

Structural Studies on Histamine Receptors

Tian Geng

Department of Life Sciences

Imperial College London

Thesis submitted for the degree of Doctor of Philosophy

2016

ABSTRACT

The work presented in this thesis concerns a sub-family of the biologically important family of G-protein coupled receptors (GPCRs) known as histamine receptors. They are recognized as important drug targets for many disorders in mankind, such as allergic rhinitis and Alzheimer's disease.

This thesis has its roots in the first structure of histamine receptor H1, determined in complex with a first generation antihistamine (doxepin) in 2011. Here, extensive biological, computational and structural investigations have been carried out for H1 in complex with two second-generation antihistamines (cetirizine and fexofenadine). Crystal structures of H1-cetirizine (3.1 Å) and H1-fexofenadine (3.4 Å) show the binding modes of second-generation antihistamines correspond well with the previous docking studies. The so-called "anion binding site" on H1 is responsible for the high specificity of the ligands. Some regions on the receptor were difficult to define in the crystal structures. These regions were characterized by radioligand binding assays, together with molecular dynamic simulations (carried out by our collaborators from Oxford University) suggested that, second generation ligands have a significantly large impact on receptor dynamics.

While H1 mediates many immune-related disorders, the histamine receptor H3 is a drug target to many mental disorders. In addition to the H1 work presented here, it is also reported the initial work done on the histamine receptor H3. With extensive screenings on various H3 constructs (from human and turkey organisms) and expression systems, it was observed that the best construct was from turkey. An optimal expression protocol in insect cells was obtained. This has established a secure starting point for future structural studies on H3.

DECLARATION

I declare that the work described in this thesis has been carried out by myself, unless otherwise stated. All sources of information have been specifically acknowledged.

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives Licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

FIGURE LIST

	Figure Title/Description	Page
Figure 1.1	Schematic of the general signal transduction mechanism of a GPCR	17
Figure 1.2	GPCR biological response level with the presence of different types of ligands	18
Figure 1.3	Conserved GPCR features	21
Figure 1.4	Conserved inter-helical contact positions of GPCR	22
Figure 1.5	Interactions between the β 2AR receptor and G protein	23
Figure 1.6	Sequence alignment of 4 histamine receptors	31
Figure 1.7	Structural key features of H1, viewed from the extracellular side	35
Figure 1.8	Alignment of residues at the 16 ligand-binding related positions of aminergic GPCRs	35
Figure 1.9	Cetirizine and Hydroxyzine	38
Figure 1.10	A simplified schematic illustration of data collection	45
Figure 1.11	A tree-like dendrogram	52
Figure 1.12	A simplified schematic representation of a protein crystallization diagram	58
Figure 1.13	Cartoon representation of the lipidic cubic phase and the crystallization process of membrane protein in the system	61
Figure 1.14	Molooolein/water phase diagram	63
Figure 2.1	H1 construct	66
Figure 2.2	SDS-PAGE gel of H1-T4L from second metal affinity chromatography, TEV cleavage and reverse IMAC	78
Figure 2.3	H1 crystal and LCP material in a cry-loop	79
Figure 2.4	Dendrogram of cluster analysis on cell parameters of H1-cetirizine	82
Figure 2.5	Merging statistics for clusters formed with H1-cetirizine datasets	83
Figure 2.6	Correlation plots of the type $CC_{1/2}$ versus resolution are displayed for three different datasets	86
Figure 2.7	X-ray diffraction of H1-cetirizine complex	87
Figure 2.8	Dendrogram of cluster analysis on cell parameters of H1-fexofenadine	90
Figure 2.9	X-ray diffraction of H1-fexofenadine complex	91
Figure 2.10	Partial electron density maps for H1-cetirizine complex and H1-fexofenadine	93
Figure 2.11	$ F_o - F_c $ omit maps for cetirizine and fexofenadine	94
Figure 2.12	Subregions of H1 binding pocket.	95
Figure 2.13	Electron density of the “lipophilic subsite” for H1-cetirizine complex and H1-fexofenadine	96
Figure 2.14	Electron density of the “D107 ^{3.32} ionic interaction region” for H1-cetirizine complex and H1-fexofenadine	96
Figure 2.15	Electron density of “anion binding site” for H1-cetirizine complex, H1-fexofenadine and H1-doxepin	97
Figure 2.16	Cetirizine in H1-cetirizine complex	98
Figure 2.17	Isomorphous difference map	99
Figure 2.18	Cetirizine binding pocket	100
Figure 2.19	pK _a based microspecies distribution (%) of cetirizine	102
Figure 2.20	Superimposition of doxepin, phosphate ion and cetirizine.	103
Figure 2.21	Electron densities for F435 ^{6.55} and benzene ring of cetirizine in structure of H1-cetirizine complex are well defined	106
Figure 2.22	Interaction of F435 ^{6.55} with doxepin vs. Cetirizine	106
Figure 2.23	Superimposition of cetirizine and fexofenadine	108
Figure 2.24	Fexofenadine binding pocket	109
Figure 2.25	Cetirizine and Pyrilamine	114
Figure 3.1	pOPINE-3C-eGFP vector	121
Figure 3.2	Snake-plot of H3 human transmembrane region with ICL3 and Helix VIII.	130
Figure 3.3	GFP fluorescence 3D image of insect cell pellets.	133
Figure 3.4	SDS-PAGE gel analysis of H3 insect cells expression screen	135
Figure 3.5	FSEC profiles of best H3 constructs.	136
Figure 3.6	SDS-PAGE gel analysis of H3 insect cells expression using P1 and P2 viruses.	139

Figure 3.7	FSEC profiles of H3 expression optimization on the factor of TOI.	141
Figure 3.8	FSEC profiles of H3 expression optimization on the factor of Time of Harvesting.	142
Figure 3.9	FSEC profiles of H3 expression optimization on cell lines.	148
Figure 3.10	FSEC profiles of H3 expression optimization on cell lines.	148
Figure 3.11	SDS-PAGE gel of H3-T4L-GFP fusion protein IMAC purification.	150
Figure 3.12	SDS-PAGE gel of H3-T4L-GFP fusion protein 3C protease cleavage test	150

TABLE LIST

	Table Title	page
Table 1.1	Four histamine receptors sequence similarity in different regions.	32
Table 1.2	Sequence similarity within the 16 ligand-binding related positions of aminergic GPCRs.	35
Table 1.3	Seven lattice systems and 14 Bravais lattices	44
Table 2.1	Parameters of thirteen datasets of H1-cetirizine	81
Table 2.2	Merging statistics for the final dataset	85
Table 2.3	Data collection, processing and refinement statistics for H1-cetirizine	88
Table 2.4	Parameters of 28 datasets of H1-fexofenadine	89
Table 2.5	Dataset collection, processing and refinement statistics for H1-fexofenadine	92
Table 2.6	K_D K_i and B_{max} values obtained from RLB assays at pH7.5 and pH5.8 for 5 different H1 mutants plus a wild type H1.	112
Table 2.7	Interactions between the nearest sidechain nitrogen atom, and oxygen atoms of cetirizine.	113
Table 2.8	Interactions between the nearest protein sidechain atom to atoms of cetirizine	114
Table 3.1	List of H3 constructs with different T4L insertions positions and various N- and C- terminus truncations formatted into a 96 well plate and number accordingly	131
Table 3.2	Sf 21 cell parameters before and after infection	143
Table 3.3	Cell parameters before infection	144
Table 3.4	Volume of the virus used in the test	146
Table 3.5	Cell parameters measured 48 and 70 hours post infection (constructG4)	146

ABBREVIATIONS

Abbreviation	Description
β2AR	β2 adrenergic receptor
aLCV	absolute Linear Cell Variation
BBB	Blood brain barrier
BMGY	Buffered glycerol complex media
BMMY	Buffered methanol complex media
BRIL	B562 RIL
CHS	Cholesteryl hemisuccinate
CNS	Central nervous system
CPM	Counts per minute
DAG	Diacylglycerol
DDM	N-dodecyl-β-d-maltopyranoside
DLS	Diamond light source
DMSO	Dimethyl sulfoxide
DPM	Disintegration per minute
dsT4L	Disulfide-stabilized T4L
ECL2	Extracellular loop 2
FCS	Foetal calf serum
FSEC	Fluorescence-detection size-exclusion chromatography
GDP	Guanosine diphosphate
GFP	Green fluorescent protein
GPCRs	G protein-coupled receptors
G-proteins	Heterotrimeric guanine nucleotide-binding proteins
GRAFS	Glutamate, Rhodopsin, Adhesion, Frizzled/Taste2 and Secretin
GTP	Guanosine triphosphate
H1R	Human histamine H1 receptor
His8	Octa-histidine
HTS	High throughput screening
ICL3	Intracellular loop
IMAC	Immobilized metal ion affinity chromatography
IMPs	Integral membrane proteins
IP3	Inositol-1,4,5-trisphosphate
LCP	Lipidic cubic phase
MD	Minimal dextrose
MD	Molecular dynamics
MOI	Multiplicity of infection
MPs	Membrane proteins
mT4L	Minimal T4L
OPPF	Oxford protein production facility
PBS	Phosphate buffered saline
PDC	Protein-detergent-complex
PDCs	Protein-detergent complexes
pfu	Plaque forming units
PIP2	Phosphatidylinositol 4,5-bisphosphate
PKC	Protein kinase c
RFU	Relative fluorescent unit
RLB	Radioactive ligand binding assay
SBDD	Structure-based drug discovery
T4L	T4 lysozyme
TEV	Tobacco etch virus
TLS	Translation-Libration-Screw rotation
TM	Transmembrane domains
TOI	Time of infection
XFEL	X-ray Free Electron Laser
YPD	Yeast extract peptone dextros

ACKNOWLEDGEMENTS

First of all, from the bottom of my heart, I would like to thank my supervisor Dr Isabel Moraes, Professor So Iwata and Dr John Barker for giving me this wonderful opportunity to fulfill this work at the Membrane Protein Laboratory (MPL) and Evotec (Abingdon), for all your support and guidance throughout this four-year period.

A massive thanks to Dr Alex Cameron, who was my co-supervisor until moved to Warwick University during my third year. Thank you for the continuous help and endless support.

A load of grateful thanks to people from Evotec, thank you Shweta and Nic for your help and advises on RLB assays. You guys made my time at Evotec so much fun. Thank you Richard, Rahul, for your precious advices on insect cells.

I would like also to thank all the OPPF members, Ray, Louise, Heather and Nahid for your knowledge and support on high throughput expression pipeline.

Thank you Prof Phil Biggin and Benjamin Morris for the MD simulations.

To Dr James Foadi, thank you for bring me to the wonderful world of BLEND, and thank you for making my last bit of PhD much easier.

To James Birch, thank you for your help on mutagenesis.

Nien-Jen, Indran, you are my best mates in the lab. You made the long hour working time (nights and weekends) with so much fun.

Uncle (Yilmaz), your Blewberry story still making me laugh.

Mutsko and Momi, thanks for all the girls' parties, sweet cards and the lovely green tea. Life is so much better with you girls.

I would like to thank all MPL members (former and present), thank you for all the coffee break company.

Thank you Steve for carefully reading some parts of the thesis and your valuable advises.

A big thank to my love Mr Renjie Zhang, none of this would have been achieved without your love and support.

Baba and Mama, thank you for your unconditional love and support.

This research was funded by Evotec (UK) Ltd. and Membrane Protein Laboratory (MPL) at Diamond Light Source.

Table of Contents

ABSTRACT	2
DECLARATION	3
FIGURE LIST	4
TABLE LIST	6
ABBREVIATIONS	7
ACKNOWLEDGEMENTS.....	8
.....	12
1 Chapter One - Introduction.....	12
1.1 Overview.....	13
1.2 The cell membrane and membrane proteins.....	13
1.3 G protein-coupled receptors (GPCRs).....	15
1.3.1 GPCR classification.....	15
1.3.2 Signal transduction mechanism of GPCRs.....	16
1.3.3 GPCR functional states and ligand types.....	18
1.3.4 GPCR structural scaffolds	19
1.3.5 Challenges and strategies in GPCR structural studies.....	24
1.3.5.1 Expression systems.....	24
1.3.5.2 Inherent structural flexibility and instability	25
1.3.5.3 Crystallization	27
1.3.6 Structure-based drug discovery (SBDD) and GPCRs.....	27
1.3.6.1 Structure-based drug discovery (SBDD)	28
1.3.6.2 SBDD for GPCRs.....	29
1.4 Histamine receptors	30
1.4.1 Human histamine H1 receptor (H1)	33
1.4.1.1 Antihistamines	36
1.4.1.1.1 Classic antihistamines	37
1.4.1.1.2 Non-classic antihistamines	37
1.4.1.1.3 Cetirizine	38
1.4.1.1.3.1 Sedating or non-sedating antihistamines.....	39
1.4.1.1.3.2 Selectivity.....	40
1.4.2 Human histamine H3 receptor (H3)	40
1.5 Macromolecular Crystallography.....	41
1.5.1 Why crystals?	42
1.5.2 X-ray diffraction theory in a nutshell	44
1.5.3 The main steps in the structure determination process.....	46
1.5.3.1 Data collection.....	47
1.5.3.2 Data processing	48
1.5.3.3 Data combination.....	52
1.5.3.4 Intial data determination	53
1.5.3.5 Model completion and model refinement.....	54
1.5.4 Protein crystallization.....	56
1.5.4.1 Crystallization of integral membrane proteins (IMPs)	59
1.5.4.1.1 Aqueous-based crystallization-Vapor Diffusion	60
1.5.4.1.2 Lipid-based crystallization-Lipidic Cubic Phase (LCP).....	60
2 Chapter Two	64
2.1 Materials and Methods	65
2.1.1 Expression.....	65
2.1.1.1 Expression construct.....	65
2.1.1.2 Generation and screening of expression colonies	66

2.1.1.3	Small-scale expression test	66
2.1.1.4	Large-scale expression of H1T-T4L	67
2.1.2	Purification of H1-T4L with antagonist.....	68
2.1.2.1	Membrane preparation and membrane wash	68
2.1.2.2	Solubilization	69
2.1.2.3	First IMAC-Talon metal affinity Chromatography.....	69
2.1.2.4	Second IMAC-Ni-sephasrose affinity chromatography.....	69
2.1.2.5	Cleavage of GFP-His ₈ region and reverse IMAC.....	70
2.1.3	Crystallization in Lipidic cubic phase (LCP)	70
2.1.4	Harvesting crystals from LCP	71
2.1.5	Data collection and processing	71
2.1.6	Phenix-refinement.....	72
2.1.7	Protein structure figure preparation	72
2.1.8	Radioactive ligand binding assay (RLB).....	72
2.1.8.1	Site-directed mutagenesis and membrane production.....	72
2.1.8.2	Saturation radioligand binding assays	73
2.1.8.3	Data analysis: saturation binding assay.....	74
2.1.8.4	Competition radioligand binding assays	75
2.1.8.5	Data analysis: competition radioligand binding assay	77
2.2	Results	77
2.2.1	H1-T4L clone selection and Large-scale expression.....	77
2.2.2	Protein purification of H1-T4L with different antagonists	78
2.2.3	Crystallization in Lipidic Cubic Phase	79
2.2.4	Data collection and processing	79
2.2.4.1	Data processing for H1-cetirizine	80
2.2.4.2	Data processing for the datasets of H1-fexofenadine	89
2.2.5	Data quality	93
2.2.6	Crystal structures of H1 in complexes with cetirizine and fexofenadine	95
2.2.6.1	Subregions of H1 ligand binding site	95
2.2.6.2	Interactions between cetirizine and H1 receptor.....	98
2.2.6.2.1	Comparison of cetirizine and doxepin in the binding pocket of H1 receptor	103
2.2.6.2.2	Position of residue F435 ^{6.55}	104
2.2.6.3	Interactions between fexofenadine and the H1 receptor	107
2.2.7	Radiolabelled ligand binding assays and Molecular Dynamics simulations	110
2.3	Discussions	116
3	Chapter Three	119
3.1	Overview.....	120
3.2	Materials and Methods	120
3.2.1	High throughput expression screening with different expression systems	120
3.2.1.1	Construct design and synthesis.....	121
3.2.1.2	Small-scale expression screening.....	121
3.2.1.2.1	E.coli expression screening	122
3.2.1.2.2	Mammalian cells (HEK 293) expression screening	122
3.2.1.2.3	Insect cells (baculovirus) expression screening.....	123
3.2.1.3	Small-scale membrane solubilisation	124
3.2.1.4	Fluorescence-Detection Size-Exclusion Chromatography (FSEC)	124
3.2.2	Expression optimization with insect cell/baculovirus system	125
3.2.2.1	Optimization of MOI, TOI and HT for H3 construct C6 and G4.	125
3.2.2.1.1	Optimization on MOI: different virus passages.....	126
3.2.2.1.2	Optimization on TOI, MOI and time of harvesting.....	126
3.2.2.2	Optimization on cell lines with a better control of virus titer.....	127
3.2.2.2.1	P2 virus amplification and titer estimation based on increased cell diameter following infection.....	127
3.2.2.2.2	Cell line optimization	127

3.2.3	Large-scale expression	128
3.2.4	Purification of turkey H3 for the purpose of obtaining H3-T4L-GFP fusion protein	128
3.3	Results	129
3.3.1	H3 Constructs	129
3.3.2	Small scale expression screening.....	132
3.3.2.1	<i>E.coli</i> expression screening.....	132
3.3.2.2	4.2.2.2 Mammalian expression screening.....	132
3.3.2.3	Insect cells expression screening.....	133
3.3.2.4	Fluorescence-Detection Size-Exclusion Chromatography (FSEC)	136
3.3.3	Expression optimization with baculovirus system	138
3.3.3.1	Optimization on MOI, TOI and time of harvesting for H3 construct C6 and G4.	138
3.3.3.1.1	Optimization on virus passages.....	138
3.3.3.1.2	Optimization on MOI, TOI and time of harvesting.....	139
3.3.3.2	Optimization on cell lines with a better control of virus titer.....	143
3.3.3.2.1	P2 virus amplification and titer estimation based on increased cell diameter following infection.....	143
3.3.3.2.2	Optimization on three cell lines	144
3.3.4	Large-scale expression and one step IMAC purification of H3.....	149
3.4	Discussions	151
4	Bibliography.....	153
5	Appendix.....	160

Chapter One - Introduction

1.1 Overview

GPCRs are an extraordinary class of membrane proteins in several respects. Their importance is well justified by the interest they stimulate in both academic and pharmaceutical circles, especially in modern times. The seed of the present doctoral work was implanted in such a stimulating atmosphere, especially after the solution, in 2011, of the first histamine receptor by So Iwata and co-workers (Shimamura et al., 2011). This work spurred an excitement of possibilities, also extended to other types of histamine receptors. The H3 receptor is particularly interesting not only for the elucidation of mechanisms dealing with the transmission of signals in brain activity, but also as a potential drug target for mental diseases such as schizophrenia, and Alzheimer's disease. The present doctoral work aims to further the understanding of the ligand binding mechanism of H1 receptor. In using the experience accumulated from the study of H1, this thesis tries also to establish useful protocols to open the way to large-scale expression and production of H3 protein, thus preparing the framework for future purification and crystallization efforts.

In this chapter, the field of membrane proteins, GPCRs and histamine receptors is briefly reviewed to provide the background for the following chapters. Also, since a very important part of this thesis centered on the use of crystallography to determine the structure of histamine receptor H1 with ligands, the main aspects of this discipline are also included.

1.2 The cell membrane and membrane proteins

Each cell in the human body has a plasma membrane surrounding its protoplasm. Overall, the plasma membrane is around 30Å thick (Wiener and White, 1992) and is formed by three major components: lipids, proteins and sugars. Its fluid lipid bilayer

is composed of phospho-lipids arranged with their polar hydrophilic heads facing towards the aqueous environment on each side whilst the non-polar hydrophobic tails form the inside of the bilayer. The cell membrane not only separates the cellular contents from the external surrounding environment as it is also involved in numerous complex functions. While lipid molecules mainly provide structural support to the cell membrane, membrane proteins (MPs) are responsible for many important biological functions including cell signaling, cell adhesion, ion channel conductance and transport of molecules in and out of the cell. Therefore, membrane proteins are of great importance making up 50% of the drug targets on the current market (Overington et al., 2006).

Membrane proteins can be broadly grouped into two categories, peripheral membrane proteins and integral membrane proteins (IMPs) (Berg et al., 2007), depending on how they associate with the lipid bilayer. Peripheral membrane proteins are only associated with the surface of either side of the membrane, while integral membrane proteins (IMPs) are embedded in the membrane bilayer. IMPs can only be separated from the membrane through the disruption of their hydrophobic interactions with the lipids in the membrane, usually by the use of detergents or denaturing agents. IMPs can be further classified into two groups: integral polytopic proteins and integral monotopic proteins. Integral polytopic proteins have one or more transmembrane domains that span the membrane lipid bilayer, whereas integral monotopic proteins attach to only one side of the membrane and do not cross it. (Blobel, 1980).

IMPs are the most common type of membrane proteins and are responsible for many essential physiological functions. This thesis is focused on one of the most important

receptor families: G-protein coupled receptors (GPCRs).

1.3 G protein-coupled receptors (GPCRs)

GPCRs are integral membrane proteins that comprise one of the largest families of IMPs. More than 900 human GPCRs have been predicted to exist from analysis of the human genome (Fredriksson et al., 2003, Takeda et al., 2002). The seven-transmembrane GPCRs play a critical role in regulating cellular responses to many types of biologically active stimuli with great prospects for medical and therapeutic interference (Zamanian et al., 2011, Takeda et al., 2002). Over 40% of the currently approved pharmaceuticals, including many of the best selling drugs target GPCRs (Eglen et al., 2007). Consequently, GPCRs 3D structures are in demand since these structures could greatly assist in drug discovery processes (Insel et al., 2007). Nevertheless, structure solution of GPCRs to atomic resolution has met with many technical issues and crystal structures have only recently started becoming available. Presently, among the 900 potential human GPCRs only 15 structures are available (White, 2016).

1.3.1 GPCR classification

The first GPCR classification system was constructed by Kolakowski two decades ago, dividing GPCRs into 6 classes (A-F) based on sequence similarity; rhodopsin-like (Class A), secretin-like (Class B), metabotropic glutamate receptors (Class C), fungal mating pheromone receptors (Class D), cAMP receptors (Class E) and frizzled/smoothed (Class F) (Kolakowski, 1994). Although this system is used most frequently, there are many disadvantages, including the fact that two of the classes cannot be found in mammals. A new human GPCR classification system, called GRAFS has been introduced and extensively used in the last ten years, GRAFS,

standing for Glutamate, Rhodopsin, Adhesion, Frizzled/Taste2 and Secretin (Schiöth and Fredriksson, 2005). Class A from the Kolakowski's system and Rhodopsin family from the GRAFS system overlap extensively, with more than 85% of all GPCRs belonging to this class/family (Bjarnadóttir et al., 2006). In this thesis, unless specially indicated, GPCRs will refer to members of the Rhodopsin family (Class A).

1.3.2 Signal transduction mechanism of GPCRs

Being a multi-functional protein family, GPCRs respond to a range of diverse ligands including light, neurotransmitters, biogenic amines, lipids, proteins, amino acids, hormones, odorants, nucleotides, chemokines, and many others (Kroeze et al., 2003). They react to extracellular stimuli by sending intracellular signals via an interaction of their intracellular domains with heterotrimeric G-proteins (Guanine nucleotide-binding proteins). The general signal transduction pathway of GPCRs is shown in Figure 1.1. G-proteins have three subunits: $G\alpha$, $G\beta$ and $G\gamma$. In the inactive state, the G-protein α subunit is associated with the $G\beta\gamma$ complex and guanosine diphosphate (GDP), (Fig. 1.1A). When a GPCR is activated, the receptor undergoes a conformational change (Fig. 1.1B). This causes the receptor to act as a guanine nucleotide exchange factor, resulting in the exchange of GDP with guanosine triphosphate (GTP) in the G-protein α subunit (Fig. 1.1C). After the GTP replaces GDP, the G-protein is activated, resulting in the dissociation of the α subunit and $\beta\gamma$ complex as well as from the receptor (Fig. 1.1D). As a consequence, both the α subunit and the $\beta\gamma$ complex are activated to bind and regulate the activities of their downstream target proteins, passing the signal to different types of second messengers such as cAMP. Following this, the α subunit hydrolyzes the GTP to GDP, and rendering the α subunit inactive (Fig. 1.1E), allowing it to bind to the $\beta\gamma$ complex

(Fig. 1.1A) and terminating thus the downstream signaling events (Pierce et al., 2002).

The diversity in the downstream signaling systems of the G-proteins and the multiple functions of GPCRs can be illustrated by the various types of G-proteins. The three subunits ($G\alpha$, $G\beta$ and $G\gamma$) of a G-protein can be divided into a variety of families. The α -subunits, which establish the fundamental properties of a G protein, have four families: $G\alpha_s$, $G\alpha_{i/o}$, $G\alpha_{q/11}$ and $G\alpha_{12/13}$, and members from each family share some similar functional properties. The $G\beta$ and $G\gamma$ subunits exist as a $\beta\gamma$ -complex, which is

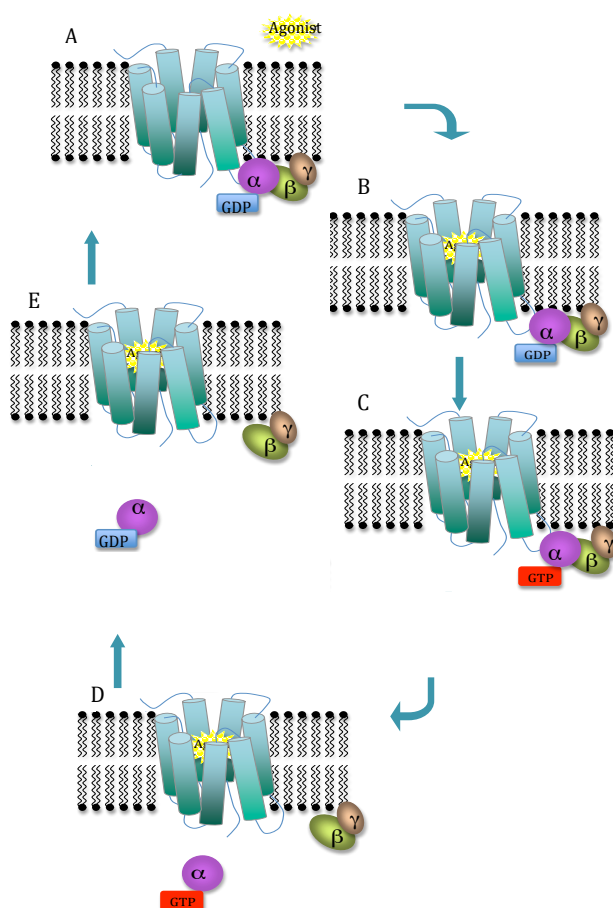


Figure 1.1 Schematic of the general signal transduction mechanism of a GPCR. A: Inactive state of GPCR and G-protein. B: An agonist binds to the receptor. C: The receptor is activated, GDP is replaced by GTP. D: G-protein α subunit and $G\beta\gamma$ complex dissociated and both components are activated. E: GTP hydrolyzed to GDP. Adapted from (Pierce et al., 2002)

combined from 5 types of G β subunits and 12 types of G γ subunits (Hermans, 2003).

The G-proteins anchor to the intracellular membrane by both the α -subunit and the γ -subunit through their covalently attached lipid tails (Bohm et al., 1997).

1.3.3 GPCR functional states and ligand types

The binding of ligands to the GPCR do not necessarily activate the protein. Different types of ligands cause diverse effects as shown in Figure 1.2. Agonists are ligands that trigger a biological response once it binds to the receptor. Partial and weak agonists refers to ligands that only trigger a partial or very weak intracellular response, while full agonists trigger a full response. Inverse agonists are ligands that hinder the receptor's constitutive activity once it binds to the receptor by stabilizing the interactions that keep the receptor in its inactive conformation. A neutral antagonist is a ligand that binds to the receptor without changing the activation state. Inverse agonists, partial agonists or neutral antagonists, which block the binding sites to prevent agonists from binding, are collectively known as antagonists (Tate, 2012).

Agonists binding to a GPCR are not always necessary for interaction with the G-protein as many GPCRs showing constitutive activity (Kudo et al., 1996).

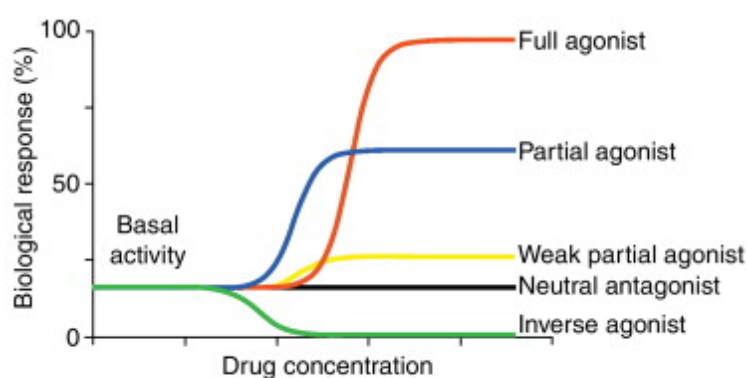


Figure 1.2 GPCR biological response level with the presence of different types of ligands. Adapted from (Tate, 2012).

1.3.4 GPCR structural scaffolds

There are currently a number of GPCR structures solved, most with antagonist bound. The currently available GPCR structures have revealed that GPCRs share very similar structures in the transmembrane domain, but vary notably in their N- terminus, C-terminus and loops. The common scaffold of GPCRs is explained in Figure 1.3 using the histamine receptor H1 as a representative example. All GPCRs have a well-conserved seven-transmembrane helix domain across the cellular membrane, where each helix is composed of about twenty-five to thirty-five residues with a high degree of hydrophobicity (Kroeze et al., 2003). These helices are connected by three extracellular loops and three intracellular loops. Within the transmembrane helical bundle, there is a conserved interaction system built up by 24 non-covalent inter-helical contacts (Venkatakrishnan et al., 2013). This network is conserved among 36 spatial positions occupied by various residues from different receptors as shown in Figure 1.4. GPCRs also have an extracellular N-terminus and an intracellular C-terminus with variations in function and length.

In Figure 1.3, several shared details of GPCRs are emphasized, including common sequence motifs, many of which are involved in structural stabilization. The most conserved motifs are the DR(E)Y, CWxP and NPxxY in helices III, VI and VII, respectively. A disulphide bond formed between two cysteines in extracellular loop 2 (ECL2) and helix III is also conserved. These highly conserved features play important roles in GPCR activation and signal transduction. Some of the interactions between these motifs are referred to as micro molecular switches, including the so-called the ionic lock, the rotamer toggle and the 3-7 lock, all of which are involved in receptor activation (Nygaard et al., 2009, Trzaskowski et al., 2012, Schwartz et al.,

2006). The ionic lock involves the DRY motif on helix III and the D/E residue on helix VI. It stabilizes the receptor via interactions between the two helices (Trzaskowski *et al.*, 2012). The rotamer toggle switch, involving the CWxP motif on helix VI with the aromatic residues Cys, Trp and Pro, modulate the domain around the Pro residue, causing the interacting ligand and its receptor to undergo conformational changes (Bhattacharya *et al.*, 2008). The 3-7 lock, consisting of variable residues in different receptors, is a switch which involves interaction between a pair of residues in TM3 and in TM7 that contributes to the activation of several GPCRs, including the Histamine receptor H1 (Shimamura *et al.*, 2011). The so called “structural hub” helix III has another unique feature, a tilt-angle, which is around 35° to the axis perpendicular to the plane of the lipid bilayer (Venkatakrisnan *et al.*, 2013). It has interactions with the bound ligand, ECL2, ICL2, G-protein and most of the

transmembrane helices. These intense interactions prove its important role in maintaining the structure and function.

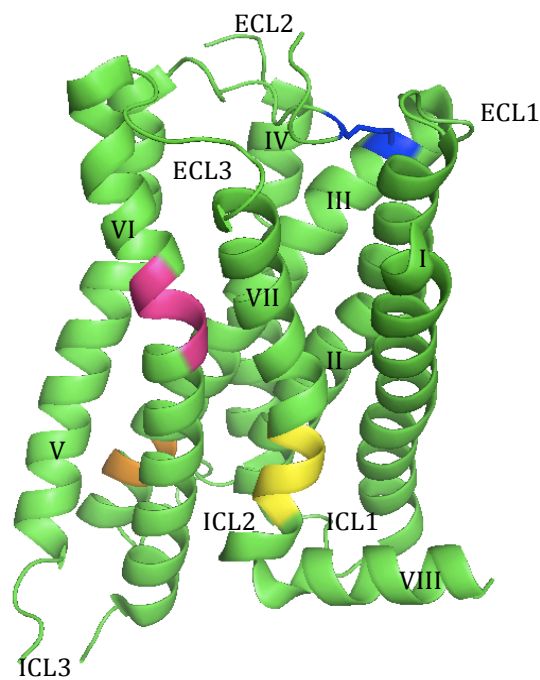


Figure 1.3 Conserved GPCR features. Structure of Histamine receptor H1 (PBD: 3RZE) as a template of GPCRs, conserved motif D(E)RY, CWxP and NPxxY are colored in orange, pink and yellow respectively; two Cystines and the disulfide bond between them are colored in blue.

A number of conserved spatial positions within the ligand binding pocket are observed in nearly all GPCRs. Equivalent residues at these positions from different receptors make contact with the ligands in nearly all class A GPCRs, regardless of residue types. These conserved positions include 3.32, 3.33, 3.36 on helix III, 6.48 and 6.51 on helix VI and 7.39 on helix VII (Venkatakrishnan et al., 2013). Spatial position numbers refer to the GPCR Ballesteros–Weinstein numbering system (Ballesteros and Weinstein, 1995). Within GPCR subfamilies, different receptors also share several conserved positions around the ligand-binding pocket. This feature is explained below by using the aminergic GPCR subfamily as a representative

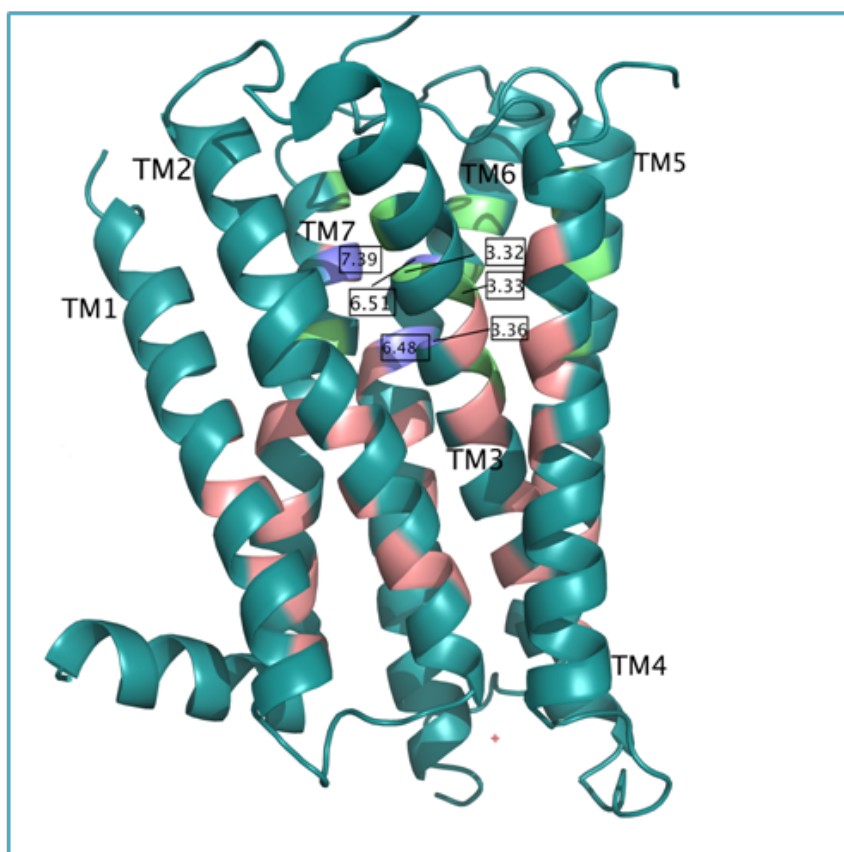


Figure 1.4 Conserved inter-helical contact positions of GPCR. Using the structure of Histamine receptor H1 (PBD: 3RZE) as a representative template of GPCRs. The 36 inter-helical contacts related positions are coloured in salmon, 16 ligand contact related positions of aminergic GPCRs are coloured in green, positions responsible for both helical and ligand contact are coloured in purple. The 6 conserved ligand contact positions across all GPCRs: 3.32, 3.33, 3.36, 6.48, 6.51 and 7.39 are marked with Ballesteros–Weinstein numbers.

template. Comparing the structures of the six receptors: $\beta 1$ AR (Warne et al., 2012), $\beta 2$ AR (Wacker et al., 2010), D3R (Chien et al., 2010), H1 M2R (Haga et al., 2012) and M3R (Kruse et al., 2012), there are 10 spatial positions that have a high probability to be involved in ligand binding events. Equivalent residues at each of these positions have non-covalent contacts with the ligands and are present in more than half of different types of receptors. These positions include 3.28, 3.37, 5.39, 5.42, 5.43, 5.46, 6.52, 6.55, 7.35 and 7.43. Therefore, together with the other six class A GPCRs conserved positions, there are 16 identical ligand binding related positions in aminergic GPCRs (Fig 1.4).

In 2011, the structure of a GPCR in complex with the G-protein was solved. The structure of the $\beta 2$ AR ternary complex with an agonist and a G protein was elucidated for the first time, which presented the non-rhodopsin GPCR transmembrane signaling of a G-protein at the structural level (Rasmussen *et al.*, 2011). The active state of $\beta 2$ AR is stabilized by extensive interactions with the G protein, with the interface between $\beta 2$ AR and the G protein formed by ICL2, helix V and helix VI (Fig. 1.5a) of the GPCR. The main transmembrane conformational changes of $\beta 2$ AR from the inactive to the active state include a 14Å outward movement of helix VI on the

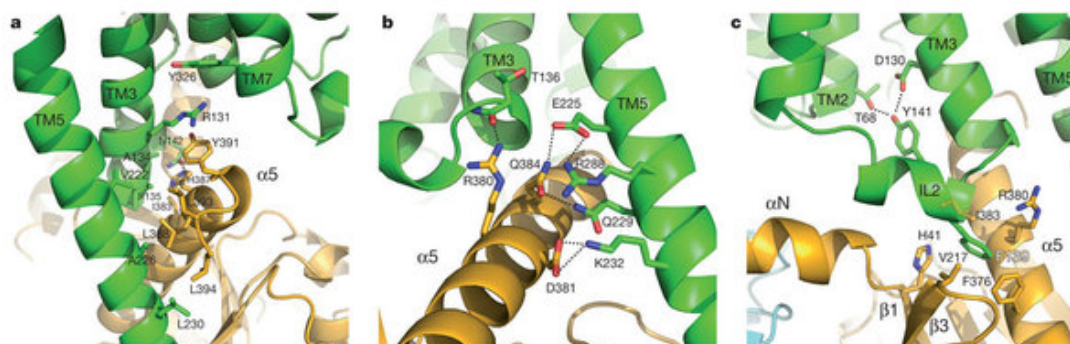


Figure 1.5 Interactions between the $\beta 2$ AR receptor and G protein. $\beta 2$ AR is coloured in green, and G protein is coloured in golden (Rasmussen *et al.*, 2011).

intracellular side and an intracellular extension of helix V by 7 residues. Together, the structural changes on helix V and helix VI lead to the formation of a binding cavity for the G protein. The majority of the sequences involved in G protein coupling are not conserved among GPCRs, possibly due to the diversity of G protein types. Interactions between the G protein and the conserved DRY motif in helix III, and between the DRY motif and the NPxxY motif (helix VII) have both been observed (Fig. 1.5a). The structure also suggests that the receptor residues Thr 68 and Asp 130 interact with Tyr 141 on ICL2 dragging Phe 139 into the hydrophobic G protein binding pocket (Fig. 1.5c), highlighting the important role of DRY motif in receptor-G protein interaction.

1.3.5 Challenges and strategies in GPCR structural studies

X-ray crystal analysis is the modern preferred technique to analyse proteins structural details. High-resolution data for GPCRs can be obtained only when enough pure and stable protein for the necessary formation of crystals is produced. Even after years of research and effort, major challenges in the expression, purification, and crystallization of proteins exist in obtaining well-ordered crystals. Several strategies and tools have been developed and used for GPCR structural studies. These have accelerated determination of GPCR crystal structures in recent years.

1.3.5.1 Expression systems

It is important to choose a heterologous expression system that can produce a sufficient yield of monodispersed protein, a prerequisite for structural analysis. The *E.coli* expression system has many advantages for protein expression, such as low cost and rapid cell growth rates. However, the difference in lipid composition of the *E.coli* membranes compared to eukaryotic cell membranes, can potentially affect the

protein folding state. Consequently, most GPCR structural studies were carried out in insect cells (Tate and Grisshammer, 1996). The advantage of this system is the high yield of homogenous GPCRs suitable for crystallization. However, there are still some drawbacks, such as long growth time and requirement of special equipment and laboratory setup. The work of Asada and Uemura suggested that *Pichia Pastoris* is a suitable alternative host for GPCRs expression, comparable to the insect cells system (Asada et al., 2011), with the structures of H1 (Shimamura *et al.*, 2011) and A2a (Hino *et al.*, 2012) solved using protein produced from the *Pichia* expression system. *Pichia Pastoris* cells can be grown in flasks or in bioreactors. The flask system is flexible and several different expression conditions can be carried out in parallel for the purpose of expression condition screening. The bioreactor system, however, often results in a greater protein yield since the cells can be grown to a much higher density (Singh *et al.*, 2008). Mammalian cells provide an environment most similar to the native membranes containing eukaryotic proteins, increasing the likelihood of expression of functional GPCRs. Despite this, it is difficult and time consuming to maintain healthy cell lines in culture, which is critical for high protein expression levels (Andrell and Tate, 2013).

1.3.5.2 Inherent structural flexibility and instability

Another obstacle to obtaining structural information is the inherent structural flexibility and instability of GPCRs. This is probably the biggest obstacle that needs to be overcome. The inherent flexibility needed to adopt different conformations of the active and inactive states of GPCRs make these receptors intrinsically dynamic, and it is the cause of their instability. GPCRs also have reduced hydrophilic areas, these typically play an important role in the formation of crystal contacts, which in turn result in the production of diffraction-quality crystals.

Protein engineering of the non-rhodopsin GPCRs has involved a series of approaches to obtain more stable constructs, including N- and C- termini truncation, point mutations and third intracellular loop (ICL3) replacement (Chun *et al.*, 2012). The ICL3 in GPCRs is poorly structured and its flexibility causes undesired movements to the transmembrane helix V and VI, which is proposed to make the crystallization difficult. Different strategies have been used to tackle this problem, including fusion body, antibody complexes and mutagenesis to increase thermostability.

In 2007, Cherezov and his co-workers replaced the ICL3 of the β 2 adrenergic receptor (β 2AR) with T4 lysozyme (T4L), the T4L stabilized flexible helix V and helix VI, yielding a 2.4 Å resolution structure of the fusion protein (Cherezov *et al.*, 2007). Since then, there have been 17 GPCR structures solved by the T4L engineering method, and T4 lysozyme-engineering has become the most successful method to date. More recently, the thermostabilized apocytochrome b562 RIL (BRIL) was also used as fusion partner in opioid nociceptin/orphanin FQ peptide receptor. The structure, which previously would not crystallize with the T4L insertion (Thompson *et al.*, 2012), was determined at 3.0 Å resolution (Chun *et al.*, 2012), suggesting BRIL is a valid alternative fusion partner for crystallization of GPCRs. In 2014, two modified versions of the T4L fusion body, dsT4L (disulfide-stabilized T4L) and mT4L (minimal T4L), have been introduced and T4L could be applied even more broadly. The fusion strategy is useful for enlarging the hydrophilic region of the receptor and to avoid the flexibility of the ICL3 domain. However, to stabilize the trans-membrane region, the receptor always has to combine with the high-affinity inverse agonists to reduce the conformational changes associated with its basal activity (Tate 2012). The disadvantage of this strategy is that the ICL3 is located at the site where the G-protein

couples, blocking the insertion of a potential fusion partner at ICL3. Therefore, this strategy cannot be used to reveal the G-protein binding stage of GPCRs (Chun *et al.*, 2012). Another strategy of using T4L is to insert T4L in the N-terminus instead of in ICL3. This strategy has enabled the successful determination of the structure of human β 2AR bonded to a G protein (Rasmussen *et al.*, 2011).

The strategy of thermostabilization by mutagenesis is well established and has been proven quite successful (Tate, 2012). By using mutagenesis, no fusion partner is needed. The receptor amino acids sequence can be modified to improve thermostability and tolerance to short-chain detergents, eventually, to facilitate crystal formation. However, this is both a time consuming and costly process.

1.3.5.3 Crystallization

Membrane proteins in their native state are embedded in a lipid bilayer, from which they are solubilized by the addition of detergents. The traditional aqueous-based crystallization methods (such as vapor diffusion) mixes the protein with various crystallization solutions, leaving it exposed to the precipitant solution, with risk of denaturation and aggregation. The solution conditions would be even more difficult for unstable proteins like GPCRs. It has been observed that Lipidic Cubic Phase (LCP) has been the most successful crystallization method for GPCRs during recent years. In this method a protein-detergent-complex (PDC) is transferred into a lipid matrix that helps the stabilization of the protein. This crystallization method leads to type I crystal packing where protein molecules make contacts via both hydrophobic and hydrophilic regions (Landau and Rosenbusch, 1996).

1.3.6 Structure-based drug discovery (SBDD) and GPCRs

As one of the most important drug target protein family, GPCRs draw a lot of interest in both industry and academic research. Especially in the last decade, extensive research has involved identification of GPCR subtypes, cloning, establishment of expression cell lines, GPCR polymorphisms and others. In spite of these efforts, for a long time, only a limited number of crystal structures have been available. Hence, drug discovery programs have been limited to functional information using tissue-based assays and high throughput screening (HTS). In the last few years, the increasing numbers of reported GPCR structures have made an enormous impact in the field of GPCR drug discovery.

1.3.6.1 Structure-based drug discovery (SBDD)

Structure-based drug discovery (SBDD) is a strategy using the three-dimensional atomic structure of a protein or a ligand-protein complex to obtain information regarding the mode of ligand binding. Design of a new potential drug is then based on this information (Blundell, 1996). There are many examples of successful drugs that were designed based on their target protein 3-D structures. The early examples of SBDD include proteases and kinases, of which the most successful example is the development of HIV (human immunodeficiency virus) protease inhibitors (Wlodawer and Vondrasek, 1998). The HIV protease was identified as a potential anti HIV drug due to its critical role in the HIV life cycle, where it is responsible for the HIV virion maturation by correctly cleaving its polypeptide precursors. The HIV protease inhibitors leave the HIV polyproteins in their immature non-infectious form. The structure of the HIV protease started to appear around the end of 20th century, followed by the development of a range of HIV protease inhibitors, including Saquinavir, Ritonavir and Indinavir, all using structural-based drug design strategies

(Lv et al., 2015).

1.3.6.2 SBDD for GPCRs

Although SBDD had been successfully applied to soluble proteins, the application of the method to GPCRs only started around the year 2000 with bovine rhodopsin as the first ever GPCR structure solved (Palczewski et al., 2000). Subsequently, the use of the method accelerated after 2007, when other GPCR crystal structures started to be solved at an increasing pace. Currently, there are 22 unique GPCR structures solved across all major classes of GPCRs including A, B, C and F (White). This success is due to the application of several methods of crystallization described earlier. The availability of these structures may be assumed to be of equal value to GPCRs drug discovery similar to the structures of soluble proteins.

There are two main strategies for structure-based drug discovery: (1) discovery of new drug series using structure-based virtual screening and (2) optimization of lead candidates guided by the 3D structure. Structure-based virtual screening is aided by docking techniques, which uses a structurally defined ligand-binding site (with or without a ligand) of a target protein to screen a large number of compounds. Information such as shape of the binding site, protein ligand favorable interactions, together with the properties of new ligands could help to find the HIT compounds for further screening. Orthosteric binding sites of GPCRs are particularly suitable for virtual screening. These are deeply buried in the protein when compared with most of the exposed and large binding sites of soluble proteins. This characteristic allows the binding sites in GPCRs fit small molecules with high efficiency. On the top of that, the large number of GPCR ligands in the library is ready to use, also contributed to the success.

With more GPCR structures being solved at a convenient resolution, i.e. better than 3 Å, SBDD of GPCRs will be very helpful in lead optimization. Hopefully, the structure of protein-ligand complexes will provide information on shape, orientation of the ligand and their favorable interactions. This will help medicinal chemists to identify the most suitable areas of the ligands for optimization in terms of solubility and selectivity while maintaining the binding affinity the ligands. Fragment screening is a another frequently applied method for many pharmaceutical companies (Chessari and Woodhead, 2009). However, the disadvantage of this method is the high hit rate, resulting in a heavy downstream workload in verifying the hits.

1.4 Histamine receptors

Histamine is an effective biogenic substance that induces various physiological effects (Axe *et al.*, 2006). Its tight connection to brain function is related to the histamine released by the neurons and projected to all the major regions of the brain (Brown *et al.*, 2001). The actions of histamine are launched by activating histamine receptors as they have histamine as their natural agonist. To date, four histamine receptors have been discovered, including H₁, H₂, H₃, and H₄, which all transduce extracellular signals through different types of G-proteins, namely G_{q/11}, G_s, G_{i/o} and G_{i/o} respectively (Parsons and Ganellin, 2006). The four histamine receptors demonstrate the diverse biologic effects of histamines. H₁ regulates severe allergic responses, H₂ triggers the secretion of gastric acid, H₃ is important for regulation of neurotransmission and H₄ mediates immunological reactions (Leurs *et al.*, 2005) (Axe *et al.*, 2006).

Sequence alignment of the four histamine receptors is shown in Figure 1.6 and table 1.1. Within the trans-membrane region (excluding N-terminus, C-terminus and intercellular loop 3) the sequence identity between H3 and H4 is 50%, while between any other pair of histamine receptors, it is only around 30%. As shown in table 1.1, all the receptors have similar structural scaffolds within the transmembrane region, play an important role in GPCR activation and signal transduction. Within the 16 aminergic GPCR ligand contact related positions, H3 and H4 share up to 81% identity.

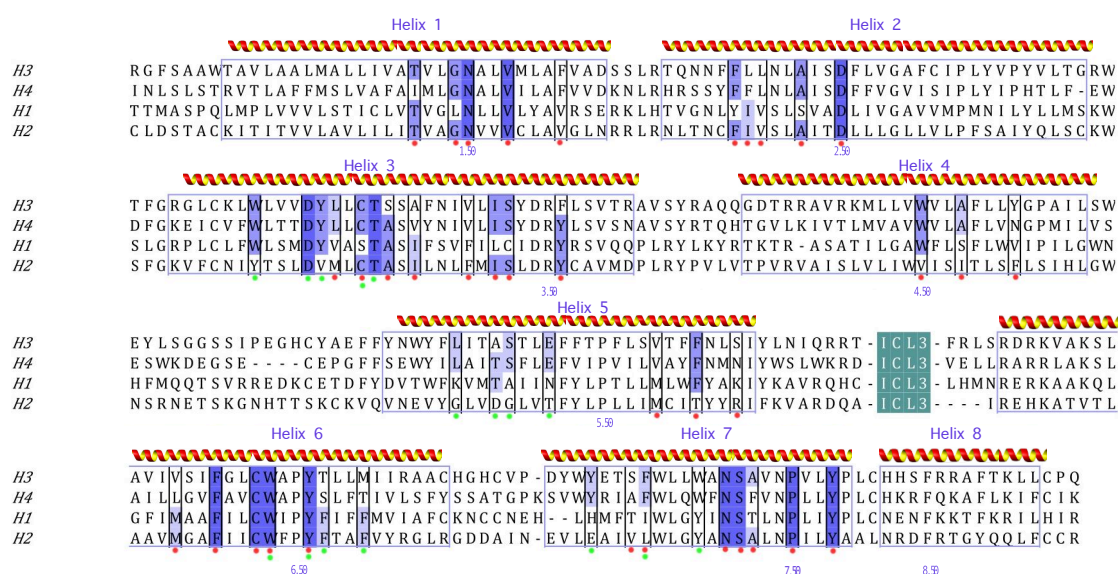


Figure 1.6 Sequence alignment of 4 histamine receptors. The alignment of human H1, H2, H3 and H4 amino acid sequences (N-terminus, C-terminus and intercellular loop 3 excluded). The 36 topologically conserved positions responsible for non-covalent inter-helical contacts are marker with red dots, while the 16 aminergic GPCR ligand-binding related positions are marked with green dots. All the positions are shown as columns in black box. Equivalent residues in the boxes are coloured according to the percentage of identical residues in each column, the darker, the more identical residues. The most conserved residues of each helix are denoted using Ballesteros-Weinstein numbering system.

Table 1.1 Four histamine receptors sequence similarity in different regions. Identical sequences $\geq 50\%$ are marked in shade.

Sequence 1	Sequence 2	Identical sequences (%)		
		Transmembrane region	Conserved inter-helical contact positions	Conserved ligand contact positions
H1	H2	33.0	58%	44%
H1	H4	28.0	44%	44%
H1	H3	30.0	47%	38%
H2	H4	25.0	56%	31%
H3	H2	27.0	53%	31%
H3	H4	50.0	72%	81%

1.4.1 Human histamine H1 receptor (H1)

As one of the most well characterized drug targets, H1 receptor's antagonists are very important in controlling a variety of diseases, especially allergic reactions (Overington et al., 2006). Activated H1 binds to G-protein subfamily $G_{q/11}$ which stimulates phospholipase C, leading to the hydrolysis of phosphatidylinositol 4,5-bisphosphate (PIP2), formation of inositol-1,4,5-trisphosphate (IP3) and diacylglycerol (DAG). IP3 causes calcium mobilization and DAG activates of Protein Kinase C (PKC) (Noubade et al., 2007). H1 also regulates other signaling pathways, but there is currently limited information regarding the cell type-specific signaling pathways (Noubade et al., 2007).

First-generation drugs, however were known to deliver unwanted side effects, including dry mouth, sedation and arrhythmia (Shimamura *et al.*, 2011). Although the x-ray structure of H1 in complex with doxepin has revealed some of the H1 specific ligand binding related residues (Shimamura *et al.*, 2011), GPCRs can accommodate diverse ligands within their binding site in general. Different ligand complexes may possibly present different and more specific binding pocket information. Indeed, more detailed structural information will facilitate new-generation antagonist design, which keeps H1 receptor selectivity while avoiding or reducing the side effects.

The H1 structure shares some common motifs with other GPCRs, such as D(E)RY, CWxP and NPxxY and the disulphide bond between Cys 180 in loop 2 and Cys 100 at the extracellular side of helix III (Shimamura *et al.*, 2011). The seven residues between the disulphide bridge and helix V in ECL2 play an important role, positioned at the entrance of the ligand-binding pocket (Fig. 1.7a). A special property of H1 is

that it contains a seven-residue region that is only a few amino acids longer than other GPCRs, forming a larger space within the ligand-binding pocket, potentially fitting larger H1 antagonists (Shimamura *et al.*, 2011). A second special feature of H1 is a conserved Pro 161 which induces a kink in helix VI leading to a tight i+3 helical turn, which can accommodate the relatively large Trp side chain, an important feature for ligand specificity of H1 (Shimamura *et al.*, 2011).

The ligand-binding pocket of H1 was characterized via the first-generation H1-antagonist Doxepin (Shimamura *et al.*, 2011). To date, the H1 ligand binding pocket is the deepest from all the known GPCR structures (Venkatakrisnan *et al.*, 2013). It was found that some residues that are common among aminergic receptors constitute the main part of the H1 binding pocket. One signature feature of the H1 binding pocket is an anion-binding site, located at the entrance of the ligand binding region. This is constituted by Tyr 431, His 450, Lys 179 and Lys 191 (Fig. 1.7b). A phosphate ion observed in the H1 structure has played a positive role in assisting ligand binding sitting in this anion-binding site (Shimamura *et al.*, 2011) (Fig. 1.7).

It has been shown that H1 antagonists work as inverse agonists on the receptor, thereby weakening its basal activity. The H1-antagonist doxepin makes hydrophobic interactions with the indole ring of Trp 428, potentially stabilizing the hydrophobic packing around helix VI (Shimamura *et al.*, 2011).

An important contraction of the extracellular ligand-binding pocket has been found in the β 2-AR. This is believed to be involved in GPCR activation. With the exception of H1, the antagonists are all much larger than the natural agonist histamine, and the size

enables them to block the activation-related contraction because it is a twisting motion (Shimamura *et al.*, 2011)

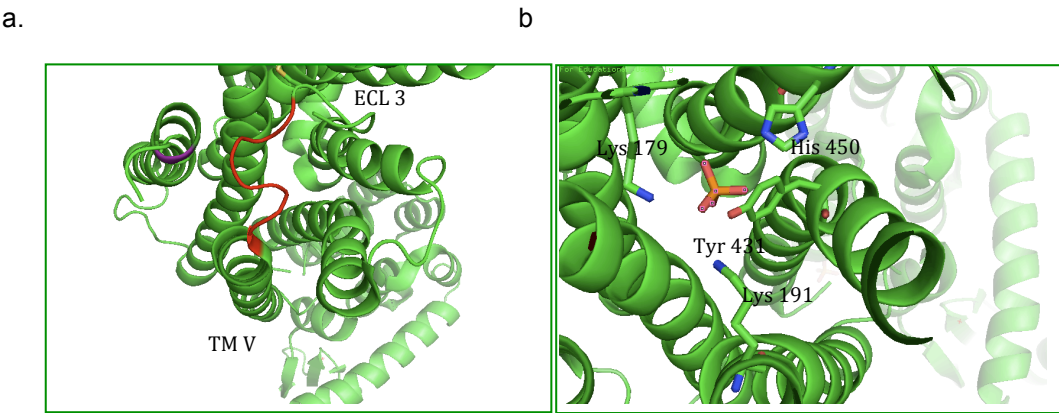


Figure 1.7 Structural key features of H1, viewed from the extracellular side. Seven residues between the disulphide bridge and the helix V in ECL2 are colored in red. (b) Anion-binding site composed of Tyr 431, His 450, Lys 179 and Lys 191, shown as sticks in green colour, a phosphate ion is bind in it, shown in orange.

The crystal structure of H1 with doxepin has revealed information about its binding pocket, as well as raised questions about specific features of the H1 ligand-binding pocket. Doxepin also can bind to H2 and the Muscarinic receptor with different affinities (Figueiredo *et al.*, 1990). A comparison of the ligand contacts in related positions of Histamine receptors are compared to other aminergic GPCRs (Fig. 1.8) shows five best similarities are found, as listed in Table 1.2. H1 and the muscarinic receptor share 56% of their sequences, while H1 and H2 share 44% of them. More

Beta2-AR	W	D	V	V	T	A	S	S	S	W	F	F	N	Y	N	Y
Betal-AR	W	D	V	V	T	A	S	S	S	W	F	F	N	F	N	Y
Dopamine-R	F	D	V	C	T	V	S	S	S	W	F	F	H	Y	T	Y
Muscarinic2-W	D	Y	S	N	T	T	A	A	W	Y	N	V	W	Y	Y	Y
Muscarinic3-W	D	Y	S	N	T	T	A	A	W	Y	N	V	W	Y	Y	Y
H1	W	D	Y	S	T	K	T	A	N	W	Y	F	F	H	I	Y
H2	Y	D	V	C	T	G	D	G	T	W	Y	F	F	E	L	Y
H3	W	D	Y	C	T	L	A	S	E	W	Y	T	M	Y	F	W
H4	W	D	Y	C	T	L	T	S	E	W	Y	S	T	Y	F	W

Figure 1.8 Alignment of residues at the 16 ligand-binding related positions of aminergic GPCRs. Equivalent residues are coloured according to the percentage of identical residues.

Sequence 1	Sequence 2	Identical residues (%)
H3	H4	81
H1	Muscarinic2 R	56
H1	Muscarinic3 R	56
H1	H2	44
H1	H4	44

Table 1.2 Sequence similarity within the 16 ligand-binding related positions of aminergic GPCRs.

detailed interactions within the binding pocket are clearly needed.

1.4.1.1 Antihistamines

Antihistamines are widely used medications for the treatment of allergic rhinitis, allergic conjunctivitis, urticaria, coughs, colds and insomnia (Church et al., 2010). They are capable of down-regulating the H1 histamine-activated reactions on the endothelial cells, the airway smooth muscle and some nerve fibers. This helps to diminish the vascular permeability, with reduced release of fluid and improved peripheral resistance, bronchodilation and ease the mucosal and cutaneous itching. (Simons, 2003)

The current antihistamines on the market could be classified into two groups: classic antihistamines (also called first-generation antihistamines or sedating antihistamines) and non-classic antihistamines (also called second-generation antihistamines or nonsedating antihistamines), depending on whether or not they have an effect on the central nervous system (CNS).

Classic antihistamines generally have small molecular weight and are highly lipophilic, like in the case of diphenhydramine and hydroxyzine. They are able to cross blood brain barrier (BBB) and affect the CNS, thereby causing sedation.

The non-classic antihistamines, such as terfenadine, astemizole, loratadine, cetirizine, azelastine and fexofenadine, on the other hand, are not lipophilic and cross the BBB to a minimal degree, rarely creating sedation.

1.4.1.1.1 Classic antihistamines

The classic antihistamines efficiently relieve allergic symptoms, emerged on the market in the 1940s-1950s (Emanuel, 1999). Unfortunately, they were also quite toxic in many ways. Firstly, they were well-known sedative antihistamines. This class of drugs can go across BBB and disrupt the histamine-regulated neurotransmission at CNS, which can cause sedation, drowsiness and fatigue, resulting in much diminished cognitive function such as learning performance and the ability to safely operate machinery. (Church et al., 2010) As well as from the sedation side effect, classic antihistamines also have anti-cholinergic side effects. The initial antihistamines share the basic chemical structure with other types of medications, such as cholinergic muscarinic antagonists, neuroleptics and antihypertensive drugs (Church et al., 2010, Emanuel, 1999). Moreover, H1 also has high sequence similarity with other aminergic receptors, having muscarinic receptors as the highest, with around 45% sequence identity (Simons and Simons, 2008). Therefore, this class of antihistamines has non-specific selectivity, interacting with e.g. the muscarinic receptor, the adrenergic receptor, the dopamine receptor, etc. (Church et al., 2010, Emanuel, 1999), causing undesirable side effects such as dry mouth, tachycardia and urinary retention (Bousquet et al., 2001).

1.4.1.1.2 Non-classic antihistamines

Since the 1980s, a new generation of antihistamines has been developed via structural modification of existing drugs (Simons and Simons, 2008). Compared to the old drugs, they have better selectivity and rarely cross the BBB because of their larger molecular weight and reduced lipophilicity. Hence, the related side effects have been

generally overcome.

Nevertheless, other side effects have emerged among some of these agents, especially the cardiotoxic effects. In 1986, it was reported that an overdose of terfenadine or astemizole could lead to ventricular arrhythmia, namely torsades de pointes (TdP). These antihistamines work as a blocker to certain type of potassium channel, which mediates the ventricular repolarization, and results in a prolongation of the QT interval leading to for example ventricular arrhythmias (Camelo-Nunes, 2006). However, the cardiotoxic effects are not class specific. There are other agents, such as loratadine, cetirizine, azelastine, and fexofenadine that are free from torsades de pointes or other ventricular arrhythmias. (DuBuske, 1999)

1.4.1.1.3 Cetirizine

As an example, cetirizine is a widely used non-classic antihistamine, with the trade names such as Zyrtec and Reactine. The compound is the carboxylated metabolite of hydroxyzine, and consists of a diphenylmethane moiety directly connected to a piperazine moiety and an alkyl group with a carboxyl group as shown in Figure 1.9. Hydroxyzine has an alcohol in place of the carboxyl group of cetirizine.

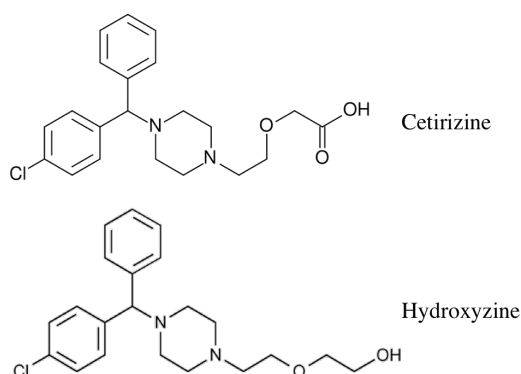


Figure 1.9 Cetirizine and Hydroxyzine

1.4.1.1.3.1 Sedating or non-sedating antihistamines

When we talk about the non-sedative effect of second-generation antihistamines, the concept of brain penetration has to be mentioned, which is the direct reason of sedation.

To determine the ability of brain penetration of a compound, Overton and Meyer's theory was applied early in the 20th century. According to this theory, the compound with a higher level of lipophilicity can go through the blood-brain barrier more easily. However, this rule has proved to be too simple, as more parameters need to be taken into account. For instance, the level of ionization of a compound at physiological pH should also be considered because only the neutral species of the compounds pass across the BBB. Another very important parameter is the hydrogen bonding capacity. A higher hydrogen bonding capacity means that a larger portion of the compound can bind to the serum protein before reaching the BBB, bringing down the level of brain penetration. The last factor is the P-glycoprotein, which plays an important role as active transporter. All the parameters should be considered together for the final assessment of the level of brain penetration.

Furthermore, molecular weight is also a key parameter to consider, large molecules have difficulties with going through the BBB. However, this argument is only valid to a certain level because the effect of size can be negligible when the molecular weight is below 500 Da.

Compared to the first generation antihistamines, cetirizine has a relatively low lipophilicity (100-fold lower than hydroxyzine) and a large hydrogen bond capacity.

At physiological pH, cetirizine is mainly in the zwitterionic form. All these physicochemical properties together with its high efflux rate render cetirizine non-sedative.

1.4.1.1.3.2 Selectivity

Both cetirizine and hydroxyzine are very potent, even though cetirizine has a 3-fold lower affinity. When it comes to selectivity, though, cetirizine is superior over other receptors (serotonin receptor, α 1-adrenergic receptor and D2 dopamine receptor), while hydroxyzine is weak in selectivity. Without a structure of H1-cetirizine complex at atomic level, the detailed interactions between the compound and the receptor remain unknown.

1.4.2 Human histamine H3 receptor (H3)

The H3 receptor inhibits histamine synthesis and release. It also negatively regulates the release of several other neurotransmitters, such as acetylcholine, dopamine and noradrenalin (Stark *et al.*, 2001). The H3 receptor antagonists might offer a cure for memory disorders, narcolepsy, ADHD, arousal and cognition (Axe *et al.*, 2006). The recently discovered receptor is located in the central and peripheral nervous systems. Its potential as a pharmaceutical drug target has invested H3 structural studies a high priority task. Structural information is also expected to improve the understanding of the specific ligand-binding behavior of other histamine receptors, such as H4, which has 81% identical residues around ligand binding positions.

Because of the important role of H3 in drug development, functional research and computational work has been carried out since H3 was first cloned in 1999 (Lovenberg *et al.*, 1999). The outcomes provide some structural related information.

Atsuo and his co-workers have demonstrated that the carboxyl-terminus of H3 also plays a role in receptor inactivation, in particular the amino acid residue F8.50, which is involved in the transition of the receptor to its intermediate conformation, which is either ready to be activated, or inactivated (Kuramasu *et al.*, 2011). F8.50 is believed to have interaction with the NPxxY motif to maintain the helix VIII at the right angle. The amphiphilic manner stimulates the RRTK polar residues in H3 to gather on one side, and the FFLL hydrophobic residues to move on the other side. The hydrophobic interface of helix VIII is responsible for the regulation of H3 activity, perhaps due to the stabilizing interaction between the interface and the plasma membrane. This region might be a potential target domain for H3 drug design. This information is useful for H3 construct design in structural studies.

1.5 Macromolecular Crystallography

Currently several techniques exist that provide spatial information on the atoms in macromolecules, like the membrane proteins studied in this thesis. After over 100 years since its creation, X-ray crystallography is still considered the best experimental technique capable of revealing structural details at an atomic level. Perhaps the most important reason to try and determine the 3D structure of a macromolecule is the understanding of its biological function. Furthermore, detailed structural information, for example, the interactions between proteins or between proteins and other types of molecules, has accelerated the pace of drug discovery for specific targets.

X-ray crystallography is a mature, but highly technical discipline. Membrane proteins introduce a further level of complication because, in general, they form less ordered crystals than soluble proteins. This feature is reflected in the acquisition of X-ray data of poorer quality and lower resolution, causing structure determination of membrane

proteins to be intrinsically more complicated than structure determination of soluble proteins.

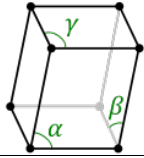
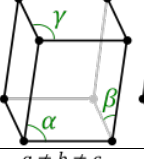
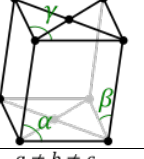
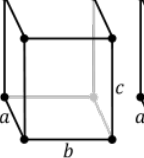
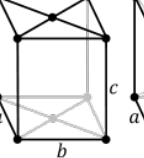
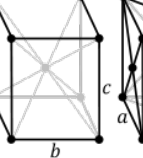
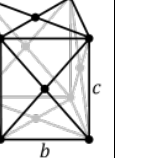
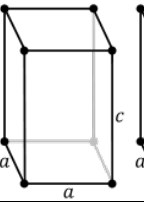
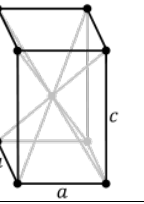
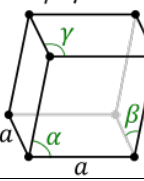
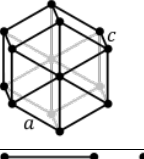
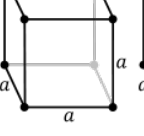
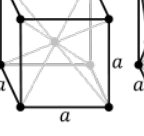
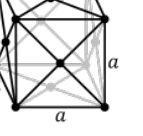
In this section the main aspects of X-ray diffraction from crystals of macromolecules will be outlined and the handling of the experimental data produced with this technique will be described through the use of some of the well-known computer programs used by the crystallographic community. The most important goal of crystal analysis is structure determination. What emerges from this analysis is a likely model of the macromolecule which can be validated both by the agreement with experimental data and with prior biological, chemical and stereo-chemical information.

1.5.1 Why crystals?

Ordinary light has a wavelength much larger than the typical size of the atoms forming a macromolecule. This is why proteins cannot be seen with conventional microscopes and, rather, need light of a much smaller wavelength. X-rays are best suited to provide atomic details for molecules and macromolecules. The scattering of X-rays by individual molecules yields weak intensities and, thus, cannot practically be used to produce detailed images. However, in recent years this is becoming more feasible thanks to the advent of powerful coherent X-ray sources like those produced by the X-ray Free Electron Lasers (XFELs). In order to bypass the limitation of the signal noise ratio by a single molecule, orderly arrangement of several billion macromolecules in a single crystal can notably amplified the desired signal. In each crystal all the identical molecules are arranged in a three-dimensional array with a periodic order, called the *lattice*. The smallest repeating unit of this lattice is known as *unit cell*. There is no real physical lattice in a crystal, but this is a convenient

geometrical framework to simplify all theories and calculations connected with molecular analysis via X-rays. In fact, the way molecules pack together to form the final 3D crystal structure involves symmetrical arrangements of the same molecules via, for instance, rotations and roto-translations. Such symmetrical constructions are described mathematically with the so-called *crystallographic space groups*. These have been studied in detail and are listed in the International Tables of Crystallography. The importance of symmetry space groups in crystallography stems from the effect they have both on diffraction data and on macromolecular models. It is beyond the scope of this thesis to describe space groups in detail, but there are several books where the argument is treated in an approachable way for biologists. Here it is important to remember that the full content of a unit cell derives from symmetry operations performed on sub-units of the unit cell itself, called *asymmetric units*. It is also important to remember that length of the sides and angles of the unit cell are restrained by the specific space group. More specifically, space group symmetry determines the type of lattice and, thus, of unit cell compatible with that symmetry. There are in total 230 space groups which are comprised of a total of 14 lattices, known commonly as *Bravais lattices*, which could be grouped into 7 lattice system types (Table 1.3).

Table 1.3 Seven lattice systems and 14 Bravais lattices

Lattice systems (7)	Bravais Lattices (14)
Triclinic	$\alpha, \beta, \gamma \neq 90^\circ$ 
Monoclinic	$\beta \neq 90^\circ$ $\alpha, \gamma = 90^\circ$  
Orthorhombic	$a \neq b \neq c$    
Tetragonal	$a \neq c$  
Orthorhombic	$\alpha = \beta = \gamma \neq 90^\circ$ 
Hexagonal	
Cubic	  

1.5.2 X-ray diffraction theory in a nutshell

When monochromatic X-rays hit a crystal, they are scattered by the electrons of the atoms within the crystal lattice. Most of the scattered waves cancel each other by

destructive interference. However, some of them interfere constructively, giving rise to so-called *diffracted intensities* that can be captured and measured by appropriate detectors (Fig. 1.10). The spatial arrangement of the spots produced by the captured intensities can be used to determine the type of crystal lattice and the size of its unit cell. The relative strength of these spots (intensities) can be used to determine the arrangement of atoms in the asymmetric unit.

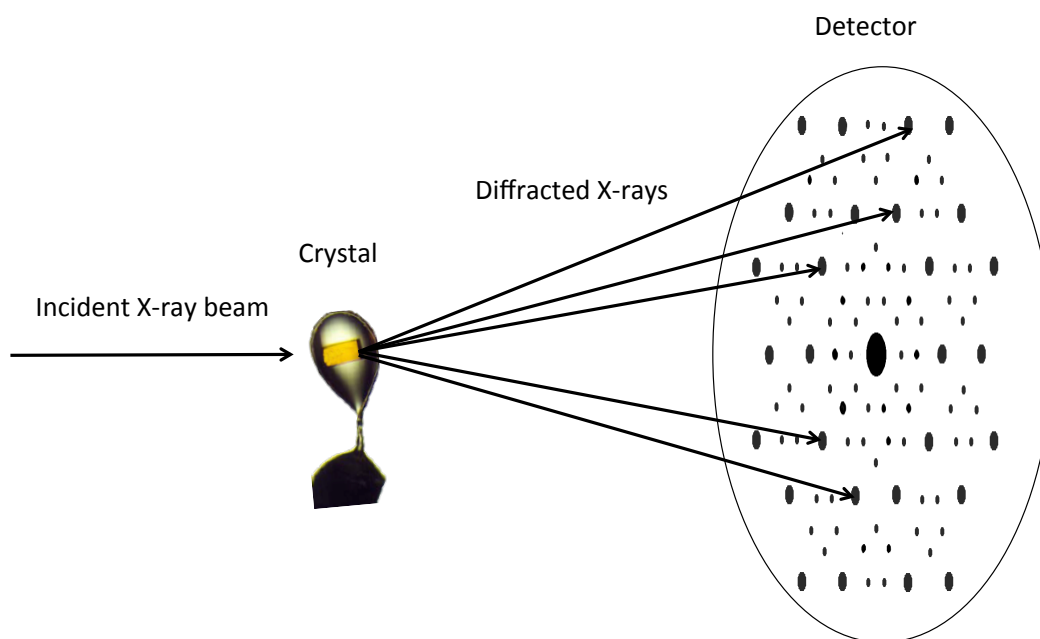


Figure 1.10 A simplified schematic illustration of data collection. A monochromatic X-ray beam is diffracted by a protein crystal, and the resultant diffractions are collected on a detector.

The whole process can be conveniently described using the concept of the Fourier transform. The 3D arrangement in space of the intensities scattered by the crystal can be geometrically described by another lattice called the *reciprocal lattice*. In mathematical terms the Fourier transform of the crystal lattice is the reciprocal lattice. If the sides and angles of the reciprocal lattice (they could be called sides and angles of the *reciprocal unit cell*) are measured from the diffraction images, the sides and angles of the direct lattice can be calculated via the Fourier transform. In fact, once the spots relative intensities are also taken into account, an *inverse Fourier transform*

can be used to determine the atomic structure forming the lattice. To be more specific, the scattered intensities (also known as *reflections*) represent only half of the information contained in the Fourier transform. The other half (phase information) is lost during the scattering process. Thus a starting model needs always to be “guessed” in order to make use of the inverse Fourier transform to determine a crystal structure. Such a “guess” can be determined by a detailed knowledge of the problem (as, for instance in this work, where an initial model for the protein was known and only the ligand molecule needed to be determined), or via probabilistic calculations whose details go beyond the scope of this thesis (for example direct methods, SAD phasing, molecular replacement, etc.).

The reciprocal lattice points are uniquely described by a triplet of integers, known as *Miller indices*, usually designated as (h,k,l). The product of the X-ray diffraction of a crystal structure is a series of reflections (lattice points) given as Miller indices, with each reflection being associated to a diffracted intensity I . In a typical diffraction process by a macromolecule, several tens of thousands of intensities are measured during data collection, resulting in multiple measurements of the same reflection. It is this data redundancy that leads to improved accuracy in the measurement of I , which greatly facilitates structure determination, even when the phase information is lost during the diffraction process.

1.5.3 The main steps in the structure determination process

All efforts detailed in the present thesis to purify the proteins and obtain crystals that diffracted to a sufficient resolution, have been made in order to collect X-ray data, ultimately leading to structure determination. The full process of structure determination consists of tasks that can be different depending on the type of

macromolecular structure to be solved. In this section, the main steps followed to determine the structures of H1-cetirizine and H1-fenoxedine will be illustrated in details. As most of the tasks have been carried out using computer programs, these will be introduced in conjunction with the different structure determination steps.

The whole structure determination work essentially consists of:

1. Data collection
2. Data processing
3. Data combination
4. Initial model determination
5. Model completion and model refinement

1.5.3.1 Data collection

When X-rays hit a macromolecular crystal, positively interfering scattered radiation in general occurs only along a small number of directions. Accordingly, only few diffraction spots are collected on the detector, when the crystal is kept still, with a fixed orientation. If the crystal rotates continuously around an axis, new scattering directions will occur at every new orientation of the crystal and many more reflections will be captured by the detector. To collect diffraction data from a crystal, it is necessary to mount the crystal in a stable frame and to rotate this frame when X-rays hit the crystal. The exact rotation range in this method of data collection (*the rotation method*) will depend on the crystal symmetry with higher symmetry space groups require smaller overall rotations to collect a complete data set. One important indicator of diffraction quality for macromolecular crystallographers is the distance from the center of each image and the position of the farthest away diffraction spots. The more distant spots provide higher resolution information. The crystals of H1-cetirizine and H1-fenoxenadine described produced diffraction images with spots at a

maximum of 3 Å resolution. This is a reasonable resolution for crystals of membrane proteins and sufficient to build an atomic model. One of the most evident visual effects on diffraction images, as data collection proceeds on a single crystal, is the gradual disappearance of diffraction spots at high resolutions. This is due to the radiation damage caused by the energetic X-rays destroying chemical bonds in the crystal creating free radicals, that in turn destroying more chemical bonds. The overall effect, as the duration of exposure to X-rays increases, is the deterioration of the orderly lattice structure in the crystal and resultant loss of resolution. After a sufficiently long exposure time, eventually the disorder caused by radiation in the crystal lattice is so widespread that the crystal does not diffract any longer. The crystals of the membrane protein used in this study survived only briefly under the intense flux of the microfocused beam at the beamline I24 (Diamond Light Source). Thus, sufficient rotation sweeps to collect a complete data set only could be achieved from a relatively few crystals. Complete and statistically meaningful sets of diffraction intensities could be obtained with the combination of many incomplete (partial) sets of data, as explained below.

1.5.3.2 Data processing

The geometric spots of the 2D images produced during data collection must be turned into intensities associated with the 3D reciprocal lattice related to the crystal lattice. This task is carried out by the so-called *integration programs*, software specifically designed for integration of X-ray data. For the data collected in this thesis the integration program used has been XDS (Kabsch, 2010). XDS select a certain number of diffraction spots across all images and use their location in the 3D reciprocal space to determine cell parameters, lattice type and crystal orientation. Based on this the program predicts the position of all potential diffraction spots in the reciprocal space.

If the majority of the observed spots match such predictions, the program assumes that the specific choice for space group and cell parameters is correct and moves on to measure the intensity of each reflection. This involves, determining the spatial extent of each diffraction spot in order to add up the numerical strength of all pixels forming the specific spot, resulting in the *integrated intensity*. After all reflections have been integrated, the program outputs a file containing a list of integrated intensities with associated Miller indices and an estimate of the error in the intensity measurement. Such files are known as *unmerged reflection files*. The symmetry of the crystal lattice is transferred to a higher symmetry in the reciprocal lattice. For this reason, in general, all intensities are partitioned in groups of equivalent reflections. Due to the peculiar nature of the data collection experiment, intensities within each group are not exactly the same. Various and specific experimental issues cause the breaking of symmetry equivalence and, as a consequence, intensities of symmetry-related reflections, supposed to have exactly the same value, are different. For this reason it is necessary to correct numerically all groups of equivalent integrated intensities so that, after this operation, they are more similar with each other. This specific operation, in crystallography, is called *scaling of data*. The programs here used to carry out data scaling are *AIMLESS* included in the CCP4 suite (Collaborative Computational Project, 1994). All scaled intensities are written out to a new type of file as *merged intensities* (indicating the fact that they contain averaged values of each intensity). Structure factor amplitudes are extracted from scaled intensities and the merged reflections files are used for structure determination during phasing, model building and model refinement.

To conclude this section, it is important to report the way in which data quality is

assessed in macromolecular crystallography. It is clear that if all symmetry-related intensities, after scaling, are very similar with each other, then their average, used as a representative scaled intensity and, thus, a representative structure factor for the specific reflection, can be determined with high precision. A measure of such precision is provided by three types of so-called *R statistics*, the R_{merge} , R_{meas} and R_{pim} (Evans, 2006). They are given by the following formulas:

$$R_{meas} = \frac{\sum_h \sum_l |I_{hl} - \langle I_h \rangle|}{\sum_h \sum_l \langle I_h \rangle}$$

$$R_{meas} = \sum_h \sqrt{\frac{n_h}{n_h - 1}} \sum_l |I_{hl} - \langle I_h \rangle| / \sum_h \sum_l \langle I_h \rangle$$

$$R_{p.i.m.} = \sum_h \sqrt{\frac{1}{n_h - 1}} \sum_l |I_{hl} - \langle I_h \rangle| / \sum_h \sum_l \langle I_h \rangle$$

Where the index l refers to all reflections with same Miller index h . I_h is the scaled observed intensity, $\langle I_h \rangle$ is the average of all scaled observed intensities with the miller index h , and n_h is the number of symmetry-related reflections for miller index h . While R_{merge} has been traditionally used by crystallographers for many years, in recent times it has been realized that this value does not take into account crystal symmetry. R_{meas} obviates to this bias. The R_{pim} , like R_{meas} , is not affected by symmetry, but, in addition, it takes into account how many symmetry-equivalent reflections, on average, were used for scaled intensities. Values of all the above statistics close to zero indicate data with a high level of precision. For soluble proteins it is likely to find values around 0.05, representing a 5% variation in measurements of the unmerged intensities, while for membrane proteins it is more likely to find values above 0.15-0.20. Data quality is also determined by the completeness in reciprocal space. This is essentially how many of the possible reflections in reciprocal space are

represented by the data set and is measured in percentage values. Data completeness is related to details and geometric distortions of the electron density. It is normally preferred to have datasets with completeness above 90%. To finish, data resolution is, quite obviously, important to convey a measure of data quality. Traditionally this was measured as signal-to-noise in resolution shells. This means that all data are partitioned in groups (shells) with similar geometric resolution and ratio of averages of intensities and their errors are calculated for each shell. Numerical values for this signal-to-noise indicator (termed $Mn(I/sd(I))$) start high and drop to small values as the geometric resolution increases. The resolution at which they reach the 1.5-2 limit, is considered the real resolution of the scaled data. For membrane proteins, and the H1 molecules investigated here are no exception, resolutions of potentially solvable structures are normally found between 3 and 4 angstroms. Recently another indicator of resolution the $CC_{1/2}$ has been shown to be a useful. This is defined in the following way. Each group of symmetry-related reflections is randomly split in two halves and intensity averages are computed for each one of the two halves. The whole set of scaled data is, thus, split in two halves and correlations between the two halves, the $CC_{1/2}$, can be computed for each resolution shell. If intensities reflecting the structure have been obtained from the scaling process, then both random halves will include such intensities and will display high correlation. When the noise starts to become comparable to the structural information (for instance, because it is measured at resolutions higher than the effective resolution of collected data), then the correlation will be low and $CC_{1/2}$ will start to decrease. AIMLESS suggests data truncation when the $CC_{1/2}$ falls below the 0.3 threshold. A more conservative choice prefers 0.5, rather than 0.3.

1.5.3.3 Data combination

As previously mentioned most of the individual integrated datasets (i.e. data from a single rotation sweep) were not sufficient for structure determination. It has been possible to obtain complete datasets starting from the many incomplete ones, using one of the programs included in the CCP4 suite, called *BLEND* (Foadi et al., 2013).

BLEND makes use of the cell parameters of each crystal to

determines how much they differ from each other. The overall scenario of distances is summarized in a tree-like structure of similarities called *dendrogram*. (Fig. 1.11). The process through which a dendrogram is calculated is called cluster analysis and the various groups connected in a dendrogram are called clusters.

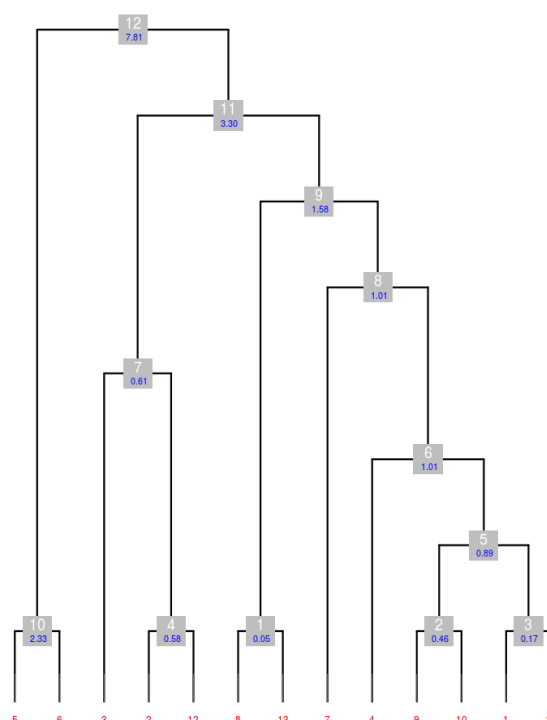


Figure 1.11 A tree-like dendrogram.

BLEND produces scaled intensities for each cluster of the dendrogram. Each cluster is more complete than the component datasets and sufficient completeness is obtained for clusters in the upper part of the tree. The scaling of each cluster, in *BLEND*, is carried out using *AIMLESS* and *POINTLESS*. All statistics indicating data quality are reported for each cluster so that it is possible to judge and select the most convenient ones. Combinations of data different from the clusters are possible, as it will be explained in the description of data combination reported for the H1 structures in

Chapter 2. An important parameter used in BLEND is also an intuitive measure of clusters isomorphism. Clustering is achieved using the full set of unit cell parameters, but these do not provide an intuitive feeling for cells' similarity, because they are more than just one number and because various linear transformations are applied to the original parameters before computing the dendrogram. To obviate this complication, an *absolute Linear Cell Variation* (aLCV) parameter has been added to the program to provide a linear absolute difference (in angstroms) between two or more unit cells. Consider two unit cells, A and B. Three diagonals, D_aA , D_bA and D_cA are computed across the three faces of unit cell A. Similarly, three diagonals, D_aB , D_bB and D_cB are computed across the three faces of unit cell B. The largest difference among D_aA , D_bA , D_cA or among D_aB , D_bB , D_cB is the aLCV. In a cluster with several unit cells, the highest aLCV computed across all couples in the cluster is the aLCV for the cluster.

1.5.3.4 Initial data determination

Molecular replacement was used to recover the information lost in the diffraction process: the phases of the structure factors. An initial model of the ligand free H-T4L structure and, accordingly, an initial set of phases, was already available (Shimamura et al., 2011). More specifically, both structures here solved are a combination of the H1-T4L complex with two different ligands. As these ligands form a very small fraction of the total structure, the solved model of H1-T4L will provide most of the information necessary to solve both H1-cetirizine and H1-fenoxenadine. What has been important for the structural work presented here is the alternation of model refinement and model completion carried out in several cycles and described in the following section.

An important assessment of the agreement between the initial model and data is provided, once more, via R-statistics. This time the differences will be calculated between structure factor amplitudes from the model and structure factor amplitudes from the experimental data. Values, before refinement, of around 0.4 or less indicate some substantial agreement between the model and the structure, certainly a good starting point for structure refinement. But after completion, the modified (and completed) model is expected to reach an agreement with R around 0.2 or better. In recent years it has been found useful to adopt cross validation methods from statistics. Essentially, a number of reflections (around 10% of the total) are kept out of all crystallographic calculations (model completion and refinement), but they are then used to calculate the R statistics. Using the jargon developed by crystallographers, the R statistics calculated with most of the reflections is called R_{work} while the one calculated using the unused reflections is called R_{free} . What is normally observed in a successful model completion / refinement process is both R_{work} and R_{free} decreasing towards low values. When the R_{work} keeps decreasing, but the R_{free} remains stable, this corresponds to an overfitting of the model parameters because the model is not entirely matching all the data and, rather, the structure is being biased towards the data used for refinement. In such instances it is better to reconsider modifying some parts of the model.

1.5.3.5 Model completion and model refinement

The initial model selected comes in the form of a so-called PDB file, an ASCII file including Cartesian coordinates of all atoms forming the structure and the corresponding isotropic vibrational motion around their centers, typically represented by positive quantities known as *B factors* or *thermal factors*. The higher the B factor,

the wider is the atom's vibrational motion. When the atoms coordinates and B-factors are used in conjunction with the experimentally observed structure factor amplitudes (indicated as F_o), an inverse Fourier transform can calculate the electron density map (indicated as ρ) that represent a first approximation of the structure to be determined. In short, atomic coordinates and B factors can be used to calculate structure factors $F = F_{\text{exp}}(i\varphi)$, where F represents the structure factor amplitude and φ the structure factor phases. From the initial atomic model, via the direct Fourier transform, calculated structure factors $F_c = F_c \exp(i\varphi_c)$ can be calculated. If an inverse Fourier transform is calculated with these quantities, the resulting electron density map exactly matches the initial model. As the purpose of the initial model is to approximate the structure to be determined, the initial electron density map needs to be somewhat different from the one reproducing the initial model. The phases φ_c always will have to be used to transfer structural information from the initial model, but the amplitudes F_c can be replaced. An option would be to use F_o , but it has been demonstrated that electron density maps calculated using $2F_o - F_c$ as structure factor amplitudes yields electron density maps least biased towards the initial model and reproducing the maximum possible amount of details of the structure to be determined. A useful, but different, electron density map is obtained using $F_o - F_c$ as structure factor amplitudes. In this map the density corresponding to the initial model is essentially completely removed and the only density visible is the one corresponding to those atoms of the structure to be determined not contained in the initial model, and also tells if atoms are in the wrong place.

The initial $2F_o - F_c$ map mostly has regions of electron density in which fits the atomic model. But there are also empty parts of density as well as parts of model not fitting

any density. This is what can be called an *incomplete model*, because it is mostly a complete representation of the final structure, but still needs some additional details. Such details must be provided manually by the macromolecular crystallographer, both trying to guess the best orientation to fit specific ligands and minor molecules, both using stereochemical expertise. The model building/model completion program used to carry out this task for the two H1 structures of this thesis is COOT (Emsley et al., 2010). At each stage of partial addition to the initial model, all atomic coordinates and B factors have to be refined to achieve the best agreement with the experimental structure factor amplitudes and sensible existing stereochemical information. This task is performed by so-called *refinement programs*. The program used for the present work is the *refine* module in the *PHENIX* platform (Adams et al., 2010).

1.5.4 Protein crystallization

Obtaining well-ordered protein crystals is essential for protein structure determination through X-ray crystallography. In a protein crystal, millions of identical protein molecules are stacked up in a periodic lattice, this lattice acting as a three-dimensional diffraction grating when exposed to X-rays. Well-ordered crystals produce diffraction to higher resolution, allowing more detailed protein structures to be determined.

In a crystallization experiment, the starting protein sample is in an aqueous environment, where proteins remain folded and functional. During the crystallization course, protein solubility is gradually changed in order to shift the system to a favorable state for crystallization. As shown in the protein solubility diagram (Fig. 1.12), protein solution starts from an undersaturated state, which is transitioned into the supersaturation state, or more precisely, the nucleation zone, followed by crystal growth.

By increasing the protein or precipitant concentration, or both at the same time, the system shifts to the nucleation zone, where proteins form small microscopic clusters, referred to as nuclei. The nucleus could either lead to disordered structures, such as aggregates and precipitates, or to crystal growth. Crystal size will increase by consuming the free protein molecules in solution, until equilibrium is reached.

Crystallization is a major barrier for crystallography, involving many challenges. Firstly, requiring a pure homogeneous protein sample: a rule of thumb being more than 97% purity. Furthermore, the whole process needs to happen under in an aqueous environment. Protein crystals are normally very fragile because of molecules are held by noncovalent bonds, such as hydrogen bonds, salt bridges, and hydrophobic interactions. This means that any movement of vibration might disturb the crystal growth. There are also many other factors that are known to affect crystal growth, including ionic strength, pH, temperature and the presence of additives. The chosen crystallization method and experimental setup could also have some influence. For membrane protein, the type of detergents is also an important consideration.

In order to identify appropriate conditions for growing well-diffracting protein crystals for a given protein, all the factors mentioned above need to be tested and combined, which involves multiple steps of screening.

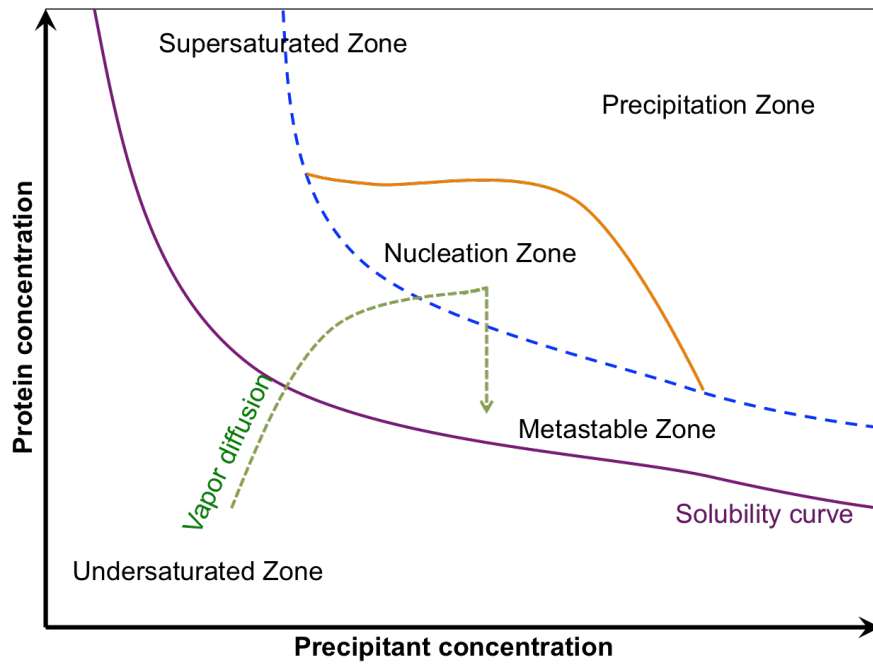


Figure 1.12 A simplified schematic representation of a protein crystallization diagram. The most commonly varied two parameters: protein and precipitant concentrations are used in the diagram. In order to reach crystal growth stage, the system starts from the undersaturated zone. Once reach the nucleation zone, then enters into the metastable zone followed by crystal growth. The route for vapor diffusion is marked by the green dash line.

1.5.4.1 Crystallization of integral membrane proteins (IMPs)

Crystallization of integral membrane proteins (IMPs) is generally considered more challenging than for soluble proteins. Integral membrane proteins comprise amphipathic surfaces, being embedded in the membrane lipid bilayer *in vivo*. Once the proteins are extracted from the membrane during purification, detergents will wrap around their hydrophobic surfaces, which can, potentially, maintain them in a functional, folded conformation (Moller and le Maire, 1993). As such, the actual molecules that are being crystallized are the protein-detergent complexes (PDCs). Selecting a detergent is therefore a critical step, which involves investigating two aspects of detergents, their effect on protein stability and on the formation of well-diffracting crystals. Normally, long chain detergents form larger micelles, and in principle could stabilize IMPs better. However, they cover more of the protein surface, and decrease the area available to form protein-protein contacts, which normally leads to poor quality crystals. Using short chain detergents therefore seems an obvious solution. However, short chain detergents are less suitable for the solubilization and stabilization of membrane proteins. Therefore, identifying a most suitable detergent involves a careful compromise of the two extremes.

Apart from the various parameters discussed for soluble proteins and detergent selection, IMPs crystallization also involves some additional considerations, such as the use of additives (Smith et al., 1998) and antibody fragments (Ostermeier et al., 1995). Fortunately, there are a large number of crystallization conditions that have been established from empirical screening rounds. Careful investigations on more parameters will give better success rate of obtaining the well-diffracting quality crystals.

1.5.4.1.1 Aqueous-based crystallization-Vapor Diffusion

The various crystallization techniques, namely vapor diffusion, micro-dialysis and batch crystallization that are being employed on soluble proteins can also be applied to membrane proteins. Vapor diffusion is the most frequently used method, and the typical experimental setup involves either a hanging or sitting drop. In both experiments, a small volume of protein solution is mixed with a precipitant solution and left to equilibrate against a reservoir solution, containing similar buffers and precipitating agents (mainly salts and PEGs) to the precipitant solution. Due to the higher concentration of precipitants in the reservoir solution, water evaporates from the protein droplet, and the protein and precipitants concentrations increase, hopefully to a point within the nucleation zone that is required for producing crystals (Rhodes, 1993). Crystals normally start to appear after a few days to weeks. The speed of crystallization mainly depends on the protein characteristics and the precipitants used. The method is normally carried out in several rounds with variations in precipitants, drop sizes and incubation temperature, etc., until the well-diffracting crystals are obtained.

1.5.4.1.2 Lipid-based crystallization-Lipidic Cubic Phase (LCP)

In the traditional aqueous-based membrane protein crystallization methods, the protein-detergent-complexes are exposed to the precipitant solution, posing a risk of denaturation and aggregation. These methods produce type II crystal packing, in which the crystal contacts are only between the hydrophilic surfaces of the molecules. With large areas of the protein surface covered by detergents, crystals have a high solvent content and consequently poor crystal packing (Ostermeier and Michel, 1997).

These problems could be overcome by lipid-based crystallization methods, including the bicelle (Faham and Bowie, 2002), vesicle (Takeda et al., 1998) and lipidic mesophase (cubic phase and sponge phase) (Landau and Rosenbusch, 1996) methods. Lipidic cubic phase (LCP), has been extremely successful for, but not limited to GPCRs (Landau and Rosenbusch, 1996). Crystallization of membrane proteins in lipidic cubic phase (LCP) is a method using LCP as a host medium, in which the protein-detergent-complex (PDC) is transferred into the lipid matrix, stabilizing the proteins and leading to type I crystal packing, in which the protein molecules make contacts through both hydrophobic and hydrophilic regions, resulting in a much better crystal packing (Landau and Rosenbusch, 1996). In the lipidic cubic phase method, lipids form a three-dimensional curved bilayer that is interpenetrated by water channels giving access to hydrophilic components as shown in Figure 1.13. As a result, this dynamic bicontinuous system can accommodate not only membrane

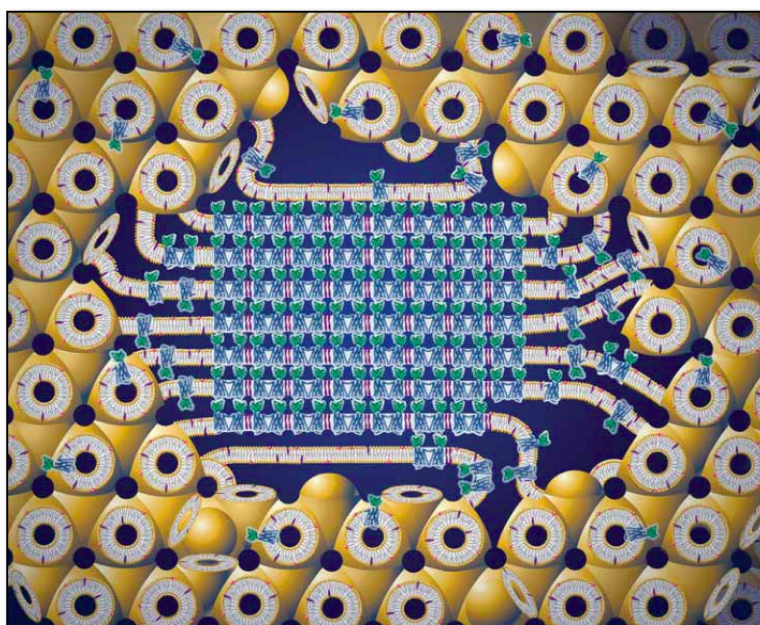


Figure 1.13 Cartoon representation of the lipidic cubic phase and the crystallization process of membrane protein in the system.

The protein molecules that reconstituted into the lipid bilayers in lipidic cubic phase are diffusing into the type I crystal packing, which is in the lamellar phase. Lipid (tan), detergent (pink), protein (blue and green), cholesterol (purple), bilayer and aqueous channels (dark blue). The lipid bilayer is around 40 Å thick. (Figure was taken from Caffrey *et al.* 2009)

proteins in their native and active conformation, but also detergents, salts, additives, etc. The huge area of the interface between curved lipid bilayer and water channels is perfect for nucleation, followed by crystal growth through the lateral diffusion of protein molecules.

The most critical step of crystallization in LCP is the formation of this bicontinuous cubic phase, which is guided by the phase diagram. Monoolein has been the most successful lipid, its phase diagram (Fig.1.14) having been elucidated many times (Briggs and Caffrey, 1994, Qiu and Caffrey, 2000, Caffrey and Cherezov, 2009).

A hypothesis for the mechanism was proposed by Caffrey: protein solution is mixed with lipid at an appropriate ratio under favored temperature (what is this), driving the system into a cubic phase (space group type $Pn3m$) (Fig. 1.14), which is in a viscous transparent form. During this process, the protein molecules are embedded into the lipid bilayer. Once the LCP bolus is overlaid by the precipitant solution (in excess volume), the components start to diffuse into the LCP matrix (molecular sponge) through the aqueous channels, triggering the phase transition towards the formation of the lamellar phase and crystal growth (Caffrey, 2003).

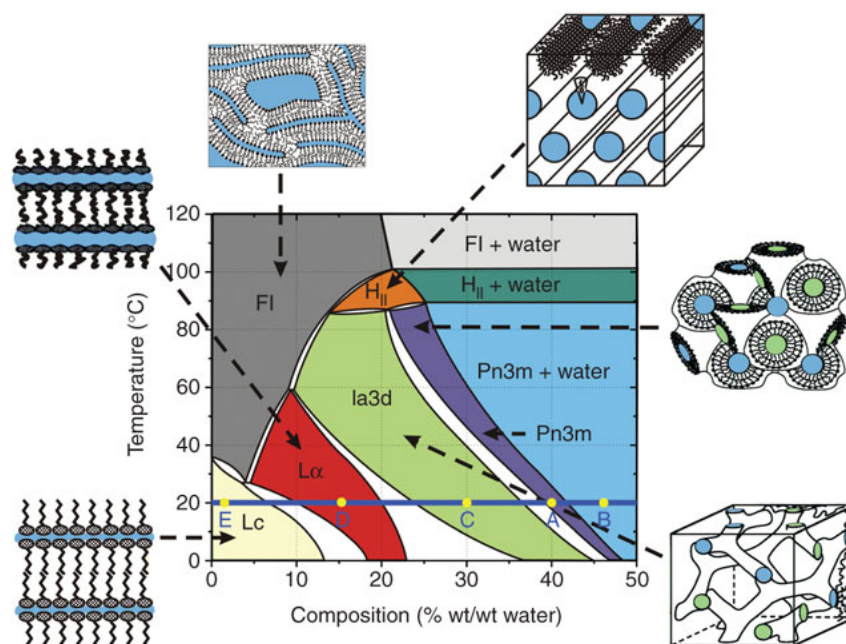


Figure 1.14 Molooloin/water phase diagram. Various phase states according to the water composition at a range of temperatures. 20 °C is the most commonly used temperature, which is highlighted as a blue line. Lc, FI, HII, Lα, Ia3d and Pn3m represent lamellar crystal phase, fluid isotropic phase, inverted hexagonal phase, lamellar liquid crystal phase, cubic-Ia3d phase and cubic-Pn3m phase respectively. Blue and green colored areas represent water. Figure taken from (Caffrey and Cherezov, 2009)

2 Chapter Two

**Structural studies of the human
Histamine receptor H1 in complex
with second-generation
antihistamines**

2.1 Materials and Methods

The protein expression, purification and crystallization steps were carried out closely following the published protocols (Shimamura et al., 2011), with modifications.

2.1.1 Expression

2.1.1.1 Expression construct

The construct used for the H1 study was engineered by (Shiroishi et al., 2011) based on the T4L method. It has been subcloned into a pPIC9K vector for the expression in *Pichia pastoris* and structural determination.

In the H1 construct (Fig. 2.1), the N-terminus has been truncated from Met1 to Lys19 as this region was suggested to be disordered by the secondary structure prediction servers, and including two N-linked glycosylation sites (Asn5 and Asn18), which are with the potential to affect the protein crystallization. The badly-structured third cytoplasmic loop (ICL3) has been replaced by the T4L mentioned above and the construct is followed by tobacco etch virus (TEV) protease cleavage site, green fluorescent protein (GFP) and octa-histidine (His8) tag. GFP tag is used to assist monitoring the target protein during the expression and purification steps, while the His8 tag is used for the immobilized metal ion affinity chromatography (IMAC) purification. The C-terminal region keeps the same length since it was supposed to be a short loop (Shimamura et al., 2011).

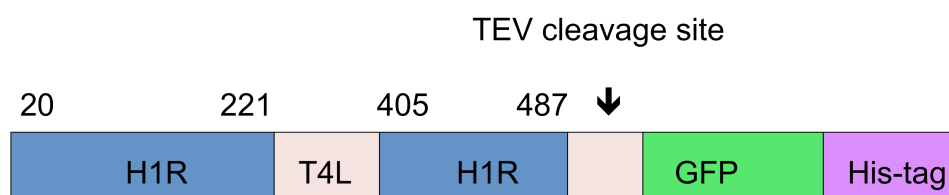


Figure 2.1. H1 construct.

2.1.1.2 Generation and screening of expression colonies

Pichia Pastoris SMD1163 strain was kindly given by Dr. Shweta Singh, MPC, Imperial College, London. *Pichia Pastoris* SMD1163 competent cells was made in-house, with protocol by (Cregg *et al.*, 1993). SMD1163 competent cells were mixed with PmeI linearized H1-T4L-GFP DNA and the plasmids transformed into SMD1163 *Pichia Pastoris* strain, using an Eppendorf electroporator (2000V). The cell suspension was plated on MD (Minimal Dextrose) agar plates (1.34% yeast nitrogen base w/o amino acids, 2% glucose, 0.00004% biotin and 1.5% agar) and incubated for 2-3 days at 30°C.

The colonies that appeared on the MD-agar plates were harvested and serial dilutions were prepared. The equivalent of 10^5 cells/plate was spread on YPD (Yeast Extract Peptone Dextrose)-agar plate (1% yeast extract, 2% dextrose, 2% peptone, 2% agar) and supplemented with 0, 0.1 or 0.25 mg/ml geneticin (G418) to select those colonies containing the target gene. Incubation was carried out for 3-5 days at 30°C.

Colonies that appeared on geneticin containing YPD-agar plate were brought to the small-scale expression trials.

2.1.1.3 Small-scale expression test

Single *Pichia Pastoris* colonies containing the H1 plasmid on YPD plates containing

0.1 mg/ml geneticin were picked. Each colony was inoculated with 5 ml of Buffered Glycerol Complex Media (BMGY) (100 mM potassium phosphate pH 6.0, 1.34% yeast nitrogen base w/o amino acids, 2% peptone, 0.00004% biotin, 1% yeast extract, 1% glycerol). The culture was grown with a shaking speed of 250 r.p.m at 30 °C until an OD₆₀₀ of 2-6 was reached. The cells were harvested by centrifugation at 3,000 g for 10 min and the cell pellet was re-suspended in Buffered Methanol Complex Media (BMMY) (100 mM potassium phosphate pH 6.0, 1.3% yeast nitrogen base without amino acids, 2% peptone, 0.00004% biotin, 1% yeast extract, 0.5% methanol) containing 2.5% dimethyl sulfoxide to achieve a starting OD₆₀₀ of 1 for all cultures. The culture was left to grow for 20 hours at 30 °C with a final OD₆₀₀ of 13, and then the cells harvested by centrifugation at 5000 g for 5 mins.

Cells were harvested by centrifugation at 3,000 g for 5 min after 20 hours and cell pellets re-suspended in 1xPBS buffer. GFP Relative Fluorescent Unit (RFU) of the cells were measured in an optical-bottom-black 96-well plate (Nunc, Rochester, NY) using a SpectraMax M2e (Molecular Devices). The machine settings were: emission 512 nm, excitation wavelength 488 nm, bottom read, and the expression level was calculated as shown in box 1 (Newstead *et al.*, 2007). The clone with the best expression level was picked and brought to large-scale expression.

2.1.1.4 Large-scale expression of H1T-T4L

The pre-culture was inoculated into BMGY in 2.5 Liters baffled flasks with 500 ml media in each flask. The culture was grown at 30°C and shaken at 250 rpm until an OD₆₀₀ of 2-6 (log-phase growth, around 16-18 hours) was reached. The cells were harvested by centrifugation and the cell pellet was re-suspended in Buffered Methanol Complex Media (BMMY) with the addition of 2.5% (v/v) dimethyl sulfoxide

(DMSO) to achieve a starting OD₆₀₀ of 1. After 22 hours, OD₆₀₀ of the culture reached ~13. Cells were harvested by centrifugation, followed by flash freezing and stored at -80 °C.

2.1.2 Purification of H1-T4L with antagonist

2.1.2.1 Membrane preparation and membrane wash

The yeast cell pellet was re-suspended in a breaking buffer containing 50mM HEPES pH7.5, 120mM NaCl, 5% (v/v) glycerol, 2mM EDTA and EDTA-free protease inhibitor cocktail tablets, by stirring at 4 °C until homogenized. Typically, 150 grams of cells were re-suspended in 360ml of ice-cold breaking buffer. The cells were broken using a cell disruptor (Constant System) by passing through the system four times at increasing pressures of 25, 30, 35 and 39Kpsi. Undisrupted cells and cell debris were separated and removed by centrifugation at 3,000 g for 20 minutes at 4°C, and then ultracentrifugation was used to collect yeast membranes at 100,000 g for 30 minutes at 4°C. The membranes were then washed by the method of dounce homogenation, ultracentrifugation and re-suspension. Those steps were repeated 6 times. The initial two washes was performed with buffer containing 10mM HEPES pH7.5, 5% (v/v) glycerol, 2mM EDTA and EDTA-free protease inhibitor cocktail tablets, and the following four washes were performed with the washing buffer

Box 1: Numerical estimate of whole-cell overexpression level

Sample RFU / 11,300RFU X (0.03 mg ml⁻¹) = GFP Concentration

11,300 RFU: Relative Fluorescent Unit (RFU) of pure yeast-enhanced green fluorescence yEGFP at concentration of 0.03 mg ml⁻¹

0.03 mg ml⁻¹: Concentration of pure GFP

supplemented with 1M NaCl, 10 mM MgCl₂, and 20mM KCl. The concentration of whole membrane proteins was estimated using the bicinchoninic acid method (Pierce). Washed membranes were flash frozen in liquid nitrogen and stored at -80 °C.

2.1.2.2 Solubilization

The membrane suspension was solubilized by incubation in the buffer containing 20mM HEPES pH7.5, 500mM NaCl, 2% (v/v) glycerol, EDTA-free protease inhibitor cocktail tablets, 1% of n-dodecyl-β-d-maltopyranoside (DDM, Anatrace), 0.2% of cholesteryl hemisuccinate (CHS, Sigma), and the addition of 5 mM ligand and 2mg/ml iodoacetamide, for 1-2 hours.

2.1.2.3 First IMAC–Talon metal affinity Chromatography

After the solubilization step, unsolubilized materials were removed using ultracentrifugation. The supernatant and pre-equilibrated TALON resin were mixed and stirred gently at 4 °C overnight. Talon resin was washed intensively in the wash buffer (20mM HEPES pH7.5, 500mM NaCl, 10% glycerol, 0.025% of n-dodecyl-β-d-maltopyranoside, 0.05% of cholesteryl hemisuccinate) in the presence of 0.1 mM Ligand to get rid of the soluble protein, and the receptor was eluted using 200 mM imidazole in several fractions. All the fractions were checked by SDS–PAGE gel, and those H1-T4L-GFP containing fractions were pooled and concentrated to 2.5 ml. Then the PD-10 column was used to remove the imidazole. From this point forward, all steps have the presence of 0.1 mM ligand.

2.1.2.4 Second IMAC–Ni-sephasrose affinity chromatography

The protein was mixed to Ni-sepharose high performance resin and incubated for 1 hour, then the flow through was collected, followed by 20 column volumes of wash

buffer (20mM HEPES pH7.5, 500mM NaCl, 10% glycerol, 0.025% of n-dodecyl- β -d-maltopyranoside, 0.05% of cholesteryl hemisuccinate) with the presence of 0.5 mM Ligand (cetirizine and fexofenadine). Three column volumes of sample were eluted using 500 mM imidazole in different fractions. All fractions were assessed by SDS-PAGE gel and the H1-T4L-GFP containing fractions were collected.

2.1.2.5 Cleavage of GFP-His₈ region and reverse IMAC

The H1-T4L-GFP containing fractions from Ni-sepharose elution was mixed and incubated with His-tagged TEV protease for cleavage, and dialysed overnight in the dialysis buffer (20mM HEPES pH7.5, 500mM NaCl, 10% glycerol, 0.025% of n-dodecyl- β -d-maltopyranoside, 0.05% of cholesteryl hemisuccinate and 1 mM Ligand) at 4 °C to get rid of Imidazole. The cleavage product was processed by reverse IMAC. In which the protein samples were applied onto the Ni-Sepharose high performance resin. The impurities that bound to the resins in the initial two IMAC steps will bind to the resin here again, while the H1-T4L, which has lost the GFP-His₈ tag during cleavage, will be in the flow through, together with the first column volume washes were collected. The protein was concentrated to ~40 mg/ml with a 100 KDa molecular weight cut-off Vivaspin concentrator and subjected to crystallization trials.

2.1.3 Crystallization in Lipidic cubic phase (LCP)

The Lipidic cubic phase mixture was prepared by mixing 40% (w/w) H1 solution, 54% monoolein (MO) and 6% cholesterol. The crystallization trials were set up using the Mosquito-LCP robot (TTP Lab Tech, UK). The plates were set up using 50nl of protein-monoolein dispensed on the glass sandwich plate with 800nl of precipitant solution dispensed on top of the lipidic cubic phase drop. All steps needed to be performed at 20-22°C, and the plates were incubated at 20°C. The precipitant

condition was: 24–33% PEG400, 10 mM MgCl₂, 100 mM Na-citrate pH 4 - pH6.5, 300 mM ammonium phosphate and 1 mM ligand (Sigma).

2.1.4 Harvesting crystals from LCP

MicroMounts loops (MiTeGen) were used to harvest LCP crystals from glass sandwich plates. The rigid design is helpful to harvest from the dense LCP material, and thinner design on loop area makes it easier to fish crystals grown on the bottom and helpful to lift the crystal from underneath. The whole design helps avoiding catching too much LCP material with the loops.

2.1.5 Data collection and processing

The crystals of H1 receptor in complexes were tested at the microfocus beamline I24 at Diamond Light Source, using a beam size of 10 μm x 10 μm and a wavelength of 0.96863 Å. The detector was a Pilatus 3 6M.

Before data collection, the diffraction grid scan was applied in order to locate the crystal in the opaque LCP material on a MiTeGen loop and identify the best diffracting part on a tile-shaped crystal. For the grid scan, the crystal was exposed for 0.01 second, with low transmission (5%) to minimize radiation damage.

Two grid scans were carried out, 90° apart, in order to center the X-ray beam accurately on the best diffracting part of the crystal.

Diffraction data for H1-cetirizine and H1-fexofenadine have been collected during several experiments. Crystals were always exposed for 0.2 second, with 50-70% transmission and an oscillation angular step of 0.2°. All data from the different crystals were integrated using the computer program XDS (Kabsch, 2010). Data that

could be successfully integrated were subsequently combined to form complete datasets, using BLEND (Foadi et al., 2013).

2.1.6 Phenix-refinement

The program Phenix (Adams et al., 2010) was used for phasing and refinement. The published structure of the H1-doxepin complex (PDB:3RZE) (Shimamura et al., 2011), deprived of all ligands and non-protein atoms, was used as starting model. Rigid body refinement was followed by TLS (Translation-Libration-Screw rotation) refinement. For both strategies, two separate regions of the model were defined: T4L region as chain A (residue 28 to 485) and the rest as chain B (residue 1002 to 1161). Accordingly, both rigid body and TLS refinement were performed separately on the two regions. The refinement strategy was further modified on refinement targets and restraints. In particular, the reference model restraints and the secondary structure restraints were used because these useful for low-resolution data. The above refinement was monitored via R-value statistics (R_{work} and R_{free}), as these reflect the quality of the agreement between data and updated model. Different protocols of refinement were tried, in order to obtain the lowest possible R-factors with the minimal difference between R_{work} and R_{free} .

2.1.7 Protein structure figure preparation

Protein structure figures were generated with The PyMOL Molecular Graphics System (Version 1.7.4 Schrödinger, LLC.) and COOT (Emsley et al., 2010).

2.1.8 Radioactive ligand binding assay (RLB)

2.1.8.1 Site-directed mutagenesis and membrane production

The work for the site-directed mutagenesis and membrane production was carried out by Mr. James Birch from the Membrane Protein Laboratory of Imperial College at Diamond Light Source. Mutagenesis was performed using the QuikChange Lightning Multi Site-Directed Mutagenesis Kit (Agilent Technologies), following supplied instructions. The human H1 residues K179E^{CL2}, K191^{5.39} and H450^{7.35} were mutated to Alanines, and five mutants were produced: K179A, K191A, H450A, K179A&K191A, K179A&K191A&H450A. In this chapter, the GPCR Ballesteros–Weinstein numbering system (Ballesteros and Weinstein, 1995) are used for the key residues. Protein expression and membrane preparation were carried out following the published protocol (Shiroishi et al., 2012). The mutants were expressed in *S. cerevisiae* in small-scale, followed by membrane preparation. Final membranes were flash-frozen and stored at -80C° for radioactive ligand binding assays.

2.1.8.2 Saturation radioligand binding assays

Saturation radioligand binding assay is a technique to measure the specific binding of a radiolabeled ligand at various concentrations when the system reaches equilibrium. From it, ligand affinity and receptor density can be derived. The affinity of a ligand for a receptor is well described by the equilibrium dissociation constant, K_D (nM). This constant describes the interaction strength between a ligand and its receptor. It is, in fact, the concentration of a ligand required to occupy 50% of the receptors. B_{max} is the maximal number of receptor binding sites occupied by the ligand (pmol/mg) and it correlates to the amount of receptor present in the binding assay (Neubig et al., 2003).

The saturation radioligand assays were carried out using the H1 ligand [³H]-pyrilamine to determine the K_D and B_{max} of [³H]-pyrilamine of five different H1-T4L mutants plus the H1-T4L. In the binding reactions, yeast membranes bearing the H1

receptors containing 20 µg of protein were incubated with an increasing concentration of [³H]-pyrilamine (0.125, 0.25, 0.5, 1.25, 2.5, 5, 10, 20, 30, 40 nM) to define total binding. Non-specific binding for each concentration was determined under the identical conditions, in the presence of 40 nM unlabeled pyrilamine. Binding was performed in a buffer containing 20mM Hepes pH 7.5, 150mM NaCl of a final volume of 200 µl in a 96 well plate at 25°C for 2 hours to reach equilibrium. Following incubation, the free radioligands were separated from the bound radioligands by extracting the suspension onto Glass Fiber Filters (Packard, Groningen) under vacuum using a cell harvester (PerkinElmer). The filters were washed at least three times with Phosphate buffered saline (PBS) buffer (137mM NaCl, 2.7mM KCl, 10mM Na₂HPO₄, 2mM KH₂PO₄, pH 7.4) to remove remaining unbound radioactive ligands. They were then transferred into an oven incubated at 40°C until they were completely dry. Dry filters were finally aligned to and mounted into the OmniFilter 96-well Microplate, 20 µl of MicroScintTM-O scintillation cocktail added to each well. The amount of bound tritium was counted using a Topcount Microplate scintillation counter (Packard/Perkin Elmer). All measurements were taken three times, simultaneously. The experiments were then repeated using a different buffer containing 20mM Na-phosphate pH5.8, 150mM NaCl to obtain the the K_D and B_{max} at a pH5.8.

2.1.8.3 Data analysis: saturation binding assay

What is measured from the scintillation counter is the amount of bound tritium for each total and non-specific binding reaction, expressed as counts per minute (CPM). From this the values of disintegration per minute (DPM) can be calculated using the equation: $DPM = CPM / \text{Counting Efficiency}$. The specific binding of radioligand in DPM was calculated using the equation: $\text{Specific bound} = \text{total bound} - \text{NSB}$. The free

(unbound) radioligand concentration for each reaction was calculated using the equation: Free=total added-total bound. DPM values were then converted to concentration units (i.e. pmol) by using the specific activity of radioligand [³H]-pyrilamine (DPM/fmol). An Y vs X data table was created where the average specific bound radioligand concentration (pmol/mg) was plotted on the Y-axis against the free radioligand (nM) on the X-axis. Data were analysed using the Graphpad Prism version 5.00 for Mac OS X, GraphPad Software, San Diego California USA, www.graphpad.com. K_D and B_{max} were obtained by fitting the equation for the one-site binding model using non-linear regression.

2.1.8.4 Competition radioligand binding assays

The competition binding assay is a method to measure the equilibrium binding of a fixed concentration radioligand in the presence of an unlabeled ligand (competitor) with various concentrations. The resultant data produce the value of IC_{50} , which is used to calculate the equilibrium dissociation constant (K_i) of the competing ligand to the target receptor using knowledge of K_D from the saturation binding assay. IC_{50} is the concentration of the competing ligand (unlabeled) required to displace 50% of the specific binding of the radioligand. It is essential to calculate the ligand affinity (K_i). K_i is the affinity, or equilibrium dissociation constant for the unlabeled ligand to the receptor that is calculated using the following Cheng-Prusoff equation (Cheng and Prusoff, 1973).

$$K_i = \frac{IC_{50}}{1 + L / K_D}$$

Where L is the concentration of radiolabelled ligand, and K_D is the radioligand dissociation constant. The GraphPad software used for this analysis could produce the value of K_i using the above equation.

The binding affinity (K_i) of H1 ligands cetirizine and fexofenadine to H1 mutants and H1-T4L were determined by competition radioligand binding assays against [3 H]-pyrilamine. In each reaction, yeast membranes bearing the H1 receptors containing 20 μ g of protein are incubated with an increasing concentration of unlabeled ligand cetirizine or fexofenadine (10^{-4} , $10^{-4.5}$, 10^{-5} , $10^{-5.5}$, 10^{-6} , $10^{-6.5}$, 10^{-7} , 10^{-8} , 10^{-9} , 10^{-10} M), and the radiolabeled ligand [3 H]-pyrilamine with a fixed concentration of 30 nM. Total binding was measured in the absence of any competitor. Non-specific binding of the radiolabeled ligand was determined under identical conditions, in the presence of 40 nM unlabeled pyrilamine and absence of competing ligand. Binding was performed in the buffer containing 20mM Hepes pH7.5, 150mM NaCl of a final volume of 200 μ l at 25°C for 2 hours to reach equilibrium. Following incubation, the free radioligands were separated from the bound radioligands by sucking the suspension onto Glass Fiber Filters (Packard, Groningen) under vacuum using a cell harvester (PerkinElmer). The filters were washed at least three times with Phosphate buffered saline (PBS) buffer (137mM NaCl, 2.7mM KCl, 10mM Na₂HPO₄, 2mM KH₂PO₄, pH 7.4) to remove remaining unbound radioactive ligand. They were then transferred into an oven incubated at 40°C until they are complete dry. Dry filters were then aligned to and mounted into the OmniFilter 96-well Microplate and 20 μ l of MicroScintTM-O cocktail were added to each well. The amount of bound tritium was counted using Topcount Microplate scintillation counter (Packard/Perkin Elmer). All measurements were taken three times, simultaneously.

2.1.8.5 Data analysis: competition radioligand binding assay

Binding data were converted from counts per minute (CPM) to disintegration per minute (DPM) using the equation: $DPM = CPM / \text{Counting Efficiency}$. Specific binding (Bound-NSB) in the presence of competing ligand at different concentrations was converted to a percentage of the maximum specific binding using the equation: $\text{Specific bound (percentage)} = (\text{Bound-NSB}) / (\text{Total bound-NSB}) \times 100$. The specific bond (%) was plotted on the Y-axis against the logarithm of the concentration (M) of the unlabeled ligand (X-axis). The data were analyzed by using the Graphpad Prism non-linear regression, choosing the one site-Fit K_i from the Competition binding equations panel. The software directly produces the value of K_i .

Binding assay results, K_D and K_i , are presented as pK_D and pK_i , which correspond to the negative logarithm of K_D and K_i respectively (the logarithm is in base 10).

Manual inspection, correction and rebuilding of the model, often in combination with cycles of refinement, was performed with the aid of the graphics program *Coot* (Emsley et al., 2010), using sigma-weighted $|2F_o - F_c|$ and $|F_o - F_c|$ maps. Mostly, the rebuilding focused on loops correction, addition of ligands and mutation of rotamers, all of this taking into account the fit with the electron density, the geometry and the ligand environment.

2.2 Results

2.2.1 H1-T4L clone selection and Large-scale expression

Seven colonies containing the H1-T4L gene were obtained from the clone selection. The clones with a higher copy number of gene insertion were screened as mentioned

in method. They were brought to the small-scale expression trials and the best expression clone was selected based on GFP expression level. Large-scale expression was performed with the best H1-T4L expression clone. A total of 19 g wet cells per 1 L culture were obtained. The GFP expression level was ~0.1mg/ml, which is considered to be a sufficient yield for membrane proteins regarding to other published results (Newstead *et al.*, 2007). The concentration of membrane proteins at this stage was estimated by the Bicinchoninic acid (BCA) assay (Pierce). From 12 L culture, 3 mg of membrane proteins were obtained.

2.2.2 Protein purification of H1-T4L with different antagonists

Purification of H1-T4L with the presence of different antagonists was carried out by IMAC and reverse IMAC after cleavage of GFP- His₈. Gel analysis (Fig. 2.2) shows that the protein has similar levels of purity as those in the published report by (Shimamura *et al.*, 2011). Around 1 mg of pure protein was obtained, a sufficient amount for LCP crystallization trials.

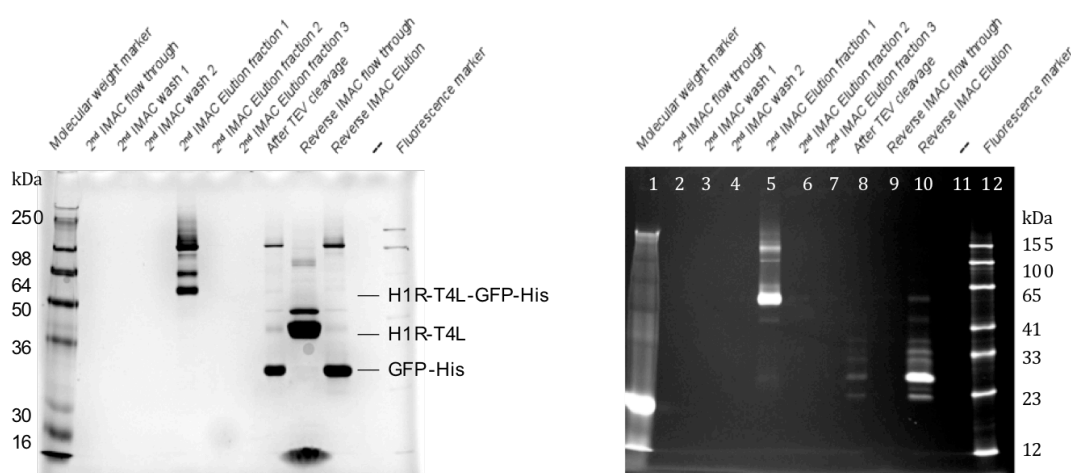


Figure 2.2. SDS-PAGE gel of H1-T4L from second metal affinity chromatography, TEV cleavage and reverse IMAC. (left) Visualized by Coomassie blue staining and (right) visualized by in-gel fluorescence.

2.2.3 Crystallization in Lipidic Cubic Phase

Lipidic Cubic phase crystallization trials of H1-T4L with the ligands cetirizine and fexofenadine were set up with freshly prepared protein. Crystals appeared after 4-6 weeks. Although these are crystals in complex with different ligands, they have very similar morphology. The largest obtained crystals are 50·50·10µm in size.

2.2.4 Data collection and processing

Crystals for both H1 complexes were grown from LCP. The common problem with the LCP crystals is that they cannot be visualized in the beamline visual setup. The grid scan function available at the Diamond Light Source (DLS) I24 beamline, has been extremely helpful to find H1 crystals in the completely opaque loop (Fig. 2.3). In addition, since different parts of the very thin tile-shaped H1 crystals diffract quite differently, the grid scan has been pivotal for maximizing the quality of the diffraction images obtained.

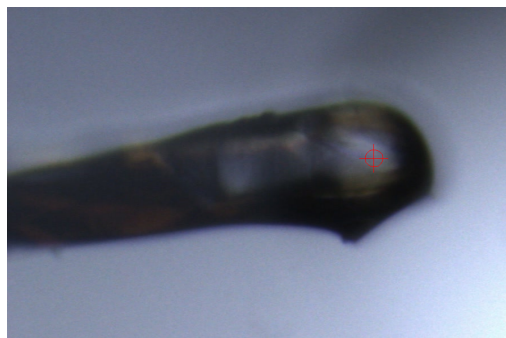


Figure 2.3 H1 crystal and LCP material in a cry-loop. The red cross indicates the beam center, where the crystal is located, found by grid scan.

It is very common to obtain fragile microcrystals from LCP trials. Useful X-ray data from H1 tiny crystals (50·50·10µm in size) were successfully obtained using the microfocus I24 beamline. Although other non-microfocus beamlines had also been tried initially, the diffraction images collected had a reduced signal-to-noise ratio compared to the microfocus beamline (data not included). I24 provided the small

focal and intense beam matching crystal size and making it possible to get the most out of the tiny H1 crystals.

The fragile crystals deteriorated during the data collection due to the radiation damage, to such an extent that no complete data set could be obtained. Thus, even though large 90-degree rotation ranges were collected for each crystal, only the initial part of each collection contained useful data at the highest available resolution. In all subsequent images the resolution fell sharply as due to radiation damage. In the case of H1-cetirizine, for instance, only one dataset (number 11, as indicated later in the text) displayed a 98% completeness at a resolution of 3.3 Å, but the corresponding electron density map did not provide convincing details for the main ligand. For the analysis work carried out in this thesis, it has been essential to assemble partial datasets into complete ones, with the aim of producing maps with details of sufficient resolution. This task was carried out with the help of the recent BLEND software (Foadi et al., 2013) in the CCP4 suite of programs for crystallography (Collaborative Computational Project, 1994). With the aid of BLEND several complete datasets were obtained for both H1-cetirizine and H1-fexofenadine, including those eventually used to build interpretable models of the proteins and their ligands, as explained in the next two sections.

2.2.4.1 Data processing for H1-cetirizine

For the case of H1-cetirizine 35 continuous sweeps of data were collected from 19 single crystals. As it turned out later, only 13 of these sweeps (from 7 single crystals) could be integrated with the XDS program. Only one of these datasets (serial number 11) was complete and could be used for generating initial map and model, but the electron density around the ligand was of insufficient quality (Table 2.1).

Table 2.1 Parameters of thirteen datasets of H1-cetirizine

Serial Number	R _{meas}	R _{pim}	Completeness (%)	Multiplicity	Resolution (Å) CC _{1/2} = 0.3
1	0.305	0.213	16.8	1.1	3.2
2	0.693	0.477	17.2	1.1	3.8
3	0.170	0.113	37.7	1.3	3.1
4	0.180	0.124	25.8	1.2	3.2
5	0.468	0.323	23.7	1.2	3.8
6	0.387	0.270	24.3	1.2	3.8
7	0.293	0.205	16.2	1.1	3.5
8	0.269	0.173	60.0	1.7	3.0
9	0.274	0.150	65.1	2.0	3.0
10	0.350	0.186	67.3	2.1	3.5
11	0.379	0.208	98.1	3.0	3.3
12	0.296	0.180	59.0	2.1	3.3
13	0.338	0.193	65.7	2.1	3.3

The only other way to obtain complete datasets, as mentioned earlier, was to combine the 13 individual datasets into complete ones with higher redundancy. This could improve the accuracy of the scaled intensities, which, in turn, could improve map resolution around crucial zones of the protein and ligand. This is obviously strongly related to structure isomorphism. It is evident that data from two crystals of the same structure reproduce its finer details only if the structures various parts are oriented and translated in a highly correlated fashion in the two crystals. This is why the strategy of assembling complete datasets out of partial ones can be successful only if the composing crystals show a good amount of isomorphism. The analysis carried out for the two H1 structures has made use of the computer program BLEND for multiple crystals data management, precisely with the purpose of assessing crystals

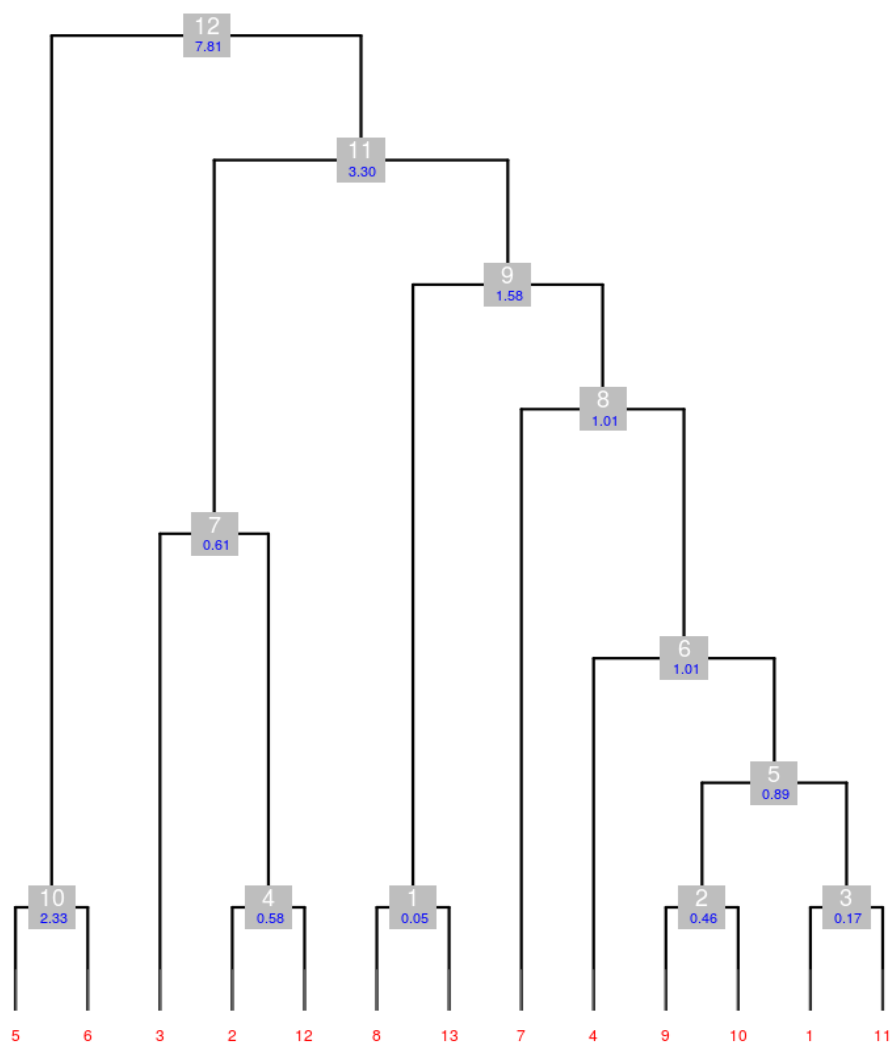


Figure 2.4 Dendrogram of cluster analysis on cell parameters of H1-cetirizine.

isomorphism. The very first assessment is done using cell parameters as similarity descriptors. In very simple terms, two crystals, or different parts of the same crystal, are similar if their cell parameters are close to each other. Isomorphism for the 13 datasets under analysis is depicted by the dendrogram (Fig. 2.4), and it is the result of a cluster analysis process carried out using cell parameters as statistical descriptors.

Twelve clusters were formed out of 13 datasets. The numbers in red are the individual datasets, while the white numbers in the grey boxes correspond to specific clusters. Numbers in blue are the absolute Linear Cell Variation (aLCV) and are measured in angstroms. They provide a measure of datasets similarity; the smaller the aLCV value, the closer the datasets. BLEND generates scaled data for every cluster displayed in the dendrogram. The working resolution for all scaling jobs was fixed at 3 Å in order to make comparisons simpler and to enable details at higher resolution to be visible as much as possible. The result is displayed in Figure 2.5.

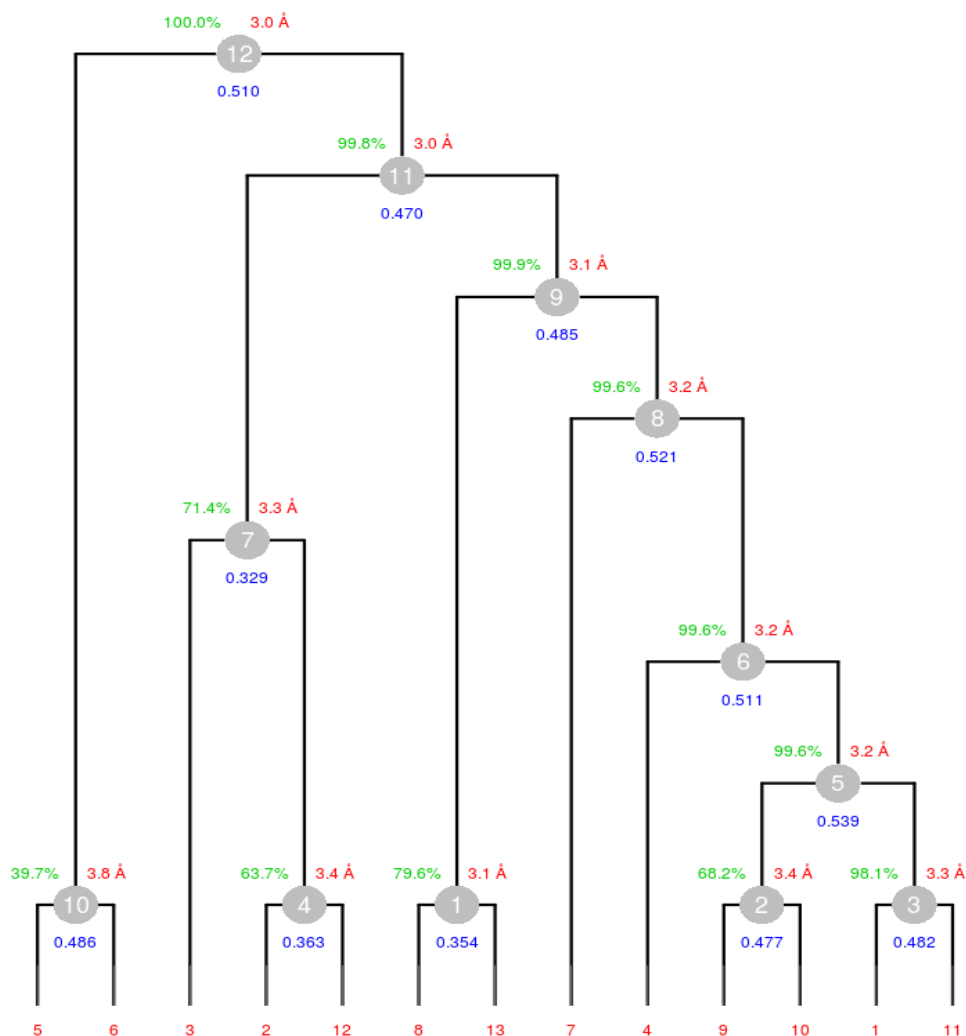


Figure 2.5 Merging statistics for clusters formed with H1-cetirizine datasets

The dendrogram in Figure 2.5 has the same structure of the dendrogram in Figure 2.4. Here, clusters are indicated by grey circles, rather than boxes, and three numbers are displayed with different colours around each circle. The number in green indicates completeness, the number in red indicates resolution as estimated by the $CC_{1/2}$ parameter, and the number in blue is the R_{meas} statistics, indicating how well intensities merge together. It is quite clear from the dendrogram that the resolution is extended by the inclusion of datasets. This is why top clusters 11 and 12 have a resolution of 3Å based on $CC_{1/2}$ being greater than 0.3. R_{meas} values, though, are not great. This should be correlated with an increasing non-isomorphism following data accumulation, as quantified in the dendrogram at Figure 2.4. These numbers can be improved by filtering out specific datasets in the BLEND process known as combination, where groups of datasets different from the clusters are tried to see if certain statistics improve. A close inspection of Figure 2.5 will convince the reader that all combinations should start by including the branch on the right, because branches on the left of the dendrogram are not complete and have lower isomorphism (higher values of the absolute Linear Cell Variation). Thus, starting from cluster 5, one or two individual datasets are removed in an attempt at improving resolution from 3.2 Å and/or improving the R_{pim} , because the last statistics takes into account both merging quality and data multiplicity. It turns out that removing datasets 1 and 10 from cluster 5 improves the R_{pim} , which goes from 0.206 to 0.181 (Table 2.1). Next, dataset 4 is added to the filtered cluster 5, so to produce a filtered cluster 6. The resolution is unchanged, while the R_{pim} improves, going from 0.181 to 0.171. Then dataset 7 is included in the filtered cluster 6, so to produce a filtered cluster 8. This time both resolution and R_{pim} essentially remain unchanged. We know that the inclusion of the remaining datasets to form clusters 9, 11 and 12 is beneficial to

resolution. Thus more datasets can be added to the filtered cluster 8, in an attempt at improving numbers. Eventually the best result comes from the filtered cluster 11, while the filtered cluster 12 provides slightly worst statistics.

The filtered cluster 11 was finally selected for the subsequent steps in crystallographic structure analysis. As the highest resolution suggested by the scaling programs AIMLESS, in terms of signal-to-noise being greater than 1.5, was 3.22 Å, it was decided to adopt a conservative approach to model building and refinement and provide input data at 3.2 Å resolution. Merging statistics for the final dataset are displayed in the first part of Table 2.2.

Table 2.2 Merging statistics for the final dataset

Starting Cluster	Datasets Filtered out	R _{meas}	R _{pim}	Completeness (%)	Multiplicity	Resolution (Å) (CC _{1/2} = 0.3)
5	1, 10	0.387	0.181	99.6	4.2	3.2
6	1, 10	0.379	0.171	99.5	4.5	3.2
8	1, 10	0.391	0.173	99.5	4.7	3.2
9	1, 10	0.409	0.148	99.8	7.1	3.1
11	1, 2, 10	0.400	0.129	100	8.8	3.0
12	1, 2, 5, 10	0.417	0.132	99.9	9.1	3.0

To conclude this section it is interesting and, in fact, illuminating to show the dramatic improvement in terms of resolution and map quality obtained with the inclusion of multiple datasets. In Figure 2.6 correlation plots of the type CC_{1/2} versus resolution are displayed for three different datasets. The black curve describes the behavior of the only complete single dataset 11; the red curve comes from cluster 11 and the green curve from the same cluster, filtered of some individual datasets. While, resolution wise, filtering does very little to improve matters, it is quite evident, and

indeed striking, the improvement when enriching dataset 11 with other isomorphous data. As mentioned earlier, this is ultimately caused by an increased multiplicity and by the elimination of bias brought in by the use of independent crystals.

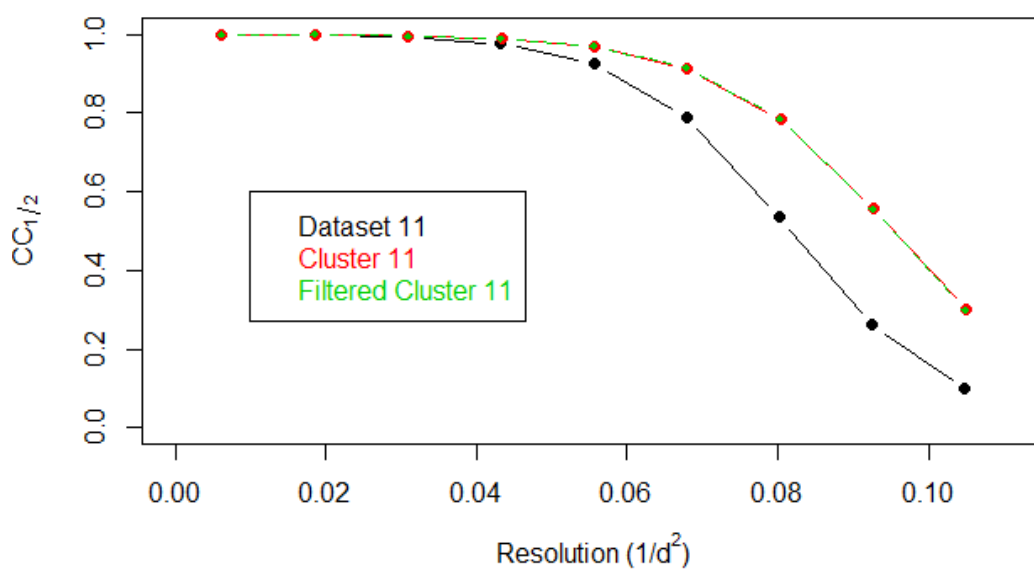


Figure 2.6 correlation plots of the type $CC_{1/2}$ versus resolution are displayed for three different datasets.

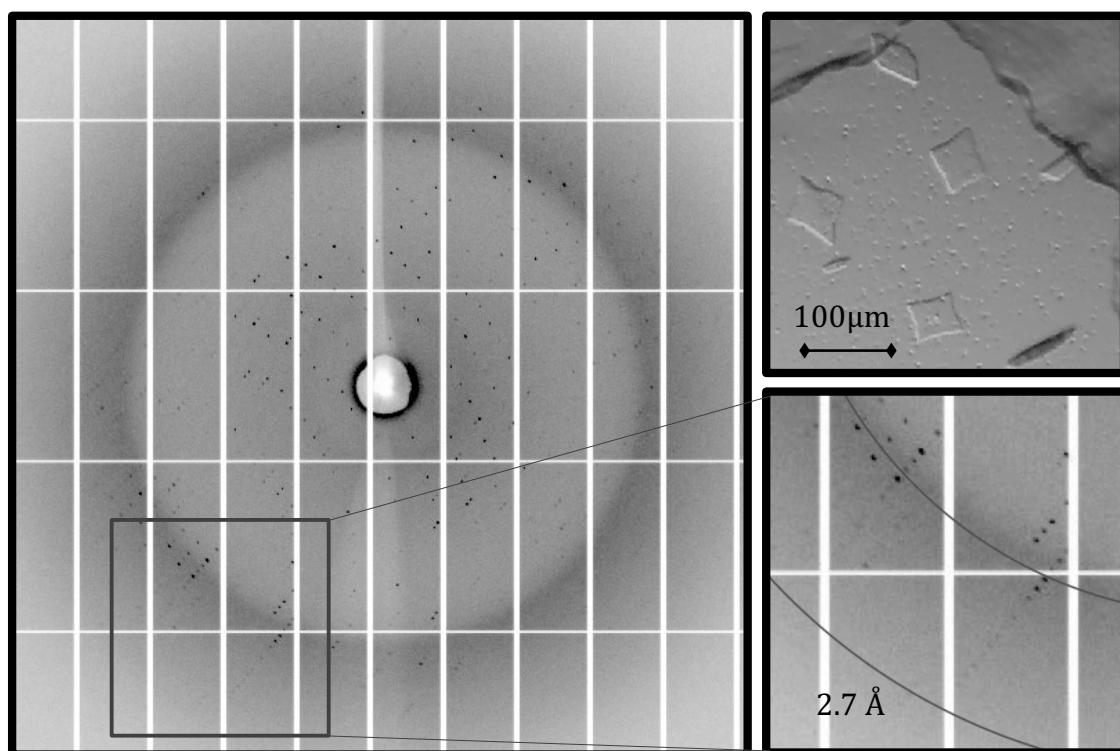


Figure 2.7 X-ray diffraction of H1-cetirizine complex (left). Crystal of H1 receptor in the complex with cetirizine grown in LCP (up right). Weak diffraction spots are visible at 2.7 Å (bottom right).

Table 2.3: Data collection, processing and refinement statistics for H1-cetirizine**Data collection and processing**

Beamline	DLS-I24
Space group	I422
Unit cell parameters	a=87.73, b=87.73, c=331.32
a, b, c (Å), α , β , γ (°)	$\alpha, \beta, \gamma = 90$
Wavelength (Å)	0.96863
Resolution range (Å)	49.68-3.2 (3.42-3.2)
Number of reflections	105443 (18344)
Number of unique reflections	11226 (1975)
R _{merge} (%)	31 (128)
R _{meas} (%)	32.8 (135)
R _{pim} (%)	10.6 (42)
Completeness (%)	100 (100)
Mean $I/\sigma(I)$	6.5 (1.8)
Multiplicity	9.4 (9.3)
Mn(I) half-set correlation CC(1/2)	0.99 (0.57)

Model Refinement

Resolution (Å)	3.2
R _{work} (%)	24.4
R _{free} (%)	28.6
Average B factor after TLS refinement (Å ²)	91.3

r.m.s.d. from ideal values

r.m.s.d. bond length (Å)	0.011
r.m.s.d. bond angles (°)	1.8

Values in parentheses refer to the highest resolution shells.

2.2.4.2 Data processing for the datasets of H1-fexofenadine

For the case of H1-fexofenadine, 28 datasets were collected from 5 crystals.

Completeness and other statistics are described in Table 2.4.

Table 2.4 Parameters of 28 datasets of H1-fexofenadine

Serial Number	R _{meas}	R _{pim}	Completeness (%)	Multiplicity	Resolution (Å) (CC _{1/2} = 0.3)
1	0.364	0.243	83.9	1.7	3.5
2	0.808	0.494	68.7	1.8	3.7
3	0.600	0.384	74.1	1.9	4.0
4	>1	> 1	79.6	1.7	6.2
5	0.622	0.341	50.5	2.6	3.9
6	> 1	> 1	63.2	2.1	4.4
7	> 1	> 1	60.2	2.1	5.5
8	0.332	0.228	32.9	1.5	4.1
9	0.255	0.164	42.7	1.9	3.3
10	> 1	> 1	51.0	2.2	5.0
11	> 1	> 1	31.2	3.3	6.0
12	0.428	0.246	49.9	2.7	4.0
13	0.226	0.150	42.0	1.4	3.2
14	0.372	0.252	36.0	1.4	3.7
15	0.637	0.353	54.7	2.2	3.8
16	0.903	0.545	56.5	2,2	4.5
17	0.607	0.363	63.0	2.0	3.8
18	> 1	> 1	67.6	2.0	4.2
19	0.513	0.275	39.8	3.0	3.7
20	0.328	0.192	54.4	2.3	3.4
21	0.572	0.295	89.8	3.1	3.5
22	0.595	0.404	84.8	1.6	3.8
23	0.236	0.142	53.2	1.9	3.1
24	0.447	0.294	78.9	1.7	4.0
25	0.772	0.427	72.6	2.7	4.2
26	> 1	> 1	83.9	1.6	5.0
27	0.222	0.134	75.1	2.2	3.2
28	> 1	> 1	72.1	2.2	4.2

As it was the case for H1-cetirizine, also in this case no single dataset has been deemed sufficiently good to calculate interpretable electron density maps. Therefore BLEND was used also in this case to try and improve both merging statistics and resolution. Cluster analysis performed on all 28 datasets resulted in the dendrogram shown at Figure 2.8.

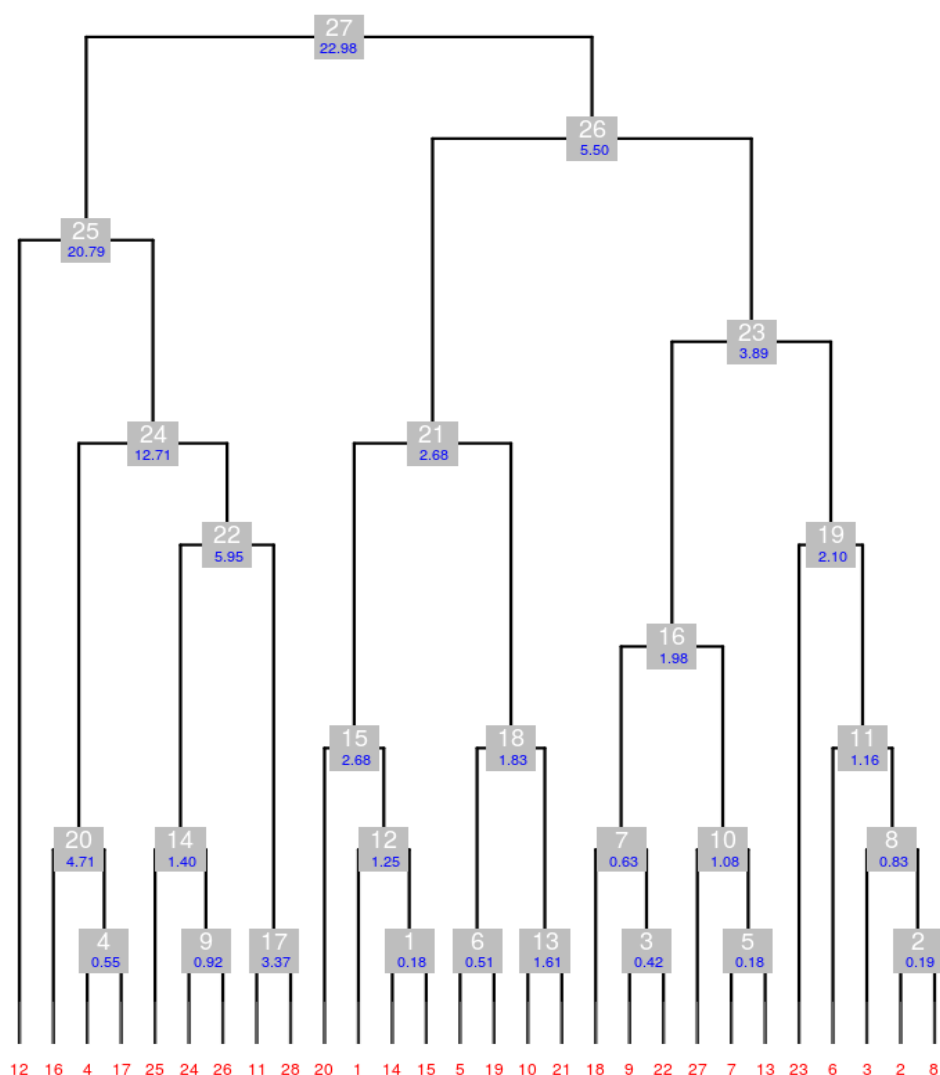


Figure 2.8 Dendrogram of cluster analysis on cell parameters of H1-fexofenadine

All 28 datasets, in general, manifest substantial non-isomorphism. In fact, the absolute Linear Cell Variation for all datasets amounts to 22.98 Å. Of the two branches

splitting from the top cluster the one on the right has lower aLCV and, therefore, is the one best placed to provide a complete set with good-quality data. To cut a long story short, after having assessed merging statistics for all clusters under cluster 26, and after having filtered out bad quality data, a final complete dataset was obtained as union of the individual datasets 2, 8, 9, 13, 22, 23 and 27. From Figure 2.8 it is immediate to notice that these numbers belong all to cluster 23. It has been in fact possible to obtain a clear electron density map for cluster 21 as well, but in this case the density for fexofenadine was less defined. Final statistics for the chosen complete dataset are displayed at Table 2.5.

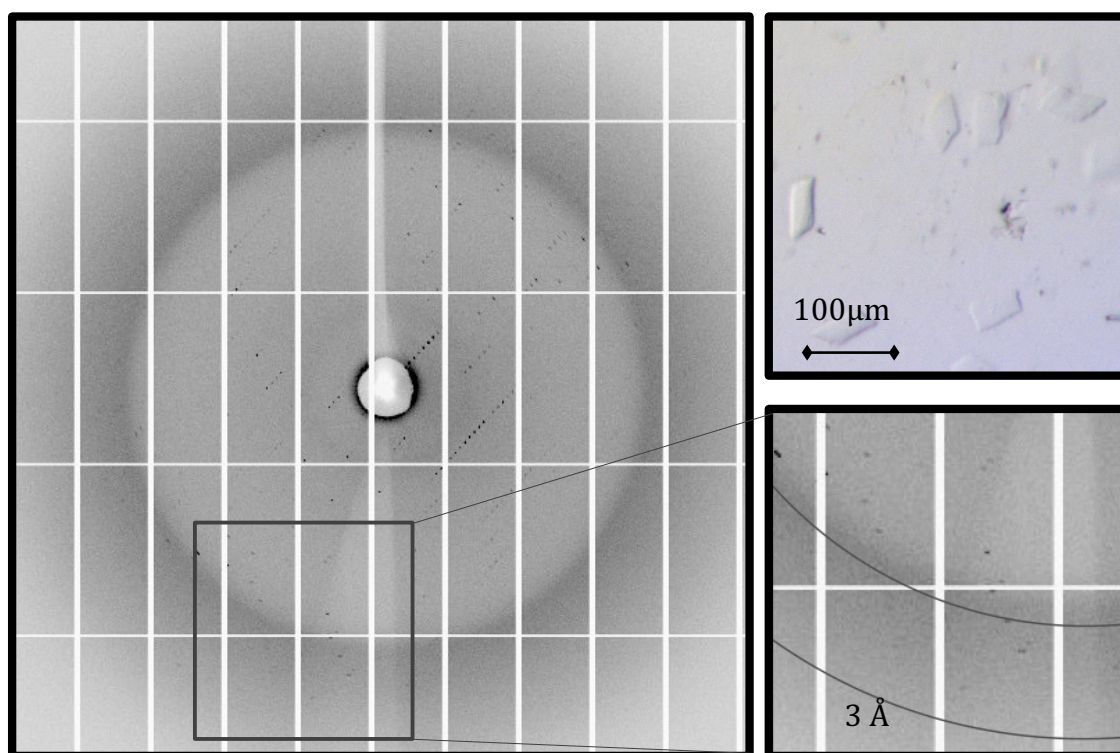


Figure 2.9 X-ray diffraction of H1-fexofenadine complex (left). Crystal of H1 receptor in the complex with fexofenadine grown in LCP (up right). Weak diffraction spots are visible at 3 Å (bottom right).

Table 2.5: Dataset collection, processing and refinement statistics for H1-fexofenadine

Data collection and processing	
Beamline	DLS-I24
Space group	I422
Unit cell parameters	a=87.4, b=87.4, c=331.22
a, b, c (Å), α , β , γ (°)	$\alpha, \beta, \gamma = 90$
Wavelength (Å)	0.96859
Resolution range (Å)	43.7-3.4 (3.67-3.4)
Number of reflections	32503 (4943)
Number of unique reflections	8477 (1460)
R _{merge} (%)	15.7 (52.6)
Completeness (%)	91.3 (78.6)
Mean $I/\sigma(I)$	6.9 (2.5)
Multiplicity	3.8 (3.4)
Mn(I) half-set correlation CC(1/2)	0.992 (0.564)
Model Refinement	
Resolution (Å)	3.2
R _{work} (%)	22.7
R _{free} (%)	28.6
Average B factor after TLS refinement (Å ²)	100.81
r.m.s.d. from ideal values	
r.m.s.d. bond length (Å)	0.012
r.m.s.d. bond angles (°)	1.7

Values in parentheses refer to the highest resolution shells.

2.2.5 Data quality

The structures of H1 in complexes with cetirizine and fexofenadine were both solved by molecular replacement using the structure of H1-doxepin complex as a model. The final resolutions of the complex structures are 3.2 Å and 3.4 Å respectively. The R-values from refinement are $R_{\text{work}}=24.4\%$, $R_{\text{free}}=28.6\%$ for H1-cetirizine and $R_{\text{work}}=22.7\%$, $R_{\text{free}}=28.6\%$ for H1-fexofenadine. They both have reasonable geometries. In the Ramachandran plot, the model of H1-cetirizine is with 95% in the preferred regions and 1.18% of outliers, while the model of H1-fexofenadine is with 95% in the preferred regions and 1.89% of outliers.

The overall protein structure and electron density quality are quite consistent among H1-cetirizine, H1-fexofenadine and the published H1-doxepin (Shimamura et al., 2011) complexes. The quality of the electron density maps of H1-cetirizine and H1-fexofenadine are illustrated in Figure 2.10.

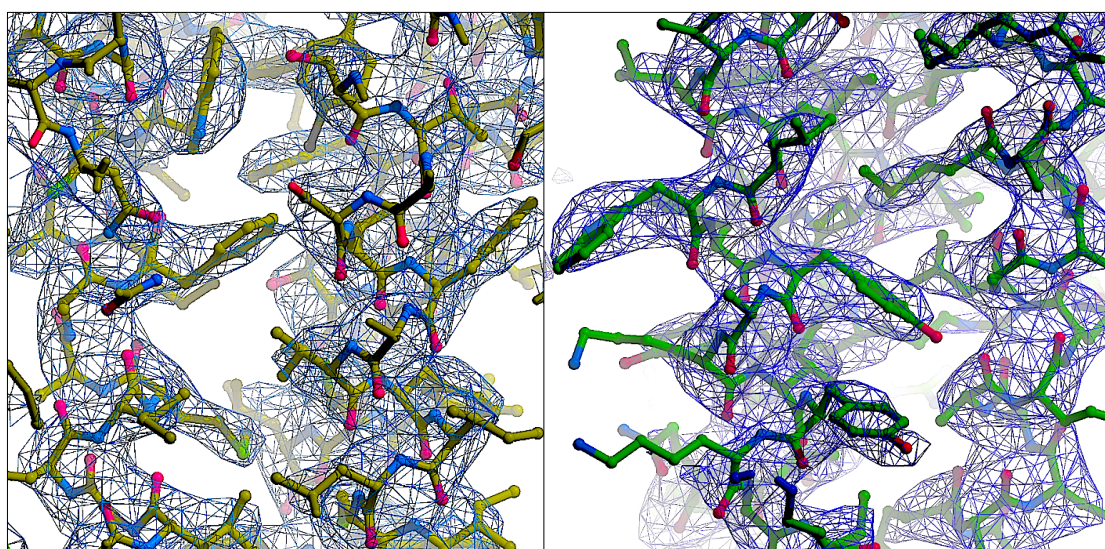


Figure 2.10 Partial electron density maps for H1-cetirizine complex (left) and H1-fexofenadine (right). $|2F_o-F_c|$ map contoured at 1σ .

Electron densities of most residues in the regions involved in the inter-helical contacts and ligand-binding site are well defined. The weaker densities were observed for residues on the flexible loop regions, as well as helix I and helix VIII, which are both not in tight contact with transmembrane core. These features of electron density quality are commonly observed for this range of resolutions in family A GPCRs (Rasmussen et al., 2007, Kruse et al., 2012, Granier et al., 2012).

For both H1-cetirizine and H1-fexofenadine complexes, most of the electron densities allowed the modeling of protein residues, apart from a section of extracellular loop 2 (residues F168-V174), which is not included since the electron density for it is not contiguous. This missing loop section also occurred in the H1-doxepin structure, which was solved at 3.1 Å (Shimamura et al., 2011). The omit maps for cetirizine and fexofenadine (Fig. 2.11) have allowed the initial fitting of the ligands into the right binding site.

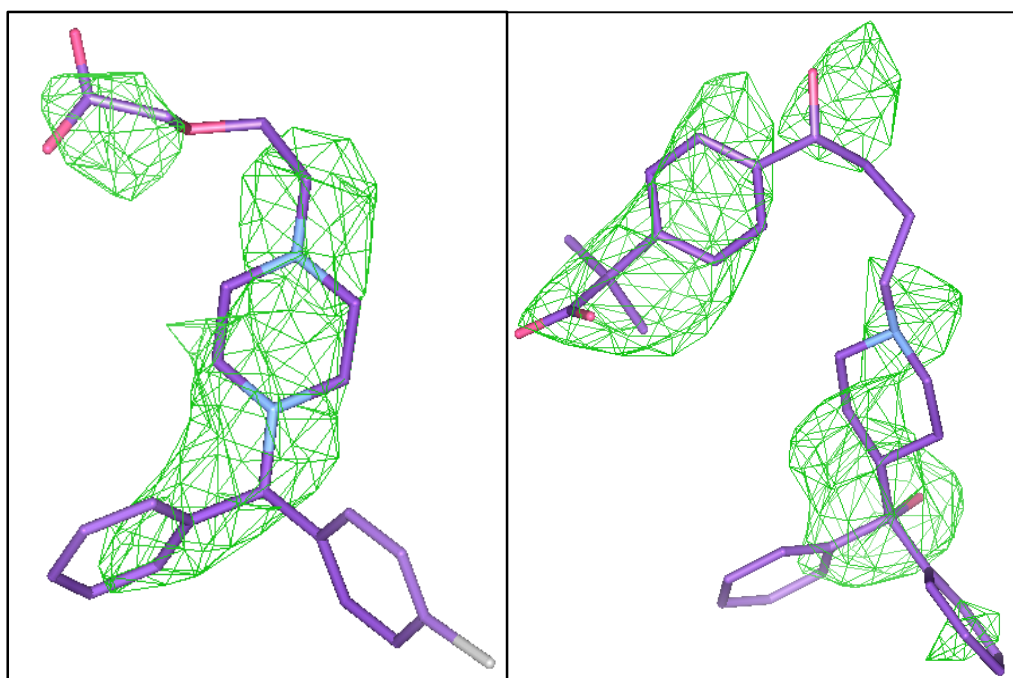


Figure 2.11 $|F_o - F_c|$ omit maps for cetirizine (left) and fexofenadine (right). $|F_o - F_c|$ omit maps are contoured at 3 σ , and shown as green mesh, cetirizine and fexofenadine are shown as purple sticks.

2.2.6 Crystal structures of H1 in complexes with cetirizine and fexofenadine

2.2.6.1 Subregions of H1 ligand binding site

The crystal structure of H1 in complexes with cetirizine and fexofenadine share the backbone structure with H1 in complex with doxepin. The main focus of the structures is on the ligand binding site. The binding pocket in H1 is a relatively deep cavity with the entrance on the extracellular side. There are some special features of the binding pocket, which are used to divide the pocket into three subregions: the lipophilic subsite, the D107^{3.32} ionic interaction region, and the anion binding site as shown in Figure 2.12. The region at the binding pocket bottom is highly hydrophobic, which comprises of a cluster of aromatic residues. This highly conserved region is

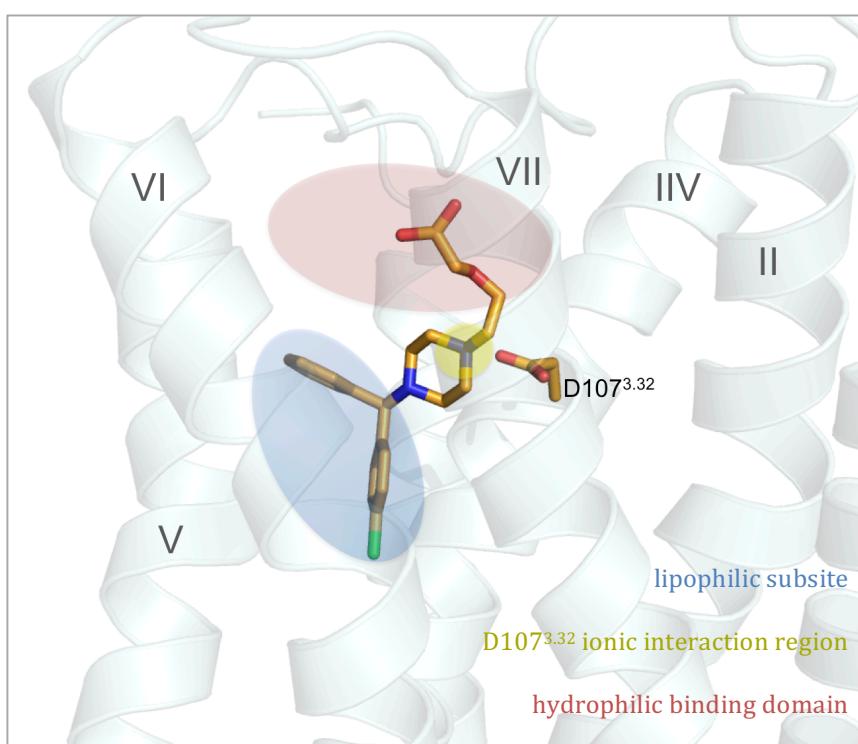


Figure 2.12 Subregions of H1 binding pocket. The three subregions of H1 binding pocket are illustrated in the H1-cetirizine structure. Lipophilic subsite, D107^{3.32} binding region, and hydrophilic binding domain are shown in blue, yellow and ruby shade respectively. H1 receptor is shown as cartoon representation, cetirizine is shown as orange sticks.

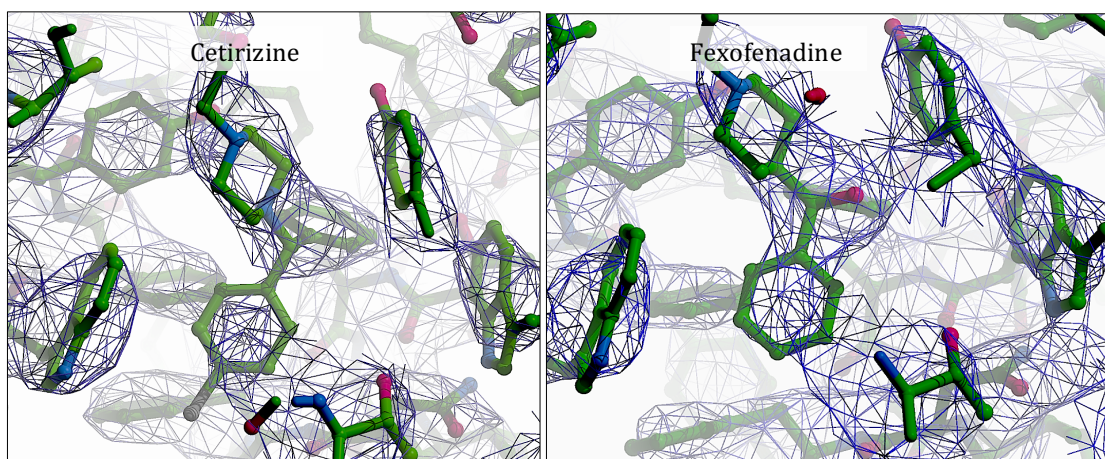


Figure 2.13 Electron density of the “lipophilic subsite” for H1-cetirizine complex (left) and H1-fexofenadine (right). $|2F_o - F_c|$ map contoured at 1σ .

called the “lipophilic subsite”. The “anion binding site” around the entrance region is with a great degree of variations across aminergic GPCRs. The acidic residue D107^{3.32}, which is highly conserved for aminergic GPCRs is the most important residue, always forms a salt bridge with the amine moiety of agonists, antagonists and the receptor’s natural ligands (Moguilevsky et al., 1998; Nonaka et al., 1998; Ohta et al., 1994).

The electron densities of the “lipophilic subsite” and the “D107^{3.32} ionic interaction region” for both complexes are quite well resolved (Fig. 2.13, Fig. 2.14), hence the residue side-chains of these two regions were modeled confidently.

However, the side-chains of the residues at the “anion binding site” on the

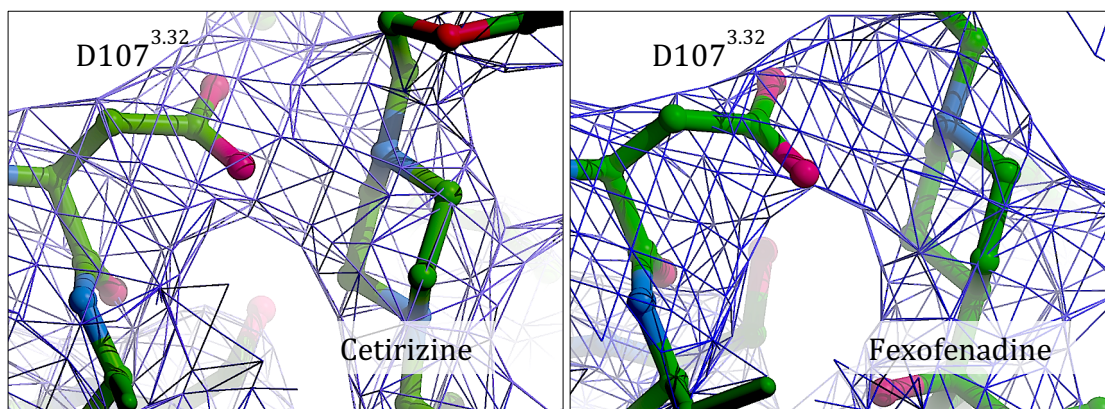


Figure 2.14 Electron density of the “D107^{3.32} ionic interaction region” for H1-cetirizine complex (left) and H1-fexofenadine (right). $|2F_o - F_c|$ map contoured at 1σ .

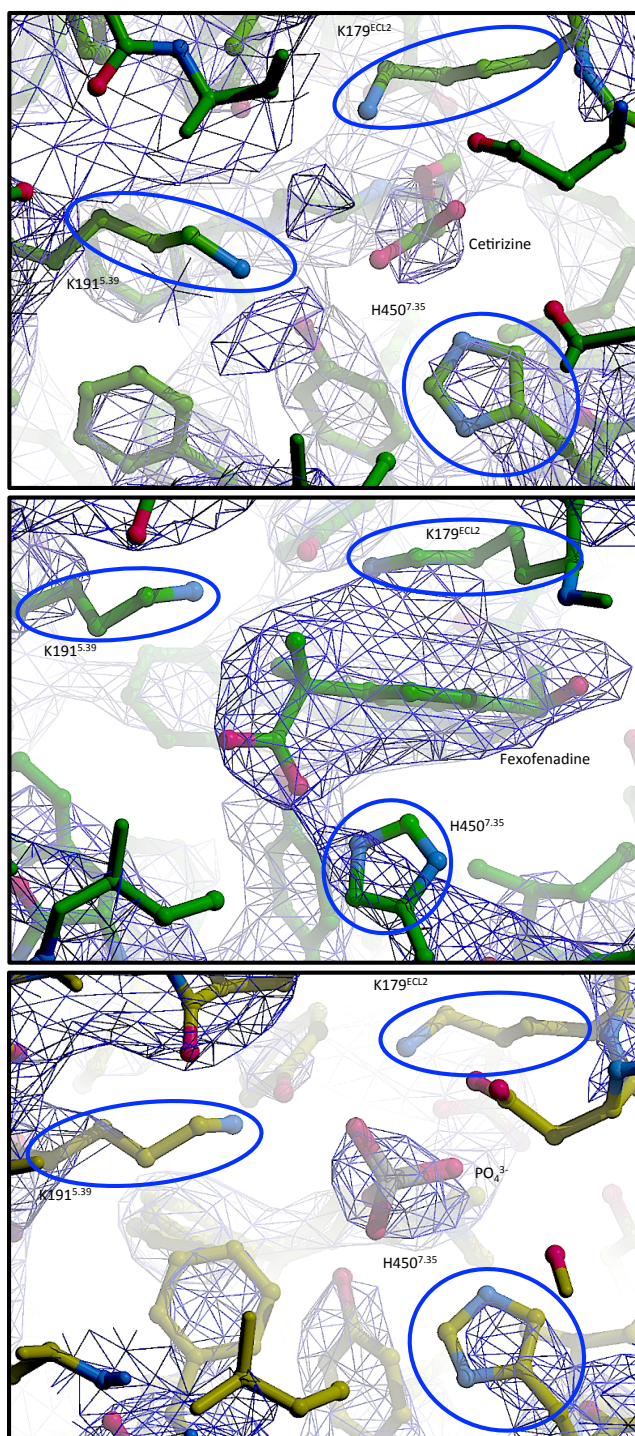


Figure 2.15 Electron density of “anion binding site” for H1-cetirizine complex (top), H1-fexofenadine (middle) and H1-doxepin (bottom). $|2F_o - F_c|$ map contoured at 1σ .

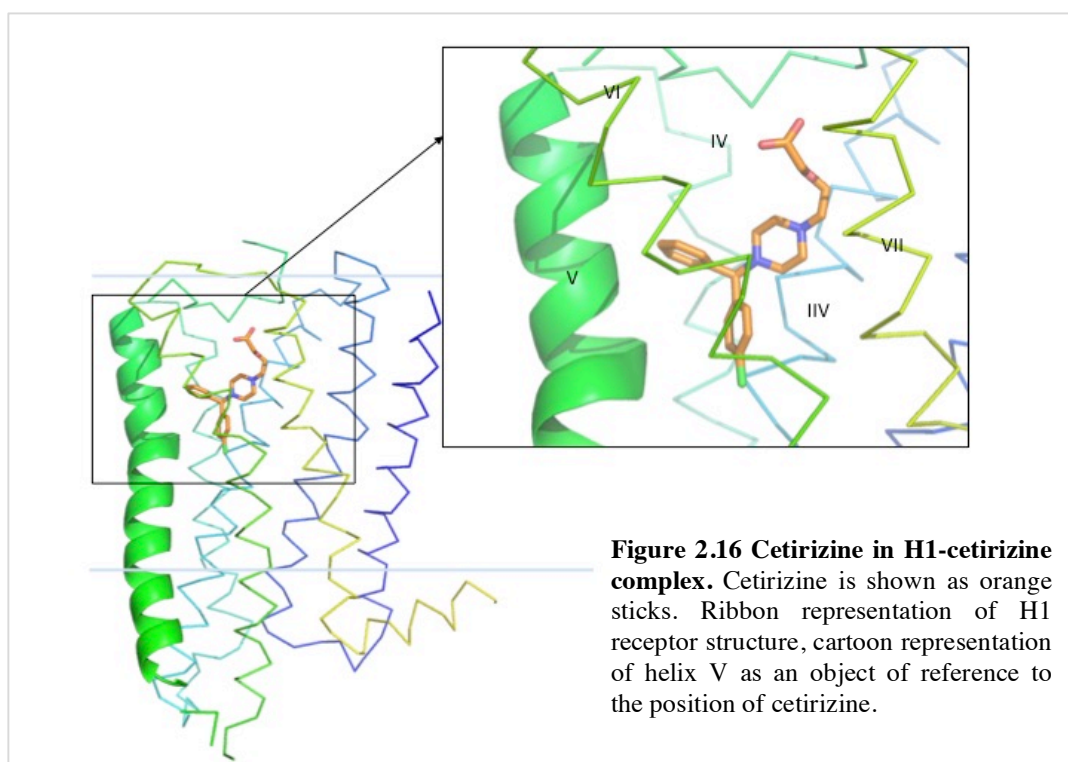
extracellular side have weak electron density for both complexes. Especially the long lysines, K179 on extracellular loop 2 and K191 on helix V, the density of the side-chains are totally missing (Fig.2.15). The “anion binding site” in crystal structure was first describe with the H1-doxepin structure, where it is coordinating with a phosphate ion. An inspection on the electron density map of the published H1-doxepin complex (Fig. 2.15 bottom) revealed that the lysines in this structure do not have any electron density either, although the binding site including residue K179, K191 and H450 was reported in the literature (Shimamura et al.,

2011). In the structures reported in this thesis, the two lysines were left as they are after refinement.

2.2.6.2 Interactions between cetirizine and H1 receptor

The second-generation antihistamine cetirizine contains a modified diphenylmethyl moiety, with a chloride attached to one of its benzene rings. The diphenyl ring system adopts a butterfly-shaped conformation in the crystal structure, having the benzene rings mimicking two spreading wings as shown in Figure 2.16. Cetirizine is a racemic mixture of R-cetirizine and S-cetirizine. However, the difference between the two configurations in the crystal structure is too small, beyond the obtained x-ray data. In theory, it would be impossible to decide which configuration to put in. An isomorphous difference map calculated from H1-cetirizine and H1-doxepin data revealed the positive electron density corresponding to the position of the chlorine, as indicated in Figure 2.17, hence the chlorobenzene ring was placed in this density, indicating only S-cetirizine has been observed in the crystal structure.

The chlorobenzene ring points almost perpendicularly down to the intracellular membrane plane, while the benzene ring projects towards helix V (Fig.2.16).



As shown in Figure 2.18, the diphenylmethyl moiety in cetirizine is responsible for the extensive hydrophobic interactions with the receptor. The benzene ring is in van der Waals contact with residues including Y108, W158, A195, Y431, F432 and F435 on helix III, V and VI. While the chlorobenzene ring only is within van der Waals contact distance of F432 and W428. The edge-to-face π -stacking interaction with F435 made the major difference on the protein, which will be described in the following section.

A piperazine moiety is directly attached to the diphenylmethyl group, via one of its nitrogens. The piperazine moiety anchors to the protein by forming a salt bridge with the most conserved residue D107^{3,32}. This important residue is the binding site for H1 natural agonist histamine. As shown in figure 2.18, it is obvious that the 2.5Å distance between D107^{3,32} and the nitrogen b (marked with blue circle in Figure 2.19) is much shorter than with nitrogen a (marked with blue circle in Figure 2.19), and forms a strong salt bridge. The two amine groups of the piperazine moiety have

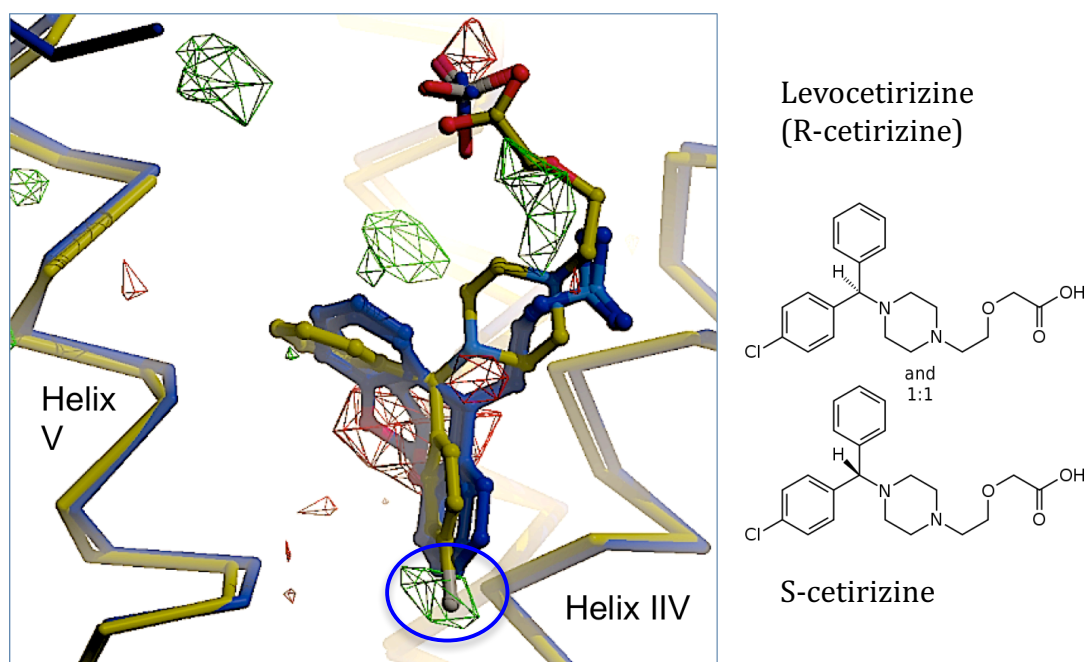


Figure 2.17 Isomorphous difference map. Calculated from H1-cetirizine data and H1-doxepin data. The electron density at the position of the chlorine is in blue circle. $|F_o - F_c|$ map contoured at 3σ .

different protonated states under various pH environments. The distribution of their protonation microspecies at different pHs were predicted using the protonation bundle of software ChemAxon (<http://www.chemaxon.com>) (Fig. 2.10). At pH 7.5, the microspecies distribution of cetirizine is 50% of 4, 15% of 5 and 35% of 1

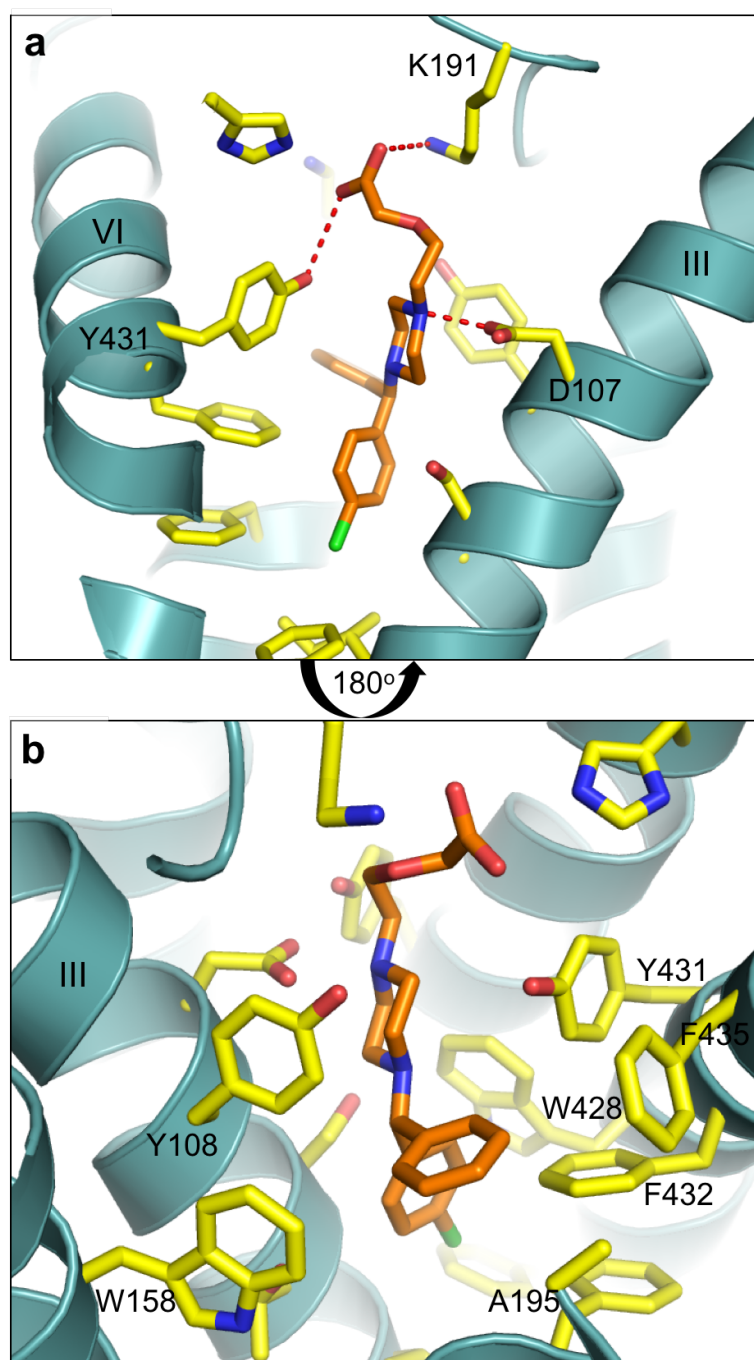


Figure 2.18 Cetirizine binding pocket. **a.** Direct view on polar contacts with the receptor; **b.** direct view on the "lipophilic subsite". Cetirizine is shown as orange sticks, H1 receptor is shown as cartoon presentation, residue side chains are shown as yellow sticks, polar contacts are shown as red dash lines.

microspecies; while at pH 5.8, the distribution is 75% of 4, 24% of 5 and 1% of 1 microspecies.

The protein was produced at pH 7.5, however the crystals were obtained at pH4.5-5.5. The exact pH along the binding pocket of the receptor is not known. It is assumed to be similar to the extracellular environment but it will more likely to be a gradient between the extracellular and intracellular pH.

The proportion of protonated amine group b (Fig. 2.19) is larger at acidic pH, where reaches the highest of 75% at pH5.8, and the overall proportion of protonated amine groups in cetirizine is larger at acidic pH, and at pH5.8 reaches a peak. This could lead to the conclusion that the ligand binds to the protein tighter at acidic pH, and results in more stable conformation, hence explains why crystals were grown at acidic pH. Or what was captured in the crystal structure was the conformation of only nitrogen b is protonated, and the salt bridge only forms between Asp107^{3.32} and nitrogen b.

The carboxyl group is connected to the piperazine moiety through a flexible ethoxy chain. This carboxyl group is protonated at all pHs within the experiment. It potentially could form salt bridges with K191. A potential hydrogen bond could be formed between Y431 and ligand carboxylate group.

The diphenyl ring system of cetirizine sits quite firmly in the “lipophilic subsite” of the binding pocket, and the piperazine moiety is anchored tightly by the D107^{3,32}, while on the other side of the ligand, the carboxyl group is sticking out towards the extracellular side, interacting with the flexible extracellular domain.

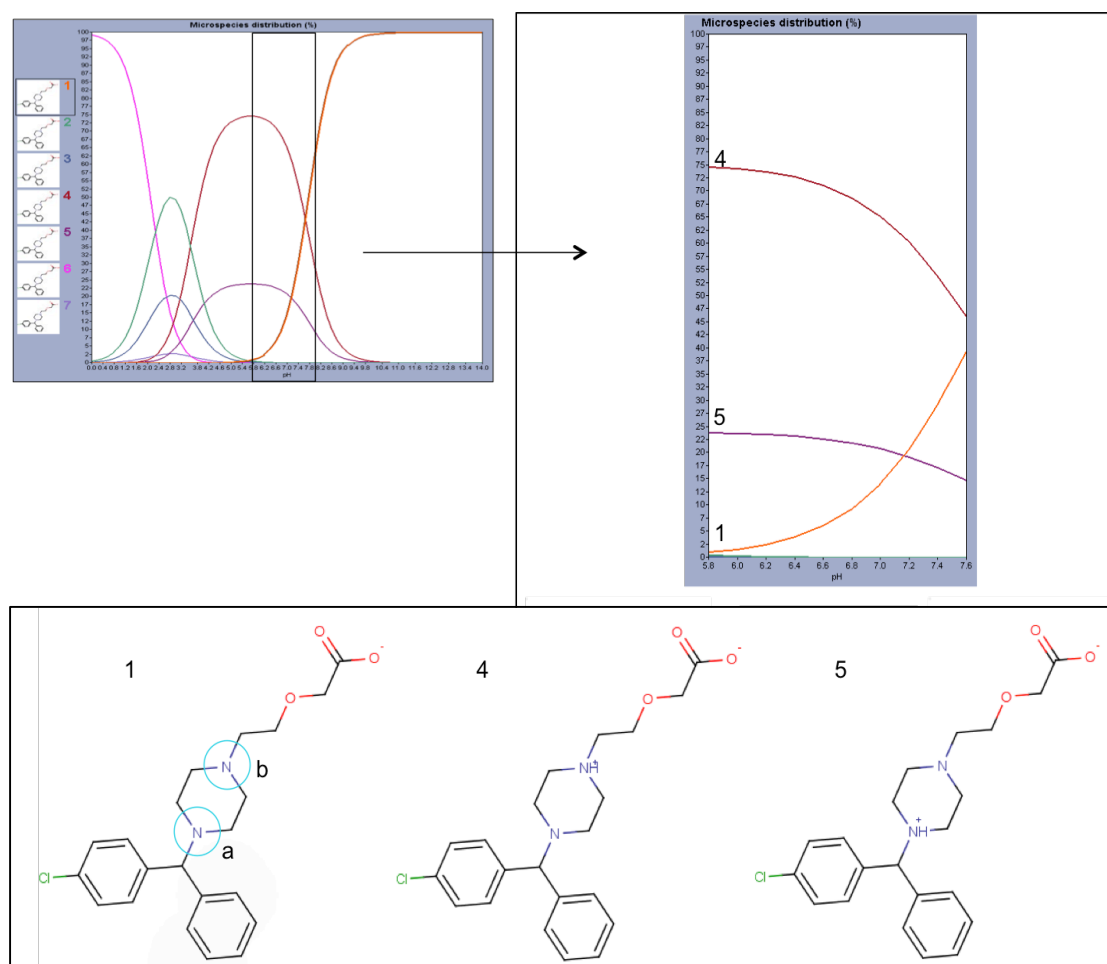


Figure 2.19 pK_a based microspecies distribution (%) of cetirizine. The microspecies 1, 4 and 5 are present in the pH 5.5 to pH 7.5.

2.2.6.2.1 Comparison of cetirizine and doxepin in the binding pocket of H1 receptor

Overall, cetirizine sits in a similar position to doxepin. Superimposition of the two ligands in H1-cetirizine and H1-doxepin complexes shows the ligands are almost overlapped at the benzene ring region and the amine moiety. Although doxepin is much shorter than cetirizine, an anion-binding site was identified in the H1-doxepin structure, and a phosphate ion was modeled into it. The carboxyl group on cetirizine is at the same binding site as phosphate in H1-doxepin structure (Fig.2.20).

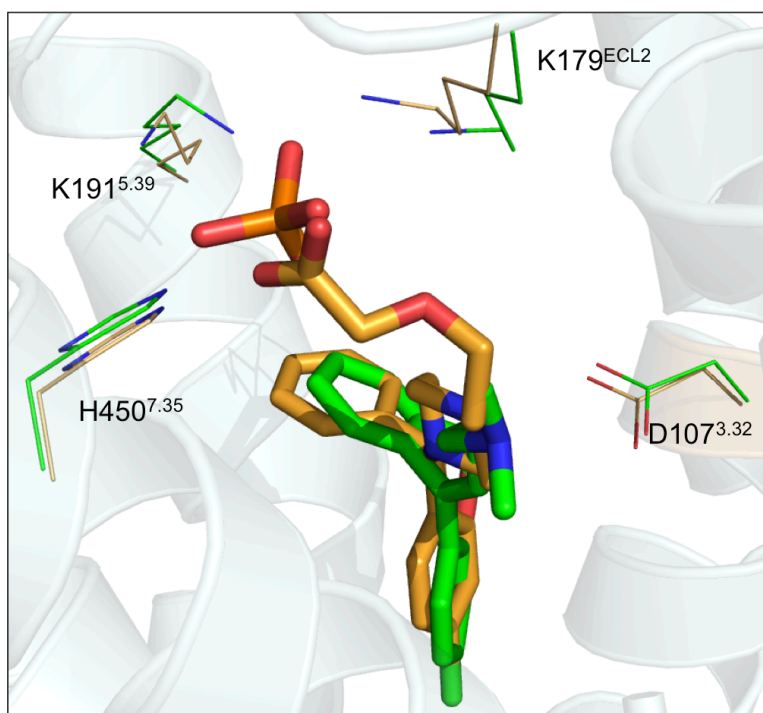


Figure 2.20 Superimposition of doxepin, phosphate ion and cetirizine. Cetirizine is shown as orange sticks, doxepin is shown as green sticks, phosphate ion is shown as golden sticks. Cartoon representation of H1 receptor structure. The residues form salt bridges with the ligands are shown as lines in the matching color with the ligands.

Doxepin is interacting with the receptor only in the “lipophilic subsite” and the highly conserved residue D107^{3,32}. Cetirizine has a more extended structure. Its carboxyl group on the other end of the ligand is sticking out to the extracellular domain, interacting with the “anion binding site”. It binds similarly as the phosphate ion identified at the same position in the H1-doxepin structure (Fig. 2.20).

Although doxepin binds very tightly with the receptor, the residues involved in binding are quite conserved in other GPCRs. Cetirizine, on the other hand, forms salt bridges with the receptor on the “hydrophilic binding domain”, which is specific to the H1 receptor, hence ensured the specificity of this second-generation antihistamine. In the structure of H1-doxepin, the ligand is too short to interpret the extracellular region of the binding pocket, but the “anion-binding site” is proposed by Iwata and his colleagues (Shimamura et al., 2011) to be responsible for the specific binding of non-classical antihistamines. Here, this assumption has been first time proved by a crystal structure. The “anion-binding site”, which in this these called “the hydrophilic binding domain”, indeed is interacting with the carboxyl group of cetirizine through H1 specific residues.

2.2.6.2.2 Position of residue F435^{6,55}

The H1 receptor in H1-doxepin complex shares the same scaffold as in H1-doxepin complex, including the conserved GPCR microdomains, such as CWxP, D(E)RY, NPxxY motifs and disulphide bonds.

Although the backbones of the receptor in both complexes can be almost entirely overlaid, some differences on the residue side chains have also been observed. The

most obvious difference between the two structures is the position of residue F435^{6.55}. Compared with H1-doxepin complex, the F435^{6.55} side chain in H1-cetirizine complex moved away from the transmembrane core, with helix VI as a rotational axis. There is a 3Å distance between the benzene ring centers in the two complexes. The benzene plane of F435^{6.55} was also adjusted to a best angle to be perpendicular to the benzene ring on cetirizine. Their positions are defined very well on the electron density maps as shown in Figure 2.21.

By taking a closer observation on this residue and its near vicinity in both complexes, a conclusion was made that the different positions of F435^{6.55} is caused by a small shift of the benzene rings in the two ligands. The benzene ring on cetirizine is closer to helix VI by about half an angstrom, and it is almost perpendicular to the F435^{6.55} benzene ring, forming a perfect face-to-edge interaction, while the counterpart F435^{6.55} benzene ring in H1-doxepin complex just forms hydrophobic interaction with the edge of the tricyclic moiety on doxepin. Doxepin contacts with Y108 via face-to-edge interaction instead. Although in the H1-cetirizine complex the residue Y108 remains at the same position as in H1-doxepin complex, benzene ring on cetirizine can only form hydrophobic contact with it.

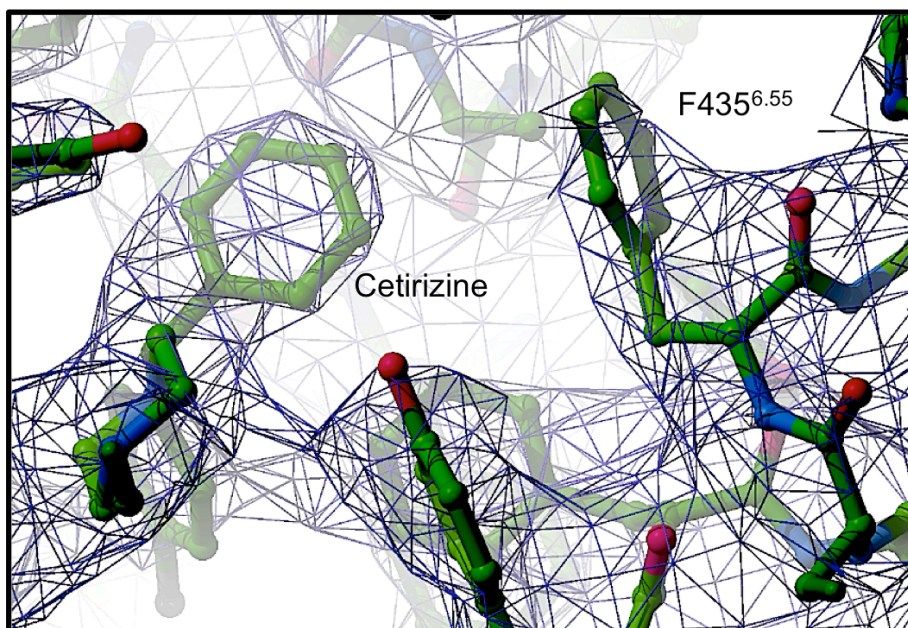


Figure 2.21 Electron densities for F435^{6.55} and benzene ring of cetirizine in structure of H1-cetirizine complex are well defined.

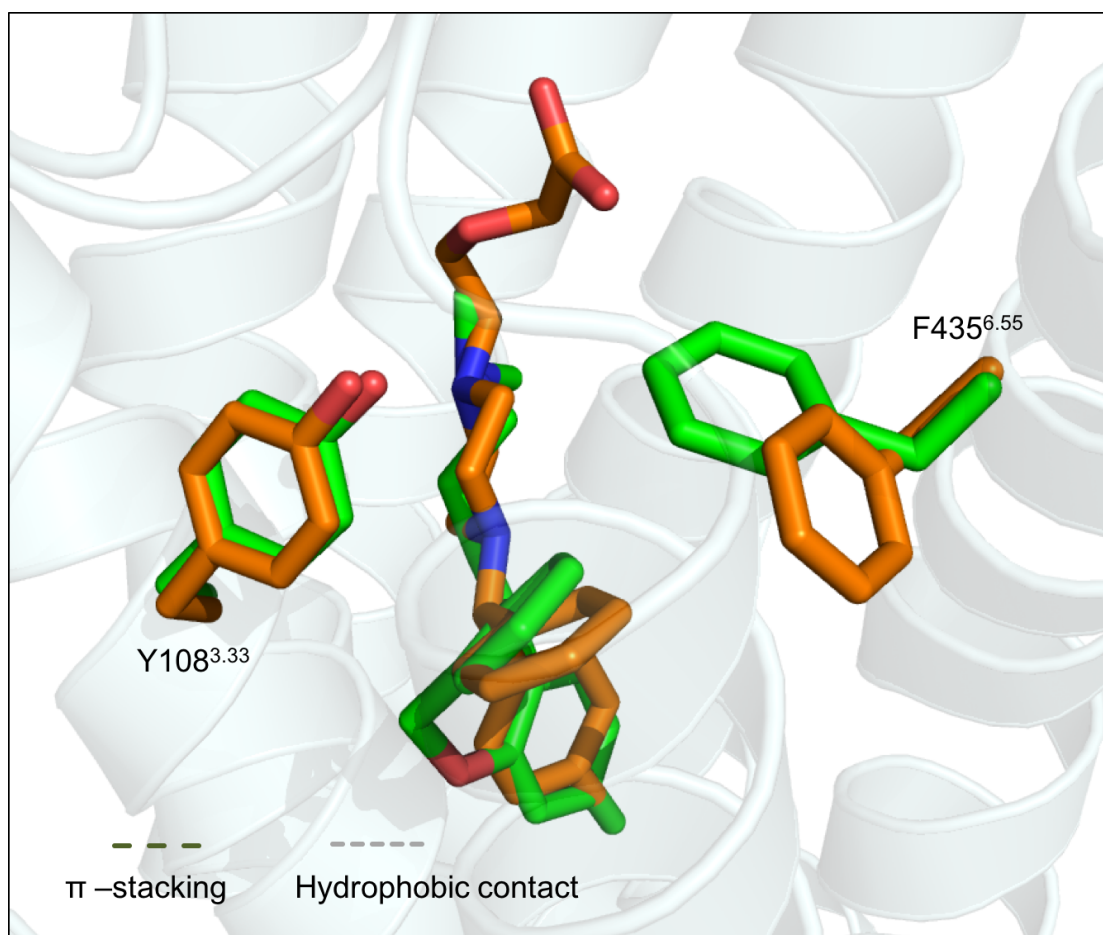


Figure 2.22 Interaction of F435^{6.55} with doxepin vs. Cetirizine. Cetirizine is shown as orange sticks, doxepin is shown in green sticks. F435^{6.55} and Y108^{3.33} are shown in the matching colours with the interaction ligands. Cartoon representation of H1 receptor structure.

This difference can illustrate that small variations on the ligand structure could lead to a different way of binding, and might lead to changes on protein conformation.

2.2.6.3 Interactions between fexofenadine and the H1 receptor

Fexofenadine is another second-generation antihistamine, with the most extended structure. It contains some very similar moieties to cetirizine, including a diphenylmethyl moiety, a piperidine ring, a flexible alkyl chain and a carboxyl group (Fig. 2.23). The main structure difference between these two ligands is that fexofenadine has a very extended structure, with a benzene ring connected to the isobutyric acid moiety in place of the carboxylic acid group in cetirizine. A part of fexofenadine, from its diphenyl end to the amine moiety in piperidine ring inherited the binding mode of its counterpart in cetirizine, which sits deeply and firmly in the “lipophilic subsite”, and anchored tightly with residue D107^{3,32}. While the rest part of the ligand structurally is longer, and binds to the receptor with some different interactions.

The superimposition of the two complexes (Fig. 2.23) presents that the two ligands almost overlapped at the diphenyl group. The diphenyl moiety of fexofenadine also adopts a butterfly-shaped conformation in the crystal structure, and it interacts with H1 within this “lipophilic subsite” in almost the same way with cetirizine: majority interactions are hydrophobic contacts. The piperidine ring of fexofenadine almost overlays on the piperazine ring of cetirizine, the amine moiety binds to D107^{3,32} through salt bridge at the same position (Fig. 2.24).

At the top of the binding site, the carboxyl group extended further towards helix VI, forms salt bridge with Lys450. The benzene ring that attached to the isobutyric acid oriented perpendicularly to H450 and forms a π -stacking interaction, the benzene ring

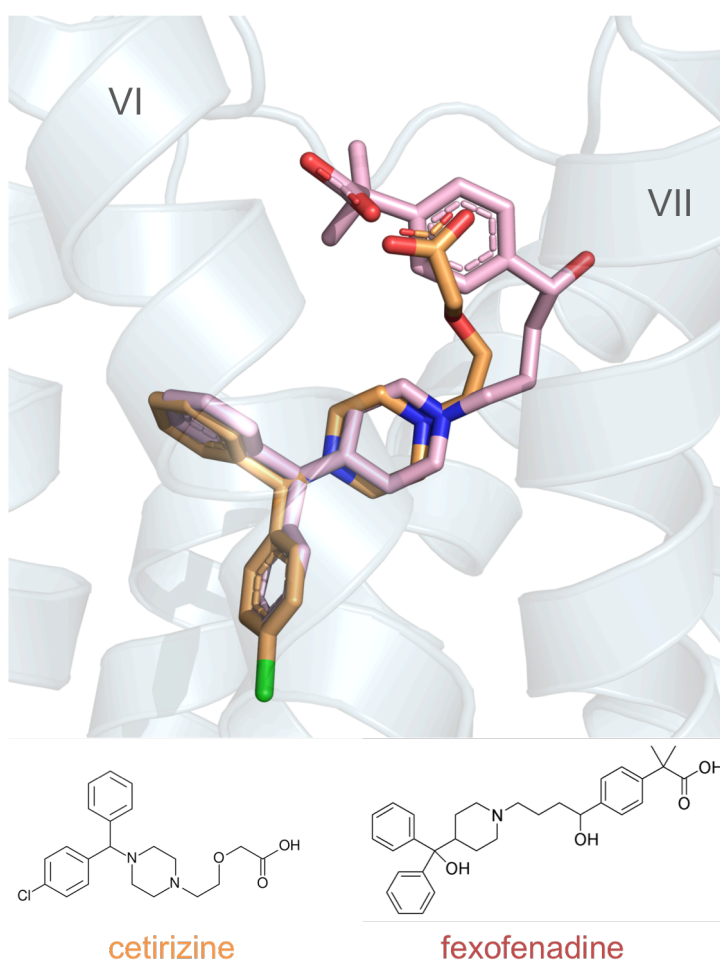


Figure 2.23 Superimposition of cetirizine and fexofenadine. Cetirizine is shown as orange sticks, fexofenadine is shown as pink sticks. Cartoon representation of H1 receptor structure.

also forms hydrophobic interactions with K179. These binding features ensure that fexofenadine bind more specifically to H1.

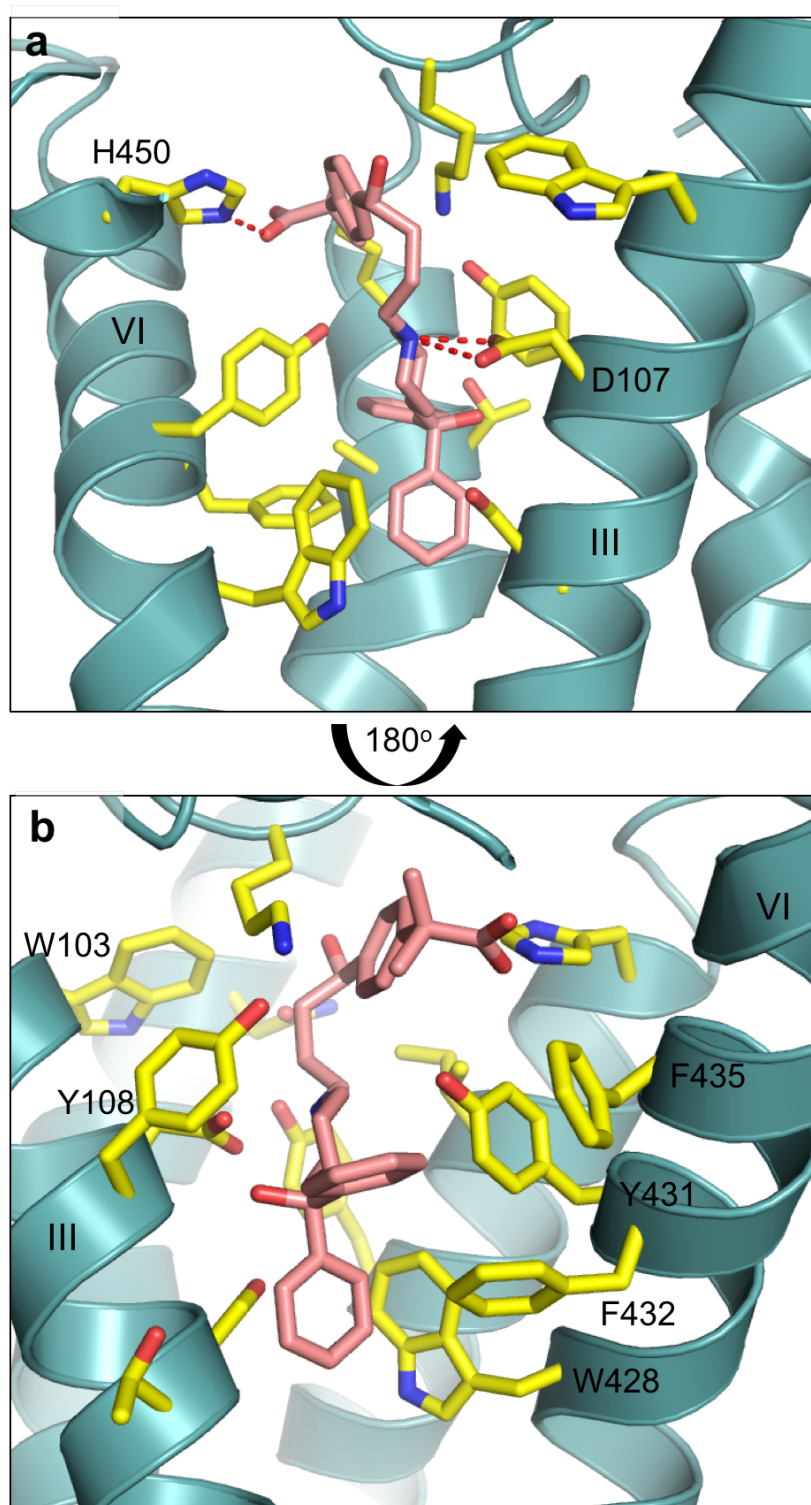


Figure 2.24 Fexofenadine binding pocket. **a.** Direct view on polar contacts with the receptor; **b.** direct view on the "lipophilic subsite". Fexofenadine is shown as salmon sticks, H1 receptor is shown as cartoon presentation, residue side chains are shown as yellow sticks, polar contacts are shown as red dash lines.

2.2.7 Radiolabelled ligand binding assays and Molecular Dynamics simulations

The crystal structures of H1 complexes reported in this thesis for the first time revealed the interactions between H1 and second-generation antihistamines. The specific binding mainly focuses at the “anion binding site”, including residues K179, K191 and H450.

However, the two lysines do not have any electron density as mentioned previously, their positions and conformations are not defined confidently. In order to explore their roles in binding, mutagenesis around these residues has been carried out and 5 mutants were produced. The binding affinities of these mutants obtained from radiolabeled ligand binding assays are shown in Table 2.6.

There was no big variation observed among mutants or between different pH systems (pH7.5 and pH5.8) for the K_D values to [3H]-pyrilamine, all in the range of 1-10 nM. Most of the mutations did not show significant influence on B_{max} values at both pHs, except the triple mutant (K179A+K191A+H450). The B_{max} value of this mutant dropped more than four-fold compare to the wild type at both pHs. The possible reason for it is that mutating all three residues at the same region to alanine simultaneously could cause non-direct effects to the protein stability and function. This mutant might be not very stable or the binding efficiency has decreased significantly.

For K_i values at pH7.5, mutant K191A has an obvious large value for cetirizine, which is 731 nM, while the value for wild type is 140 nM. This difference corresponds with other literatures. The double and triple mutants both include the K191A mutation, and reduced affinities were also observed for them. No significant difference was observed for fexofenadine. It might be due to fexofenadine having a very long and chunky moiety at the extracellular side of the binding site, which enabled the ligand has more interactions with the protein; while for cetirizine, the interaction with residue K191 has a more important role in binding.

Since the overall variations in affinity among the mutants are quite small, over interpretation should be avoided. However, the obvious difference is the values between different pHs. The crystallization condition was at acidic pHs, and the human tissue where inflammation normally takes place also has a local acidic environment. Therefore, the binding assays were also performed at pH5.8. As described in early section, at pH5.8, more proportion of amine moiety on the ligands is protonated, which should improve the affinity of the ligands in theory. And the binding assay results have proved that indeed binding affinity of all the mutants are much higher at pH5.8 than at pH7.5 for cetirizine and fexofenadine, around 10 times and 3 times higher respectively. The observed decreased K_i values for the H191A containing mutants at pH7.5 were not observed at pH5.8. This could be because of the overall much higher affinities have overwritten the mild differences triggered by H450A.

Table 2.6 K_D , K_i and B_{max} values obtained from RLB assays at pH7.5 and pH5.8 for 5 different H1 mutants plus a wild type H1.

At pH7.5

H1-PYRILAMINE

Mutant	K_D (nM)	pK_D	B_{max}
K179A	7.29	8.14	18.55
K191A	6.18	8.21	24.18
H450A	8.08	8.09	10.46
K+K	3.38	8.47	10.16
K+K+H	2.84	8.55	3.27
WT	8.21	8.09	14.67

H1-CETIRIZINE

Mutant	K_i (nM)	pK_i
K179A	103	7.02
K191A	731	6.21
H450A	158	6.84
K+K	641	6.22
K+K+H	745	6.17
WT	140	6.86

H1-FEXOFENADINE

Mutant	K_i (nM)	pK_i
K179A	141	6.86
K191A	84	7.08
H450A	202	6.73
K+K	160	6.81
K+K+H	266	6.62
WT	130	6.89

At pH5.8

H1-PYRILAMINE

Mutant	K_D (nM)	pK_D	B_{max}
K179A	3.00	8.60	18.80
K191A	3.15	8.54	18.88
H450A	2.26	8.69	9.67
K+K	2.77	8.58	14.53
K+K+H	1.64	8.81	4.73
WT	4.22	8.43	19.20

H1-CETIRIZINE

Mutant	K_i (nM)	pK_i
K179A	11.70	7.93
K191A	16.83	7.78
H450A	4.63	8.35
K+K	23.27	7.64
K+K+H	38.22	7.43
WT	16.01	7.80

H1-FEXOFENADINE

Mutant	K_i (nM)	pK_i
K179A	33.28	7.43
K191A	31.12	7.46
H450A	17.07	7.74
K+K	55.77	7.24
K+K+H	35.89	7.28
WT	47.84	7.18

However, with the data of radiolabelled ligand binding assays, it is still not clear about the roles of the three residues. In order to explain the data, molecular dynamics (MD) simulations on the crystal structure of H1 bound to cetirizine were carried out by our collaborator Mr. Benjamin Morris from Biggin's Group, University of Oxford. From the results of the MD simulations, how these interactions varied over the course of the MD simulations were observed, the data is summarized in Table 2.7. It is clear from MD simulations results that the three residues in the mutagenesis do not form constant interactions with the cetirizine.

The comparison between these results and the frequency of interactions between other parts of the ligand, and different binding site residues (Table 2.7), indicating that the interactions between aromatic residues and the aromatic parts of the ligand form very stable interactions.

Table 2.7 Interactions between the nearest sidechain nitrogen atom, and oxygen atoms of cetirizine. Where O1 and O2 are the oxygen atoms of the carboxylic acid group, and O3 is the ether.

Residue 1	Residue 2	Time spent under 4Å distance (%)			Total
		MD run 1	MD run 2	MD run 3	
K179	Cetirizine (O1/O2)	22	24	45	30
K179	Cetirizine (O3)	64	3	3	23
K191	Cetirizine (O1/O2)	4	17	0	7
H450	Cetirizine (O1/O2)	45	28	56	43

Table 2.8 Interactions between the nearest protein sidechain atom to atoms of cetirizine. A π -stacking interaction is defined as the ligand and sidechain being within 5.5Å and a hydrogen bond is defined as being within 4Å. Here, N2 is nitrogen atom at the end of the piperazine nearest the ether.

Residue 1	Residue 2	Time spent interacting (%)
Y108	Cetirizine (ring 1)	91
F109	Cetirizine (ring 2)	95
D107	Cetirizine (N2)	61
W158	Cetirizine (ring 1)	59

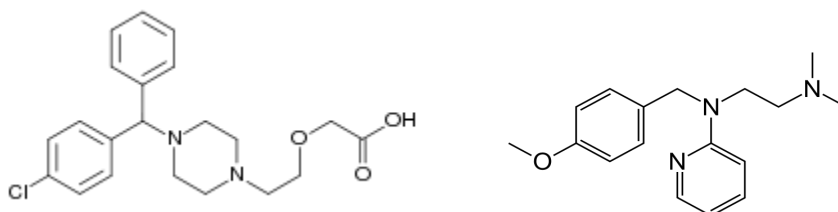


Figure 2.25 Cetirizine (left) and Pyrilamine (right).

It is clear from Tables 2.7 and 2.8 that the key interactions with cetirizine are the interactions lower in the pocket: these interactions have a much higher frequency than any of the interactions with K179, K191, or H450. The percentage contribution of the binding energy is likely much higher for the deep pocket residues than for the charged residues in the mutagenesis data. It is important to note that the ether and carboxylic acid groups of cetirizine are the part of the scaffold that differs from that of pyrilamine (Fig. 2.25), which was used in the binding assays for cetirizine and fexofenadine to compete against. That is to say that pyrilamine likely doesn't form interactions with K179, K191, or H450; or at least it can't form salt bridges or hydrogen bonds with them.

The residues in the mutagenesis are all large and on top of the binding pocket, this coupled with the fact that they likely don't have a large contribution to binding free

energy, leads us to believe that they could instead contribute mainly to binding kinetics. Mutating these residues to alanine could result in a small reduction in binding energy, but a big change in the on or off rates of large ligands such as cetirizine. The effect on binding kinetics would likely be different for smaller ligands such as pyrilamine, so competition assays would be clouded.

2.3 Discussions

Previously, the only crystal structure of H1 was in a complex with doxepin, determined at 3.1 Å (Shimamura et al., 2011). With this single structure of H1, the global impacts of other ligands on the overall protein structure or local impacts on the ligand-binding site are not well understood. Since doxepin is a small ligand and it binds quite deeply, the binding mode of entry area of the binding pocket is not known. Although one structure is sufficient for the docking of some ligands, challenges still remain for more ligands. Firstly, when docking with different ligands, it is hard to estimate the degree of conformational changes of the ligand-binding pocket. Secondly, the binding pocket is quite large, which means there are more than one binding models that could be produced by the docking software, with similar scores.

The two extra crystals structures of H1-ligand complexes described in this thesis could be potentially valuable for further improvement on *in-silico* docking or new drug discovery. Structures have provided further details of ligand binding, which could potentially assist with better analysis of the range of conformational changes different ligand could trigger.

The small structural differences observed between the three complexes demonstrate that different ligands have minor effect on the structure of the receptor. The “lipophilic subsite” and the “D107^{3.32} ionic interaction region” of the binding pocket are highly conserved. Therefore, they are not expected to relate with any selectivity functions on protein subtypes or advanced pharmacological properties (eg. BBB permeability, solubility and metabolism). Their roles are more associated with the tight binding of the ligands. The conformational changes are very small, beyond the resolutions of the X-ray data in this study.

Cetirizine and fexofenadine are both second-generation antihistamines. Previous docking studies using the H1-doxepin structure as a model suggested that they could both bind specifically to H1, via the “anion binding site” (Shimamura et al., 2011). The structures reported here well correspond with the predicted ligand binding modes, confirmed the residue H450^{7,35} is indeed forming salt bridges with both ligands. However, the conformation of the rest part of the “anion binding site” including the K179^{ECL2} and K191^{5,39} could not be modeled accurately in all three complexes. Therefore, the specific binding mode with these residues could not be identified. In order to define their roles in binding, a series of radiolabelled ligand binding assays were designed and carried out for five H1 mutants with key residues at the “anion binding site” mutated to alanines. The results were not as expected, there was no obvious decrease in affinities of the mutants with the key residues mutated to alanines. Possible explanation would be that residues at the “anion binding site” are not associated with binding affinity as expected, but more related to binding specificities. To interpret these data, the MD simulation studies were carried out by our collaborator on cetirizine, the results suggested that, second generation ligands might have minor effect on the protein structure changes and ligand affinities, but significantly larger impact on receptor dynamics.

Thank to crystallography, we get to understand more about GPCRs in the recent years. However, protein crystals normally trap molecules in their low energy conformations, which only present part of the signaling process. Now we start to realize that GPCR dynamics play very important role in their functions and can explain their complicated signaling behaviors. GPCR dynamics involves energetics in different conformational states and the transition between these states. They are the direct cause to ligand binding efficiency and downstream responses for many GPCRs.

To address this problem, molecular dynamics simulations and more biophysical studies, such as NMR and time-dependent derivatization with chemical probes should be carried out as complementary methods.

Chapter Three

Screening of Constructs and Development of Expression Protocols of the Histamine Receptor H3 for The Purpose of Structural Studies

3.1 Overview

For membrane structural studies, obtaining well-ordered diffraction-quality crystals is quite challenging. It requires a sufficient quantity of protein with a good level of homogeneity and stability. Therefore, it is necessary to have a rapid and reliable screening strategy to produce the most suitable constructs and a suitable expression system for protein production prior to crystallization trials. Here, 47 different [Histamine receptor H3 (H3)] constructs were screened by a high throughput protein expression screening platform. Further optimization was carried out by inspecting protein homogeneity using fluorescence-detection size-exclusion chromatography (FSEC). Satisfactory results were obtained as the optimal expression factors in insect cells were identified for the best construct (G4). Large-scale expression and purification of H3 was attempted. While expression was achieved, future work (eg. purification and crystallization) need to be carried out in presence of the antagonist.

3.2 Materials and Methods

3.2.1 High throughput expression screening with different expression systems

The initial high throughput protein expression screens were carried out at OPPF-UK (Oxford Protein Production Facility, Research Complex at Harwell).

3.2.1.1 Construct design and synthesis

Twelve constructs were designed based on full-length sequences of H3 from both human and turkey with different T4 Lysozyme insertion sites on ICL3. All constructs studied were synthesized by GenScript. These twelve constructs were used as templates, and 47 further constructs with various truncations at the N- and C- termini were designed and generated using Clontech In-Fusion cloning system according to the manual. The different H3 constructs were amplified by PCR and subcloned into vector pOPINE-3C-eGFP, which was developed by OPPF for multi-host expression and allows expression in bacterial, mammalian and insect cell lines (Berrow et al., 2009). It has a GFP tag and a hexa-histidine tag at the C terminus, with a 3C proteolytic cleavage site between the target sequence and GFP. GFP works as a reporter for monitoring protein expression systems since it emits a strong fluorescence signal.

3.2.1.2 Small-scale expression screening

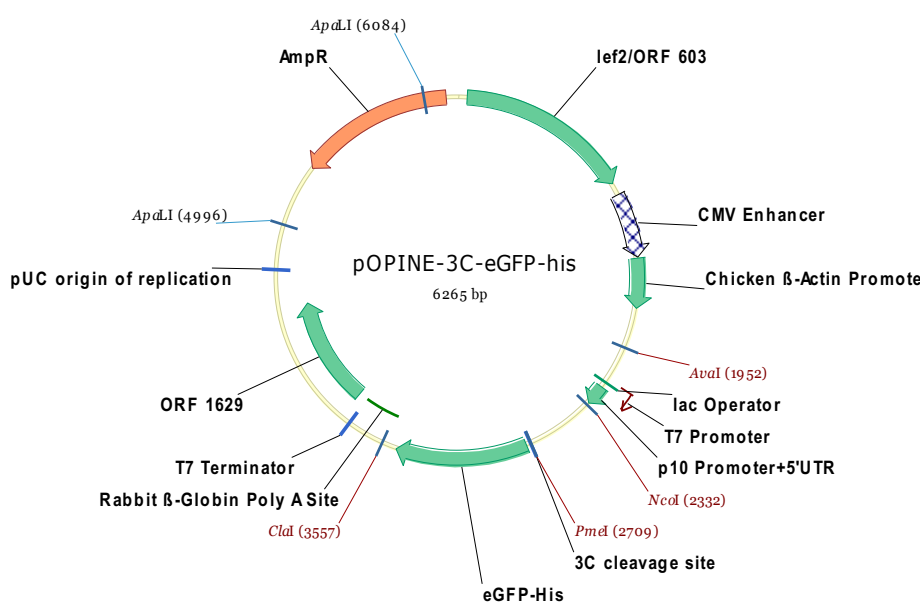


Figure 3.1 pOPINE-3C-eGFP vector (Addgene plasmid # 41125)

3.2.1.2.1 E.coli expression screening

H3 plasmids were transformed into 3 different competent cell lines: C41, BL21Lemo and C43. Positive single transformants were picked from LB agar plates supplemented with carbenicillin, and were used for inoculating 0.7 ml cell cultures using Power Broth medium supplemented with carbenicillin in 96 deep well blocks (ABgene AB-0932), sealed with gas-permeable adhesive seal (ABgene AB-0718). The filled deep well blocks were put into a shaking incubator at 37°C, 225 rpm overnight. The overnight cultures were diluted by adding 15 µl of overnight cultures to 3 ml of TB medium (auto-induction medium) and Power Broth medium supplemented with carbenicillin in 24 deep well plates. The cells were grown in a shaking incubator at 37°C, 225 rpm for 3-5 hours, until the OD₆₀₀ was around 0.5, the cells were induced by adding IPTG to a final concentration of 1mM. Growth was continued for 18 hours at 20°C, 600 rpm. The cells were harvested by centrifugation at 2000g for 10 minutes, and then frozen at -80 °C for further characterization.

3.2.1.2.2 Mammalian cells (HEK 293) expression screening

HEK293 (Human Embryonic Kidney) cell stock (provided by OPPF-UK) was grown to the concentration of $1.5-2 \times 10^5$ cells/ml. The cells were transferred into 24 well plates, 1 ml per well. The plates were incubated at 37 °C overnight under a 5% CO₂ / 95% air atmosphere. The medium was removed from the cell layer by aspiration, and then 1 ml of DMEM medium supplemented with 2% FCS (foetal calf serum) added to each well. The cells were transfected by adding 1 µg plasmid DNA of each construct and 44 µg GeneJuice cocktail (Merck) to the cells and incubated at 37 °C for 3 days under a 5% CO₂ / 95% air atmosphere. The GFP expression level was inspected by

fluorescence microscopy at this point, the supernatant was discarded and cell pellets of all constructs were frozen at -80°C for further characterization.

3.2.1.2.3 Insect cells (baculovirus) expression screening

The insect cells *Spodoptera frugiperda* (Sf9) (provided by OPPF) was propagated in Sf900II medium (Invitrogen) as suspension culture, in 100-500 ml volumes with a medium/flask volume ratio of 1/5 at 26°C. The cells were passaged to a density of 4×10^5 cells/ml every 3-4 days. The virus production and small-scale expression screening were carried out in monolayer and suspension cultures respectively.

To produce the initial virus stock (P0 virus), the Sf9 cells were grown to a density of 5×10^5 cells/ml in Sf900II medium prior to infection. Then Sf9 cells were transferred into two 24 well plates (500 µl per well) and left to attach for 30 minutes at room temperature. Cells were transfected by adding transfection mix (250 ng of bacmid, 100-500 ng of vector DNA, 50 µl of Sf900II medium and 1.5 µl of EugeneHD transfection reagent (Promega)) into each well, followed by a gentle swirl, and incubation for 5-6 days at 26°C. The supernatant was harvested without touching the cell monolayer to create the P0 virus.

To amplify the virus, the Sf9 cells in Sf900II medium were prepared at a concentration of 1×10^6 cells/ml before use. The Sf9 cells were transferred into two 24 well plates, 500 µl per well and left for 30 minutes at room temperature. 5 µl P0 virus stock was added to cells in matching format, and the cells incubated for 6 days at

26°C. The supernatant was harvested to create the second passage virus stock (P1 virus).

For the small-scale expression test, the Sf9 cells in Sf900II medium were prepared at concentration of 1×10^6 cells/ml before use. The Sf9 cells were transferred into 24 well plates with 5 ml per well. 5 or 50 μ l of P1 virus was added to each well and the cells incubated at 26°C with shaking at 120 rpm. 1ml of culture was harvested using 96 well plates after 72 and 96 hours by centrifugation for 15 minutes at 2000g, supernatant was removed and cell pellets were frozen and stored at -80°C for further characterization.

3.2.1.3 Small-scale membrane solubilisation

Cell pellets of three cell lines were disrupted by freezing at -80°C for 1 hour, defrosting for 20 minutes, then re-suspending in 200 μ l of NPI-10 buffer (50 mM NaH_2PO_4 , 300 mM NaCl, 10 mM imidazole and 1% Tween 20, pH 8), supplemented with 1 mg/ml lysozyme, protease inhibitor and DNase I. The cells were shaken at 800 rpm at 4°C for 30 minutes, 1% of DDM and 0.2% CHS was added to the suspension, which was shaken for another 2 hours, followed by ultra-centrifugation at 55k g for 15 minutes. The supernatants were collected by gentle aspiration and crude membrane samples were analyzed by SDS-PAGE gel and/or FSEC.

3.2.1.4 Fluorescence-Detection Size-Exclusion Chromatography (FSEC)

The FSEC system includes a SEC column (Superose 6 10/300 GL) connected to a fluorescence detector to detect the GFP fusion protein during elution. This system is suitable to analyse crude membranes and little protein is required.

The expression level and monodispersity of protein in solubilized membranes from different constructs were screened by FSEC. The column was equilibrated with buffer containing 0.03% DDM, 0.006% CHS, 20 mM Tris (pH7.5) and 150 mM NaCl and 100 µl of each solubilized membrane sample was applied onto the SEC column attached to a Shimadzu High Performance Liquid Chromatography System. The column was ran with a flow rate of 0.3 ml /min and GFP fluorescence monitored by excitation of 488 nm and emission of 512 nm.

3.2.2 Expression optimization with insect cell/baculovirus system

Many factors could affect the protein expression level and quality in baculovirus system, including cell line, medium, oxygen supply, bioreactor size/shape, infection strategy and harvesting time. Infection strategy is normally the first factor to be optimized for a new recombinant protein of interest. The analyzed samples were crude membranes, using GFP fluorescent signal to monitor and compare the production of each target fusion protein. The effects of different factors on the expression of recombinant protein were evaluated in two aspects: protein yield and protein quality by SDS gel or FSEC profiles.

3.2.2.1 Optimization of MOI, TOI and HT for H3 construct C6 and G4.

Initially, the infection strategy was optimized around the multiplicity of infection (MOI), time of infection (TOI), and the time of harvesting. MOI refers to the different passages/titers/quantities of the viruses, TOI refers to the measured cell density at the time of infection, and time of harvesting refers to the time post infection for cell harvesting.

For this series of optimization, factors under optimization are listed below:

- **Construct:** Histamine receptor H3, construct C6 and G4 (Table 3.1).
- **TOI:** 1×10^6 and 2×10^6 cells/ml .
- **MOI:** Virus passage: P1 and P2; Virus volume: 1, 5, 10 ul of virus in 5 ml of cell cultures.
- **Time of harvesting:** 48 and 72 hours post infection.
- **Cell line:** Sf9, Sf21 and Hi5.

3.2.2.1.1 Optimization on MOI: different virus passages

Stock viruses of four H3 constructs (C6, G4, F5 and G6 as shown in Table 3.1) were chosen to compare the P1 and P2 virus stock in order to get a general idea of virus quality. The rest of the tests were performed with constructs C6 and G4. P2 virus (generated after a third passage) was amplified from second passage as described for the P1 virus. For H3 protein expression, 5 ml of Sf9 suspension culture were infected at TOI of 1×10^6 cells/ml, with 5ul of P1 and P2 viruses and harvested at 72 hours post infection. Cell pellets were solubilized and analyzed by SDS-PAGE gel (in-gel fluorescence).

3.2.2.1.2 Optimization on TOI, MOI and time of harvesting

The tests to investigate the factors of TOI and time of harvesting were carried out sequentially, while each of them was done by orthogonal experiments with MOI (1-10 quantitatively). 5 ml of Sf9 suspension culture were infected at TOI of 1×10^6 and 2×10^6 cells/ml, with 1-10 ul of P1 viruses and harvested at 48 and 72 hours post

infection. Cells were harvested and solubilized as described before. The crude membranes were analyzed by the FSEC system.

3.2.2.2 Optimization on cell lines with a better control of virus titer

P2 virus titer was estimated based on the changes on diameters of viable cells, in order to have a better control of virus titer. Further optimization was performed for construct G4 on 3 different cell lines: Sf9, Sf21 and Hi5.

3.2.2.2.1 P2 virus amplification and titer estimation based on

increased cell diameter following infection

10ml of Sf21 cells in flasks were infected with 5 μ l of P1 viruses when the cell density was around 1×10^6 cells/ml, with the viability (above 95%), the cell diameter were measured and recorded. 85 hours post infection, the cell density, viability and cell diameter were measured and recorded. The virus titer was calculated based on the change on the size of cells, which will be described in details in results section. The viruses were collected by a slow spin at 800g for 5 minutes. The P2 virus stock was made to the final concentration of around 5×10^9 virus particles/ml by resuspending the cell pellet with freezing mix (DMSO 10%, media, serum FBS 10%). and aliquoting into 2ml tubes before storing at -80C.

3.2.2.2.2 Cell line optimization

The expression test was performed in three cell lines: Sf21, Sf9 and Hi5, with the TOI of 1×10^6 cells/ml. The tests were performed by orthogonal experiments with the MOI of 1, 5 and 10, and harvesting time of 48 hours and 70 hours post infection. 3ml of

cell culture (Sf9 and Sf21 in Sf900II medium, Hi 5 in Express Five medium, Invitrogen) was infected with P2 virus stock. The cell parameters before infection and before harvesting were monitored. Cells were harvested and solubilized as described before. The crude membranes were analyzed by the FSEC system.

3.2.3 Large-scale expression

Following the determination of the optimal factors for expression in insect cells, large-scale expression of H3 (G4 construct) was attempted. 5 Liters of Sf9 cell suspension in Sf900II medium were grown in a CellBag bioreactor-10L (GE Healthcare). The bioreactor was placed on a rocker platform, which was heated to 27°C, with the rocking speed of 25 rpm at an angle of 7° from horizontal. Infection took place when the cell density of 1×10^6 cells/ml was achieved using 10 ml of P2 recombinant virus. The bioreactor was kept inflated by an air pump with a flow rate of 0.2 liters per minute to maintain bag pressure and oxygen for cell respiration. The cells were harvested at 70 hours post infection by centrifugation for 15 minutes at 2000g.

3.2.4 Purification of turkey H3 for the purpose of obtaining H3-T4L-GFP fusion protein

The purification of turkey H3 was attempted by adapting the purification protocols used for human H1. The main difference was all the steps are performed in the absence of any antagonist. Another difference was the method for cell disruption as the insect cells are much softer than yeast cells.

The insect cell pellet was re-suspended in ice-cold breaking buffer containing 50mM

HEPES pH7.5, 120mM NaCl, 5% (v/v) glycerol, 2mM EDTA and EDTA-free protease inhibitor cocktail tablets, by stirring at 4 °C. The cells were then disrupted and washed by the method of dounce homogenation (20 strokes), ultracentrifugation and re-suspension. The wash was carried out with low salt buffer (10mM HEPES pH7.5, 5% (v/v) glycerol, 2mM EDTA and EDTA-free protease inhibitor cocktail tablets) first, and then repeated twice with the addition of 1M NaCl, 10 mM MgCl₂, and 20mM KCl to the buffer. The membranes were collected by ultracentrifugation at 100,000g for 45 minutes at 4°C. The concentration of whole membrane proteins was estimated using the bicinchoninic acid method (Pierce). The membrane suspension was solubilized by with 1% of DDM, 0.2% of CHS, and the addition of 2mg/ml iodoacetamide, for 1-2 hours, as described in the protocols for H1 in Chapter 2. After the solubilization step, the insolubilized materials were removed by ultracentrifugation. The supernatant was brought to IMAC step using TALON resin in the buffer containing 20mM HEPES pH7.5, 500mM NaCl, 10% glycerol, 0.025% of DDM, 0.05% of CHS. The fusion protein was eluted using 200 mM imidazole. All the fractions were checked by SDS–PAGE gel, and those containing H3-T4L-GFP pooled and concentrated to 2.5 ml. A small portion of the H3-T4L-GFP samples were mixed and incubated with 3C protease for cleavage to test the cleavage efficiency. The samples were assessed by SDS–PAGE gel.

3.3 Results

3.3.1 H3 Constructs

The aim of this work was to find a construct/constructs that could lead to high expression levels, establish a suitable expression system and develop the optimal protocol that could produce good quality of the H3 protein in a sufficient quantity for

the purpose of structure determination. In order to achieve this, several recombinant genes containing H3-T4L with different T4L insertion positions were designed. These H3-T4L genes were later used as templates for making further constructs with various N- and C- terminal truncations. A total number of 47 H3-T4L constructs, together with an eGFP control were formatted into a 96 well plate and numbered accordingly (Fig. 3.2 and Table 3.1), they are used for expression host screening.

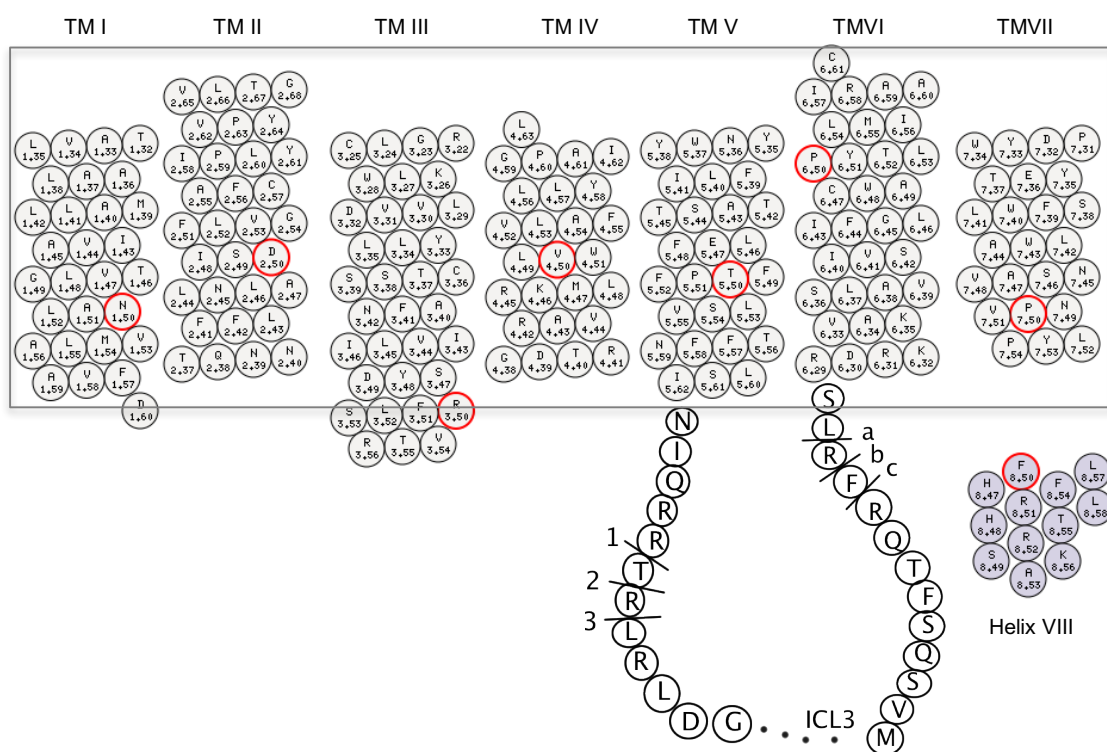


Figure 3.2 Snake-plot of H3 human transmembrane region with ICL3 and Helix VIII. Red circles mark the most conserved residues within each TM. H3 constructs were designed to have fusion body T4 Lysozyme replacing most of the region on ICL3. The various insertion sites are shown as 1, 2, and 3 on TM 5 side and a, b, c on TM 6 side. The insertion sites on H3 turkey are at the equivalent positions. Plot was generated by GPCR-SSFE server (Worth et al., 2011)

Table 3.1. List of H3 constructs with different T4L insertions positions and various N- and C- terminus truncations formatted into a 96 well plate and number accordingly.

Construct number	Template	T4L insertion site	N- truncation	C- truncation
A1	Human T4 window 1	3c	22	476
B1	Human T4 window 1	3c	27	476
C1	Human T4 window 2	2c	22	475
D1	Human T4 window 2	2c	27	475
E1	Human T4 window 3	3b	22	475
F1	Human T4 window 3	3b	27	475
G1	Human T4 window 4	2b	22	474
H1	Human T4 window 4	2b	27	474
A2	Human T4 window 5	1a	22	472
B2	Human T4 window 5	1a	27	472
C2	Human T4 window 6	3a	46	494
D2	Human T4 window 6	3a	51	494
E2	Human T4 window 6	3a	1	494
F2	Human WT	--	1	445
G2	Turkey T4 window 1	3c	47	508
H2	Turkey T4 window 1	3c	47	513
A3	Turkey T4 window 1	3c	53	508
B3	Turkey T4 window 1	3c	53	513
C3	Turkey T4 window 1	3c	59	508
D3	Turkey T4 window 1	3c	59	513
E3	Turkey T4 window 2	2c	47	507
F3	Turkey T4 window 2	2c	47	512
G3	Turkey T4 window 2	2c	53	507
H3	Turkey T4 window 2	2c	53	512
A4	Turkey T4 window 2	2c	59	507
B4	Turkey T4 window 2	2c	59	512
C4	Turkey T4 window 3	3b	47	507
D4	Turkey T4 window 3	3b	47	512
E4	Turkey T4 window 3	3b	53	507
F4	Turkey T4 window 3	3b	53	512
G4	Turkey T4 window 3	3b	59	507
H4	Turkey T4 window 3	3b	59	512
A5	Turkey T4 window 4	2b	47	506
B5	Turkey T4 window 4	2b	47	511
C5	Turkey T4 window 4	2b	53	506
D5	Turkey T4 window 4	2b	53	511
E5	Turkey T4 window 4	2b	59	506
F5	Turkey T4 window 4	2b	59	511
G5	Turkey T4 window 5	1a	47	504
H5	Turkey T4 window 5	1a	47	509
A6	Turkey T4 window 5	1a	53	504
B6	Turkey T4 window 5	1a	53	509
C6	Turkey T4 window 5	1a	59	504
D6	Turkey T4 window 5	1a	59	509
E6	Turkey WT	--	1	469
F6	Turkey WT	--	37	469
G6	Turkey WT	--	59	469
H6	eGFP (control)	--	--	--

3.3.2 Small scale expression screening

3.3.2.1 *E.coli* expression screening

From the SDS-PAGE gel analysis, despite various cell strains and medium types having been tried, there was very little expression of H3-GFP fusion protein from *E.coli* system. This just confirmed that *E.coli* is a poor host for heterologous expression of H3, which might be expected, since eukaryotic proteins like GPCRs do not usually express well in *E. coli*. In general, GPCRs do not express well in *E.coli* as the composition of membrane is different, which is important to maintain the integrity of eukaryotic protein structures. This system has been often used for screening because of the low cost, and fast growth rate. However, in this study, because of the small-scale high throughput expression screening strategy, the cost and growing time became minor concerns.

3.3.2.2 4.2.2.2 Mammalian expression screening

Inspection of the HEK293 cells for expression levels 72 hours after transfection was performed using fluorescence microscopy. There was almost no expression of GFP for any construct except for the eGFP control (data not shown). This might be caused by an unhealthy cell stock. Mammalian cell expression is quite problematic, the cell lines health is critical for expression (Andrell and Tate, 2013). However, compared to other expression hosts, mammalian cells provide the most native-like environment for eukaryotic protein expression, (for example, lipid composition) and the method is believed to be an currently under-used system because of the high cost and difficulties in maintaining healthy cell lines (Andrell and Tate, 2013). However, it is worth putting more effort into the expression optimization if one tried other expression hosts

and still expects better homogenous protein yield. Such as for setting crystallization trials, minimum of 0.5-1 mg of pure concentrated protein is needed.

3.3.2.3 Insect cells expression screening

From the whole cell GFP expression level of 72 hours post infection (Fig. 3.3), turkey constructs expressed more than the human variants. It shows a similar expression pattern when different volumes of P1 virus were used, while infection with 5 μ l per 5 ml of culture of P1 virus gave slightly higher expression level than with 50 μ l per 5 ml of culture. Extending the infection time to 96 hours did not show much difference in the yield of H3 (data not shown).

SDS-PAGE gel analysis illustrated in Figure. 3.4, shows the H3-GFP fusion proteins expressed well in insect cells. However, it was observed that large amounts of free GFP also were produced, which may result from adventitious proteolysis. The different amounts of P1 virus still gave similar results in terms of expression, while the infection time present notably different results on the gel. Cells harvested 72

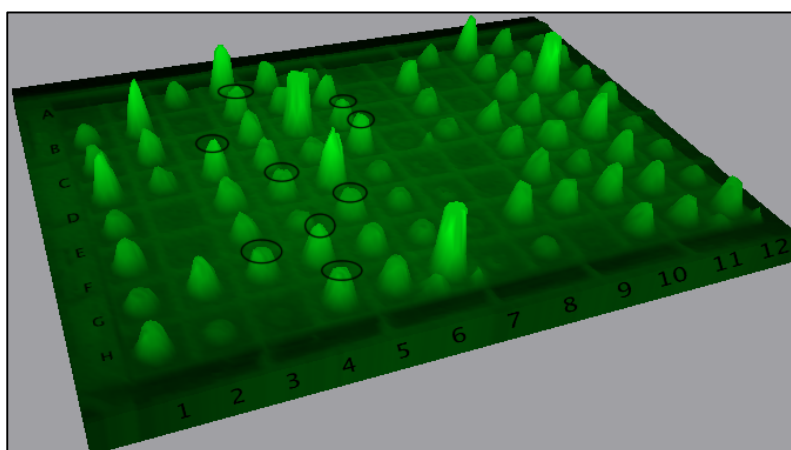


Figure 3.3 GFP fluorescence 3D image of insect cell pellets. Cells were harvested 72 hours post infection. Cell pellets in 96 well plate format. 5 μ l P1 virus expression in columns 1-6, and 50 μ l P1 virus expression in columns 7-12. Cell pellets in black circles presented better profiles on SDS-PAGE gel.

hours post infection show promising GFP-fusion protein bands, while cells harvested 96 hours post infection display an increased amount of free GFP bands and almost no target protein. This probably because of the p10 promoter on pOPINE vector (Fig. 3.1) is a late promoter, which only switches on about 24 hours after infection, the expression of target proteins driven by this promoter should start 48-72 hours after infection, and the virus cycle can cause cell death and lysis about 4-5 days post infection (Massotte, 2003). Therefore, 72 hours is the good time point that protein production has started long enough for post-translational modifications, but not too long causing cell death.

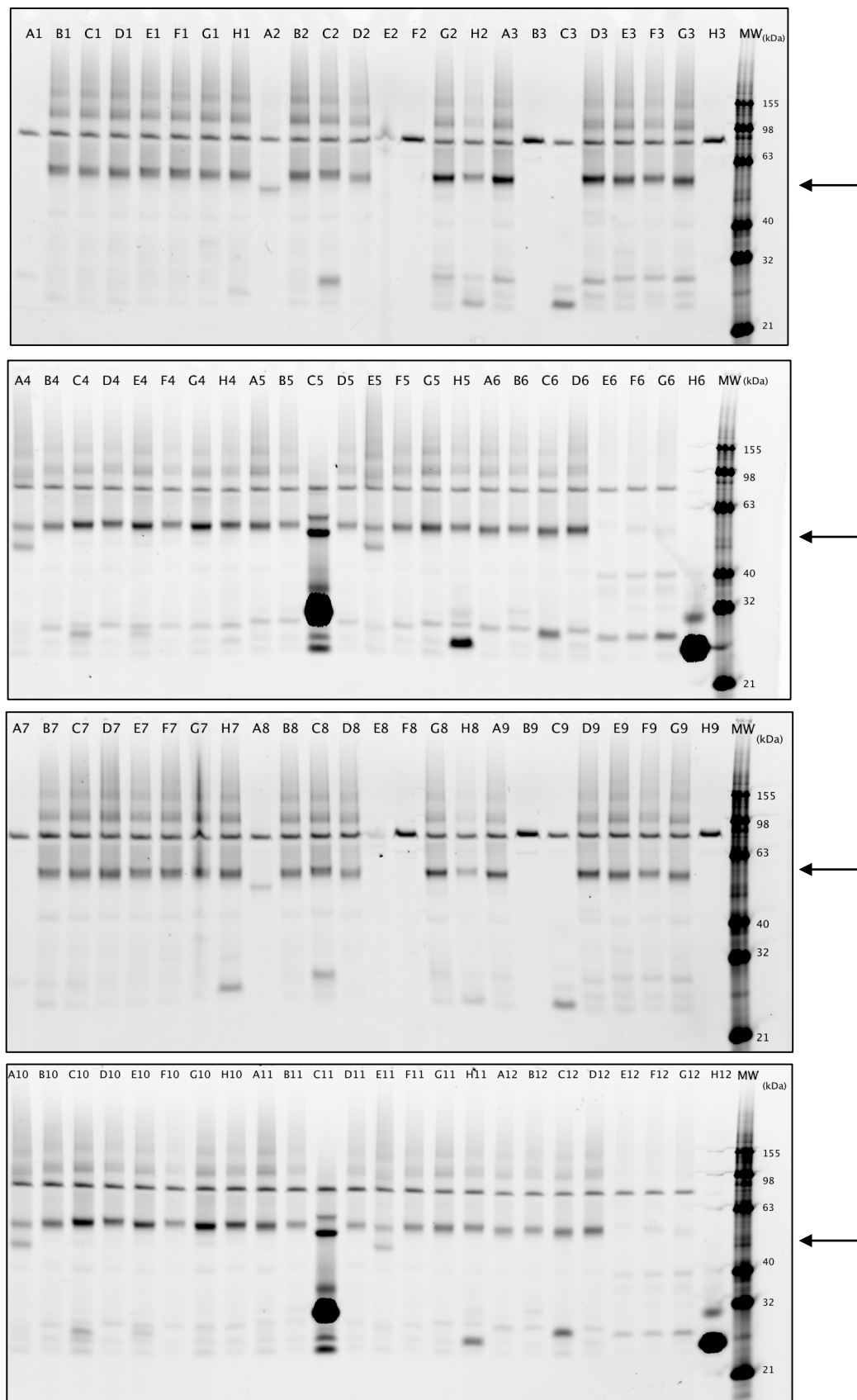


Figure 3.4 SDS-PAGE gel analysis of H3 insect cells expression screen. Solubilized crude membranes are from insect cells expression system, 72 hours post infection. 5 μ l P1 virus expression is A1-H6, and 50 μ l P1 virus expression is A7-H12. Gels were visualized by in-gel fluorescence. Arrows indicate GFP fusion H3 protein bands.

3.3.2.4 Fluorescence-Detection Size-Exclusion Chromatography (FSEC)

Solubilized membranes from insect cells with an infection time of 72 hours were further characterized using FSEC. The SEC profiles of most constructs present multiple, broad and asymmetric peaks, suggesting they were not stable in the detergent-containing buffer or incorrectly folded during expression. However, 9 constructs out of 47 gave reasonable symmetric and sharp peaks, the retention time is indicating a H3-T4L monomer as shown in Figure 3.5, Interestingly, they are all turkey constructs, indicating that turkey H3 protein sequences are more prone to express with better folding and stability.

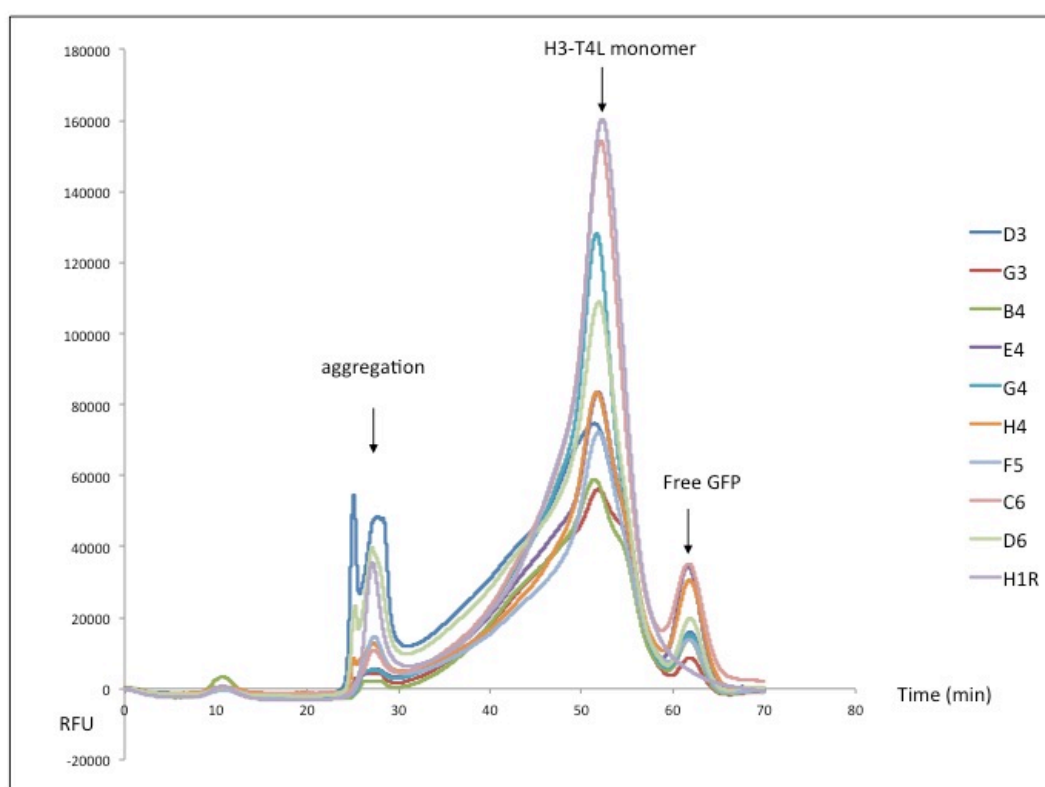


Figure 3.5 FSEC profiles of best H3 constructs. Crude membranes were from insect cells expression system, 72 hours post virus infection. H1 profile indicates the correct elution time as a positive control. The arrows indicate the estimated positions of aggregate, H3 monomer, and free GFP from left to right.

Within the 9 constructs, the FSEC profiles of C6 and G4 shown very small aggregation peaks, indicating that the proteins of those constructs were quite stable and did not tend to aggregate. Comparison of the C6 SEC curve to a standard curve produced from the H1 membranes, suggests C6 could potentially yield 2 mg of target protein from 8 litres of culture, which is a sufficient yield for crystallization trials. Compared to C6 and G4, the SEC profiles of the other 7 constructs show much lower expression levels and the aggregation peaks are quite large. They might worth further optimization on expression conditions and detergents if C6 and G4 fail to yield crystals. Currently, constructs C6 and G4 represent the most suitable candidates to carry on to large-scale expression, purification and crystallization trials. It is worth keeping in mind that, FSEC was performed on diluted unpurified samples, while in crystallization trials, the purified protein is typically much more concentrated. Therefore, further work needs to be done to monitor the characteristics of these two constructs.

From the initial screening work, the best expression system has been in insect cells, with an infection time of 72 hours. The constructs with the best bands on SDS-PAGE gel did not give nice FSEC profile, indicating in the insect cells expression system using pOPINE vector, whole cells expression level does not have much correlation with the usable yield of target protein (Fig. 3.4 and Fig. 3.5). The most promising constructs are C6 and G4, although they do not show the highest levels of GFP expression. The solubilization buffers used during the screening work did not contain any antagonists, which are very important to improve protein stability by maintaining the receptors in their inactive state. During the proper protein purification procedures,

the antagonists should be added from solubilization step and through all the procedures afterwards, by which, in principle, should improve the protein stability and yield dramatically.

3.3.3 Expression optimization with baculovirus system

3.3.3.1 Optimization on MOI, TOI and time of harvesting for H3 construct C6 and G4.

3.3.3.1.1 Optimization on virus passages

The expression level of each H3 construct was much higher when infected with P1 virus than when infected with P2 virus as shown by SDS-PAGE (Fig. 3.6). Several thick bands in the range between 23-41 kDa were observed from the P2 virus infected expression. They might be free GFP, GFP degradation products or GFP-peptide fragments. Although there was also a free GFP band observed for C6 construct infected with P1 virus, it can be easily separated during the purification procedure. No bands were observed on the lane for G4 construct (P2 virus infected), which might be due to mistakes in sample loading. However, the conclusion is that H3 expression level infected with P1 virus is better, and the factor of P1 virus was fixed during further optimization.

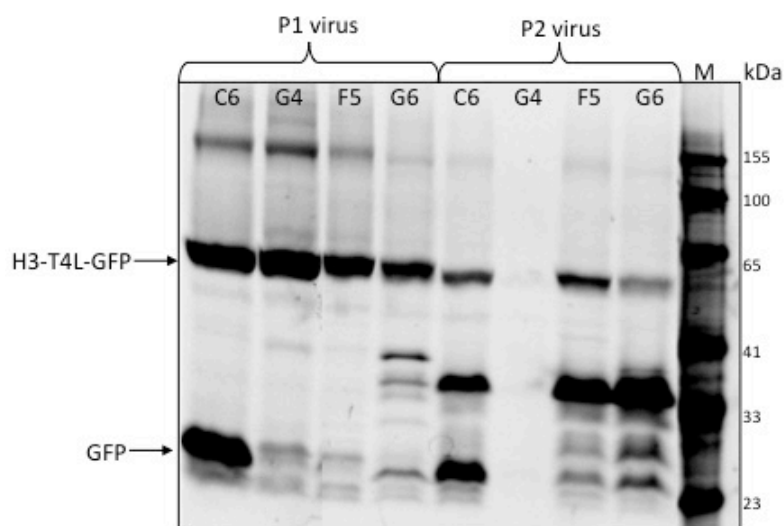


Figure 3.6 SDS-PAGE gel analysis of H3 insect cells expression using P1 and P2 viruses. Solubilized crude membranes are from insect cells expressed with the P1 and P2 viruses of H3 constructs (C6, G4, F5 and G6). Gels were visualized by in-gel fluorescence.

3.3.3.1.2 Optimization on MOI, TOI and time of harvesting

In the FSEC profiles of TOI optimization, higher peaks of H3-GFP (construct C6) fusion protein have been observed for the TOI of 2×10^6 cells/ml (Fig. 3.7, Top panel), indicating a higher expression level. In terms of MOI, MOI of 1 has a strong influence on the expression level corresponding to the different TOIs, with a nearly doubled expression level at the high TOI, while MOI of 5 and 10 did not show significant differences. This result indicates that high TOI has the most significant influence on expression level when a low MOI is used. However, the peaks of the target protein are broader in the high TOI, and the free GFP and large aggregation peaks were also observed.

On the other hand, the FSEC profiles of construct G4 presented some different results. The peaks of target protein were not significantly different between TOI of 2×10^6 cells/ml and 1×10^6 cells/ml, but some large aggregation peaks have been observed at

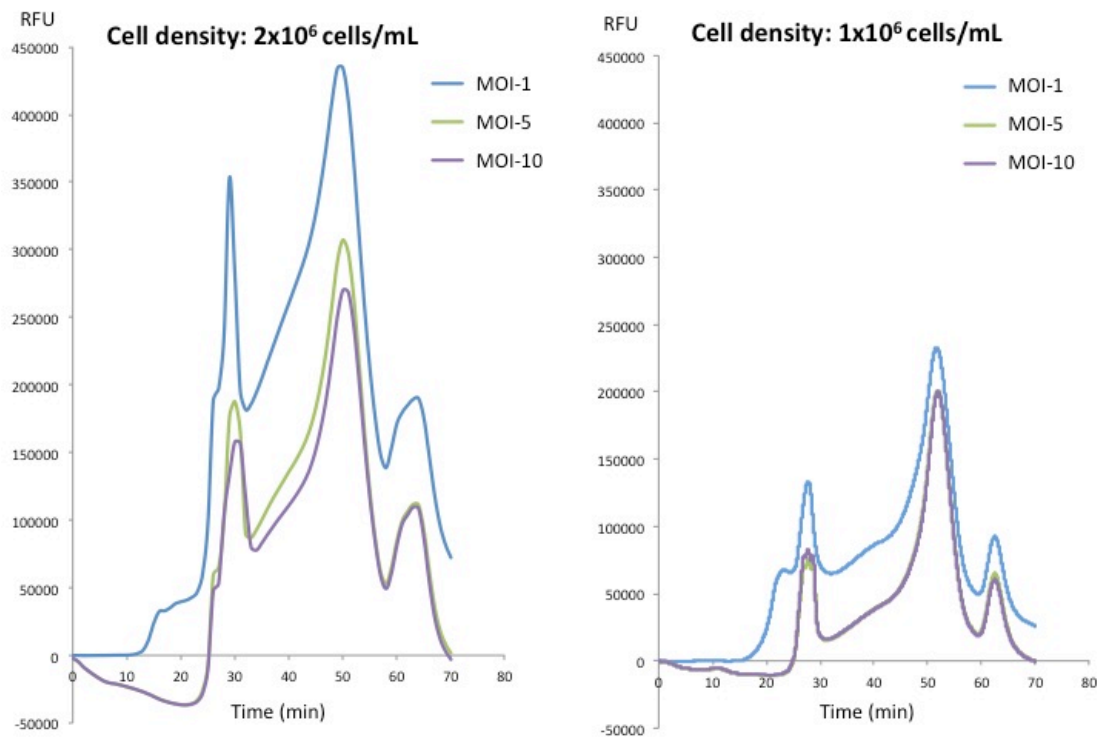
the high TOI (Fig. 3.7, Bottom panel), indicating higher expression level of target protein might have been produced, but then turned to aggregation. For both constructs, the optimal MOI is 1. Therefore, the TOI of 1×10^6 cells/ml is fixed for following optimization.

Optimization on time of harvesting of 48 and 72 hours post infection are compared in Figure 3.8. The FSEC profiles of the cells with late harvesting time have higher peaks for target protein for both constructs (C6 and G4). Harvesting time of 52 Hours post infection was also quickly checked (data not shown), and it turned up as a transition point in the increasing phase of expression level.

Construct G4 always has a much smaller free GFP peak, the target protein peak is also sharper and more symmetrical than construct C6. From this point, G4 construct is used for further optimization.

Optimization on TOI

Fixed factors: Construct C6, P1 virus, harvested at 72 hours post infection.



Optimization on TOI

Fixed factors: Construct G4, P1 virus, harvested at 72 hours post infection.

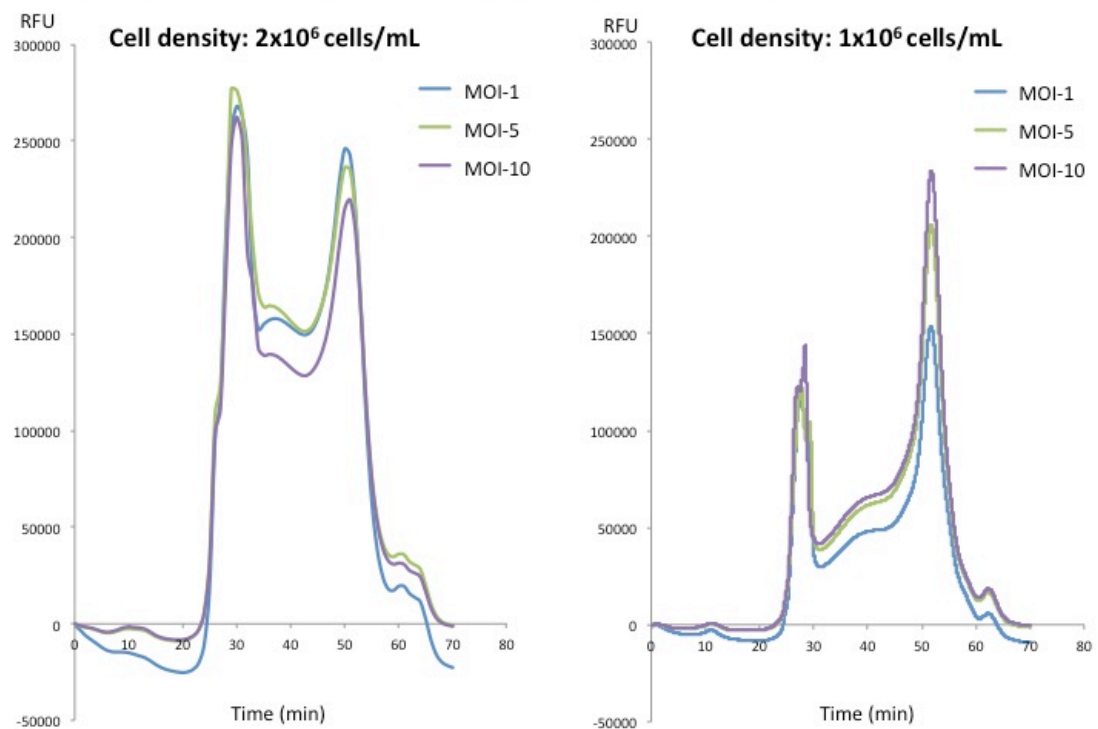
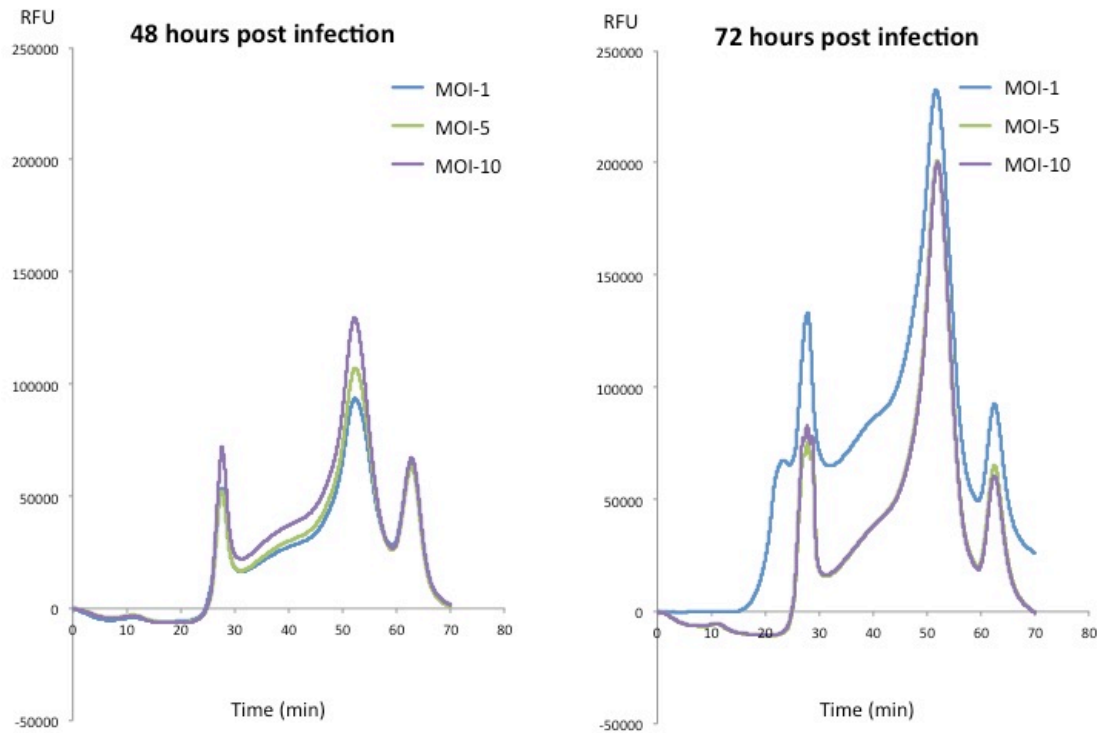


Figure 3.7 FSEC profiles of H3 expression optimization on the factor of TOI. Comparison of crude membrane FSEC profiles between TOI of 1×10^6 cells/ml and 2×10^6 cells/ml. Top panel: H3 construct C6. Bottom panel: H3 construct G4.

Optimization on Time of Harvesting

Fixed factors: Construct C6, P1 virus, TOI=1x10⁶ cells/mL



Optimization on Time of Harvesting

Fixed factors: Construct G4, P1 virus, TOI=1x10⁶ cells/mL

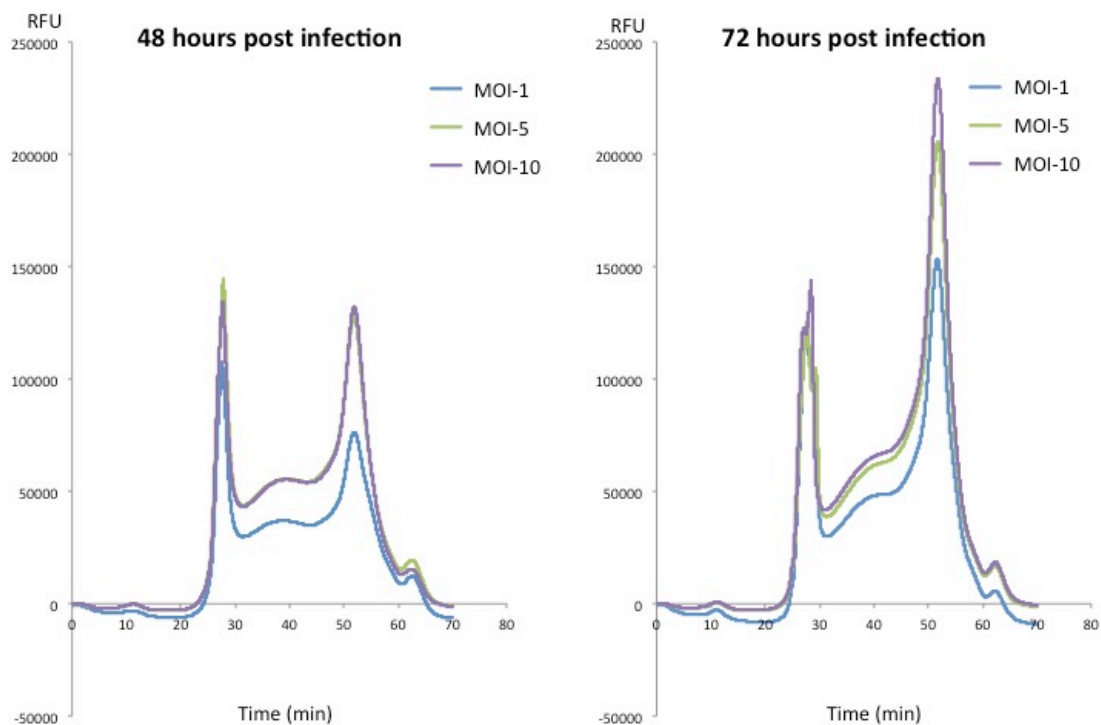


Figure 3.8 FSEC profiles of H3 expression optimization on the factor of Time of Harvesting.

Comparison of crude membrane FSEC profiles between the harvesting time of 48 and 72 hours post infection. Top panel: H3 construct C6. Bottom panel: H3 construct G4.

3.3.3.2 Optimization on cell lines with a better control of virus titer.

3.3.3.2.1 P2 virus amplification and titer estimation based on increased cell diameter following infection

Table 3.2 Sf 21 cell parameters before and after infection

Day 1: Sf21 cells before infection with P1 viruses C6 and G4			
Virus	Cell density (cells/ml)	Viability	Cell diameter (µm)
G4	1×10^6	97%	11.2
C6	0.9×10^6	96%	11.1
Day 5: 85 hours post infection			
Virus	Cell density (cells/ml)	Viability	Cell diameter (µm)
G4	3.6×10^6	96%	14.2
C6	3.4×10^6	94%	13.6

Previous viruses were produced by following a high through put pipeline at OPPF, the exact virus titer was not determined, and MOI was defined as volume of virus to volume of cells. In this section, P2 virus was amplified while cell parameters were monitored and the virus titers were estimated based on changes of cell diameters (Janakiraman et al., 2006) (Table 3.2). At the time of 85 hours post infection, the cell density has increased above 3 times and still with the viability of more than 90%,

indicating the cells are still in a good condition. Around 100 virus particles are produced per 1µm increasing in cell diameter.

In the case of G4, cell diameter increased by 14.2-11.2=3µm, 300 virus particles were produced per cell. The virus concentration is $3.6 \times 10^6 \times 300 = 1.08 \times 10^9$ virus particles per 1ml of cell culture.

In the case of C6, cell diameter increased by 13.6-11.1=2.5µm, 250 virus particles were produced per cell. The virus concentration is $3.4 \times 10^6 \times 250 = 0.85 \times 10^9$ virus particles per 1ml of cell culture.

The P2 virus stock was made to the final concentration of around 5×10^9 virus particles/ml for the following tests.

3.3.3.2.2 Optimization on three cell lines

By knowing the exact titer of the P2 virus, a formula could be used to calculate the volume of virus needed for the MOI of 1, 5 and 10. The results are listed in Table 3.4.

$$\frac{\text{MOI} \times \text{cell density} \times \text{culture volume}}{\text{Virus concentration}} = \text{volume of virus}$$

Where MOI=1, 5 and 10

Virus concentration= 5×10^9 virus particles/ml

Table 3.3 Cell parameters before infection

Cell line	Cell density (cells/ml)	Cell diameter (µm)	Viability	Note
Sf21	1.7×10^6	10.3	98%	Diluted to 1×10^6
Sf9	1.2×10^6	10.6	92%	--
Hi5	1.3×10^6	12.4	99%	--

Culture volume=3ml

Here, Multiplicity of infection (MOI) refers to the plaque forming units (pfu) of virus used for infection / number of cells.

In three cell lines, the cell parameters were measured and recorded in Table 3.5. At 48 hours post infection, the cell densities of sf9 cell increased from 1.2×10^6 cells/ml to 2.5×10^6 cells/ml at MOI=1, 3.3×10^6 cells/ml at MOI=5 and 3.5×10^6 cells/ml at MOI=10. This indicating that the cells were still growing after infection, the cell viability remained around 95%. At 70 hours post infection, cell densities at MOI of 5 and 10 remained constant, while at MOI=1, the cell density increased from 2.5 cells/ml to 2.9 cells/ml. These changes suggest that cells were all infected at 48 hours post infection for the MOI of 5 and 10, while for MOI of 1, a small part of the cells were not infected at 48 hours post infection, and were still able to divide further. Although a lower MOI result in a higher cell counts, the infection was not a synchronous infective process throughout the entire culture. The cell viabilities at 70 hours post infection were all just under 90%, indicating the cell lysis just started.

The cell densities of sf21 cell increased from 1×10^6 cells/ml to around 1.3×10^6 cells/ml at all the MOIs at 48 hours post infection, indicating that the cells were still growing after infection, the cell viability was around 86% showing they have started breaking down. At 70 hours post infection, cell densities increased to around 1.7×10^6 cells/ml at all MOIs, suggesting a further cell growth. Cell viability at this point dropped to 64% and 72% for MOI of 5 and 10, respectively. This level of low cell viabilities is a sign that harvesting was a bit late. However, for the MOI of 1, the cell viability was still good, which is 90%. These data shows the infection did not spread

Table 3.4 Volume of the virus used in the expression tests

Cell line	Volume of virus (μL)		
	MOI 1	MOI 5	MOI 10
Sf21	0.6	3	6
Sf9	0.72	3.6	7.2
Hi5	0.78	3.9	7.8

Table 3.5 Cell parameters measured 48 and 70 hours post infection (construct G4)

		48 hours post infection			70 hours post infection		
Cell line	MOI	Cell density (cells/ml)	Viability	Cell diameter (μm)	Cell density (cells/ml)	Viability	Cell diameter (μm)
Sf9	1	2.5×10^6	93%	13.4	2.9×10^6	82%	14.1
	5	3.3×10^6	95%	13.4	3.2×10^6	86%	13.5
	10	3.5×10^6	94%	13.7	3.5×10^6	72%	13.8
Sf21	1	1.3×10^6	86%	14	1.6×10^6	90%	14.3
	5	1.3×10^6	86%	14	1.8×10^6	64%	13.3
	10	1.4×10^6	87%	13.4	1.8×10^6	72%	13.2
Hi5	1	4×10^6	92%	16.4	3.4×10^6	73%	13.9
	5	3.1×10^6	72%	16.1	2.8×10^6	79%	12.2
	10	2.5×10^6	75%	15.3	2.3×10^6	66%	11.5

throughout the entire Sf21 cells cultures even after 48 hours post infection, regardless the MOIs were used.

In the case of Hi5 cells, the cell densities increased from 1.3×10^6 cells/ml to 4×10^6 cells/ml at MOI=1, 3.1×10^6 cells/ml at MOI=5 and 2.5×10^6 cells/ml at MOI=10 at 48 hours post infection, indicating that the cells remain growing after infection, the cell viability remained around 92% for MOI=1 showing they were in a very healthy states,

but below 75% for MOI of 5 and 10. A rapid cell lysis has occurred for both of them. All cell densities dropped at 70 hours post infection, as well as the cell viabilities have dropped below 80%. Indicating that all the cells have been infected, and Infected cells started lysis quite quickly.

The expression level and protein quality from different cell lines was assessed by FSEC, as shown in Figure 3.9 and Figure 3.10. Cells harvested 48 hours post infection, expression levels were not high for Sf9 and Sf21 cell lines, the FSEC profiles for them are very similar, while for Hi5 cell line, the expression level is about three fold higher than the other two cell lines. However, a shoulder was observed on the left of protein peak. The protein quality is a downside from Hi5 cell line.

Expression levels of Sf9 and Sf21 for cells harvested 70 hours post infection have increased, with Sf9 showing higher levels. The Aggregation and free GFP peaks are both small and the protein peak is relatively symmetric and sharp. The Hi5 protein peak is much smaller, with a huge free GFP peak, meaning most of the protein has degraded. Therefore, the optimal expression strategy for turkey H3 (construct G4) is Sf9 cell line, MOI of 10, TOI of 1×10^6 cells/ml, harvested at 70 hours post infection. This expression strategy could potentially yield 0.4mg fusion protein per liter of cell culture.

Optimization on Cell lines

Fixed factors: Construct G4, P2 virus, TOI=1x10⁶ cells/mL, Time of harvesting=48 hours post infection

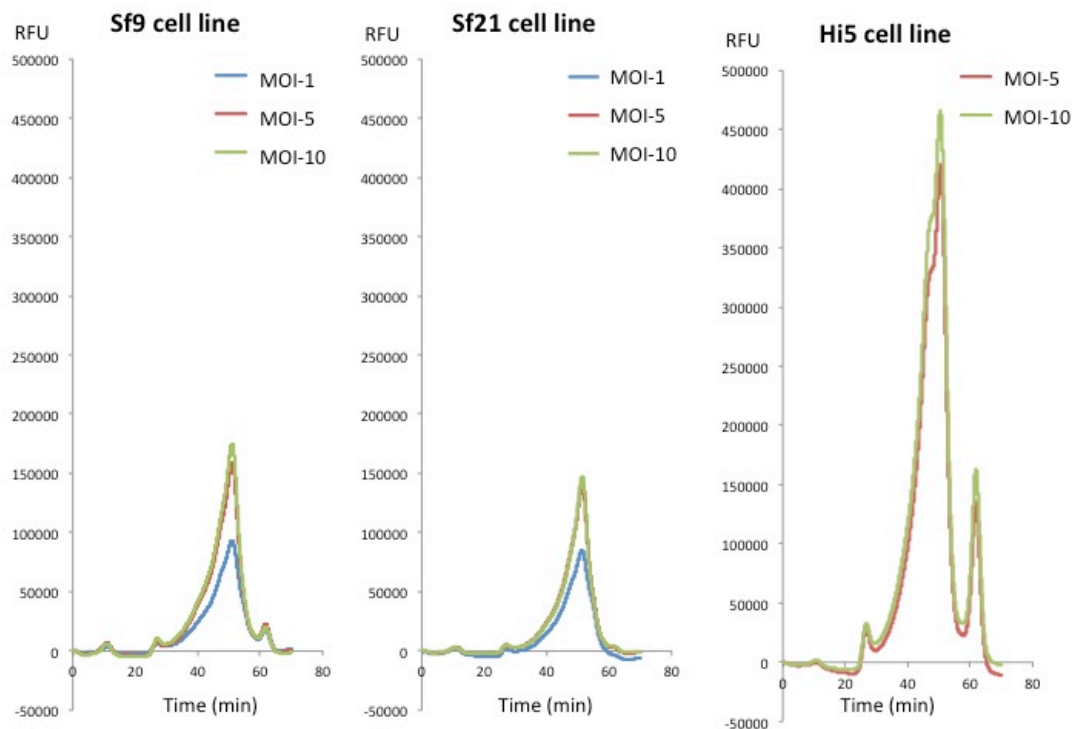


Figure 3.9 FSEC profiles of H3 expression optimization on cell lines. Comparison of crude membrane FSEC profiles between three cell lines: Sf9, Sf21 and Hi5.

Optimization on Cell lines

Fixed factors: Construct G4, P2 virus, TOI=1x10⁶ cells/mL, Time of harvesting=70 hours post infection

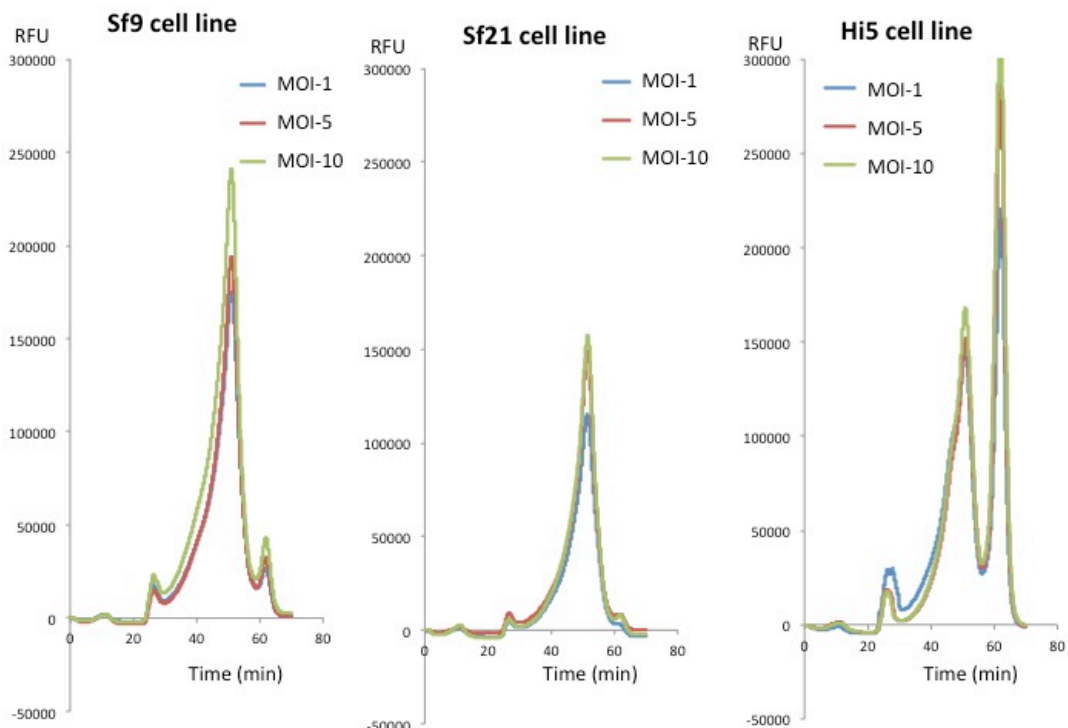


Figure 3.10 FSEC profiles of H3 expression optimization on cell lines. Comparison of crude membrane FSEC profiles between three cell lines: Sf9, Sf21 and Hi5.

3.3.4 Large-scale expression and one step IMAC purification of H3

A large-scale expression was carried out with the optimal expression strategy developed from the earlier studies. A total amount of 48g of cell pellet was obtained from 5L of cultures grown in a CellBag bioreactor. Purification of H3 was attempted by one step IMAC to obtain H3-T4L-GFP fusion protein. The initial purpose was to check the protein yield from large-scale expression, and use the H3-T4L-GFP fusion protein to check its binding profiles to some ligands, and to screen a few H3 antagonists for future purification and crystallization.

SDS-PAGE analysis (Fig. 3.11) shows that the IMAC elution containing a high yield of H3-T4L-GFP fusion protein, as well as large quantities of free GFP and other peptide fragments. A final quantity of 0.5 mg of H3-T4L-GFP fusion protein was obtained from 5L of cell cultures. Since this purification was a one-step IMAC purification, the poor purity of resultant protein was to be expected. However, the large fraction of free GFP also indicated that the H3 protein was not very stable through the purification steps, and further optimization of the purification was required, such as selecting the optimal detergent and lipid. Additionally, all the purification steps were carried out in the absence of antagonists, which potentially could be the main reason for the instability of the proteins. Cleavage of the GFP fusion with 3C protease was shown to occur with a high efficiency, and the cleavage product did not contain any fusion protein (Fig. 3.12).

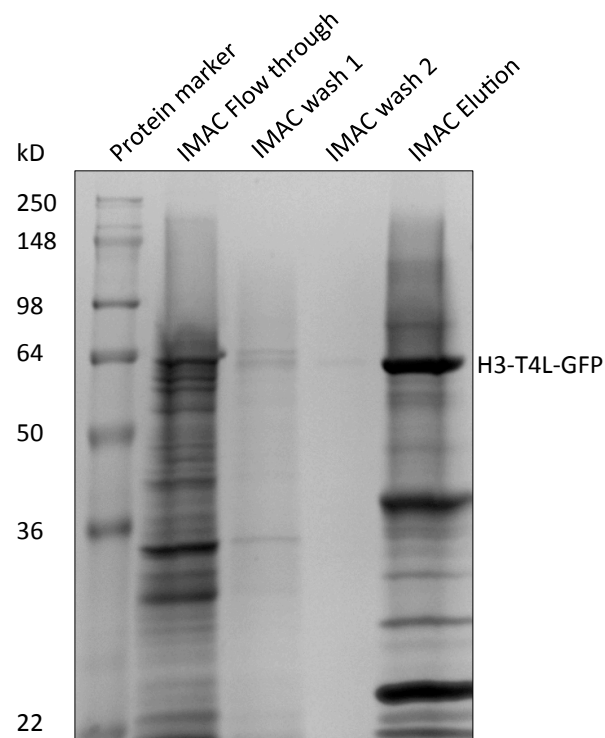


Figure 3.11 SDS-PAGE gel of H3-T4L-GFP fusion protein IMAC purification. Gel was visualized by Coomassie blue staining.

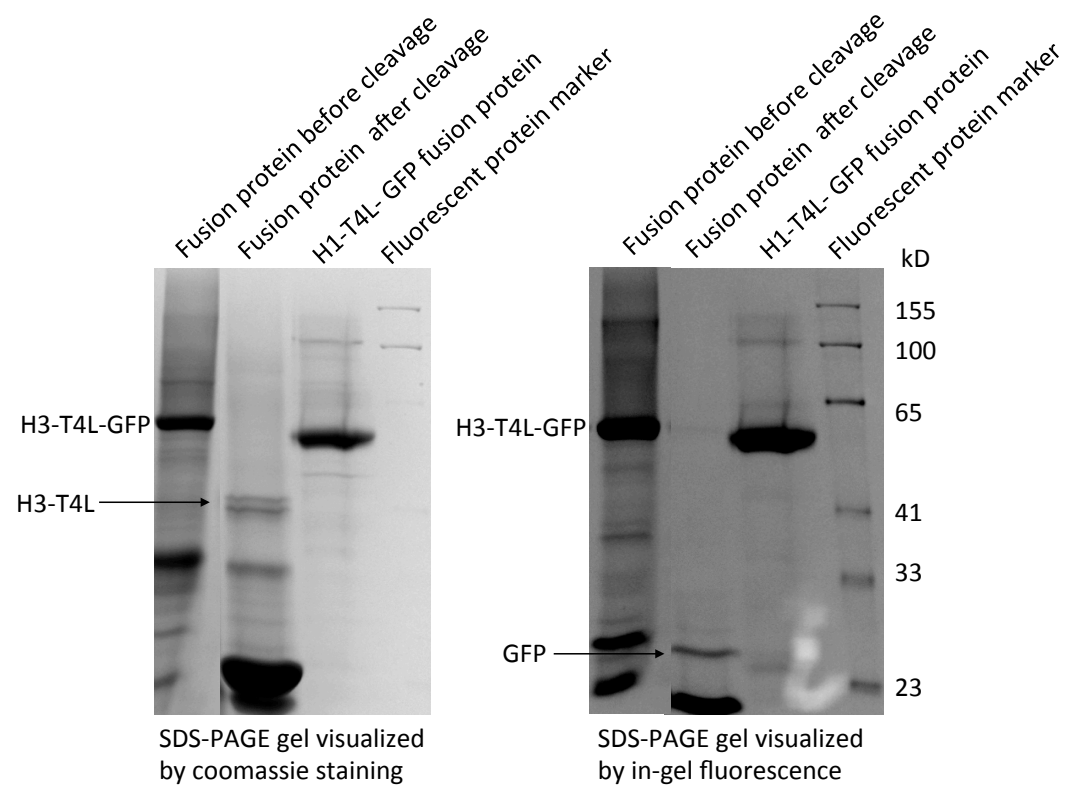


Figure 3.12 SDS-PAGE gel of H3-T4L-GFP fusion protein 3C protease cleavage test. (left) Visualized by Coomassie blue staining and (right) visualized by in-gel fluorescence

3.4 Discussions

Briefly, through a series of screenings, the turkey H3 constructs are generally more stable than their human H3 counterparts. One possible reason for this could be that bird species have higher core temperature, and potentially contain more thermostabilized genes. Sequence alignment of the human and turkey constructs (Appendix) shows that the two sequences have more than 90% identical residues at the inter-helical region, and have 100% identical residues at the conserved ligand contact positions. This demonstrates that the turkey H3 will be a good model for human H3 studies.

The best expression system (insect cell) for turkey H3 has been established, and an optimal expression regime identified. This expression system was tested many times in two different labs, showing this protocol is highly reproducible.

Protein yield from the large-scale expression (0.1mg/L) was much lower than estimated from small-scale expression tests (0.4mg/ml). Attention should be drawn to changing bioreactor types and culture volumes. The dynamics of infection and protein expression are different between small flasks and large CellBags. The time to infection onset and optimal expression to happen might be different as well. The health condition of cells is harder to control in a CellBag. For secreted proteins, lysed cells are not a problem, while for membrane proteins, the healthy state of cells and membranes at the time of harvesting is critical. Also, expression in large CellBags requires more stringent controls on the cells environment, such as, a temperature-controlled lab to mimic the incubator for small flasks to achieve optimal cell growth and protein expression. In spite of the reduced yield, H3-GFP fusion proteins have been successfully produced for functional studies and screenings of detergents and

antagonists, via techniques such as radioligand binding assays and Microscale thermophoresis. This expression protocol in insect cells could potentially lead to the production of milligram quantities of high quality H3 protein, for crystallization and structural determination.

Bibliography

- ADAMS, P. D., AFONINE, P. V., BUNKOCZI, G., CHEN, V. B., DAVIS, I. W., ECHOLS, N., HEADD, J. J., HUNG, L. W., KAPRAL, G. J., GROSSE-KUNSTLEVE, R. W., MCCOY, A. J., MORIARTY, N. W., OEFFNER, R., READ, R. J., RICHARDSON, D. C., RICHARDSON, J. S., TERWILLIGER, T. C. & ZWART, P. H. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*, 66, 213-21.
- ANDRELL, J. & TATE, C. G. 2013. Overexpression of membrane proteins in mammalian cells for structural studies. *Mol Membr Biol*, 30, 52-63.
- ASADA, H., UEMURA, T., YURUGI-KOBAYASHI, T., SHIROISHI, M., SHIMAMURA, T., TSUJIMOTO, H., ITO, K., SUGAWARA, T., NAKANE, T., NOMURA, N., MURATA, T., HAGA, T., IWATA, S. & KOBAYASHI, T. 2011. Evaluation of the *Pichia pastoris* expression system for the production of GPCRs for structural analysis. *Microb Cell Fact*, 10, 24.
- AXE, F. U., BEMBENEK, S. D. & SZALMA, S. 2006. Three-dimensional models of histamine H3 receptor antagonist complexes and their pharmacophore. *J Mol Graph Model*, 24, 456-64.
- BALLESTEROS, J. A. & WEINSTEIN, H. 1995. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci*, 25, 366-428.
- BERG, J. M., TYMOCZKO, J. L. & STRYER, L. 2007. *Biochemistry*, New York, W. H. Freeman.
- BERROW, N. S., ALDERTON, D. & OWENS, R. J. 2009. The precise engineering of expression vectors using high-throughput In-Fusion PCR cloning. *Methods Mol Biol*, 498, 75-90.
- BHATTACHARYA, S., HALL, S. E. & VAIDEHI, N. 2008. Agonist-induced conformational changes in bovine rhodopsin: insight into activation of G-protein-coupled receptors. *J Mol Biol*, 382, 539-55.
- BJARNADOTTIR, T. K., GLORIAM, D. E., HELLSTRAND, S. H., KRISTIANSSON, H., FREDRIKSSON, R. & SCHIOTH, H. B. 2006. Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse. *Genomics*, 88, 263-73.
- BLOBEL, G. 1980. Intracellular protein topogenesis. *Proc Natl Acad Sci U S A*, 77, 1496-500.
- BLUNDELL, T. L. 1996. Structure-based drug design. *Nature*, 384, 23-6.
- BOHM, A., GAUDET, R. & SIGLER, P. B. 1997. Structural aspects of heterotrimeric G-protein signaling. *Curr Opin Biotechnol*, 8, 480-7.

- BOUSQUET, J., VAN CAUWENBERGE, P., KHALTAEV, N., ARIA WORKSHOP, G. & WORLD HEALTH, O. 2001. Allergic rhinitis and its impact on asthma. *J Allergy Clin Immunol*, 108, S147-334.
- BRIGGS, J. & CAFFREY, M. 1994. The temperature-composition phase diagram and mesophase structure characterization of monopentadecenoin in water. *Biophys J*, 67, 1594-602.
- BROWN, R. E., STEVENS, D. R. & HAAS, H. L. 2001. The physiology of brain histamine. *Prog Neurobiol*, 63, 637-72.
- CAFFREY, M. 2003. Membrane protein crystallization. *J Struct Biol*, 142, 108-32.
- CAFFREY, M. & CHEREZOV, V. 2009. Crystallizing membrane proteins using lipidic mesophases. *Nat Protoc*, 4, 706-31.
- CAMELO-NUNES, I. C. 2006. New antihistamines: a critical view. *J Pediatr (Rio J)*, 82, S173-80.
- CHENG, Y. & PRUSOFF, W. H. 1973. Relationship between the inhibition constant (K₁) and the concentration of inhibitor which causes 50 per cent inhibition (I₅₀) of an enzymatic reaction. *Biochem Pharmacol*, 22, 3099-108.
- CHEREZOV, V., ROSENBAUM, D. M., HANSON, M. A., RASMUSSEN, S. G., THIAN, F. S., KOBILKA, T. S., CHOI, H. J., KUHN, P., WEIS, W. I., KOBILKA, B. K. & STEVENS, R. C. 2007. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science*, 318, 1258-65.
- CHESSARI, G. & WOODHEAD, A. J. 2009. From fragment to clinical candidate--a historical perspective. *Drug Discov Today*, 14, 668-75.
- CHIEN, E. Y., LIU, W., ZHAO, Q., KATRITCH, V., HAN, G. W., HANSON, M. A., SHI, L., NEWMAN, A. H., JAVITCH, J. A., CHEREZOV, V. & STEVENS, R. C. 2010. Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science*, 330, 1091-5.
- CHUN, E., THOMPSON, A. A., LIU, W., ROTH, C. B., GRIFFITH, M. T., KATRITCH, V., KUNKEN, J., XU, F., CHEREZOV, V., HANSON, M. A. & STEVENS, R. C. 2012. Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors. *Structure*, 20, 967-76.
- CHURCH, M. K., MAURER, M., SIMONS, F. E., BINDSLEV-JENSEN, C., VAN CAUWENBERGE, P., BOUSQUET, J., HOLGATE, S. T., ZUBERBIER, T., GLOBAL, A. & ASTHMA EUROPEAN, N. 2010. Risk of first-generation H(1)-antihistamines: a GA(2)LEN position paper. *Allergy*, 65, 459-66.
- COLLABORATIVE COMPUTATIONAL PROJECT, N. 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*, 50, 760-3.

- CREGG, J. M., VEDVICK, T. S. & RASCHKE, W. C. 1993. Recent advances in the expression of foreign genes in *Pichia pastoris*. *Biotechnology (N Y)*, 11, 905-10.
- DUBUSKE, L. M. 1999. Second-generation antihistamines: the risk of ventricular arrhythmias. *Clin Ther*, 21, 281-95.
- EGLIN, R. M., BOSSE, R. & REISINE, T. 2007. Emerging concepts of guanine nucleotide-binding protein-coupled receptor (GPCR) function and implications for high throughput screening. *Assay Drug Dev Technol*, 5, 425-51.
- EMANUEL, M. B. 1999. Histamine and the antiallergic antihistamines: a history of their discoveries. *Clin Exp Allergy*, 29 Suppl 3, 1-11; discussion 12.
- EMSLEY, P., LOHKAMP, B., SCOTT, W. G. & COWTAN, K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, 66, 486-501.
- EVANS, P. 2006. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr*, 62, 72-82.
- FAHAM, S. & BOWIE, J. U. 2002. Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure. *J Mol Biol*, 316, 1-6.
- FIGUEIREDO, A., RIBEIRO, C. A., GONCALO, M., ALMEIDA, L., POIARES-BAPTISTA, A. & TEIXEIRA, F. 1990. Mechanism of action of doxepin in the treatment of chronic urticaria. *Fundam Clin Pharmacol*, 4, 147-58.
- FOADI, J., ALLER, P., ALGUEL, Y., CAMERON, A., AXFORD, D., OWEN, R. L., ARMOUR, W., WATERMAN, D. G., IWATA, S. & EVANS, G. 2013. Clustering procedures for the optimal selection of data sets from multiple crystals in macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, 69, 1617-32.
- FREDRIKSSON, R., LAGERSTROM, M. C., LUNDIN, L. G. & SCHIOTH, H. B. 2003. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol*, 63, 1256-72.
- GRANIER, S., MANGLIK, A., KRUSE, A. C., KOBILKA, T. S., THIAN, F. S., WEIS, W. I. & KOBILKA, B. K. 2012. Structure of the delta-opioid receptor bound to naltrindole. *Nature*, 485, 400-4.
- HAGA, K., KRUSE, A. C., ASADA, H., YURUGI-KOBAYASHI, T., SHIROISHI, M., ZHANG, C., WEIS, W. I., OKADA, T., KOBILKA, B. K., HAGA, T. & KOBAYASHI, T. 2012. Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature*, 482, 547-51.
- HERMANS, E. 2003. Biochemical and pharmacological control of the multiplicity of coupling at G-protein-coupled receptors. *Pharmacol Ther*, 99, 25-44.
- HINO, T., ARAKAWA, T., IWANARI, H., YURUGI-KOBAYASHI, T., IKEDA-SUNO, C., NAKADA-NAKURA, Y., KUSANO-ARAI, O., WEYAND, S., SHIMAMURA, T., NOMURA, N., CAMERON, A. D., KOBAYASHI, T., HAMAKUBO, T.,

- IWATA, S. & MURATA, T. 2012. G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody. *Nature*, 482, 237-40.
- INSEL, P. A., TANG, C. M., HAHNTOW, I. & MICHEL, M. C. 2007. Impact of GPCRs in clinical medicine: monogenic diseases, genetic variants and drug targets. *Biochim Biophys Acta*, 1768, 994-1005.
- JANAKIRAMAN, V., FORREST, W. F. & SESHAGIRI, S. 2006. Estimation of baculovirus titer based on viable cell size. *Nat Protoc*, 1, 2271-6.
- KABSCH, W. 2010. Xds. *Acta Crystallogr D Biol Crystallogr*, 66, 125-32.
- KOLAKOWSKI, L. F., JR. 1994. GCRDb: a G-protein-coupled receptor database. *Receptors Channels*, 2, 1-7.
- KROEZE, W. K., SHEFFLER, D. J. & ROTH, B. L. 2003. G-protein-coupled receptors at a glance. *J Cell Sci*, 116, 4867-9.
- KRUSE, A. C., HU, J., PAN, A. C., ARLOW, D. H., ROSENBAUM, D. M., ROSEMOND, E., GREEN, H. F., LIU, T., CHAE, P. S., DROR, R. O., SHAW, D. E., WEIS, W. I., WESS, J. & KOBILKA, B. K. 2012. Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature*, 482, 552-6.
- KUDO, M., OSUGA, Y., KOBILKA, B. K. & HSUEH, A. J. 1996. Transmembrane regions V and VI of the human luteinizing hormone receptor are required for constitutive activation by a mutation in the third intracellular loop. *J Biol Chem*, 271, 22470-8.
- KURAMASU, A., SUKEGAWA, J., SATO, T., SAKURAI, E., WATANABE, T., YANAGISAWA, T. & YANAI, K. 2011. The hydrophobic amino acids in putative helix 8 in carboxy-terminus of histamine H(3) receptor are involved in receptor-G-protein coupling. *Cell Signal*, 23, 1843-9.
- LANDAU, E. M. & ROSENBUSCH, J. P. 1996. Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. *Proc Natl Acad Sci U S A*, 93, 14532-5.
- LEURS, R., BAKKER, R. A., TIMMERMAN, H. & DE ESCH, I. J. 2005. The histamine H3 receptor: from gene cloning to H3 receptor drugs. *Nat Rev Drug Discov*, 4, 107-20.
- LOVENBERG, T. W., ROLAND, B. L., WILSON, S. J., JIANG, X., PYATI, J., HUVAR, A., JACKSON, M. R. & ERLANDER, M. G. 1999. Cloning and functional expression of the human histamine H3 receptor. *Mol Pharmacol*, 55, 1101-7.
- LV, Z., CHU, Y. & WANG, Y. 2015. HIV protease inhibitors: a review of molecular selectivity and toxicity. *HIV AIDS (Auckl)*, 7, 95-104.
- MASSOTTE, D. 2003. G protein-coupled receptor overexpression with the baculovirus-insect cell system: a tool for structural and functional studies. *Biochim Biophys Acta*, 1610, 77-89.
- MOLLER, J. V. & LE MAIRE, M. 1993. Detergent binding as a measure of hydrophobic surface area of integral membrane proteins. *J Biol Chem*, 268, 18659-72.

- NEUBIG, R. R., SPEDDING, M., KENAKIN, T., CHRISTOPOULOS, A.,
INTERNATIONAL UNION OF PHARMACOLOGY COMMITTEE ON
RECEPTOR, N. & DRUG, C. 2003. International Union of Pharmacology
Committee on Receptor Nomenclature and Drug Classification. XXXVIII.
Update on terms and symbols in quantitative pharmacology. *Pharmacol
Rev*, 55, 597-606.
- NEWSTEAD, S., KIM, H., VON HEIJNE, G., IWATA, S. & DREW, D. 2007. High-
throughput fluorescent-based optimization of eukaryotic membrane
protein overexpression and purification in *Saccharomyces cerevisiae*.
Proc Natl Acad Sci U S A, 104, 13936-41.
- NOUBADE, R., MILLIGAN, G., ZACHARY, J. F., BLANKENHORN, E. P., DEL RIO, R.,
RINCON, M. & TEUSCHER, C. 2007. Histamine receptor H1 is required
for TCR-mediated p38 MAPK activation and optimal IFN-gamma
production in mice. *J Clin Invest*, 117, 3507-18.
- NYGAARD, R., FRIMURER, T. M., HOLST, B., ROSENKILDE, M. M. & SCHWARTZ,
T. W. 2009. Ligand binding and micro-switches in 7TM receptor
structures. *Trends Pharmacol Sci*, 30, 249-59.
- OSTERMEIER, C., IWATA, S., LUDWIG, B. & MICHEL, H. 1995. Fv fragment-
mediated crystallization of the membrane protein bacterial cytochrome
c oxidase. *Nat Struct Biol*, 2, 842-6.
- OSTERMEIER, C. & MICHEL, H. 1997. Crystallization of membrane proteins.
Curr Opin Struct Biol, 7, 697-701.
- OVERINGTON, J. P., AL-LAZIKANI, B. & HOPKINS, A. L. 2006. How many drug
targets are there? *Nat Rev Drug Discov*, 5, 993-6.
- PALCZEWSKI, K., KUMASAKA, T., HORI, T., BEHNKE, C. A., MOTOSHIMA, H.,
FOX, B. A., LE TRONG, I., TELLER, D. C., OKADA, T., STENKAMP, R. E.,
YAMAMOTO, M. & MIYANO, M. 2000. Crystal structure of rhodopsin: A
G protein-coupled receptor. *Science*, 289, 739-45.
- PARSONS, M. E. & GANELLIN, C. R. 2006. Histamine and its receptors. *Br J
Pharmacol*, 147 Suppl 1, S127-35.
- PIERCE, K. L., PREMONT, R. T. & LEFKOWITZ, R. J. 2002. Seven-transmembrane
receptors. *Nat Rev Mol Cell Biol*, 3, 639-50.
- QIU, H. & CAFFREY, M. 2000. The phase diagram of the monoolein/water
system: metastability and equilibrium aspects. *Biomaterials*, 21, 223-
34.
- RASMUSSEN, S. G., CHOI, H. J., ROSENBAUM, D. M., KOBILKA, T. S., THIAN, F.
S., EDWARDS, P. C., BURGHAMMER, M., RATNALA, V. R., SANISHVILI, R.,
FISCHETTI, R. F., SCHERTLER, G. F., WEIS, W. I. & KOBILKA, B. K. 2007.
Crystal structure of the human beta2 adrenergic G-protein-coupled
receptor. *Nature*, 450, 383-7.
- RASMUSSEN, S. G., DEVREE, B. T., ZOU, Y., KRUSE, A. C., CHUNG, K. Y.,
KOBILKA, T. S., THIAN, F. S., CHAE, P. S., PARDON, E., CALINSKI, D.,
MATHIESEN, J. M., SHAH, S. T., LYONS, J. A., CAFFREY, M., GELLMAN, S.

- H., STEYAERT, J., SKINIOTIS, G., WEIS, W. I., SUNAHARA, R. K. & KOBILKA, B. K. 2011. Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature*, 477, 549-55.
- RHODES, G. 1993. *Crystallography made crystal clear : a guide for users of macromolecular models*, San Diego, Academic Press.
- SCHIOTH, H. B. & FREDRIKSSON, R. 2005. The GRAFS classification system of G-protein coupled receptors in comparative perspective. *Gen Comp Endocrinol*, 142, 94-101.
- SCHWARTZ, T. W., FRIMURER, T. M., HOLST, B., ROSENKILDE, M. M. & ELLING, C. E. 2006. Molecular mechanism of 7TM receptor activation--a global toggle switch model. *Annu Rev Pharmacol Toxicol*, 46, 481-519.
- SHIMAMURA, T., SHIROISHI, M., WEYAND, S., TSUJIMOTO, H., WINTER, G., KATRITCH, V., ABAGYAN, R., CHEREZOV, V., LIU, W., HAN, G. W., KOBAYASHI, T., STEVENS, R. C. & IWATA, S. 2011. Structure of the human histamine H1 receptor complex with doxepin. *Nature*, 475, 65-70.
- SHIROISHI, M., KOBAYASHI, T., OGASAWARA, S., TSUJIMOTO, H., IKEDA-SUNO, C., IWATA, S. & SHIMAMURA, T. 2011. Production of the stable human histamine H(1) receptor in *Pichia pastoris* for structural determination. *Methods*, 55, 281-6.
- SHIROISHI, M., TSUJIMOTO, H., MAKYIO, H., ASADA, H., YURUGI-KOBAYASHI, T., SHIMAMURA, T., MURATA, T., NOMURA, N., HAGA, T., IWATA, S. & KOBAYASHI, T. 2012. Platform for the rapid construction and evaluation of GPCRs for crystallography in *Saccharomyces cerevisiae*. *Microb Cell Fact*, 11, 78.
- SIMONS, F. E. 2003. H1-Antihistamines: more relevant than ever in the treatment of allergic disorders. *J Allergy Clin Immunol*, 112, S42-52.
- SIMONS, F. E. & SIMONS, K. J. 2008. H1 antihistamines: current status and future directions. *World Allergy Organ J*, 1, 145-55.
- SMITH, B. S., KOBE, B., KURUMBAIL, R., BUCHANAN, S. K., VENKATRAMANI, L., VAN DER HELM, D. & DEISENHOFER, J. 1998. Crystallization and preliminary X-ray analysis of ferric enterobactin receptor FepA, an integral membrane protein from *Escherichia coli*. *Acta Crystallogr D Biol Crystallogr*, 54, 697-9.
- STARK, H., ARRANG, J. M., LIGNEAU, X., GARBARG, M., GANELLIN, C. R., SCHWARTZ, J. C. & SCHUNACK, W. 2001. The histamine H3 receptor and its ligands. *Prog Med Chem*, 38, 279-308.
- TAKEDA, K., SATO, H., HINO, T., KONO, M., FUKUDA, K., SAKURAI, I., OKADA, T. & KOUYAMA, T. 1998. A novel three-dimensional crystal of bacteriorhodopsin obtained by successive fusion of the vesicular assemblies. *J Mol Biol*, 283, 463-74.

- TAKEDA, S., KADOWAKI, S., HAGA, T., TAKAESU, H. & MITAKU, S. 2002. Identification of G protein-coupled receptor genes from the human genome sequence. *FEBS Lett*, 520, 97-101.
- TATE, C. G. 2012. A crystal clear solution for determining G-protein-coupled receptor structures. *Trends Biochem Sci*, 37, 343-52.
- TATE, C. G. & GRISSHAMMER, R. 1996. Heterologous expression of G-protein-coupled receptors. *Trends Biotechnol*, 14, 426-30.
- THOMPSON, A. A., LIU, W., CHUN, E., KATRITCH, V., WU, H., VARDY, E., HUANG, X. P., TRAPELLA, C., GUERRINI, R., CALO, G., ROTH, B. L., CHEREZOV, V. & STEVENS, R. C. 2012. Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. *Nature*, 485, 395-9.
- TRZASKOWSKI, B., LATEK, D., YUAN, S., GHOSHDASTIDER, U., DEBINSKI, A. & FILIPEK, S. 2012. Action of molecular switches in GPCRs--theoretical and experimental studies. *Curr Med Chem*, 19, 1090-109.
- VENKATAKRISHNAN, A. J., DEUPI, X., LEBON, G., TATE, C. G., SCHERTLER, G. F. & BABU, M. M. 2013. Molecular signatures of G-protein-coupled receptors. *Nature*, 494, 185-94.
- WACKER, D., FENALTI, G., BROWN, M. A., KATRITCH, V., ABAGYAN, R., CHEREZOV, V. & STEVENS, R. C. 2010. Conserved binding mode of human beta2 adrenergic receptor inverse agonists and antagonist revealed by X-ray crystallography. *J Am Chem Soc*, 132, 11443-5.
- WARNE, T., EDWARDS, P. C., LESLIE, A. G. & TATE, C. G. 2012. Crystal structures of a stabilized beta1-adrenoceptor bound to the biased agonists bucindolol and carvedilol. *Structure*, 20, 841-9.
- WHITE, S. *Membrane Proteins of Known 3D Structure* [Online]. Available: <http://blanco.biomol.uci.edu/mpstruc/> [Accessed 30 Aug 2015].
- WIENER, M. C. & WHITE, S. H. 1992. Structure of a fluid dioleoylphosphatidylcholine bilayer determined by joint refinement of x-ray and neutron diffraction data. III. Complete structure. *Biophys J*, 61, 434-47.
- WLODAWER, A. & VONDRASEK, J. 1998. Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. *Annu Rev Biophys Biomol Struct*, 27, 249-84.
- WORTH, C. L., KREUCHWIG, A., KLEINAU, G. & KRAUSE, G. 2011. GPCR-SSFE: a comprehensive database of G-protein-coupled receptor template predictions and homology models. *BMC Bioinformatics*, 12, 185.
- ZAMANIAN, M., KIMBER, M. J., MCVEIGH, P., CARLSON, S. A., MAULE, A. G. & DAY, T. A. 2011. The repertoire of G protein-coupled receptors in the human parasite *Schistosoma mansoni* and the model organism *Schmidtea mediterranea*. *BMC Genomics*, 12, 596.

5 Appendix

At pH7.5 H1-Pyramine									
Mutant	Kd (nM)	SD	SEM	pKi	SD	SEM	Bmax	SD	SEM
K179A	7.29	0.47	0.27	8.14	0.03	0.02	18.55	5.99	3.46
K191A	6.18	2.13	1.07	8.21	0.15	0.08	24.18	12.97	6.48
H450A	8.08	2.51	1.25	8.09	0.13	0.07	10.46	5.22	2.61
K+K	3.38	0.73	0.36	8.47	0.09	0.05	10.16	4.60	2.30
K+K+H	2.84	1.28	0.64	8.55	0.20	0.10	3.27	1.20	0.60
None	8.21	0.83	0.48	8.09	0.04	0.02	14.67	6.08	3.51

At pH7.5 H1-cetrizine									
Mutant	Ki (nM)	SD	SEM	pKi	SD	SEM		SD	SEM
K179A	103.2	50.3	29.0	7.02	0.22	0.13		41	24
K191A	731.0	521.3	301.0	6.21	0.32	0.19		17	10
H450A	158.2	72.5	41.8	6.84	0.22	0.13		95	55
K+K	640.9	284.5	164.2	6.22	0.19	0.11		56	40
K+K+H	744.8	437.3	309.3	6.17	0.27	0.19		157	91
None	140.5	25.8	14.9	6.86	0.08	0.05		16	9

At pH7.5 H1-fexofenadine									
	Ki (nM)	SD	SEM	pKi	SD	SEM		SD	SEM
	141	41		6.86				0.14	0.08
	84	17		7.08				0.09	0.05
	202	95		6.73				0.19	0.11
	160	56		6.81				0.15	0.11
	266	157		6.62				0.24	0.14
	130	16		6.89				0.05	0.03

At pH5.8 H1-Pyramine									
Mutant	Kd (nM)	SD	SEM	pKd	SD	SEM	Bmax	SD	SEM
K179A	3.00	2.12	1.22	8.60	0.34	0.19	18.80	6.72	3.88
K191A	3.15	1.43	0.83	8.54	0.24	0.14	18.88	2.26	1.31
H450A	2.26	1.34	0.77	8.69	0.24	0.14	9.67	0.83	0.48
K+K	2.77	1.09	0.63	8.58	0.16	0.09	14.53	3.45	1.99
K+K+H	1.64	0.71	0.41	8.81	0.17	0.10	4.73	2.71	1.56
None	4.22	2.48	1.43	8.43	0.28	0.16	19.20	8.35	4.82

At pH5.8 H1-cetrizine									
Mutant	Ki (nM)	SD	SEM	pKi	SD	SEM		SD	SEM
K179A	11.70	1.50	0.86	7.93	0.05	0.03		2.21	1.28
K191A	16.83	1.73	1.00	7.78	0.05	0.03		4.42	2.55
H450A	4.63	1.18	0.68	8.35	0.12	0.07		2.44	1.41
K+K	23.27	3.75	2.16	7.64	0.07	0.04		6.62	3.82
K+K+H	38.22	8.97	5.18	7.43	0.10	0.06		8.66	5.00
None	16.01	1.96	1.13	7.80	0.05	0.03		9.47	5.46

At pH5.8 H1-fexofenadine									
	Ki (nM)	SD	SEM	pKi	SD	SEM		SD	SEM
	33.28	2.21		7.43				0.07	0.04
	31.12	4.42		7.46				0.01	0.01
	17.07	2.44		7.74				0.04	0.02
	55.77	6.62		7.24				0.04	0.02
	35.89	8.66		7.28				0.10	0.06
	47.84	9.47		7.18				0.04	0.03

Figure 5.1 Results from radioligand binding assays

