STRUCUTRE/FUNCTION RELATIONSHIPS OF THE ESTERASE 6 PROTEIN OF DROSOPHILA MELANOGASTER

by

Mark Andrew Myers

Department of Botany,
The Australian National University,
Canberra.

and

C.S.I.R.O.,
Division of Entomology,
Canberra.

Declaration

The research carried out in the course of this investigation and the results presented in thesis are, except where acknowledged, the original work of the author.

Mark A. Myers
Acknowledgements

I would like to thank my supervisors Dr. John Oakeshott and Dr. Angela Delves for their support and friendship. In particular I thank John for his enthusiasm and for creating a fruitful and rewarding work environment. My thanks also go to Marion Healy for her help and support throughout the years.

Thanks are also due to Rollin Richmond for the gift of anti-sera and purified protein and to Jill Karotam and Lis van Papenrecht for allowing me to use their unpublished results. The numerous members of the Oakeshott laboratory, past and present, I thank for their friendship, support and help. They are; Kristen Ardlie, Alex Bortoli, Peter Campbell, Peter Cooke, Mira Dumancic, Annie Game, Marion Healy, Jill Karotam, Joanne Labate, Wendy Odgers, Anthony Parker, Robyn Russell, Marlene Saad, Kerrie Ann Smyth, Merrin Spackman and Lis van Papenrecht and all the other people who I should have thanked but can't fit on the page.

A special thankyou goes to Peter Christian, Peter Cooke, Jane Symmonds and Bill Warren for helpful discussions, meals, an ear to listen, and friendship.

I would also like to thank my parents, Kath and Eric, whose support and love have made this all possible.

Last but not least, I wish to thank Carolyn for her love, patience, support and understanding.
I dedicate this thesis to Carolyn, my best friend and future wife.
# TABLE OF CONTENTS

**SUMMARY**

<table>
<thead>
<tr>
<th>CHAPTER 1: GENERAL INTRODUCTION</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 MOLECULAR EVOLUTION AND MULTIGENE FAMILIES</td>
<td>1</td>
</tr>
<tr>
<td>1.2 THE SERINE HYDROLASE MULTIGENE FAMILY</td>
<td>3</td>
</tr>
<tr>
<td>1.2.1 The Serine Proteases</td>
<td>3</td>
</tr>
<tr>
<td>1.2.2 The Serine Esterases</td>
<td>4</td>
</tr>
<tr>
<td>1.2.3 Evolution of the Serine Hydrolases</td>
<td>5</td>
</tr>
<tr>
<td>1.3 ESTERASE 6 OF <em>DROSOPHILA MELANOGASTER</em></td>
<td>6</td>
</tr>
<tr>
<td>1.3.1 Biochemistry of Esterase 6</td>
<td>6</td>
</tr>
<tr>
<td>1.3.2 Physiology of Esterase 6</td>
<td>7</td>
</tr>
<tr>
<td>1.3.3 Genetic and Molecular Characterisation of Esterase 6</td>
<td>8</td>
</tr>
<tr>
<td>1.4 STRUCTURE/FUNCTION RELATIONSHIPS OF THE SERINE ESTERASES</td>
<td>10</td>
</tr>
<tr>
<td>1.4.1 Catalytic Residues</td>
<td>10</td>
</tr>
<tr>
<td>1.4.2 Other Structural Features</td>
<td>12</td>
</tr>
<tr>
<td>1.5 STRUCTURE/FUNCTION ANALYSIS USING PROTEIN ENGINEERING</td>
<td>13</td>
</tr>
<tr>
<td>1.5.1 Substrate Specificity</td>
<td>13</td>
</tr>
<tr>
<td>1.5.2 Catalytic Mechanism</td>
<td>16</td>
</tr>
<tr>
<td>1.6 ANALYSIS OF THE ROLE OF GLYCOSYLATION USING PROTEIN ENGINEERING</td>
<td>18</td>
</tr>
<tr>
<td>1.6.1 Glycosylation and Intracellular Transport</td>
<td>19</td>
</tr>
<tr>
<td>1.6.2 Glycosylation and Biological Function</td>
<td>21</td>
</tr>
<tr>
<td>1.7 AIMS OF THIS STUDY</td>
<td>22</td>
</tr>
</tbody>
</table>

**CHAPTER 2: ACTIVE SITE ANALYSIS**

| 2.1 INTRODUCTION | 24 |
2.1.1 Choice of Mutations 24
2.1.2 Kinetic Analysis 25

2.2 MATERIALS AND METHODS 29

2.2.1 Construction of Mutant Est6 Clones 29
  2.2.1.1 General Molecular Biological Techniques 29
  2.2.1.2 Oligo-nucleotide Directed in vitro Mutagenesis 30
2.2.2 P-element Mediated Germ-line Transformation 31
  2.2.2.1 Fly Stocks and Culture Conditions 31
  2.2.2.2 Germ-line Transformation Procedure 32
2.2.3 Protein Purification 33
  2.2.3.1 General Procedures for EST6 and Protein Detection 33
  2.2.3.2 Purification Procedure 34
2.2.4 Biochemical Comparison of Mutant and Wild Type EST6 36
  2.2.4.1 Quantitation of EST6 Active Sites in the Purified Preparations 36
  2.2.4.2 Sensitivity to DFP 36
  2.2.4.3 Esterase Assays for Kinetic Analysis 37
  2.2.4.4 Effect of Temperature and pH 39
  2.2.4.5 Analysis of Initial Reaction Rate Data 39

2.3 RESULTS 40

2.3.1 Expression of the EST6 Active Site Mutants 40
2.3.2 Purification of Wild Type and Mutant EST6 40
2.2.3 Biochemical Analysis 43
  2.2.3.1 The Effect of pH 43
  2.3.3.2 Inhibition by DFP 43
  2.2.3.3 Active Site Quantitation 44
  2.3.3.4 Substrate Specificity 45
  2.3.3.5 The Effect of Temperature 47

2.4 DISCUSSION 50

2.4.1 The Effect of pH 50
2.4.2 Substrate Interactions 52
2.4.3 Activation Energies
2.4.4 Conclusions

CHAPTER 3: ANALYSIS OF ASPARAGINE-LINKED GLYCOSYLATION

3.1 INTRODUCTION

3.1.1 Choice of Mutations

3.2 MATERIALS AND METHODS

3.2.1 Construction and Expression of Mutant Clones

3.2.1.1 Oligo-nucleotide Directed \textit{in vitro} Mutagenesis
3.2.1.2 Oligo-nucleotide Probes and Colony Hybridisation
3.2.1.3 Construction of a Clone With Four N-linked Glycosylation Site Mutations
3.2.1.4 Germ-line Transformation

3.2.2 Biochemical Analysis of the Mutant Proteins

3.2.2.1 Preparation of Homogenates
3.2.2.2 Western Blots and Antibody Detection
3.2.2.3 Concanavalin A Binding
3.2.2.5 Glycanase Digestion
3.2.2.5 Thermostability

3.2.3 Physiological Analysis

3.2.3.1 Tissue Distribution
3.2.3.2 Mating Experiments

3.3 RESULTS

3.3.1 Biochemical Characterisation of EST6 with Mutant Glycosylation Sites

3.3.1.1 Electrophoretic Mobility and Size Estimation
3.3.1.2 Concanavalin A binding
3.3.1.3 Glycanase Digestion
3.3.1.4 Thermostability

3.3.2 Physiological Characterisation
3.4 DISCUSSION

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3.2.1 Tissue Localisation</td>
<td>76</td>
</tr>
<tr>
<td>3.3.2.2 Transfer to Females</td>
<td>77</td>
</tr>
<tr>
<td>3.4.1 Structural Analysis of the Glycosylation of EST6</td>
<td>79</td>
</tr>
<tr>
<td>3.4.2 Functional Role of the EST6 Glycans</td>
<td>82</td>
</tr>
<tr>
<td>3.4.3 A Neutral Hypothesis for the Role of Glycosylation of EST6</td>
<td>84</td>
</tr>
<tr>
<td>3.4.4 Conclusions</td>
<td>87</td>
</tr>
</tbody>
</table>

CHAPTER 4: DISCUSSION

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 THE SERINE ACTIVE SITE</td>
<td>88</td>
</tr>
<tr>
<td>4.1.1 Comparison with Distantly Related Esterases</td>
<td>89</td>
</tr>
<tr>
<td>4.1.2 Comparison with EST6 Homologues from Other Drosophila Species</td>
<td>91</td>
</tr>
<tr>
<td>4.1.3 Future Work</td>
<td>93</td>
</tr>
<tr>
<td>4.2 N-LINKED GLYCOSYLATION</td>
<td>94</td>
</tr>
<tr>
<td>4.2.1 Comparison with Distantly Related Esterases</td>
<td>95</td>
</tr>
<tr>
<td>4.2.2 Comparison with EST6 Homologues from Other Drosophila Species</td>
<td>97</td>
</tr>
<tr>
<td>4.2.3 Future Work</td>
<td>99</td>
</tr>
<tr>
<td>4.3 CONCLUSIONS</td>
<td>100</td>
</tr>
</tbody>
</table>

REFERENCES                                                                 | 102 |
APPENDIX I                                                                 | 121 |
APPENDIX II                                                                | 122 |
APPENDIX III                                                               | 123 |
SUMMARY

This thesis analyses the function of two structural features of the serine hydrolase, esterase-6 (EST6), from Drosophila melanogaster; namely, the non-consensus serine active site, and the N-glycosylation of the protein. EST6 was chosen as the subject of this study for several reasons. Firstly, the population genetics of EST6 has been extensively analysed and this enzyme has been found to be polymorphic in natural populations. A latitudinal cline for the two major electromorphs of the enzyme has also been identified which suggests natural selection is acting on this polymorphism. Secondly, EST6 has been implicated in the mating biology of D. melanogaster and is transferred from the male ejaculatory duct to the female's reproductive system during mating and subsequently translocated to her haemolymph. However, EST6 appears to have other functional roles in closely related species, suggesting that this enzyme is undergoing rapid evolutionary change. Thirdly, the Est6 gene forms part of a tandem duplication of esterase genes and is a member of a multi-gene family whose members have diverse functional roles. These features make EST6 a good model system for the study of molecular evolution. Lastly, the gene encoding EST6 has been cloned and characterised and a germ-line transformation procedure for D. melanogaster is available. These features enable specific alteration of the protein's structure through manipulation of the DNA encoding the protein, and analysis of the biochemical and physiological affects of these alterations on the enzyme's function by expression of the mutated genes in vivo.

A highly conserved region containing the reactive serine has been identified by sequence comparison of a number of serine esterases. Such a comparison reveals that EST6 has a non-consensus sequence; this enzyme has histidine (position 187) adjacent to the reactive serine in place of the consensus glutamic acid. The role of this histidine residue in the function of EST6 has been investigated by using site directed in vitro mutagenesis to replace the histidine codon with either glutamine or glutamic acid codons, followed by expression, purification and biochemical characterisation of these active site mutants and comparison of their properties with the wild type protein.

Such an analysis reveals that substitution of histidine 187 for glutamine has little effect on the biochemical properties of EST6. In contrast, the presence of glutamic acid at position 187 alters the properties of the enzyme dramatically. The major effects of the histidine to glutamic acid substitution are as follows. The pH
profile of this enzyme is altered such that the optimum pH is increased from 7 to 9, which is consistent with an alteration in the catalytic mechanism of the enzyme at alkaline pH. This substitution also changes the substrate preference of EST6, such that the enzyme has gained the ability to hydrolyse acetylthiocholine, while showing decreased activity for ßnaphthylesters and p-nitrophenylacetate. Lastly, the Gibb's free energy of activation for this mutant enzyme is increased in comparison to wild type EST6. The latter results are consistent with residue 187 interacting directly with the alkyl group of the substrate and suggest that this interaction is fully realised in the transition state. From these results it is apparent that residue 187 plays a role in substrate binding and that the presence of histidine in place of the consensus glutamic acid at this position in EST6 is responsible for at least some of the functional divergence apparent between EST6 and acetylcholinesterases.

The second structural feature of EST6 analysed in this study is the N-glycosylation of the protein. EST6 has previously been shown to be a glycoprotein with four potential N-linked glycosylation sites. However, the glycosylation state of these sites and the structure and function of the attached oligosaccharides remains unknown. These aspects of the glycosylation of EST6 have been addressed in this study by substituting glutamine for the asparagine residues in the four potential N-linked glycosylation sites of EST6. This was achieved by site directed in vitro mutagenesis and expression of the mutant Est6 clones in vivo by germ-line transformation.

Five mutant EST6 proteins were produced; four lacked one each of the four potential glycosylation sites, while the fifth lacked all four sites. Analysis of these EST6 mutants revealed that each of the potential glycosylation sites has attached oligosaccharides of 1 to 2 kilodaltons in the wild type protein. However, the oligosaccharide attached to the second glycosylation site is slightly larger than the others and is sensitive to cleavage by endoglycosidase F and endoglycosidase H, which indicates it is of the simple, high mannose type. The oligosaccharides attached to the other three sites may be of a more complex structure because of their resistance to cleavage by these endoglycosidases. The second glycosylation site mutation also leads to an increase in thermostability but this effect is lost when all four of the glycosylation site mutations are present at once, presumably due to compensation by, or interaction with, one or more of the other mutations.
While the first three potential N-linked glycosylation sites are glycosylated on wild type EST6, the fourth site does not always have oligosaccharide attached to it. This results in two molecular forms of EST6 which differ in the amount of carbohydrate attached to them. This differential glycosylation at the fourth site may arise because of its proximity to the carboxyl terminal of the protein.

Functional analysis of the role of the glycans revealed that the non-glycosylated mutant was secreted efficiently into the ejaculatory duct of the males, as evidenced by the transfer of substantial amounts of this enzyme to Est6^O females during mating, and was subsequently translocated to the female’s haemolymph, as evidenced by its appearance in her thorax. This male donated mutant enzyme can still be detected in recipient females two days after mating, suggesting that the in vivo stability of the enzyme is not compromised by the lack of carbohydrate attachment.

From these results no obvious function for the attached oligosaccharides is apparent. Therefore, the possibility that the EST6 oligosaccharides are functionally neutral cannot be eliminated on the basis of this study. In view of this, it has been calculated that the number of N-linked glycosylation site consensus sequences present in EST6 is not significantly different to that expected by chance. However, more detailed analysis of these EST6 proteins, with mutations in the N-linked glycosylation sites, is required to determine if the attached oligosaccharides play a more subtle functional role.
CHAPTER 1. GENERAL INTRODUCTION

This thesis aims to define the functional role of various structural features of the serine hydrolase, esterase 6 (EST6), of *Drosophila melanogaster*. The work is in four parts; this chapter reviews the literature on the evolution of multigene families and the determination of structure/function relationships of proteins, with special reference to the serine hydrolase multigene family. Literature concerning the esterase 6 gene-enzyme system is also reviewed. Chapters two and three describe the experimental analysis of two particular structural features of EST6 and the investigation of their functional roles. The fourth chapter discusses the importance of these structural features in an evolutionary and functional context and proposes future work which may resolve some of the questions raised.

1.1 MOLECULAR EVOLUTION AND MULTIGENE FAMILIES

The duplication and subsequent divergence of genetic loci are accepted as a major mechanism for the evolution of loci encoding novel functions (MacIntyre, 1976). This view holds that when gene duplication events occur one of the gene copies is effectively freed from the constraints of natural selection (Ohno, 1970; Kimura, 1983). The redundant copy is then free to accumulate previously "forbidden" mutations, i.e. deleterious to the performance of the gene's original function, to eventually fulfil a new function which is of advantage to the organism. Ohno (1970) has proposed that multi-locus isozymes represent an early step in this process.

It has been proposed that untranslatable copies of duplicate genes, as opposed to multi-locus isozymes, are a necessary intermediate step for the rapid accumulation of mutations (Koch, 1972). After accumulation of various substitutions, deletions and insertions, reversion of the mutation responsible for suppression of translation then exposes the new gene product to the process of natural selection. Many eukaryote genomes do in fact harbour a pool of "silent" duplicated genes, the best known examples being the globin pseudogenes (Proudfoot and Maniatis, 1980; Nishioka *et al*, 1980).

While the proposals put forward by Ohno (1970) and Koch (1972) differ in the nature of the duplicate intermediates required for rapid divergence in function, they do agree that the relaxation of selective constraint through gene duplication is a requirement for the evolution of genes with novel function.

Another mechanism proposed for the evolution of novel genes is the process of exon shuffling (Gilbert, 1978; Tonegawa *et al*, 1978). This was proposed to explain the observation that exons coding for homologous protein domains are common to otherwise dissimilar proteins. One example of this is bovine thyroglobulin, the
carboxyl terminal domain of which has 30% sequence identity to the cholinesterases, while the amino terminal domains appear to be tandem duplications of another, unrelated domain (Mori et al, 1987). Another example is the proteases of blood coagulation and fibrinolysis, which have large non-catalytic regulatory domains attached to trypsin-like protease domains and which have been proposed to have arisen by exon shuffling (Patthy, 1985).

The existence of multigene families and clusters of related genes in particular chromosomal regions of eukaryotes (Dayhoff et al, 1972a; Zouros et al, 1982; Hedrich and von Deimling, 1987; Collet et al, 1990) provides evidence for the occurrence of gene duplication and subsequent divergence during the course of evolution. A widely accepted corollary of this is that functional similarity reflects evolutionary relatedness. However, such functional relatedness may also arise through convergent evolution. In this case the proteins are analogous to each other rather than homologous.

Homologous proteins can be further classed by whether they represent the product of the same gene in different species (orthologous), or the products of different but related genes in the same species (paralogous). Only the latter class necessarily arises through gene duplication so the distinction between an orthologous or paralogous relationship between the genes in question is important.

It is generally accepted that most of the divergence observed between homologous gene sequences is largely neutral to selection (Kimura, 1983). Neutral substitutions accumulate in a gene through random mutation and are not selected against because they have very little or no effect on the function of the gene or its product. For example, nucleotide changes occur more often in the untranslated regions of genes, such as introns or the intergenic regions that have no known function, than they do in the coding regions (Miyata et al, 1980). Within the coding region, nucleotide changes that do not alter the encoded amino acid sequence are more frequent than those that do (Kreitman, 1983). And in the encoded protein itself, substitution of amino acids on the surface of the protein, and which are not involved in functional interactions, are more common than substitution of residues important for structure or function of the protein, e.g. the active sites of enzymes (Dickerson, 1962). In the case of proteins, it is difficult to identify amino acid changes responsible for functional divergence if little is known about the residues important for the function of the protein.
The serine hydrolases represent one of the largest and best characterised examples of a multigene family on the basis of both functional and structural criteria and will now be discussed further as an example of some of the above principles.

1.2 THE SERINE HYDROLASE MULTIGENE FAMILY

It has been proposed that the serine proteases and serine esterases are phylogenetically related and so constitute a serine hydrolase multigene family (Augustinsson, 1968). Evidence in support of this includes their sensitivity to organophosphate and carbamate inhibitors (Augusteyn et al, 1969), overlapping substrate specificity (Heymann, 1980; Preverio et al, 1983) and irreversible inhibition by diisopropylfluorophosphate (DFP) through covalent attachment to the reactive serine (Krisch, 1971). The most compelling evidence comes from the sequence of tryptic peptides containing the serine active sites of a number of carboxyl esterases identified by the presence of bound DFP (Dayhoff et al, 1972b). Comparison of these serine active sites with those of the trypsin-like proteases (Young et al, 1978) reveals a common amino acid motif, Gly-X-Ser-X-Gly, where Gly represents glycine, Ser the reactive serine and X is variable. This match suggests a common evolutionary origin (Neurath, 1984).

Recently, full sequence data have become available for a number of serine esterases and comparison of these with the serine proteases suggests these enzymes can be divided into several distinct multigene families on the basis of sequence similarities. Furthermore the serine proteases themselves appear to fall into two distinct classes. The three main serine hydrolase families are the trypsin-like proteases, the subtilisin proteases and the esterases/lipases. The evidence for this restructuring of the multigene family is discussed below. It will also be argued in section 1.2.3 that the Gly-X-Ser-X-Gly motif may represent a relic sequence from an enzyme ancestral to all of the serine hydrolases.

1.2.1 THE SERINE PROTEASES

The common evolutionary origin of the human pancreatic serine proteases chymotrypsinogen and trypsinogen has been recognised for over twenty five years (Walsh and Neurath, 1964; Hartley et al, 1965). The number of proteases with significant sequence identity to these enzymes presently stands at thirty five, with representatives from bacteria, insects and mammals (Rogers, 1985; Greer, 1990). While some of these enzymes may be orthologous, the majority (over twenty) have
been isolated from humans and so can be considered paralogous. Many of these enzymes are involved in blood clotting and complement fixation (Young et al, 1978). Apart from the reactive serine, the existence of essential aspartic acid (Asp) and histidine (His) residues involved directly in a catalytic charge relay system is well established (Price and Stevens, 1982). The three residues involved in the catalytic mechanism and the regions of primary sequence immediately surrounding these are highly conserved amongst all the members of this family. The crystal structures of a number of these proteases have been determined and the conservation of their primary structure is reflected in similar secondary and tertiary conformations (Greer, 1990).

A second family of serine hydrolases with basically similar properties is the subtilisin proteases. On the basis of sequence similarity this family is distinct from the trypsin-like proteases but shares the same catalytic mechanism (Kraut, 1977). However, the three catalytic residues, His, Asp and Ser, appear in a different order in the primary structure (Wells et al, 1983). The serine active site also has the sequence Gly-X-Ser-X-Ala which is in contrast to the other serine hydrolases. Members of this family have also been crystallised and, apart from the relative orientation of the residues involved in the catalytic triad, there is no similarity in tertiary structure with the trypsin-like serine proteases (Kraut, 1977). The only representatives of this family so far identified have come from bacteria and there is no evidence that gene duplication has played a role in the evolution of the subtilisin proteases as all the members of the family so far identified come from different species of bacteria.

1.2.2 THE SERINE ESTERASES

A third family of serine hydrolases is the esterases. This class bears no sequence similarity to the trypsin-like or subtilisin serine proteases apart from the motif Gly-X-Ser-X-Gly around the reactive serine. The catalytic mechanism of this family of serine hydrolases is not fully understood but apart from the reactive Ser, good evidence exists for the involvement of an essential His residue (section 1.3.4.1) and for a different catalytic mechanism to the proteases (Quinn, 1987). Enzymes included in this family on the basis of sequence similarity are the cholinesterases and acetylcholinesterases (Hall and Spierrer, 1986; Doctor et al, 1990), some lipases (Shimada et al, 1989 and 1990; Kawaguchi et al, 1989), lysophospholipases and cholesterol esterases (Han et al, 1987; Kyger et al, 1989; Kissel et al, 1989), non-specific carboxylesterases (Oakeshott et al, 1987; Ozols, 1987; Long et al, 1988;
Sergeev et al., 1989) and juvenile hormone esterase (Hanzlik et al., 1989). They are found in eukaryotic microorganisms, insects and vertebrates but as yet none have been reported from prokaryotes. Numerous examples of esterase gene clusters have been provided by genetic and molecular analysis, providing strong evidence of gene duplication occurring in this family (Zouros et al., 1982; Hedrich and von Deimling, 1987; Collet et al., 1990). While two members of this family have been crystallised (Hata et al., 1979; Schrag et al., 1988) the tertiary structures have not been solved and so only limited information concerning the three dimensional structure of these enzymes is available.

1.2.3 EVOLUTION OF THE SERINE HYDROLASES

It is apparent from the above discussion that the grouping of serine hydrolases into a single superfamily of phylogenetically related enzymes, on the basis of functional criteria, is a misleading representation of the evolutionary relationships between these enzymes. Sequence comparisons reveal that at least three multigene families are actually involved. The common active site sequence, Gly-X-Ser-X-Gly/Ala, present in each of these families then is either due to convergent evolution or represents a relic sequence from a distant common ancestor.

Brenner (1988) has proposed that the serine hydrolases have descended from a precursor peptide containing cysteine, the sulphydryl group of which was capable of binding metal ions which endowed the precursor with catalytic activity. Brenner proposes that the motif Gly-X-Ser-X-Gly is a relic of this precursor "enzyme" and that during the course of evolution elaboration of the enzyme's structure has eliminated the requirement of metal ions, allowing cysteine to be replaced by serine in the active site. Each serine hydrolase family described above may represent a separate line of descent from the primitive precursor peptide. In support of this is the observation that a number of viral cysteine proteases are homologous to the trypsin-like serine proteases. In these enzymes cysteine replaces serine as the active site nucleophile but the catalytic Asp and His are absolutely conserved (Bazan and Fletterick, 1988). Furthermore, mutant forms of both acetylcholinesterase (Gibney et al., 1990) and trypsin (Higaki et al., 1989) containing a cysteine residue substituted for the reactive serine retain significant ester hydrolytic activity, albeit greatly reduced in comparison to the wild type enzymes. This demonstrates that cysteine can, to some extent, act as the active site nucleophile in some serine hydrolases.
Within families of homologous serine hydrolases the processes of gene duplication and subsequent divergence have led to a diversity of functions. An example of divergence between paralogous proteins is provided by the bovine serine proteases, trypsinogen, chymotrypsinogen A/B and elastase, which differ markedly in their substrate specificity and primary sequence although the essential catalytic residues and their overall tertiary structures are conserved (Hartley et al, 1965; Walsh and Neurath, 1964).

In contrast to this, orthologous enzymes can show a high degree of sequence divergence while still retaining the same functions. For example, the substrate preference of acetylcholinesterase from *D. melanogaster* and *T. californica*, orthologous members of the serine esterase family, is conserved, despite an overall amino acid sequence identity of only 38% (Myers et al, 1988).

As can be seen from this brief description, the processes of gene duplication and subsequent functional divergence of the paralogous enzymes have played a part in the evolution of the serine hydrolase multigene family, although conservation of function between orthologous enzymes is also apparent. The issue of neutral and functional amino acid differences between members of this family and the conservation of residues essential for activity will be discussed in detail in section 1.5.

The next section introduces esterase 6, the subject of this thesis. This enzyme belongs to the serine esterase family of serine hydrolases and has a number of characteristics which make it an interesting evolutionary case study.

1.3 ESTERASE 6 OF DROSOPHILA MELANOGASTER

Esterase 6 (Est6 = gene; EST6 = protein; E.C.3.1.1.1) is the major β-carboxyl esterase of *D. melanogaster* and has been well characterised at the biochemical, physiological, genetic and molecular levels. A review of the literature covering each of these aspects of the esterase 6 system is presented.

1.3.1 BIOCHEMISTRY OF ESTERASE 6

EST6 is classified as a β-carboxylesterase on the basis of its sensitivity to organophosphate inhibitors, resistance to the cholinesterase inhibitor eserine and a preference for βnaphthylester substrates when challenged with a mixture of *a* and βnaphthylesters. It has been proposed that the majority of esterases should be classified as "general esterases" until more information, particularly the *in vivo*
substrate, is known (Walker and Mackness, 1983). As the *in vivo* substrate for EST6 is unknown, it may also be referred to as a general esterase.

EST6 in *D. melanogaster* and its sibling species *D. simulans* and *D. mauritiana* has been shown to be a monomer, by inhibitor titration of purified enzyme, with a molecular weight of approximately 60 kilodaltons (Morton and Singh, 1985; Mane *et al*, 1983b). Amino acid analysis of purified EST6 demonstrated the presence of glucosamine residues (Mane *et al*, 1983b). This, in conjunction with the observations that EST6 binds to the lectin concanavalin A and reacts positively in a phenolsulphuric acid reagent test, demonstrate that this enzyme is a glycoprotein (Mane *et al*, 1983b). The presence of glucosamine residues and the binding to concanavalin A suggests that the carbohydrate is linked to asparagine residues, although the presence of O-linked carbohydrate cannot be ruled out. Amino acid analysis also reveals the absence of half cystine residues, suggesting that all the cysteine residues in the protein are involved in disulphide bridges (Mane *et al*, 1983b).

EST6 displays a preference for substrates with two (acetate) or three (propionate) carbon acyl groups, has reduced activity with four carbon (butyrate) acyl groups and negligible activity when the acyl chain is five carbon atoms or more (Mane *et al*, 1983b; White *et al*, 1988). Both pnitrophenyl and Bnaphthyl ester substrates are hydrolysed at a substantial rate (Mane *et al*, 1983b). The pH optimum of the enzyme is 7-8 and below pH 6 activity decreases dramatically (White *et al*, 1988). A similar pH profile has been observed for acetylcholinesterase and is indicative of the involvement of a histidine residue in catalysis (Rosenberry, 1975; see section 1.4.1).

### 1.3.2 Physiology of Esterase 6

The physiology of EST6 has been extensively analysed and a role in reproductive behaviour is implicated. The enzyme is present in the haemolymph of larvae and both male and female adults, but the major site of expression is in the anterior ejaculatory duct (AED) of adult males, where activity increases after eclosion to reach a maximum 12 to 36 hours later (Sheehan *et al*, 1979). The enzyme is transferred from the AED to the female during the first minute after mating initiation, prior to sperm transfer, and is then rapidly translocated to her haemolymph where it can be detected for up to 4 days later (Richmond and Senior, 1981; Meikle *et al*, 1990). These observations have prompted analysis of the effects that EST6 has on mating.
Analysis of the mating behaviour of females mated to EST6 null males versus EST6 producing males has been performed and a number of differences have been identified. At 18°C, males producing EST6 initiated mating seven minutes sooner and copulation duration was five minutes shorter in comparison to the EST6 null males (Gilbert and Richmond, 1982). At 25°C both of these effects were absent. It has been suggested that at the lower temperature EST6 may increase sperm motility, thus aiding in sperm transfer, storage and use, which in turn leads to the observed effects on mating behaviour (Gilbert and Richmond, 1982). As sperm motility increases with increasing temperature (Gromko et al, 1984a), the absence of an EST6 effect at 25°C may be expected.

The remating behaviour of females is also affected by the presence or absence of male donated EST6 from the first mating. Remating during the first day after an initial mating to an EST6 active male is less likely when compared to females initially mated to EST6 null males (Gilbert et al, 1981a; Scott, 1986). This effect is due to decreased receptivity of the female to the second male (Scott, 1986). Such an effect on female remating behaviour may increase the fitness of the male, as a second mating within a day of the first results in between 40 and 90% of the first male's sperm being displaced in favour of the second male's (Gilbert et al, 1981b; Gromko et al, 1984a, 1984b).

The biochemical basis for these effects has not been forthcoming. At one stage it was believed cis-vaccenyl acetate, which produces anti-aphrodisiac effects when hydrolysed to cis-vaccenol, may have been the physiological substrate of EST6 (Mane et al, 1983a; Zawistowski and Richmond, 1986). It has since been shown that EST6 is not implicated in the control of the levels of these substances in flies of either sex (Vander Meer et al, 1986; Scott and Richmond, 1987).

1.3.3 GENETIC AND MOLECULAR CHARACTERISATION OF ESTERASE 6

The Est6 gene has been genetically mapped to the third chromosome at 36.8 map units (Wright, 1963) which corresponds to cytological position 69A1 (Oakeshott et al, 1987). EST6 has been shown to be variable in natural populations by both electrophoretic (Wright, 1963; Oakeshott et al, 1981; Cooke et al, 1987) and thermostability (Cochrane and Richmond, 1979; Labate et al, 1989) criteria. Latitudinal clines have been observed for the two common allozymes of EST6 (EST6-F and EST6-S) in both D.melanogaster and its sibling species D.simulans, suggesting that this polymorphism may be under natural selection (Anderson and Oakeshott, 1984).
Within *D. melanogaster*, at least ten electromorphs can be distinguished using high resolution electrophoretic techniques (Cooke *et al.*, 1987) and sequence analysis of these variants reveals a substantial amount of electrophoretically cryptic amino acid polymorphism (Cooke, 1989).

The sequencing of clones containing the esterase 6 structural gene shows that it encodes a protein of the correct molecular weight, containing four potential N-linked glycosylation sites, a 21 amino acid signal peptide, and bearing significant sequence similarity to acetylcholinesterase (Oakeshott *et al.*, 1987) and to a number of other esterases (section 1.2.2). The *Est6* transcript is detected in all life stages but has its highest levels in adult males, which corresponds to the observed distribution of the protein product of this gene. The coding region of the gene is interrupted by a single 51 base pair (bp) intron. All of these structural features and the expression pattern are conserved in *Est6* from the sibling species *D. simulans* and *D. mauritiana*, except that the signal peptide is two amino acids shorter and the fourth potential N-linked glycosylation site is absent in both (J. Karotam pers. comm.). This same glycosylation site is polymorphic in *D. melanogaster* (Cooke *et al.*, 1989). In *D. yakuba* and *D. erecta* the sexually dimorphic expression of the enzyme is not apparent (E. van Papenrecht, pers.comm.). In both of these species the EST6 protein is a dimer and this is believed to represent the more ancient form (Morton and Singh, 1985). This significant divergence between related species suggests that EST6 is undergoing rapid evolutionary change in the *melanogaster* subgroup.

The active site serine of EST6 has been identified on the basis of sequence comparisons with cholinesterases and a number of other carboxyl esterase active sites which have been identified directly by the ability of the organophosphate, DFP, to bind to the reactive serine (Dayhoff *et al.*, 1972b; Oakeshott *et al.*, 1987). A number of esterases have now been sequenced and the reactive serine is found in an equivalent position in the primary amino acid sequence of each and is embedded in an octapeptide of conserved amino acids (from here on termed the serine active site).

Sequence analysis of the region of genomic DNA surrounding *Est6* has revealed a second esterase gene only 197 bp 3' from the termination codon of *Est6* (Oakeshott *et al.*, 1987; Collet *et al.*, 1990). This gene, called *EstP*, has 64% DNA sequence identity with *Est6* and contains an intron in the equivalent position to *Est6* which is 56 bp, 5 bp longer than the *Est6* intron. The *EstP* gene is transcribed in late larvae and to a lesser extent in adults of both sexes but a protein product has not yet been identified.
So although $Est6$ and $EstP$ are closely related, they clearly have evolved different regulatory systems and probably different functions.

Collet et al (1990) propose that the $Est6/EstP$ duplication in $D.melanogaster$ is homologous to the $Est4/Est5$ duplication of $D.mojavensis$ and $D.arizonensis$ (Zouros et al, 1982), and the $Est1/EstJ$ duplication of $D.buzzatii$ (East, 1984). This duplication event would therefore be as old as the divergence of the repleta and melanogaster groups, which has been estimated to have occurred 60-80 million years ago (Throckmorton, 1975).

1.4. STRUCTURE/FUNCTION RELATIONSHIPS OF THE SERINE ESTERASES

While the three dimensional structure of a member of the serine esterase family is yet to be solved, a certain amount of information concerning particular structural features can be deduced in its absence. One approach is to compare the amino acid sequences of a number of esterases; residues essential for catalysis and correct structural conformation will be more conserved than those with less functional significance. Using such an approach, in combination with biochemical evidence from analysis of various serine esterases, a number of structural regions of $EST6$ and other serine esterases can be assigned putative functions.

1.4.1 CATALYTIC RESIDUES

The catalytic mechanism of the serine esterases is potentially different to that of the serine proteases. Quinn (1987), on the basis of proton inventories, has proposed that acetylcholinesterase does not use a charge relay system to shuttle protons, as is seen with the Asp-His-Ser catalytic triad of the serine proteases. Instead, an acid-base mechanism involving the reactive Ser and a His residue, but not necessarily an Asp residue, appears to be the mechanism used (Quinn, 1987).

The active site Ser residue has been identified as the active site nucleophile of a number of esterases by its ability to bind covalently to the organophosphate inhibitor DFP (Krisch, 1971; Dayhoff et al, 1972b). The reactive Ser is conserved in terms of the position in the primary structure (Figure 1.1) and the sequence of the residues surrounding it is also highly conserved in each of the esterases characterised (Figure 1.2). The sequence of each of the serine active sites shown conforms to the Gly-X-Ser-X-Gly/Ala consensus typical of the serine hydrolases (section 1.2).

A feature of the sequence surrounding the reactive Ser worthy of note is the presence of an acidic glutamate (Glu) residue adjacent to the reactive Ser, on the amino
Figure 1.1  Schematic representation of the structure of selected serine esterases. The reactive serine (Ser), and the histidine (His) residue implicated in the catalytic mechanism are labelled and their position in the primary sequence of each esterase is given. Potential N-linked glycosylation sites are shown (CHO). The glycosylation sites of Human BChE, rabbit liver esterase and Torpedo AChE are known to be glycosylated. Cysteine residues shown to be involved in disulphide bridge formation in Torpedo AChE are also labelled (C) and the horizontal lines linking these cysteines represents the disulphide bridge. Equivalent cysteine residues and their deduced disulphide bridging pattern in the other esterases are also shown.
Figure 1.2 The serine active site from selected esterases. The reactive serine is asterixed. Residue 187 in EST6 is underlined. Rab.Liv. EST = rabbit liver esterase. JHE = juvenile hormone esterase. References for each of these sequences are given in the text.
terminal side, in the majority of the characterised esterases (Figure 1.2). However, there are two exceptions; the first is EST6 which has the basic residue His, and the second is juvenile hormone esterase (JHE), which has the neutral glutamine (Gln) in this position. These amino acids are conservative in terms of the size of the side chains but differ in their charge. The existence of such substitutions in an otherwise highly conserved region of the protein implies that this residue may be responsible for some of the functional divergence observed between the members of this family.

The other residue believed to be involved in the catalytic mechanism of the serine esterases is His. Several observations lend strong support to the existence of such a residue and implicate one particular His as a candidate for the residue involved. This is based in part on the observation of an increase in velocity due to the deionisation of a group with a $pK_a$ around 6 for acetylcholinesterase (AChE) (Rosenberry, 1975). The specific group involved has been assumed to be His, the imidazole side chain of which has a $pK_a$ of 6.0. The proteases, which have been shown to have a His in the catalytic triad (section 1.2.1), have a similar pH profile indicating that this assumption is valid.

Other evidence for the involvement of His in the catalytic mechanism of serine esterases comes from crystallisation of a lipase from Geotrichum candidam, the gene for which has recently been cloned and sequenced and found to have homology to other serine esterase genes (Shimada et al, 1990). The low resolution crystal structure of this enzyme has been determined but the precise structure has not been solved (Hata et al, 1979). This study identified a His residue, by its ability to bind platinum, in what is presumed to be the substrate binding pocket of the crystal structure. Other chemical agents that modify His residues were shown to eliminate activity, suggesting that an essential His is involved in catalysis (Hata et al, 1979).

Further evidence comes from a study of DFP binding to a rabbit liver microsomal esterase (Ozols, 1987). In this study tryptic peptides of the enzyme, which had previously been treated with radiolabelled DFP, were isolated and those binding the DFP were sequenced. The active site Ser reacted with DFP, as expected, but there was also a second peptide containing DFP bound to a His residue (Ozols, 1987). The sequence surrounding this His was very similar to a conserved region in the primary sequence of the serine esterases that includes an invariant His. This suggests that the DFP binding His residue is involved in catalysis and is present in or near the active site so as to allow reaction with DFP.
Recently, protein engineering of two His residues in AChE, one of which (His440) is homologous to the DFP-reactive His discussed above, has been performed (Gibney et al, 1990). The two residues, His440 and His 425, have both been proposed as candidates in the catalytic mechanism of AChE (Sikorav et al, 1986). When each of these residues was independently mutated to Gln, activity was reduced by 52% (His425Gln) and 100% (His440Gln) respectively. As substitution of His 440 eliminates activity, this may be the His residue which is involved in catalysis (Gibney et al, 1990). It remains possible that the loss of activity upon substitution of His440 reflects disruption of the tertiary structure of the protein, rather than the direct effect of having removed an amino acid side chain involved in catalysis. Despite this criticism, these data, in conjunction with the other evidence summarised above, strongly implicate His440 in the catalytic mechanism of the serine esterases. This His residue is conserved in EST6 and all the other serine esterases for which full sequence data are available (Figure 1.1).

1.4.2 OTHER STRUCTURAL FEATURES

This section considers two other features of the serine esterases, namely the pattern of disulphide bridges and N-linked glycosylation.

The disulphide bridges of AChE from Torpedo californica have been determined (MacPhee-Quigley et al, 1986) and the cysteine residues involved are largely conserved in the other serine esterase sequences (Figure 1.1). The similar position in the primary sequence of the first four cysteine residues involved in the first two disulphide bridges indicates a high degree of conservation of this structural feature. EST6 has a third pair of cysteine residues in a different position to those found in AChE. There are two lines of evidence to suggest that these do in fact form a disulphide bridge. Firstly, no free sulphydryl groups are detected upon amino acid analysis (Mane et al, 1983b) and secondly, when one of these cysteines is substituted, as has been found in a naturally occurring variant of EST6, an unstable electrophoretic phenotype is apparent (Cooke, 1989). A third pair of cysteine residues is not present in all the sequences, although pairs of cysteine residues appear in different positions in the sequence of some which may be capable of forming disulphide bridges.

Another structural feature shown in Figure 1.1 is the potential N-linked glycosylation sites present in the primary structures of the esterases. Such sites are characterised by the sequence Asn-X-Ser/Thr, where X is any amino acid except proline (Marshall, 1972). EST6 from D.melanogaster has four potential N-linked
glycosylation sites, one of which is polymorphic in natural populations and absent in the sibling species *D. simulans* and *D. mauritiana* (J. Karotam, pers.comm.). At least one of these sites must be glycosylated as EST6 has been shown to have N-linked carbohydrate attached to the mature protein (section 1.3.1). A number of other esterases are known to be glycoproteins and their oligosaccharide attachment sites have been characterised. *Torpedo californica* AChE has four potential N-linked glycosylation sites, all of which have carbohydrate attached (MacPhee-Quigley *et al*, 1986), Human butyrylcholinesterase has 9 out of ten potential sites glycosylated (Lockridge *et al*, 1987) and rabbit liver microsomal esterase has carbohydrate attached at both of its potential glycosylation sites (Korza and Ozols, 1988).

1.5. STRUCTURE/FUNCTION ANALYSIS USING PROTEIN ENGINEERING

Of the above mentioned functional and structural features, two in particular are of potential functional importance to EST6 and are amenable to analysis. Firstly, the presence of His adjacent to the Ser in the serine active site of the enzyme may impart some unique functional feature on the EST6 protein, as the majority of the other esterases have Glu in this same position. As there is no evidence for the direct involvement of this residue in the catalytic mechanism, it may be possible to substitute other residues at this position, using site directed mutagenesis, to produce proteins with altered properties but which still retain hydrolytic activity. Secondly, the glycosylation of EST6 has not been characterised and no obvious functional role is apparent for the glycans of any esterases analysed. The potential N-linked glycosylation sites of a number of other proteins have been mutated using site directed mutagenesis techniques and the effect of this on the protein analysed. EST6 is amenable to this type of analysis as well.

Some examples of site directed mutagenesis being applied to the analysis of structure/function relationships in proteins will now be reviewed. The first section below concentrates on the analysis of substrate binding and catalysis by the serine proteases, which have been the most extensively analysed enzymes using protein engineering techniques. Then section 1.6. gives a background to glycosylation and describes some examples of the application of protein engineering to the analysis of the function of glycosylation. The examples described in section 1.6 involve a variety of proteins as very little of this work has been performed on serine proteases.
1.5.1 Substrate Specificity

From knowledge of the three dimensional structure of enzymes the particular amino acid side chains present in the substrate binding cleft can be identified and their potential functions deduced. Two families of enzyme for which such information is available are the trypsin-like proteases and the subtilisin proteases. Some elegant studies using protein engineering to confirm the role of particular amino acids believed to be involved in determining substrate preference, and which have greatly enhanced our understanding of the particular mechanisms involved in this function, will now be described.

Tertiary structural analysis reveals the presence of an Asp residue (position 189 in the primary sequence) at the base of the substrate binding pocket of trypsin, which is believed to confer the preference for positively charged arginyl and lysyl substrates displayed by this enzyme (Bode et al., 1975). Chymotrypsin, on the other hand, has Ser at the corresponding position, which may be responsible for this enzyme's preference for aromatic amino acid substrates (Bode and Schwager, 1975). The substrate binding pockets of these enzymes are otherwise similar (Kraut, 1977; Gráf et al., 1987) despite the overall amino acid sequence identity being only forty percent (Walsh and Neurath, 1964). Mutation of the Asp189 residue of trypsin to Ser, using site directed mutagenesis, changes the substrate preference of trypsin from basic amino acids to a chymotrypsin-like preference for aromatic amino acid substrates (Gráf et al., 1988). This confirms the deductions made from the tertiary structures of the enzymes and is consistent with the proposition that electrostatic complementarity between Asp 189 and basic substrates is responsible for the substrate preference of wild type trypsin (Gráf et al., 1988). However, when this same residue was mutated to the basic lysine the expected change to a preference for acidic substrates, predicted on the basis of electrostatic complementarity, was not observed, due to unexpected interactions with other regions of the protein (Gráf et al., 1987). This last result illustrates that novel catalytic activities are not necessarily arrived at by simply replacing a single amino acid residue with one displaying the desired properties, as indirect structural consequences of the substitution also come into play.

Another example where differences in substrate specificity between members of a multigene family have been explored through protein engineering involves trypsin and elastase (Sawyer et al., 1978). Comparison of the related tertiary structures of these enzymes suggests that two Gly residues located in the substrate binding pocket of
trypsin, at positions 216 and 226 in the primary sequence, are important in determining the substrate specificity differences between these enzymes. Based on these observations, Craik et al (1985) chose Gly 216 and Gly 226 as appropriate targets for the engineering of the substrate specificity of trypsin. Through molecular modelling, predictions of the effect on substrate specificity of substituting these Gly residues with Ala were made. The predictions were that Ala at position 216 should show better activity for arginine than lysine as a substrate while Ala at 226 should show relatively enhanced lysine activity. Protein engineering studies demonstrated these predictions to be correct.

Subtilisin has also been the subject of protein engineering of substrate specificity. In an extensive analysis of this enzyme, a conserved Gly at the base of the substrate cleft was replaced with twelve different nonionic amino acids and the substrate preference of the mutant enzymes tested against each of four hydrophobic substrates (Estell et al, 1986). The general finding of this study was a greater catalytic efficiency for small hydrophobic substrates when the Gly was substituted with hydrophobic amino acids, but not with amino acids with hydrophilic side chains. Once the optimum binding volume of the cleft was exceeded, by enlarging either the substrate or the volume of the amino acid replacing the Gly, a precipitous drop in catalytic efficiency was observed.

A second study used a similar approach to alter the substrate specificity by introducing charged residues into the substrate binding cleft of subtilisin. The same Gly residue as discussed above and a second residue, which is Glu in the wild type enzyme, were changed to residues differing in charge (Wells et al, 1987a). For substrates complementary in charge to the amino acid substituted in the substrate binding cleft an increase in the specificity constant ($k_{cat}/K_m$ see section 2.1) of up to 2,000 fold was observed, while the mutant enzymes displayed a decrease in specificity towards like-charged substrates.

An example where the substrate specificity of one member of the subtilisin gene family was engineered, by the introduction of three amino acid substitutions, to have the properties of another member of the family with vastly different substrate preferences has also been presented (Wells et al, 1987b). The introduced changes represent less than 4% of the amino acid sequence differences between these enzymes but were shown to account for all the observed differences in substrate preference.

Other examples of altered substrate preference due to the substitution of a single or a small number of amino acids have been reported for cytochrome P450...
(Lindberg and Negishi, 1989), carboxypeptidase Y (Winther et al, 1985) and tyrosyl tRNA synthetase (Fersht et al, 1985; Lowe et al, 1985). A generalisation that can be drawn from these and the examples described above is the importance of particular hydrogen bonds and hydrophobic and electrostatic interactions between individual amino acids and the substrate in the determination of substrate specificity. Consequently, the substrate specificity differences between members of a multigene family are most often due to a small fraction of the total number of amino acid differences between the enzymes.

1.5.2 CATALYTIC MECHANISM

Another functional feature of the serine proteases that has been investigated using site directed mutagenesis is the catalytic mechanism. The trypsin-like and the subtilisin proteases both utilise a charge relay system involving a His, an Asp and the reactive Ser residues (section 1.2.1). For both trypsin and subtilisin some or all of these residues have been mutated and the catalytic consequences for the mutant enzymes investigated.

The precise catalytic role of the Asp residue in the catalytic triad of the serine proteases has been the subject of much speculation. The catalytic His and Ser have been shown to be essential for catalysis through chemical modification techniques (Dixon et al, 1956; Shaw et al, 1965) and high resolution crystal structures have shown these two residues to be within hydrogen bonding distance of each other (Stroud et al, 1974). These and other data are consistent with a catalytic mechanism where the His extracts a proton from the Ser and donates a proton to the leaving group. The crystal structure has also revealed the catalytic Asp to be very close to the His, raising the possibility of a charge relay system with the Asp residue aiding in catalysis (Stroud et al, 1974). Consistent with this is that the Asp residue is invariant in the primary sequence and crystal structures of all studied serine proteases (Greer, 1990). However, selective chemical modification of this residue is not possible, so direct evidence of a role in catalysis is lacking. This evidence has now been provided using site directed mutagenesis.

The catalytic Asp (position 102 in the primary structure) of trypsin has been substituted with the isosteric amino acid asparagine (Asn) and the three dimensional structure and the kinetic parameters of the mutant protein characterised (Sprang et al, 1987; Craik et al, 1987). The three dimensional structure of the mutant enzyme was not significantly affected by this substitution (Sprang et al, 1987). Biochemical
analysis of the mutant enzyme detected significant hydrolytic activity but it was of a greatly reduced rate in comparison to the wild type (Craik et al, 1987). The major effects of the substitution are as follows; below pH 7, the pK_a of the His is lowered by 1.5 pH units in the mutant enzyme, the maximal reaction rate is lowered 5,000 fold, while the K_m shows only a slight increase. Above pH 7, the reaction rate increases in proportion to the solvent hydroxide concentration. Also, the reactive serine is 10,000 fold less sensitive to reactive serine specific inhibitors, indicating that the nucleophilicity of this residue is greatly compromised (Craik et al, 1987). The Asp 102 residue therefore seems to have its primary effect on the reactive Ser, supporting the role of this residue in the catalytic triad. The surprising result is the pH dependence of the enzyme above pH7, which suggests that solvent hydroxyl ions are being used in the reaction at alkali pH, though the precise mechanism by which this occurs is not understood (Craik et al, 1987).

The catalytic triad of subtilisin has also been analysed using site directed mutagenesis. Each of the Asp, His and Ser residues in the catalytic triad were substituted with Ala, individually and in combination (Carter and Wells, 1988). Mutation of either the Ser or the His of the catalytic triad reduced activity 2x10^6 fold while substitution of the Asp caused a 20,000 fold decrease in activity. Interestingly, the mutant enzyme with the Asp substituted showed a significantly greater increase in activity with increasing pH than either of the other mutant enzymes. Mutation of either the Ser or the His in combination with the Asp mutation further decreased activity in comparison to substitution of the Asp alone. The effects of the substitutions on K_m were minimal. These results demonstrate that each of the residues in the catalytic triad are essential for maximal activity but that the effect of substituting the Asp is 100 fold less than the that produced by substituting the other two catalytic residues.

There are two major similarities between the two studies described above. Firstly, substitution of the Asp in the catalytic triad reduces the rate of hydrolysis but does not greatly effect the K_m of the enzymes, which demonstrates that this residue is not involved in substrate binding. Secondly, these mutant enzymes display a corresponding increase in activity with increasing pH that is not observed in the wild type enzymes, which is best explained by the mutant enzymes utilising solvent hydroxyl ions in the catalytic mechanism.

An important difference between the effect of substituting residues involved directly in the catalytic mechanism and those involved in substrate binding is brought out by comparing the results described in section 1.5.1 and 1.5.2. Substitution of
residues involved in substrate binding alters the substrate preference of the enzyme in a manner dependent on the properties of the replacement amino acid and its orientation relative to the substrate, however, the enzyme still retains catalytic activity. When residues involved directly in catalysis are substituted, drastic decreases in hydrolytic activity are the result. This is consistent with the observation that residues involved in substrate binding are often divergent between members of the serine protease multi-gene families but the catalytic triad residues are absolutely conserved in all of the enzymes. Thus, the residues essential for function are conserved throughout evolution while those involved in more specific fine-tuning of that function, in this case substrate preference, are more likely to be variable, and this depends on the particular substrates the enzyme is required to recognise.

1.6 ANALYSIS OF THE ROLE OF GLYCOSYLATION USING PROTEIN ENGINEERING

There are two broad classes of carbohydrate side chain attached to glycoproteins, N-linked or O-linked, which are distinguished on the basis of the amino acid to which the glycan is attached. O-linked oligosaccharides are attached to the hydroxyl group of either serine or threonine residues, a process which occurs in the golgi apparatus, and there appears to be no requirement for a consensus recognition sequence in the protein for this to occur (Kornfeld and Kornfeld, 1985; Farquhar, 1985). In contrast, N-linked oligosaccharides are attached to the side chain amine group of Asn residues which occur in the receptor site consensus sequence Asn-X-Ser/Thr, where X is any amino acid except proline and either Ser or Thr is allowable in the third position (Marshall, 1972). The N-linked oligosaccharide structures themselves are divided into three broad types. All three have the same characteristic core structure but the structure of the polysaccharide branches differ in their monosaccharide composition. The three types are 1) the high mannose oligosaccharides, which have branches consisting of polymers of mannose, 2) the complex type, which also include fucose, galactose and glucosamine saccharides in their structures and 3) the hybrid type, which contain both of the other branch types in one oligosaccharide (Kornfeld and Kornfeld, 1985). The attachment and processing of N-linked glycans have been well characterised and will be described below.

The first step in N-linked glycosylation of glycoproteins is the attachment of a precursor oligosaccharide by the enzyme oligosaccharyl transferase (Kaplan et al, 1987). Synthesis of the protein is directed to the rough endoplasmic reticulum (RER) by the signal peptide (Verner and Schatz, 1988) where oligosaccharyl transferase,
which resides in the RER, attaches the precursor oligosaccharide as the nascent polypeptide is being synthesised (Kaplan et al., 1987). The carbohydrate is then further modified as the glycoprotein traverses the RER and golgi apparatus, with various types of modification occurring in particular sub-compartments of the organelles (Farquhar, 1985; Hirschberg and Snider, 1987; Pfeffer and Rothman, 1987). The golgi apparatus also directs proteins to their final destination. A well characterised example of this is targeting of lysosomal enzymes. These enzymes have a unique signal that binds a specific phosphotransferase which catalyses the formation of mannose 6-phosphate on their N-linked side chains. This in turn is recognised by the mannose 6-phosphate receptor which translocates these proteins to prelysosomal compartments that then go to form lysosomes (Kornfeld, 1987). In contrast, secreted glycoproteins appear to travel through the golgi apparatus by a more passive bulk flow mechanism and are then packaged into vesicles for release into the extracellular environment, with no evidence for the involvement of specific receptors in the process (Pfeffer and Rothman, 1987).

The roles of N and O-linked carbohydrate components of secreted, membrane-bound and lysosomal glycoproteins have been the subject of much study and speculation and many different roles for the glycans attached to glycoproteins have been implicated. These include a conformational requirement where the protein does not attain its correct conformation when not glycosylated and as a consequence is not secreted (Dorner et al., 1987), intracellular targeting to specific organelles such as lysosomes (discussed above), extracellular protein targeting and biological recognition (Paulson, 1989) and metabolic stabilisation by the attached oligosaccharides through protection of the polypeptide chain from proteolysis (Olden et al., 1982). A number of examples also exist where no biological function has been assigned to the glycans of glycoproteins despite extensive analysis (Spiro, 1970; Olden et al., 1982; Kornfeld and Kornfeld, 1985).

More recently, in vitro site directed mutagenesis has been used to mutate potential N-linked glycosylation sites in DNA clones of glycoprotein genes with an aim to elucidating the function of the glycans. These are then expressed in an appropriate expression system and the resultant non-glycosylated protein analysed. The results of a number of such studies are discussed below.
1.6.1 GLYCOSYLATION AND INTRACELLULAR TRANSPORT

The glycan moieties of cell surface and secreted glycoproteins have been implicated, by the use of protein engineering, in the transport of these proteins from the RER to the golgi and the cell surface. Non-glycosylation of erythropoietin (Ep) (Dubé et al, 1988) and the heavy chain of immunoglobulin A (IgA) (Taylor and Wall, 1988) by mutation of the glycosylation sites, resulted in drastically reduced secretion of these proteins. Ep has three N-linked glycans, whose attachment was prevented by substituting Gln for Asn in the acceptor sites, and one O-linked glycan attached to a serine. This serine was mutated to Gly which prevented carbohydrate attachment at this site. Less than 10% of the protein was secreted when either the O-linked site or two of the N-linked sites were mutated; the third N-linked site had no effect on secretion efficiency (Dubé et al, 1988). The two N-linked sites of IgA were mutated in the same manner and the absence of either also led to less than 10% of the protein being secreted (Taylor and Wall, 1988). However, another secreted glycoprotein, the serine protease tissue plasminogen activator (tPA), is still secreted when all three of the N-linked acceptor sites normally glycosylated in this protein are mutated (Hansen et al, 1988).

It has been suggested that a RER-localised protein, heavy chain binding protein (BiP), has a role in preventing incorrectly folded proteins in the RER from being secreted. BiP binds incorrectly folded proteins and so prevents their passage to the golgi apparatus (Dorner et al, 1987; Pelham, 1986). BiP associates with malfolded immunoglobulin heavy chains in the RER (Bole et al, 1986); hence the decreased secretion efficiency of this protein upon mutation of the glycosylation sites may be due to disrupted folding patterns. In the case of tPA, the glycan moieties are not required for conformation or folding, so removal of the glycosylation sites does not effect secretion. However, the conformational effect leading to decreased secretion can only apply to the N-linked sites as O-linked carbohydrate attachment occurs in the golgi apparatus, removed from the RER where BiP is localised. Because of this, the decreased secretion of Ep lacking the O-linked glycan was attributed to rapid degradation in the golgi (Dubé et al, 1988).

In contrast to these findings is the effect of glycosylation on cell surface glycoproteins. The addition of N-linked glycosylation sites to a hybrid cell surface protein allows cell surface localisation (Guan et al, 1985). The protein, a form of rat growth hormone which also contains the membrane spanning domains of vesicular stomatitis virus (VSV) glycoprotein, when not glycosylated, is transported efficiently to
the golgi where it is retained (Guan et al, 1985). Introduction of potential N-linked glycosylation sites results in glycosylation of the protein and transport of this glycosylated form from the golgi to the cell surface. Either of two different sites that were introduced resulted in this effect (Guan et al, 1985). Complementary to this is that either one of two N-linked glycans on the VSV G-protein is required for transport of this glycoprotein to the cell surface (Machamer et al, 1985). These findings have prompted the suggestion that the glycans themselves play an important role as recognition signals in the intracellular transport of cell surface glycoproteins (Machamer et al, 1985; Guan et al, 1985; Adams and Rose, 1985). A conformational effect in these cases is unlikely as, in both proteins, either one of two glycans was all that was necessary to enable correct transport of the protein.

A cell surface protein, histocompatibility antigen HLA-A2, has been analysed in a similar manner and gave contradictory results (Santos-Aguado et al, 1987). The single N-linked glycosylation site of this protein was mutated by an Asn to Gin substitution, which resulted in no carbohydrate attachment and accumulation of the protein within the cell rather than on the cell surface. Interestingly, when instead the Ser in the third position of the acceptor site was mutated to Gly, glycosylation was still prevented but the protein was transported to the cell surface (Santos-Aguado et al, 1987). The defect in transport with the Asn to Gln substitution therefore was not due to lack of carbohydrate attachment but to some conformational effect of the Gln substitution. In this case it is apparent that the N-linked carbohydrate is not necessary for cell surface localisation.

A possible explanation for the contradictory results concerning glycosylation and the transport of cell surface glycoproteins is that there is more than one pathway for transport to the cell surface, one of which requires glycosylation and one which does not. Clearly, further research is required before this issue can be resolved. The secreted glycoproteins present a more comprehensible situation where it seems glycosylation is not directly necessary for secretion. Only in cases where the glycans are essential for correct folding and conformation is secretion affected.

1.6.2 GLYCOSYLATION AND BIOLOGICAL FUNCTION

In two of the above examples the biological function of the non-glycosylated proteins was also analysed. These are the tPA and Ep studies.

The protease, tPA, plays a role in breaking down blood clots by hydrolysing plasminogen in the clot to the proteolytically active form, plasmin, which then in turn
breaks down the fibrin clot (Young et al, 1978). As non-glycosylation of the tPA did not affect the secretion of this protein, the non-glycosylated form was used in assays of conversion of plasminogen to plasmin, fibrin binding and fibrin clot lysing potential (Hansen et al, 1988). The non-glycosylated mutants retained wild type activity in the plasminogen activation assay, and at concentrations above 17nM clot lysis ability was also equivalent to wild type. However, at lower concentrations the non-glycosylated form of tPA was slightly more effective at clot lysis and also bound to fibrin more efficiently than the glycosylated form (Hansen et al, 1988). Hence, non-glycosylation of tPA does not affect its proteolytic activity but does have a subtle influence on the protein's ability to interact with its biological site of action, the blood clot. Surprisingly, binding to the clot is enhanced when tPA is not glycosylated, thus potentially increasing its bioactivity.

The second example is Ep, which is required for the maintenance, proliferation, and differentiation of the stem cells that produce erythrocytes (Dubé et al, 1988 and references therein). While non-glycosylation at two of the N-linked sites and the O-linked site of this protein greatly reduced the amount of the protein secreted, one N-linked site had no effect on secretion, enabling assay of its bioactivity. Also, one of the N-linked glycosylation site mutants that showed reduced secretion was still transported from the cell in sufficient quantities to allow determination of its specific activity (Dubé et al, 1988). These two mutant proteins displayed 25% and 20% of the wild type protein's ability to stimulate erythroid colony formation from Ep-responsive erythroid colony-forming cells in culture (Dubé et al, 1988).

In both of these examples, the precise molecular mechanism by which glycosylation acts in the functioning of the protein remains unknown. The observation of a role in fibrin binding suggests that the glycan moieties influence the interaction between tPA and fibrin, but further research is required to determine precisely how this is mediated.

1.7 AIMS OF THIS STUDY

Esterase 6 has been extensively studied as an evolutionary model for a number of reasons. These include the observation of parallel geographic clines in electromorph frequency in the sibling species D. melanogaster and D. simulans and its role in the mating biology of these species, which raise the possibility that natural selection acts at this locus. Secondly, divergent biochemical and physiological properties among other related species suggests that EST6 is undergoing rapid evolutionary change within the
melanogaster subgroup. Thirdly, the presence of a duplicated esterase gene close to Est6 and the different tissue and temporal expression patterns of these two genes make them of great interest in terms of the role of gene duplication in the process of molecular evolution. Lastly, esterase 6 belongs to a multi-gene family which includes enzymes involved in a diversity of functions including neuronal transmission (AChE) and control of in vivo titres of hormones (JHE).

To gain a full understanding of the molecular basis for the divergent functions of members of this multi-gene family, the functionally important amino acids need to be identified. Once identified, the particular functions of such residues and the mechanisms by which such functions are mediated can then be investigated. The technique of in vitro mutagenesis and subsequent expression of the mutant proteins enables the specific functions of particular amino acids to be tested.

The specific aims of this thesis are to characterise the function of two particular structural features of EST6, identified by comparison of the primary sequence of several esterases, using the techniques of oligo-nucleotide directed in vitro mutagenesis and germ-line transformation of D.melanogaster embryos. Firstly, the non-consensus serine active site of this enzyme was analysed by substituting amino acids with different properties at the position adjacent to the reactive serine, where EST6 has a His residue in place of the consensus Glu. Chapter 2 describes this analysis. The second structural feature to be analysed, described in chapter 3, was the N-linked glycosylation of the enzyme. Structural and functional analyses of EST6 proteins with mutations in the N-linked glycosylation sites are described. The last chapter presents a general discussion of the results in relation to both distantly related esterases, as well as esterase 6 homologues from other Drosophila species.
CHAPTER 2. ACTIVE SITE ANALYSIS

2.1 INTRODUCTION

Amino acid residues of putative functional importance to proteins can be identified through comparison of the sequences of a family of related but divergent proteins. Amino acids involved in the general functioning of the family of proteins in question will display a high degree of conservation between sequences. Residues involved in functions specific to a particular subset of the family should be conserved in those particular proteins but not in those with dissimilar specific functions. These criteria have enabled identification of the serine active site octapeptide of the serine esterases and raised the possibility that the His residue at position 187 in the active site of EST6 (adjacent to the reactive serine) may be responsible for some divergent properties of this enzyme (section 1.4.1).

This chapter describes the investigation of the role of this amino acid residue in the functioning of EST6 by site directed mutagenesis and biochemical analysis of the mutant proteins.

2.1.1 CHOICE OF MUTATIONS

In choosing amino acids to replace candidate residues in proteins it is important to minimise disruptions to the tertiary structure that may be caused by the substitution. Replacement of residues with amino acids carrying larger side chains can cause such problems due to steric effects. Hence changes to amino acids with isosteric or smaller side chains is recommended.

As EST6 has the basic residue His at position 187 while the majority of serine esterases have Glu in the analogous position, this is an obvious substitution to make. The side chain of Glu has a similar molecular volume to the imidazole group of His but differs significantly in that it is acidic and carries a negative charge at physiological pH. Another substitution which is more conservative is replacement of this residue with glutamine (Gln). The side chain of this amino acid is similar in structure to Glu but has the hydroxyl of the carboxylic acid group replaced with an amine. These substitutions give a series of alterations at the 187 position with His and Gln being neutral at physiological pH and Glu being essentially similar to Gln but carrying a negative charge. Each of these amino acids also differ in their potential hydrogen bonding and electrostatic interactions. Hence these
two mutations allow alteration of the physicochemical properties of the group at position 187 while controlling for steric effects.

2.1.2 Kinetic Analysis

If the residue at position 187 of EST6 is important in catalysis, substitution of this residue with those described above should result in an alteration of enzymatic activity. For many enzymatically catalysed reactions the velocity (v) of the reaction can be described as a function of substrate concentration [S] by the Michaelis-Menton equation,

\[ v = V_{\text{max}} \frac{[S]}{([S]+K_m)} \]  
(2.1)

where \( V_{\text{max}} = k_{\text{cat}}E_0 \), and \( E_0 \) is the concentration of enzyme active sites. The parameters \( k_{\text{cat}} \) and \( K_m \) are otherwise known as the turn-over number and the Michaelis constant respectively, and describe the maximal rate of substrate hydrolysis and the concentration of substrate at which half the maximal rate is achieved. These parameters are functions of a number of fundamental rate constants whose individual values are not readily determined.

EST6 is a serine hydrolase and so catalyses the hydrolysis of ester substrates through the following generalised reaction pathway

\[
\begin{align*}
E+S & \xrightleftharpoons{\text{k}_1}{\text{k}^{-1}} (ES) & & \xrightarrow{\text{k}_2} EQ + P & & \xrightarrow{\text{k}_3} E+Q + P \\
& \text{ } & & & & \text{ } \\
& \text{ } & & \text{ } & & \text{ } \\
\end{align*}
\]

where \( P \) and \( Q \) are the alkyl and acyl groups of the substrate respectively. Water acts as a second substrate in the reaction but since the reaction occurs in aqueous solution water can be considered to be saturating and so ignored in the consideration of rate constants. For initial reaction conditions the concentrations of \( P \) and \( Q \) are effectively zero so the rate of the reverse reaction given by \( k_2 \) and \( k_3 \) will be negligible in comparison to the rates of the forward reaction, \( k_2 \) and \( k_3 \). Under these circumstances \( k_2 \) and \( k_3 \) can be ignored and the reaction written, in folded form, as follows
It can be seen from this equation that $k_2$ describes the rate of acylation and $k_3$ the rate of deacylation of the enzyme. The rate equation (2.1) of the reaction can be expressed in terms of the rate constants in scheme (2.3) using the schematic method of King and Altman (1956). Application of this method yields the following rate expression

$$
\frac{dQ}{dT} = \frac{dP}{dT} = \frac{k_1 k_2 k_3 [S] E_0}{(k_2 k_3 + k_{-1} k_3 + k_1 k_3 [S] + k_1 k_2 [S])}
$$  \hspace{1cm} (2.4)

Dividing the numerator and denominator through by $k_1(k_2+k_3)$ gives

$$
\frac{k_2 k_3}{(k_2 + k_3)} [S].E_0
\frac{1}{k_3 (k_{-1}+k_2)} + [S]
\frac{1}{k_1 (k_2 + k_3)}
$$  \hspace{1cm} (2.5)

which is of the same form as equation (2.1), with the following identities
The usual method of determining \( k_{\text{cat}} \) and \( K_m \) is to measure the initial rate of hydrolysis at varying substrate concentrations. This gives estimates of \( V_{\text{max}} \) and \( K_m \) which are applicable only to the particular conditions used for the reactions and so are termed "apparent" values. As \( V_{\text{max}} = k_{\text{cat}} \cdot [E]_0 \), determination of the initial enzyme concentration enables calculation of \( k_{\text{cat}} \). However, no information as to the values of the true rate constants, \( k_1, k_{-1}, k_2 \) and \( k_3 \) is forthcoming from such an analysis. This is particularly important for the interpretation of \( K_m \) as this parameter is often taken as a direct measure of the dissociation constant \( K_s \) (\( k_{-1}/k_1 \)). As can be seen from equation 2.7, \( K_m \) bears little resemblance to \( K_s \), except under the extreme circumstance of \( k_2 \) being much smaller than the other rate constants involved. In this situation \( k_2 \) will make little contribution to the value of \( K_m \) and can be ignored, while \( k_3 \) will cancel, making \( K_m \) a close approximation to the true dissociation constant.

It has been shown for the protease chymotrypsin that the ratio \( k_{\text{cat}}/K_m \) is the best estimator of effectiveness of the enzyme in hydrolysing particular substrates (Brot and Bender, 1969). This parameter, otherwise known as the specificity constant, is directly related to the free energy difference between the enzyme-substrate transition state and the free unbound substrate and enzyme and should apply to any enzymatically catalysed reaction that obeys Michaelis-Menton kinetics (Fersht, 1974; Page, 1984). From equations 2.6 and 2.7 it is apparent that

\[
\frac{k_{\text{cat}}}{K_m} = \frac{k_1 k_2}{k_{-1} + k_2} \tag{2.8}
\]
The relationship of this parameter to the free energy of interaction between the substrate and enzyme has been utilised in the analysis of altered substrate binding due to introduced mutations (Fersht and Leatherbarrow, 1987; Wells et al, 1987b) and in measuring similarities of evolutionary diverse enzymes with similar functions (Withers and Rupitz, 1990). Such an analysis of EST6 and the active site mutants described above using a variety of substrates should reveal any alterations in substrate interaction caused by the mutations.

Another approach to identifying differences in the catalytic properties of enzymes is to observe their response to varying environmental conditions. Determination of $k_{\text{cat}}$ at differing temperatures enables calculation of the activation energies involved in the reaction. If either $k_2$ or $k_3$ is the rate limiting step the calculated activation energy will relate to that particular step. On the other hand, if $k_2$ and $k_3$ are very similar in magnitude, the calculated activation energy will be a function of the energy involved in both steps. In any case, an alteration in the energy required for these processes due to the substitution of His 187 may become apparent through such an analysis. A similar approach can be taken to analyse the variation of $K_m$ with temperature. If $K_m$ is a close approximation to $K_s$ then the calculated free energy will relate to substrate binding; if not, then the quantity will be a complex function of the energies of the catalytic steps that contribute to $K_m$.

Another environmental variable that can yield useful information on differences in aspects of catalysis is pH. The state of ionisation of groups involved in catalysis may have drastic effects on the rate of enzyme catalysed reactions. The variation of an enzyme's activity with pH has been utilised to determine the pKa of the particular groups involved in the catalytic mechanisms of serine hydrolases (section 1.4.1). Varying the pH at which the assay is performed and observing the effects on activity may reveal differences in the catalytic properties of enzymes otherwise not apparent.

By using the approaches described above to analyse the properties of EST6 and how they are affected by the substitution of His 187, the role of this amino acid in the catalytic functioning of the enzyme may be revealed. This chapter describes the construction of clones of Est6 with the codon for amino acid 187 mutated to either Gln or Glu using oligonucleotide directed in vitro mutagenesis, expression of the mutant proteins in the fly through P-element mediated germ-line transformation of D.melanogaster embryos, and subsequent purification and biochemical analysis of these mutant proteins.
2.2 MATERIALS AND METHODS

The materials and methods used in the construction and expression of the mutant Est6 clones and analysis of the mutant proteins are divided into four sections; 1) the construction of the mutant clones, 2) the expression of these using P-element mediated germ-line transformation 3) purification of wild type EST6 and the mutant proteins from the transformants, and 4) the biochemical characterisation of these purified enzymes.

2.2.1 CONSTRUCTION OF MUTANT EST6 CLONES

The techniques used in the construction of mutant Est6 clones are described. General techniques used are described followed by the procedure for oligo-nucleotide directed in vitro mutagenesis.

2.2.1.1 General Molecular Biological Techniques

All restriction enzymes were purchased from Promega and Boehringer Mannheim and reactions performed according to the manufacturer's instructions. Restriction fragments were fractionated on 1% agarose gels in 0.04 M Tris Acetate pH7.8, 1mM Ethylene-diamine-tetra-acetic acid (EDTA) and the required size of DNA fragment purified using GeneClean (BIO101) according to the manufacturer's instructions.

Genomic DNA was prepared from adult flies by the method of Coen et al (1982) as modified by Cooke (1989). Plasmid DNA was prepared according to the method of Miller (1972). Ligations were performed according to Maniatis et al (1982, pp 244-46).

All plasmid and single stranded DNA was prepared from cultures of the TG-1 strain of E.coli (K12, Δ[lac-pro], supE, thi, hsdD5/F'traD36, proA+B+, lacIq, lacZΔM15). The preparation of transformation competent TG-1 cells and the transformation of these cells were according to Hanahan (1985).

DNA Sequencing

Sequencing was performed by the dideoxy nucleotide chain termination method of Sanger et al (1977), using reagents obtained from Bresatec. Reactions were performed according to the manufacturer's instructions. A series of custom made oligo-nucleotides
homologous to regions of the EST6 gene were used as sequencing primers and have been described previously (Cooke, 1989).

**Southern Analysis**

Digested DNA was electrophoresed in 1% agarose gels and transferred to nitrocellulose (Hybond C, Amersham) in 0.27 M sodium acetate, 2.7M sodium chloride (NaCl) (20xSSC). Filters were then baked for 2 hours at 80°C in vacuo and then prehybridised at 65°C for at least 4 hours in 0.9M sodium chloride, 0.05M sodium phosphate pH7, 5mM EDTA, 0.1%PVP, 0.1% bovine serum albumin (BSA), 0.1% Ficoll, 0.1% Sodium dodecyl sulphate (SDS) and 0.5 mg/ml salmon sperm DNA. Nick translated probes (Maniatis *et al*, 1982, pp109-12) of a *HindIII* fragment from a genomic clone of EST6 (see Appendix I) (Oakeshott *et al*, 1987) were hybridised, under the same conditions as for pre-hybridisation, overnight at 65°C. Filters were washed in 0.5xSSC, 0.1% SDS at 65°C and autoradiographed on Kodak X-ray film.

2.2.1.2 Oligo-nucleotide Directed *in vitro* Mutagenesis.

Mutagenesis was according to the method of Eckstein (Taylor *et al*, 1985) using reagents supplied in the Amersham *in vitro* mutagenesis kit (code RPN.2322). Twenty eight base pair oligo-nucleotides carrying the required base substitutions were custom synthesised by Bresa (Table 2.1).

The Eckstein method involves annealing the mutagenic oligo-nucleotide to the single stranded template and extending with phosphorothioated nucleotides to give a phosphorothioated mutant DNA strand. The resultant hetero-duplex is then digested with *NciI*, which selectively nicks the non-phosphorothioated template strand. This is followed by exonuclease III digestion and re-extension to give a homo-duplex mutant molecule. All reactions were performed according to the manufacturers instructions.

In the present case, a cDNA clone of *Est6* (Oakeshott *et al*, 1987) was subcloned into the *EcoRI* site of the phage M13, single stranded DNA extracted from bacterial culture supernatants (Yanisch-Perron *et al*, 1985), and this used as the template for mutagenesis. Positives were screened by sequencing and once identified were sequenced in their entirety to check for unwanted mutations. A region of the cDNA containing the mutated active site region was then subcloned into a genomic clone which was isolated...
from Dm145, a wild type strain of D. melanogaster (Oakeshott et al, 1987). This was done by digesting the cDNA with NruI and NcoI (see Appendix I for a restriction map giving the position of restriction sites in clones of Est6), purifying the fragment then subcloning this into the same two restriction sites of the genomic clone. The cDNA and the genomic clone have identical sequences so the only amino acid differences between the conceptually translated mutant and wild-type proteins remain those introduced with the oligo-nucleotides.

Table 2.1 Oligo-nucleotides used for Mutagenesis

<table>
<thead>
<tr>
<th>wild type sequence</th>
<th>Est6^{H187Q}</th>
<th>Est6^{H187E}</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTGTTGTTGGTCACTCGCTGAGGA</td>
<td>CTGTTGTTGGTCAATCGCTGAGGA</td>
<td>CTGTTGTTGGTCAATCGCTGAGGA</td>
</tr>
</tbody>
</table>

The amino acid substitution introduced by the mutation is given in the left hand column and the oligo-nucleotide used to make each mutant is shown next to it. The wild type sequence is given for comparison. The nucleotide differences introduced using the above oligo-nucleotides are underlined. The terminology for the mutations is as follows; the amino acid being replaced is given first, followed by the number of the amino acid in the primary sequence (taking the first amino acid of the mature protein as one). The amino acid replacing the wild type residue is given after this. The single letter amino acid code is used. So for example, Est6^{H187Q} has glutamine(Q) replacing the histidine(H) found at amino acid number 187 in the wild type sequence.

2.2.2 P-ELEMENT MEDIATED GERM-LINE TRANSFORMATION

Proteins were expressed by transforming the mutant Est6 clones into the germ-line of D. melanogaster embryos. Using this expression system ensures the correct processing and post-translational modification of the proteins.

2.2.2.1 Fly Stocks and Culture Conditions

Fly stocks were maintained on cornmeal/treacle/yeast media in 375ml bottles at 25°C. The stocks used in this study are listed in Table 2.2. References describing these strains are included in the table.
Table 2.2 Fly Stocks Used in this Study

<table>
<thead>
<tr>
<th>STOCK</th>
<th>REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>p[ry;Δ(2-3),99B]</td>
<td>(Robertson et al, 1988)</td>
</tr>
<tr>
<td>TM3 Sb e ryAK/ry</td>
<td>506</td>
</tr>
<tr>
<td>w;p[ry;Δ(2-3),99B]</td>
<td></td>
</tr>
<tr>
<td>w;TM3 Sb e/TM6b Tb e</td>
<td></td>
</tr>
<tr>
<td>Est6° ry</td>
<td>(Sheehan et al, 1979)</td>
</tr>
<tr>
<td>Dm145</td>
<td></td>
</tr>
<tr>
<td>SM6a/Pm;TM6b,D3/Mo Sb</td>
<td>(Lindsley and Grell, 1968)</td>
</tr>
</tbody>
</table>

Lower case letters represent recessive mutations and upper case letters represent dominant markers. TM- represents a multiply inverted third chromosome carrying dominant and recessive genetic markers (balancer) and SM- a second chromosome balancer. Mo is also a balanced third chromosome.

2.2.2.2 Germ-line Transformation Procedure

A HindIII /Scal fragment which includes 1.1 kilobases (kbp) of DNA 5' of the coding region and 200 bp 3' DNA of the coding region of the Est6H187Q mutant was gel fractionated and purified and then end-filled (Maniatis et al, 1982, pp120) in preparation for "blunt-end" ligation into the Hpal site of the P-element transformation vector Carnegie 20. This vector utilises the rosy gene as a phenotypic marker for detecting positive transformants. The vector CaSpeR, which has the white gene as the phenotypic marker, was used for the Est6H187E mutant clone. In this case a Nhel fragment from the mutant clone, which has 860bp 5' and 200bp 3' of the Est6 coding region, was cloned into the Xbal site of this vector. Xbal and Nhel have compatible ends so this fragment could be cloned into CaSpeR directly. Supercoiled DNA from these constructs was then prepared through two sequential caesium chloride gradients and aliquots precipitated with ethanol (Rubin and Spradling, 1982) and stored at -20°C. Aliquots were resuspended in injection buffer (Rubin and Spradling, 1982), to a concentration of 500μg/ml, on the day of use.

Embryos were dechorionated manually and the DNA was introduced to embryos by micro-injection according to the procedures of Rubin and Spradling (1982). The transposase source, p[ry;Δ(2-3),99B], is a stably integrated copy of the P-element.
which constitutively expresses the transposase (Robertson et al., 1988). Embryos heterozygous for this modified P-element were injected with the Est6 constructs and transformant progeny from these individuals scored on the basis of the rescue of the appropriate eye colour mutation (Figure 2.1a and b). Once putative transformants were obtained they were mapped to a chromosome and the presence of an integrated copy of the Est6 construct in the genome was confirmed by the presence of unique restriction fragments detected by Southern analysis using the probe and procedures described in section 2.2.1. Transformants were crossed genetically into an Est6 null background, by replacement of the third chromosomes with those from the Est6 strain, and established as stocks.

2.2.3 PROTEIN PURIFICATION

The procedures used for the detection of EST6, estimation of amounts of total protein and purity of preparations are described in section 2.2.3.1. This is followed by the procedure used to purify EST6 from whole fly homogenates.

2.2.3.1 General Procedures For EST6 and Protein Detection

**End Point Assay for Detection of βNaphthylacetate Hydrolysis**

Enzyme sample was added to the wells of microtitre plates and made up to a volume of 100 μl with 0.1M sodium phosphate buffer pH 7 (PO₄ buffer). βNaphthylacetate was then added to a final concentration of 1 mM and the reaction allowed to proceed at 20°C for 20 to 50 minutes. The reaction was stopped with the addition of 0.7% SDS and 0.36% w/v fast blue (diazo blue tetrazolium salt TR). The colour reaction between the fast blue and the βnaphthol was left to develop for 30 minutes after which time the absorbance at 595 nm was read using an LKB microtitre plate reader (model 3550). The amount of βnaphthol produced was estimated by comparison with βnaphthol standards of known concentration.

**Protein Determination**

Protein concentrations were determined by the method of Bradford (1976), using reagent supplied by Biorad. Assays were performed in 96 well microtitre plates according to the manufacturers specifications. Absorbance was measured at 595 nm and
Figure 2.1 Crossing programs for germ line transformation. Δ2-3,99B is an abbreviation for p[Δ2-3,99B]. For Carnegie20 putative positives are scored on the basis of wild type eye colour. Transformants using CaSpeR may have eye colours varying from yellow through to wild type (red).
protein concentration estimated by comparison to BSA standards of known concentration.

Native and Denaturing Poly Acrylamide Gel Electrophoresis (PAGE)

For native PAGE, the gel and electrode buffers were as described by Vernick et al. (1988). Gels of 7% acrylamide (40:1 acrylamide:bis-acrylamide) were cast using the Biorad "Protean II" system and electrophoresed at 20 Watts for 3 hours. After electrophoresis the gels were soaked in PO₄ buffer for 30 minutes, then stained as follows. 10mg each of α and βnaphthylacetate were dissolved in 1 to 2 ml 50% acetone and made up to 100ml with PO₄ buffer containing 10mg Fast Garnet. The gel was soaked at room temperature in the stain until bands of esterase activity were visible. Gels were then fixed in methanol:acetic acid:water(5:1:5) and heat dried with suction.

The procedure for denaturing gel electrophoresis was as described by Laemmli (1970). Gels were 10% acrylamide unless otherwise stated. Protein bands were visualised by silver staining according to the method of Wray et al. (1981).

2.2.3.2 Purification Procedure

Preparation of Homogenates

Flies of both sexes and 2 to 10 days of age were frozen in liquid nitrogen. Approximately 12 grams were homogenised in 10 to 20 ml of buffer containing 20mM Imidazole, pH 6.9, 1mM EDTA, 1mM dithiothreitol (DTT) (imidazole buffer). The resulting slurry was filtered through four layers of gauze and a plug of glass wool. The homogenate was then clarified by centrifugation at 8,000xg for 30 minutes (Sorvall SS34 rotor) at 4°C.

Ammonium Sulphate Precipitation

Solid ammonium sulphate was added slowly to the homogenate until 55% saturation was reached. This was then kept on ice with stirring for a further 30 min. The precipitate that formed was removed from the homogenate by centrifugation for 30 min. at 8,000xg at 4°C. The supernatant was brought to 100% saturation by slow addition of solid ammonium sulphate and kept on ice with stirring for a further hour. The precipitate was pelleted by centrifugation as described above and the supernatant
discarded. The pellet was taken up in an equal volume of imidazole buffer and
undissolved material removed by filtration through a 0.22μm membrane. Salt was
removed from the sample by dialysing overnight against three changes of imidazole
buffer or by repeated concentration in a centriprep 30 (Amicon) followed by dilution
in imidazole buffer. The sample was then concentrated to a smaller volume (if
necessary) using a centriprep 30 or aquacide III (Calbiochem 17852).

DEAE Sepharose Anion Exchange

A Pharmacia column (XK26/40), packed with DEAE-FF sepharose to a volume
of 80ml, was equilibrated with several column volumes of imidazole buffer. Two to 4
ml of sample, prepared as above, was adsorbed onto the column and unbound protein
washed through with 60ml of imidazole buffer. A gradient of 0 to 700mM NaCl was used
to elute the bound protein. The flow rate was kept at 4ml/min and the gradient was
linear over a volume of 308ml. The Pharmacia FPLC system was used to control the
gradients and flow rates for this and all subsequent columns and the eluted protein was
monitored by UV absorption (280nm). Six ml fractions were collected and EST6
activity was detected by native gel electrophoresis and staining with naphthylacetate or
by detection of naphthylacetate hydrolysis by the end-point assay (section 2.2.2). EST6
eluted at 370-400mM NaCl. These fractions were pooled and concentrated using the
centriprep 30 or aquacide III.

Gel Filtration

The concentrated EST6-containing fractions were applied to a prepacked
Pharmacia "Superose 6" gel filtration column. The column was equilibrated with 2
column volumes (44ml) of 50mM Tris HCl pH 8.06, 1mM EDTA, 1mM DTT and 10%
glycerol (Tris buffer). The sample was then loaded and Tris buffer run through the
column at 0.5 ml/min. Fractions of 1ml were collected and assayed for EST6 activity.
EST6 eluted at 15-17ml (73% of a column volume). EST6 containing fractions were
pooled and used directly in the next step.

MonoQ Anion Exchange

A second anion exchange step was performed using the Pharmacia "mono Q" anion
exchange column (volume 6ml). The column was equilibrated with Tris buffer and 1 to
2ml of the pooled fractions from the gel filtration step were loaded onto the column. Unbound protein was eluted with 6ml of buffer. Bound protein was eluted with a linear gradient of 0 to 500mM NaCl over a 20ml volume (flow rate of 0.5ml/min.). Fractions containing protein peaks were collected and assayed for EST6 activity. EST6 eluted at 190 to 230 mM NaCl. EST6-containing fractions were pooled and analysed by denaturing SDS gel electrophoresis and silver staining (section 2.2.2). These samples appeared to be pure (section 2.3) and so were aliquoted and stored at -70deg C.

2.2.4 BIOCHEMICAL COMPARISON OF MUTANT AND WILD TYPE EST6

This section describes the biochemical analysis of EST6 and the two active site mutant enzymes. The procedures used for the assay of ester hydrolysis of various substrates, the effect of different environmental conditions and the determination of the sensitivity of these enzymes to organophosphate inhibitors are described.

2.2.4.1 Quantitation of EST6 Active Sites in the Purified Preparations

The organophosphate Dibrom can be used to titrate the active site of EST6 and so give an estimate of the number of active sites in an enzyme preparation (White et al, 1988). Dibrom was dissolved to 20mM in acetonitrile and diluted further to working concentrations with 0.1M sodium phosphate buffer pH7 (PO₄ buffer). Enzyme was diluted in PO₄ buffer and 20ul dispensed into the wells of microtitre plates. An equal volume of the required dilution of Dibrom was added and the enzyme and inhibitor preincubated at 28°C for 20 minutes. Bnaphthylacetate was then added to a final concentration of 1mM and the reaction allowed to proceed at 28°C for 20 to 50 minutes. Activity was plotted against Dibrom concentration and the concentration of active sites estimated by extrapolation of the line to the concentration of Dibrom at which there is no Bnaphthylacetate hydrolytic activity. This gives an estimate of the concentration of active sites in the purified preparations.

2.2.4.2 Sensitivity to Di-Isopropyl-Fluoro-Phosphate (DFP)

DFP (from Merck) was diluted to 0.4mM in PO₄ buffer. The enzyme was diluted in PO₄ buffer containing 100µg/ml BSA and 50ul aliquoted into the wells of microtitre plates. An equivalent volume of the required dilution of the 0.4mM DFP stock solution was added and the inhibitor and enzyme preincubated for 30 min. at 28°C. 10ul of
Bnaphthylacetate was then added to a final concentration of 1mM and the reaction allowed to proceed for 20 to 50 min. The reaction was stopped and the amount of Bnaphthol produced determined as described (section 2.2.3.1). As irreversible inhibition mimics the behaviour of non-competitive inhibition even though its mode of action is competitive in nature (Ainsworth, 1977), inhibition constants ($K_i$) could be calculated.

### 2.2.4.3 Esterase Assays For Kinetic Analysis

**General Assay Procedure**

Assays were performed in 1ml quartz cuvettes. Purified protein was diluted in PO$_4$ buffer containing 100ug/ml BSA and 0.9ml was added to the cuvette and placed in the spectrophotometer (LKB biochrom, Ultrospec II). The reaction was maintained at the required temperature by circulating water from a thermostatically controlled water bath through a water heated cuvette holder (LKB model 4072-088). A 10mM stock solution of each substrate was diluted in PO$_4$ buffer to the required concentration and 100ul added to the cuvette. The cuvettes were left to equilibrate to the required temperature in the cuvette holder for 20 minutes prior to addition of substrate. The change of absorbance was recorded on a chart recorder (Activon) and initial reaction velocity measured graphically from the trace. Reactions were performed at each of four substrate concentrations in triplicate for each substrate. All initial reaction velocities were corrected for autohydrolysis of the substrates by performing reactions without the addition of enzyme and subtracting the observed autohydrolysis rate from the rates of hydrolysis observed in the presence of the enzymes.

**Bnaphthyl Ester Assays**

The assay method used is a modification of that described for anaphthylacetate and anaphthylbutyrate by Mastropaolo and Yourno (1981), where the production of anaphthol was monitored by measuring the change of absorbance at 235 nm. In this study the absorbance spectra of Bnaphthol and Bnaphthylacetate were determined (Figure 2.2). Bnaphthol absorbs maximally at 230-235 nm, which is similar to that reported for anaphthol (Mastropaolo and Yourno, 1981). However, Bnaphthylacetate also absorbs substantially at these wavelengths, which is in contrast to
Figure 2.2 Absorbance spectra for Bnaphthol and Bnaphthylacetate.
anaphthylacetate. For this reason, it was necessary to determine the extinction coefficients of both the βnaphthyl ester substrate and βnaphthol and subtract the former from the latter in order to relate the change in absorbance to the moles of substrate hydrolysed. A further modification of the method is that a wavelength of 230 nm was used, as the difference in the absorbance of the substrate and product is greatest at this wavelength. The extinction coefficients of βnaphthol, βnaphthylacetate, βnaphthylpropionate and βnaphthylbutyrate at 230 nm were determined using 10 concentrations of each and these values used to calculate the number of moles of substrate hydrolysed. βnaphthylester substrates were dissolved in 50% acetone to a concentration of 10 mM. These stock solutions were then diluted to the required concentrations in PO4 buffer. Assays were performed as described above.

**pNitrophenylacetate Assay**

The extinction coefficient of p-nitrophenol at 400 nm was determined using the same buffer as in the enzyme assays, since pH affects the absorbance of this molecule. pNitrophenyl acetate (181 mg) was dissolved in 5 ml methanol and diluted to a concentration of 10 mM in PO4 buffer. The p-nitrophenylacetate was added slowly to the buffer so as to prevent precipitation. This stock solution was then diluted to the required concentration and 100 μl added to the cuvette. Reactions were performed as described in General Assay Procedure.

**Acetylthiocholine Iodide Assay**

The procedure is a modification of that by Ellman et al (1961). As the rates of hydrolysis being measured were very low, the assays were performed in 96 well microtitre plates. This allowed the simultaneous assay of many samples over a period of several hours. Acetylthiocholine iodide was dissolved to 50 mM in PO4 buffer and diluted in the same buffer to the required concentration. Enzyme was diluted in PO4 buffer containing 100 μg/ml BSA and 5:5-dithiobis-2-nitrobenzoate (DTNB) was added to 3 mM. 225 μl of this diluted enzyme was aliquoted into the microtitre plate wells and placed at the reaction temperature to equilibrate. 25 μl of substrate solution was then added and the reaction rate followed by reading absorbance at regular intervals. The change in absorbance was plotted against time. The initial velocity was measured over the linear range of these plots and converted to molar quantities using the extinction
coefficient given by Ellman et al. (1961) (1.36x10^4 Abs./Molar). Initial rates of hydrolysis were measured for each of five substrate concentrations and 2 to 5 replicates of each were performed. To test for the sensitivity of detected acetylthiocholine activity to inhibition by eserine a substrate concentration of 5x10^-4 was used. Eserine was added to the enzyme sample to a final concentration of 1x10^-5 M. Assays were then performed as described above.

2.2.4.4 Effect of Temperature and pH

βnaphthylacetate was used as the standard substrate for comparing the effect of environmental variables on the activity of the EST6 mutants. Temperature effects were determined by performing the assay as described for βnaphthylacetate at each of four temperatures. Temperatures below 28°C were attained by running water, maintained at the required temperature by a refrigerated water bath (LKB model 2219), through the cuvette holder of the spectrophotometer, while 28 and 35°C were attained using a heating water bath (LKB model 4072-088) in the same manner.

A pH profile for each of the enzymes was determined by measuring the initial velocity at a variety of pHs at a substrate concentration at least 20 fold greater than the estimated \( K_m \) of each of the enzymes for βnaphthylacetate. This then gives an estimate of \( V_{max} \) for each enzyme at the various pHs used. The buffers used are as follows: for pHs of 5 and 6, 0.1M NaAcetate buffer; for pHs 6, 7 and 8, 0.1M PO\(_4\) buffer; and for pHs 8, 9 and 10, 10mM Tris-HCl.

2.2.4.5 Analysis of Initial Reaction Rate Data

The data were fitted to the Michaelis-Menton equation (equation 2.1) and the maximal velocity and Michaelis constant and their associated standard errors estimated using the nonlinear least squares regression programme from the Systat statistics package. A sample analysis is shown in Appendix II. A tolerance limit of 0.000001 was used and consensus reached in less than fourteen rounds of iteration in every case, excepting the acetylthiocholine iodide data. For this substrate a tolerance of 0.01 and initial parameter estimates were necessary, due to the low rate of hydrolysis and consequent high variance in measurements observed for this substrate.
2.3 RESULTS

2.3.1 EXPRESSION OF THE EST6 ACTIVE SITE MUTANTS

Germline transformants for both the Est6H187Q and the Est6H187E constructs were isolated. A single transformed strain was chosen for each mutant and used as the source of protein in the analysis that follows. The transformant carrying the Est6H187Q mutant, the construct which is in the transformation vector Carnegie 20, segregates with the second chromosome and the transformant chosen for study carrying the Est6H187E mutant, the construct in the transformation vector CaSpeR, segregates with the X chromosome. The CaSpeR transformation vector has the advantage that the white marker gene in the vector is sensitive to position effects, which is reflected in the eye colour phenotype of transformants. The Est6H187E transformant chosen has wild-type expression of the white gene, suggesting a high level of expression, which may also be reflected in the levels of EST6H187E expression.

Once the transformants were crossed into an EST6 null background and stocks established, crude homogenates were electrophoresed on native PAGE and stained with naphthylacetate. The Est6H187Q and Est6H187E transformed strains both displayed levels of EST6 activity similar to the reference wild type strain Dm145, from which the original clone used in this study was isolated, indicating that the amino acid substitutions at position 187 have not destroyed the hydrolytic activity of EST6. As the mutant EST6 proteins retained hydrolytic activity, it was possible to purify this activity from whole flies and biochemically characterise the mutant proteins.

2.3.2 PURIFICATION OF WILD TYPE AND MUTANT EST6

EST6 was purified from the EST6H187Q, EST6H187E and Dm145 strains to enable detailed biochemical analysis and comparison of these proteins. A summary of the purification steps and their effectiveness is shown in Table 2.3. Examples of UV absorption chromatographs from the DEAE-sepharose and Superose 6 columns, and the MonoQ chromatographs for each of the purifications, are shown in Appendix III. A point to note in comparing the fold purifications for each of the enzymes in Table 2.3 is the lack of specificity of the esterase assay. Thus, the amount of activity present in the early steps of the purifications may over-estimate the true amount of EST6 activity present, as contaminating esterases may contribute to the total activity. For example, the apparent
decrease in specific activity of the EST6$^{H187E}$ sample after the first step in the procedure may well be due to the removal of a contaminating esterase rather than loss of EST6.

The degree of purity of the preparations was judged by denaturing PAGE followed by silver staining for protein, native PAGE followed by staining for naphthylacetate hydrolysis and the UV absorption chromatographs from the MonoQ anion exchange column. The silver stained gel (Fig.2.3b) shows that, in each case, the only visible band is approximately 60 kd, which is the expected molecular weight for EST6. The native gels (Figure 2.3a) show only one band of naphthylacetate hydrolytic activity, giving further evidence that this protein is EST6 and that the preparations are not contaminated with other esterases. The last line of evidence that these preparations are close to homogeneous is the chromatographs from the last purification step, which show a single UV absorbing peak corresponding to the fractions with EST6 activity (Appendix III). These three lines of evidence indicate that the preparations essentially contain only the EST6 protein.

It is apparent from Figure 2.3 that EST6$^{H187E}$ has a faster electrophoretic mobility in comparison to Dm145 and EST6$^{H187Q}$. This is expected as EST6$^{H187E}$ has an additional negative charge due to the substitution of the neutral His for the negatively charged Glu. In contrast, Gln and His are both neutral at pH 8.8, the pH of the native gel buffer (Vernick et al, 1988), therefore the wild type and EST6$^{H187Q}$ enzymes have the same net charge and so the same electrophoretic mobility.

Each of the EST6 proteins behaved similarly throughout each of the purification steps. However, the strains did differ in the amounts of protein isolated from 12g of flies. EST6$^{H187Q}$ and Dm145 yielded similar amounts of protein and were deemed pure after 200 to 400 fold purification had been achieved. The EST6$^{H187E}$ mutant yielded much larger quantities of protein and was only purified 40 fold. This indicates relatively high levels of EST6 protein, which presumably results from high levels of Est6 expression in this strain. This is consistent with the high level of white expression of this strain, apparent from its wild type eye colour phenotype, and is reflected in the size of EST6 protein peaks relative to other protein peaks eluted from the columns (Appendix III).
<table>
<thead>
<tr>
<th>Step</th>
<th>Protein conc.</th>
<th>Esterase activity</th>
<th>Specific activity</th>
<th>% Recovery</th>
<th>Purification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mg/ml</td>
<td>umol/min</td>
<td>umol/min/mg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dm145</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>crude</td>
<td>nd</td>
<td>nd</td>
<td>0.027*</td>
<td>nd</td>
<td>X1.1</td>
</tr>
<tr>
<td>ammonium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sulphate fraction</td>
<td>6.2</td>
<td>0.24</td>
<td>0.029</td>
<td>nd</td>
<td>X1.1</td>
</tr>
<tr>
<td>DEAE Sepharose</td>
<td>1.1</td>
<td>0.132</td>
<td>0.24</td>
<td>55</td>
<td>X9</td>
</tr>
<tr>
<td>Superose6</td>
<td>0.045</td>
<td>0.136</td>
<td>1.51</td>
<td>55</td>
<td>X56</td>
</tr>
<tr>
<td>Mono Q</td>
<td>0.005</td>
<td>0.102</td>
<td>10.2</td>
<td>43</td>
<td>X378</td>
</tr>
<tr>
<td>EST6H187Q</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>crude</td>
<td>7.51</td>
<td>1.73</td>
<td>0.019</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ammonium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sulphate fraction</td>
<td>6.9</td>
<td>0.56</td>
<td>0.051</td>
<td>32</td>
<td>X2.7</td>
</tr>
<tr>
<td>DEAE Sepharose</td>
<td>0.68</td>
<td>0.26</td>
<td>0.96</td>
<td>15</td>
<td>X50</td>
</tr>
<tr>
<td>Superose6</td>
<td>0.02</td>
<td>0.084</td>
<td>2.1</td>
<td>5</td>
<td>X110</td>
</tr>
<tr>
<td>Mono Q</td>
<td>0.012</td>
<td>0.044</td>
<td>3.7</td>
<td>3</td>
<td>X195</td>
</tr>
<tr>
<td>EST6H187E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>crude</td>
<td>6.5</td>
<td>3.1</td>
<td>0.025</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ammonium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sulphate fraction</td>
<td>7.0</td>
<td>0.72</td>
<td>0.023</td>
<td>23</td>
<td>X0.92</td>
</tr>
<tr>
<td>DEAE Sepharose</td>
<td>1.66</td>
<td>0.196</td>
<td>0.32</td>
<td>6</td>
<td>X13</td>
</tr>
<tr>
<td>Superose 6</td>
<td>0.144</td>
<td>0.167</td>
<td>0.77</td>
<td>5</td>
<td>X31</td>
</tr>
<tr>
<td>Mono Q</td>
<td>0.060</td>
<td>0.12</td>
<td>1.0</td>
<td>4</td>
<td>X40</td>
</tr>
</tbody>
</table>

Table 2.3 Summary of the Purification of EST6

Esterase activity is for βnaphthylacetate hydrolysis.

*The total amount of esterase activity and protein present in the crude homogenate used as starting material for Dm145 were not determined (nd) so the average specific activity of several homogenates from the same strain is given here for reference and the fold purification calculated from this value. The percentage recovery for this protein is given as compared to the total activity in the ammonium sulphate fraction, not the crude homogenate.
Figure 2.3 a) Native PAGE (7% polyacrylamide) of the purified wild-type and mutant EST6 protein preparations used in the kinetic analysis. H187Q = ESTH187Q, H187E = EST6H187E and Dm145 = wild type EST6 from the strain Dm145. The gel has been stained with a and βnaphthylacetate/fast garnet as described in section 2.2.3.1.

b) SDS denaturing PAGE (10% polyacrylamide) of the purified wild-type and mutant EST6 protein preparations used in the kinetic analysis. The position of the 66 kilodalton molecular weight marker is shown. The gel was silver stained as described in section 2.2.3.1.
2.3.3 BIOCHEMICAL ANALYSIS

2.3.3.1 The Effect of pH

Enzyme activity assays were performed over the pH range 5 to 10 using a combination of three buffers. There was no significant effect of buffer on activity for a particular pH so the data for both buffers at each of pH 6 and pH 8 were pooled and used for a plot of log velocity versus pH (Fig 2.4). Wild type EST6 and EST6H187Q both have low activity at pH 5 which increases dramatically at pH 6. Both enzymes display an optimum pH of 7.0. The EST6H187E mutant has low activity at pH 5 which also increases at more alkali pH, but in this case the increase is more gradual and it is not until the pH is 9 that this enzyme displays maximal activity.

The pKₐ of particular groups involved in catalysis can be calculated from the pH profile by extrapolating the linear regions of the upward slope, and a horizontal line representing maximal velocity, to a point where they intersect. The pH at this point represents an estimate of the pKₐ of an ionisable group in the protein which needs to be deprotonated for catalysis of the hydrolytic reaction to occur (Alberty and Massey, 1954). For both EST6H187Q and the wild type enzyme the pKₐ of such a group is 6. For EST6H187E there is a sharp increase in activity at pH 6 indicating the deprotonation of a group with a pKₐ of 6 is still required. However, the increase in activity above pH 6 is more gradual, suggesting activity is dependent on solvent hydroxyl ion concentration (Craik et al., 1987).

2.3.3.2 Inhibition by DFP

DFP is an organophosphate which inhibits serine hydrolases by binding covalently to the reactive serine, and so exhibits irreversible inhibition.

Plots of the reciprocal of velocity versus DFP concentration are shown in Figure 2.5. Irreversible inhibitors, like DFP, mimic non-competitive inhibitors as they prevent the binding of substrate. The inhibition constant (Kᵢ) of DFP for each of the enzymes can be estimated from these plots in the same manner as Kᵢ for non-competitive inhibitors, i.e. the x-intercept is equivalent to -Kᵢ (Dixon and Webb, 1964). Values for Kᵢ so calculated are 45±7μM for the wild type, 38±6μM for EST6H187Q and 8±1μM for EST6H187E. From these values it is apparent that EST6H187E is significantly more sensitive to DFP than the other two enzymes, having a five fold lower inhibition constant.
Fraction of Maximal Activity versus pH

Figure 2.4 The fraction of maximal activity as a function of pH. Standard errors are shown by bars.
Figure 2.5  The reciprocal of velocity (umol Bnaphthol produced per minute per mg protein) plotted as a function of DFP concentration. The correlation coefficients (R) are shown.
2.3.3.3 Active Site Quantitation

The number of active sites present in each of the purified preparations was estimated by titration with the organophosphate inhibitor Dibrom (Altech associates). Since EST6 has a single catalytic site (Mane et al, 1983) and Dibrom is an irreversible inhibitor that reacts stoichiometrically with the enzyme (Fest and Schmidt, 1973), the number of active sites can be titrated with the inhibitor, if it is assumed that all molecules of the inhibitor present have bound to the enzyme. This criterion is met for inhibitors with Ki values of the order of 10⁻⁹ M (Ainsworth, 1977; Dixon and Webb, 1964, pp333). Dibrom fits this criterion for EST6 (White et al, 1988).

The plots of remaining activity against Dibrom concentration for EST6, EST6H¹87Q and EST6H¹87E are linear and the corresponding correlation coefficients are high and significant (Figure 2.6). The concentration of active sites, estimated as described above, was used to convert the estimates of maximal velocity for each of the enzymes to the fundamental rate constant, k_cat, from the equation \( V_{\text{max}} = k_{\text{cat}} \cdot E_0 \) (section 2.1.2). As EST6 has one active site (Mane et al, 1983), the apparent molecular weight for each of the esterases can be calculated from the amount of protein present and the concentration of active sites estimated from the titration, if it is assumed that all the protein present in the preparations is active. The values calculated are 62±6kd for the wild type, 66±8kd for EST6H¹87Q and 47±9kd for EST6H¹87E. The major source of error in these calculations is the protein estimations, as the concentration of protein in some samples was at the limits of detection for the assay. Despite this limitation, these estimates are clearly not significantly different from each other, or from the molecular weight of 59.7kd calculated for EST6 from the deduced amino acid sequence (Oakeshott et al, 1987). This confirms the previous finding that EST6 has a single active site (Mane et al, 1983b; White et al, 1988) and demonstrates this also to be true of EST6H¹87Q and EST6H¹87E.
3.3.3.4 Substrate Specificity

The kinds of substrate recognized by each of the enzymes against a variety of substrates are shown in Table 3.4. In general, the wild-type enzyme Dm is higher toward higher K_m values than closer ESTH145, which in turn is higher than ESTH187E, and this is mirrored in the specificity constant for kcat/Km. In terms of substrate preference, however, this shows a parallel pattern to the specificity constant for kcat/Km. From the spectrophotometer analyses, it is observed that for derivative H187Q, the substrate preference is for short chain-length acrylates and naphtalenes, being the least preferred naphtalene substrate of substrate and product equally preferred. A strong observed trend is that the specificity constant for Dm145 and ESTH187Q. This results in the loss of specificity of the substrate for these substrates as well as the enzyme.

Figure 2.6  Initial velocity (V) plotted as a function of substrate concentration.

Correlation coefficients (R) are given for each line and error bars shown.
2.3.3.4 Substrate Specificity

The kinetic parameters for each of the enzymes against a variety of substrates are shown in Table 2.4. In general, the wild type enzyme has a higher turnover number (kcat) than does EST6H187Q, which in turn is higher than EST6H187E, and the same is true for Km. In terms of substrate preference however, the most meaningful parameter is the specificity constant, kcat/Km (section 2.1.2, eq 2.8). From the specificity constants it is clear that each of the enzymes has a preference for short chain acyl groups with naphthylbutyrate being the least preferred naphthylester substrate and acetate and propionate being equally preferred. A striking observation is the similarity of the specificity constants for Dm145 and EST6H187Q. There are no significant differences between the values of kcat/Km for any of the substrates that these two enzymes hydrolyse. In contrast EST6H187E has a significantly lower specificity constant for each of these substrates than have the other enzymes.

Another interesting observation is that EST6H187E is capable of hydrolysing the cationic substrate acetylthiocholine iodide, which is not hydrolysed at detectable levels by either the wild-type or EST6H187Q enzymes. The rate of hydrolysis for EST6H187E is quite low but the Km is in the mM range. There are several lines of evidence which indicate that this activity is not attributable to contaminating acetylcholinesterase in the EST6H187E enzyme preparation. Firstly, this acetylthiocholine hydrolytic activity is not sensitive to eserine, an acetylcholinesterase inhibitor, since the initial velocity at a substrate concentration of 5X10^-4M in the presence and absence of eserine is not significantly different, being 9.6±0.2 and 9.7±0.5 nmol/min/mg respectively. Secondly, the wild type and EST6H187Q enzyme preparations, which were prepared by the same procedure, do not display any activity for this substrate. Thirdly, the Km of the measured acetylthiocholine hydrolytic activity (2.0±1.5X10^-3M) is significantly higher than the value of 79±3X10^-6M observed for acetylcholinesterase from D.melanogaster (Gnagey et al, 1987), which would not be expected if this activity was due to low level contamination by this enzyme.

Each of the enzyme preparations were also tested for juvenile hormone (JH) hydrolytic activity. The assays were kindly performed by P.Campbell using the method of Hammock and Roe (1985). None of the enzymes displayed any detectable JH hydrolytic activity.
<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Substrate</th>
<th>Parameter</th>
<th>Bnaphthyl acetate</th>
<th>Bnaphthyl propionate</th>
<th>Bnaphthyl butyrate</th>
<th>p-nitrophenyl acetate</th>
<th>acetyl-thiocholine iodide</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$V_{\text{max}}$</td>
<td>21.8(1.1)</td>
<td>44.2(14.1)</td>
<td>9.5(1.4)</td>
<td>16.4(0.6)</td>
<td>no activity</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$K_m$ (uM)</td>
<td>2.09(0.24)</td>
<td>4.7(2.6)</td>
<td>2.61(0.83)</td>
<td>2.25(0.26)</td>
<td>_</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$k_{\text{cat}}$ (sec$^{-1}$)</td>
<td>22.5(1.1)</td>
<td>46(15)</td>
<td>9.8(1.4)</td>
<td>16.9(0.6)</td>
<td>_</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$k_{\text{cat}}/K_m$ (X10$^{-6}$)</td>
<td>10.8(0.8)</td>
<td>9.7(2.3)</td>
<td>3.8(1.1)</td>
<td>7.5(0.6)</td>
<td>_</td>
<td></td>
</tr>
<tr>
<td>Dm145</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EST6$^{H187Q}$</td>
<td>$V_{\text{max}}$</td>
<td>13.3(0.9)</td>
<td>41.1(10.2)</td>
<td>5.70(0.74)</td>
<td>8.53(1.28)</td>
<td>no activity</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$K_m$ (uM)</td>
<td>1.35(0.29)</td>
<td>5.6(2.2)</td>
<td>1.70(0.58)</td>
<td>1.28(0.24)</td>
<td>_</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$k_{\text{cat}}$ (sec$^{-1}$)</td>
<td>14.6(1.1)</td>
<td>45(11)</td>
<td>6.3(0.8)</td>
<td>9.4(1.4)</td>
<td>_</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$k_{\text{cat}}/K_m$ (X10$^{-6}$)</td>
<td>10.8(1.5)</td>
<td>8.1(1.3)</td>
<td>3.69(1.3)</td>
<td>7.3(0.5)</td>
<td>_</td>
<td></td>
</tr>
<tr>
<td>EST6$^{H187E}$</td>
<td>$V_{\text{max}}$</td>
<td>3.9(0.4)</td>
<td>11.4(1.6)</td>
<td>1.17(0.13)</td>
<td>2.97(0.14)</td>
<td>0.033(0.022)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$K_m$ (uM)</td>
<td>0.55(0.26)</td>
<td>1.79(0.65)</td>
<td>0.97(0.39)</td>
<td>1.64(0.28)</td>
<td>2.0(1.5)$\times$10$^3$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$k_{\text{cat}}$ (sec$^{-1}$)</td>
<td>3.0(0.3)</td>
<td>8.8(1.2)</td>
<td>0.9(0.1)</td>
<td>2.3(0.1)</td>
<td>0.025(0.017)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$k_{\text{cat}}/K_m$ (X10$^{-6}$)</td>
<td>5.4(2.0)</td>
<td>4.9(1.2)</td>
<td>0.93(0.25)</td>
<td>1.4(0.2)</td>
<td>1.3(0.2)$\times$10$^{-5}$</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.4 Kinetic Parameters for a Variety of Substrates

$V_{\text{max}}$ is expressed in umol hydrolysed/minute/mg protein. These were then converted to $k_{\text{cat}}$ values using the data from the Dibrom quantitation of sites. Associated standard errors are given in brackets.
The degree of similarity between the active sites of enzymes can be quantified through use of linear free energy relationships as proposed by Withers and Rupitz (1990). This procedure involves plotting the logarithms of the specificity constants, obtained experimentally, for several substrates for a pair of related enzymes. The correlation coefficient for the line so derived is a direct measure of the similarity of the transition state complexes of the two enzymes. Graphs comparing the specificity constants of the two active site mutants to those of the wild-type enzyme are shown in Figure 2.7. The graph for EST6\textsuperscript{H187Q} versus Dm145 EST6 has a correlation coefficient and a slope close to 1, indicating a high degree of similarity between the transition states, and so the active sites, of these enzymes. In contrast the graph for the EST6\textsuperscript{H187E} mutant has a lower correlation coefficient (0.9) and a significantly different slope (1.68), indicating a difference in the active site structures of these enzymes.

As there are significant differences between the wild-type EST6 and EST6\textsuperscript{H187E} in their active sites, the change in free energy between the two can be calculated from the equation

\[
\Delta G_i = RT \ln \left( \frac{k_{cat}/K_m}{(k_{cat}/K_m)_{wt}} \right)
\]

where \( R \) is the gas constant and \( T \) is the temperature of the reaction in degrees Kelvin. This equation was originally formulated for the variation of a series of substrates reacting with the same enzyme but can be applied to mutated enzymes reacting with the same substrate providing that (1) the mutation is to an amino acid that is isosteric or smaller than that which it replaces and (2) the amino acid does not take part directly in catalysis (Fersht and Leatherbarrow, 1987). The first prerequisite is fulfilled since Gln and Glu have side chains of similar size to His. Given there are no significant differences between the specificity constants of Dm145 and EST6\textsuperscript{H187Q}, it can be safely assumed that the second prerequisite is also fulfilled. The differences observed for the EST6\textsuperscript{H187E} mutant enable calculation of \( \Delta G_i \) and these are shown in Table 2.5. There is a significant difference in the \( \Delta G_i \) calculated for \( p \)-nitrophenylacetate in comparison to \( \beta \)-naphthylacetate and propionate but not butyrate. However, there are no significant differences between the \( \Delta G_i \) values calculated for the three \( \beta \)-naphthyl ester substrates. The \( \Delta G_i \) value for \( \beta \)-naphthylbutyrate is substantially larger than for the other \( \beta \)-naphthyl esters but the large standard error means that this is not significant.
Figure 2.7  Natural log. of the specificity constants (kcat/Km) for H187Q and H187E plotted against natural log. (kcat/Km) for Dm145. Data are for the Bnapthyl esters and pnitrophenylacetate.
Table 2.5 Change in Interaction Energy For EST6\(^{H187E}\) in comparison to the wild type.

The energy changes are expressed in kilojoules/mol and are calculated for the temperature at which the assays were performed (28\(^\circ\)C). Standard errors are given in parentheses.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>(\beta)naphthyl acetate</th>
<th>(\beta)naphthyl propionate</th>
<th>(\beta)naphthyl butyrate</th>
<th>(p)nitrophenyl acetate</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Delta G_i) for EST6(^{H187E})</td>
<td>-1.74(0.75)</td>
<td>-1.69(0.25)</td>
<td>-3.57(1.56)</td>
<td>-4.16(0.18)</td>
</tr>
</tbody>
</table>

In summary, the EST6\(^{H187E}\) mutant protein differs significantly in substrate specificity from the wild type EST6 protein; most notably the gain of acetylthiocholine hydrolytic activity and an interaction energy significantly lower than the wild type for \(p\)nitrophenylacetate and \(\beta\)naphthyl esters. EST6\(^{H187Q}\), on the other hand, has very similar characteristics to the wild type enzyme in terms of specificity constants but does show a slight but significant decrease in \(k_{cat}\) and \(K_m\) for the majority of the substrates tested.

2.3.3.5 The Effect of Temperature

The kinetic constants calculated for the enzymes at each of four temperatures are shown in Table 2.6. There is a significant effect of temperature on \(V_{max}\) and \(k_{cat}\) for each of the enzymes; there is a monotonic increase in these parameters with increasing temperature for each of the enzymes (Figure 2.8a).

As Figure 2.8b shows, the \(K_m\) estimates display more complex and varied relationships with temperature. There is no significant effect of temperature on the \(K_m\) of Dm145. However, the \(K_m\) estimate at 35\(^\circ\)C for EST6\(^{H187Q}\) is significantly higher than at 20\(^\circ\)C. In contrast, the \(K_m\) estimates for EST6\(^{H187E}\) show a significant decrease with increasing temperature, between 10 and 20\(^\circ\)C and between 20 and 28\(^\circ\)C.

The specificity constants of each enzyme (Table 2.6) are not significantly different between 20\(^\circ\)C and 28\(^\circ\)C. There is a significant decrease in this ratio at 10\(^\circ\)C in
<table>
<thead>
<tr>
<th>Temp.</th>
<th>10°C</th>
<th>20°C</th>
<th>28°C</th>
<th>35°C</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dm145</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$V_{\text{max}}$</td>
<td>10.9(0.8)</td>
<td>17.7(1.0)</td>
<td>21.8(1.1)</td>
<td>25.7(1.5)</td>
</tr>
<tr>
<td>$K_m(\mu M)$</td>
<td>1.5(0.3)</td>
<td>1.5(0.2)</td>
<td>2.09(0.24)</td>
<td>1.4(0.2)</td>
</tr>
<tr>
<td>$k_{\text{cat}}(s^{-1})$</td>
<td>11.2(0.8)</td>
<td>18.2(1.0)</td>
<td>22.5(1.1)</td>
<td>26.5(1.5)</td>
</tr>
<tr>
<td>$k_{\text{cat}}/K_m$</td>
<td>7.6(1.0)</td>
<td>12.2(1.7)</td>
<td>10.8(0.8)</td>
<td>19.3(2.3)</td>
</tr>
<tr>
<td><strong>EST6H187Q</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$V_{\text{max}}$</td>
<td>6.61(0.45)</td>
<td>12.1(0.8)</td>
<td>13.3(0.9)</td>
<td>21.4(1.5)</td>
</tr>
<tr>
<td>$K_m(\mu M)$</td>
<td>1.1(0.3)</td>
<td>1.1(0.3)</td>
<td>1.4(0.3)</td>
<td>2.2(0.4)</td>
</tr>
<tr>
<td>$k_{\text{cat}}(s^{-1})$</td>
<td>7.29(0.50)</td>
<td>13.4(0.9)</td>
<td>14.0(1.1)</td>
<td>23.5(1.6)</td>
</tr>
<tr>
<td>$k_{\text{cat}}/K_m$</td>
<td>6.8(1.1)</td>
<td>11.9(1.8)</td>
<td>10.8(1.5)</td>
<td>10.7(1.0)</td>
</tr>
<tr>
<td><strong>EST6H187E</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$V_{\text{max}}$</td>
<td>2.04(0.20)</td>
<td>3.53(0.17)</td>
<td>3.9(0.4)</td>
<td>5.9(0.2)</td>
</tr>
<tr>
<td>$K_m(\mu M)$</td>
<td>1.8(0.4)</td>
<td>0.90(0.16)</td>
<td>0.55(0.26)</td>
<td>0.65(0.11)</td>
</tr>
<tr>
<td>$k_{\text{cat}}(s^{-1})$</td>
<td>1.57(0.15)</td>
<td>2.72(0.13)</td>
<td>3.0(0.3)</td>
<td>4.5(0.2)</td>
</tr>
<tr>
<td>$k_{\text{cat}}/K_m$</td>
<td>0.90(0.14)</td>
<td>3.0(0.4)</td>
<td>5.5(2.0)</td>
<td>7.0(0.9)</td>
</tr>
</tbody>
</table>

Table 2.6 Kinetic parameters for each of the enzymes at the four different assay temperatures. Standard errors are given in parentheses.
Figure 2.8 a) and b) The natural log of kcat and Km as functions of the reciprocal of absolute temperature. The correlation coefficients (R) for the kcat lines are shown.
each case but only for the wild type enzyme does the specificity constant differ significantly at 35°C, being almost double that observed at 28°C.

The enthalpy (ΔH) of activation for the reaction can be calculated from the Arrhenius plot, which is of the form \( \log_{10}(k_{\text{cat}}) \) vs 1/T (Dixon and Webb, 1964, pp150-156) (Fig 2.8a). The slope of the straight line so generated is then \(-E_a/2.303R\) (R=8.314 J/deg/mol), where \(E_a\) is the energy of activation and is equal to \(ΔH+RT\). The Gibb's free energy (ΔG) of activation can be calculated from the equation

\[
\Delta G = RT \log(T. k_b/k_{\text{cat}}. h)
\]  

(2.10)

where \(h\) is Planck's constant and \(k_b\) is Boltzmann's constant and T is the temperature in degrees Kelvin. The entropy (ΔS) of activation can then be calculated from the ordinary thermodynamic relationship

\[
\Delta S = (\Delta H - \Delta G)/T
\]  

(2.11)

These values are shown in Table 2.7. There are no significant differences in the enthalpy, entropy or energy of activation among the three enzymes. EST6^H187E however does have a significantly higher Gibb's free energy of activation in comparison to the other enzymes.

A similar analysis can be performed for the \(K_m\) estimates from Table 2.6 (Dixon and Webb, 1964 pp150-156). If it is assumed that acylation is the rate limiting step (section 2.1.2) then \(K_m\) is a close approximation to \(K_s\) and the energy of substrate binding can be determined from this using the equation

\[
\Delta G^0 = -RT \ln(1/K_m)
\]  

(2.12)

which gives the Gibb's free energy change for substrate binding.

The enthalpies involved with the significant differences in \(K_m\) between temperatures were calculated using the equation

\[
\Delta H^0 = RT \frac{\ln(1/\Delta K_m)}{\Delta T}
\]  

(2.13)

which gives the enthalpy of the process of substrate binding. The entropy can then be calculated using equation 2.11 above. These values are shown in Table 2.8.

The Gibb's free energy of substrate binding for each of the enzymes at 28°C do not differ significantly. However, the enthalpy and entropy of the process as calculated from the \(K_m\) estimates differ substantially for the mutant enzymes, with \(\Delta H^0\) for EST6^H187E being endothermic (negative) below 28°C and exothermic above this temperature, while the process appears to be exothermic for EST6^H187Q between 20-35°C. The same pattern is observed for the entropy of the process.
Table 2.7 Thermodynamic Quantities—Activation Energies Calculated from $k_{cat}$ Estimates

The quantities above are expressed in kilojoules/mol. (KJ/mol). The parameters are the energy measurements at 28°C (301K) as estimated from equations 2.10, 2.11 and the slope of the Arrhenius plot. The values of $k_{cat}$ used in estimating $\Delta G$ were calculated from the Arrhenius plots (Fig 2.8a). Large standard errors associated with the enthalpy of activation are because these are derived from the slope of a line (the Arrhenius plot, Fig. 2.8a) based on four points. Any non-linearity in this plot will give a larger error in the final parameter estimate. The calculation of the Gibb's free energy is estimated from a single measurement, by applying equation 2.10; it is therefore more accurate, but only a good estimate for the temperature in question.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>$E_a$ (KJ/mol)</th>
<th>$\Delta H$</th>
<th>$\Delta G$</th>
<th>$\Delta S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dm145</td>
<td>25.41(2.87)</td>
<td>22.92(2.87)</td>
<td>65.99(0.27)</td>
<td>-0.143(0.009)</td>
</tr>
<tr>
<td>EST6H187Q</td>
<td>32.23(3.73)</td>
<td>29.72(3.73)</td>
<td>66.69(0.23)</td>
<td>-0.123(0.012)</td>
</tr>
<tr>
<td>EST6H187E</td>
<td>29.12(3.37)</td>
<td>26.62(3.37)</td>
<td>70.26(0.27)</td>
<td>-0.145(0.010)</td>
</tr>
</tbody>
</table>

Table 2.8 Process Energies Involved With Substrate Binding (Assuming $K_m = K_s$)

Quantities are expressed in kilojoules/mol and standard errors are given in parentheses. No enthalpy and entropy values were calculated for Dm145 as temperature does not significantly effect the $K_m$ of this enzyme (see Table 2.7). The enthalpy and entropy values shown for the mutant enzymes were calculated from the $K_m$ differences observed over the temperature range indicated. Temperature ($T$) was taken as 28°C (301K) and the values calculated from equations 2.11, 2.12 and 2.13.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Temp. range</th>
<th>$\Delta H^0$</th>
<th>$\Delta G^0$</th>
<th>$\Delta S^0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dm145</td>
<td></td>
<td></td>
<td>-32.7(0.3)</td>
<td></td>
</tr>
<tr>
<td>EST6H187Q</td>
<td>20-35°C</td>
<td>689.0(4.6)</td>
<td>-33.7(0.5)</td>
<td>2.40(0.01)</td>
</tr>
<tr>
<td>EST6H187E</td>
<td>10-28°C</td>
<td>-580.3(4.7)</td>
<td>-36.1(1.2)</td>
<td>-1.81(0.01)</td>
</tr>
<tr>
<td></td>
<td>28-35°C</td>
<td>1,734(24)</td>
<td>-36.1(1.2)</td>
<td>5.88(0.08)</td>
</tr>
</tbody>
</table>
2.4 Discussion

The results of the biochemical analysis of the wild type and mutant EST6 enzymes indicate that substituting Glu for the His residue adjacent to the reactive Ser of EST6 has drastic effects on the catalytic properties of the enzyme. Virtually every biochemical property determined reveals a significant difference between the EST6\(^{H187E}\) mutant and the wild type. These enzymes display differences in pH optimum, DFP sensitivity, substrate specificity and Gibb's free energy of activation. The significance of these results is enhanced by the observation that Gln substituted at the same position has very little effect on the enzyme's properties. This enables a comparative analysis of the differences between the enzymes in terms of molecular interactions since the Gln substitution, having little effect on activity, effectively rules out a number of explanations for the observed differences between the EST6\(^{H187E}\) and the wild type enzyme.

2.4.1 The Effect of pH

Each of the enzymes displays low activity at pH 5 but significantly greater activity at pH 6. Increased velocity due to the deionisation of a group with a pK\(_a\) of 6-6.5 has been observed previously for EST6 (White et al., 1987) and suggests the involvement of a His residue in the catalytic mechanism (section 1.4.1). This aspect of the pH profile is conserved in each of the active site mutants. However, the optimal pHs of the enzymes differ, with EST6 and EST6\(^{H187Q}\) displaying maximal activity at pH 7 while EST6\(^{H187E}\) reaches its maximal activity at pH 9. A large increase in the rate of hydrolysis by the EST6\(^{H187E}\) mutant is observed when the pH is raised from 7 to 9, while both the wild type and EST6\(^{H187Q}\) enzymes show no significant difference in activity over this pH range. This suggests a fundamental alteration in an aspect of catalysis of the EST6\(^{H187E}\) mutant.

As the substrate concentration used in this experiment is 20 fold greater than the K\(_m\) of the enzymes, the measured rate of hydrolysis can be assumed to be dependent on changes in V\(_{\text{max}}\) and not K\(_m\). Therefore, any explanation of the altered pH dependence must relate to alterations in the rate constants k\(_2\) and k\(_3\) rather than substrate affinity.

The EST6\(^{H187Q}\) substitution acts as a control in this analysis due to the similarity of the pH profiles of this mutant and the wild type enzymes. This indicates that His 187 is not directly involved in catalysis, which leaves the possibility that the
introduction of the ionic side chain of Glu has led to an alteration of the catalytic mechanism at alkaline pH.

One broad possibility is an indirect, conformational effect of the His to Glu substitution at position 187. An alteration in tertiary structure that leads to less favourable binding of the transition state would lead to a slower rate of hydrolysis. A pH dependent conformational change, brought about by ionisation of a group elsewhere in the protein, may then lead to more favourable binding of the transition state, thus increasing the rate of hydrolysis at alkaline pH.

Another explanation is that the catalytic mechanism is altered at high pH, such that solvent hydroxyl ions are being utilised in some manner to facilitate hydrolysis of the substrate by EST6H187E. The direct involvement of solvent hydroxyl ions is indicated by the proportional increase in activity with increasing pH between pH 6 and pH 9. Whether the hydroxyl ions interact with the reactive Ser or His (section 1.4.1), or facilitate the reaction in some other manner, is not clear.

Evidence for a possible direct effect that the His to Glu substitution may have on the catalytic mechanism of EST6 comes from an analysis of the serine proteases. It is well established that the serine proteases use a charge relay involving three residues, Asp, His and Ser, in their catalytic mechanism (section 1.2.1). The His and Asp residues serve to shuttle a proton from the reactive Ser and so increase its nucleophilicity. The serine esterases, by analogy, may utilise a similar system. The anomalous pH dependence of the EST6H187E mutant then may indicate a perturbation of the charge relay system such that the nucleophilicity of the reactive serine is compromised at low pH. Electronic interaction between the carboxyl group of the Glu and the reactive Ser and/or the catalytic His may conceivably perturb the required interaction between these catalytic residues and so decrease the nucleophilicity of the reactive Ser.

A similar pH dependence to EST6H187E has been observed for mutant forms of trypsin (Craik et al, 1987) in which the catalytic mechanism was perturbed by the substitution of catalytic Asp with an Asn residue (section 1.5.2). The reduced activity of this trypsin mutant was due to reduced nucleophilicity of the reactive Ser, as evidenced by its relative insensitivity to reactive Ser-specific inhibitors (Craik et al, 1987). If a charge relay in the EST6H187E mutant was perturbed, this enzyme would be expected to be less sensitive to organophosphate inhibitors which are specific to the reactive Ser. Inhibition by the organophosphate DFP shows the opposite to be true (section 2.3.3.2). EST6H187E is in fact more sensitive to this chemical than either the
wild type or \textit{EST6}^{H187Q} enzymes. This indicates that the nucleophilicity of the reactive Ser is not reduced by this mutation, so perturbation of a charge relay system cannot explain the anomalous pH dependence of \textit{EST6}^{H187E}.

Despite this, it is still conceivable that Glu 187 is causing the altered pH dependence through a direct effect on the catalytic mechanism of the enzyme, because \textit{EST6} may not have the same catalytic mechanism as the serine proteases. It has been proposed that \textit{AChE} uses an acid-base mechanism for hydrolysis of ester substrates which involves the reactive Ser and a His residue (Quinn \textit{et al}, 1989) and it is possible that this mechanism applies to \textit{EST6} as well (section 1.3.4.1). Such a mechanism does not require increased nucleophilicity of the reactive Ser (Williams, 1984), therefore perturbation of this mechanism should not affect the sensitivity of the reactive Ser to DFP. Hence it is conceivable that some interaction between the catalytic Ser or His residues and the Glu residue introduced at position 187 of \textit{EST6}^{H187E} is interfering with this mechanism.

Clearly there are numerous possible explanations for the anomalous pH dependence of \textit{EST6}^{H187E}. Elucidation of what precisely is causing this phenomenon awaits the determination of the three dimensional structure of \textit{EST6} and the mechanism of catalysis utilised by the serine esterases. The precise role of solvent hydroxyl ions in the hydrolysis of esters by \textit{EST6}^{H187E} at alkaline pH also awaits detailed biochemical analysis of \textit{EST6}^{H187E} at alkaline pH.

Since there is a possibility of an altered catalytic mechanism used by the \textit{EST6}^{H187E} mutant at alkaline pH, comparison of the kinetic parameters for this mutant with those of the wild type and \textit{EST6}^{H187Q} enzymes at high pH may not be valid. However, it is reasonable to assume that at neutral and acidic pH the mechanism will not be altered, given that all the enzymes have the same pH dependence below pH8. As all the results being discussed in the sections below were determined at pH7, comparison of kinetic parameters between enzymes seems reasonable.

\textbf{2.4.2 SUBSTRATE INTERACTIONS}

From the data presented on the substrate specificity it is apparent that wild type \textit{EST6} has a preference for short chain acyl groups but is less sensitive to the size of the alkyl group of the substrate. Specifically, \textit{b}-napththylesters with two and three carbon acyl groups are preferred to the four carbon butyrate ester, while the single aromatic ring of \textit{p}-nitrophenol is only marginally less preferable as the alkyl group
than the bulkier βnaphthol group. Similar results have been reported for EST6 previously (Mane et al, 1983; White et al, 1988).

The EST6H^{187}Q mutant shows a similar substrate specificity to the wild type but the EST6H^{187}E mutant shows some significant differences. In particular, this enzyme has lower specificity constants for βnaphthyl esters and p-nitrophenylacetate but has gained the ability to hydrolyse acetylthiocholine.

The similarity between the free energy relationships of the wild type and EST6H^{187}Q enzymes for a variety of substrates differing in both acyl and alkyl groups indicates that the His at position 187 is either not involved directly in substrate or transition state binding or, if it is, that the same interactions can be performed by Gln in its place. As Fersht et al (1985) have pointed out, there is a sterically equivalent polar nitrogen atom in the side chains of Gln and His which allows these two residues to perform the same bonding interactions under some circumstances.

The changes in free energy demonstrated by EST6H^{187}E in comparison to wild type EST6, however, suggest that the residue at position 187 does have an influence on the bonding interactions between the enzyme and the substrate. For all the substrates tested there is a decrease in the interaction energy for EST6H^{187}E, in comparison to wild type EST6, of between -1.7 kJ/mol, for βnaphthylacetate and βnaphthylpropionate and -4.2 kJ/mol for p-nitrophenylacetate (Table 2.6).

These observations suggest that residue 187 does affect substrate specificity and that any role played in this function by the His at 187 can be fulfilled by Gln in its place. However, the presence of Glu at this position decreases the strength of the interaction between the substrate and the enzyme.

There are three lines of evidence that suggest the residue present at position 187 influences the enzyme's binding of the alkyl rather than the acyl group of the substrate. Firstly, acyl chain length has no significant effect on the change of free energy of interaction for EST6H^{187}E in comparison to wild-type, as there are no significant differences between the ΔG_i values calculated for βnaphthylacetate, propionate and butyrate. Secondly, the ΔG_i values for βnaphthylacetate and p-nitrophenylacetate are significantly different. As these two substrates have the same acyl group but different alkyl groups, it is the alkyl group of these substrates which is influencing the magnitude of the change in interaction energy. Thirdly, the EST6H^{187}E mutant has gained the ability to hydrolyse acetylthiocholine which, like βnaphthylacetate and p-nitrophenylacetate, also carries an acetate acyl group but differs in the alkyl group.
An explanation for the observed changes in substrate interactions caused by the substitution of Glu for His at position 187 is that this substitution is having indirect effects on the protein's structure which alter the substrate binding pocket. For this to be the case it is necessary to assume that Gln in the same position has no such influence on tertiary structure. This may suggest that Gln is hydrogen bonding to another residue in the protein in the same manner as His and so having little effect on conformation, or alternatively, that the charged Glu residue is distorting the tertiary structure through electronic interaction with other groups in the protein. However, it seems improbable that a tertiary structural change could lead to the altered substrate interactions described above. For this to be the case, it must be assumed that the conformational change has either introduced or removed a group that interacts with the alkyl group of the substrate while not affecting the groups which interact with the substrate acyl group.

A more likely possibility is that residue 187 interacts directly with the alkyl group of the substrate. This possibility can be tested, as specific predictions can be made concerning the changes in interaction energy for particular types of interaction between residue 187 and the substrate alkyl group.

Gráf et al (1988) have determined the changes in interaction energies for a mutant of trypsin containing the charged Asp residue in place of the polar Ser residue at the base of the substrate binding pocket in wild type trypsin. By comparing the specificity constants of this mutant to the specificity constants for wild type trypsin for a variety of substrates, the relative binding energies for various enzyme/substrate interactions were determined. The substrates were classed as either ionic (I), non-ionic but polar (P), or apolar (A). The order of relative strengths of binding interactions so established was as follows

$$E_{II} > E_{PP} > E_{PA} > E_{IP} \dashv E_{IA}$$

(2.14)

where E represents the change in interaction energy and I, P, and A are the type of interacting groups in the enzyme and the substrate respectively. As can be seen from this series, interaction between an ionic group in the enzyme with an ionic group in the substrate contributes the greatest binding energy while ionic-polar and ionic-apolar interactions between the enzyme and the substrate contribute the least (Gráf et al, 1988).

Figure 2.9 shows that the observed changes in interaction energy between wild type EST6 and EST$^{H187E}$ for β-naphthylacetate and p-nitrophenylacetate are consistent with this series, if it is assumed that residue 187 interacts directly with the alkyl
Figure 2.9 Possible interactions between the alkyl groups of the substrates tested and the residue present at position 187 in EST6. The structures of the alkyl groups and amino acid side chains are shown. WT represents wild type EST6 while H187Q and H187E represent EST6H187Q and EST6H187E respectively. The nature of the alkyl groups and the amino acid side chains are given (either polar, apolar or ionic). The change in interaction energy for each of the mutant EST6 enzymes and the type of interaction expected for each is on the right.
group of the substrate. βnaphthol is apolar as it consists of two phenolic rings, p-nitrophenyl has a polar nitro group, both His and Gln have polar groups in their side chains and Glu has an ionic side chain. Hence, the binding of wild type EST6 and EST6H187Q to βnaphthyl esters is a polar-apolar interaction between residue 187 and the alkyl group of the substrate, while their binding to p-nitrophenyl esters is a polar-polar interaction. In contrast, EST6H187E and βnaphthyl esters represents an ionic-apolar interaction, while its binding to p-nitrophenyl esters is an ionic-polar interaction between these groups.

The $\Delta G_i$ values for βnaphthylacetate hydrolysis by EST6H187E in comparison to wild type, a change from a polar-apolar interaction to an ionic-apolar interaction, is -1.7kJ/mol. This decrease in interaction energy is consistent with the results of Gráf et al (1988), as polar-apolar interactions are stronger than ionic-apolar interactions (equation 2.14). For p-nitrophenylacetate, which represents a change from a polar-polar to an ionic-polar interaction, $\Delta G_i$ is 4.2kJ/mol. The polar-polar interaction between His and p-nitrophenyl is complementary. The side chain of Glu, however, is negatively charged, which is not complementary to the negative polarity of p-nitrophenyl, so this will be a repulsive interaction. This results in a decreased $\Delta G_i$ value of a greater magnitude than that observed for βnaphthyl esters, as a complementary polar-polar interaction is being replaced by a weaker, non-complementary ionic-polar interaction (equation 2.14). For acetylthiocholine, no change in interaction energy could be calculated as the wild type enzyme does not hydrolyse this substrate. However, as choline is positively charged, the ability of EST6H187E to hydrolyse this substrate can also be explained by direct interaction between the Glu at 187 and the alkyl group of this substrate, as this represents an ionic-ionic interaction which is complementary in charge. Hence, the observed changes in interaction energy and the gain of acetylthiocholine hydrolytic activity by EST6H187E conform to the predictions made on the basis that residue 187 interacts directly with the substrate alkyl group.

A similar set of predictions can be tested for some mutants of AChE constructed by Gibney et al (1990). Wild type AChE has a Glu in the equivalent position to the His residue at position 187 in EST6 (see figure 1.2). Mutation of this residue to Gln and His reduces activity with acetylcholine as a substrate by 50% and 100% respectively (Gibney et al, 1990). Furthermore, the Glu to Gln change eliminates the substrate inhibition which is normally observed with this enzyme. Quinn (1987) have presented evidence for the existence of a carboxyl group in the active site of AChE that
interacts with the positively charged alkyl group of acetylcholine. It is reasonable to assume, from the work of Gibney et al (1990), that the carboxyl group in question is that of the Glu residue adjacent to the reactive Ser. The interaction between acetylcholine and this residue is ionic, which provides an explanation for the 50% loss of activity when it is replaced by the neutral Gln. However, the possibility that indirect structural disruption due to the substitution of this Glu residue is responsible for the observed effects on the activity of these AChE mutants cannot be ruled out. Despite this, the results of Gibney et al (1990) do support the contention that the residue adjacent to the reactive Ser in AChE and EST6 interact directly with the alkyl group of the substrate; firstly because all the alterations in activity for the various substrates can be explained by electrostatic interactions between the substrate and this amino acid residue, and secondly, if an effect on tertiary structure is the cause, then the same structural disruption must have been introduced by these substitutions in both the AChE mutants and the EST6 mutants presented here, which seems unlikely.

In summary, the data presented strongly suggest that residue 187 in EST6 and its equivalent in AChE interact directly with the substrate. Substitution of this residue to amino acids with different properties leads to altered substrate preferences which can be explained in terms of electrostatic interactions between this amino acid and the alkyl group of the substrate.

2.4.3 Activation Energies

The Gibb's free energy of activation for both wild-type EST6 and EST6H187Q is 66 to 67 kilojoule/mol at 25°C. It can be seen from Table 2.8 that the EST6H187Q mutant has a higher enthalpy of activation, which is offset by a reduced entropy value, resulting in an overall free energy change which is not significantly different to that observed for the wild type enzyme. However, as the enthalpy values were calculated from the slope of the Arrhenius plot, they are associated with high standard errors; therefore none of the differences in enthalpy and entropy between these enzymes are significant. Clearly the determination of $k_{cat}$ at several more temperatures is necessary to produce better estimates of the enthalpy and entropy values, and so determine if there are differences between these enzymes.

The Gibb's free energies, however, could be calculated from a single temperature point and so give accurate estimates of this quantity. The Gibb's free energy of activation for EST6H187E (70kJ/mol) is significantly higher than that observed for wild type EST6 and EST6H187Q. Comparison of the individual enthalpy and
entropy values reveals this change to be due mainly to an increase in enthalpy. This higher Gibb's free energy of activation indicates that a greater input of energy is required for hydrolysis to occur; in other words EST6$^{H187E}$ is a poorer catalyst.

The precise meaning of the Gibb's free energy quantity depends on what rate constants affect the parameter $k_{cat}$ because this is the parameter on which the Arrhenius plots, and so the estimates of activation energy, are based. As has been discussed earlier (section 2.1.2), $k_{cat}$ is dependent on both $k_2$, the rate of acylation, and $k_3$, the rate of deacylation. If either one of these steps is rate limiting, then the calculated energy parameters will apply to that step. If, however, both rate constants are in part rate limiting, then the activation energies described here will be a complex function of both steps. (Whichever is the case, it can be seen from figure 2.8a that the Arrhenius plots are linear, which indicates that the same step(s) remains rate limiting throughout the temperature range assayed in this study for each of the enzymes.)

Which of $k_2$ and $k_3$ is rate limiting for EST6 can in part be deduced from the substrate specificity data presented in this study; this is because the alkyl group of the substrate should have no influence on $k_3$, since the alcohol leaving group is not involved in deacylation (section 2.1.2, scheme 2.3). As has been pointed out previously (section 2.4.2), the nature of the alcohol leaving group has a significant effect on $k_{cat}$, with EST6 having a lower turn-over number for $p$nitrophenylacetate than for $\beta$napthylacetate as substrate. (Acyl chain length has an even greater effect on $k_{cat}$, which may reflect on either acylation or deacylation rate, since this group is involved in both processes.) Therefore, acylation ($k_2$) is at least partly rate limiting for EST6; whether $k_3$ is also in part rate limiting cannot be determined from the data. Hence, the larger Gibb's free energy of activation for EST6$^{H187E}$ may be due to an increased energy barrier, at least for acylation, but possibly for deacylation as well.

The theory of transition state stabilisation states that enzymatic catalysis is due to an enzyme being complementary to the transition state rather than the unreacted substrate, which results in a lower activation energy barrier to be overcome for the reaction to proceed (Pauling, 1948; Fersht, 1974). As has been pointed out by Fersht and Leatherbarrow (1987), if an amino acid side chain that binds equally well to the substrate in both the ground and the transition states is replaced by one incapable of the same interactions, then this should raise $K_s$ but not affect $k_{cat}$. However, if the side chain only binds substrate in the transition state, then replacement should leave $K_s$ unaltered but lower $k_{cat}$.
Application of this theory to the EST6 data presented is limited by the inability to estimate $K_s$. This parameter is a complex function of the rates involved in every step of the reaction (equation 2.7) and will only approximate $K_s$ when the condition $k_2 \ll k_3$ is met. Since this does not apply to EST6, the energy of substrate binding is incalculable (hence the values shown in Table 2.8 do not reflect the process energies of substrate binding). This limits the discussion of whether residue 187 binds to the substrate in the ground state or the transition state to the $k_{cat}$ parameters.

It is apparent from Table 2.4 that the major contributions to the differing specificity constants ($k_{cat}/K_m$) for EST6$^{H\text{187}E}$ are from the $k_{cat}$ values. In the previous section (section 2.4.2) it was proposed that the decreased energy of interaction with the substrate, for EST6$^{H\text{187}E}$ in comparison to the wild type enzyme, is due to altered interactions between residue 187 and the alkyl group of the substrate. Hence, by the criteria of Fersht and Leatherbarrow (1987) outlined above, this is consistent with the interaction between residue 187 and the alkyl group of the substrate being realised in the transition state. In light of this and the interpretation given for the altered substrate specificity of EST6$^{H\text{187}E}$ (section 2.4.2), the simplest explanation for the greater Gibb's free energy of activation for this mutant enzyme is that the energy barrier for acylation has been increased through the weakening of a transition state interaction between Glu 187 and the alkyl group of the substrate.

2.4.4 CONCLUSIONS

The substitution of Glu for His at 187 affects the activity of the enzyme in a number of ways. Firstly, EST6$^{H\text{187}E}$ has an altered pH profile, which has raised the possibility of an altered catalytic mechanism for this mutant at alkaline pH, which involves solvent hydroxyl ions. Secondly, this substitution affects the substrate specificity of the enzyme, with respect to the nature of the alkyl group. The data are consistent with this being due to direct interaction between residue 187 and the substrate alkyl group and that this interaction is realised in the transition state. Consistent with this is the increased Gibb's free energy of activation for EST6$^{H\text{187}E}$.

While determination of the tertiary structure of EST6 and further biochemical information concerning the rate limiting step in the catalytic mechanism are necessary for a full understanding of the effects produced by the substitutions of His 187 described in this study, several important facts have been established. Introduction of Glu at position 187 leads to the utilisation of a different mechanism of catalysis at alkaline pH and an altered substrate specificity in terms of the alkyl group.
of the substrate. These results indicate that this residue may have been important in the evolution of new enzymatic activity, and so functional divergence, after the duplication of the "ancestral esterase" gene.
CHAPTER 3 ANALYSIS OF ASPARAGINE-LINKED GLYCOSYLATION

3.1 INTRODUCTION

A major structural feature of EST6 that may be of importance to the function of this protein is the presence of attached carbohydrate (Mane et al., 1983). The N-linked carbohydrate component of secreted, membrane bound and lysosomal glycoproteins has been the subject of much study and speculation. Many different roles for the glycans attached to glycoproteins have been implicated and include; a conformational requirement where glycosylation is necessary for correct folding and stability, intracellular targeting to specific organelles such as lysosomes, extracellular protein targeting and biological recognition and metabolic stabilisation through protection of the polypeptide chain from proteolysis. Examples where no biological function has been assigned to the glycans of glycoproteins also exist (section 1.6).

One limitation of the approaches used in these studies has been the use of general inhibitors of in vivo glycosylation, introducing complications due to non-glycosylation of all cellular glycoproteins including those involved in protein transport. A second limitation is that much of the functional analysis has been carried out in yeast or mammalian cell culture. In vivo functions involving extracellular interactions with tissues or organs removed from the site of production clearly are difficult to test in such systems.

The use of recombinant DNA technology, including site-directed mutagenesis, and P-element mediated germ-line transformation techniques for the introduction of foreign genes into the genome of D.melanogaster, overcomes these two major limitations. The technique of site-directed mutagenesis provides an alternative to the use of general inhibitors of glycosylation as prevention of oligosaccharide attachment to specific N-linked sites on particular proteins can be achieved by introducing appropriate mutations into the corresponding structural gene (Hansen et al., 1988; Taylor and Wall, 1988; Dubé et al., 1988). As a germ-line transformation system is available for D.melanogaster, mutant genes constructed in vitro can be expressed in vivo in the tissue normally responsible for production of the protein in the whole organism; this overcomes the second limitation.
The latter is of particular importance in the case of EST6 because this glycoprotein is exposed to a number of *in vivo* environments, including the haemolymph of both sexes and the male ejaculatory duct (section 1.3.3). Male derived activity is also transferred to the female's reproductive system during mating and subsequently translocated to her haemolymph (Richmond and Senior, 1981). The expression of glycosylation-deficient EST6 in *D. melanogaster* using the combination of site directed *in vitro* mutagenesis and germ-line transformation will enable the analysis of the *in vivo* function of the carbohydrate component of this glycoprotein.

### 3.1.1 Choice of Mutations

N-linked oligosaccharides are invariably attached to the asparagine (Asn) residue in the glycosylation site consensus sequence, Asn-X-Ser/Thr, where X represents any amino acid except proline (Marshall, 1972). The substitution of Asn with Gln in consensus sites that are normally glycosylated has been found to prevent oligosaccharide attachment (Hansen *et al.*, 1988; Taylor and Wall, 1988; Dudé *et al.*, 1988). Such a substitution should have minimal direct effects on the tertiary structure of the protein because the side chain of Gln has only one extra carbon atom. As oligosaccharides of up to several kilodaltons in size are normally attached to these Asn residues, the replacement of Asn with a residue carrying one extra carbon atom should have no significant steric effects on protein conformation. In one exceptional case, however, the substitution of Gln for Asn successfully prevented carbohydrate attachment to the protein and disrupted the normal subunit association this protein undergoes, but this disruption was due to the substitution of Asn rather than the lack of carbohydrate attachment (Santos-Aguado *et al.*, 1987). This demonstrates that functional differences due to mutation of N-linked glycosylation sites are not always due to the absence of oligosaccharide and so caution must be used in the interpretation of such experiments.

This chapter describes the construction and expression of glycosylation-deficient mutant EST6 by the substitution of Gln for Asn in the four consensus N-linked glycosylation sites in the protein, followed by germline transformation. Mutants deficient in only one of each of the four potential N-linked glycosylation sites and a fifth mutant deficient in all four sites were constructed and expressed in
D. melanogaster. These were then used for structural characterisation of the carbohydrate attached to EST6. A preliminary functional analysis of the role of the oligosaccharides of EST6 is also described.
3.2 MATERIALS AND METHODS

The materials and methods used in the analysis of the glycosylation of EST6 are described in three sections; 1) the construction and expression of the glycosylation site mutations, 2) biochemical procedures used in the analysis of the mutant proteins and 3) procedures used for their physiological characterisation.

3.2.1 CONSTRUCTION AND EXPRESSION OF MUTANT CLONES

The procedures used for the construction of the glycosylation site mutant clones of Est6 and germ-line transformation of these into D.melanogaster are described.

The restriction enzyme digests, ligations, DNA sequencing, Southern blots and gel purifications of restriction fragments were performed as described in section 2.2.1.1.

3.2.1.1 Oligo-nucleotide Directed in vitro Mutagenesis

Mutation of Asn codons to Gln codons at each of the four potential N-linked glycosylation sites of EST6 was achieved using oligo-nucleotide directed in vitro mutagenesis. Mutagenic oligo-nucleotides were custom-synthesized commercially. As well as the required codon change, restriction sites were included or deleted in the sequence of the oligo-nucleotides, in order to facilitate preliminary identification of mutant clones. The sequences and restriction site alterations of each oligo-nucleotide are shown in table 3.1.

The template for mutagenesis was a 3.9 kilobase HindIII fragment, containing the Est6 coding region, from a genomic clone isolated from the wild type strain Dm145 (Oakeshott et al, 1987). This fragment was subcloned into the HindIII site of the plasmid vector pTZ18U, from which single stranded DNA was made using the modified M13 helper bacteriophage K07 (Veiera and Messing, 1982; Zagursky and Berman, 1984). Oligo-nucleotide directed in vitro mutagenesis was performed as described in section 2.2.1 using this single stranded DNA as a template. Preliminary screening for mutagenised clones was performed using restriction digests of plasmid DNA to test for the presence of the novel restriction sites described in Table 3.1. In cases where the mutagenic frequency was low, initial screening was performed by colony hybridisation using the mutagenic oligo-nucleotide as a hybridisation probe.
<table>
<thead>
<tr>
<th>Glycosylation site number</th>
<th>Wild type sequence (top line)</th>
<th>Oligo-nucleotide sequence (bottom line)</th>
<th>Restriction sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>-GS(^1) N21Q</td>
<td>Asn</td>
<td>TGGTGGGATAATGGAAGGCTACTACACG</td>
<td>BamHI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TGGTGGGATCAAAGGCTACTACACG</td>
<td></td>
</tr>
<tr>
<td>-GS(^2) N399Q</td>
<td>Asn</td>
<td>TTCCTTTCAAGCATAGCAGGAGGTC</td>
<td>TaqI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TTCCTTTCAAGCATGTCGAGGAGGTC</td>
<td></td>
</tr>
<tr>
<td>-GS(^3) N435Q</td>
<td>Asn</td>
<td>GGTCCCTGGCCAAGGCTGAGATT</td>
<td>TaqI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GGTCCCTGGCCAAGGCTGAGATT</td>
<td></td>
</tr>
<tr>
<td>-GS(^4) N485Q</td>
<td>Asn</td>
<td>GCTTCGAGTGAATGCTCTCATTAAAT</td>
<td>Sau3A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GCTTCGAGTGAATGCTCTCATTAAAT</td>
<td></td>
</tr>
</tbody>
</table>

**Table 3.1 Mutagenic Oligo-nucleotides for Glycosylation Sites**

Glycosylation site mutations are represented by -GS\(^x\) where x denotes their order in the primary sequence, starting at the N-terminal end. Below this is given the amino acid change (single letter code) and position in the primary sequence. The asparagine codon (AAT in every case) has been altered to the glutamine codon (CAA). Base changes in the oligo-nucleotides are underlined. Altered restriction sites are in italics and the name of the enzyme that recognises the site given. For -GS\(^1\), -GS\(^2\) and -GS\(^4\) a restriction site was introduced and for -GS\(^3\) a site removed. Nucleotide changes other than those in the Asn codons do not lead to amino acid substitutions.
3.2.1.2 Oligo-nucleotide Probes and Colony Hybridisation

Hybridisation probes were end labelled with $^{32}$P-ATP according to the method described by Maniatis et al (1982). Bacterial colonies carrying plasmids generated by the above mutagenesis procedure were transferred onto nitrocellulose filters and the DNA extracted by placing the filters on two changes of filter paper soaked in 0.5M NaOH/1.5MNaCl for 10 minutes each. This procedure was repeated with filter paper soaked in 0.5M Tris-HCl (pH8.8), after which the filters were briefly rinsed in 2XSSC. After baking at 80°C under vacuum for 2 hours the filters were washed in 2XSSC/0.1%SDS at 65°C for a further 2 hours. Hybridisation of the radio labelled oligo-nucleotide probe was performed at 37°C overnight in 1M phosphate buffer (pH6.9) containing 0.7% SDS and 0.5% Blotto. The filters were then briefly washed in 3XSSC/0.1%SDS at room temperature and autoradiographed to ensure that the probe had hybridised to the plasmid DNA from the colonies. High stringency washes were then performed at a temperature between the melting point for a perfect match in nucleotide sequence between the oligo-nucleotide and the plasmid and the melting point if mismatches are present. Melting temperatures ($T_m$) were calculated from the equation

$$T_m = 2\times(A+T) + 4\times(G+C) \quad ^\circ C$$

where $A$, $C$, $G$ and $T$ represent the number of those respective nucleotides in the probe. An A or T mismatch will reduce the $T_m$ by approximately 2°C while a C or G mismatch will reduce it by 4°C. Washing at a temperature of approximately two degrees below the $T_m$ calculated for a perfect match between the mutagenic oligo-nucleotide and the plasmid insert should distinguish between wild type and mutant clones.

The desired mutant clones identified in this way were then confirmed by restriction digests. All mutants were checked by sequencing to ensure that the required substitutions were present and that no unwanted mutations had been introduced.

3.2.1.3 Construction of a Clone With Four N-linked Glycosylation Site Mutations

The strategy for construction of an $Est\delta$ clone carrying all the mutations introduced by the four oligo-nucleotides described above was as follows. A clone carrying the second glycosylation site mutation was used as a template for mutagenesis
with the oligo-nucleotide for the third glycosylation site. A fragment, from the SalI site in the pTZ18U poly linker to a Ncol site present in the insert, spanning the region of the clone containing these 2nd and 3rd site mutations, was then ligated into the corresponding position of a clone bearing the 1st glycosylation site mutation. A Csp451 fragment carrying the first three glycosylation site mutations was then isolated from this clone and substituted for the equivalent fragment from the clone mutant for the fourth glycosylation site (see Appendix I for restriction map). At every stage restriction digests were performed to check for the presence of the desired mutations. The DNA surrounding the restriction sites involved in the construction was sequenced in the final clone to ensure no rearrangements or mutations had been introduced during the process.

3.2.1.4 Germ-line Transformation

Insertion of mutant clones into the CaSpeR transformation vector and germ line transformation were performed as described in section 2.2.2. Putative transformants that segregated with the first or second chromosome were isolated for each construct and the presence of an introduced copy of Est6 confirmed by Southern blot analysis. Transformants were then crossed into an Est60 background and established as stocks.

3.2.2 BIOCHEMICAL ANALYSIS OF THE MUTANT PROTEINS

This section describes the techniques used to determine the structural effects of mutation of the potential N-linked glycosylation sites of EST6.

3.2.2.1 Preparation of Homogenates.

Homogenates of the transformed stocks were prepared from collections of adult flies (2 to 5 day old) of mixed sexes that had been frozen in liquid nitrogen and stored at -70°C. These were then homogenised in 0.1M Phosphate pH 6.8 at 4°C and the resulting slurry centrifuged at 14,000 xg for ten minutes. The supernatant, containing the soluble protein fraction, was used as the source of EST6 protein for analysis. Homogenates were prepared on the day of use to minimise degradation of any carbohydrate present.
End-point esterase activity assays, native and denaturing polyacrylamide gel electrophoresis and protein determinations were performed as described in section 2.2.3.1.

3.2.2.2 Western Blots and Antibody Detection of EST6.

EST6 in crude homogenates was detected on denaturing SDS gels by Western blotting and binding to anti-EST6 polyclonal anti-sera. Each lane was loaded with 10 to 25μg of protein, electrophoresed for 3 to 4 hours and electroblobbed on to nitrocellulose overnight in Tris-glycine/20% methanol at 30V, 0.1mA. Filters were blocked in 50mM Tris, 0.15M NaCl, 0.1% Tween20, pH7.4 (TBST) and 2%BSA for 1 to 2 hours at room temperature. The primary antibody, supplied by R. Richmond, was generated by immunising rabbits with EST6 purified by affinity chromatography (White et al, 1988). The anti-sera, which was supplied lyophilised, was resuspended in TBST plus 2% BSA and incubated with the filters overnight at room temperature. Filters were then washed four times in TBST before incubation with alkaline phosphatase-conjugated goat anti-rabbit secondary antibody (Boehringer Mannheim) in TBST plus 2% BSA for at least 3 hours. Filters were then soaked in 100mM Tris, 100mM NaCl, 5mM MgCl₂, pH9.5 for 5 minutes, blotted dry then reacted with 160μg/ml bromo-chloro-indolyl phosphate, 330μg/ml Nitroblue tetrazolium in the same buffer. The colour development was stopped by washing in distilled water.

3.2.2.3 Concanavalin A Binding

N-linked carbohydrate was detected by the ability of the protein to bind concanavalin A (conA). ConA linked to sepharose 4B was purchased from Sigma. For each strain, aliquots of 30 μl of crude homogenates (100μg protein) were incubated with 5μl conA sepharose at room temperature, with occasional mixing, for 45 minutes. The conA sepharose was pelleted by centrifugation at 14,000xg for 5 minutes. The supernatant was collected and esterases remaining in solution fractionated by native PAGE and detected by staining with a and β-naphthylacetate. ConA treated samples were compared to untreated aliquots of the same homogenates.
3.2.2.4 Glycanase Digestion

The glycans of EST6 were tested for sensitivity to digestion by specific endoglycosidases to enable initial characterisation of their structure. An initial time course experiment was performed on partially purified wild type EST6, using endoglycosidase F, to determine the conditions and length of time for incubation with the glycanase required. Endoglycosidase F (free of N-glycosidase F), and endoglycosidase H (endoH) were purchased from Boehringer Mannheim.

The time course of digestion of denatured and undenatured EST6 with endoF was performed on partially purified EST6 kindly supplied by R.Richmond. This protein was purified by the method described by White et al (1988). 20ug of this preparation was diluted in 0.1 M NaAcetate buffer pH5.8 and denatured by the addition of SDS to 4% and 2-mercaptoethanol to 1% followed by boiling for 3 minutes. The non-ionic detergent, n-octylglucoside, was added to 10% and the protease inhibitor 1,10-phenanthroline to 5mM. N-octylglucoside was used in preference to other non-ionic detergents to sequester the SDS before glycosidase digestion as this detergent does not interfere with SDS PAGE (Haselbeck and Hosel, 1988). This denatured protein sample was digested with 20mU of endoF. Aliquots containing 4ug of protein were removed at regular time intervals and boiled to stop the reaction. This procedure was also performed on 20ug of the protein without denaturation prior to addition of the endoF. In this case, non-ionic detergent was not added before the addition of the endoF. All other conditions were the same.

Digestion of unpurified EST6 in crude whole fly homogenates was performed using the conditions identified by the above experiment. Aliquots containing 50ug of protein from crude homogenates, diluted in 0.1M NaAcetate pH5.8 buffer, were denatured as described above. The denatured protein sample was then digested overnight at room temperature with either 0.1U of endoF or 20mU of endoH and electrophoresed on SDS denaturing PAGE. The gels were then Western blotted and EST6 protein detected as described in section 3.2.2.2.

3.2.2.5 Thermostability

Homogenates prepared as described in section 3.2.2.1 were diluted to 800ug protein/ml with PO₄ buffer. Three aliquots of 160ul of each strain were incubated at 55±1°C in a water bath for each of five time intervals ranging from 2 to 14 minutes.
and then placed on ice. Esterase activity was detected by the end point assay described in section 2.2.3. As other esterases present in adults are more heat labile, the activity remaining after 2 minutes at 55°C is attributable to EST6 (Cochrane and Richmond, 1979). The natural logarithm of the fraction of remaining activity, taking the activity after 2 minutes incubation as the maximum, was plotted against time of incubation and the slope of the regression line taken as a measure of thermostability.

3.2.3 PHYSIOLOGICAL ANALYSIS

This section describes experiments for the assessment of the functional consequences of mutation of the N-linked glycosylation sites of EST6. The functions examined were the transfer of male derived enzyme to the female during mating and translocation of male donated activity to her haemolymph. Before these experiments were undertaken, the tissue distribution of EST6 in the transformants was determined to ensure they displayed the wild type expression pattern.

3.2.3.1 Tissue Distribution

Male ejaculatory ducts were dissected from the transformants to determine if the integrated Est6 genes were being expressed in the correct tissues. As EST6 is also expressed in the haemolymph of both males and females, whole females and males with their ejaculatory ducts removed were also examined for the presence of EST6 activity.

Anterior and posterior ejaculatory ducts and the ejaculatory bulb were dissected together from 4 to 5 day old virgin males. Dissections were performed in Drosophila Ringer's solution and the dissected material frozen in liquid nitrogen and stored at -70°C. The carcasses of these males and whole 4 day old virgin females were also collected and frozen. Samples were homogenised as described in section 3.2.2.1 and 30ug protein from the whole females and male carcasses and 1ug from the ejaculatory ducts were fractionated by native PAGE. The presence of EST6 was detected by staining with a and βnaphthylacetate as substrates (section 2.2.3).

3.2.3.2 Mating Experiments

Mating experiments were performed to determine if the glycosylation site mutations in EST6 affected transfer to the female during mating and subsequent translocation to her haemolymph. This was detected by the appearance of male derived
activity in the abdomens and thoraces of mated females. Virgin males from the transformed mutant strains and w;Est6\(\theta\) females, which lack endogenous EST6, were collected at intervals of less than 12 hours and left to age at 20°C. After 3 to 4 days, two virgin males were introduced into the mating vial, without anaesthetisation, and left a further day. The next morning a 3 to 4 day old virgin female was aspirated into the vial and the flies left to mate. At specified time intervals after mating had been initiated, the mating pair was collected by aspiration and frozen in liquid nitrogen.

The frozen mated females were then separated into abdomen, and head and thorax, by dissection with a scalpel on a glass plate resting in a dry ice/ethanol bath to keep the flies frozen throughout the procedure. The presence of male donated EST6 in individual w;Est6\(\theta\) female body parts from each time point was detected by native PAGE followed by staining with \(a\) and \(\beta\)naphthylacetate.
3.3 Results

Transformants were obtained for each of the constructs and two independent transformants for the -GS1234 construct and one each of the other constructs chosen for biochemical and physiological analysis. In each of these strains the eye colour phenotype, and hence the mutant Est6 gene, segregates with the first or second chromosomes, enabling replacement of the third chromosomes with a homologue carrying the Est6° allele. For all of the experiments described Dm145 was used as the reference wild type strain. This was the same strain from which the genomic clone used in constructing the glycosylation site mutations was originally isolated (Oakeshott et al, 1987), so any differences between EST6 from this strain and the proteins with mutated glycosylation sites are attributable to the introduced mutations.

3.3.1 Biochemical Characterisation of EST6 with Mutant Glycosylation Sites

3.3.1.1 Electrophoretic Mobility and Size Estimation

Native and denaturing PAGE were performed to determine the effect of the glycosylation site mutations on the molecular weight and native electrophoretic mobility of EST6.

The soluble fractions from whole fly homogenates were prepared, electrophoresed on native PAGE and stained for esterase activity. As can be seen in Figure 3.1a, all the glycosylation site mutant proteins have bands of βnaphthylacetate staining esterase activity. However, the electrophoretic mobilities of these are different to the wild type EST6. The one exception to this is the EST6-GS4 enzyme, which has an identical mobility to one form of the wild type protein. However, EST6-GS4 does differ from the wild type enzyme in another respect. Wild type EST6 appears as a doublet on native PAGE but, as can be seen in Figure 3.1a, the slower mobility isozyme is absent in EST6-GS4. EST6-GS1234 also lacks a second, slower mobility β staining esterase. Each of the other mutant proteins however, appear as doublets and both isozymes are altered in electrophoretic mobility to the same extent. Mobility can be expressed as the fractional mobility of the enzyme relative to the dye front (Rf), but for the purposes of this study, Rf will be given as relative to the major, faster
Figure 3.1  

a) Native PAGE (7% polyacrylamide) of whole fly homogenates. -GSX = EST6-GSX and Dm145 is the reference wild type strain. The gel has been stained with a and βnaphthylacetate/fast garnet as described in section 2.2.3.1.

b) Whole fly homogenates separated by SDS PAGE (10% polyacrylamide), Western blotted and stained with anti-EST6 antisera (primary anti-body) and an alkaline phosphatase conjugated second anti-body. Lanes are labelled as for a). The position of the 66 and 42 kilodalton markers is shown.
mobility isozyme of activity displayed by wild type EST6 from Dm145 homogenates. The \( R_f \) values so calculated are shown in Table 3.2.

The soluble fractions of crude homogenates of each strain were also electrophoresed on denaturing SDS gels, Western blotted and EST6 detected with the anti-EST6 polyclonal antibody (Figure 3.1b). Again there are two forms of the wild type protein present which differ in size by approximately 1.5 kd. Each of the homogenates of the transformants have protein binding the anti-EST6 antibody. However, the mutant proteins are a different size to the wild type EST6 protein in all cases, with the exception again being EST6-GS\(^4\). This strain has a protein of the same size as the smaller of the wild type forms but lacks the higher molecular weight form. EST6-GS\(^{1234}\) also has only one EST6 band. This result, in conjunction with the banding pattern of the EST6-GS\(^4\) mutant on native gels, indicates that the slower mobility isozyme is associated with the presence of the fourth glycosylation site and with the size difference in a proportion of the EST6 molecules present.

<table>
<thead>
<tr>
<th>esterase</th>
<th>relative mobility</th>
<th>size (kd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>wild type EST6</td>
<td>0.96, 1.00</td>
<td>54.5, 56</td>
</tr>
<tr>
<td>EST6-GS(^{1234})</td>
<td>1.20</td>
<td>51</td>
</tr>
<tr>
<td>EST6-GS(^1)</td>
<td>0.98, 1.03</td>
<td>53.5, 54.5</td>
</tr>
<tr>
<td>EST6-GS(^2)</td>
<td>1.04, 1.13</td>
<td>53, 54</td>
</tr>
<tr>
<td>EST6-GS(^3)</td>
<td>0.98, 1.04</td>
<td>53.5, 54.5</td>
</tr>
<tr>
<td>EST6-GS(^4)</td>
<td>1.00</td>
<td>55</td>
</tr>
</tbody>
</table>

Table 3.2 Mobilities of the Glycosylation Site Mutant Proteins
Electrophoretic mobilities (\( R_f \)) of the band(s) of activity of the mutant enzymes on native PAGE are given relative to the faster mobility isozyme of wild type EST6. Molecular weights for the glycosylation site mutant proteins are the average of estimates from three Western blots of SDS gels and are given for the two molecular weight forms where applicable.

Estimates of the size of each of the mutant EST6 proteins were made from several Western blots similar to that shown in Figure 3.1b. The averages of the values
calculated are shown in Table 3.2. Mutations in the first and third glycosylation sites each result in a 1.5 kd decrease in the size of the molecule while the second glycosylation site mutation decreases molecular weight by 2 kd. The EST6-GS\textsuperscript{1234} multiple mutant is 5kd smaller than the higher molecular weight wild type form.

A naturally occurring variant has been identified which lacks the fourth glycosylation site, due to the substitution of Ala for Ser in the third position of the N-linked consensus site tri-peptide (Cooke 1989). Homogenates from these strains were also electrophoresed on native PAGE to determine if this polymorphism affects electrophoretic mobility. As can be seen in Figure 3.2, the slower mobility form of the enzyme is absent in the strain which lacks the fourth glycosylation site, giving further evidence for the association of the slow mobility form of EST6 with the presence of the fourth glycosylation site.

As mutation of each of the first three glycosylation sites results in a decrease in the size of the EST6 molecule and more anodal relative mobility on native PAGE, each of these sites must be glycosylated in the wild type enzyme. Mutation of the fourth site results in the absence of the slower mobility form of EST6 normally observed on native PAGE but does not affect the mobility of the faster electrophoretic form. This suggests that the fourth site is glycosylated on only a portion of the molecules, which results in two different molecular forms of EST6.

3.3.1.2 Concanavalin A Binding

The presence of N-linked glycan moieties on the wild type and mutant EST6 proteins was determined by their ability to bind the mannose-specific lectin conA. Homogenates were treated with sepharose-bound conA as described in section 3.2.2.3 and subjected to native PAGE (Figure 3.3). As well as the wild type and each of the glycosylation site mutant transformants, a strain carrying both an Est6-GS\textsuperscript{1234} insertion and an endogenous copy of the wild type gene was included in this analysis.

The wild type EST6 protein and each of the mutants lacking only one of the glycosylation sites were mostly removed from the homogenate by the conA treatment. In contrast, the amount of EST6-GS\textsuperscript{1234} was not affected. This is best illustrated by the strain carrying both the EST6-GS\textsuperscript{1234} and the wild type protein as the latter acts as an internal control for this homogenate. The wild type protein binds the conA and so is precipitated, while the EST6-GS\textsuperscript{1234} does not bind and so remains in solution. Note
Figure 3.3 Native PAGE of whole fly homogenates treated with sepharose bound conA as described in section 3.2.2.3. A + indicates the homogenate has been treated with conA while - indicates a control lane of the same homogenate which has not been treated with conA. The red stain represents the enzyme’s preference for βnaphthylacetate while the brown stain represents a preference for anaphthylacetate. EST6 is the strongly staining βnaphthylacetate preferring esterase in each lane. The enzyme with a preference for anaphthylacetate that has the same electrophoretic mobility as EST6−GS\textsuperscript{1234} is EST9. This enzyme is also removed by the conA treatment. −GS\textsuperscript{x} represents EST6−GS\textsuperscript{x} and Dm145 is the reference wild type control.
that there is an a staining band that comigrates with EST6-GS$^{1234}$ which binds to the conA and so reduces the intensity of the stain in this region of the gel in the conA+ lanes, however it is apparent that the red colour of the β staining band of the EST6-GS$^{1234}$ protein is not affected (see Fig 3.3, lane 4). This indicates that mutation of the four potential N-linked glycosylation sites eliminates the attachment of N-linked carbohydrate to the EST6 protein while mutation of any one site still allows glycosylation to occur at the other sites. From this analysis it is apparent that the EST6-GS$^{1234}$ protein is non-glycosylated.

It can also be seen in Figure 3.3 that other esterases bind to conA. In particular, esterase 9 (the a staining esterase that comigrates with EST6-GS$^{1234}$) is removed from the homogenates by conA, which indicates that this esterase is also glycosylated.

3.3.1.3 Glycanase Digestion

Glycanase sensitivity can be used to structurally characterise glycan moieties due to the different substrate specificities displayed by these enzymes. EndoF hydrolyses high mannose and hybrid and complex biantennary N-linked glycans, while endoH is restricted to simple high mannose substrates (Tarentino et al, 1985).

Partially purified EST6 was digested with endoF then electrophoresed on a denaturing SDS gels and the protein visualised by silver staining (Figure 3.4). The denatured protein is hydrolysed to a form with lower molecular weight after 6 hours of digestion. In contrast, in the native, undenatured form a large proportion of the protein does not display this change in molecular weight. It has commonly been observed that endoglycosidases act more efficiently when the substrate glycoprotein has been denatured prior to digestion (Tarentino et al, 1985) and EST6 appears to be no exception. For this reason, endoglycosidase digests of crude homogenates from the glycosylation site mutant transformed strains were performed on previously denatured preparations.

Crude homogenates of the mutants were denatured and digested as described and Western transfers performed to visualise the EST6 protein. Both endoF and endoH gave similar results and one such Western blot of homogenates digested with endoF is shown in Figure 3.5. The wild type EST6 and the 1st, 3rd and 4th glycosylation site mutants show a shift in mobility after endoF digestion. The decrease in molecular weight
Figure 3.4 SDS denaturing PAGE (7% polyacrylamide) of partially purified EST6 digested with endoglycosidase F. Digests were performed on EST6 in the native or the denatured form as described in section 3.2.2.4. Two control lanes are shown (left hand lanes), in which EST6 was denatured as for digestion with endoF then diluted in SDS sample buffer, and EST6 which had not been denatured before being diluted in SDS sample buffer. The next two sets of lanes are samples of EST6 digested with endoF while in the denatured or native forms as indicated. The time of incubation with endoF for each sample is also shown. The position of the 45 and 66 kilodalton molecular weight standards is marked.
Figure 3.5 Western blot of whole fly homogenates separated by SDS PAGE (10% polyacrylamide) and stained with anti-EST6 antisera (primary anti-body) and an alkaline phosphatase conjugated second anti-body. The lanes labelled +EndoF contain samples that had been denatured and digested with endoF. The lanes labelled -EndoF contain aliquots of the same samples that were not digested with endoF. -G$^x$ represents EST6-G$^x$ and Dm145 is the reference wild type control.
observed for these proteins after digestion with endoF is approximately 1-2 kd and is similar to that observed for the denatured purified preparation in Figure 3.4. The EST6-GS\textsuperscript{1234} and EST6-GS\textsuperscript{2} mutant proteins show no size change after digestion with endoF, indicating that the size changes observed for the other glycosylation site mutants and wild type EST6 are not due to proteolysis. The resistance of EST6-GS\textsuperscript{2} to endoF and endoH digestion, in contrast to the other single glycosylation site mutants, indicates that the glycan attached to the second glycosylation site is sensitive to hydrolysis by these enzymes. Thus the glycan attached to the second glycosylation site is of the high mannose type of oligosaccharide.

3.3.1.4 Thermostability

Disruption of the conformation of a protein which results in a less energetically stable form should result in an increased rate of denaturation at elevated temperature. Therefore, determination of the rates of thermal denaturation of the glycosylation site mutant proteins will be a measure of the structural disruption caused by the introduced mutations. The rate of thermal denaturation at 55°C for each of the mutant enzymes quantifies the thermostability of the enzymes, thus allowing comparison of the mutant EST6 proteins.

The data from which the rates of thermal denaturation were calculated are shown in Figure 3.6. The rapid initial decrease in activity is attributable to other, more thermolabile esterases, while the activity remaining after 2 minutes incubation is attributable to EST6 (Cochrane and Richmond, 1979). The relative amounts of EST6 activity in each of these strains is also apparent from this figure. Most of the transformant strains have slightly less activity than Dm145, with the exception being EST6-GS\textsuperscript{1} which has a substantially greater amount.

Thermal denaturation rates were calculated as described in section 3.2.2.4 and are shown in Table 3.3. The thermal denaturation rate of the EST6-GS\textsuperscript{2} mutant protein is significantly greater than that of wild type EST6 and each of the other glycosylation site mutants. None of the other proteins, including the two EST6-GS\textsuperscript{1234} strains analysed in this experiment, are significantly different from the wild type enzyme.

The -GS\textsuperscript{2} mutation, which results in a significant increase in the rate of thermal denaturation, would also be expected to increase the rate observed for EST6-GS\textsuperscript{1234}, which also carries this mutation. Hence, the similarity of the thermal
Figure 3.6  Plot of the total remaining activity versus the time of incubation at 55 degrees celsius. The standard errors associated with each point are less than 10% in each case. Activity is umol 8naphthol produced per minute per milligram of protein.
denaturation rates for EST6-GS\textsuperscript{1234} and wild type EST6 is unexpected. Presumably, some interaction between this mutation and one or more of the other glycosylation site mutations must be compensating for the destabilising effects of the -GS\textsuperscript{2} mutation.

<table>
<thead>
<tr>
<th>Esterase</th>
<th>Thermal Denaturation Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>wild type EST6</td>
<td>-0.043±0.009</td>
</tr>
<tr>
<td>EST6-GS\textsuperscript{1234}(a)</td>
<td>-0.043±0.014</td>
</tr>
<tr>
<td>EST6-GS\textsuperscript{1234}(b)</td>
<td>-0.037±0.017</td>
</tr>
<tr>
<td>EST6-GS\textsuperscript{1}</td>
<td>-0.036±0.014</td>
</tr>
<tr>
<td>EST6-GS\textsuperscript{2}</td>
<td>-0.132±0.037</td>
</tr>
<tr>
<td>EST6-GS\textsuperscript{3}</td>
<td>-0.043±0.009</td>
</tr>
<tr>
<td>EST6-GS\textsuperscript{4}</td>
<td>-0.057±0.005</td>
</tr>
</tbody>
</table>

Table 3.3 Denaturation rate is expressed as the slope of the plot of the natural logarithm of the fraction of activity remaining versus time at 55\textdegree C. EST6-GS\textsuperscript{1234} (a) and (b) represent homogenates from independent transformant strains expressing this protein. Standard errors are taken to be those of the regression line.

3.3.2 PHYSIOLOGICAL CHARACTERISATION

The transfer of male derived EST6 to females and translocation of this into the female's haemolymph was investigated for the glycosylation defective mutants. Before these experiments were performed, the tissue expression of EST6 in the transformants was determined, to ensure that altered tissue expression would not complicate the results of the mating experiments.

3.3.2.1 Tissue Localisation

Ejaculatory ducts were dissected from each of the transformant strains and homogenates electrophoresed on native gels. The remainder of the males and whole females from the same strains were treated in the same way (Figure 3.7). The majority of the male activity in each strain is present in the ejaculatory ducts. The females and male remainders have a lesser amount of activity, which presumably corresponds to the haemolymph activity previously observed for EST6 (Sheehan et al,
Figure 3.7 Native PAGE of homogenates of male ejaculatory ducts (ej.duct.), the carcasses of the males from which these ducts were dissected (m), and whole females (f), for each of the EST6-GS\textsuperscript{x} transformants and the wild type strain, Dm145. Dissections and homogenates were prepared as described in section 3.2.3.1.
1979; Morton and Singh, 1985). The same results are seen for the wild type strain Dm145.

It has previously been shown that 1.1 kbp of DNA 5' of the coding region of Est6 is adequate for correct expression of EST6 (M. Healy, pers. comm.). This analysis demonstrates that 860 bp of 5' DNA, the amount present in the mutant constructs described here (section 2.2.2.2), is sufficient for wild type adult expression.

3.3.2.2 Transfer to Females

EST6 is normally transferred from the male ejaculatory duct to the female's reproductive system within the first minute after the initiation of copulation (Richmond and Senior, 1981) and within the next 3 minutes is translocated to her haemolymph (Meikle et al, 1990). To test if transfer of the glycosylation site mutant proteins was impaired, males from the wild type and transformant stocks were mated to Est6° females. These females were left to age for various time intervals after mating and then their abdomens were removed. Male donated activity was detected in each of the female body parts by native PAGE (Figure 3.8).

For each of the glycosylation site mutants a large proportion of the male's total EST6 is transferred to the female in less than 3 minutes after mating initiation. This is highly repeatable between replicates. This is further evidence that non-glycosylated EST6 is secreted from the cells in which it is produced, as only extracellular enzyme can be transferred to the female with the male ejaculate.

As male donated EST6 is rapidly translocated from the female reproductive tract to her haemolymph (Meikle et al, 1990), comparison of the abdomens and thoraces of the mated females should indicate if this process is occurring normally for the glycosylation site mutant EST6. Activity restricted to the female reproductive tract would be localised in the abdomen while protein translocated to the haemolymph would be expected to circulate to the thorax and head as well. The time course of transfer also gives an indication of the longevity of the various mutant proteins in the female. It can be seen in Figure 3.8 that the majority of the EST6 activity remains in the abdomens of females mated to males of the wild type and mutant strains. This presumably reflects the relative amounts of haemolymph present in each of the individual body parts, rather than retention of EST6 in the female reproductive system, since Meikle et al (1990) have demonstrated that all the male donated wild
Figure 3.8 Native PAGE of individual mated males and females body parts. The strain of the male and the length of time after mating initiation before the female was frozen, is given. Est6⁰ females that do not produce any endogenous EST6 were used so that any EST6 activity present in the females is male donated. The whole male, the females's abdomen (Ab) and the female's thorax (Th) were homogenised as described (section 3.2.2.1) and loaded onto the gel in adjacent lanes as indicated. -GSx represents EST6-GSx and Dm145 is the wild type strain.
type EST6 is translocated to the female's haemolymph. It is also apparent that male donated activity is still detectable in abdomens of mated females 48 hours after mating. This is true of the wild type and each of the glycosylation site mutants, including the more thermolabile EST6-GS² enzyme.

Twenty minutes after mating, male donated EST6 activity is apparent in the thoraces of females mated to the wild type and to each of the glycosylation site mutants, excepting the multiple mutant EST6-GS¹²³⁴. As body parts from individual females were used in this experiment it is unclear whether this is due to a lack of transfer of the enzyme to the haemolymph of the female or merely reflects a lesser amount of EST6 being transferred to females by this particular transformant strain, such that the final levels of activity in the thorax are below the limits of detection by native PAGE.

For this reason, matings between males of the EST6-GS¹²³⁴ strain and Est6⁰ females were repeated and homogenates of 10 individuals' body parts analysed (Figure 3.9). Abdomens and thoraces/heads from females mated to Dm145 males were also analysed as a control. It is apparent from this experiment that male donated EST6-GS¹²³⁴ does in fact appear in the thorax of the females within 20 minutes after mating. This indicates that N-linked glycosylation is not a requirement for the translocation of male donated EST6 to the haemolymph of mated females. However, a subtle effect on the rate of translocation remains possible.
Figure 3.9 Native PAGE of mated males and female's body parts. The wild-type strain, Dm145, is given in the left panel and EST6-GS^{1234} is in the right panel. The numbers represent the total number of individual body parts pooled and electrophoresed. Female abdomens from one or two individuals and a dilution series of 1 to 5 female thoraces were analysed.
3.4 DISCUSSION

The aims of this analysis were to structurally characterise the glycosylation of EST6 and to determine the role that glycosylation plays in the function of EST6 by comparing the physiology of the glycosylated and the non-glycosylated proteins. The results presented in section 3.3 will be discussed in terms of (1) the structure of the EST6 glycoprotein and (2) the role that glycosylation might play in the function of EST6.

3.4.1 STRUCTURAL ANALYSIS OF THE GLYCOsyLATION OF EST6

Mutation of the potential N-linked glycosylation sites of EST6 from Asn-X-Ser/Thr to Gln-X-Ser/Thr has been effective in preventing the addition of carbohydrate to the protein, as evidenced by the inability of conA to bind EST6-GS\textsuperscript{1234}. The same amino acid substitution has previously been found to eliminate carbohydrate attachment at N-linked glycosylation sites of a number of glycoproteins (Dubé et al, 1988; Taylor and Wall, 1988; Hansen et al, 1988).

Each of the N-linked glycosylation site mutations has an effect on the size of EST6, demonstrating that each of these sites has carbohydrate attached in the wild type protein. However, the fourth glycosylation site appears to be glycosylated on only a proportion of molecules, resulting in two forms of EST6 which differ in size.

Comparison of the molecular weights of the glycosylated wild type EST6 with non-glycosylated mutant EST6 indicates that the attached carbohydrate contributes approximately 5 kd to the molecular weight of this protein. Each of the four side chains is between 1 and 2 kd molecular weight. The glycan attached to the second site may be slightly larger than those present on the other sites. Molecular weight estimation suggests this glycan is approximately 0.5kd larger than the others; however the inherent inaccuracy of molecular weight determination by SDS PAGE requires that this observation remain tentative.

Partial glycosylation of the fourth glycosylation site provides an explanation for EST6 appearing as a doublet on native and SDS denaturing PAGE. Independent evidence to support this is provided by the naturally occurring variant found to have an amino acid substitution in the fourth glycosylation site tri-peptide (Cooke, 1989). When this strain is analysed on native PAGE the slower mobility EST6 form is absent (section 3.3.1.1).
Of relevance to partial glycosylation at the fourth glycosylation site of EST6 is a survey by Gavel and von Heijne (1990) that revealed a significant increase in the frequency of non-glycosylated acceptor sites at the carboxy terminal end of glycoproteins. Several factors are possible contributors to this phenomenon. Firstly, studies on the glycosylation of nascent chains of various glycoproteins have demonstrated a requirement for a further 40 to 50 amino acids to be synthesised after the glycan acceptor site for the site to be accessible to the luminal surface of the rough endoplasmic reticulum (RER) and so glycosylated (Bergman and Kuehl, 1977; Glabe et al, 1980). This, in conjunction with the higher frequency of non-glycosylated sites close to the carboxy terminal of glycoproteins, has been interpreted to mean that glycosylation is more likely to occur when the nascent polypeptide chain is still associated with the polyribosomes and spanning the RER (Gavel and von Heijne, 1990). Secondly, the length of time the acceptor site is exposed to the oligosaccharyl transferase before the protein folds into its tertiary conformation has been suggested as a contributing factor, as sites towards the amino terminal end of the protein will be exposed to the transferase for a greater period of time than those near the carboxy terminal end (Gavel and von Heijne, 1990).

Partial glycosylation of the fourth glycosylation site of EST6 can be explained by its position in the primary amino acid sequence, which is 38 residues from the carboxy terminal of the protein. Being less than 40 amino acids from the carboxy terminal leads to brief exposure of the site to the transferase, or occasional disassociation of the nascent polypeptide from the RER membrane before transfer of the glycan. This then results in no oligosaccharide attachment to this site on a proportion of the EST6 molecules.

Digestion with endoglycosidases has demonstrated the glycan on the second acceptor site of EST6 to be sensitive to cleavage by both endoF and endoH. From the known substrate preference of these enzymes it can be deduced that this glycan is of the simple high mannose class. The resistance to cleavage of the other three glycans may be due to a more complex structure which these endoglycosidases are unable to recognise as a substrate. However, steric hindrance of the endoglycosidase due to some aspect of the protein's conformation at these glycosylation sites cannot be ruled out as an explanation. Digestion of the purified wild type EST6 by endoF in the native and the denatured states illustrates that hydrolysis at the second site of the native protein is
limited, presumably by steric hindrance, but upon denaturation the side chain is readily removed by the enzyme. Denaturation prior to digestion has been shown to minimise the effects of steric hindrance for other proteins (Tarentino et al, 1985).

While steric effects cannot be eliminated, a more complex glycan structure on the first, third and fourth sites may be the cause of the endoF/H resistance displayed by these glycans. Precedents exist for attachment of structurally different oligosaccharides to different sites of the same molecule (Anderson and Grimes, 1982). Determination of the structures of the oligosaccharides on the serine protease tissue plasminogen activator (tPA) (Penninca et al, 1983), and on porcine ribonuclease (Jackson and Hirs, 1970) have revealed three glycans attached to each of these proteins, one of the high mannose type and two of the complex type. Only isolation and direct structural analysis of the oligosaccharides attached to EST6 will establish conclusively any structural differences among the glycans attached to EST6.

Mutation of the second acceptor site has the greatest effect on native PAGE electrophoretic mobility (section 3.3.1.1). As this electrophoretic technique separates on the basis of both charge and conformation, differences in mobility may be due to an alteration of either of these properties. The only known charged saccharide found in glycoproteins is sialic acid, which is not present in insects (Warren, 1963), including D.melanogaster (Corfield and Schauer, 1982). This suggests that a charge change due to the loss of the glycan is not responsible for the large difference in electrophoretic mobility and leaves conformation as the most likely explanation. The glycan attached to the second glycosylation site is slightly larger and of a different structure to the other glycans attached to EST6. As the glycans are presumably on the surface of the molecule, this difference in conformation may be detected by native electrophoresis and so explain the large difference in Rf produced by this mutation.

In summary, EST6 exists in two molecular forms that differ in their glycosylation states. The first form has four asparagine linked carbohydrate side chains while the second has three. This difference is attributable to partial glycosylation of the fourth acceptor site due to its position near the carboxyl terminal of the protein. Each of the glycans is 1 to 2 kd in size and at least one, the glycan attached to the second glycosylation site, is of the high mannose class of oligosaccharide and may be slightly larger than the other glycans present on EST6.
3.4.2 FUNCTIONAL ROLE OF THE EST6 GLYCANS

Several possible roles for the oligosaccharides of secreted glycoproteins have been proposed and instances where each of these apply are known (section 1.6). Many of these possibilities may apply to EST6. The results presented in section 3.3 are discussed below with a view to addressing the role of glycosylation in the function of EST6.

A common finding when glycosylation of glycoproteins is prevented, whether by mutation of glycosylation sites or treatment of cells with inhibitors of the glycosylation process, is impaired secretion of the resulting non-glycosylated protein (Dorner et al, 1987; Taylor and Wall, 1988). This is because the non-glycosylated protein does not attain its correct conformation and as a consequence is retained in the RER (Dorner et al, 1987). This retention is due to a RER-localised 77kd protein which has been found associated with incorrectly folded proteins (Dorner et al, 1987; Copeland et al, 1986; Gething et al, 1986). The presence of non-glycosylated EST6 in the soluble fraction of detergent-free homogenates, and the observation of transfer of the protein to females during mating, both demonstrate that non-glycosylated EST6 is secreted and so is attaining its correct conformation in the RER. Similar results have been observed for a serine protease, tissue plasminogen activator (tPA) (section 1.6.1).

Analysis of thermostability provides a test of conformational stability, as less energetically stable structures would be expected to unfold more rapidly than stable structures at a particular temperature. The thermostability of the non-glycosylated mutant EST6 demonstrates that this mutant is attaining a tertiary structure as energetically stable as the glycosylated protein. Further evidence in support of conserved tertiary structure is that the glycosylation site mutant proteins retain hydrolytic activity. As the proteins with mutated glycosylation sites have not been purified, and a specific antibody for the quantitation of EST6 molecules in crude homogenates is not available, the specific activity of these proteins cannot be determined. However, it is apparent from native PAGE and staining with βnaphthylacetate that each of the glycosylation site mutants has substantial hydrolytic activity. The major reason for choosing Gin as the amino acid to substitute for Asn in the putative acceptor sites was to minimise disruption of the proteins tertiary
structure (section 3.1) and these observations indicate that this aim has been achieved.

Mutation of the the second glycosylation site is associated with an increase in thermodelatability of EST6. If this is due to a direct destabilising effect of the Gln in place of Asn at this position, it would also be expected to destabilise the EST6-GS\textsuperscript{1234} protein, which also carries the second glycosylation site mutation. As this does not occur, one or more of the other mutations in EST6-GS\textsuperscript{1234} is presumably compensating for the decreased stability. An interaction of the glycan attached to the second glycosylation site with one or more of the other glycans may be the cause of this phenomenon. Thermostability analysis of EST6 with this mutation in combination with each of the other acceptor site mutations individually would be necessary to determine with which mutation(s) this interaction occurs.

As non-glycosylation of EST6 has no detectable effect on the secretion or in vitro stability of the enzyme, an in vivo physiological difference was sought. EST6 is exposed to several different in vivo environments and may also be required to recognise particular target tissues in the female in order to fulfill its physiological role (section 1.3.2). In particular, this study concentrates on the effect of glycosylation on the in vivo longevity and translocation to the female haemolymph of male donated enzyme.

Male donated non-glycosylated and wild type (Dm145) EST6 activity was detected in homogenates of females for two days after mating using native PAGE. Two previous studies, both using Dm145 as the wild type strain, have detected wild type male donated enzyme in mated females for a maximum of 1 to 2 hours (Richmond and Senior, 1981) and 12 hours (Meikle et al, 1990), using spectrophotometric activity assays and Western blotting respectively as the detection methods, although Meikle et al (1990) detected male donated activity in females by Western blots 96 hours after mating when males of a strain with higher EST6 activity were used. The greater longevity of Dm145 EST6 detected in this present study may reflect the sensitivity of native PAGE as the EST6 activity detection method. The presence of male donated non-glycosylated EST6 in females two days after mating suggests that glycosylation of the enzyme does not significantly affect the longevity of male donated EST6 in females.

A final possibility investigated in this study is that glycosylation of EST6 facilitates the translocation of the enzyme from the female reproductive tract to her
The observation of male donated non-glycosylated EST6 in mass homogenates of thoraces of mated females 20 minutes after mating clearly demonstrates that the N-linked glycans of EST6 are not an absolute requirement for the enzyme to enter the female haemolymph from the reproductive system. However, the possibility of a more subtle effect on transport of the enzyme cannot be excluded on the basis of these experiments. Non-glycosylated enzyme could not be detected in the thoraces of individual females after mating whereas glycosylated enzyme was apparent in female thoraces in this experiment. This may be attributed to a lesser amount of enzyme initially transferred to the female or to a genuine effect on the rate of translocation of the enzyme. To resolve this issue a more elaborate experiment would be required involving a larger number of time points after mating and preferably utilising a specific and more quantitative measure of the amount of EST6 present, for example use of a monoclonal anti-body to the protein to detect EST6 in crude homogenates. The detailed kinetics of translocation of male donated EST6 to the female haemolymph for the wild type and glycosylation deficient enzymes could then be compared and any subtle differences established.

No conclusive evidence has been presented here to suggest a functional role for the carbohydrate component of EST6. Therefore, the possibility that the glycans of EST6 have no functional role cannot be eliminated. More sensitive techniques may detect differences in the transport or translocation of the enzyme and the possibility of a role for the glycans once the enzyme has been translocated to the female's haemolymph has not been investigated at this stage. Such analysis awaits the isolation of a monoclonal anti-body specific for EST6. The final proof of a role for the glycans of EST6 is to determine if the glycosylation defective enzymes are still capable of causing altered mating behaviour of the female, as has been observed for the wild type enzyme. Such a test of biological function will be the most conclusive indicator of whether EST6 function is affected by the glycosylation site mutations.

3.4.3 A NEUTRAL HYPOTHESIS FOR THE ROLE OF GLYCOSYLATION OF EST6

If glycosylation plays no role in the function of EST6, the question is raised; why is the enzyme glycosylated at all? It has been proposed that glycosylation represents a primitive evolutionary device to enable secretion which has since been superseded, rendering carbohydrate attachment unnecessary for secretion to occur in
some cases (Eylar, 1965). Since this was proposed, a large number of examples where no biological role can be detected for the carbohydrate components of glycoproteins have been forthcoming (Olden et al, 1982 and references therein). Yet examples where glycosylation is essential for biological activity also exist (section 3.1), indicating that this is an over simplification.

An alternative hypothesis is that glycosylation is important for the secretion, stability or biological role of some glycoproteins but in other cases simply happens by accident as the protein journeys through the RER and golgi apparatus. As this is the only route for secretion to the extracellular environment (for review see Farquhar, 1985), any N-linked glycosylation site consensus sequence present in a secreted protein will be prey to carbohydrate attachment by oligosaccharyl transferase, whether or not the protein requires such attachments for biological function. In cases where the carbohydrate component is of no functional importance, the occurrence of potential N-linked glycosylation sites in the protein should be random, assuming their presence conferred no particular disadvantage. With this in mind, the expected frequency due to chance occurrences of the sequence Asn-X-Ser/Thr in the esterases was determined. The amino acid composition of all the available esterase sequences was used to increase the sample size. The amino acid composition of the eighteen esterases sequenced to date (see section 1.2.2 for references) was calculated using the Composition programme from the GCG sequence analysis package (Devereux et al, 1984). The expected frequency (F) of chance occurrences of the above glycosylation site tri-peptide sequence is given by

\[ F = \text{freq.}(\text{Asn}) \times (\text{freq.}(\text{Ser}) + \text{freq.}(\text{Thr})) \times (1 - \text{freq.}(\text{Pro})) \].

(A term accounting for the frequency of proline is included because the X in the tripeptide can be any amino acid excluding proline.) This gives a frequency of

\[ F = 0.0468 \times (0.0692 + 0.0542) \times (1 - 0.0635) = 0.0054 \pm 0.0032 \]

(where the standard error is calculated using the normal approximation to the binomial distribution and for EST6, n=523)

This corresponds to an average of one occurrence in every 185 amino acids. For EST6, the tri-peptide Asn-X-Ser/Thr would be expected to occur about three times purely by chance. The observed number of sites (4, F=0.0076) is not significantly different to this.
It would be expected that carbohydrate attachment to sites that are buried within the interior of the protein, after it folds into its tertiary structure, would prevent the protein attaining its correct conformation through steric hindrance by the oligosaccharides. Presumably, mutations that lead to glycosylation at such unfavourable sites would be strongly selected against as they would destroy the activity of the protein in question. This would result in a lower frequency of potential N-linked glycosylation site consensus sequences than expected due to chance alone. This qualification is difficult to take into account as the three dimensional structure of esterases has not been determined. However, the proportion of residues buried in the tertiary structure of 28 different proteins has been calculated from their crystal structures on the basis of amino acid side chain surface area accessible to solvent (Janin et al, 1978). For Asn, the amino acid to which the carbohydrate is attached in the acceptor site tri-peptide, the fraction buried in the interior of the protein was calculated to be 0.22. Assuming this same proportion of buried Asn residues applies to esterases the frequency of acceptor site tri-peptides where the Asn residue is not buried will be

\[ F = 0.0054 \times (1 - 0.22) = 0.0042 \pm 0.0028 \]

This again is not significantly different from the frequency of 0.0076 observed for EST6. In fact, the frequency of buried Asn residues in the esterases would have to be 0.40 for the frequency of observed acceptor site tri-peptides in EST6 to be significantly greater than that expected by chance.

Hence it seems feasible that four potential N-linked glycosylation sites whose asparagine residues are accessible to the surface of the protein could occur by chance in an esterase the length of EST6. Proving such a hypothesis is difficult, for even the most rigorous laboratory experiments may fail to find a difference in the properties of non-glycosylated EST6, yet glycosylation may still confer a selective advantage to the enzyme in the field. However, it is a possibility that should not be neglected. It also raises a very interesting evolutionary question, as unnecessarily glycosylating proteins would be expected to be disadvantageous in terms of energy expenditure. As research into the functional significance of glycosylation of secreted glycoproteins advances, it will be interesting to see how frequently the lack of assignment of definite functions to the carbohydrate component occurs and if any generalisations can be drawn.
3.4.4 CONCLUSIONS

This study has established that each of the four potential N-linked glycosylation sites in the wild type EST6 protein do carry attached carbohydrate. The fourth glycosylation site, however, is not utilised on all the EST6 molecules, resulting in two molecular forms of the enzyme that differ in the amount of attached carbohydrate. The glycan attached at the second glycosylation site is of the high mannose type while the other three sites appear to carry complex type oligosaccharides, although direct chemical determination or a lectin binding survey is required to confirm the structure of these glycans.

Glycosylation of EST6 is not required for secretion of the enzyme from the site of production, nor for transfer of the enzyme to the female from the male's ejaculatory duct during mating. Non-glycosylated EST6 also passes successfully from the female reproductive system into her haemolymph. There is no significant effect on the conformation of the protein attributable to non-glycosylation, although an anomaly does occur where mutation of the second site results in greater thermostability. This effect is lost when the second site mutation is present with the other three glycosylation site mutations.

The possibility has been raised that the N-linked glycans on EST6 perform no function and are only glycosylated as a consequence of this protein being secreted. The number of glycosylation sites present in this protein is not significantly different to that due to chance. However, a biological role for the carbohydrate component of EST6 cannot be ruled out on the basis of this study. In particular, the effects on the mating behaviour of females mated to EST6-GS1234 producing males in comparison to males producing glycosylated EST6 must be investigated.
CHAPTER 4. DISCUSSION

The main aims of this thesis were to determine the functional role of two structural features of the EST6 protein; the residue at position 187 adjacent to the reactive Ser, and the N-linked glycosylation sites of the protein. The results of the analysis of each of these structural features will be summarised, then discussed with reference to the conservation and divergence of these structural features in other serine esterases. Directions for future research will also be considered.

4.1 THE SERINE ACTIVE SITE

The results presented in chapter 2 demonstrate that residue 187 of EST6 plays a role in the determination of the enzyme's substrate preference. Substitution of His 187 with the ionic Glu results in a decreased energy of interaction with βnaphthyl esters, an even greater decrease in interaction energy with p-nitrophenylacetate, and the gain of the ability to hydrolyse acetylthiocholine. These characteristics can best be explained by direct electrostatic interactions, which are realised in the transition state complex, between residue 187 and the alkyl group of the substrate. In contrast to the Glu substitution, the presence of Gln in place of His at position 187 does not alter the specificity of the enzyme. This can be explained by the fact that both His and Gln have polar side chains which can interact with the substrate in a similar manner. EST6H187E also displays a different pH profile to the wild-type enzyme. EST6H187E displays an increase in activity with increasing pH, which is consistent with the utilisation of substrate hydroxyl ions in catalysis at alkaline pH. However, the precise mechanism by which this occurs is not clear.

The demonstration of a role for His187 in EST6 in the determination of substrate specificity, and the presence of Glu in the equivalent position in the majority of other serine esterases, suggest that this amino acid substitution may be responsible for differences in substrate specificity between them. In particular, the ability of EST6H187E to hydrolyse acetylthiocholine, whereas the wild type enzyme does not hydrolyse this substrate, implicates the substitution residue 187 as an important structural difference between the active sites of the cholinesterases and EST6.

The role of residue 187 in EST6 and its equivalent in the cholinesterases and JHE will be discussed in the next section. This will be followed by a comparison
between EST6 from *D. melanogaster* and its homologues from other species of *Drosophila*.

### 4.1.1 Comparison with Distantly Related Esterases

The cholinesterases and EST6 have diverged greatly in their primary structures, having only 30% amino acid sequence identity (Myers *et al.*, 1988). However, the substrate binding cleft of EST6 must have some similarity to those of the cholinesterases, as a single amino acid substitution enables EST6 to hydrolyse acetyltiocholine.

The active site of the AChEs has been proposed, on the basis of biochemical data, to consist of two sub-sites. The first is the esteratic site, which contains the reactive Ser and where the hydrolysis of ester bonds occurs (Rosenberry, 1975). The second is the anionic site, which binds the positively charged choline group of the substrate (Rosenberry, 1975) and is believed to include at least one carboxyl group (Quinn, 1987). When the Glu adjacent to the reactive Ser of *Torpedo californica* AChE is mutated to His, activity is lost, but when it is mutated to Gln activity is reduced and substrate inhibition is not observed (Gibney *et al.*, 1990). These observations, in conjunction with the data presented on EST6H 187E, suggest that the Glu adjacent to the reactive Ser provides one of the carboxyl groups of the anionic site.

The apparent $K_m$ of EST6H 187E for acetyltiocholine (2mM) is approximately 25 fold greater than that observed for *D. melanogaster* AChE (79uM) (Gnagey *et al.*, 1987). Although this is a substantial difference, it does indicate that the EST6H 187E enzyme has reasonable affinity for this substrate. However, the retention of 50% of wild type activity when the residue adjacent to the reactive Ser is replaced by Gln in *T. californica* AChE (Gibney *et al.*, 1990), and the lack of detectable acetyltiocholine hydrolysis by the EST6 mutant with Gln in this position, suggest that other residues present in AChE (but absent in EST6) are also involved in the binding of this substrate.

Two other regions of the primary structure of the cholinesterases have been implicated in the formation of the anionic site. Sequence analysis of the atypical form of human butyrylcholinesterase (BChE), which has an apparent $K_m$ for succiny1choline two orders of magnitude higher than the wild type enzyme, identified the substitution of Asp at position 70 in BChE with Gly as being responsible for this
difference (McGuire et al, 1989). This residue is present in an equivalent position in *T.californica* and bovine serum AChE, which suggests that this Asp residue also forms part of the anionic site of the AChEs (Chatonnet and Lockridge, 1989). In another study, on AChE from the electric eel (*Electrophorus electricus*), a peptide which bound to quaternary ammonium compounds was isolated and found to have the sequence Gly-Ser-X-Phe (Kieffer et al, 1986). An equivalent sequence of Gly-Ser-Phe-Phe is present in *T.californica* AChE and a two in four match with this sequence is found in *D.melanogaster* AChE, human BChE and bovine serum AChE (Chatonnet and Lockridge, 1989). Both an Asp at position 70 (see table 4.1) and a sequence similar to this quaternary ammonium binding peptide are absent in EST6. The absence of these residues may be responsible for the low level of acetylthiocholine hydrolytic activity of EST6^{H187E} in comparison to the true cholinesterases.

It is clear from the mutational studies presented here and by Gibney et al (1990) that the Glu residue adjacent to the reactive Ser of the cholinesterases is a major component of the anionic site which binds choline, but not the only residue in the substrate binding pocket necessary for the efficient hydrolysis of choline esters.

Another esterase which does not have Glu adjacent to the reactive serine is JHE, which has Gln in the equivalent position (Hanzlik et al, 1989). This enzyme hydrolyses juvenile hormone (JH), a lipid substrate with an acyl group 15 to 17 carbon atoms long. Both EST6 and EST6^{H187Q} are incapable of hydrolysing this substrate (section 2.3.3.4). This is consistent with the observation that EST6 has a preference for substrates with short acyl carbon chains and is incapable of hydrolysing naphthyl esters with acyl chains longer than 5 carbon atoms (White et al, 1988). The inability of EST6 to recognise JH as a substrate then, is not necessarily a consequence of having His at position 187, which influences the preferred alkyl group of the substrate, but a result of the preferred acyl group of the enzyme. The regions involved in binding the acyl group of esters have not been identified for these enzymes, so the structural differences responsible for the difference in preferred acyl chain length between EST6 and JHE cannot be identified at this stage. Hence, the role in JH hydrolysis of the Gln adjacent to the reactive serine in JHE remains unknown. Only mutational analysis using JHE will answer this question.
4.1.2 COMPARISON WITH EST6 HOMOLOGUES FROM OTHER DROSOPHILA SPECIES

Esterase S (ESTS) from *D. virilis* and EST6 from *D. melanogaster* are closely related both structurally and functionally, but have amino acids with different physicochemical properties at position 187. ESTS has phenylalanine (Phe), which has a large hydrophobic side chain, at position 187 (Sergeev *et al*, 1989), while EST6 has the smaller, basic His in this same position.

ESTS has 50% sequence identity to EST6, and plays a somewhat similar physiological role, being expressed predominantly in the ejaculatory bulb and transferred to females during mating (Korochkin *et al*, 1976). Whether EST6 and ESTS are paralogous or orthologous is unresolved; genetic evidence suggests that the genes for these enzymes are on non-homologous chromosomes (Yenikolopov *et al*, 1983), while other evidence, including their degree of sequence identity, their similar tissue distribution and the fact that the ESTS gene forms part of a tandem duplication of esterase genes with a similar molecular structure to that observed for Est6/EstP (Sergeev *et al*, 1989; section 1.3.3), indicate that ESTS is at least closely related, if not orthologous, to EST6.

It is unexpected that an esterase with a similar expression pattern and possibly a similar role in mating behaviour should have an amino acid substitution at a position which, as shown in chapter 2, influences the substrate preference of the enzyme. However, the *in vivo* substrates of EST6 and ESTS are not necessarily the same. There are differences in the mechanics of the transfer of these enzymes to females during mating; EST6 is secreted into the ejaculatory duct and transferred with the ejaculate to the female, from where it is subsequently translocated to her haemolymph (section 1.3.2). In contrast, ESTS is microsomal (Korochkin *et al*, 1973) and is localised in a waxy plug which is transferred to females during mating, where it has been proposed to play a role in the insemination reaction by inducing swelling of the vaginal cavity (Korochkin *et al*, 1976). Hence, while both EST6 and ESTS are implicated in reproductive biology and are transferred to females during mating, their roles in reproduction are likely to be different. Consequently, the substrate(s) that each enzyme is required to hydrolyse in the performance of their physiological role may also be dissimilar.
The *in vivo* substrates of these enzymes need to be determined to resolve this question. It would also be of interest to mutate His 187 in EST6 to Phe and determine the effect of this on the enzyme's preference for various artificial substrates. Given the role of the amino acid at this position in determining the preferred alcohol leaving group of the substrate (chapter 2), it may be predicted that introduction of the hydrophobic side chain of Phe would enhance hydrolysis of substrates with hydrophobic alcohol leaving groups such as naphthyl esters.

In contrast to ESTS, esterases orthologous to EST6 from species more closely related to *D.melanogaster*; namely EST6 from *D.simulans*, *D.mauritiana* (J.Karotam pers.comm.), *D.yakuba*, *D.erecta* (E van Papenrecht, pers.comm.) and EST5B from *D.pseudoobscura* (Brady and Richmond, in press), invariably have His187 in the Ser active site. Similarly, the duplicate esterase adjacent to EST6, ESTP, and the two other esterases in the EST5 cluster of *D.pseudoobscura*, EST5A and EST5C, also have His adjacent to the reactive Ser. These species are more closely related to *D.melanogaster* than to *D.virilis*; all belong to the subgenus *Sophophora* while *D.virilis* belongs to the subgenus *Drosophila*. However, the enzymes in *D.pseudoobscura*, *D.yakuba* and *D.erecta* are not expressed in significant quantities in the male ejaculatory duct (Morton and Singh, 1985) and so may not be involved in the reproductive behaviour of these species. As EST6 is expressed in the haemolymph of both sexes of *D.melanogaster*, *D.simulans*, *D.mauritiana*, *D.erecta*, *D.yakuba* and *D.pseudoobscura* (Brady and Richmond, in press; section 1.3.3), this may indicate a common physiological role for this enzyme in the haemolymph.

This raises the question of whether the presence of His instead of Glu in the serine active site of EST6 is required for the hydrolysis of the *in vivo* substrate of the enzyme. If this amino acid is required in the substrate binding pocket for the performance of the enzyme's physiological role, it would be expected to be the role performed in the haemolymph. This is because expression in the haemolymph is believed to be the ancestral state (Morton and Singh, 1985) and the esterases orthologous to EST6 that display only haemolymph expression all have His at position 187. This suggests that the occurrence of this amino acid at position 187 pre-dates the recruitment of EST6 to a role in mating behaviour in *D.melanogaster* and its sibling species.
The role that EST6 plays in the haemolymph is not known. The EST6 null mutant of *D. melanogaster* is viable under laboratory conditions and displays no unusual morphological or behavioural characteristics, apart from the differences in mating behaviour (described in section 1.3.3). A possible role for the enzyme is in general detoxification of compounds encountered by the insect. If this is the case, the enzyme would be required to have a broad substrate range. Alternatively, EST6 may play a specific but subtle role in the metabolism of the fly which, while not essential, is of some advantage to organisms in the field; in this case the enzyme may only be required to hydrolyse a specific substrate. Only determination of the *in vivo* substrate(s) of EST6 will show if the His at position 187 is essential for EST6 to carry out its physiological role in the haemolymph.

On the other hand, it is possible to test whether His 187 is essential for the effects of EST6 on mating behaviour. Use of the transformed strains described in this thesis to determine whether EST6H187Q and EST6H187E have similar effects on mating behaviour as wild type EST6 would give an indication of whether the altered substrate specificities of these enzymes affect hydrolysis of the *in vivo* substrate of EST6.

### 4.1.3 FUTURE WORK

A significant contribution to the analysis of the structure/function relationships of EST6 would be the solving of a high resolution crystal structure. This would greatly assist in the identification of amino acid residues present in the substrate binding pocket and further our understanding of the catalytic mechanism of the serine esterases by revealing what amino acids are within hydrogen bonding distance and so capable of interacting with the reactive serine. It would also be of interest to compare the three dimensional structures of widely divergent members of this multi-gene family. As the degree of sequence identity between distantly related members is only 25-30%, the amount of tertiary structure similarity displayed by such divergent proteins would be of great benefit to our understanding of the minimum sequence similarity needed to fulfill particular structural requirements. The recent availability of a number of esterase clones and the use of efficient expression systems, such as baculovirus expression vectors, should make these achievements merely a matter of time.
In terms of the mutant proteins constructed and characterised in this study, a high resolution crystal structure would be of great benefit. Such information would enable direct structural verification of the conclusions concerning the role of residue 187 in the determination of substrate specificity. Also of interest is the pH dependence of EST6H187E, which is indicative of an alternative catalytic mechanism involving solvent hydroxyl ions. The determination of the tertiary structure of this mutant protein would establish if an altered conformation changing the relative configurations of residues involved in catalysis is responsible for the anomalous pH dependence.

In the absence of the EST6 tertiary structure, a number of biochemical analyses may be performed which would lead to a greater understanding of the role of residue 187. Testing mutant EST6 proteins with a large number of different amino acid substitutions at this position with a greater diversity of substrates will further refine the precise nature of enzyme/substrate interactions in the serine esterases. Similarly, the anomalous pH dependence could be investigated by determining the response to alkali pH of a wide range of EST6 mutants with different substitutions at this position.

Assessment of the importance of His 187 to the performance of EST6's function may be tested by using germ line transformants producing EST6H187E and EST6H187Q in mating experiments. Furthermore, if a range of mutant EST6 proteins with substitutions at position 187 were generated, and the effects of those which alter in vitro substrate specificity on the mating behaviour of the females were determined, some insight may be gained as to the nature of the in vivo substrate. The presence or absence of a mating effect could be correlated with the physicochemical properties of preferred artificial substrates and so the nature of the physiological substrate become apparent.

4.2 N-LINKED GLYCOSYLATION

Mutation of the N-linked glycosylation sites of the enzyme has established that each of the four consensus acceptor sites have attached oligosaccharides in the wild type protein. The fourth glycosylation site, however, is not utilised on all of the EST6 molecules, resulting in two molecular forms of the enzyme which differ in the amount of attached carbohydrate. The lack of carbohydrate attachment at this site on a
proportion of the EST6 molecules is most likely a consequence of its proximity to the carboxy terminal of the protein. The second glycosylation site has a high mannose oligosaccharide attached while the other sites may have complex type oligosaccharides. Glycosylation is not required for secretion, transfer of ejaculatory duct EST6 to the female during mating, or translocation of the male donated activity to the female’s haemolymph from her reproductive system. The EST6 proteins with mutated glycosylation sites also retain hydrolytic activity and, with one exception, the mutations do not affect the enzymes thermostability. The exception is the second glycosylation site mutation which increases the thermostability of the protein when present alone, but not when present with the other three glycosylation site mutations. The possibility has been raised that the N-linked glycans of EST6 perform no function and are only glycosylated as a consequence of the protein being secreted. The number of glycosylation sites present in EST6 is not significantly different to the number expected by chance alone, which supports such a contention. However, a biological role of a more subtle nature cannot be ruled out on the basis of this study.

4.2.1 COMPARISON WITH DISTANTLY RELATED ESTERASES

The glycosylation of all four of the N-linked glycosylation sites of EST6 is consistent with the observed frequency of 90% for the utilisation of potential N-linked glycosylation sites in proteins known to be N-glycosylated (Gavel and von Heijne, 1990). Other serine esterases where the sites of glycosylation have been determined demonstrate a high frequency of utilisation. All four of the N-linked acceptor sites in Torpedo californica AChE (MacPhee-Quigley et al, 1986), both the sites in rabbit liver microsomal esterase (Korza and Ozols, 1988) and nine of the ten sites in human BChE (Lockridge et al, 1987) are known to have carbohydrate attached. Only one of these proteins, T. californica AChE, has a glycosylation site close to the carboxy terminal end. The fourth site of this enzyme is 40 amino acids from the carboxyl terminal but it is not clear whether this site is glycosylated on all molecules of the protein (MacPhee-Quigley et al, 1986). The site not glycosylated in BChE forms part of the sequence Asn-Asn-Ser-Thr..., with carbohydrate being attached to the second, but not the first Asn residue (Lockridge et al, 1986). The lack of glycosylation at the first site in this sequence may be due to steric hindrance of the oligosaccharyl transferase once an oligosaccharide has been attached to the second site
(Gavel and von Heijne, 1990). Therefore, the esterases appear to be no exception to the general observation that the majority of potential N-linked glycosylation sites in glycoproteins are glycosylated.

Of the N-linked glycosylation sites of these esterases, only one site shows any degree of conservation, being present in the same position in human BChE, rabbit liver microsomal esterase and *T. californica* AChE, but not in EST6 (table 4.1). This conserved site is part of a sequence that displays a high degree of similarity beyond the N-linked glycosylation site as well, which suggests that this region is conserved for reasons other than the preservation of the glycosylation site. None of the other N-linked glycosylation sites of these esterases are in comparable positions.

<table>
<thead>
<tr>
<th></th>
<th>EST6</th>
<th>Human BChE</th>
<th>Rab.Micro.EST</th>
<th>Torpedo AChE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>W S D I F D A T K t P v a C I Q w d Q f</td>
<td>W S D I W N A T K Y A N s C c Q N i D Q</td>
<td>W S h V k N t T s Y P p m C s s D a v s</td>
<td>W S g V W N A S t Y P N n C q Q y v D E</td>
</tr>
</tbody>
</table>

**Table 4.1 Comparison of the conserved glycosylation site of selected esterases**

The glycosylated Asn residues (N) are underlined. Amino acids conserved between all four sequences are in bold type and conservative substitutions are in upper case letters. Where there is no apparent conservation, the amino acids are in lower case. The numbering is for EST6. The conserved Cys residue forms a disulphide bridge in *T. californica* AChE and probably the other esterases as well (section 1.3.4.2). The Asp residue mutated in the atypical form of human BChE and believed to form part of the anionic site in the cholinesterases is at position 71 (see section 4.1.1). Rab.Micro.EST=rabbit microsomal esterase.

A functional role for the N-linked glycans of these esterases has not been investigated. However, a possible role for the glycans of human BChE is suggested by the observation that specific receptors for asialoglycoproteins exist in the livers of mammals (Ashwell and Harford, 1982). These receptors recognise galactose residues...
which are made terminal on the oligosaccharides of glycoproteins by the removal of sialic acid. As a consequence, desialylated serum glycoproteins have drastically reduced survival times in the circulation in comparison to native forms of the same proteins (Ashwell and Harford, 1982). Similar receptors for mannose and N-acetylglucosamine have also been identified. These receptors are believed to be a mechanism for the clearance of biologically active serum glycoproteins from the circulation once their required function has been performed (Ashwell and Harford, 1982). Human BChE has terminal sialic acid residues on its N-linked oligosaccharides and is a serum protein (Lockridge et al, 1986) but it has not been demonstrated that removal of these terminal sialic acids reduces the survival time of this enzyme in the circulation. However, the large number of N-linked glycosylation sites in this enzyme, which is far greater than would be expected purely by chance (see section 3.4.3), may suggest glycosylation is of some importance to this protein. A mechanism for the clearance of haemolymph glycoproteins in insect species has not been demonstrated. However, it would be of interest to test if non-glycosylated EST6 does persist longer in the haemolymph than the wild type enzyme, as a similar mechanism for the clearance of haemolymph glycoproteins may exist.

4.2.2 COMPARISON WITH EST6 HOMOLOGUES FROM OTHER DROSOPHILA SPECIES.

Comparison of the positions of potential N-linked glycosylation sites in genes homologous to EST6 from other Drosophila species reveals a great deal of variation in their positions in the primary sequence. All of these esterases have between two and four sites, none of which are shared between all the EST6 and ESTP homologues (Table 4.2).

Brady and Richmond (1990) have proposed a phylogeny for the triplicated esterases at the EST5 locus in D.pseudoobscura on the basis of sequence similarities. This proposes that EST5B is orthologous to EST6 while EST5C represents a duplication of EST5B in the D.pseudoobscura lineage that occurred after the species divergence which led to D.melanogaster and D.pseudoobscura. EST5A is orthologous to ESTP. This is supported by the observation that two glycosylation sites are conserved among EST5B, EST5C and EST6 (amino acids 399 and 485). However, there are no glycosylation sites common to EST5A and ESTP, which Brady and Richmond (1990)
propose are orthologous, while EST5A, EST5B and EST5C all share a site (amino acid 94) that is not present in EST6 or ESTP. For this to have occurred the site at 94 must have been lost independently in both EST6 and ESTP or gained independently in EST5A and EST5B. ESTS and EST6 from D. yakuba and D. erecta share this site at position 94 with EST5A, EST5B and EST5C, which may suggest that the presence of this site is the ancestral state and it has been lost independently in both EST6 and ESTP.

<table>
<thead>
<tr>
<th>Species/Esterase</th>
<th>Position of Glycosylation Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>melanogaster EST6</td>
<td>21 399 435 485</td>
</tr>
<tr>
<td>ESTP</td>
<td>58 187 435</td>
</tr>
<tr>
<td>simulans/mauritiana EST6</td>
<td>21 399 435</td>
</tr>
<tr>
<td>yakuba</td>
<td>94 399</td>
</tr>
<tr>
<td>erecta</td>
<td>94 399</td>
</tr>
<tr>
<td>pseudoobscura EST5A</td>
<td>73 94</td>
</tr>
<tr>
<td>EST5B/C</td>
<td>94 399 485</td>
</tr>
<tr>
<td>virilis ESTS</td>
<td>94 380</td>
</tr>
</tbody>
</table>

Table 4.2 Potential N-linked Glycosylation Sites in EST6 and its Homologues
The numbers represent the position in the primary amino acid sequence of the Asn residue in the acceptor site tri-peptide, according to the numbering of EST6 from D. melanogaster. The alignment of EST6, ESTP, EST5A, EST5B and EST5C sequences requires one 2 amino acid gap in EST6 (Brady and Richmond, in press) while alignment of the ESTS sequence can be made with a single 1 amino acid gap in ESTS (GCG gap programme, gap weight,3: length weight,0.1). The data for D. simulans and D. mauritiana are from J. Karotam (pers.comm.) and for D. yakuba and D. erecta from E. Van Papenrecht (pers.comm.).

Another point worthy of note is that each of these Drosophila esterases has four or fewer N-linked glycosylation sites. This small number of sites lends support to the proposal that N-linked glycosylation of EST6 may be neutral to selection, as has been suggested in chapter 3. Of course, the analysis of the effect that donation of non-glycosylated EST6 to females has on mating behaviour needs to be determined before a
role for glycosylation in this function can be ruled out, and a role in some aspect of the haemolymph function of the enzyme is also a possibility. It is also feasible that glycosylation of EST6 may play no detectable role in the functioning of EST6 in the laboratory but is still of benefit to the organism in the field. If this is the case, the lack of conservation of the locations of the N-linked glycosylation sites in the primary sequence may reflect that the presence of a site somewhere in the primary sequence, not its precise location, is important for the fulfillment of the functional role. Theoretically, in vitro mutagenesis experiments could be carried out to test such possibilities, although the subtle nature of any such role of the glycans will make it difficult to detect differences.

4.2.3 FUTURE WORK

A major priority for future work on the glycosylation of EST6 is a detailed analysis of the in vivo function of the non-glycosylated mutant enzyme. The role of glycosylation of EST6 in the transfer of male EST6 to the female and the effects on mating behaviour attributable to EST6 will best be determined by using the non-glycosylated EST6 mutant transformants in mating experiments. Once it has been established that glycosylation does or does not prevent the expected effects on mating behaviour, the reason for this can be investigated. The rate of translocation of male donated EST6 to the female haemolymph needs to be thoroughly characterised to establish whether glycosylation is involved in this process. A technical problem associated with such experiments is the detection of small quantities of EST6 in a specific manner. The isolation of a monoclonal antibody specific for the protein would overcome this, as ELISA assays using a specific antibody will assure specific and sensitive detection of the protein in small quantities of particular tissues.

A further analysis would be to perform in situ antibody staining of tissue sections of EST6 null females mated to males producing glycosylated and non-glycosylated EST6. If there is a particular target tissue for male donated enzyme in the female, accumulation of EST6 at this site may be detected in this manner. If glycosylation plays a role in such a process, accumulation of non-glycosylated EST6 at any such target tissue would be expected to be reduced.

Another possibility for a functional role of the EST6 glycans is in the survival time of the enzyme in the haemolymph. The observation that glycans attached to serum
glycoproteins in mammals can mediate their survival time in the circulation raises the possibility that a similar mechanism may exist in insects (section 4.2.1). An initial experiment to test this possibility is to determine if male-donated, non-glycosylated EST6 survives for longer than the wild-type protein in the haemolymph of Est6⁹ females. If this is the case, the mechanism by which this occurs can then be investigated in detail. One approach is to synthesise radiolabelled EST6 and EST6-GS¹²³⁴ in a cell culture expression system, then inject these proteins into the haemolymph. If the glycosylated enzyme is being sequestered from the circulation by a particular organ, for example, the fat body, the radiolabel would be expected to accumulate in that tissue. If such sequestering does occur and is mediated by the glycans of the protein, then EST6-GS¹²³⁴ would not be expected to accumulate in that particular organ.

The possibility that the EST6 glycans perform no functional role is harder to test. If no obvious difference in translocation of non-glycosylated EST6 to the female haemolymph is observable, or a particular target tissue for which the glycans of the enzyme are required for recognition is not identified, then an alternative measure of function will be required. Perhaps the only way to determine if glycosylation is necessary for the physiological function of EST6 is to perform population cage experiments under varying environmental conditions. For such experiments, a large number of independent transformants would be required to eliminate fitness differences due to the position of integration of the EST6 constructs. By following the change in frequency of non-glycosylated EST6 in such populations, it may be possible to detect differences, too subtle to be identified by conventional biochemical and physiological analysis, but still of adaptive importance for EST6 function.

4.3 CONCLUSIONS

In conclusion, the use of sequence comparisons among divergent esterases to identify an amino acid substitution in an otherwise conserved region of EST6, and investigation of the effects of replacement of this residue on the activity of the protein using protein engineering techniques, has been successful in identifying an amino acid that affects the substrate specificity of the enzyme. This demonstrates that evolutionary conservation can be used, in the absence of detailed knowledge of the three dimensional structure, to identify amino acids of potential functional
importance. However, the three dimensional structure of EST6 is still required to confirm the conclusions presented concerning the role of residue 187.

Mutation of the N-linked glycosylation sites of EST6 has been successful in identifying which sites are glycosylated and has enabled preliminary structural characterisation of one of the attached glycans. A functional role for the glycans of EST6 remains elusive and the possibility that the glycans of EST6 are functionless has been raised. However, a subtle functional role for the EST6 glycans remains feasible. The availability of the mutant EST6 transformants, whose isolation has been described in this thesis, should enable the identification of such a role for the glycans of EST6.
REFERENCES


BODE, W. and SCHWAGER, P. 1975. The refined crystal structure of bovine β-trypsin at 1.8 Å resolution. *J. Mol. Biol.* **98**: 693-717


COCHRANE, B.J. and RICHMOND, R.C. 1979. Studies of esterase 6 in *Drosophila melanogaster* II. The genetics and frequency distribution of naturally occurring variants studied by electrophoretic and heat stability criteria. *Genetics* 93: 461-478


SHAW, E., MARES-GUIA, M., and COHEN, W. 1965. Evidence of an active centre histidine in trypsin through use of a specific reagent, 1-chloro-3-tosylamido-7-amino-2-heptanone, the chloromethyl ketone derived from Na-Tosyl-L-lysine. *Biochemistry* **4**: 2219-2224


APPENDIX I

A restriction map of Est6 and the surrounding genomic region is shown. Restriction sites used in the making of mutant Est6 constructs for transformation and their positions are labelled. The positions of the N-linked glycosylation site consensus sequences and the serine active site of the EST6 protein are also marked. The scale is 1cm=200 nucleotides.
Figure Legend

CHO = N-linked glycosylation site
Ser = Serine active site
The intron is in black and the coding regions for EST6 and ESTP are hatched.
APPENDIX II

A sample data set and output of the Systat statistics non-linear least squares regression programme are shown.

<table>
<thead>
<tr>
<th>Substrate Concentration (µM)</th>
<th>Initial Velocity (umol 8naphthol produced per min per mg protein)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.00</td>
<td>14.32</td>
</tr>
<tr>
<td>5.00</td>
<td>15.95</td>
</tr>
<tr>
<td>5.00</td>
<td>15.87</td>
</tr>
<tr>
<td>3.00</td>
<td>12.97</td>
</tr>
<tr>
<td>3.00</td>
<td>12.95</td>
</tr>
<tr>
<td>3.00</td>
<td>12.45</td>
</tr>
<tr>
<td>2.00</td>
<td>10.81</td>
</tr>
<tr>
<td>2.00</td>
<td>10.75</td>
</tr>
<tr>
<td>2.00</td>
<td>10.68</td>
</tr>
<tr>
<td>1.25</td>
<td>7.41</td>
</tr>
<tr>
<td>1.25</td>
<td>8.08</td>
</tr>
<tr>
<td>1.25</td>
<td>8.88</td>
</tr>
</tbody>
</table>
THE FOLLOWING RESULTS ARE FOR:
GROUP = 1.000

<table>
<thead>
<tr>
<th>ITERATION</th>
<th>LOSS</th>
<th>PARAMETER VALUES</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.1720760D+04</td>
<td>.1000D+00 .1000D+00</td>
</tr>
<tr>
<td>1</td>
<td>.6002039D+03</td>
<td>.1355D+02 -.4172D+00</td>
</tr>
<tr>
<td>2</td>
<td>.7608666D+02</td>
<td>.2950D+02 .2743D+01</td>
</tr>
<tr>
<td>3</td>
<td>.3202260D+01</td>
<td>.2259D+02 .2236D+01</td>
</tr>
<tr>
<td>4</td>
<td>.3082730D+01</td>
<td>.2221D+02 .2205D+01</td>
</tr>
<tr>
<td>5</td>
<td>.3067707D+01</td>
<td>.2229D+02 .2210D+01</td>
</tr>
<tr>
<td>6</td>
<td>.3051668D+01</td>
<td>.2227D+02 .2195D+01</td>
</tr>
<tr>
<td>7</td>
<td>.3011383D+01</td>
<td>.2203D+02 .2134D+01</td>
</tr>
<tr>
<td>8</td>
<td>.2991981D+01</td>
<td>.2181D+02 .2088D+01</td>
</tr>
<tr>
<td>9</td>
<td>.2990453D+01</td>
<td>.2177D+02 .2084D+01</td>
</tr>
<tr>
<td>10</td>
<td>.2990446D+01</td>
<td>.2178D+02 .2085D+01</td>
</tr>
<tr>
<td>11</td>
<td>.2990446D+01</td>
<td>.2178D+02 .2085D+01</td>
</tr>
</tbody>
</table>

DEPENDENT VARIABLE IS V

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>SUM-OF-SQUARES</th>
<th>DF</th>
<th>MEAN-SQUARE</th>
</tr>
</thead>
<tbody>
<tr>
<td>REGRESSION</td>
<td>1744.787</td>
<td>2</td>
<td>872.394</td>
</tr>
<tr>
<td>RESIDUAL</td>
<td>2.990</td>
<td>10</td>
<td>0.299</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PARAMETER</th>
<th>ESTIMATE</th>
<th>A.S.E.</th>
<th>LOWER 95%</th>
<th>UPPER &lt;95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>VMAX</td>
<td>21.776</td>
<td>1.069</td>
<td>19.393</td>
<td>24.158</td>
</tr>
<tr>
<td>KM</td>
<td>2.085</td>
<td>0.242</td>
<td>1.545</td>
<td>2.624</td>
</tr>
</tbody>
</table>

ASYMPTOTIC CORRELATION MATRIX OF PARAMETERS

<table>
<thead>
<tr>
<th></th>
<th>VMAX</th>
<th>KM</th>
</tr>
</thead>
<tbody>
<tr>
<td>VMAX</td>
<td>1.000</td>
<td>0.964</td>
</tr>
<tr>
<td>KM</td>
<td>0.964</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Sample chromatographs from the DEAE sepharose anion exchange column and the Superose 6 gel filtration column are shown in Figures 1 and 2. Both of these chromatographs are from the purification of EST6 from Dm145. Chromatographs from these columns for the purification of EST6^{H187Q} and EST6^{H187E} were essentially the same. In contrast, the three preparations displayed large differences in the relative amounts of EST6 eluted from the MonoQ anion exchange column. Hence, chromatographs from the MonoQ anion exchange column for each of the proteins are shown in Figures 3 a), b) and c).

**Figure 1.** Chromatograph of the proteins eluted from the DEAE-FF sepharose anion exchange column. Protein was detected by continuous measurement of absorbance at 280nm. The NaCl gradient is represented by the straight line and the concentration is given on the right. EST6 activity, which was detected in the fractions as described (section 2.2.3.1), eluted at a NaCl concentration of 370-400 mM and is represented by the hatched box.

**Figure 2.** Chromatograph of proteins eluted from the Superose 6 gel filtration column. EST6 eluted at 73% of a column volume.
Figure 3. Chromatographs from the MonoQ anion exchange column. In each case, two major UV absorbing peaks were observed and the EST6 activity corresponded to the second peak. The relative heights of the two peaks differed in every case. For Dm145, the first peak and the EST6 peak were of a similar size. For the EST6\textsuperscript{H187Q} transformant, the first peak was much bigger than the EST6 peak, and for the
EST6H187E transformant the converse was true. In each case EST6 eluted at 190-230 mM NaCl.