http://researchspace.auckland.ac.nz

ResearchSpace@Auckland

Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author’s right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

To request permissions please use the Feedback form on our webpage. http://researchspace.auckland.ac.nz/feedback

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the Library Thesis Consent Form.
Functional and Structural Analyses of an Olfactory Receptor from *Drosophila melanogaster*

Aidan Kiely

October 2008

Supervisors: Andrew V. Kralicek
David R. Greenwood
David L. Christie
Richard D. Newcomb

A thesis submitted in fulfillment of the requirements of Doctor of Philosophy in Biological Science
Abstract

In insects, olfaction is mediated by a large family of integral membrane proteins, called olfactory receptors (ORs), that mediate the transduction of odorant binding into a neuronal signal. A functional assay for insect ORs was developed utilising calcium imaging in Sf9 cells. The *Drosophila melanogaster* OR, Or22a, was expressed using transient transfection, and its activity measured by monitoring increased intracellular calcium levels using a calcium–sensitive dye. The interaction of the odorants ethyl butyrate, pentyl acetate and ethyl acetate with Or22a were both dose–dependent and sensitive, with EC$_{50}$ values of $1.53 \times 10^{-11}$ M, $5.61 \times 10^{-10}$ M and $3.72 \times 10^{-9}$ M, respectively. Furthermore, Or22a expressed in Sf9 cells has a similar response profile to a range of odorants previously tested *in vivo*. This assay system will provide a useful tool for the investigation of insect olfactory receptor structure and function.

A consensus of eleven transmembrane (TM) domain prediction algorithms suggested a model for Or22a that contains seven TM domains, reminiscent of GPCRs. To test this model empirically, the membrane topology of Or22a was determined using epitope–tagging of predicted loops followed by immunochemistry. These experiments revealed that Or22a has seven TM domains but that its orientation in the membrane is opposite to that of GPCRs, having a cytoplasmic N–terminus. This orientation was also observed for *Epiphyas postvittana* Or1, which suggests that this inverted topology may be common to all insect ORs.

To test whether Or22a forms higher order structures, fluorescence resonance energy transfer (FRET) between cyan and yellow fluorescent proteins inserted into the intracellular loops of Or22a was employed. The third intracellular loop interacts strongly with itself
in homo-multimers, with interactions between the first and first loops and first and third
loops also observed. These experiments show that ligand binding ORs can form multimeric
structures in heterologous cells. The co-transfection of Or83b into S2 cells had no impact
on these interactions, however Or83b is likely expressed in this cell line. Finally, models of
how a ligand binding OR interacts physically with the ion channel Or83b are presented,
and approaches that could be used to distinguish between these models are discussed.
I would first like to acknowledge the guidance I have received from my supervisors: Andrew Kralicek, David Greenwood, Richard Newcomb and David Christie. Their support, encouragement and enthusiasm for all aspects of my PhD research provided the platform from which my research was built. Thank you isn’t enough!

I believe that research, like football, is a team effort. I would like to extend my gratitude to the many members of the Molecular Olfaction group at HortResearch who have helped in many ways. It has been a pleasure being part of your team! I would especially like to thank Astrid, who taught me the basics of cell culture and molecular biology with incredible patience; Colm, the best lab barista there is, with whom many ideas were generated, some of which were even vaguely scientific; and Edwige, for the constant encouragement and support when everything seemed impossible. To Melissa, Tam, Pia, Doreen, Andy, Pablo, Nadeesha, Caroline, Pia, Sylvia, Cyril and Clinton; sharing, celebrating and commiserating the many highs and lows of post-graduate research with you has been incredible. Thank you all for the support!

There are many other people at HortResearch who have helped make working there an absolute pleasure. Special thanks to Phil for pool, the Social Club, the Volleyball Crew and the Helldorados for the many enjoyable hours spent not doing research. I have made many friends during my time at Hort, and your support has been appreciated as well. Thank you all!

Coral Warr of Monash University, Australia, provided cDNA clones of the *Drosophila* olfactory receptors used in this thesis. Astrid Authier and Andrew Kralicek cloned the pIB-Or22a-dsRed, pIB-Or22a-aequorin and pIB-Or22a-Or83b expression constructs. Melissa Jordan provided *E. postvittana* Or1 used for topology analysis. Pete Murphy supplied the dsRed gene. Matt Templeton suggested Triton X–114 phase separation as a potential solution to problems I was having solubilising olfactory receptors for western blot analysis. Thanks!

Paul Sutherland and Ian Hallet of the HortResearch Microscopy Suite provided in depth guidance about the detection and imaging of immunoochemical samples. Jacqui Ross and
Hillary Holloway of the University of Auckland Biomedical Imaging Research Unit taught me the basics of confocal microscopy. Ilva Rupenthal made many important suggestions about methods for the analysis of fluorescence resonance energy transfer. The microscopy portions of this thesis would have been much more difficult without their expertise.

This thesis was prepared using \LaTeX, and I thank Marcus Davy for suggesting it as an alternative to Microsoft Word. I think it has made the process of compiling a document this size much easier than the alternative, and would recommend it to anyone starting to write their thesis.

I acknowledge the support of The Agricultural and Marketing Research and Development Trust (AGMARDT) for the provision of a Doctoral Scholarship. I thank the University of Auckland for awarding me the Duffus Lubecki Scholarship. I would like to thank the School of Biological Sciences for awarding me a Travel Scholarship to support my attendance at the International Symposium of Olfaction and Taste. My research has also been supported by grants from the Foundation for Research, Science and Technology and the Ministry of Research, Science and Technology of New Zealand.

I would like to thank my family: Barbara, Maurice, Claire, Liam and Stella for their encouragement and support in the undertaking of this PhD, and in all of my endeavours. A big thanks to Sheryl, Chris, Colin, Sarah and Sherily for your support too!

Lastly, but most importantly, I would like to thank my wife, Katrina, for her support and understanding during the course of my research, and especially over the writing up period. It simply would not have been possible without your efforts. Thank you!
# Contents

Abstract i  
Acknowledgments iii  
Table of Contents v  
List of Figures ix  
List of Tables xii  
List of Abbreviations xiii  
List of Publications xv  

1 General Introduction 1  
1.1 Morphology of the insect olfactory system 2  
1.2 Perireceptor events 5  
1.3 Olfactory system signalling 8  
1.4 G Protein-coupled receptors 9  
1.5 Olfactory receptor genes 13  
1.6 Odour coding in the *Drosophila antennae* 16  
1.7 Aims of this project 19
## Development of an *In Vitro* Olfactory Receptor Assay

### 2.1 Introduction

- **2.1.1 Receptor functional assays**
- **2.1.2 Insect cell expression systems**
- **2.1.3 The use of heterologous expression systems to study olfactory system signalling**
- **2.1.4 Aims**

### 2.2 Materials and Methods

- **2.2.1 Materials**
- **2.2.2 Expression vector construction**
- **2.2.3 Insect cell culture**
- **2.2.4 Transfection of insect cell lines**
- **2.2.5 RT–PCR**
- **2.2.6 Functional assay of olfactory receptors using calcium imaging**

### 2.3 Results

### 2.4 Discussion

## Determination of the Transmembrane Topology of Insect Olfactory Receptors

### 3.1 Introduction

- **3.1.1 Prediction of membrane topology of proteins**
- **3.1.2 Experimental techniques for topology determination**
- **3.1.3 Topology of insect ORs**
- **3.1.4 Aims**
### Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2 Materials and Methods</td>
<td>50</td>
</tr>
<tr>
<td>3.2.1 Materials</td>
<td>50</td>
</tr>
<tr>
<td>3.2.2 Transmembrane domain prediction of Or22a</td>
<td>50</td>
</tr>
<tr>
<td>3.2.3 Construction of epitope–tagged ORs</td>
<td>51</td>
</tr>
<tr>
<td>3.2.4 Activity assays on epitope tagged ORs</td>
<td>52</td>
</tr>
<tr>
<td>3.2.5 Membrane fractionation by differential centrifugation</td>
<td>53</td>
</tr>
<tr>
<td>3.2.6 Membrane protein extraction by Triton X–114 phase separation</td>
<td>53</td>
</tr>
<tr>
<td>3.2.7 Western blotting</td>
<td>54</td>
</tr>
<tr>
<td>3.2.8 Coating coverslips with poly–L–lysine</td>
<td>55</td>
</tr>
<tr>
<td>3.2.9 Preparation of formaldehyde fixative</td>
<td>55</td>
</tr>
<tr>
<td>3.2.10 Immunohistochemistry of S2 Cells</td>
<td>56</td>
</tr>
<tr>
<td>3.2.10.1 Detection of antigens under permeabilized conditions</td>
<td>56</td>
</tr>
<tr>
<td>3.2.10.2 Detection of antigens under non–permeabilized conditions</td>
<td>57</td>
</tr>
<tr>
<td>3.2.11 Imaging of immunochemistry slides</td>
<td>57</td>
</tr>
<tr>
<td>3.3 Results</td>
<td>58</td>
</tr>
<tr>
<td>3.4 Discussion</td>
<td>64</td>
</tr>
<tr>
<td>4 Investigation of Olfactory Receptor Oligomerisation</td>
<td>71</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>71</td>
</tr>
<tr>
<td>4.1.1 Techniques for investigating protein–protein interactions</td>
<td>72</td>
</tr>
<tr>
<td>4.1.2 Approaches for the measurement of FRET</td>
<td>74</td>
</tr>
<tr>
<td>4.1.3 Olfactory receptor oligomerisation</td>
<td>75</td>
</tr>
</tbody>
</table>
4.1.4 Aims ................................................................. 76

4.2 Materials and Methods .............................................. 78

4.2.1 Materials ......................................................... 78

4.2.2 Cell culture and transfection ................................. 78

4.2.3 RT–PCR ............................................................ 78

4.2.4 Construction of Or22a fusions with fluorescent proteins .... 79

4.2.5 Activity assays on fluorescent protein–tagged Or22a ....... 80

4.2.6 Fluorescent microscopy and FRET imaging ............... 80

4.3 Results ............................................................... 81

4.4 Discussion ............................................................ 84

5 Concluding Discussion .................................................. 92

5.1 Summary of Results ................................................. 92

5.1.1 In vitro olfactory receptor assay ............................... 93

5.1.2 Topology analysis of ORs ....................................... 93

5.1.3 Analysis of the oligomeric structure of ORs ................. 93

5.2 Current hypotheses and future directions ....................... 94

A Appendix A: Vector Diagrams ......................................... 102

B Appendix B: Criteria for the Determination of Responding Cells 105

C Appendix C: Topology Prediction Algorithm Description and Output 107

References ................................................................. 116
# List of Figures

1.1 Ultrastructure of the insect antenna ......................................................... 3
1.2 General morphology of the insect olfactory sensillum ................................. 3
1.3 Crystal structure of bovine rhodopsin ....................................................... 10
1.4 A model of signal transduction in olfactory systems .................................. 11
1.5 Topology of classical GPCRs and ORs ...................................................... 16
1.6 Response of *D.melanogaster* ORs ............................................................ 17
1.7 Sub-structure of the insect brain ............................................................... 18
1.8 Functional map of the neuronal structure of *Drosophila* .......................... 19

2.1 RT-PCR of Sf9 cells transfected with pIB-Or22a ..................................... 34
2.2 Response of single Sf9 cells loaded with Fluo-4 and transfected with pIB-Or22a ................................................................. 35
2.3 Comparative fluorescence of responding and non-responding Sf9 cells transfected with pIB-Or22a over the timecourse of a calcium assay. ............. 36
2.4 Transfection efficiency of four transfection reagents ................................. 37
2.5 Optimisation of the Escort transfection protocol ....................................... 38
2.6 The response of Sf9 cells transfected with Or22a or co-transfected with Or22a and Or83b. ............................................................. 38
2.7 Response profile of Or22a expressed in Sf9 cells ..................................... 39
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.8</td>
<td>Dose response of Or22a to five odorants</td>
<td>40</td>
</tr>
<tr>
<td>3.1</td>
<td>Schematic diagram outlining the principle of topology determination by epitope tagging.</td>
<td>47</td>
</tr>
<tr>
<td>3.2</td>
<td>Schematic diagram showing the distribution of <em>D. melanogaster</em> Or22a transmembrane domains predicted by eleven prediction algorithms</td>
<td>58</td>
</tr>
<tr>
<td>3.3</td>
<td>Analysis of the expression of epitope–tagged Or22a in S2 cells using western blotting.</td>
<td>61</td>
</tr>
<tr>
<td>3.4</td>
<td>Responses of epitope–tagged versions of Or22a to $10^{-5}$ M ethyl butyrate and EpOr1 to $10^{-5}$ M citral as monitored using calcium imaging in transfected Sf9 cells.</td>
<td>62</td>
</tr>
<tr>
<td>3.5</td>
<td>Fluorescence microscopy images showing the orientation of rhodopsin TM1 controls.</td>
<td>63</td>
</tr>
<tr>
<td>3.6</td>
<td>Fluorescence microscopy images showing the orientation of <em>D. melanogaster</em> Or22a and <em>E. postvittana</em> Or1.</td>
<td>63</td>
</tr>
<tr>
<td>3.7</td>
<td>Fluorescence microscopy images showing the full topology of <em>D. melanogaster</em> Or22a.</td>
<td>64</td>
</tr>
<tr>
<td>3.8</td>
<td>Experimentally verified model of the transmembrane topology of Or22a.</td>
<td>65</td>
</tr>
<tr>
<td>4.1</td>
<td>Principle of FRET in investigating protein–protein interactions.</td>
<td>74</td>
</tr>
<tr>
<td>4.2</td>
<td>Mean odorant response of fluorescent protein–tagged versions of Or22a</td>
<td>82</td>
</tr>
<tr>
<td>4.3</td>
<td>Single cell FRET studies detect interactions between CFP and YFP tagged Or22a</td>
<td>85</td>
</tr>
<tr>
<td>4.4</td>
<td>Single cell FRET assessing the impact of Or83b co–expression with OR FRET constructs</td>
<td>86</td>
</tr>
<tr>
<td>4.5</td>
<td>RT–PCR of cDNA extracted from native S2 cells</td>
<td>86</td>
</tr>
<tr>
<td>4.6</td>
<td>Models of olfactory receptor homo–oligomerisation</td>
<td>88</td>
</tr>
<tr>
<td>5.1</td>
<td>Models of olfactory receptor function</td>
<td>98</td>
</tr>
</tbody>
</table>
A.1  pIB-Or22a ................................................................. 103
A.2  pIB-Or83b ................................................................. 103
A.3  Schematic diagram and nucleotide sequence of Rhodopsin TM1 constructs.  104
C.1  Graphical output of the TMHMM program ................................. 108
C.2  Graphical output of the TMPred program ................................. 109
C.3  The output of the Predator method of secondary structure prediction when
       queried with the sequence of Or22a. ................................ 112
C.4  Output of HMMTOP ......................................................... 113
List of Tables

3.1 Summary output of topology prediction programs .......................... 59

C.1 Output of the DAS program when queried with the Or22a sequence . . . . 108
C.2 Output of the TMHMM program when queried with the Or22a sequence . 109
C.3 Output of TMPred ................................................................. 110
C.4 Predicted transmembrane segments for Or22a as determined using TMAP 110
C.5 Predicted transmembrane segments for Or22a as determined using SPLIT . 112
C.6 Output from TMR2 ................................................................. 114
C.7 Output from SOSUI. ............................................................ 115
C.8 Output from TOPPRED .......................................................... 115
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔF</td>
<td>Change in fluorescence</td>
</tr>
<tr>
<td>°C</td>
<td>Degrees celsius</td>
</tr>
<tr>
<td>AC</td>
<td>Adenylyl cyclase</td>
</tr>
<tr>
<td>ADP</td>
<td>Adenosine diphosphate</td>
</tr>
<tr>
<td>AL</td>
<td>Antennal lobe</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>BiFC</td>
<td>Bimolecular fluorescence complementation</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair(s)</td>
</tr>
<tr>
<td>BRET</td>
<td>Bioluminescence resonance energy transfer</td>
</tr>
<tr>
<td>cAMP</td>
<td>Cyclic adenosine monophosphate</td>
</tr>
<tr>
<td>CCD</td>
<td>Charge–coupled device</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary deoxyribonucleic acid</td>
</tr>
<tr>
<td>CFP</td>
<td>Cyan fluorescent protein</td>
</tr>
<tr>
<td>cGMP</td>
<td>Cyclic guanosine monophosphate</td>
</tr>
<tr>
<td>CGRP</td>
<td>Calcitonin–gene–related peptide</td>
</tr>
<tr>
<td>C–terminus</td>
<td>Carboxy–terminus</td>
</tr>
<tr>
<td>DAPI</td>
<td>4,6–diamidino–2–phenylindole</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxynucleotide triphosphate</td>
</tr>
<tr>
<td>EAG</td>
<td>Electroantennogram</td>
</tr>
<tr>
<td>EC₅₀</td>
<td>Half maximal effective concentration</td>
</tr>
<tr>
<td>FRET</td>
<td>Fluorescence resonance energy transfer</td>
</tr>
<tr>
<td>g</td>
<td>Gravitational constant</td>
</tr>
<tr>
<td>GDP</td>
<td>Guanosine diphosphate</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>GPCR</td>
<td>G protein–coupled receptor</td>
</tr>
<tr>
<td>G protein</td>
<td>Guanine nucleotide–binding proteins</td>
</tr>
<tr>
<td>GRK</td>
<td>G protein–coupled receptor kinase</td>
</tr>
<tr>
<td>GTP</td>
<td>Guanosine triphosphate</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>HEK293</td>
<td>Human embryonic kidney–293</td>
</tr>
<tr>
<td>HMM</td>
<td>Hidden Markov model</td>
</tr>
<tr>
<td>IP₃</td>
<td>Inositol triphosphate</td>
</tr>
<tr>
<td>Kᴍ</td>
<td>Michaelis constant</td>
</tr>
<tr>
<td>kDa</td>
<td>kilo Daltons</td>
</tr>
<tr>
<td>LDS</td>
<td>Lithium dodecyl sulphate</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger ribonucleic acid</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>N–termius</td>
<td>Amino–terminus</td>
</tr>
<tr>
<td>OBP</td>
<td>Odorant binding protein</td>
</tr>
<tr>
<td>ODE</td>
<td>Odorant degrading enzyme</td>
</tr>
<tr>
<td>OR</td>
<td>Olfactory receptor</td>
</tr>
<tr>
<td>ORN</td>
<td>Olfactory receptor neuron</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide electrophoresis</td>
</tr>
<tr>
<td>PBP</td>
<td>Pheromone binding protein</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate–buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>pH</td>
<td>Potential of hydrogen</td>
</tr>
<tr>
<td>PIP₂</td>
<td>Phosphatidylinositol bisphosphate</td>
</tr>
<tr>
<td>PIPES</td>
<td>Piperazine-N,N’-bis(2-ethanesulfonic acid)</td>
</tr>
<tr>
<td>PLC</td>
<td>Phospholipase C</td>
</tr>
<tr>
<td>PVDF</td>
<td>Polyvinylidene fluoride</td>
</tr>
<tr>
<td>RAMP</td>
<td>Receptor activity–modifying protein</td>
</tr>
<tr>
<td>RANTES</td>
<td>Regulated activation normal T cell expressed secreted</td>
</tr>
<tr>
<td>RCSB</td>
<td>Research Collaboratory for Structural Bioinformatics</td>
</tr>
<tr>
<td>ROI</td>
<td>Region of interest</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RTK</td>
<td>Receptor tyrosine kinase</td>
</tr>
<tr>
<td>RT–PCR</td>
<td>Reverse transcription polymerase chain reaction</td>
</tr>
<tr>
<td>S2</td>
<td>Schneider 2</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulphate</td>
</tr>
<tr>
<td>S.E.M.</td>
<td>Standard error of the mean</td>
</tr>
<tr>
<td>Sf9</td>
<td>Spodoptera frugiperda 9</td>
</tr>
<tr>
<td>SNMP</td>
<td>Sensory neuron membrane protein</td>
</tr>
<tr>
<td>TBS</td>
<td>Tris–buffered saline</td>
</tr>
<tr>
<td>TM</td>
<td>Transmembrane</td>
</tr>
<tr>
<td>Y2H</td>
<td>Yeast two–hybrid</td>
</tr>
<tr>
<td>YFP</td>
<td>Yellow fluorescent protein</td>
</tr>
</tbody>
</table>
List of Publications

The research presented in this thesis is my own work, and contributed to the following publications:


In addition, the following publications are in review or in preparation:


General Introduction

Olfaction, or the detection of volatile signals in an organism’s environment (Mombaerts, 1999), is the oldest of the senses from an evolutionary viewpoint. The common features of olfactory systems are thought to have evolved approximately 500 million years ago (Vosshall et al., 2000), and still share striking similarities in divergent species today. Insects provide a powerful model system for the study of olfactory systems, as they are amenable to genetic manipulation and are easy to study with well established molecular, electrophysiological and behavioural techniques.

Olfaction underlies a range of important insect behaviours, many of which have a significant human impact. Approximately 25% of annual agricultural losses are due to the activity of herbivorous insects (Jacquin-Joly & Merlin, 2004). Insects are also the primary vector of many diseases, including malaria, that cause millions of deaths each year in Africa alone (Jacquin-Joly & Merlin, 2004). Some insect behaviours are
desirable, such as flower pollination, which is important for fruit growth and pasture development in horticulture and agriculture. Sex pheromone reception to facilitate mating is another important role of olfaction in insects (Butler, 1970). A greater understanding of the molecular mechanisms of olfaction is therefore important for the development of behavioural modification strategies for insect pests, such as species specific–insect control.

1.1 Morphology of the insect olfactory system

Early behavioural experiments identified the antennae as the primary olfactory organs of insects (von Frisch, 1922), a finding that was confirmed with the development of the electroantennogram (EAG) by Schneider (1957). Using this strategy, Schneider was able to show that there was an increase in the electrical activity of excised antennae from a male silk moth (*Bombyx mori*) when the antenna was exposed to an extract of the “sexual attracting substance” of female silk moths. This attractant was later identified as \((E,Z)-10,12\)-hexadecadien-1-ol and christened “bombykol” by its discoverers (Butenandt *et al.*, 1959), representing the first insect sex pheromone to be characterised.

The antennae of most insects are comprised of three main parts (Figure 1.1). The most basal segment is known as the scapus, which is attached to the head by the elastic articular membrane covering a ball and socket joint. The pedicellus is the second segment and is attached to the scapus by two muscles and an elastic membrane. The long flagellum is the third part of the antenna and is divided into multiple annuli, which contain most of the sensilla (Schneider, 1964; Keil, 1999). The sensilla are the morphological units of the antennae responsible for the olfactory response. They are specialised hair–like structures that extend from the surface of the antennae and contain one or more olfactory receptor neurons (ORNs) bathed in sensillum lymph, supported by auxiliary cells (Figure 1.2). These auxiliary cells have a developmental role in the formation of the sensilla during embryogenesis, and later support the maintenance of the lymph, including the secretion of odorant binding proteins (OBPs) and odorant degrading enzymes (ODEs) (Keil, 1999).
1.1 Morphology of the insect olfactory system

Figure 1.1: Ultrastucture of a *Periplaneta americana* antenna showing the scapus (S), pedicellus (P) and the beginning of the flagellum (F). Also shown are the ocellus (O) and articular membrane (AM). Figure from Okada and Toh (2000).

Figure 1.2: A) Scanning electron micrograph of a male *Epiphyas postvittana* antenna showing the long hair–like sensilla trichodea. Scale bar = 100 µM. Image from Jordan (2006). B) Generalised schematic of an insect olfactory sensillum. The cuticle of the sensillum contains many pores which allow contact between the sensillum and the external environment. Olfactory receptor neurons (ORNs) are bipolar cells, with dendrites extending into the sensillum and axonal projections to the central nervous system. The ORNs are associated with auxiliary cells which aid in maintenance of the sensillum. The dendrite is bathed in the sensillum lymph, which contains soluble proteins such as odorant binding proteins (OBPs), pheromone binding proteins (PBPs) and odorant degrading enzymes (ODEs). The odorant receptor is expressed by the ORN and is localised to the dendrite.
Sensilla can be loosely classified into six groups based on their external characteristics, first as seen using light microscopy (Schneider, 1964), and later as seen using electron microscopy (Altner & Prillinger, 1980). There are six general functional classes of sensilla, the sensilla basiconica, sensilla trichodea, sensilla auricillica, sensilla coeloconica, sensilla chaetica and sensilla styloconica. The first four of these groups have been implicated in olfaction. The sensilla basiconica are short and have a round, blunt tip and are innervated by the branched dendrites of as many as 50 neurons (Schneider, 1964). The cuticular wall of basiconic sensilla is relatively thin and contains numerous pores (Keil, 1999). These sensilla are responsible for the detection of plant volatiles in *Spodoptera littoralis* (Anderson *et al.*, 1995) and *Cactoblastis cactorum* (Pophof *et al.*, 2005) as determined using electrophysiological measurements.

Sensilla trichodea are long and thin, with a sharply pointed tip (Schneider, 1964), and are innervated by one to four neurons with unbranched dendrites (Keil, 1999). Originally thought to be involved in mechanoreception, trichoid sensilla have since been implicated in the detection of pheromones by male insects (Almaas & Mustaparata, 1990; Hansson *et al.*, 1995). This has now been proven using single sensilla EAG recording (Shields & Hildebrand, 2000). It is therefore of no surprise that the number of trichoid sensilla is sexually dimorphic, with female insects having reduced numbers, and in some species they are absent entirely (Steinbrecht, 1999).

The sensilla auricillica are shaped like a shoehorn, and have a porous cuticle innervated by three neurons. They have been demonstrated to have a role in olfaction, responding to both plant volatile compounds (Anderson *et al.*, 2000), and in the case of *Cydia pomonella*, to sex pheromone compounds as well (Ebbinghaus *et al.*, 1998; Ansebo *et al.*, 2005).

Sensilla coeloconica are very short, double-walled sensilla that are located in pits on the antennal cuticle. They do not have pores, but the fusion of the two walls is incomplete, leaving openings through which odorants can traverse. Approximately five neurons innervate this sensilla class, and they are sensitive to plant aldehydes and aliphatic acids in *C. cactorum* (Pophof, 1997; Pophof *et al.*, 2005).
Perireceptor events

The remaining two sensilla groups have no known role in olfaction. The first, the sensilla chaetica, are modified trichoid sensilla that have a thicker cuticle and a flexible circular membrane at the base (Schneider, 1964). Sensilla chaetica are mechanoreceptors, but in some organisms they have a single pore at the tip and are believed to recognise non-volatile compounds on contact (Keil, 1999). The final group of sensilla is the sensilla styloconica, which are short and peg-like. While innervated, these sensilla do not have any pores and it is therefore unlikely they have an olfactory role, but there is some evidence that they act as hygro- and thermo-receptors (Shields & Hildebrand, 2001).

Regardless of the large morphological differences in sensilla, one key distinguishing feature of those that have an olfactory function is the presence of many pores in the cuticle wall, responsible for the access of volatile compounds to the sensilla with little water loss to the antennae (Schneider, 1964). Two distinct types of pore structure have been identified (Steinbrecht, 1969). In the pore-tubule systems of single-walled sensilla (sensilla basiconica, trichodia and auricillica), the odorant enters the pore by diffusion where it is then transferred to the lymph. The spoke-channel system as used by double walled sensilla (sensilla coeloconica) utilises lymph-filled channels through both cuticular walls to allow diffusion of an odorant into the sensillum core.

1.2 Perireceptor events

Once an odorant molecule has diffused through a sensilla pore, it enters the sensillum lymph, where it is able to come into contact with the dendritic membrane of an olfactory receptor neuron (ORN), before being removed or broken down in order to reset the system. These events have been termed perireceptor events (Getchell et al., 1984). The sensillum lymph bathes the dendrites of the ORNs, and acts to protect the dendrites and provide the necessary ionic conditions for neuron depolarisation (Keil & Steinbrecht, 1984). The sensillum lymph is also highly proteinaceous, containing soluble proteins that can reversibly bind to odorants, or degrade them (Vogt & Riddiford, 1981; Prestwich et al., 1989). The first OBP was identified in *Antheraea polyphemus* through exposure of
an antennal homogenate to a radioactive pheromone analog (Vogt & Riddiford, 1981). A 15 kDa protein was identified and called pheromone binding protein (PBP). Subsequently, OBPs have been characterised in a range of different insect species (see Vogt (2003) for a review).

The sex pheromone of *A. polyphemus* is a blend of \((E,Z)\)-6,11-hexadecadienyl acetate and \((E,Z)\)-6,11-hexadecadienal (Kochansky *et al.*, 1975), both highly hydrophobic molecules. It has therefore been suggested that the PBP binds the otherwise insoluble pheromone in order to concentrate and transport it to or from the membrane of the ORNs (Pelosi & Maida, 1995), in a role that would also protect the pheromone from degradation by other enzymes present in the sensillum lymph. A molecular mechanism for the binding and transport of odorants has been suggested from evidence provided by the structures of *B. mori* PBP. An NMR structure of the PBP at pH 6.5 (Lee *et al.*, 2002) was similar to that identified using X-ray crystallography of PBP with the *B. mori* pheromone bombykol present (Sandler *et al.*, 2000). However, the structure of PBP at pH 4.5 showed a conformational change that resulted in a new helix forming in the pheromone binding pocket (Horst *et al.*, 2001), which may suggest why bombykol is unable to bind to PBP at low pH (Wojtasek & Leal, 1999). Thus, at the physiological pH of the sensillum lymph (pH 6.5), the PBP is able to bind bombykol. As the PBP–bombykol complex nears the ORN membrane, the physiological pH is thought to decrease to approximately 4.5, inducing a conformational shift in the PBP that ejects the pheromone (Leal, 2003). In contrast to this hypothesis, the binding of compounds found in *A. mellifera* queen pheromone to the PBP ASP1 was found to be favoured at low pH (Pesenti *et al.*, 2008). These differences have yet to be resolved.

Differential expression of specific OBPs also controls odour discrimination in the fruit fly species *D. sechellia* and *D. mauritiana* (Matsuo *et al.*, 2007). *D. sechellia* expresses the *OBP57d* gene to a very high level and is attracted to hexanoic and octanoic acids (HA and OA), whereas *D. mauritiana* does not express it at all and avoids these compounds. *OBP57d* knockouts of *D. sechellia* show the same avoidance of HA and OA as *D. mauritiana*, indicating that OBP57d acts not as an adaptor molecule like LUSH (Zhou
et al., 2004; Laughlin et al., 2008), but instead targets HA and OA for breakdown or sequesters it. This prevents the interaction of HA and OA with an otherwise active OR, and hence there is no signal to the central nervous system and no avoidance behaviour. OBPs can therefore be critical in deciding not only what an organism can smell, but what it can’t.

Once an odorant molecule has entered a sensillum and activated an ORN, a mechanism must be in place to terminate the signal. ODEs are implicated in the resetting of the olfactory system through two mechanisms: the breakdown of odorants in the sensillum lymph by extracellular ODEs (Tasayco & Prestwich, 1990), and the uptake and breakdown of odorants by accessory cells in the sensillum (Maibeche-Coisne et al., 2002). These enzymes have a diverse range of chemical activities, including aldehyde oxidases, aldehyde dehydrogenases, cytochrome P450s, glutathione–S–transferases, esterases and epoxide hydrolases; activities as diverse as the odorants they are required to break down (Vogt, 2003).

Odorant degrading enzymes are particularly important in pheromone reception, as the high level of specificity in this system means even small changes in chemical structure eliminate the function of the pheromone. Antennal specific aldehyde oxidase (Rybczynski et al., 1990) and esterase (Vogt & Riddiford, 1981) enzymes have been identified in A. polyphemus. This moth species uses a blend of an ester and an aldehyde as its sex pheromone (Kochansky et al., 1975), implicating these enzymes in the breakdown of sex pheromone. A mathematical model of the pheromone detection process of A. polyphemus has predicted that the breakdown of pheromone in the sensillum must be extremely fast to cope with the spatial and temporal requirements of pheromone stream detection and navigation (Kaissling, 2001). Kinetic studies on the antennal specific esterase of this species has demonstrated that it cleaves the pheromone component (E,Z)–6,11–hexadecadienyl acetate with a $K_m$ of 1.1 nM, which is comparable with the predicted required rate (Ishida & Leal, 2005).
1.3 Olfactory system signalling

The interaction of an odorant with the antennae results in the depolarisation of ORNs and the subsequent generation of an action potential, which can be measured by EAGs (Schneider, 1957). The exact nature by which ORNs detect and then respond to odorant compounds has been only recently discovered. Early research showed that chemicals that cause the modification of protein structure by disrupting amine and sulphydryl groups eliminate the olfactory response (Villet, 1974), implicating proteins in the reception of odorant molecules. Further experiments suggested that a second messenger system, possibly using cyclic adenosine monophosphate (cAMP) as a second messenger, was used by ORNs upon activation to elicit a signal (Villet, 1978).

More recently, it has been observed that activation of an olfactory receptor (OR) resulted in generation of inositol triphosphate (IP$_3$) by phospholipase C (PLC), two components of the phosphatidyl inositol secondary messenger system (Boekhoff et al., 1990b). Stimulation of PLC in insect antennae is also GTP–dependent and modulated through guanine nucleotide regulatory proteins (G proteins), as moth antennae show no sensitivity to pheromone in the presence of pertussis toxin, a specific G protein inhibitor (Boekhoff et al., 1990a). A further experiment showed that the IP$_3$ response occurred within 50 ms, indicative of a primary role in olfactory signal transduction (Boekhoff et al., 1993). A slower but more prolonged pulse of cGMP was also detected, and these elevated levels of cGMP were demonstrated to inhibit the generation of IP$_3$, suggesting a role in signal attenuation or adaptation/desensitization.

However, *Drosophila* mutants lacking a PLC gene, and therefore unable to generate IP$_3$, were demonstrated to have normal antennal olfactory activity, but lack any form of olfactory activity in the maxillary palp, a secondary olfactory organ (Riesgo-Escovar et al., 1995). This could suggest that there is a complementary PLC gene that is involved in olfactory signalling in the antennae, or that different signalling pathways are utilised in these two organs. A further piece of evidence about the role of IP$_3$ in the insect olfactory signalling pathway comes from the observation of the olfactory capability of
IP\textsubscript{3} receptor knockout mutants (Deshpande et al., 2000). These insects exhibit normal olfactory responses, but show reduced adaptation to repeated and/or strong odour stimuli. This suggests that IP\textsubscript{3} is not a primary second messenger in olfactory signalling, yet is perhaps an important part of the feedback loop that leads to the system resetting, which is different to the results of Boekhoff et al. (1993). While it was unclear what specific signalling pathways are used by insect pheromone and olfactory receptors, it appeared likely that the response involves G proteins, leading to the suggestion that ORs could be members of the G protein–coupled receptor superfamily.

### 1.4 G Protein–coupled receptors

G protein–coupled receptors (GPCRs) are a large family of integral membrane proteins with a characteristic structure of seven transmembrane (TM) domains connected by alternating intra- and extracellular loops with an extracellular N–terminus (Figure 1.3) (Pierce et al., 2002). They are responsible for the transduction of extracellular messages and interact with an enormous variety of signal types. GPCRs are important receptors for light, neurotransmitters, hormones, calcium ions, peptides, nucleotides and chemokines (Choi et al., 2003).

The conformational change that GPCRs undergo upon activation exposes a G protein binding site on the cytosolic side of the membrane. G proteins are heterotrimeric proteins, consisting of $\alpha$, $\beta$ and $\gamma$ subunits, and have an inherent guanine triphosphate hydrolase activity (Hamm, 2001). The binding of the G protein to an activated GPCR results in the exchange of a GDP molecule for a GTP molecule on the $\alpha$ subunit, causing the dissociation of the $\alpha$ subunit from the $\beta\gamma$ dimer. These two individual units are then able to activate a wide range of secondary messenger signalling cascades, such as the phosphatidyl inositol second messenger system, ultimately resulting in a cellular response. In time, the G protein $\alpha$ subunit hydrolyses the bound GTP to GDP, causing the $\alpha$ and $\beta\gamma$ components to re–associate to form the heterotrimer, restoring the G protein and concluding the activation cycle (Figure 1.4) (Mombaerts, 1999). In vertebrate ORNs, this
response is thought to be the opening of chloride channels, through which chloride ions enter the cell and cause neuron depolarisation (Mombaerts, 2004).

Another critical part of the signal transduction cascade is the inactivation of the receptor and the resetting of the system for further activation. For GPCRs, this process begins with phosphorylation of the activated receptor by G protein receptor kinases (GRKs), which enables the binding of arrestin proteins (Krupnick & Benovic, 1998). These proteins physically prevent further activation of secondary messengers, and also causes the recycling of the receptor by promoting clathrin–mediated endocytosis (Goodman et al., 1996). Indeed, both the GRKs (Boekhoff et al., 1994) and arrestins (Merrill et al., 2004) are required for the correct desensitization of the insect olfactory system. However, their importance appears to vary for different odorants (Merrill et al., 2004), which indicates that this is not the only pathway used in olfactory desensitization.

The classical model of GPCR ligand binding, as illustrated in Figure 1.4, does not fully explain their diverse biological activities. The role of oligomerisation, either with the same receptor type (homodimerisation) or a different receptor type (heterodimerisation)
1.4 G Protein–coupled receptors

Figure 1.4: A model of signal transduction in olfactory systems (modified from Mombaerts (1999)). The top panel indicates the resting state, where the receptor is in the inactive form, the G protein has a bound GDP, adenylyl cyclase and phospholipase C are inactive and the cation and chloride channel proteins are closed. On activation of the receptor by an odorant (lower panel), the G protein exchanges the GDP for a GTP and dissociates. The \( \alpha \) subunit activates adenylyl cyclase, catalysing the conversion of ATP into cyclic AMP (cAMP), and/or phospholipase C, which catalyses the production of inositol trisphosphate (IP\(_3\)) from phosphatidylinositol bisphosphate (PIP\(_2\)). Both of these secondary messengers are able to activate ligand–gated cation channels, and the subsequent inflow of cations opens voltage–dependent chloride channels causing neuron depolarisation.

has been identified as important for their function. Heterodimerisation between the \( \gamma \)–aminobutyric acid receptor type B1 and type B2 is required for either receptor to be trafficked to the cell membrane (Margeta-Mitrovic et al., 2000). These receptors also require heterodimerisation for activity, as the type B1 receptor houses the ligand binding site, and the type B2 receptor contains the G protein binding site (Margeta-Mitrovic et al., 2001). Not only can the localisation of a GPCR be altered by oligomerisation, different dimerisation partners can alter the signalling pathway of the receptor complex. The heterodimerisation of a microreceptor and \( \delta \) opioid receptors changes the response of the receptor from the inhibitory to the excitatory G protein pathway (Charles et al.,
Ligand binding can also alter the dimerisation state of a receptor, such as the thyrotropin receptor. In the inactive state, this receptor forms an oligomer that is unable to bind a G protein. Ligand binding causes the oligomer to dissociate, enabling signal transduction (Latif et al., 2003).

The association of non-receptor proteins also has a large part to play in the function of receptors. The calcitonin receptor–like receptor (CRLR) is trafficked to the cell membrane only when associated with a member of the receptor activity modifying protein family (RAMPs) (McLatchie et al., 1998). In the same study it was demonstrated that the co-expression of RAMP–1 with CRLR caused the receptor to be activated in the presence of calitonin–gene–related peptide (CGRP), whereas co-expression of RAMP–2 caused CRLR to be unresponsive to CGRP, and instead become activated by adrenomedullin. This indicates that the receptor specificity of a ligand is not always entirely due to the specific receptor expressed, it can instead be the presence or absence of non–receptor proteins.

There are accessory proteins also known to be vital in olfaction. The localisation of olfactory receptors (ORs) to the cilia in Caenorhabditis elegans requires the presence of ODR–4, a chaperone with a single TM domain (Dwyer et al., 1998). Three proteins have been found that enhance the expression, targeting and ligand response of mouse ORs in vitro, though their specific role in vivo is unknown (Saito et al., 2004). A further protein, Hsc70t, a heat shock protein known to have a role in protein folding and specifically expressed in mouse olfactory tissue, enhances the expression of mouse ORs in vitro (Neuhaus et al., 2006). It is clear that the correct folding, targeting and ligand binding of ORs utilises a number of accessory proteins, but to date only the OR, Or83b, has been identified as having a chaperone–like activity in insects (Benton et al., 2006). No non–OR proteins, have been demonstrated to have a similar role in insects. It is interesting to note that the accessory proteins involved in all of these different systems show no sequence similarity (Saito et al., 2004), which hinders the identification of insect analogues.
1.5 Olfactory receptor genes

In light of the biochemical evidence that ORs were likely to be GPCRs, it was assumed that there would be a large family of ORs, they would have the characteristic seven transmembrane domain structure and would be localised to olfactory tissues. ORs were identified in the rat olfactory epithelium based on these assumptions in Nobel Prize winning research (Buck & Axel, 1991). ORs have subsequently been identified in a number of other vertebrate species, including fish (Ngai et al., 1993), birds (Nef et al., 1996), humans (Glusman et al., 2001; Zozulya et al., 2001), and mice (Zhang & Firestein, 2002). In these species, the OR gene families showed a high degree of conservation (Mombaerts, 1999), which aided their isolation using homology-based techniques.

The development of genome sequencing projects facilitated the identification of OR genes in other species. Over forty OR genes were identified from only 15% of the genome of the nematode worm *C. elegans*, based on the predicted presence of seven transmembrane domains, and observed to be expressed in the sensory neurons (Troemel et al., 1995). These receptors were highly divergent within *C. elegans*, and also showed very little similarity to those ORs identified in vertebrates. As many as 1300 functional OR genes have now been described in the worm (Robertson & Thomas, 2006).

Putative OR genes were identified from the early releases of *D. melanogaster* genome data largely based on the assumption that they would be a large gene family and contain more than one transmembrane domain (Clyne et al., 1999; Gao & Chess, 1999; Vosshall et al., 1999). A full set of 60 OR genes encoding 62 OR proteins was subsequently identified (Robertson et al., 2003) and named from the completed *Drosophila* genome database (Committee, 2000). *Drosophila* ORs show very little sequence similarity, ranging from 17% to 26% within this species (Vosshall et al., 2000). OR genes have subsequently been identified in further insect species, including *Anopheles gambiae* (Fox et al., 2001), *Heliothis virescens* (Krieger et al., 2002), and *B. mori* (Sakurai et al., 2004; Wanner et al., 2007a).

While homology-based OR identification strategies have generally proved unsuccessful
due to the low degree of sequence similarity within this gene family, there is one receptor for which this is an exception. The *D. melanogaster* Or83b receptor has homologues that share approximately 70-80% amino acid identity in a range of insect species including *H. virescens* (Krieger et al., 2002), *Apis mellifera*, *Tenebrio molitor*, *Calliphora erythrocephala*, *B. mori*, *A. pernyi* (Krieger et al., 2003), *Aedes aegypti* (Melo et al., 2004), *A. gambiae* (Pitts et al., 2004) and *Epiphyas postvittana* (Jordan et al., 2008). While the majority of ORNs in *Drosophila* express only one OR (Vosshall et al., 1999), Or83b is co-expressed in most ORNs and is essential for olfaction *in vivo* (Larsson et al., 2004). Olfactory ability can be rescued in Or83b mutant *Drosophila* by expression of Or83b homologues from *Ceratitis capitata*, *A. gambiae* and *H. zea*, indicating that the functionality of this receptor has been conserved through at least 250 million years of evolution (Jones et al., 2005).

Initial insight into the function of Or83b came with the observation that conventional ORs expressed in ORNs not expressing Or83b were not targeted to the dendritic membrane, and instead accumulated in the cell body (Larsson et al., 2004). It had previously been suggested that Or83b acts as a dimerisation partner for conventional ORs (Krieger et al., 2003; Vosshall et al., 2000). Evidence for dimerisation was observed using bioluminescence resonance energy transfer (BRET) to show that a conventional ligand–binding OR, Or22a, was associated with Or83b when heterologously expressed, and that this co-expression increased the sensitivity of Or22a by a thousand fold (Neuhaus et al., 2005). A similar increase in sensitivity was seen in a heterologous assay on *B. mori* pheromone receptors when co–expressed with the *B. mori* Or83b orthologue (Nakagawa et al., 2005). The heterodimerisation of an OR with Or83b occurs quickly after the translation and folding of ORs in the ORNs, and Or83b is responsible for trafficking the ligand–binding ORs to the dendritic membrane where they perform their sensory function (Benton et al., 2006).

As discussed above, conventional GPCRs, including known vertebrate and nematode ORs, have a characteristic seven transmembrane domain topology with an extracellular N–terminus and an intracellular C–terminus (Figure 1.5). However, there is data that suggests that insects ORs have the same number of transmembrane domains, but that the
orientation in the cell membrane is opposite to that of a conventional seven transmembrane
domain protein in that the N–terminus is intracellular (Benton et al., 2006). Glycosylation
scanning has been used to independently verify this observation, and to confirm the
presence of seven TM domains in Or83b (Lundin et al., 2007). This topological model
would put the most conserved parts of the OR in the cytosol (Clyne et al., 1999), and
it is one of these domains of the odorant sensitive ORs, the third intracellular loop, that
interact with Or83b (Benton et al., 2006). The authors suggest that because of their
unique structure, insect ORs may represent a family of proteins that have independently
evolved to couple with G proteins, or that they may use a distinct insect specific signalling
pathway. It is possible that Or83b itself makes up part of this novel transduction system,
as it has been observed that ORNs expressing Or83b exhibit spontaneous electrical
activity, whereas Or83b mutants do not (Dobritsa et al., 2003; Larsson et al., 2004).

Very recent research has shed further light onto the role of Or83b in insect olfactory signal
transduction (Sato et al., 2008; Wicher et al., 2008). The study of Wicher et al. (2008)
suggested that Or83b alone is able to form an ion channel, but co–expression of a non–
Or83b subunit is required to confer ligand sensitivity. Wicher et al. (2008) also provide
evidence that OR signalling in HEK293 cells occurs through two distinct pathways, a rapid
onset direct activation of the ion channel and a slower but more sensitive and sustained
activation mediated by a $G_{\alpha s}$ protein and cAMP. This contrasts with another recent
study that observed reduced sensitivity in Drosophila antennal sensory neurons in which
a $G_{\alpha q}$ protein had been knocked out, implicating phospholipase C and IP$_3$ as signal
transduction elements (Kain et al., 2008), supporting earlier studies (Boekhoff et al.,
1990a; Boekhoff et al., 1990b). On the other hand, neither Sato et al. (2008) nor Smart
et al. (2008) observed evidence for metabotropic signalling, either through $G_{\alpha q}$ or $G_{\alpha s}$
activated pathways. Both studies were able to demonstrate that insect ORs are able to
signal in the presence of common inhibitors of G protein signal transduction enzymes.
Experiments to reconcile these differences will aid in the understanding of this important
aspect of insect olfaction.
1.6 Odour coding in the *Drosophila* antennae

Each antenna of *Drosophila* contains approximately 1200 ORNs (Stocker, 1994). The spatial pattern of OR expression in these ORNs is similar in each fly, and can be mapped to individual neuronal subtypes in sensilla using electrophysiological measurement of the responses to a panel of distinct odorants (de Bruyne et al., 2001). This research identified that there is overlap in the morphological classification and sensory function of sensilla. Further research enabled the assignment of an OR, Or22a, to the neuronal subtype ab3A (Dobritsa et al., 2003). A response profile of the OR was made using electrophysiological measurements to a range of compounds which was then able to be matched to the known response of specific ORNs. Individual mis-expression of the full complement of *Drosophila* ORs in an ab3A neuron in which the Or22a gene had been deleted enabled the matching of many more ORs to their neuronal subtype (Figure 1.6) (Hallem & Carlson, 2004).

Functional expression of the calcium-sensitive fluorescent protein Cameleon in specific neuronal cells is another technique that has been used to analyse the odorant binding of ORs *in vivo* (Pelz et al., 2006). This imaging approach allowed the response of Or22a to over 100 odorants at various concentrations to be assessed at both the ORN and glomerular level, and demonstrated that the activity of this receptor in the glomerulus matched its activity in the ORN, and showed that odour coding occurs at the level of the antennal lobe.

The functional analysis of olfactory receptors has greatly increased the understanding of the insect odorant coding system. It was observed that insect ORs vary markedly in their
tuning breadth (Hallem & Carlson, 2004), with some ORs responding to many compounds and others to very few (Figure 1.6). Some compounds also activate many different ORs with differing strengths. This functional overlap expands the capacity of what odours a small set of ORs can detect. At higher concentrations, the same odorants elicited responses from a larger number of ORs, and suggests that a form of intensity coding is carried out at the OR level (Hallem & Carlson, 2006). The temporal dynamics of OR activity also varies for certain odorants at different concentrations, and that in some cases the response occurs much longer than the odorant stimulus (Hallem & Carlson, 2006). Thus, the temporal activity of ORs can also help to encode the duration and intensity of an odour pulse, an essential piece of information for navigation in the environment.

The specific OR that is expressed in each ORN is the sole mediator of a range of important neuronal characteristics, including the spontaneous firing rate, whether the neuron showed an excitatory or an inhibitory response to an odorant and the rate at which the signal is terminated (Hallem & Carlson, 2004). Furthermore, the observation that odorants can act as inhibitors of some ORs while acting as activators of others increases the complexity of the odour code, particularly in the presence of mixtures of compounds (Oka et al., 2004).

Figure 1.6: Response of *D. melanogaster* ORs. Colored dots indicate strong responses (as defined by a rate of >100 spikes/s following stimulation with a $10^{-2}$ dilution of odorant). Reproduced from Hallem and Carlson (2004).

Once activated, the ORNs transmit their signal through their axons into the central nervous system of the fly. The axons of the ORNs terminate in the glomeruli of the antennal lobe, where they connect with the projection neurons of the CNS (Figure 1.7) (Carlson & Hansson, 2003). The initial observation that the number of glomeruli was
similar to that of the number of ORN subtypes in the antennae led to the prediction that each ORN subtype innervated a single glomerulus in the antennal lobe (Vosshall, 2001). The development of a complete map of the interaction between ORNs and glomeruli has confirmed this prediction, showing that the axons of each ORN subtype converge on a spatially distinct, bilaterally symmetrical glomerulus (Vosshall, 2001; Couto et al., 2005; Fishilevich & Vosshall, 2005). The axons of spatially grouped ORN subtypes project to similar spatial areas in the antennal lobe, a pattern that is correlated to the sensilla subtypes of the antennae, indicating some level of higher organisation (Vosshall, 2001; Couto et al., 2005; Fishilevich & Vosshall, 2005).

![Figure 1.7: Three dimensional representation of the Heliothis armigera brain. Axons of the olfactory sensory neurons enter the antennal lobe (AL) and innervate the spherical glomeruli that can be seen within. Also shown are the antennal mechanosensory center (AMMC), the subesophageal ganglion (SOG) and the optic lobe (OL). The scale bar represents 200 µm. Reproduced from Skiri et al. (2005).](image)

The observation that each glomerulus represents the activity of one OR suggests that insects use a combinatorial odour code, as has been proposed for other systems (Malnic et al., 1999). A particular odorant elicits different excitatory and inhibitory responses in several receptors, which results in differential firing rates of their ORNs. These neurons then activate the glomeruli which they innervate in the antennal lobe, generating an odour specific pattern of activated glomeruli. Projection neurons transmit these signals to the higher processing centres of the insect brain, the mushroom body and the lateral horn, where the signal is interpreted and a behavioural response is formulated (Figure 1.8) (Hallem & Carlson, 2004).
1.7 Aims of this project

The recent findings that the non–canonical receptor Or83b has seven TMs but with a GPCR–atypical topology (Benton et al., 2006; Lundin et al., 2007) as well as its ability to form an ion channel (Sato et al., 2008; Wicher et al., 2008), has challenged many basic assumptions about the structure and function of insect olfactory receptors, including the secondary messenger signalling pathways that they utilise, their monomeric and higher order structure, and their mechanism of activation. The research presented in this thesis aims to investigate the structural features of the model odorant–activated OR, Or22a from *D. melanogaster*, including its orientation in the cell membrane, the number of TM domains present and its ability to form higher order structures in the cell membrane. Or22a has been chosen as the model OR for these experiments, as its structure and function has been widely studied.

In Chapter 2, an *in vitro* functional assay for insect ORs using calcium imaging will be developed to allow the questions above to be addressed. The Sf9 cell line from *S. frugiperda* will be tested for its ability to express Or22a. The response of Or22a to odorants in this cell line will be compared to data previously obtained *in vivo* in order to determine...
whether ORs function in this system as they do in the insect. The applicability of this system for generating pharmacological data for insect ORs will also be assessed.

In Chapter 3, a set of experiments are performed in order to test whether the orientation of the ligand binding ORs have an intracellular N–terminus like that of Or83b, or whether they have extracellular N–termini reminiscent of classical GPCRs. In silico predictions of the transmembrane topology of Or22a are then tested empirically to examine whether ligand binding ORs contain seven transmembrane regions, as observed for Or83b. Insertion of antibody epitopes into the predicted hydrophilic loop domains followed by localisation of the epitope using immunochemistry will be used to determine the orientation of the loops relative to the cell membrane.

In Chapter 4, the question of whether ligand binding ORs form multimeric structures is investigated through studying the interaction between certain combinations of intracellular loops of Or22a. Fluorescence resonance energy transfer was used to address this question. If Or22a does not form homomeric interactions, it can be suggested that it interacts only with Or83b. If Or22a is observed to form homo–multimers, the number and type of interactions between Or22a subunits will aid in the development of quaternary structure models of insect ORs. For example, one specific interaction might be evidence of a dimeric structure, whereas multiple interactions might suggest a higher order structure.

Finally, in Chapter 5, the general discussion, these data are synthesised and put in context with recent advances in the field of insect olfaction. Three models of how a functional OR complex are presented and a number of future experiments proposed that will enable these models to be distinguished.
2

Development of an In Vitro Olfactory Receptor Assay

2.1 Introduction

Olfactory receptors are just one class of proteins that are responsible for the transduction of extracellular signals, enabling a cell or organism to respond to its environment. Types of receptor include G protein–coupled receptors (GPCRs), receptor tyrosine kinases (RTKs) and ligand–gated ion channels (Helmreich, 2001). The GPCR superfamily of proteins is responsible for the transduction of a range of signals that include hormones, neurotransmitters, light, tastants, and chemokines (Choi et al., 2003). Early identification of these receptors began by probing tissue libraries with a specific ligand of interest and looking for activity in so called “ligand fishing” experiments (Yanasigawa et al.,
1988). However, with the advent of widespread genome sequencing, a large number of receptors have been identified for which the ligand is unknown. For example, there are over 350 GPCRs with no known ligand that have been identified from the completed human genome (Vassilatis et al., 2003). Because of this, a number of different assays have been developed to assign ligands to these so called “orphan” receptors.

2.1.1 Receptor functional assays

Assays to determine ligands for orphan receptors fall into three main categories: ligand binding assays, protein redistribution assays and second-messenger based assays (for a review, see Eglen (2005)). Ligand binding assays involve the exposure of receptors to radioactively-labelled ligand and the determination of free and bound ligand. Interacting ligands that bind tightly to the receptor are not removed by washing, and therefore the level of reduction in radioactivity indicates the affinity of the receptor for that particular ligand. These assays were used to identify the B$_{2a}$ bradykinin and V$_{1a}$ vasopressin receptors (Howl et al., 1997). A similar technique is used to monitor the activation of GPCRs by measuring the agonist-dependent binding of radioactive non-hydrolysable GTP analogs to G proteins. These analogs are unable to be cleaved and are therefore depleted from the assay solution when the GPCR is activated (White et al., 2007).

Another property that can be utilised to monitor receptor activation is the redistribution of either the receptor itself, or an accessory protein, in response to a ligand. The internalisation of receptors is a mechanism for desensitising the cell to the stimulus and halting signalling, and is carried out through the recruitment of a number of accessory proteins, such as the β-arrestins (Ferguson et al., 1996). By tracking changes in the distribution of fluorescently-tagged receptor or accessory protein upon addition of a potential ligand, receptor activators can be identified. The human parathyroid hormone receptor was identified using this technique (Conway et al., 2001). Alternatively, when fluorescently-tagged arrestin binds to a fluorescently-tagged receptor, the emitted wavelength of light changes due to the proximity of the fluorophores in a phenomenon
known as fluorescence or bioluminescence resonance energy transfer (FRET and BRET, respectively). In this way the interaction of an arrestin with a receptor can be directly measured, and the strength of the ligand–receptor interaction deduced (Bertrand et al., 2002).

In addition to direct assays of receptor–ligand interaction, a number of other events occur in order to propagate the signal. Activation of a receptor leads to the up–regulation of a number of downstream enzymes that cause an increase in the concentration of second–messenger molecules. These second–messengers, including calcium ions, cyclic adenosine monophosphate (cAMP), and inositol tri–phosphate (IP$_3$), act on a number of targets, such as other enzymes or transcriptional regulation elements, in order to elicit a cellular response (Helmreich, 2001). Changes in the concentration of these molecules can be monitored to identify receptor ligands. Other signalling pathways, such as those that utilise receptor tyrosine kinases, modulate their second messengers through phosphorylation. Measuring the phosphorylation state of certain proteins can be used to give an indication of receptor activity (Rotin et al., 1992). Measuring these second–messenger events has the advantage over direct radioligand binding assays in that the ligand must activate the receptor, rather than just bind to it (Knight et al., 2003).

Measurement of the concentration of the second–messenger cAMP using enzyme immunoassays has been used to deorphant a number of receptor proteins, including the identification of the Gpr4 receptor from the human pathogen Cryptococcus neoformans as an amino acid–sensing protein (Xue et al., 2006). Radiographic cAMP assays have similarly been applied to high–throughput screens of GPCRs (Kariv et al., 1999). IP$_3$ accumulation has also been used to determine ligand–receptor interactions, and have been used to identify the human Gpr75 receptor as a RANTES receptor, involved in chemokine reception and the inflammatory response (Ignatov et al., 2006).

Aequorin, a protein from the jellyfish Aequoria victoria that luminesces in the presence of calcium ions (Shinomura et al., 1962), has been used to measure the activation of a number of receptors for which calcium is a second messenger. Dupriez et al. (2002) used aequorin to assay the human somatostatin 2 receptor (a GPCR), the vanilloid receptor 1
(an ion channel) and the epidermal growth factor receptor (an RTK). A similar strategy was used to identify the human prostanoid receptors (Ungrin et al., 1999). These receptors all signal using calcium ions in vivo, but intracellular calcium measurements for GPCRs that do not utilise calcium as a second–messenger, such as the serotonin receptor 1A, can be performed in cell lines over–expressing non–native promiscuous G proteins that couple to this pathway (Knight & Grigliatti, 2004).

Calcium–sensitive fluorescent dyes, such as Fluo–4 and Fura2, have also been used to measure GPCR–mediated mobilisation of calcium in response to ligands. For example, the role of $\alpha_{2a}$–adrenoreceptors in nitric oxide synthesis in endothelial cells was elucidated by measuring calcium generation in response to arginine (Toshi et al., 2007). These approaches are collectively termed “calcium imaging”. Calcium imaging has also been utilised in sensory receptor biology. It has been used to measure rat ORN responses to odorants (Gautam et al., 2006). Experiments on the Drosophila gustatory receptor Gr5a expressed in S2 cells showed that this receptor is narrowly tuned to the carbohydrate trehalose (Chyb et al., 2003). Gr5a was expressed in S2 cells using an inducible promoter, and the cells exposed to trehalose or other similar saccharides. Fura2 was used to demonstrate that trehalose alone stimulated a calcium influx in Gr5a expressing cells as measured using a ratiometric analysis of the cell’s fluorescence.

### 2.1.2 Insect cell expression systems

The use of insect cell systems for the expression and functional assay of recombinant receptor proteins is well established (see Massotte (2003) for a review). Common insect cell lines used are the Sf9 cell line from the fall army worm, Spodoptera frugiperda, (Vaughn et al., 1977) and S2 cells from D. melanogaster (Schneider, 1972). These cell lines are relatively straigt–forward to culture as they can be grown in serum–free conditions and do not require controlled atmosphere incubators (Schlaeger, 1996). They also post–translationally modify proteins in a similar manner to mammalian cell systems (Altmann et al., 1999), making them a better system for the study of the specific function
of eukaryotic proteins as opposed to bacterial systems.

Infection of SF9 cells with recombinant baculovirus allows high levels of expression of a gene of interest under the control of “late” promoters (Altmann et al., 1999). Expression of these genes occurs near the end of the viral life cycle when the cells are about to lyse, at which stage processing and targeting of the protein may not be optimal. Despite these drawbacks, baculovirus expression has been used to study the function of human olfactory receptors (Matarazzo et al., 2005). Expression systems using immediate–early promoters in plasmids have also been developed to allow earlier expression of a gene of interest in a non–lytic system, and making it possible to generate stably–expressing cell lines (McCarroll & King, 1997). Such a system has yet to be used for the expression of insect ORs.

2.1.3 The use of heterologous expression systems to study olfactory system signalling

Vertebrate olfactory receptors are rhodopsin–like Class A GPCRs, and their signal transduction pathway is well understood. Interaction with an olfactory G protein–coupled receptor leads to the activation of a G protein (Malnic et al., 1999). This G protein subsequently up–regulates adenyl cyclase activity, causing an increase in cAMP concentration and the opening of cAMP–gated sodium and calcium channels (Ronnet et al., 1993). An influx of these ions causes neuronal depolarisation and subsequent signalling to the higher processing centers of the organism (Firestein & Werblin, 1989). Much of this research was carried out in painstakingly isolated primary olfactory neurons (Firestein & Werblin, 1989; Ronnet et al., 1993), but it has since provided the basis for a number of different assays in heterologous systems using techniques explained above. For example, Levasseur et al. (2003) were able to determine that human and rat OR–I7 detect medium chain aldehydes when expressed in Xenopus oocytes, HEK293 and COS–7 cells using calcium imaging. Similar assays for mammalian ORs have been developed in yeast (Minic et al., 2005) and SF9 cells (Knight et al., 2003; Matarazzo et al., 2005).
Because of the dearth of knowledge regarding insect olfactory signalling, much of our understanding of the insect olfactory system comes from \textit{in vivo} electrophysiology experiments \cite{deBruyne2001, Dobritsa2003, Hallem2004, Hallem2006}. These \textit{in vivo} electrophysiology assays measure electrical output, and the exact signal transduction pathway that the insect olfactory receptors use does not need to be understood. There is some biochemical evidence for the involvement of G proteins and another common second messenger enzyme, phospholipase C (PLC), the enzyme that generates IP$_3$. PLC levels increased in an antennal extract of cockroaches exposed to its pheromone periplanone B, an effect that was inhibited by the addition of pertussis toxin, a G protein inhibitor \cite{Boekhoff1990a, Boekhoff1990b}.

Because these pathways are not well understood, there are very few functional assays that have been developed for heterologously–expressed insect ORs. One such assay utilises voltage–clamped \textit{Xenopus} oocytes expressing \textit{D. melanogaster} Or43b and a promiscuous G$_\alpha$ protein \cite{Wetzel2001}. Experiments using this system identified cyclohexanone, cyclohexanol, benzaldehyde, and benzyl alcohol as agonists at nanomolar concentrations of Or43b, findings that were later corroborated by \textit{in vivo} electrophysiology experiments \cite{Hallem2004}. Interestingly, in this assay system Or43b coupled to the phospholipase C pathway through G proteins. However, this may not reflect what occurs \textit{in vivo}. \textit{Xenopus} oocytes have also been used to express and assay \textit{B. mori} Or1 and Or3 in the presence and absence of the Or83b orthologue BmOr2 \cite{Nakagawa2005} and to identify a receptor for the honey bee queen pheromone component 9–oxo–2–decenoic acid \cite{Wanner2007b}. The human HEK293 cell line is another heterologous system that has been used to functionally express insect ORs. Response of \textit{D. melanogaster} Or43a in HEK293 cells was detected using the Fura2 calcium–sensitive dye at substrate concentrations in the millimolar range, though this low sensitivity was increased to micromolar concentrations by the co–expression of Or83b \cite{Neuhaus2005}. Expression in HEK293 cells was used to confirm the identity of putative pheromone receptors from \textit{H. virescens} \cite{Grosse-Wilde2007}. The human HeLa cell line has also been used to assess the function of ORs from both \textit{A. gambiae} and \textit{D. melanogaster} using both electrophysiological measurements and calcium imaging \cite{Sato2008}.
2.1.4 Aims

Despite the impact of the results generated by *in vivo* electrophysiology experiments, it is not a practical approach for the decoding of ORs from all insect species, due to a lack of both the requisite genetic tools and the understanding of their antennal physiology. To this end, the research presented in this chapter aims to develop a generic insect olfactory receptor assay, and to examine whether ORs expressed in this heterologous system function as they do *in vivo*. Or22a from *D. melanogaster* will be used as a model OR to address questions of whether the Sf9 system expresses insect ORs sufficiently to detect a response, whether OR expression alone is sufficient for activity and if Or22a responds *in vitro* in a dose responsive manner and can therefore be used to collect pharmacological data on these receptors. Sf9 cells have been chosen for these experiments due to their ease of culture, their high level of adherence to the cell culture surface and the availability of both plasmid and baculovirus–based expression systems. The cell line is derived from a moth species, and will hopefully provide the appropriate conditions for the efficient expression and targeting of ORs from a range of species. They are also more adherent than the *Drosophila* S2 cell line, an important attribute for an imaging–based assay.

2.2 Materials and Methods

2.2.1 Materials

Sf9 and S2 cells were purchased from Invitrogen. All odorants were purchased from Sigma–Aldrich at the highest purity available. Stock solutions were made to a concentration of $10^{-2}$ M and stored at $-20$ °C. Odorant stock solutions were made up in water, except for odorants with low solubility in water as reported in the Merck Index, which were made in DMSO.
For each assay, odorant solutions were made fresh from the stock solution to the desired concentration in sterile assay buffer. Fluo–4 acetoxymethyl ester (excitation wavelength maxima = 494 nm, emission wavelength maxima = 516 nm) was obtained from Invitrogen as a lyophilized powder, made to 1 mM in DMSO and stored at $-20^\circ$C. All restriction enzymes and other cloning materials were from New England Biolabs, unless otherwise noted. Six and twelve well culture plates were from Nunc, and cell culture flasks were from Corning.

Assay buffer contained 21 mM KCl, 12 mM NaCl, 18 mM MgCl$_2$, 3 mM CaCl$_2$, 170 mM $D$–glucose, 1 mM probenecid (Sigma–Aldrich) and 10 mM PIPES, pH 7.2. The buffer was sterilised using 0.22 µM SteriCup–GP filter units (Millipore) and stored in the dark at 4 $^\circ$C prior to use.

### 2.2.2 Expression vector construction

The plasmid pIB-Or22a was constructed by inserting the cDNA for *D. melanogaster Or22a* (Genbank accession CG12193) into the multiple cloning site of the pIB/V5-His vector (Invitrogen) using the restriction enzymes *Kpn* I and *Sac* II. The plasmid pIB-Or83b was constructed by inserting the cDNA for *D. melanogaster Or83b* (Genbank accession CG10609) into the multiple cloning site of the pIB/V5-His vector using the restriction enzymes *Kpn* I and *Sac* II. cDNA clones of these genes were kindly provided by Dr Coral Warr (Monash University, Melbourne, Australia).

To create a double construct of Or22a and dsRed fluorescent protein (Genbank accession ABY85250; a gift from Dr Peter Murphy, HortResearch, New Zealand) under the control of separate promoters, the A1P promoter (from the pA1PC-$\beta$Gal plasmid) was inserted into the multiple cloning site of pEGFP-1 using *Bgl* II and *Xho* I. The EGFP gene was replaced by the dsRed gene using *Bam*H I and *Not* I. pIB-Or22a was digested with *BspH* I, end-filled using *Taq* polymerase (Invitrogen) and dephosphorylated using shrimp alkaline phosphorylase (Roche Diagnostics). The entire fragment A1P-dsRed-SV40-polyA-tail was cut using *Afe* I and *Afl* II, end-filled and ligated into the linearised pIB-Or22a.
2.2 Materials and Methods

A version of aequorin (Genbank accession AY601106) was synthesised with codons optimised for *Drosophila* expression and surrounded by *Bam*H I and *Not* I restriction sites. pIB-Or22a-dsRed was digested with *Not* I to free dsRed, ligated and transformed to re-circularise. This plasmid was then cut with *Bam*H I, dephosphorylated and the aequorin fragment ligated to create pIB-Or22a-Aequorin. The correct orientation of the aequorin gene was determined using restriction digestion with *Sac* II.

To create pIB-Or22a-Or83b, *Not* I sites were added onto Or83b using PCR and subcloned into the pGem-T Easy vector (Promega). pIB-Or22a-aequorin was first cut with *Stu* I to remove the aequorin fragment, ligated and transformed. This re–circularised plasmid was then cut with *Not* I, purified and ligated with the *Not* I–cut Or83b fragment from pGem-T-Or83b. Restriction digestion with *Bam*H I was used to determine the correct orientation of the Or83b fragment. All expression clones were verified by sequencing prior to use in experiments. Vector maps for the expression plasmids can be found in Appendix A.

### 2.2.3 Insect cell culture

*S. frugiperda* Sf9 cells or serum–free adapted *D. melanogaster* S2 cells were maintained either as adherent cultures in T-25 tissue culture flasks or as suspension cultures in 50 mL Erlenmeyer flasks (Nalgene). Sf9 cells were grown in Sf–900 II Serum Free Media (SFM) and S2 cells were grown in *Drosophila* media in the dark at 28 °C.

Adherent cultures were grown to approximately 90% confluence as judged under a light microscope. To subculture, the cells were dislodged from the growing surface by washing with the media contained in the flask. A small aliquot was then taken and the cell density measured using a hemocytometer. 1 x 10⁵ cells were then seeded in 6 mL of Sf–900 SFM in a new T–25 flask. Cells in suspension culture were grown until the culture was at a density greater than 1 x 10⁷ viable cells/mL, when a new flask was seeded at a density of approximately 1 x 10⁵ cells/mL. Cell viability was estimated using the trypan blue exclusion method (Sambrook & Russel, 2001).
2.2.4 Transfection of insect cell lines

1 x 10^6 cells were added to 3 mL of insect cell media and placed in each well of a Nunclone six well tissue culture plate (Nunc). Twenty four hours later the cells were washed twice and then replaced with 800 µL of fresh media. For each well to be transfected, 1 µg of the required plasmid DNA and 6 µL of transfection reagent were each added to separate 100 µL aliquots of fresh media. The DNA– and transfection reagent–containing solutions were mixed and incubated at room temperature for 15 minutes. The resultant 200 µL of solution was then added to the well containing insect cells, swirled gently and incubated at 28 °C for 7.5 hours. After this time, the DNA and transfection reagent solution was removed, the cells washed twice, replaced with 3 mL of fresh media and incubated at 28 °C until use. Solution components were scaled as appropriate for other well formats.

2.2.5 RT–PCR

One well each of Sf9 cells transfected with pIB-V5/His (empty vector) or pIB-Or22a were washed twice with 3 mL of PBS, then resuspended in 1 mL of fresh PBS and placed into 1.7 mL eppendorf tubes. Total RNA was extracted from each sample using the Trizol method (Invitrogen) following the manufacturer’s instructions, with the resultant RNA pellet resuspended in 5 µL of diethylpyrocarbonate–treated water. First strand cDNA synthesis was carried out using 1 µg of RNA as the template with Superscript III reverse transcriptase (Invitrogen) according to the manufacturer’s instructions.

cDNA from 0.1 µg of RNA was used as the template in a 50 µL PCR that included 2.5 mM dNTPs (Invitrogen), 200 µM of each primer, 1 unit of Platinum Taq polymerase (Invitrogen) and 1x PCR buffer including 1.5 mM magnesium. The thermocycling conditions were: initial denaturation at 94 °C for two minutes, then 30 cycles of 94 °C for 30 seconds, 55 °C for 30 seconds and 72 °C for two minutes, then a final extension at 72 °C for ten minutes. The PCR products were analysed by gel electrophoresis in 1% agarose including 1x SYBR Safe DNA gel stain (Invitrogen) and visualised using an ImageQuant 300 System (GE Healthcare Life Sciences). The primers
2.2 Materials and Methods

used were the pIB plasmid–specific OplE2 Forward (5’–CGCAACGATCTGGTAAACAC–3’) and OplE2 Reverse (5’–GACAATACAAAACATAGATTAGTCAG–3’) primers, and the Or22a–specific primers 22aD (5’–ATCAGCTGGACAAGATGC–3’) and 22aH (5’–GCATTGCGACCCGTCTTTTC–3’).

2.2.6 Functional assay of olfactory receptors using calcium imaging

Cells were transfected with receptor gene constructs in twelve–well tissue culture plates. The twelve-well plate containing Sf9 cells expressing the receptor of interest was tilted and the media removed with a pipette before the cells were washed with 1 mL of sterile assay buffer. Five hundred microliters of assay buffer containing 2 µM Fluo–4 acetoxymethyl ester and 0.01% pluronic acid F–127 (Sigma-Aldrich) was added to each well and the plates incubated in the dark for 20 minutes at 28 °C. The Fluo–4 solution was removed and the cells washed twice with sterile assay buffer to remove excess dye and replaced with 400 µL of fresh buffer. Each plate was then incubated in the dark at 28 °C for a further 20 minutes prior to calcium imaging.

Calcium imaging was performed using the 10x objective lens on a Leitz Fluovert FS inverted fluorescence microscope fitted with an I3 filter set (excitation filter 450-490 nm, dichroic mirror 510 nm, long pass emission filter 520 nm), an N2.1 filter set (excitation filter 515-560 nm, dichroic mirror 580 nm, long pass emission filter 590 nm) and a MicroMax CCD camera system (model RTE/CCD-1300-Y/HS; Princeton Instruments). The camera and image acquisition were controlled using MetaFluor v6.1 (Universal Imaging Corporation).

Initially, images of the Fluo–4 loaded cells were taken every 10 seconds for one minute. An internal negative control was included within each experiment to control for stretch receptors by next adding 50 µL of assay buffer and recording again for one minute at 10 second intervals. Fifty microliters of the odorant solution of interest was then added and six further images acquired at 10 second intervals. Finally, to determine maximal
fluorescence, the calcium ionophore ionomycin (Sigma) was added to a final concentration of 10 µM and images collected for a further minute at 10 second intervals. Photo-bleaching was minimized as much as possible by switching off the light source between image acquisitions.

Image data was analyzed using Scion Image v1.34 (Scion Corporation) or MetaFluor Analyst v5.1 (Universal Imaging Corporation). When using Scion Image, the image files from a particular well were imported and formatted into a stack of images based on capture order. Responding cells were manually identified by scrolling through the time series of images and looking for particular cells for which there was a visible change in fluorescence on addition of ligand. The mean grey value for such cells was recorded using the “Measure” function and background corrected by subtracting a similarly measured adjacent background area prior to further analysis.

For MetaFluor Analyst, regions of interest (ROI) defined by drawing around the margin of every cell in the field of view, and an additional three background ROIs also defined. The fluorescence intensity data for each timepoint was extracted using the “Calculate Plot” function in the software. The background fluorescence value was calculated by averaging the three background ROIs for each timepoint. This value was used to correct for differences in the background of each image by subtracting the background of the image from each cell’s fluorescence value before it was included in further analysis. Cells responding to the addition of an odorant were selected by applying a set of criteria developed from a meta-data set of cells that had been previously assigned as responding by several independent researchers. These criteria are described in Appendix B.

The change in fluorescence elicited by an odorant was calculated by measuring the mean fluorescence intensity of each cell over the twelve images before the addition of the substrate ($F_{Blank}$) and subtracting this from the highest fluorescence intensity of the same cell over six consecutive images after addition of the substrate ($F_{Substrate}$). This value was divided by the maximal fluorescence, given by the maximal fluorescent intensity of the cell after exposure to ionomycin ($F_{Max}$). The formula used for this calculation is shown in Equation 2.1. Cells that responded to saline alone were disregarded from further analysis.
The mean $\Delta F$ for a given substrate or concentration was used to indicate the degree of olfactory receptor activation. EC$_{50}$ values were calculated by plotting the mean $\Delta F$ for a given odorant at different concentrations in OriginPro v7.5 (OriginLab) and fitting a sigmoidal curve weighted to the error bars around the mean $\Delta F$ values. The Hill slope of these curves was not constrained. The EC$_{50}$ value was calculated from this curve as the concentration at which half the maximal response was obtained.

\[ \Delta F = \frac{F_{\text{Substrate}} - F_{\text{Blank}}}{F_{\text{Max}}} \]  

(2.1)

## 2.3 Results

Transient transfection of Sf9 cells and subsequent expression of a gene of interest has been previously used to express functional receptor proteins (Knight & Grigliatti, 2004). In order to check that this cell line is able to express an olfactory receptor from the pIB/V5 plasmid, reverse transcription followed by PCR on cDNA created from Sf9 cells transfected with pIB-Or22a was performed. The first primer pair used, OpIE2 forward and reverse primers (expected fragment = 1455 bp), was specific to the promoter and poly–adenylation sequence of the pIB/V5 plasmid, and the other two pairs used, 22aD and OpIE2 reverse (expected fragment = 959 bp), and 22aH and OpIE2 reverse (expected fragment = 534 bp), each contained a primer specific to Or22a along with the OpIE2 reverse primer. For all three pairs of primers, bands were observed on an agarose gel corresponding to the expected fragment sizes (1500 bp, 950 bp and 550 bp respectively) indicating that the Or22a gene was being successfully expressed (Figure 2.1). No expression of Or22a was seen in Sf9 cells transfected with empty pIB/V5 plasmid.

Sf9 cells transfected with pIB-Or22a express Or22a at the mRNA level, but it remained to be seen whether the protein is expressed and functional. To test this, Sf9 cells transfected with pIB-Or22a using FuGene 6 were loaded with the calcium–sensitive fluorescent dye, Fluo–4. When ethyl butyrate was added to a final concentration of $10^{-5}$ M, 3.9% of
2.3 Results

Figure 2.1: Reverse transcription polymerase chain reaction of Sf9 cells transfected with pIB-Or22a to detect expression of Or22a. A) OpIE2 forward and reverse primers, C) 22aH and OpIE2 reverse primers and E) 22aD and OpIE2 reverse primers used on Sf9 cells transfected with pIB-Or22a plasmid DNA. Lanes B, D and F are the corresponding PCRs performed on cDNA transcribed from total RNA extracted from Sf9 cells transfected with pIB/V5 (empty vector control).

Sf9 cells exhibited a visible increase in calcium level. Ethyl butyrate was chosen for this assay as it had been shown to activate the ORN expressing Or22a in vivo (Hallem & Carlson, 2004). The responses of transfected Sf9 cells to water and non–transfected cells to ethyl butyrate was negligible (Figure 2.2). Representative fluorescent intensity data for responding and non–responding Sf9 cells are presented in Figure 2.3.

The response rate of 3.9% was very low, and in order to increase throughput by improving the transfection rate, a number of transfection reagents were trialled. Expression of the dsRed fluorescent protein was used as a marker of successful plasmid transfection and expression. The number of cells expressing dsRed after transfection of the plasmid pIB-dsRed with Escort IV (Sigma), Cellfectin (Invitrogen) and Lipofectamine (Invitrogen) using the manufacturer’s instructions was compared to that of FuGene 6 (Figure 2.4). The transfection rate generated by Escort IV was greater than that of the other three, with a successful dsRed expression rate of nearly 40%, compared to the other three which had expression rates of approximately 10%. Optimisation of the transfection rate generated by Escort IV demonstrated that the optimal transfection conditions for Sf9 cells were using 6 µL of Escort IV and 1 µg of plasmid DNA per well of a six well plate and incubating the cells for 48 hours prior to an experiment (Figure 2.5).
2.3 Results

Figure 2.2: Response of single Fluo–4 loaded Sf9 cells transfected with pIB-Or22a to a) $10^{-5}$ M ethyl butyrate and b) water, and the response of non-transfected cells to c) $10^{-5}$ M ethyl butyrate. The fluorescence response is quantified based on the percentage of cells that increase in fluorescence after addition of the ligand in an entire field of view. The results shown are the mean ± S.E.M. of three separate experiments.

Despite a four-fold increase in the apparent transfection rate, no similar increase in the number of Or22a–expressing Sf9 cells responding to ethyl butyrate was observed. This suggested that the transfection rate was not the limiting factor in this system. A report had demonstrated that the co–expression of the chaperone olfactory receptor Or83b in a heterologous assay system increased the response rate and sensitivity through dimerisation with an odorant sensitive OR in HEK293 cells (Neuhaus et al., 2005). When Sf9 cells were transfected with both pIB-Or83b and pIB-Or22a, loaded with Fluo–4 and exposed to $10^{-5}$ M or $10^{-10}$ M ethyl butyrate, there was no difference in the number of cells responding compared to cells transfected with pIB-Or22a alone (Figure 2.6). One possibility was that the transfection of two plasmids did not necessarily mean that each Sf9 cell would express both genes. In order to rule this out, a plasmid was constructed that contained copies of both Or22a and Or83b under the control of separate promoters to ensure that each cell that was transfected would express both genes. Transfection of this plasmid in Sf9 cells also failed to elicit an increase in the response rate over the expression of Or22a alone (Figure 2.6).

Sf9 cells had been shown to express functional Or22a (Figure 2.2), but it was also important to test whether the receptor still responded to a similar range of ligands when expressed in this system as it does in vivo. Or22a expressed in Sf9 cells was
Figure 2.3: Comparative fluorescence of responding and non–responding Sf9 cells transfected with pIB-Or22a over the timecourse of a calcium assay. Fluorescent intensity of two cells is plotted and representative images of those cells shown for each of the four stages of a calcium assay: the first six frames prior to addition of buffer, frames seven to twelve after addition of buffer, frames thirteen to eighteen after addition of $10^{-5}$ M ethyl butyrate and frames nineteen to twenty four after addition of 2 µM ionomycin. Pseudocoloured images represent the fluorescence intensity (red is highest, purple is lowest).

presented with a panel of odorants that were previously used to test Drosophila ORs in vivo (Hallem & Carlson, 2004). The ligands tested included esters (ethyl butyrate, pentyl acetate, isoamyl acetate, ethyl acetate, ethyl propionate, butyl acetate, methyl salicylate), alcohols (hexan–1–ol, 1–octen–3–ol), ketones (2,3–butanedione, 2–heptanone) and an aldehyde ($E^2$–hexenal). Each chemical was exposed to Fluo–4 loaded Sf9 cells expressing Or22a at a concentration of $10^{-9}$ M, and the increase in fluorescence elicited by the substrate expressed as a fraction of the fluorescence elicited by the calcium ionophore ionomycin (Figure 2.7). Addition of saline prior to addition of the substrate was included as a negative control. Of the odorants tested, ethyl butyrate gave the strongest response for Or22a, and four of the top six ligands were short chain esters.

In order to determine if the Sf9 assay system could be utilised to collect pharmacological
data for insect ORs, five odorants were chosen to be tested over a wide range of concentrations, from $10^{-5}$ M to $10^{-13}$ M. These odorants elicited high (ethyl butyrate), medium (pentyl acetate, ethyl acetate) and low (methyl salicylate, $E2$–hexenal) responses when administered at $10^{-9}$ M (Figure 2.7). A sigmoidal curve was fitted to the $\Delta F$ values recorded for each concentration and the concentration at which each odorant elicited half of the maximal response ($EC_{50}$) was calculated (Figure 2.8). The $EC_{50}$ value for ethyl butyrate is $(1.53 \pm 1.16) \times 10^{-11}$ M, pentyl acetate is $(5.61 \pm 1.65) \times 10^{-10}$ M and ethyl acetate is $(3.72 \pm 1.47) \times 10^{-9}$ M. Neither methyl salicylate nor $E2$–hexenal elicited a sufficient response to calculate an $EC_{50}$ value, even at the highest concentrations tested.

2.4 Discussion

Using *D. melanogaster* Or22a as a model insect OR, an assay system has been developed that provides a new tool for the de–orphaning of insect ORs. Or22a is activated by a number of odorants when expressed in Sf9 cells (Figure 2.7) with similar relative affinities to when it is expressed in *vivo* (Hallem & Carlson, 2004). Of the odours tested, ethyl butyrate was the odorant that elicited the strongest response from Or22a, and odorants
2.4 Discussion

Figure 2.5: Optimisation of the Escort transfection protocol. Transfection rate was calculated for varying amounts of Escort and DNA per well measured at three timepoints, 24, 48, and 72 hours post-transfection. A) 6 µL of Escort and 1 µg of DNA. B) 6 µL of Escort and 2 µg of DNA. C) 12 µL of Escort and 2 µg of DNA. D) 12 µL of Escort and 1 µg of DNA. E) 3 µL of Escort and 1 µg of DNA. F) 3 µL of Escort and 0.5 µg of DNA. G) 6 µL of Escort and 0.5 µg of DNA. Error bars represent the S.E.M. of three fields of view per well for two wells.

Figure 2.6: The response of Sf9 cells transfected with pIB-Or22a (left), co-transfected with both pIB-Or22a and pIB-Or83b (middle) and transfected with pIB-Or22a-Or83b double construct (right) to two concentrations of ethyl butyrate. The mean ∆F was calculated for 12-15 cells from two calcium imaging experiments when exposed to $10^{-5}$ M or $10^{-10}$ M ethyl butyrate. Error bars indicate S.E.M.
Figure 2.7: Response profile of Fluo-4 loaded Sf9 cells transfected with pIB-Or22a to the range of odorants previously tested in vivo (Hallem & Carlson, 2004), each at a final concentration of $10^{-9}$ M. Water is included as a negative control. The strength of response is calculated using the following formula given in Equation 2.1. The value for $\Delta F$ given for each compound is the mean response of 6-10 responding cells for that compound from two separate experiments $\pm$ S.E.M.

with ester functional groups were four of the top six ligands. Odorants that activated Or22a moderately in vivo, such as hexanol and heptanone, did so in vitro, except 2,3–butanedione. This substrate activated Or22a much more strongly in vitro than in vivo. It is possible that the expression of odorant binding proteins that sequester 2,3–butanedione are expressed in the fly that are not present in the Sf9 cell assay system. An alternative is that the relative solubility and vapour pressure of the odorant mean that different concentrations of 2,3–butanedione are being compared relative to the other ligands.

Calcium imaging in Sf9 cells enabled the collection of pharmacological data for the interaction of five ligands with Or22a. The rank order of the EC$_{50}$ values for ethyl butyrate, pentyl acetate and ethyl acetate determined in vivo matched those obtained using the Sf9 system, but were 4–6 orders of magnitude less sensitive when measured in vivo (Pelz et al., 2006). However, it is impossible to accurately determine the actual concentration of the odorant within the sensilla, and it is possible that the factor limiting sensitivity in vivo is diffusion of the odorant in the olfactory system, rather than the sensitivity of the OR itself (Laing, 1987). As the odorant is delivered in the vapour phase for in vivo assays, the quantitation of the substrate in its initial solution does not necessarily reflect the concentration in the antenna. This is not the case for the Sf9 calcium
Figure 2.8: Dose response curves of Or22a transiently expressed in Sf9 cells to ethyl butyrate, ethyl acetate, pentyl acetate, E2–hexenal and methyl salicylate. The change in fluorescence elicited by a range of concentrations of each odorant was calculated and the data fitted with a sigmoidal curve using OriginLab. EC_{50} values were calculated for the three odorants that activated Or22a and are expressed ± 95% confidence intervals. Error bars represent the S.E.M. of 25–30 responding cells from at least three separate experiments.
imaging system described in this study as the dilutions of the odorant are provided in a solution at a known concentration with no intermediate vapour phase.

The observation that the Sf9 system mimics the \textit{in vivo} response of ORs suggests that the OR is the sole determinant of the specificity of the olfactory system. This is also supported by Hallem \textit{et al.} (2004) who demonstrate that ORs have a similar response profile even when expressed in the non–native neuron. Other heterologous assay systems have also been developed that do not require additional components for odorant discrimination (e.g. Wetzel \textit{et al.} (2001), Sato \textit{et al.} (2008) and Wicher \textit{et al.} (2008)). However, OBPs have an important role in pheromone discrimination, especially demonstrated in the role of LUSH in detection of \textit{cis}–vaccenyl acetate in \textit{Drosophila} (Xu \textit{et al.}, 2005; Laughlin \textit{et al.}, 2008). Additional accessory proteins are also important for pheromone detection, such as sensory neuron membrane proteins (SNMPs) (Benton \textit{et al.}, 2007). Further experiments on pheromone receptors in \textit{in vitro} systems will begin to resolve these differences between general olfaction and pheromone detection.

The sensitivity of Or22a expressed in Sf9 cells is remarkable. The EC$_{50}$ value for the interaction of Or22a with ethyl butyrate determined here is $1.53 \times 10^{-11}$ M, which is very low when compared to other receptors. For example, the $\alpha_{2A}$ adrenergic receptor has an EC$_{50}$ of 1.1 $\mu$M for the substrate noradrenaline (Lohse \textit{et al.}, 2003), and the D1 dopamine receptor has an EC$_{50}$ of 13 nM to dopamine (Fici \textit{et al.}, 1997). The sensitivity of ORs expressed in Sf9 cells is high compared to the expression of Or22a in the human HEK293 cell line, where micromolar amounts of ethyl butyrate were required to elicit a response (Neuhaus \textit{et al.}, 2006). Further research using the Sf9 assay system has shown that an OR from \textit{E. postvittana}, Or3, has an EC$_{50}$ of $1.1 \times 10^{-13}$ M for the odorant citral, a further 100 fold increase in sensitivity over Or22a (Jordan \textit{et al.}, 2008). High sensitivities are also observed for \textit{B. mori} ORs when expressed in the Sf9 system (Anderson \textit{et al.}, 2008). The sensitivities of insect ORs presented here are also similar to that reported for the pheromone receptor HR13 from \textit{H. virescens} in the presence of the appropriate pheromone binding protein (Grosse-Wilde \textit{et al.}, 2007). The high degree of sensitivity of insect ORs indicates the behavioural importance of being able to detect very small
quantities of odorant in the environment.

The co–expression of the co–receptor Or83b increased the sensitivity of insect ORs expressed in HEK293 cells (Neuhaus et al., 2006). The response of Bombyx pheromone receptors co–expressed with an Or83b orthologue in Xenopus oocytes was also more sensitive than expression in HEK293 cells, being able to detect a response when 100 nM of pheromone is applied (Nakagawa et al., 2005), but this is still less sensitive than the Sf9 assay system. HEK293 cells stably expressing Heliothis pheromone receptors and a G\textsubscript{a15} protein were also able to detect pheromone components at nanomolar concentrations, but were able to respond at femtomolar concentrations if the appropriate pheromone binding protein was added (Grosse-Wilde et al., 2007). However, the Sf9 assay system does not require the co–expression of any exogenous factors to obtain its high level of sensitivity.

The expression of the receptor at the cell surface can also affect the perceived sensitivity of the response. A cell with large numbers of receptor at the cell membrane that is activated will elicit a larger increase in second messengers than a similar cell with lower amounts of receptor exposed to the same ligand (Clyne et al., 2003). This can cause differences in EC\textsubscript{50} value between experiments. The expression level of Or22a in the individual Sf9 cells measured was not controlled between experiments in this study, and is one area of potential investigation in the future. It is also unknown how the expression of the OR compares between different heterologous systems, such as HEK293 cells or Xenopus oocytes. The high degree of sensitivity observed in the Sf9 cell line may be indicative of efficient expression and membrane targeting of the OR to the cell membrane.

The Sf9 cell line expresses an ortholog of the co–receptor Or83b as demonstrated by RT–PCR (Smart et al., 2008). Expression of Or83b is essential for olfaction in vivo, with Or83b mutant Drosophila being anosmic (Larsson et al., 2004). However, some heterologous assays of insect ORs have demonstrated function without the expression of Or83b (e.g. Wetzel et al., 2001). The endogenous expression of Or83b in Sf9 cells may explain why the additional transfection of Or83b failed to increase the number of cells responding to ethyl butyrate or the sensitivity of the interaction (Figure 2.6). It may also explain the high level of sensitivity exhibited by ORs expressed in this system.
Endogenous Or83b may be acting as a chaperone, increasing the amount of ligand–binding OR reaching the cell membrane (Benton et al., 2006). Alternatively, native Or83b may already have formed an ion channel structure (Sato et al., 2008; Wicher et al., 2008), and co–expressed ligand–sensitive OR is able to interact directly to form an active OR complex. In addition, the expression of Or83b suggests it is possible that Sf9 cells may be expressing other, as yet unknown, proteins that are aiding the function of insect ORs.

The functional analysis of insect ORs by calcium imaging in the Sf9 cell line is not without its practical pitfalls. The low apparent expression rate of ORs means that a large number of assays is required to obtain a large enough data set to be meaningful. Each assay field of view could contain up to 300 Sf9 cells, any of which could be an OR–expressing cell. This means that data analysis is a significant bottleneck in the throughput of this assay system. Initially, cells that were responding were selected by observing those cells that exhibited an increase in fluorescence upon addition of substrate by looking at the images from an assay. However, this proved highly variable between experimental conditions and researchers. Automated data collection was trialled using MetaMorph software, but the differential take–up of the Fluo–4 by the Sf9 cells meant that any automated cell recognition algorithm generated too many false positive or false negative results to be useful. The best system so far utilises a range of Excel macros to determine whether a cell is responding or not. This approach is impartial, but still requires the drawing of regions of interest around each of the cells in a single assay, most of which will be eliminated, and is therefore very time consuming. Whilst this system in its current state is not high throughput, expression of an insect OR in Sf9 cells has demonstrated that it is applicable to the analytical study of insect OR structure and function.
Determination of the Transmembrane Topology of Insect Olfactory Receptors

3.1 Introduction

The function of a protein is defined by the three dimensional structure of the molecule (Laskowski et al., 2003). Knowledge of the structure of a given protein can therefore provide a great deal of insight into its activity. Unfortunately, the experimental determination of high resolution structures for membrane proteins is difficult. This is demonstrated by the relative under-representation of structures of membrane proteins in the RCSB Protein Data Bank. 30% of total proteins are thought to be membrane proteins, yet they make up only 1% of solved structures in the Protein Data Bank (Krogh et al., 2001).
The over-expression, purification and crystallisation of membrane proteins are three of the major bottlenecks in the determination of membrane protein structure (Muller et al., 2008). Recently, the structure of the β2 adrenergic receptor has been solved (Rasmussen et al., 2007). This protein was crystallised using the binding of antibodies to stabilise the protein during the purification and crystallisation process. This technique could be used to increase the number of membrane protein structures known, but it is very laborious to develop suitable antibodies. Another potential approach around these bottlenecks is the computational modelling of a protein structure, either using a homologue of known structure as a template (Strahs & Weinstein, 1997), or completely de novo (Bradley et al., 2005). While advances are being made for the modelling of globular proteins, these techniques still yield poor results for membrane proteins due to their generally larger size and the complexity of modelling the cell membrane (Fleishman & Ben-Tal, 2006). Until more advanced computational methods are developed, other techniques must be used to determine structural elements of membrane proteins.

### 3.1.1 Prediction of membrane topology of proteins

There are important pieces of structural information that can be determined without a high resolution crystal structure. The topology, or the number and location of TM domains and their orientation relative to the cell membrane, can be defined using a combination of computational and experimental techniques (von Heijne, 2006). Early algorithms for the prediction of TM regions identified regions of hydrophobicity of at least 20 amino acids in length as putative TM domains (Kyte & Doolittle, 1982). Simple measures of hydrophobicity do not predict the orientation of the protein, and it was not until the observation that internal loops contained a higher proportion of positively charged amino acids (the “positive–inside” rule) that predictions of orientation could be made (Claros & von Heijne, 1994). The development of machine-learning algorithms, such as hidden Markov models (HMMs) and neural networks have allowed many structural determinants, such as hydrophobicity, charge and the statistical analysis of likely localisation from known structures, to be used in predictions concurrently. There
are now many such algorithms available, such as TMHMM (Krogh et al., 2001) and HMMTOP (Tusnady & Simon, 2001), that have greatly increased the accuracy of topology prediction. Further improvement has come through the analysis of alignments of closely related genes, such as the program TMAP (Milpetz et al., 1995), on the assumption that topology will be conserved among family members.

3.1.2 Experimental techniques for topology determination

While the degree of accuracy of TM domain prediction is increasing with each new algorithm released, the result is still only a model that requires experimental validation. One common technique used for the determination of the orientation of a hydrophilic region of a protein is the insertion of a reporter gene that functions only in a specific subcellular localisation. For example, alkaline phosphatase is only active when translocated to the periplasm of bacteria, and β–galactosidase is only active when in the cytosol. Insertion of these proteins into putative hydrophilic regions and the addition of colorimetric substrates for the given enzyme will uncover which enzyme is active, and therefore which side of the membrane that loop is on (Haardt & Bremer, 1996). However, these techniques are generally only applicable in bacterial expression systems (van Geest & Lolkema, 2000).

A technique applicable to proteins in a range of systems is cysteine scanning mutagenesis. This utilises the unique chemical properties of the sulphydryl side group of cysteine residues. By eliminating and then re–introducing individual cysteine residues in hydrophilic domains and adding membrane permeable or impermeable reagents that crosslink to the sulphydryl group, the orientation of the cysteine residue can be determined (Zhu & Casey, 2007). Glycosylation scanning mutagenesis is a similar technique. N–linked glycosylation occurs at the asparagine residue of the consensus sequence asparagine–X–threonine/serine, where X is any amino acid, on the luminal side of the endoplasmic reticulum membrane (Welpley et al., 1983). By removing endogenous glycosylation sites, re–introducing a specific glycosylation site and assessing whether it is N–glycosylated
by mobility shift on an SDS–PAGE gel, the orientation of the site can be determined (Vannier et al., 1998). Glycosylation scanning can also help define hydrophilic loop size, as glycosylation can only occur if the Asn residue is at least 12 residues upstream and 14 residues downstream from the end of TM segments (Nilsson & von Heijne, 1993). However, this restriction reduces the utility of glycosylation scanning for the assessment of small hydrophilic loops.

These techniques all require the modification of the protein which can cause structural alterations. The generation of antibodies to the putative hydrophilic regions of a protein of interest can be used to determine its topology in the native state (Toyoda et al., 2000). If the epitope is on the cytosolic side of the membrane, the antibody can only bind to its epitope if the cell membrane is permeabilised with detergent. The antibody will bind to an extracellular epitope regardless of the presence or absence of detergent. In this way the orientation of the hydrophilic loop can be determined (Figure 3.1). If antibodies to the native protein are not available or the protein is poorly antigenic, epitopes that are recognised by commercially available antibodies, such as the myc epitope or the FLAG epitope, can be inserted into the protein and used for epitope scanning (Smolka et al., 1999).

**Figure 3.1:** Schematic diagram outlining the principle of topology determination by epitope tagging. A membrane protein (gray cylinders) is expressed with an antibody epitope (red region) inserted into it and exposed to a fluorescently conjugated antibody. If the cell membrane is not permeabilised (left panel), the antibody can only bind to the epitope if it is localised to the external face of the cell membrane. However, antibody binding will occur regardless of the position of the epitope if the cell membrane is permeabilised with detergent (right panel). Thus the localisation of that region of the protein can be determined by the condition required for antibody accessibility.
3.1.3 Topology of insect ORs

Vertebrate ORs are known to be rhodopsin–like GPCRs, and many structural features can thus be assigned to them based on this homology (Malnic et al., 1999). Indeed, there are even homology models of vertebrate ORs that have predicted their odorant binding energy with a moderate degree of accuracy (Hall et al., 2004). Insect ORs show no sequence homology to this superfamily of receptors, and instead share more sequence homology with mouse potassium channels than mouse ORs (Benton et al., 2006). These differences were further highlighted by the observation that three insect ORs have an intracellular N–terminus, the opposite orientation to GPCRs (Benton et al., 2006). The co–receptor Or83b, as well as the ligand–binding ORs Or9a and Or43a, have an intracellular N–terminus, determined using a combination of β–galactosidase reporter expression, bimolecular fluorescence complementation and immunochemistry (Benton et al., 2006). These results throw doubt onto assumptions that insect ORs are GPCR–like in structure, and while TM prediction programs predict that they are likely to have seven TM domains (Benton et al., 2006), this has only recently been experimentally validated for the co–receptor Or83b using glycosylation scanning (Lundin et al., 2007). To date, the number of TM domains of the odorant–sensitive insect ORs has not yet been determined empirically.

3.1.4 Aims

The overall aim of this chapter is to determine the TM orientation and topology of insect ligand–binding ORs. Epitope tagging followed by immunochemistry has been chosen to address this question. This approach has been chosen because the insertions are smaller (6–10 amino acids) compared to reporter protein insertions, which can be several hundred amino acids in size and may therefore have a greater potential affect on the OR. This approach also requires less cloning compared with a cysteine scanning mutagenesis approach, and is more flexible than glycosylation scanning mutagenesis. It is also less time consuming and more predictable than generating antibodies to the ORs themselves.
Antibody epitopes will be inserted onto the termini of these receptors and they will be expressed in S2 cells. The accessibility of epitopes inserted into hydrophilic loop domains of ORs to the appropriate antibody will be tested in the presence and absence of a membrane permeabilising detergent using rhodopsin constructs as a postive control.

As described above, these receptors are thought to have an intracellular N–terminus. However, this has yet to be tested in an OR from another species. The orientation of the N– and C–termini of both Or22a from *D. melanogaster* and Or1 from *E. postvittana* will be tested. If the hypothesis that Or22a and Or1 have an intracellular N–terminus is true, the N–terminal epitope will only be accessible in the presence of detergent while the C–terminal epitope will be accessible regardless of the permeabilisation of the cellular membrane. On the other hand, accessibility of the N–terminal epitope in both conditions will suggest the alternate orientation. If the orientation of Or83b extends to an OR from *E. postvittana*, we can conclude that this recently discovered structure is conserved through 250 million years of evolution.

The number of TM domains that the insect ligand binding ORs contain will also be examined. *In silico* evidence suggests that they contain seven TM domains. This hypothesis will be tested for *D. melanogaster* Or22a. TM domain prediction algorithms will be used to suggest possible locations of the seven TM and six hydrophilic loop domains of this receptor. Epitopes will be inserted into each of these six putative hydrophilic loop domains, and their antibody accessibility tested. If these loop domains show an alternating out–in–out topology and the termini are on opposite sides of the membrane, then a seven TM model can be applied to this receptor family. If a different topology is observed, then an alternate model of the topology of these receptors must be hypothesised.

An important issue in using the epitope–tagging approach is whether the insertion of epitopes into ORs affects the structural integrity of the receptors. This will be tested by assessing the function of ORs containing epitopes using calcium imaging. If the function of these receptors is not significantly different to that of the wild–type receptor, the structure is unlikely to have been greatly affected by the insertion of an epitope. If the insertion of an epitope does eliminate the ability of the receptor to transduce an odorant binding
event into a calcium signal similar to that of the wild-type receptor, then some caution
must be exercised in the interpretation of epitope accessibility experiments.

3.2 Materials and Methods

3.2.1 Materials

Mouse anti–FLAG M2 monoclonal antibody, 4’,6–diamidino–2–phenylindole (DAPI),
poly–L–lysine solution, Triton X–114, fetal goat serum, saponin and paraformaldehyde
were from Sigma–Aldrich. Slowfade Gold mountant, MagicMark XP western blot
molecular weight standards, mouse anti–myc antibody, goat anti–mouse AlexaFluor 488
conjugated antibody, LDS sample buffer and 4–12% PAGE gels were from Invitrogen.
Rabbit anti–mouse antibody conjugated to alkaline phosphatase was from Abcam. Cell
culture dishes were from Nunc. Coverslips and microscope slides were from Greiner.
Polyvinylidene fluoride membrane was from Millipore. NBT–BCIP stock solution was
from Roche.

3.2.2 Transmembrane domain prediction of Or22a

The protein sequence of D. melanogaster Or22a (Genbank accession CG12193) was
submitted to the following TM domain prediction programs: DAS (Cserzo et al., 1997),
TMHMM (Krogh et al., 2001), TMPred (Hofmann & Stoffel, 1993), TMAP (Milpetz
et al., 1995), Predator (Frishman & Argos, 1995), SPLIT (Juretic et al., 1993), HMMPH
(Tusnady & Simon, 2001), PredTMR2 (Pasquier & Hamodrakas, 1999), RbDe (Campagne
& Weinstein, 2000), SOSUI (Hirokawa et al., 1998) and TOPPRED (Claros & von
Heijne, 1994). The output for each algorithm was plotted in Illustrator (v. CS2, Adobe
Systems). Each TM domain was represented by a bar drawn to scale and placed along a
line that represents the peptide chain of Or22a. This allowed the different predictions
to be compared and a consensus assignment of TM domains to be made. TOPO2
(http://www.sacs.ucsf.edu/TOPO-run/wtopo.pl) was used to draw a two dimensional representation of the consensus model of OR topology.

### 3.2.3 Construction of epitope–tagged ORs

In order to determine the topology of insect ORs, an epitope–tagging approach was chosen. Myc–tagged rhodopsin TM1 constructs were created to ensure antibody, fixative and detergent concentrations were optimal for determination of orientation. Versions of Or22a with myc epitopes on the N– and C–termini and FLAG epitopes in each putative hydrophilic region were constructed, and *E. postvittana* Or1 was also myc–tagged on the N– and C–termini in order to assess its orientation.

N–terminally myc–tagged Or22a was kindly provided by Dr Coral Warr (Monash University, Australia) and sub–cloned into pIB-V5/His vector using the restriction enzymes *Kpn* I and *Sac* II. To construct a C–terminal myc–tagged Or22a and to facilitate the insertion of FLAG (DYKDDDDK) epitopes into the predicted hydrophilic loop regions of the receptor, two *Drosophila* codon–optimised constructs were synthesised (Genscript Corporation). Each construct consisted of the Or22a open reading frame with three in–frame copies of the myc epitope on the 3’ end, all surrounded by *Kpn* I (5’) and *Sac* II (3’) restriction sites to allow for later cloning into pIB/V5-His. Both constructs contained two silent mutations that introduced additional restriction sites; a *Sal* I site by replacing 394C>G and an *Aat* II site by changing 618G>C. In addition to these, the second construct (the “donor”) had FLAG epitopes surrounded by a unique pair of restriction sites inserted directly between the following amino acids: Thr77-Phe78 (*Eco* RI), Arg123-Cys124 (*Sph* I), Asp178-Ser179 (*Nhe* I), Lys244-Thr245 (*Afl* II), Thr295-Gly296 (*Spe* I) and Ser338-Asp339 (*Bsp*E I).

These two constructs, the “donor” and “acceptor” versions of Or22a, were shuttled from the supplied pUC57 vector into pIB/V5-His using *Kpn* I and *Sac* II restriction sites. Using the restriction sites silently introduced plus four endogenous restriction sites (*Bam* HI, *Xho* I, *Bsm* I and *Mfe* I) six versions of Or22a were created that contained a FLAG
3.2 Materials and Methods

epitope in one of its predicted hydrophilic region plus a C–terminal myc epitope. The C–terminal myc epitope was then removed using the restriction enzymes Mfe I and Sac II (for FLAG epitopes in loops 1–5) or the PCR (Loop 6) as activity could not be recovered from any C–terminal myc–epitope fusion protein in Sf9–based cell assays (see Results). The PCR used the Expand High Fidelity system (Roche) according to the manufacturer’s recommendations and the two primers 5’–GGTACCATGTTAAGCAAGTTTTTTTCCCCA–3’ and 5’–AACTTTGGCCGAAAGGTTTTCAATAACCGCGG–3’, which removed the C–terminal myc–tag and added Kpn I and Sac II sites for the subsequent subcloning of the PCR product into pIB/V5-His.

N–terminally myc–tagged Epiphyas postvittana Or1 (EposOr1) was prepared by Dr Melissa Jordan after identification from an antennal EST database developed by HortResearch. The N–terminal myc–tag was removed and replaced by a C–terminal myc–tag using the PCR (reaction conditions as above) and the primers 5’–CTCGAG-ATGGATGTATTCAATTAAAAAT–3’ and 5’–CCGCGGTTAAAGGTCTTCTTCGGAGATT-AAGCTTTGTTCTGATTTGCAAATGTTCTCAG–3’. PCR product fragments were cloned into pGem–T Easy, sequenced to verify integrity of the construct and shuttled into pIB/V5-His using the Xho I and Sac II sites included in the PCR primers.

Control constructs consisting of the first 70 amino acids of rhodopsin (Benton et al., 2006) with three copies of the myc–tag at either the N– or C–terminus were codon optimised for Drosophila expression and synthesised by Genscript Corporation. They were shuttled into pIB/V5-His using Kpn I and Sac II. The sequence for these constructs is given in Appendix A.

3.2.4 Activity assays on epitope tagged ORs

Activity assays on epitope–tagged ORs were carried out using calcium imaging in Sf9 cells as described in Chapter 2. The activity of Or22a was measured with ethyl butyrate at a concentration of $10^{-5}$ M, and EposOR1 was analysed with citral at a concentration of $10^{-5}$ M.
3.2.5 Membrane fractionation by differential centrifugation

In order to detect the expression of insect ORs using western blotting, a differential centrifugation membrane fractionation technique was applied to both enrich the protein sample for ORs and to test for their expression in the cell membrane. This approach is adapted from Henningsen et al. (2002).

Forty eight hours post–transfection, S2 cells were washed twice gently with ice–cold PBS (137 mM NaCl, 10 mM Na$_2$HPO$_4$, 2.7 mM KCl, 2 mM KH$_2$PO$_4$, pH 7.4). One mL of ice-cold PBS was then used to dislodge the cells from the growing surface. The cell suspension was added to a 1.5 mL eppendorf tube and spun at 250g for 5 minutes to pellet the cells. The PBS was removed and the cells resuspended in 200 µL of lysis buffer (150 mM NaCl, 50 mM Tris, 1% ASB-14, 1x protease inhibitor cocktail, pH 7.8). The cells were then sonicated with two 30 second passes with an ultrasonic homogeniser (Kontes) on full power, with the solution rested on ice for at least two minutes between passes.

Unlysed cells and cellular debris was collected by spinning the lysate at 1,500g for five minutes. The supernatant was then spun at 7,500g to pellet cellular organelles. The supernatant from this step was then spun at 100,000g in a Beckman Airfuge for 30 minutes to pellet the cell membrane. The resultant membrane pellet was resolubilised in lithium dodecyl sulphate (LDS) sample buffer plus reducing agent and heated at 99 °C or 37 °C for 4 minutes prior to centrifugation at full speed for 30 seconds. The supernatant was then separated by 1D gel electrophoresis on a 4–12% Bis-Tris polyacrylamide gel according to the manufacturer’s instructions.

3.2.6 Membrane protein extraction by Triton X–114 phase separation

S2 cells were transfected in 6–well plates with the olfactory receptor construct of interest (see Section 2.2.4 for protocol). Forty eight hours after transfection the cells were washed twice and then dislodged with 1 mL cold PBS. The cells were pelleted by centrifugation
at 13,000g prior to extraction of membrane proteins using Triton X–114 phase separation (Bordier, 1980). The cell pellets were resuspended in 250 µL of lysis buffer (2% (v/v) pre-condensed Triton X-114 in PBS), sonicated for two, thirty second passes and incubated on ice for two hours. Insoluble debris was removed by spinning the samples at 13,000g and 4 °C for two minutes. The supernatant was incubated at 37 °C for ten minutes and then spun at 13,000g at 25 °C to separate the detergent and aqueous phases. The aqueous phase was reserved for analysis, while the detergent phase pellet was washed by adding 1 mL of fresh cold PBS and incubating on ice for ten minutes. It was then heated at 37 °C for ten minutes and spun at 13,000g at 25 °C to separate the phases. The aqueous phase was discarded, and the detergent phase pellet washed in this manner a further two times. Following the final wash step the detergent phase was resuspended in 50 µL of LDS sample buffer (including 100 mM dithiothreitol and 6 M urea), heated at 37 °C for 20 mins and spun at 13,000g for 30 seconds to remove particulates. The protein samples were separated by SDS-PAGE using 4-12% Bis-Tris polyacrylamide gels as described above.

3.2.7 Western blotting

The separated proteins were transferred to polyvinylidene fluoride (PVDF) membrane in a Trans–Blot SD semi-dry transfer device (BioRad). The protein gel and filter paper were equilibrated in Towbin transfer buffer (192 mM glycine, 25 mM Tris, 20% methanol and 0.1% SDS) for 30 minutes. Meanwhile, the PVDF membrane was wetted in 100% methanol for 30 seconds, then distilled water for two minutes and finally equilibrated in Towbin transfer buffer as above. The transfer device was set up according to the manufacturer’s instructions and the transfer conducted at 5 V for 5 hours.

After electroelution of the proteins onto the PVDF membrane, the membrane was incubated in blocking buffer (TBS–T (50 mM Tris, 150 mM NaCl, 0.1% Tween–20, pH 7.5) + 5% non-fat milk powder) for two hours. The membrane was washed three times for five minutes in fresh TBS–T, and then incubated in a 1:5000 dilution of mouse anti-myc antibody in blocking buffer for two hours. The membrane was washed three
times for five minutes in fresh TBS–T again and then incubated in a 1:2500 dilution of anti-mouse alkaline phosphatase conjugated antibody for one hour. The membrane was washed six times for five minutes each with TBS–T prior to detection.

Chemiluminescence was used to detect immunoreactive regions in the membrane. The membrane was incubated in fresh detection buffer (100 mM Tris, 100 mM NaCl, pH 9.5) for five minutes and then drained well. Approximately 3 mL of NBT–BCIP (Roche) colorimetric substrate was added to the membrane and incubated until bands became apparent. The membrane was then washed in an excess of water and left to dry prior to imaging with a digital camera.

3.2.8 Coating coverslips with poly–L–lysine

Pre–cleaned 21 mm coverslips (Greiner) were acid–cleaned to remove impurities by soaking them in a 1:1 by volume mixture of nitric and hydrochloric acids for 60 minutes. The coverslips were then rinsed in ultrapure water until the solution around the coverslips reached pH 7. The coverslips were stored in a 70% ethanol solution until coating with poly–L–lysine.

Acid–cleaned coverslips were removed from the 70% ethanol solution, dipped in 100% ethanol and flamed to ensure sterility. A number of coverslips were then placed in a 10 cm cell culture dish, covered in 0.01% poly–L–lysine solution and incubated for 60 minutes. The coverslips were rinsed three times with sterile PBS, before the PBS was removed and the coverslips left to dry. Once dry, the coverslips were rinsed once more with PBS and stored at 4 °C in PBS until use.

3.2.9 Preparation of formaldehyde fixative

2% formaldehyde solution was prepared by adding two grams of paraformaldehyde powder to 100 mL of ultrapure water in a beaker and covering it with tin foil. The solution was
heated at 70 °C with stirring for one hour until clear. Once cool, the solution was made up to 1x PBS with a 10x PBS stock solution and adjusted to pH 7.2. Fixative was stored at 4 °C prior to use, but not stored for more than one week.

3.2.10 Immunohistochemistry of S2 Cells

The following method was adapted from that described by Smolka et al. (1999). S2 cells were seeded on poly–L–lysine coverslips placed in 6 well tissue culture plates 24 hours prior to transfection. Transfection of S2 cells was carried out as described in Section 2.2.4. Forty eight hours post transfection, duplicate coverslips were antibody–stained with the following two protocols to determine the localisation of the epitope.

3.2.10.1 Detection of antigens under permeabilized conditions

All steps were carried out at room temperature unless otherwise stated. Forty eight hours after transfection, coverslips of S2 cells expressing the epitope construct of interest were transferred to a clean 6 well cell culture plate and washed three times for five minutes each with PBS. The cells were fixed in 2% formaldehyde solution for 15 minutes, rinsed in 50 mM ammonium chloride to quench excess formaldehyde prior to blocking and permeabilisation in blocking buffer (PBS containing 5% goat serum and 0.1% saponin) for 15 minutes. The cells were incubated in a 1:200 dilution of the primary antibody stock in blocking buffer and incubated for two hours. The cells were washed with PBS three times for five minutes each, before being incubated in 1:200 dilution of the secondary antibody stock in blocking buffer and incubated for one hour. The coverslips were then washed five times for five minutes with PBS, before being incubated in PBS containing 2 µg/mL of the nuclear counterstain 4’6-diamidino–2–phenylindole (DAPI) for five minutes. The cells were briefly rinsed and left to air dry prior to mounting on glass slides in SlowFade Gold mountant.
3.2.10.2 Detection of antigens under non–permeabilized conditions

The following steps were carried out at 4 °C to reduce membrane internalisation. Forty eight hours after transfection, coverslips of S2 cells expressing the epitope construct of interest were transferred to a clean 6–well cell culture plate and washed three times for five minutes with PBS. The cells were then incubated in a 1:200 dilution of the primary antibody stock in PBS + 5% goat serum for two hours. The coverslips were then washed five times for five minutes with PBS prior to fixation and quenching as above. The following steps were carried out at room temperature. The cells were washed three times for five minutes in PBS and incubated in a 1:200 dilution of the secondary antibody stock in blocking buffer for one hour. The final washing, DAPI incubation and mounting of coverslips was carried out as above.

3.2.11 Imaging of immunochemistry slides

Slides were imaged using a Vanox AHBT3 upright fluorescence microscope (Olympus) fitted with a CoolSnap CCD camera (Photometrics) and standard FITC and DAPI filter sets. Camera acquisition parameters were controlled with RS Image (v1.9.2, Roper Scientific). For the AlexaFluor 488 dye the exposure time was held constant at 200 ms, while DAPI was imaged with an exposure time of 20 ms. Images of cells positive for AlexaFluor 488 staining are representative of at least 100 positive cells from duplicate slides from three separate experiments. A result was deemed to be negative when no AlexaFluor 488 staining cells were observed in any of duplicate slides in experiments repeated at least three times. This amounts to observing at least 20,000 cells. Results indicating no staining for non–permeabilised cells were only accepted if both the Rhodopsin TM1 controls exhibited an extracellular N–terminus and intracellular C–terminus and if positively stained cells were seen in the corresponding permeabilised condition in the same experiment.
3.3 Results

A number of freely available algorithms for the prediction of the TM topology of proteins were used to develop a consensus model of the topology of *D. melanogaster* Or22a. The protein sequence of Or22a was submitted to each program in FASTA format and the predictions recorded (See Table 3.1 for a summary of results). A description of the methodology of each program and the full output can be found in Appendix C. The prediction output was plotted using Illustrator in order to allow direct comparison of the different program’s predictions (Figure 3.2). With the assumption that insect ORs contain seven TM domains (Benton et al., 2006), consensus regions for the TM domains of Or22a were able to be assigned based on the predictions of at least five algorithms.

![Figure 3.2: Schematic diagram showing the distribution of *D. melanogaster* Or22a transmembrane domains predicted by eleven prediction algorithms. Each red segment represents a putative TM domain, and is in scale according to the amino acid length as shown at bottom. The seven consensus TM domains are indicated by roman numerals, and the positions of myc epitope insertions are indicated by white triangles (N– and C–termini) and FLAG epitopes by black triangles (Loops 1 to 6).](image)

All eleven algorithms predicted TM domain 1, and all but one predicted domain 3, both in generally conserved positions (Figure 3.2). Either one large or two smaller TM domains were predicted by ten of the eleven prediction methods in the domain 5–6 region. Only
### Table 3.1: Summary of the output of the eleven transmembrane domain prediction algorithms when queried with the amino acid sequence of *D. melanogaster* Or22a. The amino acid positions and total number of predicted TMs are presented for each algorithm.

<table>
<thead>
<tr>
<th>Method</th>
<th># of TMs</th>
<th>Start</th>
<th>Finish</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMPred</td>
<td>7</td>
<td>50</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>83</td>
<td>105</td>
<td></td>
</tr>
<tr>
<td></td>
<td>136</td>
<td>164</td>
<td></td>
</tr>
<tr>
<td></td>
<td>197</td>
<td>222</td>
<td></td>
</tr>
<tr>
<td></td>
<td>267</td>
<td>290</td>
<td></td>
</tr>
<tr>
<td></td>
<td>289</td>
<td>315</td>
<td></td>
</tr>
<tr>
<td></td>
<td>354</td>
<td>377</td>
<td></td>
</tr>
<tr>
<td>Predator</td>
<td>7</td>
<td>46</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>106</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td></td>
<td>181</td>
<td>202</td>
<td></td>
</tr>
<tr>
<td></td>
<td>207</td>
<td>230</td>
<td></td>
</tr>
<tr>
<td></td>
<td>237</td>
<td>260</td>
<td></td>
</tr>
<tr>
<td></td>
<td>274</td>
<td>293</td>
<td></td>
</tr>
<tr>
<td></td>
<td>369</td>
<td>394</td>
<td></td>
</tr>
<tr>
<td>RbDe</td>
<td>7</td>
<td>50</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>87</td>
<td>107</td>
<td></td>
</tr>
<tr>
<td></td>
<td>137</td>
<td>157</td>
<td></td>
</tr>
<tr>
<td></td>
<td>183</td>
<td>203</td>
<td></td>
</tr>
<tr>
<td></td>
<td>264</td>
<td>280</td>
<td></td>
</tr>
<tr>
<td></td>
<td>287</td>
<td>304</td>
<td></td>
</tr>
<tr>
<td></td>
<td>357</td>
<td>377</td>
<td></td>
</tr>
<tr>
<td>SPLIT</td>
<td>6</td>
<td>50</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>145</td>
<td>162</td>
<td></td>
</tr>
<tr>
<td></td>
<td>197</td>
<td>219</td>
<td></td>
</tr>
<tr>
<td></td>
<td>266</td>
<td>291</td>
<td></td>
</tr>
<tr>
<td></td>
<td>297</td>
<td>320</td>
<td></td>
</tr>
<tr>
<td></td>
<td>362</td>
<td>386</td>
<td></td>
</tr>
<tr>
<td>HMMTOP</td>
<td>6</td>
<td>44</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>83</td>
<td>107</td>
<td></td>
</tr>
<tr>
<td></td>
<td>138</td>
<td>160</td>
<td></td>
</tr>
<tr>
<td></td>
<td>191</td>
<td>215</td>
<td></td>
</tr>
<tr>
<td></td>
<td>264</td>
<td>288</td>
<td></td>
</tr>
<tr>
<td></td>
<td>363</td>
<td>387</td>
<td></td>
</tr>
<tr>
<td>TMAP</td>
<td>5</td>
<td>44</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>137</td>
<td>163</td>
<td></td>
</tr>
<tr>
<td></td>
<td>262</td>
<td>282</td>
<td></td>
</tr>
<tr>
<td></td>
<td>298</td>
<td>309</td>
<td></td>
</tr>
<tr>
<td></td>
<td>364</td>
<td>387</td>
<td></td>
</tr>
<tr>
<td>TOPPRED</td>
<td>5</td>
<td>49</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>88</td>
<td>108</td>
<td></td>
</tr>
<tr>
<td></td>
<td>145</td>
<td>165</td>
<td></td>
</tr>
<tr>
<td></td>
<td>284</td>
<td>304</td>
<td></td>
</tr>
<tr>
<td></td>
<td>357</td>
<td>377</td>
<td></td>
</tr>
<tr>
<td>DAS</td>
<td>5</td>
<td>50</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>143</td>
<td>159</td>
<td></td>
</tr>
<tr>
<td></td>
<td>206</td>
<td>215</td>
<td></td>
</tr>
<tr>
<td></td>
<td>265</td>
<td>304</td>
<td></td>
</tr>
<tr>
<td></td>
<td>377</td>
<td>387</td>
<td></td>
</tr>
<tr>
<td>TMHMM</td>
<td>4</td>
<td>50</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>145</td>
<td>164</td>
<td></td>
</tr>
<tr>
<td></td>
<td>263</td>
<td>285</td>
<td></td>
</tr>
<tr>
<td></td>
<td>295</td>
<td>314</td>
<td></td>
</tr>
<tr>
<td>SOSUI</td>
<td>3</td>
<td>48</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>191</td>
<td>203</td>
<td></td>
</tr>
<tr>
<td></td>
<td>275</td>
<td>297</td>
<td></td>
</tr>
<tr>
<td>PredTMR2</td>
<td>3</td>
<td>50</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>145</td>
<td>164</td>
<td></td>
</tr>
<tr>
<td></td>
<td>273</td>
<td>290</td>
<td></td>
</tr>
</tbody>
</table>
3.3 Results

TMPred, RbDe, HMMTOP, TOPPRED and DAS predicted a TM domain in region 2, but the predicted regions were similar. TM 7 was predicted by all algorithms except TMHMM, Sosui and PredTMR2, and like TM 2 these were generally in similar positions for each algorithm. Variable predictions were seen in the TM 4 region; TMPred, SPLIT, HMMTOP and Sosui predicted a single region, whereas Predator and RbDe predicted two TM domains.

Only three prediction algorithms assigned seven TM domains: TMPred, Predator and RbDe. The consensus model was based on the prediction given by the TMPred, as the domains it predicted were generally individually predicted by other independent methods. However, TMPred has a problem in the TM 5–6 region as it predicts two TM domains that overlap. As this is very unlikely to occur, the putative positions of TM 5 and 6 were based on those predicted by SPLIT, as they are very similar to those predicted by both the TMHMM and RbDe programs.

In order to experimentally verify this model, Or22a constructs were synthesised that contained myc epitopes on the N– and C–termini. FLAG epitopes were also inserted into each putative hydrophilic loop region of a C–terminally myc–tagged Or22a to facilitate double–labelling experiments. Initial attempts to confirm the expression of these constructs by western blotting on membrane proteins isolated using a differential centrifugation protocol failed. An immunoreactive mass was later observed to be present at the very top of the protein gel, indicating that the proteins were not sufficiently solubilised and were unable to migrate into the gel. Denaturation of the proteins at either 99 °C or 37 °C yielded the same result. Extraction of the ORs using Triton X–114 phase separation (Bordier, 1980) proved successful for all constructs except C–terminally myc–tagged Or22a (Figure 3.3). Immunoreactive bands were visible running slightly faster than the 50 kDa band of the molecular weight standard. This corresponds well with a predicted mass of Or22a with epitopes of 49.6 kDa.

Epitope–tagged constructs were assessed for function using calcium imaging in Sf9 cells as described in Chapter 2. N–terminal myc fusions of Or22a were active, but all C–terminal myc fusions were inactive (Figure 3.4). Once the C–terminal myc–tag was removed from
the hydrophilic loop FLAG fusions using the PCR, the activity of these receptors was not significantly different to wild-type Or22a (Figure 3.4). *E. postvittana* Or1 was also myc–tagged, and was tested with the odorant citral (Jordan, 2006). N–terminally myc–tagged Or1 was active, but the C–terminally myc–tagged version was not (Figure 3.4).

![Figure 3.3: Analysis of the expression of epitope–tagged Or22a in S2 cells by western blotting. A) The expression of empty vector transfected S2 cells (Lane 2), N–terminally myc–tagged Or22a (Lane 3) and C–terminally myc–tagged Or22a (Lane 4) using an anti–myc antibody. B) The expression of Or22a with FLAG epitopes inserted into the putative hydrophilic loops one to six (Lanes 2–7) using an anti–FLAG antibody. Lane 1 in both Panel A and Panel B contains the MagicMark XP molecular weight standards (major band at 40 kDa, with the next slowest migrating band having a molecular weight of 50 kDa).](image)

In order to have confidence in the determination of orientation by epitope tagging in insect cells, a construct consisting of the N–terminal loop, the first TM domain and the first intracellular loop of rhodopsin was synthesised (RhoTM1). A similar construct had previously been used as a positive control in an *in vivo* topological analysis as it has a known topology (Benton *et al.*, 2006), with an extracellular N–terminus and an intracellular C–terminus. Expression of N–terminal and C–terminal myc–tagged RhoTM1 constructs was used to optimise fixative type and antibody staining conditions. Expression of these constructs in both Sf9 and S2 cells was trialled, with S2 cells showing lower autofluorescence at the wavelengths of the filter set used to detect the AlexaFluor 488–conjugated secondary antibody. When expressed in S2 cells, the RhoTM1 construct demonstrated the known topology of an extracellular N–terminus and an intracellular...
C-terminus (Figure 3.5). No antibody staining was seen in empty–vector transfected S2 cells.

![Diagram](image)

**Figure 3.4:** Mean responses of epitope–tagged versions of Or22a to $10^{-5}$ M ethyl butyrate (Panel A) and OR1 to $10^{-5}$ M citral (Panel B) as monitored using calcium imaging in Sf9 cells. “N–Term” and “C–Term” represent Sf9 cells expressing N– and C–terminal myc–tagged receptors, “L1” to “L6” represent Sf9 cells expressing Or22a–FLAG fusions in putative hydrophilic loops one to six. “WT” represents wild–type receptor, and “–ve” represents empty–vector transfected Sf9 cells. Activity is given as a $\Delta F$ value, which is the fluorescence increase observed on addition of the substrate as a proportion of the maximal fluorescence elicited by ionomycin. Values given as the mean $\Delta F \pm$ S.E.M. for two replicate experiments of three wells each. For Or22a experiments, “a” indicates no significant difference to wild–type, whereas “b” indicates no significant difference to the empty vector negative control (two–tailed t–test with Bonferroni correction for multiple testing, $p < 0.05$).

The preparation of samples for antibody labelling produced consistent results with RhoTM1 constructs, and so was applied to the determination of the topology of two ligand–sensitive ORs: *D. melanogaster* Or22a and *E. postvittana* Or1. Each of these receptors was fused to a myc epitope at either the N– or C–terminus and expressed in S2 cells. The N–terminal myc epitopes on both Or22a and Or1 were only accessible to the antibody in the presence of the detergent saponin, indicating an intracellular localisation, whereas the C–terminal epitopes were antibody–accessible irrespective of the presence of saponin (Figure 3.6). This indicates that these receptors share the same orientation as Or83b (Lundin et al., 2007).

Insect ORs have been assumed as having seven TM domains, first based on the hypothesis
3.3 Results

Figure 3.5: Fluorescence microscopy images showing the orientation of rhodopsin TM1 controls. S2 cells expressing rhodopsin TM1 constructs myc–tagged at either the N– or C–terminus were antibody labelled in the presence (Perm.) or absence (Non–perm.) of the detergent saponin. Green fluorescence is from the AlexaFluor 488–conjugated secondary antibody and indicates antibody binding, and blue fluorescence is the nuclear dye DAPI. Empty–vector transfected S2 cells were similarly stained to show there is no non–specific binding of the antibodies. Schematic diagrams indicate the position of the epitope. Scale bar represents 20 µm.

Figure 3.6: Fluorescence microscopy images showing the orientation of D. melanogaster Or22a and E. postvittana Or1. S2 cells expressing each OR tagged with a myc epitope at the N– or C– terminus and incubated with antibodies in both the presence (Perm.) and absence (Non–perm.) of the detergent saponin. Green fluorescence indicates binding of the AlexaFluor 488–conjugated secondary antibody, blue fluorescence indicates the presence of cell nuclei (DAPI). Scale bar represents 20 µm.
that they are GPCRs, and later on computational predictions of the number of TM domains (Benton et al., 2006). However, this had yet to be thoroughly experimentally tested. FLAG epitopes were inserted into the each putative hydrophilic loop region, and their orientation relative to the cell membrane assessed. FLAG epitopes inserted in the first, third and fifth putative hydrophilic regions were accessible to anti–FLAG antibodies in both the presence and absence of saponin, whereas the second, fourth and sixth putative hydrophilic regions were only accessible in the presence of saponin (Figure 3.7). These results indicate that Or22a has at least seven TM domains. An experimentally supported model of the transmembrane topology of Or22a is presented in Figure 3.8.

**Figure 3.7:** Fluorescence microscopy images of epitope–tagged versions of Or22a in the loop regions. S2 cells expressing Or22a constructs FLAG–tagged in each of its predicted hydrophilic regions (Loop 1 (L1) to Loop 6 (L6)) were antibody labelled in the presence (Perm.) or absence (Non–perm.) of the detergent saponin. Green fluorescence indicates binding of the AlexaFluor 488–conjugated secondary antibody, blue fluorescence indicates the nuclear stain DAPI. Scale bar represents 20 µm.

### 3.4 Discussion

GPCRs have seven TM domains with an extracellular N–terminus (Pierce et al., 2002). Vertebrate ORs are GPCRs and share this seven TM topology (Malnic et al., 1999). However, insect ORs were suggested to share the seven TM domains characteristic of
Figure 3.8: Experimentally verified model of the transmembrane topology of Or22a. This model was developed using prediction algorithms and verified using epitope-labelling experiments.
GPCRs but have an intracellular N–terminus (Benton et al., 2006). This topology was later experimentally supported by glycosylation scanning mutagenesis experiments on the chaperone receptor Or83b (Lundin et al., 2007). The research presented in this chapter provides further evidence that ligand–binding insect ORs share the intracellular N–terminus motif. This result raises the question of whether insect ORs signal using G proteins or signal through a different pathway altogether. Further research to identify the signal transduction pathway used by insect ORs can now be targeted to the internal loops of the ORs.

The research presented here provides the first formal test of the hypothesis that insect ORs contain seven membrane spanning domains by examining which side of the cell membrane epitopes inserted into the hydrophilic domains reside. Lundin et al. (2007) were unable to determine the localisation of the putative extracellular loops 1 and 3 of Or83b due to their small size and the inherent drawbacks of the glycosylation scanning mutagenesis technique. The assignment of the loops as extracellular therefore remains uncertain. Moreover, the effect the insertion of glycosylation sites has on the function of Or83b was not assessed. The mutations required to insert glycosylation sites are not as extensive as those required to insert epitopes, but the possibility remains that regions of the protein important for correct membrane insertion and folding were affected such that the topology was altered.

An intracellular N–terminus is also a feature of the olfactory receptor Or1 from the lepidopteran species E. postvittana. This indicates that this orientation is conserved right across these orders and through at least 250 million years of evolution since the split between lepidoptera and diptera (Jones et al., 2005). It is therefore likely that this topology is associated with species whose olfactory systems utilise an Or83b–like receptor, and that these species signal using a conserved mechanism. Analysis of species of further evolutionary divergence from dipterans/lepidopterans will further elucidate how widespread this topology and hence signalling mechanism is across evolutionary time.

The adiponectin receptor family shares the same seven TM topology and intracellular N–terminus as insect ORs (Yamauchi et al., 2003). Adiponectin receptors share very
little sequence similarity with GPCRs and signal independently of G proteins (Yamauchi et al., 2003). There are a number of adiponectin receptor homologues from distantly related species that share this topology, including yeast (Karpichev et al., 2002), plants and insects (Hsieh & Goodman, 2005), which indicates an ancient evolutionary origin. Protein structural features are often more conserved than nucleotide sequence, but the adiponectin receptors do not signal through a similar pathway to insect ORs and are more likely to have independently acquired the same orientation.

No single TM domain algorithm predicted all of the TM domains that were supported by the immunochemistry experiments. Similarly, in studies assessing the accuracy of TM prediction algorithms, there was not one algorithm that performed well across the many different types of TM protein used as a test data set (Chen et al., 2002; Cuthbertson, 2005). However, some algorithms performed accurately for many of the test protein families: Split, HMMTOP, TMHMM and TMAP (Cuthbertson, 2005). These assessment studies utilised predictions of the same protein structures as were used to train the prediction algorithms in assessing the accuracy of the predictions. It is therefore unclear how they would behave when presented with sequences for which there is no closely related proteins of known structure in the protein data bank. It is notable that none of the four better performing algorithms predicted a TM region in Or22a that was not supported by experimental data. This suggests that a consensus approach amongst these four algorithms may generate more accurate topological models of insect ORs, including their orientation, rather than using the algorithms singularly.

The model presented in Figure 3.8 relies on the prediction of TM domains to precisely identify the beginning and end of each TM domain. However, it is the boundaries of the TM domains that vary the most between prediction algorithms, even for the TM domains that were predicted by the majority of programs (e.g. TM1 and TM7). The experimental verification of the exact boundaries of the TM regions for Or22a or Or83b will enable further analysis of the structural features of these receptors. One possible technique is proteolytic cleavage of these proteins in cell membranes at high pH (Wu et al., 2003). Under these conditions the membrane forms a flat sheet, enabling cleavage of the
hydrophilic regions while leaving the TM domains in the membrane phase. Analysis of the
subsequent cleaved peptides by mass spectrometry will determine the borders of the TM
regions. This technique may identify the presence of any re-entrant pore loops that are
not clearly defined using either the glycosylation scanning or epitope tagging approaches.
This technique may yield an efficient high throughput analysis of the topology of a large
number of ORs.

Sf9 cell calcium assays were used to test the hypothesis that the insertion of an epitope
into specific hydrophilic regions had no effect on receptor function, and by extension,
receptor structure. All of the epitope–tagged OR constructs were functional, except any
construct that had a C–terminal fusion. The inactivity of C–terminally epitope–tagged
ORs has been previously noted (Benton et al., 2006). However, Lundin et al. (2007) assign
the C–terminus of Or83b as extracellular after attaching a 25 amino acid linker containing
a glycosylation site. These results suggest that the assignment of the C–terminal region
as extracellular must be made with some caution. Confirmation of the localisation of
the C–terminus of both the co–receptor Or83b and the ligand–binding ORs using a non–
mutagenic technique is required to confirm that the seventh putative TM domain traverses
the membrane.

The lack of activity of insect ORs with a C–terminal fusion also suggests that this region
is functionally important. Indeed, the C–terminus is among the most conserved region of
insect ORs (Clyne et al., 1999). This domain may be important for the interaction of ORs
in the formation of multimers, as BRET between Or22a homodimers and Or22a/Or83b
heterodimers is observed when they are tagged on the C–termini with luciferase and GFP
(Neuhaus et al., 2005). The functional importance of this interaction is as yet unknown,
though it may be important for the correct targeting and transport of the OR:Or83b
complex to the cell membrane. Alternatively, it may form part of the mechanism by
which a ligand–sensitive OR transduces an activation signal to the channel formed by
Or83b.

While the topology data as presented in the thesis shows single images to convey a negative
result, it is important to appreciate that many hundreds of individual cells were in fact
analysed during the course of these studies. Although no quantitative data was collected, this does give greater confidence in the results. Or22a constructs with FLAG epitopes inserted into the hydrophilic loop domains were initially constructed with C–terminal myc tags to facilitate double–labeling experiments. This would have increased confidence in the interpretation of negative results, as it would reduce analysis to only those cells expressing Or22a fusion constructs as well as providing an internal control for the immunolabeling reactions. Despite these shortcomings, the presentation of topology data in a similar manner to that presented here is widely accepted in the literature (see Covitz et al. (1998), Smolka et al. (1999), Lewis et al. (2001) and Yasuada et al. (2003) for example).

Since the research described here was conceived and conducted, two papers have been published that provide evidence that insect ORs form ligand–gated ion channels (Sato et al., 2008; Wicher et al., 2008). This suggests that adiponectin receptors and insect ORs have independently acquired the 7TM motif with intracellular N–termini. However, this topology has yet to be observed in other ion channels. Voltage gated ion channels generally share a six TM topology, with each functional unit made up of four domains (see Armstrong and Hille (1998) for a review). Ligand–gated ion channels have varying topologies, with most having between two and six TM domains (Landry & Gies, 2008). Some ligand–gated ion channels have seven TM domains, such as channel rhodopsin (Nagel et al., 2003), though these are predicted to have the classical GPCR orientation rather than that seen for insect ORs. It is thus likely that insect ORs form a new class of ion channel.

In addition to their activity as ion channels, insect ORs are thought to activate G\textsubscript{as} proteins in HEK293 cells and signal using cAMP as a second messenger (Wicher et al., 2008), though no evidence for this was seen in a similar study (Sato et al., 2008). Inhibitors of G proteins do not stop insect OR signalling in the HEK293 cell line, though there were differences observed in the rectification of the signal after activation and only high concentrations of odorant were applied (Smart et al., 2008). This contrasts with the finding that mutation of G\textsubscript{aq} and phospholipase C proteins causes a significant reduction in antennal signalling \textit{in vivo} (Kain et al., 2008). It is possible that a slower G protein–
mediated response is a component of insect OR signalling, though this metabotropic component of signalling has yet to be fully explored. The majority of the conserved regions of insect ORs occur in what is now the experimentally confirmed intracellular loops (Clyne et al., 1999). Confirmation of the topology of the ligand–sensitive ORs enables mutagenesis studies aimed at teasing apart the G protein–activating and Or83b–activating mechanisms of the ligand–sensitive ORs.

One feature of some ion channels, such as the potassium channels, is the presence of a loop domain that enters but does not traverse the membrane (Durrell et al., 1998). This pore loop is an important determinant of the permeability of the channel. A motif in the third intracellular loop of Or83b, TVVGYLG, has been implicated as part of the selectivity filter of the Or83b channel (Wicher et al., 2008). The third intracellular loops of Or22a and Or43a have been shown to interact with the third intracellular loop of Or83b (Benton et al., 2006). This interaction may indicate that these domains make up the pore of the OR ion channel structure. If, on the other hand, the OR channel is made up entirely of Or83b subunits, this interaction may mediate the opening and closing of the pore upon ligand binding through OR:Or83b interaction.
4

Investigation of Olfactory Receptor Oligomerisation

4.1 Introduction

Protein–protein interactions are fundamental to all biological processes. They are involved in the regulation of cell growth, DNA replication, gene expression and numerous other important activities (Kerrien et al., 2007). Protein–protein interactions can be stable over time, allowing the formation of protein complexes such as ATP synthase (Devenish et al., 2008), or they can be transient, such as the interaction of an activated G protein with phospholipase C in signal transduction (Helmreich, 2001). Identifying the targets and mechanisms through which a protein interacts with others can therefore provide a great deal of insight into its function.
4.1 Introduction

4.1.1 Techniques for investigating protein–protein interactions

Interactions between proteins have a wide range of potential affinities, from very high affinity in protein complexes, to the lower affinities observed for dynamic regulatory events (Lalonde et al., 2008). Thus a range of techniques are used to assess protein–protein interactions. Early experiments used co-localisation of fluorescence in immunochemistry experiments to imply interaction. This proved unsatisfactory, as the resolution of such experiments is measured in hundreds of nanometers, whereas protein–protein interactions occur on a scale of several nanometers (Piston & Kremers, 2007). Affinity purification coupled to mass spectrometry is one technique used to demonstrate proteins are interacting, and are not just in the same vicinity. It relies on the capture of a target protein and its interacting partners using an antibody. Non-interactive proteins can then be removed from the sample by washing the column, prior to elution of the bound proteins. The eluate is then proteolytically digested and analysed using mass spectrometry in order to identify the proteins interacting with the target (Rigaut et al., 1999).

Co-immunoprecipitation of proteins can also be used to identify interactions with a similar principle to affinity purification. An antibody for a target protein is added to a cell lysate and the antibody along with interacting proteins is purified. Polyacrylamide gel electrophoresis followed by mass spectrometry can then be used to identify interacting partners, or specific antibodies to putative partners can be used to show interaction by western blotting (Hebert et al., 1996). Direct crosslinking of proteins in close proximity using reagents that form covalent bonds between proteins can also be used to show interaction. One example is the use of disulphide trapping, which uses cupric orthophenanthroline to create disulphide bonds between cysteine residues of neighbouring proteins (Klco et al., 2003).

Yeast two-hybrid (Y2H) systems utilises two proteins of interest each tagged with either the DNA binding or activation domain of a yeast transcription factor. If the two proteins interact, a functional transcription factor will be reconstituted which can go on to activate
expression of a reporter gene (Uetz et al., 2000). This system requires that the proteins of interest are able to be transported to the nucleus to activate reporter gene expression, and is therefore of no use for investigating interactions between membrane or otherwise anchored proteins. The mating–based Split–Ubiquitin system is a development of the Y2H system that overcomes this drawback (Obrdlik et al., 2004). The two proteins of interest are each tagged with one domain of ubiquitin with a transcription factor included. Functional complementation of the ubiquitin protein upon interaction of the tagged proteins targets it for proteolytic degradation, releasing the transcription factor to enter the nucleus and activate reporter gene expression.

Protein–protein interactions can be assessed using bimolecular fluorescence complementation (BiFC). In this technique each half of a fluorescent protein such as GFP is fused to two proteins of interest. In this state, neither protein is fluorescent. If the two proteins of interest interact, the two halves of GFP interact and form a fluorescent moiety which can be visualised. BiFC has been used to study static protein–protein interactions (Ghosh et al., 2000), as well as dynamic interactions in protein folding (Cabountos et al., 2005). It has the advantage over some of the above techniques in that it not only allows detection of a protein–protein interaction, but it also enables visualisation of the cellular compartment that the interaction is occurring in (Lalonde et al., 2008).

Resonance energy transfer (RET) occurs when energy is transferred in a non–radiative manner from a donor to an acceptor molecule. In the case of fluorescent proteins, an excited donor molecule can transfer energy to an acceptor molecule with an overlapping excitation spectra (Piston & Kremers, 2007). This excitation results in the emission of fluorescence at the protein’s characteristic emission wavelength. This process is known as fluorescence resonance energy transfer (FRET). It can also occur when a bioluminescent enzyme, such as luciferase, is used as an energy donor, in which case it is termed bioluminescence resonance energy transfer (BRET).

RET is highly dependent upon the distance between and the relative orientations of the fluorophores, only occurring between fluorophores within 10 nm of each other (Wu & Brand, 1994). Protein–protein interactions can be detected by tagging each protein with
either a RET donor, such as CFP, or an acceptor, such as YFP. Interaction between the proteins can bring the fluorophores within 10 nm, causing CFP to transfer energy to YFP (Figure 4.1). FRET has been used in a number of protein–protein interactions studies, such as the identification of the domains that interact in the formation of α1b–adrenoreceptor oligomers (Carrillo et al., 2004), and the ligand–induced dissociation of metabotropic glutamate receptor dimers (Tateyama et al., 2004). FRET has also been used to analyse ion channel subunit stoichiometry (Amiri et al., 2003).

Figure 4.1: Principle of fluorescence resonance energy transfer (FRET) in investigating protein–protein interactions in and between membrane proteins. FRET is the non–radiative transfer of energy between two fluorophores in close proximity (≤10 nm) and whose emission and excitation spectra overlap. In this example, FRET occurs between cyan and yellow fluorescent proteins (CFP and YFP, respectively). If the fluorophores are too far apart, excitation of the CFP at 436 nm will result in emission from CFP at 480 nm and none from YFP at 535 nm, as no energy is transferred. If a protein–protein interaction or conformational shift results in the two fluorophores coming close together, a greater portion of the CFP energy is transferred and increased YFP emission is observed at 535 nm.

4.1.2 Approaches for the measurement of FRET

There are five general approaches for the measurement of FRET in biological systems. The first, sensitised emission, uses emission filters to measure both donor and acceptor molecule emission upon excitation of the donor (Gordon et al., 1998). However, due to the significant overlap between the excitation and emission spectra of the donor and acceptor as required for FRET, multiple controls and image processing are required to determine genuine FRET signals from background noise. Dynamic FRET experiments can be performed once the system is calibrated. The second technique, fluorescence
4.1 Introduction

lifetime imaging (FLIM), utilises changes in the decay of fluorescence molecules on a nanosecond timescale caused by energy transfer (Ng et al., 1999). FLIM is very precise, but measurement of fluorescence lifetime requires specialised equipment, and can be a slow procedure to establish.

The third technique for measuring FRET uses fluorescence polarisation. When excited with polarised light, different degrees of FRET will cause light to be emitted at different angles. The ratio of the emission at different angles indicates the degree of FRET. This technique allows high signal-to-noise data to be collected, is very fast and can be done with minimal data processing, making it suitable for high throughput screening approaches. However, polarized FRET experiments are not able to accurately quantify FRET, and is instead applicable to detecting the presence or absence of FRET in a sample.

Spectral imaging is a fourth technique for FRET detection. Instead of using multiple channels as for sensitized emission, the entire spectrum containing the donor and acceptor emission on excitation of the donor is measured (Rizzo et al., 2006). Using linear unmixing of the spectra, the contribution of both donor and acceptor molecules can be assessed, and the amount of FRET occurring can be calculated from the difference. This technique is suitable for both spectroscopic and microscopic FRET experiments, but can suffer from poor signal to noise ratios due to the processing required. Acceptor photobleaching, or donor dequenching, is a fifth technique for measuring FRET. The amount of FRET is measured by destroying the acceptor molecule’s ability to absorb energy and measuring the subsequent increase in donor emission due to the fact that FRET can no longer occur (Kenworthy & Edidin, 1998). This technique is simple, but destructive, and is therefore unsuitable for the measurement of dynamic FRET.

4.1.3 Olfactory receptor oligomerisation

As discussed in Section 1.4, it is increasingly being realised that the oligomerisation of GPCRs is an important aspect of receptor functionality. Indeed, higher order oligomers have been identified in GPCRs such as rhodopsin (Fotiadis et al., 2003). Oligomerisation
is also functionally important for ion channels, as many channels are made up of multiple subunits in order to transport substrates through a central pore. For example, the transient receptor potential C channels form a tetramer (Amiri et al., 2003), as do the acid–sensing ion channels (Gao et al., 2007).

The *Drosophila* co–receptor Or83b is essential for olfaction in vivo, as ligand sensitive ORs are not correctly targeted to the dendritic membrane in its absence (Larsson et al., 2004). A BRET approach was used to demonstrate that Or83b heterodimerises with the ligand sensitive OR, Or43a, when expressed in HEK293 cells, and that Or43a is also able to homodimerise (Neuhaus et al., 2005). These constructs were fused to GFP or luciferase at the C–terminus, and no evidence was provided that these constructs were functional. As described in Chapter 3 and in Benton et al. (2006), no OR fusions at the C–terminus are functional. BiFC was later used to corroborate the presence of OR:Or83b heterodimers and Or83b:Or83b homodimers using functional N–terminal YFP fusions in vivo (Benton et al., 2006). However, homodimerisation of ligand–sensitive ORs was not observed in the absence of Or83b. Additionally, Y2H analysis demonstrated that the third intracellular loops of both Or43a and Or22a interact with the third intracellular loop of Or83b (Benton et al., 2006). The extent of these interactions between ORs is yet to be fully investigated, though these studies suggest that the N– and C–termini, as well as the third intracellular loop region are all in close proximity with each other in insect OR oligomers.

### 4.1.4 Aims

The research presented in this chapter aims to test whether insect ligand–binding ORs form homomeric interactions, and if so, which TM domains are interacting when the OR is in the cell membrane. FRET has been chosen to address this question, as a similar approach has been utilised for insect ORs previously (Neuhaus et al., 2005), and N–terminal fluorescent protein fusions of insect ORs have been observed to be functional (Benton et al., 2006). Fluorescent proteins inserted into the intracellular domains of Or22a will be used to monitor which domains are in close proximity when co–expressed in
4.1 Introduction

a heterologous system. If the two tagged domains are in close proximity, energy transfer will occur between the fluorophores and a FRET signal will be observed. Should the fluorescent proteins be further apart than the FRET radius, no FRET signal will be observed, and it can be concluded that the domains are not in close proximity in homo–oligomers.

Or22a has been chosen as the model receptor for this work as its TM topology is now understood (see Chapter 3). The insertion of fluorescent proteins can be targeted to regions of known localisation, and there is evidence that insertion in these domains does not majorly affect protein function. The insertion of fluorescent proteins in the intracellular loops of Or22a can therefore be used to show which loops are in close proximity in OR multimers, and by extension, which TM domains may be interacting in the cell membrane.

There is evidence that ORs form homo– and hetero–oligomers both in vitro and in vivo as described in Section 4.1.3. It is therefore hypothesised that Or22a will form homomeric interactions when expressed in S2 cells, as many other receptors and ion channels do. It may form these interactions generally, with many possible interacting domains, or between specific interaction interfaces. Alternatively, Or22a may only interact with Or83b, in which case no homomeric interactions will be observed. The observation of multiple interacting interfaces will also be evidence for a quaternary structure of more than two subunits, whereas one specific interface will support the hypothesis that insect ORs form a dimeric quaternary structure.

The third intracellular loop has previously been observed to interact in heteromers of ligand–binding ORs with Or83b (Benton et al., 2006), and it is hypothesised that this interaction domain is also utilised in homomeric interactions. FRET will be used to determine whether this interaction also occurs in Or22a homomers in vitro. FRET experiments will be performed by expressing a FRET pair with and without the additional expression of Or83b in order to test whether the presence of Or83b impacts on any observed homomeric interactions. One pair that exhibits FRET between the CFP and YFP fusions and one that does not will be chosen for this preliminary experiment. If these
same interaction interfaces are also used in the interaction with Or83b, then a reduced level of FRET between tagged Or22a subunits should be observed.

4.2 Materials and Methods

4.2.1 Materials

Fluorodish glass-bottomed culture dishes were purchased from World Precision Instruments. Plasmids containing CFP and YFP (pECFP and pEYFP) were sourced from Clontech Laboratories. Other materials are as described previously in Sections 2.2.1 and 3.2.1.

4.2.2 Cell culture and transfection

Serum–free adapted S2 cells were cultured as described in Section 2.2.3. 1 x 10^6 cells were seeded into poly-L-lysine coated glass-bottomed FluoroDish culture dishes for imaging experiments and transfected, as for six–well plates, as described in Section 2.2.4. Co–transfection of FRET constructs utilised 1 µg of each expression plasmid with 12 µL of Escort IV transfection reagent for each Fluorodish required.

4.2.3 RT–PCR

RT–PCR on cDNA extracted from 1 x 10^6 S2 cells was performed as described in Section 2.2.5. The Or83b–specific primers had the sequence 5’–GACAACCTCGATGCCGGAGC–3’ and 5’–CGAGGCTGAACCTCAACCG–3’. They were designed to generate a 750 bp fragment of Or83b. The positive control for the PCR consisted of pIB-Or83b with the pIB plasmid–specific primers OpIE2 forward and reverse primers (see Section 2.2.5 for sequences). Unfortunately, no control reaction for the Or83b–specific primers was carried out on a known Or83b gene.
4.2.4 Construction of Or22a fusions with fluorescent proteins

The expression vectors containing the fluorescent proteins CFP and YFP (pIB-CFP and pIB-YFP) were constructed by cutting enhanced CFP and enhanced YFP from the vectors pECFP and pEYFP with the restriction enzymes *Not* I and *Bam*H I and ligating them into pIB/V5 that had been cut using the same enzymes. All cloning procedures were carried out as described in Section 2.2.2.

The expression vector containing CFP and YFP fused in the same open reading frame, pIB-CFP-YFP, was constructed by using PCR to insert an *Nhe* I site onto the 5’ end of the CFP gene in pIB-CFP to generate pIB-CFP(Nhe) using the primers 5’–CGG-GTACCGGTGCCACCATGGCTAGCGTGAGCAAGGGCG–3’ and 5’–GACAATACAAAC-TAAGATTGTCAG–3’. The PCR products were sub-cloned into pGem–T Easy and sequenced, before being shuttled back into pIB/V5 using the *Kpn* I and *Sac* II sites included in the primers. *Nhe* I sites were added onto the 5’ and 3’ ends of YFP and the stop codon removed to create pIB-(Nhe)YFP(Nhe) using the primers 5’–GCTAGC-GTGAGCAAGGGCGAGGAGC–3’ and 5’–GCTAGCCTTGTACAGCTCGTCCATGCCG–3’. This product was sub-cloned into pGem–T Easy and sequenced, before being cut with *Nhe* I and ligated into pIB-CFP(NheI). The sequence of the final pIB-CFP-YFP clone was verified prior to use in experiments.

N–terminally YFP–tagged Or22a was constructed by using PCR to add an *Nhe* I site onto the 5’ end of the Or22a gene using the primers 5’–GCGCGGTACCATGGCTAGCTTA-AGCAAGTTTTTCCACATAA–3’ and 5’–GACAATACAAACTAAGATTTAGTCAG–3’. This PCR product was subcloned into pGem–T Easy and sequenced, before the fragment (Nhe)YFP(Nhe) was inserted as above. The final pIB-NYFP-Or22a clone was sequenced prior to expression.

Insertion of CFP and YFP into the intracellular loops of Or22a was carried out utilising the unique restriction sites flanking the FLAG epitopes in the hydrophilic loops of Or22a used for topology analyses and described in Section 3.2.3. *Sph* I sites flank the FLAG epitope in the first intracellular loop (I1), *Afl* II in the second intracellular loop (I2) and *Xba* I in
the third intracellular loop (I3). Primers were designed that added these sites onto CFP and YFP whilst removing the start and stop codons from them. These primers had the sequence 5’–GGTCGCCACCCXXXXXGTGAGCAAGGGGCG–3’ and 5’–GCGGCCGCTXXX-XXXCTTGACAGCTCGTC–3’, where XXXXXX represents the sequence of the restriction site required (GCATGC for Sph I, CTCAAG for Afl II and TCTAGA for Xba I).

CFP and YFP PCR products generated using these primers were subcloned into pGem–T Easy and sequenced, before being shuttled into the appropriately cut pIB-Or22a-FLAG clone. Orientation of the fluorescent protein was checked by PCR using the primer 5’–GCACCATCTTCTTCAAGGACGACG–3’ and the OpIE2 reverse primer (see Section 2.2.5). Clones that appeared to have the fluorescent protein inserted in the correct orientation were sequenced prior to use. This generated the clones pIB-Or22a-I1CFP, pIB-Or22a-I1YFP, pIB-Or22a-I2YFP, pIB-Or22a-I3CFP and pIB-Or22a-I3YFP. Unfortunately, no clone of pIB-Or22a-I2CFP has yet been made, although not through lack of trying.

4.2.5 Activity assays on fluorescent protein–tagged Or22a

Activity assays on fluorescent protein–tagged ORs were carried out using calcium imaging in Sf9 cells as described in Chapter 2, using $10^{-5}$ M ethyl butyrate as substrate. Assays were performed on wild–type Or22a in all experiments as a positive control, and empty vector transfected cells were used as a negative control.

4.2.6 Fluorescent microscopy and FRET imaging

Forty eight hours after transfection with a 1:1 ratio of two differentially tagged Or22a construct, S2 cells were washed three times with assay buffer and left immersed in a further 3 mL of assay buffer (see Section 2.2.1 for components). A Leica SP2 Laser Scanning Confocal microscope was used for living cell microscopy and FRET measurements. Cells were imaged with a 40x/0.80 NA HCX Apo L water immersion objective. CFP was
excited with the 458 nm and YFP with the 514 nm line of an argon laser. A 458/514 nm beam splitter was installed in the lightpath. Detection ranges were set to measure CFP between 460 and 514 nm and YFP between 525 and 625 nm. Gain and offset settings were maintained at levels such that minimal CFP or YFP signal was seen in non–transfected S2 cells, and kept constant for all imaging experiments.

Acceptor photobleaching FRET experiments were controlled through the Leica LSM software. Once appropriately co–transfected cells were detected, CFP and YFP images were acquired sequentially, with these and all subsequent images being averaged across three frames. YFP was photobleached using the 514 nm laser at 100% power for 50 frames, before further CFP and YFP images were acquired. FRET efficiency in the samples was assessed by calculating the increase in CFP fluorescence after destruction of the YFP fluorophore using Equation 4.1 on a pixel by pixel basis to generate a FRET efficiency image. FRET efficiency is given as a ratio, with a FRET efficiency of 1 implying complete transfer of CFP energy to YFP, and 0 implying no energy transfer. Mean FRET efficiency in S2 cells was calculated first by filtering the FRET efficiency image with a low–pass filter to remove background noise. Regions of interest were then drawn around cells co–expressing constructs containing fluorophores and the FRET efficiency of that cell measured. The FRET efficiency was calculated from 14–36 cells co–expressing the given fluorophore pair and expressed as the mean ± S.E.M.

\[
FRET_{\text{eff}} = \frac{D_{\text{post–bleach}} - D_{\text{pre–bleach}}}{D_{\text{post–bleach}}} 
\]  

(4.1)

### 4.3 Results

The fluorescent proteins CFP and YFP were inserted into the N–terminus and the intracellular loops of Or22a taking advantage of restriction sites engineered into Or22a during the construction of epitope–tagged versions for TM domain analysis. First, the structural and functional integrity was assessed using a functional assay. If recombinant
OR was still able to be activated by ethyl butyrate to increase intracellular calcium levels it was assumed to have maintained structural integrity. All Or22a–YFP fusions were functional, giving increased intracellular calcium levels in response to $10^{-5}$ M ethyl butyrate (Figure 4.2). The N–terminal YFP fusion gave the same response strength as wildtype Or22a, while YFP inserted into the first, second or third intracellular loop resulted in a receptor that gave a weaker response, but still significantly greater than the empty vector negative control (two–tailed t–test, $p < 0.05$).

![Figure 4.2](image.png)

**Figure 4.2:** Mean odorant response of fluorescent protein–tagged versions of Or22a exposed to $10^{-5}$ M ethyl butyrate as assessed using calcium imaging in Sf9 cells. WT represents wild–type Or22a, NY is YFP fused to the N–terminus, I1Y is YFP inserted into the first intracellular loop, I2Y is YFP inserted into the second intracellular loop and I3Y is YFP inserted into the third intracellular loop. The negative control is empty vector transfected Sf9 cells. Response is measured as mean $\Delta F \pm$ S.E.M. of responding cells from two separate experiments. “a” indicates not significantly different from wildtype Or22a but significantly different to the negative control, “b” indicates significantly different to both positive and negative controls, and “c” indicates significantly different to the positive control but not the negative control (two–tailed t–test, $p < 0.05$).

To assess which domains were in close proximity during the formation of possible homo–oligomers, combinations of different Or22a–CFP and Or22a–YFP fusion proteins were co–transfected into S2 cells and the amount of FRET measured using a donor dequenching upon acceptor photobleaching approach. After assessing the data with a one way analysis of variance followed by Tukey’s multiple comparison test, three distinct groups were observed (Figure 4.3). The positive control cells expressing a construct of CFP and YFP in the same open reading frame (CFP–YFP) had the highest FRET efficiency.
4.3 Results

(p values ranging from 0.0008 to 0.0021). A significant amount of FRET was observed when N–terminal YFP (NYFP) and I1–CFP, NYFP and I3–CFP, I1–CFP and I1–YFP, I1–CFP and I3–YFP, and I3–CFP and I3–YFP constructs were co–expressed. The FRET efficiencies for these combinations was significantly different to both the CFP–YFP positive control and the remaining combinations for which no FRET was observed (p values ranging from 0.0026 to 0.0338). The third group, exhibiting no FRET, was observed when CFP and YFP were co–expressed in different plasmids, and when I1–CFP and I2–YFP or I2–YFP and I3–CFP fusions were co–expressed (p values ranging from 0.7164 to 0.8844).

These results demonstrate the low expression of the OR constructs in comparison to the CFP–YFP positive control construct. This created problems testing cells for FRET, as only cells expressing both constructs could be analysed. Coupled with the fact that CFP has an extinction coefficient five times less than YFP (Rizzo et al., 2006), meaning that its brightness per molecule is five times lower, this resulted in few cells that were suitable for analysis. Another possible source of error in the results is the inability to assess the stoichiometry of the expression of the CFP and YFP tagged versions in each cell (Piston & Kremers, 2007). While every care was taken to ensure that each transfection contained a 1:1 ratio of plasmid DNA, differential transfection or expression of individual constructs may cause bias in the perceived FRET efficiency.

As insect ORs have been observed to heterodimerise with Or83b in vivo, FRET experiments were performed in S2 cells expressing a FRET pair with and without the additional expression of Or83b in order to assess whether these interactions occur in the presence of Or83b (Figure 4.4). The intracellular loop 1 CFP and intracellular loop 1 YFP fusions were chosen as a representative pair that exhibited FRET, and the intracellular loop 2 YFP and intracellular loop 3 CFP fusions were chosen because no FRET was observed between this pair. No significant difference was observed when Or83b was co–expressed with the FRET partners (two–tailed t–test, p < 0.05). In order to test whether this negative result was due to the endogenous expression of Or83b by S2 cells, RT–PCR on S2 cell cDNA was performed. RT–PCR using Or83b–specific primers produced a PCR
product between 850 and 650 bp in size, corresponding to a predicted size of 700 bp. This result suggests that, like Sf9 cells (Smart et al., 2008), S2 cells express Or83b (Figure 4.5). The sequence of this fragment has yet to be determined to verify that it indeed encodes D. melanogaster Or83b.

4.4 Discussion

Understanding the structure of a ligand–activated ion channel that is an OR is an important step in elucidating the mechanism by which they transduce olfactory signals. As discussed in Section 1.4, oligomerisation has been identified as an important feature of both GPCRs and ion channels, though its role in the function of insect ORs has yet to be fully investigated. To this end, a fluorescence resonance energy transfer approach was taken to investigate whether ligand–activated ORs form homo–oligomers, and if so, which domains are in close proximity in these structures. Interaction was observed between the intracellular loops of some combinations of fluorescently–tagged Or22a proteins (Figure 4.3), indicating that this receptor does form a higher order structure when expressed in S2 cells.

Insect OR homo–oligomerisation has also been observed for the Drosophila OR, Or43a, in a heterologous system (Neuhaus et al., 2005). These results contrast with similar experiments performed in vivo where no evidence for homo–oligomerisation between ligand–binding ORs was observed and interaction only seen between a ligand–binding OR and Or83b, or between Or83b molecules (Benton et al., 2006). One possible reason for this discrepancy is that the location of the FRET partners differs between studies. Benton et al. (2006) noted that homo–oligomerisation did not occur between N–terminally fluorescently–tagged Or43a receptors in vivo using bimolecular fluorescence complementation of ORs. This may simply be because the N–termini of the ligand–binding ORs are not in close proximity in the higher order structure that ORs adopt in vivo. Alternatively, it could imply that the identification of homo–oligomers of Or22a in a heterologous system does not reflect the nature of ORs in vivo, and could further
Figure 4.3: Single cell FRET studies to detect interactions between CFP– and YFP–tagged Or22a in S2 cells. A) Representative FRET efficiency images for CFP–YFP single open reading frame control (CFP–YFP), CFP and YFP co–expressed in the same cell (CFP + YFP) and cells co–expressing different combinations of CFP– and YFP–tagged versions of Or22a. NY = N–terminal YFP fusion, I1C = intracellular loop 1 CFP fusion, I3C = intracellular loop 3 CFP fusion, I1Y = intracellular loop 1 YFP fusion, I2Y = intracellular loop 2 YFP fusion and I3Y = intracellular loop 3 YFP fusion. FRET was assessed using donor dequenching by acceptor photobleaching (See Section 4.2.6). The calibration bar illustrates the FRET efficiency of each pixel, with dark blue denoting no FRET and red denoting >25% FRET. B) Mean FRET efficiency calculated from 14–36 S2 cells and is expressed as the mean ± S.E.M. “a”, “b” and “c” indicate statistical significance groups (one way ANOVA with Tukey’s multiple comparison test, p < 0.05).
Figure 4.4: Results of single cell FRET assessing the impact of Or83b co-expression on CFP– and YFP–tagged Or22a in S2 cells. Mean FRET efficiency was calculated and expressed ± S.E.M. No significant differences were observed between cells transfected with Or83b compared with those expressed in native S2 cells without Or83b transfection (two–tailed t–test, p < 0.05).

Figure 4.5: Reverse transcription polymerase chain reaction of native S2 cells. A) 1Kb+ molecular weight standard. B) Negative control reaction with no template using Or83b–specific primers. C) Or83b–specific primers on S2 cDNA. The arrow on the right hand side indicates the major product in lane C.
be examined by expressing the constructs used in this study in an *in vivo* system. Nevertheless, the observation of specific FRET between certain pairs of tagged Or22a and not others suggests that, while these interactions may not occur *in vivo*, it may be that these domains are those that are interacting with Or83b in the insect.

The protein–protein interaction data collected for Or22a, together with data from the literature, allows some potential models of insect OR quaternary structure to be proposed (Figure 4.6). Multiple possible interaction interfaces were observed, which supports the hypothesis that insect ORs form a quaternary structure of more than two subunits. The data obtained using FRET in this study suggests that a higher order structure is more likely, as there are multiple potential interaction domains (e.g. I1–I1, I1–I3, I3–I3). However, it is possible that interactions with lower FRET efficiency are occurring between subunits in neighbouring receptor dimers, and that only the major interactions occur within a complex. A dimeric structure may still be possible if this is the case (e.g. Figure 4.6B).

Both rhodopsin and the α1b–adrenoreceptor form oligomeric chains that extend in two dimensions (Fotiadis et al., 2003; Carrillo et al., 2004). Chains of Or22a monomers could form using repeated I1–I1 and I3–I3 interactions, and possibly in the second dimension if the N–terminus interacts with the second intracellular loop (e.g. Figure 4.6C). These experiments between the N–terminus and the second intracellular loop have yet to be performed. This particular combination is given as an example. As the manner in which the TM helices of insect ORs pack in the cell membrane is unknown, unlike that of GPCRs, the geometry of these interactions can not be calculated accurately. It is therefore possible that other as yet unknown interactions also mediate the formation of oligomeric ORs.

The potential role of Or83b in this “chain” model is unclear. A functional OR structure may involve repeating elements, such as Or22a–Or83b–Or22a, in order to form a functional OR complex *in vivo*. Or83b receptors alone are able to make up a functional ion channel with a low level of basal activity (Sato et al., 2008; Wicher et al., 2008). A repeating OR:Or83b:OR may enable the ligand–binding OR to interact with multiple Or83b ion channels in order to convey ligand sensitivity. Such a model would fit with the Y2H work
Figure 4.6: Models of homo–oligomerisation of the ligand–activated OR, Or22a. A) An Or22a monomer. TM domains are numbered, and the faces of the helical bundle with intracellular loops joining them are labelled with different colours (red, intracellular loop 1; blue, intracellular loop 2; green, intracellular loop 3). The N–terminus is prior to TM1, and the C–terminus is after TM7. The helical bundle is depicted as a circle to simplify interpretation. B) Or22a as a homodimer, with I1–I1 and I3–I3 as potential, though not the only possible, interaction faces. C) Or22a as oligomers in interacting chains extending in one dimension through repeated I1–I1 and I3–I3 interactions, and in the second with a postulated N–I2 interaction. D) Or22a as a pore–forming tetramer, with I1–I3 interactions predominating. A tetramer is shown, though the data does not suggest that a trimer or pentamer is any less likely.
of Benton et al. (2006), who demonstrated that the third intracellular loops of Or83b and either Or43a or Or22a interact. An I3–I3 interaction is a feature of the model presented in Figure 4.6C.

Alternatively, ligand–sensitive insect ORs may themselves form a pore unit, perhaps through repeated I1–I3 interactions in a ring–like structure (e.g. Figure 4.6D). A tetrameric structure is illustrated, though the data does not suggest that a trimer or pentamer is any less likely. In vivo, Or83b may substitute for some of the Or22a subunits, making up a heteromeric channel. The ligand–sensitive OR may associate around this structure, interacting with the Or83b channel in order to convey odorant sensitivity, with the ligand–sensitive ORs in close enough proximity to each other for FRET to occur in a manner consistent with the results presented here. Further experiments are required to distinguish between these putative models.

The role of Or83b in multimerisation with other ORs was not directly assessed in this study. However, it was hypothesised that if both homo– and heteromeric interactions were occurring through the same domains, then the additional expression of Or83b may reduce the level of FRET seen as it would disrupt the FRET pair. No change in FRET efficiency was observed in two Or22a FRET combinations when Or83b was co–transfected with the FRET pair. While not every possible FRET pair was tested with additional transfection of Or83b, some tentative conclusions can be made. As Or83b is likely to be expressed in the S2 cell line as determined by RT–PCR, the transfection of Or83b may have no additional effect over endogenous expression. This could be because the Or22a FRET pair is associated with endogenous Or83b in the S2 cells, and that the observed FRET is taking place despite this interaction. Performing similar experiments in a null–Or83b background, such as using RNA intereference, or tagging Or83b with a fluorescent protein and measuring the FRET efficiency between it and Or22a in S2 cells will enable this hypothesis to be more thoroughly tested.

The regions of highest FRET efficiency occurred towards the cell membrane, particularly when N–terminally YFP–tagged Or22a was co–expressed with a CFP–tagged version of Or22a (Figure 4.3). However, the presence of FRET in the internal regions of the cells
suggests that protein–protein interactions between Or22a molecules form shortly after protein synthesis or as the protein is trafficked to the cell membrane. This is similar to the observed interactions between ORs and Or83b in vivo, where Or83b is required for trafficking of the receptor to the dendritic membrane, and thus the interaction between the receptors forms prior to transport to the cell membrane (Benton et al., 2006).

Lipid rafts are regions of the cell membrane that are enriched in cholesterol and sphingolipids, which causes the formation of membrane microdomains (Oh & Schnitzer, 2001). Many receptors have been observed to be specifically enriched in lipid rafts (reviewed in Becher and McIlhinney (2005)), suggesting they have a role in grouping together signalling proteins that utilise a common signalling pathway. The observation of FRET in specific regions in the cell membrane in this study, as opposed to spread evenly around the border of the cell, is consistent with the hypothesis that insect ORs are localised to membrane microdomains. The study of Wicher et al. (2008) observed calcium influx in HEK293 cells expressing Or22a localised to specific regions of the cell membrane, also consistent with a raft–like association of ORs. Further study in vivo is required to determine whether this is the case for ORs in the insect ORN.

One important experimental control is missing from the FRET experiments presented in this study. FRET measurements are susceptible to artifacts due to the non–specific aggregation of the putative interacting proteins. This is a particular problem for FRET analyses of membrane proteins, as their diffusion is further restricted to the plane of the cell membrane (Pfleger & Eidne, 2005). The measurement of FRET between Or22a and a receptor that is unlikely to interact with insect ORs, such as a mammalian GPCR, would have allowed greater confidence that the observed FRET is a result of a specific interaction between ORs and not the result of aggregation. For example, Neuhaus et al. (2005) used the β2–adrenergic receptor to confirm the specificity of the BRET signals observed in their homo– and heterodimerisation experiments on insect ORs. Despite the lack of this control, no FRET was seen when some pairs of tagged Or22a receptors were co–expressed. If the observed FRET signals were due to non–specific aggregation, similar amounts of FRET would be observed in all experimental conditions given that the visible
expression of each construct was similar. Some confidence can therefore be taken that the observed FRET signals are due to specific interactions between Or22a monomers.

The insertion of a fluorescent protein into any of the intracellular domains yielded a functional OR (Figure 4.2). An N-terminal YFP fusion was the least different to wild-type Or22a, but insertion into the first, second or third intracellular loops also generated functional receptors. These results suggest that the insertion of a protein into these domains does not alter the protein structure, indicating that any protein–protein interactions observed are not an artifact caused by protein fusion. However, the fact that insertion in the intracellular loops causes a larger reduction in receptor activity suggests that these domains are more important for protein structure than the N–terminal domain. The third intracellular loop has been implicated in interaction with Or83b (Benton et al., 2006) and in the formation of a possible pore–loop structure (Wicher et al., 2008). Roles for the first and second intracellular loops have yet to be identified.

Performing FRET experiments on homo- and hetero-oligomeric insect ORs using an in vivo expression system may prove a more useful approach than the heterologous system employed in this study. Insect ORs fused to fluorescent proteins could be expressed in the ab3A neuronal subtype that has had the native Or22a gene deleted using the GAL4–UAS system, similar to that used for in vivo electrophysiological analysis of insect ORs (Dobritsa et al., 2003; Hallem et al., 2004a). This could enable the structure of the OR complex to be determined in a system that more closely represented the native environment, as expression of the ORs will take place in the presence of any accessory proteins required for the correct expression, targeting and transport of the FRET constructs to the dendritic membrane. There may also be further proteins that are members of the OR complex in vivo that may influence their interaction. Increased expression levels may be observed in such an expression system that improve detection and analysis of FRET.
5.1 Summary of Results

The mechanisms through which insects detect and respond to volatile chemical cues in their environment are still not fully understood. Recent findings regarding the structure and function of insect ORs, as well as the signal transduction pathways they utilise, have raised a number of questions regarding the nature of insect olfactory systems (see Chapter 1). This PhD project was undertaken to address some of these questions. The aims of this project were to develop a functional assay for insect ORs, and to test the hypotheses that the ligand–binding ORs share the intracellular N–terminus and seven membrane topology of the co–receptor Or83b and that insect ORs form higher order structures in the cell membrane.
5.1 Summary of Results

5.1.1 In vitro olfactory receptor assay

A simple, robust functional assay for measuring the activation of insect ORs utilising calcium imaging in the Sf9 insect cell line was developed. When expressed in Sf9 cells, Or22a from D. melanogaster responded to the odorant, ethyl butyrate (Figure 2.2). This system did not require the addition of any exogenous factors, such as the co-receptor Or83b, to demonstrate the exquisite sensitivity of these receptors (Figure 2.6). The response profile of Or22a expressed in this system was similar to its activity in vivo (Figure 2.7), demonstrating that the Sf9 system accurately determines the range of odorant sensitivities of insect ORs. This assay system can also be applied to the collection of pharmacological data for these receptors, such as EC$_{50}$ values. The functional assay of insect ORs in the Sf9 cell line will find a number of applications in the study of the insect olfactory system, including investigation of OR signal transduction and the coding of odour mixtures.

5.1.2 Topology analysis of ORs

Previous research had indicated that the insect OR co-receptor Or83b has the same number of TM domains but the opposite orientation to GPCRs (Benton et al., 2006; Lundin et al., 2007). Using epitope–tagging and immunochemistry, this inverted GPCR orientation was also found for two ligand–binding ORs, Or22a from D. melanogaster and Or1 from E. postvittana (Figure 3.6). A full topological model of Or22a was developed by testing TM prediction data with immunochemistry on epitope–tagged fusion proteins (Figure 3.7). The topology of insect ORs is unlike that of any known receptor, and suggests that they form a novel family of receptors.

5.1.3 Analysis of the oligomeric structure of ORs

Receptors, whether they be GPCRs, ion channels or other receptor types, often function as higher order oligomers. The ability of Or22a to form higher order structures was examined
using FRET. The first intracellular loop was found in close proximity to both neighbouring first and third intracellular loops, and the third intracellular loops of neighbouring Or22a subunits are also in close proximity to each other. These results suggest that Or22a forms a quaternary structure of multiple receptors. Furthermore, this higher order structure may be more than just a dimeric structure as multiple interaction faces were observed between neighbouring Or22a receptors. However, there are still a number of experiments required to complete this data set. Significantly, there is no experiment testing the interaction between N–termini of co–expressed Or22a, nor between N–terminus and intracellular loop 2. The interactions between Or22a and Or83b have also not been investigated fully. Testing these potential interactions is vital for determining the likely quaternary structure of insect ORs.

5.2 Current hypotheses and future directions

In contrast with the detailed *in vivo* studies of Hallem et al. (2004) and Pelz et al. (2006), the assay developed here does not require an in depth understanding of the experimental organism’s anatomy. *In vivo* assays have been utilised for the screening of non–Drosophila ORs, including an OR from *A. gambiae* attuned to a human sweat volatile (Hallem et al., 2004b). However, the Sf9 assay system will provide an additional tool for the deorphaning of insect ORs that is straightforward to perform with simple and widely available reagents and equipment. To date, it has been used to identify *E. postivittana* Or1 and Or2 as plant terpene receptors (Jordan et al., 2008), and has also been used to identify ligands for female–specific *B. mori* ORs (Anderson et al., 2008).

Odours in the environment are usually present in highly heterogenous mixtures. The manner in which mixtures of odours are coded at the OR level is currently unknown, as most studies have focused on the presentation of single odorants. The use of calcium imaging in Sf9 cells would be ideal for probing whether combinations of odorants result in higher (synergistic) or lower (inhibitory) responses than would be expected from response to the pair of compounds individually. The ability to distinguish between odorants is based
upon the interpretation of the current level of glomerular activity (Hallem & Carlson, 2006). As each glomerulus represents the sum of activity of one olfactory receptor gene, it is therefore important how a mixture of compounds interacts with each OR. Oka et al. (2004) demonstrated that the odorant eugenol activated two distinct mouse glomeruli by itself, but in the presence of methyl isoeugenol, only one of these glomeruli was activated. This indicates that methyl isoeugenol inhibits, or masks, one of the ORs that eugenol normally interacts with, altering perception of the odour space. Synergistic or inhibitory interactions between odorants could prove to be an important aspect of odour coding for insects in the environment.

The modes of olfactory receptor activation are a potential target for study using the calcium imaging in Sf9 cells. Hallem et al. (2006) noted that a reduction in ORN firing rate caused by the addition of a ligand was “strikingly prevalent” for many ORs. However, it is not yet known how this is caused at the OR level. It has previously been shown that the OR itself is the sole mediator of the basal ORN firing rate (Hallem & Carlson, 2004). This suggests that the OR is in equilibrium between the fully active and inactive conformations in the resting state. The addition of an activating odorant causes a shift in equilibrium towards the active conformation increasing ORN firing rate, and an inhibitory ligand causes a shift in equilibrium towards the inactive conformation, decreasing the ORN firing rate. The ability of each odorant to cause a conformational shift may be the key functional determinant of the strength of the response. This could be investigated in the Sf9 assay system using known activating and inhibitory odorants to determine the mechanism through which this occurs.

The assay system for insect ORs described here may enable the rapid screening of potential antagonists of insect olfaction. As described in Chapter 1, insects are the primary vector of many diseases with a significant impact on humans. The topically applied insect repellent N,N-diethyl-meta-toluamide (DEET) has recently been described as acting on insect ORs (Ditzen et al., 2008). Sf9 cell–based assays could be used to investigate potential inhibitors for the ORs that detect ecologically relevant odorants, perhaps without disrupting desired behaviours or having any undesirable off–target effects.
Expression of insect ORs in the Sf9 cell line enables investigation of the second messengers used in signal transduction. For example, Smart et al. (2008) used a number of inhibitors of enzymes common in second messenger generation to show that insect ORs do not use a GPCR–like signal transduction pathway in the Sf9 cell line. The addition of the G protein inhibitor GDP–\(\beta\)-S or the adenylyl cyclase inhibitor SQ 22,536 had no effect on the maximal response of Or43b to ethyl butyrate. Similar results were seen in the HEK293 cell line, though GDP–\(\beta\)-S did result in a change in the length of time required to return to basal calcium levels. These results suggest that G proteins are not the major component of signalling in this cell line, though they may have a modulatory role. This is supported by the observation that insect ORs share no similarity to any known GPCR (Wistrand et al., 2006). There is evidence that different odorant molecules can cause rat ORs to signal through different secondary messenger pathways, changing the response of the ORN through activation of synergistic or inhibitory pathways (Duchamp-Viret et al., 2003). Calcium imaging in the Sf9 assay system could prove a great resource in determining whether similar phenomena occur in the insect odour coding system, as studying one OR in isolation will enable us to be able to distinguish mixture effects caused by the interaction of both compounds with one OR type with those caused by the interaction of the mixture with multiple ORs.

The mechanism through which the insect olfactory receptor response cycle is terminated is also currently unknown. Whether the receptors are internalised after activation or whether other systems are used to modulate receptor activity is unclear. A potential role of G proteins has been suggested, as this may explain why the PLC pathway is up–regulated in response to pheromone in vivo (Boekhoff et al., 1990a; Boekhoff et al., 1990b). Further investigation using functional assays in Sf9 cells may allow the signal transduction pathway of insect ORs to be fully tested, including the role of G proteins.

The functional expression of insect ORs in Sf9 cells also enables a number of potential approaches for understanding the domains of ORs that are functionally important. Mutational analysis, either through chimeric receptors or the generation of point mutants, could enable the identification of residues or domains involved in ligand binding, signal
transduction or maintaining receptor structure (for example, see Yin et al. (2004)). This may allow greater insight into how an odorant binds to an OR and what interactions are important for OR functionality. This assay could also be used to reconstitute the ORN lymph by adding purified OBPs or ODEs to investigate their role in peripheral signalling (Grosse-Wilde et al., 2007). This assay system provides a tool to monitor mutant or otherwise altered ORs for activity in order to confirm whether they have a deleterious effect or not on receptor function and therefore confirm the biological relevance of such regions in the OR. This application is used to support both the determination of the TM topology of Or22a and Or1, as well as the investigation of the higher order structure of Or22a as described in Chapters 3 and 4 of this thesis.

Unfortunately, the assay described in this thesis is low throughput. Several attempts were made to increase the number of responding cells per assay (see Chapter 2). In other studies, Sf9 cells stably expressing proteins have been generated that have improved receptor assays (Knight et al., 2003). However, the generation of polyclonal cell lines resulted in no increase in the number of Sf9 cells responding upon addition of ligand compared to transiently transfected cells, and the generation of clonal Sf9 cell lines stably expressing ORs proved impossible despite many attempts.

The assay presented in this thesis has a degree of subjectivity in the analysis of cell data. In fields of view with no obvious responding cells it is tempting to look harder for cells that may be responding, but which would otherwise have been ignored as not responding if there were better candidates. In order to get around this potential bias, a number of attempts were made to automate data collection. Cell recognition algorithms were trialled to collect data in order to remove this observer bias. These attempts failed due to the variability in the levels of fluorescence of the Fluo–4 loaded Sf9 cells between experiments. The best solution so far uses criteria that define a responding cell (see Appendix B for a description). Fluorescence data is collected for every cell in a field of view and these criteria applied. This has reduced variability both between experiments and between different researchers but is still quite laborious.

Baculovirus expression systems have previously been used to express human ORs for
transient assays in the Sf9 cell line (Matarazzo et al., 2005), and could prove to be useful in lifting the response rate of insect ORs. Such an expression system is yet to be trialled for the functional analysis of ORs using the Sf9 cell line. Until a higher OR expression rate has been achieved, a high throughput plate-reader type assay is not feasible. This may require a search for additional components of the olfactory system that act as chaperones to improve the heterologous expression of insect ORs similar to those seen for mice and worm ORs (Dwyer et al., 1998; Saito et al., 2004).

The precise mechanism used by insect ORs in generating neuron depolarisation is still unknown. While recent studies suggest that Or83b forms an ion channel (Sato et al., 2008; Wicher et al., 2008), the form this complex takes in the dendritic membrane is still not clear. Coupled with the results presented here, which suggest that Or22a, and by extension other ligand–binding ORs, forms homomeric interactions in the cell membrane, three models can be put forward to explain the mode of activation of insect ORs (Figure 5.1).

Firstly, the ligand–binding ORs and Or83b may interact to form a heteromeric odorant–gated ion channel. This complex is likely to contain at least two ligand–binding ORs, but may contain more, due to the multiple possible interfaces of interaction shown using
5.2 Current hypotheses and future directions

FRET. However, the observation that Or83b alone is able to form a functional cation channel both in vivo and in vitro suggests that the cation structure may be comprised of Or83b subunits alone. This second model requires the interaction of, possibly multiple, ligand–binding ORs with this cation channel in order to convey odorant sensitivity. Thirdly, a combination of direct interaction of the ligand–binding OR with the Or83b channel and modulation of the channel activity through a G protein–mediated increase in cAMP elicited by activation of the ligand–binding OR is possible. This G protein component of signalling may be the major component of signalling, as OR activity is seen in vitro in the absence of Or83b (Wetzel et al., 2001), though it remains controversial whether G proteins are utilised in vivo (Sato et al., 2008; Smart et al., 2008).

FRET has been demonstrated as a useful system for the analysis of insect OR protein–protein interactions (see Chapter 4). Experiments testing Or83b fused to a fluorescent protein and co–expressed with the Or22a constructs developed in this study may begin to shed light on the subunit stoichiometry of the OR:Or83b complex. A study by Drenan et al. (2008) used different combinations of tagged and untagged subunits of the acetylcholine receptor to determine the pentameric subunit assembly and stoichiometry of this heteromeric receptor. A similar approach taken to study insect ORs may determine the ratio of ligand binding ORs to Or83b in each functional OR unit. We can then suggest whether the functional unit of insect ORs is a heteromeric ion channel or a homomeric channel with interacting ligand–binding subunits.

Taking advantage of the fluorescently tagged versions of Or22a developed in this study, a three–colour FRET approach could be taken for distinguishing between possible dimeric and multimeric structures (Lopez-Gimenez et al., 2007). This approach uses a third FRET partner that accepts energy from YFP, such as dsRed, and measures sequential energy transfer between the three molecules (CFP to YFP to dsRed). This increases the number of interactions that can be assessed simultaneously, allowing a more detailed understanding of the subunit stoichiometry to be developed. Some receptors, such as the metabotropic glutamate receptors, exhibit changes in their dimeric structure upon addition of a ligand (Tateyama et al., 2004). Similar experiments measuring changes in
FRET elicited by odorant addition may provide further insight into insect OR structure and function. Co-immunoprecipitation of native or heterologously expressed ORs may also enable the number of OR subunits that make up a functional complex to be determined (Uberti et al., 2003).

FRET may also be applied to investigating the conformation changes that ORs undergo upon odorant binding. For example, intramolecular FRET has been used to define the conformational changes that the parathyroid hormone receptor and the α2A–adrenergic receptor undergo upon ligand binding (Lohse et al., 2003; Vilardaga et al., 2003). The fusion of CFP into the third intracellular loop and a YFP onto the C-terminus of these receptors enabled the monitoring of receptor activation. When a ligand was applied, a conformational shift resulted in the two fluorescent proteins moving further apart, thereby altering the wavelengths of light emitted by the fluorescent proteins. A similar study on insect ORs may reveal the conformational changes that occur within the ligand–binding ORs upon odorant binding, the mechanism by which they communicate this event to Or83b and the conformational changes that occur within the OR ion channel that result in changes in ion transport. The order of these events may again clarify whether this ion channel structure better fits with a homo- or heteromeric model of OR structure.

The G protein–binding of insect ORs has already been assessed using the Sf9 OR assay (Smart et al., 2008). Protein–protein interaction studies could prove useful in confirming or rejecting the hypothesis that G proteins have a role in direct OR signalling. Fluorescently–labelled G proteins could be used to identify the presence and potentially the timescale of G protein binding to either ligand–binding ORs or Or83b to investigate a role in primary signal transduction or signal modulation. It is also possible that an OR transactivates an as yet unknown G protein–coupled receptor that is responsible for signal transduction through either IP$_3$ or cAMP. Investigating these potential interactions in the insect antenna will further elucidate the mechanism of insect olfaction.

Questions regarding the signal transduction, mechanism of activation and structure of the insect olfactory system remain at the forefront of insect olfactory research. What structure does the OR complex have? How does an odorant activate an OR? How are
mixtures of odorants represented by OR activity? How does activation of an OR lead to neuron depolarisation? How is the activation cycle ended rapidly enough to enable odour discrimination in the environment? The research presented in this thesis has begun to answer some of these questions, and sets a platform from which further hypotheses can be tested. Answering these questions will further enhance our understanding of arguably the most ecologically important sensory system for insects.
Appendix A: Vector Diagrams
Assay Constructs

Figure A.1: pIB-Or22a

Figure A.2: pIB-Or83b
Rhodopsin TM1 constructs

**Figure A.3:** Schematic diagram and nucleotide sequence of Rhodopsin TM1 constructs used in immunochemistry experiments.
Appendix B: Criteria for the Determination of Responding Cells

A series of criteria were developed from 250 responding Sf9 cells and incorporated as conditional formulae into an Excel spreadsheet to both speed the analysis of assay data and to provide consistency between experiments. The criteria were as follows:

- The fluorescence reading of the cell must be relatively stable. A threshold is set based on the fluorescence value in the first frame (A1): if $A1 > 50$, then the threshold is 25 units; if $10 > A1 > 50$, then the threshold is $0.5 \times A1$. This threshold sets the maximum variability between frames. In order to pass this component, a cell’s fluorescence in neighbouring frames must NOT vary by more than the threshold. For each cell there are eleven such tests, of which a cell must pass at least nine.
• The cell must not show a large change on addition of saline control. This accounts for cells that move or disappear during the course of the assay. A threshold similar to above is set based on the fluorescence value in the first frame. If A1 > 40, then the threshold is 35 units; if 10 > A1 > 40, then the threshold is 0.5 x A1. The difference in cell fluorescence between A6 and A7 must be less than this threshold for the cell to pass. If the cell fails this test, then a second test is run to check whether the failure is due to a response or some other phenomena (floating cell, background change etc). If A8 > 1.2 x A6 then the cell would fail, if less than this then the cell will pass, regardless of the result of the first test.

• The cell must respond to substrate addition. The cell must meet at least four of the following conditions: A13 and A14 are at least 10% greater than the average of 11 and 12; A14 and A15 are at least 10% greater than the average of 11 and 12; A15 and A16 are at least 7.5% greater than the average of 11 and 12; A16 and A17 are at least 7.5% greater than the average of 11 and 12; A17 and A18 are at least 7.5% greater than the average of 11 and 12; the average of A13 to A18 is at least 10% greater than the average of A1 to A12. In addition, the cell must also meet the following condition: the maximum of A13 to A18 minus the average of A1 to A12 all divided by the maximum of A19 to A24 is greater than 0.03. That is, the calculate ΔF value must be greater than 0.03.

• The minimum fluorescence value must be greater than 10 units for the cell to be included in the analysis.
Appendix C: Topology Prediction
Algorithm Description and Output

The Dense Alignment Surface Method

The dense alignment surface (DAS) method of transmembrane topology prediction (Cserzo et al., 1997) utilises the comparison of each amino acid in the query sequence against each protein in a database of non-homologous protein sequences with known transmembrane domains. Sequence regions that have a similarity over a certain threshold are marked on a virtual dot-plot for each pair of proteins. Dot-plots for unrelated pairs of proteins have an even distribution of dots, whereas shared regions are indicative of transmembrane regions. All possible dot-plots are compared, and a consensus of likely transmembrane domains is produced. These are then given cut-off scores, which indicate
Table C.1: Output of the DAS program when queried with the Or22a sequence

<table>
<thead>
<tr>
<th>Start</th>
<th>Stop</th>
<th>Length</th>
<th>Cutoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>68</td>
<td>19</td>
<td>1.7</td>
</tr>
<tr>
<td>51</td>
<td>67</td>
<td>17</td>
<td>2.2</td>
</tr>
<tr>
<td>143</td>
<td>159</td>
<td>17</td>
<td>1.7</td>
</tr>
<tr>
<td>146</td>
<td>157</td>
<td>12</td>
<td>2.2</td>
</tr>
<tr>
<td>206</td>
<td>215</td>
<td>10</td>
<td>1.7</td>
</tr>
<tr>
<td>210</td>
<td>210</td>
<td>1</td>
<td>2.2</td>
</tr>
<tr>
<td>265</td>
<td>304</td>
<td>40</td>
<td>1.7</td>
</tr>
<tr>
<td>268</td>
<td>291</td>
<td>24</td>
<td>2.2</td>
</tr>
<tr>
<td>293</td>
<td>293</td>
<td>1</td>
<td>2.2</td>
</tr>
<tr>
<td>296</td>
<td>302</td>
<td>7</td>
<td>2.2</td>
</tr>
<tr>
<td>377</td>
<td>387</td>
<td>11</td>
<td>1.7</td>
</tr>
<tr>
<td>379</td>
<td>385</td>
<td>7</td>
<td>2.2</td>
</tr>
</tbody>
</table>

the likelihood of that particular region being a transmembrane domain compared to those found in the database. The output of the DAS program when queried with Or22a is given in Table C.1.

TMHMM

Figure C.1: Graphical output of the TMHMM program

Transmembrane domain prediction using the TMHMM (transmembrane hidden Markov modelling) method (Krogh et al., 2001) uses a range of transmembrane domain characteristics, such as hydrophobicity, charge bias, helix cap regions and helix length. It
Table C.2: Output of the TMHMM program when queried with the Or22a sequence

<table>
<thead>
<tr>
<th>Location</th>
<th>Position of Start</th>
<th>Position of End</th>
</tr>
</thead>
<tbody>
<tr>
<td>inside</td>
<td>1</td>
<td>49</td>
</tr>
<tr>
<td>TMhelix</td>
<td>50</td>
<td>72</td>
</tr>
<tr>
<td>outside</td>
<td>73</td>
<td>144</td>
</tr>
<tr>
<td>TMhelix</td>
<td>145</td>
<td>164</td>
</tr>
<tr>
<td>inside</td>
<td>165</td>
<td>262</td>
</tr>
<tr>
<td>TMhelix</td>
<td>263</td>
<td>285</td>
</tr>
<tr>
<td>outside</td>
<td>286</td>
<td>294</td>
</tr>
<tr>
<td>TMhelix</td>
<td>295</td>
<td>317</td>
</tr>
<tr>
<td>inside</td>
<td>318</td>
<td>397</td>
</tr>
</tbody>
</table>

scores each characteristic for a particular region, then combines the best fitting regions for each characteristic to give a best fit model using score difference minimization. It uses this information to assign likely transmembrane domains. Output for TMHMM is given in Figure C.1 and Table C.2.

**TMPred**

![Graphical output of the TMPred program](image)

Figure C.2: Graphical output of the TMPred program
The TMpred program (Hofmann & Stoffel, 1993) makes a prediction of membrane-spanning regions and their orientation. The algorithm is based on the statistical analysis of TMbase, a database of naturally occurring transmembrane proteins. The prediction is made using a combination of several weight-matrices for scoring after comparing the query sequence to the database. The prediction given by the TMpred program is illustrated in Figure C.2 and Table C.3.

**TMAP**

<table>
<thead>
<tr>
<th>Region</th>
<th>Start Residue</th>
<th>End Residue</th>
<th>TM Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>44</td>
<td>72</td>
<td>29</td>
</tr>
<tr>
<td>2</td>
<td>137</td>
<td>163</td>
<td>27</td>
</tr>
<tr>
<td>3</td>
<td>262</td>
<td>282</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>289</td>
<td>309</td>
<td>21</td>
</tr>
<tr>
<td>5</td>
<td>364</td>
<td>387</td>
<td>24</td>
</tr>
</tbody>
</table>

**Table C.4:** Predicted transmembrane segments for Or22a as determined using TMAP
The TMAP prediction algorithm (Milpetz et al., 1995) uses the multiple sequence alignment of homologous proteins to both identify transmembrane helix domains and predict which side will be extra- or intracellular. Input data can either be in the format of a multiple sequence alignment, or of a single protein. If a single protein is used, a search of the SWISS-PROT is used to find homologous sequences, which are then used in the prediction. TMAP was unable to find any sufficiently similar proteins to Or22a in the SWISS-PROT database, so it performed its prediction algorithms on the Or22a sequence alone. The TMAP output for Or22a is shown in Table C.

**Predator**

The Predator prediction method (Frishman & Argos, 1995) is generally a secondary structure prediction algorithm in that it is able to assign either \( \alpha \)-helical or \( \beta \)-sheet structure. It does this based on the features of each structure type, such as hydrogen bond energy and torsional angle probabilities. By calculating the probabilities of small portions of the protein, it is able to build a model of the secondary structure. The output of Predator for Or22a is shown in Figure C.3.

**SPLIT**

The SPLIT method (Juretic et al., 1993) uses a calculation of the average hydrophobicity of the neighbours of a particular amino acid and comparing this to the hydrophobicity of that particular amino acid. This data is used to calculate the preference of each amino acid for being in a transmembrane helix. Regions that reach a certain threshold of hydrophobicity are classified as transmembrane domains. The output can be seen in Table C.5.
Figure C.3: The output of the Predator method of secondary structure prediction when queried with the sequence of Or22a.

Table C.5: Predicted transmembrane segments for Or22a as determined using SPLIT HMMTOP

<table>
<thead>
<tr>
<th></th>
<th>Helix</th>
<th>N-Term</th>
<th>C-Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>145</td>
<td>162</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>197</td>
<td>219</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>266</td>
<td>291</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>297</td>
<td>320</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>362</td>
<td>386</td>
<td></td>
</tr>
</tbody>
</table>

HMMTOP

The HMMTOP (Hidden Markov Model for Topology Prediction) method (Tusnady & Simon, 2001) utilises a neural network that predicts transmembrane domains based on
the maximum divergence of the amino acid composition in a given a protein sequence. This method is based on the assumption that the localisation of transmembrane segments is determined by the difference in amino acid distribution in different structural classes of the protein. The likelihood is then maximised across the protein, and a topology prediction made. The output for HMMTOP is shown in Figure C.4.

N-terminus: IN
Number of transmembrane helices: 6

Figure C.4: Output of the HMMTOP prediction algorithm. “H” indicates helix, “I” indicates that the residue is predicted to be cytoplasmic, “O” indicates the residue is predicted to be extracellular. A capitalised prediction is more likely than an un-capitalised prediction.

PredTMR2

PredTMR2 (Pasquier & Hamodrakas, 1999) couples a hydrophobicity analysis to a detection of potential membrane helix termini using an artificial neural network. The
program also contains a pre-processing algorithm which attempts to classify proteins as either membrane or globular proteins. The output is shown in Table C.6.

<table>
<thead>
<tr>
<th>TM</th>
<th>Begin</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>69</td>
</tr>
<tr>
<td>2</td>
<td>145</td>
<td>164</td>
</tr>
<tr>
<td>3</td>
<td>273</td>
<td>290</td>
</tr>
</tbody>
</table>

Table C.6: Output from TMR2

SOSUI

There are three basic assumptions in the SOSUI system (Hirokawa et al., 1998). First, membrane proteins are characterized by at least one, particularly hydrophobic, primary transmembrane helix. Secondary hydrophilic transmembrane helices may also exist in multspanning membrane proteins even though their hydrophobicity is in fact similar to the hydrophobic segments of soluble proteins. The possible role of secondary transmembrane helices is the formation of active sites of proteins. Third, the primary transmembrane helices are stabilized by a combination of amphiphilic side chains at the helix ends as well as high hydrophobicity in the central region. When polar interactions are also found in the center of a primary helix, their existence is usually for the stabilization of transmembrane helices.

This software uses four physico-chemical parameters: the hydropathy index of Kyte and Doolittle (Kyte & Doolittle, 1982) an amphiphilicity index, an index of amino acid charges, and the length of each sequence. The second parameter expresses the amphiphilicity of polar side chains and was devised by the calculation of the transfer energy of the hydrocarbon part of a polar side chain. The output is shown in Table C.7.
<table>
<thead>
<tr>
<th>No.</th>
<th>N terminal</th>
<th>transmembrane region</th>
<th>C terminal</th>
<th>length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>48</td>
<td>YKILAFVNIVMLILLPISISIE</td>
<td>70</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>191</td>
<td>CFLMETAIYMDLCTDVCPLISML</td>
<td>213</td>
<td>23</td>
</tr>
<tr>
<td>3</td>
<td>275</td>
<td>LIGTVLGLSMINLMFFSTFWTG</td>
<td>297</td>
<td>23</td>
</tr>
</tbody>
</table>

Average of hydrophobicity : 0.215113

**Table C.7:** Output from SOSUI.

**TOPPRED**

The TOPPRED prediction algorithm (Claros & von Heijne, 1994) establishes a hydrophobicity profile for the protein using a number of hydrophobicity scales in order to determine the transmembrane domains. The topology is then determined using the “positive–inside” rule. Output shown in Table C.8.

<table>
<thead>
<tr>
<th>Helix</th>
<th>Begin</th>
<th>End</th>
<th>Score</th>
<th>Certainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>49</td>
<td>69</td>
<td>1.923</td>
<td>Certain</td>
</tr>
<tr>
<td>2</td>
<td>88</td>
<td>108</td>
<td>0.679</td>
<td>Putative</td>
</tr>
<tr>
<td>3</td>
<td>145</td>
<td>165</td>
<td>1.271</td>
<td>Certain</td>
</tr>
<tr>
<td>4</td>
<td>284</td>
<td>304</td>
<td>2.090</td>
<td>Certain</td>
</tr>
<tr>
<td>5</td>
<td>357</td>
<td>377</td>
<td>1.163</td>
<td>Certain</td>
</tr>
</tbody>
</table>

**Table C.8:** Output from TOPPRED. The specific parameters used were: Full window size = 21, core window size = 11, wedge window size = 5, cutoff for certain transmembrane segments = 1.00, cut off for putative transmembrane segments = 0.60, critical distance between 2 transmembrane segments = 2, critical loop length = 60.
References


REFERENCES


REFERENCES


REFERENCES


References


REFERENCES


