A Combined Computational Strategy of Sequence and Structural Analysis Predicts the Existence of a Functional Eicosanoid Pathway in *Drosophila melanogaster*

Michael Scarpati

*The Graduate Center, City University of New York*

How does access to this work benefit you? Let us know!

Follow this and additional works at: https://academicworks.cuny.edu/gc_etds

Part of the Bioinformatics Commons, Computational Biology Commons, Molecular Biology Commons, and the Structural Biology Commons

Recommended Citation


https://academicworks.cuny.edu/gc_etds/2386

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
A Combined Computational Strategy of Sequence and Structural Analysis Predicts
the Existence of a Functional Eicosanoid Pathway in *Drosophila melanogaster*

by

Michael Scarpati

A dissertation submitted to the Graduate Faculty in Biology in partial fulfillment of the
requirements for the degree of Doctor of Philosophy, The City University of New York

2017
This manuscript has been read and accepted for the Graduate Faculty in Biology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Date
___________________
Chair of Examining Committee
Dr. Shaneen Singh, Brooklyn College

Date
___________________
Executive Officer
Dr. Cathy Savage-Dunn

Dr. Shubha Govind, City College

Dr. Robert Dottin, Hunter College

Dr. Luis Quadri, Brooklyn College

Dr. Evros Vassiliou, Kean University

Dr. Raj Wadgaonka, SUNY Downstate

Supervising Committee

The City University of New York
Abstract

With increased understanding of their roles in signal transduction and metabolism, eicosanoids have emerged as important players in human health and disease. Mammalian prostanoids and related lipid mediators perform varied functions in different tissues and organs. Synthesized through the oxygenation of C20 polyunsaturated fatty acids, mammalian eicosanoids are both pro- and anti-inflammatory. The physiological contexts in which eicosanoid family members act at the cellular level are not well understood. In this study, we examined whether the genome of *Drosophila melanogaster*, a powerful model for innate immunity and inflammation, codes for the enzymes required for eicosanoid biosynthesis. We report the existence of putative eicosanoid biosynthesis enzymes in *Drosophila melanogaster* which may function together as a pathway similar to the mammalian eicosanoid synthesis pathway. Standard sequence-based search methods failed to identify high confidence orthologs for a majority of the mammalian eicosanoid synthesis enzymes in *D. melanogaster*, and in insects generally. Using sensitive sequence analysis techniques, we identified candidate orthologs in the *Drosophila* genome that share low global sequence identities with their human counterparts. The *Drosophila* sequences were further scrutinized by modeling and structural analyses. We generated and evaluated full-length models for top-scoring *Drosophila* candidates corresponding to each human eicosanoid synthesis enzyme and identified potentially equivalent functional residues. This combination of sensitive sequence and structural analyses revealed that the existence of eight high confidence, five mid-range and eight low confidence candidates. Four predicted cyclooxygenases and two potential lipoxygenase activating proteins, highly divergent from their human counterparts, were identified, although similar methods failed to identify putative lipoxygenase enzymes. Tertiary structures of a majority of identified candidate fly proteins are very similar to the corresponding human target enzymes and appear to possess the necessary
catalytic residues. These results, in combination with other recent biochemical studies alluding to eicosanoid activity in insects by other groups, suggest that *D. melanogaster* may indeed possess biosynthesis pathways for eicosanoid or eicosanoid-like biolipids. However, the predominant view in the field is that an eicosanoid synthesis pathway does not exist in *Drosophila* primarily because to date clear homologs of the enzymes of this pathway have not been identified. Our study challenges this currently held view. Molecular-genetic and biochemical analyses of individual biosynthetic enzymes in *D. melanogaster*, a model organism with low genetic redundancy will reveal if the fly enzymes are functionally equivalent to their mammalian counterparts; their *in vivo* interactions will allow construction of pathways and networks in a physiological context. Our findings predict that classical or novel eicosanoids or eicosanoid-like lipid mediators regulate biological functions in insects. Eicosanoids are known to play important roles in insect immunity. The identification of these lipid mediators will therefore provide new insect control measures or the means of improving the health of beneficial insects.
TABLE OF CONTENTS

ABSTRACT ........................................................................................................................................ iv
TABLE OF CONTENTS .................................................................................................................. vi
LIST OF ABBREVIATIONS ........................................................................................................ xiv
LIST OF TABLES ........................................................................................................................ xvi
LIST OF FIGURES .................................................................................................................... xviii
CHAPTER 1: INTRODUCTION ................................................................................................... 1
Overview of the eicosanoid biosynthesis pathway ................................................................ 1
The COX branch of the eicosanoid pathway ......................................................................... 2
The LOX branch of the eicosanoid pathway ......................................................................... 3
Non-classic Eicosanoids ........................................................................................................... 4
Eicosanoid biosynthesis and signaling in insects .................................................................. 7

CHAPTER 2: METHODS .......................................................................................................... 9
Overview ..................................................................................................................................... 9
Selection and Characterization of Human Eicosanoid Synthesis Enzymes ......................... 10
Identification of Potential Eicosanoid Synthesis Enzymes in D. melanogaster .................... 12
In silico Modeling of D. melanogaster Candidates ................................................................ 13

CHAPTER 3: RESULTS .......................................................................................................... 15
Identification and Characterization of Candidate Eicosanoid Synthesis Enzymes in D.
melanogaster ............................................................................................................................... 15
Group 1: The High Scoring Matches ..................................................................................... 16
Prostaglandin E Synthase (PTGES) ...................................................................................... 16
A. Predicted Secondary Structure ..................................................................................... 17
B. Domain Architecture ................................................................................................. 17
C. Model Generation and Validation ............................................................................. 18
   1. Generation ........................................................................................................... 18
   2. Validation .......................................................................................................... 18
D. Structural Analyses .................................................................................................. 19
D. Structural Analyses ................................................................................................... 41
  1. Secondary Structure ............................................................................................. 41
  2. Physiochemical Properties ................................................................................... 42
E. Analysis of Functional Residues ............................................................................... 42
F. Summary Table........................................................................................................... 43

Gamma-glutamyltranspeptidase (GGT1) ........................................................................ 44
  A. Predicted Secondary Structure .................................................................................. 45
  B. Domain Architecture ................................................................................................. 46
  C. Model Generation and Validation ............................................................................. 46
    1. Generation ............................................................................................................ 46
    2. Validation ............................................................................................................. 48
D. Structural Analyses ................................................................................................... 49
  1. Secondary Structure ............................................................................................. 49
  2. Physiochemical Properties ................................................................................... 50
E. Analysis of Functional Residues ............................................................................... 51
F. Summary Table........................................................................................................... 51

Renal Dipeptidase (DPEP1) .............................................................................................. 51
  A. Predicted Secondary Structure .................................................................................. 52
  B. Domain Architecture ................................................................................................. 53
  C. Model Generation and Validation ............................................................................. 53
    1. Generation ............................................................................................................ 53
    2. Validation ............................................................................................................. 54
D. Structural Analyses ................................................................................................... 55
  1. Secondary Structure ............................................................................................. 56
  2. Physiochemical Properties ................................................................................... 57
E. Analysis of Functional Residues ............................................................................... 57
F. Summary Table........................................................................................................... 58

Glutathione Peroxidase (GPX1) ....................................................................................... 58
  A. Predicted Secondary Structure .................................................................................. 59
  B. Domain Architecture ................................................................................................. 59
  C. Model Generation and Validation ............................................................................. 60
    1. Generation ............................................................................................................ 60
2. Validation............................................................................................................. 60
D. Structural Analyses ............................................................................................. 61
  1. Secondary Structure ......................................................................................... 62
  2. Physiochemical Properties ............................................................................. 62
E. Analysis of Functional Residues ........................................................................ 63
F. Summary Table.................................................................................................. 64

**Carboxypeptidase A1 (CPA1)................................................................. 64**

A. Predicted Secondary Structure ........................................................................ 65
B. Domain Architecture ......................................................................................... 65
C. Model Generation and Validation .................................................................... 66
  1. Generation ....................................................................................................... 66
  2. Validation ....................................................................................................... 67
D. Structural Analyses ............................................................................................. 68
  1. Secondary Structure ......................................................................................... 68
  2. Physiochemical Properties ............................................................................. 69
E. Analysis of Functional Residues ........................................................................ 69
F. Summary Table.................................................................................................. 70

**Group 2: The Midrange Candidates......................................................... 71**

**Hematopoietic Prostaglandin D Synthase (HPGDS)................................. 71**

A. Predicted Secondary Structure ........................................................................ 72
B. Domain Architecture ......................................................................................... 72
C. Model Generation and Validation .................................................................... 73
  1. Generation ....................................................................................................... 73
  2. Validation ....................................................................................................... 74
D. Structural Analyses ............................................................................................. 74
  1. Secondary Structure ......................................................................................... 75
  2. Physiochemical Properties ............................................................................. 75
E. Analysis of Functional Residues ........................................................................ 76
F. Summary Table.................................................................................................. 77

**Prostaglandin E Synthase 3 (PTGES3)..................................................... 77**

A. Predicted Secondary Structure ........................................................................ 78
B. Domain Architecture ......................................................................................... 78
C. Model Generation and Validation ................................................................. 78
  1. Generation .................................................................................................. 78
  2. Validation .................................................................................................. 79
D. Structural Analyses ....................................................................................... 81
  1. Secondary Structure .................................................................................. 81
  2. Physiochemical Properties ........................................................................ 82
E. Analysis of Functional Residues ................................................................. 82
F. Summary Table .............................................................................................. 83

Carbonyl Reductase 1 (CBR1) ................................................................. 83
A. Predicted Secondary Structure ................................................................. 84
B. Domain Architecture .................................................................................. 84
C. Model Generation and Validation ............................................................. 85
  1. Generation ................................................................................................ 85
  2. Validation ................................................................................................ 85
D. Structural Analyses ..................................................................................... 87
  1. Secondary Structure ................................................................................ 87
  2. Physiochemical Properties ...................................................................... 88
E. Analysis of Functional Residues ................................................................. 88
F. Summary Table ............................................................................................ 89

15-Hydroxyprostaglandin Dehydrogenase NAD(+) (HPGD) ...................... 90
A. Predicted Secondary Structure ................................................................. 91
B. Domain Architecture ................................................................................ 91
C. Model Generation and Validation ............................................................. 91
  1. Generation ................................................................................................ 91
  2. Validation ................................................................................................ 92
D. Structural Analyses ..................................................................................... 94
  1. Secondary Structure ................................................................................ 94
  2. Physiochemical Properties ...................................................................... 94
E. Analysis of Functional Residues ................................................................. 95
F. Summary Table ............................................................................................ 96

Thromboxane A Synthase 1 (TBXAS1) ...................................................... 96
A. Predicted Secondary Structure ................................................................. 97
B. Domain Architecture ................................................................................................. 98
C. Model Generation and Validation ............................................................................. 98
  1. Generation ............................................................................................................ 98
  2. Validation ............................................................................................................. 99
D. Structural Analyses ................................................................................................... 100
  1. Secondary Structure ............................................................................................. 102
  2. Physiochemical Properties ................................................................................... 103
E. Analysis of Functional Residues ............................................................................... 103
F. Summary Table .......................................................................................................... 104

**Group 3: The Most Distant Candidates** ................................................................. 105

**Cyclooxygenase 1/2 (PTGS1 / PTGS2)** ................................................................. 105
  A. Predicted Secondary Structure .............................................................................. 108
  B. Domain Architecture ............................................................................................. 109
  C. Model Generation and Validation ........................................................................ 109
    1. Generation ........................................................................................................... 109
    2. Validation ............................................................................................................ 112
  D. Structural Analyses ............................................................................................... 113
    1. Secondary Structure ............................................................................................ 114
    2. Physiochemical Properties .................................................................................. 116
  E. Analysis of Functional Residues .......................................................................... 116
  F. Summary Table ....................................................................................................... 117

**Prostaglandin D Synthase (PTGDS)** ................................................................. 118
  A. Predicted Secondary Structure .............................................................................. 118
  B. Domain Architecture ............................................................................................. 119
  C. Model Generation and Validation ........................................................................ 119
    1. Generation ........................................................................................................... 119
    2. Validation ............................................................................................................ 120
  D. Structural Analyses ............................................................................................... 121
    1. Secondary Structure ............................................................................................ 121
    2. Physiochemical Properties .................................................................................. 122
  E. Analysis of Functional Residues .......................................................................... 122
  F. Summary Table ....................................................................................................... 123
Prostacyclin Synthase (PTGIS) .......................................................... 124
  A. Predicted Secondary Structure .......................................................... 125
  B. Domain Architecture ........................................................................... 126
  C. Model Generation and Validation .......................................................... 126
      1. Generation ......................................................................................... 126
      2. Validation ........................................................................................ 127
  D. Structural Analyses ................................................................................. 128
      1. Secondary Structure ........................................................................ 129
      2. Physiochemical Properties .............................................................. 130
  E. Analysis of Functional Residues ............................................................... 130
  F. Summary Table ...................................................................................... 131

Leukotriene C4 Synthase (LTC4S) .............................................................. 132
  A. Predicted Secondary Structure ............................................................. 132
  B. Domain Architecture ............................................................................. 133
  C. Model Generation and Validation .......................................................... 133
      1. Generation ......................................................................................... 133
      2. Validation ........................................................................................ 133
  D. Structural Analyses ................................................................................. 134
      1. Secondary Structure ........................................................................ 135
      2. Physiochemical Properties .............................................................. 135
  E. Analysis of Functional Residues ............................................................... 136
  F. Summary Table ...................................................................................... 137

Arachidonate 5-Lipoxygenase-Activating Protein (ALOX5AP) ...................... 137
  A. Predicted Secondary Structure ............................................................. 138
  B. Domain Architecture ............................................................................. 138
  C. Model Generation and Validation .......................................................... 138
      1. Generation ......................................................................................... 138
      2. Validation ........................................................................................ 139
  D. Structural Analyses ................................................................................. 140
      1. Secondary Structure ........................................................................ 141
      2. Physiochemical Properties .............................................................. 141
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>Arachidonic acid</td>
</tr>
<tr>
<td>ALA</td>
<td>Alpha-linoleic acid</td>
</tr>
<tr>
<td>ALOX5AP</td>
<td>Arachidonate 5-Lipoxygenase Activating Protein</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Local Alignment Search Tool</td>
</tr>
<tr>
<td>CBR</td>
<td>Carbonyl reductase</td>
</tr>
<tr>
<td>COX</td>
<td>Cyclooxygenase</td>
</tr>
<tr>
<td>cPGES</td>
<td>Cytosolic prostaglandin E synthase</td>
</tr>
<tr>
<td>cPLA2</td>
<td>Cytosolic phospholipase A2</td>
</tr>
<tr>
<td>CYP</td>
<td>Cytochrome</td>
</tr>
<tr>
<td>CYP450</td>
<td>Cytochrome P450</td>
</tr>
<tr>
<td>DHET</td>
<td>Diol dihydroxyeicosatrienoic acid</td>
</tr>
<tr>
<td>Duox</td>
<td>Dual oxidase</td>
</tr>
<tr>
<td>EET</td>
<td>Epoxyeicosatrienoic acid</td>
</tr>
<tr>
<td>EFOXs</td>
<td>Exo-derivatives</td>
</tr>
<tr>
<td>ExPaSy</td>
<td>Expert Protein Analysis System</td>
</tr>
<tr>
<td>FLAP</td>
<td>Five-Lipoxygenase Activating Protein</td>
</tr>
<tr>
<td>GSH</td>
<td>Glutathione</td>
</tr>
<tr>
<td>GST</td>
<td>Glutathione S-transferase</td>
</tr>
<tr>
<td>HETE</td>
<td>Hydroxyeicosatetraenoic acid</td>
</tr>
<tr>
<td>HMMs</td>
<td>Hidden Markov models</td>
</tr>
<tr>
<td>HPETE</td>
<td>Hydroperoxyeicosatetraenoic acid</td>
</tr>
<tr>
<td>Irc</td>
<td>Immune regulated catalase</td>
</tr>
<tr>
<td>LA</td>
<td>Linoleic acid</td>
</tr>
<tr>
<td>LOX</td>
<td>Lipoxygenase</td>
</tr>
<tr>
<td>LPS</td>
<td>Lipopolysaccharide</td>
</tr>
<tr>
<td>LT</td>
<td>Leukotriene</td>
</tr>
<tr>
<td>LTCS</td>
<td>Leukotriene C4 synthase</td>
</tr>
<tr>
<td>MAPEG</td>
<td>Membrane-associated proteins involved in eicosanoid and glutathione metabolism</td>
</tr>
<tr>
<td>MBD</td>
<td>Membrane-binding domain</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>Mgst</td>
<td>Microsomal glutathione S-transferase</td>
</tr>
<tr>
<td>mPGES</td>
<td>Microsomal prostaglandin E synthase</td>
</tr>
<tr>
<td>NADPH</td>
<td>Nicotinamide adenine dinucleotide phosphate</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Center for Biotechnology Information</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>PG</td>
<td>Prostaglandin</td>
</tr>
<tr>
<td>PGD2</td>
<td>Prostaglandin D2</td>
</tr>
<tr>
<td>PGE2</td>
<td>Prostaglandin E2</td>
</tr>
<tr>
<td>PGES</td>
<td>Prostaglandin E synthase</td>
</tr>
<tr>
<td>PGF2α</td>
<td>Prostaglandin F2α</td>
</tr>
<tr>
<td>PGFS</td>
<td>Prostaglandin F synthase</td>
</tr>
<tr>
<td>PGH2</td>
<td>Prostaglandin H2</td>
</tr>
<tr>
<td>PGHS</td>
<td>Prostaglandin H synthase</td>
</tr>
<tr>
<td>PGI2</td>
<td>Prostacyclin/prostaglandin I2</td>
</tr>
<tr>
<td>PGIS</td>
<td>Prostacyclin synthase</td>
</tr>
<tr>
<td>POX</td>
<td>Peroxidase</td>
</tr>
<tr>
<td>PRS-BLAST</td>
<td>Reverse position-specific Basic Local Alignment Search Tool</td>
</tr>
<tr>
<td>PSI-BLAST</td>
<td>Position-Specific Iterated Basic Local Alignment Search Tool</td>
</tr>
<tr>
<td>PSSMs</td>
<td>Pre-calculated position-specific scoring matrices</td>
</tr>
<tr>
<td>PUFA</td>
<td>Polyunsaturated fatty acid</td>
</tr>
<tr>
<td>Pxn</td>
<td>Proxidasin</td>
</tr>
<tr>
<td>Pxt</td>
<td>Peroxinectin-like</td>
</tr>
<tr>
<td>RMSD</td>
<td>Root Mean Square Deviation</td>
</tr>
<tr>
<td>SIB</td>
<td>Swiss Institute of Bioinformatics</td>
</tr>
<tr>
<td>SMART</td>
<td>Simple modular architecture research tool</td>
</tr>
<tr>
<td>Su(2)P</td>
<td>Suppressor of ref(2)P sterility</td>
</tr>
<tr>
<td>TMPD</td>
<td>Tetramethyl-p-phenylenediamine</td>
</tr>
<tr>
<td>TX</td>
<td>Thromboxane</td>
</tr>
<tr>
<td>TXA2</td>
<td>Thromboxane A2</td>
</tr>
</tbody>
</table>
List of Tables

Table 1  Summary of the mammalian eicosanoid synthesis enzymes selected as representatives for this study. 11

Table 2  Summary of D. melanogaster genes encoding potential orthologs for the mammalian eicosanoid synthesis enzymes examined during this study. 15-16

Table 3  Summary of features shared by PTGES and potential D. melanogaster ortholog CG1742. 22

Table 4  Summary of features shared by PTGES2 and potential D. melanogaster ortholog CG4086. 29

Table 5  Summary of features shared by AKR1A1 and potential D. melanogaster ortholog CG6084. 36

Table 6  Summary of features shared by LTA4H and potential D. melanogaster ortholog CG10602. 43

Table 7  Summary of features shared by GGT1 and potential D. melanogaster ortholog CG6461. 51

Table 8  Summary of features shared by DPEP1 and potential D. melanogaster ortholog CG6154. 58

Table 9  Summary of features shared by GPX1 and potential D. melanogaster ortholog CG12013. 64

Table 10 Summary of features shared by CPA1 and potential D. melanogaster ortholog CG18585. 70

Table 11 Summary of features shared by HPGDS and potential D. melanogaster ortholog CG8938. 77

Table 12 Summary of features shared by PTGES3 and potential D. melanogaster ortholog CG16817. 83

Table 13 Summary of features shared by CBR1 and potential D. melanogaster ortholog CG11200. 89

Table 14 Summary of features shared by HPGD and potential D. melanogaster ortholog CG18814. 96

Table 15 Summary of features shared by TBXAS1 and potential D. melanogaster ortholog CG3616. 104
<table>
<thead>
<tr>
<th>Table 16</th>
<th>Templates selected for compositing modeling of CG4009 using I-TASSER.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 17</td>
<td>Summary of features shared by PTGS1 and potential <em>D. melanogaster</em> ortholog CG4009.</td>
</tr>
<tr>
<td>Table 18</td>
<td>Summary of features shared by PTGDS and potential <em>D. melanogaster</em> ortholog CG33126.</td>
</tr>
<tr>
<td>Table 19</td>
<td>Summary of features shared by PTGIS and potential <em>D. melanogaster</em> ortholog CG3466.</td>
</tr>
<tr>
<td>Table 20</td>
<td>Summary of features shared by LTC4S and potential <em>D. melanogaster</em> ortholog CG33178.</td>
</tr>
<tr>
<td>Table 21</td>
<td>Summary of features shared by ALOX5AP and potential <em>D. melanogaster</em> ortholog CG33177.</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1  The canonical human eicosanoid synthesis pathway.  
Figure 2  Summary of the workflow and tools used to carry out the present study.  
Figure 3  Predicted secondary structure of PTGES and CG1742 calculated using PROMALS3D.  
Figure 4  Predicted domain architecture of PTGES and CG1742 determined using Interpro and Pfam.  
Figure 5  HMM profile of CG1742 aligned against chain A of 4YL0, a template structure selected for modeling CG1742.  
Figure 6  Validation of the CG1742 model: ProQ2 quality score mapped to a 3D model of CG1742 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).  
Figure 7  Pairwise alignment of CG1742 and 4AL0 generated from structural superposition with shared secondary structure elements and conserved residues highlighted.  
Figure 8  Pairwise alignment of CG1742 and 4AL0 generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.  
Figure 9  PTGES (4AL0, cyan-blue) superimposed on the predicted structure of CG1742 (green-red).  
Figure 10  PTGES (4AL0, cyan-blue) superimposed on the predicted structure of CG1742 (green-red), with potential matches for conserved functional residues highlighted.  
Figure 11  Predicted secondary structure of PTGES2 and CG4086 calculated using PROMALS3D.  
Figure 12  Predicted domain architecture of PTGES2 and CG4086 determined using Interpro and Pfam.  
Figure 13  HMM profile of CG4086 aligned against chain A of 1Z9H, a template structure selected for modeling CG4086.  
Figure 14  Validation of the CG4086 model: ProQ2 quality score mapped to a 3D model of CG4086 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).
Figure 15  Pairwise alignment of CG4086 and 1Z9H generated from structural superposition with shared secondary structure elements and conserved residues highlighted.  

Figure 16  Pairwise alignment of CG4086 and 1Z9H generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.  

Figure 17  Truncated PTGES2 (1Z9H, cyan-blue) superimposed on the predicted structure of CG4086 (green-red).  

Figure 18  Truncated PTGES2 (1Z9H, cyan-blue) superimposed on the predicted structure of CG4086 (green-red), with potential matches for conserved functional residues highlighted.  

Figure 19  Predicted secondary structure of AKR1A1 and CG6084 calculated using PROMALS3D.  

Figure 20  Predicted domain architecture of AKR1A1 and CG6084 determined using Interpro and Pfam.  

Figure 21  HMM profile of CG6084 aligned against chain A of 2ALR, a template structure selected for modeling CG6084.  

Figure 22  Validation of the CG6084 model: ProQ2 quality score mapped to a 3D model of CG6084 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).  

Figure 23  Pairwise alignment of CG6084 and 2ALR generated from structural superposition with shared secondary structure elements and conserved residues highlighted.  

Figure 24  Pairwise alignment of CG6084 and 2ALR generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.  

Figure 25  AKR1A1 (2ALR, cyan-blue) superimposed on the predicted structure of CG6084 (green-red).  

Figure 26  AKR1A1 (2ALR, cyan-blue) superimposed on the predicted structure of CG6084 (green-red), with potential matches for conserved functional residues highlighted.  

Figure 27  Predicted secondary structure of LTA4H and CG10602 calculated using PROMALS3D.  

Figure 28  Predicted domain architecture of LTA4H and CG10602 determined using Interpro and Pfam.
Figure 29  HMM profile of CG10602 aligned against chain X of 3B7U, a template structure selected for modeling CG10602.

Figure 30  Validation of the CG10602 model: ProQ2 quality score mapped to a 3D model of CG10602 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).

Figure 31  Pairwise alignment of CG10602 and 3B7U generated from structural superposition with shared secondary structure elements and conserved residues highlighted.

Figure 32  Pairwise alignment of CG10602 and 3B7U generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

Figure 33  LTA4H (3B7U, cyan-blue) superimposed on the predicted structure of CG10602 (green-red).

Figure 34  LTA4H (3B7U, cyan-blue) superimposed on the predicted structure of CG10602 (green-red), with potential matches for conserved functional residues highlighted.

Figure 35  Predicted secondary structure of GGT1 and CG6461 calculated using PROMALS3D.

Figure 36  Predicted domain architecture of GGT1 and CG6461 determined using Interpro and Pfam.

Figure 37  HMM profile of CG6461 aligned against chains A and B of 4GDX, a template structure selected for modeling CG6461.

Figure 38  Validation of the CG6461 model: ProQ2 quality score mapped to a 3D model of CG6461 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).

Figure 39  Pairwise alignment of CG6461 and 4GDX generated from structural superposition with shared secondary structure elements and conserved residues highlighted.

Figure 40  Pairwise alignment of CG6461 and 4GDX generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

Figure 41  GGT1 (4GDX, cyan-blue) superimposed on the predicted structure of CG6461 (green-red).

Figure 42  GGT1 (4GDX, cyan-blue) superimposed on the predicted structure of CG6461 (green-red), with potential matches for conserved
functional residues highlighted.

Figure 43  Predicted secondary structure of DPEP1 and CG6154 calculated using PROMALS3D.

Figure 44  Predicted domain architecture of DPEP1 and CG6154 determined using Interpro and Pfam.

Figure 45  HMM profile of CG6154 aligned against chain A of 1ITQ, a template structure selected for modeling CG6154.

Figure 46  Validation of the CG6154 model: ProQ2 quality score mapped to a 3D model of CG6154 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).

Figure 47  Pairwise alignment of CG6154 and 1ITQ generated from structural superposition with shared secondary structure elements and conserved residues highlighted.

Figure 48  Pairwise alignment of CG6154 and 1ITQ generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

Figure 49  DPEP1 (1ITQ, cyan-blue) superimposed on the predicted structure of CG6154 (green-red).

Figure 50  DPEP1 (1ITQ, cyan-blue) superimposed on the predicted structure of CG6154 (green-red), with potential matches for conserved functional residues highlighted.

Figure 51  Predicted secondary structure of GPX1 and CG12013 calculated using PROMALS3D.

Figure 52  Predicted domain architecture of GPX1 and CG12013 determined using Interpro and Pfam.

Figure 53  HMM profile of CG12013 aligned against chain A of 2F8A, a template structure selected for modeling CG12013.

Figure 54  Validation of the CG12013 model: ProQ2 quality score mapped to a 3D model of CG12013 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).

Figure 55  Pairwise alignment of CG12013 and 2F8A generated from structural superposition with shared secondary structure elements and conserved residues highlighted.

Figure 56  Pairwise alignment of CG12013 and 2F8A generated from structural superposition with conserved residues highlighted using the
CLUSTALX color scheme.

**Figure 57** GPX1 (2F8A, cyan-blue) superimposed on the predicted structure of CG12013 (green-red).

**Figure 58** GPX1 (2F8A, cyan-blue) superimposed on the predicted structure of CG12013 (green-red), with potential matches for conserved functional residues highlighted.

**Figure 59** Predicted secondary structure of CPA1 and CG18585 calculated using PROMALS3D.

**Figure 60** Predicted domain architecture of CPA1 and CG18585 determined using Interpro and Pfam.

**Figure 61** HMM profile of CG18585 aligned against chain A of 2V77, a template structure selected for modeling CG18585.

**Figure 62** Validation of the CG18585 model: ProQ2 quality score mapped to a 3D model of CG18585 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).

**Figure 63** Pairwise alignment of CG18585 and 2V77 generated from structural superposition with shared secondary structure elements and conserved residues highlighted.

**Figure 64** Pairwise alignment of CG18585 and 2V77 generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

**Figure 65** Truncated CPA1 (2V77, cyan-blue) superimposed on the predicted structure of CG18585 (green-red).

**Figure 66** Truncated CPA1 (2V77, cyan-blue) superimposed on the predicted structure of CG18585 (green-red), with potential matches for conserved functional residues highlighted.

**Figure 67** Predicted secondary structure of HPGDS and CG8938 calculated using PROMALS3D.

**Figure 68** Predicted domain architecture of HPGDS and CG8938 determined using Interpro and Pfam.

**Figure 69** HMM profile of CG8938 aligned against chain A of 1IYH, a template structure selected for modeling CG8938.

**Figure 70** Validation of the CG8938 model: ProQ2 quality score mapped to a 3D model of CG8938 (top); ProSA global quality score ranking.
Figure 71  Pairwise alignment of CG8938 and 1IYH generated from structural superposition with shared secondary structure elements and conserved residues highlighted.  75

Figure 72  Pairwise alignment of CG8938 and 1IYH generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.  75

Figure 73  HPGDS (1IYH, cyan-blue) superimposed on the predicted structure of CG8938 (green-red).  76

Figure 74  HPGDS (1IYH, cyan-blue) superimposed on the predicted structure of CG8938 (green-red), with potential matches for conserved functional residues highlighted.  76

Figure 75  Predicted secondary structure of PTGES3 and CG16817 calculated using PROMALS3D.  78

Figure 76  Predicted domain architecture of PTGES3 and CG16817 determined using Interpro and Pfam.  78

Figure 77  HMM profile of CG16817 aligned against chain A of 1EJF, a template structure selected for modeling CG16817.  80

Figure 78  Validation of the CG16817 model: ProQ2 quality score mapped to a 3D model of CG16817 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).  80

Figure 79  Pairwise alignment of CG16817 and 1EJF generated from structural superposition with shared secondary structure elements and conserved residues highlighted.  81

Figure 80  Pairwise alignment of CG16817 and 1EJF generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.  82

Figure 81  PTGES3 (1EJF, cyan-blue) superimposed on the predicted structure of CG16817 (green-red).  82

Figure 82  PTGES3 (1EJF, cyan-blue) superimposed on the predicted structure of CG16817 (green-red), with potential matches for conserved functional residues highlighted.  84

Figure 83  Predicted secondary structure of CBR1 and CG11200 calculated using PROMALS3D.  84
Figure 84  Predicted domain architecture of CBR1 and CG11200 determined using Interpro and Pfam.

Figure 85  HMM profile of CG11200 aligned against chain A of 3BHJ, a template structure selected for modeling CG11200.

Figure 86  Validation of the CG11200 model: ProQ2 quality score mapped to a 3D model of CG11200 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).

Figure 87  Pairwise alignment of CG11200 and 3BHJ generated from structural superposition with shared secondary structure elements and conserved residues highlighted.

Figure 88  Pairwise alignment of CG11200 and 3BHJ generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

Figure 89  CBR1 (3BHJ, cyan-blue) superimposed on the predicted structure of CG11200 (green-red).

Figure 90  CBR1 (3BHJ, cyan-blue) superimposed on the predicted structure of CG11200 (green-red), with potential matches for conserved functional residues highlighted.

Figure 91  Predicted secondary structure of HPGD and CG18814 calculated using PROMALS3D.

Figure 92  Predicted domain architecture of HPGD and CG18814 determined using Interpro and Pfam.

Figure 93  HMM profile of CG18814 aligned against chain A of 2GDZ, a template structure selected for modeling CG18814.

Figure 94  Validation of the CG18814 model: ProQ2 quality score mapped to a 3D model of CG18814 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).

Figure 95  Pairwise alignment of CG18814 and 2GDZ generated from structural superposition with shared secondary structure elements and conserved residues highlighted.

Figure 96  Pairwise alignment of CG18814 and 2GDZ generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

Figure 97  HPGD (2GDZ, cyan-blue) superimposed on the predicted structure of CG18814 (green-red).
Figure 98  HPGD (2GDZ, cyan-blue) superimposed on the predicted structure of CG18814 (green-red), with potential matches for conserved functional residues highlighted.

Figure 99  Predicted secondary structure of TBXAS1 and CG3616 calculated using PROMALS3D.

Figure 100  Predicted domain architecture of TBXAS1 and CG3616 determined using Interpro and Pfam.

Figure 101  HMM profile of CG3616 aligned against chain A of 4DGZ, a template structure selected for modeling CG3616.

Figure 102  Validation of the CG3616 model: ProQ2 quality score mapped to a 3D model of CG3616 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).

Figure 103  Pairwise alignment of CG3616 and 4DGZ generated from structural superposition with shared secondary structure elements and conserved residues highlighted.

Figure 104  Pairwise alignment of CG3616 and 4DGZ generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

Figure 105  TBXAS1 (4DGZ, cyan-blue) superimposed on the predicted structure of CG3616 (green-red).

Figure 106  TBXAS1 (4DGZ, cyan-blue) superimposed on the predicted structure of CG3616 (green-red), with potential matches for conserved functional residues highlighted.

Figure 107  Predicted secondary structure of PTGS1 and CG4009 calculated using PROMALS3D.

Figure 108  Predicted domain architecture of PTGS1 and CG4009 determined using Interpro and Pfam.

Figure 109  Multiple Sequence Alignment illustrating CG4009 aligned against template structures selected by I-TASSER for multi-template modeling of CG4009.

Figure 110  Validation of the CG4009 model: ProQ2 quality score mapped to a 3D model of CG4009 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).

Figure 111  Pairwise alignment of CG4009 and 3N8V generated from structural superposition with shared secondary structure elements and conserved residues highlighted.
conserved residues highlighted.

Figure 112  Pairwise alignment of CG4009 and 3N8V generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

Figure 113  PTGS1 (3N8V, cyan-blue) superimposed on the predicted structure of CG4009 (green-red).

Figure 114  PTGS1 (3N8V, cyan-blue) superimposed on the predicted structure of CG4009 (green-red), with potential matches for conserved functional residues highlighted.

Figure 115  Predicted secondary structure of PTGDS and CG33126 calculated using PROMALS3D.

Figure 116  Predicted domain architecture of PTGDS and CG33126 determined using Interpro and Pfam.

Figure 117  HMM profile of CG33126 aligned against chain A of 5WY9, a template structure selected for modeling CG33126.

Figure 118  Validation of the CG33126 model: ProQ2 quality score mapped to a 3D model of CG33126 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).

Figure 119  Pairwise alignment of CG33126 and 5WY9 generated from structural superposition with shared secondary structure elements and conserved residues highlighted.

Figure 200  Pairwise alignment of CG33126 and 5WY9 generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

Figure 201  PTGDS (5WY9, cyan-blue) superimposed on the predicted structure of CG33126 (green-red).

Figure 202  PTGDS (5WY9, cyan-blue) superimposed on the predicted structure of CG33126 (green-red), with potential matches for conserved functional residues highlighted.

Figure 203  Predicted secondary structure of PTGIS and CG3466 calculated using PROMALS3D.

Figure 204  Predicted domain architecture of PTGIS and CG3466 determined using Interpro and Pfam.

Figure 205  HMM profile of CG3466 aligned against chain A of 4DGZ, a
template structure selected for modeling CG3466.

Figure 206 Validation of the CG3466 model: ProQ2 quality score mapped to a 3D model of CG3466 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).

Figure 207 Pairwise alignment of CG3466 and 2IAG generated from structural superposition with shared secondary structure elements and conserved residues highlighted.

Figure 208 Pairwise alignment of CG3466 and 2IAG generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

Figure 209 PTGIS (2IAG, cyan-blue) superimposed on the predicted structure of CG3466 (green-red).

Figure 210 PTGIS (2IAG, cyan-blue) superimposed on the predicted structure of CG3466 (green-red), with potential matches for conserved functional residues highlighted.

Figure 211 Predicted secondary structure of LTC4S and CG33178 calculated using PROMALS3D.

Figure 212 Predicted domain architecture of LTC4S and CG33178 determined using Interpro and Pfam.

Figure 213 HMM profile of CG33178 aligned against chain A of 2H8A, a template structure selected for modeling CG33178.

Figure 214 Validation of the CG33178 model: ProQ2 quality score mapped to a 3D model of CG33178 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).

Figure 215 Pairwise alignment of CG33178 and 2PNO generated from structural superposition with shared secondary structure elements and conserved residues highlighted.

Figure 216 Pairwise alignment of CG33178 and 2PNO generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

Figure 217 LTC4S (2PNO, cyan-blue) superimposed on the predicted structure of CG33178 (green-red).

Figure 218 LTC4S (2PNO, cyan-blue) superimposed on the predicted structure of CG33178 (green-red), with potential matches for conserved functional residues highlighted.
Figure 219  Predicted secondary structure of ALOX5AP and CG33177 calculated using PROMALS3D.

Figure 220  Predicted domain architecture of ALOX5AP and CG33177 determined using Interpro and Pfam.

Figure 221  HMM profile of CG33177 aligned against chain A of 2H8A, a template structure selected for modeling CG33177.

Figure 222  Validation of the CG33177 model: ProQ2 quality score mapped to a 3D model of CG33177 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).

Figure 223  Pairwise alignment of CG33177 and 2Q7M generated from structural superposition with shared secondary structure elements and conserved residues highlighted.

Figure 224  Pairwise alignment of CG33177 and 2Q7M generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

Figure 225  ALOX5AP (2Q7M, cyan-blue) superimposed on the predicted structure of CG33177 (green-red).

Figure 226  ALOX5AP (2Q7M, cyan-blue) superimposed on the predicted structure of CG33177 (green-red), with potential matches for conserved functional residues highlighted.

Figure 227  A theoretical D. melanogaster eicosanoid synthesis pathway.
Chapter 1: Introduction

Overview of the eicosanoid biosynthesis pathway

The eicosanoids are a family of biologically active lipids that have been implicated in various signaling pathways, with a central role as mediators of inflammation [1-3]. The canonical eicosanoid biosynthesis pathway begins with the release of fatty acids, primarily arachidonic acid (AA), from membrane phospholipids in response to activation of phospholipase A$_2$ [4, 5]. At this point the canonical pathway diverges, depending on whether the fatty acid substrate is processed by a cyclooxygenase (COX), lipoxygenase (LOX) or a P450 epoxygenase (P450E). The COX enzymes generate prostanoids (prostaglandins, prostacyclins and thromboxanes) and resolvins, whereas the LOX enzymes produce the leukotrienes, hydroperoxyeicosatetraenoic acids (HPETEs), hydroxyeicosatetraenoic acids (HETEs) and lipoxins. There is some cross-talk between the COX and LOX pathways, as both are known to produce hydroxyoctadecadienoic acids (HODEs) The P450 epoxygenase pathway yields epoxyeicosatrienoic acids (EETs). All these downstream products of the canonical eicosanoid pathway branches regulate a diverse variety of signaling pathways, with the COX and LOX subfamilies emerging as important mediators of inflammation and immunity in mammals [6-9].
FIG. 1. The canonical human eicosanoid synthesis pathway. Prostaglandins are highlighted in red, leukotrienes in green and thromboxanes in blue. HETE mediators, which form part of the non-classic eicosanoid pathway, are highlighted in yellow.

The COX branch of the eicosanoid pathway

The COX and LOX branches of the eicosanoid pathway are well-characterized in humans. The COX enzymes (prostaglandin G/H synthases) are encoded by two genes: COX-1 (PTGS1) and COX-2 (PTGS2) [10]. COX-1 is constitutively expressed at a low level by most cells. In contrast, COX-2 is upregulated at sites of inflammation and during tumor progression [10, 11]. COX-1 and COX-2 are membrane-associated heme-containing homodimers that catalyze the conversion of an 18-22 carbon polyunsaturated fatty acid (typically AA, an ω-6, 20-carbon fatty acid) into prostaglandins G₂ and H₂ [12]. Although AA is the preferred substrate, and COX signaling tends to be pro-inflammatory, the COX enzymes are also known to oxygenate eicosapentaenoic acid (EPA, 20:5 n-3), linoleic acid (LA, 18:2 n-6), dihomo-γ-linolenic acid (DHLA) and docosahexaenoic acid (DHA), with varying efficiencies [12-14]. This
initial substrate choice is significant, as it determines whether the downstream prostanoid will be pro- or anti-inflammatory. Fatty acid with two double bonds (e.g., AA) are converted into series-2 prostanoids (e.g., PGH₂, PGE₂), which are associated with pro-inflammatory signaling. In contrast, fatty acid substrates with three or four double bonds (e.g., EPA, DHLA, respectively) yield series-1 and 3 prostanoids (PGH₁, PGE₃), which are associated with anti-inflammatory signaling. Anti-inflammatory resolvins are produced from EPA and DHA, whereas the pro- and anti-inflammatory 9- and 13-HODEs are generated from LA by COX enzymes [15,16].

Non-steroidal anti-inflammatory drugs (NSAIDs) can inhibit the ability of COX enzymes to generate pro-inflammatory downstream products [17]. PGH₂ is subsequently converted to a variety of downstream eicosanoids, including PGD₂, PGE₂, PGF₁α, PGF₂α, PGI₂ (prostacyclin), and the thromboxanes TXA₂ and TXB₂. The downstream processing route depends on the enzymatic machinery present in a given cell type. For example, endothelial cells primarily produce PGI₂, whereas platelets mainly produce TXA₂ [18].

**The LOX branch of the eicosanoid pathway**

The LOX pathway operates independently from the COX pathway, diverging at the initial fatty acid processing stage. In humans, LOX enzymes are encoded by several genes (ALOX5, ALOX12, ALOX12B, ALOX15). Mammalian LOX enzymes catalyze the stereospecific dehydrogenation and dioxygenation of a polyunsaturated fatty acid (PUFA), leading to the formation of a hydroxyperoxide containing a 1-hydroxyperoxy-2-trans-4-cis-pentadiene fragment [19, 20]. The mammalian LOXs have been named according to the position of the carbon at which the oxygen molecule is inserted into AA, the primary fatty acid substrate, e.g., as 5-, 12-, or 15-LOX) [21]. The primary products are 5S-, 12S-, or 15S-HPETE, which can be further reduced by glutathione peroxidase to hydroxy forms (5-, 12-, 15-HETE), respectively.
[21, 23]. 5-LOX is the only LOX that produces leukotrienes, and it is notable that its catalytic activity requires the presence of a second protein referred to alternatively as Arachidonate 5-Lipoxygenase Activating Protein (ALOX5AP) or Five-Lipoxygenase Activating Protein (FLAP), though the precise function of FLAP is unknown at this time [19, 22]. Following the conversion of a fatty acid (e.g., AA) into 5-HPETE, processing by 5-LOX and other downstream leukotriene syntheses enzymes results in the formation of leukotrienes LTB₄, LTC₄, LTD₄, and LTE₄. The leukotrienes are clinically significant, as they are implicated in the allergic response and maintenance of chronic inflammation. 15-LOX (and to a lesser extent 5-LOX) produce the lipoxins LXA and LXB [23].

**Non-classic Eicosanoids**

The cyclooxygenase and lipoxygenase branches of the eicosanoid synthesis pathway have been thoroughly studied in human and other mammalian systems. As a result, the major end products of these pathways (i.e., prostaglandins, thromboxanes and leukotrienes) are referred to as canonical eicosanoids. However, in recent years additional classes of oxygenated PUFA derivatives have been identified as minor or alternative products of the eicosanoid synthesis pathways, including eoxins, hepoxilins, resolvins, lipoxins, epi-lipoxins, levuglandins, oxoeicosanoids, hydroxyeicosatrienoic acids (HETEs), epoxyeicosatrienoic acids (EETs) and endocannabinoids. Together, this group represents a large and growing family of lipid mediators generally referred to as non-classic eicosanoids.

Eoxins are pro-inflammatory eicosanoids derived from arachidonic acid by arachidonate 15-lipoxygenase (ALOX15) [24]. The eoxins derive their name from eosinophils, the cell type in which they were originally discovered in abundance [25]. Eoxins resemble the cysteinyi
leukotrienes produced by ALOX5 but have a different three-dimensional structure (i.e., eoxins are 14,15-leukotriene analogs) and are generated by cells in response to different stimuli [26].

Hepoxilins are hydroxy-epoxy eicosanoids (epoxyalcohols) which contain both hydroxyl and epoxy groups, the latter spanning the C11-C12 double bond, and unlike leukotrienes and lipoxins, none of the double bonds are conjugated [27]. Hepoxilins are produced by the epidermal subfamily of mammalian lipoxygenase enzymes ALOX12B and ALOXE3 (12-R-LOX and eLOX-3, respectively), which are preferentially expressed in the skin and several other epithelial tissues [28]. These enzymes act in concert to convert fatty acid substrates via R-HPETEs to specific epoxyalcohol derivatives [29]. 12R-LOX and eLOX-3 are unique compared to all other currently identified LOXs in that 12R-LOX is the only mammalian LOX that forms products with R-chirality and eLOX-3 functions as a hydroperoxide isomerase by exhibiting only latent dioxygenase activity [30].

Lipoxins, epi-lipoxins and resolvins are potent anti-inflammatory eicosanoids which modulate the inflammation response in mammals. Lipoxins and epi-lipoxins (stereoisomers of lipoxins) are derived from arachidonic acid and contain three hydroxyl residues and four double bonds [31]. Lipoxins synthesis typically involves a LOX enzyme which adds a 15S-hydroxyl residue to arachidonic acid, whereas synthesis of the epi-lipoxins involves aspirin-pretreated COX-2 or a CYP450 enzyme capable of adding a 15R-hydroxyl residue to arachidonic acid [32]. Resolvins are a structurally diverse class of oxygenated derivatives of the omega-3 PUFAs (EPA, DHA, DPA), which derive their name from the fact that this family of lipid mediators is known to be produced during the resolution phase of the inflammation response [33, 34].

Levuglandins are reactive γ-ketoaldehydes formed by the spontaneous rearrangement of prostaglandin H (PGH) [35, 36]. Levuglandins are highly effective at cross-linking proteins and
DNA, which makes detection difficult as newly-produced levuglandins are rapidly sequestered by available nucleophiles, potentially generating heterogeneous protein aggregates which may have pathologic consequences if prostanoids are aberrantly overexpressed [35].

Oxo-eicosanoids refer to oxygenated derivatives of eicosatetraenoic acid (e.g., 5-Oxo-6,8,11,14-eicosatetraenoic acid, “5-oxo-ETE”), which are produced by a dehydrogenase acting on HETE intermediates produced by LOX enzymes [37]. Oxo-eicosanoids are known to produce pro-inflammatory effects similar to leukotrienes, but interact with different cell surface receptors [38, 39].

Epoxycosatrienoic acids (EETs) and hydroxyeicosatrienoic acids (HETEs) are derived from arachidonic acid by a subset of CYP450 enzymes [40]. The CYP450 superfamily includes various heme-containing enzymes that are essential for basic metabolism in humans and ubiquitous among all five kingdoms of life [41]. In the context of eicosanoid synthesis, CYP450 enzymes are necessary for the synthesis of prostacyclins and thromboxanes (i.e., prostacyclin synthase and thromboxane A synthase are both CYP450 family members) [42]. However, CYP450 enzymes are further relevant to arachidonic acid metabolism given that two distinct subfamilies (ω-hydroxylases and epoxygenases) are also able to form metabolites derived from arachidonic acid [40, 43, 44]. Arachidonic acid is metabolized by the CYP450 ω-hydroxylase family of enzymes in mammalian cells to 7-, 10-, 12-, 13-, 15-, 16-, 17-, 18-, 19-, and 20-HETEs, with the pro-inflammatory 20-HETE being the primary metabolite [45]. In contrast, the epoxygenase CYP450 enzymes metabolize arachidonic acid by olefin epoxidation, resulting in four regioisomeric epoxycosatrienoic acids (EETs): 5,6-EET, 8,9-EET, 11,12-EET, and 14,15-EET [46]. Each of these regioisomers can be formed as either an R,S or S,R enantiomer given that the epoxide group can attach to the double bonds in two separate configurations [47].
CYP450 enzymes also convert EPA into epoxy-derivatives, and endocannabinoids containing 11,12- and 14,15-EETs. [48].

The various non-classic eicosanoids discussed above includes clinically important autocrine and paracrine mediators of inflammation and homeostasis that operate in parallel with the classic eicosanoids.

**Eicosanoid biosynthesis and signaling in insects**

Extensive scientific literature from the last 25 years, based primarily in mammalian cell culture and rodent models, has established the central roles of eicosanoids in the response to infection (acute inflammation) as well as in chronic diseases in humans and various animal models [47-49]. A number of vertebrate and cell culture models that mimic human diseases are now available for studying inflammation. Despite an overwhelming increase in our knowledge regarding mechanisms underlying inflammation, it has been difficult to genetically map the signaling pathways targeted by eicosanoids. A primary reason for this gap is the large number of prostanoids and lipid mediators and the complexity of the mammalian system itself.

In order to better understand the COX and LOX pathways, identification of suitable model organisms is a priority. The COX family enzymes are highly conserved across the animal kingdom, with orthologs found in the primitive marine corals as well as the higher vertebrates [50]. Traditional genomic analyses have failed to identify COX orthologs in the known genomes of insects, unicellular organisms, or plants, although prostaglandins, their primary products, have been found in some of these organisms [51, 52]. Recent studies have suggested that an insect cyclooxygenase may exist in orthologs of the gene *Pxt* [53, 54]. LOX enzymes display an even broader degree of conservation, with orthologs found across the animal kingdom and in a variety of plants [55]. Interestingly enough, sequence and biochemical analyses have failed to identify
any insect LOX orthologs, with the sole exception being a report of LOX activity in the primitive insect, *Thermobia domestica* [56].

As mediators of the immune and inflammatory response in mammals, eicosanoids remain an important area of research. A better understanding of these enzymes (and their biosynthesis), may yield therapeutics that mitigate inflammation or compensate for defects in the immune response. The canonical eicosanoid synthesis pathway (prostanoids and leukotrienes) is well-characterized in humans and in mammalian models. However, researchers would benefit from having access to a low cost, high volume model to study this pathway, *i.e.*, *D. melanogaster*. In order to better understand the *in vivo* roles of mammalian COX and LOX pathways, we examined if enzymes of these pathways are encoded in the *Drosophila* genome. Flies do not produce C20 PUFAs, but nevertheless possess the ability to produce lipid mediators [57]. While an initial BLAST searches based on sequence analysis alone failed to identify likely fly orthologs for these enzymes, a more rigorous approach using iterative sequence searches combined with structural modeling revealed a surprising degree of similarity and apparent conservation of catalytic as well as other key functional residues. Our study suggests that insects possess a functional eicosanoid pathway and open up possibilities of utilizing a powerful model organism for eicosanoid research.
Chapter 2: Materials and Methods

Overview

The flowchart provided below summarizes the general protocol of the present study with the various tools used to identify and characterize putative D. melanogaster eicosanoid synthesis enzymes which may be orthologs for the classic eicosanoid synthesis enzymes discussed in detail below.

**FIG. 2.** Summary of the workflow and tools used to carry out the present study.
Selection and Characterization of Human Eicosanoid Synthesis Enzymes

A canonical eicosanoid synthesis pathway in humans was obtained from the NCBI BioSystems Database by compiling enzymes identified in the Eicosanoid Synthesis (BSID: 198888) and Arachidonic Acid pathway maps (BSID: 829971) [58]. Protein sequence(s) for each gene associated with the canonical human eicosanoid synthesis pathway were obtained from the NCBI RefSeq database [59]. In each case, either the sole protein product or the major isoform was selected as the representative sequence for each enzyme. Known and high-confidence predicted orthologs for each of the canonical mammalian eicosanoid synthesis enzymes were identified using the KEGG Orthology database [60]. A multiple sequence alignment (MSA) for each target enzyme with the selected orthologs was then generated using PROMALS3D, followed by manual refinement of the alignment [61]. Domains and active sites on each query and putative D. melanogaster ortholog enzyme were annotated (or predicted, in the case of the putative orthologs) using the Pfam and InterPro databases [62, 63]. Consensus secondary structure profiles were generated using the PsiPred, JPred and PSSPred servers [64-66] to confirm identified functional domains and to ascertain domain boundaries. The mammalian eicosanoid synthesis enzymes selected as targets for this study are summarized by Table 1.
TABLE 1: Human Eicosanoid Synthesis Enzymes

<table>
<thead>
<tr>
<th>NCBI Gene Accession (UniProt ID)</th>
<th>NCBI Protein RefSeq Accession (length)</th>
<th>Common Name</th>
<th>Representative Crystal Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALOX5 (P09917)</td>
<td>NP_000689.1 (674 aa)</td>
<td>Arachidonate 5-lipoxygenase</td>
<td>3O8Y (2.39 Å)</td>
</tr>
<tr>
<td>ALOX12 (P18054)</td>
<td>NP_000688.2 (663 aa)</td>
<td>Arachidonate 12-lipoxygenase, S-type</td>
<td>3D3L (2.60 Å)</td>
</tr>
<tr>
<td>ALOX12B (O75342)</td>
<td>NP_001130.1 (701 aa)</td>
<td>Arachidonate 12R-lipoxygenase, R-type</td>
<td>None available.</td>
</tr>
<tr>
<td>ALOX15 (P16050)</td>
<td>NP_001131.3 (662 aa)</td>
<td>Arachidonate 15-lipoxygenase</td>
<td>2P0M (2.40 Å)</td>
</tr>
<tr>
<td>ALOX5AP (P20292)</td>
<td>NP_001620.2 (161 aa)</td>
<td>Arachidonate 5-lipoxygenase Activating Protein</td>
<td>2Q7M (4.00 Å)</td>
</tr>
<tr>
<td>LTA4H (P09960)</td>
<td>NP_000886.1 (611 aa)</td>
<td>Leukotriene A4 hydrolase</td>
<td>3CHO (1.80 Å)</td>
</tr>
<tr>
<td>LTC4S (Q168730)</td>
<td>NP_665874.1 (150 aa)</td>
<td>Leukotriene C4 synthase</td>
<td>3PCV (1.90 Å)</td>
</tr>
<tr>
<td>GGT1 (P19440)</td>
<td>NP_001275762.1 (569 aa)</td>
<td>Leukotriene C4 hydrolase</td>
<td>4GDX (1.67 Å)</td>
</tr>
<tr>
<td>DPEP1 (P16444)</td>
<td>NP_001121613.1 (411 aa)</td>
<td>Dipeptidase 1</td>
<td>1ITQ (2.30 Å)</td>
</tr>
<tr>
<td>PTGS1 (P23219)</td>
<td>NP_000953.2 (599 aa)</td>
<td>Cyclooxygenase 1</td>
<td>1Q4G (2.00 Å)</td>
</tr>
<tr>
<td>PTGS2 (P35354)</td>
<td>NP_000954.1 (604 aa)</td>
<td>Cyclooxygenase 2</td>
<td>3NT1 (1.73 Å)</td>
</tr>
<tr>
<td>PTGDS (P41222)</td>
<td>NP_000945.3 (190 aa)</td>
<td>Prostaglandin D synthase</td>
<td>4IMO (1.88 Å)</td>
</tr>
<tr>
<td>PTGIS (Q16647)</td>
<td>NP_000952.1 (500 aa)</td>
<td>Prostacyclin synthase</td>
<td>2IAG (2.15 Å)</td>
</tr>
<tr>
<td>PTGES (O14684)</td>
<td>NP_004896.1 (152 aa)</td>
<td>Microsomal prostaglandin E synthase 1</td>
<td>4AL0 (1.16 Å)</td>
</tr>
<tr>
<td>PTGES2 (Q9H7Z7)</td>
<td>NP_079348.1 (377 aa)</td>
<td>Microsomal prostaglandin E synthase 2</td>
<td>2PBJ (2.80 Å)</td>
</tr>
<tr>
<td>PTGES3 (Q15185)</td>
<td>NP_006592.3 (160 aa)</td>
<td>Cytosolic prostaglandin E synthase</td>
<td>1EJF (2.49 Å)</td>
</tr>
<tr>
<td>TBOXAS1 (P24557)</td>
<td>NP_001052.2 (534 aa)</td>
<td>Thromboxane A2 synthase</td>
<td>None available.</td>
</tr>
<tr>
<td>AKR1C3 (P42330)</td>
<td>NP_003730.4 (323 aa)</td>
<td>Prostaglandin F synthase</td>
<td>1RY0 (1.69 Å)</td>
</tr>
<tr>
<td>CBR1 (P16152)</td>
<td>NP_001748.1 (277 aa)</td>
<td>Carbonyl reductase 1</td>
<td>3BHM (1.80 Å)</td>
</tr>
<tr>
<td>HPGD (P15428)</td>
<td>NP_000851.2 (266 aa)</td>
<td>15-hydroxyprostaglandin dehydrogenase NAD(+)</td>
<td>2GDZ (1.65 Å)</td>
</tr>
</tbody>
</table>
Identification of Potential Eicosanoid Synthesis Enzymes in *D. melanogaster*

We used several different sequence analysis algorithms to identify candidate *Drosophila* orthologs for each human enzyme. Specifically, each human enzyme reference sequence was queried against a subset of the *D. melanogaster* NCBI non-redundant sequence database using PSI-BLAST, DELTA-BLAST or JACKHMMER [67-69]. DELTA-BLAST is a more sensitive variant of the traditional BLAST algorithm that incorporates domain information from pre-constructed position specific scoring matrices (PSSMs) in order to improve detection of homology. JACKHMMER is an alternative sequence analysis and alignment tool that relies on hidden Markov model (HMM) profiles in order to boost sensitivity compared to the traditional BLAST algorithms. JACKHMMER converts a query sequence into a HMM profile using a substitution matrix and gap penalties, which is then searches against a sequence database (*e.g.*, the NCBI refseq database). Sequences that score above the inclusion threshold from this first search are then aligned and used to construct a second HMM profile, which may then be queried once again against the same database [70]. The iterative nature of this filtering process typically results in a HMM profile that is able to detect distant homologs that may not be detected by the traditional BLAST algorithms. For this study, BLAST and HMMer algorithms were leveraged as part of a combined approach in order to identify candidates which may be remote homologs for the human eicosanoid synthesis enzymes.

Each search was conducted with three iterations using the default parameters. *D. melanogaster* candidates with statistically significant full-length alignments to the query were selected after each round. These candidates were then analyzed using the same secondary structure and domain analysis tools as discussed in the section above. Each candidate was then aligned against the previously generated MSA (containing the human enzyme and
known/predicted orthologs) in order to screen out hits that lacked known functional residues and/or those with substantially dissimilar secondary structure or domain architecture.

**In silico Modeling of D. melanogaster Candidates**

Full-length three-dimensional models were generated for each top-scoring *D. melanogaster* candidate protein identified in the previous stage using the MODELLER, LOMETS and I-TASSER software packages [71-73]. MODELLER, a homology modeling tool, predicts the tertiary structure of a protein sequence based on satisfaction of spatial restraints derived from sequentially similar templates whose structure is known (*i.e.*, a sequence-sequence comparison). In contrast, I-TASSER utilizes a threading algorithm (*i.e.*, a sequence-structure comparison) and fragment assembly based on replica-exchange Monte Carlo simulations; it also builds unaligned primary sequence regions using *ab initio* methods. LOMETS is a meta-server that generates and ranks modeling results produced using several alternative threading algorithms which are executed in parallel. The accuracy of homology modeling depends largely on the target-template alignment, with ≥40% sequence identity generally yielding high quality models and ≥30% being the lower threshold for an acceptable model [74]. However, threading based modeling tools are generally less sensitive to sequence identity differences and some packages (*e.g.*, I-TASSER) are known to generate high quality models using templates sharing as little as 20-30% sequence identity with the target sequence [75, 76].

Each *D. melanogaster* candidate sequence was searched against the Protein Data Bank database of published structures using PSI-BLAST and JACKHMMER [67, 69]. Several potential templates with ≥30%, full-length sequence identity were aligned against each candidate using PROMALS3D and evaluated [61]. In instances where a single full-length template was available, MODELLER was used to generate a three-dimensional model of the candidate. If a
suitable single full-length template was not found, either due to a lack of coverage or low sequence identity within the aligned range, I-TASSER was used to generate a composite model based on several templates. In each case, the initial model was used to generate a refined full atomic model using ModRefiner, which also optimizes sidechain placement [77]. All of the refined models were then evaluated using the PROCHECK, ProSa, ProQ2 and MODFOLD4 servers [78-80]. Loops and unstructured N- and C-terminal regions in these models were optimized using ab initio methods (e.g., MODELLER’s loop routines or QUARK), refined using ModRefiner and reevaluated as necessary [81]. Unsuitable models were rejected and rebuilt using alternatives templates according to the preceding protocol. Models were analyzed using the surface property tools in the Chimera and PyMol visualization and structural modeling packages, in addition to online tools including, COFACTOR, and COACH [82-87].
Chapter 3: Results

Identification and Characterization of Candidate Eicosanoid Synthesis Enzymes in *D. melanogaster*

A purely sequence-based analysis of the *D. melanogaster* genome (Release 6.04, February 24th, 2015) using traditional sequence analysis tools (e.g., BLASTp) fails to identify likely orthologs for a majority of the canonical human eicosanoid synthesis enzymes. However, using a more sensitive approach based on iterative HMMER searches and structural modeling, we have uncovered potential candidates in *D. melanogaster* that may function as eicosanoid synthesis enzymes. This analysis is supported by the fact that a majority of these candidates share a highly similar tertiary structure with the human target enzymes and also appear to possess the necessary catalytic residues. A summary of our findings is provided as Table 2.

**TABLE 2. Summary of *D. melanogaster* genes encoding potential orthologs for the mammalian eicosanoid synthesis enzymes examined during this study.**

<table>
<thead>
<tr>
<th>Human Gene</th>
<th><em>D. melanogaster</em> Candidate(s)</th>
<th>CG Identifier</th>
<th>Flybase Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTGS1, PTGS2</td>
<td>CG4009, isoform B</td>
<td>CG4009</td>
<td>FBgn0038469</td>
</tr>
<tr>
<td></td>
<td>Pxt (“Peroxinectin-like”)</td>
<td>CG7660</td>
<td>FBgn0261987</td>
</tr>
<tr>
<td></td>
<td>Cardinal</td>
<td>CG6969</td>
<td>FBgn0263986</td>
</tr>
<tr>
<td>PTGDS</td>
<td>Nlaz (“Neural Lazarillo”)</td>
<td>CG33126</td>
<td>FBgn0053126</td>
</tr>
<tr>
<td>HPGDS</td>
<td>GST S1 (“Glutathione S transferase S1”)</td>
<td>CG8938</td>
<td>FBgn0010226</td>
</tr>
<tr>
<td>PTGIS</td>
<td>CYP450-4D2</td>
<td>CG3466</td>
<td>FBgn0011576</td>
</tr>
<tr>
<td>PTGES</td>
<td>MGST-like, isoform A (“microsomal glutathione S-transferase-like”)</td>
<td>CG1742</td>
<td>FBgn0025814</td>
</tr>
<tr>
<td>PTGES2</td>
<td>SupRef(2)p [“Suppressor of Ref(2)p”]</td>
<td>CG4086</td>
<td>FBgn0004465</td>
</tr>
<tr>
<td>PTGES3</td>
<td>CG16817, isoform A</td>
<td>CG16817</td>
<td>FBgn0037728</td>
</tr>
<tr>
<td>TBXAS1</td>
<td>CYP450-9H1</td>
<td>CG17577</td>
<td>FBgn0033775</td>
</tr>
<tr>
<td>AKR1A1, AKR1B1, AKR1C3</td>
<td>CG6084, isoform D</td>
<td>CG6084</td>
<td>FBgn0086254</td>
</tr>
<tr>
<td>CBR1</td>
<td>CBR, isoform B (“carbonyl reductase”)</td>
<td>CG11200</td>
<td>FBgn0034500</td>
</tr>
<tr>
<td>ALOX5, ALOX12, ALOX12B, ALOX15</td>
<td>None identified</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>ALOX5AP</td>
<td>CG33177</td>
<td>CG33177</td>
<td>FBgn0053177</td>
</tr>
<tr>
<td>LTA4H</td>
<td>CG10602</td>
<td>CG10602</td>
<td>FBgn0032721</td>
</tr>
<tr>
<td>LTC4S</td>
<td>CG33178, isoform A</td>
<td>CG33178</td>
<td>FBgn0053178</td>
</tr>
<tr>
<td>GGT1</td>
<td>GGT, isoform A (“gamma-glutamyl transpeptidase”)</td>
<td>CG6461</td>
<td>FBgn0030932</td>
</tr>
<tr>
<td>DPEP1</td>
<td>CG6154, isoform C</td>
<td>CG6154</td>
<td>FBgn0039420</td>
</tr>
<tr>
<td>HPGD</td>
<td>CG18814</td>
<td>CG18814</td>
<td>FBgn0042137</td>
</tr>
<tr>
<td>GPX1</td>
<td>PHGPx, isoform A</td>
<td>CG12013</td>
<td>FBgn0035438</td>
</tr>
<tr>
<td>CPA1</td>
<td>CG18585, isoform A</td>
<td>CG18585</td>
<td>FBgn0031929</td>
</tr>
</tbody>
</table>

**Group 1: The High Scoring Matches**

Eight of the putative orthologs identified in this search are notable for being particularly high confidence matches: CG1742, CG4086, CG6084, CG10602, CG6461, CG6154, CG12013 and CG18585. Members of this set share a minimum of 30% sequence identity with the human enzyme, the same overall fold, and display conservation of one or more functional residues.

**Prostaglandin E Synthase (PTGES)**

The first gene in this set, CG1742 (“MGST-like, isoform A,” NP_524696.1), encodes a 152 amino acid protein predicted to have a single pfam domain (MAPEG, PF01124) spanning the 18 to 148 amino acid region. This protein displays 34% identity and 53% similarity to the sequence of human prostaglandin E synthase (PTGES, NP_004869.1). PTGES is an integral membrane protein that operates as a homotrimer, and is known to catalyze the oxidoreduction of prostaglandin endoperoxide H2 (PGH2) to prostaglandin E2 (PGE2). PTGES is also characterized by the presence of a single MAPEG domain, which spans the region 16-146. PTGES (e.g., PDB 4AL0). The predicted structure of CG1742 are nearly identical and display an RMSD 1.15 Å when superimposed, as shown in Fig. 9 below.
PTGES is a glutathione-binding protein, and residues R38, R70, E77, R110, Y117, R126, and Y130 are associated with this activity [88]. Structural superposition reveals that CG1742 has an aligned match for each of these residues, R40, R71, E78, R111, Y118, R128 and F132. With respect to the last residues, a tyrosine to phenylalanine substitution is likely to be functionally equivalent as phenylalanine and tyrosine have side chains with a single aromatic ring of similar volume. Phenylalanine differs only in that it lacks the hydroxyl group in the ortho position on the benzene ring. Previous studies report that the aromatic ring of Y130 is likely important for PGH2-binding and an analog for this structural element is provided by F132 [89].

A. Predicted Secondary Structure

B. Domain Architecture

FIG. 3. Predicted secondary structure of PTGES and CG1742 calculated using PROMALS3D.

FIG. 4. Predicted domain architecture of PTGES and CG1742 determined using Interpro and Pfam.
C. Model Generation and Validation

1. Generation

HHpred, Phyre2 and LOMETS were used to identify suitable templates for modeling CG1742. All three tools identify chain A of 4YL0 (human prostaglandin E synthase) as a high-scoring match suitable for single-template modeling.

<table>
<thead>
<tr>
<th>CG1742 Range</th>
<th>Template Range</th>
<th>% Identity</th>
<th>% Similarity</th>
<th>E-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-151</td>
<td>2-149</td>
<td>36%</td>
<td>58%</td>
<td>5.3e-22</td>
</tr>
</tbody>
</table>

![FIG. 5. HMM profile of CG1742 aligned against chain A of 4YL0, a template structure selected for modeling CG1742.](image)

2. Validation

The quality of the CG1742 model was evaluated using ProQ2 and ProSA. The per-residue S-Score plot generated by ProQ2 was mapped to the CG1742 model by saving the S-Score assigned to each residue as a B-factor field parameter and then rendering a color-coded ribbon view in Chimera based upon this parameter. As illustrated by the image below, overall model quality following refinement is high. The quality of the CG1742 model was also evaluated using ProSA and found to display a quality level comparable to experimentally determined structures having a comparable length (i.e., a global Z-score of -4.07). The ProSA per-residue energy score graph is also favorable, showing a score that is consistently negative over a window size of 40 residues.
FIG. 6. Validation of the CG1742 model: ProQ2 quality score mapped to a 3D model of CG1742 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).

D. Structural Analyses

A structural superposition reveals that the CG1742 model and human prostaglandin E synthase (4AL0) display a similar secondary structure architecture with substantial full-length overlap. The pairwise alignment generated by this structural superposition also indicates conservation of residues having similar physiochemical properties.
1. Secondary Structure

![Pairwise alignment of CG1742 and 4AL0 generated from structural superposition with shared secondary structure elements and conserved residues highlighted.](image)

FIG. 7. Pairwise alignment of CG1742 and 4AL0 generated from structural superposition with shared secondary structure elements and conserved residues highlighted.

2. Physiochemical Properties

![Pairwise alignment of CG1742 and 4AL0 generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.](image)

FIG. 8. Pairwise alignment of CG1742 and 4AL0 generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.
E. Analysis of Functional Residues

<table>
<thead>
<tr>
<th>PTGES (PDB: 4AL0) and CG1742 Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PTGES Structure</strong></td>
</tr>
<tr>
<td><img src="image1" alt="PTGES Structure" /></td>
</tr>
</tbody>
</table>

**FIG. 9.** PTGES (4AL0, cyan-blue) superimposed on the predicted structure of CG1742 (green-red). RMSD: 1.015 Å.

**FIG. 10.** PTGES (4AL0, cyan-blue) superimposed on the predicted structure of CG1742 (green-red), with potential matches for conserved functional residues highlighted.
F. Summary Table

<table>
<thead>
<tr>
<th></th>
<th>Length (AA)</th>
<th>Domain Architecture (Pfam, range)</th>
<th>Functional Residues (aligned matches in D. melanogaster)</th>
<th>Sequence ID%</th>
<th>Structural Overlap (RMSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prostaglandin-E-synthase (PTGES, NP_004869.1, PDB: 4AL0)</td>
<td>152</td>
<td>MAPEG (PF01124) 16-146</td>
<td>R38, E77, R70, R110, Y117, R126, Y130</td>
<td>34% ID</td>
<td>1.015 Å</td>
</tr>
<tr>
<td>MGST-like (CG1742, NP_524696.1)</td>
<td>152</td>
<td>MAPEG (PF01124) 18-148</td>
<td>R40, E78, R71, R111, Y118, R128, F132</td>
<td>53% SIM</td>
<td></td>
</tr>
</tbody>
</table>

Prostaglandin Synthase E 2 (PTGES2)

The second gene in this set, CG4086 (“Su(P) Suppressor of ref(2)P sterility,” NP_524116.2) encodes a 417 amino acid protein. CG4086 displays 33% identity and 49% similarity to the sequence of human prostaglandin E synthase 2 (PTGES2, NP_079348.1). PTGES2 is membrane-associated, as opposed to PTGES which is an integral membrane protein, though both catalyze the same reaction. Despite the functional similarities, PTGES2 possesses a substantially different domain architecture and tertiary structure. PTGES2 possesses GST domains (PF13417 and PF14497) spanning amino acids 104-175 and amino acids 201-368. CG4086 shares a highly similar domain architecture with a Glutaredoxin domain (PF00462), which is functionally related to the GST domain, spanning amino acids 125-184 and a GST domain (PF14497) spanning amino acids 248-396. PTGES2’s catalytic activity requires C110 [90]. The crystal structure of truncated PTGES2 has been published (1Z9H), with the N-terminal membrane-associated region omitted (residues 1 to 99). Superposition of this truncated structure against the predicted structure of CG4086 reveals an RMSD of 0.814 Å. CG4086’s C133 overlaps with PTGES2’s C110, suggesting a match for this catalytic residue.
## A. Predicted Secondary Structure

![Image of predicted secondary structure]

**FIG. 11.** Predicted secondary structure of PTGES2 and CG4086 calculated using PROMALS3D.
B. Domain Architecture

FIG. 12. Predicted domain architecture of PTGES2 and CG4086 determined using Interpro and Pfam.

C. Model Generation and Validation

1. Generation

HHpred, Phyre2 and LOMETS were used to identify suitable templates for modeling CG4086. All three tools identify chain A of 1Z9H (Macaca fascicularis prostaglandin E synthase-2) as a high-scoring match suitable for single-template modeling. The human and M. fascicularis orthologs of prostaglandin E synthase-2 are membrane-bound enzymes with a short N-terminal transmembrane domain. The cytosolic portion of M. fascicularis prostaglandin E synthase-2 begins at approximately position 80 of the full-length sequence and includes the remainder of the protein. The 1Z9H structure provides data for the majority of this cytosolic domain (i.e., 100-373) of the full-length M. fascicularis prostaglandin E synthase-2 sequence. The cytosolic domain of M. fascicularis prostaglandin E synthase-2 is identified as a template suitable for modeling 120-397 of the CG4086 sequence. The N-terminus of CG4086 is predicted to have a transmembrane or membrane-associated region, suggesting a similar shared architecture compared to human and M. fascicularis prostaglandin E synthase-2.
FIG. 13. HMM profile of CG4086 aligned against chain A of 1Z9H, a template structured selected for modeling CG4086.

2. Validation

The quality of the CG4086 model was evaluated using ProQ2 and ProSA, as described above in the discussion of CG1742. Again, overall model quality is high (a ProSA global Z-score of -6.52). The ProSA per-residue energy score graph is also favorable, showing a score that is consistently negative over a window size of 40 residues. The N-terminal region of the model is indicated to have a comparatively low quality. However, this may be explained by the fact that this segment connects to the predicted transmembrane or membrane-associated N-terminal domain which was for which no suitable template exists at this time. The N-terminal segment may be a disordered linker to the membrane-bound/associated N-terminus. In any event, this segment cannot be reliably modeled using existing homology or ab initio methods.
FIG. 14. Validation of the CG4086 model: ProQ2 quality score mapped to a 3D model of CG4086 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).

D. Structural Analyses

A structural superposition reveals that the CG4086 model and *M. fascicularis* prostaglandin E synthase-2 (1Z9H) display a similar secondary structure architecture across the modeled region with substantial full-length overlap. The pairwise alignment generated by this structural superposition also indicates substantial conservation of residues having similar physiochemical properties.
FIG. 15. Pairwise alignment of CG4086 and 1Z9H generated from structural superposition with shared secondary structure elements and conserved residues highlighted.

1. Physiochemical Properties
FIG. 16. Pairwise alignment of CG4086 and 1Z9H generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

E. Analysis of Functional Residues

<table>
<thead>
<tr>
<th>PTGES (PDB: 1Z9H) and CG4086 Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PTGES Structure</strong></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

FIG. 17. Truncated PTGES2 (1Z9H, cyan-blue) superimposed on the predicted structure of CG4086 (green-red). RMSD: 0.814 Å.
PTGES2 (PDB: 1Z9H) and CG4086 Model Superimposed

FIG. 18. Truncated PTGES2 (1Z9H, cyan-blue) superimposed on the predicted structure of CG4086 (green-red), with potential matches for conserved functional residues highlighted.

F. Summary Table

TABLE 4. Summary of features shared by PTGES2 and potential D. melanogaster ortholog CG4086.

<table>
<thead>
<tr>
<th></th>
<th>Length (AA)</th>
<th>Domain Architecture (Pfam, range)</th>
<th>Functional Residues (aligned matches in D. melanogaster)</th>
<th>Sequence ID%</th>
<th>Structural Overlap (RMSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prostaglandin-E-synthase 2 (PTGES2, NP_079348.1, PDB: 1Z9H)</td>
<td>377</td>
<td>GST-N3 domain (PF13417) 104-175</td>
<td></td>
<td>C110</td>
<td>33% ID 49% SIM 0.814 Å</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GST-C3 (PF14497) 201-368</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suppressor of ref(2)P sterility (CG4086, NP_524116.2)</td>
<td>417</td>
<td>Glutaredoxin domain (PF00462) 125-184</td>
<td></td>
<td>C133</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GST-C domain (PF14497) 248-396</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
AKR1A1

The third gene in this set, CG6084, encodes a 316 amino acid protein that is a high confidence match for the human AKR1A1 [“alcohol dehydrogenase NADP(+)”, NP_001619.1]. This enzyme, along with enzymes encoded by the related genes in this family (e.g., AKR1B5 and AKR1C3), catalyzes the NADPH-dependent reduction of a variety of aromatic and aliphatic aldehydes to their corresponding alcohols. In particular, AKR1B1 and AKR1B5 are known to function as prostacyclin F synthases, converting prostaglandin endoperoxide H₂ (PGH₂) to prostaglandin F₂α (PGF₂α) [91]. CG6084 and AKR1A1 share identical domain architectures [a single aldo/keto reductase family (PF00248) domain], as well as 48% identity and 62% similarity at the sequence level. AKR1A1 function requires the presence of three residues for its catalytic activity, Y49, K78 and H111 [92]. CG6084 possesses analogs for this triad of residues in the form of Y50, K79 and H112. In addition, the structural overlap of the predicted structure for CG6084 and human AKR1A1 displays an RMSD of 0.847 Å.
A. Predicted Secondary Structure

FIG. 19. Predicted secondary structure of AKR1A1 and CG6084 calculated using PROMALS3D.

B. Function/Domain Analysis

FIG. 20. Predicted domain architecture of AKR1A1 and CG6084 determined using Interpro and Pfam.
C. Model Generation and Validation

1. Generation

HHpred, Phyre2 and LOMETS were used to identify suitable templates for modeling CG6084. All three tools identify chain A of 2ALR (human aldehyde reductase, AKR1A1) as a high-scoring match suitable for single-template modeling.

<table>
<thead>
<tr>
<th>CG6084 Range</th>
<th>Template Range</th>
<th>% Identity</th>
<th>% Similarity</th>
<th>E-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-316</td>
<td>5-324</td>
<td>51%</td>
<td>73%</td>
<td>1.3e-98</td>
</tr>
</tbody>
</table>

**FIG. 21.** HMM profile of CG6084 aligned against chain A of 2ALR, a template structure selected for modeling CG6084.

2. Validation

The quality of the CG6084 model was evaluated using ProQ2 and ProSA, as described above in the discussion of CG1742. Again, overall model quality is high (a ProSA global Z-score...
of -10.94). The ProSA per-residue energy score graph is also favorable, showing a score that is consistently negative over a window size of 40 residues.

**FIG. 22.** Validation of the CG6084 model: ProQ2 quality score mapped to a 3D model of CG6084 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).

**D. Structural Analyses**

A structural superposition reveals that the CG6084 model and human aldehyde reductase (2ALR) display a similar secondary structure architecture across the modeled region with
substantial full-length overlap. The pairwise alignment generated by this structural superposition also indicates substantial conservation of residues having similar physiochemical properties.

1. Secondary Structure

![Pairwise alignment of CG6084 and 2ALR generated from structural superposition with shared secondary structure elements and conserved residues highlighted.](image)

**FIG. 23.** Pairwise alignment of CG6084 and 2ALR generated from structural superposition with shared secondary structure elements and conserved residues highlighted.
2. Physiochemical Properties

FIG. 24. Pairwise alignment of CG6084 and 2ALR generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

E. Analysis of Functional Residues

![Image of AKR1A1 (PDB: 2ALR) and CG6084 Model]

FIG. 25. AKR1A1 (2ALR, cyan-blue) superimposed on the predicted structure of CG6084 (green-red). RMSD: 0.847 Å.
AKR1A1 (PDB: 2ALR) and CG6084 Model Superimposed

**FIG. 26.** AKR1A1 (2ALR, cyan-blue) superimposed on the predicted structure of CG6084 (green-red), with potential matches for conserved functional residues highlighted.

### F. Summary Table

**TABLE 5.** Summary of features shared by AKR1A1 and potential *D. melanogaster* ortholog CG6084.

<table>
<thead>
<tr>
<th>Length (AA)</th>
<th>Domain Architecture (Pfam, range)</th>
<th>Functional Residues (aligned matches in <em>D. melanogaster</em>)</th>
<th>Sequence ID%</th>
<th>Structural Overlap (RMSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prostacyclin F synthase (AKR1B1, NP_001619.1, PDB: 1RY0)</td>
<td>316</td>
<td>Aldo/keto reductase family (PF00248) 4-289</td>
<td>Y49, K78, H111</td>
<td>48% ID 62% SIM</td>
</tr>
<tr>
<td>Uncharacterized protein (CG6084, NP_648484.1)</td>
<td>316</td>
<td>Aldo/keto reductase family (PF00248) 9-282</td>
<td>Y50, K79, H112</td>
<td></td>
</tr>
</tbody>
</table>

**LTA4H**

The fourth gene in this set, CG10602 encodes a 613 amino acid protein that displays 44% identity and 58% similarity to human leukotriene A4 hydrolase (LTA4H, NP_000886.1).

LTA4H is an epoxide hydrolase that catalyzes the hydrolysis of the epoxide LTA4 to the diol,
LTB4. This catalytic activity requires the presence of a zinc ion, which is coordinated by the catalytic triad of H296, H300 and E319. Residues E297, D376 and Y384 are also reportedly essential for efficient catalysis [93]. LTA4H’s domain architecture consists of Peptidase family M1 (PF01433) spanning 13-387 and a Leukotriene A4 hydrolase, C-terminal domain (PF09127) spanning amino acids 464-608. CG10602 shares an identical domain organization as well as the critical residues at H293, H297, E316 (the zinc triad) and E294, D374 and Y382. The tertiary structure similarity is readily apparent, as evidenced by a superposition of human LTA4H (1H19) and the predicted stature of CG10602, which results in an RMSD of 0.891 Å.

A. Predicted Secondary Structure

**FIG. 27.** Predicted secondary structure of LTA4H and CG10602 calculated using PROMALS3D.
B. Domain Architecture

**FIG. 28.** Predicted domain architecture of LTA4H and CG10602 determined using Interpro and Pfam.

C. Model Generation and Validation

1. **Generation**

HHpred, Phyre2 and LOMETS were used to identify suitable templates for modeling CG10602. All three tools identify chain X of 3B7U (human leukotriene A4 hydrolase) as a high-scoring match suitable for single-template modeling.
<table>
<thead>
<tr>
<th>CG10602 Range</th>
<th>Template Range</th>
<th>% Identity</th>
<th>% Similarity</th>
<th>E-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-611</td>
<td>9-614</td>
<td>45%</td>
<td>66%</td>
<td>4.1e-164</td>
</tr>
</tbody>
</table>

Query:  
```
v dw s p + t+h l +df + g+ + nl +ld +d+ + + + g a + + + +
```  
Target:  
```
9VDCSLSFAPV-3CFHRLHAGCVDFARTLTLQPAAALTQVGEDLNLALSLVLDKLFILKEV-VINQES-VKNALQEG 84
PP
686888887515689998765..5799999999
```  

Query:  
```
v dd + l +t gq + +t+ +k+ + +i + + g+ + q g+ + qg + + + + g+ + qg + + + +
```  
Target:  
```
OSVRGSHIEG-PLAUKSNPKLE16ISFETDSPSSALQDLTLEQGKHPYFSDQCQAIHCRALTFCQDTPSVLYTLTA 164
PP
999***********99735889*********98622211567888876669
```  

FIG. 29. HMM profile of CG10602 aligned against chain X of 3B7U, a template structure selected for modeling CG10602.
2. Validation

The quality of the CG10602 model was evaluated using ProQ2 and ProSA, as described above in the discussion of CG1742. Again, overall model quality is high (a ProSA global Z-score of -9.42). The ProSA per-residue energy score graph is also favorable, showing a score that is consistently negative over a window size of 40 residues, except for minor high-energy regions at the N-terminus and two loop regions.

FIG. 30. Validation of the CG10602 model: ProQ2 quality score mapped to a 3D model of CG10602 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).
D. Structural Analyses

A structural superposition reveals that the CG10602 model and human leukotriene hydrolase (3B7U) display a similar secondary structure architecture across the modeled region with substantial full-length overlap. The pairwise alignment generated by this structural superposition also indicates substantial conservation of residues having similar physiochemical properties.

1. Secondary Structure
2. Physiochemical Properties

FIG. 32. Pairwise alignment of CG10602 and 3B7U generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

E. Analysis of Functional Residues

![LTA4H (PDB: 1H19) and CG10602 Model](image)

- **LTA4H Structure**
- **D. melanogaster Model**
- **Superimposed**

**FIG. 33.** LTA4H (3B7U, cyan-blue) superimposed on the predicted structure of CG10602 (green-red). RMSD: 0.891 Å.
LTA4H (PDB: 3B7U) and CG10602 Model Superimposed

FIG. 34. LTA4H (3B7U, cyan-blue) superimposed on the predicted structure of CG10602 (green-red), with potential matches for conserved functional residues highlighted.

F. Summary Table

TABLE 6. Summary of features shared by LTA4H and potential *D. melanogaster* ortholog CG10602.

<table>
<thead>
<tr>
<th></th>
<th>Length (AA)</th>
<th>Domain Architecture (Pfam, range)</th>
<th>Functional Residues (aligned matches in <em>D. melanogaster</em>)</th>
<th>Sequence ID%</th>
<th>Structural Overlap (RMSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leukotriene A4 hydrolase (LTA4H, NP_000886.1, PDB: 1H19)</td>
<td>611</td>
<td>Peptidase family M1 (PF01433) 13-387</td>
<td>H296, E297, H300, E319, D376 and Y384</td>
<td>44% ID 58% SIM</td>
<td>0.891 Å</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LTA4H, C-term (PF09127) 464-608</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncharacterized protein (CG10602, NP_724139.1)</td>
<td>613</td>
<td>Peptidase family M1 (PF01433) 8-420</td>
<td>H293, E294, H297, E316, D374 and Y382</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LTA4H, C-term (PF09127) 464-610</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
GGT1

The fifth gene in this set, CG6461, encodes a 579 amino acid protein ("Gamma-glutamyltranspeptidase," NP_573303.1) that shares 39% identity and 54% similarity to human Gamma-glutamyltranspeptidase 1 (GGT1, NP_001275762.1). GGT1 cleaves the gamma-glutamyl bond of extracellular glutathione (gamma-Glu-Cys-Gly), glutathione conjugates, and other gamma-glutamyl compounds. With respect to eicosanoid synthesis, GGT1 catalyzes the conversion of leukotriene C4 (LTC4) to leukotriene D4 (LTD4). GGT1 operates as a heterodimer consisting of small and large subunits which are produced from the same precursor that undergoes autocatalytic cleavage to produce the mature form of the enzyme. The only reported catalytic residue is T381, which corresponds to T382 in CG6461. GGT1 residues R107, T399 and E420 are believed to play a role in glutamate binding [94]. CG6461 possesses putative analogs for each of these residues, namely R107, T400 and E421, respectively. CG6461 and GGT1 also share identical domain architecture, with Gamma-glutamyltranspeptidase (PF01019) domains spanning the majority of each protein. Superimposed, the human structure (4GDX) and CG6461’s predicted structure display an RMSD of 1.126 Å.
A. Predicted Secondary Structure

Conservation:

Ggt1_1-569  1 MK-----KLVGLVLAVLVVLVGLCLWLPSASEPKDHNVTYRAAADAQCSIRKGRDARIDGSAYV
CG6461_1-579  1 MK1VWSKLLWWWLLAAALMVT----ALTGLGFSL-KRNDTLYIGAVSNICANAVGEMTDGSGAYV
Consensus:  
Consensus:  

Ggt1_1-569  67 AATAALLCGLMHSMGIGGFLITLTYSTTRKAEVIRANVRAFILAAMTFNSSEEGQQGKGLVSAPVF
CG6461_1-579  67 AAATALLLCEGLMLHSMIGIGGGFVAVTITRSRKYETVIAKESAPAAAAHMKMFGEE--SI1TGAAGSNV
Consensus:  
Consensus:  

Ggt1_1-569  137 EIRGTEILRAQRCHGLPWLRFQPSIQDLARQGFPSKGLALSLENHKTTPVQQPVLCVCRD--DRKVLREG205
CG6461_1-579  136 EILGWMWMMRYYGILKRFLEFSLIKARHGVSRYUALAIAGLSKDLKIKAEFLGhAVLTNAGTLEG205
Consensus:  
Consensus:  

Ggt1_1-569  206 ERLLPQPLACTYETIAEGAQAfYN-GSITQVXDIQAAVGIAVDTAEDMNHPLNISLG-----DV272
CG6461_1-579  206 DYMKRPALADTLIERAENGKRYDEXEDG78TGRKDQV1DKNGGILTEQDRLQDVRWES65DHGSHFFV5GFTY
Consensus:  
Consensus:  

Ggt1_1-569  273 VLYMPASALGLPVLAILLILKCNFSRESVE5PQQKLGTRYHLRVEAFRAAYAKRKLTDGKFV----D337
CG6461_1-579  273 TLYSTPMKS6GVPFALPI1MNLDYTD---------NPIYIWQRVQVAFKHAYQORTNIGMAYMPP5SAAS337
Consensus:  
Consensus:  

Ggt1_1-569  338 VTEVFMNTSPEFFLAQISDDTHPISYY-KPEFTYPDDDCCHQSVLVAEDSGAVSAT2STNYFGS406
CG6461_1-579  338 INATLEEMKPEFLVESRKILNEDHNSKDYLGTYGAPMTVEEDHQTAMHNOVLATGDVG1TS1NNYFGS407
Consensus:  
Consensus:  

Ggt1_1-569  407 KVRSPVSGLIFQNNEMDFDSSPTINEGQ7FPASQNPQGQLSMPCTIMVDQGRVCRMVGAAAGGQ
CG6461_1-579  408 KVASTQTIGLISNEDMDPSSQVINGVPGEFAFAPNYFQFKPSSMSPSCLVEHCNVRVL2GAGA6STR477
Consensus:  
Consensus:  

Ggt1_1-569  477 IVTATLLAIYMLAFGYDVRKAEVEEPRHQRQLLFPVTVERNDQAVTAAELTRHHHTQ1ASTFIAVQA546
CG6461_1-579  478 ITSTYVARVIMYLRKESITLAVNNSSLHQLQQHMLFYEQECDVMSVYVQLKVQHMEMYEPVFGSA5AV
Consensus:  
Consensus:  

Ggt1_1-569  547 IVRTA-6GWAAD6SRKGQPEPAG---656
CG6461_1-579  548 TAIGALEQEP6PD6RD6CGSALT6ATK6N6MQ6
Consensus:  
Consensus:  

FIG. 35. Predicted secondary structure of GGT1 and CG6461 calculated using PROMALS3D.
B. Domain Architecture

![Diagram showing domain architecture of GGT1 and CG6461]

FIG. 36. Predicted domain architecture of GGT1 and CG6461 determined using Interpro and Pfam.

C. Model Generation and Validation

1. Generation

HHpred, Phyre2 and LOMETS were used to identify suitable templates for modeling CG6461. All three tools identify chains A and B of 4GDX (human gamma-glutamyl-transpeptidase 1 heavy and light chains, respectively) as high-scoring matches suitable for composite modeling. Human gamma-glutamyl-transpeptidase 1 is expressed as a single polypeptide which is cleaved during post-translational processing to produce a heavy chain and a light chain which form the GGT1 holoenzyme complex. Chains A and B of the 4GDX PDB were joined using Chimera to form a single composite structure which was then used as a single template to model CG6461.
FIG. 37. HMM profile of CG6461 aligned against chains A and B of 4GDX, a template structure selected for modeling CG6461.
2. Validation

The quality of the CG6461 model was evaluated using ProQ2 and ProSA, as described above in the discussion of CG10602. Again, overall model quality is high (a ProSA global Z-score of -9.54). The ProSA per-residue energy score graph is also favorable, showing a score that is consistently negative over a window size of 40 residues.

FIG. 38. Validation of the CG6461 model: ProQ2 quality score mapped to a 3D model of CG6461 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).
D. Structural Analyses

A structural superposition reveals that the CG6461 model and human gamma-glutamyltranspeptidase 1 (4GDX) display a similar secondary structure architecture across the modeled region with substantial full-length overlap. The pairwise alignment generated by this structural superposition also indicates substantial conservation of residues having similar physiochemical properties.

1. Secondary Structure
FIG. 39. Pairwise alignment of CG6461 and 4GDX generated from structural superposition with secondary structure elements and conserved residues highlighted.

2. Physiochemical Properties

FIG. 40. Pairwise alignment of CG6461 and 4GDX generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

E. Analysis of Functional Residues

GGT1 (PDB: 4GDX) and CG6461 Model

<table>
<thead>
<tr>
<th>GGT1 Structure</th>
<th>D. melanogaster Model</th>
<th>Superimposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Image of GGT1 structure]</td>
<td>[Image of D. melanogaster model]</td>
<td>[Image of superimposed structures]</td>
</tr>
</tbody>
</table>

FIG. 41. GGT1 (4GDX, cyan-blue) superimposed on the predicted structure of CG6461 (green-red). RMSD: 1.126 Å.
FIG. 42. GGT1 (4GDX, cyan-blue) superimposed on the predicted structure of CG6461 (green-red), with potential matches for conserved functional residues highlighted.

F. Summary Table

TABLE 7. Summary of features shared by GGT1 and potential D. melanogaster ortholog CG6461.

<table>
<thead>
<tr>
<th></th>
<th>Length (AA)</th>
<th>Domain Architecture (Pfam, range)</th>
<th>Functional Residues (aligned matches in D. melanogaster)</th>
<th>Sequence ID%</th>
<th>Structural Overlap (RMSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gamma-glutamyltranspeptidase 1 (GGT1, NP_001275762.1, PDB: 4GDX)</td>
<td>569</td>
<td>Gamma-glutamyltranspeptidase (PF01019) 55-564</td>
<td>T381</td>
<td>39% ID 54% SIM</td>
<td>1.126 Å</td>
</tr>
<tr>
<td>Gamma-glutamyltranspeptidase (CG6461, NP_573303.1)</td>
<td>579</td>
<td>Gamma-glutamyltranspeptidase (PF01019) 56-566</td>
<td>T382</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DPEP1

The sixth gene in this set, CG6154 (“DPEP,” NP_733146.2), encodes a 434 amino acid metallopeptidase that shares 42% identity and 56% similarity with human DPEP1 (“Dipeptidase,
renal,” NP_004404.1). Both enzymes contain a single Membrane dipeptidase (PF01244) domain spanning the majority of the protein. DPEP1 is known to regulate leukotriene activity by catalyzing the conversion of leukotriene D₄ (LTD₄) to leukotriene E₄ (LTE₄) using zinc as a cofactor. To that end, the residues H36, D38, E141, H214 and H235 are reported to be essential for coordinating the zinc ions required for catalysis. Aligned against CG6154, these residues correspond to H71, D73, E184, H257 and H278. This high degree of similarity extends to the tertiary structure, as the superimposed structure of human DPEP1 (1ITQ) and the predicted structure of CG6154 display a RMSD of 0.523 Å.

### A. Predicted Secondary Structure

<table>
<thead>
<tr>
<th>Conservation:</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>99</th>
<th>999</th>
<th>9999</th>
<th>99999</th>
<th>999999</th>
<th>9999999</th>
<th>99999999</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPEP1_1-411</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>M-</td>
<td>-WGW-</td>
<td>-WFLV-</td>
<td>-AVCTADFFRDGEAERIMRDSFDV</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>CG6154_1-434</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>WSRK</td>
<td>FLGFGVLCLNLCLFFYFLFLIFLCA</td>
<td>70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus aa:</td>
<td></td>
<td>M.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus ss:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

```plaintext
<table>
<thead>
<tr>
<th>Conservation:</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>99</th>
<th>999</th>
<th>9999</th>
<th>99999</th>
<th>999999</th>
<th>9999999</th>
<th>99999999</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPEP1_1-411</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>36</td>
<td>HNDLP</td>
<td>WQQLLIMFNRK</td>
<td>QDERA-----N</td>
<td>LTTTLAGTSTNIPKLRAGFGV</td>
<td>97</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG6154_1-434</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>71</td>
<td>HNDLPW</td>
<td>RIKFLKIQGKDFFAGGLR</td>
<td>ELAPWSSSAWSH</td>
<td>DLRKLKEGWSAQ</td>
<td>140</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus aa:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus ss:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Conservation:

<table>
<thead>
<tr>
<th>Conservation:</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>99</th>
<th>999</th>
<th>9999</th>
<th>99999</th>
<th>999999</th>
<th>9999999</th>
<th>99999999</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPEP1_1-411</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>96</td>
<td>TLEQCMCVHDRMMYFC</td>
<td>MPF</td>
<td>LYT</td>
<td>SSAGIRQAF</td>
<td>1LGVEGCH</td>
<td>HSDSLLGVLRL</td>
<td>YQGMRLTLT</td>
<td>67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG6154_1-434</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>141</td>
<td>TLEQ</td>
<td>DRLR</td>
<td>RGLVQ</td>
<td>YPHNLALV</td>
<td>TAVS</td>
<td>GIKQ</td>
<td>RGYK</td>
<td>5LGVEGCH</td>
<td>HSDSLLGVLRL</td>
<td>YQGMRLTLT</td>
<td>210</td>
</tr>
<tr>
<td>Consensus aa:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus ss:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Conservation:

<table>
<thead>
<tr>
<th>Conservation:</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>99</th>
<th>999</th>
<th>9999</th>
<th>99999</th>
<th>999999</th>
<th>9999999</th>
<th>99999999</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPEP1_1-411</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>168</td>
<td>HSCTNPWADNLLV</td>
<td>DGSDFQGCLSFFQYVQ</td>
<td>KIVGRL</td>
<td>LSRL</td>
<td>HAV</td>
<td>VTMT</td>
<td>AIQ</td>
<td>LG</td>
<td>4LSFGP</td>
<td>237</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG6154_1-434</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>211</td>
<td>HTCNPW</td>
<td>DMCNDY</td>
<td>QGKL</td>
<td>VVEMNR</td>
<td>LGMLVSH</td>
<td>YPF</td>
<td>TMT</td>
<td>DL</td>
<td>4LSFGP</td>
<td>280</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus aa:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus ss:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Conservation:

<table>
<thead>
<tr>
<th>Conservation:</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>99</th>
<th>999</th>
<th>9999</th>
<th>99999</th>
<th>999999</th>
<th>9999999</th>
<th>99999999</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPEP1_1-411</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>238</td>
<td>AYTSC</td>
<td>VSNR</td>
<td>LVQ</td>
<td>RQIA</td>
<td>ING</td>
<td>LG</td>
<td>WMA</td>
<td>VPH</td>
<td>pH</td>
<td>CSQG</td>
</tr>
<tr>
<td>CG6154_1-434</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>281</td>
<td>AHAIC</td>
<td>NRN</td>
<td>LVQ</td>
<td>RQIA</td>
<td>ING</td>
<td>LG</td>
<td>WMA</td>
<td>VPH</td>
<td>pH</td>
<td>CSQG</td>
</tr>
<tr>
<td>Consensus aa:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus ss:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Conservation:

<table>
<thead>
<tr>
<th>Conservation:</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>99</th>
<th>999</th>
<th>9999</th>
<th>99999</th>
<th>999999</th>
<th>9999999</th>
<th>99999999</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPEP1_1-411</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>308</td>
<td>VVFVFGD</td>
<td>SVKL</td>
<td>YD</td>
<td>DELL</td>
<td>RNL</td>
<td>NHYW</td>
<td>EAV</td>
<td>EQ</td>
<td>ASN</td>
<td>QPA</td>
<td>EEEIPDLQ</td>
<td>372</td>
<td></td>
</tr>
<tr>
<td>CG6154_1-434</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>351</td>
<td>VNLVP</td>
<td>SGDL</td>
<td>YD</td>
<td>DELL</td>
<td>RNL</td>
<td>NHYW</td>
<td>EAV</td>
<td>EQ</td>
<td>ASN</td>
<td>QPA</td>
<td>EEEIPDLQ</td>
<td>420</td>
<td></td>
</tr>
<tr>
<td>Consensus aa:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus ss:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conservation:</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>9</th>
<th>99</th>
<th>999</th>
<th>9999</th>
<th>99999</th>
<th>999999</th>
<th>9999999</th>
<th>99999999</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPEP1_1-411</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>411</td>
<td>LGGSCRMHYGYS</td>
<td>SSGL</td>
<td>LHRH</td>
<td>WGG</td>
<td>LLAS</td>
<td>ALPL</td>
<td>CLS</td>
<td>411</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG6154_1-434</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>434</td>
<td>IMGSR</td>
<td>CYG</td>
<td>QPR</td>
<td>434</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus aa:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus ss:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

h 84

```
FIG. 43. Predicted secondary structure of DPEP1 and CG6154 calculated using PROMALS3D.

B. Domain Architecture

![Diagram of DPEP1 and CG6154 domain architecture]

FIG. 44. Predicted domain architecture of DPEP1 and CG6154 determined using Interpro and Pfam.

C. Model Generation and Validation

1. Generation

HHpred, Phyre2 and LOMETS were used to identify suitable templates for modeling CG6154. All three tools identify chain A of 1ITQ (human renal dipeptidase) as a high-scoring match suitable for single-template modeling.
<table>
<thead>
<tr>
<th>CG6154 Range</th>
<th>Template Range</th>
<th>% Identity</th>
<th>% Similarity</th>
<th>E-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>53-425</td>
<td>2-361</td>
<td>50%</td>
<td>72%</td>
<td>1.5e-113</td>
</tr>
</tbody>
</table>

**FIG. 45.** HMM profile of CG6154 aligned against chain A of 1ITQ, a template structure selected for modeling CG6154.

2. **Validation**

The quality of the CG6154 model was evaluated using ProQ2 and ProSA, as described above in the discussion of CG1742. Again, overall model quality is high (a ProSA global Z-score of -7.99). The ProSA per-residue energy score graph is also favorable, showing a score that is consistently negative over a window size of 40 residues.
D. Structural Analyses

A structural superposition reveals that the CG6154 model and human renal dipeptidase (1ITQ) display a similar secondary structure architecture across the modeled region with substantial full-length overlap. The pairwise alignment generated by this structural superposition also indicates substantial conservation of residues having similar physiochemical properties.
FIG. 47. Pairwise alignment of CG6154 and 1ITQ generated from structural superposition with shared secondary structure elements and conserved residues highlighted.
2. Physiochemical Properties

FIG. 48. Pairwise alignment of CG6154 and 1ITQ generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

E. Analysis of Functional Residues

<table>
<thead>
<tr>
<th>DPEP1 (PDB: 1ITQ) and CG6154 Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DPEP1 Structure</strong></td>
</tr>
</tbody>
</table>

FIG. 49. DPEP1 (1ITQ, cyan-blue) superimposed on the predicted structure of CG6154 (green-red). RMSD: 0.523 Å.
FIG. 50. DPEP1 (1ITQ, cyan-blue) superimposed on the predicted structure of CG6154 (green-red), with potential matches for conserved functional residues highlighted.

F. Summary Table

**TABLE 8.** Summary of features shared by DPEP1 and potential *D. melanogaster* ortholog CG6154.

<table>
<thead>
<tr>
<th>Dipeptidase, renal (DPEP1, NP_004404.1, PDB: 1ITQ)</th>
<th>411</th>
<th>Membrane dipeptidase (PF01244)</th>
<th>H36, D38, E141, H214 and H235</th>
<th>42% ID</th>
<th>56% SIM</th>
<th>0.523 Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPEP (CG6154, NP_733146.2)</td>
<td>434</td>
<td>Membrane dipeptidase (PF01244)</td>
<td>H71, D73, E184, H257 and H278</td>
<td>42% ID</td>
<td>56% SIM</td>
<td>0.523 Å</td>
</tr>
</tbody>
</table>

**GPX1**

The seventh gene in this set, CG12013 (“glutathione peroxidase,” NP_728870.1), encodes a 169 amino acid selenoprotein that is 32% identical and 44% similar to human GPX1 (“glutathione peroxidase 1,” NP_000572.2). As part of the eicosanoid pathway, GPX-1 converts
the 12(S)-HPETE produced by lipoxygenase-12 (12-LOX) into 12(S)-HETE. GPX1 requires a selenocysteine residue (U49) for catalysis. CG12013’s amino acid sequence indicates that a cysteine residue is present at the corresponding position in this enzyme. However, this residue may in fact be a selenocysteine, as this substitution is unlikely to be detected by standard sequencing techniques. A structural alignment of human GPX1 (2F8A) and the predicted structure of CG12013 reveals an RMSD of 0.699 Å.

A. Predicted Secondary Structure

![Predicted secondary structure of GPX1 and CG12013 calculated using PROMALS3D.](FIG. 51)

B. Domain Architecture

![Predicted domain architecture of GPX1 and CG12013 determined using Interpro and Pfam.](FIG. 52)
C. Model Generation and Validation

1. Generation

HHpred, Phyre2 and LOMETS were used to identify suitable templates for modeling CG12013. All three tools identify chain A of 2F8A (human glutathione peroxidase 1) as a high-scoring match suitable for single-template modeling.

<table>
<thead>
<tr>
<th>CG12013 Range</th>
<th>Template Range</th>
<th>% Identity</th>
<th>% Similarity</th>
<th>E-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>7-122</td>
<td>20-140</td>
<td>39%</td>
<td>62%</td>
<td>2.3e-17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query</th>
<th>Target</th>
<th>PP</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 yknaaerief+tvkdtg ndvrelykg+tvklv+tvslc +k+knvgekt+dklyg+tvzl+fpconqfgtep+85++</td>
<td>FSOSQSVGAFSEARPL +GFPSLGLSGRLK+VLTTEHVASLGG+TVRDVTQKRNQRGLPRLCLVLGLFPONPGCHR -NAK</td>
<td>131</td>
</tr>
<tr>
<td>86 gesevchrdskadig ...e+tvkvdvngd+aplyk +k +e +l + g +f +k +vng a pl +1+</td>
<td>NREILNLSKVRQPGGfepntMLFKeKEVNGAGHPLFALR</td>
<td>140</td>
</tr>
<tr>
<td>131-169</td>
<td>168-206</td>
<td>49%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query</th>
<th>Target</th>
<th>PP</th>
</tr>
</thead>
<tbody>
<tr>
<td>169 + wnf kflv +gvp+ ry+ +di die 11</td>
<td>NDVAWNFEFLVGPPLRLLRYSRRRFQFIDIEPFDAIL</td>
<td>206</td>
</tr>
<tr>
<td>5789*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**FIG. 53.** HMM profile of CG12013 aligned against chain A of 2F8A, a template structure selected for modeling CG12013.

2. Validation

The quality of the CG12013 model was evaluated using ProQ2 and ProSA, as described above in the discussion of CG1742. Again, overall model quality is high (a ProSA global Z-score of -7). The ProSA per-residue energy score graph is also favorable, showing a score that is consistently negative over a window size of 40 residues.
FIG. 54. Validation of the CG12013 model: ProQ2 quality score mapped to a 3D model of CG12013 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).

D. Structural Analyses

A structural superposition reveals that the CG12013 model and human glutathione peroxidase 1 (2F8A) display a similar secondary structure architecture across the modeled region with substantial full-length overlap. The pairwise alignment generated by this structural
superposition also indicates substantial conservation of residues having similar physiochemical properties.

1. Secondary Structure

![Secondary Structure Diagram](image1)

**FIG. 55.** Pairwise alignment of CG12013 and 2F8A generated from structural superposition with shared secondary structure elements and conserved residues highlighted.

2. Physiochemical Properties

![Physiochemical Properties Diagram](image2)

**FIG. 56.** Pairwise alignment of CG12013 and 2F8A generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.
E. Analysis of Functional Residues

<table>
<thead>
<tr>
<th>GPX1 (PDB: 2F8A) and CG12013 Model</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>GPX1 Structure</td>
<td><em>D. melanogaster</em> Model</td>
<td>Superimposed</td>
</tr>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
</tbody>
</table>

**FIG 57.** GPX1 (2F8A, cyan-blue) superimposed on the predicted structure of CG12013 (green-red). RMSD: 0.699 Å.

<table>
<thead>
<tr>
<th>GPX1 (PDB: 2F8A) and X Model Superimposed</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image4.png" alt="Image" /></td>
<td></td>
</tr>
</tbody>
</table>

**FIG. 58.** GPX1 (2F8A, cyan-blue) superimposed on the predicted structure of CG12013 (green-red), with potential matches for conserved functional residues highlighted.
F. Summary Table

**TABLE 9. Summary of features shared by GPX1 and potential *D. melanogaster* ortholog CG12013.**

<table>
<thead>
<tr>
<th></th>
<th>Length (AA)</th>
<th>Domain Architecture (Pfam, range)</th>
<th>Functional Residues (aligned matches in <em>D. melanogaster</em>)</th>
<th>Sequence ID%</th>
<th>Structural Overlap (RMSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glutathione peroxidase 1 (GPX1, NP_000572.2, PDB: 2F8A)</td>
<td>203</td>
<td>GSHPx (PF00255) 16-130</td>
<td>C49</td>
<td>32% ID 44% SIM</td>
<td>0.699Å</td>
</tr>
<tr>
<td>Glutathione peroxidase (CG12013, NP_728870.1)</td>
<td>169</td>
<td>GSHPx (PF00255) 13-122</td>
<td>C45</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**CPA1**

The eighth and final gene in this set, CG18585 ("CPA," NP_609132.1) encodes a 422 amino acid protein that is 34% identical and 54% similar to human CPA1 ("Carboxypeptidase A1," NP_001859.1). CPA1 is a metallocarboxypeptidase that requires a zinc atom as a cofactor for catalysis. CPA1 has been shown to convert the potent leukotriene C4 (LTC4) to the less potent leukotriene F4 (LTF4) by hydrolysis of an amide bond, suggesting that CPA1 serves to reduce inflammation. The residues H179, E182, H306 and E380 are reportedly required for catalysis. These residues correspond to H178, E181, H305 and E382 in CPA. Aligned, the structure of human CPA1 (3FJU) and the predicted structure of CG18585 display an RMSD of 0.753 Å.
A. Secondary Structure Prediction

FIG. 59. Predicted secondary structure of CPA1 and CG18585 calculated using PROMALS3D.

B. Domain Architecture
FIG. 60. Predicted domain architecture of CPA1 and CG18585 determined using Interpro and Pfam.

C. Model Generation and Validation

1. Generation

HHpred, Phyre2 and LOMETS were used to identify suitable templates for modeling CG18585. All three tools identify chain A of 2V77 (human carboxypeptidase A1) as a high-scoring match suitable for single-template modeling. Human carboxypeptidase A1 undergoes posttranslational processing, which results in cleavage of the N-terminal portion of the polypeptide spanning from position 1 to 110 (i.e., including a signal peptide spanning 1-16 and an activation peptide spanning 17-110). The mature form of the enzyme consists of the C-terminal portion of the polypeptide (i.e., position 111-419). CG18585 is predicted to share a similar architecture, as the N-terminal regions align against the sequence of the human carboxypeptidase A1 activation peptide. The C-terminal region of CG18585 which appears to correspond to the sequence encoding mature human carboxypeptidase A1 was modeled for this study.

<table>
<thead>
<tr>
<th>CG18585 Range</th>
<th>Template Range</th>
<th>% Identity</th>
<th>% Similarity</th>
<th>E-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>114-417</td>
<td>4-305</td>
<td>40%</td>
<td>65%</td>
<td>3.8e-69</td>
</tr>
</tbody>
</table>

```plaintext
<table>
<thead>
<tr>
<th>Query</th>
<th>Target</th>
<th>PP</th>
<th>% Identity</th>
<th>% Similarity</th>
<th>E-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>114-417</td>
<td>4-305</td>
<td>40%</td>
<td>65%</td>
<td>3.8e-69</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CG18585 Range</th>
<th>Template Range</th>
<th>% Identity</th>
<th>% Similarity</th>
<th>E-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>271-350</td>
<td>300-380</td>
<td>40%</td>
<td>65%</td>
<td>3.8e-69</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CG18585 Range</th>
<th>Template Range</th>
<th>% Identity</th>
<th>% Similarity</th>
<th>E-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>351-417</td>
<td>380-482</td>
<td>40%</td>
<td>65%</td>
<td>3.8e-69</td>
</tr>
</tbody>
</table>
```
FIG. 61. HMM profile of CG18585 aligned against chain A of 2V77, a template structure selected for modeling CG18585.

2. Validation

The quality of the CG18585 model was evaluated using ProQ2 and ProSA, as described above in the discussion of CG1742. Again, overall model quality is high (a ProSA global Z-score of -7.75). The ProSA per-residue energy score graph is also favorable, showing a score that is consistently negative over a window size of 40 residues.

FIG. 62. Validation of the CG18585 model: ProQ2 quality score mapped to a 3D model of CG18585 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).
D. Structural Analyses

1. Secondary Structure

A structural superposition reveals that the CG18585 model and human carboxypeptidase A1 (2V77) display a similar secondary structure architecture across the modeled region with substantial full-length overlap. The pairwise alignment generated by this structural superposition also indicates substantial conservation of residues having similar physicochemical properties.
FIG. 63. Pairwise alignment of CG18585 and 2V77 generated from structural superposition with shared secondary structure elements and conserved residues highlighted.

2. Physiochemical Properties

FIG. 64. Pairwise alignment of CG18585 and 2V77 generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

E. Analysis of Functional Residues

<table>
<thead>
<tr>
<th>CPA1 (PDB: 2V77) and CG18585 Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPA1 Structure</td>
</tr>
</tbody>
</table>

FIG. 65. Truncated CPA1 (2V77, cyan-blue) superimposed on the predicted structure of CG18585 (green-red). RMSD: 0.753 Å.
FIG. 66. Truncated CPA1 (2V77, cyan-blue) superimposed on the predicted structure of CG18585 (green-red), with potential matches for conserved functional residues highlighted.

F. Summary Table

TABLE 10. Summary of features shared by CPA1 and potential *D. melanogaster* ortholog CG18585.

<table>
<thead>
<tr>
<th></th>
<th>Length (AA)</th>
<th>Domain Architecture (Pfam, range)</th>
<th>Functional Residues (aligned matches in D. <em>melanogaster</em>)</th>
<th>Sequence ID%</th>
<th>Structural Overlap (RMSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carboxypeptidase A1 (CPA1, NP_001859.1, PDB: 3FJU)</td>
<td>419</td>
<td>Carboxypeptidase activation peptide (PF02244) 26-100</td>
<td>H179, E182, H306 and E380</td>
<td>34% ID 54% SIM</td>
<td>0.753 Å</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Zinc carboxypeptidase (PF00246) 128-406</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPA (CG18585, NP_609132.1)</td>
<td>422</td>
<td>Carboxypeptidase activation peptide (PF02244) 33-106</td>
<td>H178, E181, H305 and E382</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Zinc carboxypeptidase (PF00246) 128-408</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Group 2: The Midrange Candidates

Five candidates were identified in the search based more on structural similarity than underlying sequence conservation. Members of this set share 20-30% sequence identity with the human target, but share similarity at the fold level and possess at least one conserved functional residue.

HPGDS

The first gene in this set, CG8938 (GST-S1, NP_725653.1) encodes a 249 amino acid glutathione-S-transferase enzyme, which displays 27% identity and 43% similarity to human HPGDS, (“Hematopoietic Prostaglandin D synthase,” NP_055300.1). HPGDS, alternatively known as Prostaglandin-H2 D-isomerase, is a glutathione-requiring prostaglandin D synthase. HPGDS operates as a homodimer, which can optionally be activated by Ca\(^{2+}\) and Mg\(^{2+}\) to increase catalytic efficiency to 150% of the basal level [86]. Coordination of the metallic ion ligands is coordinated by D93, D96 and D97. CG8938 and GST-S1 share an identical domain architecture, with an N-terminal GST-N domain (PF02798) and a C-terminal GST-C domain (PF00043). Similarly, CG8938 possesses aligned matches for two of the three aspartic residues requires for enhanced catalytic activity, at positions D139 and D143 (corresponding to D93 and D97 in HPGDS). At the corresponding position aligned to D96, CG8938 instead has an asparagine (N142). However, mutagenesis studies by Inouie et al. have shown that a D to N substitution at this same position in HPGDS actually increases PGD\(_2\) synthesis [95]. In addition to this sequence-based evidence, superposition of the predicted CG8938 structure and human HPGDS (1IYI) reveals an RMSD of 1.045 Å.
A. Secondary Structure Prediction

![Secondary Structure Prediction](image)

**FIG. 67.** Predicted secondary structure of HPGDS and CG8938 calculated using PROMALS3D.

B. Domain Architecture

![Domain Architecture](image)

**FIG. 68.** Predicted domain architecture of HPGDS and CG8938 determined using Interpro and Pfam.
C. Model Generation and Validation

1. Generation

HHpred, Phyre2 and LOMETS were used to identify suitable templates for modeling CG8938.

All three tools identify chain A of 1IYH (human hematopoietic prostaglandin D synthase) as a high-scoring match suitable for single-template modeling.

<table>
<thead>
<tr>
<th>CG8938 Range</th>
<th>Template Range</th>
<th>% Identity</th>
<th>% Similarity</th>
<th>E-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>49-248</td>
<td>2-197</td>
<td>36%</td>
<td>61%</td>
<td>7.9e-36</td>
</tr>
</tbody>
</table>

**FIG. 69. HMM profile of CG8938 aligned against chain A of 1IYH, a template structure selected for modeling CG8938.**

2. Validation

The quality of the CG8939 model was evaluated using ProQ2 and ProSA, as described above in the discussion of CG1742. Again, overall model quality is high (a ProSA global Z-score of -6.16). The ProSA per-residue energy score graph is also favorable, showing a score that is consistently negative over a window size of 40 residues, after the initial N-terminal segment that was generated using \textit{ab initio} methods. This proline rich N-terminal segment may include a signal peptide or other processing-related motif.
FIG. 70. Validation of the CG8938 model: ProQ2 quality score mapped to a 3D model of CG8938 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).

D. Structural Analyses

A structural superposition reveals that the CG8938 model and human hematopoietic prostaglandin D synthase (1IYH) display a similar secondary structure architecture across the modeled region with substantial full-length overlap. The pairwise alignment generated by this structural superposition also indicates substantial conservation of residues having similar physiochemical properties.
1. Secondary Structure

FIG. 71. Pairwise alignment of CG8938 and 1IYH generated from structural superposition with shared secondary structure elements and conserved residues highlighted.

2. Physiochemical Properties

FIG. 72. Pairwise alignment of CG8938 and 1IYH generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.
E. Analysis of Functional Residues

TABLE 11. Summary of features shared by HPGDS and potential D. melanogaster ortholog CG8938.

<table>
<thead>
<tr>
<th>HPGDS (PDB: 1IYI) and CG8938 Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPGDS Structure</td>
</tr>
</tbody>
</table>

FIG. 73. HPGDS (1IYI, cyan-blue) superimposed on the predicted structure of CG8938 (green-red). RMSD: 1.045 Å.

FIG. 74. HPGDS (1IYH, cyan-blue) superimposed on the predicted structure of CG8938 (green-red), with potential matches for conserved functional residues highlighted.
The second gene in this set, CG16817 (NP_649925.1) encodes a 184 amino acid protein, which displays 24% identity and 40% similarity to human PTGES3 (“prostaglandin E synthase 3,” NP_006592.3). In contrast to PTGES1 and PTGES2 discussed above, PTGES3 is a cytosolic protein. However, all three enzymes share the same catalytic function, *i.e.*, oxidoreduction of prostaglandin endoperoxide H₂ (PGH₂) to prostaglandin E₂ (PGE₂). PTGES3 and CG16817 share an identical domain architecture, with a CS Domain (PF04969) spanning the N-terminus to the middle of the protein. The C-terminal region of PTGES3 is notable for showing a compositional bias towards aspartic acid and glutamic acid residues in the range spanning 108 to 160. At this time, only one crystal structure for PTGES3 has been published (1EJF) and this structure is limited to a truncated version of the protein spanning residue 1 to 110, *i.e.*, the acidic C-terminal region has not been crystalized. An alignment of PTGES3 and CG16817 shows that CG16817 also displays a compositional bias towards acidic residues at its C-terminus, though it is unclear what role this serves in either protein. We were unable to generate a full-length model of CG16817 due to a lack of any suitable template structure for the C-terminal acidic region.
However, the truncated CG16817 model (residues 1 to 113) shows a RMSD of 0.823 Å compared to PTGES3.

A. Secondary Structure Prediction

![Image](75.png)

**FIG. 75.** Predicted secondary structure of PTGES3 and CG16817 calculated using PROMALS3D.

B. Domain Architecture

![Image](76.png)

**FIG. 76.** Predicted domain architecture of PTGES3 and CG16817 determined using Interpro and Pfam.

C. Model Generation and Validation

1. Generation

HHpred, Phyre2 and LOMETS were used to identify suitable templates for modeling CG16817. All three tools identify chain A of 1EJF (human prostaglandin E synthase 3) as a high-scoring match suitable for single-template modeling of the first two-thirds of CG16817 (i.e., the region spanning 1-110), which includes the CS domain mapped to the segment spanning
10-85. The crystal structure of the acidic (E/D-rich) C-terminus of human prostaglandin E synthase 3 was not included in 1EJF and is currently unknown. Sequence analysis of CG16817 reveals that a similar E/D-rich segment at the C-terminus of the polypeptide, providing further support for CG16817 being orthologous to or at least functionally similar to human prostaglandin E synthase 3. The modeled portion of CG17817 is predicted to be structured as a β-sandwich consisting of two sheets of anti-parallel β-strands, which matches the structure of the corresponding portion of human prostaglandin E synthase 3. Attempts to model the acidic C-terminal segment using ab initio methods were unable to generate a high-confidence structure. It should be noted that this segment is predicted to be an intrinsically disordered region (e.g., as predicted by the IUPRed server, http://iupred.enzim.hu/). As a result, the present structural analysis of CG16817 is limited to the segment spanning 1-110 of the full-length polypeptide sequence.

<table>
<thead>
<tr>
<th>CG16817 Range</th>
<th>Template Range</th>
<th>% Identity</th>
<th>% Similarity</th>
<th>E-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>9-127</td>
<td>3-121</td>
<td>26%</td>
<td>46%</td>
<td>8.7e-10</td>
</tr>
</tbody>
</table>

**FIG. 77.** HMM profile of CG16817 aligned against chain A of 1EJF, a template structure selected for modeling CG16817.

2. Validation

The quality of the CG16817 model was evaluated using ProQ2, and ProSA as described above in the discussion of CG1742. Again, overall model quality is high (a ProSA global Z-score of -2.96). The ProSA per-residue energy score graph is also favorable, showing a score that is
consistently negative over a window size of 40 residues, except for the loop regions connecting
the β-sandwich halves. As noted above, this model is truncated and excludes the acidic C-
terminus of CG16817, which may contribute to the tertiary structure of the full-length protein in
a manner that is not accurately represented by the current model. However, the modeled structure
has a reasonable quality score and appears to be structurally analogous to the crystalized portion
of human prostaglandin E synthase 3, which is similarly truncated.

FIG. 78. Validation of the CG16817 model: ProQ2 quality score mapped to a 3D model of
CG16817 (top); ProSA global quality score ranking (lower left) and per-residue quality
graph (lower right).
D. Structural Analyses

A structural superposition reveals that the CG16817 model and human prostaglandin E synthase 3 (1EJF) display a similar secondary structure architecture across the modeled region with substantial full-length overlap. The pairwise alignment generated by this structural superposition also indicates substantial conservation of residues having similar physiochemical properties.

1. Secondary Structure

**FIG. 79.** Pairwise alignment of CG16817 and 1EJF generated from structural superposition with shared secondary structure elements and conserved residues highlighted.
2. Physiochemical Properties

FIG. 80. Pairwise alignment of CG16817 and 1EJF generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

E. Analysis of Functional Residues

<table>
<thead>
<tr>
<th>PTGES3 (PDB: 1EJF) and CG16817 Model (Partial)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truncated PTGES3 Structure</td>
</tr>
</tbody>
</table>

FIG. 81. Truncated PTGES3 (1EJF, cyan-blue) superimposed on the predicted structure of CG16817 (green-red). RMSD: 0.823 Å.

FIG. 82. Truncated PTGES3 (1EJF, cyan-blue) superimposed on the predicted structure of CG16817 (green-red), with potential matches for conserved functional residues highlighted.
F. Summary Table

**TABLE 12. Summary of features shared by PTGES3 and potential *D. melanogaster* ortholog CG16817.**

<table>
<thead>
<tr>
<th></th>
<th>Length (AA)</th>
<th>Domain Architecture (Pfam, range)</th>
<th>Functional Residues (aligned matches in <em>D. melanogaster</em>)</th>
<th>Sequence ID%</th>
<th>Structural Overlap (RMSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prostaglandin E synthase 3 (PTGES3, NP_006592.3, PDB: 1EJF)</td>
<td>160</td>
<td>CS Domain (PF04969) 4-79</td>
<td>N/A</td>
<td>24% ID 40% SIM</td>
<td>0.823 Å</td>
</tr>
<tr>
<td>Uncharacterized protein (CG16817, NP_649925.1)</td>
<td>184</td>
<td>CS Domain (PF04969) 10-85</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**CBR1**

The third gene in this set, CG11200 (“CBR,” NP_611471.1) encodes a 355 amino acid protein that displays 20% identity and 33% similarity to human CBR1 (“carbonyl reductase 1,” NP_001748.1). CBR1 is a NADPH-dependent reductase with broad substrate specificity, *e.g.*, it can convert PGE<sub>2</sub> to PGF<sub>2α</sub>. It has been reported that CBR1 binds to NADP via N90 and that Y194 serves as a proton acceptor for the reaction. A corresponding match for both of these residues can be found in CG11200 at positions N154 and Y233. An alignment of human CBR1 (3BHJ) against the predicted model for CG11200 shows an RMSD of 1.153 Å.
A. Secondary Structure Prediction

<table>
<thead>
<tr>
<th>Conservation:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CBSR1-1-277</strong></td>
<td>1 MLGTTLLM1FFVGLLCAFLFSKTKEF8KSWEFKTE FRYQ1LGIVGLVHDAQYKARDRAVLKQDRIA</td>
</tr>
<tr>
<td><strong>CG11200-1-355</strong></td>
<td>1 MLGTTLLM1FFVGLLCAFLFSKTKEF8KSWEFKTE FRYQ1LGIVGLVHDAQYKARDRAVLKQDRIA</td>
</tr>
<tr>
<td><strong>Consensus_aa</strong></td>
<td>hhhhhhhhhhhhhhh</td>
</tr>
<tr>
<td><strong>Consensus_ss</strong></td>
<td>eee</td>
</tr>
</tbody>
</table>

**Conservation:**

| **CBSR1-1-277** | 9 LVTQGKNSGLGAI1VVDGLCLGLSVDGTVL1TARDVTGQAAVQQLEAE---GLSPRFHQLD1DDQSIRAILRD |
| **CG11200-1-355** | 71 VITQGKNSGLGAI1VVDGLCLGLSVDGTVL1TARDVTGQAAVQQLEAE---GLSPRFHQLD1DDQSIRAILRD |
| **Consensus_aa** | hhhhhhhhhhhhhhh |
| **Consensus_ss** | eee |

**Conservation:**

| **CBSR1-1-277** | 9 FLRKEYGGLVDLNNAG1AFLKVAAPDFIHIATVIMTKIFGFRDCCELFLPLIKFQ-----19VVNS |
| **CG11200-1-355** | 140 LIKERYSKVSDDLNAG1AFLKVAAPDFIHIATVIMTKIFGFRDCCELFLPLIKFQ-----19VVNS |
| **Consensus_aa** | hhhhhhhhhhhhhhhhh |
| **Consensus_ss** | eee |

**Conservation:**

| **CBSR1-1-277** | 140 SIVSVALNKSDPESQPQRFQSETITEEEELVLMKSFVEDDKGQKEGQPSAYGTVKGGTVLST1RH |
| **CG11200-1-355** | 208 SCVNLGNYKRDIN-----------------------CTKYHPTGATYSQSKLAQLFLTRHL |
| **Consensus_aa** | hhhhhhhhhhhhhhhhh |
| **Consensus_ss** | eee |

**Conservation:**

| **CBSR1-1-277** | 9 RKLSEQRKGD1ILNACC8PGWRTMDAG----------PKATKSPEEGAETPYVL1LAPPDDEAPMPQ3 |
| **CG11200-1-355** | 249 TLLDAE-----KSHQVNNYVFIVDOTLHEHATSSTVPIFKFLKFLFPTMSGKSTVFFPAIDPSP1EQGQT |
| **Consensus_aa** | hhhhhhhhhhhhhhhhh |
| **Consensus_ss** | eee |

**Conservation:**

| **CBSR1-1-277** | 269 YSLNQGKGPDDAKKPAKCEQLPQFSCLUDL1QYNYGNY |
| **CG11200-1-355** | 315 YSLNQGKGPDDAKKPAKCEQLPQFSCLUDL1QYNYGNY |
| **Consensus_aa** | hhhhhhhhhhhhhhhhh |
| **Consensus_ss** | eee |

**FIG. 83.** Predicted secondary structure of CBSR1 and CG11200 calculated using PROMALS3D.

B. Domain Architecture

**FIG. 84.** Predicted domain architecture of CBSR1 and CG11200 determined using Interpro and Pfam.
C. Model Generation and Validation

1. Generation

HHpred, Phyre2 and LOMETS were used to identify suitable templates for modeling CG11200. All three tools identify chain A of 3BHJ (human carbonyl reductase 1 complexed with glutathione) as a high-scoring match suitable for single-template modeling.

<table>
<thead>
<tr>
<th>CG11200 Range</th>
<th>Template Range</th>
<th>% Identity</th>
<th>% Similarity</th>
<th>E-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>68-206</td>
<td>5-143</td>
<td>34%</td>
<td>56%</td>
<td>1.6e-12</td>
</tr>
</tbody>
</table>

FIG. 85. HMM profile of CG11200 aligned against chain A of 3BHJ, a template structure selected for modeling CG11200.

2. Validation

The quality of the CG11200 model was evaluated using ProQ2 and ProSA, as described above in the discussion of CG1742. Again, overall model quality is high (a ProSA global Z-score of -6.55). The ProSA per-residue energy score graph is also favorable, showing a score that is consistently negative over a window size of 40 residues, except for the N-terminus (which is predicted to be a signal peptide) and a single loop region.
FIG. 86. Validation of the CG11200 model: ProQ2 quality score mapped to a 3D model of CG11200 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).
D. Structural Analyses

1. Secondary Structure

![Pairwise alignment of CG11200 and 3BHJ generated from structural superposition with secondary structure elements and conserved residues highlighted.](image)

**FIG. 87.** Pairwise alignment of CG11200 and 3BHJ generated from structural superposition with secondary structure elements and conserved residues highlighted.
2. Physiochemical Properties

FIG. 88. Pairwise alignment of CG11200 and 3BHJ generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

E. Analysis of Functional Residues

<table>
<thead>
<tr>
<th>CBR1 (PDB: 3BHJ) and CG11200 Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBR1 Structure</td>
</tr>
</tbody>
</table>

FIG 89. CBR1 (3BHJ, cyan-blue) superimposed on the predicted structure of CG11200 (green-red). RMSD: 1.153 Å.
FIG. 90. CBR1 (3BHJ, cyan-blue) superimposed on the predicted structure of CG11200 (green-red), with potential matches for conserved functional residues highlighted.

F. Summary Table

TABLE 13. Summary of features shared by CBR1 and potential D. melanogaster ortholog CG11200.

<table>
<thead>
<tr>
<th></th>
<th>Length (AA)</th>
<th>Domain Architecture (Pfam, range)</th>
<th>Functional Residues (aligned matches in D. melanogaster)</th>
<th>Sequence ID%</th>
<th>Structural Overlap (RMSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbonyl reductase 1 (CBR1, NP_001748.1, PDB: 3BHJ)</td>
<td>277</td>
<td>Short chain dehydrogenase domain (PF16152) 6-151</td>
<td>N90 and Y194</td>
<td>20% ID</td>
<td>1.153 Å</td>
</tr>
<tr>
<td>CBR (CG11200, NP_611471.1)</td>
<td>355</td>
<td>Short chain dehydrogenase domain (PF16152) 68-252</td>
<td>N154 and Y233</td>
<td>33% SIM</td>
<td></td>
</tr>
</tbody>
</table>
The fourth gene in this set, CG18814 (NP_652673.2), encodes a 267 amino acid protein that displays 26% identity and 47% similarity to human HPGD (“15-hydroxyprostaglandin dehydrogenase NAD(+)”, NP_000851.2). HPGD catalyzes the conversion of the 15-hydroxyl group of prostaglandins into a keto group, which strongly reduces the biologic activity of these molecules. As a result, HPGD is considered the primary enzyme responsible for degradation of prostaglandins. It has been reported that N91, S138, Q148 and Y151 are required for catalysis. [96]. CG18814 residues N89, S137 and Y150 of CG18814 align with these catalytic residues and are oriented at overlapping or directly adjacent positions in close proximity in the structural alignment of the CG18814 model against the structure of HPGD. CG18814 lacks an obvious analog for Q148, which is located in a flexible loop region. However, CG18814 possesses a structurally-aligned loop region which contains Q142. Q148 and Q142 are located within approximately 5Å of each other and oriented similarly in a static superposition of the predicted model and crystal structure and may be functionally analogous. Moreover, both proteins have an identical domain architecture, with a Short-chain dehydrogenase domain (PF00106) spanning from the N-terminus to approximately residue 180. Superposition of human HPGD (2GDZ) and the predicted structure of CG18814 reveals an RMSD of 0.715 Å.
A. Secondary Structure Prediction

**FIG. 91.** Predicted secondary structure of HPGD and CG18814 calculated using PROMALS3D.

<table>
<thead>
<tr>
<th>Conservation:</th>
<th>HPGD_1-266</th>
<th>CG18814_1-267</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consensus aa</td>
<td>MipGKshlhhG.GIG+...</td>
<td>hhp.LLb.K.Khhbl...</td>
</tr>
<tr>
<td>Consensus s</td>
<td>eeeeee</td>
<td>hhhhhhhhhhhhh</td>
</tr>
</tbody>
</table>

**FIG. 92.** Predicted domain architecture of HPGD and CG18814 determined using Interpro and Pfam.

**FIG. 93.** Predicted domain architecture of HPGD and CG18814 determined using Interpro and Pfam.

B. Domain Architecture

**FIG. 92.** Predicted domain architecture of HPGD and CG18814 determined using Interpro and Pfam.

C. Model Generation and Validation

1. Generation

HHpred, Phyre2 and LOMETS were used to identify suitable templates for modeling CG18814. All three tools identify chain A of 2GDZ (human 15-hydroxyprostaglandin
dehydrogenase 1 complexed with NAD\(^+\)) as a high-scoring match suitable for single-template modeling.

<table>
<thead>
<tr>
<th>CG18814 Range</th>
<th>Template Range</th>
<th>% Identity</th>
<th>% Similarity</th>
<th>E-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-242</td>
<td>6-249</td>
<td>31%</td>
<td>59%</td>
<td>1.3e-34</td>
</tr>
</tbody>
</table>

**FIG. 93.** HMM profile of CG18814 aligned against chain A of 2GDZ, a template structure selected for modeling CG18814.

### 2. Validation

The quality of the CG18814 model was evaluated using ProQ2 and ProSA, as described above in the discussion of CG1742. Again, overall model quality is high (a ProSA global Z-score of -6.94). The ProSA per-residue energy score graph is also favorable, showing a score that is consistently negative over a window size of 40 residues, except for the putative active site and a C-terminal loop.

---

92
FIG. 94. Validation of the CG18814 model: ProQ2 quality score mapped to a 3D model of CG18814 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).
D. Structural Analyses

1. Secondary Structure

![Secondary Structure Diagram]

FIG. 95. Pairwise alignment of CG18814 and 2GDZ generated from structural superposition with CG18814 with shared secondary structure elements and conserved residues highlighted.
2. Physiochemical Properties

FIG. 96. Pairwise alignment of CG18814 and 2GDZ generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

E. Analysis of Functional Residues

<table>
<thead>
<tr>
<th>HPGD (PDB: 2GDZ) and CG18814 Model</th>
<th>D. melanogaster Model</th>
<th>Superimposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPGD Structure</td>
<td>D. melanogaster Model</td>
<td>Superimposed</td>
</tr>
</tbody>
</table>

FIG. 97. HPGD (2GDZ, cyan-blue) superimposed on the predicted structure of CG18814 (green-red). RMSD: 0.715 Å.

FIG. 98. HPGD (2GDZ, cyan-blue) superimposed on the predicted structure of CG18814 (green-red), with potential matches for conserved functional residues highlighted.
**Summary Table**

**TABLE 14. Summary of features shared by HPGD and potential *D. melanogaster* ortholog CG18814.**

<table>
<thead>
<tr>
<th>Length (AA)</th>
<th>Domain Architecture (Pfam, range)</th>
<th>Functional Residues (aligned matches in <em>D. melanogaster</em>)</th>
<th>Sequence ID%</th>
<th>Structural Overlap (RMSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>15-hydroxyprostaglandin dehydrogenase NAD(+) (HPGD, NP_000851.2, PDB: 2GDZ)</td>
<td>266</td>
<td>Short chain dehydrogenase domain (PF00106) 6-170</td>
<td>Y151</td>
<td>26% ID 47% SIM 0.715 Å</td>
</tr>
<tr>
<td>Uncharacterized protein (CG18814, NP_652673.2)</td>
<td>267</td>
<td>Short chain dehydrogenase domain (PF00106) 7-166</td>
<td>Y150</td>
<td></td>
</tr>
</tbody>
</table>

**TBXAS1**

The fifth and final gene in this set, CG3616 (“CYP450-9c1,” NP_523850.1) encodes a cytochrome p450 oxidase, which displays 27% identity and 45% similarity to TBXAS1 (“thromboxane A synthase,” NP_001052.2). TBXAS1 localizes to the endoplasmic reticulum membrane, where it catalyzes the conversion of PGH\(_2\) to thromboxane A\(_2\) (TBA\(_2\)). TBXAS1 is categorized as a cytochrome p450 family (CYP450) enzyme based on sequence similarity. This categorization is illustrated by the fact that the only pfam domain identified in TBXAS1 has a Cytochrome P450 domain (PF00067) spanning the majority of the protein. CYP450 enzymes are highly conserved in eukaryotes, where they catalyze many reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids. However, this ubiquitous nature of CYP450 enzymes makes it difficult to identify potential orthologs for any one specific enzyme within the family. CG3616 was selected as a candidate from the set of eight candidates with similar sequences after taking into account the initial sequence comparison, the
transmembrane helix pattern (calculated using TOPCONS) and the alignment score against a multiple sequence alignment of known or high confidence TBXAS1 orthologs

A. Secondary Structure Prediction

FIG. 99. Predicted secondary structure of TBXAS1 and CG3616 calculated using PROMALS3D.
B. Domain Architecture

FIG. 100. Predicted domain architecture of TBXAS1 and CG3616 determined using Interpro and Pfam.

C. Model Generation and Validation

1. Generation

HHpred, Phyre2 and LOMETS were used to identify suitable templates for modeling CG3616. All three tools identify chain A of 4DGZ (human CYP450 3A4) as a high-scoring match suitable for single-template modeling.
FIG. 101. HMM profile of CG3616 aligned against chain A of 4DGZ, a template structure selected for modeling CG3616.

2. Validation

The quality of the CG3616 model was evaluated using ProQ2 and ProSA, as described above in the discussion of CG1742. Again, overall model quality is high (a ProSA global Z-score of -8). The ProSA per-residue energy score graph is also favorable, showing a score that is consistently negative over a window size of 40 residues, except for the N-terminus which is predicted to be a signal peptide and a single loop region.
D. Structural Analyses

A crystal structure is currently unavailable for human thromboxane A synthase. However, based on sequence analysis this enzyme is predicted to display the typical CYP450 fold and high quality models (based on CYP450 crystal structures) are available. For example, a
A high confidence model of human thromboxane A synthase is available from the SWISS MODEL database and was used in this study as a stand-in for the likely actual structure of human thromboxane A synthase.

A structural superposition reveals that the CG3616 model and the predicted human thromboxane A synthase model display a similar secondary structure architecture across the modeled region with substantial full-length overlap. The most notable difference occurs at segment 282-304 of CG3616, which consists of a single short α-helix bordered by loops on either side (α-helix 14). The corresponding segment of the TBXAS1 model (286-325) consists of two α-helices joined by a short loop. The pairwise alignment generated by this structural superposition also indicates substantial conservation of residues having similar physiochemical properties.
FIG. 103. Pairwise alignment of CG3616 and 4DGZ generated from structural superposition with shared secondary structure elements and conserved residues highlighted.
2. Physiochemical Properties

FIG. 104. Pairwise alignment of CG3616 and 4DGZ generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

E. Analysis of Functional Residues

<table>
<thead>
<tr>
<th>TBXAS1 (SWISS-MODEL) and CG3616 Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBXAS1 Model</td>
</tr>
<tr>
<td>[Diagram showing TBXAS1 model]</td>
</tr>
</tbody>
</table>

FIG. 105. TBXAS1 (SWISS-MODEL, cyan) superimposed on the predicted structure of CG3616 (green). RMSD: 0.699 Å.
FIG. 106. TBXAS1 (4DGZ, cyan-blue) superimposed on the predicted structure of CG3616 (green-red), with potential matches for conserved functional residues highlighted.

F. Summary Table

TABLE 15. Summary of features shared by TBXAS1 and potential D. melanogaster ortholog CG3616.

<table>
<thead>
<tr>
<th></th>
<th>Length (AA)</th>
<th>Domain Architecture (Pfam, range)</th>
<th>Functional Residues (aligned matches in D. melanogaster)</th>
<th>Sequence ID%</th>
<th>Structural Overlap (RMSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thromboxane A synthase</td>
<td>534</td>
<td>Cytochrome P450 domain (PF00067) 44-530</td>
<td>C479</td>
<td>27% ID</td>
<td>0.699 Å</td>
</tr>
<tr>
<td>(TBXAS1, NP_001052.2, PDB: N/A (modeled))</td>
<td></td>
<td></td>
<td></td>
<td>45% SIM</td>
<td></td>
</tr>
<tr>
<td>Cytochrome P450-9c1</td>
<td>518</td>
<td>Cytochrome P450 domain (PF00067) 34-513</td>
<td>C462</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(CG3616, NP_610820.1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Group 3: The Most Distant Candidates

Many of the putative orthologs identified in *D. melanogaster* although divergent in sequence from the human target, are similar enough that traditional methods for analyzing and aligning sequences can be utilized. However, for a small subset, the sequences are so divergent (*i.e.*., 20-20% identical) that a more intensive analysis was required to determine whether further study was justified.

PTGS1/PTGS2

The search for a cyclooxygenase ortholog proved to be one of these difficult cases. As noted above, it is widely accepted that *D. melanogaster* lacks a cyclooxygenase ortholog. However, there have been sporadic reports of prostaglandin detected in *D. melanogaster* extracts over the years. The human genome encodes three major cyclooxygenase isozymes via two loci. PTGS1 encodes Cox-1 (a constitutive cyclooxygenase), whereas PTGS2 encodes Cox-2 (which is inducible and expressed in a tissue-specific manner) A truncated and poorly characterized isoform of PTGS1, Cox-3 is also known to be expressed in certain cells, and reportedly demonstrates reduced prostaglandin synthesis activity relative to COX-1 [97]. Cox-1 and Cox-2 operate as homodimers and are characterized by two domains: an N-terminal EGF-like domain (IPR000742) and an animal heme peroxidase domain (PF03098). The former allows for association with the cell and nuclear membrane while the latter is responsible for the cyclooxygenase activity. The Cox enzymes initiate the prostaglandin synthesis pathway by converting arachidonic acid, or other polyunsaturated fatty acid (PUFA) substrates, into the unstable intermediate PGG₂ followed by PGH₂. During this process, it has been reported that a small amount of the PUFA substrate is converted into a racemic mixture of 15-Hydroxyicosatetraenoic acids (*i.e.*, 15-HETEs), which may be processed into lipoxins, a poorly understood class of anti-inflammatory eicosanoids. The Cox-1 and Cox-2 enzymes contains two
active sites: a heme with peroxidase activity, responsible for the reduction of PGG$_2$ to PGH$_2$, and a cyclooxygenase site, where arachidonic acid is converted into the hydroperoxy endoperoxide prostaglandin G$_2$ (PGG$_2$). Extensive studies have been performed on the sheep ortholog of Cox-1, (NP_001009476.1; PDB: 1CQE), which is a proxy for the human ortholog [98]. These studies reveal that cyclooxygenase activity is mediated by Y385, which forms a radical capable of abstracting a hydrogen from the PUFA substrate (e.g., carbon-13 of arachidonic acid). H207 and H388 are also required for the dual peroxidase and cyclooxygenase functions performed by these enzymes [99, 100]. Other functional residues identified in the literature include R120, Q203, V349, and S530 [101]. R120 interacts with C-1 of the PUFA substrate, arachidonic acid [102]. Q203 is conserved among mammalian Cox enzymes, though the function of this residue is currently unknown. V349 is believed to play a role in substrate specificity; mammalian Cox enzymes share this residue and show a substrate preference for arachidonic acid, while invertebrate Cox enzymes that have a leucine at this position display specificity for linoleic acid [103]. and S530 acetylation is the basis for Cox-2 inhibition by aspirin [104].

An initial screen of D. melanogaster using BLASTP failed to identify meaningful hits. However, iterative HMMER searches were able to identify four potential candidates: CG4009 (NP_650588.2), CG6969 (“Cardinal,” NP_651081.1), CG7660 (“Pxt,” NP_650648.3) and CG3477 (“Pxd,” NP_996223.1). Each of these candidates was added to a previously generated multiple sequence alignment containing PTGS1 and known or high confidence predicted orthologs using MAFFT. The presence of absence of functional residues at or near aligned positions was noted. Similarly, the candidates were analyzed to determine the secondary structure and domain architecture of each protein. Following this filtering step, only two proteins remained as viable candidates for further study: CG4009 and CG6969. Both appear to have the requisite catalytic triad and share a similar domain architecture, though they each lack the N-
terminal EGF-like domain. Notably, putative analogs for the H207, Y385 and H388 catalytic triad appear to be present on CG4009 at H163, Y399, H401 and on CG7660 at H222, Y564, H568, respectively.

Structural models of CG4009 and CG6969, and CG7660 were built and validated as described above. PTGS1 (3N8V) and the predicted structure for CG4009 both display a cyclooxygenase fold, and the predicted catalytic domains are superimposable with a RMSD of 1.875 Å. CG7660 and CG6969 each display a similar tertiary structure and with predicted catalytic domains that overlap with PTGS1 with a RMSD of 2.810 and 2.942 Å, respectively. The predicted structures differ from PTGS1 at their N-terminus, which is expected given the absence of the EGF-like domain, though the CG4009 structure is marginally closer. In any event, these three structures present the necessary catalytic residues in a similar orientation as in PTGS1 and so likely possess cyclooxygenase activity.
A. Secondary Structure Prediction

FIG. 107. Predicted secondary structure of PTGS1 and CG4009 calculated using PROMALS3D.
B. Domain Architecture

![Predicted domain architecture of PTGS1 and CG4009 determined using Interpro and Pfam.]

C. Model Generation and Validation

1. Generation

HHpred, Phyre2 and LOMETS were used to identify suitable templates for modeling CG4009. All three tools identified human COX-1 and COX-2, as well as the mouse and sheep orthologs, as potential templates for modeling a substantial portion of the animal haem peroxidase domain which spans 95-617 of CG4009 (e.g., 5F1A, 1IGZ). Lactoperoxidases, alapha-dioxygenase and myeloperoxidase structures are also identified as templates for this portion of CG4009. However, none of these templates aligns well enough with CG4009 to support high confidence single-template modeling. As a result, I-TASSER was used to generate a multi-template model of CG4009 based on six templates: 4HHR, 3FAQ, 3Q9K, 1CXP, 2GJ1, and 1CVU.

### TABLE 16. Templates selected for compositing modeling of CG4009 using I-TASSER.

<table>
<thead>
<tr>
<th>Template</th>
<th>Organism</th>
<th>Protein</th>
<th>% Identity</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>4HHR</td>
<td>A.thaliana</td>
<td>Fatty acid dioxygenase</td>
<td>20%</td>
<td>90%</td>
</tr>
<tr>
<td>3FAQ</td>
<td>B. bubalis</td>
<td>Lactoperoxidase</td>
<td>30%</td>
<td>85%</td>
</tr>
<tr>
<td>3Q9K</td>
<td>B. taurus</td>
<td>Lactoperoxidase</td>
<td>28%</td>
<td>85%</td>
</tr>
<tr>
<td>1CXP</td>
<td>Human</td>
<td>Myeloperoxidase</td>
<td>29%</td>
<td>83%</td>
</tr>
<tr>
<td>2GJ1</td>
<td>B. taurus</td>
<td>Lactoperoxidase</td>
<td>28%</td>
<td>84%</td>
</tr>
<tr>
<td>1CVU</td>
<td>M. musculus</td>
<td>COX-2</td>
<td>18%</td>
<td>76%</td>
</tr>
</tbody>
</table>
The multi-template alignment used by I-TASSER to generate a full-length model of CG4009 is shown below as Fig. 109. The full-length model is predicted to have a TM-SCORE of 0.73±0.11. A TM-SCORE of >0.50 is typically indicative of a correct global fold assignment. This model was subjected to additional refinement (e.g., sidechain packing and loop refinement) as described in the methods section above.
FIG. 109. Multiple Sequence Alignment illustrating CG4009 aligned against template structures selected by I-TASSER for multi-template modeling of CG4009.
2. Validation

The quality of the CG4009 model was evaluated using ProQ2 and ProSA, as described above in the discussion of CG1742. Again, overall model quality is high (a ProSA global Z-score of -6.59). The ProSA per-residue energy score graph is also favorable, showing a score that is consistently negative over a window size of 40 residues except for loop regions and the N-terminus, which is predicted to be a signal peptide.

FIG. 110. Validation of the CG4009 model: ProQ2 quality score mapped to a 3D model of CG4009 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).
D. Structural Analyses

A structural superposition reveals that the CG4009 model and the full-length structure of sheep COX-1 (3N8V) display a similar secondary structure architecture across the modeled region. Conserved secondary structure elements are readily apparent upon visual inspection of the superimposed structures, as illustrated by Fig. 113 below. However, the 3D position of many of these conserved secondary structure elements is shifted in three-dimensional space due to the alternative placement of intervening loop regions. Thus, while many analogous secondary structure elements are indeed present when the CG4009 model and COX-1 (3N8V) are superimposed, the conservation of these elements is not as clear from a pairwise structural alignment as is the case for many of the other enzymes analyzed in this study.
1. Secondary Structure
FIG. 111. Pairwise alignment of CG4009 and 3N8V generated from structural superposition with CG4009 with shared secondary structure elements and conserved residues highlighted.
2. Physiochemical Properties

![Pairwise alignment of CG4009 and 3N8V generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.](image1)

**FIG. 112.** Pairwise alignment of CG4009 and 3N8V generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

E. Analysis of Functional Residues

<table>
<thead>
<tr>
<th>PTGS1 (PDB: 3N8V) and CG4009 Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PTGS1 Structure</strong></td>
</tr>
</tbody>
</table>

![PTGS1 (3N8V, cyan-blue) superimposed on the predicted structure of CG4009 (green-red). RMSD (predicted catalytic domain): 2.660 Å.](image2)
PTGS1 (PDB: 3N8V) and CG4009 Model Superimposed

FIG. 114. PTGS1 (3N8V, cyan-blue) superimposed on the predicted structure of CG4009 (green-red), with potential matches for conserved functional residues highlighted.

F. Summary Table

**TABLE 17.** Templates selected by I-TASSER to model CG4009 using a composite modeling protocol.

<table>
<thead>
<tr>
<th></th>
<th>Length (AA)</th>
<th>Domain Architecture (Pfam, range)</th>
<th>Functional Residues (aligned matches in <em>D. melanogaster</em>)</th>
<th>Sequence ID%</th>
<th>Structural Overlap (RMSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyclooxygenase 1</td>
<td>599</td>
<td>Animal heme peroxidase domain (PF03098) 142-581</td>
<td>H207, Y385 and H388</td>
<td>14% ID</td>
<td>2.660 Å</td>
</tr>
<tr>
<td>(PTGS1, NP_000953.2, PDB: 3N8V)</td>
<td></td>
<td></td>
<td></td>
<td>26% SIM</td>
<td></td>
</tr>
<tr>
<td>Uncharacterized protein</td>
<td>649</td>
<td>Animal heme peroxidase domain (PF03098) 95-617</td>
<td>H163, Y399, and H401</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(CG4009, NP_650588.2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
PTGDS

The search for a potential ortholog for PTGDS also proved to be difficult. PTGDS is a second prostaglandin-D-synthase, which operates in conjunction with the isozyme HPGDS (discussed above). BLAST-based searching failed to identify any \textit{D. melanogaster} candidates for this enzyme (\textit{i.e.}, an apparent ortholog for NP\_000945.3). However, iterative HMMER searches suggested CG33126 (\textit{“NLaz,” NP\_787960.1}) merits further scrutiny despite having only 12\% identity and 31\% similarity to PTGDS. Both proteins share similar domain architecture, namely a Lipocalin (PF00061) domain, which is known to take the form of a beta barrel structure. A model was generated for PTGDS, as described above. Comparative studies revealed that human PTGDS (5WY9) and CG33126 overlap with an RMSD 2.084 \textAA{} and that CG33126 has an aligned match for PTGDS’s catalytic cysteine (C65) at position C67. CG33126 contains an N-terminal alpha helix that is absent from PTGDS, however this can helix is predicted to be a signal peptide. PTGDS contains a similar N-terminal signal peptide that is cleaved during processing (\textit{i.e.}, omitted from the 5WY9 crystal structure).

A. Secondary Structure Prediction

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{predicted_secondary_structure.png}
\caption{Predicted secondary structure of PTGDS and CG33126 calculated using PROMALS3D.}
\end{figure}
B. Domain Architecture

FIG. 116. Predicted domain architecture of PTGDS and CG33126 determined using Interpro and Pfam.

C. Model Generation and Validation

1. Generation

HHpred, Phyre2 and LOMETS were used to identify suitable templates for modeling CG33126. All three tools identify chain A of apolipoprotein D (2HZQ) and chain A of human prostaglandin D synthase (5WY9) as high-scoring matches suitable for single-template modeling. 2HZQ and 5WY9 are β-barrel proteins with a substantially similar tertiary structure. As a result, models generated using either template are substantially similar. For this study, 5WY9 was selected as a template for modeling CG33126.

<table>
<thead>
<tr>
<th>CG33126 Range</th>
<th>Template Range</th>
<th>% Identity</th>
<th>% Similarity</th>
<th>E-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>34-184</td>
<td>26-178</td>
<td>18%</td>
<td>58%</td>
<td>5.0e-07</td>
</tr>
</tbody>
</table>

FIG. 117. HMM profile of CG33126 aligned against chain A of 5WY9, a template structure selected for modeling CG33126.
2. Validation

The quality of the CG33126 model was evaluated using ProQ2 and ProSA, as described above in the discussion of CG1742. Again, overall model quality is high (a ProSA global Z-score of -5.47). The ProSA per-residue energy score graph is also favorable, showing a score that is largely negative over a window size of 40 residues.

FIG. 118. Validation of the CG33126 model: ProQ2 quality score mapped to a 3D model of CG33126 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).
D. Structural Analyses

A structural superposition reveals that the CG33126 model and human prostaglandin D synthase (5WY9) display a similar secondary structure architecture across the modeled region with substantial full-length overlap. The pairwise alignment generated by this structural superposition also indicates substantial conservation of residues having similar physiochemical properties.

1. Secondary Structure

![Pairwise alignment of CG33126 and 5WY9 generated from structural superposition with CG33126 with shared secondary structure elements and conserved residues highlighted.](image)

**FIG. 119.** Pairwise alignment of CG33126 and 5WY9 generated from structural superposition with CG33126 with shared secondary structure elements and conserved residues highlighted.
2. Physiochemical Properties

![Image of pairwise alignment of CG33126 and 5WY9](image)

FIG. 200. Pairwise alignment of CG33126 and 5WY9 generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

E. Analysis of Functional Residues

<table>
<thead>
<tr>
<th>PTGDS (PDB: 5WY9) and CG33126 Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTGDS Structure</td>
</tr>
<tr>
<td><strong>D. melanogaster</strong> Model</td>
</tr>
<tr>
<td>Superimposed</td>
</tr>
</tbody>
</table>

![Image of PTGDS (5WY9, cyan-blue) superimposed](image)

FIG. 201. PTGDS (5WY9, cyan-blue) superimposed on the predicted structure of CG33126 (green-red). RMSD: 2.084 Å.
FIG. 202. PTGDS (5WY9, cyan-blue) superimposed on the predicted structure of CG33126 (green-red), with potential matches for conserved functional residues highlighted.

F. Summary Table

TABLE 18. Summary of features shared by PTGDS and potential *D. melanogaster* ortholog CG33126.

<table>
<thead>
<tr>
<th></th>
<th>Length (AA)</th>
<th>Domain Architecture (Pfam, range)</th>
<th>Functional Residues (aligned matches in <em>D. melanogaster</em>)</th>
<th>Sequence ID%</th>
<th>Structural Overlap (RMSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prostaglandin-D-synthase (PTGDS, NP_000945.3, PDB: 5WY9)</td>
<td>190</td>
<td>Lipocalin domain (PF00061) 40-184</td>
<td>C65</td>
<td>12% ID</td>
<td>2.084 Å</td>
</tr>
<tr>
<td>Neural Lazarillo (CG33126, NP_787960.1)</td>
<td>224</td>
<td>Lipocalin domain (PF00061) 48-190</td>
<td>C67</td>
<td>31% SIM</td>
<td></td>
</tr>
</tbody>
</table>
PTGIS

PTGIS ("prostacyclin synthase") is another enzyme that does not have a clear ortholog in *D. melanogaster*. PTGIS catalyzes the isomerization of PGH$_2$ to prostacyclin, the only prostaglandin with a bicyclic ring structure. PTGIS encodes a 500 amino acid enzyme characterized by a cytochrome p450 family domain ranging from position 30 to 494. The N-terminus is predicted to contain a 200 amino acid signal sequence. BLAST and iterative HMMER searches reveal multiple high-scoring full length matches among the various CYP450 enzymes encoded in the *D. melanogaster* genome. However, CG3466 ("CYP450-4d2," NP_525043.1) was identified as the top candidate based upon similarity of its secondary structure profile. CG3466 possesses an aligned match for PTGIS’s heme axial ligand (C441) at position C449. PTGIS (2IAG) and CG3466 are superimposable with an RMSD of 1.213 Å, despite sharing only 14% identity and 30% similarity at the sequence level.
A. Secondary Structure Prediction

FIG. 203. Predicted secondary structure of PTGIS and CG3466 calculated using PROMALS3D.
B. Domain Architecture

**FIG. 204.** Predicted domain architecture of PTGIS and CG3466 determined using Interpro and Pfam.

C. Model Generation and Validation

1. Generation

HHpred, Phyre2 and LOMETS were used to identify suitable templates for modeling CG3466. All three tools identify various CYP450 structures as suitable for high confidence single-template modeling of full-length CG3466. Chain A of human CYP450 3A4 (4DGZ) was a particularly high-scoring match and selected as a template for generating a model of CG3466 for this study. Human prostacyclin synthase (2IAG), which is also a member of the CYP450 superfamily, was identified as a template. However, it was indicated as having a lower quality score for full-length modeling of this target and so was not selected as a template.
FIG. 205. HMM profile of CG3466 aligned against chain A of 4DGZ, a template structure selected for modeling CG3466.

2. Validation

The quality of the CG3466 model was evaluated using ProQ2 and ProSA, as described above in the discussion of CG1742. Again, overall model quality is high (a ProSA global Z-score of -8.59). The ProSA per-residue energy score graph is also favorable, showing a score that is substantially negative over a window size of 40 residues except for in a local strip region.
FIG. 206. Validation of the CG3466 model: ProQ2 quality score mapped to a 3D model of CG3466 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).

D. Structural Analyses

A structural superposition reveals that the CG3466 model and human prostacyclin synthase (2IAG) display a similar secondary structure architecture across the modeled region with substantial full-length overlap. The pairwise alignment generated by this structural
superposition also indicates substantial conservation of residues having similar physiochemical properties.

1. Secondary Structure

FIG. 207. Pairwise alignment of CG3466 and 2IAG generated from structural superposition with secondary structure elements and conserved residues highlighted.
2. Physiochemical Properties

FIG. 208. Pairwise alignment of CG3466 and 2IAG generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.

E. Analysis of Functional Residues

| PTGIS (PDB: 2IAG) and CG3466 Model |
|-----------------|-----------------|-----------------|
| **PTGIS Structure** | **D. melanogaster Model** | **Superimposed** |

FIG. 209. PTGIS (2IAG, cyan-blue) superimposed on the predicted structure of CG3466 (green-red). RMSD: 1.213 Å.
FIG. 210. PTGIS (2IAG, cyan-blue) superimposed on the predicted structure of CG3466 (green-red), with potential matches for conserved functional residues highlighted.

F. Summary Table

**TABLE 19.** Summary of features shared by PTGIS and potential *D. melanogaster* ortholog CG3466.

<table>
<thead>
<tr>
<th></th>
<th>Length (AA)</th>
<th>Domain Architecture (Pfam, range)</th>
<th>Functional Residues (aligned matches in <em>D. melanogaster</em>)</th>
<th>Sequence ID%</th>
<th>Structural Overlap (RMSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prostacyclin synthase (PTGIS, NP_000952.1, PDB: 2IAG)</td>
<td>500</td>
<td>Cytochrome p450 family (PF00067) 30-494</td>
<td>C441</td>
<td>14% ID</td>
<td>1.213 Å</td>
</tr>
<tr>
<td>Cytochrome p450-4d2 (CG3466, NP_525043.1)</td>
<td>501</td>
<td>Cytochrome p450 family (PF00067) 31-495</td>
<td>C449</td>
<td>30% SIM</td>
<td></td>
</tr>
</tbody>
</table>
**LTC4S**

Leukotriene C4 synthase (LTC4S), is a 150 amino acid enzyme that catalyzes the conjugation of leukotriene A4 with reduced glutathione to form leukotriene C4. BLAST searches against *D. melanogaster* fail to identify an obvious ortholog. However, an iterative HMMER search identified the 165 amino acid protein CG33178 (NP_788904.1) as a potential candidate. The two proteins share full length identity of 18% and similarity of 31% and have an identical domain architecture in the form of a single MAPEG domain (PF01124) spanning the bulk of the protein. Studies of the crystal structure of LTC4S (2PNO) have revealed that it is an integral membrane protein composed of four transmembrane helices and that it functions as a homotrimer. The predicted model of CG33178 displays a similar fold, consisting of a bundle of four alpha helices, and it superimposes on 2PNO with an RMSD of 1.257 Å. CG33178 also displays potential equivalents for each of LTC4S’s catalytically relevant residues (R30, R31 and R104) at positions R51, R53 and R139.

### A. Secondary Structure Prediction

![FIG. 211. Predicted secondary structure of LTC4S and CG33178 calculated using PROMALS3D.](image-url)
B. Domain Architecture

FIG. 212. Predicted domain architecture of LTC4S and CG33178 determined using Interpro and Pfam.

C. Model Generation and Validation

1. Generation

HHpred, Phyre2 and LOMETS were used to identify suitable templates for modeling CG33178. All three tools identify chain A of 2H8A (R. norvegicus Microsomal Glutathione S-transferase 1) as a high-scoring match suitable for single-template modeling. Human leukotriene C4 synthase is also identified as a potential template, but has a lower full-length alignment score.

<table>
<thead>
<tr>
<th>CG33178 Range</th>
<th>Template Range</th>
<th>% Identity</th>
<th>% Similarity</th>
<th>E-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-155</td>
<td>7-145</td>
<td>42%</td>
<td>64%</td>
<td>1.0e-24</td>
</tr>
</tbody>
</table>

FIG. 213. HMM profile of CG33178 aligned against chain A of 2H8A, a template structure selected for modeling CG33178.

2. Validation

The quality of the CG33178 model was evaluated using ProQ2 and ProSA, as described above in the discussion of CG1742. Again, overall model quality is high (a ProSA global Z-score
of -4.04). The ProSA per-residue energy score graph is also favorable, showing a score that is consistently negative over a window size of 40 residues except for a single loop region.

**FIG. 214.** Validation of the CG33178 model: ProQ2 quality score mapped to a 3D model of CG33178 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).
D. Structural Analyses

A structural superposition reveals that the CG33178 model and human leukotriene C4 synthase (2PNO) display a similar secondary structure architecture across the modeled region with substantial full-length overlap. The pairwise alignment generated by this structural superposition also indicates substantial conservation of residues having similar physiochemical properties.

1. Secondary Structure

![Pairwise alignment of CG33178 and 2PNO generated from structural superposition with shared secondary structure elements and conserved residues highlighted.](image)

FIG. 215. Pairwise alignment of CG33178 and 2PNO generated from structural superposition with shared secondary structure elements and conserved residues highlighted.

2. Physiochemical Properties

![Pairwise alignment of CG33178 and 2PNO generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.](image)

FIG. 216. Pairwise alignment of CG33178 and 2PNO generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.
E. Analysis of Functional Residues

<table>
<thead>
<tr>
<th>LTC4S (PDB: 2PNO) and CG33178 Model</th>
<th>D. melanogaster Model</th>
<th>Superimposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTC4S Structure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Superimposed</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**FIG. 217.** LTC4S (2PNO, red) superimposed on the predicted structure of CG33178 (green). RMSD: 1.257 Å.

**FIG. 218.** LTC4S (2PNO, cyan-blue) superimposed on the predicted structure of CG33178 (green-red), with potential matches for conserved functional residues highlighted.
F. Summary Table

<p>| TABLE 20. Summary of features shared by LTC4S and potential D. melanogaster ortholog CG33178. |</p>
<table>
<thead>
<tr>
<th>Length (AA)</th>
<th>Domain Architecture (Pfam, range)</th>
<th>Functional Residues (aligned matches in D. melanogaster)</th>
<th>Sequence ID%</th>
<th>Structural Overlap (RMSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leukotriene C4 synthase (LTC4S, NP_665874.1, PDB: 2PNO)</td>
<td>150</td>
<td>MAPEG domain (PF01124) 8-131</td>
<td>R30, R31 and R104</td>
<td>18% ID 31% SIM</td>
</tr>
<tr>
<td>Uncharacterized protein (CG33178, NP_788904.1)</td>
<td>165</td>
<td>MAPEG domain (PF01124) 29-159</td>
<td>R51, R53 and R139</td>
<td></td>
</tr>
</tbody>
</table>

ALOX5AP

Arachidonate 5-lipoxygenase-activating protein (ALOX5AP, NP_001620.2), the final target protein in this set, is required for leukotriene synthesis by arachidonate 5-lipoxygenase (ALOX5). It has been reported that ALOX5AP functions by binding ALOX5 to the membrane and loading polyunsaturated fatty acid substrates (e.g., arachidonic acid) onto ALOX5 for conversion into leukotriene A₄. Like LTC4S above, ALOX5AP is a MAPEG family protein that functions as a membrane-associated homotrimer. BLAST searches for a potential ortholog do not identify a clear match in D. melanogaster, though iterative HMMER searches reveal CG33177 and CG33178 as potential candidates. Both are MAPEG family proteins with highly similar sequences and predicted tertiary structures. CG33178 displays greater sequential similarity to LTC4S and analogs for the catalytically relevant residues and so was assigned as the top candidate for LTC4S. As a result, CG33177 was selected as the top candidate for a potential ortholog for ALOX5AP. The two proteins display 13% identity and 28% similarity and overlap with an RMSD of 1.024 Å.
A. Secondary Structure Prediction

FIG. 219. Predicted secondary structure of ALOX5AP and CG33177 calculated using PROMALS3D.

B. Domain Architecture

FIG. 220. Predicted domain architecture of ALOX5AP and CG33177 determined using Interpro and Pfam.

C. Model Generation and Validation

1. Generation

HHpred, Phyre2 and LOMETS were used to identify suitable templates for modeling CG33177. All three tools identify chain A of 2H8A (R. norvegicus Microsomal Glutathione S-transferase 1) as a high-scoring match suitable for single-template modeling.
<table>
<thead>
<tr>
<th>CG33177 Range</th>
<th>Template Range</th>
<th>% Identity</th>
<th>% Similarity</th>
<th>E-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>23-156</td>
<td>8-144</td>
<td>44%</td>
<td>63%</td>
<td>7.8e-28</td>
</tr>
</tbody>
</table>

---

**FIG. 221.** HMM profile of CG33177 aligned against chain A of 2H8A, a template structure selected for modeling CG33177.

2. Validation

The quality of the CG33177 model was evaluated using ProQ2 and ProSA, as described above in the discussion of CG1742. Again, overall model quality is high (a ProSA global Z-score of -4.33). The ProSA per-residue energy score graph is also favorable, showing a score that is consistently negative over a window size of 40 residues, except for the flexible C-terminal α-helix.
FIG. 222. Validation of the CG33177 model: ProQ2 quality score mapped to a 3D model of CG33177 (top); ProSA global quality score ranking (lower left) and per-residue quality graph (lower right).

D. Structural Analyses

A structural superposition reveals that the CG33177 model and human ALOX5AP (2Q7M) display a similar secondary structure architecture across the modeled region with full-length overlap. The pairwise alignment generated by this structural superposition also indicates substantial conservation of residues having similar physiochemical properties.
1. Secondary Structure

FIG. 223. Pairwise alignment of CG33177 and 2Q7M generated from structural superposition with shared secondary structure elements and conserved residues highlighted.

2. Physiochemical Properties

FIG. 224. Pairwise alignment of CG33177 and 2Q7M generated from structural superposition with conserved residues highlighted using the CLUSTALX color scheme.
E. Analysis of Functional Residues

<table>
<thead>
<tr>
<th>ALOX5AP (PDB: 2Q7M) and CG33177 Model</th>
<th>ALOX5AP Structure</th>
<th>D. melanogaster Model</th>
<th>Superimposed</th>
</tr>
</thead>
</table>

FIG. 225. ALOX5AP (2Q7M, cyan) superimposed on the predicted structure of CG33177 (green). RMSD: 1.024 Å.

FIG. 226. ALOX5AP (2Q7M, cyan-blue) superimposed on the predicted structure of CG33177 (green-red), with potential matches for conserved functional residues highlighted.
### F. Summary Table

**TABLE 21. Summary of features shared by ALOX5AP and potential *D. melanogaster* ortholog CG33177.**

<table>
<thead>
<tr>
<th>Length (AA)</th>
<th>Domain Architecture (Pfam, range)</th>
<th>Functional Residues (aligned matches in <em>D. melanogaster</em>)</th>
<th>Sequence ID%</th>
<th>Structural Overlap (RMSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arachidonate 5-lipoxygenase-activating protein (ALOX5AP, NP_001620.2, PDB: 2Q7M)</td>
<td>161</td>
<td>MAPEG domain (PF01124) 5-136</td>
<td>N/A</td>
<td>13% ID 1.024 Å</td>
</tr>
<tr>
<td>Uncharacterized protein CG33177, NP_788903.1)</td>
<td>167</td>
<td>MAPEG domain (PF01124) 30-161</td>
<td>N/A</td>
<td>28% SIM</td>
</tr>
</tbody>
</table>
Chapter 4. Discussion

Our results suggest that *D. melanogaster* may possess a set of eicosanoid synthesis enzymes similar to the canonical eicosanoid synthesis enzymes found in humans and mammals generally. At this stage, it is unknown whether these enzymes form a functional eicosanoid synthesis pathway similar to the mammalian pathway or if *D. melanogaster* possesses a unique eicosanoid synthesis pathway (e.g., structural differences in the *D. melanogaster* enzymes may result in the processing of different PUFA substrates and synthesis of alternative lipid mediators other than the canonical eicosanoids). Notwithstanding these caveats, a schematic of a putative *D. melanogaster* eicosanoid synthesis pathway based on our analysis is shown as Fig. 227 below. Notably, the proposed pathway appears to account for a full complement of prostaglandin synthesis enzymes. A functional thromboxane synthesis pathway may also be present, as potential orthologs for thromboxane A synthase have been identified. However, given the structural and sequential similarity of TBXA synthase to the numerous functionally unrelated cytochrome p450 oxidase in the *D. melanogaster* genome, speculation must be reserved pending experimental validation. The existence of a leukotriene synthesis arm of the pathway could not be fully resolved since a thorough search of the *D. melanogaster* genome has failed to identify a potential lipoxygenase, which is critical for the initial processing of the PUFA substrate into leukotriene intermediates. Interestingly, potential orthologs exist for each of the downstream leukotriene processing enzymes. Based upon our current understanding, a functional lipoxygenase is necessary for HETE intermediates and the corresponding final products (e.g., lipoxins and other non-canonical eicosanoids).
FIG 227. A theoretical *D. melanogaster* eicosanoid synthesis pathway.
Studies using conventional sequence analysis techniques alone have failed to demonstrate the possible existence of a functional eicosanoid synthesis pathway and these negative results suggested that flies do not possess the ability to synthesize eicosanoids [105, 106]. However, our data challenges this view and suggest that the fly genome may in fact possess potential orthologs for a majority of the eicosanoid synthesis enzymes. These candidates have thus far eluded detection because of the high degree of sequential divergence compared to the mammalian enzymes. However, when compared at the structural level it is clear that these candidates share the same overall fold and matches for the known or predicted catalytic residues. The existence of eicosanoid synthesis enzymes would partially explain recent studies that have identified prostaglandins or prostaglandin-like compounds in *D. melanogaster* when AA was provided in their diet [106-108]. However, questions remain as to the exact role played by these enzymes and the intermediate and final end products of the pathway.

We speculate that the apparently missing lipoxygenase may be partially explained by *D. melanogaster* expressing a cyclooxygenase with lipoxygenase activity. Dual functional cyclooxygenases have been identified which are able to generate prostaglandins and lipoxygenase products (e.g. HPETEs). For example, mammalian cyclooxygenases have been shown to produce limited amounts of 11- and 15-HPETE as a by-product in addition to PGH, [109-112], and aspirin-acetylated COX-2 has been shown to produce 15R-HETEs [113]. Similar activity by one of the putative *D. melanogaster* COX orthologs could provide upstream processing needed for 11- and 15-HPETE, though this fails to provide the 5-HPETE intermediate needed for downstream leukotriene synthesis. Alternatively, 5-HETEs may be produced using an unconventional mechanism by the COX candidates or perhaps by an unidentified enzyme. The existence of structurally conserved downstream leukotriene synthesis enzymes suggests that the upstream intermediates are likely present in some form. *D. melanogaster* may possess a CYP450
enzyme capable of synthesizing 5-HETE; lipoxygenase-like CYP450s have been identified that produce 5-, 8-, 9-, 11-, 12- and 20-HETEs [114, 115]. Recent studies have reported on bacterial lipoxygenases which are highly divergent from animal and plant lipoxygenases, both in terms of overall sequence conservation and the presence of expected catalytic residues [116]. There is also a possibility that D. melanogaster expresses a functional 5-LOX that radically differs from the canonical profile of a lipoxygenase.

The existence of a functional eicosanoid synthesis pathway requires both the presence of the requisite synthetic enzymes and suitable substrates. The mammalian eicosanoids are primarily derived from 20-carbon PUFAs (arachidonic acid, eicosapentaenoic acid and dihomo-γ-linolenic acid), as well as from 18-carbon PUFAs (γ-linolenic acid, α-linolenic acid and linoleic acid). It has been reported that the D. melanogaster lacks homologs for the mammalian Δ5 and Δ6 desaturases based upon sequence analyses, and so it should be unable to synthesize 20–22 carbon PUFAs from essential fatty acid precursors [117]. Studies by the same group have also shown that flies raised in the absence of 20-carbon PUFAs for several generations remain healthy, implying that 20-carbon PUFAs are non-essential [106]. Based on these findings it is generally accepted that D. melanogaster relies on PUFAs obtained from its vegetarian diet, which is generally limited to the essential 18-carbon fatty acids linoleic acid (18:2n-6) and α-linolenic acid (18:3n-3).

In the vertebrate pathway, the 18-carbon fatty acids linoleic acid and α-linolenic acid can be processed into the 20-carbon PUFAs arachidonic acid and eicosapentaenoic acid, respectively, via a three-step process. First, a double bond is removed by a Δ6 desaturase and then an elongase extends the PUFA chain by the addition of two carbons. Finally, a Δ5 desaturase removes a double bond, completing the conversion. These 20-carbon PUFAs are then capable of being processed by COX or LOX to generate the canonical eicosanoids. 18-carbon
PUFAs (e.g., linoleic acid) may also be processed by 15-LOX into hydroxyoctadecadienoic acids (9- and 13-HODEs), a subfamily of non-canonical eicosanoids. Based on the current understanding that *D. melanogaster* lacks Δ5 and Δ6 desaturase activity, it would be expected that the 18-carbon PUFAs obtained from its diet are processed into HODEs as opposed to eicosanoid derivatives of the 20-carbon PUFAs. Recent studies have validated this hypothesis, confirmed the presence of 9- and 13-HODEs in *D. melanogaster* extracts [118]. However, given the reports of prostaglandin-like molecules in *D. melanogaster* extracts, it is probable that additional lipid processing is taking place to produce compounds similar to the canonical eicosanoids. Contrary to the published literature, our HMMER-based sequence analysis of the *D. melanogaster* genome has identified a potential Δ5 and Δ6 desaturase which is not identified by typical BLAST searches. The uncharacterized genes CG17928 and CG13279 (NP_609810.1 and NP_477154.1, respectively) encode proteins with reasonable full-length alignment to the human Δ5 and Δ6 desaturase. These *D. melanogaster* proteins display domain architecture similar to the human enzymes and possess the critical His box 1, 2 and 3 motifs essential for desaturase activity [119]. Based on these findings, we suggest that *D. melanogaster* may in fact have a mechanism of converting C18 to C20 PUFAs. The existence of this functionality and the nature and extent of expression of these putative desaturases will need to be experimentally validated, though biochemical evidence may be difficult to obtain if expression is occurring at a low level or in a limited population of tissue. Most of the biochemical studies reporting prostaglandin-like compounds in *D. melanogaster* have identified these compounds in a limited subset of tissue (e.g., in reproductive tissue during oogenesis) [120]. However, based on our findings it appears that the existence of Δ5 and Δ6 desaturases cannot be conclusively ruled out at this time.

The present study suggests that *D. melanogaster* has a set of eicosanoid synthesis enzymes, and in fact may possess a functional equivalent for the mammalian eicosanoid
synthesis pathway. Due to the limitations of traditional sequence analysis techniques we have explored advanced approaches to identify orthologs and demonstrate that structural models generated for several of these distant matches appear to display a high degree of structural similarity with members of the mammalian eicosanoid synthesis pathway and possess putative analogs for known functional residues. These initial findings, combined with other recent biochemical studies, suggest the possibility that *D. melanogaster* likely utilizes of eicosanoid signaling, raising the possibility that it may be useful as a model for the study of eicosanoid signaling and inflammation. Our studies with *Drosophila*, a system with low genetic redundancy and remarkable biological conservation will be highly relevant and transferrable to designing therapies for chronic disorders in humans. Our findings will also be relevant to the development of novel insect control measures.
References


151


43. Kikuta Y, Kusunose E, Kusunose M. Prostaglandin and leukotriene ω-hydroxylases.

44. Roman RJ. P-450 metabolites of arachidonic acid in the control of cardiovascular function.
Physiological reviews. 2002 Jan 1;82(1):131-85.


47. Pulichino AM, Rowland S, Wu T, Clark P, Xu D, Mathieu MC, Rieudeau D, Audoly LP.


49. Wallace JL, Bak A, McKnight W, Asfaha S, Sharkey KA, MacNaughton WK.

EVENTS IN VERTEBRATE AND INVERTEBRATE LINEAGES. Journal of Biological Chemistry. 2004 Apr 2;279(14):13624-33.


79. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in


109. Thuresson ED, Lakkides KM, Smith WL. Pgg211r-Hpete and 15r/S-Hpete are Formed From Different Conformers of Arachidonic Acid in the Prostaglandin Endoperoxide H
Synthase-1 Cyclooxygenase Site. In Eicosanoids and Other Bioactive Lipids in Cancer, Inflammation, and Radiation Injury, 5 2002 (pp. 67-72). Springer US.


163


118. Watson M. Identification of lipid mediators using Drosophila to dissect function and role in inflammatory disease (Doctoral dissertation, University of Birmingham).
