Structural and Functional Characterisation of Human Carboxylesterases

Victoria Arena De Souza, Corpus Christi College, University of Oxford
D.Phil, Nuffield Department of Clinical Medicine
Michaelmas Term 2014

Carboxylesterases are glycosylated general detoxification enzymes belonging to the serine esterase superfamily and play a critical role in the hydrolysis of numerous ester- and amide- containing molecules, including active metabolites, drugs and prodrugs. Three functionally active carboxyleterases have been identified in man (CES1-3), which all show differential tissue expression and critically overlapping, yet specific substrate selectivities. Elucidating the basis of their exact substrate preference would help facilitate the design of clinical prodrugs which are activated by carboxylesterases. Because of their widespread applications, carboxylesterases have attracted much attention in recent years, with CES1 being the most extensively studied human carboxylesterase to date.

The work presented here addresses the structure-function relationship of the three human carboxylesterases using a combination of X-ray crystallography, kinetic analysis and biophysical techniques. Recombinant proteins were successfully produced using a mammalian expression system in high yield (5.0 – 84.0 mg/ L cell culture). Analytic ultracentrifugation and size-exclusion chromatography coupled to multi-angle laser light scattering were used to investigate the proteins in solution. These results showed CES1 exists primarily in a trimeric arrangement, whilst CES2 and CES3 are monomeric. Interestingly, atypical mechanisms of substrate inhibition, positive cooperativity and biphasic kinetics were observed for both CES1 and CES2.

Three structures of CES1 were solved: wild type, an aglycosylated form and a catalytically inactive form, to 1.48, 1.86 and 2.01 Å respectively. The novel structure of CES2 was solved to 2.04 Å, which revealed that the enzyme forms a strand exchange dimer in contrast to the trimeric CES1.

To summarise, this thesis documents a platform that has been generated for the production, characterisation and crystallization of human carboxylesterases. This will aid future structural work for protein ligand binding studies.

Declaration of work

The work described herein is my own, except as described below.

Chapters 3, 4 and 5: The analytical ultracentrifugation (AUC) experiments and analysis, and the multi-angle laser light scattering (MALLS) analysis were conducted by Dr. David J Scott, Associate Professor and Reader in Physical Biochemistry of the University of Nottingham.

Acknowledgements

First and foremost it I would like to express my greatest thanks to my supervisors, Prof. Ray Owens and Dr. Martin Walsh, for their constant support, patience and knowledge throughout this project. I would not have achieved the results that I did if it were not for your guidance and enthusiasm. I cannot thank you both enough. I am very also grateful to Diamond Light Source and Chroma Therapeutics and the University of Oxford Nuffield Department of Medicine for providing the funds for this studentship.

I would like to acknowledge the contributions made to this project by Dr. Michael Charlton and Dr. David Krige, former advisors from Chroma Therapeutics Ltd, for their advice and expertise, especially in the first year of the project. Also to Tricia Kirwin-Jones for her help with the original activity assays at Chroma.

A big thankyou to Dr. David Scott for help with all the biophysical experiments and analysis, and to Gemma Harris for assistance too.

I would like to take this opportunity to thank everyone in the OPPF-UK who has contributed to this project and helped me during my time here, making it such an enjoyable and memorable experience. In particular, I would like to thank Heather Rada for all her help with tissue culture and preparing the activity assays, and Anil Verma for his assistance in protein purification, crystallization, putting up with my constant questions and just being there as a friend. Special friendships were made in this lab! Special thanks also to those that helped on proofreading this thesis!

Finally, none of this would have been achieved without the love and support of my friends and family (and cats!), especially my parents who are simply fantastic, and my amazing husband, Jon. Thank you for being my rock over the last four years, especially in the last year, when it has been tougher than we would have liked!

CONTENTS 4

LIST OF FIGURES 10

LIST OF TABLES 12

LIST OF ABBREVIATIONS 14

1. GENERAL INTRODUCTION

1.1. Properties and Functions of Carboxylesterases 16

1.1.1. General Overview 16

1.1.2. Human Carboxylesterases 17

1.1.3. Expression and Localisation of Human Carboxylesterases 21

1.1.3.1. Expression Profile 21

1.1.3.2. Cellular Localisation 22

1.1.3.3. Protein Trafficking 23

1.1.4. Altered Carboxylesterase Expression and Activity in Disease

1.1.4.1. Cancer 24

1.1.4.2. Metabolism 25

1.1.4.3. Hepatitis C Virus 25

1.1.5. Substrates of Human Carboxylesterases 26

1.1.5.1. Clinically Administered Drugs 29

Selectivity 29

Oncology Drugs 30

1.1.5.2. Narcotics 30

1.1.5.3. Environmental Chemicals 31

1.1.5.4. Endogenous Substrates

Cholesterol Metabolism 32

Vitamin A 33

1.1.6. Inhibitors of Carboxylesterases 33

1.1.6.1. Natural Inhibitors 35

1.2. Structure and Catalytic Function of Carboxylesterases 36
1.2.1. The Three Domains 36
1.2.2. Catalytic Mechanism 39
1.2.2.1. Second Conserved Serine 41
1.2.3. Multiple Binding Sites 41
1.2.3.1. The Z-Site 41
1.2.3.2. The Side-Door 43
1.2.4. Mutations in Human Carboxylesterase Genes 44
1.2.5. Glycosylation 45
1.2.6. Oligomerisation 46
1.3. Importance of Studying Carboxylesterases 47
1.3.1. Scope and Objectives of the Project 48

2. MATERIALS AND METHODS

2.1. Overview 50
2.2. Protein Expression and Purification
2.2.1. Protein sequence analysis 51
2.2.2. Synthetic Gene Design 51
2.2.3. Construction of Expression Vectors
2.2.3.1. PCR 52
2.2.3.2. Agarose Gel Electrophoresis 53
2.2.3.3. PCR Purification 53
2.2.3.4. Cloning with In-Fusion™ Technology 54
2.2.3.5. Transformations 55
2.2.3.6. Plasmid Preparation (Miniprep) 56
2.2.3.7. Construct Verification 56
2.2.3.8. Primer Details 58
2.2.3.9. Plasmid DNA Preparation (Megaprep) 60
2.2.3.10. Mutagenesis 61
2.2.4. Protein Production
2.2.4.1. Transient Expression in HEK293T Cells 62
2.2.4.2. Large-Scale Transient Expression in HEK293T Cells 63

2.2.4.3. Construction of Stable HEK293 Gnt $^{-/-}$ Cell Lines Using G418 Selection 65
2.2.4.4. SDS-PAGE and Western Blot Protein Analysis 66
2.2.4.5. Protein Purification 67
2.2.4.6. Optimisation of Purification For CES2 68
2.2.4.7. Protein Concentration 68
2.3. Biophysical Characterisation
2.3.1. Kinetic Studies Using 4- Nitrophenyl Acetate (4-NPA) as a Substrate 70
2.3.1.1. Preparation of Assay Plate 70
2.3.1.2. Data Analysis 72
Equations 72
2.3.1.3. Specific Activity Calculation 73
2.3.2. Analytical Ultracentrifugation 73
2.3.3. Multi- Angle Light Scattering 74
2.4. Crystallization, X-ray Data Collection, and Structure Determination
2.4.1. Screening 75
2.4.2. CES1
2.4.2.1. Crystallization 77
2.4.2.2. Data Collection and Refinement 77
2.4.3. CES2
2.4.3.1. Post Protein Production Modifications and Strategies Utilised to Obtain Crystals 78
A. Varying Concentration and Temperature 78
B. De-Glycosylation 79
C. Co-Crystallization 79
D. Reductive Methylation of Lysines 79
E. In Situ Proteolysis 80
F. Seeding 80
G. Additive Screen 80
2.4.3.2. Crystallization of CES2 3-4 and 4-4
A. 3-Row Optimisation 81
B. 96-Well Optimisation 82
C. Hanging Drops 82
D. CES2 3-4 Cryo-Protection 82

E. Data Collection 83
2.4.3.3. Structure Determination and Refinement 84
2.4.4. Presentation of Crystallographic Data 84

3. PRODUCTION, CHARACTERIZATION AND CRYSTALLIZATION OF CES1

3.1. Introduction 85
3.2. Protein Production
3.2.1. Small Scale Expression 89
3.2.2. Large Scale Protein Production
3.2.2.1. Stable Expression 90
3.2.2.2. Transient Expression 90
3.2.3. Purification 90
3.3. Activity Analysis
3.3.1. Observations of CES1 Kinetics 94
3.3.2. Positive Cooperativity and the Hill Plot 98
3.3.3. Specific Activity 99
3.3.4. Ping Pong Mechanism 100
3.4. Biophysical Analysis
3.4.1. AUC 101
3.4.2. SEC-MALLS 102
3.5. Crystallization and Structural Analysis
3.5.1. Crystallization 103
3.5.2. Data collection and refinement 104
3.5.2.1. CES1 Wild Type 104
3.5.2.2. CES1 Null 104
3.5.2.3. CES1 N79Q 105
3.5.3. Structural Analysis
3.5.3.1. Comparison of CES1 and N79Q 109
3.5.3.2. Non-Specific Binding Sites 111
3.6. Discussion
3.6.1. Protein Production 114
3.6.2. Activity 114
3.6.3. Protein Structure 117

4. PRODUCTION, CHARACTERISATION AND STRUCTURAL DETERMINATION OF HUMAN CARBOXYLESTERASE 2

4.1. Introduction 118

4.2. Protein Engineering of Human CES2

4.2.1. Construct Design 120

4.2.2. Glycosylation Knock-Out Mutants 122

4.2.3. Oligomerisation Mutants 123

4.3. Production of Human Recombinant CES2

4.3.1. Expression of Human Recombinant CES2 125

4.3.2. Purification of Human Recombinant CES2 126

4.3.3. Optimisation of Purification 128

4.4. Activity Analysis 132

4.4.1. Observation of Non-Michaelis Menten Kinetics 134

4.4.2. Explanation for Atypical Kinetics 136

4.4.2.1. Substrate Inhibition 136

4.4.2.2. Cooperative Kinetics/ Allosterism 136

4.4.2.3. Biphasic Kinetics 137

4.4.3. Specific Activity 138

4.4.3.1. Effect of Glycosylation 139

4.4.3.2. CES2 3-4 and 4-4 139

4.5. Biophysical Analysis 142

4.6. Crystallization of CES2 145

4.6.1. Post Protein Production Modifications 145

4.6.2. Early Crystal Hits: Some Hope 146

4.6.3. Re-thinking Crystallization: Protein Engineering 147

4.6.4. Co-Crystallization Using CPT-11 148

4.6.5. Crystallization of CES2 3-4 and CES2 4-4 150

4.6.5.1. Initial Data Collection 150

4.6.6. Crystal Optimization

4.6.6.1. Initial Optimization 151

4.6.6.2. Additive Screen 151

4.6.6.3. Hanging Drop Experiment 152

4.6.6.4. Cryo-Protection Optimization 153

4.6.6.5. Data Collection Optimization 153

4.6.6.6. Crystals of CES2 4-4 154

4.7. Structural Analysis of CES2 156
4.7.1. Active Site 162
4.7.2. *N*-Glycosylation 164
4.7.3. Mutated Residues 165
4.7.4. Mapping of Single Nucleotide Polymorphisms onto CES2 165
4.8. Discussion
4.8.1. Protein Production 167
4.8.2. Kinetics 167
4.8.2.1. Glycosylation 168
4.8.3. Structure 169

5. PRELIMINARY STUDIES ON HUMAN CES3

5.1. Introduction 171
5.2. Protein Production 172
5.3. Activity and Biophysical Analysis
5.3.1. Activity Analysis 176
5.3.2. Analytical Ultra Centrifugation 176
5.4. Crystallization 178
5.5. Summary 179

6. GENERAL DISCUSSION

6.1. Comparison of Human CES1, CES2 and CES3 180
6.1.1. Protein Production 180
6.1.2. Oligomerisation 181
6.1.3. Activity 182
6.1.4. Structural Comparisons 183
6.1.5. Comparison of Rabbit and Human CES Structures 186
6.2. Future Work 188

APPENDICES 191

REFERENCES 198

LIST OF FIGURES

Figure 1.1.	Phylogenetic tree of mammalian carboxylesterases.	19
Figure 1.2.	Sequence alignment of CES1, CES2 and CES3.	20
Figure 1.3.	Schematic flow diagram illustrating the broad substrate specificity that human carboxylesterases exhibit.	26
Figure 1.4.	Structures of compounds hydrolysed by CES2 and CES1.	27
Figure 1.5.	Domain organisation as observed in CES1.	38
Figure 1.6.	Active site of CES1.	38
Figure 1.7.	Proposed mechanism of catalysis demonstrated by carboxylesterases.	40
Figure 1.8.	Graph showing the numbers of CES1 and CES2 papers published since 1996.	48
Figure 2.1.	Diagram to show the workflow of methods used in this project.	50
Figure 2.2.	Diagram to show how the Infusion™ Cloning Technology works.	55
Figure 2.3.	Site-directed mutagenesis, using the strand overlap extension PCR method.	62
Figure 2.4.	A simplified diagram showing the two different forms of homogenous *N*-linked glycosylation found on proteins produced in this project.	64
Figure 2.5.	Breakdown of 4-NPA.	78
Figure 2.6.	Layout of the 3-row optimisation experiment.	81
Figure 2.7.	Layout of a 96-deep well plate, formulated by hand, for the optimisation of CES2 3-4.	82
Figure 3.1.	Western blot analysis of protein expression from three different constructs of wild type CES1.	89
Figure 3.2.	Overlay of the size exclusion elution profiles of the three forms of CES1.	91
Figure 3.3.	Purified protein run on an SDS-PAGE gel, under de-naturing conditions.	92
Figure 3.4.	Kinetic Analysis of CES1 and CES1 N79Q.	96
Figure 3.5.	Graphical representations of statistic parameters.	98
Figure 3.6.	Graphical AUC analysis for CES1 N79Q.	101
Figure 3.7	Molar mass vs volume SEC-MALLS plots for CES2 wt and CES2 3-4.	102
Figure 3.8	Representative images of CES1 crystals and the resulting diffraction quality.	103
Figure 3.9.	Overall quaternary structure of CES1 in its trimeric form.	107

Figure 3.10.	Detailed views of CES1 structural elements.	108
Figure 3.11.	Image of previous pairing thought to be within salt bridge distance of other in the trimeric arrangement of CES1.	110
Figure 3.12.	Positive density observed at Cys 389.	112
Figure 3.13.	Overlay of disordered region on the surface of CES1.	113
Figure 4.1.	Output from the XtalPred server.	121
Figure 4.2.	Western blot analysis for protein expression of different constructs of CES2 wt.	125
Figure 4.3.	Western blot analysis of 18 stable cell line clones secreting CES2 wt.	126
Figure 4.4.	CES2 size exclusion elution profile.	128
Figure 4.5.	CES2 3-4 gel filtration elution profiles, before and after addition of nickel sulphate to the cell media.	129
Figure 4.6.	Kinetic analysis of CES2.	135
Figure 4.7.	3-D column chart showing specific activities of the CES2 mutants.	141
Figure 4.8.	Graphical AUC analysis for CES2 3-4.	142
Figure 4.9.	Molar mass vs volume SEC-MALLS plots for CES2 wt and CES2 3-4.	143
Figure 4.10.	Native glycosylated CES2 crystal from condition PACT C7.	147
Figure 4.11.	Hydrolysis of CPT-11 into SN-38 and 4-PP.	149
Figure 4.12.	First crystals of CES2 3-4 co-crystallized with CPT-11.	150
Figure 4.13.	Hanging-drop crystals of CES2 3-4 +CPT-11.	152
Figure 4.14.	Image from the interface of the SynchLink application.	154
Figure 4.15.	Average B-factors of chains A-D in CES2.	158
Figure 4.16.	Secondary structure (a) and topology (b) view of CES2 3-4.	159
Figure 4.17.	Cartoon view of overall quaternary structure of CES2 3-4.	161
Figure 4.18.	Inter-molecular disulphide bridge.	163
Figure 4.19.	Active site of CES2 with 4-PP covalently bound.	163
Figure 4.20.	Glycosylation of Asn 276.	164
Figure 4.21.	Mapping of documented SNPs onto CES2.	166
Figure 5.1.	Western blot analysis for expression of CES3.	176
Figure 5.2.	Gel filtration profile and SDS-Page gel of purified CES3.	177
Figure 5.3.	Graphical AUC analysis for CES3 wt.	179
Figure 6.1.	Structural overlay of CES1 and CES2.	186
Figure 6.2.	Structural alignment of the active sites of CES1, CES2 and rCE.	188

LIST OF TABLES

Table 1.1.	A summary of human CES1-5 genes.	18
Table 1.2.	Differential tissue expression of human carboxylesterases.	22
Table 2.1.	Summary of all the proteins produced in this project.	52
Table 2.2.	Standard buffers used in protein purification and in preparation of the nickel column.	67
Table 2.3.	Methods used for the optimization of purification of CES2.	68
Table 2.4.	Biophysical parameters of CES1-3.	69
Table 2.5.	96-well plate format for carboxylesterase activity analysis.	71
Table 2.6.	Summary of protein concentrations used in initial crystallization screens, based on the Pre-Crystallization Tests.	76
Table 2.7.	Commercially available crystallization screens used in the OPPF.	76
Table 3.1.	Details of published structures of CES1 to date.	87
Table 3.2.	Comparison of protein yields obtained from transient expression of CES1 in HEK293T and stable expression in HEK203 $Gnt^{-/-}$.	92
Table 3.3.	Initial velocity rates in absorbance units per minute (AU min^{-1}) for CES1 native and CES1 N79Q.	95
Table 3.4.	Kinetic parameters obtained using GraphPad Prism 6.	97
Table 3.5.	Specific activity values of CES1 mutants expressed in nmol min^{-1} μg^{-1}.	100
Table 3.6.	Data collection and refinement statistics for all three structures of CES1.	106
Table 3.7.	Structure alignment results from PDBeFold.	108
Table 4.1.	Localised sequence alignment of the residues critical in salt bridge formation between monomers in CES1.	123
Table 4.2.	Comparison of protein yields obtained from transient expression in HEK293T and stable expression in HEK293 $Gnt^{-/-}$.	131
Table 4.3.	Initial velocity rates for all CES2 mutants given in absorbance units per minute (AU min^{-1}).	133
Table 4.4.	Specific activity values of CES2 mutants expressed in nmol min^{-1} μg^{-1}.	140
Table 4.5.	Initial diffraction data obtained for CES2 3-4 + CPT-11.	151

Table 4.6. Final diffraction data and refinement statistics obtained for CES2 3-4 + CPT-11. 155

Table 5.1. Localised sequence alignment of CES1 and 3. 174

Table 6.1. Structure alignment results from PDBeFold. 185

LIST OF ABBREVIATIONS

aa(s)	Amino acid(s)
ABPP	Activity based protein profiling
AcChE	Acetylcholinesterase
ADHD	Attention Deficit Hyperactivity Disorder
AUC	Analytical ultra-centrifugation
AU min^{-1}	Absorbance units per minute
CES(s)	Human carboxylesterase(s)
CPT-11	7-ethyl-10-[4-(1-piperidino)-1-piper- idino] /carbonyloxycamptothecin / irinotecan
CRP	C-reactive protein
Dere	Surface entropy reduction directed mutation as proposed by Derewenda[1]
DLS	Diamond Light Source
DMSO	Dimethyl Sulfoxide
EB	Elution buffer
EDTA	Ethylenediaminetetraacetic acid
Endo H	Endoglycosidase H
ER	Endoplasmic reticulum
FABP	Fatty acid binding protein
FCS	Fetal calf serum
G418	Geneticin
GF	Gel filtration
GFP	Green fluorescent protein
GlcNAc	*N*-Acetylglucosamine /NAG
gpCE	Guinea pig carboxylesterase
GPT	GlcNAc phosphotransferase
h	The Hill coefficient
hBuChe	Human butryl-cholinesterase
HCC	Hepatocellular carcinoma
HCV	Hepatitus C virus
HLM	Human liver microsomes
IMAC	Immobilized metal ion affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kDa	Kilodalton
LB	Lysogeny broth

MALS	Multi-Angle Light Scattering
Man	Mannose
mRNA	Messenger ribonucleic acid
mES1	Mouse carboxylesterase 1
NAG	N-acetylglucosamine
OD	Optical density
O/N	Overnight
OP Nerve Agents	Organophosphorus nerve agents
OPPF	Oxford Protein Production Facility
OPTIC	Oxford Protein Target Information Collection
PBS(T)	Phosphate buffered saline (+0.1% Tween)
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PTCN	Potassium thiocyanate
rCE	Rabbit carboxylesterase
RD	Regulatory domain
Rmsd	Root-mean-square-deviation
SAXS	Small Angle X-ray Scattering
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC-MALLS	Size exclusion chromatography- multi-angle laser light scattering
SER	Surface entropy reduction
SNPs	Single nucleotide polymorphisms
v	Velocity
wt	Wild type
4-NPA	4-nitrophenyl acetate
4-PP	4-piperidino-piperidine

1. GENERAL INTRODUCTION

1.1. Properties and Functions of Carboxylesterases

1.1.1. General Overview

Carboxylesterases are a family of enzymes that hydrolyse ester, amide, or thioester bonds into their corresponding alcohol, amine or thiol and free acid in a diverse range of chemically distinct compounds including both exogenous (e.g. cocaine, heroin) and endogenous (e.g. acyl-CoA esters) substrates. A subset of the esterase superfamily, carboxylesterases are ubiquitously expressed from bacteria to man and are involved in a range of seemingly unrelated biological processes. These include xenobiotic and narcotic metabolism, detoxification of organophosphorus (OP) compounds, acting as phase-I drug metabolic enzymes, processing of cholesterol and fatty acids, forming crucial interactions with proteins involved in trafficking within the cell[2], and roles in tumour cell surveillance[3]. During the beginning of the last century, carboxylesterase in monocytes, then simply known as an esterase, was visualized in cells by the hydrolysis of α-naphthyl acetate or butyrate, releasing naphthol groups that were trapped by colourless substrates producing visible compounds. Since this activity was restricted to monocytes, it led to the development of a simple cytochemical staining technique that could specifically identify this cell lineage. This monocyte-specific staining has also allowed for recognition of the monocytic phenotype of some acute leukaemias[4].

Historically CESs have been given different names (*table 1.1*). The lack of a consistent nomenclature has made interpreting the results for different mammalian carboxylesterases potentially confusing. However, four years ago, a systematic classification was proposed based on the identification of homologous gene families, in human, mouse and rat[5]. “CES” (human) and “Ces” (mouse and rat) root symbols are used followed by the family number (e.g. human CES1), and this system will be adopted throughout this thesis.

1.1.2. Human Carboxyesterases

Genes coding for five carboxylesterases (CES) have been identified in the human genome (CES1-5)[5], and are located on chromosome 16q *(table 1.1)*. CES1, CES2 and CES3 appear to be the functionally significant enzymes, with CES1 and CES2 accounting for the majority of carboxylesterase activity within human tissue *(table 1.2)*. CES1 shares 46.7% sequence identity with CES2[6] and intriguingly, although substrate specificity for these enzymes is overlapping, they do in fact show individual substrate preference. CES3 seems to be brain specific[7], but to date, there has been little characterization of either this protein, CES4 or CES5. CES2 and CES3 have the highest amino acid sequence identity at 47.0%, whilst CES1 and CES3 have the lowest at 44.2% *(figure 1.2)*. From an examination of a multi-sequence alignment, the most notable features are that CES2 and CES3 both have a deletion of 15 and 14 amino acids (aas) respectively compared to CES1 starting at residue 302 in CES1, and that CES3 has a small insertion of 8 aas compared to CES1 and CES2 at residue 483 (CES1 numbering).

Carboxylesterases are widely distributed in nature, appearing intermittently in bacteria and lower eukaryotes, but expanding in number in higher eukaryotes[8]. Considerable research has been carried out into mammalian carboxylesterases, with evolutionary studies concluding that the CES gene family is ancient in its genetic origins and was established long before the appearance of mammals during evolution[9]. Subsequent tandem gene duplications events have occurred in mammalian evolution (*see figure 1.1)* leading to the generation of a large number of genes for some families, particularly in rodent families[5], with mouse and rat possessing at least 20 and 15 carboxylesterase genes[10] in their genomes respectively. A comprehensive study into Ces expression in the mouse[11] has shown that they are present in a wide variety of tissues and cells, being both intracellular and secreted.

In terms of using mouse models to look at human carboxylesterase function, it is important to understand that because of these tandem gene duplication events, no single mouse carboxylesterase is identical in structure or function to any one human gene. Therefore, it would be best to create a mouse model in which the Ces genes are silenced or knocked out, and individual human CES are genes knocked in specifically in order to study the effects of human CESs in a rodent model *in vivo*.

Human gene	Aliases	Chromosome 16q coordinates	Gene size (bp)	Exons strand
CES1	Human liver carboxylesterase, CE-1, hCE, hCE1, CES1A1, egasyn, human monocyte/macrophage serine esterase 1 (SES1), acyl-coenzyme A:cholesterol acyltransferase, HU1, EST1	54,394,465-54,424,468	30,008	14 -ve
CES2	Human intestinal CE, hiCE, hCE2, human liver CE-2, Carboxylesterase-2, iCE, intestinal carboxylesterase, cocaine esterase, HU2	65,527,040-65,535,426	8,387	12 +ve
CES3	Brain carboxylesterase, carboxylesterase 3, ES31, CE3	65,552,712-65,564,450	11,739	13 +ve
CES4	ESTHL, CES8, CE5	65,580,177-65,600,543	20,367	14 +ve
CES5	CES7, CE4	54,437,867-54,466,634	28,768	13 -ve

Table 1.1. A summary of human CES1-5 genes.
All known aliases of the five human carboxylesterases taken from published work[12]. Information on gene location and size was extracted and modified from Holmes et al[5]., along with other sources.

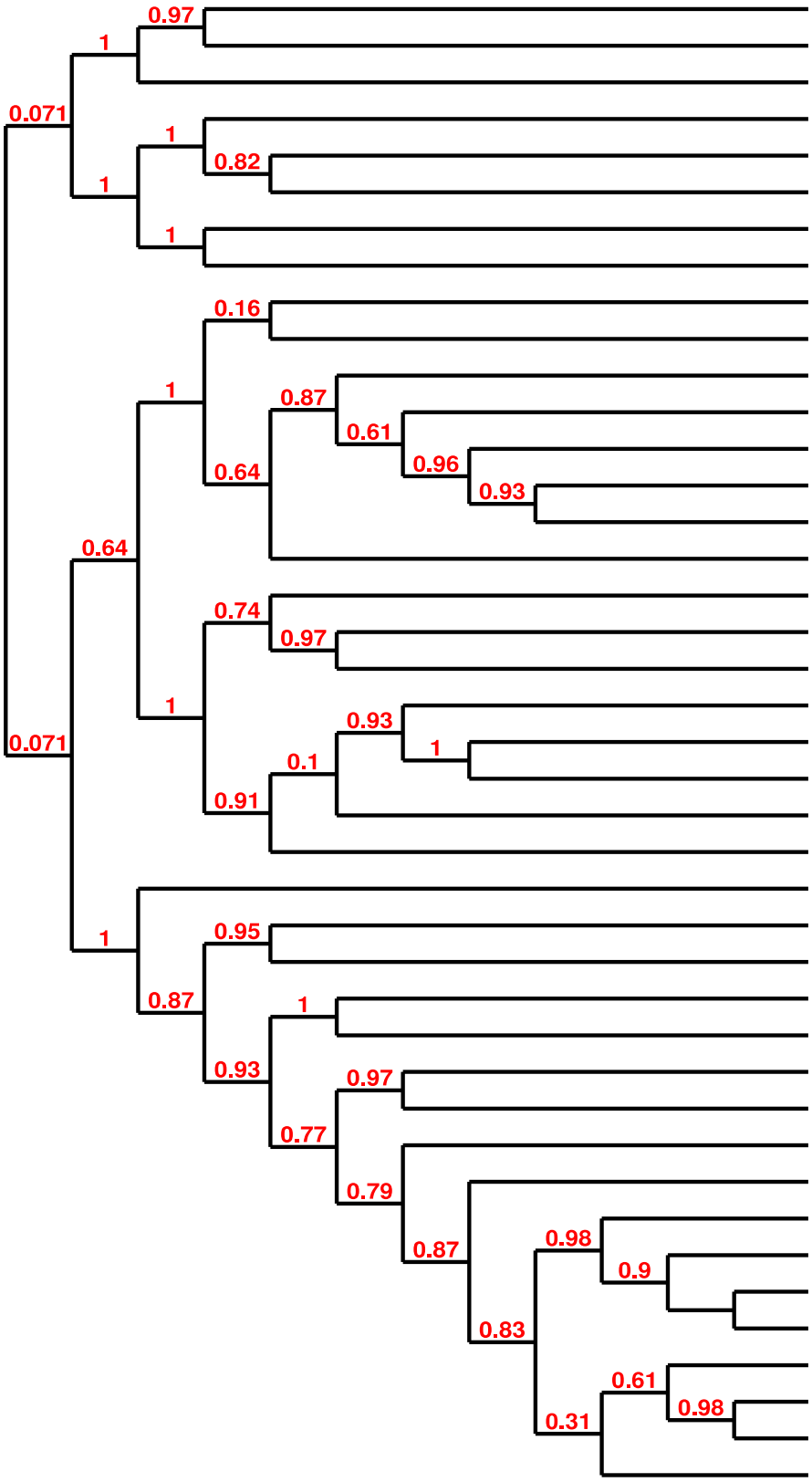
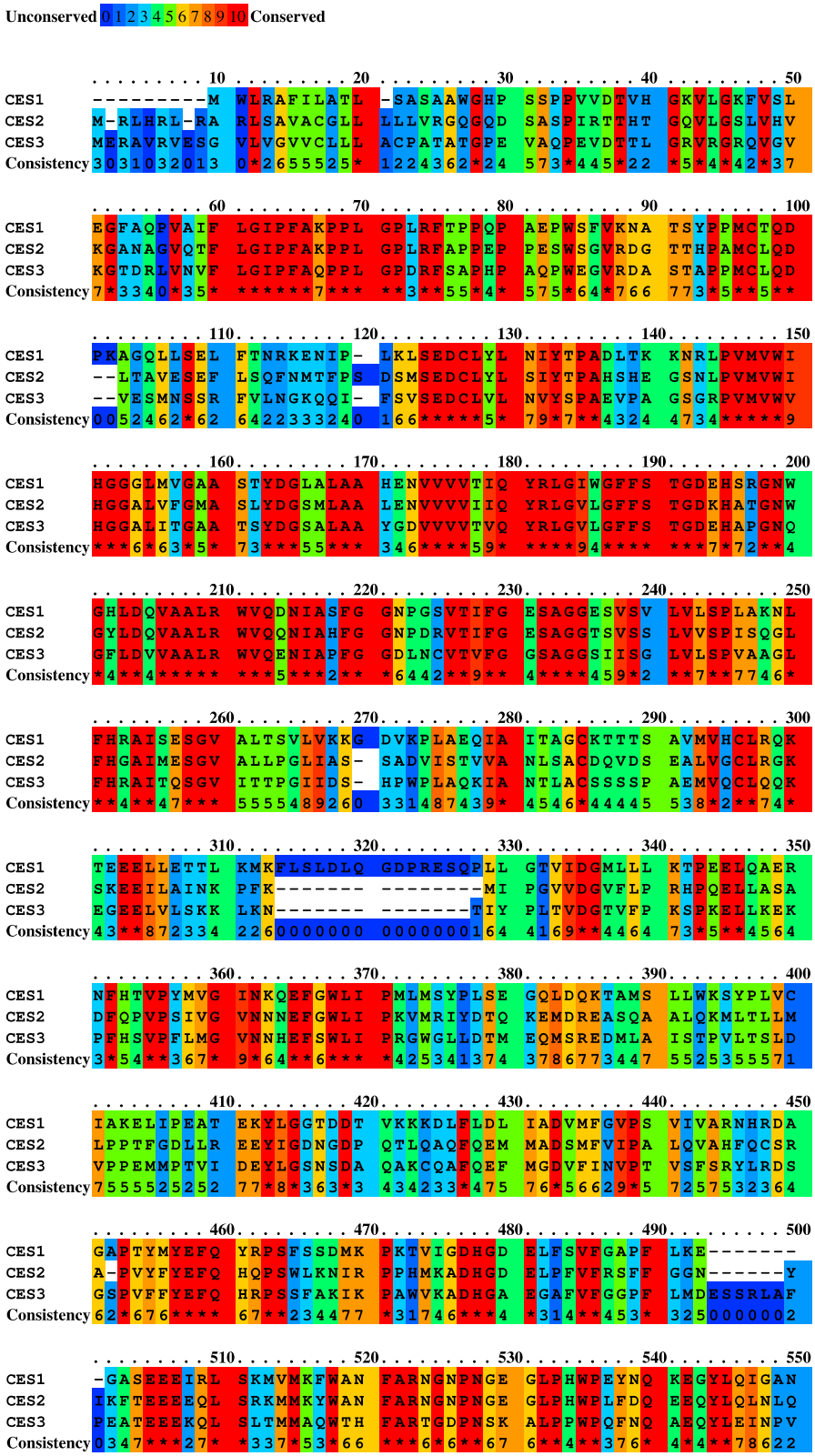


Figure 1.1. Phylogentic tree of mammalian carboxylesterases.
Constructed using the Phlylogeny.fr web service[13] with a selection of mammalian carboxylesterase sequences obtained from the UniProtKB database[14] (all chosen sequences had been reviewed). Branch support values (confidence levels) are shown in red.



	% Identity	Residues Overlap	Score	Gap frequency
CES1 and CES2	46.7	553	1137	4.2
CES2 and CES3	47.0	553	1205	1.6
CES3 and CES1	44.2	548	1079	4.6

Figure 1.2. Sequence alignment of human CES1, CES2 and CES3.
Residues are colour coded for amino acid conservation and alignment was performed using PRALINE[15]. Protein identity between CES1-3 is shown in the table, calculated using the ExPASy alignment tool, using protein sequences listed in *Appendix 1.1*.

1.1.3. Expression and Localisation of Human Carboxylesterases

1.1.3.1. Expression Profile

The expression of human CES proteins has been confirmed in a broad range of tissues, and is summarized in *table 1.2.* CES1 is predominantly expressed in liver tissue and macrophages, with some research identifying low level expression in human lung epithelia[16], small intestine, heart, spleen, colon[6] and adipocytes. CES2 is primarily expressed in the small intestine and colon, with lower levels detected in the liver, kidney, heart and skeletal muscle.

However, the mRNA transcript profiles of both CES1 and CES2 (from mRNA detection) indicate a much wider tissue distribution throughout the body and although the five enzymes show some tissue-specific expression, there is considerable overlap in their mRNA profiles. Gene expression profiling has indicated that CES3 is present in the brain[7]. The presence of CESs in primary organs of first pass metabolism suggests that this protein family has evolved to have a universal detoxification function within the body.

Many historical assays of carboxylesterase activity have been performed on whole cell lysates (e.g. from hepatocytes), and given their overlapping tissue expression, the results of such assays are likely to have come from the contribution of more than one carboxylesterase. This presents a difficulty in both interpreting existing data on specific CES expression and activity, as well as performing comprehensive functional assays on whole cells. Therefore, when attempting to understand the properties of individual CESs, it is important to analyse data from experiments using pure recombinant protein rather than whole cell lysates.

The expression of both CES1 and CES2 is known to be developmentally regulated, and increases with age. Even in normal, non-diseased states, great inter person variability is seen in both mRNA and protein levels[17], indicating that CES expression profile varies from person to person.

Gene	Expression							
	Macrophages	Liver	Gastrointestinal Tract	Blood-Brain Barrier	Heart	Kidney	Lung	Testis
CES1	***	***	*		*		*	
CES2		*	***		*	*		
CES3	*	*	*	*				
CES4				*		*	*	*
CES5				*		*	*	*

Table 1.2. Differential tissue expression of human carboxylesterases.
Summary of relative CES protein expression from papers published. ***High expression, **Medium expression, *Low expression. Expression here is under normal conditions. Data used to in the making of this table came from a collection of sources[18, 19], and is based on expression at the protein level using antibody detection. mRNA detection has been omitted for clarity.

1.1.3.2. Cellular Localisation

The localisation of mammalian carboxylesterases can differ between species and isoform. CES1-3 are known to be situated in the lumen of the endoplasmic reticulum (ER), the site of synthesis and maturation of proteins destined for secretion. Their residence is regulated by two factors; *i*) an NH_2-terminal hydrophobic signal peptide for the localisation of these proteins to the ER (*Appendix 1.1*) and *ii*) a COOH-terminal retention sequence that prevents secretion from the cell[20]. It was first identified by Robbi and Beaufay et al. in 1991 that mammalian carboxylesterases contained a tetra peptide COOH-terminal microsomal retention sequence that is usually a variation of 'HXEL' for stable ER residence. This retention motif binds to a KDEL receptor, a seven-transmembrane-domain protein, and this interaction holds the proteins in the ER. If this sequence is removed by the protein processing machinery or is absent from the protein sequence, the enzyme is not recognized by the KDEL retrieval pathway and is secreted from the cell via the Golgi apparatus. The tetra peptide motif of CES1 is HIEL and that of CES2 is HTEL. It was later discovered that CES3 had an alternative *C*-terminal retention signal of 'QEDL'[21], which supports the proposal that the COOH-terminal dipeptide 'EL' is the major, if not the only stringent, requirement for ER retention[22].

CES4 and CES5 are believed not to reside in the ER, appearing to be secreted enzymes that have a general circulatory detoxification function. CES5 is thought to be involved in catalyzing lipid transfer reactions within male reproductive fluids and in protecting neural tissue from xenobiotics[23, 24], but very little research has been carried out on either of them. The exact role of CES4 is currently unknown.

1.1.3.3. Protein Trafficking

Carboxylesterases have been reported to have a role in the retention of proteins in the ER, which does not depend on the catalytic activity of the enzymes[25, 26]. β-glucuronidases, which are phase II drug metabolising enzymes, produce glucuronic acid conjugates from the hydrolysis of xenobiotic compounds. They are retained in the ER by forming a CES1-β-glucuronidase complex through interactions in the area between residues 228-279[25], adjacent to the catalytic pocket.

C-reactive protein (CRP), which is secreted in response to infection or cellular damage, and is the most sensitive marker identified in the development of atherosclerosis[27], is retained in the ER by CES1. This enzyme holds the CRP protein in place before a reduction in binding affinity causes it to be released into the plasma when required[26]. The CRP binding site in CES1 is located on the surface of the enzyme and contained within a sequence of 23 residues, from 477-499, near the COOH- terminus[28]. This site is spatially distinct from the active site and the ER retention signal.

1.1.4. Altered Carboxylesterase Expression and Activity in Disease

Although normal carboxylesterase expression is widespread and variable under normal conditions (*table 1.2)*, it is becoming more apparent that these enzymes are up regulated and overexpressed in certain disease states.

1.1.4.1. Cancer

CES1 appears to have a role in tumour cell killing or surveillance. It has been shown to be a potent and specific marker in distinguishing hepatocellular carcinoma (HCC) from other liver disease, including cirrhosis and hepatitis[3]. Clinical data showed that CES1 protein levels were significantly higher in the earlier stages of HCC (I–II), indicating that it could be used to identify patients in early stages of the disease. A dual-marker system, combining CES1 and alpha-fetoprotein (AFP, the serological marker currently used) has the potential for increasing sensitivity and specificity of results in the early diagnosis of HCC.

Unlike the up-regulation of CES1 protein levels in certain cancers, CES2 appears to be an inverse biomarker of disease. Tang et al. showed, through Western blotting, that local expression of CES2 protein is down-regulated in colorectal cancer[29, 30]. Comparison of CES2 expression between normal and tumor tissues at stages I to IV was significantly different, and decreased in tumor tissue following progression of the disease[30]. As well as colon cancer, a decrease in the levels of CES2 has been observed in ovarian cancer. Clinical studies with ovarian cancer patients have shown that both the serum level and the expression of CES2 in the tumor mass from stages I through to IV decreased significantly with the advancement of the disease. It was also noted that CES2 was more sensitive than the cancer antigen 125 in detecting ovarian cancer in its early stages. These findings suggest that CES2 may be a useful inverse biomarker in the diagnosis of the early stages of ovarian cancer[31], and a number of other cancers. The level of CES2 expression may

also help in patient stratification by identifying those most likely to benefit from treatment, for example with the oral pro-drug of gemcitabine, LY2334737[32] (*figure 1.4, a*), a chemotherapeutic nucleoside analogue, which is activated by CES2.

1.1.4.2. Metabolism

Research shows that the CES1 gene is tightly regulated in human adipose tissue, with increased mRNA levels seen in obese (when compared to lean) subjects, as well as decreased levels during weight loss[33]. Dominguez et al[75]. demonstrated that levels of CES1 were increased approximately twofold in adipose tissue of humans with obesity and Type 2 diabetes. Expression of Ces3, a murine orthologue of CES1, was also shown to increase in a mouse model of diabetes-obesity[34]. This paper concluded that an increase in CES1 activity and the consequent release of fatty acids into the circulation may be a previously unrecognised feature of the pathogenesis of obesity related diabetes. Ces3 (triacylglycerol hydrolase (TGH)) has been reported to be important for basal lipolysis[34, 35] and may be a suitable pharmacological target for lowering fatty acid efflux from adipose tissue.

1.1.4.3. Hepatitis C Virus

Activity-based protein profiling (ABPP) has shown that CES1 has an important role in Hepatitis C virus (HCV) propagation and virion release, and that HCV can induce CES1 expression to modulate the cellular environment to its own advantage[36]. The up-regulation of CES1 is thought to result in a favorable host cell environment by promoting both lipid droplet formation and growth that subsequently aids HCV pathogenesis, rather than directly interacting with the HCV proteins themselves.

The enhanced carboxylesterase activity observed in these diseases opens up another potential therapeutic field for these important proteins.

1.1.5. Substrates of Human Carboxylesterases

Human carboxylesterases are involved in the hydrolysis of a wide range of substrates, which can be divided into four categories; *i*) clinically administered drugs, *ii*) narcotics, *iii*) environmental chemicals and *iv*) endogenous substrates (*figure 1.3*). The enzymes function in both the activation and inactivation of these diverse molecules, and will be discussed in the following paragraphs.

It has been proposed that CES1 preferentially cleaves small, non-bulky substrates[37] that contain a small alcohol group and large acyl moiety (e.g. meperidine), whereas CES2 can hydrolyse much larger, complex molecules and shows a preference for substrates with a small acyl moiety and large alcohol group (e.g. CPT-11 and cocaine benzoyl ester)[38]. Comparable studies on CES3-5 have not been reported.

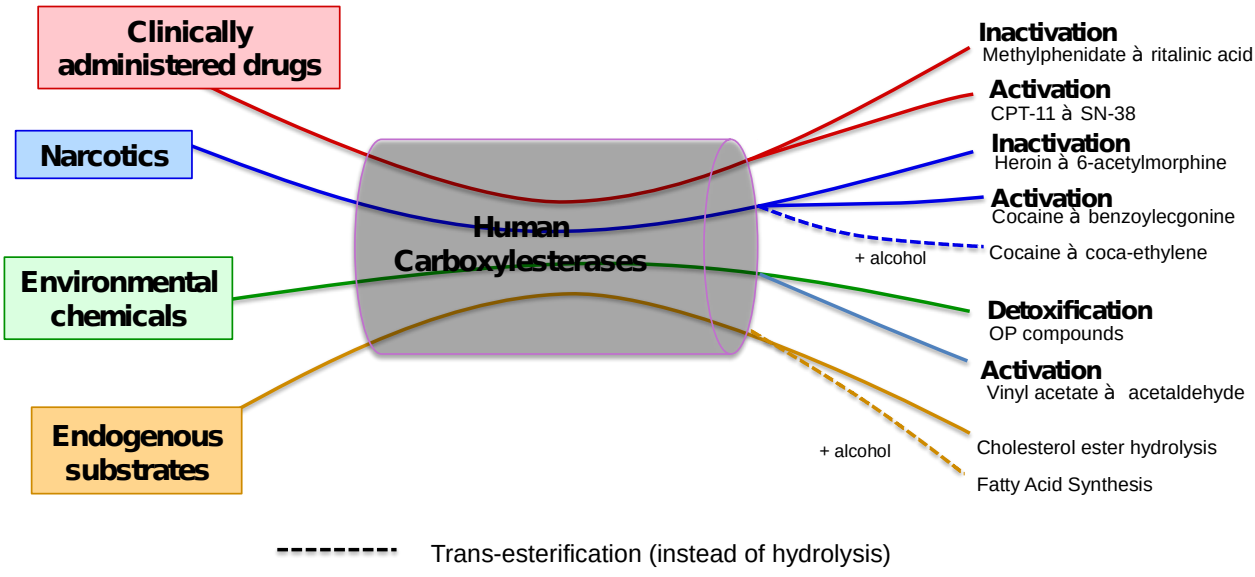
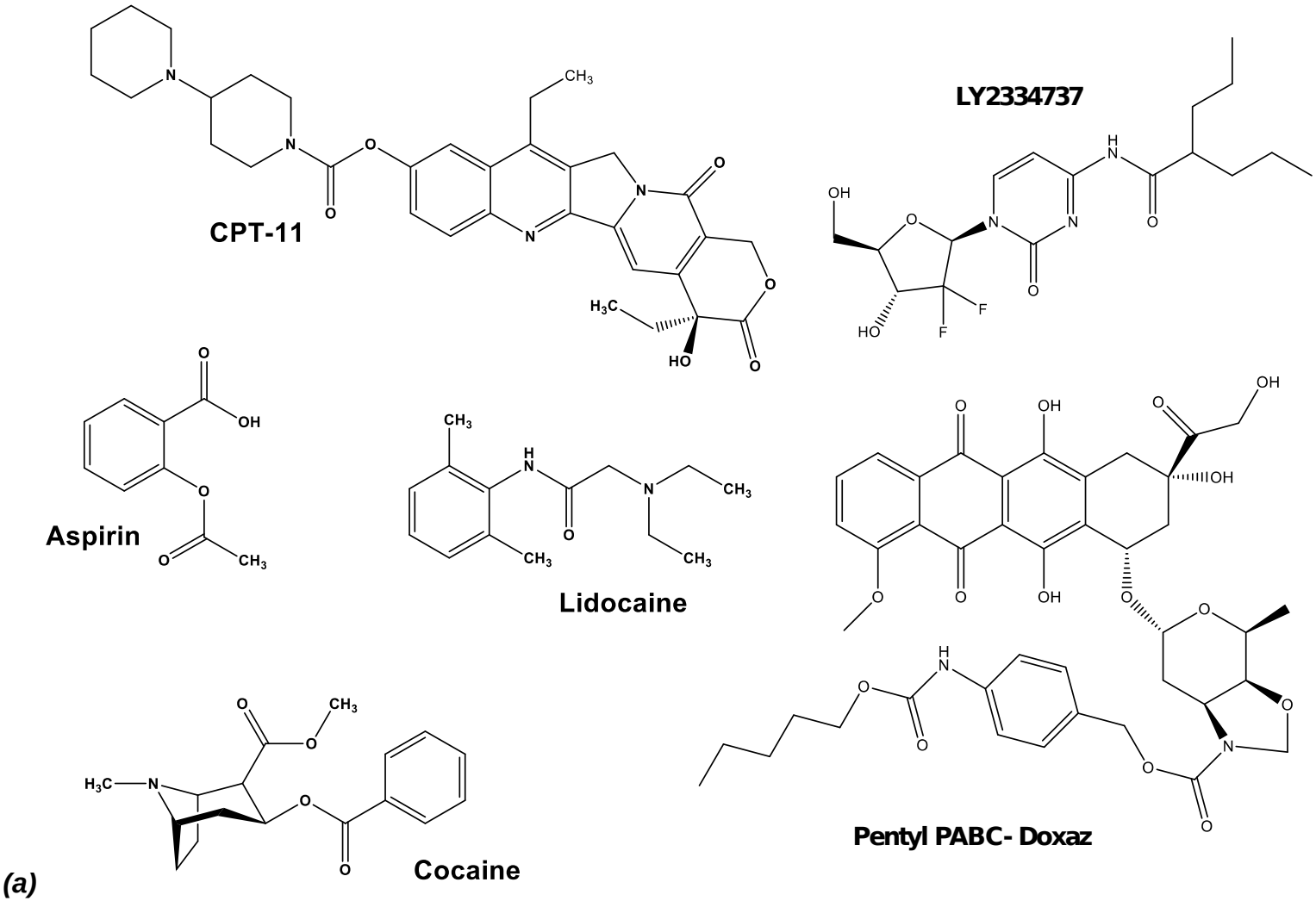


Figure 1.3. Schematic flow diagram illustrating the broad substrate specificity that human carboxylesterases exhibit.
Examples of groups of substrates that CESs are known to act on are given on the left-hand side, with the roles they play in numerous processes displayed on the right.


CPT-11
LY2334737
Aspirin
Lidocaine
Cocaine
Pentyl PABC- Doxaz

(a)

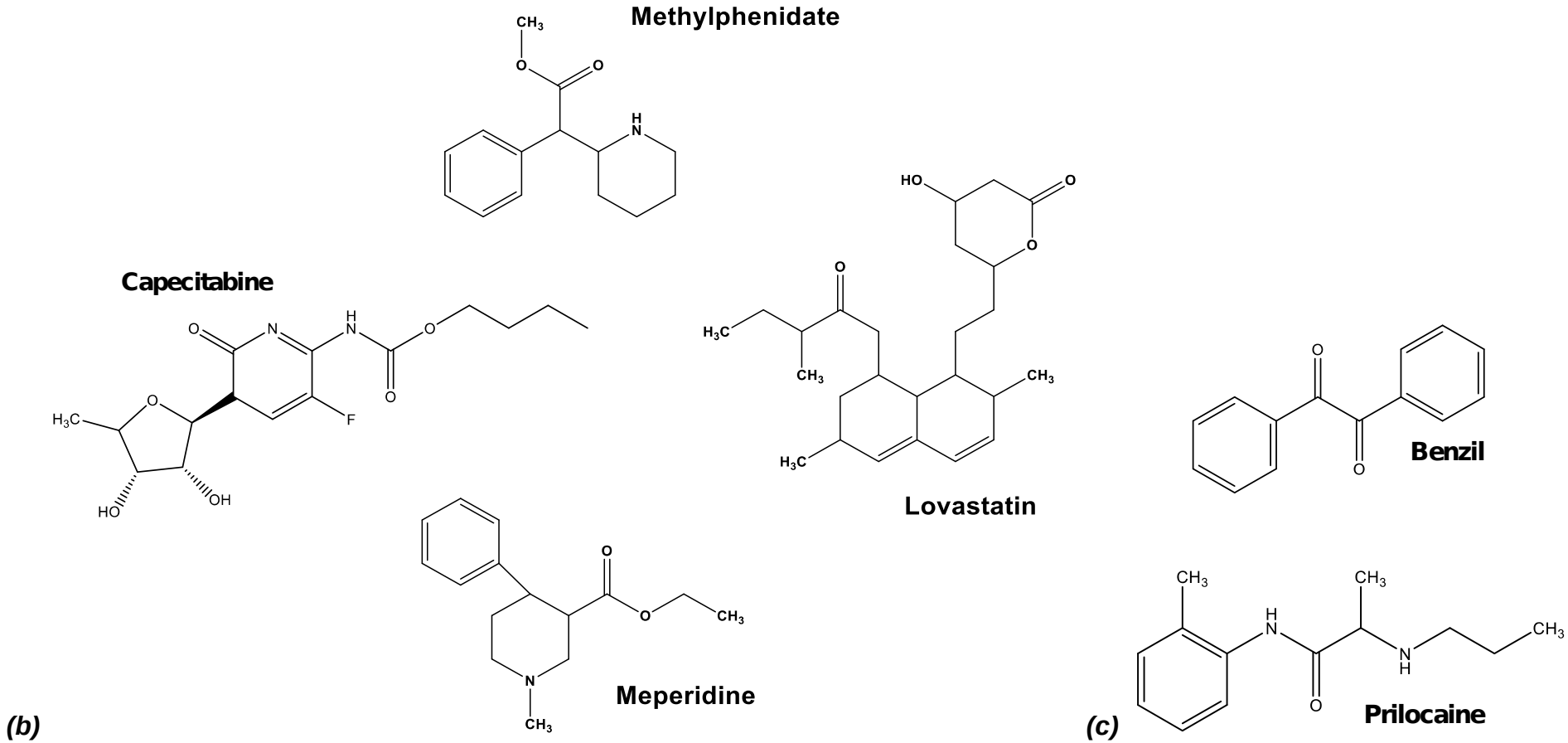


Figure 1.4. Chemical structures of compounds hydrolysed by CES2 and CES1.
(a) Drugs known to be selectively hydrolysed by CES2 include CPT-11, lidocaine, aspirin (also known as acetylsalicylic acid), cocaine, pentyl PABC doxaz (PPD) and LY2334737 (the prodrug of gemcitabine). ***(b)*** Drugs known to be hydrolysed solely by CES1 include methylphenidate, lovastatin, meperidine (also known as pethidine or Demerol™) and capecitabine. ***(c)*** Prilocaine is hydrolysed by both CES1 and CES2, and benzil acts as an inhibitor to both. Structures were constructed manually using ChemBioDraw.

1.1.5.1. Clinically Administered Drugs

Esterases contribute to the hydrolysis of approximately 10% of clinically administered drugs[39], with CES1 and CES2 being the major source of serine esterases that are involved[40]. The products of hydrolysis are generally more polar than the initial ester, resulting in increased water solubility and enhanced clearance by excretion. Therefore CESs are categorized as phase-I drug metabolizing and can both activate (e.g. tamiflu™[41]) or inactivate (e.g. aspirin) a range of drugs including anticoagulants, angiotensin receptor blockers, anti-hyperlipidemic agents, antiviral drugs, anti-tumour drugs, anti-obesity drugs, anesthetics and phospho-nonsteroidal anti-inflammatory drugs[42]. Given this extensive list, it is clear that CESs play a key role in a wide range of clinically significant drugs [43].

Selectivity

Studies *in vitro* using recombinant enzymes have shown that CESs are selective for different drugs. Structures of the compounds discussed in the following paragraphs can be found in *figure 1.4.* Methylphenidate, a psychostimulant prescribed to treat Attention Deficit Hyperactivity Disorder (ADHD), is hydrolysed only by CES1 with neither CES2 or CES3[38] showing any catalytic activity on this substrate. Methylphenidate (an ester) is the active drug, and is hydrolysed in the liver to produce the inactive carboxylate, ritalinic acid[38, 44]. CES2 hydrolyses ester-linkages in oxybutynin and aspirin, whereas CES1 does not[45]. Conversely CES1 hydrolyses meperidine, whereas CES2 is inactive[46]. An example of where the substrate specificity of CES1 and CES2 overlaps is with the two structurally related compounds prilocaine and lidocaine, both of which are local anesthetics. Prilocaine is hydrolysed by both CES1 and CES2. By contrast, lidocaine, which differs by two additional methyl side-chains, is hydrolysed solely by CES2[47].

Oncology Drugs

Carboxylesterases are known to activate certain cancer drugs. The chemotherapeutic agent, irinotecan (CPT-11), is converted from a pro-drug to its active metabolite, SN-38[48], a potent topoisomerase I poison[49], primarily by CES2[50]. The ester group has been purposely engineered into CPT-11 to improve its solubility for oral absorption, a common strategy in drug design. CES2 is also responsible for the hydrolysis of the pro-drug of gemcitabine, designated as LY2334737[32], which is used in the treatment of several types of cancer. The pro-drug has been produced by adding an amide-linked valproate (2-propylpentanoic acid) group to gemcitabine, altering the properties of the drug by improving its stability and allowing for oral administration. Oral bioavailability is important as it permits frequent low dosing that may enhance antitumor efficacy. CES2 is also responsible for the activation of the prodrug of doxorubicin[51], pentyl PABC-doxaz (PPD), an anthracycline antitumor drug that functions primarily as a topoisomerase 2 poison to induce cancer cell death. Capecitabine is a prodrug widely used in the treatment of breast and colorectal cancer, and is converted to 5-fluorouracil (5-FU) in three enzymatic steps, the first of which is catalyzed by CES1 primarily in the liver[52].

1.1.5.2. Narcotics

Carboxylesterases are involved in the metabolism of narcotics and it has been widely reported that CES1 and CES2, together with serum butyryl-cholinesterase (hBuChE), catalyze the hydrolysis of cocaine[53-55]. It was believed that CES1 catalyzed the methyl ester hydrolysis of *R*-cocaine producing the primary metabolite, benzoylecgonine, whereas CES2 hydrolysed the larger benzoyl ester group on cocaine[53, 56]. Controversially, a recent publication by Hatfield et al., indicates that CES1 is unlikely to play any role in the

metabolism of cocaine and that hydrolysis of it by this class of enzymes is via CES2 *in vivo*[57].

When alcohol is consumed in conjunction with cocaine, CES1 is thought to be responsible for catalyzing the trans-esterification of cocaine to coca-ethylene, with the alcohol forming an acyl-enzyme intermediate rather than water. Coca-ethylene is a toxic cocaine metabolite found in humans when alcohol and cocaine are abused together[58]. The CES1, CES2 and hBuChE trio of enzymes also act collectively to hydrolyse the inactive recreational drug heroin to 6-acetylmorphine[54, 57].

1.1.5.3. Environmental Chemicals

Carboxylesterases perform vital roles in the detoxification of environmental chemicals in conjunction with other enzymes including glucuronosyltransferases and mammalian microsomal cytochrome P450. CES1 has been extensively studied with respect to its involvement in the inactivation of organophosphorus (OP) nerve agents (alkylphosphonic esters), with long-term aims of designing novel forms of CES1 as a protein-based therapy for nerve agent detoxification[59]. OP nerve agents covalently modify a variety of enzymes, namely AcChEs, by phosphonylation of the catalytic serine[60], and due to their similarity in structure and function, CESs are modified in the same manner. It has been shown that CES1 stereo-selectively binds the nerve agent cyclosarin and spontaneously hydrolyses sarin[61], and it has been demonstrated in rats that irreversible binding of the nerve agent, soman, to CES1 accounts for a significant portion of its rapid detoxification[62]. It has also been shown that after the formation of the acyl-enzyme intermediate, the first step of the serine hydrolase catalytic mechanism, nerve agent-carboxylesterase complexes are capable of reactivation after incubation with sarin, restoring the enzyme's normal function[61].

Research has been conducted on CES1 to try and improve its OP nerve agent hydrolase activity by introducing mutations into the active site of the enzyme. The United States military has been looking into the development of mutant forms of CES1 as a viable prophylactic treatment for exposure to chemical weapons including tabun, sarin, soman and VX gas[59, 63].

OPCs are also widely used in pesticides and herbicides. CESs act as a bioscavangers providing protection against OPCs by detoxification via hydrolysis[64-66].

CES enzymes also play a role in the conversion of pro-carcinogens into carcinogens. For example, vinyl acetate, which is used in the paint, adhesive, and the paper-board industry, is metabolized by CES enzymes into acetaldehyde in the liver[67].

1.1.5.4. Endogenous Substrates

Cholesterol Metabolism

Carboxylesterases appear to be important in the processing of cholesterol derivatives and fatty acids in the liver. They hydrolyse endogenous compounds, such as short- and long-chain acyl-glycerols, long-chain acylcarnitine, and long-chain acyl-CoA esters. CES1 has been implicated in a variety of endogenous cholesterol metabolism pathways including cholesterol ester hydrolysis (CEH), fatty acyl Coenzyme A hydrolysis (FACoAH), acyl-Coenzyme A:cholesterol acyltransfer (ACAT), fatty acyl ethyl ester synthesis (FAEES) and testosterone synthesis[68]. Interestingly, CES3 has been found to have a functional compensatory role, being able to restore intracellular cholesteryl ester hydrolysis when CES1 expression was greatly reduced[19].

CES1 is also known to act as a fatty acyl ethyl ester synthase, trans-esterifying ethanol to a fatty acid. This occurs when the fatty acid acyl intermediate is covalently attached to the catalytic serine of CES1 and there is an abundance of an alcohol-containing

compound[69]. Fatty acyl ethyl esters (FAEEs) are toxic byproducts of alcohol abuse and a build-up of these compounds in the tissues of alcoholics is a hallmark of this disease process.

Therefore, CES1 appears capable of catalysing both the formation and elimination of cholesteryl esters using trans-esterification and hydrolysis reactions respectively[70], and these activities are vital in regulating and trafficking cholesterol substrates among cells and tissues in the body[71].

Vitamin A

Long chain acyl esters of retinol are the principle form of vitamin A stored in the body, mainly in the liver. It has been shown in mouse models that carboxylesterases are capable of catalyzing the hydrolysis of these retinyl esters to free retinol in the ER[72], and it is proposed that CES1 takes on this role in humans.

1.1.6. Inhibitors of Carboxylesterases

Alongside recent research into the substrates of carboxylesterases, work to identify non-toxic selective carboxylesterase inhibitors has intensified. Possible applications of such inhibitors include prevention of drug toxicity, increasing the half-life of drugs that are inactivated by CESs and modulating the metabolism of esterified drugs *in vivo*[73]. For example, flestolol, a short acting beta-adrenergic blocking agent is rapidly hydrolysed by CESs to an inactive metabolite and hence its biological activity is rapidly lost[74]. Inhibition of CESs would increase the *in vivo* stability of this molecule and likely improve its

therapeutic utility. By administering a specific inhibitor in combination with the drug of choice, prolonged plasma concentrations and hence greater efficacy might be achieved.

Benzil and benzil analogues have been identified as general CES inhibitors, acting as ester mimics[75] (*figure 4.1*). It has been found that for these inhibitors to be effective, the dione moiety must be in the 1,2-configuration, and adjacent to an un-substituted aromatic group[75 73]. Predicting whether or not the enzymes will hydrolyse a compound is difficult because esters can act either as substrates or inhibitors for CESs. From activity analysis of CESs, it has been noted that substrates can act as inhibitors when they are present at high concentrations. A possible use for a general CES inhibitor would be to treat a heroin overdose, where it could be administered to decrease the conversion of heroin to its active metabolite, morphine.

Diarrhea that is correlated with CPT-11 treatment is thought to result from the hydrolysis of the drug in the intestine by CES2. Since this is the dose limiting toxicity for this highly effective anticancer agent, approaches that improve this side effect would improve patient quality of care and potentially allow drug dose intensification. This could potentially be achieved by an inhibitor that targets CES2 within the gut[76]. Benzene sulfonamides are specific CES2 inhibitors[77], and based upon this scaffold, a series of highly potent fluorene analogues have been developed. These are currently undergoing further testing to see if they are able to inhibit CES2 intracellularly, with the long term goal of developing an inhibitor suitable for the modulation of CPT-11 induced toxicity that can be administered to patients undergoing treatment with this drug[76].

Interestingly, two drugs, mevastatin and tamoxifen have been found to have a weak inhibitory effect on CES1 *in vitro*[78], acting as partially non-competitive inhibitors. The binding of tamoxifen to CES1 could affect the efficacy of this drug. It may delay the clearance of the drug from the liver, therefore using a CES1 selective inhibitor in

combination with the drug could improve tamoxifen efficacy by preventing it from binding to CES1[78].

Another potential use for a CES1 specific inhibitor may be in managing cholesterol homeostasis in humans[70].

1.1.6.1 Natural Inhibitors

Carboxylesterases have been shown to be inhibited by a series of flavonoids present in grapefruit juice, affecting their hydrolysis of the prodrugs lovastatin and enalapril[79] used in the treatment of hypertension. A study into the hydrolysis of the antiviral ester prodrug bis (POC)-PMPA (also known as Tenofovir) showed that carboxylesterases can also be inhibited using extracts of strawberry and banana[80]. These findings show that interactions between ester containing drugs and naturally occurring flavonoids should be considered in the clinic.

1.2. Structure and Catalytic Mechanism

Mammalian carboxylesterases belong to the serine esterase superfamily, also known as the *α,β*-hydrolase-fold family[81]. They are one of the largest and most widely distributed enzyme classes in all three kingdoms of life, and in humans they constitute ~1% of the proteome[82]. Other members of this family include serine esterases (e.g. trypsin and thrombin), lipases (e.g. cutinase and cytosolic phospholipase A2), peptidases (e.g. dipeptidyl peptidase IV), amidases (e.g. FAAH) and acetylcholinesterases. A number of these related proteins also have differing substrate specificities among homologues.

The first mammalian carboxylesterase structure to be published was that of rabbit carboxylesterase (rCE) (1K4Y)[12]. This was followed by the structure of the human enzyme CES1, which remains the only structure of a human carboxylesterase that has been determined and published (*see table 3.1 in Chapter 3*). Rabbit CE and human CES1 share 81% sequence identity[83] and their structures revealed that they are structurally homologous to the other classes of esterases.

1.2.1. The Three Domains

Carboxylesterases are comprised of three distinct domains; a central catalytic, a regulatory (RD) and an α/β domain (*figure 1.5*). Crystallographic studies show that in CES1, the catalytic triad, comprising S221 (a nucleophile), H467 (a base) and E354 (an orientating acid), sits in the large and conformable catalytic domain (*figure 1.6)*, which consists of a central anti-parallel β sheet surrounded by several α helices. The active site occupies a 10-15Å deep hydrophobic pocket at the interface of the three domains with all three catalytic residues arranged such that a proton transfer chain can be established. The substrate binding gorge contains a large, flexible pocket on one side of the serine, and a small, rigid pocket on the opposite side that is adjacent to the oxyanion hole[84]. It also

contains the *C*-terminal helix of the enzyme. The active site of CES1 contains both specific and promiscuous regions, which enables the enzyme to act on structurally distinct substrates. The promiscuity of CES1 is enhanced by its ability to interact with multiple ligands in multiple orientations at once, for example tacrine has been seen to bind in multiple orientations within the large active site of each monomer in the CES1 hexamer[85].

The RD is mainly helical; containing two disordered loops and has been proposed to regulate substrate binding and product release. This domain is so-called because it is believed to regulate access to the catalytic gorge[12]. Published structures show that this part of the enzyme exhibits high thermal displacement parameters, potentially indicating dynamic mobility within this region[12, 69, 84].

The αβ domain or hydrolase fold, typical of hydrolases, lies adjacent to both the catalytic and regulatory domain. This domain is common to a number of hydrolytic enzymes of differing catalytic functions and phylogenetic origin[81].

Sequence alignment (*figure 1.1*) shows that the catalytic triad in CES2 consists of S228, E354 and H457. The lack of a crystallographic structure of CES2 means that it is not possible to rationalise why this enzyme is able to cleave much larger, complex molecules than CES1. It has been proposed that CES2 must contain flexible domains within its active site[86] or presents a different three dimensional structure to allow for accommodation of these esters.

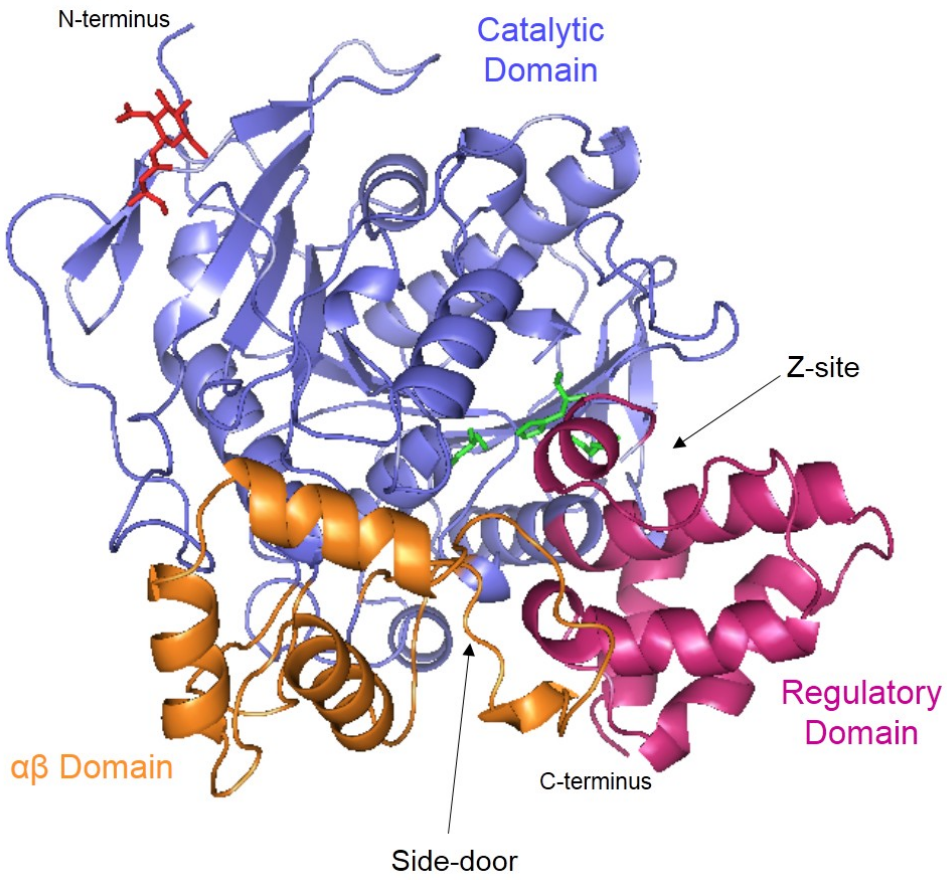


Figure 1.5. Domain organisation as observed in CES1.
A cartoon view of the CES1 monomer with the catalytic (blue), regulatory (pink) and αβ (orange) domains colour coded. The *N*- and *C*- termini (in the catalytic and RD domains respectively) are indicated and the proposed Z-site and side-door are labelled. The catalytic triad is composed of S 221, E354 and H467, shown in green. One NAG molecule is present on N79, shown in red.

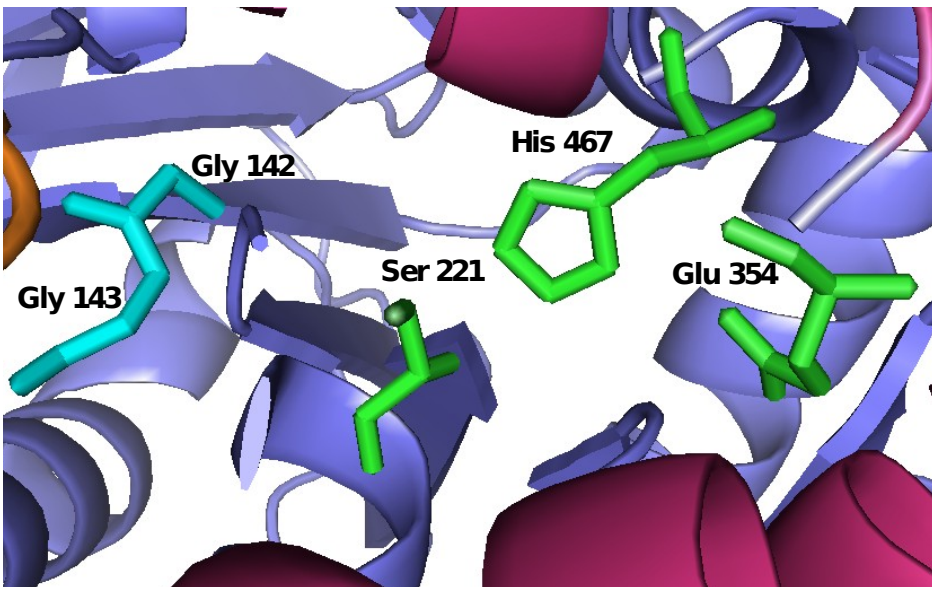
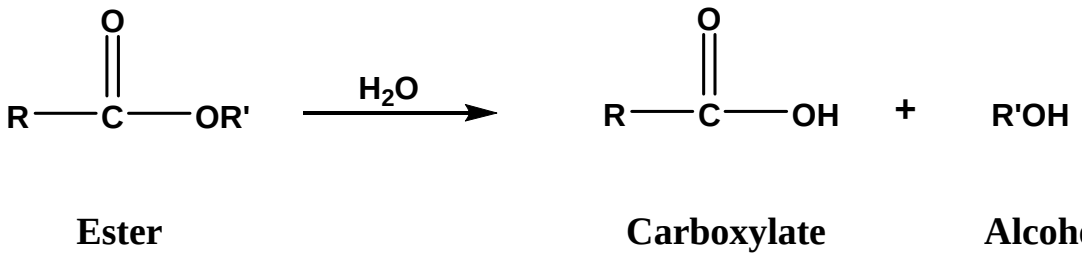


Figure 1.6. Active site of CES1.
The arrangement of the amino acid residues of the catalytic triad is highlighted in green. On the left of the diagram, the conserved residues G142 and G143 (shown in cyan) define the oxyanion hole that stabilizes the tetrahedral intermediates during the two-step hydrolysis reaction (*see figure 1.7 for more details)*.

1.2.2. Catalytic Mechanism

Structural enzymology studies have provided an insight into the catalytic mechanism of CESs. They catalyze the breakdown of chemicals that contain an amide, ester or thioester linkage, resulting in the formation of the corresponding alcohol and carboxylic acid:



The enzymes cleave via a two-step ping pong mechanism that includes the formation and degradation of an acyl-enzyme intermediate, using water as a transitional nucleophile[86] *(figure 1.7)*. A nucleophilic attack of the catalytic serine on the carbonyl carbon atom of the ester initiates hydrolysis[12], with the assistance of the histidine acting as a general base. The protonated histidine is stabilized via hydrogen bonding to the glutamic acid, and it is this crucial His-Glu pairing that function together as a 'charge relay' system to activate the serine. The primary nucleophilic attack produces the first of two tetrahedral intermediates that are stabilized by the presence of two conserved glycines in the active site, known as the oxyanion hole (*figure 1.6*). This intermediate rapidly collapses, aided by the protonated-histidine acting as an acid to produce the acyl-enzyme complex, while the alcohol component is released. The acyl-enzyme complex then undergoes an attack by a water molecule, producing the second tetrahedral intermediate. Here, the water molecule acts as the second substrate in this bi-substrate reaction. Following rapid rearrangement of this intermediate, the enzyme is regenerated and the acid component is released.

Compounds other than water can attack the acyl-intermediate. When alcohol replaces a water molecule as the second nucleophile, this results in a trans-esterification reaction rather than hydrolysis (as discussed previously).

There is much speculation concerning the exact nature of this mechanism; not only is the active site able to bind ligands, but so are two other known locations on the enzymes as well (*see 1.2.3)*.

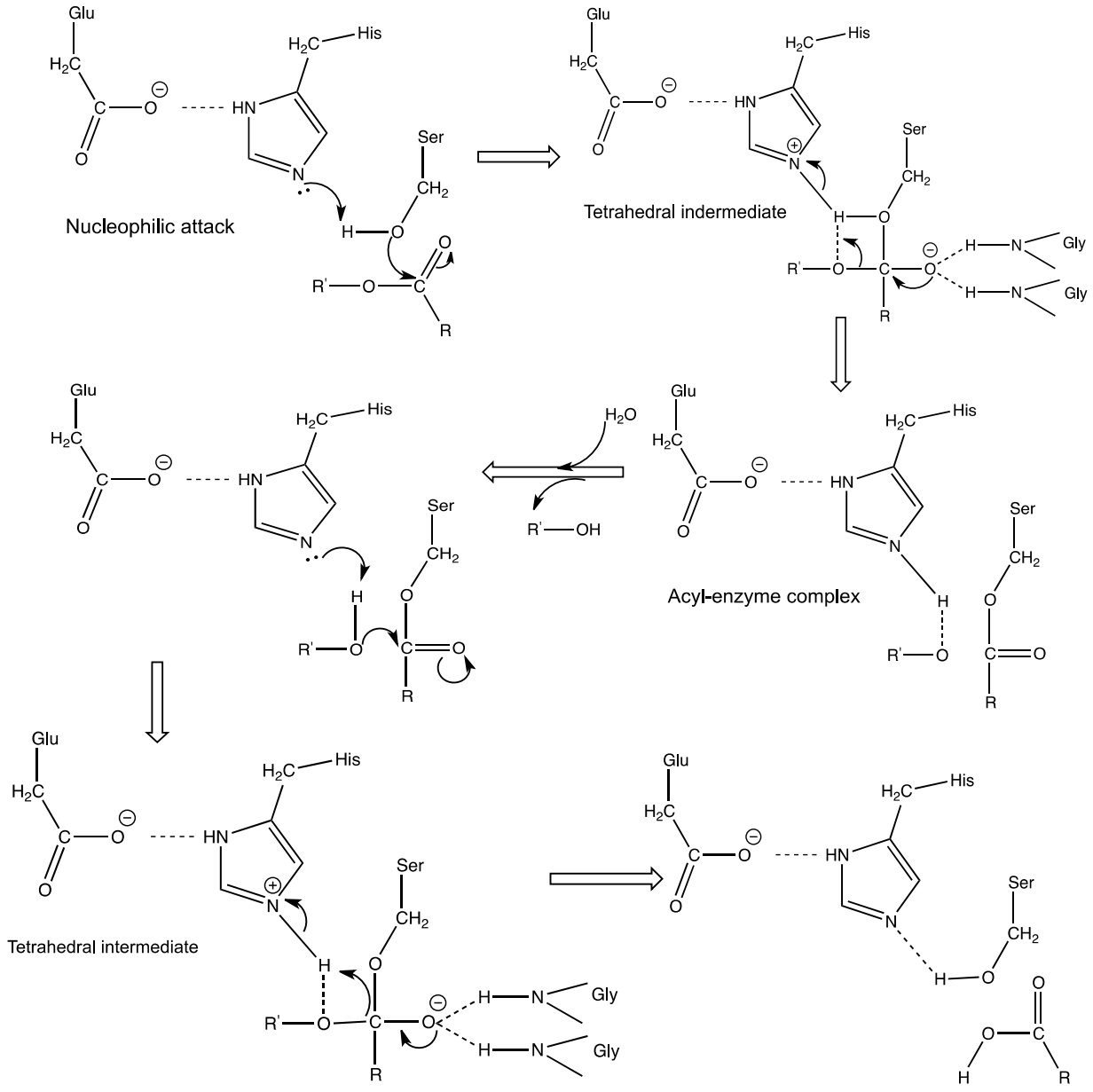


Figure 1.7. Proposed mechanism of catalysis demonstrated by carboxylesterases.

1.2.2.1. Second Conserved Serine

A second absolutely conserved serine was identified in a comparative study of several esterases and lipases[65]. This residue has been found to provide structural support for the spatial orientation of the glutamic acid thus stabilizing the catalytic triad via hydrogen-bonding rather than acting as an essential catalytic residue[45, 87]. Studies in which this serine was mutated revealed that K_{cat} was reduced rather than K_m[87], indicating that this second serine is not directly involved in substrate binding. Published structures of CES1 indicate that this second conserved serine, S247, is structurally conserved, in the correct orientation and within hydrogen-bonding distance (2.49 Å) to the catalytic glutamic acid[85]. Looking at the sequence alignment of CES1-3 (*figure 1.2*), the conserved 'SGV' motif containing this conserved serine is also present in CES2 (S254) and CES3 (S255).

1.2.3. Multiple Binding Sites

Crystallographic structures of CES1 (*table 3.1 in Chapter 3*) and rCE reveal the presence of two other non-specific binding sites: a 'Z-site' that is located within the regulatory domain of the enzyme[84, 85], that is thought to control the trimer–hexamer equilibrium of the enzyme and a 'side-door' secondary pore that leads into the active site from the surface of the enzyme (*figure 1.5)*.

1.2.3.1. The Z-Site

The Z-site was first described by Bencharit et al.[84] and is formed by the inter-digitation of two loops adjacent to the active site that give rise to a Z-shaped dimer interface. Residues involved in this site include N353, G356, W357, K414, T462 and V463. In the crystal structure of CES1 in complex with the cocaine homologue, homatrophine (1MX5),

the substrate was found to be bound to this Z- site[84] a second, low-affinity surface ligand-binding region. The CES1-homotrophine complex only contained trimers in the asymmetric unit, which led to the suggestion that binding of the ligand to this site prevented the packing of the hexamer, shifting the trimer-hexamer equilibrium towards the trimeric state, facilitating substrate binding and enzyme catalysis[84]. The authors proposed that the surface of the Z-site served as a platform for the formation of a hexamer, thus, it is only available to bind to ligands when the enzyme is in its trimeric form. This model has been supported by other studies[69]. The Z-site has also been seen to bind cholate (a bile salt and water soluble cholesterol analogue), a feature that may be crucial in the activity of the enzyme in terms of processing cholesterol esters and other endogenous substrates[69]. One possible theory for involvement of the Z-site in the allosteric activation of catalysis is that the association of compounds at this site may facilitate correct positioning of the catalytic glutamic acid. This is because in the related rabbit carboxylesterase structure, the region that corresponds to the Z-site was disordered, and the equivalent glutamic acid residue was observed in a position not amenable to contribute to catalysis[12].

Structural observations, as well as kinetic data, indicate that inhibitors which associate with both the active site and Z-site are mixed-type inhibitors of CES1[78]. This strongly suggests that the Z-site functions allosterically in CES1 activity, although this has not been explored.

Another mammalian carboxylesterase, the guinea pig homolog of CES1 (gpCE), which shares 71% sequence identity with the human enzyme, also contains a second substrate binding site that functions allosterically and contributes to the positive cooperativity displayed by the enzyme[88]. Therefore it would be reasonable to imagine that both CES1 and CES2 would show positive cooperativity through the binding of substrates to their allosteric sites. Because this has not been investigated previously, this avenue has been explored in this project, and the results are presented in Chapters 3 and 4.

1.2.3.2. The Side-Door

Published structures also revealed the existence of a 'side- or back- door' adjacent to the large, flexible active site, proposed to act as an alternative opening for the trafficking of substrates and products[12, 69], for example ethanol[84]. This pore has three common features which are also observed in rCE and a related bovine lipase[89]: *i*) a 'gate' (M425), *ii*) a 'switch' (F426) and *iii*) a 'releasing valve' (F551). The entrance to the side-door on the surface of CES1 is lined by structurally flexible residues including T252, V254, and L388 that form β-strands and random coils[84], and is separated from the active site by a thin wall. This side door is related in function to the back door observed in acetylcholinesterases[70].

Research on *p*-Nitrobenzyl esterase from *Bacillus subtilis,* an enzyme that exhibits close structural homology and a similar side-door feature to mammalian carboxylesterases, has probed the role of the specific gate residue at the side door (L362). L362 is analogous to the 'gate' residue M425 in CES1[69], and it has been found that substituting this residue with other amino acids[90] impacted kinetic parameters of hydrolysis, demonstrating the importance of the side-door to overall carboxylesterase function. Examination of the parameters for each substitution demonstrated that the most detrimental effect on enzyme activity was caused by a positively charged mutant (L362R) which significantly compromised both the rate of substrate association (K_1) and turnover (K_{cat}). The chemistry of the catalytic triad is not altered to any significant extent in the L362R mutant, thus the nucleophilic potency of the catalytic serine residue is maintained.

The fatty acid palmitate has been observed bound to this side-door in CES1[69] and crystallographic studies of rCE show 4-PP, a product of CPT-11 hydrolysis, bound to the surface of the enzyme adjacent to the Asn-linked glycosylation site (*see 1.2.5)*, supporting

the theory of a 'back-door' product exit site[12]. However, it should be noted that in their methodology, they state that the enzyme was set up for co-crystallization with 4-PP at a 1000-fold molar excess relative to protein concentration. Thus 4-PP binding to this region of rCE may have been fortuitous due to the very high concentration of it present in the crystallization experiment.

Since ethanol is proposed to access the active site of CES1 through the side-door, this alternative passageway to the active site may play an important part in trans-esterification reactions. In terms of cholesteryl ester hydrolysis, the taurocholate-CES1 structure (2DR0) implies that hydrolysis would entail extrusion of the fatty acyl tail from the active site via this side-door. In this case, CESs may interact with a fatty acid binding protein (FABP), a lipid cargo protein that may facilitate the departure of the free fatty acid through this side door[69]. However, fundamentally, the exact function of the side door in catalysis is not fully understood.

The variety of compounds observed to bind to the CES1 Z-site[69, 78] and side-door firmly establishes the promiscuity of these regions of the enzyme, a trait that mirrors the active site.

1.2.4. Mutations in Human Carboxylesterase Genes

With the vast number of therapeutic agents subject to carboxylesterase hydrolysis, there has been an increase in documented mutations of human CES genes that alter the activity of these enzymes, and have an impact on drug performance. Variants in CES1 have been reported to affect the metabolism of imidapril, oseltamivir, clopidogrel, medoxomil and methylphenidate, while polymorphisms in CES2 have been observed to affect CPT-11, orlistat, gemcitabine and aspirin activity. Not only are these genetic variants clinically

relevant in interpreting the variation in individual response to pharmacologic therapies, they may also aid the understanding of human CES mode of action.

In studies into the hydrolysis of trandolapril, an angiotensin-converting-enzyme inhibitor used to treat hypertension, two CES1 mutations, p.G143E (point mutation) and D260fs (frame shift), were identified. These mutant forms were essentially dysfunctional with regard to hydrolysis of the drug[91]. Investigating published structures helps to understand why these mutations are so severe: G143 defines the oxyanion hole (along with G142), and it is the amide nitrogens of these two crucial residues that stabilize the transition state after nucleophilic attack by S221[84]. D260fs generates a frameshift, resulting in a truncated mutant form of the protein after a premature stop codon, which has a detrimental effect on enzyme activity. This single amino acid substitution in CES1, G143E, resulting from the polymorphism 428G>A, also impairs Tamiflu™ bio-activation in humans[92].

1.2.5. Glycosylation

Human CESs possess conserved high-mannose *N*-linked glycosylation sites. CES1 has one site, N79, CES2 has two, N111 and N276, and CES3 has one, N105. Interestingly, all of these glycosylation sites are located in the catalytic domain apart from N276 in CES2, which is present in the αβ domain. On alignment of the CES1-3 sequences, none of these sites align, suggesting that they may have arisen independently and are in fact enzyme specific (*figure 1.2*).

Over 20 years ago, Kroetz et al. provided evidence that *N*-linked glycosylation was essential for maximal catalytic activity of CES1 towards simple aromatic and aliphatic esters[93]. Since then, it has also been suggested that these glycans are essential in protein folding, secretion[3], stabilization of the 3D structure[94] and in trimer formation[78]. However, there has been no thorough investigation into the true importance of glycosylation in

human carboxylesterases, and this is an area that has been explored in this project, with the findings presented in Chapters 3 and 4.

1.2.6. Oligomerisation

CES1 has always been reported to exist in a trimer- hexamer equilibrium[84], whereas CES2 exists primarily as a monomer[95]. Structural studies have shown that the trimer interface in CES1 is formed through intermolecular charge-clamps from R186 and E183 of one monomer to E72 and K78 respectively, of the next monomer. The hexameric unit is formed by two trimers stacked together with their substrate binding gorges facing towards one another[85]. Recently Lamego et al. reported that CES2 is able to form larger oligomers[96] that run at approximately 120 – 150 kDa under native, non-denaturing conditions, but no further work has been carried out on this.

1.3. Importance of Studying Human Carboxylesterases

Research into human CES1 and CES2 has accelerated dramatically over the last ten years (*figure 1.8)* as their widespread involvement in numerous unrelated biological processes has become increasingly apparent. This is partly due to whole genome sequencing of many species alongside continued efforts at trying to understand the structural elements of their active sites, as well as their emergence in a clinical context. Since a vast number of drugs are metabolized by CES enzymes, it is important to clarify their structure, substrate selectivity, tissue distribution, and species specificity. A full understanding of the enzymes structural and kinetic behaviour has implications for a variety of applications, including tissue/cell specific drug targeting, potential treatment of drug overdose, the design of herbicides with selective toxicity[64], use as prophylactic agents against toxic esters[97] and in clinical pro-drug activation.

Carboxylesterases are a well recognized and documented, yet incompletely characterized family of enzymes. Their involvement in such a diversity of processes calls for a much clearer understanding of their specific structures, which will help to rationalize the basis of their individual substrate preference.

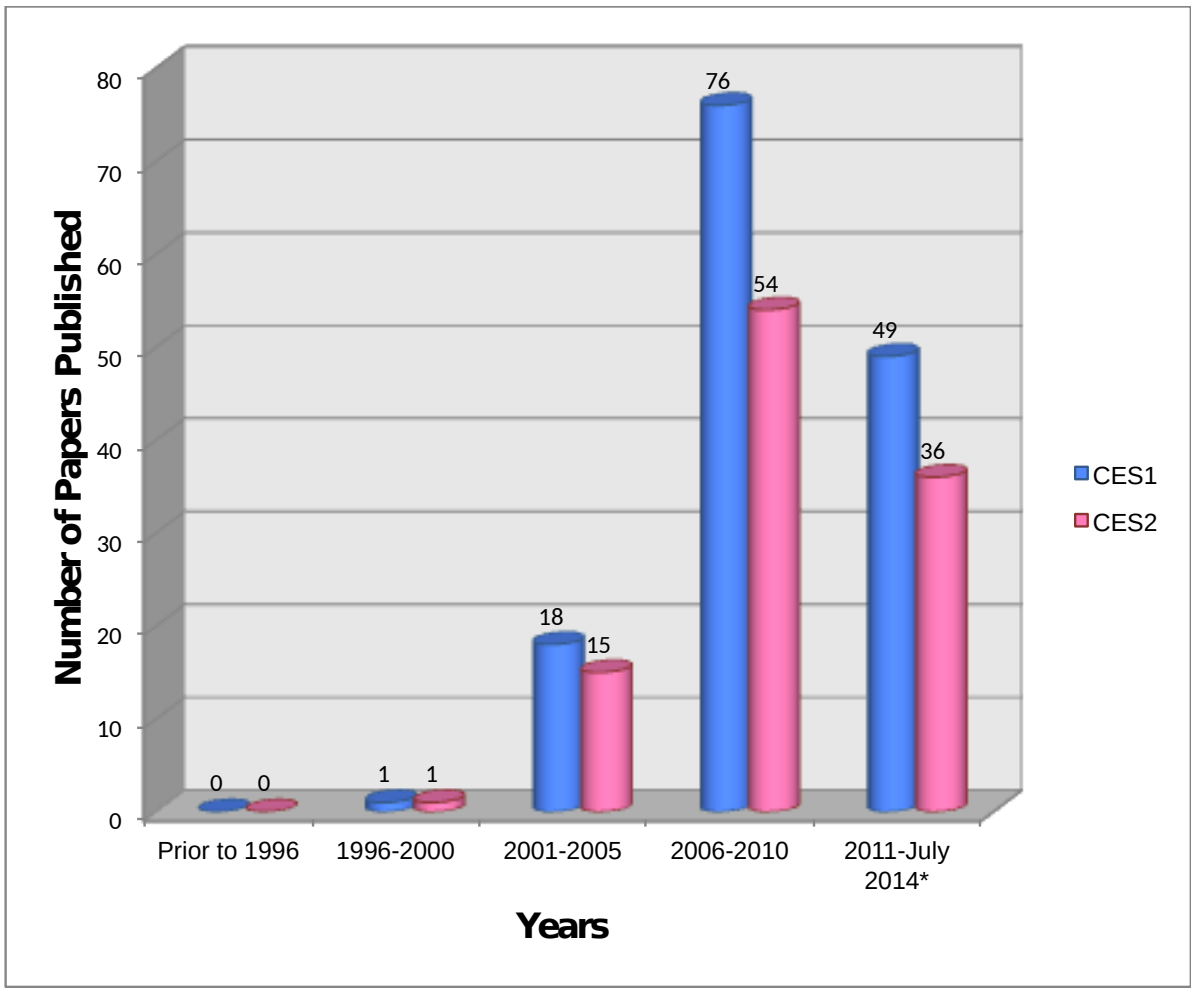


Figure 1.8. Graph showing the numbers of CES1 and CES2 papers published since 1996. Numbers shown in five year time frame windows. Information was sourced from PubMed NCBI. Publications prior to 2000 may not be shown here, as there was no standard nomenclature of human carboxylesterases. Data here was collected in July 2014.

1.3.1. Scope and Objectives of This Project

The primary focus of the D. Phil project presented here was to expand the current knowledge of human CESs, in terms of their structure-function relationship, using a structural enzymology approach. Because the only deposited structures of human CESs are of CES1, and the majority of enzymatic work to date has been performed predominantly using this enzyme, the depth of understanding of the remaining CESs is severely limited. Many questions remain as to their overlapping yet distinct substrate specificities, which can only really be elucidated from obtaining and examining high-resolution crystallographic structures.

To bridge this knowledge gap, four main objectives were set out in this project. Firstly, to produce sufficient quantities of native and mutant forms of CES1 and CES2, using the OPPF-UK (https://www.oppf.rc-harwell.ac.uk/OPPF/) high-throughput pipeline, for downstream functional and structural studies. Secondly, to probe the importance of *N*-linked glycosylation in these enzymes enzymatically, biophysically and structurally. Thirdly, to gain a higher resolution structure of CES1 than those already published using X-ray crystallography for structural studies to aid drug discovery. And lastly, to determine the novel structure of CES2, and to compare this to that of CES1. Ultimately it will be the information gained from this novel structure that will allow a detailed understanding of the molecular basis for the selectivity and specificity of these promiscuous enzymes.

As the project progressed, there was an opportunity to begin work on CES3, through the same approaches as above. This will not be discussed in the main section of this thesis, but the preliminary work obtained so far will be presented in Chapter 5.

2. MATERIALS AND METHODS

2.1. Overview

This chapter summarizes the experimental methods used during the course of this project. All of the work summarized in the flow diagram (*figure 2.1)*, apart from data collection, was carried out at the Oxford Protein Production Facility UK (OPPF-UK), which is located in the Research Complex at Harwell, adjacent to The Diamond Light Source (DLS). The OPPF-UK was established in 2001 and is a structural proteomics facility in the Division of Structural Biology, part of the Department of Medicine, Oxford University. X-ray diffraction data and Small Angle X-ray Scattering experiments were carried out at DLS.

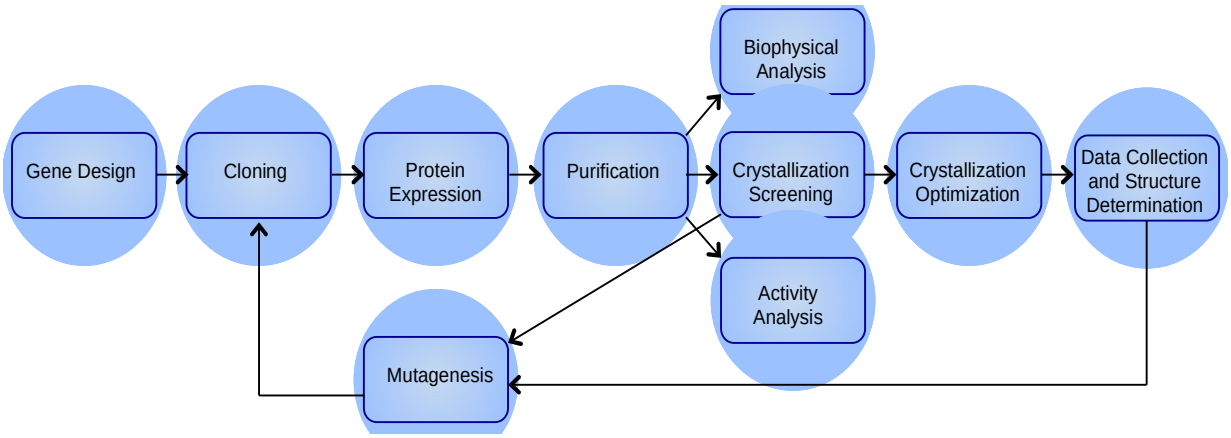


Figure 2.1. Diagram to show the workflow of experimental processes used in this project.

2.2. Protein Expression and Purification

2.2.1. Protein Sequence Analysis

Protein identity was calculated on http://web.expasy.org/sim/, using the parameters: comparison matrix: BLOSUM62, number of alignments computed: 20, gap open penalty: 12, gap extension penalty: 4.

Other bioinformatics tools used in this thesis include:

- http://services.mbi.ucla.edu/SER/
- http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index
- https://www.strubi.ox.ac.uk/RONN
- http://www.ebi.ac.uk/msd-srv/ssm/
- http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html

2.2.2. Synthetic Gene Design

Synthetic cDNAS encoding CES1, CES1 null, CES2, CES2 null, CES2 Dere, CES2 Phyre and CES3 were designed, optimized for combined *Homo sapiens* and *Spodoptera fugiperda* codon usage (see *Appendix 1.3*) and supplied by GeneArt® (*table 2.1).*

Protein Name	Description	Production
CES1	Full length	Ordered as a synthetic gene.
CES1 null	Null mutant (S221A)	
CES2	Full length	
CES2 null	Null mutant (S228A)	
CES2 Dere	Designed using the Derewenda SER prediction (513-515 QEE → AAA, 360-362 QKE → AAA)	
CES2 phyre	PHYRE disorder prediction (7 amino acids deleted from *C*-terminus)	
CES3	Full length	
CES1 N79Q	Glycosylation knockout	Made through overlap PCR.
CES2 N111Q	Single glycosylation knockout	
CES2 N276Q	Single glycosylation knockout	
CES2 N111Q+ N276Q (double K/O)	Double glycosylation knockout N111Q and N276Q	
CES2 3-4	R86K, K190E, D281K	
CES2 4-4	R86K, K190E, D281K, K297E	
CES3 2-3	R88K, S282K	
CES3 3-3	R88K, S282K, G298E.	
CES3_trunc	Deletion of 14 disordered residues from *C*-terminus, so sequence ended TLPS.	

Table 2.1. Summary of all the proteins produced in this project.

Constructs made through overlap PCR will be explained in the *Mutagenesis* section. Protein sequences can be found in *Appendix 1.1*.

2.2.3. Construction of Expression Vectors

2.2.3.1. PCR

PCR primers were designed in The Oxford Protein Target Information Collection (OPTIC) (*see H. Primer Details)* using an implementation of Primer3[98]. Primers were lyophilized when delivered, and were made up to a primer stock concentration of 100 pM/ µL using nuclease-free water. Working primer stocks (10 pM/ µL) were then prepared from the original stock. Genes were amplified from the synthetic cDNAs using KOD

Xtreme™ Hot Start DNA Polymerase (Novagen). The standard 50 µL PCR reaction and thermal cycling parameters were:

KOD Xtreme™ Hot Start DNA Polymerase reaction setup

	µL
2 x Xtreme Buffer	25.0
dNTPs (2 mM each)	10.0
PCR Grade Water	10.0
Forward Primer (10 µM)	1.5
Reverse Primer (10 µM)	1.5
Template DNA (10 ng/ µL)	1.0
KOD Xtreme™ Polymerase	1.0
Total reaction volume	***50.0***

Polymerase activation	94°C	2 min	
Denaturation	98°C	10 s	} 35 cycles
Annealing	*° C	30 s	
Extension	68°C	1 min/kbp	
Hold	4°C	∞	

**Lowest primer time*

2.2.3.2. Agarose Gel Electrophoresis

PCR products were analysed using gel electrophoresis. Gels of 1.5% w/v agarose (Sigma) were made using 1 x TBE buffer (Sigma) with 5 µL of SYBRsafe DNA stain (Invitrogen) in 50 mL. 8 µL of PCR product was loaded with 2 µL of 6 x DNA loading dye (0.25% w/v Bromophenol Blue, 30% w/v glycerol) onto the gel. 5 µL of Hyperladder I (Bioline) was run alongside as a protein standard in each gel, and 1 x TBE was used as the running buffer. Gels were run at 100 V for approximately 1 hour, and were then visualized using a Biorad ChemiDoc™ MP System.

2.2.3.3. Purification of PCR products

Products were treated with *Dpn1* (NEB) and purified using Agencourt® AMPure® magnetic beads (Beckman Coulter) on a SPRIP 96R magnetic plate according to the manufacturer's instructions, and eluted in 30 µL of elution buffer (EB). *Dpn1* was employed to remove any remaining template DNA in the reaction mix, as it only cuts methylated DNA, leaving the PCR products un-restricted.

2.2.3.4. Cloning with In-Fusion™ Technology

Genes were cloned into 5 vectors using ligation-independent cloning: pOPINE, pOPING, pOPINTTGneo, pOPINTTGneo CD4 and pOPINTTGneo HALO. Detailed vector maps of can be found in Appendix 1.4. All vectors contained *C*-terminal His tags for detection and purification, and promoter sequences for mammalian cell expression. pOPINE and pOPING also contained the T7 promoter for expression in *E.coli* and the p10 promoter and baculovirus recombination sequences for virus construction and expression in insect cells. pOPINE and pOPING were used to compare secretion from the native CES leader sequence (pOPINE) with the human µ-phosphatase leader sequence resident in the vector (pOPING). pOPINTTGneo, pOPINTTGneo CD4 and pOPINTTGneo HALO contained a *Neo* selectable maker gene for the generation of stable cells line. CD4 is a *C*-terminal fusion protein that has been used to enhance expression of secreted proteins[99]. The HALO vector contained a *C*-terminal Halotag® which provides an alternative to the His tag for detection and purification[100].

In-Fusion™ reactions were performed using Clontech Infusion™ (Takara) as previously described[101, 102]. A description of how this technology works is shown in *figure 2.2.* 1 µL of linearised pOPIN vector (see *appendix 1.4* for restriction sites details) was mixed with 2 µL of purified insert (obtained from PCR amplification) and 7 µL of water in wells in a PCR plate. This 10 µL mix was then transferred to a 50 µL tube containing a dry down

In-Fusion™ pellet. Contents were mixed briefly, transferred back to the PCR plate, sealed with a foil lid and incubated at 42°C for 30 minutes in a PCR block (Bio-Rad). Immediately after 30 minutes, the reaction was stopped by adding 40 µL of TE buffer.

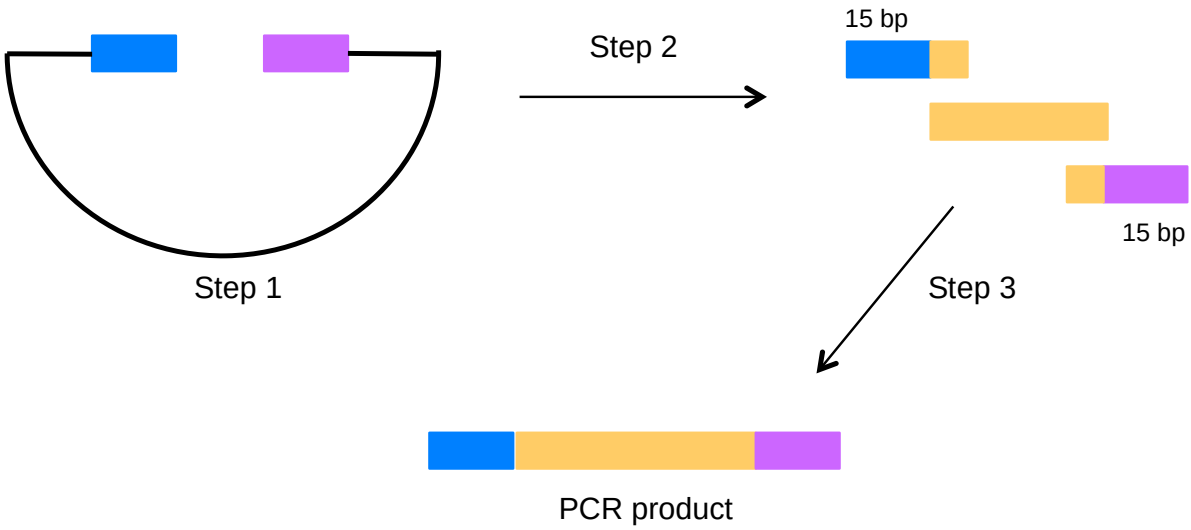


Figure 2.2. Diagram to show how the Infusion™ Cloning Technology works.
This system allows you to fuse the ends of the PCR fragment to the homologous ends of a linearized vector. The 3' and 5' regions of homology are generated by adding 15 bp extensions to both PCR primers that precisely match the ends of the linearized vector. When the vector is combined with your insert, the In-Fusion enzyme converts the double-stranded extensions into single-stranded DNA and fuses these regions to the corresponding ends of the linearized vector. **Step 1.** Generation of a linearised vector. **Step 2.** Design of gene-specific primers with 15 bp extensions homologous to vector ends. **Step 3**. Amplification of gene of interest to generate PCR product.

2.2.3.5. Transformation of *E. coli*

50 µL of OmniMAX™ T1 phage resistant chemically competent cells (Life Technologies) were defrosted on ice for 20 minutes (while the Infusion reaction was taking place). 5 µL of diluted In-Fusion™ reaction was used per aliquot of competent cells. They were incubated on ice for 30 minutes, and then subjected to heat shock treatment at 42°C for 45 s in a water bath. The cells were then directly returned to ice, 300 µL SOC medium (Life Technologies) was added, and the cells were then incubated at 37°C for 1 hour (without shaking). 60 µL of transformed cells were then plated onto blue/ white selection agar plates, spread and left to dry. The plates were prepared using molten

2% w/v agar in LB broth, supplemented with the appropriate antibiotic (Kanamycin or Carbenicillin) at 50 µg/ mL, X-Gal at 20 mg/ mL and Isopropyl β-D-1-thiogalactopyranoside (IPTG) at 1 mM. 3 mL was poured into each well of a 6-well plate (Corning). After plating the cells, the plates were tipped left to right and up and down to evenly spread the cells, inverted (with lids on) and incubated at 37°C overnight.

2.2.3.6. Plasmid DNA Preparation (Miniprep)

Single white colonies were picked using a pipette tip and were grown in a 24-well deep block (Qiagen) with 2 mL of lysogeny broth (LB) and carbenicillin at a final concentration of 50 µg/ mL (from a 50 mg/ mL stock). Only white colonies were picked as the blue colonies were derived from inefficiently linearized parental plasmid and were non-recombinant. Carbenicillin was used as it is more stable than ampicillin. Blocks were sealed with a gas permeable lid (Abgene), and were left to grow overnight in a shaking incubator at 37°C and 220 rpm. Cultures were transferred to 2 mL Eppendorf tubes and spun down at 13,000 g for 10 minutes. The supernatant was discarded and plasmid DNA was extracted from the pellets using the QIAcube (QIAGEN) with the standard DNA mini-prep program.

2.2.3.7. Construct Verification

Constructs were firstly checked by PCR screening using Phusion Flash (Thermo Scientific), pOPIN forward, and gene specific reverse primers. PCR positive clones were subjected to DNA sequence analysis using the primers T7fwd and Triexdown (for pOPINE and G) or pTT fwd and rvs (for TTG vectors) for validation that the desired constructs were obtained. This was performed by Source BioScience, using their Sanger sequencing service (http://www.lifesciences.sourcebioscience.com/genomic-services/sanger-sequencing-

service/). Gene specific internal primers were also designed and used in sequencing reactions when a full coverage of the insert was not obtained by simply using forward and reverse primers. Details of all primers can be found in *Primer Details*.

Phusion Flash High Fidelity DNA Polymerase reaction setup:

	µl
Phusion Flash PCR Mastermix	12.5
pOPIN forward	1.50
Specific reverse primer	1.50
Vector	0.50
dH_2O	9.00
Total reaction volume	***25.0***

Initial denaturation	98°C	10 s		
Denaturation	98°C	1 s	}	30-35 cycles
Annealing	55° C	5 s		
Extension	72°C	15 s/kb		
Final extension	72°C	1 min		
Hold	4°C	∞		

2.2.3.8. Primer Details

PCR Primers

Gene	Description	Given I.D.	Cloning Vector	Forward Primer	Reverse Primer
CES1	Full length	8187	pOPINE	8187 fwd	8187 rvrse
		8188	pOPING	8188 fwd	8187 rvrse
		8856	pOPINTTG neo	8188 fwd	8187 rvrse
		8857	pOPINTTG neo cd4	8188 fwd	8857 rvrse
		8858	pOPINTTG neo halo	8188 fwd	8857 rvrse
CES1	Null mutant	8189	pOPINE	8187 fwd	8187 rvrse
		8190	pOPING	8188 fwd	8187 rvrse
		8859	pOPINTTG neo	8188 fwd	8187 rvrse
		8860	pOPINTTG neo cd4	8188 fwd	8857 rvrse
		8861	pOPINTTG neo halo	8188 fwd	8857 rvrse
CES2	Full length	8193	pOPINE	8193 fwd	8193 rvrse
		8194	pOPING	8194 fwd	8193 rvrse
		8862	pOPINTTG neo	8194 fwd	8193 rvrse
		8863	pOPINTTG neo cd4	8194 fwd	8863 rvrse
		8864	pOPINTTG neo halo	8194 fwd	8863 rvrse
CES2	Phyre	8201	pOPINE	8193 fwd	8201 rvrse
		8202	pOPING	8194 fwd	8201 rvrse
		8877	pOPINTTG neo	8194 fwd	8193 rvrse
		8878	pOPINTTG neo cd4	8194 fwd	8878 rvrse
		8879	pOPINTTG neo halo	8194 fwd	8878 rvrse
CES2	Null mutant	8197	pOPINE	8193 fwd	8193 rvrse
		8198	pOPING	8194 fwd	8193 rvrse
		8865	pOPINTTG neo	8194 fwd	8193 rvrse
		8866	pOPINTTG neo cd4	8194 fwd	8863 rvrse
		8867	pOPINTTG neo halo	8194 fwd	8863 rvrse
CES2	Dere	8195	pOPINE	8193 fwd	8193 rvrse
		8196	pOPING	8194 fwd	8193 rvrse
		8868	pOPINTTG neo	8194 fwd	8193 rvrse
		8869	pOPINTTG neo cd4	8194 fwd	8863 rvrse
		8870	pOPINTTG neo halo	8194 fwd	8863 rvrse
CES3	Normal	8199	pOPINE	8199 fwd	8199 rvrse
		8200	pOPING	8200 fwd	8199 rvrse
		8871	pOPINTTG neo	8200 fwd	8199 rvrse
		8872	pOPINTTG neo cd4	8200 fwd	8872 rvrse
		8873	pOPINTTG neo halo	8200 fwd	8872 rvrse

The given I.D. is an OPPF number that uniquely defines different constucts made for a specific target.

Primer Sequences for PCR

Primer	Sequence (5'-3')
8187 fwd	aggagatataccatgTGGCTGCGCGCCTTCATCC
8187 rvrse	gtgatggtgatgtttCTCGGTCTGGGGGGGCTTTTC
8188 fwd	gcgtagctgaaaccggcCACCCTTCCTCTCCCCCTG
8857 rvrse	cagaacttccagtttCTCGGTCTGGGGGGGCTTTTC
8193 fwd	aggagatataccatgCGCCTGCACAGGCTGAGGG
8194 rvrse	gcgtagctgaaaccggcCAGGACTCCGCCTCCCCAATC
8863 rvrse	cagaacttccagtttGCGTTCCTCGGGTTCTTCCAG
8878 rvrse	cagaacttccagtttTTCCTGGATCTTCTGGGGCAGG
8201 rvrse	gtgatggtgatgtttTTCCTGGATCTTCTGGGGCAGG
8199 fwd	aggagatataccatgGAAAGGGCCGTGCGCGTG
8199 rvrse	gtgatggtgatgtttGGCCTTGCGGTTCTTCTGCTTC
8200 fwd	gcgtagctgaaaccggcACCGGACCTGAGGTGGCCC
8872 rvrse	cagaacttccagtttGGCCTTGCGGTTCTTCTGCTTC

Primers For PCR verification

Gene	Sequence (5'-3')
pOPIN fwd	GACCGAAATTAATACGACTCACTATAGGG

Primers For DNA sequencing

Gene	Sequence (5'-3')
T7 Fwd	TAATACGACTCACTATAGGG
TriEx Down	TCGATCTCAGTGGTATTTGTG
pTT fwd	TCCACAGGTGTCCACTCC
pTT rvs	TCCTTTATTAGCCAGAGG

Gene Specific Internal primers

Gene	Internal Primer (5'-3')
CES1	CCATCCAGTACCGCCTGGGC
CES2	GTTCCTGTCCCAGTTCAACATGACC
CES3	TACGGCGACGTGGTGGTGG

Primers for Mutagenesis

Enzyme	Primer name	Primer Sequence (5'-3')	Length (bp)
CES1	N79Q	ggtccttcgtgaagcaggccacctcctaccc	31
	N79Q_antisense	gggtaggaggtggcctgcttcacgaaggacc	31
CES2	N111Q	agagtttcttagccagttccaaatgaccttcccttccgac	40
	N111Q_antisense	gtcggaagggaaggtcatttggaactggctaagaaactct	40
	N276Q	ccacggtggtggcccaactgtctgcctgtg	30
	N276Q_antisense	cacaggcagacagttgggccaccaccgtgg	30
	R86K	gtcttggagcggagtgaaggacggaaccacccacc	35
	R86K_antisense	ggtgggtggttccgtccttcactccgctccaagac	35
	K190E	gcaccggcgacgagcacgccacc	23
	K190E_antisense	ggtggcgtgctcgtcgccggtgc	23
	D281K	cctgtccgcctgcaagcaggtggactccg	29
	D281K_antisense	cggagtccacctgcttgcaggcggacagg	29
	K297E	tgcgcggcaagtccgaagaagagatcctg	29
	K297E_antisense	caggatctcttccttggacttgccgcgca	29
	K297E opt	tgcgcggcaagtccgaggaagagatcctg	29
	K297E_antisense opt	caggatctcttcctcggacttgccgcgca	29
CES3	R88K	ccttgggaaggcgtgaaggatgcttccaccgcc	33
	R88K_antisense	ggcggtggaagcatccttcacgccttcccaagg	33
	S282K	aacaccctggcctgcaagtccagctccccagcc	33
	S282K_antisense	ggctggggagctggacttgcaggccagggtgtt	33
	G298E	ctgcagcagaaggaagaagaggaactggtcctga	34
	G298E_antisense	tcaggaccagttcctcttcttccttctgctgcag	34
	G298E opt	ctgcagcagaaggaagaggaggaactggtcctga	34
	G298E_antisense opt	tcaggaccagttcctcctcttccttctgctgcag	34
	CES3_trunc_antisense	gtgatggtgatgtttggagggcagggtttcggacc	35

2.2.3.9. Plasmid DNA Preparation (Megaprep)

To produce milligram quantities of plasmid DNA, megapreps were prepared using a PureLink® HiPure Plasmid Megaprep Kit (Invitrogen). Cells harvested from a 1 L overnight *E.coli* culture were re-suspended thoroughly in 100 mL of Resuspension buffer. 100 mL of Lysis Buffer was added, mixed gently and left at room temperature for no longer than 5 minutes. 100 mL of Precipitation buffer was added to the mixture and inverted 10-15 times. This lysate was then loaded onto 2 x Lysate filter cartridges attached to 1 L Duran bottles and allowed to stand for 5-10 minutes to settle, before vacuum filtration. Pressures for each stage were taken from the manufacturer's protocol. 50 mL Wash buffer was added to each filter and the vacuum re-applied. The filtered lysate from the two bottles was then pooled and applied to a single pre-equilibrated DNA

binding cartridge (manufacturer's protocol) and vacuum applied until all the lysate had passed through. Following 2 x 175 mL washes with Wash buffer, the DNA was eluted into a fresh Duran bottle by applying 2 x 50 mL Elution buffer under soft vacuum. The eluent (100 mL) was transferred to a 500 mL centrifuge bottle, and 70 mL of 2-propanol (isopropanol) was added. This was centrifuged at 6000 g for 1 hour. The 2-propanol was carefully tipped off, leaving the fragile pellet undisturbed. 20 mL of 70% ethanol was then used to wash the pellet. This was centrifuged for a further 50 minutes. The ethanol was discarded, and the pellet was left to air-dry to ensure complete evaporation of any residual ethanol. The pellet was re-suspended in 2.5 mL TE buffer and quantified by absorbance at 260/280 nm using a NanoDrop 1000 (Thermo Scientific) spectrophotometer. For storage, DNA was frozen at -20°C.

2.2.3.10. Mutagenesis

Variants of CES1, CES2 and CES3 were prepared by site-directed mutagenesis, using the strand overlap extension PCR method[103], see *figure 2.3* for details. Briefly, complementary oligonucleotides were synthesized to introduce a substitution. The primers were annealed to the corresponding wild type construct and subjected to a thermocycler for a total of 15 cycles. The resultant PCR-amplified constructs were then digested with *DpnI* to remove the non-mutated parent backbone. The mutated PCR-amplified constructs were used to transform XL1-Blue bacteria. All mutated constructs were subjected to sequencing analysis to confirm the desired mutation being made, without any secondary mutations.

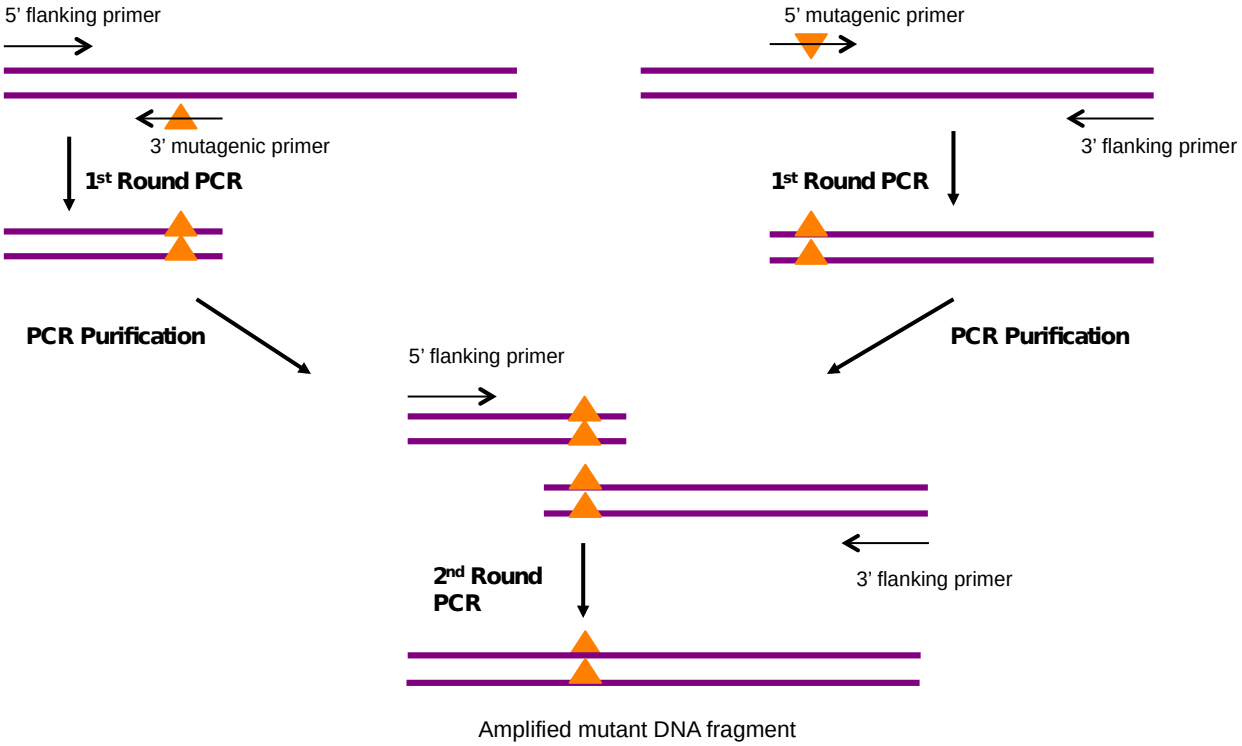


Figure 2.3. Site-directed mutagenesis, using the strand overlap extension PCR method[103]. This method required four oligonucleotide primers and three separate amplification reactions. Two complementary mutagenic primers (*see primers for mutagenesis),* induced the mutation into the desired sequence of DNA, and two flanking primers (*Appendix 1.3.*) amplified the mutant fragment and facilitated cloning of the PCR fragment into a suitable vector. Using this approach, a variety of mutations could be created, sequentially. First, two separate PCR reactions were set up in parallel. One reaction contained the 5' mutant primer and the 3' flanking primer. The other reaction contained the 3' mutant primer and the 5' flanking primer. The two amplified fragments contained mutations at the 5′ or 3′ terminus, respectively. In the second round of amplification, after gel purification, these two PCR fragments were used as templates for amplification, using only the flanking primers. After amplification, the mutation was contained within the target DNA segment, which was cloned into appropriate vectors for DNA sequencing and subsequent functional studies.

2.2.4. Protein Production

2.2.4.1. Transient Expression in HEK293T Cells

HEK 293T cells were maintained in Dulbecco's modified Eagle media (DMEM Gibco) supplemented with 10% heat inactivated fetal bovine serum (Gibco), glutamine and non-essential amino acids, and incubated at 37°C, 5% CO_2. Cells were passaged every 3 days. Spent medium was removed, and 3 mL/ 5 mL of Trypsin, for a T75/ T175 flask (Corning) were added. Flasks were gently rocked back and forth, for 5 minutes until the cells had

dissociated. 9 mL/ 15 mL of DMEM with 10% FCS was added to neutralize the trypsin and 1.5 mL/ 4 mL of solution was added to a new flask with 15 mL/ 25 mL of fresh media. Expression screening in 24-well plate format was carried out as described previously[101]. Briefly, the day before transfection, cells were seeded in 24-well plates at 1.5 x 10^5/ mL in a total volume of 1 mL DMEM plus supplements. The following day, 1 µg of each DNA construct was mixed with 2 µL Genejuice™ (Novagen) in a total volume of 60 µL serum-free DMEM, and then following incubation for 10 minutes, was added to the cells. One well of the 24 well plates was left un-transfected as a negative control and one well was transfected with a GFP expression vector (transfection control). After 72 hours, both the media and whole cell lysates (see below) were harvested and analysed by SDS-PAGE and Western blotting.

2.2.4.2. Large-Scale transient expression in HEK 293T cells

Production was scaled–up to 1.0 L of culture for transient expression in roller bottles (Greiner Bio-One) as described previously[101, 104]. Briefly, four x T175 flasks containing confluent HEK293T cells were resuspended, and added to four individual roller bottles containing 250 mL fully supplemented DMEM. These were left to incubate at 37°C for 3 days in a roller rig. Old media was carefullu removed, and replaced with 200 mL DMEM with 2% FCS and returned to the rig for 1 hour. In a flask, 100 mL serum-free DMEM was mixed thoroughly with 3.5 mL 1mg/ mL PEI. In a separate flask, 100 mL serum-free DMEM was mixed with 2 mg DNA. These two solutions were then combined, and left to incubate at room temperature for a maximum of 10 minutes. If left too long, the efficiency of transfection dropped. 50 mL of this transfection cocktail was added to each of the four roller bottles and returned to the incubator until the media started to turn orange (approximately 6 days).

Secreted proteins were harvested by removing the media from the roller bottles and centrifuging at 6000 g for 30 minutes. The supernatant was then filtered through a 0.45 µM bottle top filter into a clean 1 L Duran bottle. This was stored at 4°C until purification.

To simplify the glycoforms of the expressed proteins, kifunensine (Sigma) was added to the transient expression cultures[105] to a final concentration of 2 mg/ mL. This compound inhibits α-mannisidase I, yielding glycans of the form $GlcNac_2Man_9$, attached to the expressed proteins (*figure 2.4*).

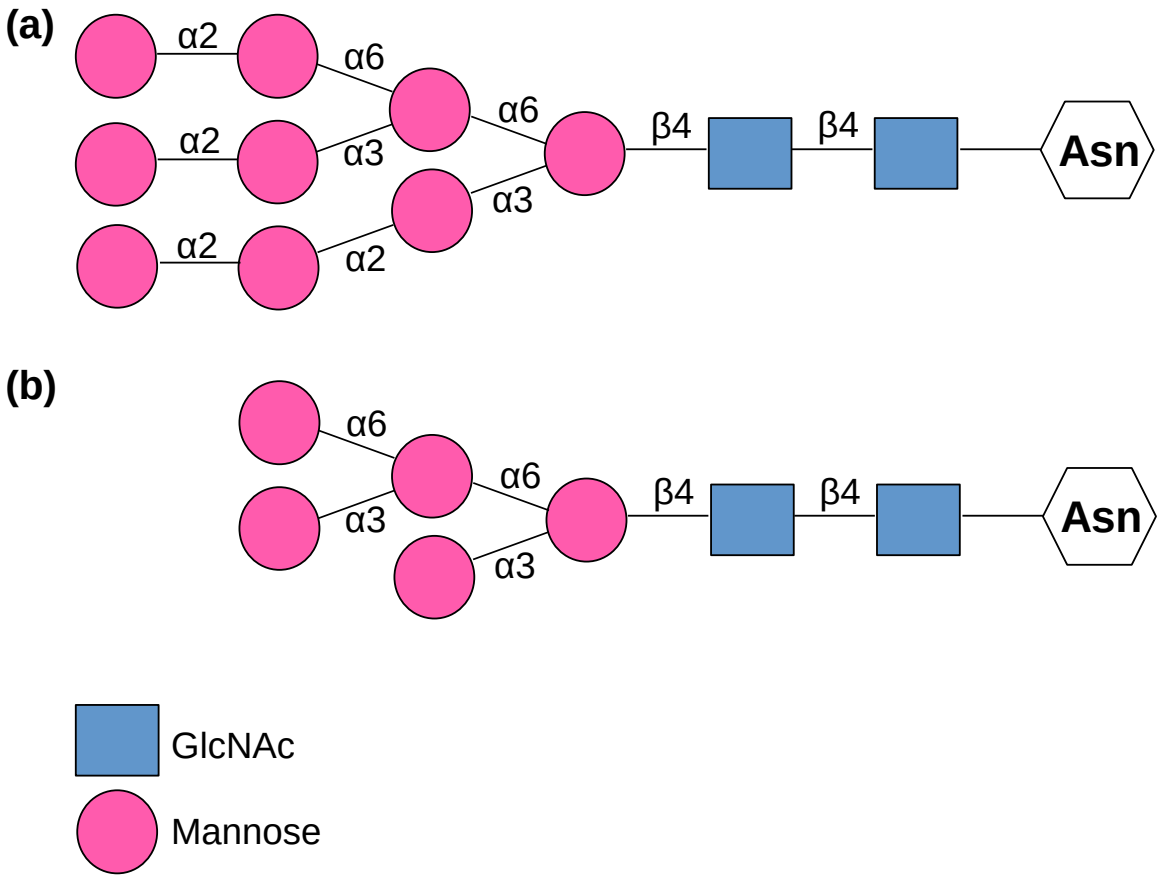


Figure 2.4. A simplified diagram showing the two different forms of homogenous *N*-linked glycosylation found on proteins produced in this project.
a) $GlcNAc_2Man_9$ is the glycan form that is present on proteins that are produced in HEK 293T cells with the addition of kifunensine, and **b)** $GlcNA_2Man_5$ is present on the proteins produced using the stable cell line HEK293 $Gnt^{-/-}$.

2.2.4.3. Construction of Stable HEK293 Gnt $^{-/-}$ Cell Lines Using G418 Selection

A mutant cell line, HEK293 Gnt$^{-/-}$, deficient in *N*-acetylglucosaminyltransferase I (GnT1), was used for the generation of stable cell lines. In the absence of GnT1, glycoproteins expressed by these cells have glycans of the form $GlcNac_2Man_5$ (*figure 2.4.*). Stable cell lines were produced for CES1 (in both pOPINTTG neo CD4 and pOPINTTG neo) and for CES1 null, CES2, CES2 null and CES3 (all in pOPINTTG neo). For each plasmid, DNA (25 μg) was mixed with 75 μL Genejuice (Merck Millipore) in a total volume of 2 mL serum-free DMEM, and incubated at room temperature for 30 minutes. This transfection mixture was added to HEK293 Gnt$^{-/-}$[106] (2 x 10^5 cells/ mL in 50 mL DMEM containing Glutamine and 10 % FCS) in a T75 flask. Cells were incubated with supplemented media for 48 hours. Serial dilutions of cells (1:500, 1:1000, 1:5000, 1:10000) were plated into 96-well flat-bottomed tissue culture plates with fresh media, and incubated overnight. The following day, G418 was added to give a final concentration of 0.8 mg/ mL per well. Plates were left to incubate at 37°C, 5% CO_2. From three weeks onwards, single colonies appeared, and were picked into 24-well tissue culture plates. Western blots of media samples were performed to test expression of all single colonies picked. The best expressing cell lines as assessed by intensity of their Western blot signal, were frozen in complete DMEM containing 10 % DMSO at 10^6 cells/ mL, as previously described[107], and stored in liquid nitrogen. To revive cell lines, a vial containing 10^6 cells was rapidly thawed in a water bath at 37°C for 2 minutes and then propagated in a T75 flask in complete DMEM without G418, until confluent. For large scale protein production, cells were expanded in T175 flasks until confluent and used to seed roller bottles as described for large-scale transient expression. Typically, volumes of 0.5 L were grown for each protein.

2.2.4.4. SDS-PAGE and Western Blot Protein Analysis

To examine protein expression and purity, samples were run on an SDS-PAGE gel and then either stained to get a rapid result, or subjected to Western blot analysis.

Whole cell lysate samples were prepared by firstly rinsing with 1 mL PBS after removal of supernatant. Then 30 μL of lysis buffer (NPI-10 : 50 mM NaH_2PO_4, 300 mM NaCl, 10 mM Imidazole, pH 8.0) (Qiagen), was added to each well. Stocks of lysis buffer were made up in 10 mL aliquots containing 80 μL DNAse (Sigma DN-25-5g, made up to a stock of 100,000 units/ mL in 20 mL water) and 50 μL protease inhibitor liquid stock (Sigma – P8340). This was then left to shake at 37°C for 40 minutes. To prepare both secreted and whole cell samples, 10 μL were boiled with equal amounts of sample buffer (4 x stock, 4% w/v SDS, 0.2% w/v Bromophenol blue, 20% v/v glycerol, 100 mM Tris-HCl pH 6.8) for 10 minutes at 95°C. Precision Plus Protein Dual™ Color Standards (Bio-Rad) and BenchMark™ Protein Ladder (Invitrogen) markers were run in parallel (2 μL of each) as the molecular weight standards for the Western blots.

Samples were loaded into individual wells of NuPAGE® Novex® 10% Bis-Tris Midi Gels (Invitrogen) and run in MOPS (3-(N-morpholino) propane sulfonic acid) SDS Running Buffer (1x) NuPAGE® (Invitrogen) at 200 V for ~ 45 minutes, until the blue dye reached the bottom of the gel. Gels were either stained with InstantBlue™ (Expedeon), following the manufacturer's instructions, or subjected to western blotting using an iBlot™ (Invitrogen). Following transfer of proteins, nitrocellulose membranes were blocked with 5% non-fat dry milk in PBS containing 0.1% Tween-20 (PBST) overnight at room temperature. Membranes were incubated with primary antibody, Anti-His_6 (Roche) (1:500 dilution in the blocking solution) for 2 hours followed by 3 x 20 min washes in PBST, then subsequently incubated with the secondary antibody, Anti-Mouse IgG (whole molecule) Peroxidase (Sigma) (1:5000 dilution in the blocking solution) for 2 hours, followed by 3 x 20min washes in PBST. Blots were developed using the ECL system (GE

Healthcare). Stained gels and Western blots were visualised and imaged using a Biorad ChemiDoc™ MP System.

2.2.4.5. Protein Purification

Medium from both transiently transfected cells and stable cell lines was collected by centrifugation at 6000 g. Secreted proteins were purified using an automated protocol[107] consisting of nickel affinity chromatography using a HiScreen™ Ni FF (GE Healthcare) column (stripped and re-charged with 100 mM $NiSO_4$), followed by size exclusion chromatography using a HiLoad 16/60Superdex 200 column (GE Healthcare) on an ÄKTAxpress unit with standard buffers (*table 2.2.*). Prior to each purification run, the gel filtration (GF) column was equilibrated overnight with GF buffer. 2 mL eluent fractions were collected in a 96 deep-well block in a serpentine fashion. Fractions containing the protein were analysed by SDS-PAGE, pooled and concentrated before use in crystallization, biophysical and kinetic studies.

Proteins were generally used immediately following purification. When proteins needed to be frozen, they were flash-frozen in vials, in liquid nitrogen at a concentration >10mg/ mL, and were stored at -80°C.

Purification Buffer	Composition
Gel filtration	200 mM NaCl, 20 mM Tris-HCl, pH 7.5
IMAC wash	500 mM NaCl, 50 mM Tris-HCl, 30 mM Imidazole, pH 7.5
IMAC elution	500 mM NaCl, 50 mM Tris-HCl, 500 mM Imidazole, pH 7.5
Strip buffer	20 mM $NaPO_4$, 0.5 M NaCl, 50 mM EDTA, pH 7.4

Table 2.2. Standard buffers used in protein purification and in preparation of the nickel column.
All buffers were filtered prior to use.

2.2.4.6. Optimisation of Purification For CES2

Using the protocol described above, recovery of all forms of CES2 and CES3 was incomplete, with protein still present in the flow through from the IMAC step. To increase the percentage of protein yield recovered from the supernatant, efforts were put into optimising the purification (*table 2.3*).

Troubleshooting Strategy	Details	Improvement on Purification?
4.7 mL HiScreen™ Ni FF column (GE Healthcare).	Same purification protocol, just substituted the standard nickel column.	×
Reduced imidazole in wash buffer.	i) The concentration of imidazole in the IMAC wash buffer was lowered from 30 mM to 10 mM. *ii)* Imidazole was omitted from the IMACwash buffer entirely. Imidazole prevents non-specific binding to the column, so by lowering the amount used, it was hoped that the His-tag on the protein would bind with a higher affinity.	×
5 mL HisTrap excel column (GE Healthcare).	Same purification protocol, just substituted the standard nickel column.	×
Increased loading time of supernatant.	Continually re-loaded the supernatant onto the nickel column overnight, to try and maximize binding of the secreted protein.	×
Dilution of supernatant.	Diluted supernatant 1:1 with PBST.	✓
Addition of nickel sulphate $NiSO_4(H_2O)_6$.	Added $NiSO_4(H_2O)_6$ to the supernatant, so the final concentration was 3-5 mM. Left to mix at room temperature for 2 hours, then loaded onto a standard nickel column.	✓✓✓

Table 2.3. Methods used for the optimization of purification of CES2.

2.2.4.7. Protein Concentration

As the proteins used in this project were approximately 60 kDa in size, they were all concentrated using an Amicon® Ultra 15 mL centrifugal filter unit with an Ultracel-30 membrane (30 kDa cut-off) (EMD Millipore) at 4°C and 3000 g for approximately 10

minutes, or until they reached an appropriate concentration for downstream processes. Prior to their use, filters were pre-rinsed using PBS.

Protein concentration was determined by measuring the UV absorbance at 280 nm from a 2 μL sample, using a NanoDrop 1000 (Thermo Scientific) spectrophotometer after blanking with 2 μL GF buffer. Extinction coefficients for all enzymes were calculated using the online server http://web.expasy.org/protparam/ after removal of their signal sequences. These were then used to calculate the correct concentration of protein using the Beer-Lambert Law, where;

$$\log_{10}\frac{I_0}{I}=\varepsilon cl$$

(*I* = Intensity, ε = extinction coefficient, c = concentration, *l* = Path length that light passes through (=1))

	CES1	CES2	CES3
Theoretical pI	6.06	5.33	5.40
Molecular Weight (kDa)	60.45	58.96	59.65
Extinction Coefficient (M^{-1} cm^{-1})	76110	67630	79785
Abs 0.1% (=1 g/ L)	1.26	1.15	1.34

Table 2.4. Biophysical parameters of CES1-3.
Extinction coefficients were calculated using ProtParam in the ExPASy server[108]. For simplicity, the concentration of all mutant forms of each protein were calculated using the wild type parameters. The pI is the isoelectric point, the pH at which the protein carries no net electrical charge.

2.3. Biophysical Characterisation

2.3.1. Kinetic Studies Using 4- Nitrophenyl Acetate (4-NPA) as a Substrate

The experimental design for activity analysis was based around documented carboxylesterase activity assays[73, 93, 109-111], but was adapted to include a wider range of both substrate and enzyme concentrations. CES1 native, CES1 null, CES1 N79Q, CES2 native, CES2 null, CES2 N111Q, CES2 N276Q, CES2 N111Q+N276Q, CES2 3-4, CES2 4-4 and CES3 were all assayed using 4-NPA (Sigma: N8130-5G, m.w.181.15 g) as the substrate[112]. Carboxylesterase activity was detected by the hydrolysis of 4-NPA into its corresponding alcohol by measuring the production of the nitrophenolate anion (*figure 2.5*) at 405 nm in the Paradigm Plate Reader (Beckman Coulter).

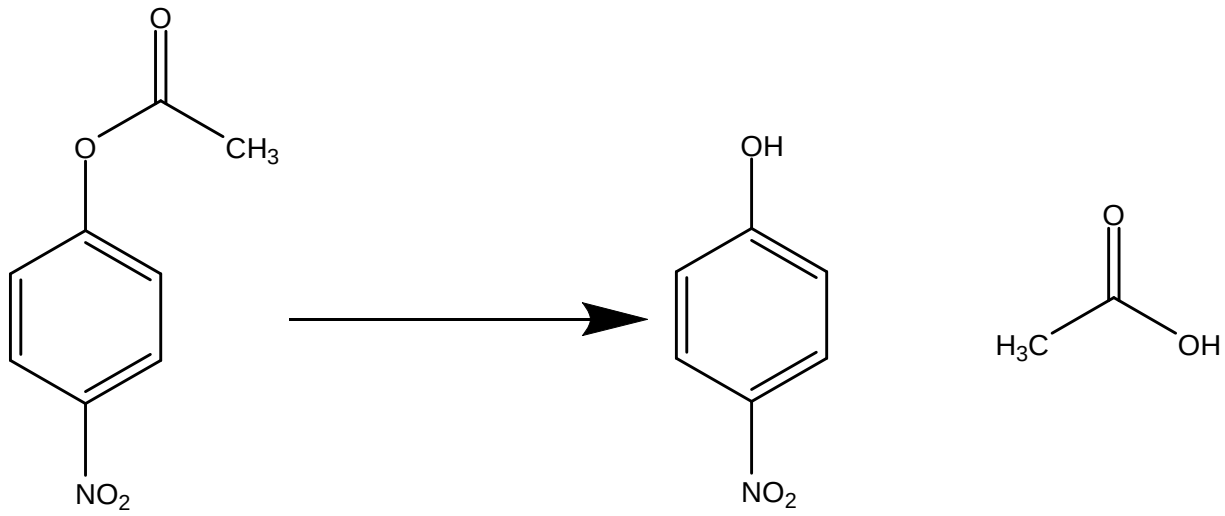


Figure 2.5. Breakdown of 4-NPA.
4-NPA is hydrolysed (water not shown) into 4-nitrophenol and acetate.

2.3.1.1. Preparation of Assay Plate

The final plate layout was developed to include every dilution of protein and substrate within one plate, with two enzymes run in a single experiment. Enzyme stocks were

prepared in 5 mL of assay buffer (50 mM Tris-HCl, pH 7.5) at a final concentration of 20 µg/ mL. A 50 mL stock of 0.1 M 4-NPA was prepared in ethanol.

50 µL of assay buffer was pipetted into all wells in rows B-H. 100 µL of enzyme 1 was pipetted into row A, columns 1-6, and 100 µL of enzyme 2 was pipetted into row A, columns 7-12. Serial dilutions using a multichannel pipette were carried out by transferring 50 µL from row A to row B, mixing 10-20 times, and so on down the plate until row G. The final 50 µL from row G was discarded, and row H contained assay buffer only as it was used as a blank.

The dilution and addition of 4-NPA needed to be executed rapidly since the reaction proceeded as soon as the enzyme was added. Therefore it was critical to get the plate into the plate reader as soon as possible to ensure that the initial rate of reaction was captured. 3 mL of 6 mM 4-NPA was made by diluting 180 µL of 0.1 M stock into 2.82 mL of assay buffer. Serial dilutions were made in 15 mL falcon tubes by diluting 1 mL of this into 1 mL of assay buffer to give concentrations of 3 mM, 1.5 mM, 0.75 mM, 0.375 mM, 0.1875 mM and 0.0938 mM. 50 µL of the highest concentration of 4-NPA (3 mM) was added to all wells in columns 1 and 7 using a repeat pipette. Then, 50 µL of 1.5 mM 4-NPA was added to columns 2 and 8, and so on. The layout of the reaction plate is shown in *table 2.5.* In the final plate set up, a serial dilution of 4-NPA had been established across the plate, and protein down the plate. The final reaction volume was 100 µL.

Plates were shaken and absorbance was read every 30 seconds at 37°C, on the kinetic setting 405 nM for 30 minutes. Two enzymes were run on the same plate, and experiments were repeated in triplicate.

[4-NPA] (mM)						[4-NPA] (mM)					
3	1.5	0.75	0.375	0.1875	0.0938	3	1.5	0.75	0.375	0.1875	0.0938
165.44	165.44	165.44	165.44	165.44	165.44	165.44	165.44	165.44	165.44	165.44	165.44
82.72	82.72	82.72	82.72	82.72	82.72	82.72	82.72	82.72	82.72	82.72	82.72
41.36	41.36	41.36	41.36	41.36	41.36	41.36	41.36	41.36	41.36	41.36	41.36
20.68	20.68	20.68	20.68	20.68	20.68	20.68	20.68	20.68	20.68	20.68	20.68
10.34	10.34	10.34	10.34	10.34	10.34	10.34	10.34	10.34	10.34	10.34	10.34
5.17	5.17	5.17	5.17	5.17	5.17	5.17	5.17	5.17	5.17	5.17	5.17
2.59	2.59	2.59	2.59	2.59	2.59	2.59	2.59	2.59	2.59	2.59	2.59
B	B	B	B	B	B	B	B	B	B	B	B
Enzyme 1						Enzyme 2					

Table 2.5. 96-well plate format for carboxylesterase activity analysis.
Two enzymes were run per plate, and all experiments were repeated in triplicate. B was a blank well, containing just assay buffer. Substrate concentration is shown on the top row. Amount of protein in µg is shown in each well.

2.3.1.2. Data Analysis

4-NPA was spontaneously hydrolysed in the assay buffer, therefore values determined for spontaneous 4-NPA hydrolysis, from the blank wells in each column, were subtracted from the data. The average of the triplicate experiments was taken, and any outlying values were omitted. For each concentration of 4-NPA, change in optical density (ΔOD) observed at each enzyme concentration over the 30 minute time course was plotted (ΔOD 405 nm vs time). From these, initial velocities were calculated from the linear part of the curve. All r^2 exceeded 0.99, indicating that the data fit the line well. Graphs of activity units (AU min^{-1}) vs substrate concentration (mM), Lineweaver-Burk and Eadie-Hoftsee plots were constructed to determine kinetic parameters.

For CES1, the allosteric sigmoidal substrate-velocity model (nonlinear regression) was fitted to the data and statistical analysis was performed using GraphPad Prism version 6.00 for Windows, GraphPad Software, La Jolla California USA, www.graphpad.com.

Equations

Allosteric Sigmoidal Model (to fit K_{half}):

$$Y = \frac{Vmax * X^h}{(Khalf^h + X^h)}$$

Allosteric Sigmoidal Model (to fit K'):

$$Y = \frac{Vmax * X^h}{(Kprime + X^h)}$$

Hill Plot:

$$\upsilon = \frac{V_1 A^n}{K_1 + A^n}$$

2.3.1.3. Specific Activity Calculation

Firstly, AU min^{-1} were converted to AU min^{-1} per µg. From that, the rate of product formation was expressed as nanomole of pNP produced per minute per microgram of protein (nmol min^{-1} μg^{-1}) using the extinction coefficient for pNP at 405 nm (18,000 M^{-1} cm), and adjusting for the reaction volume (100 µL). The path length of solution that the light passed through was not taken into account because it was automatically adjusted for by the plate reader during the experiment.

2.3.2. Analytical Ultracentrifugation

All AUC experiments were performed at the Research Complex at Harwell by Professor David Scott. Sedimentation velocity experiments were carried out on a Beckman XL-I analytical ultracentrifuge (Beckman-Coulter, CA, USA). Sedimentation velocity was performed at 20°C using 2 channel centerpieces, with protein loading concentrations of 2 mg/ mL, 1.0 mg/ mL and 0.50 mg/ mL. Data were obtained at 40 000 rpm, using a Beckman 50Ti rotor, with the cells scanned radially with interference optics and with absorbance optics at a wavelength of 280 nm. Scans were obtained every 10 minutes and data were analyzed using the program SEDFIT[113] (www.analyticalultracentrifugation.com). Sedimentation coefficient distributions were

obtained using the c(S) methodology[114], and figures were created in GUSSI 1.0.3. Solution densities and viscosities were measured directly using an Anton Paar DMA5000 densitometer/viscometer.

2.3.3. Multi-Angle Light Scattering

All size exclusion chromatography coupled with multi-angle laser light scattering (SEC-MALLS) experiments were performed at the Research Complex at Harwell with the assistance of Professor David Scott, who also processed the data.

500 µL of protein sample was made up in GF buffer at a final concentration of 5 mg/ mL. SEC-MALLS data was analysed on a Wyatt Dawn 17 angle light scatterer, connected to an ÅKTA Purifier running a Sephadex 100 column (300 mm x 10 mm). Flow rate was set at 0.3 ml^{-1}. Data were analysed in the ASTRA v6.0 software using the Debye-Zimm method: a refractive index increment of 0.185 was used.

2.4. Crystallization, X-ray Data Collection, and Structure Determination

2.4.1. Screening

Following purification, initial Pre-Crystallization Tests (PCT™) (Hampton Research) were carried out on all proteins to determine appropriate concentrations for crystallization screening, as concentration is a significant crystallization variable (*table 2.6*). Crystallization trials were then set up using 200 nL drops (100 nL protein plus 100 nL reservoir solution) in 96 well format[115] by sitting drop vapour diffusion (Greiner plates) with standard commercially available crystallization screening kits (*table 2.7*), at both 4°C and 20°C. 100 µL of crystal screen was transferred from a storage block (all stored at 4°C) to the Greiner plate using a Hydra liquid handler. Protein and reservoir drops were dispensed using a Honeybee X8 pipetting instrument (Digilab Inc, MA, USA). The plates were tightly sealed with View seals (Greiner), placed into a Rock Imager (Formulatrix) and imaged at regular time intervals. Images were accessed via https://www.oppf.rc-harwell.ac.uk/xtalpims/, which allowed for many different conditions to be visually screened and annotated rapidly. To aid in the correct detection of potential protein crystal hits as opposed to salt crystals, there was also an option to illuminate individual crystal drops with UV light and use the fluorescence generated from aromatic amino acids (primarily tryptophan) to create an image.

Protein	Initial Protein Concentration (mg/ mL)
CES1 wt	5.0
CES1 null	5.0
CES1 N79Q	6.0
CES2 wt	8.0
CES2 null	9.5
CES2 Phyre	11.0
CES2 Dere	9.0
CES2 N111Q	6.0
CES2 N276Q	6.5
CES2 N111Q+N276Q	6.5
CES3 3-4	11.0
CES2 4-4	11.5
CES3 wt	12.0

Table 2.6. Summary of protein concentrations used in initial crystallization screens, based on results from the Pre-Crystallization Tests.

Screen	Manufacturer	Features
Morpheus™	Molecular Dimensions	3D protein crystallization screen. All conditions are cryo-protected.
Index HT	Hampton Research	Sparse matrix and grid screen and incomplete factorial.
JCSG-*plus*™	Molecular Dimensions	Sparse matrix screen with reduced redundancy.
PACT *premier*™	Molecular Dimensions	Tests the effect of pH, anions and cations, using PEG as the precipitant.
Wizard I & II	Emerald BioSystems	Sparse matrix
Wizard III & IV	Emerald BioSystems	Sparse matrix

JBScreen Pentaerythritol	Jena Bioscience	Based on pentaerythritol polymers as precipitants, which also function as cryo-protectants.
PEG/ Ion HT™	Hampton Research	Polymer, salt and pH matrix screen.
Crystal Screen™	Hampton Research	Sparse matrix screen.

Table 2.7. Commercially available crystallization screens used in the OPPF[116-118].
These screening blocks were used during the course of this project. When protein was limited and a comprehensive screen could not be conducted, the top four screens were used first.

Depending on the propensity of proteins to crystallize, and the quality of the crystals obtained, different optimization techniques were employed for each protein. These are summarized in the following paragraphs.

2.4.2. CES1

2.4.2.1. Crystallization

All forms of CES1 crystallized readily in many different conditions, after approximately 4 hours to 3 days, of which all were tested for diffraction. The structures presented in this thesis came from the best-diffracting crystals that were harvested from the following conditions; **CES1 wild type**- grown at 20°C in Morpheus E3 (0.03 M each of diethylene glycol, triethylene glycol, tetraethylene glycol, pentaethylene glycol, 0.1 M Morpheus Buffer 1 at pH 6.5 [0.1 M MES/imidazole pH 6.5], 10% w/v polyethylene glycol 4000, 20% w/v glycerol), **CES1 S221A** grown at 4°C in Morpheus E11 (0.03 M each of diethylene glycol, triethylene glycol, tetraethylene glycol, pentaethylene glycol, 0.1 M Morpheus Buffer 3 pH 8.5 [0.1 M Bicine/Trizma base], 10% w/v PEG 4000, 20% v/v glycerol)**, CES1 N79Q** grown at 4°C in PACT G3 (0.1 M Bis-Tris Propane, 20% w/v polyethylene glycol 3350, 0.2 M Sodium Iodide pH 7.5). For the crystals that were harvested from the Morpheus® screen[116], no cryo-protection was needed. For the N79Q crystal grown in the PACT[117] condition, 25% glycerol was used directly as a cryo-

protectant. CES1 crystals were harvested directly from the cryo-protected Morpheus condition and immediately cryo-cooled in liquid nitrogen before data collection.

2.4.2.2. Data Collection and Refinement

Diffraction data were collected at the Diamond Light Source (DLS), on beamlines I03 and I04. The *apo* structure of CES1 was solved by the molecular replacement method with PHASER[119] using the structure of CES1 in complex with Coenzyme A determined to 2.0 Å resolution[69] (PDB code: 2h7c) as a search model. Phases calculated from this initial model were used for manual completion of the structure using COOT[120] followed by iterative cycles of refinement with REFMAC[121]. This high resolution structure of CES1 was then used as the search model for molecular replacement of the other two structures presented here. In the case of CES1 null, diffraction data extended to beyond 1.5 Å resolution, allowing for anisotropic refinement of the atomic displacement parameters. MolProbity4 was used for structure validation.

2.4.3. CES2

2.4.3.1. Post Protein Production Modifications and Strategies Utilised to Obtain Crystals

As CES2 wt did not crystallize, a number of approaches to achieve this were employed, and are summarized in the following paragraphs. At this stage in the project the mutants that were undergoing the protein modification and rescue strategies were CES2 wt, null, Phyre, Dere, N111Q, N276Q and N111Q+N276Q (see *table 2.1* for details). CES2 3-4 and CES2 3-4 were made at a later point in the project.

A. Varying Concentration and Temperature

All proteins were put into crystallization screening at their original PCT determined concentration (*table 2.6.*), and at 2, 5, 15 and 30 mg/ mL. As protein crystallization requires reaching a super saturated solution, the initial protein concentration must be high enough to reach this state during the experiment without forming aggregate. An observation of ~50% clear drops in crystallization trials indicated that the protein was in the correct concentration window. All crystallization plates were duplicated; one was stored at 20°C and the other at 4°C.

B. De-Glycosylation

The glycans were cleaved enzymatically using endoglycosidase H (endo h) (NEB). This enzyme cleaves between the two *N*-acetylglucosamine residues in the core of the oligosaccharide, generating a truncated sugar molecule with one *N*-acetylglucosamine residue remaining on the asparagine. However, glycosylated forms of CES2 crashed out of solution after this process. Therefore, it was decided to produce aglycosylated proteins by mutagenesis of the acceptor sites in the polypepetide.

C. Co-Crystallization

Co-crystallization experiments were set up with CPT-11 (Tocris Bioscience) and CHR-7150, an inhibitor of CES2 that was identified from studies at Chroma Therapeutics Ltd, Milton Park,Abingdon, Oxon,OX14 4RZ (http://www.chromatherapeutics.com). A stock solution of CPT-11 was made up to 15 mM in dH_2O and was gently warmed in a 42°C water bath to ensure all the compound had gone into solution. 20 µL aliquots were stored at -20°C and were only freeze-thawed once. CHR-7150 was made up to 50 mM in dimethyl sulphoxide (DMSO). A 3-5 molar excess of each compound was added to 200 µL of protein and incubated on ice for 1 hour prior to setting up of the crystallization screens.

D. Reductive Methylation of Lysines

The standard protocol for reductive methylation using formaldehyde from Hampton Research was used (HR2-434), according to the manufacturer's instructions. The protein was buffer exchanged into HEPES buffer, as the protocol stated to avoid Tris/ HCl buffers. Mass spectrometry could not be used to verify methylation, as the protein did not ionise effectively in the instrument.

E. In Situ Proteolysis

A JBS Floppy-Choppy kit (Jena Bioscience) was used, and contained the proteases Trypsin, Subtilisin, Papain and α-chymotrpsin. The manufacturer's protocol was followed. By adding a cocktail of proteases directly into the crystallization drops, it was hoped to digest any flexible regions of the protein, and increase the success rate of crystallization.

F. Seeding

The Seed Bead kit (Hampton Research) was used to generate crystal seeds for microseeding experiments. Crystals were added to a 1.5 mL Eppendorf that contained a seed bead and 80 µL GF buffer. This was vortexed for 1 minute, creating a seed stock that was used to set up crystallization drops automatically 300 nL (100 nL protein plus 100 nL reservoir solution plus 100 nL seed stock). The concentration of proteins used in seeding experiments, was increased by roughly a third from their original crystallization concentration, to compensate for the dilution with the seed bead solution.

This technique was used to build on an initial hit for a single crystal of CES2 wt as well as for for cross-seeding experiments where seed stocks of CES1 crystals were generated, and added to CES2 crystallization experiments.

G. Additive Screen

The additive screen (Hampton Research) is a library of 96 unique reagents that can affect the solubility and crystallization of proteins, and is used when an initial crystal screen hit condition has been identified. Within the 96-well plate, all reservoirs contained the hit condition. The additives (and protein at its identified optimal crystallization concentration) were added directly into the crystallization drop using the Honeybee X8 pipetting instrument. This technique was used around the original CES2 hit and also later on when optimizing CES2 3-4 + CPT-11.

2.4.3.2. Crystallization of CES2 3-4 and 4-4

Crystallization trials for CES2 3-4 were initially carried out at both 4°C and 20°C, as well as co-crystallized with a 3 molar excess of CPT-11 (Tocris Bioscience) (details are described in co-crystallization above). Crystals in several conditions started to appear from day 5 onwards. After these initial hits were verified as protein crystals by diffraction screening on a beam line, optimization was carried out around the condition that gave the best diffracting crystals, PACT G4 condition (0.2 M Potassium Thiocynate, 0.1 M Bis-Tris Propane, 20% PEG 3350 pH 7.5)[117].

A. 3-Row Optimisation

Initial optimization of primary crystal hits was carried out using the 3-row protocol previously described[122], in which the pH of the crystallization condition is titrated by the addition of acid and base across the last three rows of a 96-well crystallization plate, whilst the ratio of precipitant to protein is varied between the rows (*see figure 2.6*). The screen was prepared manually.

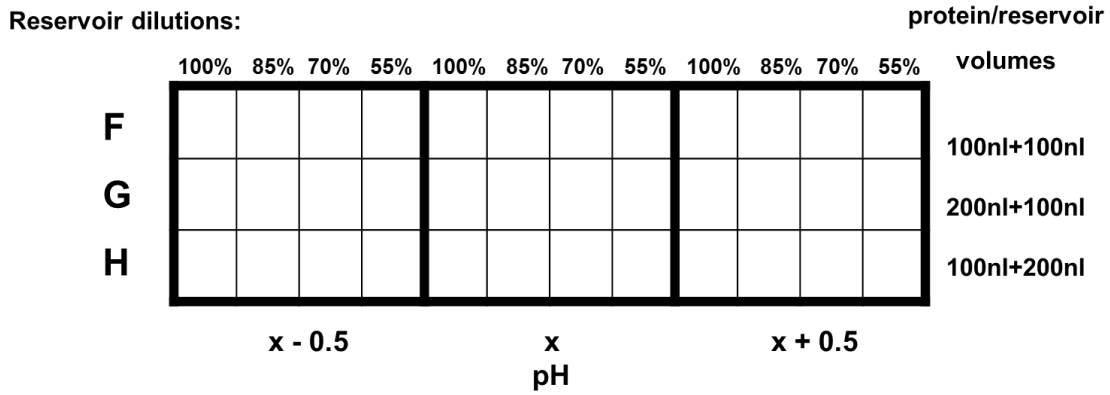


Figure 2.6. Layout of the 3-row optimisation experiment.

B. 96-Well Optimisation

Optimisation was performed around PACT *premier* F4/G4 **(**0.2 M Potassium Thiocyanate, 0.1 M Bis-Tris Propane, 20% PEG3350, pH 6.5/7.5). Based on these conditions, a 96-well optimisation grid was designed and prepared manually, varying pH and the concentrations of both PEG 3350 and potassium thiocyanate (*figure 2.7*). This was done to obtain better crystals, by potentially increasing nucleation and/ or crystal size.

	% PEG 3350											
pH	**14**	**16**	**18**	**20**	**22**	**24**	**14**	**16**	**18**	**20**	**22**	**24**
6.5	0.05 M						0.1 M					
6.5	0.15 M						0.2 M					
6.5	0.25 M						0.3 M					
6.5	0.35 M						0.4 M					
7.5	0.05 M						0.1 M					
7.5	0.15 M						0.2 M					
7.5	0.25 M						0.3 M					
7.5	0.35 M						0.4 M					

Figure 2.7. Layout of a 96-deep well plate, formulated by hand, for the optimisation of CES2 3-4.
Concentration of potassium thiocyanate is shown in the wells. Variation in concentration of PEG 3350 is shown at the top. p.H is indicated on the left hand side.

C. Hanging Drops

These experiments were performed in an EasyXtal 15-Well Tool plate (Qiagen) with greaseless screw-in DropGuard X-Seal crystallization supports (Qiagen) to try and increase the size of the crystals. 500 µL of solution was added to each well. The

following ratios of protein solution: reservoir solution were set up on each support: 1 µL: 1 µL, 0.5 µL: 1 µL and 1 µL: 0.5 µL. The supports were then sealed tightly, without disrupting the crystallization droplets.

D. CES2 3-4 Cryo-Protection

Two different cryo-protectants were used when freezing the crystals, to test their effect on the crystals, and to see if diffraction could be improved. The cryoprotectants were added to crystallization solutions based around the original screen hit, as follows:

- Ethylene glycol (25% ethylene glycol, 24% PEG 3350, 0.2 M PTCN, 0. 1 M Bis-Tris Propane pH 7.5),
- Glycerol (20% glycerol, 18% PEG 3350, 0.2 M PTCN, 0.1 M Bis-Tris Propane pH 7.5).

From analysis of preliminary diffraction patterns and the crystal characterisation, it was evident that CES2 crystallized with a relatively high solvent content. Therefore, rather than directly cryo-protecting the crystals in the original condition with 20% glycerol, 4 solutions (containing 5, 10, 15 or 20% glycerol) of the condition were made, and the crystals were serially cryo-protected. 2 µL of each solution was pipetted separately onto a cover slip. The crystals was harvested from the hanging drop condition using a 0.3 µm loop (Hampton Research) and placed in each drop for approximately 30 seconds, from the lowest to the highest concentration of glycerol. They were then immediately cryo-cooled in liquid nitrogen by directly plunging the sample pins with the harvested crystals into a sample puck ready for automounting at the beamline.

E. Data Collection

Diffraction data were collected at Diamond Light Source (DLS), on beamline I04 using remote access. Robotic handling and automation of crystals meant that two pucks (24 crystals) could be screened rapidly in 2 hours. The best diffracting crystal of CES2 3-4 + CPT-11 was grown at 14 mg/ mL in 18% PEG 3350, 0.2 M PTCN, 0.1 M Bis-Tris Propane pH 7.5. To mitigate radiation damage and to enable the collection of a complete high resolution data set from a single crystal, the data were collected via a series of 20° wedges along the length of the crystal, commonly referred to as a line scan. Data were integrated and scaled using xia2[123].

2.4.3.3. Structure Determination and Refinement

The structure of CES2 was solved by molecular replacement using PHASER[119] and the high resolution structure of CES1 null determined to 1.48 Å resolution (obtained in this project) was used as a search model. Phases calculated from this initial model were used for manual completion of the structure using COOT[120] with iterative cycles of refinement with REFMAC[121]. The strand extension and exchange region of the protein had to be built manually. 4-PP molecules were manually docked and fitted into well defined difference positive density within the active site. MolProbity4 was used for structure validation - http://molprobity.biochem.duke.edu/

2.4.4. Presentation of Crystallographic Data

The MacPyMOL Molecular Graphics System, Version 1.7rc1 was used to generate all images within this thesis. Tools within the CCP4 crystallography suite of programs for software were also used for data analysis[124].

3. PRODUCTION, CHARACTERIZATION AND CRYSTALLIZATION OF CES1

3.1. Introduction

CES1 was the first human carboxylesterase to be identified and characterized, and is by far the most extensively studied mammalian carboxylesterase to date. However, the nature of its relatively broad substrate specificity, the role of its putative glycosylation site and its allosteric kinetic behaviour[109] remain to be fully addressed and are the focus of this chapter.

Recombinant CES1 proteins for functional and structural studies have been produced using the *Spodotera frugiperda* (Sf21)[125] insect expression system, infecting the cells with a baculovirus and purifying the enzyme from the supernatant[61, 85, 93, 126-128]. The yields of purified enzymes have been reported as 7.5-12.5 mg/ L cell culture[129]. As an alternative to production in insect cells in culture, Greenblatt et al., reported using *Trichoplusia ni* (whole cabbage looper) larvae with a yield of 9 mg protein from ~ 1 kg of infected caterpillar larvae[130]. In this project, the use of mammalian cells for producing CESs was explored in order to improve protein yields. Another advantage of expressing mammalian glycoproteins in mammalian cells is that they will be authentically glycosylated[131].

In 2003, the first crystal structures of CES1 in complex with narcotic analogues (1MX5[85] and 1MX9[84]) were deposited at resolutions of 2.8 and 2.9 Å respectively (*table 3.1*). Over the last 10 years, many more structures of CES1 have been determined with the highest resolution structure reported at 2.0 Å (2H7C[69]). Obtaining a higher resolution, crystallographic structure of CES1 for downstream protein ligand binding studies was a key objective in this project. In addition, since no structure of CES1 in the absence of substrates or inhibitors had been published, a second objective was to solve the structure of the *apo* enzyme. As described in this chapter both these objectives were successfully

achieved and the *apo* enzyme solved to 1.86 Å. While this work was in progress, the structure of an *apo* form of CES1 to 2.20 Å from insect cells was reported[130]. This chapter also describes new data concerning the apparent redundancy of glycosylation for both the structure and activity of CES1, and the observation of sigmoidal non-Michaelis Menten kinetics in both wild type CES1 and a N79Q aglycosylation mutant.

Table 3.1. Details of published structures of CES1 to date.
The first three entries are structures of CES1 presented in this thesis. All other deposited structures are listed in order of resolution, from highest to lowest. Data was extracted from the PDB (http://www.rcsb.org/pdb/home/home.do).

ID	Complexed With	Date	Res. (Å)	Crystallization Conditions	Source Organism	Space Group	*R*-Value (obs) %	*R*-Free %	Molecules in unit cell	Structure Factors Deposited
	Null	October 2013	1.48	0.03 M each of diethylene glycol, triethylene glycol, tetraethylene glycol, pentaethylene glycol, 0.1 M Morpheus Buffer 3 pH 8.5 (0.1 M bicine/Trizma base), 10% w/v PEG 4000, 20% v/v glycerol), at 4°C	HEK 293 Gnt$^{-/-}$	*H* 3	12.96	17.19	1	
	Wild type	October 2011	1.86	0.03 M each of diethylene glycol, triethylene glycol, tetraethylene glycol, pentaethylene glycol, 0.1 M Morpheus Buffer 1 at pH 6.5 (0.1 M MES/imidazole pH 6.5), 10.0% w/v polyethylene glycol 4000, 20.0% w/v glycerol, at 20°C	HEK 293 Gnt$^{-/-}$	*H* 3	15.83	18.27	1	
	De-Glycosylated	March 2014	2.01	0.1 M Bis-Tris Propane, 20% w/v polyethylene glycol 3350, 0.2 M Sodium Iodide pH 7.5, at 4°C	HEK 293T	*H* 3	16.67	21.17	1	
2H7C	Coenzyme A	August 2006	2.00	8% PEG 3350, 0.4M Li2SO4, 0.1M NaCl, 0.1M LiCl, 0.1M citrate (pH 5.5), 5% glycerol, pH 5.6, VAPOR DIFFUSION, SITTING DROP, temperature 298K	*Spodoptera frugiperda*	*P* 1 2_1 1	18.3	22.1	6	No
4AB1	*apo*	March 2012	2.20	DROP WAS 45% PROTEIN SOLUTION (11mg/mL/) OF 50 MM TRIS, PH7.6, 150 MM NACL, MIXED WITH 45% 2 M AMMONIUM SULFATE, 0.1 M BIS-TRIS, PH 5.5, AND 10% 2 M NASCN. pH 5.5	*Trichoplusia ni*	*P* 6_3 2 2	18.1	23.6	1	Yes
1MX1	Tacrine	April 2003	2.40	PEG3350, glycerol, lithium sulfate, sodium chloride, lithium chloride, sodium citrate, pH 5.5, VAPOR DIFFUSION, SITTING DROP, temperature 298.0K	*Spodoptera frugiperda*	*P* 1 2_1 1	16.2	20.7	6	No
2HRQ	Soman	May 2007	2.70	0.1M Citrate pH 5.5, 8-11% PEG 3350, 0.1-0.4M Lithium Sulfate, 0.1M Lithium Chloride, 0.1M Sodium Chloride, 5% Glycerol, VAPOR DIFFUSION, SITTING DROP, temperature 295K	*Spodoptera frugiperda*	*P* 1 2_1 1	17.0	22.5	6	Yes

2HRR	Tabun	May 2007	2.70	0.1M Citrate pH 5.5, 8-11% PEG 3350, 0.1-0.4M Lithium Sulfate, 0.1M Lithium Chloride, 0.1M Sodium Chloride, 5% Glycerol, VAPOR DIFFUSION, SITTING DROP, temperature 295K	*Spodoptera frugiperda*	*P* 2_1 2_1 2_1	16.7	23.2	3	Yes
2DQZ	Homatropine, coenzyme A, and palmitate	August 2006	2.80	8% PEG 3350, 0.4M Li2SO4, 0.1M NaCl, 0.1M LiCl, 0.1M citrate (pH 5.5), 5% glycerol, pH 5.6, VAPOR DIFFUSION, SITTING DROP, temperature 298K	*Spodoptera frugiperda*	*P* 2_1 2_1 2_1	19.3	24.4	3	No
1MX5	Homatropine	April 2003	2.80	PEG3350, Glycerol, lithium sulfate, lithium chloride, sodium chloride, sodium citrate, pH 5.5, VAPOR DIFFUSION, SITTING DROP, temperature 298K pH 5.5	*Spodoptera frugiperda*	*P* 1 2_1 1	15.8	22.1	6	No
1YA8	Cleavage products of Mevastatin	August 2005	3.00	PEG3350, Sodium Chloride, Lithium Chloride, Lithium Sulfate, Citrate, Glycerol, pH 5.5, VAPOR DIFFUSION, SITTING DROP, temperature 295K	*Spodoptera frugiperda*	*P* 2_1 2_1 2_1	18.7	24.7	3	Yes
1YAH	Ethyl acetate	August 2005	3.00	PEG3350, Sodium Chloride, Lithium Chloride, Lithium Sulfate, Citrate, Glycerol, pH 5.5, VAPOR DIFFUSION, SITTING DROP, temperature 295K	*Spodoptera frugiperda*	*P* 2_1 2_1 2_1	19.0	22.9	3	Yes
2DQY	Cholate and Palmitate	August 2006	3.00	8% PEG 3350, 0.4M Li2SO4, 0.1M NaCl, 0.1M LiCl, 0.1M citrate (pH 5.5), 5% glycerol, pH 5.6, VAPOR DIFFUSION, SITTING DROP, temperature 298K	*Spodoptera frugiperda*	*P* 2_1 2_1 2_1	22.6	27.1	3	No
3K9B	Cyclosarin	January 2010	3.10	10% PEG 3350, 0.3 M Li2SO4, 0.1 M Citrate pH 5.5, 0.1 M NaCl, 0.1 M LiCl, 5% Glycerol, VAPOR DIFFUSION, SITTING DROP, temperature 295K	*Spodoptera frugiperda*	*P* 2_1 2_1 2_1	27.0	29.9	3	Yes
2DR0	Taurocholate	August 2006	3.20	8% PEG 3350, 0.4M Li2SO4, 0.1M NaCl, 0.1M LiCl, 0.1M citrate (pH 5.5), 5% glycerol , pH 5.6, VAPOR DIFFUSION, SITTING DROP, temperature 298K	*Spodoptera frugiperda*	*P* 2_1 2_1 2_1	21.9	25.5	3	No
1YA4	Tamoxifen	August 2005	3.20	PEG3350, Sodium Chloride, Lithium Chloride, Lithium Sulfate, Citrate, Glycerol, pH 5.5, VAPOR DIFFUSION, SITTING DROP, temperature 295K	*Spodoptera frugiperda*	*P* 2_1 2_1 2_1	20.3	25.2	3	Yes
1YAJ	Benzil	August 2005	3.20	PEG3350, Sodium Chloride, Lithium Chloride, Lithium Sulfate, Citrate, Glycerol, pH 5.5,	*Spodoptera frugiperda*	*P* 1	20.7	28.7	12	Yes

				VAPOR DIFFUSION, SITTING DROP, temperature 295K						

3.2. Protein Production

3.2.1. Small Scale Expression

Protein production was initially tested by small-scale transient expression in HEK293T cells and the expression levels of different constructs and mutants were compared. As part of the expression vector design, the *C*-terminal endoplasmic retention motif, HIEL-COOH, was deleted from all CES1 sequences. This ensured that proteins were secreted into the cell culture media and not retained in the luminal site of the ER where they usually reside. Expressed proteins were detected in both cells and media by western blotting using an anti-His6 antibody. The effect of replacing the CES1 native signal sequence (pOPINE vector) with a μ-phosphatase leader sequence (pOPING vector) was investigated. No marked differences in expression from the two vectors was observed (*figure 3.1*). Addition of the CD4 tag (pOPINTTGneo CD4) reduced CES1 expression, whilst addition of the halo tag knocked out all protein expression (not shown in image).

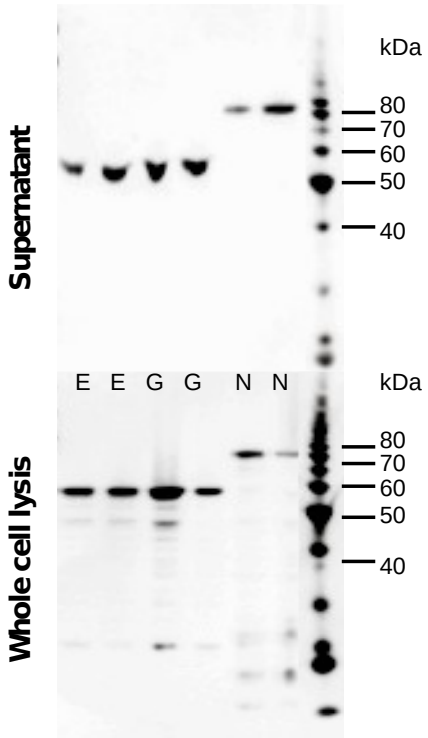


Figure 3.1. Western blot analysis of protein expression from three different constructs of wild type CES1.
Small- scale transient expression in HEK293T. DNAs were transfected in duplicate and protein expression analysed by SDS-PAGE/Western blotting using an anti-His6 monoclonal antibody. E = pOPINE, G = pOPING, N = pOPINTTG neo CD4. CES1 runs at ~60kDa; the CD4 tag adds ~30kDa to the size of the construct.

3.2.2. Large Scale Production

3.2.2.1. Stable Expression

To produce sustainable levels of protein for downstream processes, stable HEK293 gnt$^{-/-}$ cell lines were produced for CES1 wild type (wt) and null (S221A) enzymes using co-selection with a neomycin resistance gene (*neo*) incorporated into the vector. For CES1 wt, 38 G418-resistant colonies expressing protein were picked, and for CES1 null, there were only 3 that were recovered. Those giving the highest signals were frozen down and stored at -80 °C. Clonal cell lines were expanded in roller bottles and secreted proteins were purified from the cell media.

3.2.2.2. Transient Expression

The CES1 N79Q mutant was produced through site-directed mutagenesis via the overlap PCR method (*see Mutagenesis in Chapter 2*) to investigate the need for glycosylation in terms of protein production, activity and structure. Site-directed mutagenesis is preferential over treating the protein enzymatically to remove the glycans as it produces a single species of pure homogeneous un-glycosylated protein. This mutant was produced by large scale transient expression only.

3.2.3. Purification

Recombinant CES1 enzymes were recovered from the media of transfected HEK293T cells by IMAC and further purified by size exclusion chromatography. All three CES1 enzymes (wild type, null and N79Q) eluted from the size exclusion chromatography stage of purification around ~ 180 kDa corresponding to a trimeric form. Smaller peaks corresponding to putative hexamers (~ 360 kDa) and monomeric species (~ 60 kDa) were also observed (*figure 3.2*). These molecular weights were deduced from standard

calibration curves using proteins of known mass (ferritin, aldolase, conalbumin and ovalbumin).

Eluted fractions containing the protein were analysed by SDS-PAGE, pooled and concentrated before use in subsequent studies. All three proteins were highly soluble in aqueous buffer and could readily be concentrated to 25 mg/ mL. Although these proteins were flash-frozen and stored at -80°C for long-term storage, they were stable at 4°C for periods of up to a week, with no degradation or precipitation observed.

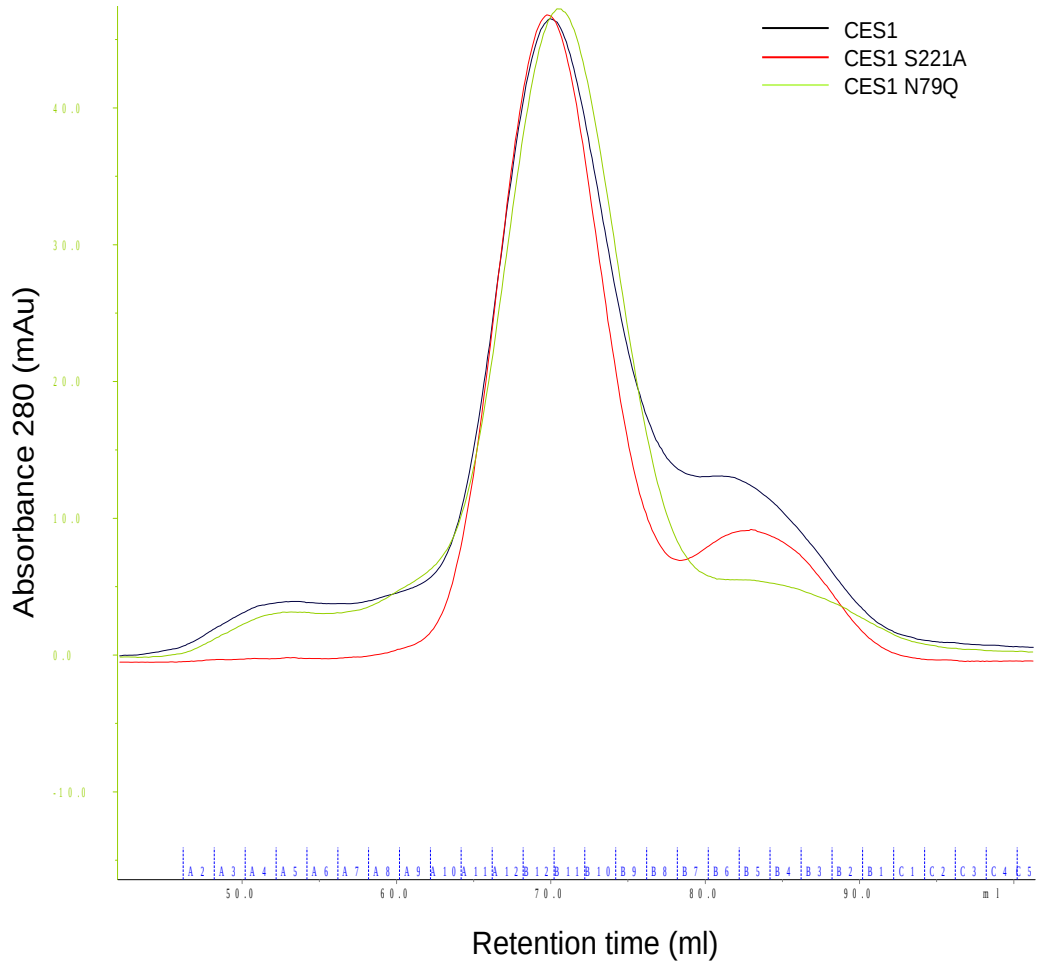
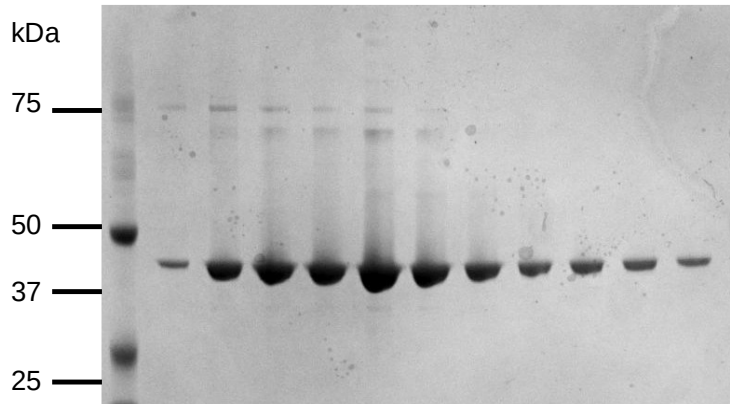
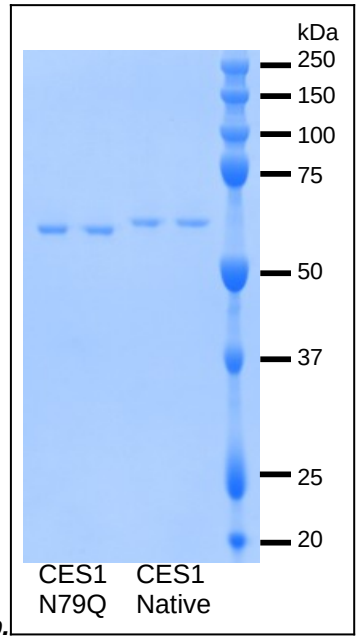


Figure 3.2. Overlay of UV trace of the size exclusion elution profiles of the three forms of CES1.
All three proteins were eluted from a 16/60 Superdex 200 column (GE Healthcare) on the same ÄKTAxpress system. All the CES1 enzymes behaved as trimers, with some association to hexamers, and dissociation to monomers.



a.



b.

Figure 3.3. Purified protein run on an SDS-PAGE gel, under de-naturing conditions.
a. 11 size exclusion fractions of CES1 wt, loaded immediately after GF elution. A10-B8 correspond to the trimeric peak, B7-B5 correspond to the monomeric peak.
b. 0.2 µgs of protein was loaded onto the gel, and then stained with Instant Blue™ after electrophoresis. CES1 N79Q ran at ~58 kDa (1st two lanes), and CES1 native (2nd two lanes) ran at ~60 kDa. It can clearly be seen that the un-glycosylated mutant runs at a lower molecular weight compared to the wild type form, indicating the absence of any glycosylation at N79.

Protein	Transient Expression	Yield (mg/ L)	Stable Cell Expression	Yield (mg/ L)
CES1 wt	-	-	Yes	64.2
CES1 null	-	-	Yes	68.0
CES1 N79Q	Yes	38.6	-	-

Table 3.2. Comparison of protein yields obtained from transient expression of CES1 in HEK293T and stable expression in HEK293 Gnt^{-}/$^{-}$.
CES1 native and null were only produced using stable cell lines whereas CES1 N79Q was produced transiently.

The yields of the recombinant CES1 enzymes are shown in Table 3.2, and analysis of samples by SDS-PAGE in Figure 3.3. Protein yields were almost five to ten-fold higher than previously reported for production from insect cells, showing the benefits of using mammalian cells. In particular clonal selection yielded highly productive cell lines. The CES1 N79Q mutant was produced transiently in HEK293T cells in the presence of kifunensine which gives a comparable high mannose containing *N*-glycoform to the HEK293Gnt$^{-/-}$ cells used for the stable CES1 native and null mutant cell lines.

The results also show that removal of the N79 glycosylation site had no significant effect on protein expression and secretion, as yields obtained from transient expression of CES1 N79Q were comparable to both glycosylated CES1 native and null mutant enzymes (38.6 mg/ L, 64.2 mg/ L and 68.0 mg/ L respectively). This is consistent with an earlier observation that treatment of insect cells with tunicamycin, which blocks *N*-glycosylation does not significantly affect expression of CES1 produced intracellularly using the baculovirus system[93]. Tunicamycin inhibits the enzyme GlcNAc phosphotransferase (GPT), which catalyzes the transfer of *N*-acetylglucosamine-1-phosphate from UDP-N-acetylglucosamine to dolichol phosphate in the first step of glycoprotein synthesis.

3.3. Measurement of Enzyme Activity

3.3.1. Observations of CES1 Kinetics

The single serine knockout mutant demonstrated no catalytic activity as expected, validating the importance of the presence of the conserved catalytic serine within the active site. CES1 native and N79Q hydrolysed the substrate 4-NPA demonstrating their functional enzymatic activity. 41.36 nM (obtained from serial dilutions of the protein) was the highest concentration of enzyme that could be used for analysis, as the reactions for double this concentration proceeded too quickly to obtain accurate values on the plate reader.

Using the initial velocity rates in activity units per minute (AU min^{-1}) (*table 3.3*), the plots of substrate vs velocity (*v*) for both CES1 native and N79Q were sigmoidal, and not classically hyperbolic. Double reciprocal Lineweaver-Burk plots, which should represent $K_a/V_{1,}$ gave non-linear concave slopes (*figure 3.4*). These results demonstrate that CES1 does not follow classical Michaelis Menten kinetics, where *v*, is equal to $VA/(K_a + A)$[132], but instead show positive cooperativity. This contradicts the majority of published CES1 kinetic analysis papers, which report Michaelis Menten kinetics[38, 111, 133] for the enzyme, but are consistent with the results of Takahashi et al. who observed non Michaelis Menten Kinetics for CES1[109]. However, these findings were based on pooled human liver microsomes (HLMs) as opposed to pure recombinant proteins.

(a)

	Enzyme Concentration (nM)				
4-NPA (mM)	**41.36**	**20.68**	**10.34**	**5.17**	**2.59**
0.0938	0.0281	0.0145	0.0081	0.0048	0.0018
0.1875	0.0584	0.0317	0.0158	0.0086	0.0044
0.375	0.107	0.0572	0.0318	0.0178	0.0094
0.75	0.2435	0.1356	0.0787	0.0418	0.0216
1.5	0.4212	0.2107	0.1158	0.0601	0.0306
3	0.5549	0.2657	0.1343	0.0669	0.0332

(b)

	Enzyme Concentration (nM)				
4-NPA (mM)	**41.36**	**20.68**	**10.34**	**5.17**	**2.59**
0.0938	0.0505	0.0248	0.0132	0.0071	0.0036
0.1875	0.0694	0.0421	0.0218	0.0112	0.0059
0.375	0.1166	0.0703	0.0394	0.0208	0.0102
0.75	0.2882	0.1782	0.1062	0.0519	0.0261
1.5	0.438	0.2607	0.1558	0.0728	0.0389
3	0.584	0.3084	0.1751	0.0761	0.0392

Table 3.3. Initial velocity rates in absorbance units per minute (AU min^{-1}) for CES1 native (*a*) and CES1 N79Q (*b*).
All r^2 exceeded 0.99. See *2.3.1.2* for details.

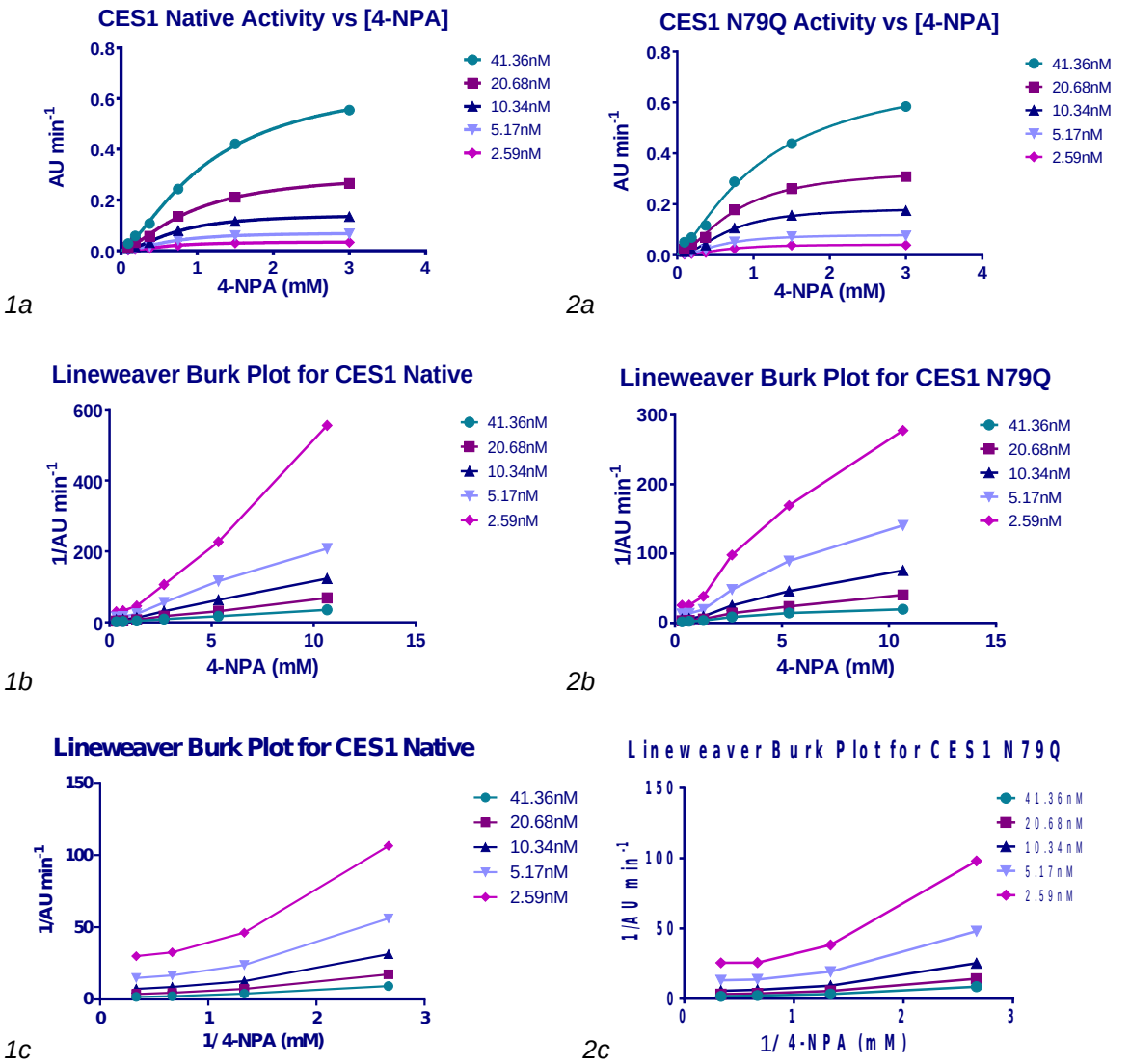
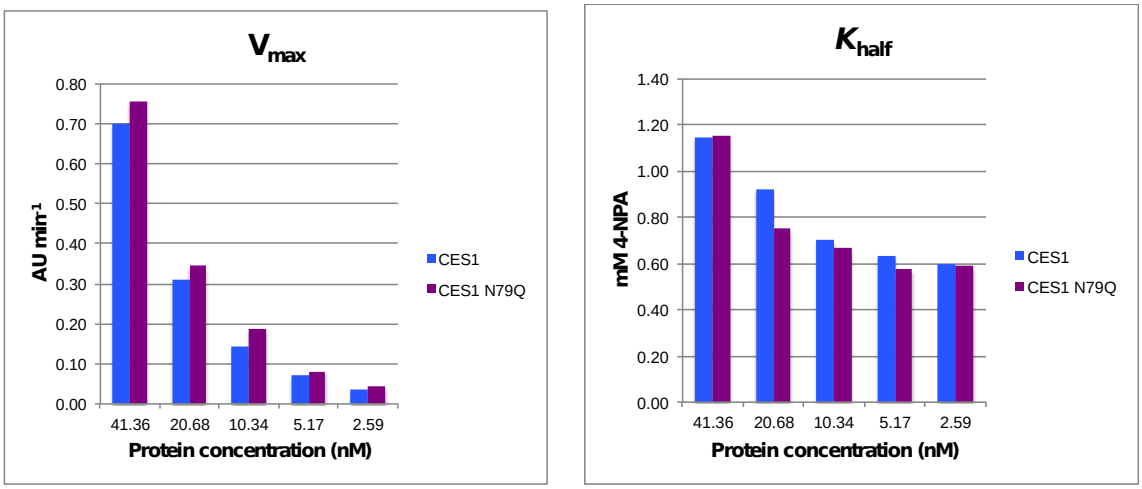


Figure 3.4. Kinetic Analysis of CES1 and CES1 N79Q.
1a and *2a* show substrate vs initial velocity plots. *1b,c* and *2b,c* are Lineweaver Burk plots (double reciprocal). In *b*, the reciprocals of all substrate concentrations are plotted but in *c*, the two lowest substrate concentrations (0.1875 and 0.0938 mM 4-NPA) are omitted to show much clearer parallel lines between protein amounts in the plot. The varying concentrations of enzyme that were assayed are indicated in the keys on the right hand side of each graph.

	Enzyme Concentration (nM)									
	41.36		20.68		10.34		5.17		2.59	
Allosteric sigmoidal										
	CES1	CES1 N79Q	CES1	CES1 N79Q	CES1	CES1 N79Q	CES1	CES1 N79Q	CES1	CES1 N79Q
V_{max}	0.698	0.753	0.312	0.344	0.145	0.187	0.072	0.081	0.035	0.042
h	1.434	1.300	1.490	1.603	1.771	1.886	1.805	1.958	1.908	2.033
K_{half}	1.147	1.154	0.919	0.749	0.700	0.668	0.635	0.573	0.594	0.587
K'	1.217	1.205	0.882	0.629	0.532	0.468	0.441	0.336	0.370	0.339
Std. Error										
V_{max}	0.044	0.102	0.019	0.028	0.007	0.013	0.004	0.006	0.001	0.004
h	0.112	0.199	0.135	0.249	0.190	0.303	0.216	0.383	0.194	0.474
K_{half}	0.131	0.300	0.101	0.115	0.062	0.079	0.059	0.076	0.042	0.088
K'	0.183	0.378	-0.153	-0.191	-0.111	-0.148	-0.107	-0.141	-0.079	-0.168
95% Confidence Intervals										
V_{max}	0.557 to 0.839	0.428 to 1.079	0.253 to 0.371	0.254 to 0.434	0.123 to 0.168	0.148 to 0.227	0.0602 to 0.0833	0.0626 to 0.0997	0.0308 to 0.0395	0.0310 to 0.0535
h	1.076 to 1.791	0.668 to 1.932	1.060 to 1.920	0.810 to 2.397	1.165 to 2.377	0.921 to 2.852	1.118 to 2.491	0.738 to 3.178	1.292 to 2.524	0.524 to 3.542
K_{half}	0.731 to 1.563	0.198 to 2.110	0.597 to 1.241	0.384 to 1.114	0.503 to 0.897	0.416 to 0.921	0.447 to 0.823	0.332 to 0.814	0.459 to 0.729	0.306 to 0.869
K'	0.636 to 1.799	0.003 to 2.406	0.395 to 1.369	0.0214 to 1.236	0.177 to 0.886	-0.004 to 0.939	0.0100 to 0.782	-0.113 to 0.785	0.119 to 0.621	-0.194 to 0.872
Goodness of Fit										
R^2	0.9985	0.994	0.998	0.993	0.997	0.993	0.996	0.988	0.997	0.983
Points Analysed	6	6	6	6	6	6	6	6	6	6

Table 3.4. Kinetic parameters obtained using GraphPad Prism 6.
Values are given for CES1 native (blue) and N79Q (purple). V_{max} is the maximum enzyme velocity given in AU min^{-1}. It is the velocity of the enzyme extrapolated to very high concentrations of substrate, and therefore is almost always higher than any velocity measured in an experiment. K_{half} is the concentration of 4-NPA that produces a half-maximal enzyme velocity (the EC50). ***h*** is the Hill slope. When *h*=1, this equation is identical to the standard Michaelis Menten equation. When it is >1.0, the curve is sigmoidal due to positive cooperativity. *h* does not always equal the number of interacting binding sites, but its value cannot exceed the number of interacting sites. The value of *h* is an empirical measure of the steepness of the curve. ***K'*** is related to the K_m. It is computed as K_{half}^{h}, and is expressed in the same units as the substrate, 4-NPA[134]. R^2 values for both enzymes indicate that the allosteric sigmoidal model fitted the data well as all values were > 0.99, apart from the two values for CES1 N79Q at the lowest concentration, which were still > 0.98.



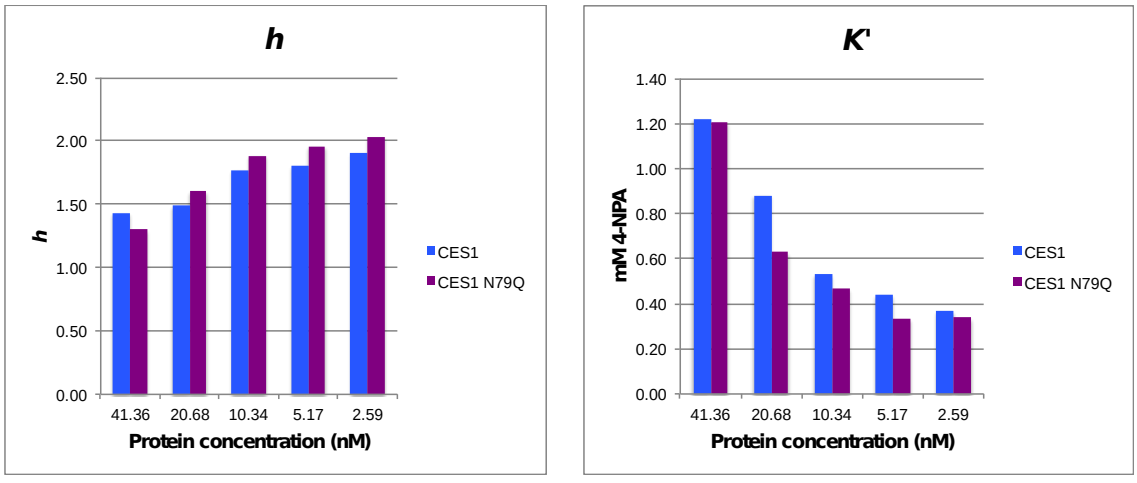


Figure 3.5. Graphical representations of statistic parameters.
Plotted as 2-D column charts. Values are taken from *table* 2.5.

3.3.2. Positive Cooperativity and the Hill Plot

The Hill Plot is an empirical plot that is used to quantify the extent of cooperativity shown by an enzyme, with a slope of greater than 1 indicating positive cooperativity within a system. For both enzymes at all concentrations, the Hill coefficient (*h*) was

>1.00, and this increased as the amount of enzyme used decreased. These values went from 1.434 → 1.908 for CES1 native and from 1.300 → 2.033 for CES1 N79Q (*table 3.4).* This demonstrates that the proteins behaved in a more 'cooperative' manner at lower substrate concentrations and that the affinity of 4-NPA for both CES1 native and N79Q increased as the concentration of 4-NPA increased. This can be seen at all concentrations of enzyme, but is more apparent at lower concentrations, and is in agreement with previous observations[109] of Hill coefficients of around 1.5, using imidaprilat as the substrate[109].

Because the enzymes do not show Michaelis Menten kinetics, accurate K_m values cannot be determined. Instead K_{half} values give a concentration of substrate that produces a half-maximal enzyme velocity, for a given amount of enzyme. Because the K_{half} value will vary with the amount of enzyme used, values need to be compared between proteins at the same concentration. At the highest protein concentration (41.36 nM), K_{half} values were 1.147 mM and 1.154 mM 4-NPA for CES1 native and N79Q respectively, and at the lowest concentration (2.59 nM) values were 593.80 μM and 587.4 μM.

3.3.3. Specific Activity

The rate of product formation was expressed as nanomole of pNP produced per minute per microgram of protein (nmol min^{-1} μg^{-1}) using the extinction coefficient for pNP at 405 nm (18,000 M^{-1} cm^{-1}) (*table 3.5*). At all substrate concentrations, apart from 3.0 mM, the specific activity for CES1 wt was measured as 7.50 ± 0.43 nmol min^{-1} μg^{-1}, and for CES1 N79Q, 10.6 ± 1.40 nmol min^{-1} μg^{-1}. This demonstrates that without the *N*-linked glycans attached, CES1 N79Q is slightly more active than the wild type.

At the highest concentration of 4-NPA, these values decreased to 3.99 and 4.65 nmol min^{-1} μg^{-1} for CES1 native and N79Q, respectively. This could indicate substrate inhibition at high concentrations.

Enzyme	[4-NPA]/ mM	Specific Activity (nmol min^{-1} μg^{-1})	Enzyme	[4-NPA]/ mM	Specific Activity (nmol min^{-1} μg^{-1})
CES1 native	0.0938	7.62 ± 1.05	**CES1 N79Q**	0.0938	12.56 ± 0.84
	0.1875	7.88 ± 0.44		0.1875	10.53 ± 0.52
	0.375	7.60 ± 1.08		0.375	9.62 ± 0.26
	0.75	7.63 ± 0.58		0.75	11.27 ± 1.73
	1.5	6.74 ± 0.48		1.5	9.03 ± 0.36
	3.0	3.99 ± 0.72		3.0	4.65 ± 0.33

Table 3.5. Specific activity values of CES1 mutants expressed in nmol min^{-1} μg^{-1}.
Values are the mean ± standard deviation of 3-5 determinations from differing protein concentrations, after normalization.

3.3.4. Ping Pong Mechanism

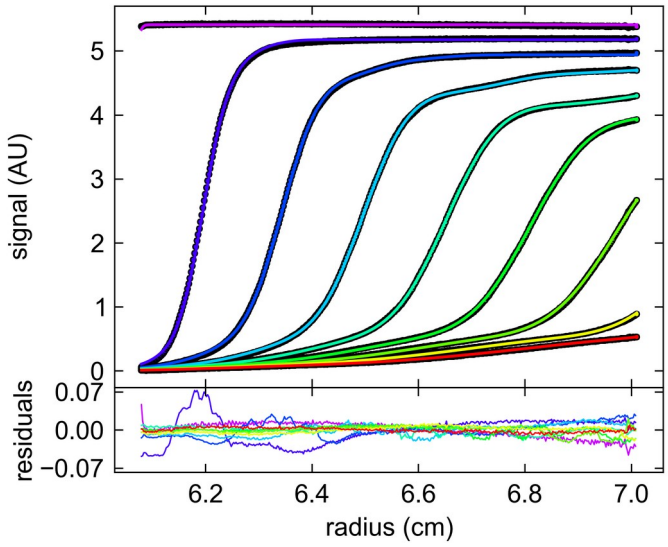
In the double reciprocal plots for both enzymes, where the lowest substrate concentrations were omitted, it appears that the lines for different protein concentrations lie almost parallel to each other, and would not converge if extrapolated. This is indicative of a ping pong or double-displacement mechanism, a characteristic of CES1 that has been seen previously[97], as well as in carboxylesterases from other organisms[135]. In this bireactant mechanism, the first substrate binds to the enzyme, transfers a constituent group, and then dissociates as the first product before the second reactant binds, picks up

the transferred group, and dissociates as the second part. CES1 cleaves via this two-step mechanism that includes the formation and degradation of an acyl-enzyme intermediate, using water as a transitional nucleophile[86]. It is the water molecule here that acts as the second substrate in the mechanism.

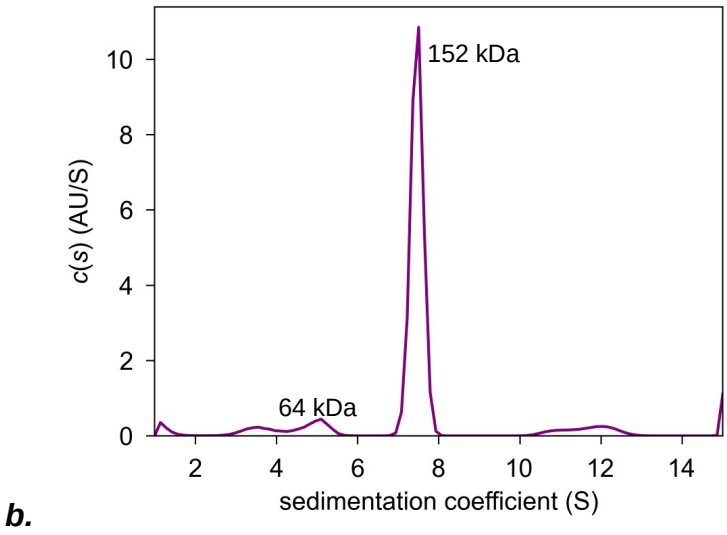
3.4. Biophysical Characterization

3.4.1. AUC

Sedimentation velocity experiments showed that all three forms of CES1 (wt, null and N79Q mutants) were mainly in the trimeric form (*figure 3.6*), with an apparent molecular weight of 152 kDa. A small amount of monomer was also identified (64 kDa), as well as some aggregation at a higher molecular weight. The sedimentation coefficient (*s*) was calculated to be 7.5, and the frictional ratio (*f*) was 1.52.



a.

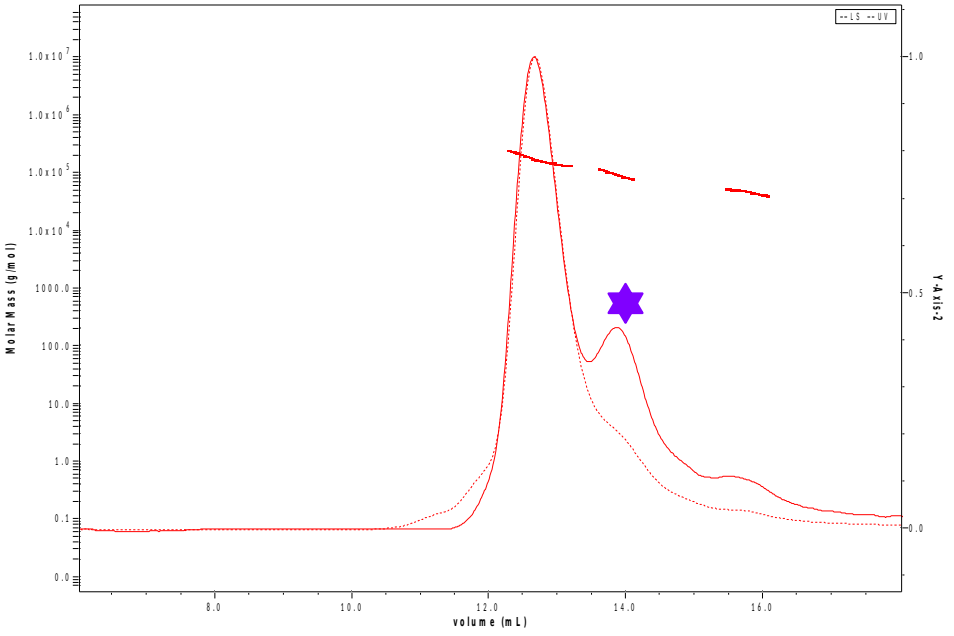


b.

Figure 3.6. Graphical AUC analysis for CES1 N79Q.
The plot of the residuals is shown in ***a,*** and the sedimentation coefficient population distribution (c(s)) against s plot is shown in ***b.*** Data for all three CES1 mutants gave the same profiles. Each clearly defined peak represents a single species. The main peak corresponds to a protein with a molecular weight of ~152 kDa. A small, minor peak of 64 kDa can also be fitted. Small amounts of aggregation is also seen.

3.4.2. SEC-MALLS

SEC-MALLS is an analytical technique for determining absolute molar masses of proteins by detecting how they scatter light. The results are generally in agreement with the AUC data, assigning the main peak a molecular weight of 158 kDa. However, in CES1 wt, a smaller peak corresponding to the monomeric species is seen (star), which is absent from the CES1 N79Q profile, where some hexamer is seen (triangle).



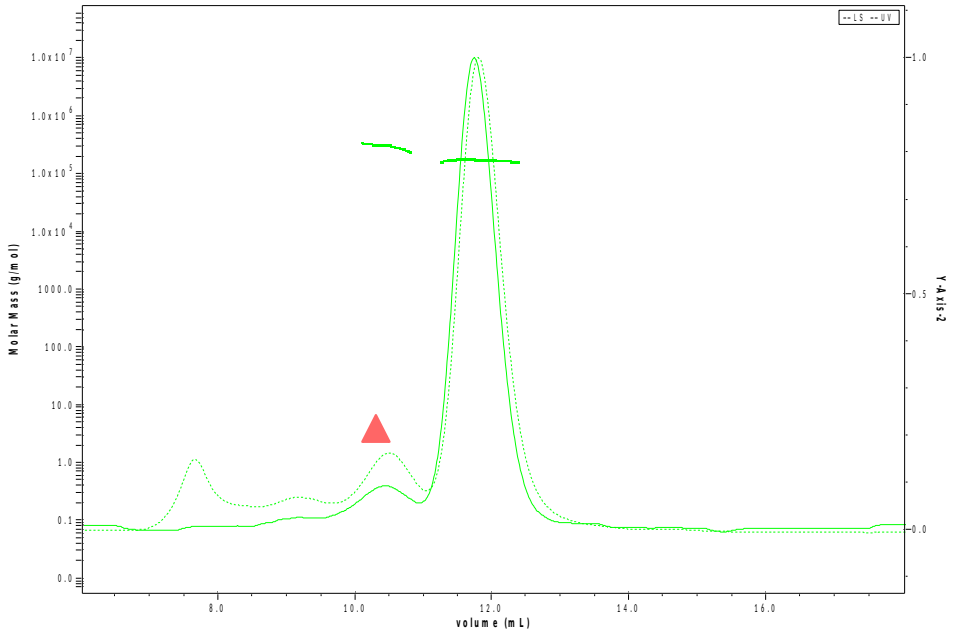
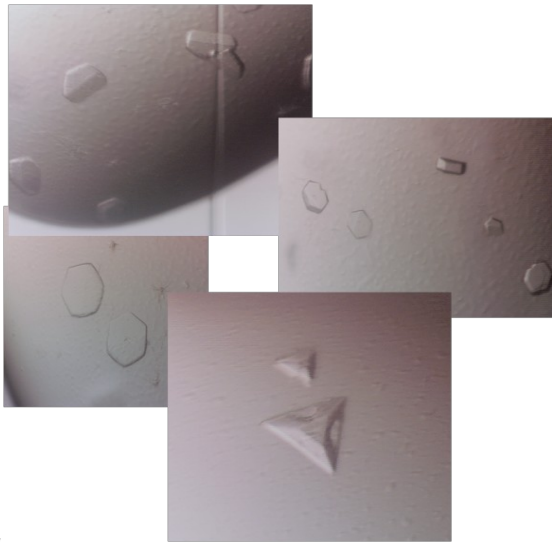


Figure 3.7. Molar mass vs volume SEC-MALLS plots for CES1 wt and CES2 N79Q. CES1 wt is the top graph, N79Q is the bottom. Both were run at 25 °C with a flow rate of 0.3 mL/ min. The main peak in CES1 corresponds to a molecular weight of 161 kDa, with the main peak in CES1 N79Q at 169 kDa.

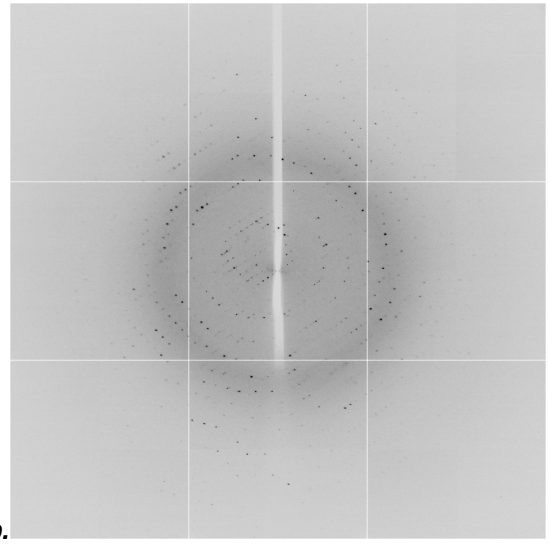
3.5. Crystallization and Structural Analysis

3.5.1. Crystallization

All of the CES1 constructs designed here crystallized readily in sitting drop vapor diffusion experiments, demonstrating different crystal morphologies in a number of conditions from several commercially available crystallization screens (*figure 3.8, a, c*). Nucleation and growth of crystals occurred rapidly, from approximately 4 hours to 7 days after setting up of the crystal trays, at both 4°C and 20°C. Further optimization around original hits was not required.



a.



b.

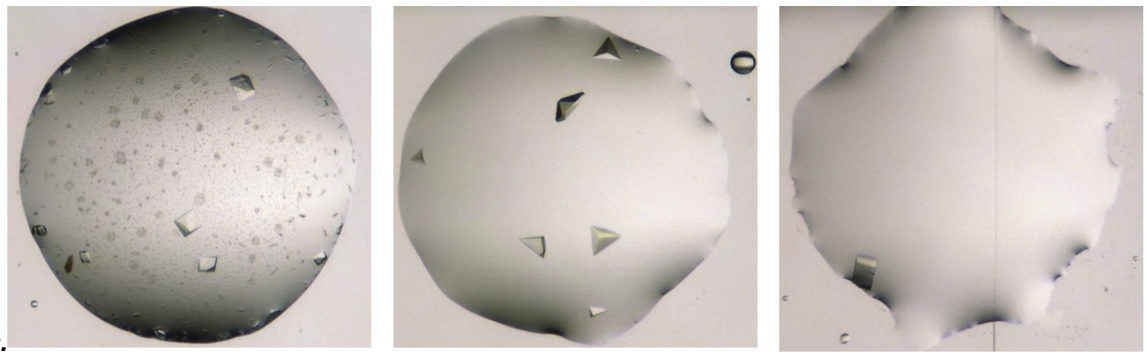
c. 

Figure 3.8. Representative images of CES1 crystals and the resulting diffraction quality.
a. A subset of the various morphologies that CES1 native crystallized in. The conditions that they were grown in are listed from the top, in a clockwise fashion: JCSG+ C4, JCSG+ D10, Morpheus E3, Morpheus H6[116, 117]. The smallest crystals were approximately 40 x 40 x 10 μm.
b. Diffraction pattern from the best diffracting CES1 native crystal collected on beamline I04.). The edge of the ADSC 315 CCD detector was set to a resolution of 1.5 Å.
c. *In* situ crystal images taken directly from the crystallization plates prior to harvesting of the best diffracting crystals, from which the data sets for the structures described here were collected. From left to right: CES1 wild type, CES1 null and CES1 N79Q. The largest crystals obtained were for CES1 null and were approximately 160 μm in diameter.

3.5.2. Data Collection and Refinement

A number of different crystal hits were tested for diffraction at Diamond Light Source. Although relatively small (~30 μm largest dimension), in general all forms of CES1 diffracted very well. A summary of the data collection and refinement statistics for the three refined structures of CES1 are listed in *table 3.6.*

3.5.2.1. CES1 Wild Type

A single crystal grown in the Morpheus E3 condition at 20°C of CES1 wt in the *apo* form diffracted to 1.86 Å and was characterized as belonging to space group *H*3 with unit cell dimensions a = 114.71 Å, b = 114.71 Å, c = 117.78 Å and α = 90°, β = 90°, γ = 120°. Data were collected on beamline I04 with an ADSC 315 CCD detector. A total of 261 images were collected using a 0.35° oscillation angle per image. The data were processed with the HKL2000 suite of programs[136]. The final structure consisted of 4155 protein atoms, 1 NAG molecule and 270 water molecules, and was refined to an *R*-factor of 15.83% and *R*-free of 18.76 % (*table 3.6*).

3.5.2.2. CES1 Null

A single crystal grown in the Morpheus E11 condition at 4°C of CES1 null diffracted to beyond 1.58 Å resolution. The crystals were characterized as belonging to space group *H*3 with unit cell dimensions, a = 115.39 Å, b 115.39 Å, c = 128.14 Å and α = 90°, β = 90°, γ = 120°. Data were collected using shutterless data collection in fine slice mode (0.15° oscillation angle) at beamline I03 with a DECTRIS P6M pixel array detector. Data were integrated and scaled automatically using the xia2 pipeline to a resolution of 1.58 Å. These data were subsequently reprocessed manually using xia2 to a resolution of 1.48 Å. The structure was refined using REFMAC5[121] within the CCP4 suite and the final model consisted of 4119 protein atoms, 1 NAG molecule and 696 water molecules (*table 3.6*).

3.5.2.3. CES1 N79Q

A single crystal grown in the PACT G3 condition at 4°C of CES1 null diffracted to 2.01 Å and was characterized as belonging to space group *H* 3 with unit cell dimensions a = 115.47 Å, b = 115.47 Å, c = 127.28 Å and α = 90°, β = 90°, γ = 120°. Data were collected using shutterless data collection in fine slice mode (0.15° oscillation angle) at beamline I04 with a DECTRIS P6M pixel array detector. Data were integrated and scaled automatically using the xia2 pipeline to a resolution of 2.18 Å. These data were subsequently reprocessed manually with xia2 to a resolution of 2.01 Å. The structure was refined using REFMAC5[121] within the CCP4 suite and the final model consisted of 4130 protein atoms, 8 iodide ions and 221 water molecules (*table 3.6*).

In all three structures, the asymmetric unit contained one monomer of CES1 (as expected from estimation of the Matthews correlation coefficient and validated by analysis of the experimental data), and crystal symmetry generated the biological trimer that is seen in solution and supported by AUC, MALS and gel filtration profiles. The high resolution structures presented here were obtained from a new crystal form of CES1, *H*3 (*H* centred

trigonal), despite being essentially identical to previous published structures of CES1 produced from insect expression systems and co-crystallized in the presence of various ligands (*figure 3.9*). Previously, CES1 has crystallized most commonly in the orthorhombic space group $P2_1\ 2_1\ 2_1$ (*table 3.1*).

In the glycosylated crystal structures of CES1 wt and null, one GlcNAc molecule can clearly be modeled onto the high-mannose *N*-linked glycosylation site, N79 (*figure 3.10, a*). In all three structures, two intact disulphide linkages, C87- C116 and C274 - C285, were present.

	CES1 WT	**CES1 S221A**	**CES1 N79Q**
Data Collection			
X-Ray Source	DLS, I04	DLS, I03	DLS, I04
Wavelength (Å)	0.9795	0.9762	0.9795
Space Group	*H*3	*H*3	*H*3
Cell Constraints			
a, b, c [Å]			
α, β, γ [°]	114.72, 114.72, 117.78 90, 90, 120	115.39, 115.39, 128.14 90, 90, 120	115.47, 115.47, 127.28 90, 90, 120
Resolution range [Å]	50.00-1.86 (1.88 – 1.86)	53.94-1.48 (1.52 - 1.48)	78.63 – 2.01 (2.06 - 2.01)
Completeness (%)	97.0 (71.5)	98.4 (91.8)	98.3 (98.0)
R_{symm}[a]	6.5 (43.3)	4.0 (69.2)	6.3 (61)
<I>/□(<I>)	17.4 (1.8)	11.1 (1.0)	9.7 (1.3)
Multiplicity	2.8 (2.3)	2.9 (2.0)	3.0 (3.0)
Data Refinement			
No. of reflections	132,450	303,901	125,184
No. of unique reflections	46, 879	104,388	41,414
R-factor (%)[b]	15.83	12.96	16.67
R_{free} (%)	18.76	17.19	21.17
Rmsds			
Bond lengths (Å)	0.013	0.0117	0.0141
Bond angles (°)	1.630	1.5133	1.616
Wilson *B*-factor (Å^2)	21.0	19.2	34.4

Mean B-factor (Å^2)	18.7	31.0	31.3

Table 3.6. Data collection and refinement statistics for all three structures of CES1.
[a]Values in parentheses refer to data in the highest resolution shell for each protein.
$R_{symm} = \Sigma_{hkl}\Sigma_j|I_{j,hkl} - \langle I_{hkl}\rangle|/\Sigma_{hkl}\Sigma_j I_{j,hkl}$ where $\langle I_{hkl}\rangle$ is the average of the intensity $I_{j,hkl}$ over $j = 1,..., N$ observations of symmetry equivalent reflections *hkl*.
[b]R-factor = Σ (||F(obs)| - |F(calc)||)/ Σ|F (obs)|; Rfree = R factor for a selected subset (5%) of the reflections that was not included in prior refinement calculation.

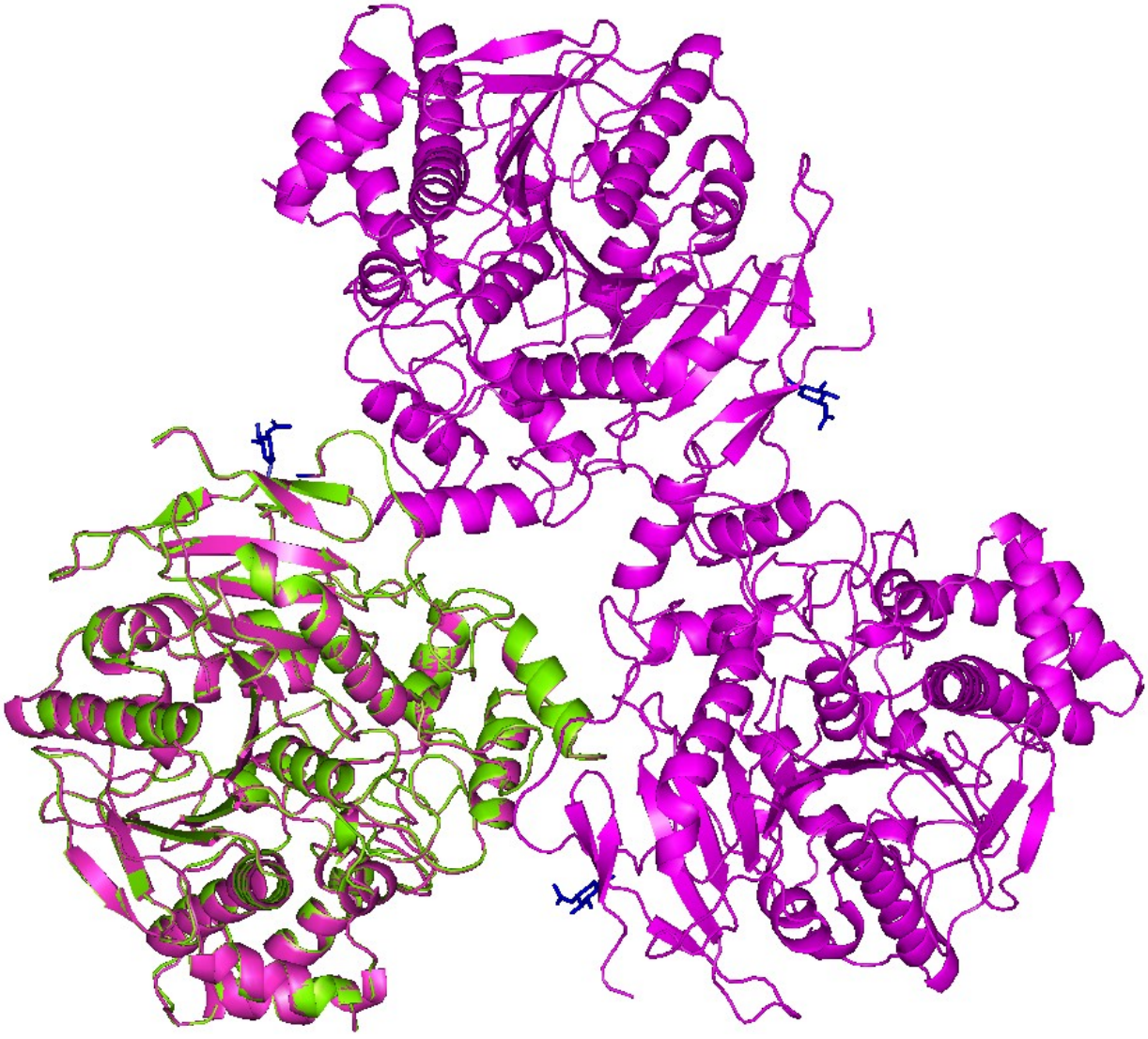


Figure 3.9. Overall quaternary structure of CES1 in its trimeric form.
Trimeric arrangement of CES1 is shown in magenta, with an overlay of a monomer of CES1 N79Q shown in green. There is no overall conformational change in the 3-D structure observed between CES1 and CES1 N79Q. It is clear that the *N*-linked glycans (blue) are not involved in trimer formation or protein stabilization.

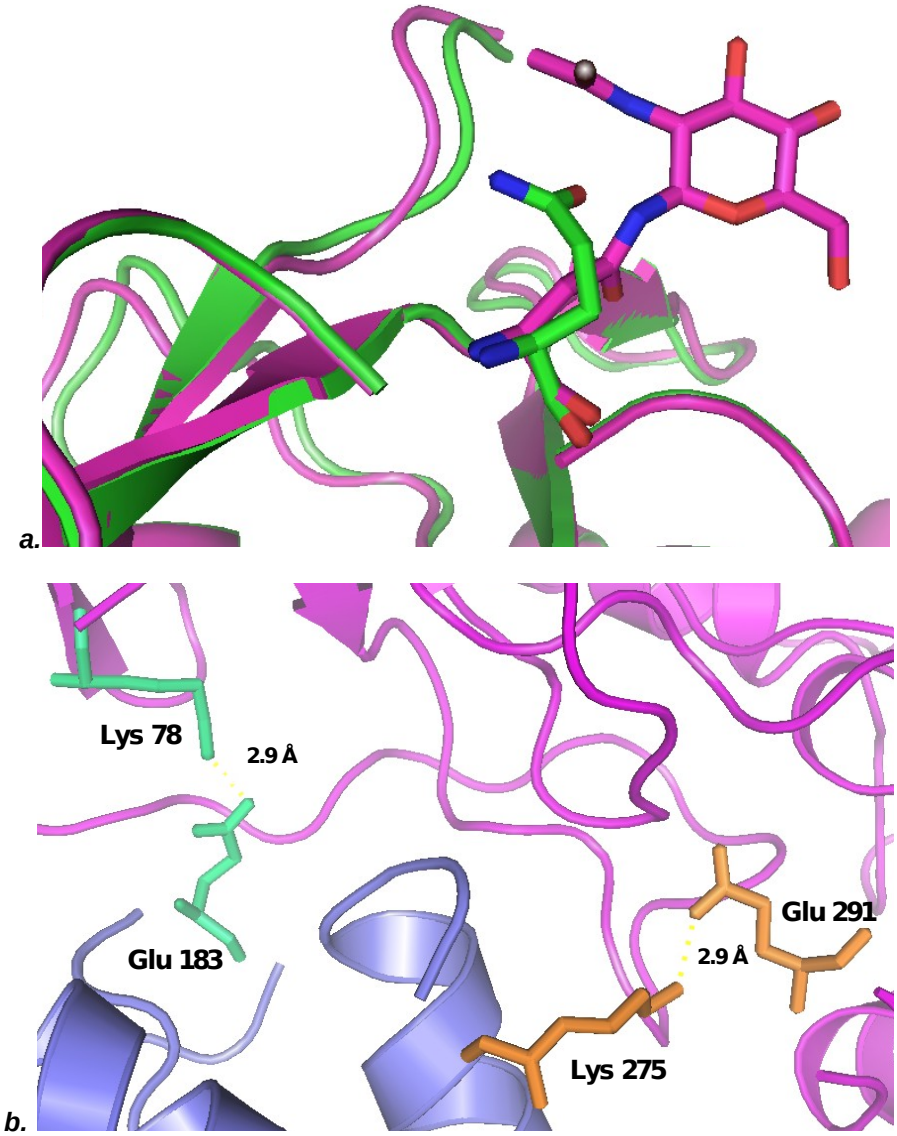


Figure 3.10. Detailed views of CES1 structural elements.
a. Overlay of the conserved glycosylation site. CES1 N79Q is shown in green, and CES1 wt is shown in magenta with one GlcNAc molecule modeled. ***b.*** The positioning of the salt bridges involved in trimer formation between two monomers of CES1.

	RMSD (Å)		
	Wt	**Null**	**N79Q**
Wt	-	0.44	0.47
Null	-	-	0.30
2H7C*	0.50	0.48	0.48
4AB1	0.52	0.57	0.56
1YA8*	0.72	0.70	0.68

Table 3.7. Structure alignment results from PDBeFold.
The table shows comparative overall RMSD values using the structures presented here, along with three other deposited structures. (*average from all chains in the asymmetric unit). 2H7C was used as it is the highest resolution structure of CES1 currently deposited with the PDB. RMSD was calculated between Ca-atoms of matched residues at best 3D superposition of the query and target structures.

All CES1 structures reported here (*table 3.5)* are almost identical to that of the search model used for molecular replacement (2H7C), indicating that binding of ligands to CES1 does not induce any gross overall conformational change in the protein from its *apo* form (*table 3.7)*. However, localised structural differences between the monomeric CES1 reported here and the ligand bound hexamer 2H7C were observed. Residues A337 to H442 were in a different structural conformation (*figure 3.13*) compared to the search model (2H7C) and after refinement of this region, it was evident that this loop region was on the surface of the protein molecule. The 2H7C CES1 monomer search model was derived from a crystal of CES1 that crystallized as a hexamer. N340 appeared to be a key residue involved in the crystal packing interaction in 2H7C, whereas in the structures reported here with one molecule in the asymmetric unit, N340 is buried, and not exposed on the protein's surface.

3.5.3. Structural Analysis

3.5.3.1. Comparison of CES1 Wild Type and N79Q

Analysis of the CES1 wt and N79Q structures shows that glycosylation does not have an effect on trimer formation, contrary to what has been proposed previously[78], as the de-glycosylated form of CES1 overlays perfectly with the glycosylated form and no conformational change is seen (*figure 3.9, figure 3.10, table 3.6*). In previous published work, it had been postulated that trimer formation in CES1 was mediated by two charged clamps across the trimer interface, from R186 and E183 of one monomer to E72 and K78 respectively, of the next monomer[78]. These charged clamps are effectively salt bridges, which are actually a combination of two non-covalent interactions: hydrogen bonding and electrostatic interactions. Examination of the interfaces of the CES1 structures reported

here, along with other deposited structures of CES1 (1MX1, 1YAH, 2H7C), confirms that the E183 and K78 pairing does indeed form a salt bridge between two monomers. However, in the CES1 wt, N79Q and null structures no salt bridge between R186 and E72 is observed (*figure 3.11*) as these residues are positioned 6.7 Å apart, and it is assumed this was erroneously identified in the previous publication describing the structure of CES1. However a salt bridge formed between K275 and E291 on two adjacent molecules, 2.9 Å apart (*figure 3.10, b*) has been identified, which is highly likely to be involved in the oligomerisation of CES1, along with the original K78 and E183 pairing. Therefore in CES1, the two salt bridges involved in stabilizing the tertiary structure of enzyme arise from the anionic carboxylate ($RCOO^-$) of a glutamic acid and the cationic ammonium (RNH_3^+) from a lysine. In previous structures, it is highly likely this newly identified salt bridge was not detected due to incorrect rotamer assignement for the respective residues (K78 and E183) in these deposited structures.

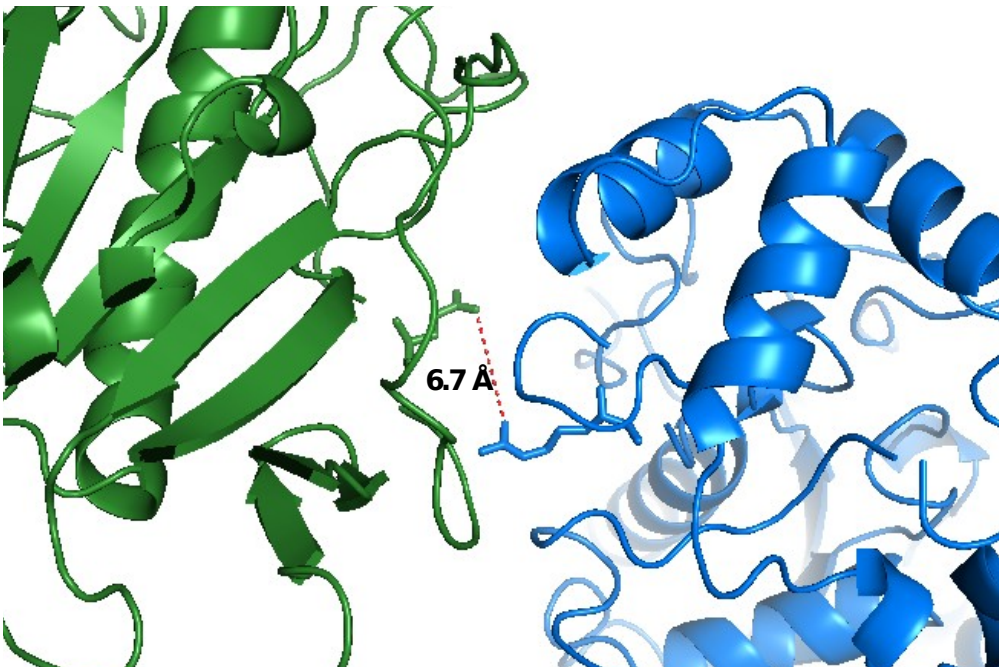


Figure 3.11. Detail showing the location of residues Glu 72 and Arg 186 which have previously been proposed to be involved in stabilization of the CES1 trimer.
Glu 72 is indicated in sticks on the monomer in green, and Arg 186 is indicated on the monomer in blue. This pairing clearly does not form a salt bridge between two monomers, as the residues are positioned 6.7 Å apart. This was observed in all three CES1 structures presented in this chapter, and in reviewing all previous publishes structures of CES1.

Since there is no overall conformational change in the 3-D structure of CES1 wt and CES1 N79Q (*figure 3.9 and figure 3.10*), it is clear that the *N*-linked glycans do not appear to assist in the stabilization of the CES1 trimer. This contrasts with what was observed by Bencharit et al.[85] who proposed that the *N*-glycans packed into adjacent monomers in the oligomer. They also identified sialic acid (SIA) moieties within their structure, which they believed to be portions of the extended glycoform at N79, although they were not connected by electron density. In the *N*-glycoslyation of proteins, the addition of sialic acids occurs in the Golgi following transit from the ER. Since native CES1 is resident in the ER, the proposed structural role of sialic acids is not physiologically relevant.

3.5.3.2. Non-Specific Binding Sites

In all three structures, no ligands from the crystallization screens were observed bound to either the side-door or the Z-site, an alternative binding site originally described by Bencharit et al.[84]. These observations challenge the idea that binding of a ligand to the Z-site prevents the packing of two trimers to form a hexamer[84], as hexamers are seen when further hexagonal symmetry is generated. Because of this, it is unlikely that this Z-site plays a role in regulating the trimer-hexamer equilibrium, and in fact the position of this dynamic equilibrium is concentration dependent.

Cysteine 389

Interestingly, in all of the CES1 structures, positive density was observed, within covalent bonding distance, of the cysteine residue at 389 (*figure 3.12*). Binding to, or modification of this residue, has not been reported in any of the published structures to date. This cysteine residue is located in the lining of the entrance to the side-door, a feature that has been exposed in previous structures of CES1 and is proposed to act as an alternative opening for the trafficking of substrates and products[12, 69]. The entrance to the

side-door is lined by structurally flexible residues, and is separated from the active site by a thin wall. It was not possible to identify the compound bound to C389, but because these are the only structures of CES1 produced in HEK cells, it is possible that the adjunct could either be a component from the complex DMEM media that the cells were grown in, or a compound that was picked up from within the cells themselves.

This observation fuels further questions as to this enzymes promiscuous nature, and highlights another potential non-specific binding site on the protein. As this cysteine residue lines the entrance to the active site, it could also be speculated that covalent modification of this residue could result in an inhibitory effect by blocking access into or out of the catalytic core.

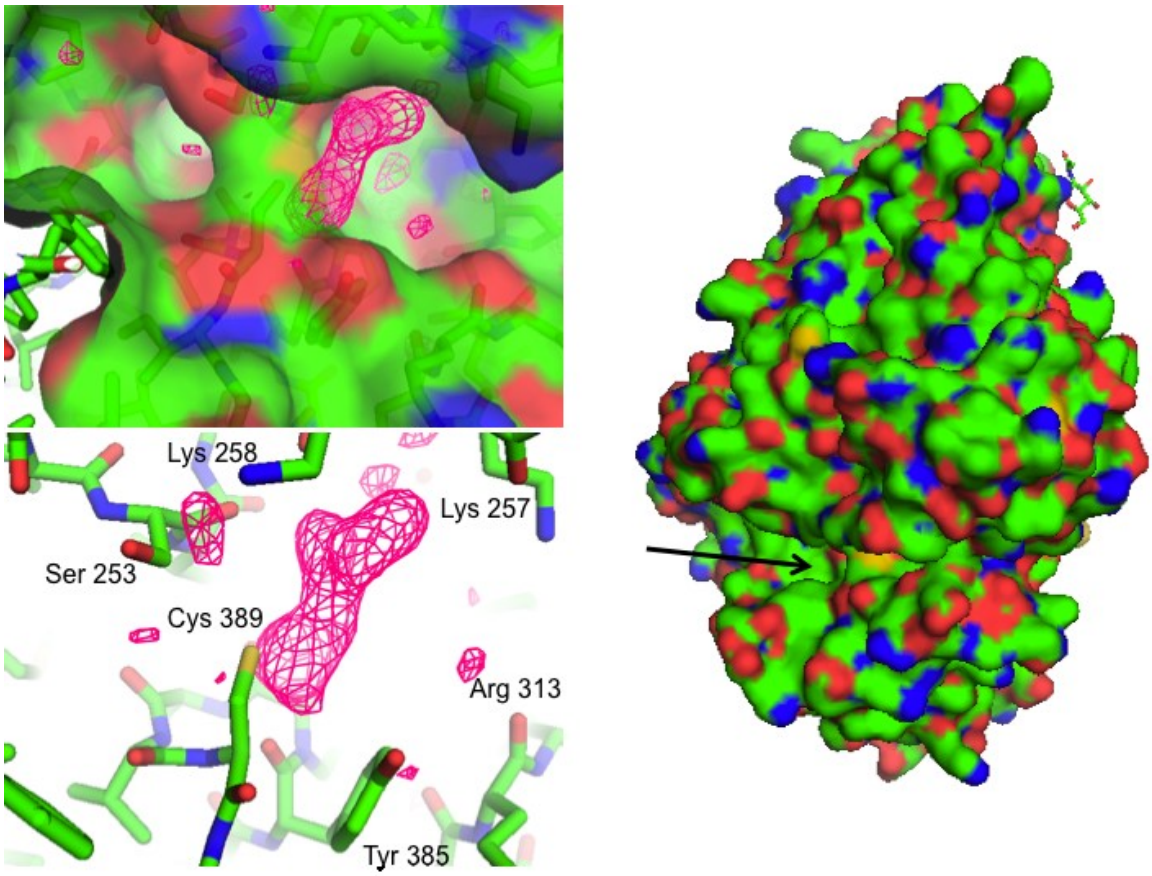


Figure 3.12. Positive density observed at Cys 389.
The Fo-Fc positive electron density map (left) is shown in pink mesh contoured at a level of 3σ l (with and without surface representation). It is evident that the sizeable blob of positive density is within covalent bonding distance of Cys 389. On the right is a surface representation of CES1, with the *N*-linked glycan visible on the right-hand side of the molecule at the top. The arrow indicates the entrance to the side-door on the surface of the molecule, where positive density is observed.

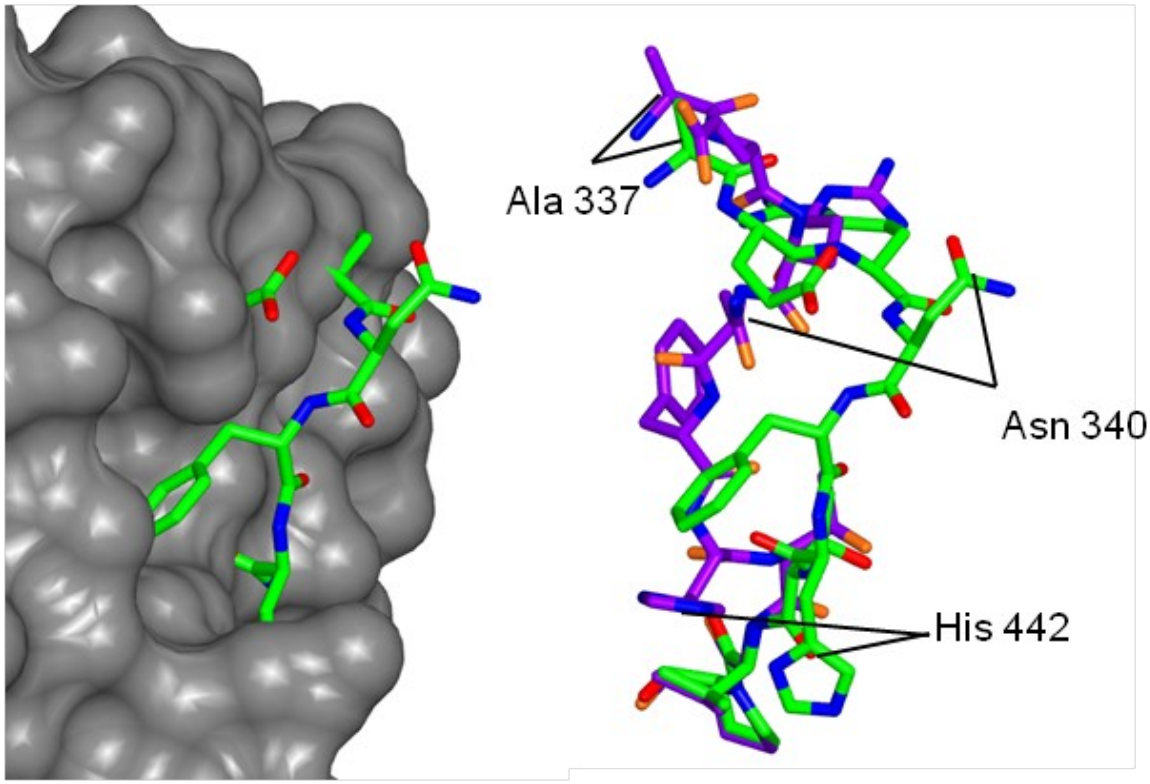


Figure 3.13. Overlay of disordered region on the surface of CES1.
Overlay of structure from this report (purple) and the equivalent part of the CES1 polypeptide chain from 2H7C (green) with the position of Asn 340 in each molecule indicated.

3.6. Discusison

3.6.1. Protein Production

The current results confirm that CES1 produced from HEK293 gnt$^{-/-}$ stable cell lines is essentially identical to previous examples of this enzyme isolated from cultured insect cells, validating the use of the mammalian system as a source for recombinant proteins in structure determination studies. The secretion of histidine-tagged CES1 has previously been demonstrated in insect cells, but no previous studies have reported the production of recombinant histidine-tagged CES1, from mammalian systems in sufficient quantity and quality for comparative kinetic and structural studies. Mammalian expression systems are now rapidly becoming the system of choice for protein production for structural studies[137].

Activity

Cooperativity refers to the observation that binding of a substrate molecule at one site affects the binding affinity of another site on the enzyme. This occurrence is usually seen in oligomeric proteins that undergo a concerted transition in each subunit from a low to a high affinity form[132] or in proteins that possess an allosteric binding site, other than the active site, where the modifier adds and causes an effect. Three possible explanations for the kinetic behavior of CES1 are as follows:

1. *Oligomer cooperativity* - The binding of a substrate molecule to the active site in one monomer may induce a slight conformational change in another monomer, thereby increasing affinity of the adjacent monomer in the trimer for the substrate.
2. *Allosteric cooperativity* – The substrate may bind non-specifically to the Z-site in the regulatory domain of CES1, or another non-selective binding site, and cause a conformational change that shifts the enzyme from the inactive T form to the active R form. The T "tense" state is the low-activity, low-affinity form, and

the R "relaxed" state is the high-activity, high-affinity form. This explanation is not mutually exclusive to cooperactivity via binding at the active site. An allosteric site is one other than the active site where the modifier adds and causes an effect.

3. *Additional Effector* – Water, which is the additional reactant in the ping pong mechanism, may act as an allosteric effector during the formation of acyl-enzyme complex. This seems the least likely explanation as water is presumed not limiting in the reaction, though it is not possible to measure the amount of water in the system.

Substrate concentration plays an important part in allosterism. At high concentrations of substrate, more enzymes are found in the R state. The T state is favoured when there is an insufficient amount of substrate to bind to the enzyme. In other words, the T and R state equilibrium depends on the concentration of the substrate. This is the first study to observe sigmoidal kinetics in recombinant CES1, both natively and aglycosylated, and supports observations made by Takahashi et al.[109], contrary to the majority of CES1 kinetic publications. Glycosylation may also affect the equilibrium between the T and R states, as CES1 N79Q appears slightly more cooperative than the wild type. Interestingly, it has been observed that CES2, a homologue of CES1 sharing 49% sequence identity, also does not follow classical Michaelis Menten kinetics[111] (this will be discussed further in *Chapter 4*). A possible explanation as to why this has been overlooked by so many groups that have performed kinetic analysis on the carboxylesterases is that assays have all been carried out at high substrate concentrations (e.g. ~3 mM for 4-NPA). At high substrate concentrations, almost all of the enzyme will be in the R form, and the kinetics will be of the normal Michaelis Menten type. Positive cooperativity is vastly more prominent at lower concentrations. Importantly though, *in vivo* localized drug

concentrations may be low, so these results provide a much clearer picture of the pharmokinetics of CES1.

Interestingly, homotropic positive cooperativity has been shown in the guinea pig homolog of CES1 (gpCE), which shares 71% identity[88] and also exists primarily as a trimer. The findings presented here are the first that examines this positive cooperativity observed in CES1.

In rCE, it has been established that glycosylation is important in enzyme function, and so was proposed to be crucial for the activity of CES1. However, results here show that mutating the N to Q at residue 79, preventing glycosylation, has no detrimental effect on the catalytic activity of human CES1. This is consistent with results from enzymatic studies on the organophosphatase, human paraoxonase (PON1), where the substitution of the glycosylation sites from N to Q showed no effect on enzyme activity, demonstrating their insignificance in PON1 secretion and activity[138].

3.6.3. Protein Structure

The crystal structures of CES1 native, CES1 S221A and CES1 N79Q were all refined to a resolution greater than 2.0 Å, and are the first crystal structures of CES1 produced using a HEK expression system. The structures revealed a new site of interaction, C389, that is capable of binding molecules covalently, which raises the question as to whether this site has a functional role. Follow-on mutagenesis studies would be required to explore the exact role of C389 in the function of the enzyme, but the location of this residue at the entrance to the substrate-binding pocket is suggestive of some redox control of substrate access.

Contrary to previous presumptions, which were in fact based on only limited data, *N*-glycosylation of CES1 does not appear to play either a structural or functional role at least *in vitro*. Because there are no major conformational changes between CES1 wt and CES1 N79Q in their crystals structures, it is very unlikely that the carbohydrate links the central region of the protein to the *C*-terminal helix, unlike previously reported[12], and neither does a sialic acid from a glycosylation site on one monomer stack adjacent to a helix in another monomer to stabilise it[78]. Glycosylation is therefore highly unlikely to play a role in the stabilization of the protein's 3D-structure or in the formation of the trimer, unlike originally thought[78, 94]. It may be that glycosylation of CES1 has a role *in vivo*, for example protecting the enzyme from proteolysis *in situ*. This could be explored by comparing the turnover rate of the wild type and aglycosylated mutant in cell-based pulse-chase experiments.

The work described in this chapter has raised new questions about the structure-function of human CES1 and in particular the possibility of allosteric regulation of activity. A platform has been developed based on mammalian cell expression for the rapid production of the enzyme which could be exploited in follow-on work to explore the allosteric mechanism. Future work could include deletion of the salt bridge residues implicated in the trimer assembly using site directed mutagenesis, followed by kinetic analysis to see if the auto-activation kinetics described depends on oligomerisation of CES1.

4. PRODUCTION, CHARACTERISATION AND STRUCTURAL DETERMINATION OF HUMAN CARBOXYLESTERASE 2

4.1 Introduction

Studies on the structure-function relationship of CES1 have dominated the human carboxylesterase field over the last fifteen years. By contrast, research into human CES2 has not matched the widespread biological roles of this enzyme, which have only recently been fully appreciated. The majority of CES2 publications concern its substrate specificity, clinical importance and evolutionary history. Structural studies of CES2, which have been attempted by several groups, have been hampered by the enzyme's apparent inherent inability to crystallize. Although it has been widely proposed that CES2 can hydrolyze much larger, complex molecules than CES1, showing a preference for substrates with a small acyl moiety and large alcohol group (e.g. CPT-11, aspirin and cocaine benzoyl ester)[38], the absence of a high resolution crystal structure of CES2 limits an understanding of its exact selectivity and specificity. As there have been no structural data of CES2 available, computational approaches have been taken to predict the hydrolysis of esters by the enzyme, using a model based on the structure of CES1[139]. A crystal structure of CES2 would help to understand its drug specificity and aid in the design of CES2 specific pro-drugs. In addition, up until now, the structural and functional role of the two *N*-glycosylation sites in CES2 has not been explored. Therefore, these became the key objectives of the work on CES2, with an emphasis and focus on determining the structure by utilizing the facilities and technologies available at the OPPF-UK.

As the project developed, and it became evident that the original constructs designed for CES2 did not yield crystals (*see 4.6.1 for details*), re-thinking the construct design process became vital. Protein engineering, using site-directed mutagenesis[103], was carried out to biochemically mutate out the *N*-linked glycosylation sites, and to promote salt bridge formation.

Historically, the majority of studies on CES2 have used protein purified from whole cell microsome lysates[8, 111, 140, 141]. This has been superseded by the use of recombinant CES2 proteins for functional work produced mainly using the *Spodotera frugiperda* (Sf21) insect expression system[7, 21, 38, 45, 48, 73, 75, 77, 93, 112, 127, 130, 133]. Recently, the expression of CES2 in HEK293T cells has been reported[96] in parallel with work detailed here.

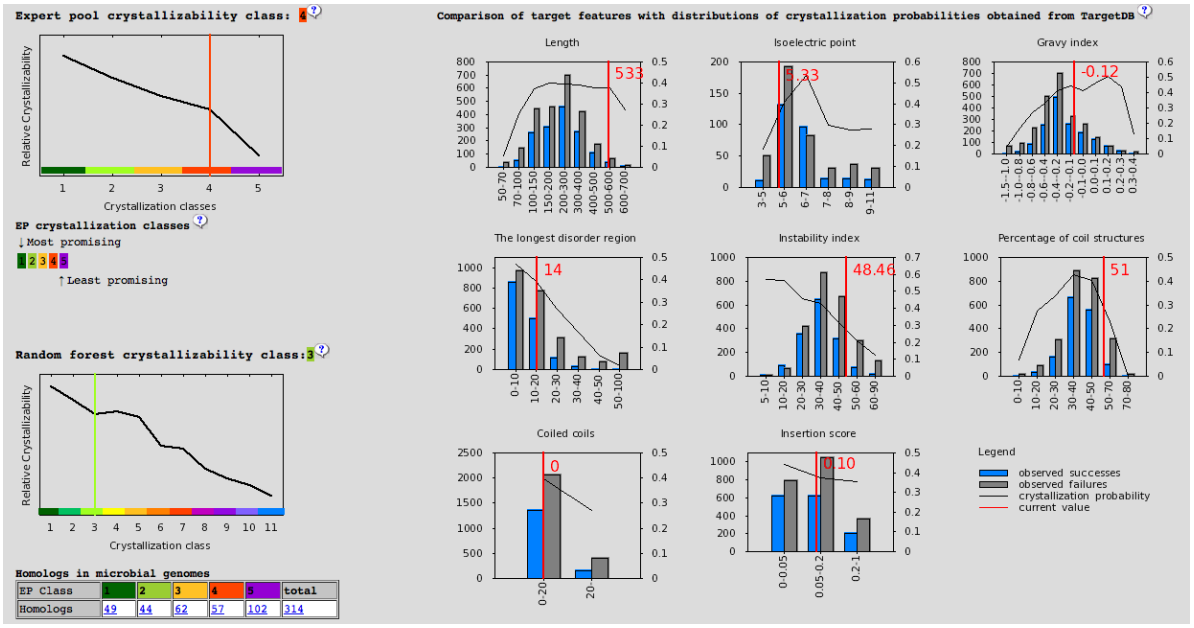
Chapter four summarizes the production, characterization and structure determination of CES2. The structure of CES2 in complex with the byproduct of CPT-11 hydrolysis, 4-PP, was solved to a resolution of 2.04 Å, providing the first crystallographic structure of CES2. This chapter also details new observations of the atypical kinetics of CES2 and an analysis of the effect of *N*-glycosylation on enzyme activity.

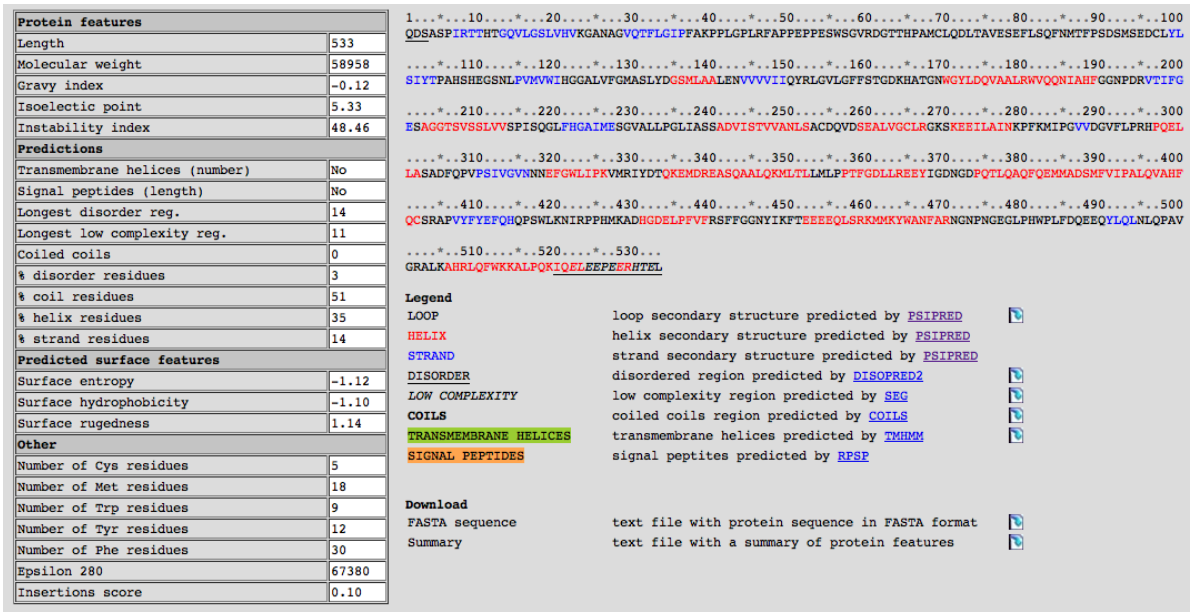
4.2. Protein Engineering of Human CES2

4.2.1. Construct Design

As part of the expression vector design process, the *C*-terminal endoplasmic retention motif, HTEL, was deleted from all CES2 sequences. This ensured that proteins were secreted into the cell culture media, and not retained in the luminal side of the ER where they usually reside.

As obtaining the crystal structure of CES2 was the main focus of this part of the project, and analysis of the secreted sequence using the XtalPred server[142] showed that CES2 fell into the 'difficult' crystallization class (*see figure 4.1*), a series of constructs were designed for recombinant expression in mammalian cells in order to improve the likelihood of obtaining diffraction quality crystals of CES2.





Protein features	
Length	533
Molecular weight	58958
Gravy index	-0.12
Isoelectic point	5.33
Instability index	48.46
Predictions	
Transmembrane helices (number)	No
Signal peptides (length)	No
Longest disorder reg.	14
Longest low complexity reg.	11
Coiled coils	0
% disorder residues	3
% coil residues	51
% helix residues	35
% strand residues	14
Predicted surface features	
Surface entropy	-1.12
Surface hydrophobicity	-1.10
Surface rugedness	1.14
Other	
Number of Cys residues	5
Number of Met residues	18
Number of Trp residues	9
Number of Tyr residues	12
Number of Phe residues	30
Epsilon 280	67380
Insertions score	0.10

```
1...*...10....*...20....*...30....*...40....*...50....*...60....*...70....*...80....*...90....*..100
QDSASPIRTTHTGQVLGSLVHVKGANAGVQTFLGIPFAKPPLGPLRFAPPEPPESWSGVRDGTTHPAMCLQDLTAVESEFLSQFNMTFPSDSMSEDCLYL

....*..110....*..120....*..130....*..140....*..150....*..160....*..170....*..180....*..190....*..200
SIYTPAHSHEGSNLPVMVWIHGGALVFGMASLYDGSMLAALENVVVVIIQYRLGVLGFFSTGDKHATGNWGYLDQVAALRWVQQNIAHFGGNPDRVTIFG

....*..210....*..220....*..230....*..240....*..250....*..260....*..270....*..280....*..290....*..300
ESAGGTSVSSLVVSPISQGLFHGAIMESGVALLPGLIASSADVISTVVANLSACDQVDSEALVGCLRGKSKEEILAINKPFKMIPGVVDGVFLPRHPQEL

....*..310....*..320....*..330....*..340....*..350....*..360....*..370....*..380....*..390....*..400
LASADFQPVPSIVGVNNNEFGWLIPKVMRIYDTQKEMDREASQAALQKMLTLLMLPPTFGDLLREEYIGDNGDPQTLQAQFQEMMADSMFVIPALQVAHF

....*..410....*..420....*..430....*..440....*..450....*..460....*..470....*..480....*..490....*..500
QCSRAPVYFYEFQHQPSWLKNIRPPHMKADHGDELPFVFRSFFGGNYIKFTEEEEQLSRKMMKYWANFARNGNPNGEGLPHWPLFDQEEQYLQLNLQPAV

....*..510....*..520....*..530...
GRALKAHRLQFWKKALPQKIQELEEPEERHTEL
```

Legend

LOOP	loop secondary structure predicted by PSIPRED
HELIX	helix secondary structure predicted by PSIPRED
STRAND	strand secondary structure predicted by PSIPRED
DISORDER	disordered region predicted by DISOPRED2
LOW COMPLEXITY	low complexity region predicted by SEG
COILS	coiled coils region predicted by COILS
TRANSMEMBRANE HELICES	transmembrane helices predicted by TMHMM
SIGNAL PEPTIDES	signal peptites predicted by RPSP

Download

FASTA sequence	text file with protein sequence in FASTA format
Summary	text file with a summary of protein features

Figure 4.1. Output from the XtalPred server[142], with the input of the sequence of secreted CES2.
The most notable point from this output is that CES2 is classified as having a crystallization class of 4 (difficult), on a 1-5 scale. It also predicted it to be a highly helical structure.

Originally, four CES2 mutants were conceived and produced: wild type CES2, null, Dere and Phyre. Full details of these are given in *Chapter 2, table 2.1.* The rationale for the CES2 Dere and CES2 Phyre variants, derived from the Derewenda[1], and Phyre[143] analyses respectively, was that either reducing the surface charge of the protein or removing the potentially un-structured *C*-terminus, might facilitate crystallization. For the

CES2 Dere mutant, the surface-entropy reduction (SER) server[144] was used to identify surface residues with long polar side-chains and hence high entropic freedom, and to replace them with alanines. By mutating large, polar and often flexible, solvent exposed residues (e.g. lysine (K), glutamic acid (E) and glutamine (Q)) that are typically located on the surface of a protein , to smaller amino acids such as alanine (A)[1] with lower entropy side-chains, it was hoped that surface patches would be created that were conducive to forming thermodynamically favorable interactions, favouring crystal nucleation and growth[145]. Therefore, two sets of three highly entropic residues situated next to each other (Q360, K361, E362) and (Q513, E514, E515) were identified and mutated to alanines.

Using the Phyre server, a protein structure prediction tool[143] that involves the detection of homologs of known 3D structure, seven highly disordered amino acids were identified on the *C*-terminal end of CES2. These were deleted with the aim that the truncated protein would be more favorable to crystallization. The RONN disorder prediction tool[146] was in agreement with the Phyre server, also identifying seven disordered residues at the *C*-terminus of the protein.

4.2.2. Glycosylation Knock-Out Mutants

The two conserved high mannose *N*-linked glycans present in CES2, N111 and N276 (*see appendix 1.1*), could not be removed enzymatically using endoglycosidase H (endoH) without causing the protein to precipitate. The endoH enzyme cleaves between the two *N*-acetylglucosamine residues in the core of the oligosaccharide, generating a truncated sugar molecule with one *N*-acetylglucosamine residue remaining on the asparagine. Thus removal of the glycans after protein production and secretion destabilised the overall structure of the mature protein in solution. Therefore, genetic engineering was used to remove the *N*-glycosylation acceptor asparagines (N111 and N276) by mutating them to

glutamine residues to block any glycosylation of the protein. This resulted in three mutants designed to assess the role of glycosylation in CES2: CES2 N111Q, CES2 N276Q and the double mutant CES2 N111Q+N276Q, which were used in crystallization, as well as in kinetic and biophysical experiments.

4.2.3. Oligomerization Mutants

In contrast to CES1 which forms trimers and readily crystallizes, CES2 behaves as a monomer as shown by work documented here (*figure 4.4. and biophysical characterization*), as well as in previous protein production publications[96], and has proved difficult to crystallize. Therefore it was decided to engineer the two salt bridges critical in the CES1 trimer formation, which were identified in this study (*Chapter 3*), into CES2. The reasoning for doing this was to attempt to promote trimer formation, which might ultimately drive forward the crystallization process. By examining the trimeric arrangement of CES1, it was shown that oligomerisation was mediated by two charge clamps across the trimer interface. These charge clamps, or salt bridges, were formed between K78 + E183 and K275 + E291. From alignment with CES1, to introduce these salt bridges into CES2, assuming a similar structural disposition of side-chains, required the following mutations in the CES2 sequence R86K, K190E, D281K and K297E (*table 4.1*).

	Residues involved in CES1 trimer formation			
	Lys 78	**Glu 183**	**Lys 275**	**Glu 291**
CES1	FV**K**NA	GD**E**HS	GC**K**TT	KT**E**EE
CES2	GV**R**DG	GD**K**HA	AC**D**QV	KS**K**EE

Table 4.1. Localised sequence alignment of the residues critical in salt bridge formation between monomers in CES1.
Resides involved in CES1 salt bridge formation that stabilised the trimeric interaction are highlighted and shown in bold. Residues in CES2 that were mutated to induce salt bridge formation are shown in blue.

Whilst producing these new constructs, the fourth mutation in CES2, K297E, proved problematic to engineer, and even though the mismatched PCR primers generated amplified products, sequencing revealed that no mutations had been introduced and only the wild type parent sequence was obtained. On the basis that introducing a glutamate residue next to two others in the sequence was causing the mutagenesis problem, primers were re-designed, and the glutamate codon altered from gaa → gag (forward) and ctt → ctc (reverse) (*see Primers for mutagenesis, Chapter 2*). This small change enabled the K297E mutant to be generated.

In the meantime, it was decided to take the triple mutant that had been created (R86K, K190E, D281K), assigned the name **CES2 3-4,** through into transient protein production, as it contained one CES1 salt bridge pairing that may have been enough to drive forward the trimer formation. When the mutant with the four mutations had been created, **CES2 4-4**, this too was also taken forward to protein production.

4.3. Production of Human Recombinant CES2

4.3.1. Expression of Human Recombinant CES2

Mutants of CES2 investigated in this thesis were produced both through transient expression in HEK293T and stable expression in HEK293 Gnt$^{-/-}$. Protein production was initially tested by small-scale transient expression in HEK293T and the expression levels of different constructs and mutants were compared. Expressed proteins were detected in both cells and media by western blotting using an anti-His6 antibody. The effect of replacing the CES2 wt signal sequence (pOPINE vector) with a μ-phosphatase leader sequence (pOPING vector) was investigated. No marked differences in expression from the two vectors was observed (*figure 4.2*), and both produced large yields of protein. Addition of both the CD4 tag (pOPINTTGneo CD4) and HALO tag (pOPINTTGneo HALO) reduced CES2 expression, as for CES1 (see *figure 3.1*).

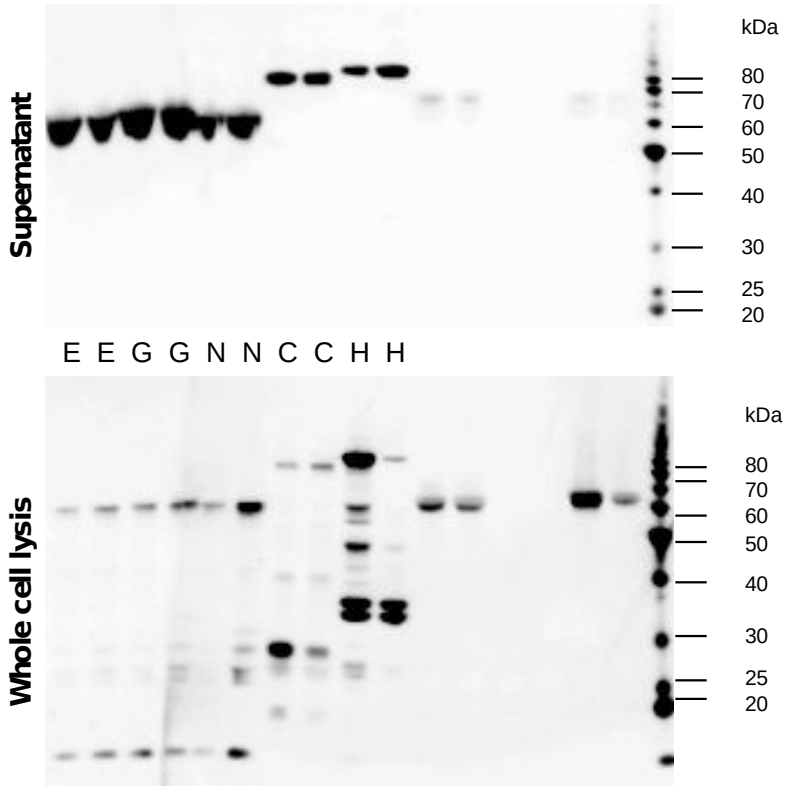


Figure 4.2. Western blot analysis for protein expression of different constructs of CES2 wt. Small- scale transient expression in HEK293T. DNAs were transfected in duplicate and protein expression analysed by SDS-PAGE/Western blotting using an anti-His6 monoclonal antibody. The top image shows detection from the cell media, the bottom is from whole cell lysates. E = pOPINE, G = pOPING, N = pOPINTTG neo, C = pOPINTTG neo CD4 and H = pOPINTTG neo HALO. CES2 runs at ~58kDa; the CD4 tag adds ~30kDa and the halo tag adds ~35kDa to the size of the constructs. These blots show that the majority of CES2 was secreted into the media. Intra-cellularly, cleavage of the CD4 and halo tags was observed as well as protein degradation.

Although transient expression of CES2 was sufficient for downstream processes, stable HEK293 gnt$^{-/-}$ cell lines were still produced for CES2 wt and null (S228A) to prevent the need for multiple transient transfections during the course of the project. For CES2 wt, 44 G418-resistant colonies expressing protein were picked (*figure 4.3*), and for CES2 null, 5 were picked. Those giving the highest expression signals were frozen down and stored at -80 °C. Clonal cell lines were expanded in roller bottles and secreted proteins were purified from the cell media.

During the progression of the project, when several CES2 mutants needed to be created, these were produced transiently, as stable cell lines for production would have taken too much time to generate.

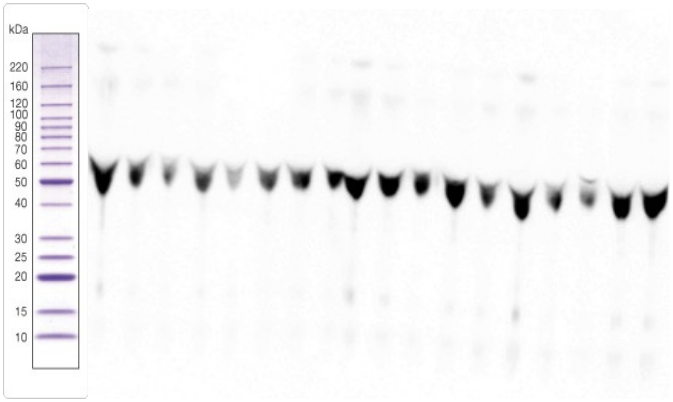


Figure 4.3. Western blot analysis of 18 stable cell line clones secreting CES2 wt cloned into pOPINTTG neo.
Due to the significant amount of secreted protein produced from a 24-well plate and loaded onto the gel, the protein bands are distorted and don't run at accurate weights. The original protein marker ladder is not shown as it was overloaded. In all clonal expansions, the protein secreted was extremely pure and no degradation was observed.

4.3.2. Purification of Human Recombinant CES2

All recombinant CES2 enzymes were recovered from the media of transfected HEK293 cells by IMAC and further purIfied by size exclusion chromatography. All mutant forms of CES2 (wt, null, Dere, Phyre, N111Q, N276Q, N111Q+N276Q, 3-4 and 4-4) eluted

from the size exclusion chromatography stage as a major peak of approximately 60 kDa, corresponding to a monomeric species. Smaller peaks corresponding to larger aggregates (> 500 kDa), and an oligomer heavier than the monomeric species that eluted broadly over 10 mLs were also observed (*figure 4.4*). This eluted between 60 and 70 mls, which corresponds to a molecular weight from 500 to 150 kDa, therefore an accurate size could not be assigned to this peak. It is important to note that this 'middle' peak was not observed in any of the three glycosylation knock out mutants. These molecular weights were deduced from standard calibration curves using proteins of known mass (ferritin, aldolase, conalbumin and ovalbumin). Interestingly, Lamego et al.[96] reported that they had also observed CES2 form oligomers, potentially dimers, under non-denaturing conditions.

Eluted fractions containing the protein were analysed by SDS-PAGE, pooled and concentrated before use in subsequent studies. All proteins were highly soluble in aqueous buffer, apart from the glycosylation mutants, and CES2 wt could readily be concentrated to 55 mg/ mL. The glycosylation mutants only reached a maximum concentration of ~ 20 mg/ mL, indicating that glycosylation modulates protein solubility. All proteins were flash-frozen and stored at -80°C for long-term storage. After one freeze-thaw cycle, all three glycosylation mutants showed substantial degradation and a heavy cream-coloured precipitate was observed. The other mutants were more stable, but small amounts of degradation were still seen.

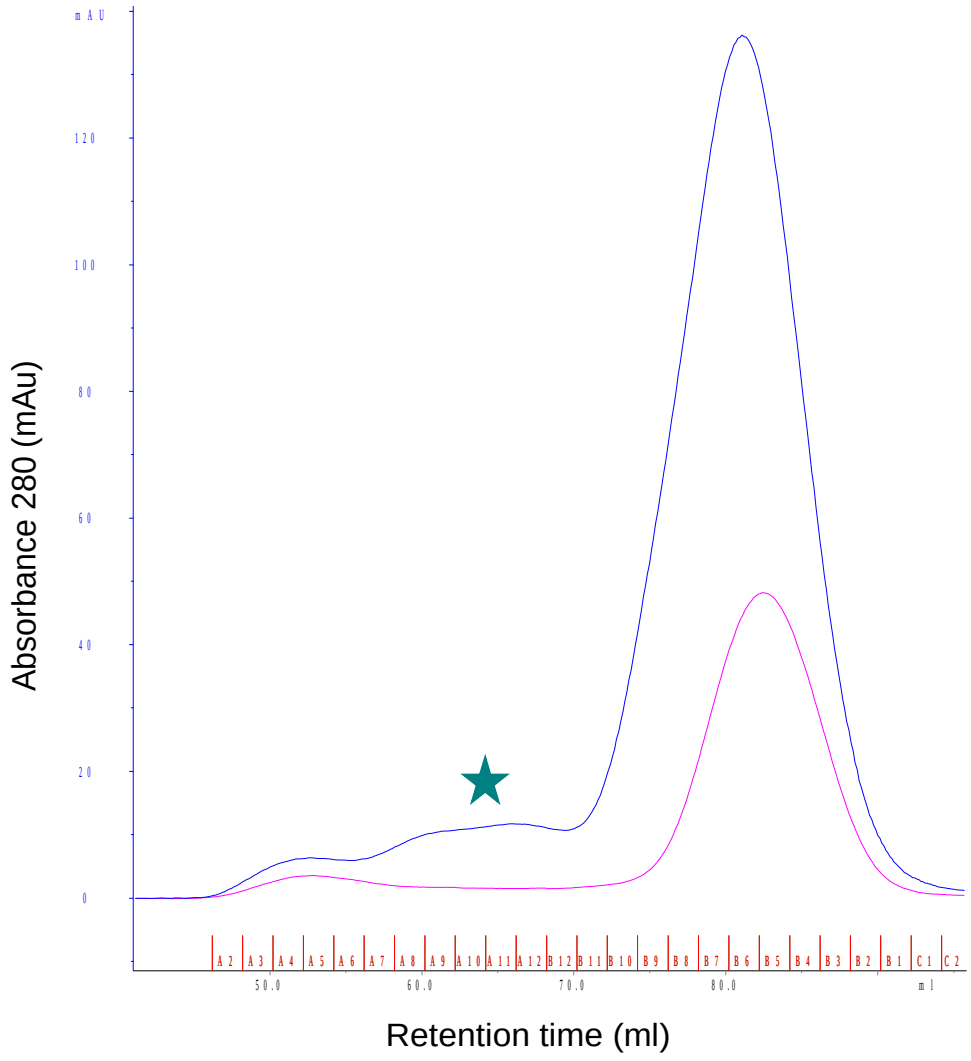


Figure 4.4. CES2 size exclusion elution profile from a 16/60 Superdex 200 column (GE Healthcare.
Elution profiles of CES2 native (blue) and CES2 N276Q (magenta) have been overlaid as they were run on the same ÄKTAxpress purification system. The main species eluted for both mutants was a monomer (~60 kDa, retention time of ~82ml). Two other possible species of higher molecular weight eluted prior to this main peak. Interestingly, with the N276Q mutant, the middle species was not observed (indicated with a star). All other forms of CES2 gave the same elution profile, but have not been aligned as they were run on different purification systems. Wells that the fractions were eluted in are shown in red on the bottom axis.

4.3.3. Optimisation of Purification

All CES2 enzymes that were produced presented problems during their two-step purification. Very low protein yields were recovered from the media using the standard purification protocol (see *Chapter 2, 2.2.4.5.)*, and upon testing the flow through, large quantities of protein were still present (*see figure 4.5*). This indicated that there was a

problem with the binding of the histidine tag to the nickel column. After multiple consecutive loadings of the flow-through, protein binding improved and the yield recovery increased. However, this was a time-consuming procedure and therefore a series of troubleshooting experiments were carried out during the course of the research project with different batches of protein (*these are summarized in Chapter 2, table 2.3*). It was particularly important to optimise this process as many of the mutants were produced through transient expression, which produces lower yields than stable cell lines, and recovery of as much protein as possible was necessary for downstream processes.

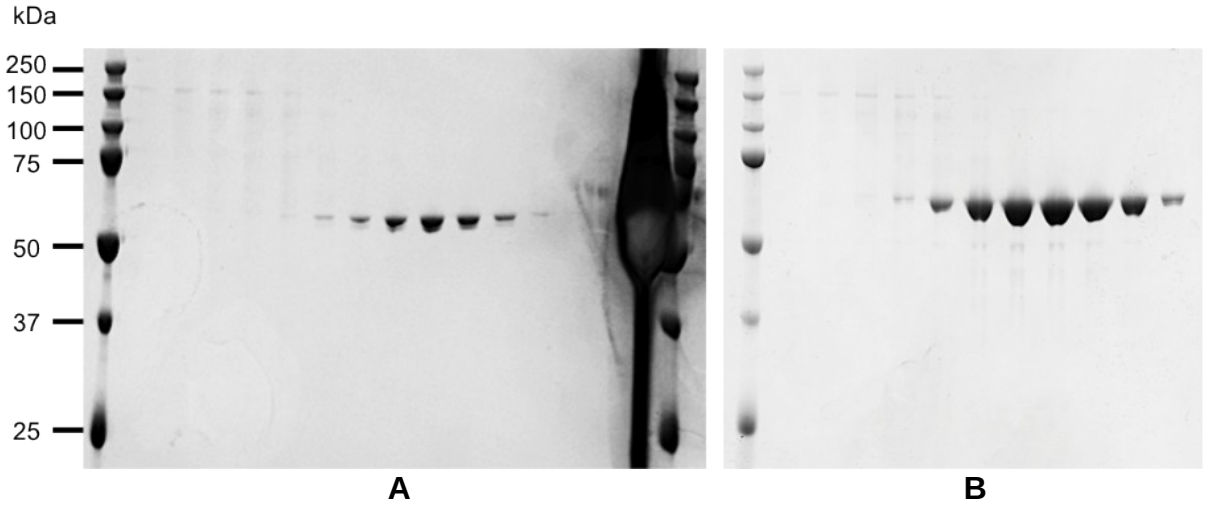


Figure 4.5. CES2 3-4 gel filtration elution profiles, before and after the addition of nickel sulphate to the cell media.
In the left hand gel (A), the diluted flow through is loaded in the last lane on the right. The right hand gel (B) shows the elution profile after optimization with nickel. In this gel, the flow through is loaded in the first lane on the left, after the ladder. The addition of nickel sulphate clearly improves binding of CES2 to the nickel column.

Of all the trouble shooting strategies carried out, only the addition of nickel sulphate (3-5 mM) coupled with the lowering of the imidazole content (15 mM → 5 mM) of the wash buffer significantly improved the protein yield recovered. Dilution of the DMEM media with PBST (1:1) prior to IMAC made a slight improvement. Presumably, the complex DMEM media in which the HEK cells were grown, contained metal chelators that

competed for binding to the nickel column. By adding small amounts of nickel to the harvested media, these chelators are neutralized (although small amounts of protein may be lost), resulting in higher binding rates of the protein and hence an increase in yield. Decreasing the imidazole concentration in the wash buffer also contributed to higher yields presumably by preventing loss of bound CES2 during the wash step.

The yields of the recombinant CES2 enzymes are shown in Table 4.2. Protein production yields from clonally selected stable cell lines were up to 84 fold greater than previously reported for production from insect cells (1mg/ L)[57] showing the benefits of investing time in producing stable cell lines. From transient expression, yields were between 5 to 36 fold better than in insect cells, demonstrating the benefits of using mammalian cells in general. The CES2 glycosylation mutants along with 3-4 and 4-4 were produced transiently in HEK293T cells in the presence of kifunensine which gives a comparable high mannose containing *N*-glycoform as the HEK293Gnt$^{-/-}$ cells used for the stable CES2 native and null mutant cell lines.

The results also show that removal of both of the glycosylation sites had no significant effect on protein expression and secretion, as yields obtained from transient expression were comparable to both glycosylated CES2 native and null mutant enzymes, similar to what was observed for the unglycosylated CES1 mutant (*Chapter 3*).

Protein	Transient	Yield (mg/ L)	Stable Cell	Yield (mg/ L)

	Expression		Expression	
CES2 wt	Yes	5.4*	Yes	84.0**
CES2 null	Yes	5.6*	Yes	80.0**
CES2 Dere	Yes	28.0**	-	-
CES2 Phyre	Yes	29.6**	-	-
CES2 N111Q	Yes	26.0**	-	-
CES2 N276Q	Yes	21.2**	-	-
CES2 N111Q+N276Q	Yes	23.7**	-	-
CES2 3-4	Yes	36.0***(22.8**)	-	-
CES2 4-4	Yes	24.8**	-	-

Table 4.2. Comparison of protein yields obtained from transient expression in HEK293T and stable expression in HEK293 Gnt^{-}*I*$^{-}$.
*Media run only once through ÄKTAxpress purification system. **Media run multiple times after discovering protein was still present at high quantities in the flow through. *** After addition of $NiSO_4(H_2O)_6$ combined with lowering the amount of imidazole in the wash buffer.

4.4. Activity Analysis

There is very limited data on the kinetics of CES2 available in the public domain. The data that have been reported come from assays using both purified recombinant protein[133] and whole cell lysates[17, 96, 147], with the majority from the later. Therefore, in this project, a comprehensive activity analysis was performed to fully understand the kinetic mechanism of CES2, along with investigating the role of the two *N*-glycans present on the enzyme. Activity analysis experiments were carried out on the following CES2 mutants: wt, null, N111Q, N276Q, N111Q+N276Q, 3-4 and 4-4.

The single serine knockout mutant, S228A, demonstrated no catalytic activity as expected, as was seen with CES1 null. All the other CES2 mutants hydrolyzed the substrate 4-NPA but activity varied between the different forms of the enzyme, and some interesting observations were made. Initial velocity rates (AU min^{-1}) are presented in Table 4.3.

(a)

	Enzyme Concentration (nM)				
4-NPA (mM)	**42.40**	**21.20**	**10.60**	**5.31**	**2.65**
0.0938			0.1021	0.0415	0.0235
0.1875		0.1745	0.1256	0.0652	0.0299
0.375		0.2111	0.1423	0.0913	0.0442
0.75	0.5088	0.2909	0.1736	0.0962	0.0406
1.5	0.6309	0.4232	0.2053	0.1013	0.0441
3	0.5191	0.4048	0.2197	0.1016	0.0458

(b)

	Enzyme Concentration (nM)					
4-NPA (mM)	**84.80**	**42.40**	**21.20**	**10.60**	**5.31**	**2.65**
0.0938		0.0844	0.0632	0.0298	0.0122	0.0049
0.1875		0.1188	0.0801	0.0357	0.0151	0.006
0.375		0.1497	0.0827	0.0409	0.0159	0.0071
0.75	0.4199	0.1647	0.0969	0.0434	0.022	0.0087
1.5	0.4908	0.2243	0.1253	0.0547	0.0246	0.0122
3	0.5192	0.2719	0.137	0.0597	0.0275	0.0124

(c)

	Enzyme Concentration (nM)				
4-NPA (mM)	**42.40**	**21.20**	**10.60**	**5.31**	**2.65**
0.75	0.373	0.226	0.116	-	-
1.5	0.4025	0.2501	0.1437	0.0653	0.0316
3	0.4448	0.2712	0.1313	0.0636	0.0301

(d)

	Enzyme Concentration (nM)					
4-NPA (mM)	**84.80**	**42.40**	**21.20**	**10.60**	**5.31**	**2.65**
0.375	0.198	0.1109	0.0373	-	-	-
0.75	0.387	0.1475	0.0486	-	-	-
1.5	0.5022	0.1584	0.0617	0.0335	0.0178	

(e)

	Enzyme Concentration (nM)				
4-NPA (mM)	**42.40**	**21.20**	**10.60**	**5.31**	**2.65**
0.0938		0.1012	0.0711	0.0315	0.0153
0.1875		0.1435	0.0879	0.0392	0.016
0.375	0.328	0.1931	0.1004	0.0463	0.0185
0.75	0.4186	0.2423	0.1211	0.0557	0.0238
1.5	0.641	0.3047	0.1593	0.0732	0.0314
3	0.6488	0.3726	0.1907	0.0854	0.0378

(f)

	Enzyme Concentration (nM)				
4-NPA (mM)	**42.40**	**21.20**	**10.60**	**5.31**	**2.65**
0.0938			0.0736	0.0349	0.0169
0.1875		0.1329	0.0797	0.0414	0.0186
0.375	0.3234	0.194	0.1046	0.0507	0.0233
0.75	0.3836	0.235	0.1306	0.0607	0.028
1.5	0.5165	0.323	0.1611	0.0769	0.0363
3	0.597	0.3762	0.1946	0.0965	0.0423

Table 4.3. Initial velocity rates for all CES2 mutants given in absorbance units per minute (AU min^{-1}).
CES2 wt (*a*), CES2 N111Q (*b*), CES2 N276Q *(c),* CES2 N111Q+N276Q *(d),* CES2 3-4 *(e)* and CES2 4-4 *(f).* All r^2 values exceeded 0.99. Values that have been omitted had an r^2 of less than 0.95. r^2 is the coefficient of determination, and it is a measure of the goodness-of-fit of simple linear regression, in this case.

4.4.1. Observation of Non-Michaelis Menten Kinetics

For all of the functionally active CES2 mutants assayed, a classic Michaelis Menten hyperbolic curve was not seen when plotting substrate vs initial velocity (AU min^{-1}) graphs (*figure 4.6, a*). Interestingly, neither was the sigmoidal shaped plot that was shown by CES1 (*Chapter 3*) observed for CES2. Both the Lineweaver Burk and Eadie-Hofstee plots, linear rearrangements of the Michaelis Menten equation, do not resemble classical plots, reinforcing the idea that this enzyme does not exhibit classic kinetics.

In the double reciprocal Lineweaver Burk plots, the two highest protein concentrations shown (21.2 nM and 10.6 nM) gave plots that curved upwards, and the two lowest concentrations (5.31 nM and 2.65 nM) gave plots that curved downwards. In the Eadie-Hofstee plot, a graphical representation of kinetics in which velocity is plotted as a function of the ratio between velocity and substrate concentration, the opposite pattern is seen. There are sharp transitions between apparently linear phases in all of the plots.

Using GraphPad, no single statistical model could be fitted successfully to the data for any of the CES2 enzymes. Models tried included Michaelis Menten, allosteric sigmoidal, two-site hyperbola, substrate inhibition and biphasic.

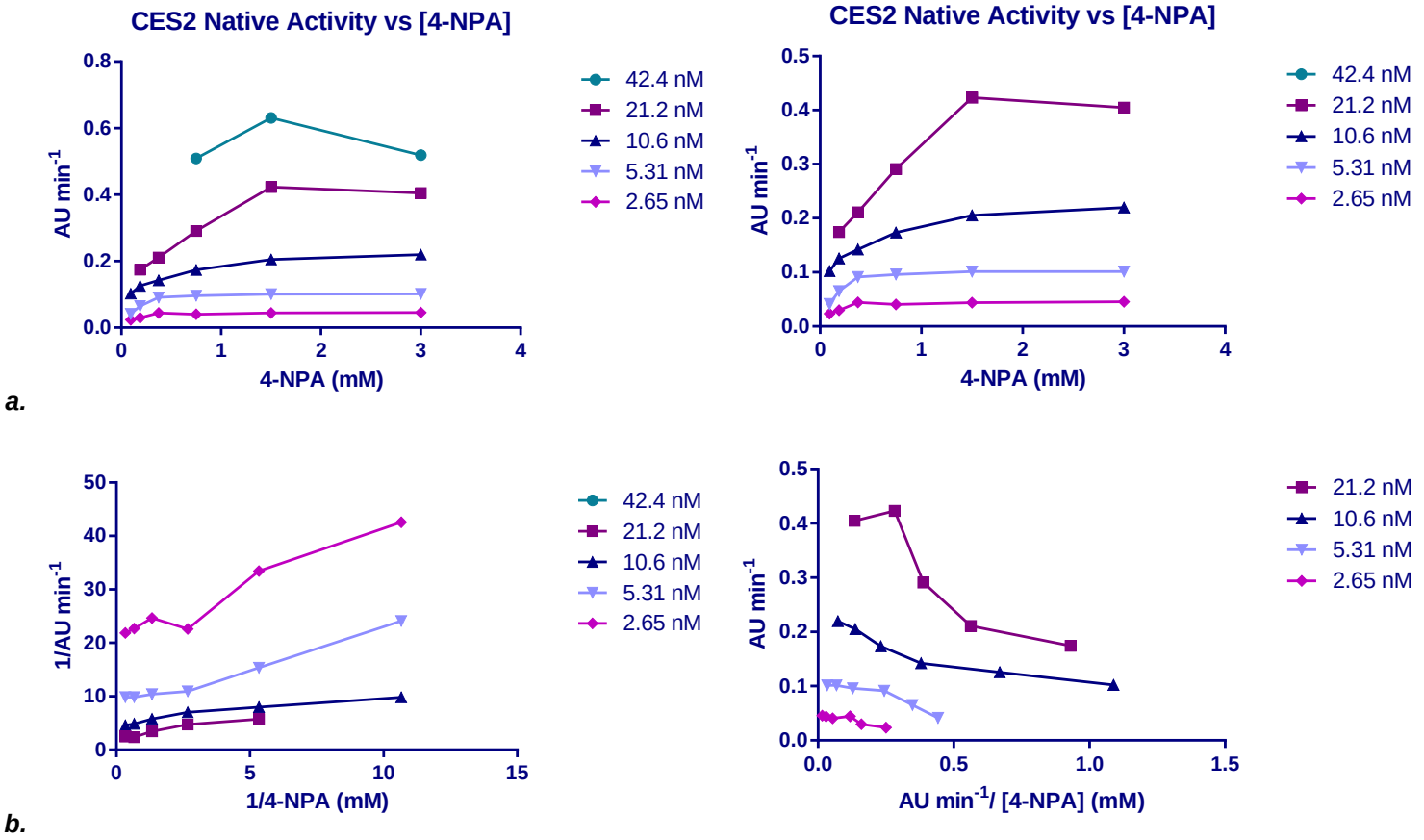


Figure 4.6. Kinetic Analysis of CES2.
a) Both figures show substrate vs velocity plots for CES2 wt. The left hand plot shows all concentrations of protein, but the right hand plot does not contain the highest concentration of enzyme (42.4 nM) for clarity. ***b)*** The plot on the left is a Lineweaver Burk (double reciprocal) and the plot on the right is an Eadie-Hofstee. Values for the highest enzyme concentration (42.4 nM) were omitted. Because all mutant forms gave similar profiles, only graphs for CES2 native are shown for simplicity. The varying concentrations of enzyme that were assayed are indicated on the right hand side of each graph. Although the straight lines joining the data points together do not fully show the unusual shapes of the graphs, no computational model was able to fit the data and so smooth curves could not be generated.

4.4.2. Explanations for Atypical Kinetics

Possible explanations for these atypical kinetic profiles include substrate inhibition, cooperative kinetics or biphasic kinetics, some of which were proposed by Korzekwa et al[148] in 1998 from studies of cytochrome P450. This issue is further complicated by the fact that CES2 may exhibit more than one of these phenomena, giving rise to the rather complex Lineweaver Burk and Eadie-Hofstee plots that are seen.

4.4.2.1. Substrate Inhibition

At a particular substrate concentration, the reaction velocity actually decreases as the concentration of the substrate is further increased, resulting in a convex reaction profile. This is seen for the highest enzyme concentrations, 42.4 and 21.2 mM (*figure 4.6, b*). This is generally only seen at high substrate concentrations, and in the data presented here, this is seen progressively as substrate concentration increases. Although substrate inhibition complicates the graphical presentation of initial velocity data, its observation is an excellent diagnostic tool for the elucidation of the kinetic mechanism. It has been shown in other enzymes that substrate inhibition has important biological functions[149], including acetylcholinesterases.

4.4.2.2. Cooperative Kinetics/ Allosterism

Interaction between subunits in an oligomeric protein or between subunits in a monomeric unit could lead to increased substrate affinity (positive cooperativity), or decreased substrate affinity (negative cooperativity). This may be the case if CES2 contains a regulatory binding site for the substrate, as has been observed for CES1 (discussed in Chapter 3).

4.4.2.3. Biphasic Kinetics

As with CES1, in the Lineweaver-Burk plots of CES2, it appears that the lines for different protein concentrations lie almost parallel to each other, and would not converge if extrapolated indicative of ping-pong reaction mechanism. As discussed in *Chapter 3*, this has been noted for CES1 before, but not for CES2. Examples of biphasic kinetics are becoming more prevalent, with several P450 isoforms exhibiting this behaviour[150]. The shape of the curves for 21.2 nM and 10.6 nM in the Eadie-Hofstee plot, is also an indicator of biphasic kinetics.

With such a variety of potential explanations for non-linear Lineweaver Burk plots, distinction between the alternatives can prove difficult. From examining the Eadie-Hofstee plots, the kinetics of CES2 may be explained by a combination of all three of these: substrate inhibition[150], positive cooperative and biphasic kinetics, as occurs with CES1. However, these three phenomena may be masking each other, leading to the uncharacteristic and ambiguous kinetic profiles.

There are some data in the scientific literature that lends support to the complex CES2 kinetics described above. In the 2001 Williams et al. paper[111], the substrate specificities of both CES1 and CES2 from human, the beagle dog and the cynomolgus monkey were analysed and the authors concluded that the enzymes tested obeyed Michaelis Menten kinetics, with a few exceptions. Both human and monkey CES2 exhibited substrate inhibition at high concentrations and hydrolysis of one substrate, 4-nitrophenyl dimethylacetate (4-NPDMA) but not 4-NPA, was best fitted to biphasic or two-site kinetics. Substrate inhibition was observed using a single concentration, 600 μM, which was five times less than the maximum concentration used in the experiments presented

here. The authors also described CES2 as demonstrating 'non-saturable kinetics' for which they used the Hill equation.

CES2 wt and mutant activity assays were carried out on several different batches of enzyme, and at different times during the project. All processed data gave the atypical data and Lineweaver Burk Plots presented above, strongly indicating that the observed kinetics are indeed real, and reflect the complex enzymatic mechanism of CES2.

4.4.3. Specific Activity

The rate of product formation was expressed as nanomole of pNP produced per minute per microgram of protein (nmol min^{-1} μg^{-1}), using the extinction coefficient for pNP at 405 nm (18,000 M^{-1} cm^{-1}). Unlike with CES1 where specific activity calculated was essentially the same regardless of substrate or protein concentration, the specific activity of CES2 declined as substrate concentration increased (*table 4.4, figure 4.7*). This unexpected observation is further evidence of concentration dependent substrate inhibition of CES2. Therefore when comparing activity of the different CES2 mutants, it is essential to look at activity rates at the same substrate concentration.

In the paper by Wang et al.[133], they report specific activity rates of 0.682 μmol mg^{-1} min^{-1} for 4-NPA, and 13.8 μmol mg^{-1} min^{-1} for fluorescein diacetate (FD) using recombinant CES2, with a substrate concentration of 0.5 mM. However their data differs significantly from data presented in this thesis. Their CES2 protein is 10 fold less active than the recombinant protein produced in this work, and they state that for 4-NPA, they observed that their recombinant CES1 was more than twice as active as their CES2, opposite to what is observed here. The specific activity of CES2 cannot be compared with previous

kinetic data published by Williams et al.[111], as they only reported K_m values at a single substrate concentration.

4.4.3.1. Effect of Glycosylation

Using 3 mM 4-NPA, CES2 native demonstrated a hydrolytic rate of 6.0 nmol min^{-1} μg^{-1}. This was over 3 fold greater than the N111Q mutant, and 1.5 fold greater than N276Q mutant. This pattern was consistent between substrate concentrations, and at 0.75 mM 4-NPA, the catalytic activity of the native enzyme was 4 fold greater than N111Q, 1.6 x N276Q and 4.6 x N111Q+N276Q.

These results indicate that the first conserved glycosylation site on CES2, N111, has a greater involvement in the catalytic activity of the enzyme compared to the second site, N276, when mutated to Q. These two sites together have an additive effect on the activity of the enzyme, as the double knock out mutant showed the largest decrease in activity. Occupancy of both *N*-glycosylation sites appears to be required for maximal CES2 catalytic activity. These effects on activity seen by knocking out one/ both glycosylation sites in CES2 are in agreement with previous studies on other drug metabolizing enzymes [151].

4.4.3.2. CES2 3-4 and 4-4

At the highest concentration of 4-NPA (3 mM), activity between CES2 native, 3-4 and 4-4 was essentially the same. As substrate concentration increases, the specific activity of CES2 wt, 3-4 and 4-4 mutants decreases. At lower substrate concentrations the native enzyme appears to be marginally more active than the two mutants, for example at 1.5 mM 4-NPA, the native CES2 was 1.2 x more active and at 0.375 mM, it was 1.7 x more active than the 3-4 and 4-4 CES2 mutants. This may indicate that one or more of the three

mutations, R86K, K190E or D281K has an inhibitory effect on catalysis at low substrate concentrations.

Mutant	[4-NPA]/ mM	Specific Activity (nmol min^{-1} μg^{-1})	Mutant	[4-NPA]/ mM	Specific Activity (nmol min^{-1} μg^{-1})
CES2 native	0.0938	88.17 ± 9.16	**CES2 3-4**	0.0938	63.50 ± 5.50
	0.1875	59.35 ± 2.47		0.1875	39.39 ± 3.22
	0.375	39.64 ± 5.15		0.375	23.34 ± 0.66
	0.75	20.87 ± 1.76		0.75	14.36 ± 0.004
	1.5	11.79 ± 0.91		1.5	9.32 ± 0.21
	3.0	6.00 ± 0.44		3.0	5.59 ± 0.09
CES2 N111Q	0.0938	24.97 ± 5.15	**CES2 4-4**	0.0938	66.65 ± 2.84
	0.1875	16.74 ± 2.35		0.1875	38.49 ± 0.99
	0.375	8.53 ± 1.35		0.375	23.48 ± 1.16
	0.75	5.22 ± 0.71		0.75	14.27 ± 0.92
	1.5	3.29 ± 0.37		1.5	9.21 ± 0.45
	3.0	1.87 ± 0.17		3.0	5.68 ± 0.10
CES2 N276Q	0.0938	-	**CES2 N111Q +N276Q**	0.0938	-
	0.1875	-		0.1875	-
	0.375	-		0.375	5.62 ± 1.10
	0.75	12.73 ± 1.46		0.75	4.33 ± 1.43
	1.5	7.79 ± 0.50		1.5	2.40 ± 0.76
	3.0	3.89 ± 1.28		3.0	-

Table 4.4. Specific activity values of CES2 mutants expressed in nmol min^{-1} μg^{-1}.
Values are the mean ± standard deviation of 3-5 determinations from differing protein concentrations, after normalization. Values omitted could not be accurately verified, either because specific activity was too low to be measured, or because a minimum of two replicate readings could not be obtained.

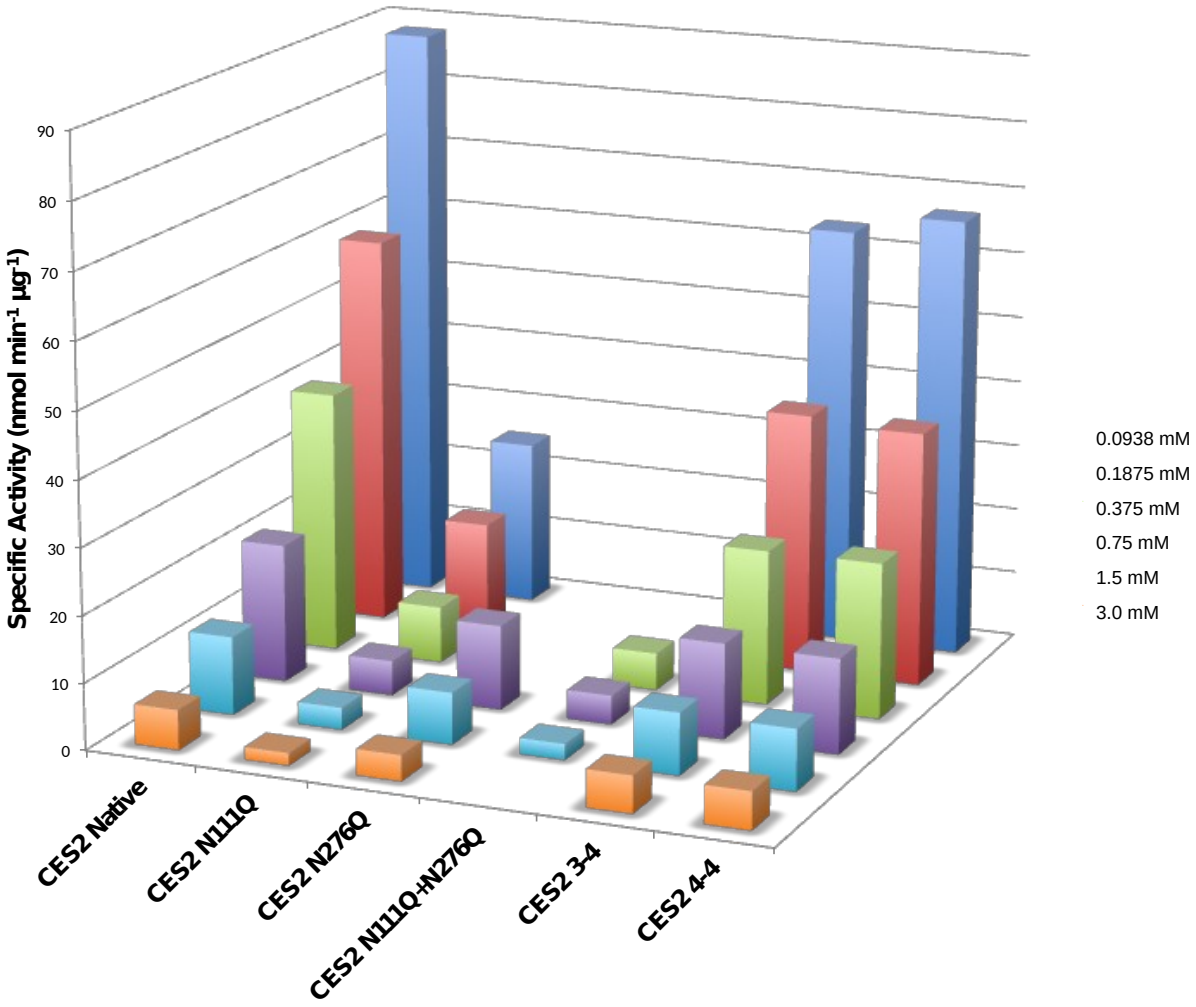
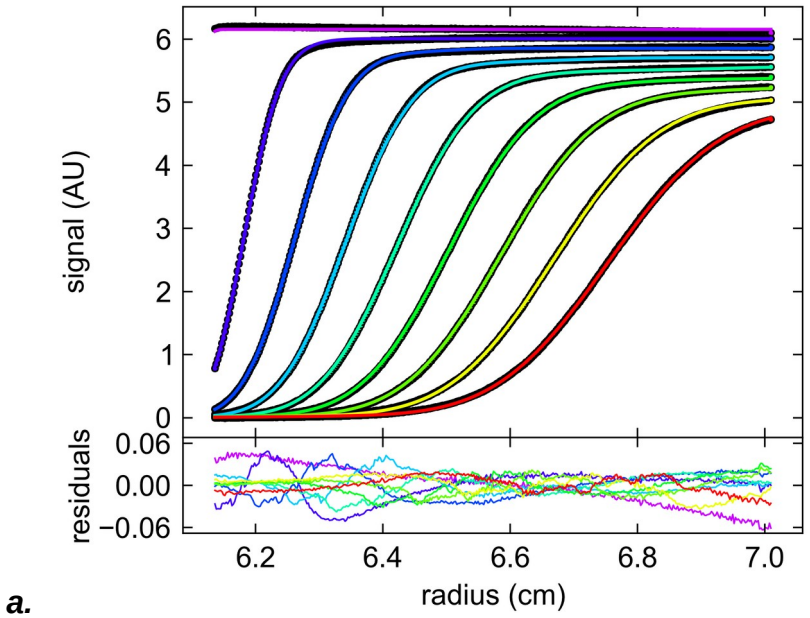


Figure 4.7. 3-D Column Chart Showing Specific Activities of the CES2 mutants. Concentrations of 4-NPA are shown in the figure legend on the right of the graph.

4.4. Biophysical Analysis

Sedimentation velocity experiments showed that all forms of CES2 were predominantly monomeric (*figure 4.8*), with an apparent molecular weight of 59 kDa, which contrasts with the trimeric assemble of CES1. There is also the possible association of a dimer at higher molecular weight. The sedimentation coefficient (*s*) was 4, and the frictional ratio (*f*) was 1.34.



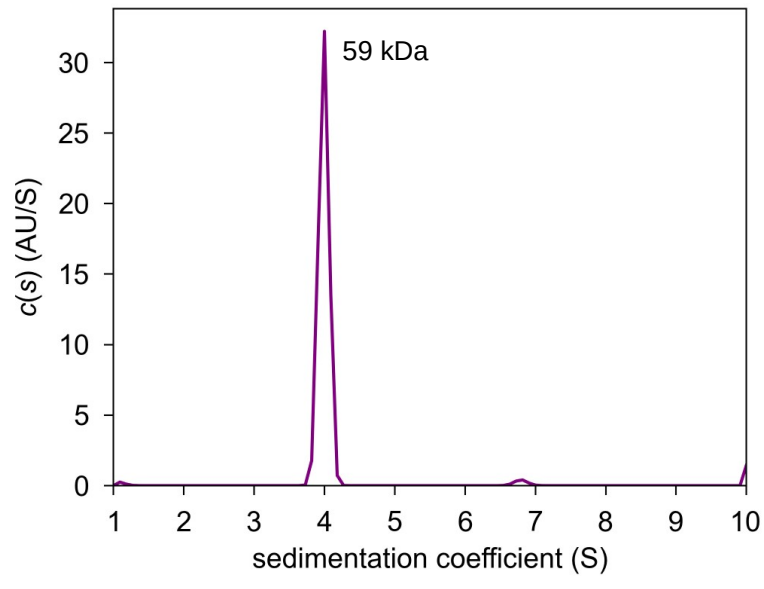


Figure 4.8. Graphical AUC analysis for CES2 3-4.
The plot of the residuals is shown in ***a,*** and the sedimentation coefficient population distribution (c(s)) against s plot is shown in ***b.*** Data for all CES2 mutants gave the same profiles. The main peak corresponds to a protein with a molecular weight of 59 kDa.

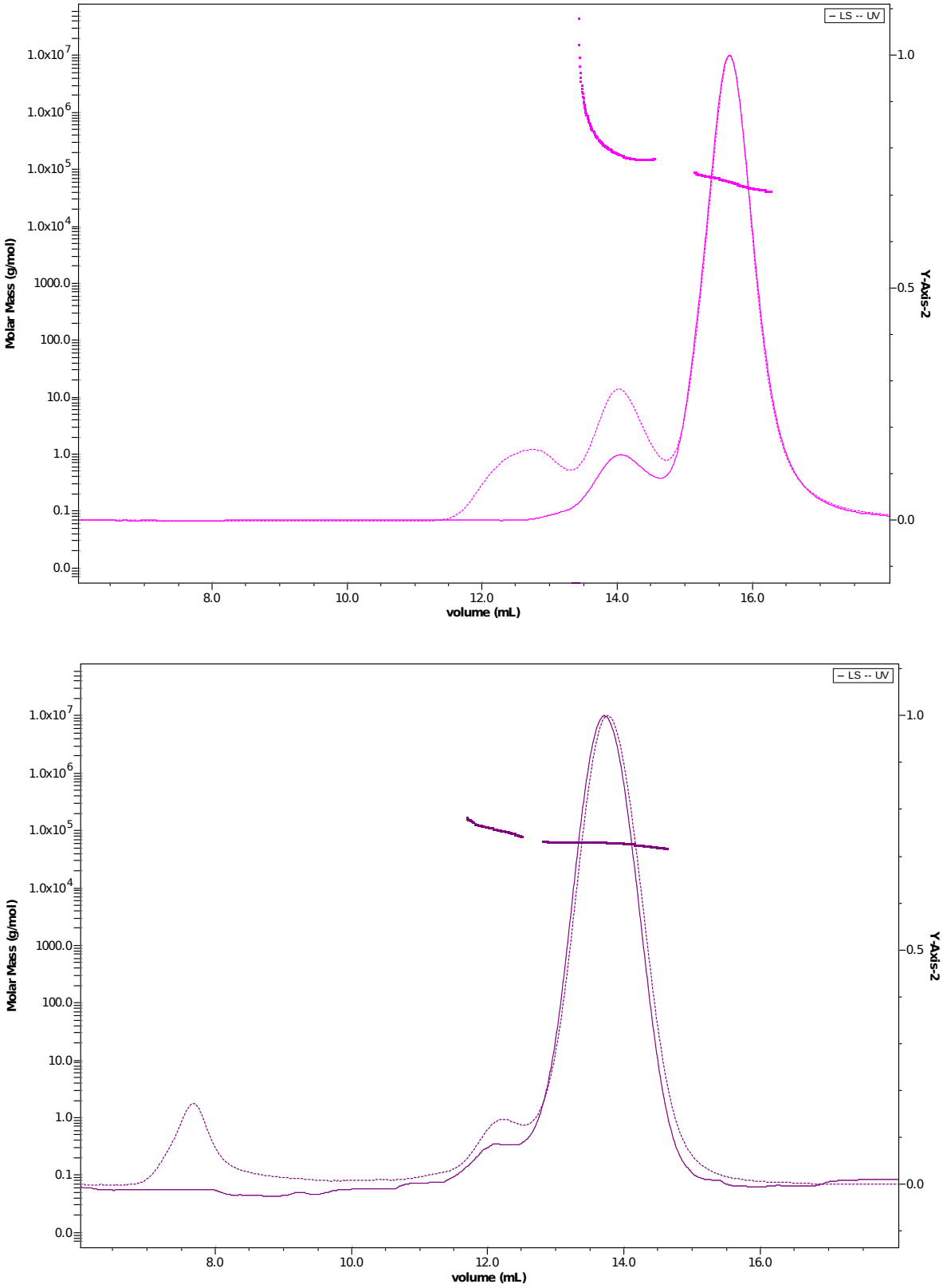


Figure 4.9. Molar mass vs volume SEC-MALLS plots for CES2 wt and CES2 3-4. CES2 wt is the top graph, 3-4 Is the bottom. Both were run at 25 °C with a flow rate of 0.3 mL/ min. The main peak in CES1 wt corresponds to a molecular weight of 57.9 kDa, with the main peak in CES1 N79Q at 58.8 kDa.

The SEC-MALLS results (*figure 4.9*) are consistent with the AUC data as well as the gel filtration elution profiles, assigning the main peak a molecular weight of 57.9 kDa. Again, association of the monomers to form a dimeric subunit was also observed. In CES2 wt, protein aggregation at a much higher molecular weight is seen.

4.6. Crystallization of CES2

At the start of the project, there had been no reports in the scientific literature of the crystallization of human CES2. Therefore, rather than using only the wild type enzyme, a number of mutants were created with the help of bioinformatics tools (*Chapter 2, 2.2.1*), with the aim of promoting crystallization of the protein. However none of the engineered CES2 proteins (Dere and Phyre, and unglycosylated mutants) gave crystals so additional post-protein production modifications were evaluated as described in the following section.

4.6.1. Post- Protein Production Modifications

A variety of commonly used protein modifications and rescue remedies were applied to CES2 to try and render it more amenable to crystallization. These experiments were performed over a period of two years on CES2 wt and all other mutants (excluding CES2 3-4 and 4-4) in many combinations, so a vast number of crystallization conditions were set up (over 500 96-well crystallization trays). Full details are described in Chapter 2, *2.4.3.1*.

To summarize, disappointingly, this large undertaking of work produced not one single crystal hit. Limited proteolysis, lysine methylation, cross-seeding with CES1 crystals and co-crystallization with inhibitors and CPT-11 all failed to provide data that could direct further crystallization experiments that could lead to optimised crystals. This really did illustrate that CES2 is refractory to crystallization. However, all screens were periodically assessed, and finally some hope was obtained from a single small crystal that started to appear approximately 5 to 6 months after being set up from a native glycosylated form of CES2.

4.6.2. Early Crystal Hits: Some Hope

Potential crystals of native, glycosylated CES2 started to appear five to six months after the crystallization trials were set up at 8mg/ mL at room temperature, in two conditions; PACT *premier*™ C7 (0.2 M sodium chloride, 0.1 M HEPES, 20% w/v PEG 6000, pH 7.0) and JCSG-*plus*™ G2 (0.02 M Magnesium chloride hexahydrate, 0.1 M HEPES pH 7.5, 22% w/v Poly (acrylic acid sodium salt) S100)[117]. No crystals were observed in the same screening conditions for any other CES2 mutants, or in the same conditions which were set up at 4°C. These crystals were cryo-protected using 20% glycerol and were screened for diffraction on beam line I04 at Diamond Light Source. The crystal from PACT C7 (*figure 4.10*) showed diffraction spots to ~ 9.0 - 10 Å, which allowed it to be characterized potentially as presenting trigonal or hexagonal symmetry with a unit cell of a = b = 169.1 Å, c = 92.6 Å $\alpha = \beta = 90°$, $\gamma = 120°$, with a possible hexamer or trimer respectively in the asymmetric unit, as estimated from the Matthews coefficient. However, the crystal died after 3 characterisation exposures, so a full data set could not be collected.

With the crystal taking so long to appear, there was also the likelihood that it was not the full length protein, but instead a degraded stable fragment that had crystallized. However, the crystal was used for microseeding experiments with CES2 wt protein that had given this original hit, in the hope that the seeds would act as heterogenous nucleation points at low supersaturation.

Although these results were extremely promising, this one-off crystal hit proved to be neither reproducible, nor optimisable, and the microseeding experiments did not promote the growth of new crystals.

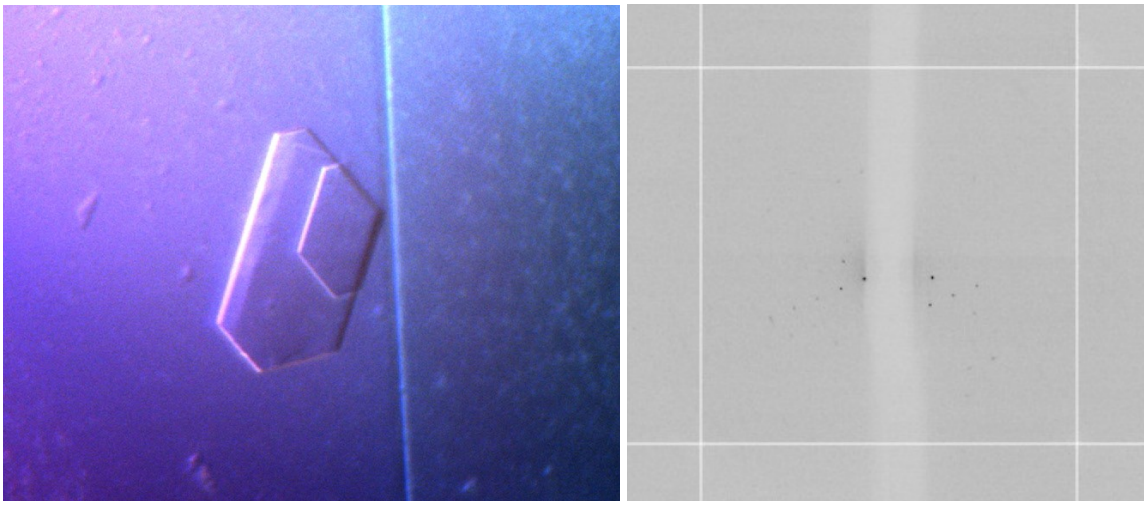


Figure 4.10. Native glycosylated CES2 from condition PACT C7.

The single, stacked plate-like crystal of CES2 native is shown on the left. It measured approximately 80 x 40 x 5 µm. The resulting diffraction pattern from the crystal is shown on the right. The data were collected on beamline I04, with an ADSC 315 CCD detector.

4.6.3. Re-thinking The Crystallization Strategy: Protein Engineering

After nearly three years with very little success in the crystallization of CES2 or any of the mutant forms, the inability to reproduce the very few sporadic crystals that diffracted extremely weakly combined with the failure of all of the techniques that were tried to aid crystallization, it was clear that this protein behaved very differently to CES1. Either CES2 was lacking suitably located surface residues to form intermolecular interactions between molecules or it possessed a structural feature, such as a flexible domain or a large termini that was preventing it from inherently crystallizing.

It was decided to go back to the stage of construct design and use the knowledge obtained from the CES1 structure to help to design a more 'crystallizable' version of CES2, as described in *4.2.3, Oligomerization Mutants*. As well as putting these two mutants into crystallization natively, it was also decided to attempt co-crystallization with a known CES2 substrate. As CPT-11 was readily available, experiments commenced with this prodrug, using a 3 molar excess.

4.6.4. Co-Crystallization Using CPT-11

CPT-11, also known as 7-ethyl-10-[4-(1-piperidino)-1-piper- idino], carbonyloxycamptothecin and irinotecan was chosen for co-crystallization experiments with CES2 native, 3-4 and 4-4. It is a water soluble prodrug that is activated *in vivo* to the active metabolite SN-38 (an analog of camptothecin), a potent topoisomerase I inhibitor[152], and the by-product 4-piperidino-piperidine (4-PP) (*figure 4.11*). It is a clinically approved anticancer drug used for the treatment of advanced colorectal carcinoma[153] and other solid tumours. CES2 has much higher specific activity for CPT-11 than CES1, which is significantly less active against this substrate[48, 83, 128, 140, 141, 154], with values of 92 pmol min^{-1} $mg^{-}1$ and 3.5 pmol min^{-1} mg^{-1} being reported, respectively[141]. It is assumed that CES2 is the major CES responsible for CPT-11 hydrolysis in cancer patients. In a comparative study, the relative activity of the three mammalian carboxylesterases was shown to be CES2> CES1>> CES3[21]. However, CPT-11 is not efficiently hydrolysed by carboxylesterases, as up to 64% of intravenous doses administered are excreted in faeces as the parent drug[155].

The most efficient mammalian carboxylesterase known to activate CPT-11 is rabbit liver carboxylesterase (rCE), which has 1.5 fold higher activity than human CES2[141]. In the first structural insights into CPT-11 activation by mammalian carboxylesterases, the crystal structure of rCE was presented with 4-PP observed bound adjacent to its Asn-linked glycosylation site (N389) on the surface of the enzyme, with nothing present in the active site[12]. Interestingly, even though rCE is considered to be the rabbit homologue of CES1, with 81% sequence identity, the human enzyme cannot process CPT-11. By contrast, CES2, which is able to hydrolyse CPT-11, only shares 47% sequence identity with the rabbit enzyme. It has also been noted that carbamates in general and CPT-11 in particular can act as serine carboxylesterase inhibitors, due to the stability of the active site serine-carbamate intermediates formed during the course of CPT-11 hydrolysis[141, 156].

Therefore CPT-11 seemed a sensible compound to use to aid the crystallization of this difficult enzyme.

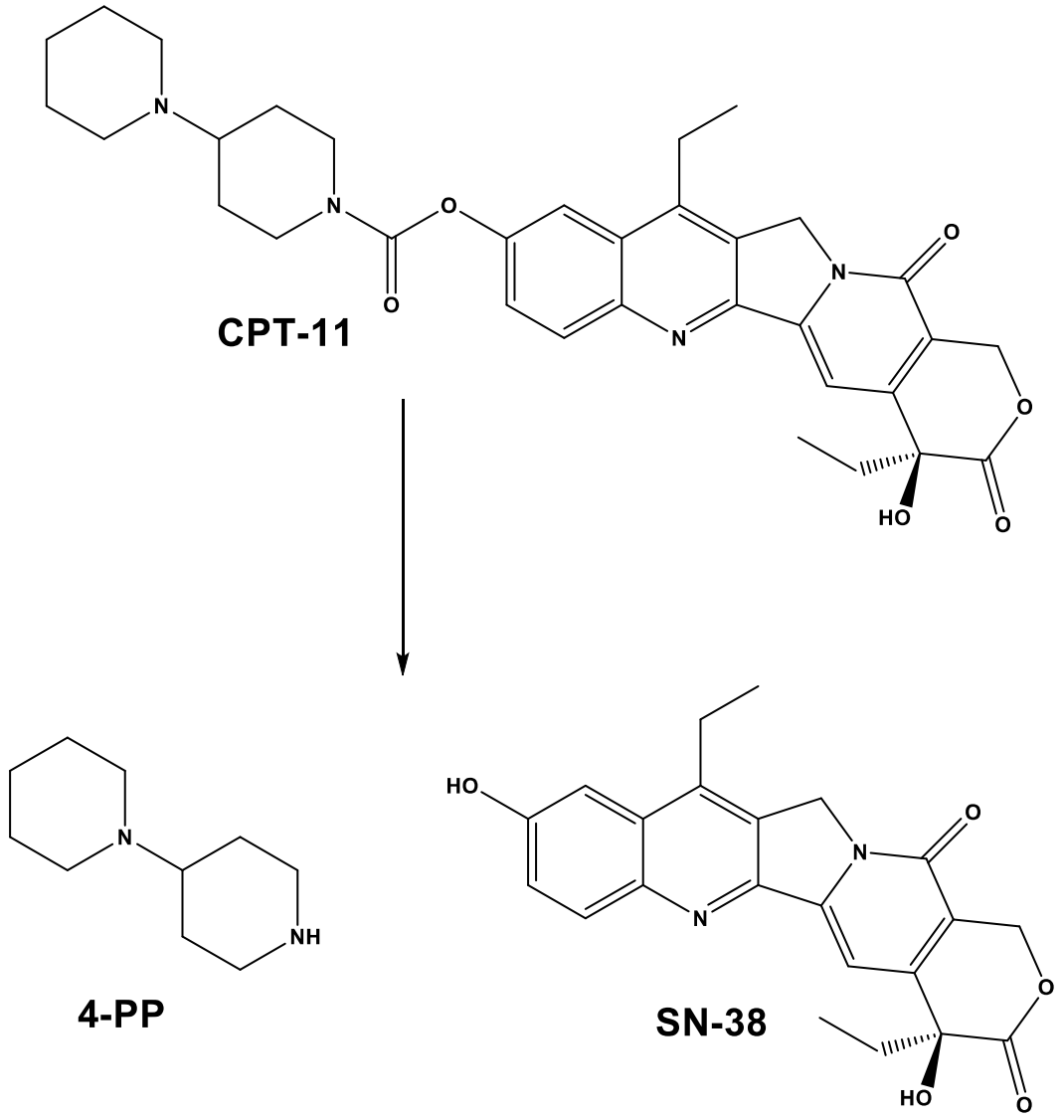


Figure 4.11. Hydrolysis of CPT-11 into SN-38 and 4-PP.
The prodrug CPT-11 is hydrolyzed *in vivo* primarily by CES2 into the active compound SN-38 and the byproduct 4-PP. SN-38 is eventually inactivated by UDP-glucuronosyltransferase isoform 1A1 to SN-38 glucuronide, which is excreted in bile and faeces.

4.6.5. Crystallization of CES2 3-4 and CES2 4-4

After setting up the initial crystallization trials, from day five onwards, crystals of CES2 3-4 co-crystallized with CPT-11 began to appear in several conditions in the PACT *premier* screen, at 20°C (*figure 4.12*), with similar plate-like crystal morphologies. Again, no crystals were observed in the same conditions set up at 4°C or in the absence of CPT-11. This was a tremendously promising breakthrough as the crystals were; *i)* appearing very quickly so presumably were not a stable fragment from proteolysis of the protein, *ii)* appearing in several similar conditions and *iii)* growing out of a heavy protein precipitate that was acting as a nucleation point.

Figure 4.12. First crystals obtained of CES2 3-4 co-crystallized with CPT-11.
The crystals grown at 20°C were clearly developing from protein precipitate. Protein precipitate appeared after 24 hours, and crystals begun to appear after 5 days. The crystals had sharp edges, in all three dimensions

4.6.5.1. Initial Data Collection

The CES2 3-4 crystals obtained above were cryo-protected with 20% glycerol, and screened for diffraction on Beamline I03 at Diamond Light Source. The best crystal diffracted to a resolution of 3.63 Å, and grew from the PACT G4 condition (0.2 M Potassium Thiocynate, 0.1 M Bis-Tris Propane, 20% PEG 3350 pH 7.5) (*table 4.5*).

Data Collection Statistics	
Protein	CES2 3-4 + CPT-11
X-ray Source	Diamond Light Source I03
Wavelength (Å)	0.9763
Space Group	*P*1
Cell Constraints	
a, b, c [Å]	96.25, 96.49, 96.92
α, β, γ [°]	72.38 77.38 64.6
Resolution range [Å]	84.7 – 3.63 (3.72 – 3.63)
R_{symm} [a]	16.0 (94.12)
<I>/□(<I>	7.9 (2.1)
Completeness (%)	98.7 (98.7)
Multiplicity	4.3 (4.3)

Table 4.5. Initial diffraction data obtained for CES2 3-4 + CPT-11.
[a]Values in parentheses refer to data in the highest resolution shell.
[b]$R_{sym} = \Sigma_{hkl}\Sigma_j|I_{j,hkl} - \langle I_{hkl}\rangle|/\Sigma_{hkl}\Sigma_j I_{j,hkl}$ where $\langle I_{hkl}\rangle$ is the average of the intensity $I_{j,hkl}$ over j = 1,..., N observations of symmetry equivalent reflections *hkl*. Numbers in parentheses refer to values in the highest resolution shell.

4.6.6. Crystal Optimization

4.6.6.1. Initial Optimization

The initial rounds of optimization involved using the 3-row optimization technique as well as varying the concentrations of precipitant and potassium thiocyanate (PTCN). Details of these can be found in Chapter 2, *2.4.3.2*. The 3-row optimization did not improve crystal growth or size. With varying the conditions of the original screen, it was found that the best crystals formed in similar conditions to the original screen, around 18-20% PEG 3350 and 0.2 M PTCN.

4.6.6.2. Additive Screen

Formulations in the additive screen containing the compounds sodium thiocyanate, L-proline, glycyl-glycyl-glycine and spermine tetrahydrochloride (B9, C2, C11, D1 and D3)

improved the size of the crystals, which were much easier to handle. However, on testing of these large crystals on the beamline, no significant improvement in diffraction quality or resolution limit was observed.

4.6.6.3. Hanging Drop Experiments

CES2 3-4 + CPT-11 crystals grown in 2 µL hanging drops (1:1 ratio of protein to reservoir solution) at an initial protein concentration of 14.0 mg/ mL were much larger than the crystals grown using 200 nL (100 nL protein plus 100 nL reservoir solution) sitting-drops in 96 well format[115], and were visible by eye. These crystals were grown at room temperature. Micrographs of the crystals taken with a visible light microscope are shown in *figure 4.13*.

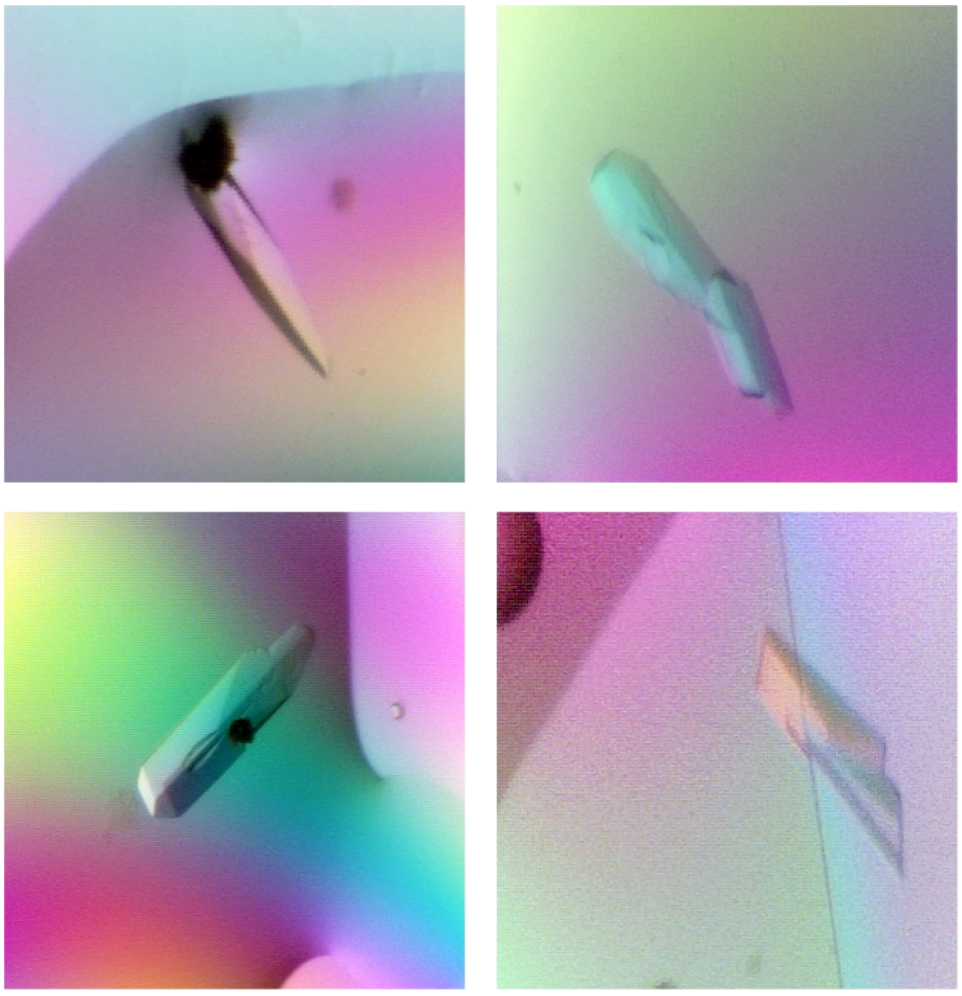


Figure 4.13. Hanging-drop crystals of CES2 3-4 +CPT-11.
Crystals were obtained by the vapour diffusion method (hanging drop) at room temperature by manually mixing 1 L of protein solution with 1 L of well solution (see main text for details).

4.6.6.4. Cryo-Protection Optimization

Crystals from the hanging drop experiments were cryo-protected in two different cryo-protectants: 25% ethylene glycol in 24% PEG 3350, 0.2 M PTCN, 0.1 M Bis-Tris Propane pH 7.5 or 20% glycerol in 18% PEG 3350, 0.2 M PTCN, 0.1 M Bis-Tris Propane pH 7.5. These two solutions were based around the original hit conditions. Twelve of the crystals were tested on beamline I03, and diffracted to a much higher resolution than previously achieved. Using glycerol as the cryoprotectant produced the best diffracting crystals with data recorded to 2.87 Å. Again, the crystal was characterized as belonging to the space group *P*1 with cell dimensions of 95.98, 96.51, 96.66 (a, b, c [Å]) and 77.45, 72.09, 64.65 (α, β, γ [°]).

In a final attempt to drive down the resolution, the cryo-protection method was optimized. Instead of directly placing the crystal into the solution with 20% glycerol (20% glycerol, 18% PEG 3350, 0.2 M PTCN, 0.1 M Bis-Tris Propane pH 7.5), four solutions (5, 10, 15 and 20% glycerol) of the condition were made up, and the crystals were serially cryo-protected, starting from the lowest concentration of glycerol, up to the highest. The rationale behind this was to try and minimise the stress impact on the crystals.

4.6.6.5. Data Collection Optimization

Because the MX beamlines at DLS are highly automated, data from cryo-optimised crystals were collected remotely, and visualised using the SynchLink application, developed for iPhones and iPads (*figure 4.14*). As it had been established that CES2 3-4 crystallized in space group *P*1, a helical line scan was used to enable the collection of 180° of data over the entire length of the crystal to mitigate the effects of radiation damage.

Assessment of 24 CES2 3-2 + CPT-11 crystals revealed crystals that were diffracting to between 2.6 – 2.8 Å resolution. Exceptionally, one crystal with dimensions of approximately 140 x 90 x 40 µm, diffracted significantly better than all others. A complete data set was collected again using a line scan and the data was processed automatically with xia2 to 2.18 Å. Manual reprocessing allowed the data to be extended to a resolution of 2.04 Å, as judged by the use of $CC_{1/2}$ as described recently by Karplus and Diederichs[157].

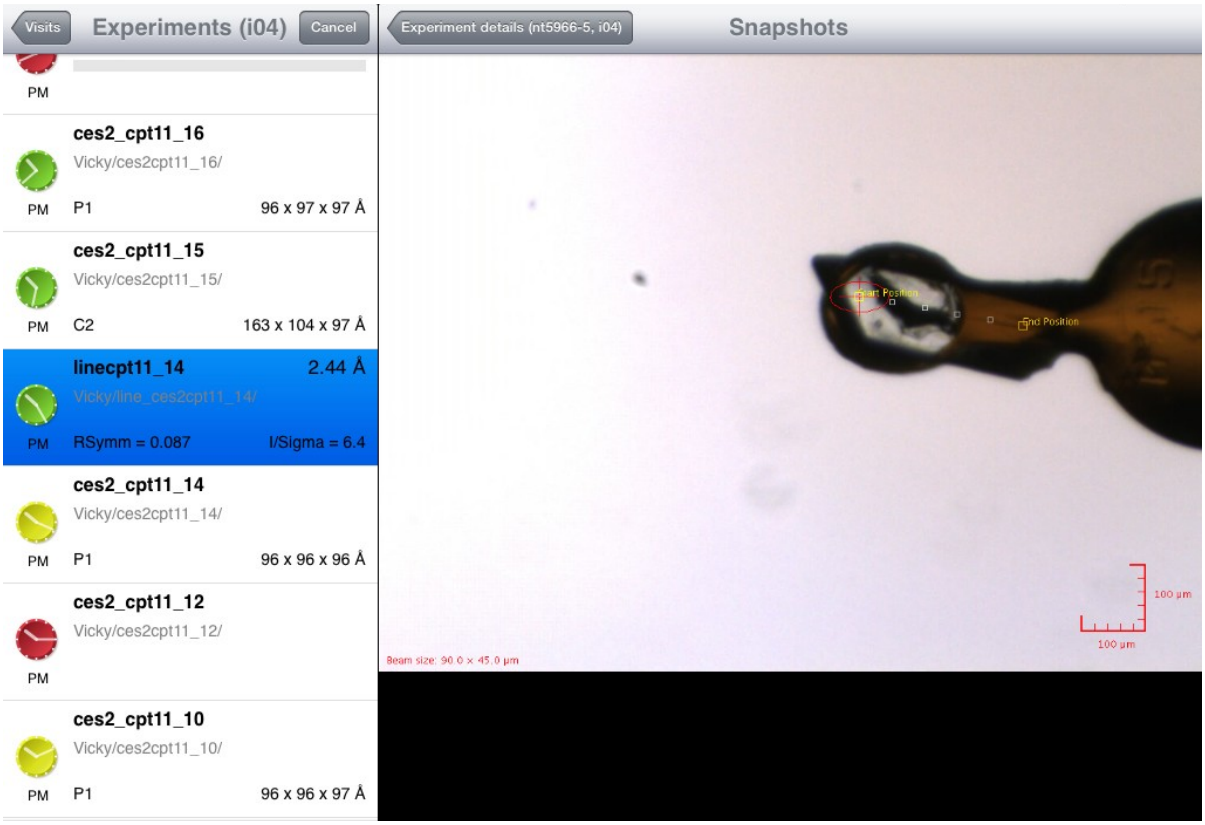


Figure 4.14. Snapshot from the interface of the SynchLink application, as data were being collected remotely at Diamond Light Source.
The final data set for CES2 3-4 + CPT-11 was collected remotely. The crystal is shown mounted in a Hampton loop on the beamline goniometer. Experiment details including unit cell dimensions, space groups and data collection times are shown in the left hand panel.

4.6.6.6. Crystals of CES2 4-4

When crystallization trials of CES2 4-4 + CPT-11 were set up, crystals appeared in the same conditions as CES2 3-4, however they had a thin plate-like morphology and appeared to lack a third dimension. When attempting to fish these delicate crystals, they broke up easily, and unfortunately attempts to mount these crystals for characterization and data collection at the beamline failed. This fourth mutation that was created in CES2, K297E, seemed to have reduced the quality of CES2 crystals.

Data Collection Statistics	
Protein	CES2 3-4 + CPT-11
X-ray Source	Diamond Light Source I04
Wavelength (Å)	0.97
Space Group	*P*1
Cell Constraints	
a, b, c [Å]	96.23, 97.13, 97.25
α, β, γ [°]	77.06, 64.49, 71.92
Resolution range [Å]	91.82 – 2.04 (2.09 - 2.04)
<I> half set correlation coefficient ($CC_{1/2}$)	0.99 (0.57)
R_{symm} [a](%)	5.6 (120)
<I>/□(<I>	7.4 (1.0)
Completeness (%)	97.4 (96.9)
Multiplicity	2.4 (2.4-2.4)

Data Refinement Statistics	
Protein	CES2 3-4 + CPT-11
No. of reflections	360,270
No. of unique reflections	152,937
***R*-factor (%)[b]**	17.19
R_{free} (%)	21.17
Rmsds	
Bond lengths (Å)	0.0188
Bond angles (°)	1.9667
Wilson *B*-factor (Å^2)	53.5
Mean B-factor (Å^2)	46.8

Table 4.6. Final diffraction data and refinement statistics obtained for CES2 3-4 + CPT-11.
[a]Values in parentheses refer to data in the highest resolution shell.

[b]$R_{sym} = \Sigma_{hkl}\Sigma_j|I_{j,hkl} - \langle I_{hkl}\rangle|/\Sigma_{hkl}\Sigma_j I_{j,hkl}$ where $\langle I_{hkl}\rangle$ is the average of the intensity $I_{j,hkl}$ over j = 1,..., N observations of symmetry equivalent reflections *hkl*. Numbers in parentheses refer to values in the highest resolution shell.
[b]R-factor = Σ (||F(obs)| - |F(calc)||)/ Σ|F (obs)|; Rfree = R factor for a selected subset (5%) of the reflections that was not included in prior refinement calculation.

4.7. Structural Analysis of CES2

CES2 3-4 crystallized in the triclinic space group *P*1, and contained four molecules in the asymmetric unit, made up of two offset dimers (*figure 4.17*), with a Matthews coefficient[158] of 3.29 $Å^3$ Da^{-1}, and solvent content of 62.62%. 1,286 water molecules were modelled in the final structure. This relatively high solvent content probably explains why the majority of crystals of CES2 3-4 did not diffract to below 2.4 Å high resolution. After refinement, the model converged with an R factor of 0.1719 (R free 0.2117) (*table 4.6*).

The resolved protein chain runs from residues 28 to 520. There is no density available for residues 444 to 457, and the *C*-terminus of the chain after residue 520 is essentially disordered, in all chains in the tetrameric structure (*figure 4.16*). In the rCE structure (1KAY), this first loop (450-466 in the rabbit sequence) also could not be modeled. There are two conserved intra-molecular disulphide linkages formed between C123 and C95, C280 and C291 that help to stabilise the structure. There is also an unexpected inter-molecular disulphide bridge formed between C428 on both molecules of CES2 in the dimer (*figure 4.18*). The dimeric subunit in the structure is stabilised by the presence of a salt bridge (distance of ~3.0 Å) formed between R390 of one molecule, and D98 of the other.

At around N342, there is a kink in the protein chain, and this marks the beginning of an unexpected feature of a considerable strand extension and exchange. The sequence at this special region in the protein is NNNEF and on sequence alignment, this motif is not conserved in CES1 (NKQEF). Because polar asparagine side chains can readily form hydrogen bonds with the peptide backbone, they are often found near the beginning of alpha helices and in turn motifs, and the three consecutive asparagines in CES2 appear to be involved in the turn motif immediately before the beginning of the helical strand

extension in this structure. The residue E345 that follows the run of asparagines is actually the orientating acid in the catalytic triad, and it is positioned on the edge of the strand extension.

Interestingly in the rCE structure, there is another region for which no electron density was observed. This runs from residues 355 to 370, from the conserved E**F**GW motif, which on alignment is **F**346 in CES2, and is located in the strand helix extension portion of the molecule. Potentially these residues, 355-370, could form a small extended region in the rabbit structure that is particularly flexible, and similar to the extension in CES2.

Areas that are poorly defined or absent in the electron density map of the CES2 monomers are generally flexible loops that connect secondary structural elements, or the carboxyl-termini of the protein chain. The absence of well-ordered electron density in these areas presumably reflects mobile regions of CES2. In the ER, the *C*-terminal endoplasmic retention signal would be presumably bound to the trans-membrane KDEL receptor, and flexibility in this region would be required for the protein to move and act on intracellular substrates, whilst anchored to the membrane.

The average B-factors (all atoms) for each chain are: Chain A 43.88, chain B, 44.86, chain C 43.05 and chain D 44.56 (*figure 4.15).*

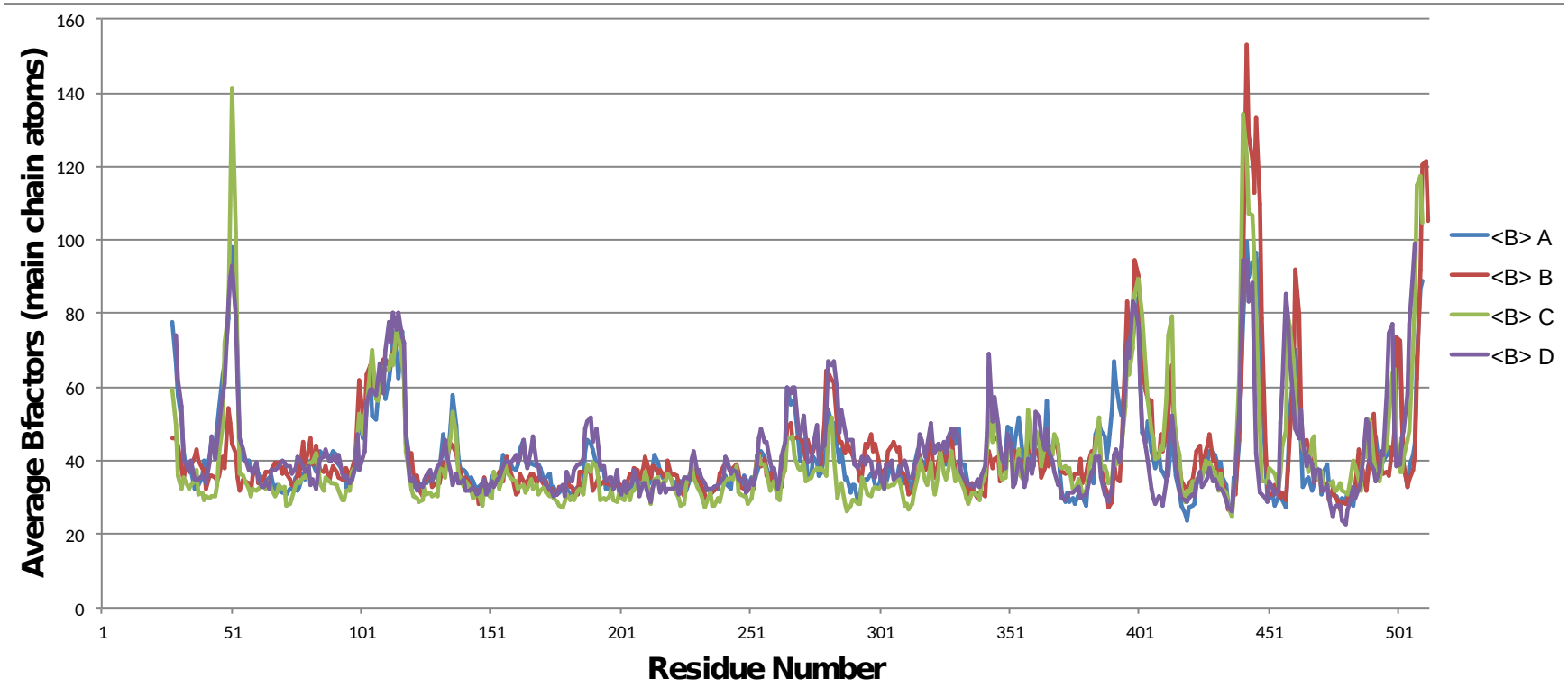
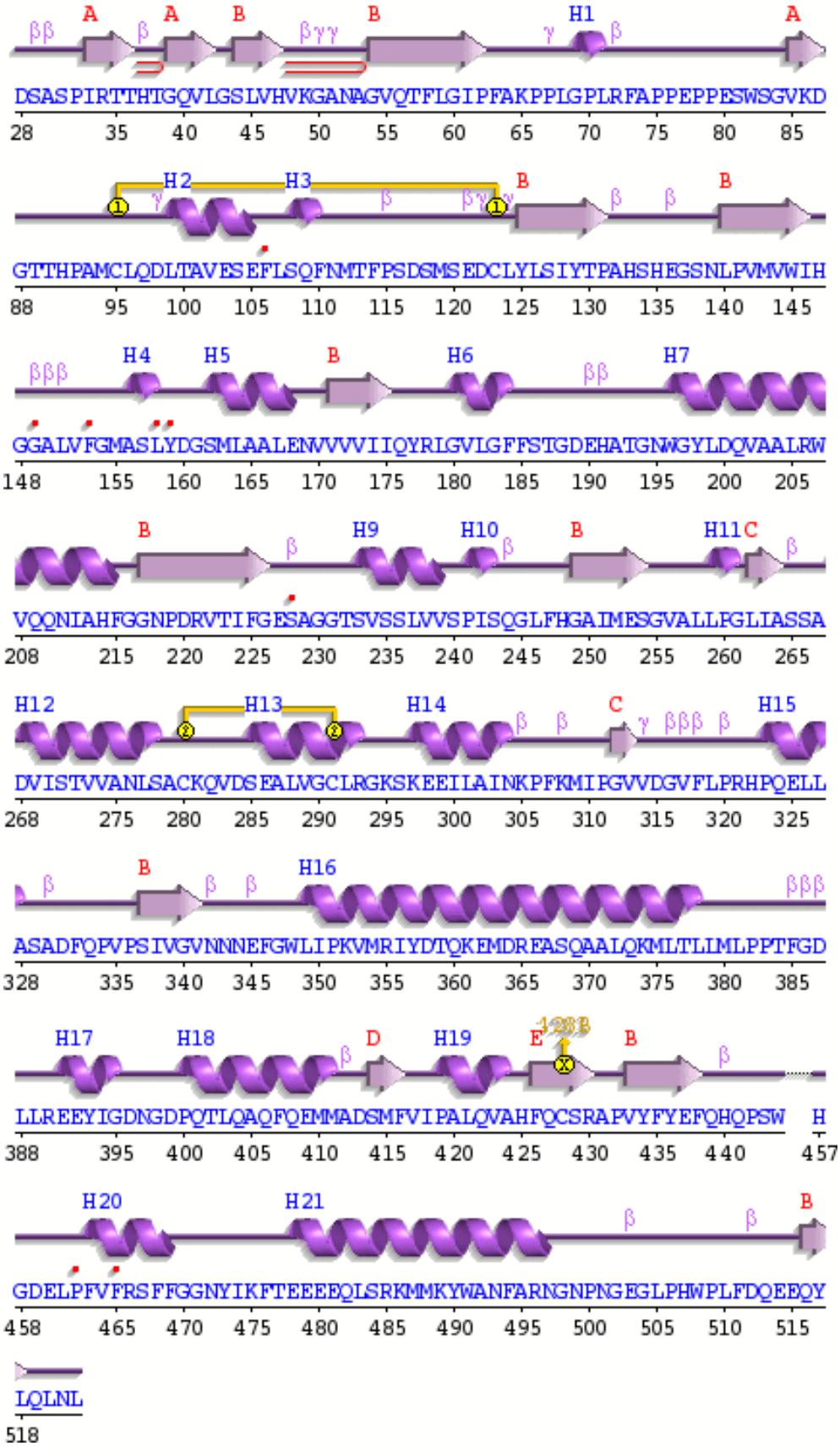


Figure 4.15. Average B-factors of Chains A-D in CES2.
Only the B-factors for the main chains are shown (no side chains). Residue numbers are shown on the *x* axis. The key for each chain is on the right hand side.

a.


DSASPIRTTHTGQVLGSLVHVKGANAGVQTFLGIPFAKPPLGPLRFAPPEPPESWSGVKD
28 35 40 45 50 55 60 65 70 75 80 85
GTTHPAMCLQDLTAVESEFLSQFNMTFPSDSMSEDCLYLSIYTPAHSHEGSNLPVMVWIH
88 95 100 105 110 115 120 125 130 135 140 145
GGALVFGMASLYDGSMLAALENVVVVIIQYRLGVLGFFSTGDEHATGNWGYLDQVAALRW
148 155 160 165 170 175 180 185 190 195 200 205
VQQNIAHFGGNPDRVTIFGESAGGTSVSSLVVSPISQGLFHGAIMESGVALLPGLIASSA
208 215 220 225 230 235 240 245 250 255 260 265
DVISTVVANLSACKQVDSEALVGCLRGKSKEEILAINKPFKMIPGVVDGVFLPRHPQELL
268 275 280 285 290 295 300 305 310 315 320 325
ASADFQPVPSIVGVNNNEFGWLIPKVMRIYDTQKEMDREASQAALQKMLTLIMLPPTFGD
328 335 340 345 350 355 360 365 370 375 380 385
LLREEYIGDNGDPQTLQAQFQEMMADSMFVIPALQVAHFQCSRAPVYFYEFQHQPSW H
388 395 400 405 410 415 420 425 430 435 440 457
GDELPFVFRSFFGGNYIKFTEEEEQLSRKMMKYWANFARNGNPNGEGLPHWPLFDQEEQY
458 465 470 475 480 485 490 495 500 505 510 515
LQLNL
518

b.

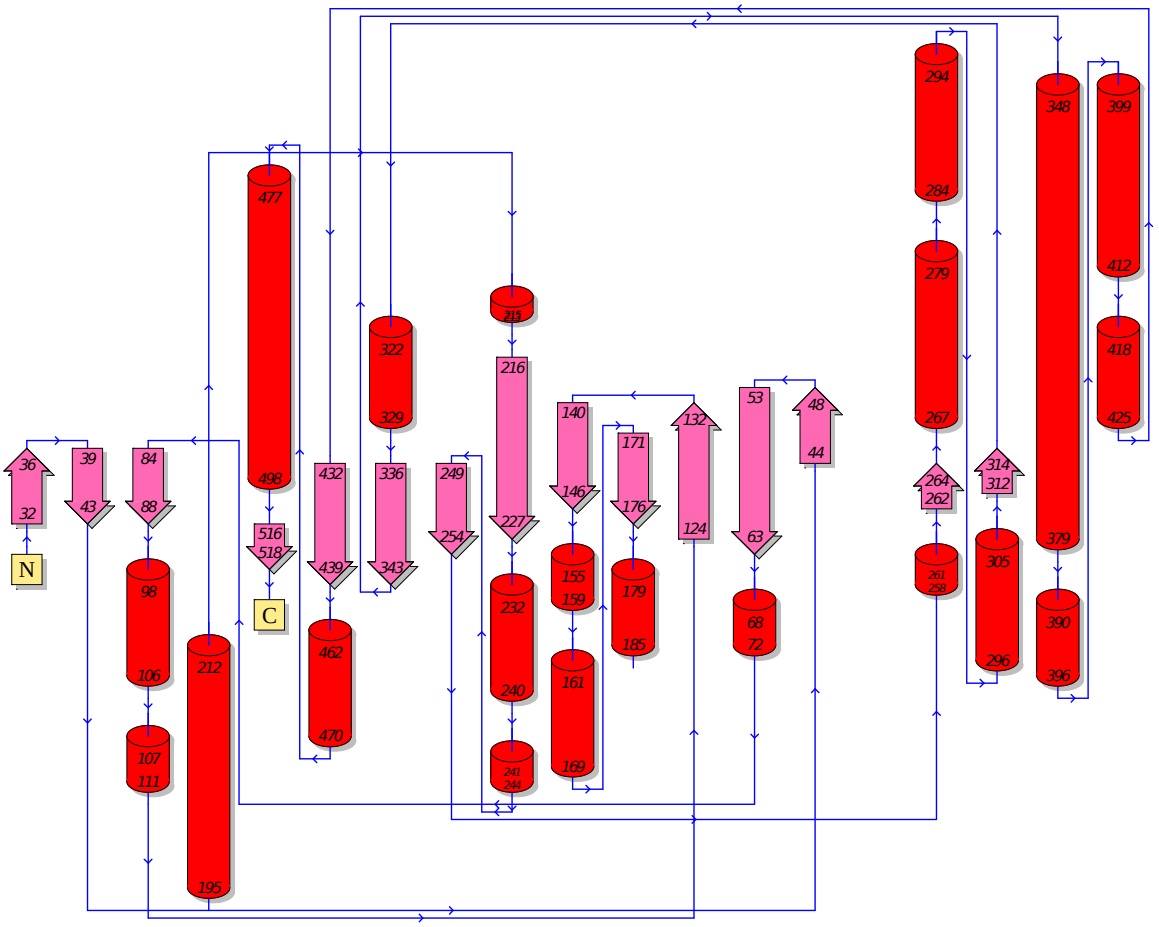


Figure 4.16. Secondary structure (a) and topology (b) view of CES2 3-4.
Chain A from the final structure was used for analysis. Images were generated using the PDBsum[159] online server. It can be seen that the secondary structure is mainly helical, with a total of 21 helices. In ***(a)*** The position of the residues involved in disulphide linkages are shown in yellow. The dotted line that proceeds residue 457 indicates the absence of structure from lack of diffraction data. ***(b)*** Indicates the structural features of the protein.

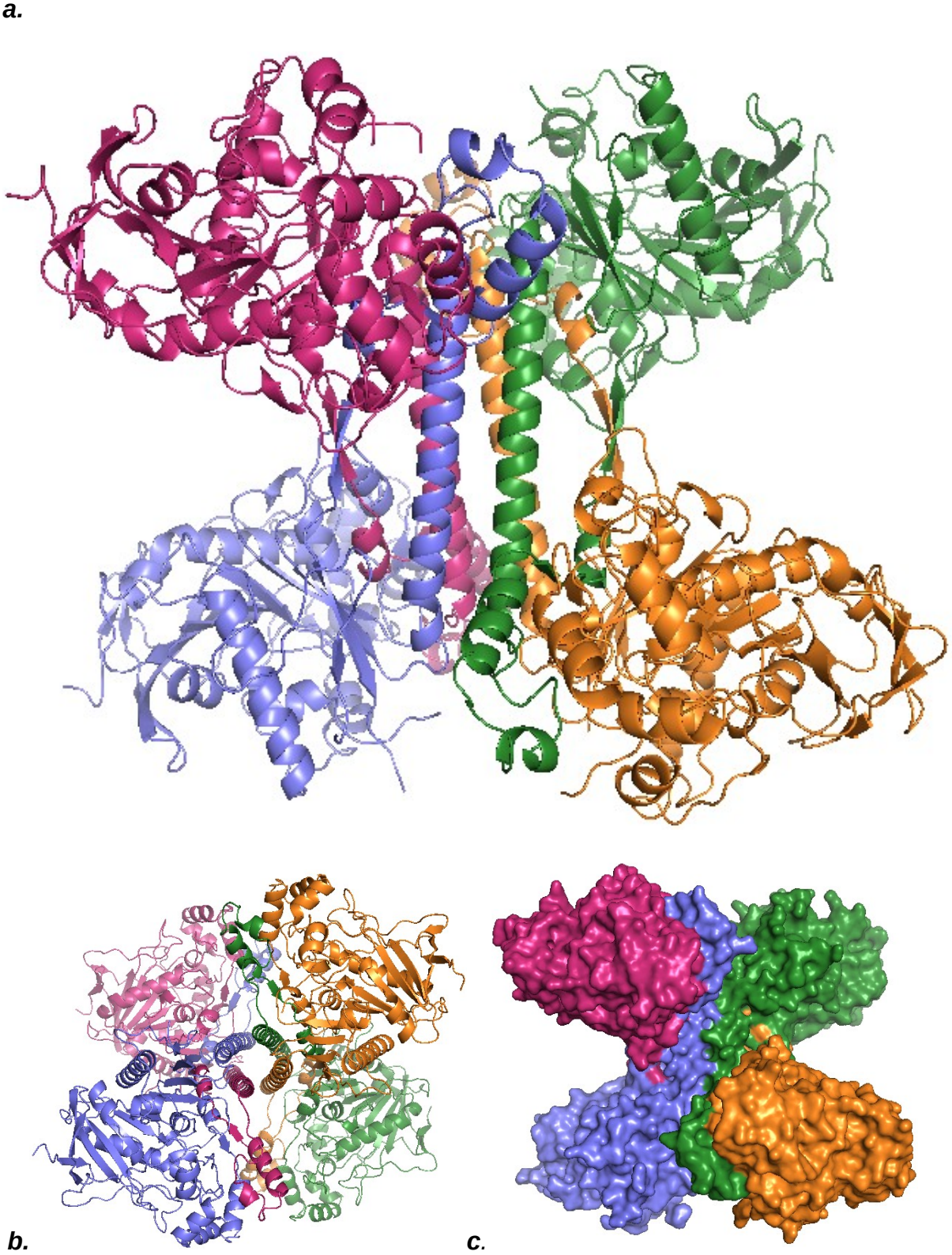


Figure 4.17. Cartoon view of overall quaternary structure of CES2 3-4 generated in MacPyMOL.
a) The tetrameric assembly of CES2 is formed by two off set dimers stacked together, with their strand extension exchange helices running anti-parallel with each other, buried deep in the core of the structure. ***b***) 90° View of figure *a* as illustrated. ***c***) Surface representation of figure *a*. Depth cue is turned on in all images.

4.7.1. Active Site

In all four active sites of the tetramer, 4-PP, the by-product of CPT-11 hydrolysis (*figure 4.11*) is observed covalently bound to each of the catalytic S228 residues (*figure 4.19*). The catalytic E354 is positioned on the hinge at the beginning of the helix extension that forms the strand exchange with the other protein chain in the dimer assembly. This 'extension' makes up the regulatory domain of the enzyme; it extends from the catalytic domain and also returns to the catalytic domain and is structurally very different from the regulatory domain in CES1.

The two conserved residues, G148 and G149, that comprise the oxyanion hole are present in the active site, adjacent to the catalytic serine, and approximately 6.0 and 8.0 Å away from the 4-PP molecule. The second conserved serine, S254, is located in the catalytic pocket at the end of a beta sheet that runs parallel with the catalytic serine, 10.5 Å away. In this structure, it is positioned 14.0 Å away from the catalytic glutamate (E354), clearly not in hydrogen-bonding distance and unlike in the structures of CES1[85, 87]. However, S254 may re-locate during catalysis. This additional conserved serine was first identified in serine proteases[160] and the most recent opinion on this residue present in all human carboxylesterases is that is not directly involved in catalysis, but is likely to be a structural support which stabilises the catalytic triad[87].

Interestingly, on comparing the structure of human CES2 with that of rCE[12], the 4-PP in rCE is not bound covalently in the active site, but on the well documented 'Z-site'. However, the rCE protein was crystallized in the presence of 4-PP, not CPT-11. It would be interesting to see if the rCE enzyme was co-crystallized with the parent compound, whether the 4-PP would be observed bound in the active site.

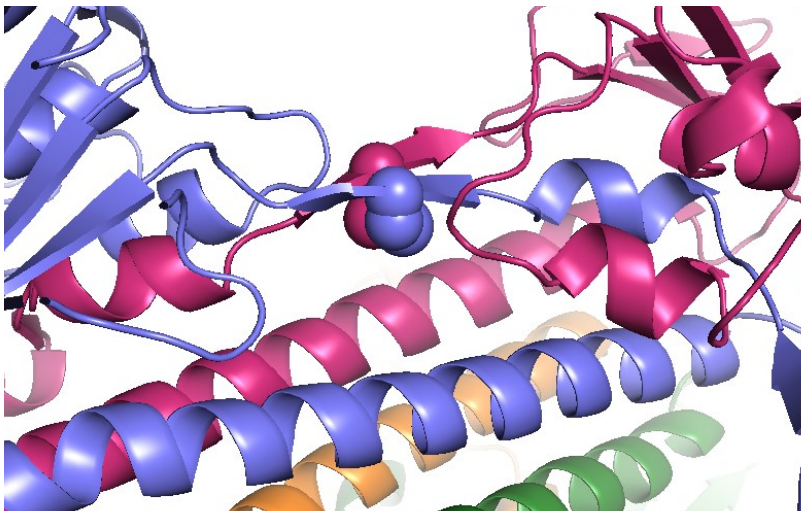


Figure 4.18. Inter-molecular disulphide bridge.
The disulphide bridge is represented by spheres between C428 residues on two monomers of the protein, running anti-parallel to each other. They are present on the part of the molecule that returns from the strand exchange and lie adjacent to the four helices in the core of the tetramer.

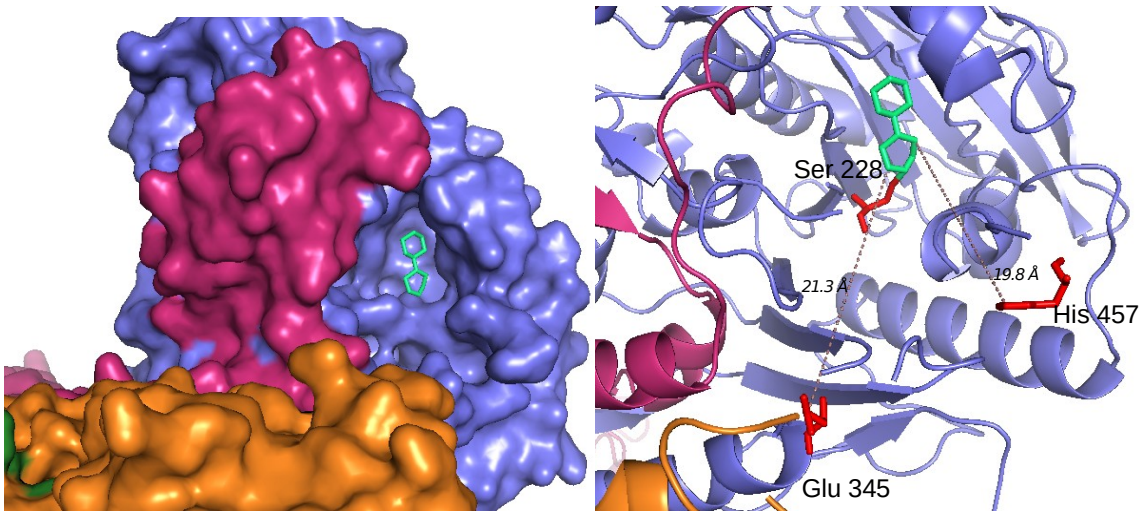


Figure 4.19. Active site of CES2 with 4-PP covalently bound.
On the left is a surface representation of the active site. It can clearly be seen that the active site in a monomer (slate coloured) of the tetramer is exposed and on the surface of the molecule. It is also partly surrounded by the strand extension region of the other monomer (pink) that together form the dimeric subunit. On the right, the residues involved in the catalytic triad are indicated. A 4-PP molecule is covalently bound to the Ser residue. This is observed in all tetramers in the structure. The His and Glu residues are 19.8 Å and 21.3 Å away from the 4-PP molecule respectively. B factors of the residues are: Ser 35.11, Glu 65.85 and His 96.61, indicating that the His residue is the most mobile.

4.7.2. *N*-Glycosylation

CES2 contains two *N*-linked glycosylation sites: N111 and N276. In the crystal structure, one *N*-acetylglucosamine (NAG) molecule can be modelled into the positive density on N276 (*figure 4.20*). The full sugar expected on this residue is $GlcNac_2Man_9$ as this protein was produced transiently in HEK293T, with the addition of kifunensine to modify the glycoforms. As can be seen from *figure 4.20*, it is likely that this sugar motif interacts with the strand exchange region of the other monomer in the dimer. Positive density extending further than the first NAG molecule was also observed at the N111 glycosylation site, but further work is needed to correctly model these sugars. These observations confirm that both *N*-glycosylation sites are occupied in the crystallized protein and that the presence of the core mannose oligosaccharide has not affected crystallization, which is sometimes seen with glycosylated proteins.

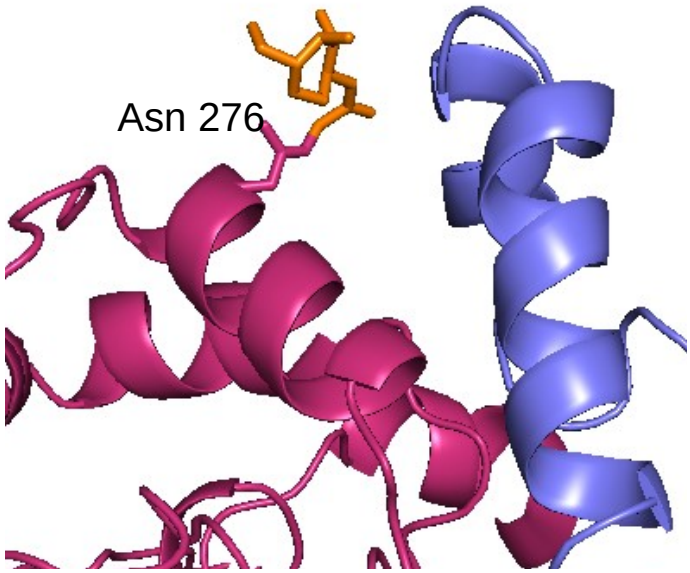


Figure 4.20. Glycosylation of Asn 276.
One NAG molecule is shown in orange. It is within hydrogen bonding distance of the other protein chain within the dimer.

4.7.3. Mutated Residues

Because the CES2 protein had been modified with three amino acid substitutions, (R86K, K190E, D281K), it was important to see where these were positioned within the structure, to try and elucidate why this mutant crystallized as opposed to the native species. Both K86 and E190, although located on the surface of the protein core, are not involved in crystal contacts. K281 is in hydrogen bonding distance to the N111 glycosylation site, and it is proposed that it could anchor this sugar moiety, which in turn may stabilise the catalytic triad. This may explain the reason for this CES2 3-4 mutant crystallizing much more readily than the wild type.

Interestingly, the fourth mutation that was introduced into CES2 4-4 but not CES2 3-4 (Lys→ Glu) appears to actually knock out a crystal contact that is observed in the CES2 3-4 structure between K297 and E298 of another molecule not in the tetrameric arrangement. Therefore, it could be speculated that CES2 with the single engineered mutation Asp 281 → Lys would yield a crystallizable form of the protein, and that the other two introduced mutations had no impact on this process.

4.7.4. Mapping of Single Nucleotide Polymorphisms Onto CES2

Single nucleotide polymorphisms (SNPs) can significantly influence the metabolism and disposition of many therapeutic drugs. There have been many naturally occurring SNPs in CES2 published, found in both introns and exons. Two CES2 SNPs that were identified in CPT-11-administered Japanese cancer patients and in the coding region of the gene, were found to almost completely knock out enzymatic activity towards CPT-11, 4-NPA and 4-methylumbelliferyl acetate[161]. These were single base changes 100C>T, which resulted in the substitution R34W and 424G>A which resulted in V142M. A third SNP, IVS8-2A>G,

in the splice acceptor site in intron 8 was also identified, and this resulted in a truncated form of CES2.

Mapping the two coding SNPs onto the CES2 structure explains their effect on the activity of the enzyme. R34 forms a hydrogen bond with Q56, which stabilises the longest beta sheet in a series of six consecutive sheets (*figure 4.21*). Mutation of this residue to a tryptophan would abolish this hydrogen bonding, and potentially disrupt the stability of the overall protein structure. V142 is also located on one of these beta sheets, and this region is 100% conserved between CES1 and 2 (P**V**MVWIHGG). It is likely that substitution of this valine residue with a bulkier side chain residue such as methionine would disrupt the beta sheet formation, and again de-stabilize the protein.

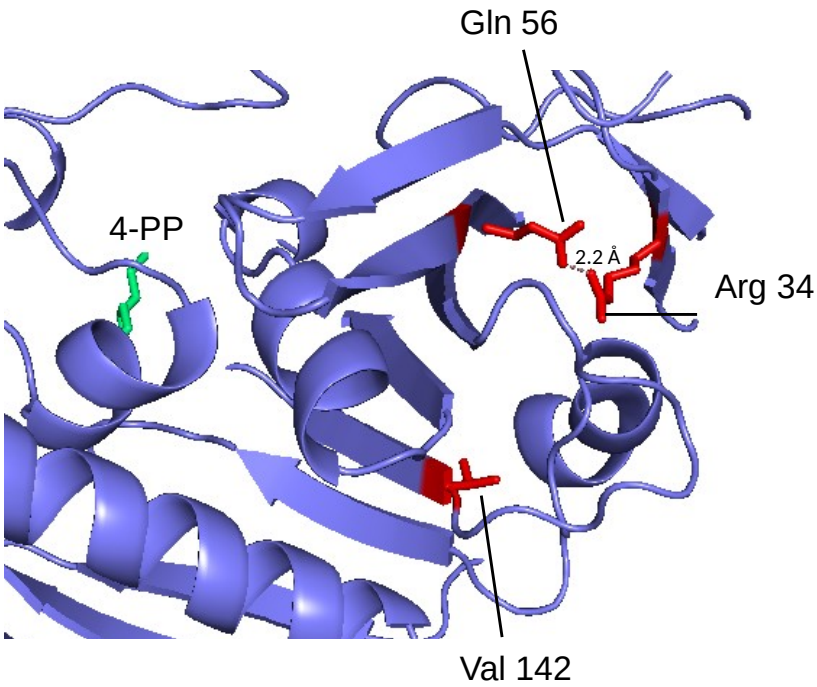


Figure 4.21. Mapping of documented SNPs onto CES2.
On the surface of the protein, hydrogen bonding (2.2 Å) between Arg 34 and Gln 56 is indicated. This is on the opposite side of the protein to the active site, so is not directly involved with the catalytic triad.

4.8. Discussion

4.8.1. Protein Production

The results documented here confirm that CES2 produced from HEK293T and HEK Gnt$^{-/-}$ cells is essentially identical to previous examples of this enzyme isolated from mammalian cells[37] validating the use of the mammalian system as a source for recombinant proteins in structure determination studies

In terms of purification of CES2, further optimization would be required in the future to maximize the yield. A possibility that may have been affecting the binding of the protein to the nickel column is that the polyhistidine-tag was partially hidden or buried in CES2, especially if CES2 was forming dimers or heavier oligomers that were masking the *C*-terminus. If this was the case, purification of CES2 could further be improved by extending the length of the tag from His6 to His9 in the construct design to increase binding ability. Also, the supernatant could be dialysed prior to purification to remove all chelators present in the DMEM media to prevent non-specific binding to the column. Possible techniques could include using dialysis tubing or a cross-flow machine, e.g. LV Centramate™ Lab Tangential Flow Filtration System to remove all chelators present in the DMEM media.

4.8.2. Kinetics

This is the first study to fully investigate the kinetics of CES2 and the effect of glycosylation on catalytic activity. It is likely that CES2 exhibits substrate inhibition, positive cooperativity and a Ping Pong mechanism, like CES1, but to a greater degree,

which makes evaluating the kinetic data relatively complex. Substrate inhibition has also been documented in other proteins closely related to CESs, including actelycholinesterases[149]. If CES2 does exhibit positive cooperativity and possesses an alternative binding site, further work would need to be carried out to identify this site on the protein.

Historically, atypical kinetic profiles for any enzyme were often misinterpreted, or ignored. In terms of drug metabolism, it is important to understand the various factors that affect the activity and behavior of CES2. It is crucial not to ignore or truncate non-hyperbolic data as it can lead to erroneous kinetic parameters, as classical Michaelis Menten are forced through the data, rather than choosing a more appropriate model. Therefore it is imperative that the correct kinetic model is used to estimate accurate parameters from the data collected. For the comprehensive activity analysis presented here, no model could be fitted to the data as not enough was known about the enzyme mechanism. However, it could be argued that this is preferable to trying to fit the data to an incorrect model and thus generate inaccurate kinetic parameters. Further studies need to be conducted on the kinetic activity of CES2 to determine whether its observed atypical kinetics are relevant *in vivo*, or are an *in vitro* artifact. However, it is still important to fully understand and appreciate these atypical kinetics *in vitro* as they may impact *in vitro-in vivo* correlations of drug metabolism[150].

4.8.2.1. Glycosylation

From findings presented here, it is evident that the two glycosylation sites in CES2 are fully occupied in the protein but have very different roles from the single glycosylation in CES1. The *N*-linked glycans appear to be critical both to the activity of the enzyme and to the stabilization of its quaternary structure, although these are ultimately related. It could be argued that glycosylation is critical for the stability of CES2. The residue N111, which

when possesses no glycans, has the biggest affect on reducing the catalytic activity of the enzyme. This residue is positioned in close proximity to the catalytic pocket on the surface of the protein, in a disordered region with poorly defined electron density. It could be speculated that native sugars on this residue contribute to the stabilization of the active site, and removal of them affects the correct positioning of the catalytic triad.

In the aglycosylated mutants, catalytic activity was compromised, and in general they were less stable than the wild type during protein production, degrading rapidly. An interesting follow-on experiment would be to genetically mutate out the glycosylation sites in CES2 individually, on the CES2 3-4 backbone, and see how this enzyme behaved in crystallization experiments.

4.8.3. Structure

The first known crystallographic structure of CES2 to 2.04 Å presented in this thesis shows marked differences to the reported structures of CES1, rCE, bacterial cocaine esterase[162] and predicted models[139]. The structure of CES2, obtained by mutating three residues was shown to be of a catalytically active enzyme through both an activity assay using 4-NPA as the substrate, and the fact that a product of CPT-11 cleavage (4-PP) was observed bound to the active site in the structure. It is important to note that it was not just the three mutations that enabled the protein to crystallize as no crystals were obtained for the *apo* form. Hence co-crystallization with CPT-11 was also critical. The high levels of 4-PP produced from CPT-11 cleavage during the crystallization experiments may have inhibited the enzyme and trapped it in its tetrahedral intermediate state, and this 'locked' state of CES2 may have allowed for nucleation, along with the three point mutations, with one potentially contributing to intermolecular crystal contacts as well as stabilising the glycosylation site at N111.

Although the intermolecular disulphide bond observed in the CES2 quaternary structure that seems to be central in stabilizing the dimer formation through strand exchange may appear unusual, it is important to consider the *in vivo* location of this enzyme. As discussed in *Chapter 1*, CES2 resides in the ER lumen and this residence is regulated by two factors; *i)* an NH_2-terminal hydrophobic signal peptide for the localization of CES2 to the ER and *ii)* the 'HTEL' COOH-terminal retention sequence that prevents secretion from the cell[20]. This retention motif binds to a KDEL receptor, and this interaction effectively anchors the protein to the ER. The environment in the ER is well established as an oxidizing environment to promote and enable disulphide bond formation and therefore the intermolecular disuphide bridge observed in the crystal structure is highly likely to exist *in vivo*, and is not necessarily an artifact of crystallization.

The question arises as to whether the tetrameric assembly of CES2 observed in the crystal structure is relevant to the protein in solution and more importantly to the biologically active species *in vivo*. The results of the size exclusion profile, AUC and SEC-MALLS show that CES2 behaves primarily as a monomer, with some evidence that is self associates to form a dimeric subunit at increasing protein concentrations.

The human CES2 structure provides the first view at the molecular level of the most efficient CPT-11 hydrolysing human carboxylesterase. This structure helps to rationalize why CES2 is able to hydrolyse much larger, complex molecules than CES1. The active site is clearly exposed on the surface of the enzyme and also the flexible catalytic histidine thumb region could allow for accommodation of esters in the catalytic pocket.

Papers that have identified SNPs in CES2[24, 161, 163] have all called for further research on this enzyme and a better understanding of how genetic polymorphisms of CES2 influence drug response, which may ultimately lead to personalized tailoring of pharmacological therapies. Having the crystallographic structure could also help in the development and design of specific CES2 inhibitors.

5. PRELIMINARY STUDIES ON HUMAN CES3

5.1 Introduction

During the course of this DPhil project, publications and interest into human CES3 increased[19, 164], and therefore work was initiated to produce and characterise this third active human carboxylesterase. The gene encoding CES3 was first identified (GI: 7019977) during the New Energy and Industrial Technology Development Organization human cDNA sequencing project (http://www.ncbi.nlm.nih.gov/nuccore/AK000105.1). It is important to note that there has also been several publications on Ces3[34, 165], but this is the mouse carboxylesterase 3 gene, which is in fact more similar to CES1.

In a 2012 publication, Zhao et al.[19] proposed that an increase in expression of CES3 was a compensatory mechanism for re-establishing intracellular carboxylesterase activity and free cholesterol (FC) efflux based on experiments using cell lysates from CES1 deficient cells. Gene compensation is a recognized mechanism whereby the loss of a gene is compensated for by increased expression of a related one. This is specifically predicted to occur in enzymes of the same family with overlapping substrate specificities[166], such as carboxylesterases.

This chapter describes the preliminary work that has been carried out on CES3, with comparisons drawn between CES1 and 2, and possible routes for further research is outlined.

5.2. Protein Production

The same approach that was followed for engineering CES1 and CES2 was applied to CES3. Thus the wild type sequence was truncated at the *C*-terminus to remove the ER retention motif, QEDL (*see appendix 1.1*), and then three mutants were constructed using this sequence as a template. The first was a *C*-terminal truncated mutant (CES3_trunc). As with CES2, the RONN[146] and PHYRE[143] disorder prediction tools identified a stretch of fourteen disordered residues, and these were truncated so that the protein ended with **TLPS**, prior to the His6 tag. These were deleted with the aim that the truncated protein might be more amenable to crystallization. The second two mutants were created in parallel with CES2. Sites were engineered into CES3 with the aim of promoting a trimeric state similar to CES1, which might favour crystallization. From the alignment of CES3 and CES1 (*table 5.1*) one of the glutamates (E191) required to introduce one of the two potential inter-subunit CES1 salt bridges into CES3 was conserved. This left three changes, R88K, S282K and G298E, to be made to the wt CES3 sequence. Again, as with CES2, there were problems generating the third mutant (G298E) which were overcome by re-designing the primers so that the Glu codon was changed from gaa → gag (forward) and ctt → ctc (reverse) (*see Primers for mutagenesis, Chapter* 2). Thus, two constructs were made: **CES3 2-3** (R88K, S282K) and **CES3 3-3** (R88K, S282K, G298E).

	Residues involved in CES1 trimer formation			
Gene	**Lys 78**	**Glu 183**	**Lys 275**	**Glu 291**
CES1	FV**K**NA	GD**E**HS	GC**K**TT	KT**E**EE
CES3	GV**R**DA	GD**E**HA	AC**S**SS	KE**G**EE

Table 5.1. Localised sequence alignment of CES1 and 3.
Resides involved in CES1 salt bridge formation that stabilize the trimeric interaction are highlighted and shown in bold. Residues in CES3 that were mutated to induce salt bridge formation are shown in blue. The residue shown in red is conserved between CES1 and 3.

Protein production was initially tested by small-scale transient expression in HEK293T, however only expression of wild type CES3 was detected in the cell media by anti-His Western blotting, and only at an extremely low level (*figure 5.1*). Surprisingly, none of the mutant proteins were secreted, and due to the constraints of time, these were not pursued further.

To produce sustainable levels of protein for downstream processes, stable HEK293 $Gnt^{-/-}$ cell lines were produced for wild type CES3 using co-selection with a neomycin resistance gene (*neo*) incorporated into the vector, for which 27 colonies expressing protein were picked. Those giving the highest signals were frozen down and stored at -80 °C. Clonal cell lines were expanded into roller bottles, and secreted proteins were purified from the cell media.

Recombinant CES3 was recovered from the media of transfected HEK293 cells by IMAC, and further purfied by size exclusion chromatography. CES3 gave a very similar elution profile to CES2 (*figure 5.2, a*) and eluted broadly from the size exclusion chromatography stage of purification at around ~ 60 kDa, corresponding to a monomeric species. Smaller peaks corresponding to larger aggregates (> 500 kDa) were also observed eluting from the S200 column prior to the main peak. Eluted fractions containing the protein were analysed by SDS-PAGE, pooled and concentrated before use in subsequent studies (*figure 5.2, b*). CES3 was highly soluble in aqueous buffer and could readily be concentrated to 48 mg/ mL. The protein was flash-frozen and stored at -80°C for long-term storage.

Recombinant CES3 proteins for functional studies have been produced previously only using the Sf21 insect expression system[21], with low yields of purified enzymes being reported at 0.7 mg/ L cell culture. The stable cell lines that were created for CES3 produced protein yields of 33.2 mg/ L, nearly 50 times the amount previously produced from the insect systems.

As with CES2, high expression levels were associated with problems of protein recovery during purification of CES3. Although yields obtained from the harvested cell media after the standard purification protocol were high, upon testing the flow through, large quantities of protein were still present (*figure 5.2, b*).

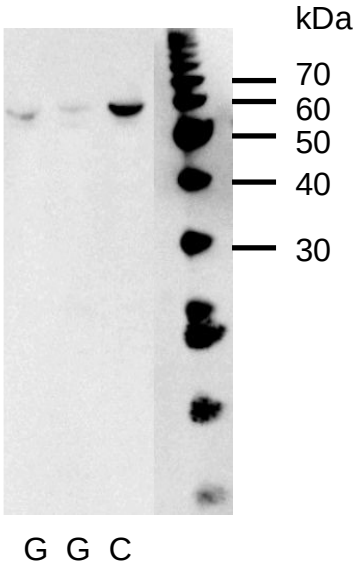
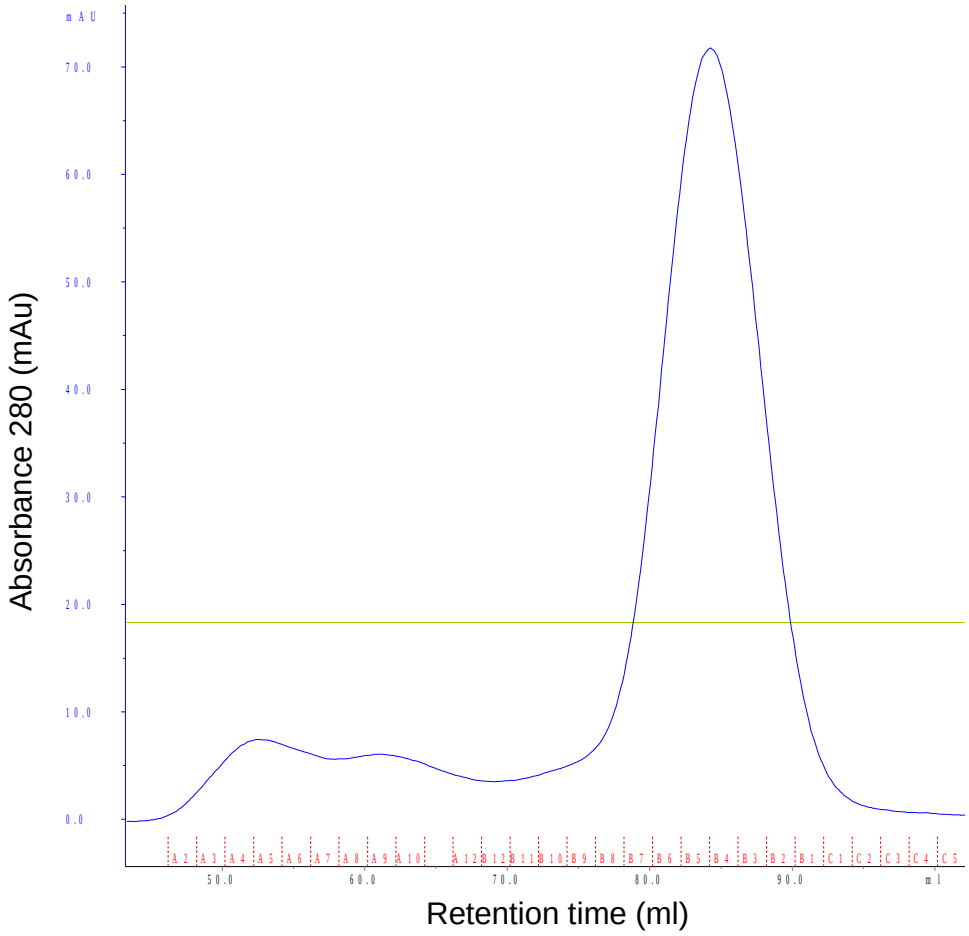


Figure 5.1. Western blot analysis for expression of CES3.
Small- scale transient expression in HEK293T. DNAs were transfected in duplicate and protein expression analysed by SDS-PAGE/Western blotting using an anti-His6 monoclonal antibody. The image shows detection from the cell media. G = CES3 in pOPINTTG neo, C = CES1 in pOPINTTGneo (used as a positive control). This image has been cropped as other protein samples were also run on this gel.

a.



b.

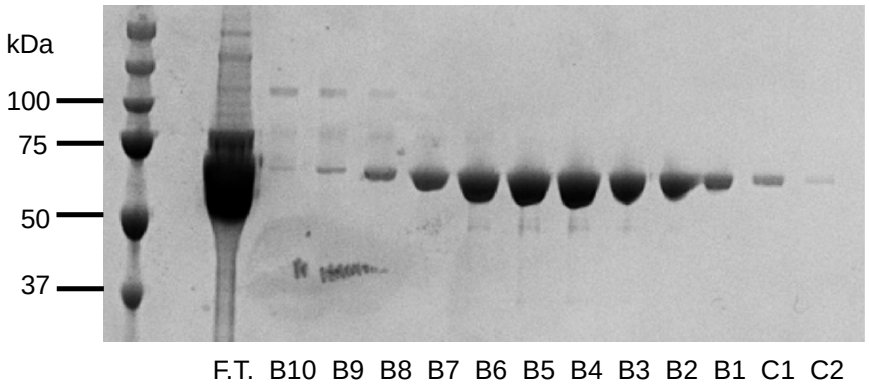


Figure 5.2. Gel filtration profile and SDS-PAGE gel of purified CES3.
a. The size exclusion gel filtration profile of CES3. The green horizontal line shows the steady flow-rate. A single UV trace peak that corresponds to approximately 60 kDa, with some aggregation at a much higher molecular weight is seen. ***b.*** The twelve fractions (B10–C2) that the peak corresponds to were run on an SDS-PAGE gel, along with the flow through (F.T.). A Precision Plus Protein™ Dual Color Standard ladder was run in the first lane. It can clearly be seen that although CES3 elutes as pure, monomeric protein, a considerable amount of it remains in the flow-through, and clearly did not bind efficiently to the nickel column.

5.3. Activity and Biophysical Analysis

5.3.1. Activity Analysis

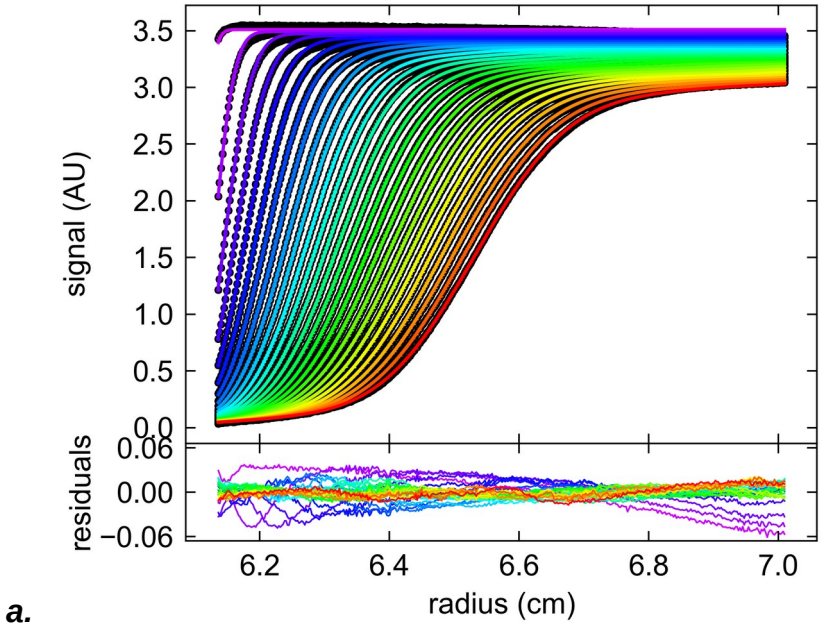
Because the only kinetic data available for CES3 is from assays on CPT-11 hydrolysis, for which it has minimal activity[21, 167], nothing is known about its kinetic mechanism or substrate specificity. Therefore, CES3 native was included in the activity assays along with CES1 and CES2.

No activity values could be obtained using substrate concentrations of less than 1.5 mM as CES3 showed an extremely low catalytic rate using 4-NPA as the substrate. Because very few data points were generated, initial velocity (AU min-1) and double reciprocal plots could not be constructed. The rate of product formation (pNP) was calculated to be 19.6 pmol min^{-1} μg^{-1} (± 0.5 SD). At the same substrate concentration, 1.5 mM, this was over 300 fold less active than CES1 native and 600 fold less than CES2 native. From these results, it can be seen that CES3 has significantly reduced enzymatic activity in comparison to the other two carboxylesterases, which has been noted previously[21, 111, 167, 168]. However, this does not necessarily mean that CES3 has no overall catalytic activity as 4-NPA and CPT-11 may not be preferred substrates of this enzyme. Carboxylesterase specificity experiments have predominantly been carried out on CES1, CES2, rCE and mES1, but very little has been done on CES3.

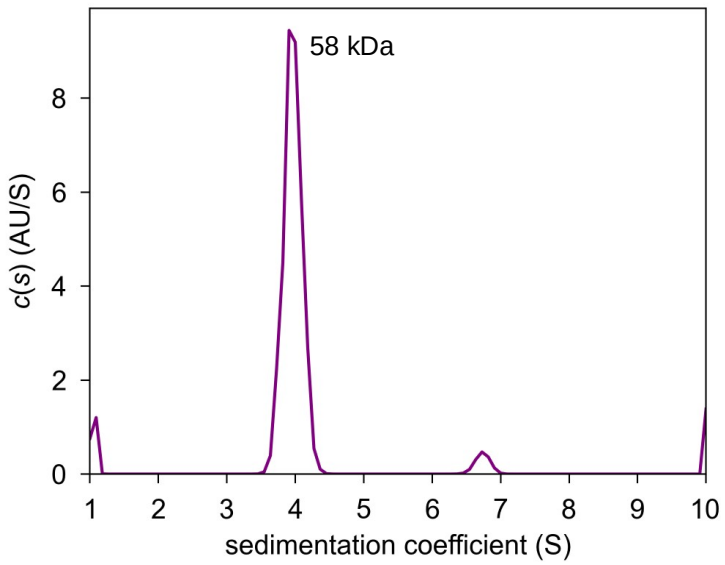
5.3.2. Analytical Ultra Centrifugation

Sedimentation velocity experiments showed that CES3 wt was predominantly monomeric (*figure 5.3)*, with an apparent molecular weight of 58 kDa. There is also the possible association of a dimer at higher molecular weight. The sedimentation coefficient

(*s*) was 4, and the frictional ratio (*f*) was 1.3. These results indicate that CES3 behaves in the same way as CES2 in solution.



a.



b.

Figure 5.3. Graphical AUC analysis for CES3 wt.
The plot of the residuals is shown in ***a,*** and the sedimentation coefficient population distribution (c(s)) against s plot is shown in ***b.*** The main peak corresponds to a protein with a molecular weight of 58 kDa.

5.4. Crystallization

CES3 was put into crystallization experiments at 12.0 mg/ mL using the standard screening pipeline described in *2.4.1*. This gave no crystals hits, therefore screening was repeated at different protein concentrations (4.0, 8.0, 16.0 and 25.0 mg/ mL), and co-crystallization with CPT-11. However no crystals were obtained, reminiscent of experience with the wt CES2. Unfortunately unlike for CES2, reversing the charge of the equivalent surface residues resulted in a lack of protein expression.

Like CES2, CES3 also fell into the 'difficult' crystallization class (4) using the XtalPred server[142], and combined with the fact that it primarily exists as a monomer, this furthers the possibility that structurally it is indeed more like CES2 than the readily crystallizable CES1, which falls into the second most promising crystallization class (2).

As discussed in *Chapter 4*, CES2 possesses the NNNEF motif, which signals the beginning of the unexpected strand extension exchange part of the molecule, which is absent in CES1. Interestingly, CES3 posseses a NNHEF motif (*appendix 1.2*), which could indicate that this may too have a strand extension, and be structurally more similar to CES2 than CES1. However, on structural alignment, CES3 does not possess the critical Cys residue in CES2 (C428) that is involved in the inter-molecular disulphide bridge in the core of the molecule.

5.5. Summary

Production of highly purified human CES3 using stable mammalian cell lines has been established, providing sufficient amounts of purified enzyme for downstream studies. From the preliminary work on CES3 presented here, it is evident that it actually behaves more like CES2 in solution, existing primarily as a monomer, sharing the highest percentage sequence identity (47%) out of all three functionally active human carboxylesterases (*Chapter 1, figure 1.2*). The wild type enzyme also proved similarly refractive to crystallization.

6. GENERAL DISCUSSION

6.1. Comparison of Human CES1, CES2 and CES3

The overall goal of this D.Phil project was to expand current knowledge on the structure and function of human carboxylesterases, and their ability to act on a wide range of substrates. The work described in this thesis addresses many of the key objectives set out for the project, and the significance and implications of these findings are addressed in the following paragraphs.

At the commencement of this project, there were only structures of one human carboxylesterase, CES1, in complex with compounds, along with one other carboxylesterase, rCE, deposited in the PDB. Over the past four years, only one more structure has been deposited; the *apo* form of human CES1 solved to a resolution of 2.20 Å. However, in this time, the number of publications concerning human carboxylesterases now exceeds ninety, many of which have called for a need for structural information on CES2.

Although CES1 and CES2 share 49% sequence identity and have overlapping substrate specificities, the findings presented here show clear differences between them.

6.1.1. Protein Production

These results demonstrate that using a mammalian expression system is superior to using insect cells both for protein quantity (at least tenfold higher) and quality in terms of authentic *N*-glycosylation. By controlling *N*-glycan processing using either a mutant cell line (stable $gnt^{-/-}$ cells) or by adding kifuenesine to cells, enzymes were produced with glycoforms that are characteristic of proteins resident in the endoplasmic reticulum. By contrast, insect cell *N*-glycoforms typically comprise short mannose/ fucosylated oligosaccharides not found on human glycoproteins. It is not obvious what the major

factors were in the significant increase in protein yield from mammalian compared to insect cells. However a contributing factor may have been the use of codon optimized sequences (*Appendix 1.3*). This could be tested by evaluating the expression of the native sequences of these three human carboxylesterases, and comparing them to the yields documented here.

6.1.2 Oligomerisation

A combination of AUC and SEC-MALLS were used to investigate how the three human carboxylesterases behaved in solution, and to complement the crystallographic experiments. It is important to study enzymes in solution at varying concentrations, to gain insight as to how they may behave *in vivo*. From these biophysical techniques, as well as from size exclusion chromatography profiles, it was shown that purified CES1 exists primarily as a trimer. By contrast, CES2 and CES3 behave primarily as monomers, although there is a hint of a higher order state both in the AUC and SEC-MALLS, consistent with previous analysis of purified CES2 by native gel electrophoresis[96]. It is important to note that in AUC and SEC-MALLS experiments, the concentration of protein is operationally limited to 2mg/ mL and 5mg / mL respectively.

To investigate the oligomerisation of proteins in solution at concentrations that are similar to those used in the crystallization experiments is not trivial. However, Small Angle X-Ray Solution Scattering (SAXS) would enable the oligomeric state of the protein to be assessed at these high protein concentrations (up to 30 mg/ mL). To this end, preliminary measurements have been made on beamline B21 at DLS for a subset of the CES constructs presented here which indicates that this will be a viable approach for assessing whether protein concentration affects the oligomeric state of the proteins in solution.

6.1.3. Activity

Although there are numerous publications that explore human carboxylesterase activity, the results here are the first to question the accepted view that both CES1 and CES2 follow Michaelis Menten kinetics. In the activity assays, at the lowest substrate concentration used (0.0938 mM 4-NPA), CES2 wt is nearly 12 x more active than CES1 wt. These results demonstrate that the relative activity of the wild type forms of these three carboxylesterases was CES2> CES1>> CES3, a pattern that has been seen before using different substrates[21, 32].

The role of *N*-glycosylation is clearly very different between CES1 and CES2. Full catalytic activity in CES2 requires the presence of both conserved glycans. The absence of each of these individually, and in combination, has a detrimental effect on activity, with the absence of N111+N276 causing the biggest decrease in activity, followed by N111, then N276. Conversely, the absence of the single glycosylation site in CES1, N79Q, does not compromise the enzymes activity, and it could be argued that it actually slightly improves the catalytic efficiency. In a very recent publication by Boonyuen et al.[110], in which CES1 was produced in *E .coli* and therefore not glycosylated, it was shown that glycosylation was not necessary for catalytic activity of CES1. Thus the role of glycosylation seems to be different depending on the enzyme, even within the same protein family.

Both CES1 and CES2 exhibit atypical enzyme kinetics, with CES2 showing the more pronounced departure from standard Michealis Menten kintetics such that a single kinetic model could not be fitted to the data generated for this enzyme. From the data presented here, both enzymes were shown to exhibit substrate inhibition, positive cooperativity and biphasic kinetics. For CES1, substrate inhibition only occurred at the highest concentration that was used in the assay (3.0 mM), whereas for CES2, inhibition occurred

progressively, from the lowest to highest concentration. This indicates that CES2 is more readily inhibited by excess substrate.

6.1.4. Structural Comparisons

In structural biology, a high resolution data set obtained from diffracting protein crystals is still the most important and strived for goal. In terms of human carboxylesterases, possessing crystal structures of them all is vital in understanding their differences in selectivity and mode of action, and only when this is known it may be possible to predict and potentially improve therapy, as well as designing enzyme specific drugs and inhibitors.

In this project, the structure of human CES2 has been determined for the first time. In addition, several high resolution structures of three forms of human CES1: wt, null and N79Q were solved. Apart from specific new details, such as the potential role for C389 substrate binding, the CES1 structures are generally consistent with the lower resolution structures that have previously been reported. However the CES2 structure was not as expected[139], and contained an unusual feature comprising a large helical extension region of approximately 90 residues stabilized by a strand exchange between two monomers which corresponds to the entire regulatory domain in CES1.

When overlaying the structures of CES1 and CES2 (*figure 6.1, table 6.1*), it is evident that the protein core consisting of the central catalytic and α/β domains is structurally conserved between the two enzymes. It is only the arrangement of the regulatory domain that differs between the two. A possible cause for this structural difference is the run of the three asparagine residues adjacent to the catalytic glutamate (**NNN**EF). These residues are upstream of this helical motif, and could induce this extension. The absence of these residues in the sequence of CES1 (NKQEF) corroborates this hypothesis. This could be tested by mutating the relevant residues in CES1 to the corresponding ones in CES2, and

assuming this does not have a detrimental effect on crystallization, determining the structure of the mutant enzyme.

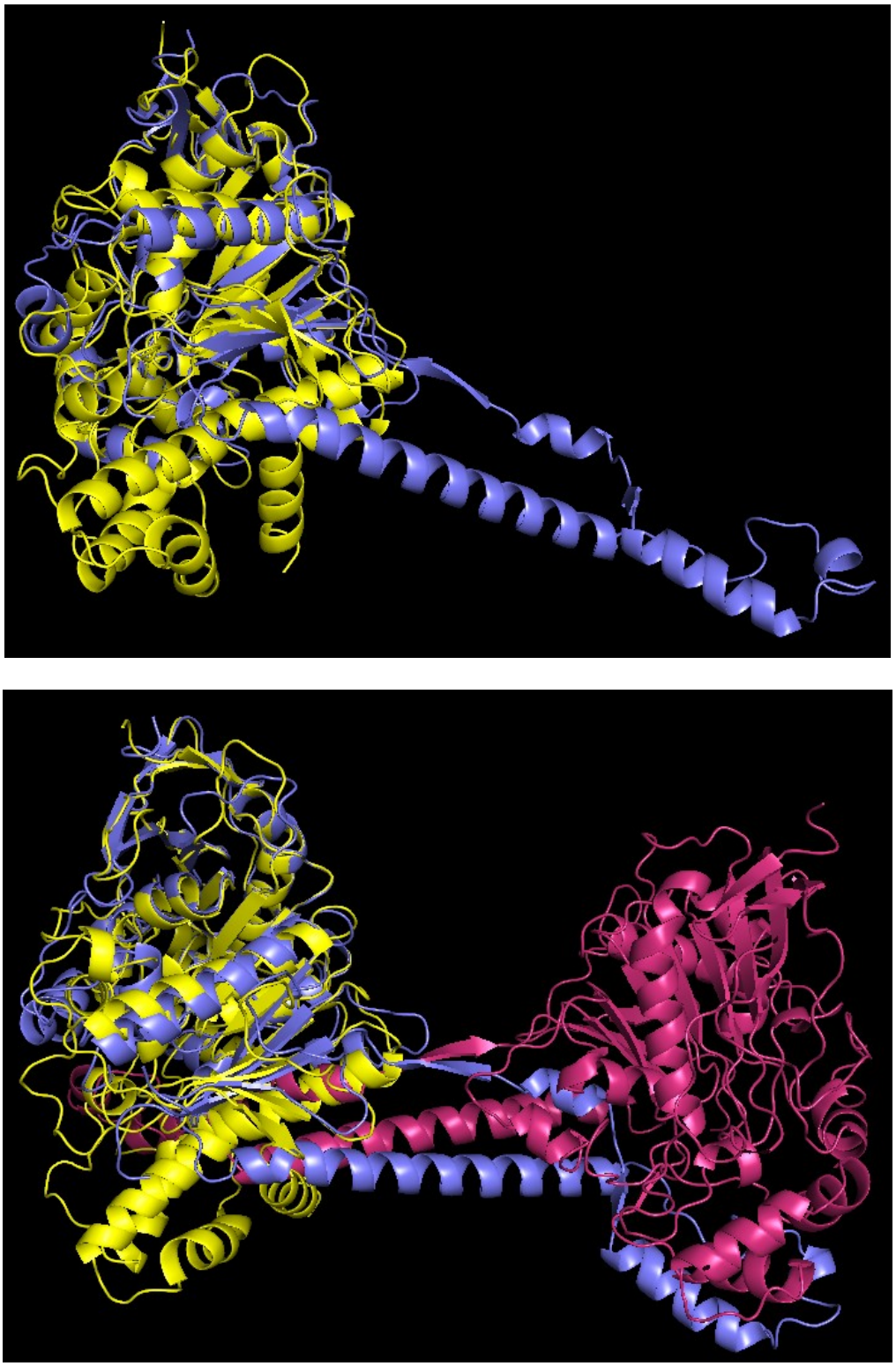


Figure 6.1. Structural overlay of CES1 and CES2.
CES1 wt is shown in yellow. CES2 wt is shown in slate (monomer) and pink (other monomer that comprises the dimer). The top image shows CES1 overlayed with chain A in CES2. The bottom image includes chain B from CES2, to show the tertiary structure of the dimer.

CES2 Chain	RMSD (Å)	N_{align}	N_{gaps}

A	1.47	347	15
B	1.52	348	15
C	1.44	345	15
D	1.42	348	15

Table 6.1. Structure alignment results from PDBeFold.
Aligned against CES1 wt, from the structure presented in *Chapter 3*. RMSD was calculated between Ca-atoms of matched residues at best 3D superposition of the query and target structures. The larger the RMSD, the more distant the matched structures are. The length of alignment, N_{align}, or number of matched residues, is calculated at best 3D superposition of query and target structures. The residues are aligned in 3D on the basis of their spatial closeness. The orientation of the structures is optimized such as to minimize RMSD and to maximize the number of aligned residues. Number of gaps, $N_{gaps,}$ is defined as a consecutive sequence section (containing 1 or more residues) of query chain that cannot be matched to any part of target chain, found between two sections of matched residues.

Because of the presence of multiple binding sites in CES1, it is important to identify if any of these promiscuous sites, or other non-specific binding sites are also present in CES2. From the kinetic data and the possible positive cooperativity observed, it is likely that CES2 also possesses alternative binding sites. Looking at the residues involved in the Z-site interface in CES1, which are located in the regulatory domain of the enzyme, most of these are not conserved in CES2, apart from G347 and W458, which are situated on the helical strand extension. Thus the direct equivalent of the CES1 Z-site is not found in the CES2 structure.

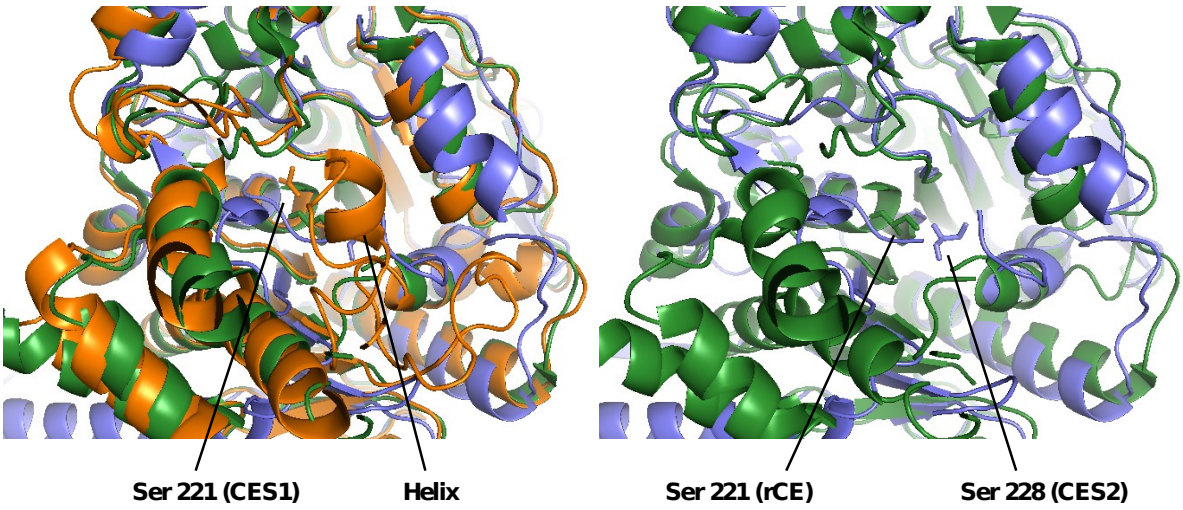
By contrast, two residues involved in the side-door binding site on CES1 (M425, F426) are also conserved in CES2 (M415, F416). This may indicate that CES2 also possesses a side door, like the putative product exit pore in rCE, where 4-PP was observed to be bound[12]. This is a feature that has long been postulated to facilitate the release of small products from the active site of AcChE[169].

6.1.5. Comparison of Rabbit and Human CES Structures

Rabbit CE (rCE) is currently the most efficient mammalian enzyme known for hydrolysing CPT-11[20], and CES2 is the most efficient human carboxylesterase to do so (approximately 100 fold more active than CES1)[140]. To investigate whether rCE and human CES2 share any structural similarities not found in CES1, the structures of the three enzymes were superimposed. As discussed in Chapter 6, the active site of CES2 is highly exposed and easily accessible, and as shown in *figure 6.2*, this is also the case for rCE. This 'open' and large active site explains why both CES2 and rCE are able to hydrolyse large, bulky compounds such as CPT-11. In the rCE structure, it has been proposed that the two disordered 16-residue loops (355-370 and 450-466), not observed in the structure, partially close over the entrance to the active site. The corresponding loops in CES2 are also un-resolved implying similar structural flexibility.

Overlaying the structure of CES1, it can be seen that the active site is blocked by a small helix, which starts from residue 357 (**WLIPM**), which is not the case in rCE and CES2. This may explain why CES1 is unable to hydrolyse large compounds, as they physically cannot gain access into the catalytic site. In the sequence alignment of CES2 and rCE, this conserved helix motif has one substituted residue in CES2 (WLIP**K**), and one in rCE (W**I**IPM). In the structure of CES2, this helix is at the beginning of the helical strand extension, and in rCE, there is no density present for this region, potentially indicating that it too may have a flexible extended helix.

a.



b

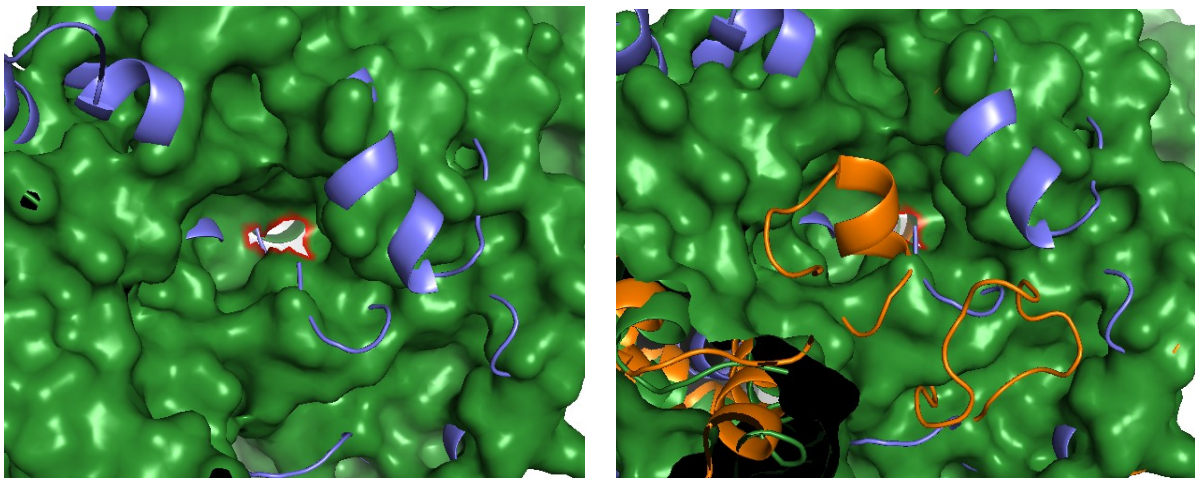


Figure 6.2. Superposition of the active sites of CES1, CES2 and rCE.
CES1 wt is shown in orange, CES2 wt is shown in slate and rCE is shown in green. In ***a,*** the catalytic serines of each enzyme are shown as sticks, and indicated. The short helix motif in CES1 that covers the entrance to the active site is also shown. In ***b***, the surface representation of rCE is shown with its catalytic Ser shown in red. A cartoon view of CES2 is also shown. On the left, it is evident that the active site is exposed and easily accessible. On the right, an overlay of CES1 is shown, and it can be seen that the short helix covers the entrance to the active site.

In conclusion, the successful structure determination of CES2 provides a platform for comprehensive structure function analysis of this important human enzyme. Potential applications include tissue/cell specific drug targeting, potential treatment of drug overdose, the design of herbicides with selective toxicity[64], use as prophylactic agents against toxic esters[97] and in clinical pro-drug activation.

6.2. Future Work

The work described in this thesis documents the establishment of a platform for the robust production and crystallization of the key human carboxylesterases. CES1 has been studied extensively with respect to its structure and substrate specificity, but there is an immense field to be explored concerning the structure and function of other human and mammalian carboxylesterases,

Now that the methodology for the successful crystallization of CES2 has been secured, co-crystallization of this enzyme with other compounds to explore its selectivity should be pursued, as well as determining the crystal structure of CES3. Apart from CES1, CES2 and rCE, there are no structures of any other mammalian carboxylesterase available. Because these enzymes seem to be omnipresent in mammals, with some species having more than ten homologues in their genome, this severely limits the understanding of this vast family of promiscuous proteins. A focused effort could be put into obtaining more crystallographic structures from a variety of species to aid drug metabolism studies.

Production and crystallization of the single mutant in CES2, D281K, should be pursued to see if, as speculated in Chapter 4, this single amino acid substitution is all that is needed to aid crystallization of CES2. Efforts should also be made in generating stable cell lines for CES3 2-3, 3-3 and CES3_trunc to produce recombinant protein for crystallization studies.

During the purification of CES2, it would be interesting to do this in an oxidizing environment to see if this would promote disulphide formation, and subsequent oligomerisation of this enzyme in solution.

Molecular modelling experiments should be carried on CES2, with a variety of compounds that are *i*) hydrolysed by CES2 alone, *ii*) hydrolysed by both CES1 and CES2 and *iii*) hydrolysed by CES1 alone. These protein-ligand docking experiments would help to rationalise the exact specificity of CES2, and further understanding into the overlapping specificity of these enzymes. The determination of the structure of CES2 provides the platform for these experiments.

The kinetic data presented in this thesis opens up a whole range of future experiments that need to be conducted to fully dissect the complex catalytic mechanism displayed by these enzymes. Because CES enzymes metabolize such a vast number of prescribed drugs, it is important to understand their kinetic characteristics. To really understand how human carboxylesterases function allosterically, research needs to be carried out on identifying the exact allosteric binding sites that contribute to the positive cooperativity that they exhibit. Individually mutating out the residues in the Z-site, side-door and C389 and then performing kinetic analysis would help to rationalize this in CES1. However, CES2 is more complex, and firstly alternative binding sites need to be identified, as well as dissecting the complex catalytic mechanism that it exhibits.

Elucidating the exact function of the 'side-door' in catalysis, e.g. by mutating residues that line this entrance, and then performing kinetic analysis, would give an insight into the exact role of this region in enzyme catalysis.

No work has yet been carried out on the involvement of glycosylation in CES3. It possesses one glycosylation site: Asn 105 and on alignment with CES2, this is seven

amino acids downstream of the first site, Asn 111, which is the most important glycan in terms of activity. It would be interesting to see if knocking out glycosylation in CES3 impacts catalytic ability, like with CES2, or has no detrimental effect on it, like CES1. It would also be curious to see if the native form of the enzyme exhibits any atypical kinetic profiles. Identifying compounds or potentially lipids that are efficiently hydrolysed by CES3 would help to determine its catalytic activity, elucidate its substrate specificity and provide further insight into its exact role biologically. It would also aid further kinetic analysis and help with co-crystallization studies.

APPENDICES

Appendix 1.1

The *N*-terminal signal peptide and the *C*-terminal endoplasmic reticulum retention signal are both shown in *italics* for each sequence.
Residues in the catalytic triad.
***N*-linked glycosylation.**

CES1 native

566 residues (with signal sequence). Mature secreted protein 548 residues.

*MWLRAFILATLSASAAWG*HPSSPPVVDTVHGKVLGKFVSLEGFAQPVAIFLGIPFAKPPLGPLRFT
PPQPAEPWSFVKNATSYPPMCTQDPKAGQLLSELFTNRKENIPLKLSEDCLYLNIYTPADLTKKNR
LPVMVWIHGGGLMVGAASTYDGLALAAHENVVVVTIQYRLGIWGFFSTGDEHSRGNWGHLDQV
AALRWVQDNIASFGGNPGSVTIFGESAGGESVSVLVLSPLAKNLFHRAISESGVALTSVLVKKGDV
KPLAEQIAITAGCKTTTSAVMVHCLRQKTEEELLETTLKMKFLSLDLQGDPRESQPLLGTVIDGMLL
LKTPEELQAERNFHTVPYMVGINKQEFGWLIPMLMSYPLSEGQLDQKTAMSLLWKSYPLVCIAKE
LIPEATEKYLGGTDDTVKKKDLFLDLIADVMFGVPSVIVARNHRDAGAPTYMYEFQYRPSFSSDMK
PKTVIGDHGDELFSVFGAPFLKEGASEEEIRLSKMVMKFWANFARNGNPNGEGLPHWPEYNQKE
GYLQIGANTQAAQKLKDKEVAFWTNLFAKKAVEKPPQTE*HIEL*

CES2 native

559 residues (with signal sequence). Mature secreted protein 533 residues.

*MRLHRLRARLSAVACGLLLLLVRGQG*QDSASPIRTTHTGQVLGSLVHVKGANAGVQTFLGIPFAK
PPLGPLRFAPPEPPESWSGVRDGTTHPAMCLQDLTAVESEFLSQFNMTFPSDSMSEDCLYLSIYT
PAHSHEGSNLPVMVWIHGGALVFGMASLYDGSMLAALENVVVVIIQYRLGVLGFFSTGDKHATGN
WGYLDQVAALRWVQQNIAHFGGNPDRVTIFGESAGGTSVSSLVVSPISQGLFHGAIMESGVALLP
GLIASSADVISTVVANLSACDQVDSEALVGCLRGKSKEEILAINKPFKMIPGVVDGVFLPRHPQELL
ASADFQPVPSIVGVNNNEFGWLIPKVMRIYDTQKEMDREASQAALQKMLTLLMLPPTFGDLLREE
YIGDNGDPQTLQAQFQEMMADSMFVIPALQVAHFQCSRAPVYFYEFQHQPSWLKNIRPPHMKAD
HGDELPFVFRSFFGGNYIKFTEEEEQLSRKMMKYWANFARNGNPNGEGLPHWPLFDQEEQYLQ
LNLQPAVGRALKAHRLQFWKKALPQKIQELEEPEER*HTEL*

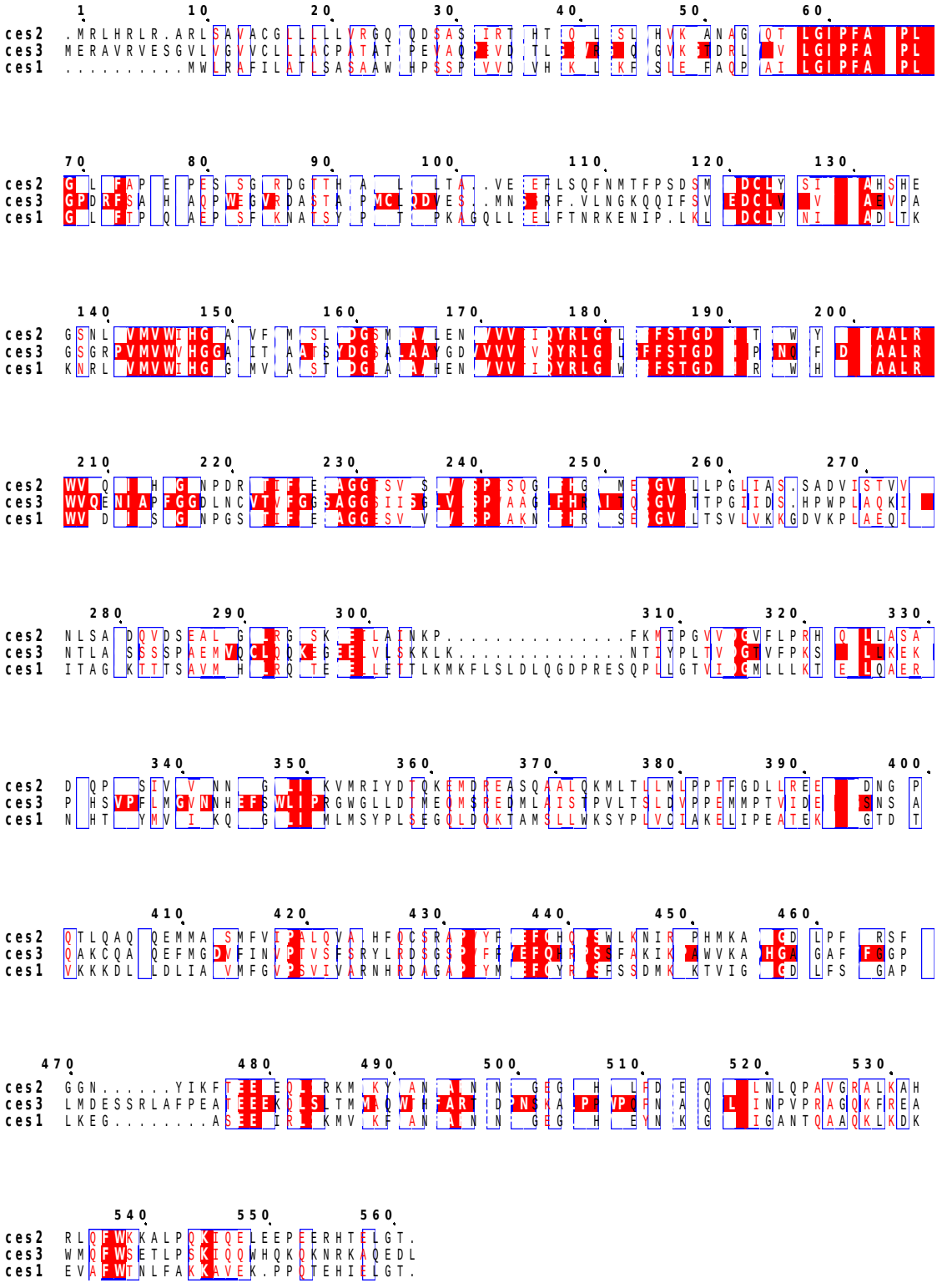
CES3 native

571 residues (with signal sequence). Mature secreted protein 545 residues.

*MERAVRVESGVLVGVVCLLLACPATA*TGPEVAQPEVDTTLGRVRGRQVGVKGTDRLVNVFLGIP
FAQPPLGPDRFSAPHPAQPWEGVRDASTAPPMCLQDVESMNSSRFVLNGKQQIFSVSEDCLVL
NVYSPAEVPAGSGRPVMVWVHGGALITGAATSYDGSALAAYGDVVVVTVQYRLGVLGFFSTGDE
HAPGNQGFLDVVAALRWVQENIAPFGGDLNCVTVFGGSAGGSIISGLVLSPVAAGLFHRAITQSG
VITTPGIIDSHPWPLAQKIANTLACSSSSPAEMVQCLQQKEGEELVLSKKLKNTIYPLTVDGTVFPK
SPKELLKEKPFHSVPFLMGVNNHEFSWLIPRGWGLLDTMEQMSREDMLAISTPVLTSLDVPPEM
MPTVIDEYLGSNSDAQAKCQAFQEFMGDVFINVPTVSFSRYLRDSGSPVFFYEFQHRPSSFAKIK
PAWVKADHGAEGAFVFGGPFLMDESSRLAFPEATEEEKQLSLTMMAQWTHFARTGDPNSKALP
PWPQFNQAEQYLEINPVPRAGQKFREAWMQFWSETLPSKIQQWHQKQKNRKA*QEDL*

µ-phosphatase leader sequence in pOPING
MGILPSPGMPALLSLVSLLSVLLMGCVAET

Appendix 1.2



Sequence alignment of CES1, 2 and 3

Appendix 1.3

Human CES1

```
native          ATGTGGCTCCGTGCCTTTATCCTGGCCACTCTCTCTGCTTCCGCGGCTTGGGGGCATCCG
optimised       ATGTGGCTGCGCGCCTTCATCCTGGCCACCCTGTCTGCCTCTGCCGCCTGGGGACACCCT
                ******** ** ***** ********** ** ***** ** ** ** ***** ** **

native          TCCTCGCCACCTGTGGTGGACACCGTGCATGGCAAAGTGCTGGGGAAGTTCGTCAGCTTA
optimised       TCCTCTCCCCCTGTGGTGGACACCGTCCACGGCAAGGTGCTGGGCAAGTTCGTGTCCCTG
                ***** ** ***************** ** ***** ******** ********   * *

native          GAAGGATTTGCACAGCCTGTGGCCATTTTCCTGGGAATCCCTTTTGCCAAGCCGCCTCTT
optimised       GAAGGCTTCGCCCAGCCCGTGGCTATCTTCCTGGGCATCCCCTTCGCCAAGCCCCCCCTG
                ***** ** ** ***** ***** ** ******** ***** ** ******** ** **

native          GGACCCCTGAGGTTTACTCCACCGCAGCCTGCAGAACCATGGAGCTTTGTGAAGAATGCC
optimised       GGCCCTCTGAGGTTCACCCCTCCACAGCCTGCCGAGCCCTGGTCCTTCGTGAAGAACGCC
                ** ** ******** ** ** ** ******** ** ** ***  *** ******** ***

native          ACCTCGTACCCTCCTATGTGCACCCAAGATCCCAAGGCGGGGCAGTTACTCTCAGAGCTA
optimised       ACCTCCTACCCCCCCATGTGCACCCAGGACCCCAAGGCCGGACAGCTCCTGTCCGAGCTG
                ***** ***** ** ********** ** ******** ** *** * ** ** *****

native          TTTACAAACCGAAAGGAGAACATTCCTCTCAAGCTTTCTGAAGACTGTCTTTACCTCAAT
optimised       TTCACCAACCGCAAGGAAAACATCCCCCTGAAGCTGTCCGAGGACTGCCTGTACCTGAAC
                ** ** ***** ***** ***** ** ** ***** ** ** ***** ** ***** **

native          ATTTACACTCCTGCTGACTTGACCAAGAAAAACAGGCTGCCGGTGATGGTGTGGATCCAC
optimised       ATCTACACCCCTGCCGACCTGACCAAGAAGAACCGCCTGCCCGTCATGGTCTGGATCCAC
                ** ***** ***** *** ********** *** * ***** ** ***** *********

native          GGAGGGGGGCTGATGGTGGGTGCGGCATCAACCTATGATGGGCTGGCCCTTGCTGCCCAT
optimised       GGCGGAGGCCTGATGGTCGGAGCCGCCTCTACCTACGACGGACTGGCCCTGGCCGCCCAC
                ** ** ** ******** ** ** ** ** ***** ** ** ******** ** *****

native          GAAAACGTGGTGGTGGTGACCATTCAATATCGCCTGGGCATCTGGGGATTCTTCAGCACA
optimised       GAGAACGTGGTGGTGGTGACCATCCAGTACCGCCTGGGCATCTGGGGATTCTTCAGCACC
                ** ********************* ** ** ****************************

native          GGGGATGAACACAGCCGGGGGAACTGGGGTCACCTGGACCAGGTGGCTGCCCTGCGCTGG
optimised       GGCGACGAGCACTCCCGCGGAAACTGGGGCCACCTGGATCAGGTCGCCGCTCTGCGCTGG
                ** ** ** ***  *** ** ******** ******** ***** ** ** *********

native          GTCCAGGACAACATTGCCAGCTTTGGAGGGAACCCAGGCTCTGTGACCATCTTTGGAGAG
optimised       GTGCAGGACAACATCGCTTCCTTCGGCGGCAACCCCGGCTCCGTGACCATCTTCGGAGAG
                ** *********** **   *** ** ** ***** ***** *********** ******

native          TCAGCGGGAGGAGAAAGTGTCTCTGTTCTTGTTTTGTCTCCATTGGCCAAGAACCTCTTC
optimised       TCTGCCGGCGGAGAGTCCGTGTCCGTGCTGGTGCTGTCCCCCCTGGCCAAGAACCTGTTC
                ** ** ** *****    ** ** ** ** **  **** **  ************** ***

native          CACCGGGCCATTTCTGAGAGTGGCGTGGCCCTCACTTCTGTTCTGGTGAAGAAAGGTGAT
optimised       CACCGCGCCATCTCCGAGTCCGGCGTGGCCCTGACCTCTGTGCTGGTGAAGAAGGGCGAC
                ***** ***** ** ***   *********** ** ***** *********** ** **

native          GTCAAGCCCTTGGCTGAGCAAATTGCTATCACTGCTGGGTGCAAAACCACCACCTCTGCT
optimised       GTCAAGCCCCTGGCCGAGCAGATCGCCATCACCGCCGGCTGCAAGACCACCACCTCCGCT
                ********* **** ***** ** ** ***** ** ** ***** *********** ***

native          GTCATGGTTCACTGCCTGCGACAGAAGACGGAAGAGGAGCTCTTGGAGACGACATTGAAA
optimised       GTGATGGTGCACTGCCTGCGCCAGAAGACCGAGGAAGAACTGCTGGAAACTACCCTGAAG
                ** ***** *********** ******** ** ** ** **  **** ** **  ****

native          ATGAAATTCTTATCTCTGGACTTACAGGGAGACCCCAGAGAGAGTCAACCCCTTCTGGGC
optimised       ATGAAGTTCCTGAGCCTGGACCTGCAGGGCGACCCTAGAGAGTCCCAGCCCCTGCTGGGC
                ***** *** *    ****** * ***** ***** ******   ** ***** ******
```

```
native          ACTGTGATTGATGGGATGCTGCTGCTGAAAACACCTGAAGAGCTTCAAGCTGAAAGGAAT
optimised       ACCGTGATCGACGGCATGCTGCTGCTGAAGACCCCCGAGGAACTGCAGGCCGAGCGCAAC
                ** ***** ** ** ************** ** ** ** ** ** ** ** **  * **

native          TTCCACACTGTCCCCTACATGGTCGGAATTAACAAGCAGGAGTTTGGCTGGTTGATTCCA
optimised       TTCCACACCGTGCCCTACATGGTCGGAATCAACAAGCAGGAATTCGGCTGGCTGATCCCC
                ******** ** ***************** *********** ** ****** **** **

native          ATGCAGTTGATGAGCTATCCACTCTCCGAAGGGCAACTGGACCAGAAGACAGCCATGTCA
optimised       ATGC---TGATGTCCTACCCCCTGTCCGAGGGCCAGCTGGACCAAAAGACTGCCATGTCC
                ****   *****  *** ** ** ***** ** ** ******** ***** ********

native          CTCCTGTGGAAGTCCTATCCCCTTGTTTGCATTGCTAAGGAACTGATTCCAGAAGCCACT
optimised       CTGCTGTGGAAGTCCTACCCTCTGGTCTGCATCGCCAAGGAACTGATCCCCGAGGCCACC
                ** ************** ** ** ** ***** ** *********** ** ** *****

native          GAGAAATACTTAGGAGGAACAGACGACACTGTCAAAAAGAAAGACCTGTTCCTGGACTTG
optimised       GAGAAGTACCTGGGCGGAACCGACGACACCGTGAAGAAGAAGGACCTGTTCCTGGACCTG
                ***** *** * ** ***** ******** ** ** ***** *************** **

native          ATAGCAGATGTGATGTTTGGTGTCCCATCTGTGATTGTGGCCCGGAACCACAGAGATGCT
optimised       ATCGCCGACGTGATGTTCGGCGTGCCCTCCGTGATCGTGGCCCGCAACCACAGGGACGCT
                ** ** ** ******** ** ** ** ** ***** ******** ******** ** ***

native          GGAGCACCCACCTACATGTATGAGTTTCAGTACCGTCCAAGCTTCTCATCAGACATGAAA
optimised       GGCGCCCCTACCTACATGTACGAGTTCCAGTACAGGCCCTCCTTCAGCTCCGACATGAAG
                ** ** ** *********** ***** ****** * **   ****   ** ********

native          CCCAAGACGGTGATAGGAGACCACGGGGATGAGCTCTTCTCCGTCTTTGGGGCCCCATTT
optimised       CCCAAGACCGTGATCGGTGACCACGGCGACGAACTGTTCTCCGTGTTCGGAGCCCCATTC
                ******** ***** ** ******** ** ** ** ******** ** ** ********

native          TTAAAAGAGGGTGCCTCAGAAGAGGAGATCAGACTTAGCAAGATGGTGATGAAATTCTGG
optimised       CTGAAGGAAGGCGCCTCCGAGGAAGAGATCCGCCTGTCCAAGATGGTGATGAAGTTCTGG
                 * ** ** ** ***** ** ** ****** * **   *************** ******

native          GCCAACTTTGCTCGCAATGGAAACCCCAATGGGGAAGGGCTGCCCCACTGGCCAGAGTAC
optimised       GCCAACTTCGCCCGCAACGGCAACCCCAACGGCGAGGGACTGCCCCACTGGCCCGAGTAC
                ******** ** ***** ** ******** ** ** ** ************** ******

native          AACCAGAAGGAAGGGTATCTGCAGATTGGTGCCAACACCCAGGCGGCCCAGAAGCTGAAG
optimised       AACCAGAAGGAAGGCTACCTGCAGATCGGCGCCAACACCCAGGCCGCCCAGAAGCTGAAG
                ************** ** ******** ** ************** ***************

native          GACAAAGAAGTAGCTTTCTGGACCAACCTCTTTGCCAAGAAGGCAGTGGAGAAGCCACCC
optimised       GACAAGGAAGTGGCCTTCTGGACCAACCTGTTCGCCAAGAAGGCTGTCGAAAAGCCCCCC
                ***** ***** ** ************** ** *********** ** ** ***** ***

native          CAGACAGAACACATAGAGCTG
optimised       CAGACCGAGCACATCGAGCTG
                ***** ** ***** ******
```

Human CES2

```
native          ATGACTGCTCAGTCCCGCTCTCCTACCACACCCACCTTTCCCGGCCCAAGCCAGCGCACC
optimised       ATGACCGCCCAGTCCAGGTCCCCCACCACCCCTACCTTCCCCGGACCTTCCCAGAGGACC
                ***** ** ****** * ** ** ***** ** ***** ***** **   **** * ***

native          CCGCTGACTCCCTGCCCAGTCCAAACTCCAAGGCTGGGCAAGGCACTGATCCACTGCTGG
optimised       CCCCTGACTCCTTGCCCTGTGCAGACCCCTAGGCTGGGCAAGGCCCTGATCCACTGCTGG
                ** ******** ***** ** ** ** ** ************** ***************

native          ACAGACCCGGGGCAGCCTCTGGGTGAACAGCAGCGTGTCCGCCGGCAGCGAACCGAGACC
optimised       ACCGACCCCGGACAGCCCCTGGGAGAACAGCAGAGAGTGCGCCGCCAGCGCACCGAAACC
                ** ***** ** ***** ***** ********* * ** ***** ***** ***** ***

native          AGCGAGCCGACCATGCGGCTGCACAGACTTCGTGCGCGGCTGAGCGCGGTGGCCTGTGGG
optimised       TCCGAGCCTACCATGCGCCTGCACAGGCTGAGGGCCAGGCTGTCTGCCGTGGCTTGCGGA
                  ****** ******** ******** **  * **  *****   ** ***** ** **

native          CTTCTGCTGCTTCTTGTCCGGGGCCAGGGCCAGGACTCAGCCAGTCCCATCCGGACCACA
optimised       CTGCTGCTGCTGCTGGTCCGTGGACAGGGACAGGACTCCGCCTCCCCAATCAGGACCACC
                ** ******** ** ***** ** ***** ******** ***   ** *** *******

native          CACACGGGGCAGGTGCTGGGGAGTCTTGTCCATGTGAAGGGCGCCAATGCCGGGGTCCAA
optimised       CACACCGGACAGGTGCTGGGATCCCTGGTGCACGTGAAGGGCGCCAACGCTGGCGTGCAG
                ***** ** ***********    ** ** ** ************** ** ** ** **

native          ACCTTCCTGGGAATTCCATTTGCCAAGCCACCTCTAGGTCCGCTGCGATTTGCACCCCCT
optimised       ACCTTCCTGGGCATCCCTTTCGCCAAGCCCCCCCTGGGACCTCTGAGATTCGCCCCTCCC
                *********** ** ** ** ******** ** ** ** ** *** **** ** ** **

native          GAGCCCCCTGAATCTTGGAGTGGTGTGAGGGATGGAACCACCCATCCGGCCATGTGTCTA
optimised       GAGCCCCCCGAGTCTTGGAGCGGAGTGCGCGACGGAACCACCCACCCTGCCATGTGCCTG
                ******** ** ******** ** *** * ** *********** ** ******** **

native          CAGGACCTCACCGCAGTGGAGTCAGAGTTTCTTAGCCAGTTCAACATGACCTTCCCTTCC
optimised       CAAGACCTGACCGCCGTGGAATCCGAGTTCCTGTCCCAGTTCAACATGACCTTCCCTTCC
                ** ***** ***** ***** ** ***** **   *************************

native          GACTCCATGTCTGAGGACTGCCTGTACCTCAGCATCTACACGCCGGCCCATAGCCATGAA
optimised       GACTCCATGTCCGAGGACTGCCTGTACCTGTCCATCTACACTCCAGCCCACTCCCACGAG
                *********** *****************   ********* ** *****   *** **

native          GGCTCTAACCTGCCGGTGATGGTGTGGATCCACGGTGGTGCGCTTGTTTTTGGCATGGCT
optimised       GGCTCCAACCTGCCCGTGATGGTCTGGATCCACGGTGGTGCCCTGGTGTTCGGCATGGCC
                ***** ******** ******** ***************** ** ** ** ********

native          TCCTTGTATGATGGTTCCATGCTGGCTGCCTTGGAGAACGTGGTGGTGGTCATCATCCAG
optimised       TCCCTGTACGACGGCTCCATGCTGGCCGCCCTGGAAAACGTGGTGGTGGTGATCATCCAG
                *** **** ** ** *********** *** **** ************** *********

native          TACCGCCTGGGTGTCCTGGGCTTCTTCAGCACTGGAGACAAGCACGCAACCGGCAACTGG
optimised       TACCGCCTGGGCGTGCTGGGCTTCTTCAGCACCGGCGACAAGCACGCCACCGGCAACTGG
                *********** ** ***************** ** *********** ************

native          GGCTACCTGGACCAAGTGGCTGCACTACGCTGGGTCCAGCAGAATATCGCCCACTTTGGA
optimised       GGCTACCTGGATCAGGTGGCCGCTCTGCGCTGGGTGCAGCAGAACATCGCCCACTTCGGC
                *********** ** ***** ** ** ******** ******** *********** **

native          GGCAACCCTGACCGTGTCACCATTTTTGGCGAGTCTGCGGGTGGCACGAGTGTGTCTTCG
optimised       GGCAACCCCGACCGCGTGACCATCTTCGGAGAGTCTGCCGGCGGAACCTCCGTGTCCTCC
                ******** ***** ** ***** ** ** ******** ** ** **    ***** **

native          CTTGTTGTGTCCCCCATATCCCAAGGACTCTTCCACGGAGCCATCATGGAGAGTGGCGTG
optimised       CTGGTGGTGTCCCCCATCTCCCAGGGACTGTTCCACGGCGCCATCATGGAATCCGGCGTG
                ** ** *********** ***** ***** ******** ***********    ******




native          GCCCTCCTGCCCGGCCTCATTGCCAGCTCAGCTGATGTCATCTCCACGGTGGTGGCCAAC
optimised       GCCCTGCTGCCCGGACTGATCGCCTCCTCCGCCGACGTGATCTCCACCGTGGTGGCCAAC
```

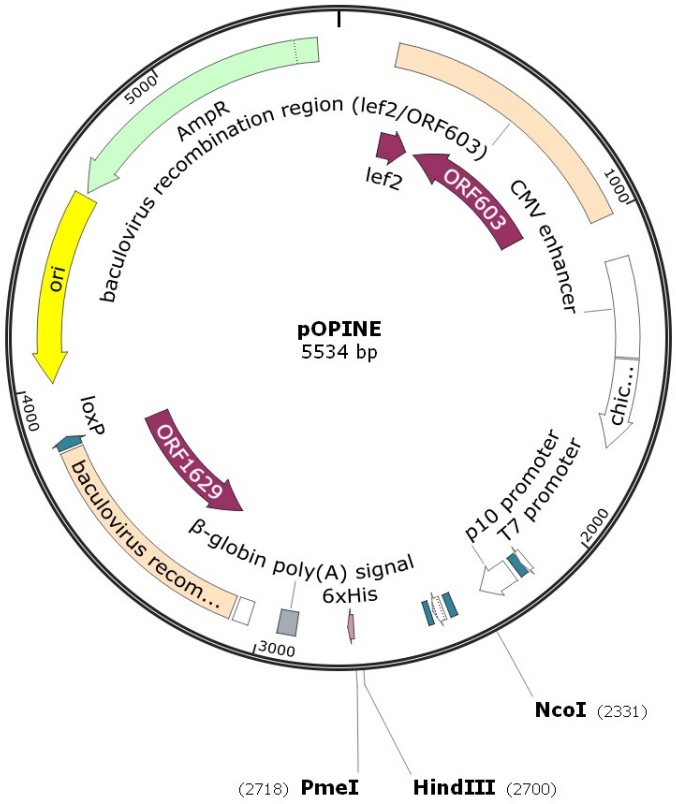
```
                ***** ******** ** ** ***  *** ** ** ** ******** ************

native          CTGTCTGCCTGTGACCAAGTTGACTCTGAGGCCCTGGTGGGCTGCCTGCGGGGCAAGAGT
optimised       CTGTCCGCCTGCGACCAGGTGGACTCCGAGGCTCTGGTCGGATGCCTGCGCGGCAAGTCC
                ***** ***** ***** ** ***** ***** ***** ** ******** ******

native          AAAGAGGAGATTCTTGCAATTAACAAGCCTTTCAAGATGATCCCCGGAGTGGTGGATGGG
optimised       AAGGAAGAGATCCTGGCCATCAACAAGCCCTTCAAGATGATCCCCGGCGTGGTGGACGGC
                ** ** ***** ** ** ** ******** ***************** ******** **

native          GTCTTCCTGCCCAGGCACCCCCAGGAGCTGCTGGCCTCTGCCGACTTTCAGCCTGTCCCT
optimised       GTGTTCCTGCCTAGGCACCCCCAGGAACTGCTGGCCTCTGCCGACTTCCAGCCCGTGCCC
                ** ******** ************** ******************** ***** ** **

native          AGCATTGTTGGTGTCAACAACAATGAATTCGGCTGGCTCATCCCCAAGGTCATGAGGATC
optimised       TCCATCGTGGGCGTGAACAACAACGAGTTCGGCTGGCTGATCCCCAAGGTCATGCGCATC
                  *** ** ** ** ******** ** *********** *************** * ***

native          TATGATACCCAGAAGGAAATGGACAGAGAGGCCTCCCAGGCTGCTCTGCAGAAAATGTTA
optimised       TACGACACCCAAAAGGAAATGGACCGCGAGGCCTCCCAGGCCGCCCTGCAGAAGATGCTG
                ** ** ***** ************ * ************** ** ******** *** *

native          ACGCTGCTGATGTTGCCTCCTACATTTGGTGACCTGCTGAGGGAGGAGTACATTGGGGAC
optimised       ACCCTGCTGATGCTGCCCCCCACCTTCGGCGACCTGCTGCGCGAAGAGTACATCGGCGAC
                ** ********* **** ** ** ** ** ********* * ** ******** ** ***

native          AATGGGGATCCCCAGACCCTCCAAGCGCAGTTCCAGGAGATGATGGCGGACTCCATGTTT
optimised       AACGGCGACCCCCAGACCCTGCAGGCCCAGTTCCAAGAAATGATGGCCGACTCTATGTTC
                ** ** ** *********** ** ** ******** ** ******** ***** *****

native          GTGATCCCTGCACTCCAAGTAGCACATTTTCAGTGTTCCCGGGCCCCTGTGTACTTCTAC
optimised       GTGATCCCCGCTCTGCAGGTGGCCCACTTCCAGTGCTCCCGCGCTCCCGTGTACTTCTAC
                ******** ** ** ** ** ** ** ** ***** ***** ** ** ************

native          GAGTTCCAGCATCAGCCCAGCTGGCTCAAGAACATCAGGCCACCGCACATGAAGGCAGAC
optimised       GAGTTCCAGCACCAGCCCTCTTGGCTGAAGAACATCCGCCCTCCCCACATGAAGGCCGAC
                *********** ******   ***** ********* * ** ** *********** ***

native          CATGGTGATGAGCTTCCTTTTGTTTTCAGAAGTTTCTTTGGGGGCAACTACATTAAATTC
optimised       CACGGCGACGAGCTGCCCTTCGTGTTCCGCTCCTTCTTCGGCGGAAACTACATCAAGTTC
                ** ** ** ***** ** ** ** *** *    ***** ** ** ******** ** ***

native          ACTGAGGAAGAGGAGCAGCTAAGCAGGAAGATGATGAAGTACTGGGCCAACTTTGCGAGA
optimised       ACCGAGGAAGAGGAACAGCTGTCCCGCAAGATGATGAAGTACTGGGCCAACTTCGCCCGC
                ** *********** *****   * * ************************** **  *

native          AATGGGAACCCCAATGGCGAGGGTCTGCCACACTGGCCGCTGTTCGACCAGGAGGAGCAA
optimised       AACGGCAACCCTAACGGCGAGGGACTGCCCCACTGGCCCCTGTTCGACCAGGAAGAACAG
                ** ** ***** ** ******** ***** ******** ************** ** **

native          TACCTGCAGCTGAACCTACAGCCTGCGGTGGGCCGGGCTCTGAAGGCCCACAGGCTCCAG
optimised       TACCTGCAGCTGAACCTGCAGCCCGCCGTGGGAAGGGCCCTGAAGGCTCACCGCCTGCAG
                ***************** ***** ** *****  **** ******** *** * ** ***

native          TTCTGGAAGAAGGCGCTGCCCCAAAAGATCCAGGAGCTCGAGGAGCCTGAAGAGAGACAC
optimised       TTCTGGAAGAAGGCCCTGCCCCAGAAGATCCAGGAACTGGAAGAACCCGAGGAACGCCAC
                ************** ******** *********** ** ** ** ** ** **  * ***

native          ACAGAGCTG
optimised       ACCGAGCTG
                ** ******
```

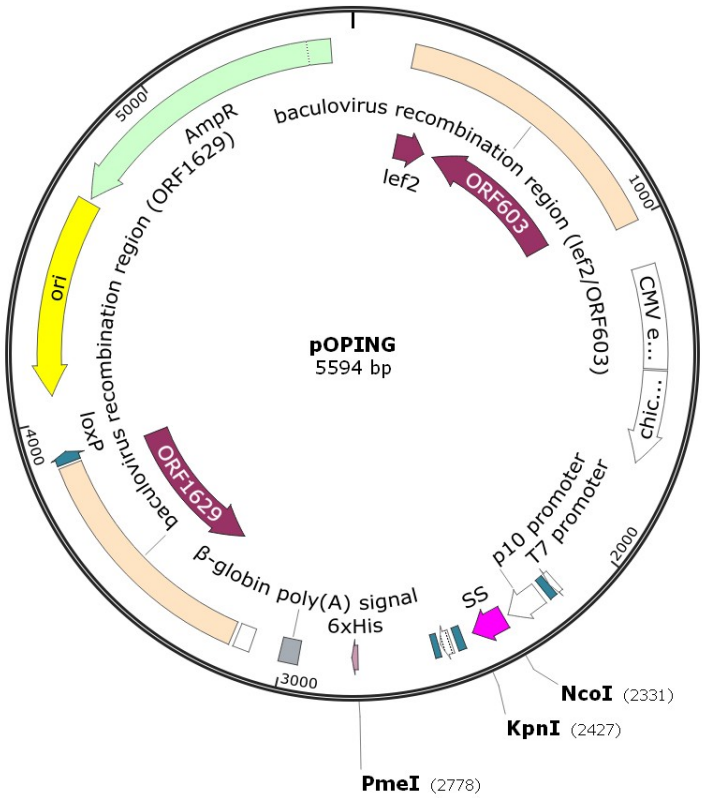
Human CES3

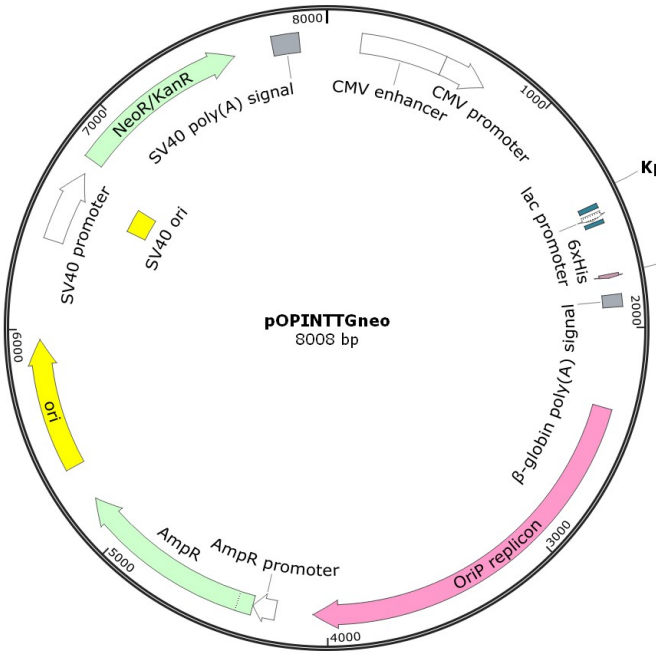
```
native          ATGGAGAGAGCAGTGAGAGTGGAGTCCGGGGTCCTGGTCGGGGTGGTCTGTCTGCTCCTG
optimised       ATGGAAAGGGCCGTGCGCGTGGAATCCGGAGTGCTGGTCGGAGTGGTCTGCCTGCTGCTG
                ***** ** ** *** * ***** ***** ** ******** ******** ***** ***

native          GCATGCCCTGCCACAGCCACTGGGCCCGAAGTTGCTCAGCCTGAAGTAGACACCACCCTG
optimised       GCCTGCCCTGCTACCGCTACCGGACCTGAGGTGGCCCAGCCTGAGGTGGACACCACCCTG
                ** ******** ** ** ** ** ** ** ** ******** ** ************

native          GGTCGTGTGCGAGGCCGGCAGGTGGGCGTGAAGGGCACAGACCGCCTTGTGAATGTCTTT
optimised       GGAAGAGTGCGTGGACGCCAAGTGGGAGTGAAGGGAACCGACCGCCTGGTGAACGTGTTC
                **  * ***** ** ** ** ***** ******** ** ******** ***** ** **

native          CTGGGCATTCCATTTGCCCAGCCGCCACTGGGCCCTGACCGGTTCTCAGCCCCACACCCA
optimised       CTGGGCATCCCCTTCGCCCAGCCCCCTCTGGGCCCTGACAGGTTCTCTGCCCCTCACCCT
                ******** ** ** ******** ** ************ ******* ***** *****

native          GCACAGCCCTGGGAGGGTGTGCGGGATGCCAGCACTGCGCCCCCAATGTGCCTACAAGAC
optimised       GCCCAGCCTTGGGAAGGCGTGCGCGATGCTTCCACCGCCCCTCCAATGTGCCTGCAGGAC
                ** ***** ***** ** ***** *****   *** ** ** ********** ** ***

native          GTGGAGAGCATGAACAGCAGCAGATTTGTCCTCAACGGAAAACAGCAGATCTTCTCCGTT
optimised       GTGGAATCCATGAACTCCTCCCGCTTCGTGCTGAACGGCAAGCAGCAGATCTTCTCCGTG
                *****   *******  *  * * ** ** ** ***** ** ****************

native          TCAGAGGACTGCCTGGTCCTCAACGTCTATAGCCCAGCTGAGGTCCCCGCAGGGTCCGGT
optimised       TCCGAGGACTGCCTGGTGCTGAACGTGTACTCCCCTGCCGAAGTGCCAGCCGGATCTGGA
                ** ************** ** ***** **   *** ** ** ** ** ** ** ** **

native          AGGCCGGTCATGGTATGGGTCCATGGAGGCGCTCTGATAACTGGCGCTGCCACCTCCTAC
optimised       AGGCCTGTGATGGTCTGGGTGCACGGCGGAGCCCTGATCACCGGCGCTGCCACCTCTTAC
                ***** ** ***** ***** ** ** ** ** ***** ** ************** ***

native          GATGGATCAGCTCTGGCTGCCTATGGGGATGTGGTCGTGGTTACAGTCCAGTACCGCCTT
optimised       GACGGATCTGCCCTGGCCGCCTACGGCGACGTGGTGGTGGTGACCGTGCAGTACAGACTG
                ** ***** ** ***** ***** ** ** ***** ***** ** ** ****** * **

native          GGGGTCCTTGGCTTCTTCAGCACTGGAGATGAGCATGCACCTGGCAACCAGGGCTTCCTA
optimised       GGCGTGCTGGGCTTCTTCAGCACCGGCGACGAGCACGCTCCTGGCAACCAGGGATTCCTG
                ** ** ** ************** ** ** ***** ** ************** *****

native          GATGTGGTAGCTGCTTTGCGCTGGGTGCAAGAAAACATCGCCCCCTTCGGGGGTGACCTC
optimised       GACGTGGTGGCTGCCCTGCGCTGGGTGCAGGAAAACATCGCCCCCTTCGGCGGCGACCTG
                ** ***** *****  ************* ******************** ** *****

native          AACTGTGTCACTGTCTTTGGTGGATCTGCCGGTGGGAGCATCATCTCTGGCCTGGTCCTG
optimised       AACTGCGTGACCGTGTTCGGCGGCTCCGCTGGCGGCTCCATCATCTCCGGACTGGTGCTG
                ***** ** ** ** ** ** ** ** ** ** **   ********* ** ***** ***

native          TCCCCAGTGGCTGCAGGGCTGTTCCACAGAGCCATCACACAGAGTGGGGTCATCACCACC
optimised       TCCCCCGTGGCCGCTGGACTGTTCCACAGGGCCATCACCCAGTCTGGCGTGATCACCACC
                ***** ***** ** ** ********** ******** ***  *** ** *********

native          CCAGGGATCATCGACTCTCACCCTTGGCCCCTAGCTCAGAAAATCGCAAACACCTTGGCC
optimised       CCCGGCATCATCGACTCCCACCCTTGGCCCCTGGCCCAGAAGATCGCCAACACCCTGGCC
                ** ** ********** ************** ** ***** ***** ****** *****

native          TGCAGCTCCAGCTCCCCGGCTGAGATGGTGCAGTGCCTTCAGCAGAAAGAAGGAGAAGAG
optimised       TGCTCCTCCAGCTCCCCAGCCGAGATGGTGCAGTGCCTGCAGCAGAAGGAAGGCGAGGAA
                ***  ************ ** ***************** ******** ***** ** **

native          CTGGTCCTTAGCAAGAAGCTGAAAAATACTATCTATCCTCTCACCGTTGATGGCACTGTC
optimised       CTGGTCCTGAGCAAGAAGCTGAAGAACACCATCTACCCCCTGACCGTGGACGGCACCGTG
                ******** ************** ** ** ***** ** ** ***** ** ***** **

native          TTCCCCAAAAGCCCCAAGGAACTCCTGAAGGAGAAGCCCTTCCACTCTGTGCCCTTCCTC
optimised       TTCCCCAAGTCCCCCAAGGAACTGCTGAAGGAAAAGCCCTTCCACTCCGTGCCATTCCTC
                ********   ************ ******** ************** ***** ******
```

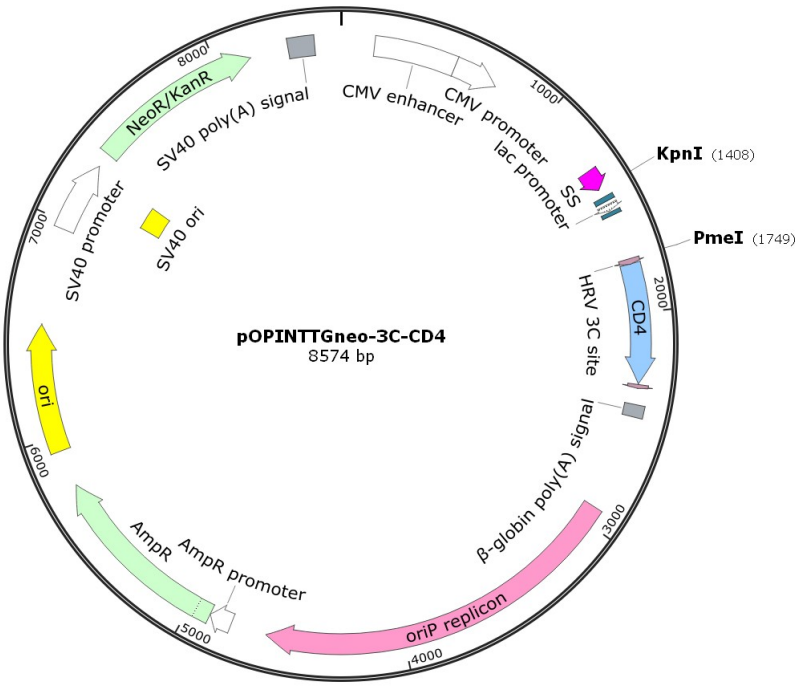
```
native          ATGGGTGTCAACAACCATGAGTTCAGCTGGCTCATCCCCAGGGGCTGGGGTCTCCTGGAT
optimised       ATGGGCGTGAACAACCACGAGTTCTCTTGGCTGATCCCCAGGGGCTGGGGCCTGCTGGAC
                ***** ** ******** ******   ***** ***************** ** *****

native          ACAATGGAGCAGATGAGCCGGGAGGACATGCTGGCCATCTCAACACCCGTCTTGACCAGT
optimised       ACCATGGAACAGATGTCCCGCGAGGATATGCTGGCCATCTCCACCCCCGTGCTGACCTCC
                ** ***** ******  *** ***** ************** ** *****  *****

native          CTGGATGTGCCCCCTGAGATGATGCCCACCGTCATAGATGAATACCTAGGAAGCAACTCG
optimised       CTGGACGTGCCCCCCGAGATGATGCCCACCGTGATCGACGAGTACCTGGGCTCCAACTCC
                ***** ******** ***************** ** ** ** ***** **   ******

native          GACGCACAAGCCAAATGCCAGGCGTTCCAGGAATTCATGGGTGACGTATTCATCAATGTT
optimised       GACGCCCAGGCCAAGTGCCAGGCCTTCCAAGAATTCATGGGCGACGTCTTCATCAACGTG
                ***** ** ***** ******** ***** *********** ***** ******** **

native          CCCACCGTCAGTTTTTCAAGATACCTTCGAGATTCTGGAAGCCCTGTCTTTTTCTATGAG
optimised       CCAACCGTGTCCTTCTCCCGCTACCTGCGCGACTCCGGCTCCCCCGTGTTCTTCTACGAG
                ** *****    ** **  * ***** ** ** ** **   *** ** ** ***** ***

native          TTCCAGCATCGACCCAGTTCTTTTGCGAAGATCAAACCTGCCTGGGTGAAGGCTGATCAT
optimised       TTCCAGCACAGGCCCTCCAGCTTCGCCAAGATCAAGCCCGCCTGGGTGAAGGCCGACCAC
                ********  * ***      ** ** ******** ** ************** ** **

native          GGGGCCGAGGGTGCTTTTGTGTTCGGAGGTCCCTTCCTCATGGACGAGAGCTCCCGCCTG
optimised       GGCGCTGAGGGCGCCTTCGTGTTCGGAGGACCCTTCCTGATGGACGAGTCCTCCCGCCTG
                ** ** ***** ** ** *********** ******** *********  **********

native          GCCTTTCCAGAGGCCACAGAGGAGGAGAAGCAGCTAAGCCTCACCATGATGGCCCAGTGG
optimised       GCCTTCCCCGAGGCCACCGAGGAAGAGAAGCAGCTGTCCCTGACCATGATGGCCCAGTGG
                ***** ** ******** ***** ***********   *** ******************

native          ACCCACTTTGCCCGGACAGGGGACCCCAATAGCAAGGCTCTGCCTCCTTGGCCCCAATTC
optimised       ACCCACTTCGCCAGGACCGGCGACCCCAACTCCAAGGCTCTGCCCCCCTGGCCCCAGTTC
                ******** *** **** ** ********   ************ ** ******** ***

native          AACCAGGCGGAACAATATCTGGAGATCAACCCAGTGCCACGGGCCGGACAGAAGTTCAGG
optimised       AACCAGGCCGAGCAGTACCTGGAAATCAACCCCGTGCCCAGGGCCGGACAGAAGTTCCGC
                ******** ** ** ** ***** ******** *****  ***************** *

native          GAGGCCTGGATGCAGTTCTGGTCAGAGACGCTCCCCAGCAAGATACAACAGTGGCACCAG
optimised       GAGGCCTGGATGCAGTTCTGGTCCGAAACCCTGCCCTCCAAGATCCAGCAGTGGCACCAG
                ************************ ** ** ** ***  ****** ** ************

native          AAGCAGAAGAACAGGAAGGCCCAGGAGGACCTC
optimised       AAGCAGAAGAACCGCAAGGCCCAGGAAGATCTG
                ************ * *********** ** **
```

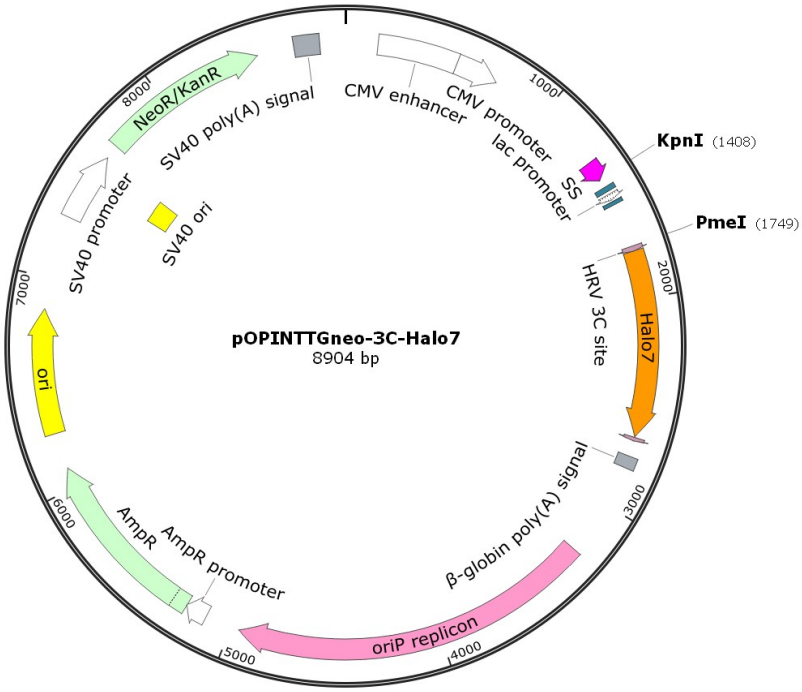
Appendix 1.4


pOPINE
5534 bp
AmpR
baculovirus recombination region (lef2/ORF603)
lef2
ORF603
CMV enhancer
ori
chic...
loxP
ORF1629
baculovirus recom...
β-globin poly(A) signal
6xHis
p10 promoter
T7 promoter
5000
1000
4000
2000
3000
NcoI (2331)
(2718) PmeI
HindIII (2700)


pOPING
5594 bp
AmpR
baculovirus recombination region (ORF1629)
baculovirus recombination region (lef2/ORF603)
lef2
ORF603
CMV e...
chic...
ori
loxP
ORF1629
baculovirus recom...
β-globin poly(A) signal
6xHis
SS
p10 promoter
T7 promoter
5000
1000
4000
2000
3000
NcoI (2331)
KpnI (2427)
PmeI (2778)


8000
1000
2000
3000
4000
5000
6000
7000
NeoR/KanR
SV40 poly(A) signal
CMV enhancer
CMV promoter
KpnI (1408)
lac promoter
6xHis
PmeI (1759)
SV40 promoter
SV40 ori
pOPINTTGneo
8008 bp
ori
β-globin poly(A) signal
AmpR
AmpR promoter
OriP replicon


8000
1000
2000
3000
4000
5000
6000
7000
NeoR/KanR
SV40 poly(A) signal
CMV enhancer
CMV promoter
lac promoter
SS
KpnI (1408)
PmeI (1749)
SV40 promoter
SV40 ori
HRV 3C site
CD4
pOPINTTGneo-3C-CD4
8574 bp
ori
β-globin poly(A) signal
AmpR
AmpR promoter
OriP replicon



REFERENCES

1. Derewenda, Z. S., It's all in the crystals…. *Acta Crystallogr D Biol Crystallogr* **2011,** *67* (Pt 4), 243-8.
2. Ellgaard, L.; Molinari, M.; Helenius, A., Setting the standards: quality control in the secretory pathway. *Science* **1999,** *286* (5446), 1882-8.
3. Na, K.; Jeong, S. K.; Lee, M. J.; Cho, S. Y.; Kim, S. A.; Song, S. Y.; Kim, H.; Kim, K. S.; Lee, H. W.; Paik, Y. K., Human liver carboxylesterase 1 outperforms alpha-fetoprotein as biomarker to discriminate hepatocellular carcinoma from other liver diseases in Korean patients. *Int J Cancer* **2013,** *133* (2), 408-15.
4. Markey, G. M., Carboxylesterase 1 (Ces1): from monocyte marker to major player. *J Clin Pathol* **2011,** *64* (2), 107-9.
5. Holmes, R. S.; Wright, M. W.; Laulederkind, S. J.; Cox, L. A.; Hosokawa, M.; Imai, T.; Ishibashi, S.; Lehner, R.; Miyazaki, M.; Perkins, E. J.; Potter, P. M.; Redinbo, M. R.; Robert, J.; Satoh, T.; Yamashita, T.; Yan, B.; Yokoi, T.; Zechner, R.; Maltais, L. J., Recommended nomenclature for five mammalian carboxylesterase gene families: human, mouse, and rat genes and proteins. *Mamm Genome* **2010,** *21* (9-10), 427-41.
6. Satoh, T.; Taylor, P.; Bosron, W. F.; Sanghani, S. P.; Hosokawa, M.; La Du, B. N., Current progress on esterases: from molecular structure to function. *Drug Metab Dispos* **2002,** *30* (5), 488-93.
7. Mori, M.; Hosokawa, M.; Ogasawara, Y.; Tsukada, E.; Chiba, K., cDNA cloning, characterization and stable expression of novel human brain carboxylesterase. *FEBS Lett* **1999,** *458* (1), 17-22.
8. Landry, D. W.; Zhao, K.; Yang, G. X.; Glickman, M.; Georgiadis, T. M., Antibody-catalyzed degradation of cocaine. *Science* **1993,** *259* (5103), 1899-901.
9. Holmes, R. S.; Chan, J.; Cox, L. A.; Murphy, W. J.; VandeBerg, J. L., Opossum carboxylesterases: sequences, phylogeny and evidence for CES gene duplication events predating the marsupial-eutherian common ancestor. *BMC Evol Biol* **2008,** *8*, 54.
10. Williams, E. T.; Wang, H.; Wrighton, S. A.; Qian, Y. W.; Perkins, E. J., Genomic analysis of the carboxylesterases: identification and classification of novel forms. *Mol Phylogenet Evol* **2010,** *57* (1), 23-34.
11. Jones, R. D.; Taylor, A. M.; Tong, E. Y.; Repa, J. J., Carboxylesterases are uniquely expressed among tissues and regulated by nuclear hormone receptors in the mouse. *Drug Metab Dispos* **2013,** *41* (1), 40-9.

12. Bencharit, S.; Morton, C. L.; Howard-Williams, E. L.; Danks, M. K.; Potter, P. M.; Redinbo, M. R., Structural insights into CPT-11 activation by mammalian carboxylesterases. *Nat Struct Biol* **2002,** *9* (5), 337-42.
13. Dereeper, A.; Audic, S.; Claverie, J. M.; Blanc, G., BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC Evol Biol* **2010,** *10*, 8; Dereeper, A.; Guignon, V.; Blanc, G.; Audic, S.; Buffet, S.; Chevenet, F.; Dufayard, J. F.; Guindon, S.; Lefort, V.; Lescot, M.; Claverie, J. M.; Gascuel, O., Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* **2008,** *36* (Web Server issue), W465-9; Edgar, R. C., MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **2004,** *32* (5), 1792-7; Castresana, J., Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **2000,** *17* (4), 540-52; Guindon, S.; Gascuel, O., A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **2003,** *52* (5), 696-704; Anisimova, M.; Gascuel, O., Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* **2006,** *55* (4), 539-52; Chevenet, F.; Brun, C.; Bañuls, A. L.; Jacq, B.; Christen, R., TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* **2006,** *7*, 439.
14. Consortium, U., Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res* **2014,** *42* (Database issue), D191-8.
15. Simossis, V. A.; Kleinjung, J.; Heringa, J., Homology-extended sequence alignment. *Nucleic Acids Res* **2005,** *33* (3), 816-24; Pirovano, W.; Feenstra, K. A.; Heringa, J., PRALINETM: a strategy for improved multiple alignment of transmembrane proteins. *Bioinformatics* **2008,** *24* (4), 492-7; Heringa, J., Local weighting schemes for protein multiple sequence alignment. *Comput Chem* **2002,** *26* (5), 459-77.
16. Munger, J. S.; Shi, G. P.; Mark, E. A.; Chin, D. T.; Gerard, C.; Chapman, H. A., A serine esterase released by human alveolar macrophages is closely related to liver microsomal carboxylesterases. *J Biol Chem* **1991,** *266* (28), 18832-8.
17. Yang, D.; Pearce, R. E.; Wang, X.; Gaedigk, R.; Wan, Y. J.; Yan, B., Human carboxylesterases HCE1 and HCE2: ontogenic expression, inter-individual variability and differential hydrolysis of oseltamivir, aspirin, deltamethrin and permethrin. *Biochem Pharmacol* **2009,** *77* (2), 238-47.
18. Imai, T., Human carboxylesterase isozymes: catalytic properties and rational drug design. *Drug Metab Pharmacokinet* **2006,** *21* (3), 173-85; Xu, G.; Zhang, W.; Ma, M. K.; McLeod, H. L., Human carboxylesterase 2 is commonly expressed in tumor tissue and is correlated with activation of irinotecan. *Clin Cancer Res* **2002,** *8* (8), 2605-11; Sanghani, S. P.; Quinney, S. K.; Fredenburg, T. B.; Sun, Z.; Davis, W. I.; Murry, D. J.; Cummings, O. W.; Seitz, D. E.; Bosron, W. F., Carboxylesterases expressed in human colon tumor tissue and their role in CPT-11 hydrolysis. *Clin Cancer Res* **2003,** *9* (13), 4983-91.
19. Zhao, B.; Bie, J.; Wang, J.; Marqueen, S. A.; Ghosh, S., Identification of a novel intracellular cholesteryl ester hydrolase (carboxylesterase 3) in human macrophages: compensatory increase in its expression after carboxylesterase 1 silencing. *Am J Physiol Cell Physiol* **2012,** *303* (4), C427-35.
20. Potter, P. M.; Wolverton, J. S.; Morton, C. L.; Wierdl, M.; Danks, M. K., Cellular localization domains of a rabbit and a human carboxylesterase: influence on irinotecan (CPT-11) metabolism by the rabbit enzyme. *Cancer Res* **1998,** *58* (16), 3627-32.
21. Sanghani, S. P.; Quinney, S. K.; Fredenburg, T. B.; Davis, W. I.; Murry, D. J.; Bosron, W. F., Hydrolysis of irinotecan and its oxidative metabolites, 7-ethyl-10-[4-N-(5-aminopentanoic acid)-1-piperidino] carbonyloxycamptothecin and 7-ethyl-10-[4-(1-piperidino)-1-amino]-carbonyloxycamptothecin, by human carboxylesterases CES1A1, CES2, and a newly expressed carboxylesterase isoenzyme, CES3. *Drug Metab Dispos* **2004,** *32* (5), 505-11.

22. Robbi, M.; Beaufay, H., The COOH terminus of several liver carboxylesterases targets these enzymes to the lumen of the endoplasmic reticulum. *J Biol Chem* **1991,** *266* (30), 20498-503.
23. Holmes, R. S.; Cox, L. A.; Vandeberg, J. L., Mammalian carboxylesterase 5: comparative biochemistry and genomics. *Comp Biochem Physiol Part D Genomics Proteomics* **2008,** *3* (3), 195-204.
24. Merali, Z.; Ross, S.; Paré, G., The pharmacogenetics of carboxylesterases: CES1 and CES2 genetic variants and their clinical effect. *Drug Metabol Drug Interact* **2014**.
25. Islam, M. R.; Waheed, A.; Shah, G. N.; Tomatsu, S.; Sly, W. S., Human egasyn binds beta-glucuronidase but neither the esterase active site of egasyn nor the C terminus of beta-glucuronidase is involved in their interaction. *Arch Biochem Biophys* **1999,** *372* (1), 53-61.
26. Macintyre, S.; Samols, D.; Dailey, P., Two carboxylesterases bind C-reactive protein within the endoplasmic reticulum and regulate its secretion during the acute phase response. *J Biol Chem* **1994,** *269* (39), 24496-503.
27. Labarrere, C. A.; Lee, J. B.; Nelson, D. R.; Al-Hassani, M.; Miller, S. J.; Pitts, D. E., C-reactive protein, arterial endothelial activation, and development of transplant coronary artery disease: a prospective study. *Lancet* **2002,** *360* (9344), 1462-7.
28. Yue, C. C.; Muller-Greven, J.; Dailey, P.; Lozanski, G.; Anderson, V.; Macintyre, S., Identification of a C-reactive protein binding site in two hepatic carboxylesterases capable of retaining C-reactive protein within the endoplasmic reticulum. *J Biol Chem* **1996,** *271* (36), 22245-50.
29. Cecchin, E.; Corona, G.; Masier, S.; Biason, P.; Cattarossi, G.; Frustaci, S.; Buonadonna, A.; Colussi, A.; Toffoli, G., Carboxylesterase isoform 2 mRNA expression in peripheral blood mononuclear cells is a predictive marker of the irinotecan to SN38 activation step in colorectal cancer patients. *Clin Cancer Res* **2005,** *11* (19 Pt 1), 6901-7.
30. Tang, X.; Wu, H.; Wu, Z.; Wang, G.; Wang, Z.; Zhu, D., Carboxylesterase 2 is downregulated in colorectal cancer following progression of the disease. *Cancer Invest* **2008,** *26* (2), 178-81.
31. Cai, L.; Tang, X.; Guo, L.; An, Y.; Wang, Y.; Zheng, J., Decreased serum levels of carboxylesterase-2 in patients with ovarian cancer. *Tumori* **2009,** *95* (4), 473-8.
32. Pratt, S. E.; Durland-Busbice, S.; Shepard, R. L.; Heinz-Taheny, K.; Iversen, P. W.; Dantzig, A. H., Human carboxylesterase-2 hydrolyzes the prodrug of gemcitabine (LY2334737) and confers prodrug sensitivity to cancer cells. *Clin Cancer Res* **2013,** *19* (5), 1159-68.
33. Jernås, M.; Olsson, B.; Arner, P.; Jacobson, P.; Sjöström, L.; Walley, A.; Froguel, P.; McTernan, P. G.; Hoffstedt, J.; Carlsson, L. M., Regulation of carboxylesterase 1 (CES1) in human adipose tissue. *Biochem Biophys Res Commun* **2009,** *383* (1), 63-7; Steinberg, G. R.; Kemp, B. E.; Watt, M. J., Adipocyte triglyceride lipase expression in human obesity. *Am J Physiol Endocrinol Metab* **2007,** *293* (4), E958-64.
34. Dominguez, E.; Galmozzi, A.; Chang, J. W.; Hsu, K. L.; Pawlak, J.; Li, W.; Godio, C.; Thomas, J.; Partida, D.; Niessen, S.; O'Brien, P. E.; Russell, A. P.; Watt, M. J.; Nomura, D. K.; Cravatt, B. F.; Saez, E., Integrated phenotypic and activity-based profiling links Ces3 to obesity and diabetes. *Nat Chem Biol* **2014,** *10* (2), 113-21.
35. Wei, E.; Gao, W.; Lehner, R., Attenuation of adipocyte triacylglycerol hydrolase activity decreases basal fatty acid efflux. *J Biol Chem* **2007,** *282* (11), 8027-35.
36. Blais, D. R.; Lyn, R. K.; Joyce, M. A.; Rouleau, Y.; Steenbergen, R.; Barsby, N.; Zhu, L. F.; Pegoraro, A. F.; Stolow, A.; Tyrrell, D. L.; Pezacki, J. P., Activity-based protein profiling identifies a host enzyme, carboxylesterase 1, which is differentially active during hepatitis C virus replication. *J Biol Chem* **2010,** *285* (33), 25602-12.
37. Wadkins, R. M.; Morton, C. L.; Weeks, J. K.; Oliver, L.; Wierdl, M.; Danks, M. K.; Potter, P. M., Structural constraints affect the metabolism of 7-ethyl-10-[4-(1-

piperidino)-1-piperidino]carbonyloxycamptothecin (CPT-11) by carboxylesterases. *Mol Pharmacol* **2001,** *60* (2), 355-62.
38. Sun, Z.; Murry, D. J.; Sanghani, S. P.; Davis, W. I.; Kedishvili, N. Y.; Zou, Q.; Hurley, T. D.; Bosron, W. F., Methylphenidate is stereoselectively hydrolyzed by human carboxylesterase CES1A1. *J Pharmacol Exp Ther* **2004,** *310* (2), 469-76.
39. Fukami, T.; Yokoi, T., The emerging role of human esterases. *Drug Metab Pharmacokinet* **2012,** *27* (5), 466-77.
40. Imai, T.; Taketani, M.; Shii, M.; Hosokawa, M.; Chiba, K., Substrate specificity of carboxylesterase isozymes and their contribution to hydrolase activity in human liver and small intestine. *Drug Metab Dispos* **2006,** *34* (10), 1734-41.
41. Shi, D.; Yang, J.; Yang, D.; LeCluyse, E. L.; Black, C.; You, L.; Akhlaghi, F.; Yan, B., Anti-influenza prodrug oseltamivir is activated by carboxylesterase human carboxylesterase 1, and the activation is inhibited by antiplatelet agent clopidogrel. *J Pharmacol Exp Ther* **2006,** *319* (3), 1477-84.
42. Wong, C. C.; Cheng, K. W.; Xie, G.; Zhou, D.; Zhu, C. H.; Constantinides, P. P.; Rigas, B., Carboxylesterases 1 and 2 hydrolyze phospho-nonsteroidal anti-inflammatory drugs: relevance to their pharmacological activity. *J Pharmacol Exp Ther* **2012,** *340* (2), 422-32.
43. Matsuzaki, T.; Yokokura, T.; Mutai, M.; Tsuruo, T., Inhibition of spontaneous and experimental metastasis by a new derivative of camptothecin, CPT-11, in mice. *Cancer Chemother Pharmacol* **1988,** *21* (4), 308-12.
44. Patrick, K. S.; Caldwell, R. W.; Ferris, R. M.; Breese, G. R., Pharmacology of the enantiomers of threo-methylphenidate. *J Pharmacol Exp Ther* **1987,** *241* (1), 152-8.
45. Satoh, T.; Hosokawa, M., The mammalian carboxylesterases: from molecules to functions. *Annu Rev Pharmacol Toxicol* **1998,** *38*, 257-88.
46. Zhang, J.; Burnell, J. C.; Dumaual, N.; Bosron, W. F., Binding and hydrolysis of meperidine by human liver carboxylesterase hCE-1. *J Pharmacol Exp Ther* **1999,** *290* (1), 314-8.
47. Higuchi, R.; Fukami, T.; Nakajima, M.; Yokoi, T., Prilocaine- and lidocaine-induced methemoglobinemia is caused by human carboxylesterase-, CYP2E1-, and CYP3A4-mediated metabolic activation. *Drug Metab Dispos* **2013,** *41* (6), 1220-30.
48. Danks, M. K.; Morton, C. L.; Pawlik, C. A.; Potter, P. M., Overexpression of a rabbit liver carboxylesterase sensitizes human tumor cells to CPT-11. *Cancer Res* **1998,** *58* (1), 20-2.
49. Tanizawa, A.; Fujimori, A.; Fujimori, Y.; Pommier, Y., Comparison of topoisomerase I inhibition, DNA damage, and cytotoxicity of camptothecin derivatives presently in clinical trials. *J Natl Cancer Inst* **1994,** *86* (11), 836-42.
50. Humerickhouse, R.; Lohrbach, K.; Li, L.; Bosron, W. F.; Dolan, M. E., Characterization of CPT-11 hydrolysis by human liver carboxylesterase isoforms hCE-1 and hCE-2. *Cancer Res* **2000,** *60* (5), 1189-92.
51. Barthel, B. L.; Torres, R. C.; Hyatt, J. L.; Edwards, C. C.; Hatfield, M. J.; Potter, P. M.; Koch, T. H., Identification of human intestinal carboxylesterase as the primary enzyme for activation of a doxazolidine carbamate prodrug. *J Med Chem* **2008,** *51* (2), 298-304.
52. Ait-Tihyaty, M.; Rachid, Z.; Larroque-Lombard, A. L.; Jean-Claude, B. J., ZRX1, the first EGFR inhibitor-capecitabine based combi-molecule, requires carboxylesterase-mediated hydrolysis for optimal activity. *Invest New Drugs* **2013,** *31* (6), 1409-23; Shindoh, H.; Nakano, K.; Yoshida, T.; Ishigai, M., Comparison of in vitro metabolic conversion of capecitabine to 5-FU in rats, mice, monkeys and humans--toxicological implications. *J Toxicol Sci* **2011,** *36* (4), 411-22; Tabata, T.; Katoh, M.; Tokudome, S.; Hosakawa, M.; Chiba, K.; Nakajima, M.; Yokoi, T., Bioactivation of capecitabine in

human liver: involvement of the cytosolic enzyme on 5'-deoxy-5-fluorocytidine formation. *Drug Metab Dispos* **2004,** *32* (7), 762-7.
53. Pindel, E. V.; Kedishvili, N. Y.; Abraham, T. L.; Brzezinski, M. R.; Zhang, J.; Dean, R. A.; Bosron, W. F., Purification and cloning of a broad substrate specificity human liver carboxylesterase that catalyzes the hydrolysis of cocaine and heroin. *J Biol Chem* **1997,** *272* (23), 14769-75.
54. Brzezinski, M. R.; Spink, B. J.; Dean, R. A.; Berkman, C. E.; Cashman, J. R.; Bosron, W. F., Human liver carboxylesterase hCE-1: binding specificity for cocaine, heroin, and their metabolites and analogs. *Drug Metab Dispos* **1997,** *25* (9), 1089-96.
55. Kamendulis, L. M.; Brzezinski, M. R.; Pindel, E. V.; Bosron, W. F.; Dean, R. A., Metabolism of cocaine and heroin is catalyzed by the same human liver carboxylesterases. *J Pharmacol Exp Ther* **1996,** *279* (2), 713-7.
56. Dean, R. A.; Christian, C. D.; Sample, R. H.; Bosron, W. F., Human liver cocaine esterases: ethanol-mediated formation of ethylcocaine. *FASEB J* **1991,** *5* (12), 2735-9.
57. Hatfield, M. J.; Tsurkan, L.; Hyatt, J. L.; Yu, X.; Edwards, C. C.; Hicks, L. D.; Wadkins, R. M.; Potter, P. M., Biochemical and molecular analysis of carboxylesterase-mediated hydrolysis of cocaine and heroin. *Br J Pharmacol* **2010,** *160* (8), 1916-28.
58. Brzezinski, M. R.; Abraham, T. L.; Stone, C. L.; Dean, R. A.; Bosron, W. F., Purification and characterization of a human liver cocaine carboxylesterase that catalyzes the production of benzoylecgonine and the formation of cocaethylene from alcohol and cocaine. *Biochem Pharmacol* **1994,** *48* (9), 1747-55.
59. Broomfield, C. A.; Kirby, S. D., Progress on the road to new nerve agent treatments. *J Appl Toxicol* **2001,** *21 Suppl 1*, S43-6.
60. Jokanović, M., Current understanding of the mechanisms involved in metabolic detoxification of warfare nerve agents. *Toxicol Lett* **2009,** *188* (1), 1-10.
61. Hemmert, A. C.; Otto, T. C.; Wierdl, M.; Edwards, C. C.; Fleming, C. D.; MacDonald, M.; Cashman, J. R.; Potter, P. M.; Cerasoli, D. M.; Redinbo, M. R., Human carboxylesterase 1 stereoselectively binds the nerve agent cyclosarin and spontaneously hydrolyzes the nerve agent sarin. *Mol Pharmacol* **2010,** *77* (4), 508-16.
62. Wahlländer, A.; Szinicz, L., Detoxification of soman in the perfused rat liver: quantitative uptake and stereoisomer metabolism. *Arch Toxicol* **1990,** *64* (7), 586-9.
63. Maxwell, D. M.; Brecht, K. M., Carboxylesterase: specificity and spontaneous reactivation of an endogenous scavenger for organophosphorus compounds. *J Appl Toxicol* **2001,** *21 Suppl 1*, S103-7; Sweeney, R. E.; Maxwell, D. M., A theoretical model of the competition between hydrolase and carboxylesterase in protection against organophosphorus poisoning. *Math Biosci* **1999,** *160* (2), 175-90.
64. Gershater, M.; Sharples, K.; Edwards, R., Carboxylesterase activities toward pesticide esters in crops and weeds. *Phytochemistry* **2006,** *67* (23), 2561-7.
65. Thomas, B. A.; Church, W. B.; Lane, T. R.; Hammock, B. D., Homology model of juvenile hormone esterase from the crop pest, Heliothis virescens. *Proteins* **1999,** *34* (2), 184-96.
66. Cashman, J. R.; Perotti, B. Y.; Berkman, C. E.; Lin, J., Pharmacokinetics and molecular detoxication. *Environ Health Perspect* **1996,** *104 Suppl 1*, 23-40.
67. Kuykendall, J. R.; Taylor, M. L.; Bogdanffy, M. S., Cytotoxicity and DNA-protein crosslink formation in rat nasal tissues exposed to vinyl acetate are carboxylesterase-mediated. *Toxicol Appl Pharmacol* **1993,** *123* (2), 283-92.
68. Mikhailov, A. T.; Torrado, M., Carboxylesterases moonlight in the male reproductive tract: a functional shift pivotal for male fertility. *Front Biosci* **2000,** *5*, E53-62.
69. Bencharit, S.; Edwards, C. C.; Morton, C. L.; Howard-Williams, E. L.; Kuhn, P.; Potter, P. M.; Redinbo, M. R., Multisite promiscuity in the processing of endogenous substrates by human carboxylesterase 1. *J Mol Biol* **2006,** *363* (1), 201-14.

70. Redinbo, M. R.; Bencharit, S.; Potter, P. M., Human carboxylesterase 1: from drug metabolism to drug discovery. *Biochem Soc Trans* **2003,** *31* (Pt 3), 620-4.
71. Ghosh, S., Cholesteryl ester hydrolase in human monocyte/macrophage: cloning, sequencing, and expression of full-length cDNA. *Physiol Genomics* **2000,** *2* (1), 1-8.
72. Harrison, E. H., Lipases and carboxylesterases: possible roles in the hepatic metabolism of retinol. *Annu Rev Nutr* **1998,** *18*, 259-76.
73. Hyatt, J. L.; Moak, T.; Hatfield, M. J.; Tsurkan, L.; Edwards, C. C.; Wierdl, M.; Danks, M. K.; Wadkins, R. M.; Potter, P. M., Selective inhibition of carboxylesterases by isatins, indole-2,3-diones. *J Med Chem* **2007,** *50* (8), 1876-85.
74. Stampfli, H. F.; Quon, C. Y., Polymorphic metabolism of flestolol and other ester containing compounds by a carboxylesterase in New Zealand white rabbit blood and cornea. *Res Commun Mol Pathol Pharmacol* **1995,** *88* (1), 87-97.
75. Wadkins, R. M.; Hyatt, J. L.; Wei, X.; Yoon, K. J.; Wierdl, M.; Edwards, C. C.; Morton, C. L.; Obenauer, J. C.; Damodaran, K.; Beroza, P.; Danks, M. K.; Potter, P. M., Identification and characterization of novel benzil (diphenylethane-1,2-dione) analogues as inhibitors of mammalian carboxylesterases. *J Med Chem* **2005,** *48* (8), 2906-15.
76. Hicks, L. D.; Hyatt, J. L.; Stoddard, S.; Tsurkan, L.; Edwards, C. C.; Wadkins, R. M.; Potter, P. M., Improved, selective, human intestinal carboxylesterase inhibitors designed to modulate 7-ethyl-10-[4-(1-piperidino)-1-piperidino]carbonyloxycamptothecin (Irinotecan; CPT-11) toxicity. *J Med Chem* **2009,** *52* (12), 3742-52.
77. Wadkins, R. M.; Hyatt, J. L.; Yoon, K. J.; Morton, C. L.; Lee, R. E.; Damodaran, K.; Beroza, P.; Danks, M. K.; Potter, P. M., Discovery of novel selective inhibitors of human intestinal carboxylesterase for the amelioration of irinotecan-induced diarrhea: synthesis, quantitative structure-activity relationship analysis, and biological activity. *Mol Pharmacol* **2004,** *65* (6), 1336-43.
78. Fleming, C. D.; Bencharit, S.; Edwards, C. C.; Hyatt, J. L.; Tsurkan, L.; Bai, F.; Fraga, C.; Morton, C. L.; Howard-Williams, E. L.; Potter, P. M.; Redinbo, M. R., Structural insights into drug processing by human carboxylesterase 1: tamoxifen, mevastatin, and inhibition by benzil. *J Mol Biol* **2005,** *352* (1), 165-77.
79. Li, P.; Callery, P. S.; Gan, L. S.; Balani, S. K., Esterase inhibition by grapefruit juice flavonoids leading to a new drug interaction. *Drug Metab Dispos* **2007,** *35* (7), 1203-8.
80. Van Gelder, J.; Annaert, P.; Naesens, L.; De Clercq, E.; Van den Mooter, G.; Kinget, R.; Augustijns, P., Inhibition of intestinal metabolism of the antiviral ester prodrug bis(POC)-PMPA by nature-identical fruit extracts as a strategy to enhance its oral absorption: an in vitro study. *Pharm Res* **1999,** *16* (7), 1035-40.
81. Ollis, D. L.; Cheah, E.; Cygler, M.; Dijkstra, B.; Frolow, F.; Franken, S. M.; Harel, M.; Remington, S. J.; Silman, I.; Schrag, J., The alpha/beta hydrolase fold. *Protein Eng* **1992,** *5* (3), 197-211.
82. Rawlings, N. D.; Barrett, A. J.; Bateman, A., MEROPS: the peptidase database. *Nucleic Acids Res* **2010,** *38* (Database issue), D227-33.
83. Danks, M. K.; Morton, C. L.; Krull, E. J.; Cheshire, P. J.; Richmond, L. B.; Naeve, C. W.; Pawlik, C. A.; Houghton, P. J.; Potter, P. M., Comparison of activation of CPT-11 by rabbit and human carboxylesterases for use in enzyme/prodrug therapy. *Clin Cancer Res* **1999,** *5* (4), 917-24.
84. Bencharit, S.; Morton, C. L.; Xue, Y.; Potter, P. M.; Redinbo, M. R., Structural basis of heroin and cocaine metabolism by a promiscuous human drug-processing enzyme. *Nat Struct Biol* **2003,** *10* (5), 349-56.
85. Bencharit, S.; Morton, C. L.; Hyatt, J. L.; Kuhn, P.; Danks, M. K.; Potter, P. M.; Redinbo, M. R., Crystal structure of human carboxylesterase 1 complexed with the Alzheimer's drug tacrine: from binding promiscuity to selective inhibition. *Chem Biol* **2003,** *10* (4), 341-9.

86. Potter, P. M.; Wadkins, R. M., Carboxylesterases--detoxifying enzymes and targets for drug therapy. *Curr Med Chem* **2006,** *13* (9), 1045-54.
87. Stok, J. E.; Goloshchapov, A.; Song, C.; Wheelock, C. E.; Derbel, M. B.; Morisseau, C.; Hammock, B. D., Investigation of the role of a second conserved serine in carboxylesterases via site-directed mutagenesis. *Arch Biochem Biophys* **2004,** *430* (2), 247-55.
88. Suzuki-Kurasaki, M.; Yoshioka, T.; Uematsu, T., Purification and characterization of guinea-pig liver microsomal deacetylase involved in the deacetylation of the O-glucoside of N-hydroxyacetanilide. *Biochem J* **1997,** *325 (Pt 1)*, 155-61.
89. Wang, X.; Wang, C. S.; Tang, J.; Dyda, F.; Zhang, X. C., The crystal structure of bovine bile salt activated lipase: insights into the bile salt activation mechanism. *Structure* **1997,** *5* (9), 1209-18; Terzyan, S.; Wang, C. S.; Downs, D.; Hunter, B.; Zhang, X. C., Crystal structure of the catalytic domain of human bile salt activated lipase. *Protein Sci* **2000,** *9* (9), 1783-90.
90. Streit, T. M.; Borazjani, A.; Lentz, S. E.; Wierdl, M.; Potter, P. M.; Gwaltney, S. R.; Ross, M. K., Evaluation of the 'side door' in carboxylesterase-mediated catalysis and inhibition. *Biol Chem* **2008,** *389* (2), 149-62.
91. Zhu, H. J.; Appel, D. I.; Johnson, J. A.; Chavin, K. D.; Markowitz, J. S., Role of carboxylesterase 1 and impact of natural genetic variants on the hydrolysis of trandolapril. *Biochem Pharmacol* **2009,** *77* (7), 1266-72.
92. Tarkiainen, E. K.; Backman, J. T.; Neuvonen, M.; Neuvonen, P. J.; Schwab, M.; Niemi, M., Carboxylesterase 1 polymorphism impairs oseltamivir bioactivation in humans. *Clin Pharmacol Ther* **2012,** *92* (1), 68-71.
93. Kroetz, D. L.; McBride, O. W.; Gonzalez, F. J., Glycosylation-dependent activity of baculovirus-expressed human liver carboxylesterases: cDNA cloning and characterization of two highly similar enzyme forms. *Biochemistry* **1993,** *32* (43), 11606-17.
94. Collet, X.; Fielding, C. J., Effects of inhibitors of N-linked oligosaccharide processing on the secretion, stability, and activity of lecithin:cholesterol acyltransferase. *Biochemistry* **1991,** *30* (13), 3228-34.
95. Ross, M. K.; Borazjani, A., Enzymatic activity of human carboxylesterases. *Curr Protoc Toxicol* **2007,** *Chapter 4*, Unit 4.24.
96. Lamego, J.; Cunha, B.; Peixoto, C.; Sousa, M. F.; Alves, P. M.; Simplício, A. L.; Coroadinha, A. S., Carboxylesterase 2 production and characterization in human cells: new insights into enzyme oligomerization and activity. *Appl Microbiol Biotechnol* **2012**.
97. Redinbo, M. R.; Potter, P. M., Mammalian carboxylesterases: from drug targets to protein therapeutics. *Drug Discov Today* **2005,** *10* (5), 313-25.
98. Rozen, S.; Skaletsky, H., Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **2000,** *132*, 365-86.
99. Brown, M. H.; Barclay, A. N., Expression of immunoglobulin and scavenger receptor superfamily domains as chimeric proteins with domains 3 and 4 of CD4 for ligand analysis. *Protein Eng* **1994,** *7* (4), 515-21.
100. Los, G. V.; Encell, L. P.; McDougall, M. G.; Hartzell, D. D.; Karassina, N.; Zimprich, C.; Wood, M. G.; Learish, R.; Ohana, R. F.; Urh, M.; Simpson, D.; Mendez, J.; Zimmerman, K.; Otto, P.; Vidugiris, G.; Zhu, J.; Darzins, A.; Klaubert, D. H.; Bulleit, R. F.; Wood, K. V., HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol* **2008,** *3* (6), 373-82.
101. Berrow, N. S.; Alderton, D.; Sainsbury, S.; Nettleship, J.; Assenberg, R.; Rahman, N.; Stuart, D. I.; Owens, R. J., A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res* **2007,** *35* (6), e45.

102. Berrow, N. S.; Alderton, D.; Owens, R. J., The precise engineering of expression vectors using high-throughput In-Fusion PCR cloning. *Methods Mol Biol* **2009,** *498*, 75-90.
103. Ho, S. N.; Hunt, H. D.; Horton, R. M.; Pullen, J. K.; Pease, L. R., Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **1989,** *77* (1), 51-9.
104. Aricescu, A. R.; Lu, W.; Jones, E. Y., A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr D Biol Crystallogr* **2006,** *62* (Pt 10), 1243-50.
105. Weng, S.; Spiro, R. G., Demonstration that a kifunensine-resistant alpha-mannosidase with a unique processing action on N-linked oligosaccharides occurs in rat liver endoplasmic reticulum and various cultured cells. *J Biol Chem* **1993,** *268* (34), 25656-63.
106. Reeves, P. J.; Kim, J. M.; Khorana, H. G., Structure and function in rhodopsin: a tetracycline-inducible system in stable mammalian cell lines for high-level expression of opsin mutants. *Proc Natl Acad Sci U S A* **2002,** *99* (21), 13413-8.
107. Nettleship, J. E.; Rahman-Huq, N.; Owens, R. J., The production of glycoproteins by transient expression in Mammalian cells. *Methods Mol Biol* **2009,** *498*, 245-63.
108. Wilkins, M. R.; Gasteiger, E.; Bairoch, A.; Sanchez, J. C.; Williams, K. L.; Appel, R. D.; Hochstrasser, D. F., Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* **1999,** *112*, 531-52.
109. Takahashi, S.; Katoh, M.; Saitoh, T.; Nakajima, M.; Yokoi, T., Allosteric kinetics of human carboxylesterase 1: species differences and interindividual variability. *J Pharm Sci* **2008,** *97* (12), 5434-45.
110. Boonyuen, U.; Promnares, K.; Junkree, S.; Day, N. P.; Imwong, M., Efficient in vitro refolding and functional characterization of recombinant human liver carboxylesterase (CES1) expressed in E. coli. *Protein Expr Purif* **2014**.
111. Williams, E. T.; Bacon, J. A.; Bender, D. M.; Lowinger, J. J.; Guo, W. K.; Ehsani, M. E.; Wang, X.; Wang, H.; Qian, Y. W.; Ruterbories, K. J.; Wrighton, S. A.; Perkins, E. J., Characterization of the expression and activity of carboxylesterases 1 and 2 from the beagle dog, cynomolgus monkey, and human. *Drug Metab Dispos* **2011,** *39* (12), 2305-13.
112. Potter, P. M.; Pawlik, C. A.; Morton, C. L.; Naeve, C. W.; Danks, M. K., Isolation and partial characterization of a cDNA encoding a rabbit liver carboxylesterase that activates the prodrug irinotecan (CPT-11). *Cancer Res* **1998,** *58* (12), 2646-51.
113. Brown, P. H.; Schuck, P., Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. *Biophys J* **2006,** *90* (12), 4651-61.
114. Schuck, P.; Rossmanith, P., Determination of the sedimentation coefficient distribution by least-squares boundary modeling. *Biopolymers* **2000,** *54* (5), 328-41.
115. Walter, T. S.; Diprose, J. M.; Mayo, C. J.; Siebold, C.; Pickford, M. G.; Carter, L.; Sutton, G. C.; Berrow, N. S.; Brown, J.; Berry, I. M.; Stewart-Jones, G. B.; Grimes, J. M.; Stammers, D. K.; Esnouf, R. M.; Jones, E. Y.; Owens, R. J.; Stuart, D. I.; Harlos, K., A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization. *Acta Crystallogr D Biol Crystallogr* **2005,** *61* (Pt 6), 651-7.
116. Gorrec, F., The MORPHEUS protein crystallization screen. *Journal of Applied Crystallography* **2009,** *42*, 1035-1042.
117. Newman, J.; Egan, D.; Walter, T. S.; Meged, R.; Berry, I.; Ben Jelloul, M.; Sussman, J. L.; Stuart, D. I.; Perrakis, A., Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy. *Acta Crystallogr D Biol Crystallogr* **2005,** *61* (Pt 10), 1426-31.
118. Gulick, A. M.; Horswill, A. R.; Thoden, J. B.; Escalante-Semerena, J. C.; Rayment, I., Pentaerythritol propoxylate: a new crystallization agent and cryoprotectant

induces crystal growth of 2-methylcitrate dehydratase. *Acta Crystallogr D Biol Crystallogr* **2002,** *58* (Pt 2), 306-9.
119. McCoy, A. J.; Grosse-Kunstleve, R. W.; Storoni, L. C.; Read, R. J., Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* **2005,** *61* (Pt 4), 458-64.
120. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **2004,** *60* (Pt 12 Pt 1), 2126-32.
121. Murshudov, G.; Vagin, A.; Dodson, E., Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D-Biological Crystallography* **1997,** *53*, 240-255.
122. Walter, T. S.; Diprose, J. M.; Mayo, C. J.; Siebold, C.; Pickford, M. G.; Carter, L.; Sutton, G. C.; Berrow, N. S.; Brown, J.; Berry, I. M.; Stewart-Jones, G. B.; Grimes, J. M.; Stammers, D. K.; Esnouf, R. M.; Jones, E. Y.; Owens, R. J.; Stuart, D. I.; Harlos, K., A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization. *Acta Crystallogr D Biol Crystallogr* **2005,** *61* (Pt 6), 651-7.
123. Winter, G., xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* **2010,** *43*, 186-190.
124. Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S., Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **2011,** *67* (Pt 4), 235-42.
125. Vaughn, J. L.; Goodwin, R. H.; Tompkins, G. J.; McCawley, P., The establishment of two cell lines from the insect Spodoptera frugiperda (Lepidoptera; Noctuidae). *In Vitro* **1977,** *13* (4), 213-7.
126. Fleming, C. D.; Edwards, C. C.; Kirby, S. D.; Maxwell, D. M.; Potter, P. M.; Cerasoli, D. M.; Redinbo, M. R., Crystal structures of human carboxylesterase 1 in covalent complexes with the chemical warfare agents soman and tabun. *Biochemistry* **2007,** *46* (17), 5063-71; Wadkins, R. M.; Hyatt, J. L.; Edwards, C. C.; Tsurkan, L.; Redinbo, M. R.; Wheelock, C. E.; Jones, P. D.; Hammock, B. D.; Potter, P. M., Analysis of mammalian carboxylesterase inhibition by trifluoromethylketone-containing compounds. *Mol Pharmacol* **2007,** *71* (3), 713-23; Nishi, K.; Huang, H.; Kamita, S. G.; Kim, I. H.; Morisseau, C.; Hammock, B. D., Characterization of pyrethroid hydrolysis by the human liver carboxylesterases hCE-1 and hCE-2. *Arch Biochem Biophys* **2006,** *445* (1), 115-23.
127. Parkinson, E. I.; Jason Hatfield, M.; Tsurkan, L.; Hyatt, J. L.; Edwards, C. C.; Hicks, L. D.; Yan, B.; Potter, P. M., Requirements for mammalian carboxylesterase inhibition by substituted ethane-1,2-diones. *Bioorg Med Chem* **2011,** *19* (15), 4635-43; Crow, J. A.; Bittles, V.; Herring, K. L.; Borazjani, A.; Potter, P. M.; Ross, M. K., Inhibition of recombinant human carboxylesterase 1 and 2 and monoacylglycerol lipase by chlorpyrifos oxon, paraoxon and methyl paraoxon. *Toxicol Appl Pharmacol* **2012,** *258* (1), 145-50.
128. Satoh, T.; Hosokawa, M.; Atsumi, R.; Suzuki, W.; Hakusui, H.; Nagai, E., Metabolic activation of CPT-11, 7-ethyl-10-[4-(1-piperidino)-1-piperidino]carbonyloxycamptothecin, a novel antitumor agent, by carboxylesterase. *Biol Pharm Bull* **1994,** *17* (5), 662-4.
129. Morton, C. L.; Potter, P. M., Comparison of Escherichia coli, Saccharomyces cerevisiae, Pichia pastoris, Spodoptera frugiperda, and COS7 cells for recombinant gene expression. Application to a rabbit liver carboxylesterase. *Mol Biotechnol* **2000,** *16* (3), 193-202.

130. Greenblatt, H. M.; Otto, T. C.; Kirkpatrick, M. G.; Kovaleva, E.; Brown, S.; Buchman, G.; Cerasoli, D. M.; Sussman, J. L., Structure of recombinant human carboxylesterase 1 isolated from whole cabbage looper larvae. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **2012,** *68* (Pt 3), 269-72.
131. Yin, J.; Li, G.; Ren, X.; Herrler, G., Select what you need: a comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. *J Biotechnol* **2007,** *127* (3), 335-47.
132. Paul F. Cook, W. W. C., *Enzyme Kinetics and Mechanism*. Taylor and Francis Group, LLC: 2007.
133. Wang, J.; Williams, E. T.; Bourgea, J.; Wong, Y. N.; Patten, C. J., Characterization of recombinant human carboxylesterases: fluorescein diacetate as a probe substrate for human carboxylesterase 2. *Drug Metab Dispos* **2011,** *39* (8), 1329-33.
134. Copeland, R. A., Enzymes: A Practical Introduction to Structure, Mechanism and Data Analysis. Second Edition ed.; Wiley-VCH: 2000; p. 390.
135. Ejima, K.; Liu, J.; Oshima, Y.; Hirooka, K.; Shimanuki, S.; Yokota, Y.; Hemmi, H.; Nakayama, T.; Nishino, T., Molecular cloning and characterization of a thermostable carboxylesterase from an archaeon, Sulfolobus shibatae DSM5389: non-linear kinetic behavior of a hormone-sensitive lipase family enzyme. *J Biosci Bioeng* **2004,** *98* (6), 445-51.
136. Otwinowski, Z.; Minor, W., Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Pt a* **1997,** *276*, 307-326.
137. Aricescu, A. R.; Owens, R. J., Expression of recombinant glycoproteins in mammalian cells: towards an integrative approach to structural biology. *Curr Opin Struct Biol* **2013,** *23* (3), 345-56.
138. Josse, D.; Xie, W.; Renault, F.; Rochu, D.; Schopfer, L. M.; Masson, P.; Lockridge, O., Identification of residues essential for human paraoxonase (PON1) arylesterase/organophosphatase activities. *Biochemistry* **1999,** *38* (9), 2816-25.
139. .
140. Hatfield, M. J.; Tsurkan, L.; Garrett, M.; Shaver, T. M.; Hyatt, J. L.; Edwards, C. C.; Hicks, L. D.; Potter, P. M., Organ-specific carboxylesterase profiling identifies the small intestine and kidney as major contributors of activation of the anticancer prodrug CPT-11. *Biochem Pharmacol* **2011,** *81* (1), 24-31.
141. Senter, P. D.; Beam, K. S.; Mixan, B.; Wahl, A. F., Identification and activities of human carboxylesterases for the activation of CPT-11, a clinically approved anticancer drug. *Bioconjug Chem* **2001,** *12* (6), 1074-80.
142. Slabinski, L.; Jaroszewski, L.; Rychlewski, L.; Wilson, I. A.; Lesley, S. A.; Godzik, A., XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics* **2007,** *23* (24), 3403-5.
143. Kelley, L. A.; Sternberg, M. J., Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **2009,** *4* (3), 363-71.
144. Goldschmidt, L.; Cooper, D. R.; Derewenda, Z. S.; Eisenberg, D., Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. *Protein Sci* **2007,** *16* (8), 1569-76.
145. Longenecker, K. L.; Garrard, S. M.; Sheffield, P. J.; Derewenda, Z. S., Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of RhoGDI. *Acta Crystallogr D Biol Crystallogr* **2001,** *57* (Pt 5), 679-88.
146. Yang, Z. R.; Thomson, R.; McNeil, P.; Esnouf, R. M., RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* **2005,** *21* (16), 3369-76.
147. Charasson, V.; Bellott, R.; Meynard, D.; Longy, M.; Gorry, P.; Robert, J., Pharmacogenetics of human carboxylesterase 2, an enzyme involved in the activation of irinotecan into SN-38. *Clin Pharmacol Ther* **2004,** *76* (6), 528-35.

148. Korzekwa, K. R.; Krishnamachary, N.; Shou, M.; Ogai, A.; Parise, R. A.; Rettie, A. E.; Gonzalez, F. J.; Tracy, T. S., Evaluation of atypical cytochrome P450 kinetics with two-substrate models: evidence that multiple substrates can simultaneously bind to cytochrome P450 active sites. *Biochemistry* **1998,** *37* (12), 4137-47.
149. Reed, M. C.; Lieb, A.; Nijhout, H. F., The biological significance of substrate inhibition: a mechanism with diverse functions. *Bioessays* **2010,** *32* (5), 422-9.
150. Hutzler, J. M.; Tracy, T. S., Atypical kinetic profiles in drug metabolism reactions. *Drug Metab Dispos* **2002,** *30* (4), 355-62.
151. Muta, K.; Fukami, T.; Nakajima, M.; Yokoi, T., N-Glycosylation during translation is essential for human arylacetamide deacetylase enzyme activity. *Biochem Pharmacol* **2014,** *87* (2), 352-9.
152. Kunimoto, T.; Nitta, K.; Tanaka, T.; Uehara, N.; Baba, H.; Takeuchi, M.; Yokokura, T.; Sawada, S.; Miyasaka, T.; Mutai, M., Antitumor activity of 7-ethyl-10-[4-(1-piperidino)-1-piperidino]carbonyloxy-camptothec in, a novel water-soluble derivative of camptothecin, against murine tumors. *Cancer Res* **1987,** *47* (22), 5944-7.
153. O'Reilly, S.; Rowinsky, E. K., The clinical status of irinotecan (CPT-11), a novel water soluble camptothecin analogue: 1996. *Crit Rev Oncol Hematol* **1996,** *24* (1), 47-70; Herben, V. M.; Ten Bokkel Huinink, W. W.; Schellens, J. H.; Beijnen, J. H., Clinical pharmacokinetics of camptothecin topoisomerase I inhibitors. *Pharm World Sci* **1998,** *20* (4), 161-72; Vanhoefer, U.; Harstrick, A.; Achterrath, W.; Cao, S.; Seeber, S.; Rustum, Y. M., Irinotecan in the treatment of colorectal cancer: clinical overview. *J Clin Oncol* **2001,** *19* (5), 1501-18; Rothenberg, M. L., Irinotecan (CPT-11): recent developments and future directions--colorectal cancer and beyond. *Oncologist* **2001,** *6* (1), 66-80; Masuda, N.; Kudoh, S.; Fukuoka, M., Irinotecan (CPT-11): pharmacology and clinical applications. *Crit Rev Oncol Hematol* **1996,** *24* (1), 3-26.
154. Khanna, R.; Morton, C. L.; Danks, M. K.; Potter, P. M., Proficient metabolism of irinotecan by a human intestinal carboxylesterase. *Cancer Res* **2000,** *60* (17), 4725-8; Humerickhouse, R.; Lohrbach, K.; Li, L.; Bosron, W. F.; Dolan, M. E., Characterization of CPT-11 hydrolysis by human liver carboxylesterase isoforms hCE-1 and hCE-2. *Cancer Res* **2000,** *60* (5), 1189-92.
155. Slatter, J. G.; Schaaf, L. J.; Sams, J. P.; Feenstra, K. L.; Johnson, M. G.; Bombardt, P. A.; Cathcart, K. S.; Verburg, M. T.; Pearson, L. K.; Compton, L. D.; Miller, L. L.; Baker, D. S.; Pesheck, C. V.; Lord, R. S., Pharmacokinetics, metabolism, and excretion of irinotecan (CPT-11) following I.V. infusion of [(14)C]CPT-11 in cancer patients. *Drug Metab Dispos* **2000,** *28* (4), 423-33.
156. Rivory, L. P.; Bowles, M. R.; Robert, J.; Pond, S. M., Conversion of irinotecan (CPT-11) to its active metabolite, 7-ethyl-10-hydroxycamptothecin (SN-38), by human liver carboxylesterase. *Biochem Pharmacol* **1996,** *52* (7), 1103-11.
157. Karplus, P. A.; Diederichs, K., Linking crystallographic model and data quality. *Science* **2012,** *336* (6084), 1030-3.
158. Matthews, B. W., Solvent content of protein crystals. *J Mol Biol* **1968,** *33* (2), 491-7.
159. de Beer, T. A.; Berka, K.; Thornton, J. M.; Laskowski, R. A., PDBsum additions. *Nucleic Acids Res* **2014,** *42* (Database issue), D292-6; Laskowski, R. A.; Hutchinson, E. G.; Michie, A. D.; Wallace, A. C.; Jones, M. L.; Thornton, J. M., PDBsum: a Web-based database of summaries and analyses of all PDB structures. *Trends Biochem Sci* **1997,** *22* (12), 488-90.
160. Krem, M. M.; Di Cera, E., Molecular markers of serine protease evolution. *EMBO J* **2001,** *20* (12), 3036-45; McGrath, M. E.; Vásquez, J. R.; Craik, C. S.; Yang, A. S.; Honig, B.; Fletterick, R. J., Perturbing the polar environment of Asp102 in trypsin: consequences of replacing conserved Ser214. *Biochemistry* **1992,** *31* (12), 3059-64.

161. Kubo, T.; Kim, S. R.; Sai, K.; Saito, Y.; Nakajima, T.; Matsumoto, K.; Saito, H.; Shirao, K.; Yamamoto, N.; Minami, H.; Ohtsu, A.; Yoshida, T.; Saijo, N.; Ohno, Y.; Ozawa, S.; Sawada, J., Functional characterization of three naturally occurring single nucleotide polymorphisms in the CES2 gene encoding carboxylesterase 2 (HCE-2). *Drug Metab Dispos* **2005,** *33* (10), 1482-7.
162. Bosron, W. F.; Hurley, T. D., Lessons from a bacterial cocaine esterase. *Nat Struct Biol* **2002,** *9* (1), 4-5.
163. Kim, S. R.; Nakamura, T.; Saito, Y.; Sai, K.; Nakajima, T.; Saito, H.; Shirao, K.; Minami, H.; Ohtsu, A.; Yoshida, T.; Saijo, N.; Ozawa, S.; Sawada, J., Twelve novel single nucleotide polymorphisms in the CES2 gene encoding human carboxylesterase 2 (hCE-2). *Drug Metab Pharmacokinet* **2003,** *18* (5), 327-32.
164. Holmes, R. S.; Cox, L. A.; VandeBerg, J. L., Mammalian carboxylesterase 3: comparative genomics and proteomics. *Genetica* **2010,** *138* (7), 695-708.
165. Feng, X. M.; Xiong, J.; Qin, H.; Liu, W.; Chen, R. N.; Shang, W.; Ning, R.; Hu, G.; Yang, J., Fluoxetine induces hepatic lipid accumulation via both promotion of the SREBP1c-related lipogenesis and reduction of lipolysis in primary mouse hepatocytes. *CNS Neurosci Ther* **2012,** *18* (12), 974-80; Lian, J.; Quiroga, A. D.; Li, L.; Lehner, R., Ces3/TGH deficiency improves dyslipidemia and reduces atherosclerosis in Ldlr(-/-) mice. *Circ Res* **2012,** *111* (8), 982-90; Lian, J.; Wei, E.; Wang, S. P.; Quiroga, A. D.; Li, L.; Di Pardo, A.; van der Veen, J.; Sipione, S.; Mitchell, G. A.; Lehner, R., Liver specific inactivation of carboxylesterase 3/triacylglycerol hydrolase decreases blood lipids without causing severe steatosis in mice. *Hepatology* **2012,** *56* (6), 2154-62.
166. Gonzalez, F. J., The study of xenobiotic-metabolizing enzymes and their role in toxicity in vivo using targeted gene disruption. *Toxicol Lett* **1998,** *102-103*, 161-6.
167. Sanghani, S. P.; Sanghani, P. C.; Schiel, M. A.; Bosron, W. F., Human carboxylesterases: an update on CES1, CES2 and CES3. *Protein Pept Lett* **2009,** *16* (10), 1207-14.
168. Quinney, S. K.; Sanghani, S. P.; Davis, W. I.; Hurley, T. D.; Sun, Z.; Murry, D. J.; Bosron, W. F., Hydrolysis of capecitabine to 5'-deoxy-5-fluorocytidine by human carboxylesterases and inhibition by loperamide. *J Pharmacol Exp Ther* **2005,** *313* (3), 1011-6.
169. Gilson, M. K.; Straatsma, T. P.; McCammon, J. A.; Ripoll, D. R.; Faerman, C. H.; Axelsen, P. H.; Silman, I.; Sussman, J. L., Open "back door" in a molecular dynamics simulation of acetylcholinesterase. *Science* **1994,** *263* (5151), 1276-8.