNA: A model derived from DNA-binding studies of mutant proteins by surface plasmon resonance. *Biochemistry* 39(39):11989–11999.

319. Gasteiger E, et al. (2005) Protein identification and analysis tools on the ExPASy server. *The Proteomics Protocols Handbook*, ed Walker J (Humana Press, Totowa, NJ), pp 571–607. 1st Ed.

320. Yang J, Zhang Y (2015) I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res* 43(W1):W174–W181.

321. Guerrero FD (2000) Cloning of a horn fly cDNA, *HiαE7*, encoding an esterase whose transcript concentration is elevated in diazinon-resistant flies. *Insect Biochem Mol Biol* 30(11):1107–1115.

322. Wang L, et al. (2015) Overexpression of two α-esterase genes mediates metabolic resistance to malathion in the oriental fruit fly, *Bactrocera dorsalis* (Hendel). *Insect Mol Biol* 24(4):467–479.

323. Poupardin R, Srisukontarat W, Yunta C, Ranson H (2014) Identification of carboxylesterase genes implicated in temephos resistance in the dengue vector *Aedes aegypti*. *PLoS Negl Trop Dis* 8(3):e2743.

324. Fraser NJ, et al. (2015) Evolution of protein quaternary structure in response to selective pressure for increased thermostability. *J Mol Biol* 428(11):2359–2371.

325. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10(6):845–858.

326. Cozzetto D, Kryshtafovych A, Ceriani M, Tramontano A (2007) Assessment of predictions in the model quality assessment category. *Proteins Struct Funct Bioinforma* 69(S8):175–183.

327. Davies AG, et al. (1996) Scalloped wings is the *Lucilia cuprina* notch homologue and a candidate for the modifier of fitness and asymmetry of diazinon resistance. *Genetics* 143(3):1321–1337.

328. Casida JE, Quistad GB (2004) Why insecticides are more toxic to insects than people: the unique toxicology of insects. *J Pestic Sci* 29(2):81–86.

329. Nedergaard M, Takano T, Hansen AJ (2002) Opinion: beyond the role of glutamate as a neurotransmitter. *Nat Rev Neurosci* 3(9):748.

330. Martyn JAJ, Fagerlund MJ, Eriksson LI (2009) Basic principles of neuromuscular transmission. *Anaesthesia* 64(s1):1–9.

331. Silverman RB (2002) *The organic chemistry of enzyme-catalyzed reactions* (Academic Press, San Diego, CA). 1st Ed.

332. Lilienfeld S (2002) Galantamine - a novel cholinergic drug with a unique dual mode of action for the treatment of patients with Alzheimer's disease. *CNS Drug Rev* 8(2):159–176.

333. Harel M, et al. (1993) Quaternary ligand binding to aromatic residues in the active-site gorge of acetylcholinesterase. *Proc Natl Acad Sci U S A* 90(19):9031–9035.

334. Ordentlich A, et al. (1993) Dissection of the human acetylcholinesterase active center determinants of substrate specificity. Identification of residues constituting the anionic site, the hydrophobic site, and the acyl pocket. *J Biol Chem* 268(23):17083–17095.

335. Ordentlich A, et al. (1998) Functional characteristics of the oxyanion hole in human acetylcholinesterase. *J Biol Chem* 273(31):19509–19517.

336. Radić Z, Pickering NA, Vellom DC, Camp S, Taylor P (1993) Three distinct domains in the cholinesterase molecule confer selectivity for acetyl- and butyrylcholinesterase inhibitors. *Biochemistry* 32(45):12074–12084.

337. Kitz RJ, Braswell LM, Ginsburg S (1970) On the question: is acetylcholinesterase an allosteric protein? *Mol Pharmacol* 6(2):108–121.

338. Reiner E, Aldridge N, Simeon V, Radić Z, Taylor P (1991) Mechanism of substrate inhibition of acetylcholinesterase. *Cholinesterases: Structure, Function, Mechanism, Genetics and Cell Biology,* eds Massoulié J, et al. (American Chemical Society, Washington, DC), pp 227–228. 1st Ed.

339. Shafferman A, et al. (1992) Substrate inhibition of acetylcholinesterase: residues affecting signal transduction from the surface to the catalytic center. *EMBO J* 11(10):3561–3568.

340. Bajda M, et al. (2013) Structure-based search for new inhibitors of cholinesterases. *Int J Mol Sci* 14(3):5608–5632.

341. Colovic MB, Krstic DZ, Lazarevic-Pasti TD, Bondzic AM, Vasic VM (2013) Acetylcholinesterase inhibitors: pharmacology and toxicology. *Curr Neuropharmacol* 11(3):315–335.

342. Bachurin SO (2003) Medicinal chemistry approaches for the treatment and prevention of Alzheimer disease. *Med Res Rev* 23(1):48–88.

343. Colombres M, Sagal J, Inestrosa N (2004) An overview of the current and novel drugs for Alzheimer's disease with particular reference to anti-cholinesterase compounds. *Curr Pharm Des* 10(25):3121–3130.

344. Giacobini E (2004) Cholinesterase inhibitors: new roles and therapeutic alternatives. *Pharmacol Res* 50(4):433–440.

345. Bierer LM, et al. (1995) Neurochemical correlates of dementia severity in Alzheimer's disease: relative importance of the cholinergic deficits. *J Neurochem* 64(2):749–760.

346. Fisher A (2012) Cholinergic modulation of amyloid precursor protein processing with emphasis on M1 muscarinic receptor: perspectives and challenges in treatment of Alzheimer's disease. *J Neurochem* 120(s1):22–33.

347. Farlow MR (2001) Pharmacokinetic profiles of current therapies for Alzheimer's disease: Implications for switching to galantamine. *Clin Ther* 23:A13–A24.

348. Selkoe DJ, et al. (2001) Alzheimer's disease: genes, proteins, and therapy. *Physiol Rev* 81(2):741–766.

349. Stoddard S V, Hamann MT, Wadkins RM (2014) Insights and ideas garnered from marine metabolites for development of dual-function acetylcholinesterase and amyloid-β aggregation inhibitors. *Mar Drugs* 12(4):2114–2131.

350. Inestrosa NC, et al. (1996) Acetylcholinesterase accelerates assembly of amyloid-β-peptides into Alzheimer's fibrils: possible role of the peripheral site of the enzyme. *Neuron* 16(4):881–891.

351. Stahl SM (2000) The new cholinesterase inhibitors for Alzheimer's disease, part 2: illustrating their mechanisms of action. *J Clin Psychiatry* 61(11):813–814.

352. Racchi M, Mazzucchelli M, Porrello E, Lanni C, Govoni S (2004) Acetylcholinesterase inhibitors: novel activities of old molecules. *Pharmacol Res* 50(4):441–451.

353. Lanari A, Amenta F, Silvestrelli G, Tomassoni D, Parnetti L (2006) Neurotransmitter deficits in behavioural and psychological symptoms of Alzheimer's disease. *Mech Ageing Dev* 127(2):158–165.

354. Stepankova S, Komers K (2008) Cholinesterases and cholinesterase inhibitors. *Curr Enzym Inhib* 4(4):160–171.

355. Mehta M, Adem A, Sabbagh M (2012) New acetylcholinesterase inhibitors for Alzheimer's disease. *Int J Alzheimers Dis* 2012:1–8.

356. Camps P, Munoz-Torrero D (2002) Cholinergic drugs in pharmacotherapy of Alzheimer's disease. *Mini Rev Med Chem* 2(1):11–25.

357. Wang B, et al. (2009) Efficacy and safety of natural acetylcholinesterase inhibitor huperzine A in the treatment of Alzheimer's disease: an updated meta-analysis. *J Neural Transm* 116(4):457–465.

358. Bai D (2007) Development of huperzine A and B for treatment of Alzheimer's disease. *Pure Appl Chem* 79(4):469–479.

359. Orhan G, Orhan I, Subutay-Oztekin N, Ak F, Sener B (2009) Contemporary anticholinesterase pharmaceuticals of natural origin and their synthetic analogues for the treatment of Alzheimer's disease. *Recent Pat CNS Drug Discov* 4(1):43–51.

360. Melnikova I (2007) Therapies for Alzheimer's disease. *Nat Rev Drug Discov* 6(5):341–342.

361. Birks JS (2006) Cholinesterase inhibitors for Alzheimer's disease. *The Cochrane Library* (John Wiley & Sons, Inc). 1st Ed.

362. Thomsen T, et al. (1991) Inhibition of acetylcholinesterase activity in human brain tissue and erythrocytes by galanthamine, physostigmine and tacrine. *Clin Chem Lab Med* 29(8):487–492.

363. Ibach B, Haen E (2004) Acetylcholinesterase inhibition in Alzheimer's disease. *Curr Pharm Des* 10(3):231–251.

364. Kojima J, Nakajima K, Ochiai M, Nakayama K (1997) Effects of NIK-247 on cholinesterase and scopolamine-induced amnesia. *Methods Find Exp Clin Pharmacol* 19(4):245–251.

365. Zhu X, et al. (1996) Effects of MF-268, a new cholinesterase inhibitor, on acetylcholine and biogenic amines in rat cortex. *J Neurosci Res* 43(1):120–126.

366. Enz A, Boddeke H, Gray J, Spiegél R (1991) Pharmacologic and clinicopharmacologic properties of SDZ ENA 713, a centrally selective acetylcholinesterase inhibitor. *Ann N Y Acad Sci* 640(1):272–275.

367. Enz A, Amstutz R, Boddeke H, Gmelin G, Malanowski J (1993) Brain selective inhibition of acetylcholinesterase: a novel approach to therapy for Alzheimer's disease. *Prog Brain Res* 98:431–438.

368. Anand R, Gharabawi G, Enz A (1996) Efficacy and safety results of the early phase studies with Exelon (TM)(ENA-713) in Alzheimer's disease: An overview. *J drug Dev Clin Pract* 8(2):109–116.

369. Birks JS, Grimley Evans J (2015) Rivastigmine for Alzheimer's disease. *Cochrane Libr* 10(4):CD001191.

370. Watkins PB, Zimmerman HJ, Knapp MJ, Gracon SI, Lewis KW (1994) Hepatotoxic effects of tacrine administration in patients with Alzheimer's disease. *Jama* 271(13):992–998.

371. Lahiri DK, Lewis S, Farlow MR (1994) Tacrine alters the secretion of the beta-amyloid precursor protein in cell lines. *J Neurosci Res* 37(6):777–787.

372. Ros E, et al. (2001) Effects of bis (7)-tacrine on spontaneous synaptic activity and on the nicotinic ACh receptor of *Torpedo* electric organ. *J Neurophysiol* 86(1):183–189.

373. McGleenon BM, Dynan KB, Passmore AP (1999) Acetylcholinesterase inhibitors in Alzheimer's disease. *Br J Clin Pharmacol* 48(4):471.

374. Thomsen T, Zendeh B, Fischer JP, Kewitz H (1991) In vitro effects of various cholinesterase inhibitors on acetyl-and butyrylcholinesterase of healthy volunteers. *Biochem Pharmacol* 41(1):139–141.

375. Zhao Q, Tang XC (2002) Effects of huperzine A on acetylcholinesterase isoforms in vitro: comparison with tacrine, donepezil, rivastigmine and physostigmine. *Eur J Pharmacol* 455(2):101–107.

376. Skau KA, Shipley MT (1999) Phenylmethylsulfonyl fluoride inhibitory effects on acetylcholinesterase of brain and muscle. *Neuropharmacology* 38(5):691–698.

377. Siek GC, Katz LS, Fishman EB, Korosi TS, Marquis JK (1990) Molecular forms of acetylcholinesterase in subcortical areas of normal and Alzheimer disease brain. *Biol Psychiatry* 27(6):573–580.

378. Rakonczay Z (2003) Potencies and selectivities of inhibitors of acetylcholinesterase and its molecular forms in normal and Alzheimer's disease brain. *Acta Biol Hung* 54(2):183–189.

379. Grossberg GT (2003) Cholinesterase inhibitors for the treatment of Alzheimer's disease: getting on and staying on. *Curr Ther Res* 64(4):216–235.

380. Zemek F, et al. (2014) Outcomes of Alzheimer's disease therapy with acetylcholinesterase inhibitors and memantine. *Expert Opin Drug Saf* 13(6):759–774.

381. Heinrich M, Teoh HL (2004) Galanthamine from snowdrop - the development of a modern drug against Alzheimer's disease from local Caucasian knowledge. *J Ethnopharmacol* 92(2):147–162.

382. Paskov DS, et al. (1962) Clinical experience with nivalin as anticholinesterase drug in anaesthesiological practice. *Proc First Eur Congr Anesthesiol*, pp 3–9.

383. Parys W (1998) Development of reminyl (R) (galantamine), a novel acetylcholinesterase inhibitor, for the treatment of Alzheimer's disease. *Alzheimers Reports* 1:S19–S20.

384. Morris JC (2001) Therapeutic continuity in Alzheimer's disease: switching patients to galantamine. *Clin Ther* 23:A1–A24.

385. Bartolucci C, Perola E, Pilger C, Fels G, Lamba D (2001) Three-dimensional structure of a complex of galanthamine (Nivalin®) with acetylcholinesterase from *Torpedo californica*: implications for the design of new anti-Alzheimer drugs. *Proteins Struct Funct Bioinforma* 42(2):182–191.

386. Pilger C, Bartolucci C, Lamba D, Tropsha A, Fels G (2001) Accurate prediction of the bound conformation of galanthamine in the active site of *Torpedo californica* acetylcholinesterase using molecular docking. *J Mol Graph Model* 19(3):288–296.

387. Kitisripanya N, Saparpakorn P, Wolschann P, Hannongbua S (2011) Binding of huperzine A and galanthamine to acetylcholinesterase, based on ONIOM method. *Nanomedicine Nanotechnology, Biol Med* 7(1):60–68.

388. Woodruff-Pak DS, Vogel RW, Wenk GL (2001) Galantamine: effect on nicotinic receptor binding, acetylcholinesterase inhibition, and learning. *Proc Natl Acad Sci U S A* 98(4):2089–2094.

389. Ago Y, Koda K, Takuma K, Matsuda T (2011) Pharmacological aspects of the acetylcholinesterase inhibitor galantamine. *J Pharmacol Sci* 116(1):6–17.

390. Marco-Contelles J, do Carmo Carreiras M, Rodríguez C, Villarroya M, García AG (2006) Synthesis and pharmacology of galantamine. *Chem Rev* 106(1):116–133.

391. Marco L, do Carmo Carreiras M (2006) Galanthamine, a natural product for the treatment of Alzheimer's disease. *Recent Pat CNS Drug Discov* 1(1):105–111.

392. Bastida J, et al. (1990) *Narcissus nivalis*: a new source of galanthamine. *Planta Med* 56(1):123–124.

393. Shieh W-C, Carlson JA (1994) Asymmetric transformation of either enantiomer of narwedine via total spontaneous resolution process, a concise solution to the synthesis of (-)-galanthamine. *J Org Chem* 59(18):5463–5465.

394. Czollner L, et al. (1998) New kilogram-synthesis of the anti-alzheimer drug (-)-galanthamine. *Tetrahedron Lett* 39(15):2087–2088.

395. Greenblatt HM, Kryger G, Lewis T, Silman I, Sussman JL (1999) Structure of acetylcholinesterase complexed with (-)-galanthamine at 2.3 resolution. *FEBS Lett* 463(3):321–326.

396. Cheung J, et al. (2012) Structures of human acetylcholinesterase in complex with pharmacologically important ligands. *J Med Chem* 55(22):10282–10286.

397. Atanasova M, Yordanov N, Dimitrov I, Berkov S, Doytchinova I (2015) Molecular docking study on galantamine derivatives as cholinesterase inhibitors. *Mol Inform* 34(6):394–403.

398. Martinez A, Castro A (2006) Novel cholinesterase inhibitors as future effective drugs for

the treatment of Alzheimer's disease. *Expert Opin Investig Drugs* 15(1):1–12.

399. Viegas C, Bolzani VDS, Barreiro EJ, Fraga CAM (2005) New anti-Alzheimer drugs from biodiversity: the role of the natural acetylcholinesterase inhibitors. *Mini Rev Med Chem* 5(10):915–926.

400. Sangnoi Y, et al. (2008) Acetylcholinesterase-inhibiting activity of pyrrole derivatives from a novel marine gliding bacterium, *Rapidithrix thailandica*. *Mar Drugs* 6(4):578–586.

401. Kanjana-opas A, Panphon S, Fun H-K, Chantrapromma S (2006) 4-Methyl-3H-pyrrolo [2, 3-c] quinoline. *Acta Crystallogr Sect E Struct Reports Online* 62(7):2728–2730.

402. Srisukchayakul P, et al. (2007) *Rapidithrix thailandica* gen. nov., sp nov., a marine gliding bacterium isolated from samples collected from the Andaman sea, along the southern coastline of Thailand. *Int J Syst Evol Microbiol* 57(10):2275–2279.

403. Okanya PW, Mohr KI, Gerth K, Jansen R, Müller R (2011) Marinoquinolines A−F, pyrroloquinolines from *Ohtaekwangia kribbensis* (Bacteroidetes). *J Nat Prod* 74(4):603–608.

404. Carroll AR, Duffy S, Avery VM (2010) Aplidiopsamine A, an antiplasmodial alkaloid from the temperate Australian ascidian, *Aplidiopsis confluata*. *J Org Chem* 75(23):8291–8294.

405. Schwalm CS, Correia CRD (2012) Divergent total synthesis of the natural antimalarial marinoquinolines A, B, C, E and unnatural analogues. *Tetrahedron Lett* 53(36):4836–4840.

406. Ellman GL, Courtney KD, Andres V, Featherstone RM (1961) A new and rapid colorimetric determination of acetylcholinesterase activity. *Biochem Pharmacol* 7(2):88–95.

407. SchuÈttelkopf AW, Van Aalten DMF (2004) PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. *Acta Crystallogr Sect D Biol Crystallogr* 60(8):1355–1363.

408. Trott O, Olson AAJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J Comput Chem* 31(2):455–461.

409. Wiesner J, Kříž Z, Kuča K, Jun D, Koča J (2007) Acetylcholinesterases–the structural similarities and differences. *J Enzyme Inhib Med Chem* 22(4):417–424.

410. Cooke PH, Oakeshott JG (1989) Amino acid polymorphisms for esterase-6 in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* 86(4):1426–1430.

411. Hamulakova S, et al. (2014) Synthesis and biological evaluation of novel tacrine derivatives and tacrine-coumarin hybrids as cholinesterase inhibitors. *J Med Chem* 57(16):7073–7084.

412. Mao F, Li J, Wei H, Huang L, Li X (2015) Tacrine-propargylamine derivatives with improved acetylcholinesterase inhibitory activity and lower hepatotoxicity as a potential lead compound for the treatment of Alzheimer's disease. *J Enzyme Inhib Med Chem* 30(6):995–1001.

Appendices

Appendix A. Supporting Information for the Research Article in Chapter 2.
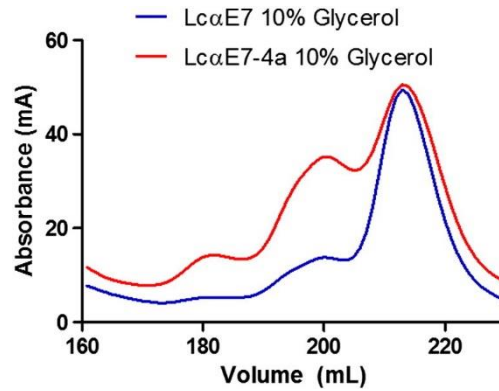
SI Fig. 1. The effect of viscosity on oligomerization. Size exclusion chromatograms of *Lc*αE7-4a and *Lc*αE7-4a WT in the presence of 10% glycerol show that for *Lc*αE7-4a, the tetrameric and dimeric fractions (peak heights 15 mAU and 35 AU) comprise approximately 50% of the total protein (monomer peak height 51 mAU). In contrast, for *Lc*αE7-4a WT, the higher-molecular-weight species (tetramer and dimer peak heights 6 mAU and 15 mAU versus 50 for monomer) comprise less than 30% of the total protein.
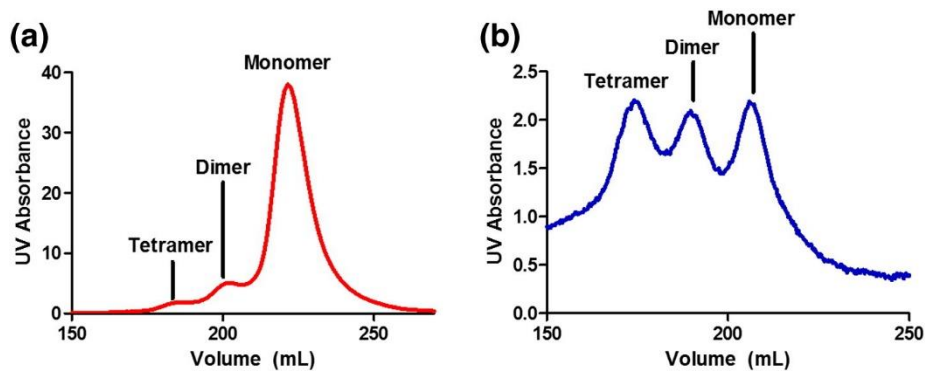
SI Fig. 2. (a) Re-equilibration of monomeric fraction (single 5 mL fraction from middle of monomeric peak) into dimeric and tetrameric peaks after 5 h at 4 °C. (b) Re-equilibration of tetrameric fraction (single 5 mL fraction from middle of tetrameric peak) into dimeric and monomeric species after 24 h at 4 °C.
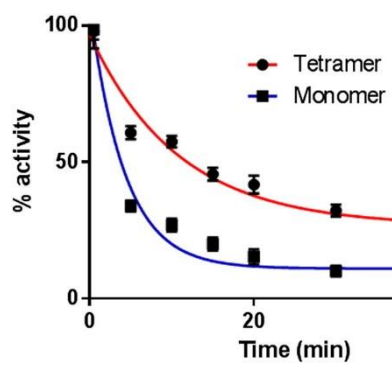
SI Fig. 3. Activity decay over time at 46 °C for monomeric and tetrameric species of *Lc*αE7-4a. Aliquots were taken over a series of time points when tetrameric and monomeric species were incubated at 46 °C. Specific activity was measured using 4-nitrophenyl butyrate, 200 mM NaCl and 20 mM HEPES (pH 7.5).
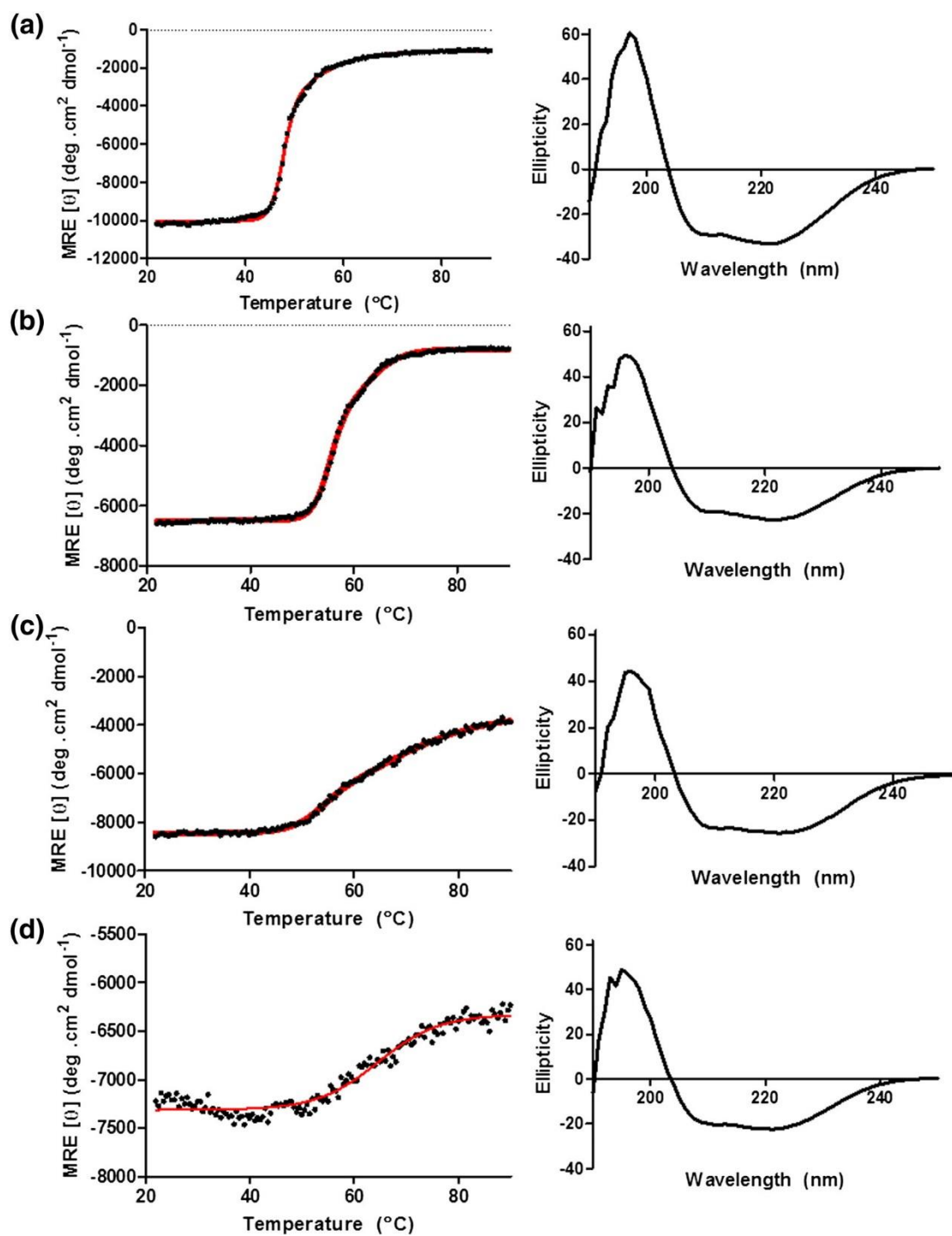
SI Fig. 4. Thermally induced unfolding of *Lc*αE7 and the *Lc*αE7-4a oligomeric species. Ellipticity as a function of temperature was monitored at 208 nm and curves were fit by non-linear regression with Eq. (1). For the *Lc*αE7-4a monomer and dimer data, a three-state model was used to fit the data. Spectra of the protein before denaturation are shown next to the denaturation curves. (a) *Lc*αE7. (b) *Lc*αE7-4a monomer. (c) *Lc*αE7-4a dimer. (d) *Lc*αE7-4a tetramer.
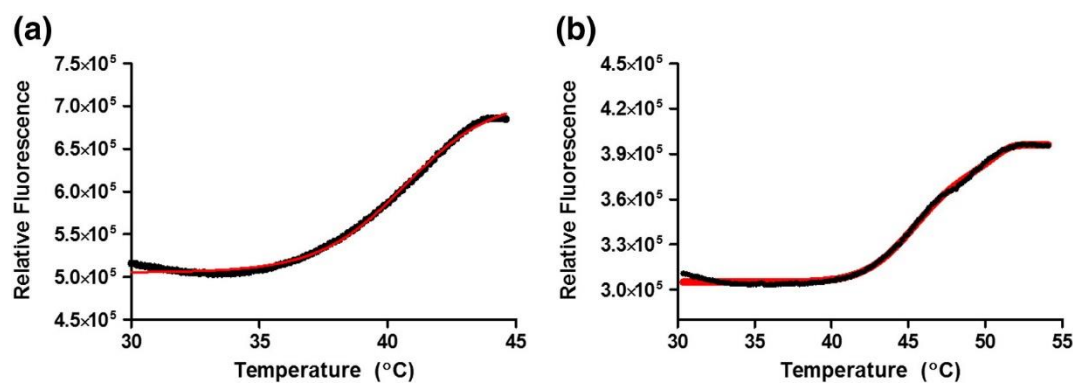
SI Fig. 5. DSF analysis of pre-SEC samples of *Lc*αE7 (2 mg/mL). (a) WT *Lc*αE7 undergoes a single transition, with an apparent melting point of 40.6 ± 0.6 °C, whereas *Lc*αE7-4a (b) is composed of one species with an apparent melting point of 44.8 ± 0.5 °C and a second species with an apparent melting point of 50.9 ± 0.9 °C.
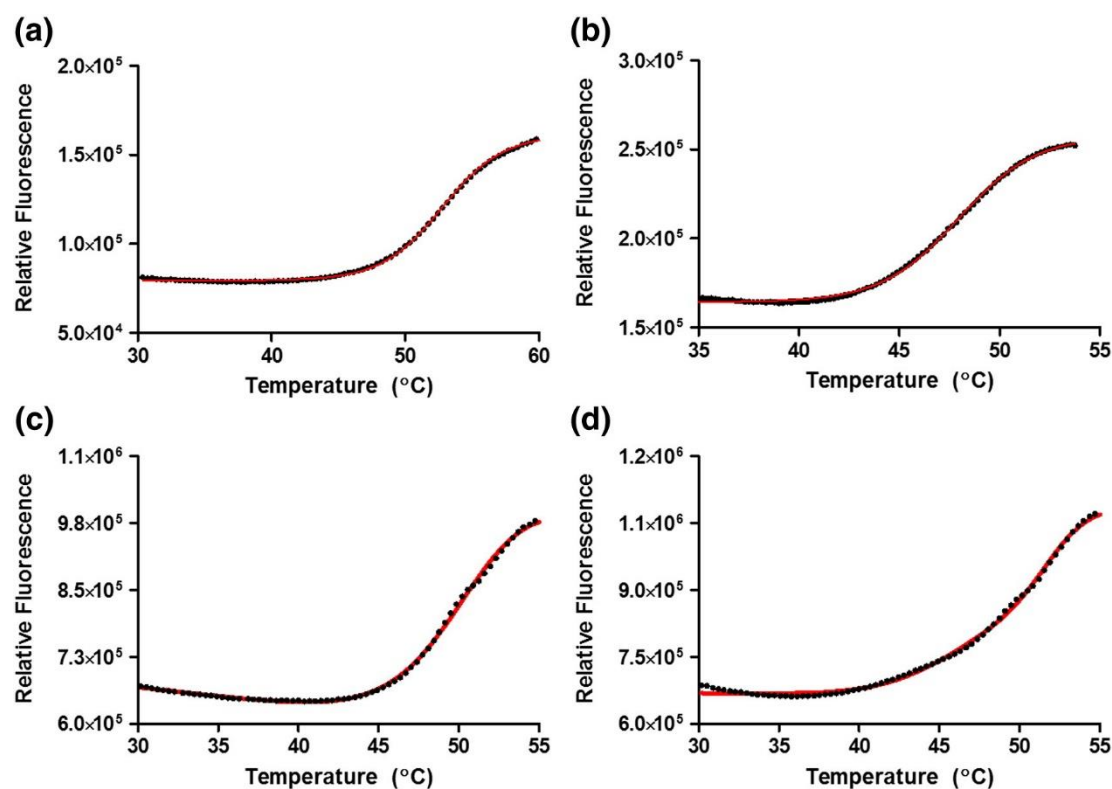
SI Fig. 6. DSF analysis of monomeric and tetrameric fractions of *Lc*αE7-4a. (a) Tetrameric fraction of *Lc*αE7-4a at 0.3 mg/mL shows a single apparent melting point at 52.8 °C, whereas monomeric fraction of *Lc*αE7-4a (b) at 0.3 mg/mL shows a single transition at 47.9 °C. After concentration to 3 mg/mL, some re-equilibration occurred and both tetrameric and monomeric samples (c and d) showed two transitions, one corresponding to monomer (~ 47 °C) and the other corresponding to tetramer (~ 53 °C).
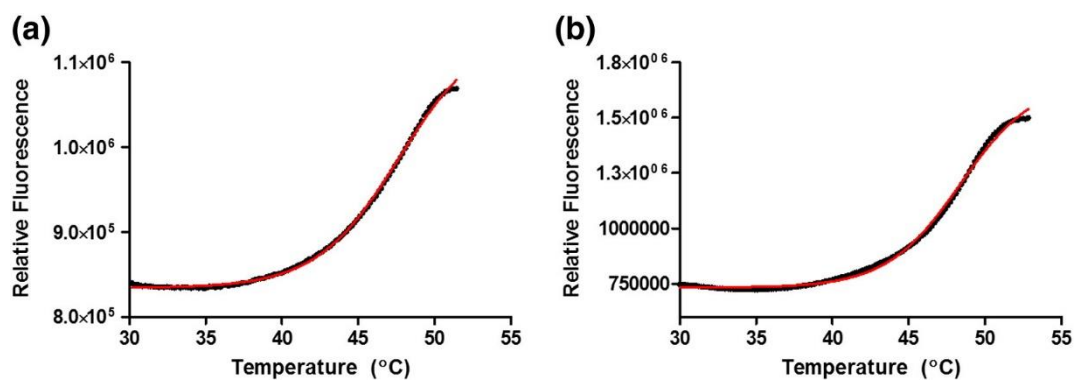
SI Fig. 7. Thermostability effects of the A285S and F478L mutations. Both mutations were introduced in to *Lc*αE7-4a. Thermostability was estimated by DSF. (a) Monomeric fraction of *Lc*αE7-4a (A285S) at 0.3 mg/mL shows a single apparent melting point at 47.6 ± 1.3 °C, within error of *Lc*αE7-4a (47.9 ± 0.2 °C). (b) Monomeric fraction of *Lc*αE7-4a (F478L) at 0.3 mg/mL shows a two apparent melting points: a major species at 48.2 ± 0.9 °C, within error of *Lc*αE7-4a (47.9 ± 0.2 °C), and a minor species at 51.2 ± 0.2 °C, which was not higher than the higher-order species in *Lc*αE7-4a (52.5 ± 0.3 °C). Thus, A285S and F478L are both likely to be selectively neutral in terms of thermostability to *Lc*αE7.

Multiple sequence alignment (positions 10–570):

```
              10        20        30        40        50        60        70        80        90
L. cuprina   MNFNVSLMEKLKWKIKCIENKFLNYRLTTNETVVAETEYGKVKGVKRLTVYDDSYYSFEGIPYAQPPVGELRFKAPQRPTPWDGVRDCCNHKDKS
C. stygia    ....a....r.........y..q.............i....i.....................l.............e....nr...
C. hominivorax ...k..qv.......f.......s....a.........i.....................l................n....
M. domestica ...k..q..r.s..l..mv..yt...s....qiid...qi.....m..........s....k...........v..e.....gpanr.
H. irritans  ......fl...r....v...i...i...hivd....i.......f.......k..........v.....k...haasr.

              100       110       120       130       140       150       160       170       180       190
L. cuprina   VQVDFITGKVCGSEDCLYLSVYTNNLNPETKRPVLVYIHGGGFIIGENHRDMYGPDYFIKKDVVLINIQYRLGALGFLSLNSEDLNVPGNAGLKD
C. stygia    ..r..................t.................lm.....v.......v......t...........
C. hominivorax ...............a.........d.v.....ey..........t.....v.......e........
M. domestica ..t...s..pt.........n....d...dk...m.f....d..f..an.nwf.....m..p...vtv...v.....k..n........
H. irritans  ..t...s.nss........n.......td.......ff.......c..an.ny..a.......f.tv...v........n........

              200       210       220       230       240       250       260       270       280
L. cuprina   QVMALRWIKNNCANFGGNPDNITVFGESAGAASTHYMMLTEQTRGLFHRGILMSGNAICPWANTQCCQHRAFTLAKLAGYKGEDNDKDVLEFLMKA
C. stygia    ..............................g..................v...........s.e....yni.........n...........l.v
C. hominivorax ............................g..a.................v....is.n....yai...n.....n.........
M. domestica .......v.s..dv.......g.......i.........m....sm.s..s.e..s..l.m..rv.....e.i......
H. irritans  ...........s...d..c..l........i...a.....av....t.m.i.h.....gy.i..ri.....n.....yd.....

              290       300       310       320       330       340       350       360       370       380
L. cuprina   KPQDLIKLEEKVLTLEERTNKVMFPFGPTVEPYQTADCVLPKHPREMVKTAWGNSIPTMMGNTSYEGLFFTSILKQMPMLVKELETCVNFVPSEL
C. stygia    .s.......d....p.hk.....a....i....s.e..i...............l.y..i.......avl..md..a...p..
C. hominivorax .ah......d....p.hv.....a..........................l...p.v.....a.l....a....t..
M. domestica n.y.....e.pq...p..mq...........v..pi....s.........li.......l.k..a..y.ev.....s...y..w..
H. irritans  n.y..are.h....n..lrd...a....t...e.p......pn...l.........li.......l.i.vg..n.h.i.....fecy..g..

              390       400       410       420       430       440       450       460       470
L. cuprina   ADAERTAPETLEMGAKIKKAHVTGETPTADNFMDLCSHTYFWFPMHRLLQLRFNHTSGTPVYLYRFDFDSEDLINPYRIMRSGRGVKGVSHADEL
C. stygia    ...................t......................i.........y...i.......
C. hominivorax ..s..ssa....l.........n...l....f........i.k.........ei.......h.........
M. domestica ..s..s......ra.iv.....d....l....e...yf.l....f........a...i......ei......f.........
H. irritans  vved.ss...s..iasil..ly.r.....les.te...df..y....f.......v.s.i.........ei.....y......a..t...

              480       490       500       510       520       530       540       550       560       570
L. cuprina   TYFFWNQLAKRMPKESREYKTIERMTGIWIQFATTGNPYSNEIEGMENVSWDPIKKSDEVYKCLNISDELKMIDVPEMDKIKQWESMFEKHRDLF
C. stygia    ..l...l...............s.v..............la.............vv.....e....1..n....
C. hominivorax ..l...a....l............v....t.........d.l..ii..............i.....e.......ly..rk...
M. domestica ..l...i.s..l..........v..te.....k...d.a...lt.....d.......g....vm.l....a.i.d.kke..
H. irritans  ..l..tm.s.....d........i..t..........p..n....tt..sl........m..g...f..l....e.l.v.q.v.n.k.e..
```


	L. cuprina	C. stygia	C. hominivorax	M. domestica	H. irritans	
L. cuprina	—	95.2	93.8	87.7	84.7	% similarity
C. stygia	89.1	—	93.3	86.4	83.6	
C. hominivorax	87.5	86.1	—	87.3	84.2	
M. domestica	76.3	73.5	75.6	—	86.8	
H. irritans	73.8	71.2	73.1	75.0	—	
% identity						

SI Fig. 8. Multiple sequence alignment of *Lc*αE7 with αE7 genes from other fly species (*Calliphora stygia*, *Cochliomyia hominivorax*, *Musca domestica,* and *Haematobia irritans*) reveals significant variation at the positions that were mutated during the directed evolution experiment. Specifically, the I419F, A472T, and I505T sequence differences already exist across the five species compared.

SI Table 1. Calculated Transition Temperature (TT_{50}) for the LcαE7 proteins as measured by DSF. The 3 mg/mL sampled were concentrated prior to analysis.

Protein	Concentration	TT_{50} (° C)
LcαE7-4a Monomer	0.3 mg/mL	47.92 ± 0.2
	3mg/mL	46.0 ± 0.9, 52.5 ± 0.3
LcαE7-4a Tetramer	0.3 mg/mL	52.8 ± 0.3
	3mg/mL	48.0 ± 0.4, 53.4 ± 0.7

Appendix B. Supporting Information for the Research Article in Chapter 3.

Supplementary material:

Molecular basis for the behavioral effects of the odorant degrading enzyme Esterase 6 in *Drosophila*

Faisal Younus [1,2], Nicholas J. Fraser [2], Chris W. Coppin [1], Jian-Wei Liu [1], Galen J. Correy [2], Thomas Chertemps [3], Gunjan Pandey [1], Martine Maïbèche [3], Colin J. Jackson [2], John G. Oakeshott [1]

[1] CSIRO Land and Water, Black Mountain, Canberra, ACT, 2601, Australia

[2] Research School of Chemistry, Australian National University, Canberra, ACT, 2601, Australia

[3] Université Pierre et Marie Curie, Institut d'Ecologie et des Sciences de l'Environnement de Paris, 75252, Paris, France

a
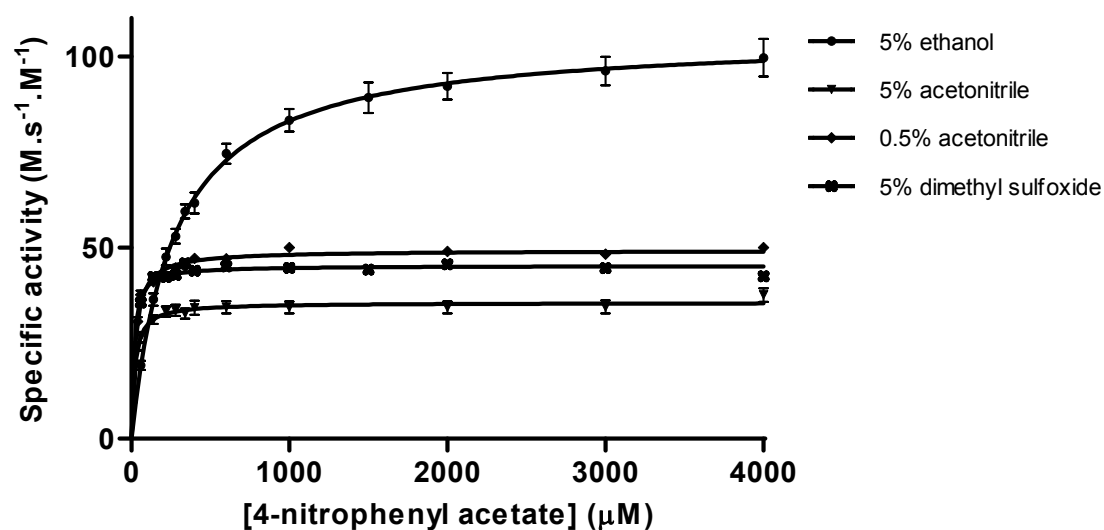
b
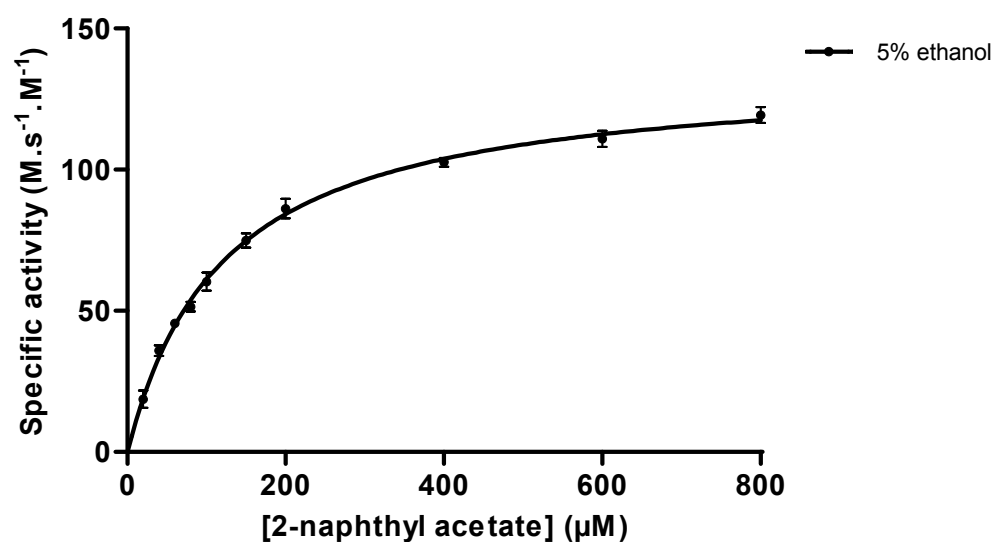
c

Substrate	Solvent conditions	k_{cat} M.s^{-1}	K_M µM	k_{cat}/K_M M^{-1} s^{-1} × 10^3
4-Nitrophenyl acetate	5% ethanol	105 ± 2	269 ± 18	392
	5% acetonitrile	36 ± 1	19 ± 3	1828
	0.5% acetonitrile	49 ± 0	23 ± 1	2111
	5% dimethyl sulfoxide	45 ± 0	14 ± 1	3253
2-Naphthyl acetate	5% ethanol	135 ± 2	121 ± 6	1118

Supplementary Figure S1. EST6 kinetic data towards the model substrates 4-nitrophenyl acetate and 2-naphthyl acetate. **(a)** EST6 Michaelis-Menten plot for 4-nitrophenyl acetate under various solvent conditions. **(b)** EST6 Michaelis-Menten plot for 2-naphthyl acetate. **(c)** EST6 kinetic parameters.

165

| Variant | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | Amino acid position | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2 | 15 | 50 | 55 | 72 | 75 | 76 | 94 | 98 | 106 | 117 | 119 | 145 | 151 | 180 | 182 | 208 | 219 | 225 | 229 | 237 | 242 | 246 | 247 | 263 | 267 | 271 | 273 | 274 | 285 | 290 | 292 | 302 | 305 | 313 | 329 | 331 | 335 | 336 | 351 | 355 | 357 | 378 | 381 | 382 | 383 | 398 | 400 | 416 | 450 | 456 | 459 | 461 | 462 | 464 | 483 | 484 | 485 | 492 | 496 | 498 | 503 | 508 | 511 | 515 | 517 |
| Wt | D | K | K | I | A | A | N | N | N | I | W | N | V | D | Q | V | R | D | K | G | N | S | S | T | T | K | F | Y | V | P | D | I | K | K | V | K | R | I | V | L | R | K | Q | D | I | E | K | S | A | F | F | D | E | M | P | S | D | N | E | K | N | K | R | I | Q | R |
| #1-2 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | R | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | S | . | . | . | . | . |
| #1-3 | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | S | . | . | . | . | . |
| #1-4 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | R | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | S | . | . | . | . | . |
| #1-5 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | S | . | . | . | . | . |
| #1-6 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | S | . | . | . | . | . |
| #2-1 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | . | . | S | . | . | . | . | . |
| #2-2 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | S | . | . | T | . | . |
| #2-3 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | D | . | . | S | . | . | . | . | . |
| #2-4 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | S | . | . | . | . | . |
| #2-5 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | S | S | . | . | . | . | . | . | . | S | . | . | . | . | . |
| #2-6 | . | . | . | . | . | . | . | . | . | N | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | S | . | . | . | . | . |
| #2-7 | . | M | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | S | . | . | T | . | . |
| #2-8 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | S | . | . | . | . | S | . | . | . | . | . |
| #2-9 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | S | . | . | T | . | . |
| #2-10 | . | . | . | . | . | . | . | . | . | . | . | . | D | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | V | . | . | . | . | . | S | . | . | . | . | . |
| #3-1 | V | . | . | . | . | . | . | . | . | . | . | . | L | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | S | . | . | . | . | . | . | . | . | S | . | . | T | . | . |
| #3-2 | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . | . | V | S | S | . | . | . | . | . | . | . | . | S | . | . | . | . | . |
| #3-3 | . | M | . | . | . | . | . | . | . | . | . | . | . | . | . | L | . | . | . | . | . | . | . | . | . | . | R | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . | . | . | S | S | . | G | . | . | . | . | . | . | . | . | . | T | . | . |
| #3-4 | A | . | . | . | . | . | . | . | . | . | . | N | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | D | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| #3-5 | . | . | . | . | . | . | . | D | . | . | . | K | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | V | . | S | . | . | . | . | . | . | . | . | . | . | S | . | . | T | . | . |
| #3-6 | V | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| #3-7 | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | L | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | V | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| #3-8 | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | S | . | . | . | . | . | . | . | T | . | K |
| #3-9 | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | K | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | Q | . | . | . | . | . | . | . | . | . | . | . | . | S | . | . | . | . | . | . | . | . | . | R |
| #3-10 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | R | Q | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . | S | . | . | . | . | . | . | . | . | D | . | . | . | . | . | . | . | . |
| #4-2 | . | . | . | . | . | . | . | . | . | . | . | . | . | K | . | . | . | . | . | A | . | . | . | . | . | . | G | F | . | . | . | . | . | V | . | . | . | . | . | . | . | . | . | . | . | G | . | G | . | S | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | K |
| #4-3 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | Q | . | . | . | . | . | . | . | . | I | . | . | . | . | . | . | . | . | G | V | S | S | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | R | . | . |
| #4-4 | . | . | . | . | . | . | . | . | . | . | . | . | L | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . | S | S | . | . | . | . | . | . | . | . | . | . | . | . | . | . | R | . | . |
| #4-5 | A | V | . | . | . | . | . | V | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | S | . | . | . | D | . | . | . | . | . | . | . | T | . | . |
| #4-6 | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . | . | G | . | S | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . |
| #4-7 | . | . | . | . | . | . | . | . | . | . | . | . | L | . | . | I | K | . | . | A | . | . | . | . | R | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | V | S | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . |
| #4-8 | . | . | E | . | . | . | . | . | . | . | . | . | . | . | . | L | . | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . | . | T | . | . |
| #4-9 | . | . | . | . | . | T | . | . | . | . | . | . | R | . | K | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | I | . | . | D | . | L | . | . | . | . | . | S | . | . | G | . | D | . | . | . | T | . | . |
| #4-10 | . | . | . | . | T | . | . | . | . | . | . | . | K | . | . | . | . | . | . | . | . | . | G | F | . | . | . | . | . | . | . | . | . | . | . | . | . | G | R | G | V | S | S | . | . | L | . | . | G | . | . | G | T | . | G |
| #4-11 | A | M | . | . | . | T | . | . | . | . | . | . | L | . | . | . | . | . | . | P | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . | . | G | V | S | . | . | V | . | . | . | Q | . | T | . | . |
| #5-1 | . | . | . | . | . | . | . | . | . | . | . | . | L | . | . | . | . | . | . | A | . | . | . | . | . | G | . | . | . | . | . | . | . | . | G | . | . | . | V | S | S | . | . | . | . | . | D | . | . | . | T | . | . |
| #5-2 | . | . | . | . | . | . | . | D | L | . | . | . | . | . | . | A | . | S | . | . | G | . | . | . | . | I | . | . | . | G | . | G | V | S | S | . | . | . | D | . | . | Q | . | T | . | E |
| #5-3 | . | . | . | T | . | . | . | . | L | . | K | . | . | L | . | A | . | . | . | G | . | . | . | . | . | . | . | G | V | S | . | . | . | . | . | D | . | . | . | T | . | E |
| #5-4 | . | . | . | . | T | D | . | D | . | . | L | . | . | K | . | . | . | A | . | . | . | . | . | . | V | . | . | . | G | . | . | G | V | S | . | . | . | . | D | G | . | . | T | . | K |
| #5-5 | . | . | . | . | . | . | . | . | L | . | . | . | . | . | . | A | . | . | . | G | . | . | . | . | . | G | . | . | G | V | S | S | . | . | . | . | D | . | . | . | T | R | . |
| #5-6 | . | . | . | V | . | . | . | D | L | . | . | K | . | . | E | . | A | . | . | . | . | . | . | . | G | V | S | . | . | . | L | . | . | . | . | . | T | . | . |
| #5-7 | . | V | . | . | . | . | . | D | L | . | . | K | . | . | . | A | . | . | . | . | . | . | . | G | V | S | S | . | . | . | . | . | . | . | . |
| #5-8 | . | V | . | . | . | . | . | . | . | . | L | . | . | K | . | E | E | . | A | . | . | . | . | . | V | . | . | . | G | . | . | G | V | S | . | . | . | . | . | . | T | . | G |
| #5-9 | . | . | . | . | . | . | . | . | L | . | . | . | . | E | . | A | . | . | . | G | . | . | . | . | . | G | V | S | S | . | V | . | . | . | G | . | . | T | . | . |
| #5-10 | . | M | . | . | T | . | . | . | L | . | K | . | . | . | A | . | . | . | . | A | . | . | . | . | G | . | . | G | V | S | . | . | . | . | . | D | . | . | . | T | . | . |
| #6-1 | . | V | . | . | . | . | . | . | . | . | L | . | . | K | . | E | S | . | A | . | . | . | . | G | F | . | . | . | . | V | . | . | . | . | . | G | . | G | V | S | S | . | . | . | . | . | . | D | . | . | . | . | . | T | . | . |

Supplementary Figure S2. Sample of EST6 replacement changes observed over six rounds of directed evolution. Variable amino acid positions and the status of EST6 Wt are indicated along the top with amino acids numbered from the first residue of the mature EST6 protein as it would be processed in its native form within the fly. Sequence variation in the six rounds of random mutagenesis and directed evolution are indicated below where a dot "." indicates no change. Variant #6.1, which was used for the crystallography, is simply labelled EST6-1 in the main text.

Supplementary Figure S3. Comparison of the catalytic activities of wildtype and evolved EST6 against 11 substrates. Specific activities at 200 μM substrate of wildtype EST6 compared to the evolved EST6-1.
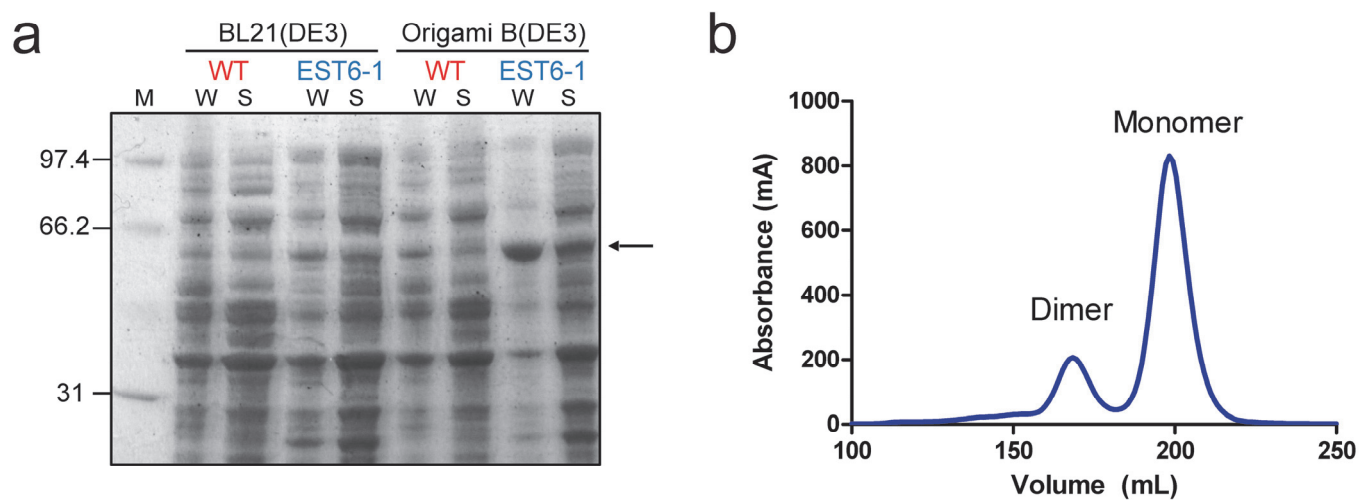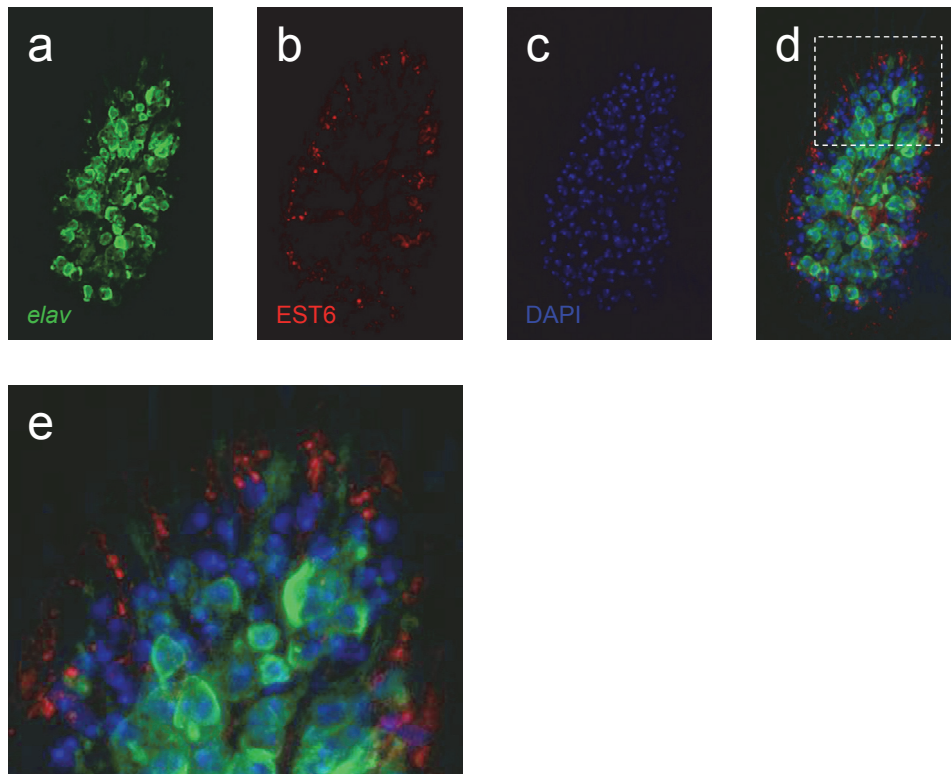
Supplementary Figure S4. Expression and purification of EST6-1. **(a)** Expression gel of whole cell (W) and soluble (S) EST6WT and EST6-1 in the BL21 expression strain and the Origami B expression strain with soluble EST6-1 marked with an arrow. M, molecular markers with sizes indicated in kDa. **(b)** Size exclusion chromatogram of EST6-1 displaying the predominant monomer species used for crystallization.
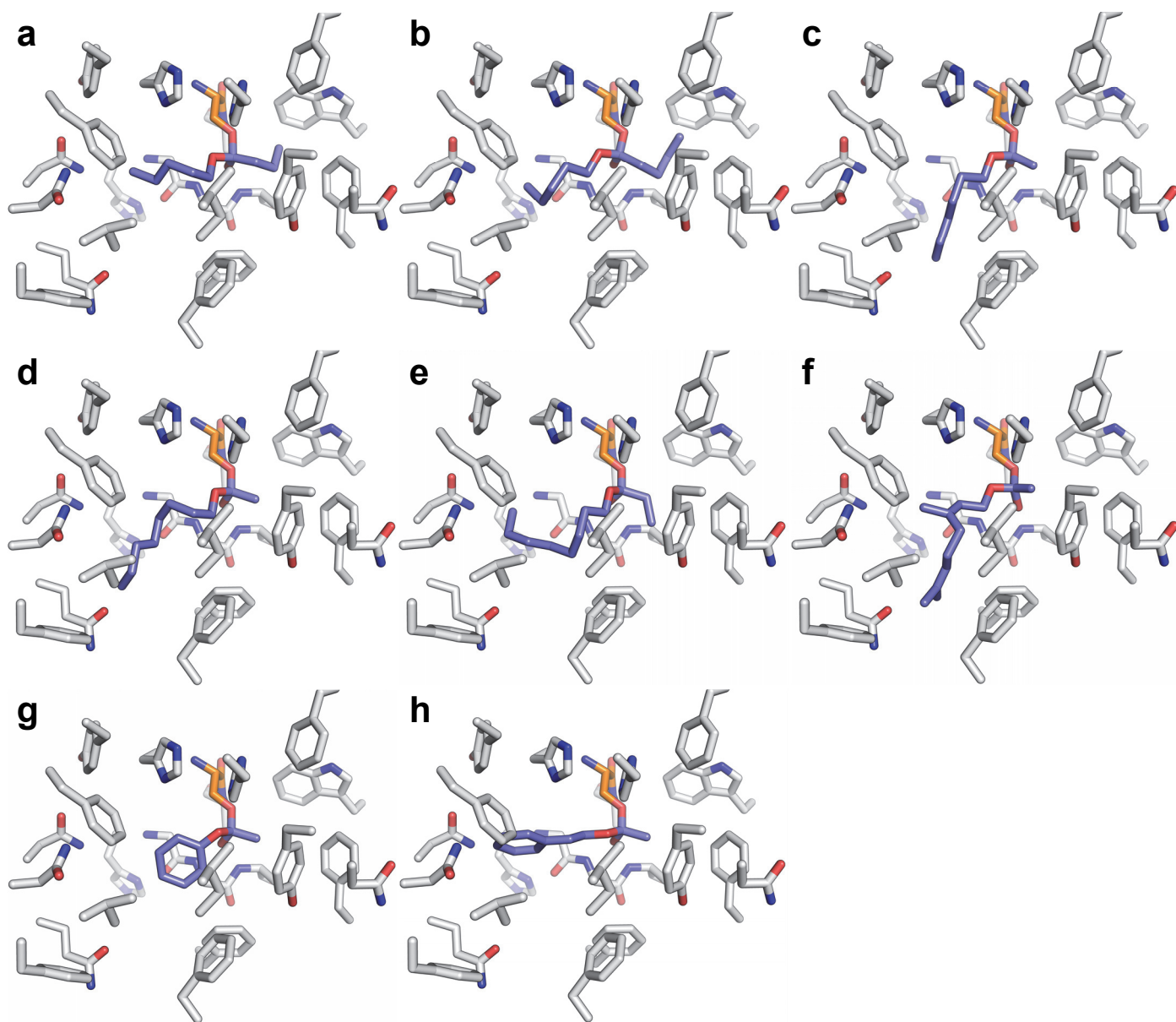
Supplementary Figure S5. Docking of representative substrates in EST6. A series of representative substrates/acyl intermediates were covalently docked into EST6: **(a)** pentyl butyrate, **(b)** pentyl pentanoate, **(c)** hexyl acetate, **(d)** octyl acetate, **(e)** octyl propionate, **(f)** geranyl acetate, **(g)** phenethyl acetate, **(h)** phenyl acetate.
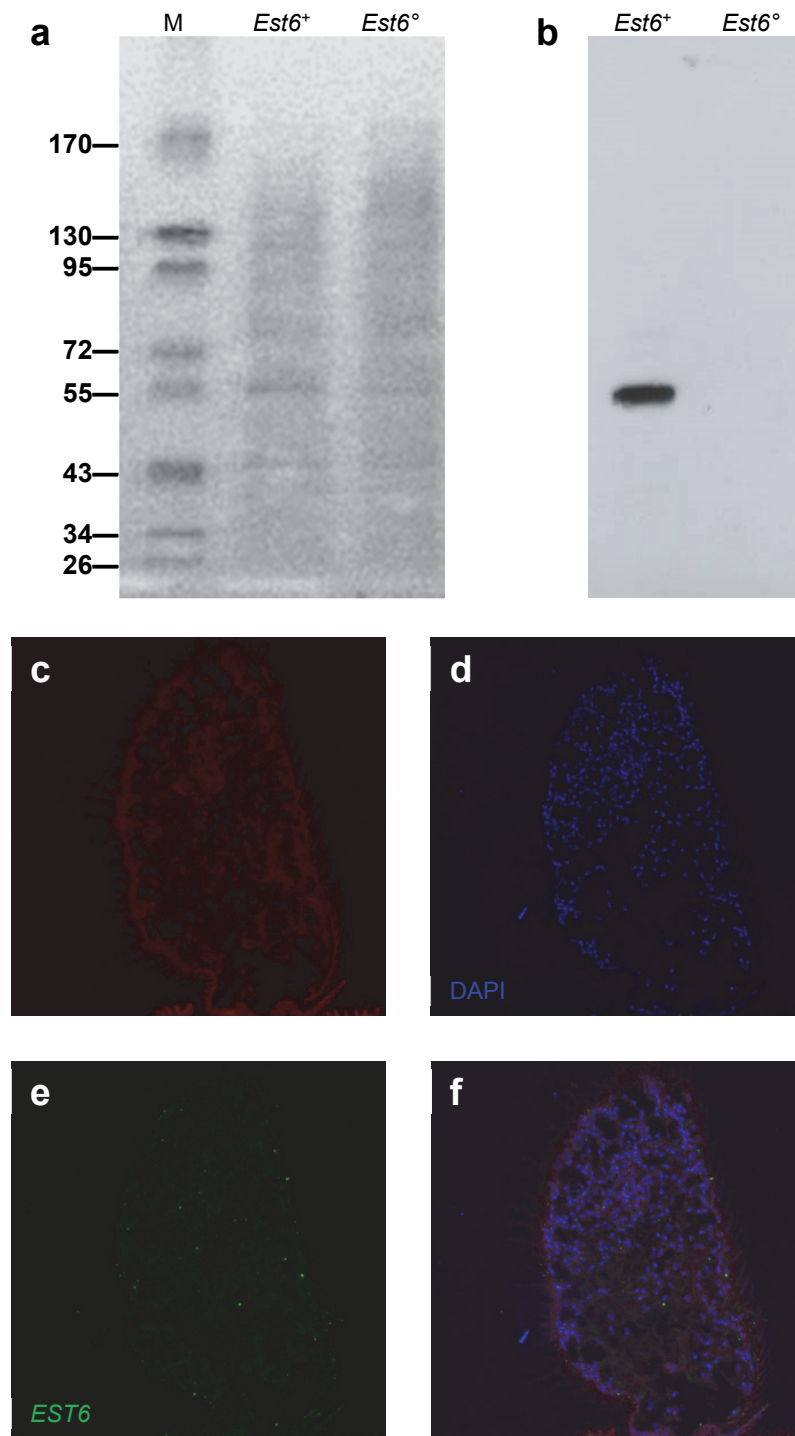
Supplementary Figure S6. Western blot and immunohistochemistry with anti-EST6 antibody showing its specificity for EST6. **(a)** Protein staining of head extracts from wildtype (*Est6*⁺) and *Est6* null mutant (*Est6°*) flies separated by SDS –PAGE. M, molecular markers with sizes indicated in kDa. **(b)** Same samples showing the location of the bound anti-EST6 antibody. **(a**) Section from third antennal segment of *Est6* null mutant showing autofluorescence of the cuticle. **(b)** DAPI staining of the nuclei. **(c)** EST6 localization on the same section: no labelling with anti-EST6 antibody was detected. **(d)** Merge image of (a), (b) and (c).

Supplementary Figure S7. EST6 and *elav* expression in the third antennal segment. **(a)** Membrane-tethered GFP expressed with the *elav* promoter (*elav^LexA^/LexAOP-mCD8::GFP* transgenic flies). **(b)** EST6 protein localization in the same section. **(c)** DAPI staining of the nuclei. **(d)** Merge image of (a), (b) and (c): Est-6 and *elav* are not expressed in the same cells. **(e)** Higher magnifications of (c): EST6 protein surrounds the *elav⁺* dendrites.
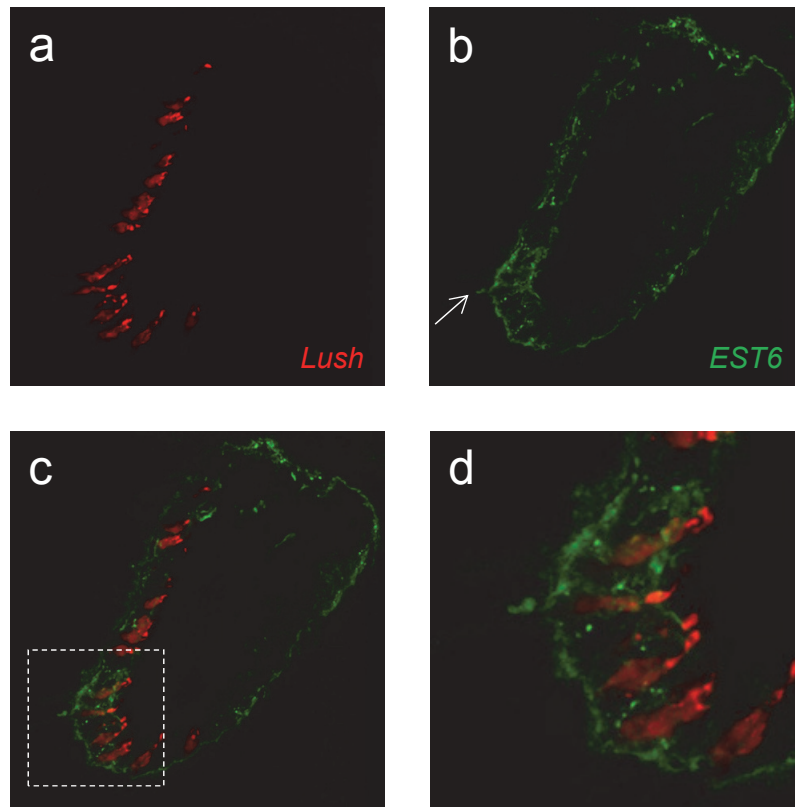
Supplementary Figure S8. EST6 and *Lush* expression in the third antennal segment, transversal sections. **(a)** Membrane-tethered RFP expressed with the *lush* promoter (*lush^Gal4^/UAS-mCD8::RFP* transgenic flies). **(b)** EST6 protein localization in the same section. Arrow indicate sensillum. **(c)** Merge image of (a) and (b): EST6 and *lush* are not expressed in the same cells. **(d)** Higher magnifications of (c): EST6 protein is secreted in the sensillum lymph.

Supplementary Table S1. EST6 kinetic parameters.

Substrate	$k_{cat}/K_M^{Est‡}$ M^{-1}.s^{-1} × 10^3	K_i μM
Straight chain esters		
Methyl decanoate	22 ± 5	
Methyl myristate	15 ± 8	
Propyl acetate	226 ± 30	
Propyl propionate	897 ± 9	
Propyl butyrate	90 ± 20	
Propyl hexanoate	276 ± 25	
Propyl decanoate	28 ± 10	
Butyl formate	472 ± 38	
Butyl acetate	425 ± 34	
Butyl propionate	966 ± 39	
Butyl butyrate	226 ± 24	
Butyl pentanoate	210 ± 18	
Butyl decanoate	109 ± 13	
Pentyl formate	416 ± 38	
Pentyl acetate	309 ± 49	
Pentyl butyrate	185 ± 15	
Pentyl pentanoate	150 ± 20	
Pentyl hexanoate	271 ± 31	
Hexyl formate	309 ± 27	
Hexyl acetate	540 ± 21	
Hexyl propionate	1055 ± 15	
Hexyl butyrate	332 ± 23	
Hexyl hexanoate	72 ± 36	
Heptyl acetate	579 ± 42	
Heptyl propionate	577 ± 124	
Heptyl butyrate	200 ± 15	
Octyl acetate	416 ± 23	
Octyl propionate	1343 ± 12	
Octyl butyrate	365 ± 28	
Nonyl acetate	698 ± 38	
Decyl acetate	740 ± 19	
Cyclic alcohol moieties		
Cyclohexyl acetate	102 ± 8	
Benzyl formate	156 ± 42	
Benzyl acetate	328 ± 22	430 ± 40
Phenyl acetate	298 ± 13	190 ± 0
Phenethyl acetate	142 ± 9	880 ± 40
4-Nitrophenyl acetate*	224 •	
2-Naphthyl acetate*	399 •	
Branched and unsaturated alcohol moieties		
Isobutyl acetate	369 ± 48	
Isopentyl acetate	315 ± 38	
2-Methylbutyl acetate	213 ± 18	
cis-3-Hexenyl acetate	295 ± 32	
E2-hexenyl acetate	689 ± 32	
Geranyl formate	375 ± 42	
Geranyl acetate	378 ± 83	
Neryl acetate	421 ± 105	
Modified acid moieties		
Butyl 2-methylbutyrate†	26 ± 8	
Hexyl 2-methylbutyrate	220 ± 29	

† k_{cat}/K_M^{Est} was taken from group assay results.
‡ k_{cat}/K_M^{Est} was below 1.5 x 10^4 for the following substrates: methyl propionate, methyl butyrate, methyl pentanoate, methyl hexanoate, methyl heptanoate, methyl octanoate, methyl laurate, methyl palmitate, ethyl propionate, ethyl butyrate, ethyl pentanoate, ethyl hexanoate, ethyl heptanoate, ethyl octanoate, ethyl decanoate, linalyl acetate, cis-vaccenyl acetate, ethyl lactate, methyl 3-hydroxyhexanoate, ethyl (R)-3-hydroxybutyrate, ethyl (S)-3-hydroxybutyrate, ethyl 3-hydroxybutyrate, ethyl 3-hydroxyhexanoate, ethyl 2-methylbutyrate, ethyl isovalerate, butyl isovalerate, ethyl trans-2-butenoate, hexyl trans-2-butenoate, ethyl tiglate, methyl benzoate, ethyl benzoate, methyl salicylate, ethyl cinnamate, methyl jasmonate, isopentyl tiglate, dimethyl carbonate, dimethyl glutarate, diethyl succinate and diethyl phthalate. All except cis-vaccenyl acetate were assayed in group assays only.
* Model substrates for which direct UV/vis spectrophotometric assays exist.
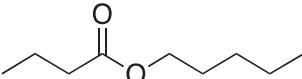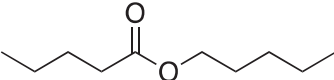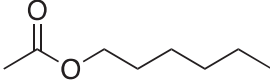• Value back calculated from full Michaelis-Menten kinetic data (Supplementary Fig. S1).
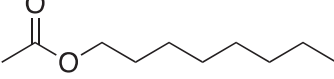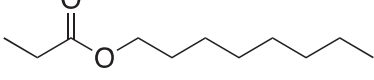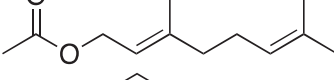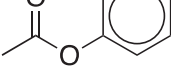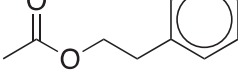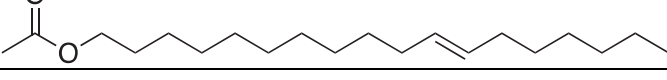
Supplementary Table S2. Compounds from Fig. 1 for which there is evidence of bioactivity in Drosophila.

Substrate	Nature of Evidence		References
	In vitro activity against known OR	Behavioural differences	
Straight chain esters			
Methyl myristate	✓	✓	[1]
Propyl acetate	✓	✓	[2]
Propyl propionate	✓		[2]
Propyl butyrate	✓		[3]
Propyl hexanoate	✓		[2]
Butyl acetate	✓	✓	[2]
Butyl butyrate	✓		[3]
Butyl pentanoate	✓		[2]
Pentyl acetate	✓	✓	[2]
Pentyl pentanoate	✓		[3]
Hexyl acetate	✓	✓	[2]
Hexyl butyrate	✓	✓	[2]
Hexyl hexanoate	✓	✓	[2]
Cyclic alcohol moieties			
Benzyl acetate	✓		[2]
Phenethyl acetate	✓	✓	[2]
Branched and unsaturated alcohol moieties			
Heptyl acetate	✓		[2]
Octyl acetate	✓		[2]
Nonyl acetate	✓		[2]
Decyl acetate	✓		[2]
Isobutyl acetate	✓		[2]
Isopentyl acetate	✓	✓	[2]
2-Methylbutyl acetate	✓		[2]
cis-3-Hexenyl acetate	✓		[2]
E2-Hexenyl acetate	✓	✓	[2]
cis-Vaccenyl acetate	✓	✓	[2]
Geranyl formate	✓		[2]
Geranyl acetate	✓		[2]
Neryl acetate	✓		[2]

Supplementary Table S3. Proteins closest in structure to EST6 using the Q-score calculated by the SALAMI server accessed on May 2016[4].

Protein Name	PDB Code	Seq ID (%)	Q Score
Manducta sexta JHE	2fj0A	22	0.9435
Lucilia cuprina αE7	4fnmA	26	0.9321
Drosophila melanogaster AChE	1qo9A	25	0.9001
Torpedo californica AChE	3m3dA	27	0.9000
Homo sapiens AChE	4m0eA	26	0.8985
Homo sapiens BChE	2j4cA	25	0.8967
Bos taurus Lipase	1aqlB	28	0.8961
Mus musculus AChE	2c0pA	26	0.8912
Ophiostoma piceae Esterase	4be4A	22	0.8861
Homo sapiens Lipase	1jmyA	29	0.8843

Supplementary Table S4. DOCKovalent[5] binding scores for the representative range of potential substrates docked into the active site of EST6.

Ester	Structure	Score*
Pentyl butyrate		-24.8
Pentyl pentanoate		-6.04
Hexyl acetate		-29.1
Octyl acetate		-22.0
Octyl propionate		-22.4
Geranyl acetate		-18.3
Phenyl acetate		0.63
Phenethyl acetate		-26.5
cis-Vaccenyl acetate		NA[†]

* Total score is calculated according to the physics-based scoring function implemented in DOCK3.7. The scoring function evaluates the ligand-protein electrostatic and van der Waals interactions and corrects for desolvation[5].

[†] No suitable binding poses (without steric clashes) obtained.

Supplementary Table S5. Data collection and refinement statistics for EST6-1. The diffraction data were indexed, integrated and scaled using Aimless and XDS[6,7]. Molecular replacement was used to obtain phases using the structure of LcαE7 with MOLREP[8]. Model building was carried out via iterative cycles of automated model building with BUCCANEER[9] and manual rebuilding with COOT[10]. Refinement was undertaken using REFMAC V 5.7.0032 accessed through the CCP4 suite of programs[11,12] and phenix.refine[13].

	EST6-1
Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
a, b, c (Å)	67.7, 80.6, 107
α, β, γ (°)	90, 90, 90
Resolution (Å)	40.32-2.15
	(2.22-2.15)
R_{sym} or R_{merge}	0.114 (0.331)
$CC_{1/2}$	0.996
CC*	0.995
$I / \sigma / I$	18.1 (4.51)
Completeness (%)	96.6 (73.4)
Redundancy	11.7 (10.4)
Refinement	
Resolution (Å)	40.32-2.15
	(2.52-2.15)
No. reflections	31400
R_{work} / R_{free}	0.1554/0.2104
	(0.1698/0.2325)
No. atoms	
Protein	520
Ligand/ion	1
Water	353
B-factors	
Protein	19.44
Ligand/ion	25.13
Water	22.11
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	0.95
PDB ID	5THM

*Values in parentheses are for highest-resolution shell.

Supplementary Text

Supplementary Results

In order to corroborate the results of the immunohistochemical experiment showing different localization of EST6 and LUSH within the antennae we performed a set of male-male courtship experiments on *Est6* knock-down flies bearing RNAi constructs driven by *Est6* or *lush* Gal4 lines to demonstrate that *Est6* is not co-expressed with *lush*. Previous work has shown that *Est6°* null mutant flies exhibit lower levels of male-male courtship[14]. As expected, we observed a significant reduction in the male-male courtship index (CI) when the *UAS-RNAi* transgene was driven by *Est6*[Gal4] (12.20 ± 3.27% for *Est6*[Gal4]/*UAS-RNAi Est6*), whereas the CI when the *UAS-RNAi* transgene was driven by *lush*[Gal4] (29.29 ± 4.48 for *lush*[Gal4]/*UAS-RNAi Est6*) was not significantly different from the various parental genotypes (27.24 ± 4.88 for *lush*[Gal4], *25.60 ± 4.75 for Est6*[Gal4] and 23.00 ± 4.76 for *UAS-RNAi Est6).* This result strongly suggests that EST6 and LUSH are not produced by the same cells of the sensilla since even an overlap of cells expressing both proteins would be expected to result in an intermediate phenotype.

Supplementary Methods

Flies. Additional lines used are *UAS-RNAi Est6* flies from the NIG Stock Center (Japan) and transgenic *Est6*[Gal4] flies described in Chertemps *et al.*[14]. Both stocks were maintained as described in the main text.

Courtship assays. Male-male courtship assays were performed as described in Chertemps *et al.*[14]. Briefly, all experiments were done under dim red lights at 25 ˚C (50% to 60% relative humidity). A single male (5 to 7 days old) was placed in a test chamber (30 mm diameter, 5 mm height) for 10 minutes before introducing a decapitated *CS* 'target' male. Courtship behaviour was observed over 10 minutes and a courtship index (CI) was calculated as the fraction of time spent in courtship activity over the observation period. The CIs of the parental control strains *lush*[Gal4], *Est6*[Gal4] and *UAS-RNAi Est6* were compared to the CIs of the *lush*[Gal4]/*UAS-RNAi Est6* and *Est6*[Gal4]/*UAS-RNAi Est6* strains, in which *Est6* inhibition is driven by either the *lush* or *Est6* promoters. At least 22 flies of each genotype were tested.

Supplementary References

1   Dweck, H. *et al.* Pheromones mediating copulation and attraction in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **112**, doi:10.1073/pnas.1504527112 (2015).

2   Dweck, H. *et al.* Olfactory Preference for Egg Laying on Citrus Substrates in Drosophila. *Curr. Biol.* **23**, doi:10.1016/j.cub.2013.10.047 (2013).

3   Galizia, C., Münch, D., Strauch, M., Nissler, A. & Ma, S. Integrating heterogeneous odor response data into a common response model: A DoOR to the complete olfactome. *Chem. Senses* **35**, doi:10.1093/chemse/bjq042 (2010).

4   Margraf, T., Schenk, G. & Torda, A. The SALAMI protein structure search server. *Nucleic Acids Res.* **37**, doi:10.1093/nar/gkp431 (2009).

5   London, N. *et al.* Covalent docking of large libraries for the discovery of chemical probes. *Nat. Chem. Biol.* **10**, 1066-1072, doi:10.1038/nchembio.1666 (2014).

6   Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, doi:10.1107/S0907444905036693 (2006).

7   Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, doi:10.1107/S0907444909047337 (2010).

8   Vagin, A. & Teplyakov, A. MOLREP: an Automated Program for Molecular Replacement. *J. Appl. Cryst.* **30**, doi:10.1107/S0021889897006766 (1997).

9   Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, doi:10.1107/S0907444906022116 (2006).

10   Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).

11   Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, doi:10.1107/S0907444911001314 (2011).

12   Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235-242, doi:10.1107/s0907444910045749 (2011).

13   Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352-367, doi:10.1107/s0907444912001308 (2012).

14   Chertemps, T. *et al.* A carboxylesterase, Esterase-6, modulates sensory physiological and behavioral response dynamics to pheromone in Drosophila. *BMC Biol.* **10**, doi:10.1186/1741-7007-10-56 (2012).