

COMPUTATIONAL MODELLING OF
SUPRAMOLECULAR HUMAN AND ANIMAL
STRUCTURES: APPLICATIONS TO ENZYMES
RELEVANT IN COMPARATIVE PHYSIOLOGICAL
STUDIES.

CHIEMELA STEVE ODOEMELAM

A thesis submitted in partial fulfilment of the requirements of
Nottingham Trent University for the degree of Doctor of Philosophy

January 2023

Copyright statement

The copyright in this work is held by the author and supervisors. You may copy up to 5% of this work for private study, or personal, non-commercial research. Any re-use of the information contained within this document should be fully referenced, quoting the author, title, university, degree level and pagination. Queries or requests for any other use, or if a more substantial copy is required, should be directed to the author.

Acknowledgments

Firstly, I would like to thank Philippe. Words cannot express my profound gratitude for all the time you dedicated to supervising me. Very many thanks for your wisdom, guidance, and encouragement. Thank you for reading through my endless drafts and replying to my emails, which I sometimes send at ungodly hours. I am proud to have been his PhD Student.

To my supervisors, Dawn, and Sam, a big thank you for their considerable time, commitment, and support to this research. Elena, thank you for reading through my drafts and for endless support throughout this journey. Thanks to Sunday & Blessing for all your help and encouragement, for reading my drafts and for offering your advice. To all Wilson group members, thank you for your support throughout this journey. Group meetings and outings would not be the same without all of you.

Thank you to the people who have heard more than their fair share about GPCRs, cytochrome P450s, nuclear magnetic resonance and cassava root; my friends, I believe you all are experts in this subject area now. Thank you for all the phone calls, encouragement, food, plants, flowers, messages, cards, and the belief in me that I could do it, all of which have kept me going, so thank you to each of you.

I appreciate my family members' affection, concern, and contributions, especially my parents Professor & Professor Stevens. A. Odoemelam. Ezinne, thank you for some of the work on Excel. Papa (My Grandpa), thank you for constantly checking on this PhD and me.

Finally, I want to thank me. I want to thank me for believing in me. I want to thank me for doing all this hard work. I want to thank myself for having no days off. I want to thank myself for never quitting.

Abstract

Currently, there is a need for improved drugs to treat diseases such as diabetes and cancer. This study used different computational methods to investigate a range of scientific problems that require rapid solutions to improve health and wellbeing. Ligand docking, in silico ADMET, protein-ligand binding affinity, density functional theory (DFT), and analytical techniques were applied to study the allosteric binding pocket of the GLP-1 receptor, the interactions between flavonoids and cytochrome P450s, the binding of phorbol diesters to the CYP19A1 enzyme and nuclear magnetic resonance (NMR) shielding. Findings from the study on ligand binding to the glucagon-like peptide-1 receptor (GLP-1R) showed that the allosteric binding pocket of the GLP-1R is located near the transmembrane (TM) domain 6 of the receptor. This finding would enable the development of new allosteric modulators which can target the allosteric binding pocket of GLP-1R identified from this study. The next study which explored the binding of different ligands into cytochrome P450s showed that specific amino acid residues in the cytochrome P450s (CYP1A1 and CYP1B1) interact with different EROD flavonoids (Asp 313 and Phe 224 upon docking into CYP1A1 and Ala 330 upon docking into CYP1B1). The flavonoids isorhamnetin and pedalitin had the lowest binding energy upon docking into the crystal structures of 6DWM and 6IQ5. The results suggest the flavonoids isorhamnetin and pedalitin as potential precursors for natural product-derived therapeutics. Additionally, the study also explored the interactions between phorbol diesters and CYP19A1. The results showed that the phorbol diesters had higher binding energy than commercial aromatase inhibitors. This suggests that phorbol diesters have the potential to modify the activity of the aromatase enzyme. A study investigating the lack of rotational invariance of some density functional theory (DFT) grids on DFT computed NMR spectra of a low vibrational frequency reaction showed that the lack of rotational invariance in some DFT grids does not impact DFT-calculated NMR spectra of low vibrational frequency reactions.

Table of Contents

COPYRIGHT STATEMENT	2
ACKNOWLEDGMENTS	3
ABSTRACT	4
LIST OF PUBLICATIONS	7
LIST OF FIGURES	8
LIST OF TABLES	11
LIST OF ABBREVIATIONS	13
1. INTRODUCTION	17
1.2 NEED FOR THE RESEARCH	17
1.3 PROJECT AIM, HYPOTHESIS, AND OBJECTIVES	17
1.3.1 <i>Hypothesis</i>	18
1.3.2 <i>Objectives of the research</i>	18
1.4 THESIS OUTLINE	18
2. METHODS	20
2.1 DENSITY FUNCTIONAL THEORY (DFT)	20
2.1.1 <i>Basis sets</i>	20
2.2 MOLECULAR DOCKING	22
2.3 IN SILICO ADMET	23
3. G-PROTEIN COUPLED RECEPTORS: STRUCTURE AND FUNCTION IN DRUG DISCOVERY	24
3.1 BACKGROUND	24
3.2 FAMILY A (RHODOPSIN-LIKE RECEPTORS)	26
3.3 FAMILY B (SECRETIN RECEPTOR FAMILY)	29
3.4 FAMILY C (METABOTROPIC GLUTAMATE RECEPTORS)	32
3.5 STRUCTURAL DIFFERENCES	34
3.6 ALLOSTERIC BINDING AND COOPERATIVITY	35
3.7 GPCR SIGNALLING VIA G-PROTEINS	38
3.8 DIELECTRIC CONSTANT	38
3.9 COMPUTATIONAL BIOLOGY TECHNIQUES IN GPCR RESEARCH	40
3.10 CONCLUSION AND PROSPECTS	42
4. EXPLORING PROTEIN-LIGAND BINDING OF THE GLUCAGON-LIKE PEPTIDE 1 RECEPTOR.	43
4.1 INTRODUCTION	43
4.2 COMPUTATIONAL METHODS	46
4.2.1 <i>Selection of ligands and receptors</i>	46
4.2.2 <i>Ligand and Receptor Preparation, Molecular docking, and Binding affinity analysis</i>	48
4.2.3 <i>In silico ADME/Tox prediction</i>	49
4.3. IN SILICO LIGAND DOCKING APPROACHES TO CHARACTERISE THE BINDING OF KNOWN ALLOSTERIC MODULATORS TO THE GLUCAGON – LIKE PEPTIDE 1 RECEPTOR AND PREDICTION OF ADME/TOX PROPERTIES.	49
4.3.1 <i>Results</i>	49
4.3.2 <i>Discussion</i>	68
4.4 EVALUATION OF THE ALLOSTERIC BINDING SITE OF THE GLUCAGON-LIKE PEPTIDE RECEPTOR: LIGAND DOCKING AND ADMET PREDICTION STUDIES.	75
4.4.1 <i>Results and Discussion</i>	75
4.5 CONCLUSIONS AND FUTURE DIRECTIONS	89
5. CYTOCHROME P450 PROTEIN-LIGAND BINDING STUDY.	92

5.1	INTRODUCTION	92
5.2	COMPUTATIONAL METHODS.....	93
5.3	COMPUTATIONAL INVESTIGATION OF LIGAND BINDING OF FLAVONOIDS IN CYTOCHROME P450 RECEPTORS.....	94
5.3.1	RESULTS AND DISCUSSION	94
5.3.1.1	<i>Docking Study</i>	94
5.3.1.2	<i>Pilot ADMET studies</i>	111
5.4	THE ROLE OF PHORBOL-DIESTERS IN MEDIATING HUMAN PLACENTAL AROMATASE CYTOCHROME P450 ACTIVITY.....	118
5.4.1	<i>Results and Discussion</i>	119
5.5	CONCLUSION.....	125
6.	ON THE CALCULATION OF NMR CHEMICAL SHIFTS USING DENSITY FUNCTIONAL THEORY.....	128
6.1	INTRODUCTION	128
6.2	COMPUTATIONAL DETAILS	130
6.3	RESULTS AND DISCUSSIONS.....	131
6.4	CONCLUSIONS	134
7.	OVERALL CONCLUSIONS AND FUTURE WORK.....	136
7.1	CONCLUSIONS	136
7.2	FUTURE WORK	136
	REFERENCES.....	138
	APPENDIX.....	151
	CHAPTER 4 - SUPPLEMENTARY INFORMATION	151
	CHAPTER 9 - SUPPLEMENTARY INFORMATION	157

List of Publications

Some parts of this thesis are based on the following papers below.

1. G-Protein coupled receptors: structure and function in drug discovery – RSC Advances, 2020 10(60):36337-36348. [Thesis Chapter 2]
DOI: <https://doi.org/10.1039/D0RA08003A>
2. In silico ligand docking approaches to characterise the binding of known allosteric modulators to the glucagon-like peptide 1 receptor and prediction of ADME/Tox properties – MDPI Applied Biosciences, 2022 1(2):143-162. [Thesis Chapter 3]
DOI: <https://doi.org/10.3390/applbiosci1020010>
3. Computational investigation of ligand binding of flavonoids in cytochrome P450 receptors – Current Pharmaceutical Design, 2022 28(45), 3637-3648. [Thesis Chapter 5]
DOI: <https://doi.org/10.2174/1381612829666221121151713>
4. The Role of Phorbol-Diesters in mediating Human Placental Aromatase Cytochrome P450 Activity – MDPI Applied Biosciences, 2022 1(3), 279-288 [Thesis Chapter 6]
DOI: <https://doi.org/10.3390/applbiosci1030017>

List of Figures

- Figure 3.1. A schematic representation of a GPCR showing the transmembrane domains, N-terminus, C-terminus, the intracellular and extracellular loops.
- Figure 3.2. Schematic diagram showing the structure of family A GPCRs generated using ClustalW.
- Figure 3.3. Illustration showing the modification of rhodopsin and its orientation in membranes.
- Figure 3.4. Schematic diagram showing the structure of family B GPCRs generated using ClustalW.
- Figure 3.5. Graphical Illustration of family C GPCR structure.
- Figure 3.6. Mechanism of action of Allosteric Modulators.
- Figure 4.1.** A schematic representation of GLP-1R showing the transmembrane domains, the N- and C-termini, and the intracellular and extracellular loops generated using G-protein-coupled receptor.
- Figure 4.2.** Illustration showing the breakdown of three forms of GLP-1 (1 – 36 amide) by DPP – IV at different time intervals.
- Figure 4.3.** Chemical structures of known Glucagon-like peptide 1 receptor allosteric modulators.
- Figure 4.4.** Chemical structures of known Glucagon-like peptide 1 receptor allosteric modulators and an electrophilic compound.
- Figure 4.5.** Schematic showing ligand (2,6,7-trichloro-3-(trifluoromethyl) quinoxaline) – GLP-1R complexes. (A) 1:5VAI, (B) 1:5VEW, (C) 1:6B3J.
- Figure 4.6.** The protein-ligand interaction between 2,6,7-trichloro-3-(trifluoromethyl) quinoxaline and the GLP 1 receptor generated using BIOVIA Discovery Studio Visualiser. (A) 1:5VAI, (B) 1:5VEW, and (C) 1:6B3J.
- Figure 4.7.** Schematic showing ligand (1-(5-(4-(tert-butyl) phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one) - GLP-1R complexes.
- Figure 4.8.** The protein-ligand interaction between 1-(5-(4-(tert-butyl) phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one and the GLP 1 receptor generated using BIOVIA Discovery Studio Visualiser.

- Figure 4.9.** Schematic showing ligand (2-((4-chlorophenyl) thio)-3-(trifluoromethyl) quinoxaline) - GLP-1R complexes.
- Figure 4.10.** The protein-ligand interaction between 2-((4-chlorophenyl) thio)-3-(trifluoromethyl) quinoxaline and the GLP 1 receptor generated using BIOVIA Discovery Studio Visualiser.
- Figure 4.11.** Schematic showing 3-(8-chloro-6-(trifluoromethyl) imidazo[1,2-a] pyridin-2-yl) phenyl cyclohexanecarboxylate bound to the crystal structures of GLP-1R and the residues present in the active site.
- Figure 4.12.** The protein-ligand interaction between 3-(8-chloro-6-(trifluoromethyl) imidazo[1,2-a] pyridin-2-yl) phenyl cyclohexanecarboxylate and the GLP 1 receptor generated using Schrodinger Maestro.
- Figure 4.13.** Schematic image of the protein-2-((6,7-dichloro-3-(trifluoromethyl) quinoxalin-2-yl) thio)-5-methyl-1,3,4-thiadiazole interactions occurring at the GLP-1R binding site.
- Figure 4.14.** Schematic image protein-5-(6,7-dichloro-3-((1-(1-(1-methylpiperidin-4-yl) ethyl)-1H-tetrazol-5-yl) thio) quinoxalin-2-yl) thiazol-2-ol interactions occurring at the GLP-1R binding site.
- Figure 4.15.** Schematic showing representation of protein-1-([1,1'-biphenyl]-4-carbonyl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one interactions occurring at the GLP-1R binding site.
- Figure 4.16.** Schematic image protein-5-chloro-3-phenyl- [1,2,3] triazolo[1,5-a] quinazoline interactions occurring at the GLP-1R binding site.
- Figure 5.1.** Chemical structure of the dietary flavonoids used for the study.
- Figure 5.2A.** Schematic representation generated using Discovery studio visualiser showing the protein-ligand interactions occurring at the CYP1A1 (6DWM) binding site.
- Figure 5.2B.** Schematic representation generated using Discovery studio visualiser showing the protein-ligand interactions occurring at the CYP1A1 (6DWM) binding site.
- Figure 5.2C.** Schematic representation generated using Discovery studio visualiser showing the protein-ligand interactions occurring at the CYP1A1 (6DWM) binding site.

- Figure 5.2D.** Schematic representation generated using Discovery studio visualiser showing the protein-ligand interactions occurring at the CYP1A1 (6DWM) binding site.
- Figure 5.2E.** Schematic representation generated using Discovery studio visualiser showing the protein-ligand interactions occurring at the CYP1A1 (6DWM) binding site.
- Figure 5.3A.** Schematic representation generated using Discovery studio visualiser showing the protein-ligand interactions occurring at the CYP1B1 (6IQ5) binding site.
- Figure 5.3B.** Schematic representation generated using Discovery studio visualiser showing the protein-ligand interactions occurring at the CYP1B1 (6IQ5) binding site.
- Figure 5.3C.** Schematic representation generated using Discovery studio visualiser showing the protein-ligand interactions occurring at the CYP1B1 (6IQ5) binding site.
- Figure 5.3D.** Schematic representation generated using Discovery studio visualiser showing the protein-ligand interactions occurring at the CYP1B1 (6IQ5) binding site.
- Figure 5.3E.** Schematic representation generated using Discovery studio visualiser showing the protein-ligand interactions occurring at the CYP1B1 (6IQ5) binding site.
- Figure 5.4.** Schematic image the chemical structures of the phorbol diesters used in the study.
- Figure 5.5.** Schematic showing the protein ligand interactions (A) P-12,13-diAcPh (B) P-12,13-diiBu (C) P-12AcPh-13iBu.
- Figure 5.6.** Schematic showing the protein ligand interactions (A) P-12Ang-13iBu (B) P-20Ac-12AcPh-13iBu (C) P-20Ac-12Ang-13iBu.
- Figure 5.7.** Schematic showing the protein ligand interactions (A) Anastrozole (B) Letrozole (C) Exemestane.
- Figure 6.1.** Schematic showing 4,4'-dimethyl-1,1'-biphenyl rotated at different angles on the X cartesian axis. (A). 0 degrees (B). 20 degrees (C). 40 degrees (D). 70 degrees.

List of Tables

- Table 3.1.** Table showing some characteristics of family A – C GPCRs.
- Table 3.2.** Some rules for the definition of dielectric constants in proteins.
- Table 4.1.** Table showing amino acids interacting with 2,6,7-trichloro-3-(trifluoromethyl)quinoxaline in the GLP-1R structure.
- Table 4.2.** Amino acids which interact in the binding sites with 1-(5-(4-(tert-butyl)phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one.
- Table 4.3.** Amino acids present in the binding site of GLP-1R interacting with 2-((4-chlorophenyl)thio)-3-(trifluoromethyl) quinoxaline.
- Table 4.4.** Amino acids interacting with 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate in the GLP-1R binding pocket.
- Table 4.5.** Binding affinities of the docked complexes of GLP-1R and the ligands.
- Table 4.6.** Physicochemical properties of the selected allosteric modulators generated using ADMET Predictor 9.5.
- Table 4.7.** Effect of the selected allosteric modulators generated using ADMET Predictor 9.5 on transport proteins.
- Table 4.8.** Predicted metabolism of the selected allosteric modulators via cytochrome P450 isoforms.
- Table 4.9.** Toxicity parameters predicted using ADMET Predictor 9.5.
- Table 4.10.** Amino acids present in the 2-((6,7-dichloro-3-(trifluoromethyl)quinoxalin-2-yl)thio)-5-methyl-1,3,4-thiadiazole (Compound 1) binding site.
- Table 4.11.** Amino acids present in the 5-(6,7-dichloro-3-((1-(1-methylpiperidin-4-yl)ethyl)-1H-tetrazol-5-yl)thio)quinoxalin-2-yl)thiazol-2-ol binding site.
- Table 4.12.** Amino acids present in the 1-([1,1'-biphenyl]-4-carbonyl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one binding site.
- Table 4.13.** Amino acids present in the 5-chloro-3-phenyl-[1,2,3]triazolo[1,5-a]quinazoline binding site.
- Table 4.14.** Physicochemical properties of the selected allosteric modulators.
- Table 4.15.** *In silico* prediction of the pharmacokinetic properties of the ligands.
- Table 4.16.** Excretion and Toxicology parameters.

Table 5.1. The dG, rank score and virtual screening score of the ligands upon docking onto CYP1A1.

Table 5.2. The dG, rank score and virtual screening score of the ligands upon docking onto CYP1B1.

Table 5.3. Physicochemical parameters of the lead compounds.

Table 5.4. Pharmacokinetics of the lead compounds.

Table 5.5. Table showing the binding energy of the compounds and binding affinities of the docked complexes.

Table 6.1. Minimum and maximum relative NMR shielding values (ppm) of optimised 4,4'-dimethyl-1,1'-biphenyl for four pruned integration grids across different orientations (x axis rotations).

Table 6.2. Minimum and maximum relative NMR shielding values (ppm) of transition state 4,4'-dimethyl-1,1'-biphenyl for four pruned integration grids across different orientations (x axis rotations).

List of Abbreviations

A ₃ AR	A ₃ Adenosine Receptor
ADMET	Absorption, Distribution, Metabolism, And Excretion - Toxicity in Pharmacokinetics
AIMD	Ab Initio Molecular Dynamics
AO	Atomic Orbitals
B3LYP	Becke, 3-parameter, Lee–Yang–Parr
B97	Becke97
B97D	Grimme’s functional including dispersion
BBB	Blood-Brain Barrier
BCRP	Breast Cancer Resistance Protein
BSEP	Bile Salt Export Pump
C-terminal	Carboxyl-Terminal
CALCR, CALCRL	Calcitonin and Calcitonin-Like Receptors
CaS receptors or CASR	Ca ²⁺ Sensing Receptors
CRD	Cysteine Rich Domain
CRF1	Corticotropin-Releasing Factor Receptor 1
CRHR1, CRHR2	Corticotropin-Releasing Hormone Receptors
cryo-EM	Cryogenic Electron Microscopy
CYP 450	Cytochrome P450
DFT	Density Functional Theory
ECD	Extracellular N-Terminal Domain
EGFR	Epidermal Growth Factor Receptor
EIE	Equilibrium Isotope Effects
EL	Extracellular Loop
EPR	Electron Paramagnetic Resonance
ER+	Estrogen (Oestrogen) Receptor Positive
EROD	Ethoxyresorufin O-Deethylase Activity
ER α	Estrogen (Oestrogen) Receptor Alpha
GABA _B receptors	γ -Aminobutyric Acid B Receptors
GCGR	Glucagon Receptor
GDP	Guanine Diphosphate

GHRHR	Growth-Hormone-Releasing Hormone Receptor
GI Abs	Gastrointestinal Absorption
GIPR	Gastric Inhibitory Polypeptide Receptor
GLP1R, GLP2R	Glucagon-Like Peptide Receptors
GPCRs	G-Protein Coupled Receptors
GPRC6A	GPCR Class C Group 6 Member A
GRKs	G-Protein-Coupled Receptor Kinases
GTO	Gaussian-type Orbitals
GTP	Guanine Triphosphate
HA	Human Aromatase
HAS	Human Serum Albumin
HBA	Hydrogen Bond Acceptors
HBD	Hydrogen Bond Donors
HCTH/*	Handy's Family of Functionals Including Gradient-Corrected Correlation
HER2	Human Epidermal Growth Factor 2
hERG	Human Ether-à-go-go-Related Gene
HF	Hartree-Fock
IL	Intracellular Loop
KIE(s)	Kinetic Isotope Effect(s)
LBDD	Ligand-Based Drug Design
LCAO	Linear Combination of Atomic Orbitals
LCP	Lipidic Cubic Phase
Log K_p	Skin Permeation
Log S (ESOL)	Water Solubility
LogD	Octanol-Water Partition Coefficient as A Function of Ph
MD	Molecular Dynamics
mGlu Receptors (GRM)	Metabotropic Glutamate Receptors
MlogP	Octanol-Water Partition Coefficient Based on Moriguchi's Model
MM	Molecular Mechanics
MP2	Second Order Møller–Plesset Perturbation Theory
MP3	Third Order Møller–Plesset Perturbation Theory

MP4	Fourth Order Møller–Plesset Perturbation Theory
MPPT	Møller–Plesset Perturbation Theory
MRI	Magnetic Resonance Imaging
MW	Molecular Weight
NAO	Numerical Atomic Orbitals
OATP1B1	Organic Anion Transporting Polypeptide 1B1
OCT2	Organic Cation Transporter 2
P-gp	P-Glycoprotein
P.E.	Petroleum Ether
PAC1/ADCYAP1R1	Adenylate Cyclase Activating Polypeptide Receptor
PBE	The 1996 Functional of Perdew, Burke and Ernzerhof
PCM	Polarized Continuum Model
PD	Pharmacodynamics
P_{eff}	Predicted Human Jejunal Permeability
PK	Pharmacokinetics
PR	Progesterone Receptor
PTHr1, PTHr2	Parathyroid Hormone Receptors
QM	Quantum Mechanics
QM/MM	Quantum Mechanical/Molecular Mechanical
REs	Regulatory Elements
Ro5	Lipinski's Rule of 5
ROHF	Restricted Open-Shell Hartree-Fock
RRF	Rhodopsin Receptor Family
SBDD	Structure-Based Drug Design
SBVS	Structure Based Virtual Screening
SCTR	Secretin Receptor
SlogP	Octanol-Water Partition Coefficient Based on Simulation Plus's Model
SPSS	Statistical Package fFor Social Sciences
STO	Slater-Type Orbitals
SUSHI	Short Consensus Repeats
Sw	Aqueous Solubility in Pure Water
TAS1R1-3	Sweet And Umami Taste Receptors

TFs	Transcription Factors
TGF α	Transforming Growth Factor Alpha
TM	Transmembrane
TMD	Transmembrane Domain
TPSA	Topological Polar Surface Area
TPSS	The T-Dependent Gradient-Corrected Functional of Tao, Perdew, Staroverov, And Scuseria
TS	Transition State
UHF	Unrestricted Hartree-Fock
VD	Volume Of Distribution
VFT/VFTM	Venus Flytrap Module
VIPR1, VIPR2	Vasoactive Intestinal Peptide Receptors
VSXC	Van Voorhis And Scuseria's T-Dependent Gradient-Corrected Correlation Functional
XED	eXtended Electron Distribution
β_2 AR	B ₂ -Adrenergic Receptor

1. Introduction

The use of advanced computational techniques play an important role in the progression of science. Therefore, the concept of modelling a molecule, enzyme or receptor as well as mechanisms of these models and their limitations need to be understood.¹ An understanding of the underlying principles of computational methods would significantly improve the outcomes achieved by combining those and experimental techniques.

Methods such as molecular dynamics (MD), the hybrid quantum mechanics/molecular mechanics (QM/MM) approach, the quantum mechanics approach, and molecular docking are usually employed in understanding reactions that occur within enzyme catalysed reactions and receptor binding sites.^{2,3}

1.2 Need for the research

Computational techniques provide a fast, reliable, easier, and cost-effective way of solving chemical, biological or mathematical problems with the aid of computers. The application of computational techniques such as ligand docking, in silico ADME/Tox prediction, and QSAR in the drug discovery has been proven to be an effective strategy in accelerating and reducing the costs associated with drug discovery and development processes.⁴ Although, there are drugs available to treat ailments, there is the need to discover new drug targets as well as novel compounds which would target known binding sites and metabolic pathways. Another challenge considered in this research is the chronic cyanide exposure and lack of vitamin A leads to several dysfunctions in the body which led to the biofortification of cassava root. Therefore, the current research explores fundamentally different computational and experimental tools to provide insight into several areas of interest such as the allosteric binding site of GLP1R, cytochrome P450 and levels of residual cyanide and beta-carotene in cassava root that can also improve health and well-being through findings of the thesis.

1.3 Project aim, hypothesis, and objectives

The overall aim of this research is to use different experimental and computational techniques to provide insight into selected biological and chemical problems.

1.3.1 Hypothesis

The current study explores if the computational and experimental methods used in this research can be applied to answer selected chemical and biological questions.

1.3.2 Objectives of the research

The overall aim will be achieved via following objectives:

1. By using structure-based drug design techniques, allosteric binding pocket of glucagon-like peptide-1 receptor will be characterised. Subsequently residues interacting in a binding site would be identified. The characterisation of the binding pocket of selected cytochrome P450s (CYP1A1, CYP1B1, CYP19A1) and identification of the residues which interact in the binding site will be performed by using structure-based drug design tools.
2. The quantification of the physicochemical and toxicological parameters of selected molecules will be achieved by using in silico ADMET prediction software.
3. In order to understand the lack of rotational invariance on NMR spectra of selected reactions, DFT will be applied in the prediction of NMR shielding.

1.4 Thesis Outline

This thesis consists of 7 chapters. The first chapter introduces the research. The second chapter provides an overview of the methods used in the research. The next two chapters are focused on GPCRs starting with a review of literature on GPCRs including an overview of family A-C GPCRs, their structural differences, GPCR signalling, allosteric binding and cooperativity. The dielectric constant (relative permittivity) of proteins is also explored in the context of site-specific environmental effects. Furthermore, the second chapter of the thesis was published in RSC Advances (2020 10(60):36337-36348). Besides a literature review on GPCRs in chapter 3, Chapter 4 explores the allosteric binding site of the glucagon-like peptide-1 receptor (GLP-1R) (a family B GPCR) using structure-based drug design techniques such as ligand docking, in silico ADMET, protein-ligand binding affinity and allosteric modulators of GLP-1R. In addition, A part of Chapter 4 was published in MDPI Applied Bioscience (2022 1(2):143-162).

Using computational methods described in Chapter 2 (molecular docking and in silico ADME/Tox prediction), Chapter 5 provides information on interactions of selected

cytochrome P450s and 22 ligands. It elaborates on the interactions of selected ethoxyresorufin O-deethylase activity (EROD) flavonoids and the cytochrome P450s - CYP1A1 and CYP1B1. These two cytochromes play a role in xenobiotics and steroid hormone biosynthesis. Additionally, chapter 5 also explores the interactions between phorbols bearing different chains at C-12, C-13 and C-20 and known aromatase inhibitors anastrozole, exemestane and letrozole and the human placental aromatase cytochrome P450 (CYP19A1). Chapter 5 has been published Current Pharmaceutical Design, 2022 28(45), 3637-3648 and MDPI Applied Biosciences, 2022, 1(3), 279-288)

Chapters 6 of this thesis is focused on NMR. This chapter explores the effects of the lack of rotational invariance of some DFT integration grids on DFT computed NMR shielding. The last chapter of the thesis concludes the residues and provides suggestions for future work.

2. Methods

This chapter provides an introductory overview to the different methods used in this project.

2.1 Density Functional Theory (DFT)

In recent years, density functional theory (DFT) has become a standard tool in the computation of electronic structure in materials science, physics, and chemistry.⁵ DFT is also used to characterise biologically relevant molecular systems accurately.⁶ However, the application of DFT in biological systems is hindered primarily due to the unfavourable scaling of the computational effort with system size.⁵ DFT was birthed in an exceptional paper published in 1964 by Hohenberg & Kohn^{7, 8} and the principal implementation of their method was later published by the same authors in 1965.^{7, 9} DFT excels greatly in computational cost/performance ratio in comparison to the electron-correlated wave function-based methods such as Møller–Plesset perturbation theory (MPPT) or coupled cluster.¹⁰ The speed of DFT is also being utilised to perform various energy and gradient calculations for a system to study its evolution over a period of time; thus, making the DFT-based MD method called ab initio molecular dynamics (AIMD) gain increasing popularity and being applied in areas such as biomolecular, physical, chemical and material sciences.¹⁰

Research into the development of novel and accurate DFT functionals is active because some calculations require more specific and non-general functionals such as B97D¹¹ (Grimme’s functional including dispersion), VSXC¹² (van Voorhis and Scuseria’s τ -dependent gradient-corrected correlation functional) and HCTH/*¹³⁻¹⁵ (Handy’s family of functionals including gradient-corrected correlation).¹⁰ However, the development of ‘general-use’ functionals is ongoing.¹⁰ In the last 30 years, many semi-empirical DFT functionals (B3LYP,¹⁶ B97¹⁷ etc.) and non-empirical (PBE,^{18, 19} TPSS²⁰ etc.) DFT functionals have been developed by chemists and physicists.²¹ This has been influenced heavily by the efforts of John Perdew and Axel Becke in the development of non-empirical and semi-empirical density functionals.²¹

2.1.1 Basis sets

Basis sets mathematically describes the orbitals of a system being investigated; it is used for an approximate theoretical calculation or modelling.²² Basis sets are used in electronic structure calculations to specify the degrees of freedom allowed for the one-particle states.²³ A good basis set should strike a balance between accuracy, computational efficiency, and

flexibility. It should provide reliable and accurate results for a wide range of molecular systems and properties, while being computationally feasible for practical applications.

Majority of the all electron quantum chemical calculations currently reported in literature utilise basis sets created from atomic orbitals (AOs) through the linear combination of atomic orbitals (LCAO) approach.²³ There are three types of LCAO basis sets often used, namely, Slater-type orbitals (STOs),²⁴⁻²⁶ also known as Exponential type orbitals,²³ Gaussian-type orbitals (GTOs)²³ and Numerical atomic orbitals (NAOs).²³ Polarisation and diffusion functions are additional basis functions that are included to improve the accuracy of calculations, particularly for systems involving atoms with diffuse electron densities or those with significant polarization effects.²⁷⁻²⁹ Diffuse functions are basis functions that are designed to capture the behaviour of electrons that are spread out over a large region, away from the atomic nucleus.²⁸ Polarization functions, also known as polarization or higher angular momentum functions, are basis functions that describe the electron distribution in response to an electric field or in regions with high electron density gradients. The inclusion of polarisation and diffusion functions helps to better describe the electron density distribution and improves the accuracy of computed properties such as energies, molecular structures, and charge distributions for systems with diffuse electron densities or strong polarization effects. Basis sets have been broadly classified into the following:

1. Minimal basis sets: STO-3G,^{30, 31} STO-4G, STO-6G, STO-3G*. These basis sets are derived from a minimal Slater-type orbital basis set.²³
2. Pople basis sets: 3-21G,³²⁻³⁷ 3-21G*. 3-21+G, 3-21+G*, 6-31G, 6-31G*, 6-31+G*, 6-31G (3df, 3pd), 6-311G, 6-311G*, 6-311+G*.²² (* = polarisation function, + = diffusion function).
3. Correlation consistent basis sets: the basis sets in this group were developed by Dunning.³⁸ They were specifically designed for post-Hartree-Fock (HF) calculations.²² These basis sets comprise shells of polarisation (correlating) functions (d, f, g, etc.) which can lead to the convergence of the electronic energy to complete the basis set limit.²² Examples of these are cc-pVDZ (Double-zeta), cc-pVTZ (Triple-zeta), cc-pVQZ (Quadruple-zeta), cc-pV5Z (Quintuple-zeta) cc-pV6Z (Sextuple-zeta), aug-cc-pVDZ (Augmented versions of cc-pVDZ etc).³⁸⁻⁴²
4. Other split valence basis sets: these basis sets were initially developed by Ahlrichs and co-workers; these basis sets include SV, SVP, TZV, TZVP.^{43, 44} These basis sets, however, have been redefined by Weigend and Ahlrichs.^{45, 46} The redefined basis

sets include Def2SV, Def2SVP, Def2SVPP, Def2TZV, Def2TZVP, Def2TZVPP, Def2QZV, Def2QZVP, Def2QZVPP, and QZVP.

5. Plane-wave basis sets: plane-wave basis sets are used in quantum chemical simulations in addition to localised basis sets. A specified number of plane-wave functions are used for the calculation. This is usually below a set cut-off energy chosen for the calculation of interest.²²

2.2 Molecular Docking

Molecular docking is essential in computer-aided drug design and structural molecular biology.⁴⁷ The technique explores the behaviour of a small molecule/ligand in a binding site of the target protein.^{48, 49} The advances in determining protein structures using X-ray crystallography or NMR spectroscopy have made molecular docking an essential tool in drug discovery.⁴⁸ Docking software can perform a thorough search evaluating the conformation of the small molecule/ligand repeatedly until the minimum energy structure is reached. A scoring function (ΔG kcal/mol) is used to rank the docking poses as the sum of the electrostatic and van der Waals energies.⁴⁷⁻⁴⁹

Molecular docking has several applications in the process of drug discovery. Its applications include providing binding hypotheses to facilitate predictions for mutagenesis studies, virtual screening to identify potential leads, combinatorial library design, structure-activity relationship lead optimisation and chemical mechanism studies.^{47, 50} Molecular docking is currently applied in the food industry to understand food safety issues on biotoxins, drug residues and foodborne pathogens.⁵¹ The major protein of interest in the food industry is human serum albumin (HAS) and protease.⁵¹ Over 60 molecular docking tools have been reported in literature.⁵¹ However, the most commonly used software includes GOLD,⁵² FlexX,⁵³ DOCK/UCSF Dock,⁵⁴ AutoDock,⁵⁵ Affinity,⁵⁶ Surflex,⁵⁷ AutoDock Vina,⁵⁸ Cresset Flare,⁵⁹ Rosetta,⁶⁰⁻⁶³ rDock,⁶⁴ MOE-Dock,⁶⁵ etc.

There are three different types of molecular docking according to the degree of simplification, namely, flexible docking, semi-flexible docking and rigid docking.⁵¹ The rigid docking, as the name implies, means that the conformation of the protein of interest and ligand does not change.⁵¹ This type of docking does not require multiple calculations; hence being the most straightforward method.⁵¹ This type of molecular docking is best suited for large systems such as protein-protein,^{51, 66} and a protein-nucleic acid.^{51, 67} In the semi-flexible approach, the conformation of the receptor/protein of interest is fixed while the

ligand or small molecular conformation can be altered.^{48, 51} This method considers the influence of changes to a ligand structure.⁵¹ Semi-flexible docking is suitable for docking small molecules and macromolecules like nucleic acids or proteins and ligands.^{51, 68} The flexible docking approach allows the conformation of the ligand and receptor to be readily modified.⁵¹ Flexible docking approach is used in the accurate evaluation of molecular interactions.^{51, 69} The theory underlying molecular docking generally encompasses the following stages: target protein preparation, ligand preparation, scoring function implementation, search algorithm utilization (examples include Monte Carlo, genetic algorithm, simulated annealing, or Lamarckian genetic algorithm), scoring and ranking of poses, and subsequent validation and refinement steps.⁴⁷

2.3 *In silico* ADMET

The drug development and discovery processes are a complex and capital-intensive venture. These processes include disease selection, target identification and validation, lead discovery and optimisation, as well as preclinical and clinical trials.^{70, 71} The advances in *in silico* methods have led to several new drugs in the market.⁷⁰ Two of the major pitfalls in the drug discovery process are the lack of efficacy and safety.⁷⁰ This implies that the absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties of chemicals play vital roles in the individual stages of drug discovery and development.⁷⁰ An effective and safe drug should demonstrate an optimised combination of pharmacodynamics (PD) and pharmacokinetics (PK) parameters. This includes affinity, high potency, and selectivity against the molecular target, as well as an adequate ADMET profile.⁷² A variety of tools are available for the *in silico* ADMET assessment of compounds. These include QuikProp,⁷³ MetaSite,⁷⁴ DataWarrior,⁷⁵ StarDrop,⁷⁶ and MetaTox.^{72, 77} Furthermore, there are web resources used in the *in silico* ADMET prediction of compounds, namely, ADMETlab,⁷⁸ CypReact,⁷⁹ FAF-Drugs,^{80, 81} MetStabOn,⁸² SwissADME,⁸³ and vNN web server.⁸⁴

3. G-protein coupled receptors: structure and function in drug discovery

This chapter provides an introductory review detailing the computational advances in GPCR pharmacology and drug discovery. It provides an overview on family A-C GPCRs; their structural differences, GPCR signalling, allosteric binding and cooperativity. The dielectric constant (relative permittivity) of proteins is also discussed in the context of site-specific environmental effects. This chapter was published in RSC Advances 2020 10(60):36337-36348.

3.1 Background

The G-protein coupled receptor (GPCR) superfamily consists of structurally similar proteins arranged into families (classes), and is one of the most abundant protein classes in the mammalian genome.⁸⁵⁻⁸⁹ GPCRs undertake a plethora of essential physiological functions and are targets for numerous novel drugs.^{88, 89} Their ligands are structurally heterogeneous, including natural odorants, nucleotides, amines, peptides, proteins, and lipids.⁸⁸ The conserved structure of GPCRs consists of seven TMD of approximately 25–35 successive amino acid residues that express moderately high levels of hydrophobicity⁸⁸ and are characterised by α -helices which span the plasma membrane.⁸⁸ The primary function of GPCRs is the transduction of extracellular stimuli into intracellular signals.⁸⁶ Currently, approximately thirty to forty percent of marketed pharmaceuticals target GPCRs.^{85, 90-94} Hence, there is enormous potential for the development of new drugs targeting these receptors.⁸⁷ Examples of drugs targeting GPCRs include histamine receptor blockers, opioid agonists, β -blockers and angiotensin receptor blockers.⁸⁹ Computational biology methods are currently being employed to understand GPCRs as such drug targets.^{90, 95, 96} Breakthroughs in GPCR crystallography has facilitated novel discovery through virtual screening as well as better off-target rationalisation.⁹⁰ Recently, the Tikhonova group developed a computational protocol which combines concepts from statistical mechanics and cheminformatics to explore the flexibility of the bioamine receptors as well as to identify the geometrical and physicochemical properties which characterise the conformational space of the bioamine family.⁹⁷ Multiple-microsecond timescale molecular dynamics (MD) simulations have been used in capturing the process of several drugs binding to β_1 - and β_2 -adrenergic receptors.⁹⁸ Molecular docking is one of the most commonly used methods in GPCR structure-based drug design (SBDD).⁹⁸ Esguerra et al. developed GPCR-ModSim, a

web-based portal designed specifically for the homology modelling and MD simulation of GPCRs.⁹⁹

It was historically assumed that GPCRs exist in two conformations: active and inactive.¹⁰⁰⁻¹⁰² The long-established extended ternary-complex model of GPCR-driven signalling was based on this concept.^{100, 103, 104} This model suggested that the active GPCR conformation opted for by G-protein-coupled receptor kinases (GRKs), arrestins and G proteins is uniform.¹⁰⁰ Nevertheless, biophysical investigations with a refined fluorescent-labelled β_2 -adrenergic receptor (β_2 AR) demonstrated that a receptor can exist in numerous conformations and that the conformational equilibrium is influenced both by the bound ligand and the proximity to the related G protein.¹⁰⁰

The human genome alone contains approximately 800 GPCRs, making it the largest family of membrane proteins.^{89, 105} GPCRs have been classified based on structural and physiological features.⁸⁸ Some systems of classification have grouped these based on location of the ligand binding pocket, while some have utilised both the structural and physiological properties.^{88, 106} The A-F classification system was the first system of classification to be introduced.¹⁰⁷ This was first introduced in 1994 as A-F, and O for the (now obsolete) GCRDb database by Kolakowski.¹⁰⁷ The defunct GCRDb system was further developed, leading to the GPCRDB^{108, 109} database by Horn et al. with the rhodopsin family (Class A) being the largest and consisting of four main groups: α , β , γ , and δ , and 13 sub-branches.^{88, 107, 108} All GPCRs comprise of seven TMD helices (Figure 3.1), alongside an eight helix and a palmitoylated cysteine at the C terminal tail.¹¹⁰

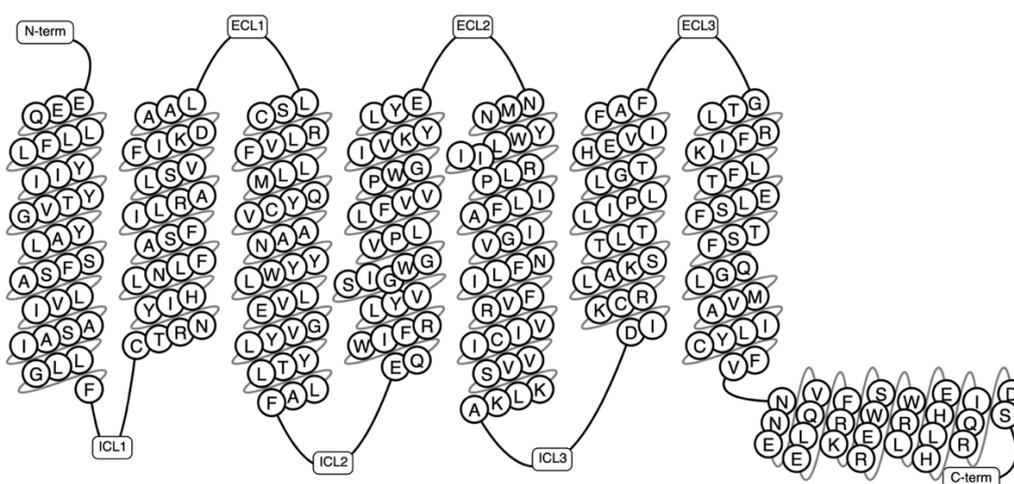


Figure 3.1. A schematic representation of a GPCR showing the transmembrane domains, N-terminus, C-terminus, the intracellular and extracellular loops (generated using GPCRDB Tools, <https://gpcrdb.org/>).¹¹¹

The diversity of GPCRs has resulted in a perceived difficulty in developing a comprehensive classification system.⁸⁹ The A-F system orders the GPCRs into six classifications on the basis of their sequence homology and functional similarity, namely: family A (Rhodopsin-like receptors), family B (Secretin receptor family), family C (Metabotropic glutamate receptors), family D (Parasitic mating pheromone receptors), family E (Cyclic AMP receptors) and family F (Frizzled and smoothed receptors).⁸⁹ Based on phylogenetic studies, human GPCRs have been classified under a system called “GRAFS”, and this system comprises of five main families namely; Glutamate (G), Rhodopsin (R), Adhesion (A), Frizzled/Taste2 (F), and Secretin (S).^{88, 105, 110} The major difference between the two systems concerns the additional division of family B into the Adhesion and Secretin families within GRAFS.¹¹⁰ This division was based on early findings describing a distinctive evolutionary history between both families.¹¹⁰ The review focused on Family A-C, emphasising their importance in drug discovery.⁹⁰ These families were chosen specifically because certain receptors within them can be found in mammals.¹¹²

3.2 Family A (rhodopsin-like receptors)

The Rhodopsin receptor family (RRF) is the largest of the GPCR families, comprising of approximately 680 members, and accounts for 80% of receptors in humans.^{88, 113} The RRF is classified into four groups (α , β , γ , δ) and 13 main subdivisions,^{88, 114} and it has numerous characteristics which indicate a common ancestry.^{88, 114} These characteristics include the DRY motif situated at the border between TM3 and intracellular loop (IL) 2 and NSxxNPxxY motif in TM7 (Figure 3.2).^{88, 114} The N-terminal region of the family A GPCR receptors are situated extracellularly,^{114, 115} while the C-terminal is located within the cytoplasm (Figure 3.3).^{114, 115} The ligand binding site is located within the extracellular region of the TMD bundle.¹¹⁴

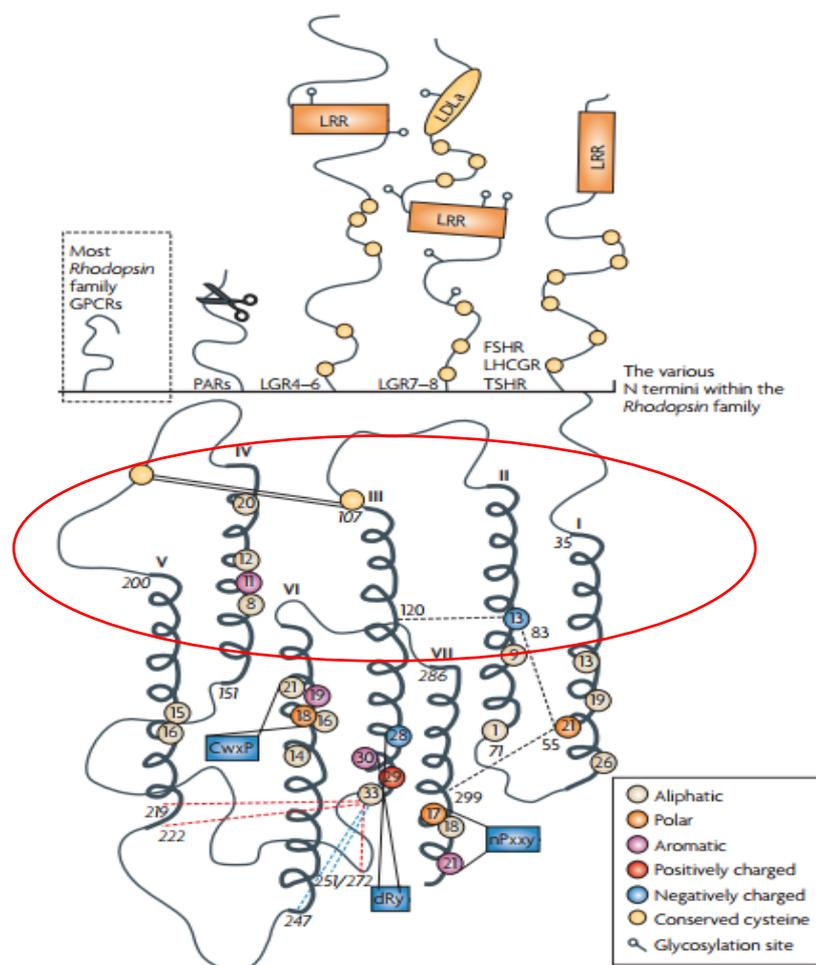


Figure 3.2. Schematic diagram showing the structure of family A GPCRs generated using ClustalW.¹¹⁶ Reprinted with permission from Springer Nature: Springer Nature, Nature Reviews Drug Discovery, Structural diversity of G protein-coupled receptors and significance for drug discovery, M. C. Lagerström and H. B. Schiöth, Copyright (2008). The upper section of Figure 2 shows the differences in the secondary structure of the N termini of the family A receptors.¹¹⁶ The scissor image indicates the cleavage site of the protease activated receptors whilst in the lower part of the image, the schematic TMD regions show the consensus of an alignment generated using ClustalW 1.82.¹¹⁶ In addition, the area circled in red describes the elliptical orientation.¹¹⁶ Residues conserved in all eight sequences are displayed as circles in which conserved aromatic residues are shown in purple, polar in orange, aliphatic residues are shown in beige, positively charged in red and negatively charged in blue.¹¹⁶

According to Palczewski, the arrangement of the seven TMD helices which vary in length from 20 to 30 residues is responsible for the overall elliptic, cylindrical shape of rhodopsin (Figure 3.3).¹¹⁵ The family A GPCRs vary greatly when their ligand preference and primary structure are considered.¹¹⁶ However, there is homogeneity in the N-termini of family A GPCRs, but heterogeneity within the TMD regions.¹¹⁶ However, some of the family A GPCRs share specific sequence motifs within the TMD region.¹¹⁶

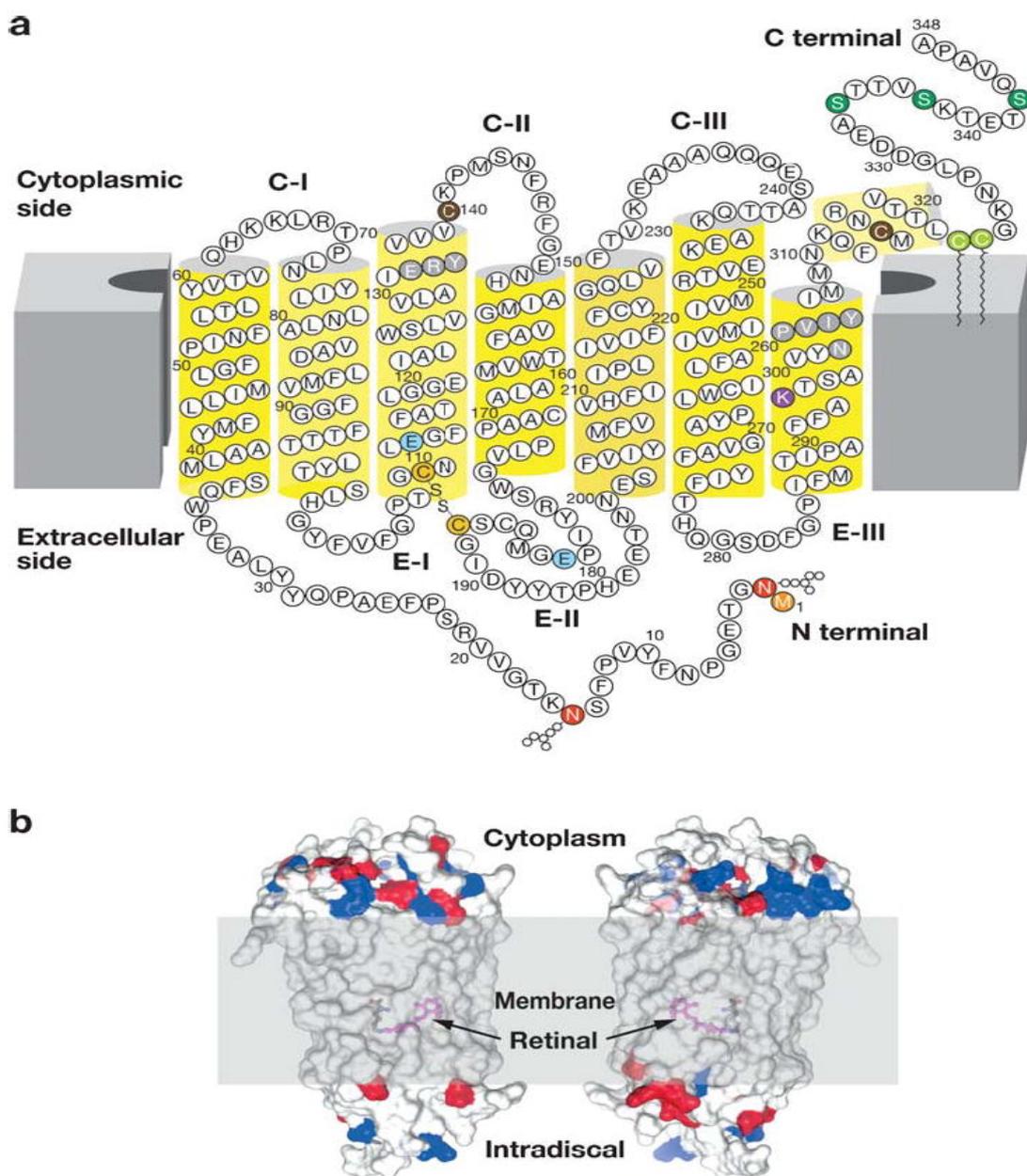


Figure 3.3. Illustration showing the modification of rhodopsin and its orientation in membranes.¹¹⁵ Reprinted with permission from Annual Reviews: Annual

Reviews, Annual review of biochemistry, G protein-coupled receptor rhodopsin, K. Palczewski, Copyright (2006). (a) Two-dimensional illustration of rhodopsin. The polypeptide of rhodopsin is seen to cross the membrane seven times with C-I, C-II, C-III comparable to the cytoplasmic loops and E-I, E-II, E-III to the extracellular loops. The yellow cylinders represent the transmembrane region (b) Depicts the location of the chromophore and the charges on the extracellular and cytoplasmic surface of rhodopsin. Red and blue colours represent negative and positive charged residues respectively, while the location of the chromophore is revealed by deleting fragments of the transmembrane helices.¹¹⁵

Palczewski reported the dimensions of rhodopsin as an ellipsoid of approximately 35x48x75 Å, with the long axis perpendicular to the membrane in the standard view.¹¹⁵ The surface area of the section protruding from the membrane is approximately 1200 Å², with cytoplasmic projection being larger in surface area and volume than the extracellular surface (Figure 3.3b).¹¹⁵ The TMD helices of rhodopsin are irregularly shaped due to the conformational changes associated with the Gly-Pro residues; they also incline at several angles in correspondence to the anticipated membrane surface.¹¹⁷ Teller et al. reported that helix 1 tilted from the membrane plane at 25° and contains a 12° kink within it as a result of Pro53 residues being present.¹¹⁷ Helix 2 kinked at an angle of 30° around Gly89 and Gly90 and the most significant bend being at Helix 6 at angle of 36° due to the presence of Pro267.¹¹⁷

3.3 Family B (secretin receptor family)

The family B GPCRs form a small group, and with an extracellular hormone-binding site, and they bind to large peptides.¹¹⁶ The family name “secretin” derives from the secretin receptor, which was the first to be cloned in this family.⁸⁷ In 1975, Sasaki et al.¹¹⁸ solved the first X-ray crystal structure of glucagon, a family B GPCR.¹¹⁹ The family corresponds to group B of the A-F system of classification,⁸⁷ and comprises 15 members including: vasoactive intestinal peptide receptors (VIPR1, VIPR2), glucagon-like peptide receptors (GLP1R, GLP2R), adenylate cyclase activating polypeptide receptor (PAC1/ADCYAP1R1), growth-hormone-releasing hormone receptor (GHRHR), calcitonin and calcitonin-like receptors (CALCR, CALCRL), gastric inhibitory polypeptide receptor

(GIPR), secretin receptor (SCTR), corticotropin-releasing hormone receptors (CRHR1, CRHR2), glucagon receptor (GCGR), and parathyroid hormone receptors (PTH1R, PTH2R).^{87, 116} These 15 receptors share between 21 and 67% sequence identity, and a large portion of the dissimilarity is identified in the N-terminal sequence.^{116, 120} These receptors contain conserved cysteine residues in the first and second extracellular loops of the TMD regions (Figure 3.4).¹¹⁶ However, the majority of the receptors within this family contain conserved cysteine residues that make up a cluster of cysteine bridges in the N-terminus¹¹⁶ The binding profile of the secretin receptors is outlined by three binding domains comprising of the proximal region and the juxta membrane region of the N-terminus, as well as the extracellular loops, together with TM6 (Figure 3.4).¹¹⁶ The ligand is thought to activate the receptor by spanning the N-terminal and the TMD extracellular loops, this way mediating the active conformation of the receptor, which increases the probability of activation of the signalling units.¹²¹

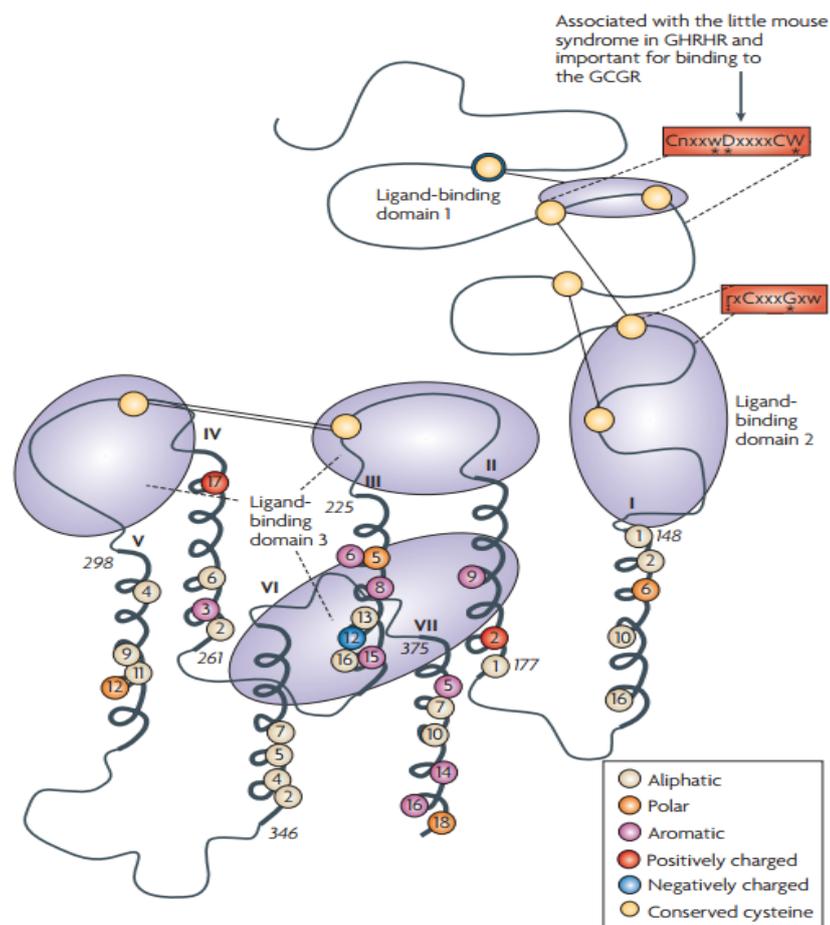


Figure 3.4. Schematic diagram showing the structure of family B GPCRs generated using ClustalW.¹¹⁶ Reprinted with permission from Springer Nature:

Springer Nature, Nature Reviews Drug Discovery, Structural diversity of G protein-coupled receptors and significance for drug discovery, M. C. Lagerström and H. B. Schiöth, Copyright (2008). The residues conserved in all 15 sequences are displayed as circles, the conserved polar residues are shown in orange, the aromatic residues in purple, the aliphatic residues in beige, the positively and negatively charged residues are shown in red and blue respectively.¹¹⁶ The uppercase letters show the completely conserved positions, the lowercase letters show the well-conserved positions (>50%) while the letter “x” show the variable positions. The conserved sequence motifs which are found in the TMD of the family B GPCRs are surrounded by red boxes.¹¹⁶ The conserved cysteine residues are depicted as yellow circles, the cysteine bridges between EL1 and EL2 are shown as two straight lines while the N-terminal cysteine bridges are drawn as lines.¹¹⁶

In addition to the presence of an extracellular N-terminal domain (ECD) of 120-160 residues, three intracellular (IL) and extracellular (EL) loops interconnect seven TMD (TM1-TM7) of 310-420 residues that are structurally similar and are thus members of the Family B GPCR.^{122, 123} According to Parthier et al. hormonal recognition in family B GPCRs is believed to follow the ‘two-domain’ binding mode, the N- and C-terminal regions of the peptides interact with the J- and N-domains of the receptors respectively, i.e. the C terminus of the peptide initiates a peptide recognition with the ECD, thus allowing the peptide N terminus to bind the TMD ligand-binding pocket activating the receptor and prompting a downstream signalling cascade.^{119, 123-125} The presence of a conserved ECD structure and the ‘two-domain’ binding mode across the family B GPCRs suggest a similar receptor activation across the GPCR family.¹²³

The secretin receptors have immense potential in drug discovery due to their importance in fundamental homeostatic functions.^{116, 123} To date, three of these hormones (glucagon, parathyroid hormone and calcitonin) are used clinically for the treatment of hypoglycaemia, osteoporosis and hypercalcaemia individually.¹¹⁶ Glucagon-like peptide-1 receptor (GLP1-R) and Glucagon-like peptide-2 receptor (GLP2-R) are particularly relevant targets, as a result of their part in appetite control and the treatment of type 2 diabetes.¹¹⁶

3.4 Family C (Metabotropic glutamate receptors)

The family C GPCRs comprise of the two γ -aminobutyric acid_B receptors (GABA_B receptors), odorant receptors in fish, eight metabotropic glutamate receptors (mGlu receptors or GRM), pheromone receptors, Ca²⁺-sensing receptors (CaS receptors or CASR), sweet and umami taste receptors (TAS1R1-3), GPCR Class C Group 6 Member A (GPRC6A) and seven orphan receptors.^{87, 88, 116, 126} The taste receptors in this GPCR family are targeted by the taste additives used in the food industry.¹²⁶ The CaS, mGlu and GABA_B receptors belong to a novel category of drug targets that are essential for considering conditions which affect the central nervous system and calcium homeostasis.¹²⁷ Currently, family C GPCRs are targeted by two therapeutic drugs in the market. One is Cinacalcet,¹²⁶⁻¹³⁰ the first GPCR allosteric modulator to be marketed, which targets the CaS receptor. The other is Baclofen, (now sold under the brand names Lioresal, Liofen, Gablofen, etc.) which is a GABA_B agonist used in the treatment of muscle spasticity.^{126, 127, 129-131}

The family C GPCRs differ from others by possessing a large extracellular domain, distal to the TMD receptors, and containing the orthosteric sites; they also form constitutive dimers with unique activation systems in comparison with other GPCR families.¹²⁶ Similarly, to their related families, family C GPCRs exhibit a typical motif of seven TMD helices however differ structurally from other GPCR families in their possession of an unusually large extracellular domain, an intracellular carboxyl-terminal (C-terminal) domain and a hepta-helical TMD (Figure 3.5a).¹²⁶ The family C GPCRs are structurally distinct from other GPCR families as a result of their extracellular domain including a cysteine rich domain (CRD, with the exception of GABA_B receptor) and Venus flytrap module (VFT).^{116, 126} The TM domain of family C GPCRs contain only the allosteric binding sites differing from other families with their TM domains conserved while the orthosteric sites are situated in the VFT module.^{126, 129} Domains present in the family C GPCRs provide numerous ligand sites of action, bar the intracellular C-terminal domain; this is highly variable and plays an essential role in signalling protein coupling and scaffolding.¹²⁶ The family C GPCRs are unique due to their compulsory dimerization, either as heterodimers (GABA_B receptor and TIRs) or homodimers (mGlu and CaS receptors) (Figure 3.5b).^{126, 132, 133}

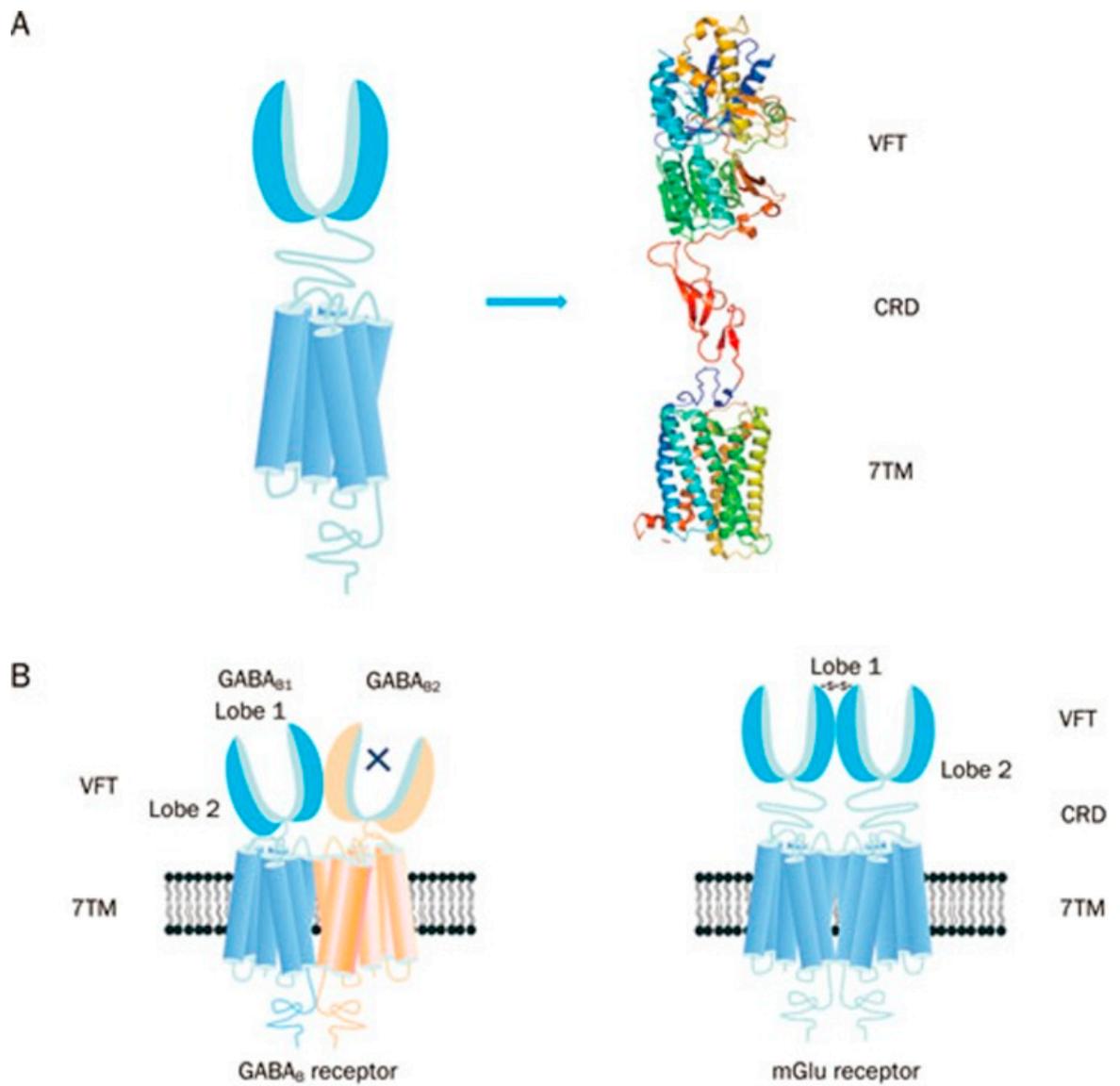


Figure 3.5. Graphical Illustration of family C GPCR structure.¹²⁶ Reprinted with permission from Springer Nature: Springer Nature, *Acta Pharmacologica Sinica*, Structure and ligand recognition of class C GPCRs, L. Chun, W.-h. Zhang and J.-f. Liu (2012). (a) Represents the structural organisation of family C GPCRs. Family C GPCRs have a peculiar structure which comprises of VFT with two lobes separated by an orthosteric binding pocket, a CRD and a TMD except for GABA_B receptor. (b) Graphical illustration of two members family C GPCRs; GABA_B receptor (heterodimer) and mGlu receptor (homodimer). There is a direct link between VFT and TMD in the GABA_B receptors and the two subunits, GABA_{B1} and GABA_{B2} make an obligatory heterodimer while the VFT connects to TMD using CRD in the

mGlu receptors. The mGlu receptors form homodimers which can potentially offer two other orthosteric binding pocket per dimer.¹²⁶

3.5 Structural Differences

GPCRs share a common structural characteristic, the TMD region, with its intracellular C-terminus and extracellular N-terminus, which exhibits the greatest homology.^{105, 113, 134, 135} The intracellular loops which span TM5 and 6, the amino terminus and the carboxyl terminus are among the most irregular structures in GPCRs with a substantial variation observed in the amino terminus (N-terminus).^{105, 136} The sequence is relatively short for peptide and monoamine receptors comprising of about 10-50 amino acids,^{105, 136} and larger for glutamate family receptors and glycoprotein hormone receptors (350-600 amino acids).^{105, 136} The largest amino terminal domains were observed in the adhesion family receptors.^{105, 136}

Bortolato *et al.* compared crystal structures of family B and family A GPCRs using receptors in the various classes (Glucagon receptors, Corticotropin-releasing factor receptor 1 (CRF₁) and Dopamine D₃ receptor).¹³⁷ The comparison of the CRF₁ and Glucagon receptor crystal structure to Dopamine D₃ receptor, a family A GPCR, showed that their cytoplasmic regions superimposed well.¹³⁷ However, the TM6 regions of both Glucagon receptors and CRF₁ extend outwardly while the cytoplasmic moieties are situated in proximity to the TM3 regions in sites similar to the dopamine, as well as other class A receptors.¹³⁷ The family B GPCRs lack the direct connectivity between TM3 and TM6 which is regarded as the classical ‘*ionic lock*’, playing an important role in family A GPCR activation.^{137, 138} The family C GPCRs structurally differ from family A and B due to their remarkably large extracellular domain which comprises of a cysteine-rich domain and VFT; an intracellular carboxyl-terminal (C-terminal) domain. The TMD regions in family A and B GPCRs are conserved however family C GPCRs have the allosteric binding site within the TMD region.¹²⁶ Table 3.1 shows some of the characteristics of the GPCR families discussed in this review.

Table 3.1. Table showing some characteristics of family A – C GPCRs.

Feature	Family A	Family B	Family C	Reference
Transmembrane domains	All families possess seven transmembrane domains			116, 139, 140
Orthosteric binding site	TM region	Extracellular loops, Extracellular N-terminus, TM6	Extracellular N-terminus (VFTM, SUSHI)	96, 126, 140, 141
Number of approved and marketed drugs	33	16	22	142
Motifs	All GPCRs share the D/E-R-Y/W motifs			115, 134, 139
Number of conserved residues in TMD regions	25	33	94	140
Type of ligand	Small molecules, proteins, peptides	Proteins, peptides	Small molecules, cations, amino acids	116
Suitable as drug targets?	Yes, except the sensory receptors	Yes	Yes, except the sensory receptors	116

TM: Transmembrane, GPCR(s): G-protein coupled receptor(s), VFTM: Venus Fly trap module, SUSHI: short consensus repeats.

3.6 Allosteric Binding and Cooperativity

Allostery is a widespread biological process, which is defined as the ability of interactions occurring at a particular site on a molecule to modulate actions on a different binding site on the same molecule.^{143, 144} For example, the binding of an allosteric modulator on a molecule allosterically changes the conformation of its binding pocket as shown in Figure 3.6. Currently, there are two types of marketed pharmaceuticals: allosteric modulators, which bind at the allosteric binding site on the receptor and allosterically change the structural conformation of the receptor binding site, and orthosteric modulators, which

bind at the active site of the receptor.¹⁴⁵ Orthosterically-binding drugs must overcome a major challenge in mediating the potential side effects arising from binding to homologous proteins sharing similar binding sites.¹⁴⁵ Hence an orthosterically-binding drug must have a very high affinity for its target, in order for a small dose to selectively achieve the goal of target-only binding.¹⁴⁵ The binding of transcription factors (TFs) to DNA regulatory elements (REs) provides a good example illustrating the specificity in orthosteric drugs.¹⁴⁵

The process of GPCR signalling initiates when an endogenous extracellular signal interacts with the orthosteric binding site of a GPCR, resulting in a conformational change which passes on the signal through the plasma membrane traversing the TMD region, and eventually activating intracellular signalling cascades through heterotrimeric G proteins and other adjunct proteins.^{143, 146, 147} A different approach, demonstrated for ligand-gated particle channels, is the advancement of allosteric modulators of the receptor subtypes, these small molecules do not bind to the traditional orthosteric binding site, instead interacting with the allosteric binding site to either enhance or inhibit receptor activation.¹⁴⁸

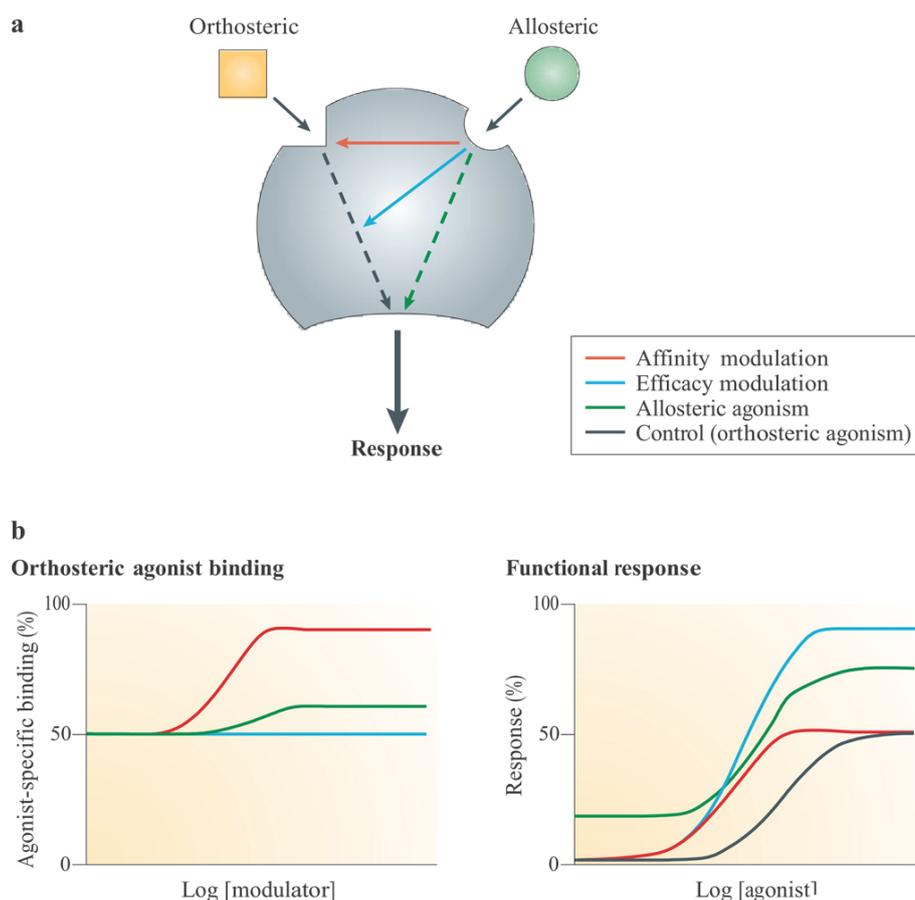


Figure 3.6. Mechanism of action of Allosteric Modulators.¹⁴⁸ Reprinted with permission from Springer Nature: Springer Nature, Nature Reviews Drug Discovery,

Allosteric modulators of GPCRs: a novel approach for the treatment of CNS disorders, P. J. Conn, A. Christopoulos and C. W. Lindsley, Copyright (2009). (a) Allosteric ligands bind to an alternative binding site on a receptor to modulate the activities of an orthosteric ligand efficacy (blue) and/or affinity (red). A number of allosteric ligands can also directly disrupt signalling in their own right (green) (b) Results from simulation show the effects on the function (right) or binding (left) of an orthosteric agonist mediated by three allosteric potentiators depicted in red, blue and green; red enhanced orthosteric agonist affinity only, blue enhanced only the efficacy, green was observed to modestly enhance both efficacy and affinity, as well as showing allosteric agonism.¹⁴⁸

Allosteric GPCR modulators show at least one of the outlined pharmacological properties (Figure 2.6). Agonism/reverse agonism: the allosteric modulator disrupts receptor signalling in either a positive (agonism) or negative (antagonism) manner, notwithstanding the presence or absence of an orthosteric ligand.¹⁴⁸ Efficacy modulation: the effect of allosterism causes changes in intracellular responses, leading to alterations in the inherent efficacy of an orthosteric ligand.¹⁴⁸ Affinity modulation: conformational change influences the orthosteric binding pocket, resulting in dissociation or association rate (or sometimes both) of the ligand being modified (Figure 2.6).¹⁴⁸ Some known allosteric modulators of family B GPCRs include Novo Nordisk compounds 1–6¹⁴⁸: T-0632, which blocks the GLP-1 induced cAMP production^{148, 149} (GLP 1 Receptor); DMP696, which blocks the CRF-stimulated adenylyl cyclase activity in cell line expressing CRF₁ receptor;^{148, 150} NBI 27914, which blocks the CRF₁ receptor;^{148, 151} NBI 35965;¹⁴⁸ Antarlamin¹⁴⁸ (CRF 1 Receptor).¹⁴⁸

Cooperativity is a thermodynamic term which has varying meanings in different biochemical contexts.^{152, 153} It is used to explain the complex interactions of identical ligands with a receptor at multiple binding sites.¹⁵² Cooperativity also describes the thermodynamics of macromolecular conformational transitions, which include nucleic acid helix-coil transitions and protein folding.¹⁵² Positive cooperativity is defined as the increase of binding affinity at one site of a receptor when a ligand is bound elsewhere.¹⁵⁴ A classic example of positive cooperativity is the binding of oxygen to haemoglobin; the binding of one oxygen molecule to the ferrous iron of the heme molecule increases the affinity of deoxyhaemoglobin for oxygen.¹⁵⁴ Negative cooperativity is observed when 2,3

bisphosphoglycerate binds to an allosteric binding site of haemoglobin and the affinity for oxygen is reduced.^{152, 154}

3.7 GPCR signalling via G-Proteins

G-proteins consist of several families of varied cellular proteins which perform several cellular functions, such as contractility and angiogenesis, learning and memory.^{155, 156} These proteins bind to the guanine nucleotides (guanine diphosphate (GDP) and guanine triphosphate (GTP) and also have inherent GTPase activity.¹⁵⁶ They play a principal role in a many cellular processes, including protein synthesis and cell development, vesicular transport, and cytoskeleton assembly, in addition to signal transduction.¹⁵⁶ G-proteins are trimers comprising of two functional components: a β -gamma dimer (35 and 8 kDa) which closely relates with the alpha subunit upon binding with GDP, and an alpha subunit (39 - 52 kDa) which is a catalyst for GTPase activity.¹⁵⁷ Human G proteins are classified into two classes, namely small (monomeric), and heterotrimeric G proteins.^{156, 157}

GPCRs are the largest superfamily of cell-surface receptors involved in TMD signalling, usually transmitting signals into cells via their response to a range of extracellular stimuli, such as glycoproteins, polypeptides and ions, and hence regulating a wide variety of physiological and developmental functions.¹⁵⁸ The intracellular-signalling cascades activated by GPCRs have been proven to be remarkably complex.^{158, 159} The binding of a ligand to the GPCR binding site leads to a conformational change in the receptor, in turn promoting the binding of the heterotrimeric G proteins, consisting of G_{α} -GDP and $G_{\beta\gamma}$ -subunits, within the intracellular moiety of the receptor.¹⁵⁹ The exchange of GTP for GDP on the G_{α} -subunit results in the reversible dissociation of the G protein subunits, initiating a downstream signalling via G_{α} -GTP and $G_{\beta\gamma}$.^{158, 159}

3.8 Dielectric constant

The most effective way of correlating the structure and function of macromolecules is through the examination of their electrostatic energies.¹⁶⁰ The intermolecular interactions present are affected by the effective dielectric constant (relative permittivity, ϵ_r),¹⁶¹ which differs according to the size and composition of the protein.¹⁶² The accuracy of the method of determination is important in understanding various biochemical interactions such as protein-ligand and protein-protein interactions, charge separation, ion channel selectivity and electron and proton transfer, signal transduction and macromolecular assembly;^{162, 163}

these interactions are influenced by the electrostatic potential of the protein surface.¹⁶²⁻¹⁶⁴ The dielectric constant of dry proteins ranges from 2.5 to 3.5 obtained from direct measurement.¹⁶³ The theoretical calculation of local dielectric constant of lone proteins based on their amino acid composition yielded an average of 2.7.¹⁶⁵ The polarity of the residues which make up the structural motifs within a protein have been shown to affect its dielectric constant values, and these findings were based on computational studies based on continuum electrostatics and molecular dynamics simulations.^{162, 163}

According to Warshel and Åqvist, the value of the dielectric constant of proteins is dependent on the property used to define it. They highlighted several possible ways of defining the dielectric constant in proteins, as outlined in Table 3.2,¹⁶⁶ where Q_1 and Q_2 are charges on ionisable groups separated by distance r , μ is a group dipole moment (in units of electron Ångström), ΔG is the electrostatic Gibbs free energy, \bar{a} is the effective radius of charge, and ϵ_B is the effective dielectric constant associated with a given interaction.

Table 3.2. Some rules for the definition of dielectric constants in proteins.

Definition	Value	Comments
Polar = ϵ large Nonpolar = ϵ small	$\epsilon = \text{large}$	Protein sites are always polar near small radii ions.
$\epsilon(r) = 332 \frac{Q_1 Q_2}{r \Delta G}$	$\epsilon(r) > 10$ often $\epsilon(r) \geq 40$	The value of ϵ is large for charge-charge interactions.
$1 - \frac{1}{\epsilon_B} = - \frac{\bar{a} \Delta G}{166 Q^2}$	$\epsilon_B > 10$	Proteins can provide as much solvation as water for ionised groups with small radii.
$\epsilon(r) = -332 \frac{Q_1 \mu_2 \cos \theta}{r^2 \Delta G}$	$\epsilon \geq 4$	For functionally important charge-dipole interactions, the value of ϵ could be as small as 4. Such a low value, however, requires relatively fixed dipoles with little energy for reorganisation.

Li *et al.* reported that the average dielectric constant inside a protein is relatively low, about 6-7, but this figure reaches about 20 – 30 on the surface of the protein.¹⁶⁷ The high

average local dielectric constant values are often linked to the charged residues while the low values are assigned automatically to the regions comprised of mostly hydrophobic residues.¹⁶⁷

According to Wilson et al. solvent effects on mechanisms of reactions have been established, but its effect on kinetic isotope effects (KIEs) are rather well less comprehended.³ A change in solvent can alter the KIE indirectly by changing the transition-state (TS) structure. It can also affect KIE by affecting isotopically sensitive vibrational frequencies directly, notwithstanding the TS structure or identity of the rate-determining step.³ Wilson et al. investigated the medium effects on KIE for S_N2 methyl transfer using UFF or UAO cavity method within the polarized continuum model (PCM) and a hybrid quantum mechanical/molecular mechanical (QM/MM) method.³ Their findings showed that the majority of variation in the equilibrium isotope effects (EIE) occur within the same range of dielectric constants ($1 \leq \epsilon \leq 10$) as is considered to occur with enzyme active sites and proteins.³ There is a possibility that any reaction which involves separation, neutralisation or charge distribution within an enzyme active site could indicate variations in KIEs, between a wildtype and mutant form of an enzyme, which originates as a result of changes in the local dielectric response within the diverse protein environment.³ The use of UFF or UAO cavity method within the polarized continuum model (PCM) and a hybrid QM/MM method to characterise ligand binding in GPCRs would further assist in understanding the interactions which occur in the both the active and inactive states of GPCRs, as well the changes which occur during the transition from inactive state to active state upon ligand activation.

3.9 Computational biology techniques in GPCR research

The first major breakthrough in human GPCR structural biology took place in 2007 as the solving of the β_2 -adrenergic receptor (β_2 AR with a diffusible ligand) using a modified lipidic cubic phase (LCP) produce to produce β_2 AR-TCL crystals which diffracted to a resolution of 2.2 Å, the structure was further refined at a 2.4 Å resolution.⁹⁶ Presently 64 structures of unique GPCRs with varying resolutions have been solved using spectroscopic methods such as fluorescence, electron paramagnetic resonance (EPR) and nuclear magnetic resonance (NMR) spectroscopy and structural techniques such as cryogenic electron microscopy (cryo-EM). This provides opportunities in employing computational biology techniques such as molecular modelling, and molecular docking in drug discovery

research.^{168, 169} The milestones achieved in GPCR structural studies have provided insights on the arrangements of the transmembrane domains,^{85-89, 95, 96} the location of the orthosteric,^{96, 116, 126} allosteric,^{96, 116, 126} bitopic,⁹⁶ as well as biased ligand binding sites⁹⁶, the homo- or hetero-oligomerization of receptors⁹⁶ and the structural rearrangements associated with conformational changes upon GPCR activation and inactivation.⁹⁶ This base of structural information on GPCRs is vital for SBDD,^{96, 170} ligand-based drug design (LBDD),⁹⁶ and integrated models which complement drug discovery efforts.⁹⁶

In 2012, Sosei Heptares published a detailed account on the use of A_{2A}R structure in identifying series of agents as potential antagonists, this became the first published GPCR SBDD discovery.¹⁷¹ In this drug discovery program, *in silico* screening of compound libraries was employed to identify diverse chemical scaffolds. Subsequently, structure-based drug design (SBDD) utilizing multiple X-ray structures of ligand-receptor complexes was employed to refine the clinical compound.¹⁷¹ The X-ray structures provided insights into an unexplored region within the orthosteric binding site, enabling the expansion of lead compounds into this binding site, and resulting in a more compact and efficient chemical series.¹⁷¹ Furthermore, optimization of selectivity for A_{2A}R over A_{1R} was achieved by targeting the same region.¹⁷¹ In a research carried out by de Graaf *et al.* using structure based virtual screening (SBVS), they identified allosteric modulators of two family B receptors namely; glucagon receptor and glucagon-like peptide receptor.¹⁷² SBDD approaches have also lead to the development of new agonists of the A₃ adenosine receptor (A₃AR).¹⁷³

Ballante *et al.*, (2020) investigated the possibility of identifying ligands for therapeutically relevant G-protein-coupled receptors through structure-based virtual screening of novel chemical compounds. They performed molecular docking screens using crystal structures of the A_{2A} adenosine and D₄ dopamine receptors and evaluated 53 molecules that ranked highest. Subsequently, experimental assessments were conducted, discovering two ligands for each receptor, with the most potent ligands demonstrating affinities in the sub-micromolar range. The analysis of the bioactivity data revealed that these ligands exhibited no significant activity at numerous off-targets, including several associated with adverse effects. Their findings showcase the efficacy of virtual screening as a valuable method for tapping into uncharted chemical territory, potentially contributing to the development of safer drug options.¹⁷⁴

Bassani *et al.* (2022) assessed the impact of sodium ions on the precision of pose prediction in docking GPCR antagonists. They investigated the effectiveness of three distinct docking programs in self-docking GPCR antagonists to address their research question. The

study's findings revealed that when the crystal structure used as the target includes resolved sodium ions, these ions must also be considered in the docking calculations. Nevertheless, if the crystallographic studies failed to resolve the sodium ion, manually inserting it into the virtual target would not provide any advantages.¹⁷⁵

Jaiteh *et al.* (2020) in their study aimed to improve rational drug design for G protein-coupled receptors (GPCRs) by utilising homology modelling and virtual screening techniques. Homology models of the D₂ dopamine and serotonin 5-HT_{2A} receptors were generated based on crystal structures of various GPCRs. The accuracy of these models was assessed by comparing them to the crystal structures of the D₂ dopamine and 5-HT_{2A} receptors. Their research findings showed that accurate predictions could be achieved, although not always with the most closely related template. Virtual screening performance was evaluated through molecular docking, indicating that multiple models based on different templates should be considered to identify the optimal binding site structure. Models based on aminergic GPCRs exhibited significant ligand enrichment and improved virtual screening performance. These findings provide valuable guidelines for successful structure-based ligand discovery using GPCR homology models.¹⁶⁸

3.10 Conclusion and prospects

GPCRs are multifaceted proteins which exist in varying conformations, and that the conformational equilibrium of this group of receptors is influenced both by the bound ligand and the proximity to the related G protein. Their structure is highly conserved comprising of seven TMD. These receptors possess different binding domains, namely, allosteric and orthosteric binding domains. The progress in GPCR structural biology has substantially accelerated our understanding of GPCRs as potential drug targets using SBDD and LBDD approaches. Further computational studies assessing nuclear quantum effects on ligand receptor binding, as well as hybrid QM/MM and empirical valence bond theory in the mechanistic studies of GPCRs would allow for further insight into the interactions which occur in both the active and inactive states of GPCRs, as well the changes which occur during the transition from these states upon ligand activation. This chapter has aimed to provide an accessible and introductory perspective on advances in GPCR-based drug discovery approaches; many reviews on the topic highlighted herein are indeed highly detailed and authoritative but may not provide as accessible an account for a less specialised or more general audience in the chemical sciences.

4. Exploring Protein-Ligand Binding of the Glucagon-Like Peptide 1 Receptor.

Following a preliminary review of literature exploring GPCR structure and function in drug discovery. This chapter explores the allosteric binding site of the glucagon-like peptide-1 receptor (GLP-1R) (a family B GPCR) using structure-based drug design techniques such as ligand docking, in silico ADMET, protein-ligand binding affinity (which were introduced in chapter 1 of this thesis) and allosteric modulators of GLP-1R. Some parts of chapter 4 has been published by *MDPI Applied Bioscience* 2022 1(2):143-162.

4.1 Introduction

Glucagon-like peptide1 receptor (GLP-1R) is of particular interest due to its role in the treatment of type 2 diabetes mellitus (T2D) and appetite regulation.¹¹⁶ This receptor belongs to the small family (or class) B of G-protein-coupled receptors (GPCRs), which consist structurally of seven transmembrane (TM) domains of 310-420 residues, interconnected by three intracellular (IL) and extracellular (EL) loops, and an extracellular N-terminal domain (NTD) of 120-160 residues (Figure 4.1).^{122, 123}

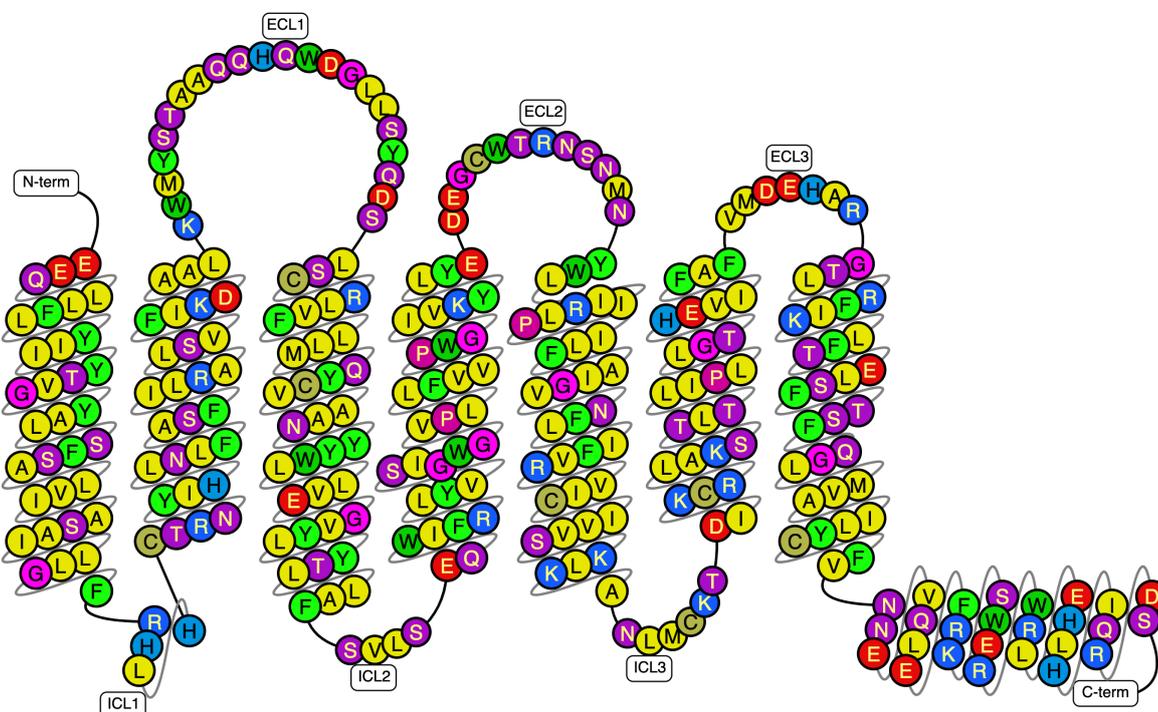


Figure 4.1. A schematic representation of GLP-1R showing the transmembrane domains, the N- and C-termini, and the intracellular and extracellular loops generated using G-protein-coupled receptor (GPCR) database tools.¹¹¹

Glucagon-like peptide 1 (GLP-1) is the endogenous ligand of GLP-1R. In response to food intake, it is produced from the gastrointestinal tract and has vital roles in regulating insulin secretion, appetite control, and carbohydrate metabolism.¹⁷⁶ As a result of the distinctive processing of its precursor glucagon, GLP-1 occurs in two active forms: GLP-1 (7 – 37 amide) and GLP-1 (7 – 36 amide). GLP-1 (7 – 36 amide) is the primary circulatory form, which exerts insulinotropic and glucoregulatory functions. However, within 1 to 4 minutes of secretion, both active forms are degraded by dipeptidyl peptidase 4 (DPP-IV) through cleavage of NH₂-terminal amino acids to their respective principal metabolites; 9 – 37 amide and 9 – 36 amide (Figure 4.2), of which each remains in circulation for about 30 minutes.¹⁷⁷ Contrary to the antecedent belief that GLP-1 (9 – 36 amide) is pharmacologically inactive as a result of its weak or no insulinotropic activity, recent findings have demonstrated that GLP-1 (9 – 36 amide) possess distinctive extra-pancreatic insulin-like actions in the heart, liver and vasculature, which are autonomously mediated irrespective of the GLP-1R.¹⁷⁸⁻¹⁸¹ GLP-1 (7-36 amide) is a 30-amino-acid peptide hormone released from intestinal L-cells following supplement ingestion.¹⁸² The peptide GLP-1 has numerous functions, including potentiation of the glucose-actuated release of insulin from pancreatic beta cells, heightening insulin articulations, obstruction of beta-cell apoptosis, progression of beta-cell neogenesis, diminishing glucagon emission, conceding gastric discharging, supporting satiety, and intensifying peripheral glucose disposal. It is also a physiological regulator of appetite and food intake.¹⁸³

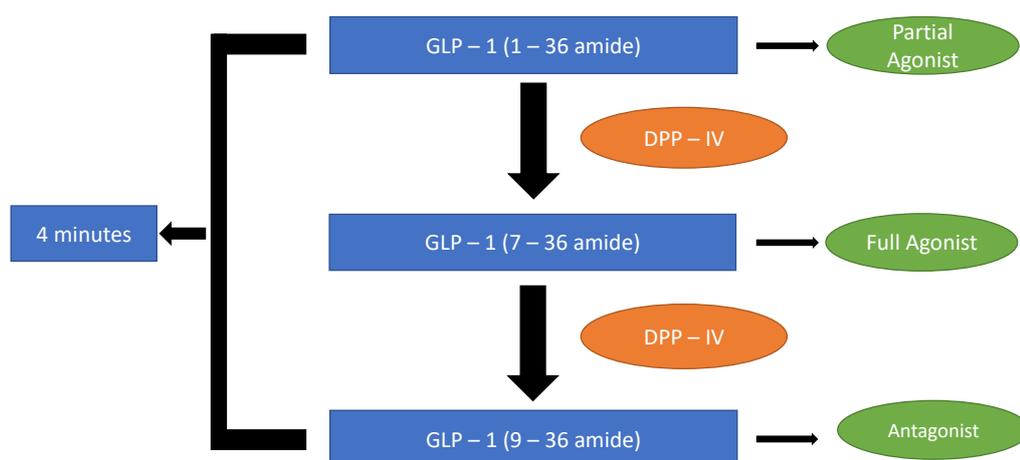


Figure 4.2. Illustration showing the breakdown of three forms of GLP-1 (1 – 36 amide) by DPP – IV at different time intervals. DPP – IV is depicted in Orange and GLP-1 in blue.

The green ovals show the forms of GLP-1 and their characteristic when degraded by DPP-IV. (It shows if that form is an agonist/antagonist to the GLP-1R).

Fasting blood concentration of GLP-1 (9 – 36 amide) in humans usually ranges from 5 to 15 pmol/L, and there is usually a two- to four-fold increase after ingesting food. There is an increase in blood GLP-1 concentration 15 minutes after food ingestion, and peak concentration is reached after 60 minutes. The blood GLP-1 concentration decreases gradually in the second hour until the next time food is ingested.¹⁷⁷ It is clear from these varied activities that GLP-1 plays a central role in controlling postprandial glucose levels and, in that capacity, drugs that stimulate the GLP-1 receptor, such as dipeptidyl peptidase 4 inhibitors or GLP-1 analogues, have been manufactured for use in the treatment of type 2 diabetes mellitus (T2D).^{184, 185} The excessive secretion of GLP-1 has been hypothesised to be responsible for postprandial reactive hypoglycaemia, while diminished secretion might lead to obesity.¹⁸² T2D treatment needs the positive allosteric modulation of GLP-1R to inhibit glucagon secretion, thus stimulating insulin secretion in a glucose-dependent routine.¹⁸⁶ Findings from previous studies have proposed a two-step, two-domain mechanism of receptor activation.^{123, 186, 187} According to this ligand-binding mechanism, the C terminus of the peptide ligand forges a complex with the NTD (1); thus, allowing the N terminus of the peptide ligand to interact with the 7TM domain and activate the family B GPCR to couple G proteins and other effectors to mediate intracellular signalling processes (2).¹⁸⁶⁻¹⁸⁹ According to Zhang et al.¹⁸⁷, the mechanism by which the ligand interacts with the 7TM domain and subsequently activates the receptor is unknown.

A structural study carried out by Song et al.¹⁸⁶ showed similarity in the transmembrane domain (TMD) architecture of GLP-1R and Glucagon receptor (GCGR), which was also consistent with the overlap in their primary sequences (45% similar in their TMDs). The human GLP-1R TMD was crystallised with two negative allosteric modulators, NNC0640 and PF-06372222, respectively, at 3.0 and 2.7Å resolution. The crystallised structures of GLP-1R and GCGR showed a common binding pocket for the negative allosteric modulators, which is located outside helices V–VII, close to the receptor's intracellular domain.¹⁸⁶ A molecular-modelling and mutagenesis study has shown that agonist positive allosteric modulators also target the same region but in a clear-cut sub-pocket at the interface between helices V and VI, which may aid the formation of an intracellular binding site that enhances G-protein coupling.¹⁸⁶ The secretin receptors have immense potential in drug discovery due to their importance in fundamental homeostatic

functions. To date, three of these hormones are used clinically: glucagon, parathyroid hormone and calcitonin, for the treatment of hypoglycaemia, osteoporosis, and hypercalcaemia, respectively.⁸⁸ This chapter determines the allosteric binding site and molecular mechanism of allosteric binding to GLP1-R, using allosteric modulators identified through a literature survey. In silico evaluation of the ADME/Tox properties of the allosteric modulators was also performed.

4.2 Computational Methods

4.2.1 Selection of ligands and receptors

Seven known positive allosteric modulators of GLP-1R and an electrophilic compound respectively were selected for this study because of their GLP-1R activating activity, as reported by Bueno et al.¹⁹⁰, these are 2,6,7-trichloro-3-(trifluoromethyl)quinoxaline (1, Figure 4.4), 1-(5-(4-(tert-butyl)phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one (2, Figure 4.4), 2-((4-chlorophenyl)thio)-3-(trifluoromethyl)quinoxaline (3, Figure 4.4), 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate (4, Figure 4.4), 2-((6,7-dichloro-3-(trifluoromethyl) quinoxalin-2-yl) thio)-5-methyl-1,3,4-thiadiazole (1, Figure 4.3), 5-(6,7-dichloro-3-((1-(1-methylpiperidin-4-yl) ethyl)-1H-tetrazol-5-yl) thio) quinoxalin-2-yl) thiazol-2-ol (2, Figure 4.3), 1-([1,1'-biphenyl]-4-carbonyl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one (3, Figure 4.3) and 5-chloro-3-phenyl- [1,2,3] triazolo[1,5-a] quinazoline (4, Figure 4.3). These structures were drawn using ChemDraw version 18.0 (PerkinElmer Informatics, Inc). The crystallographic coordinates of the active and inactive GLP-1R structures were obtained from the protein data bank with the following PDB IDs: 5VAI (resolution: 4.1 Å)¹⁸⁷ (active), 5VEW (resolution: 2.7 Å)¹⁸⁶ (inactive) and 6B3J (resolution: 3.3 Å)¹⁹¹ (active). Newer structures of the GLP-1R (PDB IDs: 6KJV (resolution: 2.8 Å)¹⁹², 6VCB (resolution: 3.3 Å)¹⁹³, 6XOX (resolution: 3.1 Å)¹⁹⁴) have been crystallised using methods such as electron microscopy and X-ray diffraction. These recent structures differ from the crystal structures used in this study and from each other in terms of the ligand bound for the crystallisation, method of crystallisation, and resolution (Appendix 1 [Table S1]).

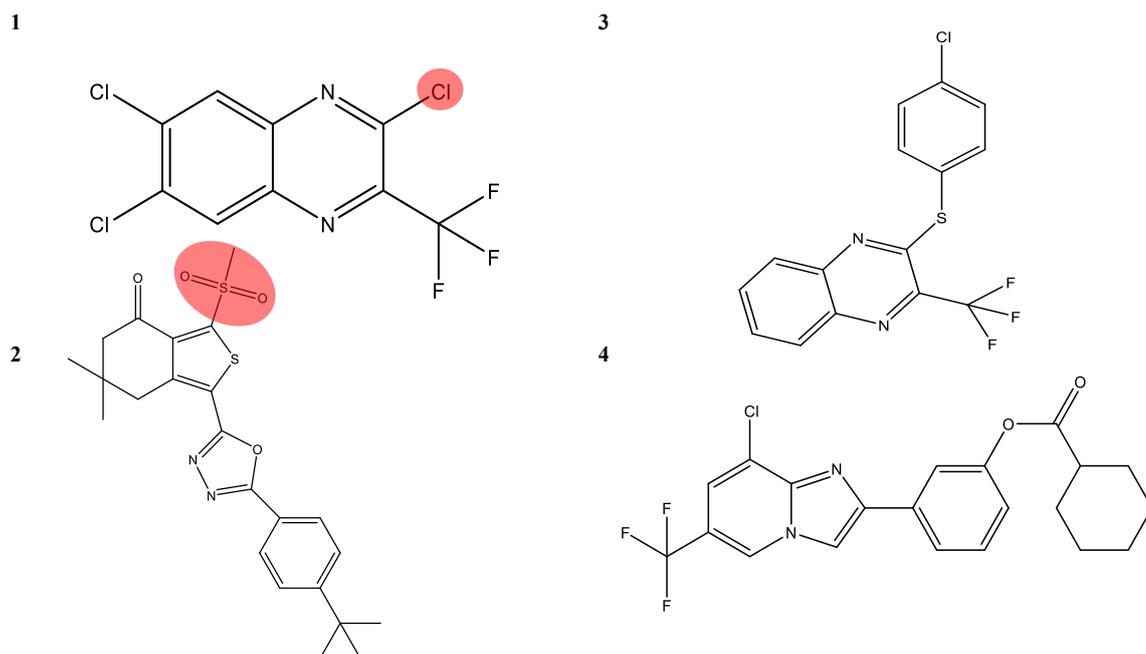


Figure 4.3. Chemical structures of known Glucagon-like peptide 1 receptor allosteric modulators. Adopted from Bueno et al.¹⁹⁰. The groups which leave in a nucleophilic attack in compounds 1 and 2 are shaded in red. 1: 2,6,7-trichloro-3-(trifluoromethyl)quinoxaline (Compound, EC₅₀ Wildtype (WT) (S.D., n) 4700 (1000,4) nm)¹⁹⁰, 2: 1-(5-(4-(tert-butyl)phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one (Compound, EC₅₀ Wildtype (WT) (S.D., n) 1500 (1000,4) nm)¹⁹⁰, 3: 2-((4-chlorophenyl)thio)-3-(trifluoromethyl)quinoxaline (Compound, EC₅₀ Wildtype (WT) (S.D., n) >30000 (NA,4) nm)¹⁹⁰, 4: 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate (Compound, EC₅₀ Wildtype (WT) (S.D., n) >30000 (NA,4) nm).¹⁹⁰

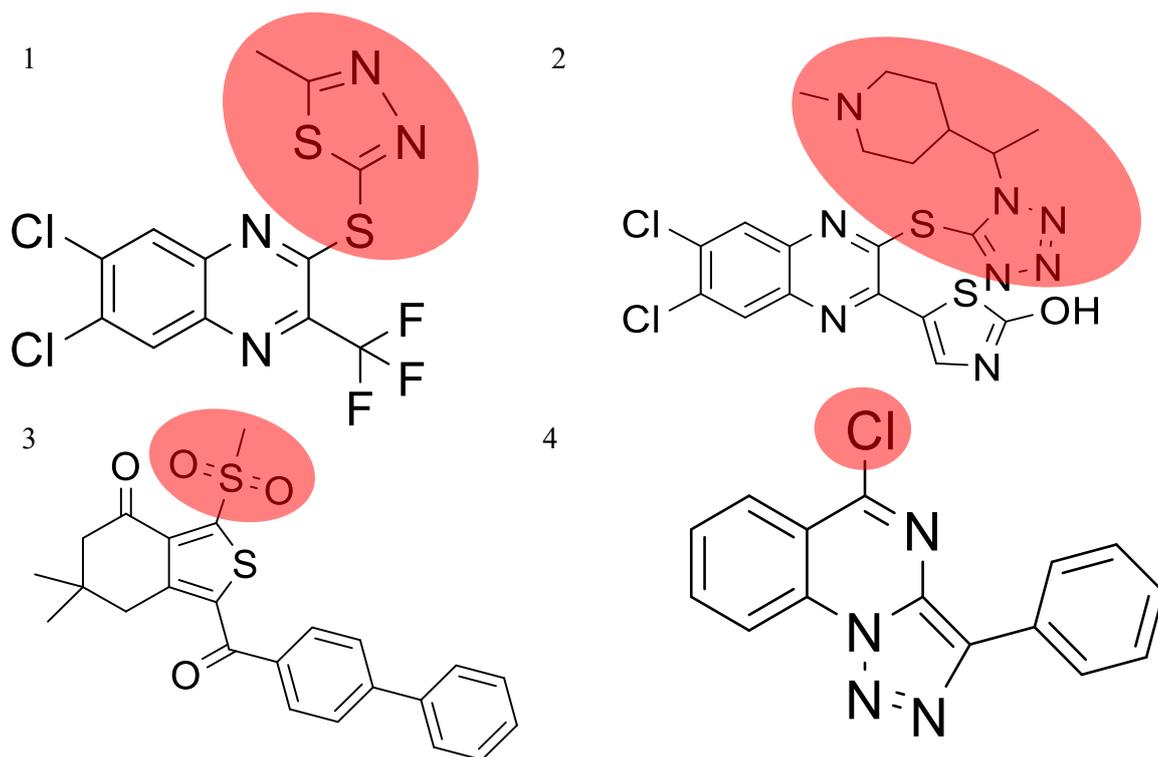


Figure 4.4. Chemical structures of known Glucagon-like peptide 1 receptor allosteric modulators and an electrophilic compound (Adopted from Bueno et al.¹⁹⁰). The groups which leave in a nucleophilic attack are circled. Key; 1: 2-((6,7-dichloro-3-(trifluoromethyl)quinoxalin-2-yl)thio)-5-methyl-1,3,4-thiadiazole. 2: 5-(6,7-dichloro-3-((1-(1-(1-methylpiperidin-4-yl)ethyl)-1H-tetrazol-5-yl)thio)quinoxalin-2-yl)thiazol-2-ol. 3: 1-([1,1'-biphenyl]-4-carbonyl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one and 4: 5-chloro-3-phenyl-[1,2,3]triazolo[1,5-a]quinazoline

4.2.2 Ligand and Receptor Preparation, Molecular docking, and Binding affinity analysis

Ligand docking into the active (PDB IDs: 5VAI¹⁸⁷, 6B3J¹⁹¹ – Chain R) and the inactive (PDB ID: 5VEW¹⁸⁶ – Chain A) structures of the GLP-1R crystal structures was performed using Flare, version 5.0 (Cresset Software, Litlington, Cambridgeshire). The protein and ligand structures were prepared using the Flare software default settings. (The full preparation of the protein and ligand at a pH of 7.0 and active site size of 6.00 Å adds missing hydrogens to proteins and cofactors and assigns optimal ionization states to the protein residues. It optimises the spatial positions of polar hydrogen atoms to maximise

hydrogen bonding and minimise steric clashes. The side chain orientation of His, Asn and Gln are optimized, then the residues with unsolved side chains are detected and reconstructed). The protein and ligands were minimized using the eXtended Electron Distribution (XED) accurate method on Flare. The very accurate but slow option for the docking calculation was selected. The grid box was set by picking amino acid residues present in the protein's TM5, TM6 and TM7. The poses with the highest binding energy were selected and visualised using BIOVIA Discovery Studio Visualiser, version 19.1 (Dassault Systemes, San Diego, United States).

The binding affinity of the protein-ligand complexes was analysed using the PRODIGY web server.^{195, 196}

4.2.3 *In silico* ADME/Tox prediction

In silico ADMET studies were performed using the ADMET predictor v.9.5 (Simulations Plus, Inc., Lancaster, CA, USA), webservers ADMETLab, and SwissADME developed by Dong *et al.*⁷⁸ and Daina *et al.*⁸³, respectively. The SMILES code for each the compounds was uploaded into the software for evaluations. The ADMET properties were calculated at pH 7.4, evaluating the compounds' physicochemical, metabolic, and toxicity properties.

4.3. *In silico* ligand docking approaches to characterise the binding of known allosteric modulators to the glucagon – like peptide 1 receptor and prediction of ADME/Tox properties.

4.3.1 *Results*

Interaction of 2,6,7-trichloro-3-(trifluoromethyl) quinoxaline with the GLP-1 Receptor

The findings of the *in silico* allosteric modulators docking studies showed 2,6,7-trichloro-3-(trifluoromethyl) quinoxaline being juxtaposed to TM6 of the active structure with PDB ID: 5VAI¹⁸⁷ (binding energy -6.46 Kcal/mol) (Figure 4.5A), TM7 of the inactive structure with PDB ID: 5VEW¹⁸⁶ (binding energy -8.50 Kcal/mol) (Figure 4.5B), and both TM6 of the active structure with PDB ID: 6B3J¹⁹¹ (binding energy -6.70 Kcal/mol) (Figure 4.5C). In poses 2 and 3 of the 5VAI-2,6,7-trichloro-3-(trifluoromethyl) quinoxaline complex, the ligand docked at the TM6 of the receptor with binding energies of 6.34

Kcal/mol and 6.25 Kcal/mol respectively. In the pose 2 of the 5VEW-2,6,7-trichloro-3-(trifluoromethyl) quinoxaline complex (binding energy -8.18 Kcal/mol) the ligand docked at the TM6 of the allosteric binding site. In the third pose, the ligand juxtaposed at TM7 of the receptor (binding energy -6.26 Kcal/mol). In the 6B3J-2,6,7-trichloro-3-(trifluoromethyl) quinoxaline complex, the ligand was bound to the TM6 of the receptor with bindings energies of -6.59 Kcal/mol and -6.43 Kcal/mol for poses 2 and three respectively. A 2D schematic of the allosteric modulators' protein-ligand interaction and respective receptors is presented (Figure 4.6). A summary of the amino acids interacting with the ligand at the allosteric binding site is presented in Table 4.1, Appendix 2 [Table S3] and Appendix 3 [Table S4].

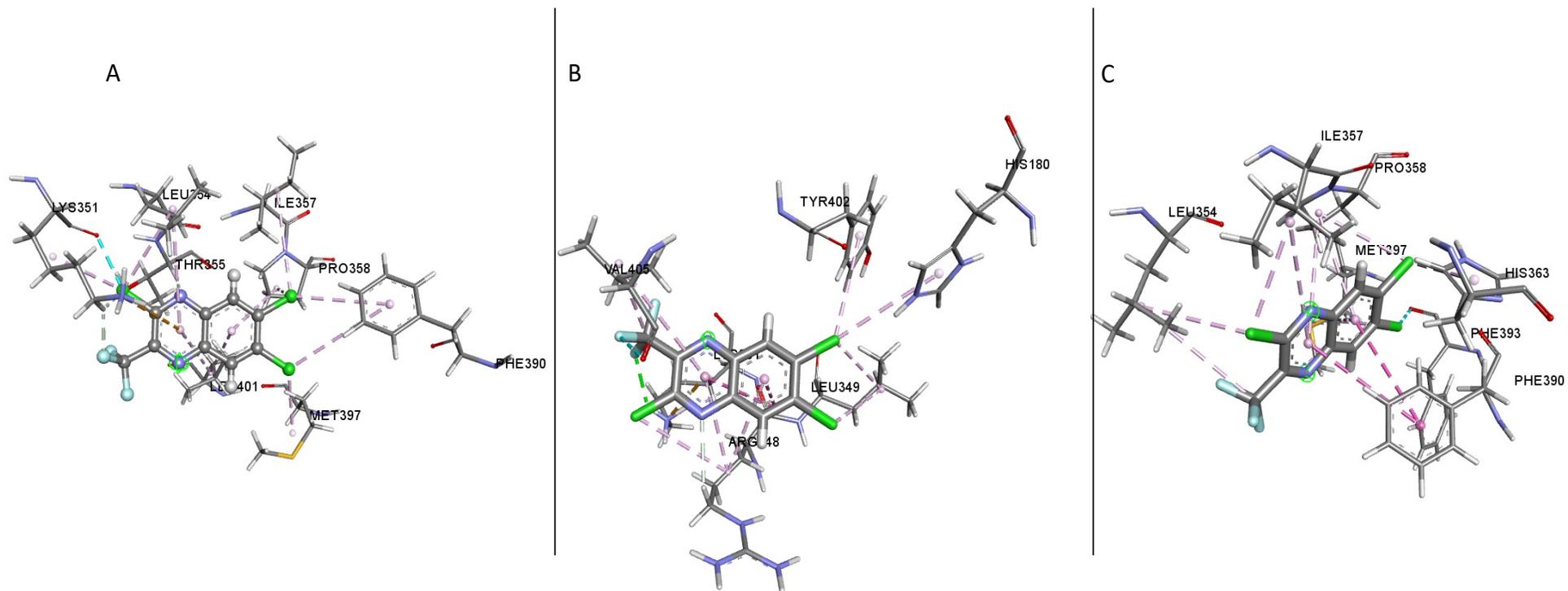


Figure 4.5. Schematic showing ligand (2,6,7-trichloro-3-(trifluoromethyl) quinoxaline) – GLP-1R complexes. (A) 1:5VAI, (B) 1:5VEW, (C) 1:6B3J.

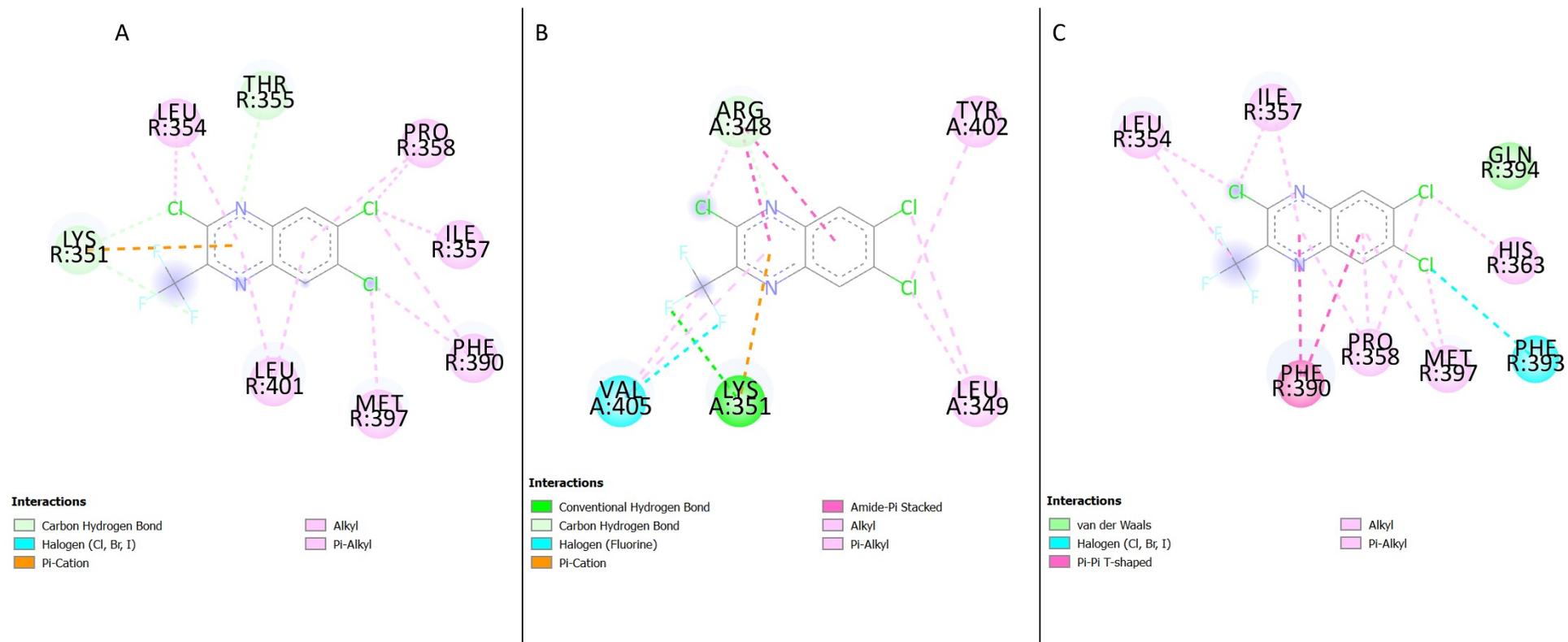


Figure 4.6. The protein-ligand interaction between 2,6,7-trichloro-3-(trifluoromethyl) quinoxaline and the GLP 1 receptor generated using BIOVIA Discovery Studio Visualiser. (A) 1:5VAI, (B) 1:5VEW, and (C) 1:6B3J.

Table 4.1. Table showing amino acids interacting with 2,6,7-trichloro-3-(trifluoromethyl)quinoxaline in the GLP-1R structure.

S/N	5VAI	5VEW	6B3J
1	LEU 354	ARG 348	LEU 354
2	THR 355	TYR 402	ILE 357
3	PRO 358	VAL 405	GLN 394
4	ILE 357	LYS 351	HIS 363
5	PHE 390	LEU 349	PHE 393
6	MET 397	HIS 180	MET 397
7	LEU 401		PRO 358
8	LYS 351		PHE 390

Interaction of 1-(5-(4-(tert-butyl) phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one with the GLP-1 receptor

Ligand docking studies were performed to understand the binding of 1-(5-(4-(tert-butyl) phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one to the active (PDB IDs: 5VAI and 6B3J) and inactive (5VEW) of GLP-1R structures. The results also showed 1-(5-(4-(tert-butyl) phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one being juxtaposed to TM6 of 5VAI ¹⁸⁷ (binding energy -7.55 Kcal/mol) (Figure 4.7A), to TM6 of 5VEW ¹⁸⁶ (binding energy -7.189 Kcal/mol) (Figure 4.7B), and TM6 of 6B3J (binding energy -7.67Kcal/mol) (Figure 4.7C). In the other poses analysed the ligand docked at the TM6 of all the GLP1R structure used for this study (5VAI, pose 2 binding energy (-7.45 Kcal/mol) and pose 3 binding energy (-7.37 Kcal/mol); 6B3J, pose 2 binding energy (-7.59 Kcal/mol) and pose 3 binding energy (-7.45 Kcal/mol); 5VEW, pose 2 binding energy (-7.066 Kcal/mol) and pose 3 binding energy (-6.873 Kcal/mol). A 2D schematic of the protein-ligand interaction of the allosteric modulators and respective receptors are presented in Figures 4.8A-C. The amino acid residues interacting with the ligand in the binding site of 1-(5-(4-(tert-butyl) phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one can be observed in Tables 3.2, S5 (Appendix 4) and S6 (Appendix 5).

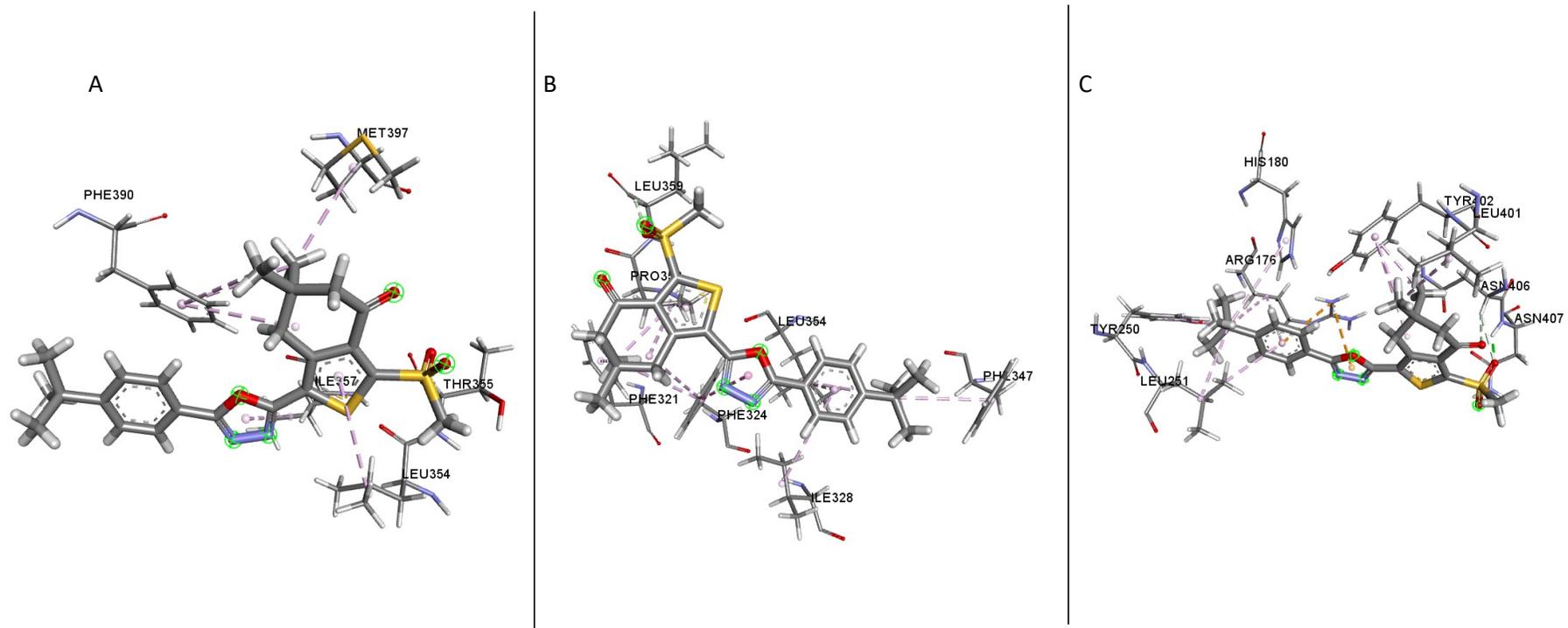


Figure 4.7. Schematic showing ligand (1-(5-(4-(tert-butyl) phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one) - GLP-1R complexes. (A) 2:5VAI, (B) 2:5VEW, (C) 2:6B3J.

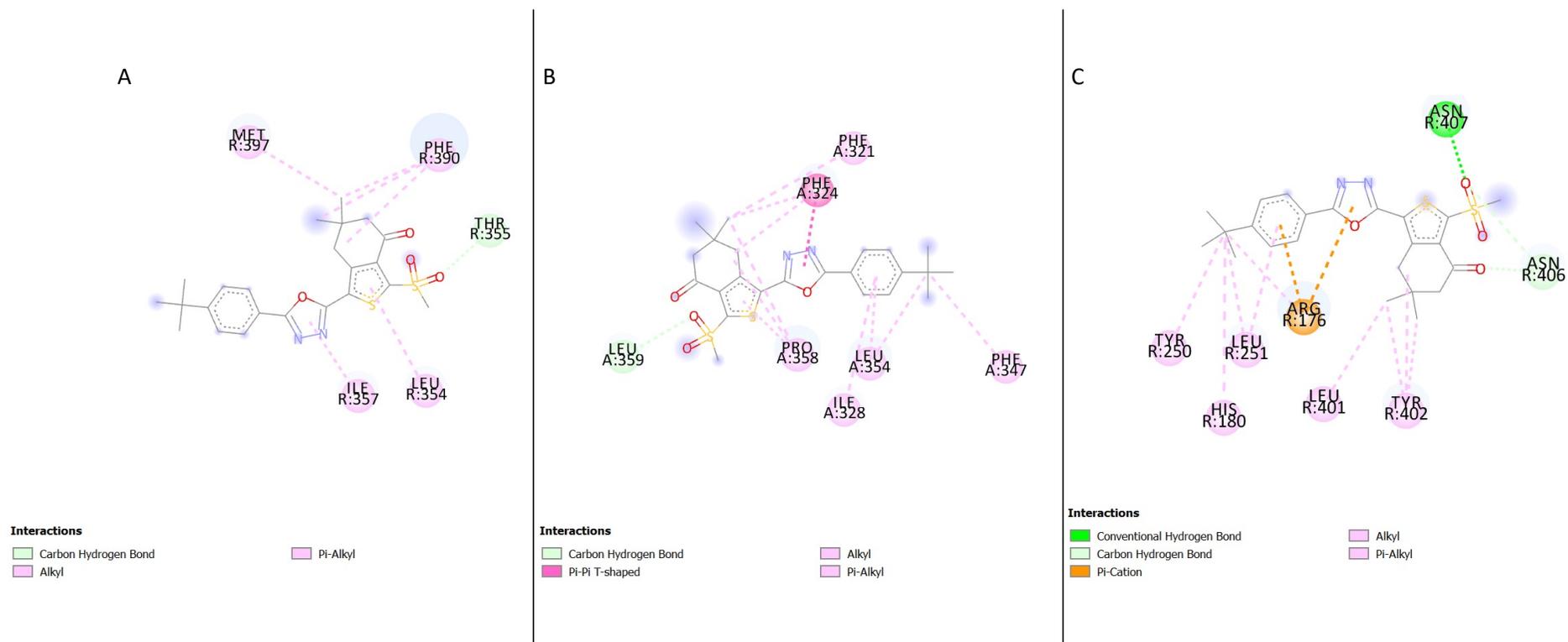


Figure 4.8. The protein-ligand interaction between 1-(5-(4-(tert-butyl) phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one and the GLP 1 receptor generated using BIOVIA Discovery Studio Visualiser. (A) 2:5VAI, (B) 2:5VEW, and (C) 2:6B3J.

Table 4.2. Amino acids which interact in the binding sites with 1-(5-(4-(tert-butyl)phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one.

S/N	5VAI	5VEW	6B3J
1	MET 397	LEU 359	ASN 407
2	PHE 390	PHE 324	ASN 406
3	THR 355	PHE 321	TYR 402
4	LEU 354	PHE 347	LEU 401
5	ILE 357	LEU 354	ARG 176
6		ILE 328	LEU 251
7		PRO 358	HIS 180
8			TYR 250

Interaction of 2-((4-chlorophenyl) thio)-3-(trifluoromethyl) quinoxaline with the GLP-1 Receptor

Docking studies carried out for the GLP-1R crystal structures showed 2-((4-chlorophenyl) thio)-3-(trifluoromethyl) quinoxaline being juxtaposed to a pocket behind TM6 in 5VAI (binding energy 6.44 Kcal/mol) (Figure 4.9A), in the inactive structure (5VEW), it juxtaposed to TM6 of the receptor (binding energy 7.877 Kcal/mol) (Figure 4.9B). In the other active structure (6B3J), 2-((4-chlorophenyl) thio)-3-(trifluoromethyl) quinoxaline juxtaposed at TM6 (binding energy -7.27Kcal/mol) (Figure 4.9C). A 2D schematic of the ligands' protein-ligand interaction and respective receptors is presented (Figures 4.10). In poses 2 and 3 of the 5VAI-2-((4-chlorophenyl) thio)-3-(trifluoromethyl) quinoxaline complex, the ligand docked at the TM6 of the receptor with binding energies of 6.31 Kcal/mol and 6.28 Kcal/mol, respectively. In the pose 2 of the 5VEW-2-((4-chlorophenyl) thio)-3-(trifluoromethyl) quinoxaline complex (binding energy -6.86 Kcal/mol) the ligand docked at the TM6 of the allosteric binding site. In the third pose, the ligand juxtaposed at TM6 of the receptor (binding energy -6.82 Kcal/mol). In the 6B3J-2-((4-chlorophenyl) thio)-3-(trifluoromethyl) quinoxaline complex, the ligand was bound between TM6 and TM7 (pose 2) of the receptor with binding energy of -7.12 Kcal/mol. In pose 3, the ligand bound to TM7 of the receptor (binding energy -6.92 Kcal/mol). Table 4.3

below shows the amino acid residues interacting in the binding sites with 2-((4-chlorophenyl) thio)-3-(trifluoromethyl) quinoxaline. Tables S7 (Appendix 6) and S8 (Appendix 7) shows amino acid residues interacting with 2-((4-chlorophenyl) thio)-3-(trifluoromethyl) quinoxaline in poses 2 and 3.

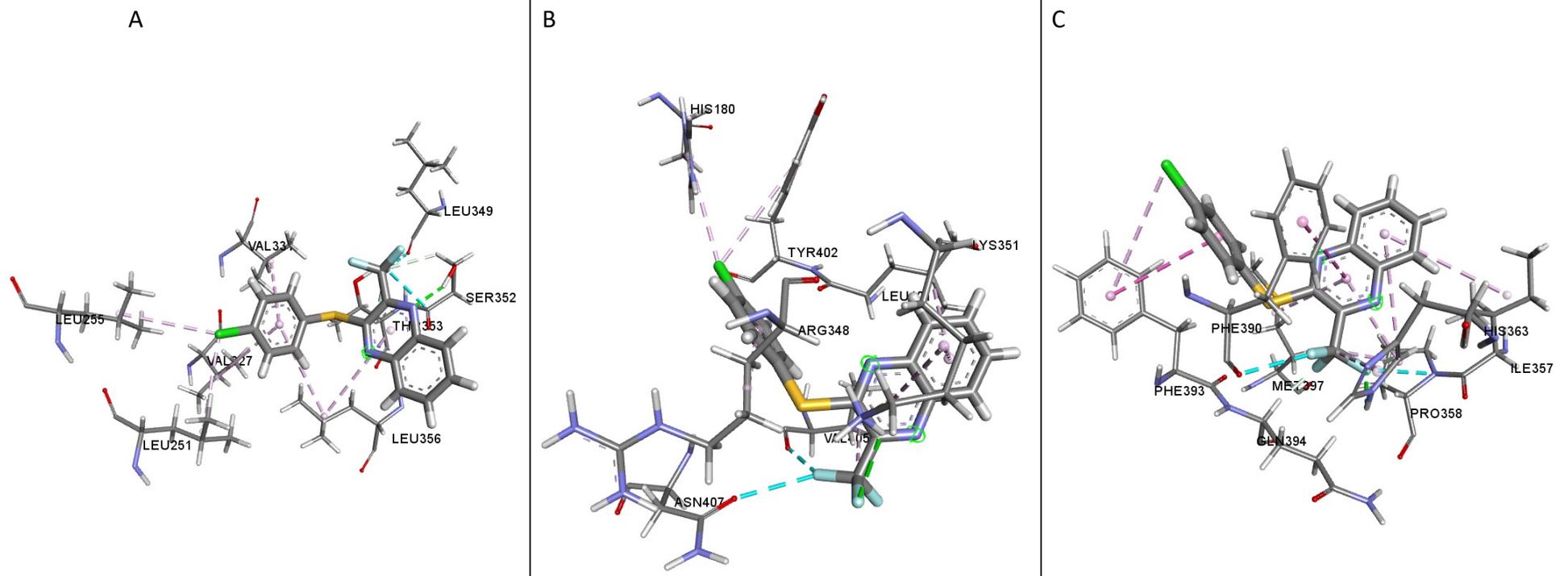


Figure 4.9. Schematic showing ligand (2-((4-chlorophenyl) thio)-3-(trifluoromethyl) quinoxaline) - GLP-1R complexes. (A) 3:5VAI, (B) 3:5VEW, (C) 3:6B3J.

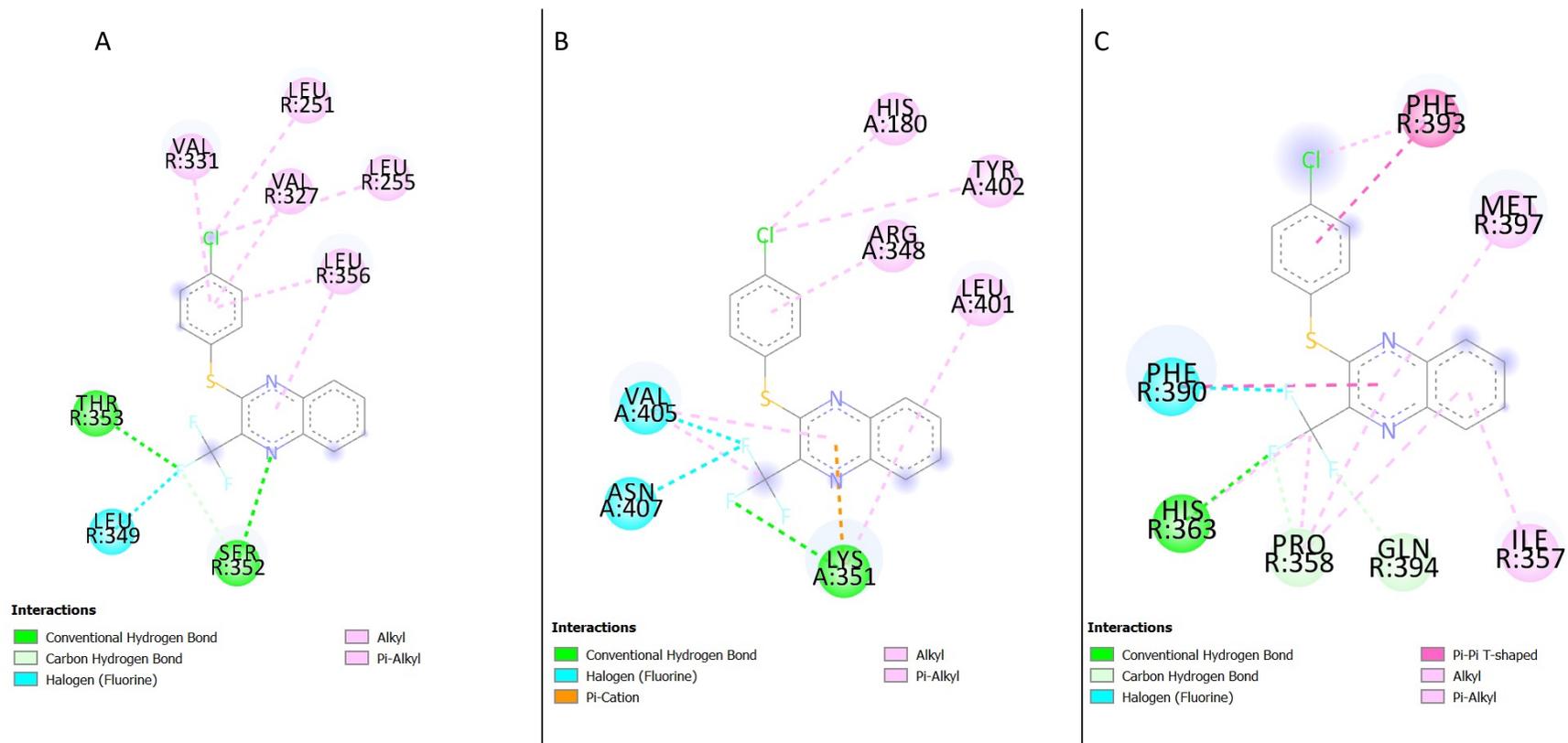


Figure 4.10. The protein-ligand interaction between 2-((4-chlorophenyl) thio)-3-(trifluoromethyl) quinoxaline and the GLP 1 receptor generated using BIOVIA Discovery Studio Visualiser. (A) 3:5VAI, (B) 3:5VEW, and (C) 3:6B3J.

Table 4.3. Amino acids present in the binding site of GLP-1R interacting with 2-((4-chlorophenyl) thio)-3-(trifluoromethyl) quinoxaline.

S/N	5VAI	5VEW	6B3J
1	VAL 331	HIS 180	PHE 393
2	LEU 251	TYR 402	MET 397
3	VAL 327	ARG 348	ILE 357
4	LEU 255	LEU 401	GLN 394
5	LEU 356	LYS 351	PRO 358
6	SER 352	ASN 407	HIS 363
7	LEU 349	VAL 405	PHE 390
8	THR 353		

Interaction of 3-(8-chloro-6-(trifluoromethyl) imidazo[1,2-a] pyridin-2-yl) phenyl cyclohexanecarboxylate with the GLP-1 Receptor

Ligand docking studies were performed to understand how 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate binds to active (5VAI¹⁸⁷ and 6B3J¹⁹¹) and inactive (5VEW¹⁸⁶) GLP-1R structures. The findings showed 3-(8-chloro-6-(trifluoromethyl) imidazo[1,2-a] pyridin-2-yl) phenyl cyclohexanecarboxylate being juxtaposed to TM6 in 5VAI (binding energy -7.25 Kcal/mol) (Figure 4.11A), to TM6 in 5VEW (binding energy -9.48 Kcal/mol) (Figure 4.11B) and TM6 and TM7 in 6B3J (binding energy -7.95 Kcal/mol) (Figure 4.11C). In the other poses analysed the ligand docked at the TM6 of the active GLP1R structure used for this study (5VAI, pose 2 binding energy (-7.159 Kcal/mol) and pose 3 binding energy (-7.157 Kcal/mol); 6B3J, pose 2 binding energy (-7.77 Kcal/mol) and pose 3 binding energy (-7.73 Kcal/mol)). In the inactive structure 5VEW, the ligand docked at TM6 in the second pose (binding energy -9.27 Kcal/mol) and in pose 3, 3-(8-chloro-6-(trifluoromethyl) imidazo[1,2-a] pyridin-2-yl) phenyl cyclohexanecarboxylate juxtaposed between TM6 and TM7 (binding energy -8.22 Kcal/mol). A 2D schematic of the protein-ligand interaction of the ligands and respective receptors are presented in Figures 4.12A-C. Table 4.4 below shows the amino acid residues interacting in the binding sites with 3-(8-chloro-6-(trifluoromethyl) imidazo[1,2-a] pyridin-2-yl) phenyl cyclohexanecarboxylate. Tables S9 (Appendix 8) and S10 (Appendix 9) shows

amino acid residues interacting with 3-(8-chloro-6-(trifluoromethyl) imidazo[1,2-a] pyridin-2-yl) phenyl cyclohexanecarboxylate in poses 2 and 3.

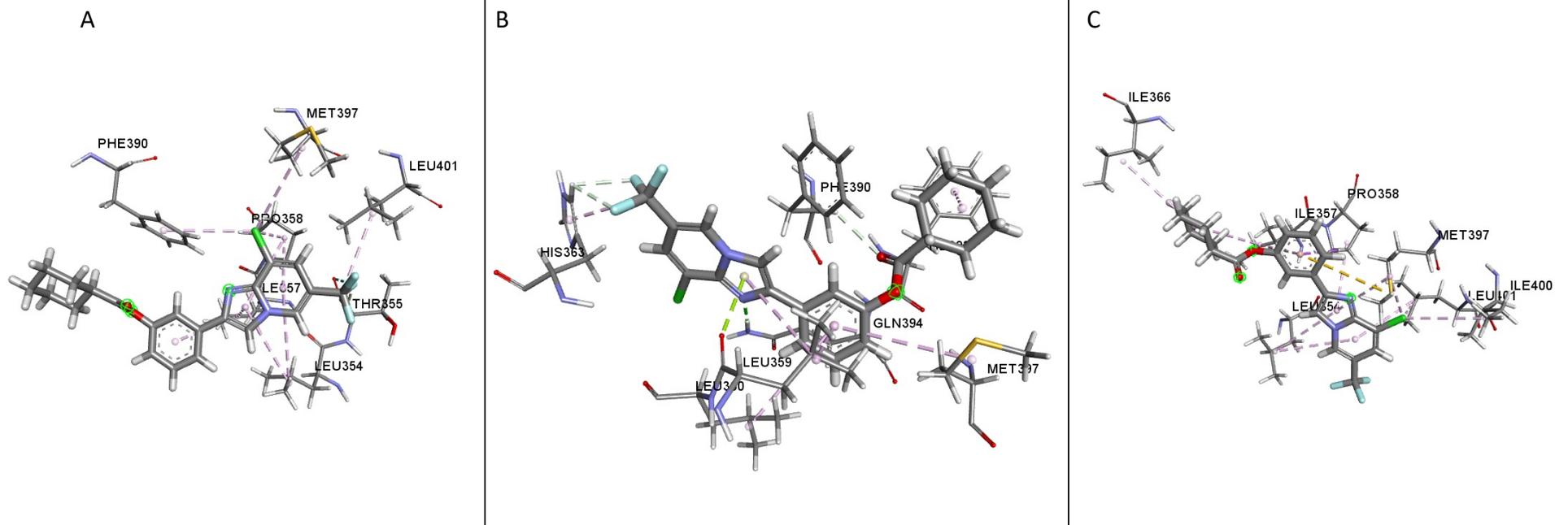


Figure 4.11. Schematic showing 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate bound to the crystal structures of GLP-1R and the residues present in the active site. (A) 4:5VAI, (B) 4:5VEW, (C) 4:6B3J.

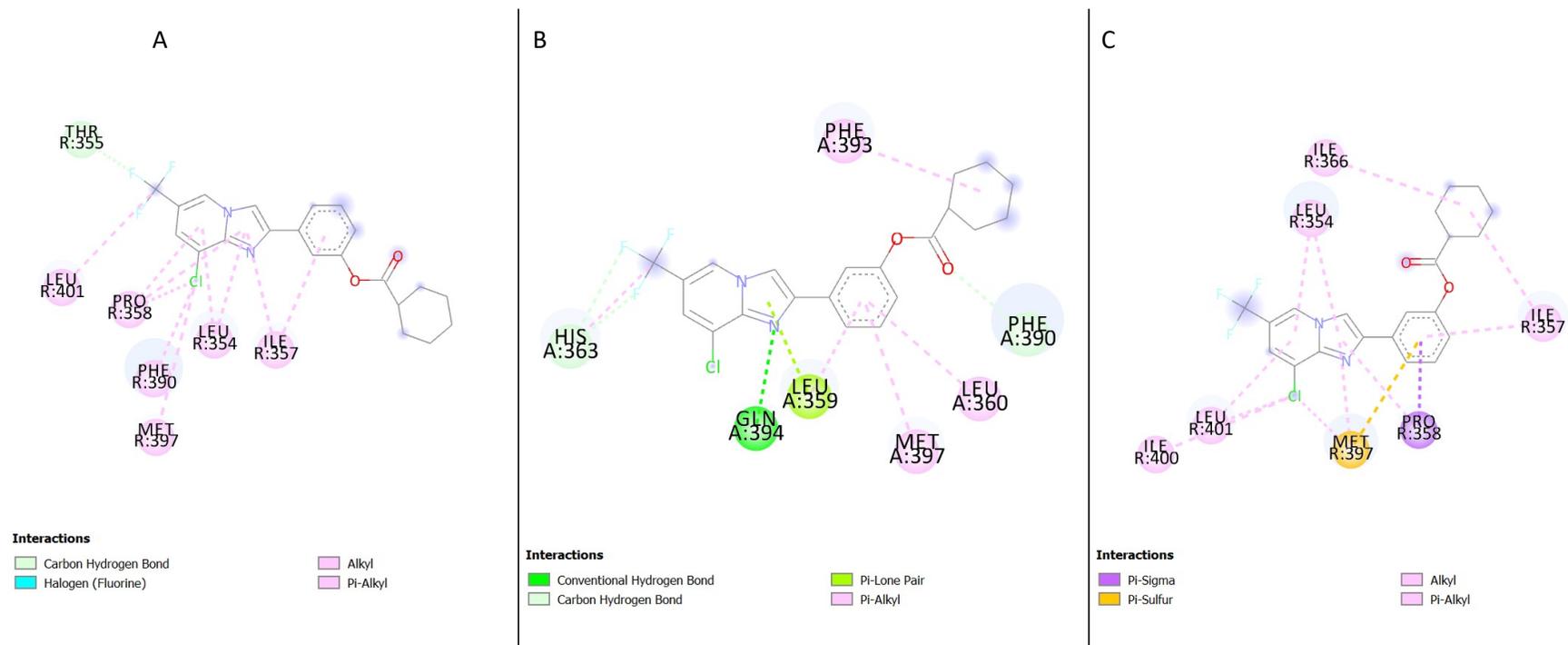


Figure 4.12. The protein-ligand interaction between 3-(8-chloro-6-(trifluoromethyl) imidazo[1,2-a] pyridin-2-yl) phenyl cyclohexanecarboxylate and the GLP 1 receptor generated using Schrodinger Maestro. (A) 4:5VAI (B) 4:5VEW and (C) 4:6B3J.

Table 4.4. Amino acids interacting with 3-(8-chloro-6-(trifluoromethyl) imidazo[1,2-a] pyridin-2-yl) phenyl cyclohexanecarboxylate in the GLP-1R binding pocket.

S/N	5VAI	5VEW	6B3J
1	THR 355	PHE 393	ILE 366
2	LEU 401	PHE 390	ILE 357
3	PRO 358	LEU 360	PRO 358
4	PHE 390	MET 397	MET 397
5	MET 397	LEU 359	LEU 401
6	LEU 354	GLN 394	ILE 400
7	ILE 357	HIS 363	LEU 354

Binding Affinity Analysis

Binding affinity analysis is used to measure the strength of biomolecular interactions. The binding affinity of any complex in thermodynamic terms is crucial in determining the feasibility of an interaction occurring in a cell or not at specified conditions^{195, 197}. The findings of the binding affinity analysis of the protein-ligand complexes analysed using PRODIGY-LIG webserver are shown in the table below (Table 4.5).

Table 4.5. Binding affinities of the docked complexes of GLP-1R and the ligands.

Complexes	Binding Affinity ΔG_{noelec} (Kcal/mol)
5VAI-2,6,7-trichloro-3-(trifluoromethyl) quinoxaline	-6.8
5VEW-2,6,7-trichloro-3-(trifluoromethyl) quinoxaline	-7.2
6B3J-2,6,7-trichloro-3-(trifluoromethyl) quinoxaline	-7.2
5VAI-1-(5-(4-(tert-butyl) phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one	-8.5
5VEW-1-(5-(4-(tert-butyl) phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one	-8.1
6B3J-1-(5-(4-(tert-butyl) phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one	-8.8

5VAI-2-((4-chlorophenyl) thio)-3-(trifluoromethyl) quinoxaline	-7.5
5VEW-2-((4-chlorophenyl) thio)-3-(trifluoromethyl) quinoxaline	-7.9
6B3J-2-((4-chlorophenyl) thio)-3-(trifluoromethyl) quinoxaline	-7.9
5VAI-3-(8-chloro-6-(trifluoromethyl) imidazo[1,2-a] pyridin-2-yl) phenyl cyclohexanecarboxylate	-8.5
5VEW-3-(8-chloro-6-(trifluoromethyl) imidazo[1,2-a] pyridin-2-yl) phenyl cyclohexanecarboxylate	-8.6
6B3J-3-(8-chloro-6-(trifluoromethyl) imidazo[1,2-a] pyridin-2-yl) phenyl cyclohexanecarboxylate	-9.2

***In silico* ADME/Tox Prediction**

The findings of the *in silico* ADME/Tox evaluation of the allosteric modulators of GLP-1R are shown in the tables below. Table 4.6 shows the physicochemical properties of the selected allosteric modulators used for the study, including the following parameters: the octanol-water partition coefficient based on Moriguchi's model (MlogP), the octanol-water partition coefficient based on Simulation plus's model (SlogP), the octanol-water partition coefficient as a function of pH (LogD), the predicted human jejunal permeability (P_{eff}), the blood-brain barrier filter (BBB filter), aqueous solubility in pure water (S_w), volume of distribution (VD), Lipinski's rule of 5 (Ro5), topological polar surface area (TPSA), hydrogen bond acceptors (HBA), hydrogen bond donors (HBD) and molecular weight (MW).

The results of the effects of the allosteric modulators on transport proteins predicted *in silico* are presented in Table 4.7. Table 4.8 shows the results of the activity of the allosteric modulators on the various isoforms of hepatic CYP450. Table 4.9 shows the toxicity parameters predicted using ADMET Predictor 9.5.

Table 4.6. Physicochemical properties of the selected allosteric modulators generated using ADMET Predictor 9.5.

Lig.	MW	MlogP	SlogP	logD	P _{eff} cm/s x 10 ⁴	BBB Filter	Sw mg/mL	VD L/Kg	% Unbound	Ro5	TPSA	HBA	HBD
1	301.48	3.15	4.25	4.25	8.86	High	0.001	4.75	8.00	0	25.78 Å ²	5	0
2	458.59	3.39	4.42	4.42	1.75	High	0.0002	0.86	7.03	0	126.75 Å ²	6	0
3	340.75	3.37	5.08	5.08	8.15	High	0.0002	2.99	3.34	0	51.08 Å ²	5	0
4	422.83	4.34	6.43	6.43	4.27	High	4.5E-05	3.48	2.79	1	43.60 Å ²	6	0

Table 4.7. Effect of the selected allosteric modulators generated using ADMET Predictor 9.5 on transport proteins.

Lig.	Pgp Substrate	Pgp Inhibitor	OATP1B1 Inhibitor	OCT2 Inhibitor	BSEP Inhibitor	BCRP substrate
1	No	No	No	Yes	No	Yes
2	No	Yes	No	Yes	Yes	No
3	No	No	No	Yes	Yes	Yes
4	No	Yes	No	Yes	Yes	Yes

1: 2,6,7-trichloro-3-(trifluoromethyl)quinoxaline, 2: 1-(5-(4-(tert-butyl)phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one, 3: 2-((4-chlorophenyl)thio)-3-(trifluoromethyl)quinoxaline, 4: 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate

Table 4.8. Predicted metabolism of the selected allosteric modulators via cytochrome P450 isoforms.

Lig.	CYP1A2	CYP2A6	CYP2B6	CYP2C8	CYP2C9	CYP2C19	CYP2D6	CYP2E1	CYP3A4
1	+/-	+	+	NS	NI/+	NI/+	NI/NS	+	NI/+
2	-/NS	NS	NS	+	-/NS	NI/NS	-/NS	NS	+/-
3	+/-	+	+	NS	NI/+	NI/+	NI/+	NS	NI/+
4	-/NS	NS	NS	NS	-/NS	-/NS	NI/NS	NS	NI/+

Key; Inhibition of CYP isoform: (-), substrate of CYP isoform: (+), NS = Non-substrate, NI = non-inhibitor.

Table 4.9. Toxicity parameters predicted using ADMET Predictor 9.5.

Lig.	AMES Toxicity	Skin Sens	hERG Filter	Repro Tox	Ser AlkPhos	Ser GGT	Ser AST	Ser ALT
1	Negative	+	No	Non-Toxic	Elevated	Elevated	Elevated	Elevated
2	Negative	-	No	Toxic	Elevated	Normal	Normal	Normal
3	Negative	+	No	Toxic	Elevated	Elevated	Elevated	Elevated
4	Negative	+	No	Toxic	Normal	Normal	Elevated	Elevated

Key; Sensitizer: (+), non-sensitizer: (-).

1: 2,6,7-trichloro-3-(trifluoromethyl)quinoxaline, 2: 1-(5-(4-(tert-butyl)phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one, 3: 2-((4-chlorophenyl)thio)-3-(trifluoromethyl)quinoxaline, 4: 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate

4.3.2 Discussion

A recent study has shown that T2D treatment needs the positive allosteric modulation of GLP-1R to inhibit glucagon secretion, thus stimulating insulin secretion.¹⁸⁶ It has also been suggested that the novel agonist human monoclonal antibody IRAB-A binds allosterically to the insulin receptor and thereby activates and enhances the signalling of insulin.¹⁹⁸ Molecular docking is a virtual screening method used to discover new ligands for GPCRs. It ranks active molecules high and produce poses which will inform chemists which compounds to purchase for further screening.¹⁴¹

Molecular docking has been used extensively in GPCR drug discovery to identify compounds (hit and lead generation) that target different receptors in the GPCR family.¹⁷¹ Jenkins *et al.*, utilised Flare to study 2-mercaptoacetamide (2MA) a structural analogue of urea, their findings showed that 2-MA is a competitive inhibitor and flare is a robust software for performing docking simulations.¹⁹⁹ Egorov *et al.*, in their in silico docking study performed using flare showed that the synthesised compounds would play a significant role in the treatment of ailments such as breast cancer, neurodegenerative diseases etc.²⁰⁰ Carlsson *et al.*, applied molecular docking to perform a screening of over six million commercially available compounds against the active like conformations of A_{2A}AR. Their findings showed that nine of the 20 predicted agonists were confirmed to be A_{2A}AR ligands.²⁰¹ Docking based programs can generate 3D conformations of binding structures that is very useful for function and drug-based analysis.²⁰² Hou *et al.*, used techniques such as homology, molecular dynamics, and molecular docking to assess prediction accuracy of ligand-binding poses and screening power of docking-based virtual screening. Their findings showed that the crystal structures outperformed the homology models before any refinement through molecular dynamics. However, the optimised homology models show a similar performance to the crystal structures following a docking assessment.²⁰³ Shoichet *et al.*, applied ligand docking to screen a large library of compounds to identify compounds with joint activity against on-targets and selectivity versus anti-targets using selected GPCRs (dopamine D₂, serotonin 5-HT_{2A}, histamine H₁, κ-opioid and μ-opioid receptors).²⁰⁴ Their findings showed a hit range of 40% to 65% for the on-targets with very reliable calculated binding affinities.²⁰⁴ Docking into a crystal structure produces an accurate ligand binding pose prediction without any refinement.²⁰⁵ The widescale application of molecular docking in drug development makes it a preferred method for this study; this has been paired with binding affinity predictions to determine the feasibility of the complexes.

The conformational transitions observed from the inactive to the active structures were similar to the findings of Liang *et al.*,¹⁹¹ who reported that the movement of TM 6 in the inactive state upon signalling created a binding pocket. The docking of the ligands to the inactive structure (PDB ID: 5VEW) is corroborated by Song *et al.*¹⁸⁶; they reported that positive allosteric modulators (PAM) of the GLP-1R bind outside helices 5-7 near the intracellular part of the receptor, but in a distinct sub-pocket between helices 5 and 6.

The docking result of 2,6,7-trichloro-3-(trifluoromethyl) quinoxaline showed several residues present in the binding sites of the different GLP-1R crystal structures; in the structure 5VAI, the following residues were identified (Table 4.1); Leu354, Lys351, Leu401, Thr355, Met397, Pro355, Phe390, Ile357 and Pro358 (Figure 4.6A). Most of these residues were hydrophobic amino acids except for Thr355 and Lys351; both are polar and positively charged. In the inactive structure 5VEW, the following residues were present in the binding site (Table 4.1), Arg348, Tyr402, Val405, Lys351, Leu349 and His180 (Figure 4.6B, Table 4.1). All the residues present in the binding site were polar. Upon examination of the structure 6B3J, the following residues interacted in the binding site (Table 4.1); Leu354, Ile357, Gln394, His363, Phe393, Met397, Pro358 and Phe390 (Figure 4.6C, Table 4.1). His363 and Gln394 were polar, while the remaining residues were hydrophobic. An analysis of the ligand's second and third top poses and the protein structures showed the residues Thr355 and Leu401 were present in the first three poses of the 5VAI-2,6,7-trichloro-3-(trifluoromethyl) quinoxaline complex (Table 4.1, Appendix 2 [Table S3] and Appendix 3 [Table S4]). The other protein structures used in the study did not have any similar interacting residues across the three poses analysed.

The presence of polar residues in the binding sites of the three structures examined suggests that these polar residues play a vital role in ligand binding.¹⁹¹ The presence of these polar residues in the crystal structures of the receptor indicates the presence of a significant polar binding network around the peptide bindings site.¹⁹¹ Bueno *et al.*¹⁹⁰ reported the capability of 2,6,7-trichloro-3-(trifluoromethyl)quinoxaline to potentiate GLP-1(9-36)-NH₂-mediated cAMP accumulation in GLP-1R-expressing cells. The study results showed that 2,6,7-trichloro-3-(trifluoromethyl) quinoxaline potentiated the activity of GLP-1(9-36)-NH₂ on the wild-type receptor but failed to exert the same effect in cells expressing the mutant GLP-1R, which lacks the cysteine-347 residue.¹⁹⁰ A comparative molecular dynamics simulation showed that the cysteine-347.36bF (C347.36bF) mutant maximises van der Waals interactions with all the three negative allosteric modulators (NAMs) PF-06372222, NNC0640 and MK-0893 through the stabilisation of the aliphatic side chain of

Lysine-3516.40b (K3516.40b) in an optimal conformation for hydrophobic interactions with NAMs.^{186, 192} An analysis of the findings from this study showed the residue Lys351(K351) forming hydrogen bonds and pi-cation interactions (Figure 4.6A, Figure 4.8, Figure 4.10B). The mutation S352→A terminates the inhibition of GLP-1R by NAMs while the T355→A eliminates the inhibition by NAMs, NNC0640 and PF-06372222 but does not inhibit the activity of MK-0893.¹⁸⁶ The findings from previous research stated that positions 352-355 play a crucial role in binding an allosteric inhibitor to GLP-1R.¹⁸⁶

The examination of the docking results of 1-(5-(4-(tert-butyl)phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one and the GLP-1R structures demonstrated several residues present in the active site of the crystal structures. In the active structure 5VAI, the following hydrophobic residues were present in the active site; Met397, Ile357, Leu354, Phe390. Thr355 was identified as the polar residues (Table 4.2, Figure 4.8A). All the residues interacting with the ligand in the binding site of the inactive crystal structure 5VEW were hydrophobic (Table 4.2, Figure 4.8B). Three of the residues, namely: Tyr402, His180 and Arg176 identified in the binding site of 6B3J, were polar, whilst the other residues were hydrophobic (Table 4.2, Figure 4.8C). The amino acid residue Phe390 interacted with compound 2, and the structure 5VAI in the top three poses was analysed (Table 4.2, Appendix 4 [Table S5] and Appendix 5 [Table S6]). The analysis of the top three poses of the 1-(5-(4-(tert-butyl)phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one - 5VEW complex showed the amino acid residues Phe324, Leu354 and Pro 358 present in all the poses analysed (Table 4.2, Appendix 4 [Table S5] and Appendix 5 [Table S6]).

The presence of polar residues in the binding site of the GLP-1R crystal structures suggests critical hydrogen bond interactions which maintain receptor integrity and apo state.¹⁹¹ A previous study has demonstrated the capacity of 1-(5-(4-(tert-butyl) phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one to potentiate GLP-1(9–36)-NH₂-mediated cAMP accumulation in GLP-1R-expressing cells. The study showed that 1-(5-(4-(tert-butyl) phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one potentiated the activity of GLP-1(9–36)-NH₂ on the wild-type receptor but failed to exert the same effect in cells expressing the mutant GLP-1R, which lacks the Cysteine-347 residue.¹⁹⁰

In the active structure (PDB ID: 5VAI¹⁸⁷-2-((4-chlorophenyl)thio)-3-(trifluoromethyl)quinoxaline complex), eight residues were identified in the active site, most of which were hydrophobic residues. The active site consists of hydrophobic residues

Val327, Val331, Leu251, Leu255, Leu356, and Leu349. Thr353 and Ser352 are the residues responsible for the polar region of the binding site (Table 4.3, Figure 4.10A). Upon examination of the inactive structure (PDB ID: 5VEW¹⁸⁶), there was an uneven distribution of polar (Tyr402, Asn407, His180, Arg348 and Lys351) and hydrophobic (Leu401 and Val405) amino acid residues (Table 4.3, Figure 4.10B). The active site of the other structure (PDB ID: 6B3J¹⁹¹) examined had the majority of hydrophobic residues present (Ile357, Phe393, Ile356, Pro358, Phe390, Met397) while the remaining residues (His363, Gln394) were polar (Table 4.3, Figure 4.10C). The presence of polar residues in the active sites of the structures examined implies a critical hydrogen bonding between the ligand and the residues.¹⁹¹

An earlier study carried out by Gong *et al.*²⁰⁶ reported that compounds which possess a 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridine-2-yl)phenyl acetate moiety, including 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridine-2-yl) phenylcyclo propane carboxylate, are selective GLP-1R agonists, and have potential as anti-diabetic treatment agents.

In the active structure (PDB ID: 5VAI¹⁸⁷), seven residues were found in the 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate binding site; the majority of the residues were hydrophobic except Thr355 (Figure 4.12A). The docking result of the inactive structure (PDB ID: 5VEW¹⁸⁶) showed the majority of the residues found in the active site being hydrophobic (Table 4.4, Figure 4.12B), His363 and Gln394 were the only polar residues identified in the binding site. Examination of the docking results with the crystal structure 6B3J (PDB ID: 6B3J¹⁹¹) showed that the residues interacting with the ligand were hydrophobic (Table 4.4, Figure 4.12C).

A study carried out by Gong *et al.*²⁰⁶ which screened heterocyclic small molecules, they discovered a novel skeleton of 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridine-2-yl)phenyl acetate derivative (3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate), which on further characterisation proved to be selective GLP-1R agonists which potentially had a therapeutic effect on diabetes. In a later study, performed by Bueno *et al.*¹⁹⁰, their findings inferred that 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate failed to activate GLP-1R through a covalent mechanism as it remained unaltered after 2 hours and thus was considered non-reactive. An analysis of the top three binding poses of the all the protein-ligand complexes showed that the residues interacting with the ligands were slightly different albeit docking at TM6 for most of the poses (Tables 4.1-4.4, Appendix 2-9 [Tables

S3-S10]). This would suggest a flexibility of the ligand within the binding pocket of the receptor.

The binding affinity of any complex in thermodynamic terms is crucial in determining the feasibility of an interaction occurring in a cell or not at specified conditions.^{195, 197} Hence, the binding affinity of the docked complexes was analysed using the PRODIGY webserver. The binding affinity (ΔG_{noelec}) values for the 5VAI-2,6,7-trichloro-3-(trifluoromethyl)quinoxaline, 5VEW-2,6,7-trichloro-3-(trifluoromethyl)quinoxaline, 6B3J-2,6,7-trichloro-3-(trifluoromethyl)quinoxaline, 5VAI-1-(5-(4-(tert-butyl)phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one, 5VEW-1-(5-(4-(tert-butyl)phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one, 6B3J-1-(5-(4-(tert-butyl)phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one, 5VAI-2-((4-chlorophenyl)thio)-3-(trifluoromethyl)quinoxaline, 5VEW-2-((4-chlorophenyl)thio)-3-(trifluoromethyl)quinoxaline, 6B3J-2-((4-chlorophenyl)thio)-3-(trifluoromethyl)quinoxaline, 5VAI-3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate, 5VEW-3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate, 6B3J-3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate were revealed to be -6.8 Kcal/mol, -7.2 Kcal/mol, -7.2 Kcal/mol, -8.5 Kcal/mol, -8.1 Kcal/mol, -8.8 Kcal/mol, -7.5 Kcal/mol, -7.9 Kcal/mol, -7.9 Kcal/mol, -8.5 Kcal/mol, -8.6 Kcal/mol and -9.2 Kcal/mol, respectively (Table 4.5). The results showed that the dockings were feasible energetically; this was demonstrated by the negative Gibbs free energy (ΔG) values (Table 4.5). The findings showed that 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate-protein structure (PDB IDs: 5VAI¹⁸⁷, 6B3J¹⁹¹, 5VEW¹⁸⁶) complexes had the highest predicted binding affinities. This suggests that 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate would be bound tightly to the allosteric binding site of the receptor. The findings in Table 4.5, also show that 1-(5-(4-(tert-butyl)phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one and 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate would be bound strongly to the allosteric binding site of the GLP-1R.

The *in silico* prediction of ADME/Tox properties of any new drug candidate is essential in drug development; thus allowing scientists to investigate its properties to

optimise the drug candidates to acceptable ADME/Tox standards.²⁰⁷ The findings in Table 4.6 showed that the allosteric modulators complied with Lipinski's rule of five except 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate. According to the ADMET Predictor software, 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate violated the rule because of a high logP value of over 5.6. This translates to 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate not being a likely drug candidate according to the traditional method of evaluating drug-likeness; Lipinski's rule of 5. The molecular descriptors HBA and HBD were found to comply with the cut off limits of Ro5 (Table 4.6). The ligands evaluated in this study complied with the Veber drug-likeness filter (rotatable bonds ≤ 10 , TPSA ≤ 140).²⁰⁸ Based on the rule of three for fragment-based drug discovery (molecular weight < 300 , ClogP < 3 , the number of hydrogen bond donors and acceptors < 3 and the number of rotatable bonds < 3), the ligands violated all the rules.²⁰⁹ Lipophilicity is often expressed as the distribution coefficient in water/octanol (logD); this parameter influences some processes like plasma protein binding, oral absorption and VD.²¹⁰ Nevertheless, higher logD values translate to higher vulnerability to P450 metabolism leading to higher clearance²¹⁰. The predicted logD values presented in Table 4.5 showed 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate with the highest logD value while 2,6,7-trichloro-3-(trifluoromethyl) quinoxaline had the least. However, all the ligands also possessed high logD values of over 3.5 which leads to low aqueous solubility. This makes these ligands potentially promiscuous as high lipophilicity often leads to low metabolic clearance and toxicity. The predicted human jejunal permeability (P_{eff}) values (Table 4.6) shows that the allosteric modulators had values of over $1.5 \text{ cm/s} \times 10^4$. This indicates that the allosteric modulators would be absorbed entirely. This finding is corroborated by previous research, which reports that drug candidates with a P_{eff} value of >1.5 would be wholly absorbed irrespective of transport mechanism(s) being utilised.²¹¹

The volume of distribution (VD) is essential in ADME studies²¹². It relates to the amount of a drug in the body to the measured concentration in a suitable biological fluid.²¹² The findings for the VD parameter (Table 4.6) showed 1-(5-(4-(tert-butyl)phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one having the lowest value (0.86 L/kg) while 2,6,7-trichloro-3-(trifluoromethyl) quinoxaline had the highest. Hassan *et al.* reported that VD values of $<5.5 \text{ L/kg}$ guaranteed decreased deep tissue penetration.²¹³ Nonetheless, compounds that enter tissues and bind extensively

will show VD above the total body water (i.e. any value greater than 1 L/kg).²¹⁴ The allosteric modulators would remain in the bloodstream for an extended period, thus exerting their effects over a longer period due to their predicted high plasma protein binding (% unbound) of over 90% (Table 4.6). The ADME/Tox software predicted high BBB penetration for all the allosteric modulators investigated. This implies that the allosteric modulators investigated may potentially treat ailments affecting the brain (Table 4.6). The predicted TPSA values (Table 4.6) shows 2,6,7-trichloro-3-(trifluoromethyl)quinoxaline, 2-((4-chlorophenyl)thio)-3-(trifluoromethyl)quinoxaline, and 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate with TPSA values of less than 60 Å², while the compound 2 TPSA value was above 60 Å² but less than 140 Å². In 2009, Fernandes and Gattass²¹⁵ reported that molecules with PSA values greater than 140 Å² are believed to have low cell membrane penetrating capacity, while those with PSA ≤ 60 Å² are easily absorbed.

All the allosteric modulators investigated were not substrates of P-glycoprotein (P-gp); 2-((4-chlorophenyl)thio)-3-(trifluoromethyl)quinoxaline and 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate were identified as inhibitors of P-gp (Table 4.7). Even though most compounds do not inhibit P-gp, there is a possibility that the allosteric modulators can be transported out of the cell by it.²¹⁶ The allosteric modulators were not inhibitors of organic anion transporting polypeptide 1B1 (OATP1B1), but the prediction showed that the allosteric modulators were inhibitors of organic cation transporter 2 (OCT2) (Table 4.7). All the ligands inhibited bile salt export pump (BSEP) except 2,6,7-trichloro-3-(trifluoromethyl)quinoxaline, while the rest were substrates of breast cancer resistance protein (BCRP) except the allosteric modulator 2-((4-chlorophenyl)thio)-3-(trifluoromethyl)quinoxaline. Cytochrome P450 (CYP450) are the enzymes that catalyse the oxidation of organic substances.²¹³ The evaluation of the allosteric modulators on the various isoforms of hepatic CYP450 was also predicted; the findings are shown in Table 4.8. Table 4.8 demonstrates that all the allosteric modulators were substrates of CYP3A4, albeit 2-((4-chlorophenyl)thio)-3-(trifluoromethyl)quinoxaline was an isoform inhibitor. Allosteric modulators showed inhibitory tendencies and substrate specificities for the various isoforms investigated (Table 4.8). The ADMET predictor software inferred that the clearance pathway for the allosteric modulators studied is via metabolism. Drug toxicity has resulted in the failure of drug candidates in clinical trials, hence, the use of in silico models to predict the potential toxicity of new drug candidates.²¹²

The main aim of a toxicology study in drug discovery is to evaluate the safety of potential drug candidates.²¹⁷ Preclinical toxicity testing is carried out to determine a drug candidate's organ, dose and species-specific toxic effects.^{218, 219} The predicted toxicity parameters were shown in Table 4.9; all allosteric modulators evaluated reported negative for AMES toxicity. The hERG filter parameter returned negative, implying that the allosteric modulators do not have an affinity for the hERG potassium channel in humans. All the allosteric modulators studied could cause potential reproductive/developmental toxicity except 2,6,7-trichloro-3-(trifluoromethyl)quinoxaline (Table 4.9). The allosteric modulators were all skin sensitizers except 2-((4-chlorophenyl)thio)-3-(trifluoromethyl)quinoxaline (Table 4.9). The liver function parameters (Table 4.9) showed some of the allosteric modulators causing elevations in levels of liver enzymes studied. 2,6,7-trichloro-3-(trifluoromethyl)quinoxaline and 1-(5-(4-(tert-butyl)phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one elevated the levels of the liver functions studied (Table 4.9), resulting in them being classed as hepatotoxic by the prediction software.

4.4 Evaluation of the allosteric binding site of the glucagon-like peptide receptor: ligand docking and ADMET prediction studies.

4.4.1 Results and Discussion

Presently, there is an extensive treatment option for T2D ranging from biguanides (metformin), meglitinides, alpha-glucosidase inhibitors, and the current therapeutic options mimic GLP-1R agonists. For instance, exenatide, albiglutide, liraglutide, lixisenatide. The major challenge faced by these therapeutic options is the degradation of GLP-1 by DPP-IV. GLP-1R agonists and other drugs which target GPCRs bind on the orthosteric binding site.^{220, 221} Molecular docking is a virtual screening method used to discover new ligands for receptors (e.g., GPCRs). It ranks the active molecules higher and produces poses which informs on the decisions to be taken in the process of drug discovery.^{222, 223} Previous studies have shown the application of molecular docking on a large scale in the drug development process.^{168, 199, 200, 203, 204, 222} In the current study, the interaction of three known GLP-1R allosteric modulators and an electrophilic compound was shown to modulate GLP-1R activity to three different GLP1-1R crystal structures to understand the allosteric binding site better. In addition, the findings of an *in silico* ADME/Tox evaluation of the ligands used in the study was also presented.

Interaction of 2-((6,7-dichloro-3-(trifluoromethyl) quinoxalin-2-yl) thio)-5-methyl-1,3,4-thiadiazole with GLP – 1 receptor

A molecular docking research was carried out to understand the allosteric binding site of GLP-1R using known allosteric modulators and an electrophilic compound that is known to modulate the activity of the receptor. The results showed that the ligand 2-((6,7-dichloro-3-(trifluoromethyl) quinoxalin-2-yl) thio)-5-methyl-1,3,4-thiadiazole was bound behind TM6 (-7.979 Kcal/mol) in the active structure with PDB ID: 5VAI¹⁸⁷. The non-polar residues Leu251, Leu359, Val331 formed alkyl/pi-alkyl interactions with 2-((6,7-dichloro-3-(trifluoromethyl) quinoxalin-2-yl) thio)-5-methyl-1,3,4-thiadiazole. The residue Leu356 formed a hydrogen bond with 2-((6,7-dichloro-3-(trifluoromethyl) quinoxalin-2-yl) thio)-5-methyl-1,3,4-thiadiazole. The polar residues interacting with the ligand, formed hydrogen bonds (Figure 4.13A). In the other active structure (PDB ID: 6B3J¹⁹¹), the ligand was juxtaposed at TM6 and TM7 (-7.7 Kcal/mol). 2-((6,7-dichloro-3-(trifluoromethyl) quinoxalin-2-yl) thio)-5-methyl-1,3,4-thiadiazole formed a halogen bond interaction with Leu354. The other amino acid residues interacting with the ligand were non-polar, they formed hydrogen, pi-sulphur, p-pi stacked, alkyl and pi-alkyl interactions (Figure 4.13C). While in the inactive structure (PDB ID: 5VEW¹⁸⁶), the ligand was juxtaposed at TM6 and TM7 of the receptor (-9.843 Kcal/mol), the residues Val405 and Asn407 formed a halogen bond interaction with the fluorine atom on 2-((6,7-dichloro-3-(trifluoromethyl) quinoxalin-2-yl) thio)-5-methyl-1,3,4-thiadiazole (Figure 4.13B). A summary table of all residues interacting with the ligand can be found below (Table 4.10).

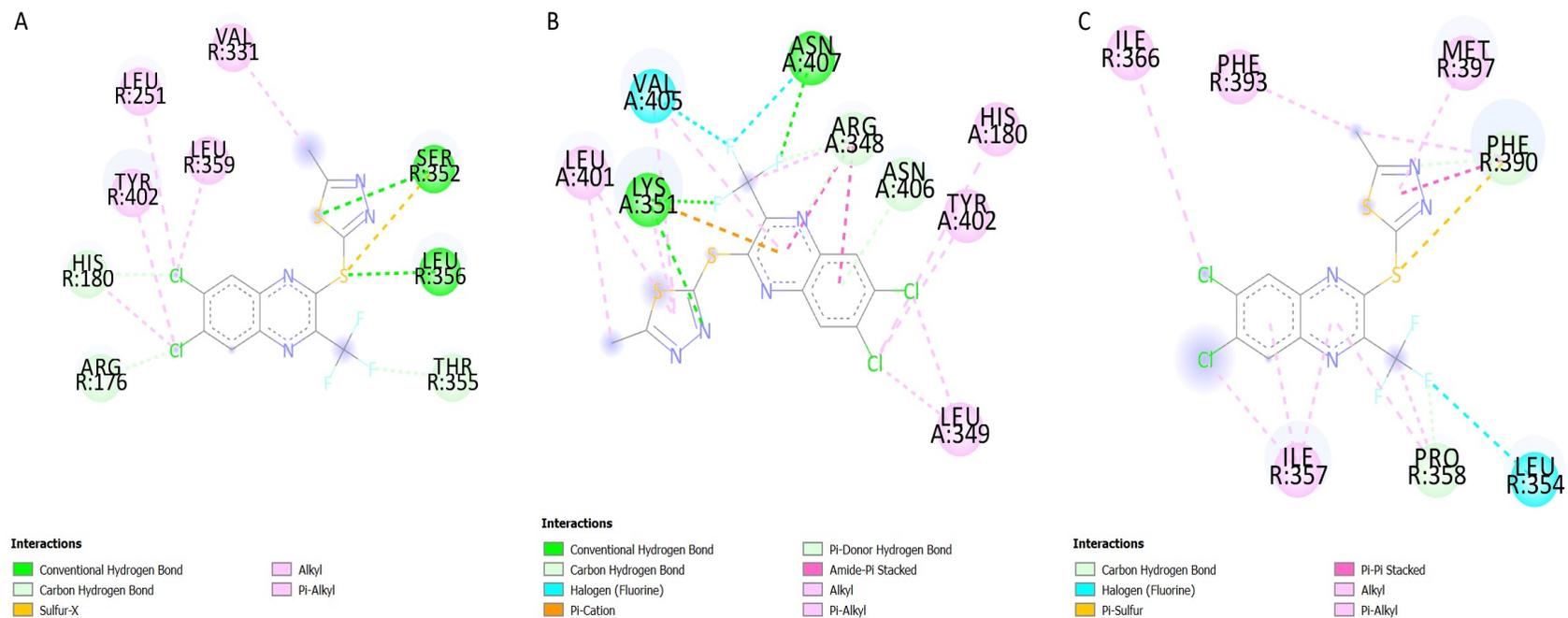


Figure 4.13. Schematic image of the protein-2-((6,7-dichloro-3-(trifluoromethyl) quinoxalin-2-yl) thio)-5-methyl-1,3,4-thiadiazole interactions occurring at the GLP-1R binding site. (A) 1:5VAI, (B) 1:5VEW, (C) 1:6B3J.

Table 4.10. Amino acids present in the 2-((6,7-dichloro-3-(trifluoromethyl) quinoxalin-2-yl) thio)-5-methyl-1,3,4-thiadiazole (Compound 1) binding site.

5VAI	5VEW	6B3J
VAL 331	VAL 405	ILE 366
SER 352	ASN 407	PHE 393
LEU 356	ARG 348	MET 397
THR 355	ASN 406	PHE 390
ARG 176	HIS 180	ILE 357
HIS 180	TYR 402	PRO 358
TYR 402	LEU 349	LEU 364
LEU 359	LEU 401	
LEU 251	LYS 351	

Interaction of 5-(6,7-dichloro-3-((1-(1-(1-methylpiperidin-4-yl) ethyl)-1H-tetrazol-5-yl) thio) quinoxalin-2-yl) thiazol-2-ol with GLP – 1 receptor

The findings of the docking study showed 5-(6,7-dichloro-3-((1-(1-(1-methylpiperidin-4-yl) ethyl)-1H-tetrazol-5-yl) thio) quinoxalin-2-yl) thiazol-2-ol juxtaposed at TM6 (-8.739 Kcal/mol) in the active structure with PDB ID: 5VAI¹⁸⁷. Hydrogen bond interaction was observed with Leu354 (Figure 4.14A). The residue Phe390 had a pi-pi t-shaped interaction with 5-(6,7-dichloro-3-((1-(1-(1-methylpiperidin-4-yl) ethyl)-1H-tetrazol-5-yl) thio) quinoxalin-2-yl) thiazol-2-ol (Figure 4.13A). In the other active structure (PDB ID: 6B3J¹⁹¹) (-9.979 Kcal/mol) investigated, 5-(6,7-dichloro-3-((1-(1-(1-methylpiperidin-4-yl) ethyl)-1H-tetrazol-5-yl) thio) quinoxalin-2-yl) thiazol-2-ol was docked at TM6 and TM7; the ligand formed a hydrogen bonds with Ser352, Val405, and Arg176 (Figure 4.14C). In the inactive structure (PDB ID: 5VEW¹⁸⁶) (-8.946 Kcal/mol), 5-(6,7-dichloro-3-((1-(1-(1-methylpiperidin-4-yl) ethyl)-1H-tetrazol-5-yl) thio) quinoxalin-2-yl) thiazol-2-ol was juxtaposed at TM6; 5-(6,7-dichloro-3-((1-(1-(1-methylpiperidin-4-yl) ethyl)-1H-tetrazol-5-yl) thio) quinoxalin-2-yl) thiazol-2-ol formed hydrogen bond interactions with the following residues; Asn406, Arg348, Val405 and Lys351 (Figure 4.14B). Table 4.11 summarises all residues interacting with the ligand.

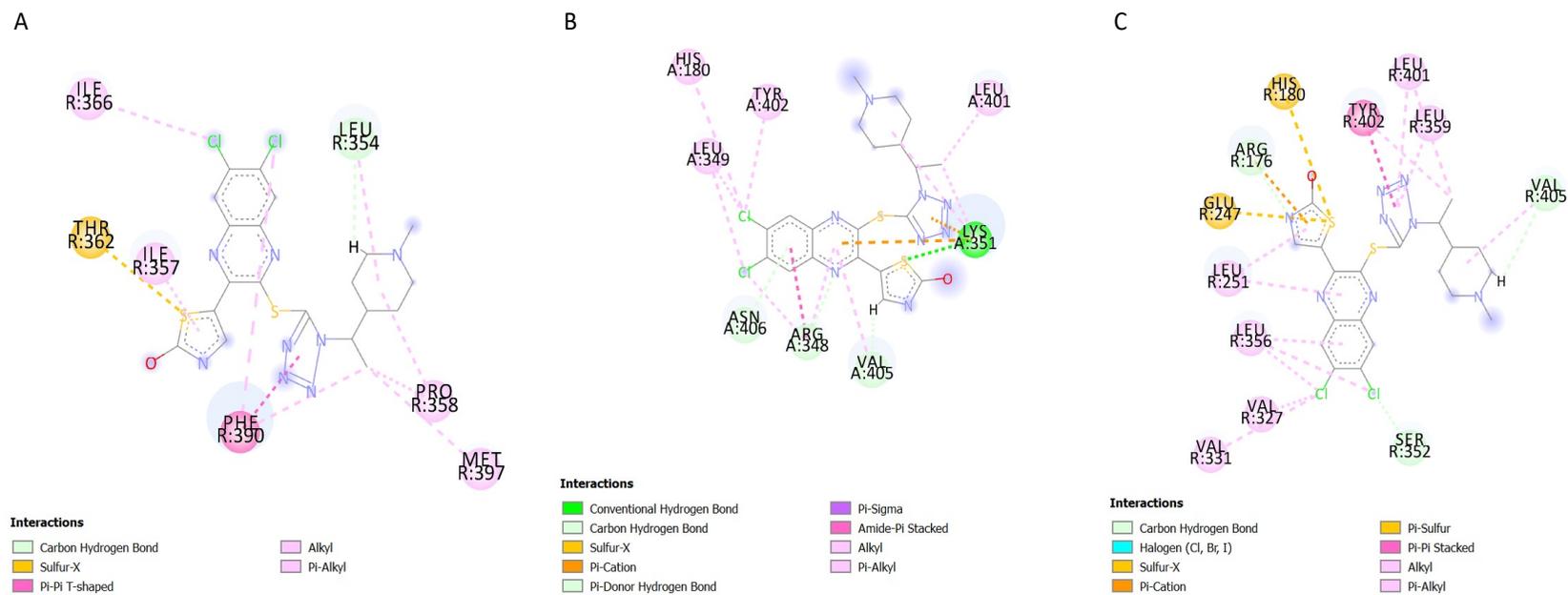


Figure 4.14. Schematic image protein-5-(6,7-dichloro-3-((1-(1-(1-methylpiperidin-4-yl) ethyl)-1H-tetrazol-5-yl) thio) quinoxalin-2-yl) thiazol-2-ol interactions occurring at the GLP-1R binding site. (A) 2:5VAI, (B) 2:5VEW, (C) 2:6B3J.

Table 4.11. Amino acids present in the 5-(6,7-dichloro-3-((1-(1-(1-methylpiperidin-4-yl) ethyl)-1H-tetrazol-5-yl) thio) quinoxalin-2-yl) thiazol-2-ol binding site.

5VAI	5VEW	6B3J
ILE 366	HIS 180	HIS 180
LEU 354	TYR 402	TYR 402
PRO 358	LEU 349	LEU 401
MET 397	LEU 401	LEU 359
PHE 390	LYS 351	VAL 405
ILE 357	ASN 406	SER 352
THR 362	ARG 348	VAL 331
	VAL 405	VAL 327
		LEU 356
		LEU 251
		GLU 247
		ARG 176

Interaction of 1-([1,1'-biphenyl]-4-carbonyl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one with GLP – 1 receptor

The results of the docking study showed 1-([1,1'-biphenyl]-4-carbonyl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one juxtaposed at TM6 and TM7 in the active structure with PDB ID: 5VAI¹⁸⁷ (-7.474 Kcal/mol). The residue Lys351 formed a hydrogen bond, pi-cation, and an amide-pi cation interaction with 1-([1,1'-biphenyl]-4-carbonyl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one. The residues Pro358, Leu354, and Leu401 formed alkyl and pi-alkyl interactions with the ligand (Figure 4.15A). In the inactive structure (PDB ID: 5VEW¹⁸⁶), 1-([1,1'-biphenyl]-4-carbonyl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one is juxtaposed at TM6 (-7.88 Kcal/mol). The residue Gln394, formed a hydrogen bond with 1-([1,1'-biphenyl]-4-carbonyl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one. 1-([1,1'-biphenyl]-4-carbonyl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one also formed pi-alkyl interactions with His363 and Pro358 (Figure 4.15B). In the other active structure (PDB ID: 6B3J¹⁹¹) (-8.882 Kcal/mol) investigated, 1-([1,1'-biphenyl]-4-carbonyl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one juxtaposed at TM6 and TM7; carbon-hydrogen bonds were formed with Gln394 and Phe390. Phe390 also formed a pi-sulphur interaction 1-([1,1'-biphenyl]-4-carbonyl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-

dihydrobenzo[c]thiophen-4(5H)-one (Figure 4.15C). Table 4.12 shows a summary of all residues interacting with the ligand.

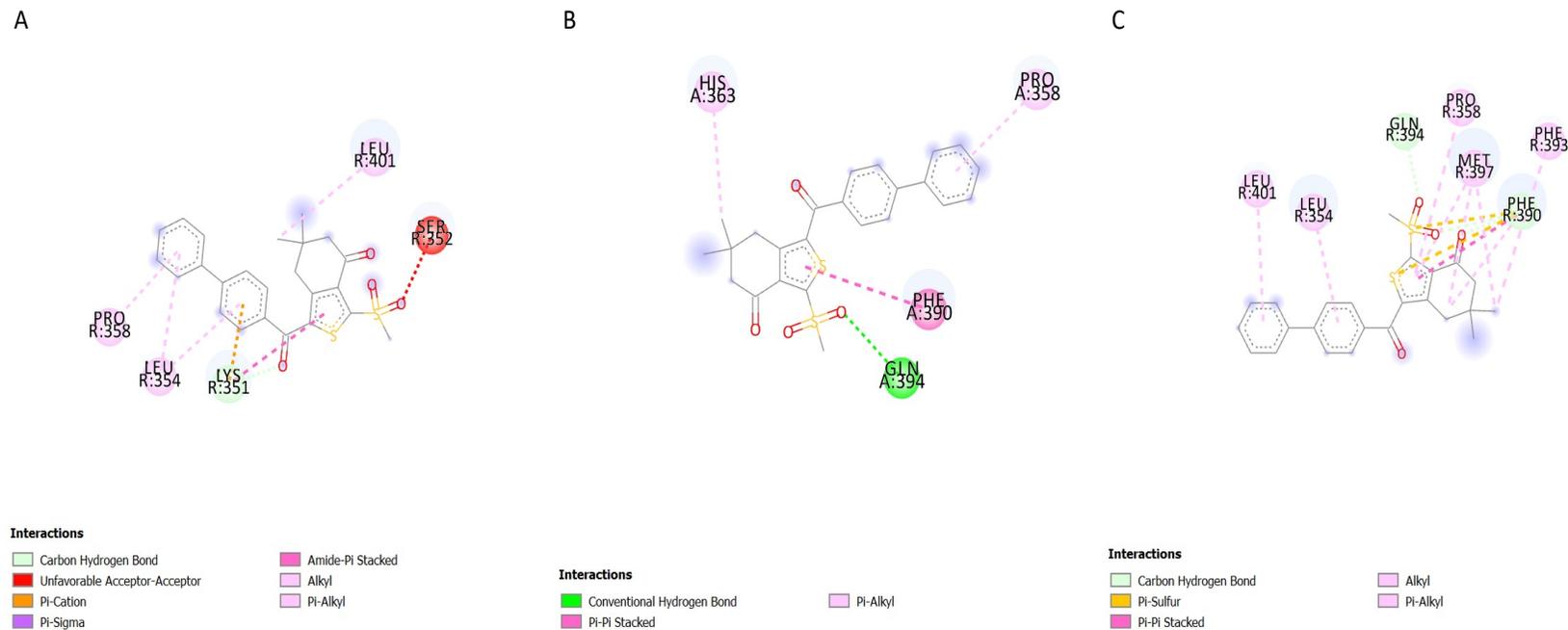


Figure 4.15. Schematic showing representation of protein-1-([1,1'-biphenyl]-4-carbonyl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one interactions occurring at the GLP-1R binding site. (A) 3:5VAI, (B) 3:5VEW, (C) 3:6B3J.

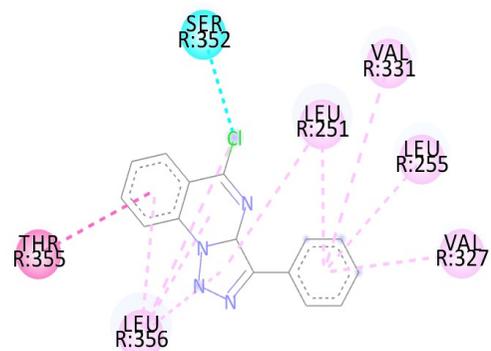
Table 4.12. Amino acids present in the 1-([1,1'-biphenyl]-4-carbonyl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one binding site.

5VAI	5VEW	6B3J
LEU 401	HIS 363	LEU 401
SER 352	PRO 358	LEU 354
LYS 351	PHE 390	GLN 394
LEU 354	GLN 394	PRO 358
PRO 358		MET 397
		PHE 390
		PHE 393

Interaction of 5-chloro-3-phenyl- [1,2,3] triazolo[1,5-a] quinazoline with GLP – 1 receptor

The findings of the docking study showed that 5-chloro-3-phenyl- [1,2,3] triazolo[1,5-a] quinazoline docked behind TM6 (-5.802 Kcal/mol) in the active structure with PDB ID: 5VAI¹⁸⁷. The residue Ser352 formed a halogen bond with the chlorine atom on 5-chloro-3-phenyl- [1,2,3] triazolo[1,5-a] quinazoline. An amide-pi stacked interaction was formed with the residue Thr355, whilst the other residues interacting with the ligand formed alkyl and pi-alkyl interactions with the ligand (Figure 4.16A). In the other active structure (PDB ID: 6B3J¹⁹¹) investigated (-6.475 Kcal/mol), 5-chloro-3-phenyl- [1,2,3] triazolo[1,5-a] quinazoline docked at TM6 and TM7. The residue Met397 formed a halogen bond interaction with the chlorine atom on 5-chloro-3-phenyl- [1,2,3] triazolo[1,5-a] quinazoline. Alkyl and pi-alkyl interactions were also formed with the residue Met397 and 5-chloro-3-phenyl- [1,2,3] triazolo[1,5-a] quinazoline. A pi-sigma interaction was formed with the Pro358 and compound 4 (Figure 4.16C). In the inactive structure (PDB ID: 5VEW¹⁸⁶) (-6.171 Kcal/mol), compound 4 is juxtaposed at TM6 and TM7. The residue Asn407 formed a hydrogen bond with 5-chloro-3-phenyl- [1,2,3] triazolo[1,5-a] quinazoline. A pi-cation interaction was formed with the residue Lys351 (Figure 4.16B). Table 4.13 shows a summary of all residues interacting with the ligand.

A

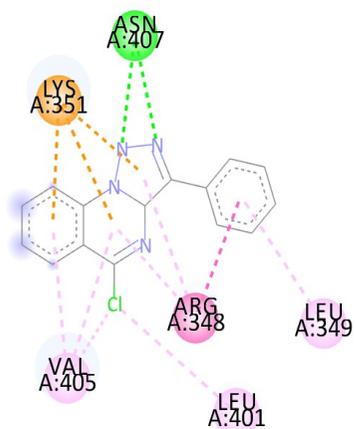


Interactions

Halogen (Cl, Br, I)
Amide-PI Stacked

Alkyl
PI-Alkyl

B

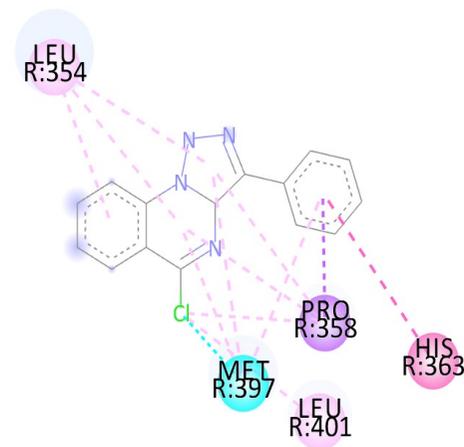


Interactions

Conventional Hydrogen Bond
PI-Cation
Amide-PI Stacked

Alkyl
PI-Alkyl

C



Interactions

Halogen (Cl, Br, I)
PI-Sigma
PI-PI Stacked

Alkyl
PI-Alkyl

Figure 4.16. Schematic image protein-5-chloro-3-phenyl-[1,2,3] triazolo[1,5-a] quinazoline interactions occurring at the GLP-1R binding site. (A) 4:5VAI, (B) 4:5VEW, (C) 4:6B3J.

Table 4.13. Amino acids present in the 5-chloro-3-phenyl- [1,2,3] triazolo[1,5-a] quinazoline binding site.

5VAI	5VEW	6B3J
SER 352	ASN 407	LEU 354
LEU 251	LYS 351	PRO 358
VAL 331	VAL 405	HIS 363
LEU 255	ARG 348	LEU 401
VAL 327	LEU 401	MET 397
LEU 356	LEU 349	
THR 355		

Bueno *et al.*¹⁹⁰ reported that 2-((6,7-dichloro-3-(trifluoromethyl)quinoxalin-2-yl)thio)-5-methyl-1,3,4-thiadiazole, compound 2, and compound 3 potentiated the activity of GLP-1(9–36)-NH₂ on the wild-type receptor but failed to exert the same effect in cells that expressed the mutant GLP-1R, which lacks the cysteine-347 residue. Knudsen *et al.*²²⁴ reported that 2-((6,7-dichloro-3-(trifluoromethyl)quinoxalin-2-yl)thio)-5-methyl-1,3,4-thiadiazole does not activate closely related receptors within family (class) B GPCRs (Glucagon-like peptide-2 (GLP-2), glucagon, and Glucose-dependent insulinotropic polypeptide (GIP) receptors).

The presence of polar residues in the binding sites of the three structures examined suggests that these polar residues play a vital role in ligand binding.¹⁹¹ The presence of these polar residues in the crystal structures of the receptor indicates the presence of a significant polar binding network around the peptide bindings site.¹⁹¹ A study carried out by Song *et al.*¹⁸⁶ and Xu *et al.*¹⁹² showed that the cysteine-347^{6.36bF} (C347^{6.36bF}) mutant maximises van der Waals interactions with all the three negative allosteric modulators (NAMs) PF-06372222, NNC0640 and MK-0893 through the stabilisation of the aliphatic side chain of Lysine-351^{6.40b} (K351^{6.40b}) in an optimal conformation for hydrophobic interactions with NAMs.^{186, 192} The findings from a research carried out by Song *et al.*, showed that positions 352-355 play a crucial role in binding an allosteric inhibitor to GLP-1R.¹⁸⁶ The residue Thr 355 was also observed to interact with 2-((6,7-dichloro-3-(trifluoromethyl)quinoxalin-2-yl)thio)-5-methyl-1,3,4-thiadiazole and 5-chloro-3-phenyl- [1,2,3] triazolo[1,5-a]

quinazoline upon docking into the active structure 5VAI. This residue (Thr 355) was also observed to interact with allosteric modulators in a previous study forming hydrogen bond interactions with the ligands.²²²

***In silico* ADME/Tox prediction**

The webservers ADMETLab and SwissADME developed by Dong *et al.*⁷⁸ and Daina *et al.*⁸³ used in this study are online web resources that predict ADME/Tox properties of potential drug candidates. Poor ADME/Tox parameters account for the failure of new drug candidates in clinical trials; therefore, it is of utmost importance to evaluate these parameters to decrease the potential risks.²²⁵ The results of the *in silico* ADME/Tox evaluation of the ligands used in the study are reported in the tables (4.14, 4.15 and 4.16). Table 4.14 highlights the following parameters: octanol-water partition coefficient based on Moriguchi's model (logP), octanol-water partition coefficient as a function of pH (logD), intrinsic aqueous solubility (logS), Lipinski's rule of 5 (Ro5), topological polar surface area (TPSA), hydrogen bond acceptors (HBA), hydrogen bond donors (HBD), and molecular weight (MW).

The allosteric modulators investigated in this study satisfied Lipinski's rules of 5. This includes log P-value lower than 5, less than 10 HBAs, less than 5 HBDs, molecular weight lower than 500 g/mol, and the number of rotatable bonds less than ten except for the allosteric modulator 5-(6,7-dichloro-3-((1-(1-(1-methylpiperidin-4-yl)ethyl)-1H-tetrazol-5-yl)thio)quinoxalin-2-yl)thiazol-2-ol, which has a molecular weight of over 500 g/mol (523.46 g/mol). At the same time, the other ligands studied met the drug-likeness criteria (Table 4.14). The allosteric modulators 2-((6,7-dichloro-3-(trifluoromethyl)quinoxalin-2-yl)thio)-5-methyl-1,3,4-thiadiazole and 1-([1,1'-biphenyl]-4-carbonyl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one have higher molecular weight values than 5-chloro-3-phenyl- [1,2,3] triazolo[1,5-a] quinazoline, potentially affecting the overall absorption of the allosteric modulators in the body.^{225, 226} TPSA has been widely accepted as a good index of drug absorption in the small intestine (less than 140 Å²) and blood-brain barrier penetration (less than 60 Å²).²²⁷ The computational TPSA values predicted for the allosteric modulators were less than 140 Å² which means they have good intestinal absorption except compound 2 (159.28 Å²) (Table 4.14). However, 5-chloro-3-phenyl- [1,2,3] triazolo[1,5-a] quinazoline showed adequate blood-brain barrier penetration, as its TPSA value (43.08 Å²) is less than 60 Å² (Table 4.14).

Table 4.14. Physicochemical properties of the selected allosteric modulators.

Lig.	LogP	LogD	LogS	HBA	HBD	MW	TPSA	Ro5
1	3.20	2.802	-6.91	7	0	397.23 g/mol	105.10 Å ²	0
2	2.95	1.56	-7.74	8	1	523.46 g/mol	159.28 Å ²	1
3	3.20	2.742	-7.37	4	4	438.56 g/mol	104.90 Å ²	0
4	3.55	2.809	-4.34	3	0	280.71 g/mol	43.08 Å ²	0

Key: 1: 2-((6,7-dichloro-3-(trifluoromethyl)quinoxalin-2-yl)thio)-5-methyl-1,3,4-thiadiazole, 2: 5-(6,7-dichloro-3-((1-(1-(1-methylpiperidin-4-yl)ethyl)-1H-tetrazol-5-yl)thio)quinoxalin-2-yl)thiazol-2-ol, 3: 1-([1,1'-biphenyl]-4-carbonyl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one, 4: 5-chloro-3-phenyl-[1,2,3]triazolo[1,5-a]quinazoline.

Lipophilicity is an essential parameter in the process of drug development.²²⁸ In drug development, high lipophilicity has been shown to contribute to attrition.²²⁸⁻²³¹ The determination of partition coefficient (LogP) and its pH-dependent variant, distribution (LogD), is a critical aspect of drug development. These measurements hold substantial importance in the quest for enhanced and optimized drug formulations.²²⁸ The LogP values of the compounds were between 2.95 and 3.55 (Table 4.14) with 5-chloro-3-phenyl- [1,2,3] triazolo[1,5-a] quinazoline having the highest LogP value. This is an indication that the ligands used in this study would likely have optimal physicochemical and ADME properties if they were to be oral drugs because their predicted LogP value is greater than 1 but less than 4.²³² According to Landry and Crawford (2020), the optimal lipophilicity of drugs is within the range of LogP = 1-3.²²⁸ LogD values of over 3.5 often leads to low aqueous solubility, thus making the potential drug candidate promiscuous.²¹⁰ However, this was not the case for the ligands studied in this research. The predicted LogD values for the ligands showed that 5-chloro-3-phenyl- [1,2,3] triazolo[1,5-a] quinazoline has the highest LogD value at 2.809 while compound 2 had the lowest (Table 4.15).

The findings of the pharmacokinetic parameters such as gastrointestinal absorption (GI Abs), blood-brain barrier permeation (BBB), P-glycoprotein substrate (Pgp), isoforms of cytochrome P450, and skin permeation (Log K_p) predicted using swissADME are presented in Tables 4.15 and 4.16. The allosteric modulators studied had low gastrointestinal absorption and no blood-brain barrier permeation (BBB) except 5-chloro-3-phenyl- [1,2,3]

triazolo[1,5-a] quinazoline. The allosteric modulators were not substrates of P-glycoprotein, as shown in Table 4.15. The prediction results showed 2-((6,7-dichloro-3-(trifluoromethyl)quinoxalin-2-yl)thio)-5-methyl-1,3,4-thiadiazole and 5-(6,7-dichloro-3-((1-(1-(1-methylpiperidin-4-yl)ethyl)-1H-tetrazol-5-yl)thio)quinoxalin-2-yl)thiazol-2-ol as inhibitors of the CYP isoforms; CYP1A2, CYP2C19, CYP2C9, and non-inhibitors of the isoforms CYP2D6 and CYP3A4. The allosteric modulator 1-([1,1'-biphenyl]-4-carbonyl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one inhibits CYP2C19, CYP2C9, and CYP3A4. Furthermore, the prediction results suggested that 1-([1,1'-biphenyl]-4-carbonyl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one is a non-inhibitor of CYP1A2 and CYP2D6. 5-chloro-3-phenyl- [1,2,3] triazolo[1,5-a] quinazoline inhibits CYP1A2 and CYP2C19, while the prediction results identified it as non-inhibitors of CYP2C9, CYP2D6, and CYP3A4.

Table 4.15. *In silico* prediction of the pharmacokinetic properties of the ligands.

Lig.	GI Abs	BBB	Pgp	CYP1A2	CYP2C19	CYP2C9	CYP2D6	CYP3A4	Log K _p
1	Low	No	No	+	+	+	-	-	-5.20 cm/s
2	Low	No	No	+	+	+	-	-	-6.18 cm/s
3	Low	No	No	-	+	+	-	+	-5.13 cm/s
4	High	Yes	No	+	+	-	-	-	-5.36 cm/s

Key: Inhibitor of CYP isoform: (+), non – inhibitor of CYP isoform: (-), 1: 2-((6,7-dichloro-3-(trifluoromethyl)quinoxalin-2-yl)thio)-5-methyl-1,3,4-thiadiazole, 2: 5-(6,7-dichloro-3-((1-(1-(1-methylpiperidin-4-yl)ethyl)-1H-tetrazol-5-yl)thio)quinoxalin-2-yl)thiazol-2-ol, 3: 1-([1,1'-biphenyl]-4-carbonyl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one, 4: 5-chloro-3-phenyl-[1,2,3]triazolo[1,5-a]quinazoline.

The findings of the excretion and toxicology parameters evaluated are presented in Table 4.16. The parameters investigated include half-life (T_{1/2}), clearance (CL), hERG blocker, Ames mutagenicity (AMES), skin sensitisation (Skin Sen), LD₅₀ of acute toxicity (LD₅₀), and drug-induced liver injury (DILI). The findings showed that the half-life and clearance values of the allosteric modulators were below 3h and 5 mL/min/kg, respectively, according to Dong *et al.*⁷⁸. The result for hERG blocking showed the allosteric modulators 2-((6,7-dichloro-3-(trifluoromethyl)quinoxalin-2-yl)thio)-5-methyl-1,3,4-thiadiazole and 5-

chloro-3-phenyl- [1,2,3] triazolo[1,5-a] quinazoline as non – blockers of hERG while other allosteric modulators were blockers of hERG. All the allosteric modulators studied were AMES negative except for the allosteric modulators 5-chloro-3-phenyl- [1,2,3] triazolo[1,5-a] quinazoline. The allosteric modulators studied are non – skin sensitizers. The LD50 figures show that the allosteric modulators are highly toxic as the values fall below 50g/kg, leading to drug-induced liver injury (Table 4.16).

Table 4.16. Excretion and Toxicology parameters

Lig.	T1/2 (h)	CL (mL/min/kg)	hERG	AMES	SkinSen	LD ₅₀ (log mol/kg)	DILI
1	2.026	0.664	-	-	-	4.449	+
2	1.508	1.301	+	-	-	2.866	+
3	1.862	0.989	+	-	-	3.309	+
4	2.188	0.601	-	+	-	2.449	+

Key; HERG Non - blocker: (-), blocker: (+), AMES negative: (-), AMES positive: (+), non-skin sensitizer: (-), skin sensitizer: (+), DILI negative (-), DILI positive (+), 1: 2-((6,7-dichloro-3-(trifluoromethyl)quinoxalin-2-yl)thio)-5-methyl-1,3,4-thiadiazole, 2: 5-(6,7-dichloro-3-((1-(1-(1-methylpiperidin-4-yl)ethyl)-1H-tetrazol-5-yl)thio)quinoxalin-2-yl)thiazol-2-ol, 3: 1-([1,1'-biphenyl]-4-carbonyl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one, 4: 5-chloro-3-phenyl-[1,2,3]triazolo[1,5-a]quinazoline.

4.5 Conclusions and future directions

The glucagon-like peptide 1 receptor (GLP-1R) is a member of the family (or class) B G-protein-coupled receptor (GPCR). The receptor is a regulator of insulin and a key target in treating Type 2 diabetes mellitus. In this chapter, computational chemistry techniques such as molecular docking were combined with *in silico* ADME/Tox predictions to determine the position and structure of the allosteric binding site, as well as to examine how the allosteric modulators bind to the binding site. *In silico* evaluation was used to evaluate the ADME/Tox properties of the allosteric modulators.

The allosteric modulators (2,6,7-trichloro-3-(trifluoromethyl)quinoxaline, 1-(5-(4-(tert-butyl)phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one, 2-((4-chlorophenyl)thio)-3-(trifluoromethyl)quinoxaline, 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-

yl)phenyl cyclohexanecarboxylate, 2-((6,7-dichloro-3-(trifluoromethyl)quinoxalin-2-yl)thio)-5-methyl-1,3,4-thiadiazole, 5-(6,7-dichloro-3-((1-(1-(1-methylpiperidin-4-yl)ethyl)-1H-tetrazol-5-yl)thio)quinoxalin-2-yl)thiazol-2-ol, 1-([1,1'-biphenyl]-4-carbonyl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one, and 5-chloro-3-phenyl-[1,2,3]triazolo[1,5-a]quinazoline) used in the study docked at various positions across TM5, TM6 and TM7 in both the active and inactive structures. Based on these findings, it can be inferred that the allosteric binding site is situated around TM6 of the receptor. Analysis of the top three binding poses of the all the protein-ligand complexes showed that the residues interacting with the ligands were slightly different even though the ligands docked at TM6 for most of the poses (Tables 4.1-4.4, 4.10-4.13, Appendix 2-9 [Tables S3-S10]). This however would suggest a flexibility of the ligand within the binding pocket of the receptor. The results of the research are in line with the findings of a previous study. It was reported that positive and negative allosteric modulators of GLP-1R bind to the same region just outside TM5-TM7 near the intracellular section of the receptor.¹⁸⁶ However, the PAMs bind in a well-defined sub-pocket at the interface between TM5 and TM6, thus facilitating the formation of an intracellular binding site that enhances G-protein coupling.¹⁸⁶ Further mutagenesis studies to understand which residues play a significant role in ligand binding.

The binding analysis showed that the dockings were feasible energetically; this was demonstrated by the negative Gibbs free energy (ΔG) values (Table 4.5). The findings showed that 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate-protein structure (PDB IDs: 5VAI¹⁸⁷, 6B3J¹⁹¹, 5VEW¹⁸⁶) complexes had the highest predicted binding affinities. The binding affinity results suggest that 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate may be an agonist of the receptor, a biased agonist or perhaps an antagonist of the receptor because of its high *in silico* predicted binding affinity. This needs to be further characterised experimentally using assays such as radioligand binding assay or fluorescent ligand binding assay.

The ADMET predictions of the allosteric modulators studied showed the compounds possessing drug-like properties, except for 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate due to its high LogP value 5-(6,7-dichloro-3-((1-(1-(1-methylpiperidin-4-yl)ethyl)-1H-tetrazol-5-yl)thio)quinoxalin-2-yl)thiazol-2-ol, which has a molecular weight of 523.46 g/mol which is over 500 g/mol. The predicted human jejunal permeability (P_{eff}) values (Table 4.6) show that the allosteric modulators had

over 1.5 cm/s x 10⁴. This indicates that the allosteric modulators would be absorbed entirely. The predicted TPSA values (Table 4.6) shows 2,6,7-trichloro-3-(trifluoromethyl)quinoxaline, 2-((4-chlorophenyl)thio)-3-(trifluoromethyl)quinoxaline and 3-(8-chloro-6-(trifluoromethyl)imidazo[1,2-a]pyridin-2-yl)phenyl cyclohexanecarboxylate with TPSA values of less than 60 Å², while the 1-(5-(4-(tert-butyl)phenyl)-1,3,4-oxadiazol-2-yl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one TPSA value was above 60 Å² but less than 140 Å². These findings were corroborated by research carried out by Fernandes and Gattass in 2009.²¹⁵ and Wang *et al.* in 2014.²¹¹ Additionally, the LogP values of some ligands (2-((6,7-dichloro-3-(trifluoromethyl)quinoxalin-2-yl)thio)-5-methyl-1,3,4-thiadiazole, 5-(6,7-dichloro-3-((1-(1-(1-methylpiperidin-4-yl)ethyl)-1H-tetrazol-5-yl)thio)quinoxalin-2-yl)thiazol-2-ol, 1-([1,1'-biphenyl]-4-carbonyl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one, 5-chloro-3-phenyl-[1,2,3]triazolo[1,5-a]quinazoline) evaluated were between 2.95 and 3.55 (Table 4.14). This falls into the optimal lipophilicity of oral drugs (LogP = 1-3 239, LogP = >1 and <4 243). The LogD values of some (2-((6,7-dichloro-3-(trifluoromethyl)quinoxalin-2-yl)thio)-5-methyl-1,3,4-thiadiazole, 5-(6,7-dichloro-3-((1-(1-(1-methylpiperidin-4-yl)ethyl)-1H-tetrazol-5-yl)thio)quinoxalin-2-yl)thiazol-2-ol, 1-([1,1'-biphenyl]-4-carbonyl)-6,6-dimethyl-3-(methylsulfonyl)-6,7-dihydrobenzo[c]thiophen-4(5H)-one, 5-chloro-3-phenyl-[1,2,3]triazolo[1,5-a]quinazoline) ligands were below 3.5, this is an indication that the ligands would be soluble in aqueous solution (Table 4.14). However, the ADME predictions of the other set compounds presented showed that they have high LogD (Table 4.6) values and low aqueous solubility. High lipophilicity often leads low metabolic clearance and toxicity.

Conclusively, docking simulations provide insights into potential allosteric binding sites and possible interactions. The results of an *in silico* ADME/Tox study are essential in developing a potential drug candidate. Further studies could be carried out using known allosteric modulators sourced from a literature search to characterise the allosteric binding pocket of GLP-1R further using molecular dynamics (MD) simulations,²³³⁻²³⁶ and multiscale quantum mechanics (QM)/molecular mechanics (MM) molecular simulations.¹⁷⁰

5. Cytochrome P450 protein-ligand binding study.

This chapter utilized molecular docking, a technique described in Chapter 2 of this thesis, to investigate the interactions between phorbols with different chains at C-12, C-13, and C-20, known aromatase inhibitors (anastrozole, exemestane, and letrozole), and the human placental aromatase cytochrome P450 (CYP19A1). The findings of this chapter have been published in *Applied Biosciences* 2022 1(3), 279-288.

Furthermore, employing computational methods outlined in Chapter 2 of this thesis, this chapter delves into the interactions between selected cytochrome P450s (CYP1A1 and CYP1B1) and specific ligands. These cytochromes play crucial roles in xenobiotics and steroid hormone biosynthesis. The results of this chapter have been published in *Current Pharmaceutical Design* 2022 28(45), 3637-3648.

5.1 Introduction

The cytochrome P450 (CYP) enzymes, crucial haemoproteins localised in the cell membrane, play a significant role in regulating diverse cellular and physiological processes, activating endogenous compounds, and participating in detoxification reactions.^{237, 238} Inhibition of CYP enzymes is a key mechanism in metabolism-based drug-drug interactions,²³⁷ with CYP1A1 and CYP1B1 classified into xenobiotics and sterols, respectively, and implicated in steroid hormone biosynthesis and the metabolism of various compounds.²³⁸ Flavonoids, dietary phytochemicals, have been reported to inhibit CYP enzymes and are potential agents for cancer prevention.²³⁹ This study investigates the inhibitory effects of specific flavonoids on CYP1A1 and CYP1B1 enzymes, assessing their ADME-Tox properties. Understanding the interactions between flavonoids and CYP enzymes can contribute to the development of novel strategies for cancer prevention and treatment, addressing the significant global health challenge posed by breast cancer, its prevalence, mortality rates, and the urgent need for effective prevention and treatment strategies. Early detection, through diagnostic methods such as mammography and MRI, along with molecular markers like oestrogen receptor alpha (ER), progesterone receptor (PR), and human epidermal growth factor 2 (HER2), plays a crucial role in improving patient outcomes.²⁴⁰⁻²⁴² Aromatase inhibitors, targeting the enzyme encoded by the CYP19 gene, have shown promising results as hormonal therapies for ER+ breast cancer, preventing oestrogen biosynthesis and demonstrating improved clinical outcomes.^{243, 244} By

investigating the interaction between phorbols and aromatase inhibitors, this research aims to enhance our understanding of breast cancer pathogenesis and develop novel compounds for treatment strategies.

5.2 Computational Methods

In this chapter, protein-ligand docking was performed to identify interactions between different cytochrome P450 structures (CYP19A1, CYP1A1, CYP1B1) and phorbols diesters and dietary flavonoids (Figure 5.1, Figure 5.4) obtained from literature due to their known inhibitory activities. The Flare software was used to prepare the protein and ligand structures (version 2.0, Cresset Software, Litlington, UK). The full preparation of the protein and ligand at a pH of 7.0 and active site size of 6.00 Å adds missing hydrogens to the protein(s) and cofactor(s), it also assigns the best ionization states to the protein residues. It maximizes hydrogen bonding and reduces steric interference by optimizing the spatial locations of polar hydrogen atoms. His, Asn, and Gln side chains are optimized for side chain orientation, and then residues with unresolved side chains are found and reconstructed. During the preparation step, the endogenous ligand(s) and heme were extracted from the protein structure. The XED accurate approach on Flare was then used to minimize the protein and ligands. The docking calculation was set to the "very accurate but slow" option (this option in the software is based on the genetic algorithm, it performs three independent docking runs to achieve the lowest possible calculated binding energy for the protein-ligand complex). The grid box was set by selecting all the amino acid residues present in the protein. The docking was validated by redocking the endogenous ligand testosterone and the heme into the binding pocket. The BIOVIA Discovery Studio Visualizer, version 19.1 was used to visualize the poses with the lowest binding energy (Dassault Systemes, San Diego, United States). The PRODIGY web server was used to analyse the protein-ligand complexes' binding affinities.^{195, 196} The study also employed various tools and software, including ADMETLab, SwissADME, ChemDraw 18.0 and BIOVIA Discovery Studio Visualiser.

5.3 Computational investigation of ligand binding of flavonoids in cytochrome p450 receptors

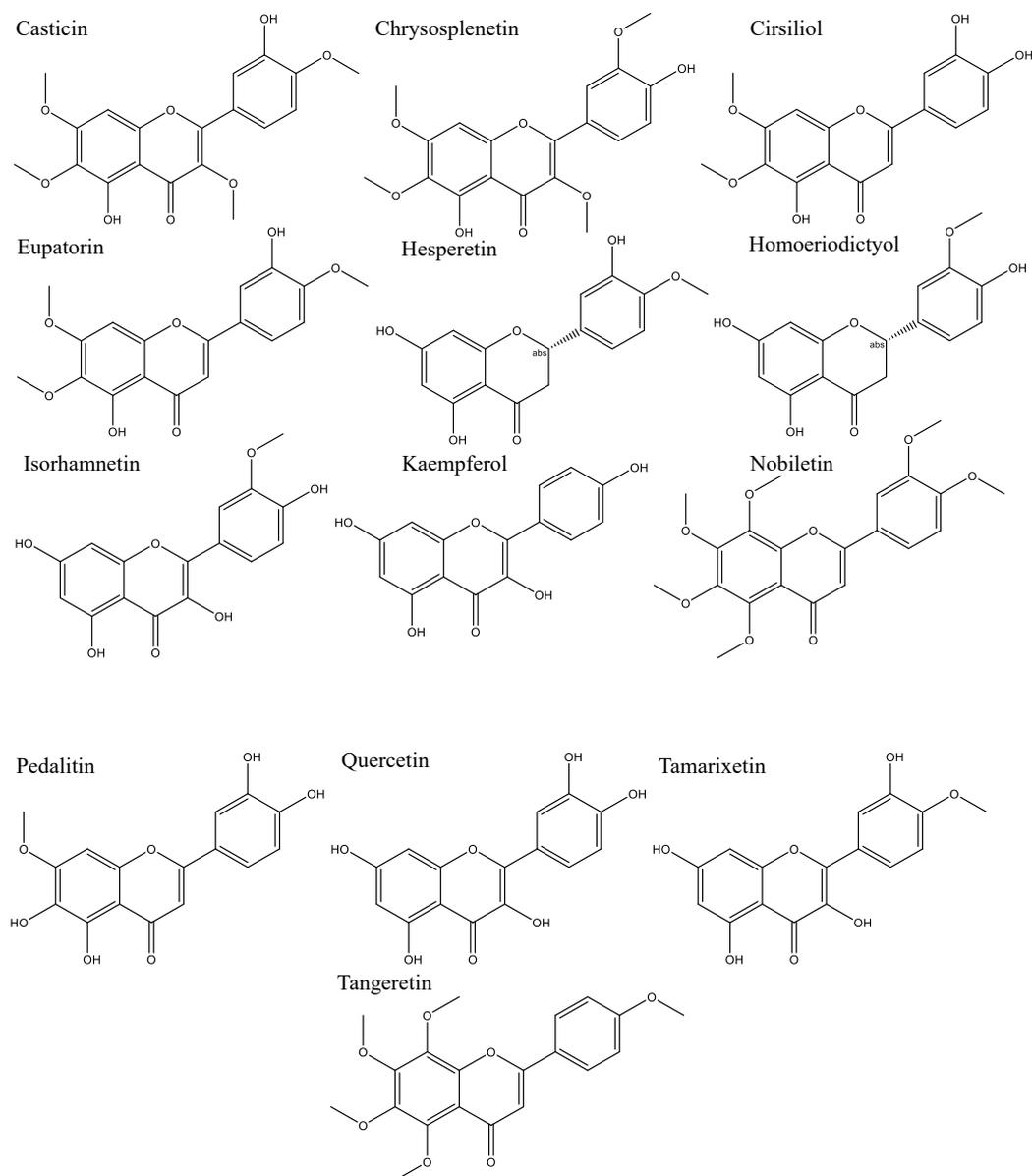


Figure 5.1. Chemical structure of the dietary flavonoids used for the study.

5.3.1 Results and Discussion

5.3.1.1 Docking Study

Human cytochrome (CYP) P450 are predominant enzymes that play essential roles in the biotransformation of a wide variety of drugs and other chemical compounds.²⁴⁵ The CYP 1 family comprises CYP1A1, CYP1A2 and CYP1B1.²⁴⁶ CYP1A1 and CYP1B1 are essential enzymes that are intricate in xenobiotic metabolism, but they do not partake much

in drug metabolism.²⁴⁶ These enzymes are significantly inhibited by ligands, which bind to the Aryl Hydrocarbon receptor while acting as metabolic enzymes for many substrates.^{246, 247} Molecular docking has been regarded as an essential and powerful tool in drug development as it is capable to predict the intermolecular framework formed between two proteins or a protein and a small molecule(s) and to indicate binding modes that elicit action.²⁴⁸

Table 5.1 shows the binding energies along with the dG score, rank score and virtual screening score of the ligands upon docking onto the crystal structure CYP1A1. The Flare lead finder ranking score/rank score is used in the ranking of ligand poses obtained during a docking run.^{59, 249, 250} The dG-scoring function, or dG score estimates an accurate free energy value of protein-ligand binding for a given protein-ligand complex, the coefficients for this function have been derived by fitting computed binding energies to the experimental values for a set of 100 protein-ligand complex with known 3D structures and binding constants measured experimentally.^{59, 249, 250} The vs-score or virtual scoring function is used to rank ligand poses generated during a docking run.^{59, 249, 250} Rof5 represents Lipinski's rule of five. Three compounds with the highest dG scores were selected and named the lead compounds for further in silico ADME-Tox screening. The ligand Isorhamnetin had the lowest binding free energy of -10.152 KCal/Mol (Table 5.1). Isorhamnetin upon docking into the crystal structure of CYP1A1 formed a π -stacked interaction with the benzene ring on the amino acid residue Phe 224 (Figure 5.2C). The carbon atom on Gly 316 also had a π -stacked interaction with the CD-ring of the ligand isorhamnetin. The methyl group on Ala 317 had a π -alkyl interaction with the D-ring of the ligand. Hydrogen bond interaction was observed between the oxygen atoms on Asp 313, 320 and Leu 217 and hydrogen atoms on isorhamnetin. Unfavourable donor-donor interaction was observed between a hydrogen atom on the residue Asn 222. The residues Phe 224 and Gly 316 interacted just at the centre of the CD ring (Figure 5.2C).

Table 5.1. The dG, rank score and virtual screening score of the ligands upon docking onto CYP1A1.

Compound	dG (Kcal/mol)	Rank score (Kcal/mol)	VS score (Kcal/mol)	Rof5
Chrysosplenetin	-9.818	-9.755	-10.972	0
Casticin	-10.101	-10.686	-11.131	0
Cirsiliol	-9.471	-10.367	-10.941	0
Eupatorin	-9.540	-9.968	-10.716	0
Hesperetin	-9.740	-9.693	-10.385	0
Homoeriodictyol	-10.107	-9.848	-10.485	0
Isorhamnetin	-10.152	-9.331	-10.546	0
Kaempferol	-9.880	-9.646	-10.714	0
Nobiletin	-9.162	-8.945	-10.000	0
Pedalitin	-10.150	-9.234	-10.660	0
Quercetin	-9.834	-8.909	-10.797	0
Tamarixetin	-10.065	-9.697	-10.628	0
Tangeretin	-8.454	-7.807	-9.319	0

The schematic representation of the protein-ligand interaction occurring at the binding site of CYP1A1 generated using DS visualiser are presented in Figures 5.2A-E.

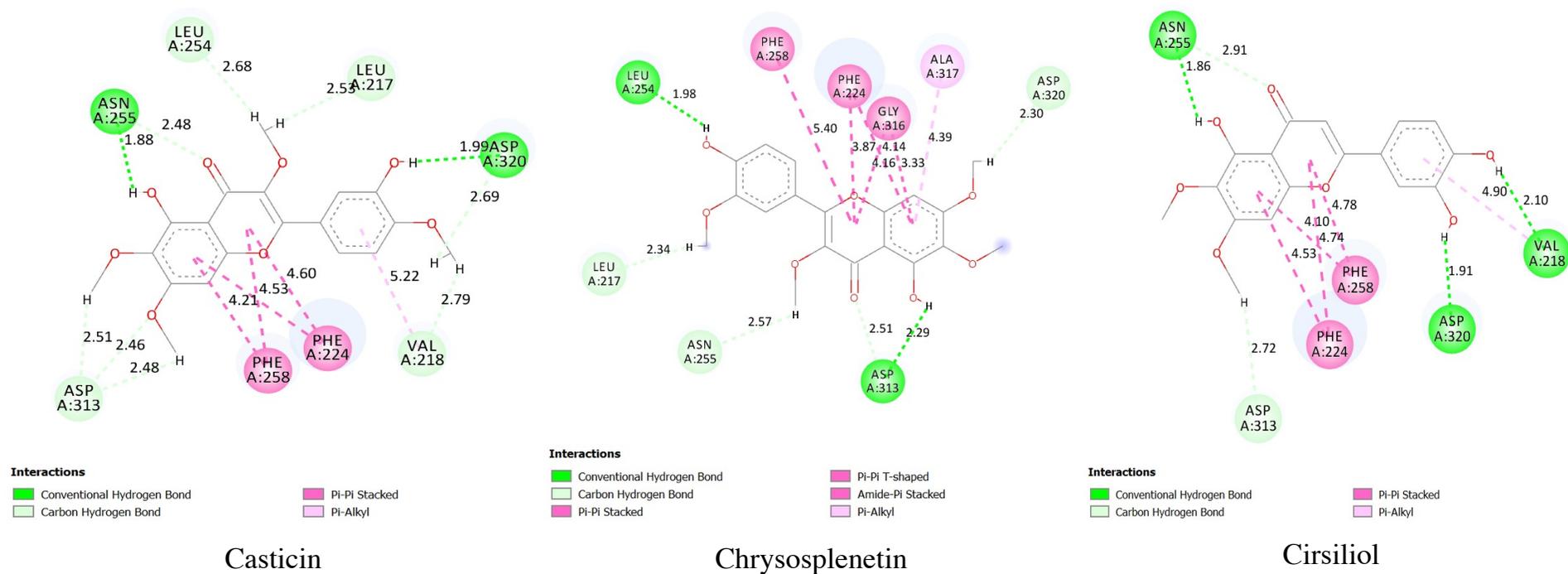


Figure 5.2A. Schematic representation generated using Discovery studio visualiser showing the protein-ligand interactions occurring at the CYP1A1 (6DWM) binding site.

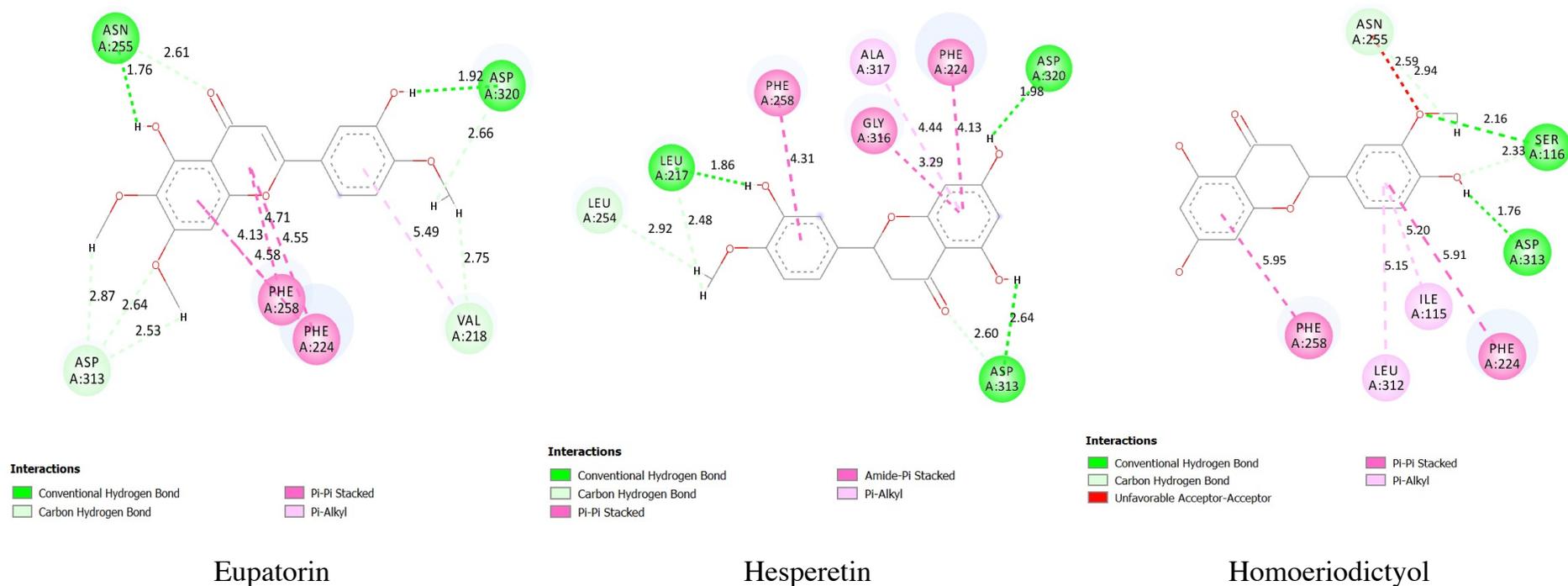


Figure 5.2B. Schematic representation generated using Discovery studio visualiser showing the protein-ligand interactions occurring at the CYP1A1 (6DWM) binding site.

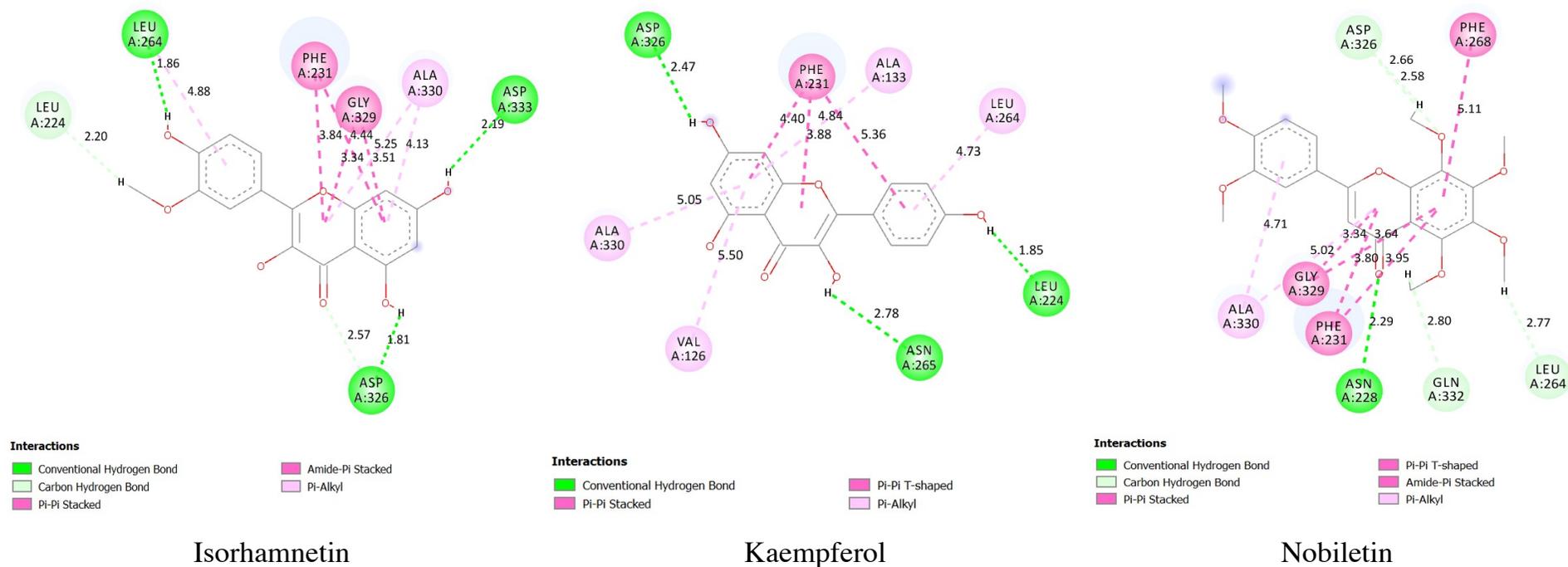
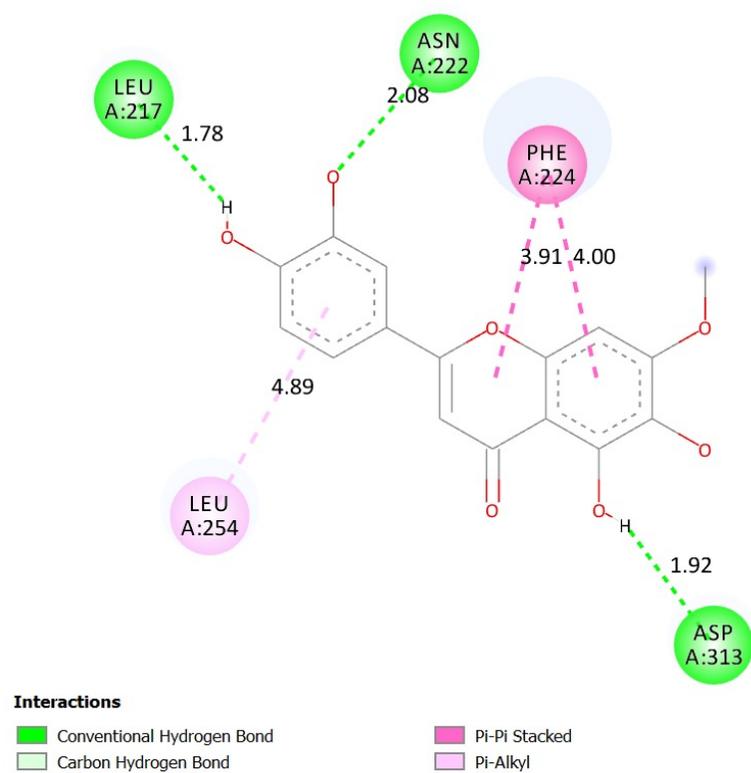
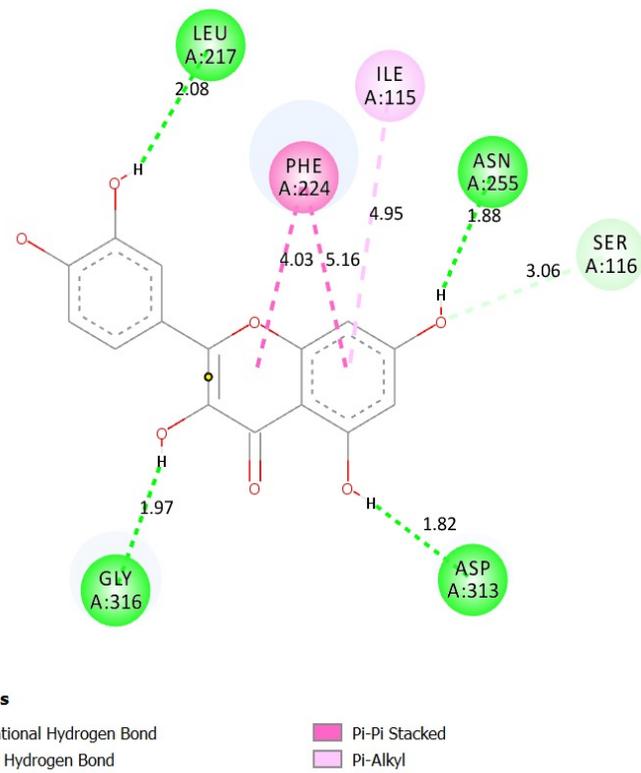


Figure 5.2C. Schematic representation generated using Discovery studio visualiser showing the protein-ligand interactions occurring at the CYP1A1 (6DWM) binding site.

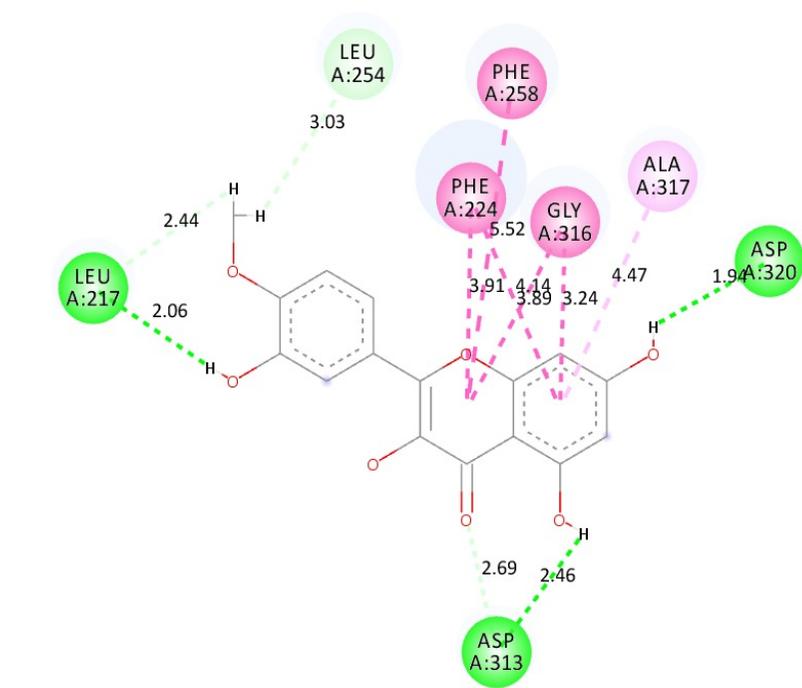


Pedalitin



Quercetin

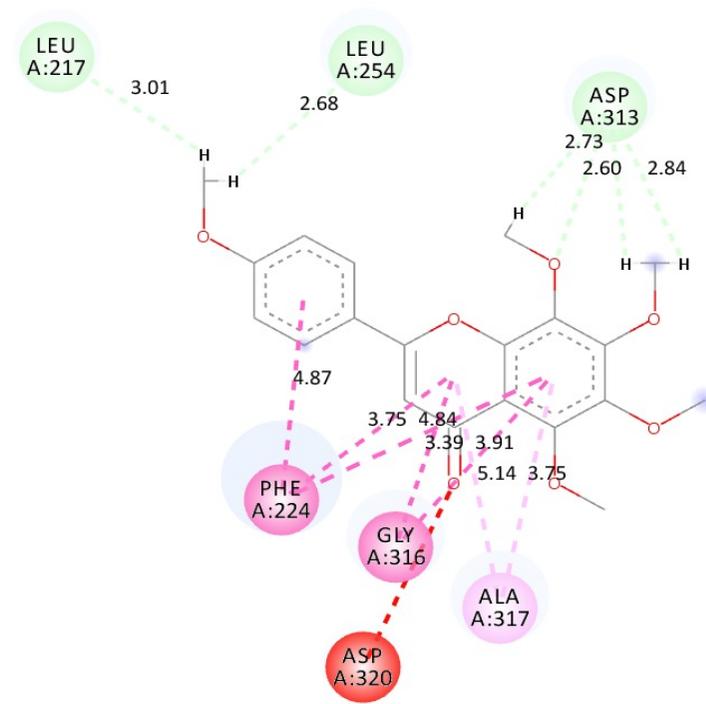
Figure 5.2D. Schematic representation generated using Discovery studio visualiser showing the protein-ligand interactions occurring at the CYP1A1 (6DWM) binding site.



Interactions

- Conventional Hydrogen Bond
- Carbon Hydrogen Bond
- Pi-Pi Stacked
- Pi-Pi T-shaped
- Amide-Pi Stacked
- Pi-Alkyl

Tamarixetin



Interactions

- Carbon Hydrogen Bond
- Unfavorable Acceptor-Acceptor
- Pi-Pi Stacked
- Pi-Pi T-shaped
- Amide-Pi Stacked
- Pi-Alkyl

Tangeretin

Figure 5.2E. Schematic representation generated using Discovery studio visualiser showing the protein-ligand interactions occurring at the CYP1A1 (6DWM) binding site.

Table 5.2 shows the binding energies along with the dG score, rank score and virtual screening score of the ligands upon docking onto the crystal structure CYP1B1. Three compounds with the highest dG scores were selected and named the lead compounds for further in silico ADME-Tox screening. The ligand Pedalitin had the lowest binding free energy of -10.823 Kcal/Mol (Table 5.2). As shown in Figure 6.3D, the sulphur atom on Met 152 had π -sulphur with the benzene ring on the ligand pedalitin. The methyl groups on the following residues (Met 152, Ile 327 and Ala 330) had π -alkyl interactions with the CD-ring of the ligand. In contrast, another π -alkyl interaction was also observed between Cys 470 and the C-ring of the ligand. Hydrogen bond interactions between oxygen atoms in the ligand and hydrogen atoms on the following residues (Arg 145, Ser 331, Gln 479) were observed. Moreover, the sulphur atom on the residue Cys 470 formed a hydrogen bond interaction with a hydrogen atom on the ligand pedalitin (Figure 5.2D). The methyl groups on the residues (Leu 475, Met 205, Val 208) formed π -alkyl interactions with the benzene ring on the ligand. As shown in Figure 6.3D, the following residues (Leu 475, Met 205, Val 208, Met 152) interacted in the centre of the benzene ring while Ile 327, Ala 330, Met 152 in the centre of the CD ring.

Table 5.2. The dG, rank score and virtual screening score of the ligands upon docking onto CYP1B1.

Compound	dG (Kcal/mol)	Rank score (Kcal/mol)	VS score (Kcal/mol)	Rof5
Chrysosplenetin	-10.216	-10.243	-11.157	0
Casticin	-10.404	-10.054	-11.442	0
Cirsiliol	-9.260	-8.100	-9.579	0
Eupatorin	-9.191	-8.158	-9.793	0
Hesperetin	-9.360	-8.458	-9.828	0
Homoeriodictyol	-8.951	-8.371	-9.628	0
Isorhamnetin	-10.177	-9.680	-11.172	0
Kaempferol	-10.059	-9.386	-10.730	0
Nobiletin	-9.496	-9.213	-10.496	0
Pedalitin	-10.823	-8.222	-11.150	0
Quercetin	-10.434	-10.199	-10.947	0
Tamarixetin	-9.776	-8.682	-10.173	0
Tangeretin	-8.440	-7.725	-9.490	0

The schematic representation of the protein-ligand interaction occurring at the binding site of CYP1A1 generated using DS visualiser are presented in Figures 5.3A-E.

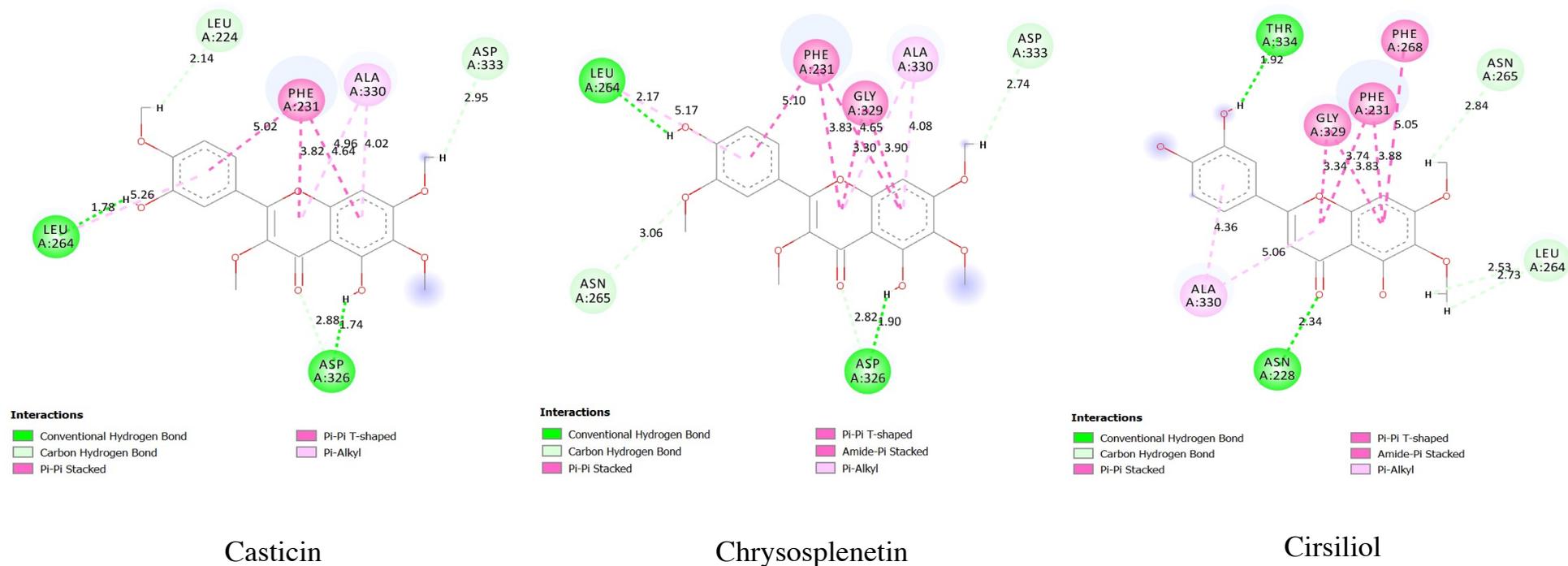


Figure 5.3A. Schematic representation generated using Discovery studio visualiser showing the protein-ligand interactions occurring at the CYP1B1 (6IQ5) binding site.

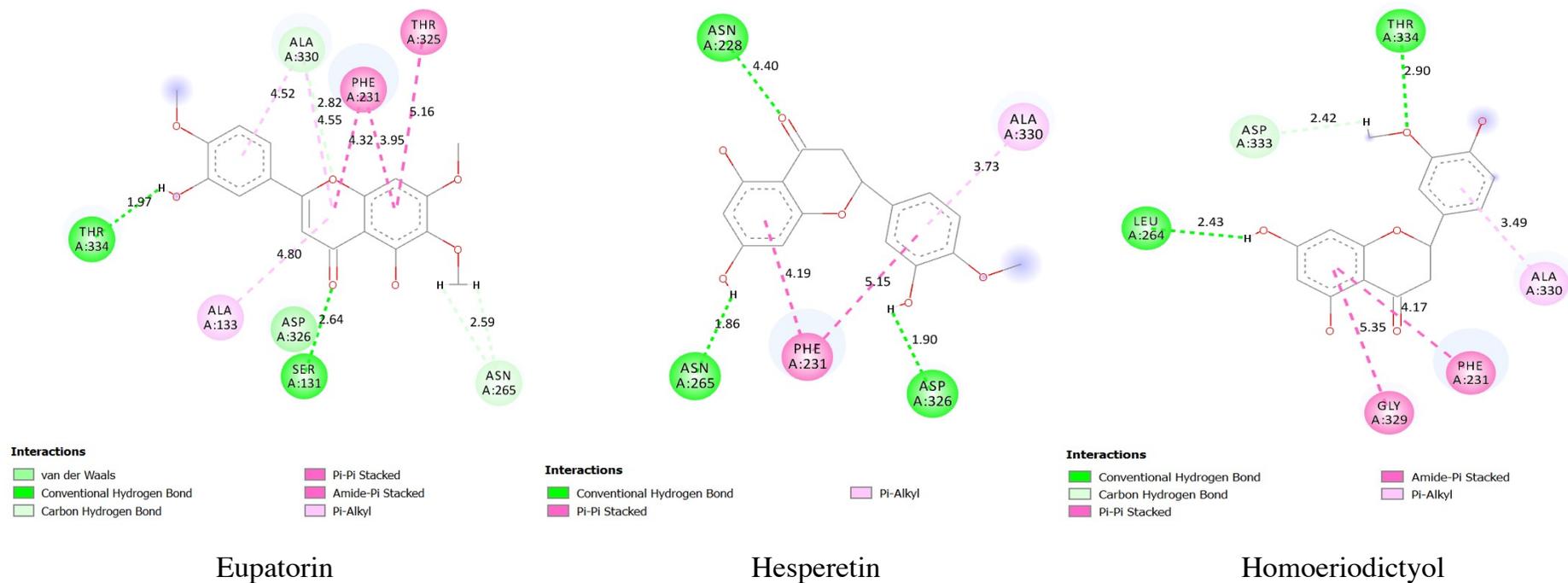


Figure 5.3B. Schematic representation generated using Discovery studio visualiser showing the protein-ligand interactions occurring at the CYP1B1 (6IQ5) binding site.

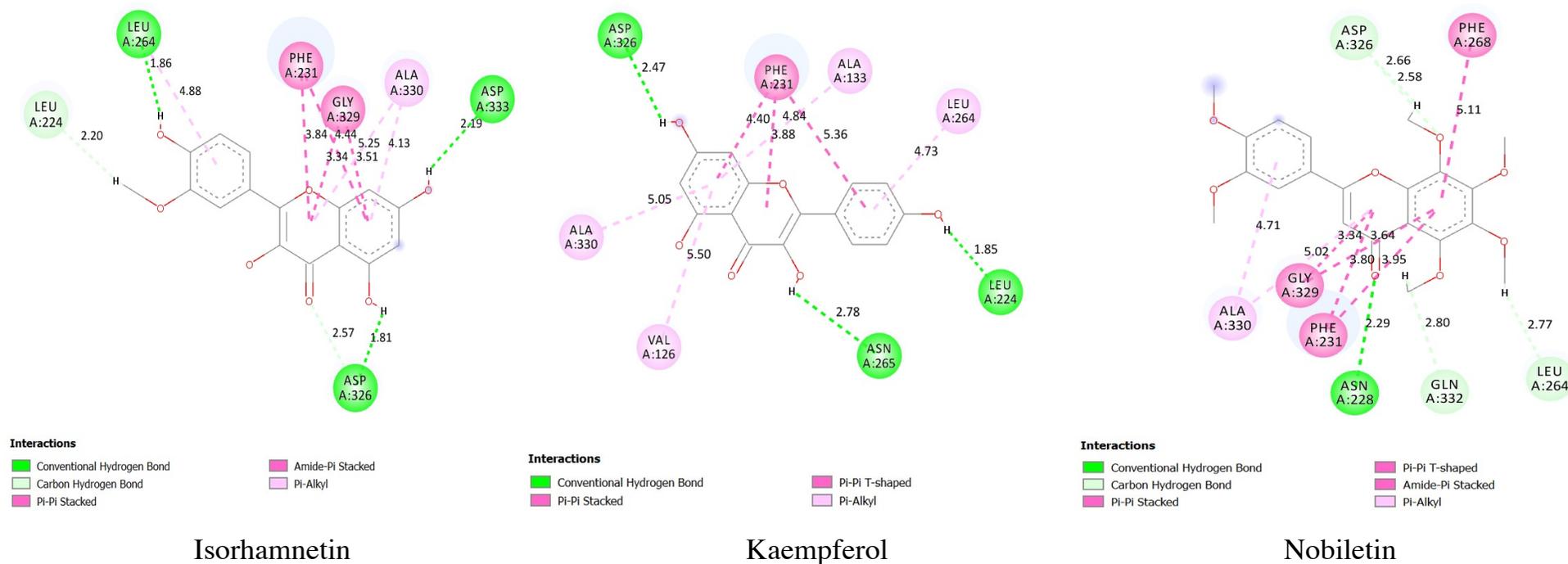


Figure 5.3C. Schematic representation generated using Discovery studio visualiser showing the protein-ligand interactions occurring at the CYP1B1 (6IQ5) binding site.

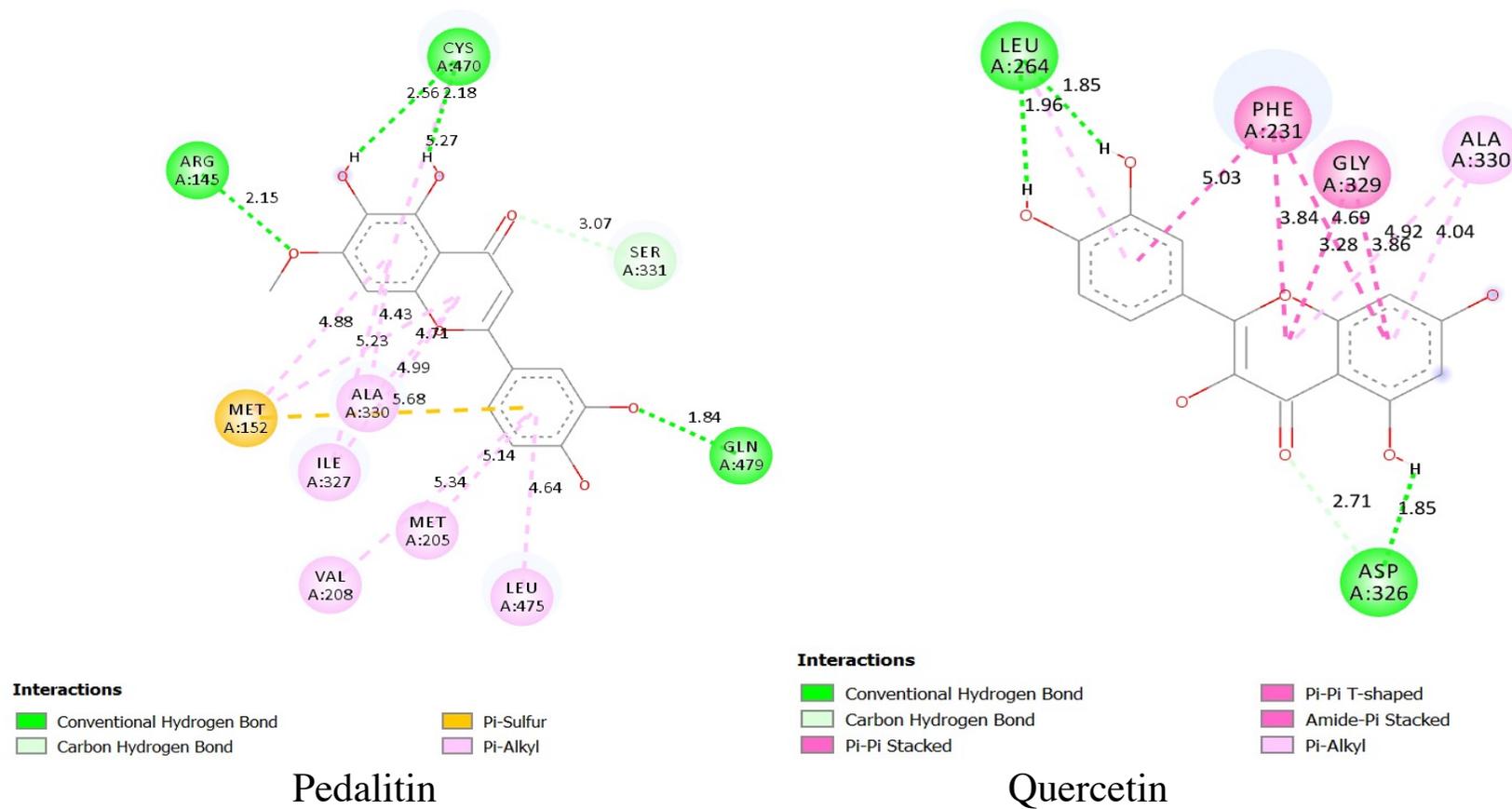
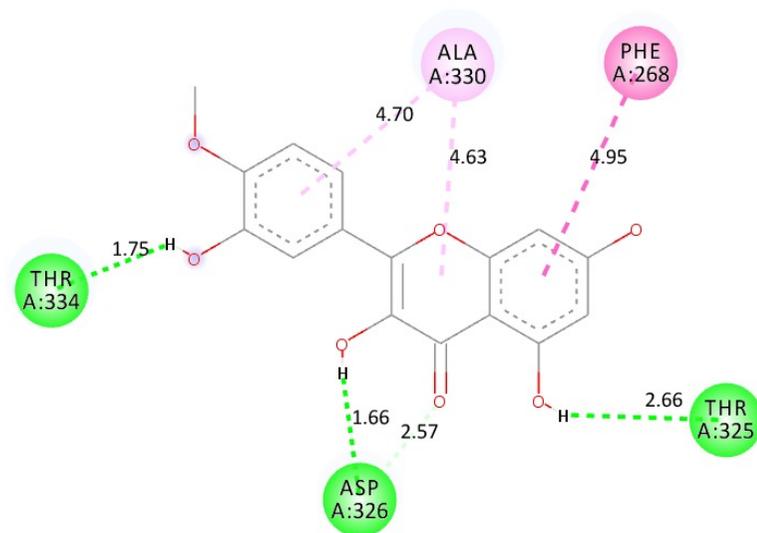


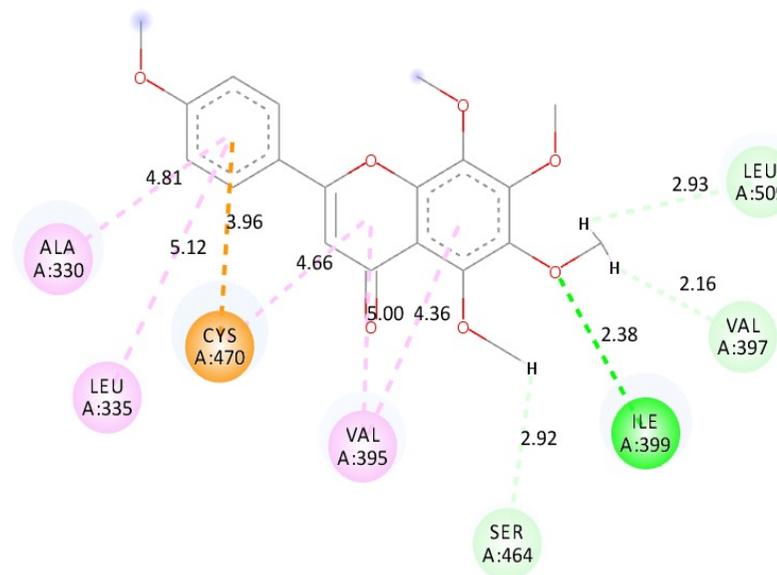
Figure 5.3D. Schematic representation generated using Discovery studio visualiser showing the protein-ligand interactions occurring at the CYP1B1 (6IQ5) binding site.



Interactions

- Conventional Hydrogen Bond
- Carbon Hydrogen Bond
- Pi-Pi T-shaped
- Pi-Alkyl

Tamarixetin



Interactions

- Conventional Hydrogen Bond
- Carbon Hydrogen Bond
- Pi-Anion
- Pi-Alkyl

Tangeretin

Figure 5.3E. Schematic representation generated using Discovery studio visualiser showing the protein-ligand interactions occurring at the CYP1B1 (6IQ5) binding site.

The findings of this study showed that the residue Phe 224 interacted with all the ligands forming a π - π stacked interaction when it was docked onto CYP1A1 (6DWM) (Figures 5.2A-E). Most of the ligands studied upon docking into CYP1B1, formed a π - π stacked interaction with Phe 231 (Figures 5.3A-E) except for pedalitin and tangeretin, which did not interact with any phenylalanine residue. These phenylalanine residues present in the binding site of CYP1A1 and CYP1B1 are conversed and interacted with the flavonoid α -naphthoflavone (ANF) upon docking into the same binding pocket.²⁵¹ In another study carried out by Kubo, Yamamoto, and Itoh²⁵², the residue Phe 231 had a π - π stacked interaction with ANF in the CD-ring. Walsh, Szklarz, and Scott²⁴⁵ also confirmed the presence of a π - π stacked interaction with Phe 224 in their study. The residues Phe 224 and Phe 231 almost converge identically on the binding site of CYP1A1 and CYP1B1, respectively.^{245, 246}

The findings on figures 5.2A-E showed that some ligands: chrysopterin, casticin, hesperetin, tamarixetin and tangeretin had a hydrogen bond interaction with the amino acid residue Leu 254. The ligand pedalitin, had a π -alkyl interaction with the residue Leu 254. In the other crystal structure (CYP1B1) studied the residue Leu 264 formed a hydrogen bond interaction with the ligands except for the following eupatorin, hesperetin, pedalitin, tamarixetin, tangeretin (Figure 5.3A-E). The ligand kaempferol interacted with Leu 264 differently forming an π -alkyl interaction. In a study carried out by Walsh, Szklarz, and Scott²⁴⁵, it was reported that the residue Leu 254 forms part of the side chains lining the active site of CYP1A1. Their findings showed that the CYP1A1 and 1B1 cavity is similar even though the 1A1 cavity is slightly larger than 1B1 due to Leu 254/Leu 264 side-chain orientation on the G-Helix.

The docking result of CYP1A1 and the ligands investigated presented in figures 5.2A-E shows the residue Ala 317 interacting with some of the ligands, namely, chrysopterin, hesperetin, isorhamnetin, nobiletin, tamarixetin and tangeretin. These ligands had π -alkyl interactions with the residue Ala 317 except for nobiletin which had a carbon hydrogen bond interaction (Figure 6.2C). Santes-Palacios et al.²⁵³ demonstrated the importance of some amino acid residues, including Ala 317, after a site-directed mutagenesis study that showed altered kinetic parameters. In a review written by Scotti et al.²⁵⁴, it was asserted that Ala 133 is crucial for the inhibitory activity of methoxy flavonoids, i.e. it is essential for the binding of methoxy flavonoids such as isorhamnetin and chrysoeriol. In addition, Sehgal et al.²⁵¹, in their study also identified Ala 317 in two binding pockets of the

enzyme CYP1A1. Based on the result shown in figures 5.1A-E, the ligands formed a π -alkyl interaction with Ala 330. However, eupatorin and kaempferol formed an π -alkyl with Ala 133 (Figure 5.3B and Figure 5.3C). The residue Ala 133 is one of the non-conserved residues in the active site of CYP1B1 with ANF bound to the active site.²⁴⁶

The residue Asp 333 had hydrogen bond interactions with the following ligands chrysosplenetin, casticin, homoeriodictyol and isorhamnetin (Figures 5.3A-E). The residue Asp 326 also had hydrogen bond interaction with the ligands chrysosplenetin, casticin, hesperetin, isorhamnetin, kaempferol, nobiletin, quercetin and tamarixetin (Figures 5.3A-E), while Van der Waals bonds were observed with eupatorin (Figures 5.3A-E). These residues were found in the active site of CYP1B1 in a previous study.²⁴⁵ The docking results of CYP1A1 showed the ligands having a hydrogen bond interaction with the residue Asp 313. Another residue, Asp 320, also interacted with some of the ligands namely, chrysosplenetin, casticin, cirsiolol, eupatorine, hesperetin, isorhamnetin, nobiletin, tangeretin and tamarixetin (Figures 5.2A-C and E). Dutkiewicz and Mikstacka²⁴⁶ in their research also confirmed Asp 313 and Asp 320 as potential hydrogen bond donors in the binding site of CYP1A1. Santes-Palacios et al.²⁵³, also confirmed that the residue Asp 313 is of importance in the CYP1A1 binding site. The following ligands: chrysosplenetin, cirsiolol, eupatorin, hesperetin, kaempferol formed hydrogen bonds with the residue Asn 265 upon docking with the crystal structure of CYP1B1 (6IQ5) (Figures 5.3A-C) The residue Asn 228 had hydrogen bond interactions with the ligands cirsiolol, hesperetin and nobiletin (Figures 5.3A-C). The residue Asn 255 had hydrogen bond interactions with chrysosplenetin, casticin, cirsiolol, eupatorin, homoeriodictyol, kaempferol, quercetin (Figures 5.2A-D), Asn 222 had an unfavourable donor-donor interaction with isorhamnetin and hydrogen bond interaction with pedalitin (Figures 5.2C-D). These residues were also identified in the active site of CYP1B1 in a different study carried out by Walsh, Szklarz and Scott.²⁴⁵ The amino acid residues Asn 222 and 255 constitute the polar residues found in the active site of CYP1A1; both residues tend to play structural roles instead of direct interactions with ANF.²⁴⁵

The ligands kaempferol, pedalitin and tangeretin formed π -alkyl interactions with the following residues Val 126, Val 208, Val 395, respectively, upon docking into CYP1B1 (Figures 5.3C-E). The residue Val 397 formed a hydrogen bond with the ligand tangeretin (Figure 5.3E). Walsh et al. in their research findings reported that Val 126 and 395 were in the active site of CYP1B1.²⁴⁵ The amino acid residue Gly 329 formed π -stacked interacted interactions with the following ligands: chrysosplenetin, cirsiolol, homoeriodictyol, isorhamnetin, nobiletin and quercetin (Figures 5.3A-D). The docking result of CYP1A1

showed the residue Gly 316 interacting with most of the ligands. Gly 316 had amide π -stacked interactions with the ligands such as chrysosplenetin, hesperetin, isorhamnetin, nobiletin, tamarixetin and tangeretin. Furthermore, Gly 316 had a hydrogen bond interaction quercetin (Figure 5.2D). Analysis of the binding sites of CYP1A1 and CYP1B1 revealed that the residues Gly 316 and Gly 329 were present in both active sites and could be potential hydrogen bond donors.²⁴⁶ The residue Thr 334 had hydrogen bond interactions with four ligands, namely, cirsiolol, eupatorin, homoeriodictyol, tamarixetin (Figures 5.3A, B and E). Thr 325 interaction was also observed. It had an amide π -stacked interaction with eupatorin and a hydrogen bond interaction with tamarixetin (Figure 5.3B, Figure 6.3E). According to Kubo et al. Thr 334 and two other residues, Val 395 and Thr 510 formed a small hydrophobic site.²⁵² This was subsequently confirmed using three different docking models of human CYP1B1.²⁵²

The ligands eupatorin, pedalitin and tangeretin formed hydrogen bonds with the following serine residues: 131, 331 and 464, respectively, on the structure of CYP1B1 (Figures 5.3B, D & E). Examination of the docking result of CYP1A1 showed Ser 116 interacting with just homoeriodictyol and quercetin (Figure 5.2B, Figure 5.2D). CYP1A1 formed hydrogen bond interactions with both ligands. Based on the interactions observed with the serine residues, it can also be inferred that homoeriodictyol and quercetin also donated hydrogen bonds just like Ser 122, which was found in the active site of CYP1A1 upon docking with ANF.²⁴⁶ The residue Gln 332 and 479 formed hydrogen bonds with nobiletin and pedalitin, respectively (Figure 5.2C, Figure 5.2D) upon docking with CYP1B1. In a study carried out by Dutkiewicz and Mikstacka²⁴⁶, their docking results show that Gln 332 formed an active site hydrogen bond with the ligands. Amino acid residues such as Ile 386, Lys 250, Lys 253, Ile 115, Ser 116, and Met 121 were found in the binding site of CYP1A1 and CYP1A2, respectively, due to their structural similarities mirror structures of the aforementioned residues can also be found in CYP1B1.^{245, 246}

5.3.1.2 Pilot ADMET studies

The webserver ADMETLab and SwissADME developed by Dong et al.⁷⁸ and Daina et al.⁸³ respectively were used in the *in silico* ADME-Tox screening of the selected compounds (casticin, homoeriodictyol, isorhamnetin, pedalitin, quercetin, and chrysosplenetin) identified from the docking study. The findings of this study are shown in Tables 5.3 and 5.4. The lead-likeness parameter on SwissADME is calculated based on the

research findings of Teague et al²⁵⁵ which defined lead-like leads as low-affinity compounds which have low molecular weight (< 350) and LogP (< 3) values

Table 5.3. Physicochemical parameters of the lead compounds.

Parameters	Chrysosplenetin	Casticin	Homoeriodictyol	Isorhamnetin	Pedalitin	Quercetin
Molecular weight (MW)	374.34 g/mol	374.34 g/mol	302.28 g/mol	316.26 g/mol	316.26 g/mol	302.24 g/mol
No. of hydrogen bond acceptors	8	8	6	7	7	7
No. of hydrogen bond donors	2	2	3	4	4	5
TPSA	107.59 Å ²	107.59 Å ²	96.22 Å ²	120.36 Å ²	120.36 Å ²	131.36 Å ²
Lipophilicity, Log P (Consensus)	2.49	2.51	1.91	1.65	1.56	1.23
Log S (ESOL)	-4.24	-4.24	3.62	-3.36	-3.76	-3.16
Drug-likeness	Yes	Yes	Yes	Yes	Yes	Yes
Lead-likeness	No; MW > 350	No; MW > 350	Yes	Yes	Yes	Yes

TPSA: Topological polar surface area, Log S (ESOL): Water solubility.

The values for absorption, distribution, metabolism, excretion and toxicity (ADMET) play a significant role in developing a new drug.⁷⁰ A drug candidate of much high quality should not just possess adequate efficacy against the drug target, but it needs to show good ADMET properties at its therapeutic dose.⁷⁰

The examination of the physicochemical properties of the lead compounds showed four of the leads; namely, homoeriodictyol, isorhamnetin, pedalin and quercetin, have soluble properties (Table 5.3). Moreover, chrysopterin and casticin were moderately soluble. The ligand quercetin had the highest total polar surface area (TPSA) due to the presence of extra polar carbonyl oxygen.²⁵⁶ The ligands did not violate any drug-likeness parameters. However, chrysopterin and casticin would not be considered as potential leads because their molecular weights were greater than 350 g/mol.⁸³

Table 5.4. Pharmacokinetics of the lead compounds.

Compound	HIA	F 20%	F 30%	P-gp	PPB (%)	BBB	hERG Blockage	H-HT	Ames
Chrysosplenetin	Positive	Positive	Positive	Non-inhibitor/non-substrate	78.302	Negative	Positive	Positive	Negative
Casticin	Positive	Positive	Positive	Non-inhibitor/non-substrate	78.585	Negative	Positive	Positive	Negative
Homoeriodictyol	Negative	Positive	Negative	Non-inhibitor/non-substrate	87.996	Positive	Negative	Negative	Negative
Isorhamnetin	Negative	Positive	Negative	Inhibitor/non-substrate	90.707	Negative	Negative	Positive	Negative
Pedalitin	Positive	Positive	Negative	Inhibitor/non-substrate	89.844	Positive	Negative	Positive	Positive
Quercetin	Negative	Positive	Negative	Non-inhibitor/non-substrate	94.987	Negative	Negative	Positive	Positive

Table 5.4. Pharmacokinetics of the lead compounds. (Continued)

Compound	CYP1A2	CYP3A4	CYP2C9	CYP2C19	CYP2D6
Chrysofenetin	Inhibitor/non-substrate	Non-inhibitor/substrate	Non-inhibitor/substrate	Non-inhibitor/substrate	Non-inhibitor/substrate
Casticin	Inhibitor/substrate	Non-inhibitor/substrate	Non-inhibitor/substrate	Non-inhibitor/substrate	Non-inhibitor/non-substrate
Homoeriodictyol	Inhibitor/substrate	Inhibitor/substrate	Non-inhibitor/substrate	Non-inhibitor/substrate	Non-inhibitor/substrate
Isorhamnetin	Inhibitor/non-substrate	Inhibitor/non-substrate	Non-inhibitor/substrate	Non-inhibitor/substrate	Non-inhibitor/non-substrate
Pedalitin	Inhibitor/substrate	Inhibitor/non-substrate	Non-inhibitor/substrate	Non-inhibitor/substrate	Non-inhibitor/substrate
Quercetin	Inhibitor/non-substrate	Inhibitor/non-substrate	Non-inhibitor/non-substrate	Non-inhibitor/non-substrate	Non-inhibitor/non-substrate

HIA: Human intestinal absorption, F: Bioavailability, P-gp: p-glycoprotein, PPB: Plasma protein binding, BBB: Blood brain barrier, hERG: human Ether-à-go-go-Related Gene (Potassium ion channel), H-HT: Human hepatotoxicity, Ames: Ames Mutagenicity

As seen in Table 5.4, the pharmacokinetic predictions carried out using ADMETLab showed that three of the leads namely, chrysosplenetin, casticin and pedalitin were positive for HIA. Furthermore, the model predicts that over 30% of the compounds are absorbed in the intestines. In contrast, the other compound, which was negative, translates to less than 30% of them absorbed intestinally. According to the prediction model, the ligands tested positive for 20% bioavailability. Based on the same model, chrysosplenetin and casticin tested positive for F30%, which translates to over 20% and 30% being bioavailable, respectively (Table 5.4). The findings of the ADMETLab predictions shows that none of the ligands was P-gp substrates. According to the prediction model, chrysosplenetin, casticin, homoeriodictyol and quercetin were non-inhibitors of P-gp while isorhamnetin and pedalitin are inhibitors of P-gp (Table 5.4).

The predictions for the PPB showed the ligands isorhamnetin and quercetin being at an optimum of over 90% (Table 5.4), whilst pedalitin had a prediction value of about 89.844%, which is also close to the optimum figures. Most of the ligands (chrysosplenetin, casticin, isorhamnetin and quercetin) tested negative for blood-brain barrier permeability (Table 5.4). The toxicity prediction showed four of the ligands (homoeriodictyol, isorhamnetin, pedalitin and quercetin) tested negative for the hERG channel blockage, as shown in Table 5.4. Moreover, chrysosplenetin, casticin, isorhamnetin, pedalitin and quercetin tested positive for H-HT (Table 5.4). Four of the ligands (chrysosplenetin, casticin, homoeriodictyol and isorhamnetin) tested negative for Ames mutagenicity. In addition, all the tested leads were found to be negative for skin sensitisation, and all the lead compounds tested were positive for DILI (Table 5.4).

The predictions showed that all the ligands inhibited the enzyme CYP1A2, while casticin, homoeriodictyol and pedalitin were found to be substrates of the enzyme inhibiting it. As shown in Table 5.4, both casticin and homoeriodictyol are inhibitors and substrates of CYP3A4, whereas chrysosplenetin is a non-inhibitor and non-substrate of CYP3A4. As seen in Table 5.4, isorhamnetin, pedalitin and quercetin inhibit CYP3A4, but they are not substrates of the enzyme CYP3A4. All the ligands, except quercetin, were found to be neither non-inhibitors nor substrate of CYP2C9. However, quercetin, which is not a substance of the enzyme, did not inhibit CYP2C9 (Table 5.4). Therefore, quercetin is a non-inhibitor and non-substrate of the enzyme CYP2C9. Regarding other ligands, chrysosplenetin, casticin, homoeriodictyol, isorhamnetin and pedalitin are non-inhibitors but substrates of CYP2C9 (Table 5.4). According to the prediction results, the ligands such as casticin, isorhamnetin and quercetin are non-inhibitors and non-substrate of the enzyme

CYP2D6. In contrast, the other ligands were non-inhibitors and substrates of the enzyme CYP2D6 (Table 5.4).

5.4 The role of phorbol-diester in mediating human placental aromatase cytochrome p450 activity.

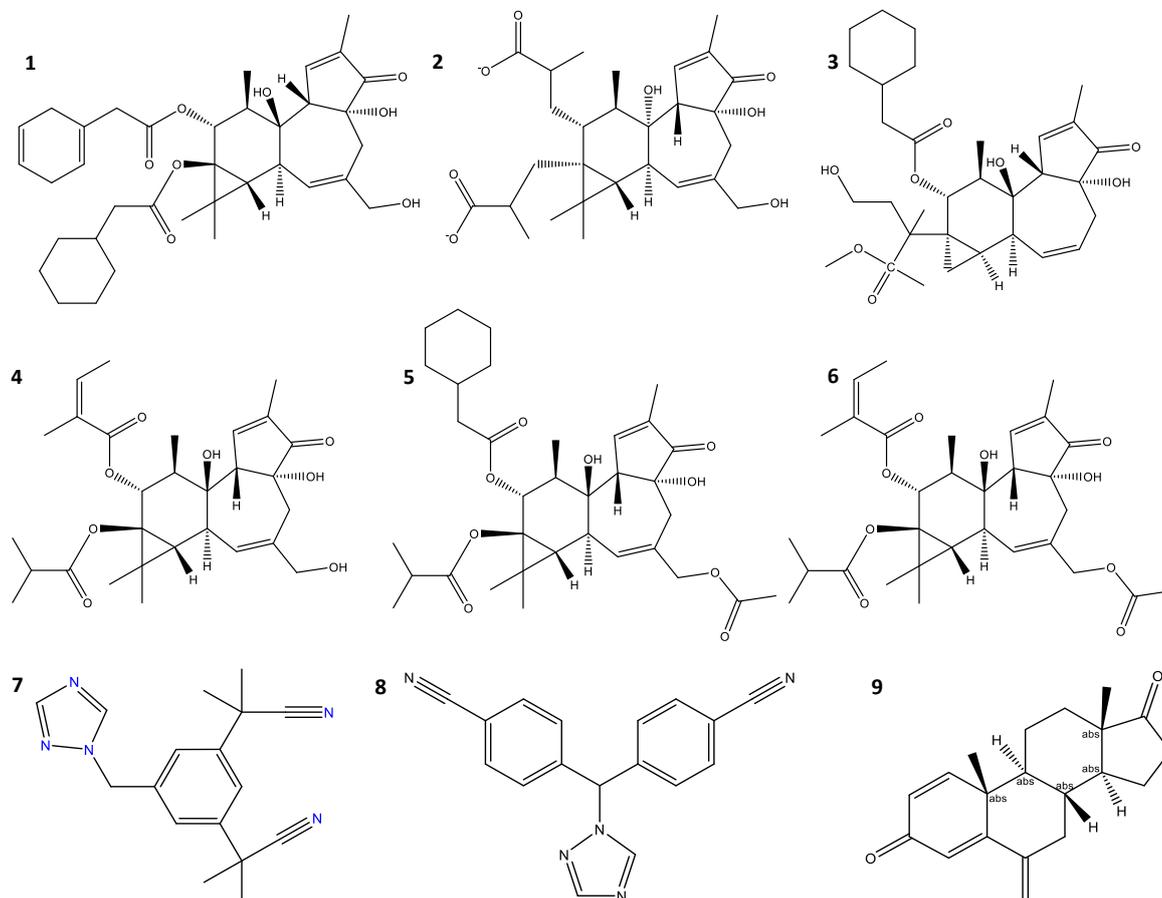


Figure 5.4. Schematic image the chemical structures of the phorbol diesters used in the study. 1. P-12,13-diAcPh: (1aR,1bS,4aR,7aS,7bS,8R,9R,9aS)-9-(2-(cyclohexa-1,4-dien-1-yl)acetoxy)-4a,7b-dihydroxy-3-(hydroxymethyl)-1,1,6,8-tetramethyl-5-oxo-1,1a,1b,4,4a,5,7a,7b,8,9-decahydro-9aH-cyclopropa[3,4]benzo[1,2-e]azulen-9a-yl 2-cyclohexylacetate. 2. P-12,13-diiBu: (1aR,1bS,4aR,7aS,7bS,8R,9R,9aS)-4a,7b-dihydroxy-3-(hydroxymethyl)-1,1,6,8-tetramethyl-5-oxo-1,1a,1b,4,4a,5,7a,7b,8,9-decahydro-9aH-cyclopropa [3,4] benzo[1,2-e] azulene-9,9a-diyl bis(2-methylpropanoate). 3. P-12AcPh-13iBu: (1aR,1bS,4aR,7aS,7bS,8R,9R,9aS)-9-(2-cyclohexylacetoxy)-4a,7b-dihydroxy-3-(hydroxymethyl)-1,1,6,8-tetramethyl-5-oxo-1,1a,1b,4,4a,5,7a,7b,8,9-decahydro-9aH-cyclopropa [3,4] benzo[1,2-e] azulen-9a-yl isobutyrate. 4. P-12Ang-13iBu: (1aR,1bS,4aR,7aS,7bS,8R,9R,9aS)-4a,7b-dihydroxy-3-(hydroxymethyl)-9a-(isobutyryloxy)-1,1,6,8-tetramethyl-5-oxo-1a,1b,4,4a,5,7a,7b,8,9,9a-decahydro-1H-cyclopropa [3,4] benzo[1,2-e] azulen-9-yl (Z)-2-methylbut-2-enoate. 5. P-20Ac-12AcPh-13iBu: (1aR,1bS,4aR,7aS,7bS,8R,9R,9aS)-3-(acetoxymethyl)-9-(2-(cyclohexa-

2,5-dien-1-yl)acetoxyl)-4a,7b-dihydroxy-1,1,6,8-tetramethyl-5-oxo-1,1a,1b,4,4a,5,7a,7b,8,9-decahydro-9aH-cyclopropa[3,4]benzo[1,2-e]azulen-9a-yl isobutyrate. 6. P-20Ac-12Ang-13iBu: (1aR,1bS,4aR,7aS,7bS,8R,9R,9aS)-3-(acetoxymethyl)-4a,7b-dihydroxy-9a-(isobutyryloxy)-1,1,6,8-tetramethyl-5-oxo-1a,1b,4,4a,5,7a,7b,8,9,9a-decahydro-1H-cyclopropa [3,4] benzo[1,2-e] azulen-9-yl (Z)-2-methylbut-2-enoate. 7. Anastrozole. 8. Letrozole. 9. Exemestane.

5.4.1 Results and Discussion

Table 5.5 shows the ligand binding energy upon docking into CYP19A1. Flare lead finder module ranked the ligand poses produced during the docking run.^{249, 250} The dG-score provides a free energy value of protein-ligand binding for any protein-ligand complex being investigated.^{249, 250} Presented in Table 5.5 is the results of the binding affinity analysis.

Table 5.5. Table showing the binding energy of the compounds and binding affinities of the docked complexes.

Compound	dG (Kcal/mol)	Binding affinity ΔG_{noelec} (Kcal/mol)
P-12,13-diAcPh	-13.993	-13.7
P-12,13-diiBu	-12.114	-11.5
P-12AcPh-13iBu	-12.944	-12.9
P-12Ang-13iBu	-12.172	-11.8
P-20Ac-12AcPh-13iBu	-13.379	-13.0
P-20Ac-12Ang-13iBu	-12.686	-12.3
Anastrozole	-7.301	-8.4
Letrozole	-7.909	-8.2
Exemestane	-8.338	-10.3

The result in the Table 5.5. showed that the phorbol diesters had higher binding energies than the commercially available aromatase inhibitors. The diester P-12,13-diAcPh had the highest binding energy value (-13.993 Kcal/mol) in comparison to the aromatase inhibitor Anastrozole with binding energy of -7.301 Kcal/mol. The phorbol diesters having higher binding energies implies a more stable protein-ligand complex than the commercially available aromatase inhibitors. This was in line with the findings of Rampogu *et al.*, where their results showed that the 81 compounds they looked at had interaction energies that were higher than those of known drug candidates.²⁵⁷ The ligands evaluated in this study docked in the same binding pocket interacting with similar amino acid residues. This pose was validated by redocking the endogenous ligand testosterone. The binding pocket identified in

this is also consistent with the binding pocket identified in previous studies.^{257, 258} A previous study also showed that amino acid residues like A307, V370, T310, F134, F221, W224, A306, S478, M374, D309, L372 and L477 play significant roles in interaction of aromatase inhibitors and the aromatase enzyme.²⁵⁸

An important pharmacodynamic endpoint in the drug discovery process is protein-ligand binding affinity.²⁵⁹ In molecular docking and rational drug discovery, predicting protein-ligand binding affinity is crucial.²⁶⁰ Binding affinity is a measure of the strength of between binding drug molecule(s) and protein(s).²⁶¹ In thermodynamic terms, the binding affinity of any complex is essential in determining whether or not a reaction will take place within a cell under certain conditions.²²² The PRODIGY-LIG webserver determines a complex's binding affinity in Kcal/mol by utilizing atomic contacts between the protein and the ligand within a distance of 10.5 Å.¹⁹⁵ The findings of the binding affinity analysis showed that the phorbol diesters (P-12,13-diAcPh, -13.7 Kcal/mol; P-12,13-diiBu, -11.5 Kcal/mol; P-12AcPh-13iBu, -12.9 Kcal/mol; P-12Ang-13iBu, -11.8 Kcal/mol; P-20Ac-12AcPh-13iBu, -13.0 Kcal/mol and P-20Ac-12Ang-13iBu, -12.3 Kcal/mol) had higher binding affinities than the commercial inhibitors (Letrozole, -8.2 Kcal/mol; Anastrozole, -8.4 Kcal/mol and Exemestane, -10.3 Kcal/mol) studied. The binding affinity results shown in Table 5.5 demonstrates the thermodynamic feasibility of the docked complexes. The results in Table 5.5, suggests that P-12,13-diAcPh would bind strongly to the binding pocket of the human aromatase enzyme.

Most of the interactions observed in the P-12,13-diAcPh-5JKW complex were alkyl and pi-alkyl. The residues Arg 145, Ala 438 and Glu 302 formed hydrogen bonds with the ligand (Figure 5.5A). Most of the residues present in the binding site were non-polar (Trp 141, Ile132, Ala 438, Ala 306, Val 370, and Phe 430). The residues Arg 435 and Arg 145 are polar and positively charged. The residue Arg 435 plays a significant role in heme binding in the active site of CYP19A1.²⁶² The residues Glu 302 and Cys 437 are negatively charged and polar respectively (Figure 6.2A). The residue Glu 302 is a highly conserved residue in the active site of CYP19A1, it plays a significant role in the inhibition of aromatase.^{263, 264} P-12,13-diiBu interacted with mostly non-polar residues except the residue Cys 437 (Figure 5.5B). Hydrogen bonds were formed between Cys 437, Ala 438, Ala 307, Gly 439, Ala 306 and Met 303, the rest of the residues formed alkyl and pi-alkyl interactions. The ligand P-12AcPh-13iBu interacted with many non-polar residues. The residues His 480, Arg 115 and Cys 437 were the only polar residues interacting with this ligand (Figure 5.5C). Alkyl and pi-alkyl interactions were observed with most of the residues except Ala 438, Arg 115 and Phe 430 which formed hydrogen bonds with the ligand P-12AcPh-13iBu.

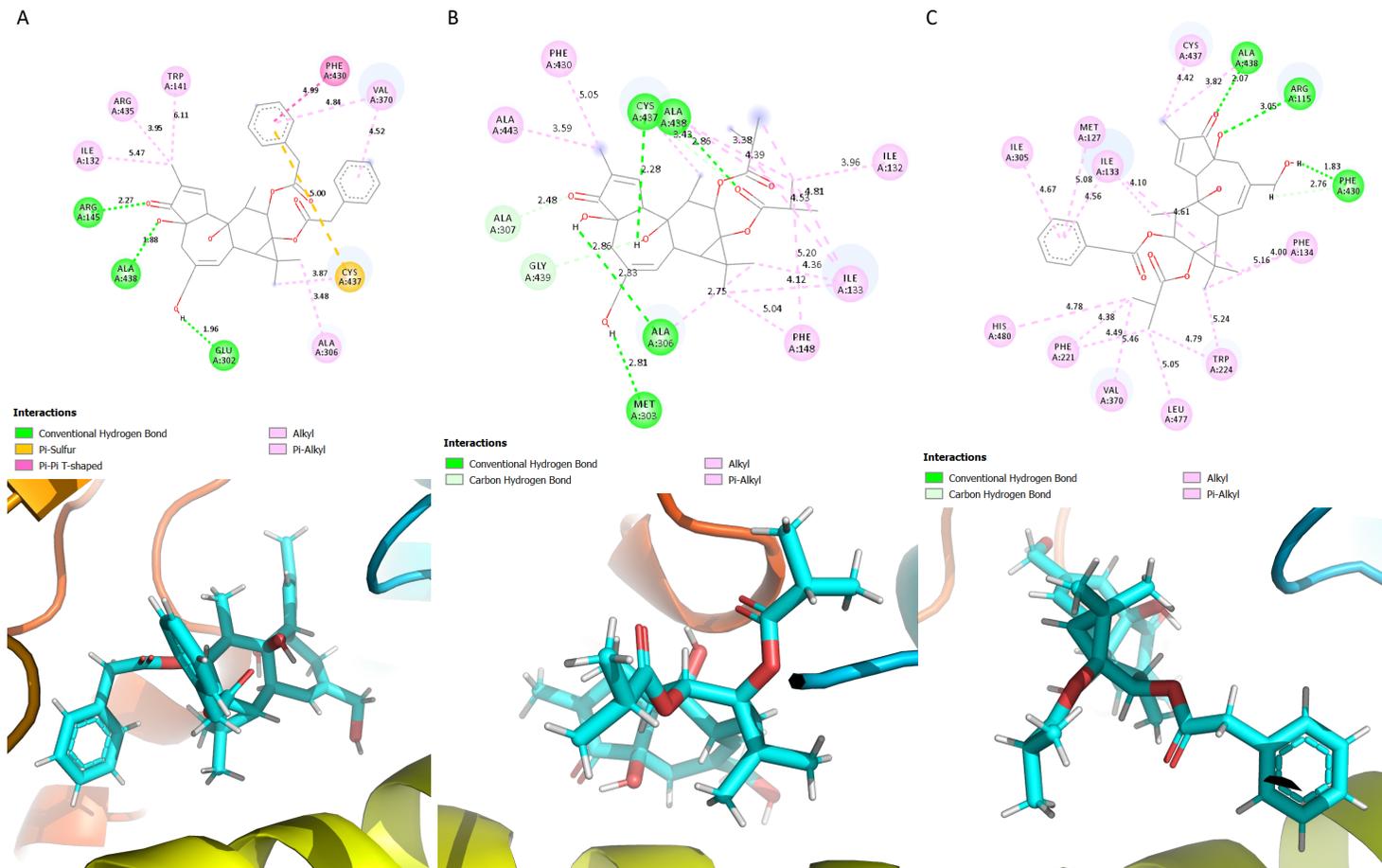


Figure 5.5. Schematic showing the protein ligand interactions (A) P-12,13-diAcPh (B) P-12,13-diiBu (C) P-12AcPh-13iBu.

The ligands P-12Ang-13iBu, P-20Ac-12AcPh-13iBu and P-20Ac-12Ang-13iBu interacted with non-polar residues except the residue Cys 437 which is a polar residue (Figure 6.3). The ligand P-12Ang-13iBu formed a carbon-hydrogen interaction with the residues Ala 307 and Gly 439 (Figure 5.6A). Hydrogen bonds were also formed with Cys 437, Ala 438, Ala 306 and Met 303 (Figure 5.6A). Hydrogen bonds were formed with Ala 438, Cys 437 and Ala 306 and the phorbol diester P-20Ac-12AcPh-13iBu (Figure 5.6B). Ala 307 and Gly 439 formed carbon hydrogen interactions with the ligand, the rest of the interacting residues formed alkyl and pi-alkyl interactions (Figure 5.6B). Ala 307 and Gly 439 formed carbon hydrogen bonds with the ligand, Ala 438, Cys 437 and Ala 306 formed hydrogen bonds with the phorbol diester P-20Ac-12Ang-13iBu (Figure 5.6C).

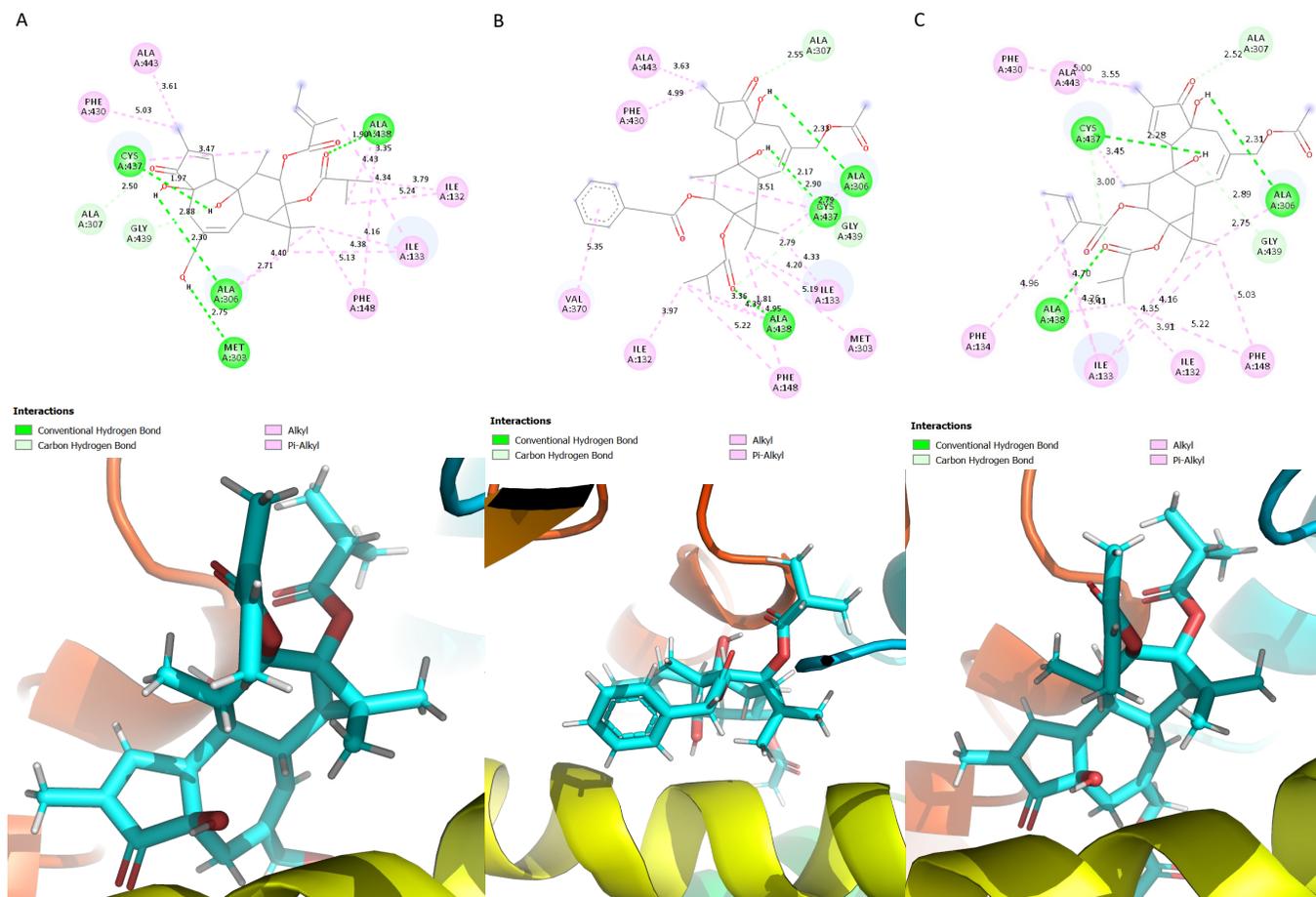


Figure 5.6. Schematic showing the protein ligand interactions (A) P-12Ang-13iBu (B) P-20Ac-12AcPh-13iBu (C) P-20Ac-12Ang-13iBu.

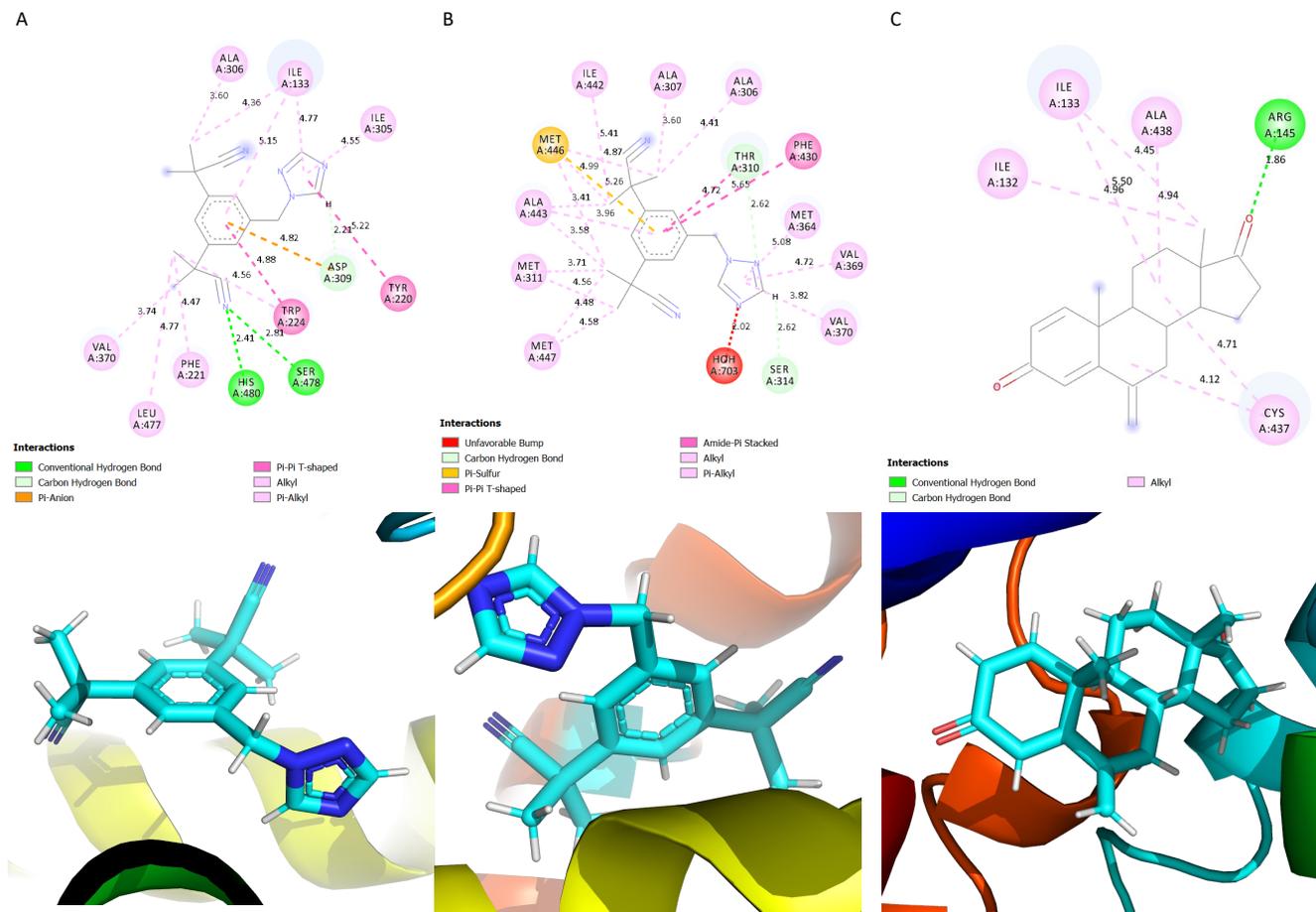


Figure 5.7. Schematic showing the protein ligand interactions (A) Anastrozole (B) Letrozole (C) Exemestane.

The aromatase inhibitor anastrozole interacted with mostly non-polar residues except for Tyr 220, Ser 478 and His 480 which were the only polar residues interacting with the ligand (Figure 5.7A). A carbon-hydrogen bond and pi-anion interaction was observed between the benzene ring of the ligand anastrozole and the amino acid residue Asp 309 (Figure 5.7A). Hydrogen bonds were formed between His 480 and Ser 478 and the residues Trp 224 and Tyr 220 formed a pi-pi t-shaped interaction with the anastrozole (Figure 5.7A). Letrozole interacted with non-polar residues apart from Ser 314 (Figure 5.7B). Carbon-hydrogen bonds were formed with Thr 310 (a pi-pi t-shaped interaction was also formed) and Ser 314 and a Pi-sulphur interaction was formed with the ligand Letrozole (Figure 5.7B). Phe 430 formed an amide-pi stacked interaction with the ligand (Figure 6.4B). The ligand exemestane interacted with only five residues: Ile 133, Ala 438, Arg 145, Cys 437 and Ile 132 (Figure 5.7C). The residue Arg 145 formed a hydrogen bond with the ligand exemestane, the other interacting residues formed alkyl interactions with the ligands. The findings of this study are in line with those of a previous study regarding the residues that were identified.²⁵⁸

The endogenous ligand testosterone interacted with the following residues: Met 374, Arg 115, Val 373, Val 370, Asp 309, Ala 306, Trp 224, Leu 477 and the heme molecule. The phorbol diesters also interacted with some of the same amino acid residues (Ala 306, Val 370, Leu 477, Trp 224, Arg 115, Asp 309) the endogenous ligand testosterone (TST) interacted within the CYP19A1 binding site. This confirmed that a number of the previously described binding residues of TST (the endogenous ligand), interacted with the phorbol diesters and commercially available drugs (Letrozole, Exemestane, and Anastrozole) upon docking to the aromatase enzyme binding site.²⁶⁵ Hydrophobic side chains Trp 224, Phe 221, Phe 134, Ile 133, Val 370, and Met 374 sur-rounded the ligand binding site, shaping the active site cleft to precisely match the TST structure.²⁶⁵ In addition to the heme group cofactor, these residues are important for ligand binding.^{265, 266}

5.5 Conclusion

This study was focused on the interactions between CYP1A1 and CYP1B1 and some flavones. The ADMET properties of some of the ligands were also investigated *in silico*. The binding energies generated from the docking showed that the ligands Isorhamnetin and

Pedalitin had the lowest binding energies in the crystal structures 6DWM and 6IQ5 with energies -10.152 and -10.823, respectively. The amino acid residues Asp 313 and Phe 224 present in 6DWM interacted with all the ligands investigated. However, in the crystal structure 6IQ5, only the amino acid residue Ala 330 interacted with all the ligands investigated. This would suggest that these residues may play a crucial role in the activation of these enzymes. The ligands did not violate any drug-likeness parameters. However, chrysosplenetin and casticin would not be considered potential leads because their molecular weights were greater than 350 g/mol.

The physicochemical properties of the lead compounds showed that the lead compounds (homoeriodictyol, isorhamnetin, pedalitin and quercetin) were soluble (Table 5.3). The ligand quercetin had the highest total polar surface area (TPSA) due to the presence of polar carbonyl oxygen.²⁵⁶ The result of pharmacokinetic predictions carried out using ADMETLab showed that three of the leads (chrysosplenetin, casticin and pedalitin) were positive for HIA (human intestinal absorption) (Table 5.4). Furthermore, the model predicts that over 30% of the compounds are absorbed in the intestines (Table 5.4). The ADMET results showed that the ligands inhibited the enzyme CYP1A2, while casticin, homoeriodictyol and pedalitin were also substrates of the enzyme. The ligands (casticin, isorhamnetin and quercetin) are non-inhibitors and non-substrate of the enzyme CYP2D6 (Table 5.4). Therefore, the ligands isorhamnetin and pedalitin would need to be further characterised to understand their mechanism of action. To understand further which of the residues play vital role in the substrate-binding site, additional structural studies of the active sites of CYP1A1 and CYP1B1 would be carried out.

The aromatic C18 oestrogens (estrone and oestradiol) are produced by the enzyme aromatase in the steroid biosynthesis pathway from C19 androgens. The inhibition of CYP19A1 activity would potentially lead to high androgen levels which is usually linked to some disorders such as ovarian cancer, infertility, and prostate cancer. Previous research has described the functions of protein kinase C (PKC) and phorbol diesters. Cytotoxic 4-phorbol esters were used to target cancerous tissues. The findings demonstrated that the phorbol prodrug effectively eliminated peptidase-positive and -negative cells and activated PKC.²⁶⁷ Tsai *et al.*, studied the activation of PKC by phorbol esters using platelet aggregation as a

model. The results of the study by Tsai and group, showed that platelets are a useful model to study natural PKC activators.²⁶⁸ The interactions between the aromatase and phorbol diesters have been investigated herein using molecular docking as a coarse-grained computational analytical technique. The findings of this computational study suggest that phorbol-CYP19A1 complexes can display higher binding energies than currently available drugs.

The residue Ala 438 formed hydrogen bond interactions with the phorbol diesters. However, no interaction with the commercially available treatment options except for the ligand exemestane which formed an alkyl interaction with the residue. The phorbol diesters in this study formed alkyl interactions with the residue Phe 430 except P-12AcPh-13iBu which formed a hydrogen bond interaction with the residue. A pi-pi T-shaped interaction was however, observed with P-12,13-diAcPh and Letrozole. The reoccurrence of these residues would suggest that they play a role in binding of the phorbol diesters to CYP19A1. The results of the binding affinity analysis showed the ligand P-12,13-diAcPh had the highest binding affinity value (-13.7 Kcal/mol). This suggests a very strong binding between P-12,13-diAcPh and the human aromatase enzyme. The ligand P-12,13-diAcPh, also showed a high binding energy (-13.993 Kcal/mol) from the docking calculation.

Analysis of the ligand binding pocket showed that majority of the residues interacting with the ligands were all hydrophobic. This is because the highest region of conservation in cytochrome P450 enzymes is the binding pocket, and this binding pocket contains the heme moiety and hydrophobic residues which accommodate the heme group.²⁶⁹ Anam *et al.* reported that the binding pocket of the aromatase enzyme needs to comprise of primarily non-polar residues to facilitate the conversion of hydrophobic androstenedione to estrone.²⁷⁰ In conclusion, our findings show that using a coarse-grained technique, the phorbol diesters were bound to the accurate binding pocket and may be able to potentially mediate the activity of CYP19A1. This is subject to further characterisation using advanced methods such as molecular dynamics simulations, ligand binding assays and breast cancer cell lines to determine the mechanism of action of these phorbol diesters.

6. On the calculation of NMR chemical shifts using density functional theory.

This chapter explores the effects of the lack of rotational invariance of some DFT integration grids on DFT computed NMR shielding. Based on the work of Bootsma and Wheeler, the lack of rotational invariance of a number of popular integration grids could affect the relative free energies of molecules with low vibrational frequencies.

6.1 Introduction

Nuclear magnetic resonance (NMR) spectroscopy in condensed matter was independently discovered by Nobel Laureates Edward Purcell and Felix Bloch in 1946.^{271, 272} They observed that magnetic nuclei such as ^1H and ^{31}P can absorb energy in a radio-frequency band upon exposure to magnetic energy of a known strength.²⁷² Prior to Purcell and Bloch's discovery, Wolfgang Pauli hypothesised the existence of nuclear spin in 1924 with the additional concept of electron spin theorised by Uhlenbeck and Goudsmit a year later.²⁷³ Subsequently, in 1937, Rabi built upon the Stern-Gerlach experiment to measure a nuclear magnetic moment at Columbia university.²⁷³ These concepts developed the foundation of chemical analysis through NMR used today.²⁷²

NMR spectroscopy has become an essential tool with diverse applicability in the structural characterisation and determination of a chemical's structure based upon the magnetic properties of their atomic nuclei.²⁷⁴⁻²⁷⁷ NMR spectroscopy is focused on the exact nuclei which demonstrate magnetic dipoles as a result of the juxtaposition of mechanical spin and electric charge.²⁷² NMR-active nuclei are those whose spin quantum number (I) $>$ 0; thus, yielding NMR spectra.²⁷² The exact spin quantum number of any given nucleus is dependent on the atomic number and mass (Z) number.²⁷⁴ If the atomic and mass numbers of an element are even, $I = 0$,^{272, 274} when Z is odd and the atomic numbers are even, ($I = \frac{n}{2}$; where n denotes an integer), and ^{272, 274} when Z is even, and the atomic number is odd ($I=n$).^{272, 274}

NMR chemical shifts contain an abundance of information required for structural determination, including calculated magnetic shielding constants.²⁷⁸ The comparison of calculated and experimental chemical shifts for a molecule can be utilised in the

determination and understanding of its specific geometry, as well as its interactions with neighbouring nuclei.²⁷⁸

Ab initio techniques and density functional theory are utilised in the calculation of NMR chemical shifts and can provide insight into the factors that may impact an NMR chemical shift value.^{276, 278-280} In line with the IUPAC convention, the chemical shift is defined as the following:

$$\delta_{\mathbf{x}} = \frac{(v_{\mathbf{x}}^{\text{S}} - v_{\mathbf{x}}^{\text{R}})}{v_{\mathbf{x}}^{\text{R}}} = \frac{(\sigma_{\mathbf{x}}^{\text{R}} - \sigma_{\mathbf{x}}^{\text{S}})}{(1 - \sigma_{\mathbf{x}}^{\text{R}})} \approx (\sigma_{\mathbf{x}}^{\text{R}} - \sigma_{\mathbf{x}}^{\text{S}})^{281} \quad (1)$$

The NMR chemical shift is highly responsive to the electronic environment which surrounds the nucleus being investigated. It is influenced by the intramolecular and intermolecular effects, including hydrogen bond distance and molecular geometry.^{276, 278} As chemical shifts are sensitive to chemical geometry, a shielding space needs to be determined prior to the calculation of average chemical shifts. This space represents the plot of a chemical shift as a function of the geometric parameter.²⁸²

When determining all the possible conformations that are present at a given time in each geometrical arrangement in solution, two distinct methodologies can be utilised. The first method calculates the energy of all possible conformations and determines the subsequent prevalence of each species according to the Boltzmann distribution.²⁷⁸ Alternatively, the second method bases itself in the determination of the distribution of molecules among the various conformations through molecular dynamics (MD) simulation.²⁷⁸ Thus, a mean of all possible conformations through the Boltzmann distribution corresponds to an ensemble average of molecular conformations, whereas a mean of snapshots obtained from an MD simulation represents a time average.²⁷⁸

In the evaluation of chemical shifts and coupling constants, molecular conformers are weighted to scale their contributions to the magnetic properties being calculated commonly done by employing a Boltzmann analysis based on the relative free energies of these molecular conformers.²⁸³

In 2019, Bootsma and Wheeler, reported concerns with respect to the evaluation of computed free energies for stereoselective and regioselective reactions using variable DFT integration grids.^{283, 284} Bootsma and Wheeler evaluating a small sample size, sampled a selection of grid points in DFT calculation of torsions (electronic energies of 2-butyne

relative to 1,3-butadiene), diastereomers (human immunodeficiency virus integrase inhibitor), and transition structures (torsion of 4,4'-dimethyl-1,1'-biphenyl), and selective reactions requiring the estimation of relative free energies.²⁸⁴ The findings of their study showed that the qualitative change in predictions of DFT-computed free energies for systems featuring low-frequency vibrational modes can be influenced by the molecular orientation.²⁸⁴ They identified differences in relative free energies up to 5 Kcal mol⁻¹; a number which exceeds what is often described as discriminatory for theoretical property determination and barriers in catalytic cycles.^{283, 284} These discrepancies in the computed free energies can pose a crucial impact on the outcome of calculations which determine reactivity property relationships.²⁸³ A 1 Kcal mol⁻¹ ΔG in free energy barriers for regioselective and stereoselective reactions can potentially arbitrate selectivity.^{283, 284}

The potential issues of solely employing DFT methods have been highlighted in the past without an extensive assessment of the errors.^{283, 285, 286} However, Bootsma and Wheeler provided a robust insight into the precision and accuracy of computed free energies using different grids.^{283, 284} Bootsma and Wheeler reported that these errors were as a result of how DFT handles molecules with low vibrational frequencies and can be mitigated by using denser integration grids such as 99,590 as a minimum.^{284, 287}

This chapter investigates the lack of rotational invariance of some DFT integration grids on DFT computed NMR shielding. This is based on a research carried out by Bootsma and Wheeler which reported that the lack of rotational invariance of popular DFT integration grids shows large unreliability in computed free energies for some isomerisation, regio- and stereoselective and torsional barrier reactions.²⁸⁴

6.2 Computational Details

Geometry optimisation to a minimum and transition state of 4,4'-dimethyl-1,1'-biphenyl (Cartesian coordinates were obtained from the supplementary information of this paper by Bootsma and Wheeler.²⁸⁴) (Figure 6.1) were performed on Gaussian 16²⁸⁸ using four integration grids: (50,194), (75,302), (99,590), and (175,974) assessable in Gaussian using the following grid keywords 'SG1grid', 'Fine', 'Ultrafine' and 'Superfine'.

All the DFT calculations were carried out using water as solvent using B97D-Becke's exchange functional¹⁷ with Grimme's D2 dispersion¹¹ and the basis set

Def2TZVP.^{45, 46} The NMR shielding was calculated using the gauge independent atomic orbitals method (GIAO) and spin-spin coupling enabled.

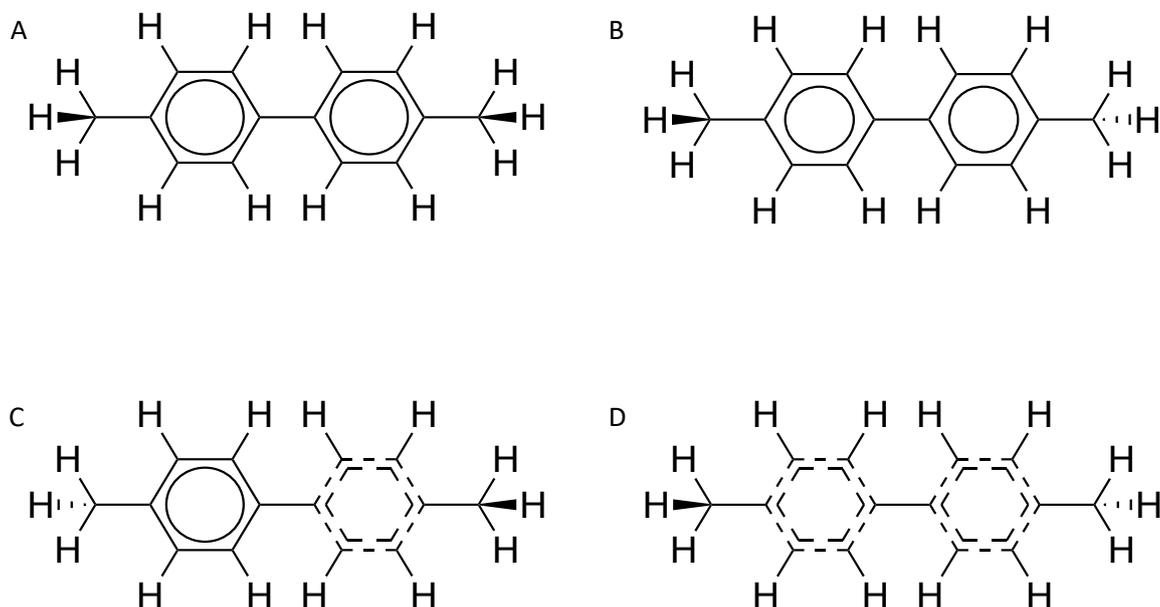


Figure 6.1. Schematic showing 4,4'-dimethyl-1,1'-biphenyl rotated at different angles on the X cartesian axis. (A). 0 degrees (B). 20 degrees (C). 40 degrees (D). 70 degrees.

6.3 Results and Discussions

The energy of a molecule ideally should not be dependent on its orientation relative to the lab-fixed cartesian coordinate system in the absence of an external field.²⁸⁴ DFT computed energies quintessentially lack rotational invariance.²⁸⁴ This is because the atom centred integration grids used in most quantum chemistry software are anchored to the cartesian axis.²⁸⁴ Bootsma and Wheeler reported that the lack of rotational invariance can be negligible in the relative electronic energy of some small systems and DFT functionals.²⁸⁴ Bootsma and Wheeler showed that variation of relative energy in the B97-D/def2-TZVP electronic energy (E) of 2-butyne relative to 1,3-butadiene as a function of the rotation of 2-butyne around an axis perpendicular to the molecule using a pruned (75,302) grid is 0.01kcal mol⁻¹.²⁸⁴

Bootsma and Wheeler, quantifying the rate of reaction of 2-butyne relative to 1,3-butadiene, which requires the torsional free energy barrier (ΔG^\ddagger) of 4,4'-dimethyl-1,1'-biphenyl.²⁸⁴ Their findings showed that using the (75,302) grid, the computed torsional free

energy barrier can be anywhere from 0.8 kcal mol⁻¹ to 3.5 kcal mol⁻¹.²⁸⁴ The same calculation using (50,194) or SG1 grid showed that the computed torsional free energy barrier can be between 1.0 kcal mol⁻¹ to 3.3 kcal mol⁻¹.²⁸⁴

Table 6.1. Minimum and maximum relative NMR shielding values (ppm) of optimised 4,4'-dimethyl-1,1'-biphenyl for four pruned integration grids across different orientations (x axis rotations)

NMR	50,194		75,302		99,590		175,974	
	Min	Max	Min	Max	Min	Max	Min	Max
¹ H	0.9901	1.0093	0.9999	1.0000	0.9999	1.0000	0.9999	1.0000
¹³ C	0.9984	1.0024	0.9468	1.0000	0.9999	1.0000	0.9999	1.0000

Table 6.1 shows the minimum and maximum values for DFT calculated NMR shielding which was obtained by varying the molecular orientation of 4,4'-dimethyl-1,1'-biphenyl in increments of 10 degrees on the x axis using four popular integration grids. The findings of this study showed that using the (50,194) grid, the relative DFT calculated ¹H NMR shielding for 4,4'-dimethyl-1,1'-biphenyl across different orientations can range from 0.990109566 ppm to 1.00936252 ppm (Table 6.1, Table S11). The DFT calculated ¹³C NMR shielding for 4,4'-dimethyl-1,1'-biphenyl across different orientations can range from 0.998432039 ppm to 1.002434307 ppm using the (50,194) grid (Table 6.1, Table S11). Using the (75,302) grid, the relative shielding value for DFT calculated ¹H for 4,4'-dimethyl-1,1'-biphenyl across different orientations ranged from 0.99997622 ppm to 1.0000668 ppm (Table 6.1, Table S12). ¹³C NMR shielding for 4'-dimethyl-1,1'-biphenyl across different orientations ranged from 0.9468023 ppm to 1.00003293 ppm (Table 6.1, Table S12).

In this study, the result also showed that using the (99,590) grid the relative DFT calculated ¹H NMR shielding for 4,4'-dimethyl-1,1'-biphenyl across different orientations can range from 0.999972738 ppm to 1.000023866 ppm (Table 6.1, Table S13). The DFT calculated ¹³C NMR shielding for 4,4'-dimethyl-1,1'-biphenyl across different orientations can range from 0.999976579 ppm to 1.00000706 ppm using the (99,590) grid (Table 6.1, Table S13). Using the (175,974) grid, the relative shielding value for DFT calculated ¹H for 4,4'-dimethyl-1,1'-biphenyl across different orientations ranged from 0.999979553 ppm to

1.000017047 ppm (Table 6.1, Table S14). ¹³C NMR shielding for 4'-dimethyl-1,1'-biphenyl across different orientations ranged from 0.999985942 ppm to 1.000005627 ppm using the (175,974) grid (Table 6.1, Table S14).

The findings on Table S11 (50,194) and Table S12 (75,302) showed a difference in shielding values for the different atoms across the different orientations albeit not being significant. Using the (99,590) and (175,974) grid, the shielding constants for the different atoms remained consistent across the different orientations (Table S13 and Table S14). No increment in value was observed after rotating 4,4'-dimethyl-1,1'-biphenyl in increments of 10 degrees on the x axis (Table S13).

Table 6.2. Minimum and maximum relative NMR shielding values (ppm) of transition state 4,4'-dimethyl-1,1'-biphenyl for four pruned integration grids across different orientations (x axis rotations).

NMR	50,194		75,302		99,590		175,974	
	Min	Max	Min	Max	Min	Max	Min	Max
¹ H	0.9908	1.0128	0.9910	1.0092	0.9911	1.0097	0.9911	1.0090
¹³ C	0.9984	1.0013	0.9993	1.0004	0.9997	1.0002	0.9998	1.0001

The findings in table 6.2 shows the minimum and maximum relative NMR shielding values (ppm) of transition state 4,4'-dimethyl-1,1'-biphenyl for four pruned integration grids across different orientations (x axis rotations). These rotations were performed in increments of 10 degrees. Using the (50,194) grid, the relative shielding value for DFT calculated ¹H for 4,4'-dimethyl-1,1'-biphenyl across different orientations ranged from 0.990838998 ppm to 1.012884074 ppm (Table 6.2, Table S15). ¹³C NMR shielding for 4'-dimethyl-1,1'-biphenyl across different orientations ranged from 0.998418085 ppm to 1.001335414 ppm using the (50,194) grid (Table 6.2, Table S15). The findings of this study showed that using the (75,302) grid, the relative DFT calculated ¹H NMR shielding for 4,4'-dimethyl-1,1'-biphenyl across different orientations can range from 0.991015935 ppm to 1.00923651 ppm (Table 6.2, Table S16). The DFT calculated ¹³C NMR shielding for 4,4'-dimethyl-1,1'-biphenyl across different orientations can range from 0.999372477 ppm to 1.000444354 ppm using the (75,302) grid (Table 6.2, Table S16).

Using the (99,590) grid, the relative shielding value for DFT calculated ^1H for 4,4'-dimethyl-1,1'-biphenyl across different orientations ranged from 0.991108008 ppm to 1.009782732 ppm (Table 6.2, Table S17). ^{13}C NMR shielding for 4'-dimethyl-1,1'-biphenyl across different orientations ranged from 0.999732885 ppm to 1.000201448 ppm using the (99,590) grid (Table 6.2, Table S17). In this study, the findings also showed that using the (175,974) grid the relative DFT calculated ^1H NMR shielding for 4,4'-dimethyl-1,1'-biphenyl across different orientations can range from 0.991104633 ppm to 1.00907832 ppm (Table 6.2, Table S18). The DFT calculated ^{13}C NMR shielding for 4,4'-dimethyl-1,1'-biphenyl across different orientations can range from 0.999833529 ppm to 1.000131789 ppm using the (175,974) grid (Table 6.2, Table S18).

The results in tables 6.2, S15, S16, S17 and S18 showed a difference in the relative shielding values for the different atoms across the different orientations. The relative shielding, though being different across the different orientations for the different atoms, the difference between them is not statistically significant (p value > 0.5). The relative shielding values were not consistent with rotations in increments of 10 degrees on the x axis.

The results also showed that the relative NMR shielding of optimised 4,4'-dimethyl-1,1'-biphenyl did not vary by a large number when the (99,590 and 175,974) grid was used. This was also observed when 4,4'-dimethyl-1,1'-biphenyl was optimised to a transition state using the (99,590 and 175,974) grid. The findings of Bootsma and Wheeler showed that the variations in free energies are larger and gradually converge in relation to the grid size.²⁸⁴

6.4 Conclusions

The calculation of chemical shifts provides a plethora of information regarding the characterisation of molecules and complex compounds present in a sample. However, prior to these calculations there are several factors that must be considered and accommodated for.^{278, 289}

The lack of rotational invariance in DFT calculations has been recognised.²⁸⁴ Even though quantum chemistry software documentation outlines the limitations of different integration grids on calculations involving molecules with low vibrational frequencies.^{284, 287} However, this advice is not heeded by the users.^{284, 287} Accurate relative free energy values are crucial for the application of quantum chemistry to (bio)organic and

organometallic systems.²⁸⁴ The findings of Bootsma and Wheeler showed that the lack of rotational invariance in popular integration grids can lead to large variations in relative free energies of some systems.²⁸⁴ The findings of Bootsma and Wheeler also showed that these errors were reduced by using a (99,590) grid in many cases.

Wilson, Grootveld and Kamerlin,²⁸³ expressed their concerns on the potential impact of the findings of Bootsma and Wheeler on computational NMR predictions because the estimation of NMR shielding tensors and magnetic susceptibilities which is based on averaging the weights from a Boltzmann analysis of molecular conformers. Wilson, Grootveld and Kamerlin,²⁸³ also emphasized that if the errors identified by Bootsma and Wheeler are not linked to just integration grids, but also to the molecular orientation contributing to free energy changes altering the relative distribution of conformers, then DFT predicted spectra properties can be subject to scrutiny.

The findings of this study showed a difference in the relative shielding values for the different atoms across the different orientations. The relative shielding though being different across the different orientations for the different atoms, the difference between them is not significant (p value > 0.5). It was also observed that using the (99,590) and (175,974) grid, the shielding constants for the different atoms remained consistent across the different orientations (Tables S13 and S14). The relative shielding value of the hydrogen atom with number 13 on the DFT calculated ^1H NMR pruned using (99,590) grid just increased from 1 ppm to 1.000004174 ppm. No increment in value was observed after rotating 4,4'-dimethyl-1,1'-biphenyl in increments of 10 degrees on the x axis (Table S13). The findings of this study addresses the concerns of Wilson, Grootveld and Kamerlin.²⁸³ Herein, it is highlighted that the lack rotational invariance of some DFT integration grids does not have an impact on the DFT computed spectra of a molecule with low vibrational frequency.

7. Overall Conclusions and future work

7.1 Conclusions

The research presented in this thesis utilises both computational and experimental techniques to advance understanding of selected biological and chemical problems. The work provided an insight into the residues which are key in protein ligand interactions, the effects of cartesian axis rotations on NMR shielding.

In conclusion, the characterisation of the allosteric binding pocket of GLP-1R inferred that the allosteric binding pocket of the receptor is in the TM6 of the receptor. The *in silico* ADME/Tox results in chapter 3 showed that some of the ligands exhibited high LogD values and low aqueous solubility. This implies that the ligands would undergo modifications to improve its oral solubility. However, some ligands used for the study in chapter 4 possessed optimum LogP and LogD values of LogP = 1-3²²⁸, LogP = >1 and <4²³² and less 3.5, respectively. The results from the study involving selected EROD flavonoids showed that the ligands isorhamnetin and pedalitin had the lowest binding energies in the enzyme structures 6DWM and 6IQ5, respectively. The results also showed the amino acid residues Asp 313, Phe 224 and Ala 330 interacted with the EROD flavonoids.

Investigating the interactions between phorbol diesters with different chains at C-12, C-13, and C-20, known aromatase inhibitors anastrozole, exemestane, and letrozole, and the human placental aromatase cytochrome P450, the findings indicated that the phorbols (P-12,13-diAcPh, P-12,13-diiBu, P-12AcPh-13iBu, P-12Ang-13iBu, P-20Ac-12AcPh-13iBu, P-20Ac-12Ang-13iBu) exhibited higher binding energies than commercially available drugs.

Furthermore, this study assessed the impact of rotational invariance on DFT computed NMR spectra of low vibrational frequency reactions. The results showed differences in relative shielding values. However, it can be concluded that the lack of rotational invariance does not significantly affect DFT computed NMR spectra.

7.2 Future Work

Based on the findings of this study, the following recommendations for future research are discussed below.

In light of the results presented in Chapter 5, further characterization of the ligands isorhamnetin and pedalitin is warranted to gain a better understanding of their mechanism of action. Additionally, conducting additional structural studies on the active sites of CYP1A1 and CYP1B1 is essential to elucidate the specific residues that play vital roles in the substrate-binding site.

The findings in Chapter 5 also demonstrate that phorbol diesters can bind to the enzyme CYP19A1 and influence its activity. However, further characterization of these ligands is required to enhance our understanding of their impact on CYP19A1.

To further explore the allosteric binding pocket of GLP-1R, future studies could incorporate molecular dynamics (MD) simulations²³³⁻²³⁶, and multiscale quantum mechanics (QM)/molecular mechanics (MM) molecular simulations,¹⁷⁰ using known allosteric modulators obtained from a literature search. Additionally, mutagenesis studies can be conducted to identify the specific residues that significantly contribute to ligand binding. Additionally, mutagenesis studies can be conducted to identify the specific residues that significantly contribute to ligand binding.

However, it should be noted that large-scale molecular dynamics simulations, meta-dynamics, and QM/MM simulations involving enzymes and receptors often encounter limitations due to insufficient computing resources. Therefore, there is a need to develop new methods that can simulate these systems on a larger scale and for longer periods of time to accurately characterize them.

References

1. G. W. Brodland, *Seminars in Cell & Developmental Biology*, 2015, **47-48**, 62-73.
2. C. S. Odoemelum, B. Percival, H. Wallis, M.-W. Chang, Z. Ahmad, D. Scholey, E. Burton, I. H. Williams, C. L. Kamerlin and P. B. Wilson, *RSC Advances*, 2020, **10**, 36337-36348.
3. P. B. Wilson, P. J. Weaver, I. R. Greig and I. H. Williams, *The Journal of Physical Chemistry B*, 2015, **119**, 802-809.
4. S.-s. Ou-Yang, J.-y. Lu, X.-q. Kong, Z.-j. Liang, C. Luo and H. Jiang, *Acta Pharmacologica Sinica*, 2012, **33**, 1131-1140.
5. D. J. Cole and N. D. M. Hine, *Journal of Physics: Condensed Matter*, 2016, **28**, 393001.
6. S. M. LaPointe and D. F. Weaver, *Current Computer-Aided Drug Design*, 2007, **3**, 290-296.
7. R. G. Parr, *Annual Review of Physical Chemistry*, 1983, **34**, 631-656.
8. P. Hohenberg and W. Kohn, *Physical Review*, 1964, **136**, B864-B871.
9. W. Kohn and L. J. Sham, *Physical Review*, 1965, **140**, A1133-A1138.
10. T. van Mourik, M. Bühl and M.-P. Gaigeot, *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 2014, **372**, 20120488-20120488.
11. S. Grimme, *Journal of computational chemistry*, 2006, **27**, 1787-1799.
12. T. Van Voorhis and G. E. Scuseria, *The Journal of Chemical Physics*, 1998, **109**, 400-410.
13. F. A. Hamprecht, A. J. Cohen, D. J. Tozer and N. C. Handy, *The Journal of Chemical Physics*, 1998, **109**, 6264-6271.
14. A. D. Boese, N. L. Doltsinis, N. C. Handy and M. Sprik, *The Journal of Chemical Physics*, 2000, **112**, 1670-1678.
15. A. D. Boese and N. C. Handy, *The Journal of Chemical Physics*, 2001, **114**, 5497-5503.
16. A. D. Becke, *The Journal of Chemical Physics*, 1993, **98**, 5648-5652.
17. A. D. Becke, *The Journal of Chemical Physics*, 1997, **107**, 8554-8560.
18. J. P. Perdew, K. Burke and M. Ernzerhof, *Physical Review Letters*, 1996, **77**, 3865-3868.
19. J. P. Perdew, K. Burke and M. Ernzerhof, *Physical Review Letters*, 1997, **78**, 1396-1396.
20. J. Tao, J. P. Perdew, V. N. Staroverov and G. E. Scuseria, *Physical Review Letters*, 2003, **91**, 146401.
21. N. Mardirossian and M. Head-Gordon, *Molecular Physics*, 2017, **115**, 2315-2372.
22. in *Computational Chemistry and Molecular Modeling: Principles and Applications*, eds. K. I. Ramachandran, D. Gopakumar and K. Namboori, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 115-138.
23. S. Lehtola, *International Journal of Quantum Chemistry*, 2019, **119**, e25968.
24. C. Zener, *Physical Review*, 1930, **36**, 51-56.
25. J. C. Slater, *Physical Review*, 1930, **36**, 57-64.
26. C. Eckart, *Physical Review*, 1930, **36**, 878-892.
27. R.-Q. Zhang and W.-J. Fan, *International Journal of Quantum Chemistry*, 2015, **115**, 570-577.

28. E. Papajak and D. G. Truhlar, *Journal of Chemical Theory and Computation*, 2010, **6**, 597-601.
29. G. Marrazzini, T. Giovannini, M. Scavino, F. Egidi, C. Cappelli and H. Koch, *J Chem Theory Comput*, 2021, **17**, 791-803.
30. W. J. Hehre, R. F. Stewart and J. A. Pople, *The Journal of Chemical Physics*, 1969, **51**, 2657-2664.
31. J. B. Collins, P. von R. Schleyer, J. S. Binkley and J. A. Pople, *The Journal of Chemical Physics*, 1976, **64**, 5142-5151.
32. J. S. Binkley, J. A. Pople and W. J. Hehre, *Journal of the American Chemical Society*, 1980, **102**, 939-947.
33. M. S. Gordon, J. S. Binkley, J. A. Pople, W. J. Pietro and W. J. Hehre, *Journal of the American Chemical Society*, 1982, **104**, 2797-2803.
34. W. J. Pietro, M. M. Francl, W. J. Hehre, D. J. DeFrees, J. A. Pople and J. S. Binkley, *Journal of the American Chemical Society*, 1982, **104**, 5039-5048.
35. K. D. Dobbs and W. J. Hehre, *Journal of Computational Chemistry*, 1986, **7**, 359-378.
36. K. D. Dobbs and W. J. Hehre, *Journal of Computational Chemistry*, 1987, **8**, 861-879.
37. K. D. Dobbs and W. J. Hehre, *Journal of Computational Chemistry*, 1987, **8**, 880-893.
38. T. H. Dunning, *The Journal of Chemical Physics*, 1989, **90**, 1007-1023.
39. R. A. Kendall, T. H. Dunning and R. J. Harrison, *The Journal of Chemical Physics*, 1992, **96**, 6796-6806.
40. K. A. Peterson, D. E. Woon and T. H. Dunning, *The Journal of Chemical Physics*, 1994, **100**, 7410-7415.
41. A. K. Wilson, T. van Mourik and T. H. Dunning, *Journal of Molecular Structure: THEOCHEM*, 1996, **388**, 339-349.
42. D. E. Woon and T. H. Dunning, *The Journal of Chemical Physics*, 1993, **98**, 1358-1371.
43. A. Schäfer, H. Horn and R. Ahlrichs, *The Journal of Chemical Physics*, 1992, **97**, 2571-2577.
44. A. Schäfer, C. Huber and R. Ahlrichs, *The Journal of Chemical Physics*, 1994, **100**, 5829-5835.
45. F. Weigend and R. Ahlrichs, *Physical Chemistry Chemical Physics*, 2005, **7**, 3297-3305.
46. F. Weigend, *Physical Chemistry Chemical Physics*, 2006, **8**, 1057-1065.
47. G. M. Morris and M. Lim-Wilby, in *Molecular modeling of proteins*, Springer, 2008, pp. 365-382.
48. N. S. Pagadala, K. Syed and J. Tuszynski, *Biophysical reviews*, 2017, **9**, 91-102.
49. J. Fan, A. Fu and L. Zhang, *Quantitative Biology*, 2019, 1-7.
50. L. Pinzi and G. Rastelli, *International Journal of Molecular Sciences*, 2019, **20**.
51. X. Tao, Y. Huang, C. Wang, F. Chen, L. Yang, L. Ling, Z. Che and X. Chen, *International Journal of Food Science & Technology*, 2020, **55**, 33-45.
52. M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray and R. D. Taylor, *Proteins: Structure, Function, and Bioinformatics*, 2003, **52**, 609-623.
53. M. Rarey, B. Kramer, T. Lengauer and G. Klebe, *Journal of Molecular Biology*, 1996, **261**, 470-489.

54. W. J. Allen, T. E. Balius, S. Mukherjee, S. R. Brozell, D. T. Moustakas, P. T. Lang, D. A. Case, I. D. Kuntz and R. C. Rizzo, *Journal of computational chemistry*, 2015, **36**, 1132-1156.
55. G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *Journal of computational chemistry*, 2009, **30**, 2785-2791.
56. C. S. Ring, E. Sun, J. H. McKerrow, G. K. Lee, P. J. Rosenthal, I. D. Kuntz and F. E. Cohen, *Proceedings of the National Academy of Sciences of the United States of America*, 1993, **90**, 3583-3587.
57. A. N. Jain, *Journal of Medicinal Chemistry*, 2003, **46**, 499-511.
58. O. Trott and A. J. Olson, *Journal of computational chemistry*, 2010, **31**, 455-461.
59. Cresset[®], *Journal*, 2019.
60. J. Meiler and D. Baker, *Proteins: Structure, Function, and Bioinformatics*, 2006, **65**, 538-548.
61. S. J. Fleishman, A. Leaver-Fay, J. E. Corn, E.-M. Strauch, S. D. Khare, N. Koga, J. Ashworth, P. Murphy, F. Richter and G. Lemmon, *PloS one*, 2011, **6**, e20161.
62. G. Lemmon and J. Meiler, in *Computational Drug Discovery and Design*, Springer, 2012, pp. 143-155.
63. R. Moretti, B. J. Bender, B. Allison and J. Meiler, *Methods Mol Biol*, 2016, **1414**, 47-62.
64. S. Ruiz-Carmona, D. Alvarez-Garcia, N. Foloppe, A. B. Garmendia-Doval, S. Juhos, P. Schmidtke, X. Barril, R. E. Hubbard and S. D. Morley, *PLOS Computational Biology*, 2014, **10**, e1003571.
65. C. R. Corbeil, C. I. Williams and P. Labute, *Journal of Computer-Aided Molecular Design*, 2012, **26**, 775-786.
66. W. Liu, G. Liu, H. Zhou, X. Fang, Y. Fang and J. Wu, *BioMedical Engineering OnLine*, 2016, **15**, 152.
67. Z. Yan and J. Wang, *PloS one*, 2013, **8**, e74443-e74443.
68. W. Jiang, W. Li, Y. Hong, S. Wang and B. Fang, *Applied Biochemistry and Biotechnology*, 2016, **178**, 739-752.
69. S. Roy, B. K. Narang, M. K. Gupta, V. Abbot, V. Singh and R. K. Rawal, *Drug research*, 2018, **68**, 395-402.
70. L. Guan, H. Yang, Y. Cai, L. Sun, P. Di, W. Li, G. Liu and Y. Tang, *Medchemcomm*, 2019, **10**, 148-157.
71. I. Kola and J. Landis, *Nature Reviews Drug Discovery*, 2004, **3**, 711-716.
72. L. L. G. Ferreira and A. D. Andricopulo, *Drug Discovery Today*, 2019, **24**, 1157-1165.
73. A. Laoui and V. R. Polyakov, *Journal of Computational Chemistry*, 2011, **32**, 1944-1951.
74. M. Trunzer, B. Faller and A. Zimmerlin, *Journal of Medicinal Chemistry*, 2009, **52**, 329-335.
75. T. Sander, J. Freyss, M. von Korff and C. Rufener, *Journal of Chemical Information and Modeling*, 2015, **55**, 460-473.
76. M. Segall, *Expert Opinion on Drug Discovery*, 2014, **9**, 803-817.

77. A. V. Rudik, V. M. Bezhentsev, A. V. Dmitriev, D. S. Druzhilovskiy, A. A. Lagunin, D. A. Filimonov and V. V. Poroikov, *Journal of Chemical Information and Modeling*, 2017, **57**, 638-642.
78. J. Dong, N.-N. Wang, Z.-J. Yao, L. Zhang, Y. Cheng, D. Ouyang, A.-P. Lu and D.-S. Cao, *Journal of cheminformatics*, 2018, **10**, 29-29.
79. S. Tian, Y. Djoumbou-Feunang, R. Greiner and D. S. Wishart, *Journal of chemical information and modeling*, 2018, **58**, 1282-1291.
80. D. Lagorce, L. Bouslama, J. Becot, M. A. Miteva and B. O. Villoutreix, *Bioinformatics*, 2017, **33**, 3658-3660.
81. M. A. Miteva, S. Violas, M. Montes, D. Gomez, P. Tuffery and B. O. Villoutreix, *Nucleic acids research*, 2006, **34**, W738-W744.
82. S. Podlowska and R. Kafel, *International journal of molecular sciences*, 2018, **19**, 1040.
83. A. Daina, O. Michielin and V. Zoete, *Scientific Reports*, 2017, **7**, 42717.
84. P. Schyman, R. Liu, V. Desai and A. Wallqvist, *Frontiers in pharmacology*, 2017, **8**, 889.
85. P. M. Dijkman, O. K. Castell, A. D. Goddard, J. C. Munoz-Garcia, C. de Graaf, M. I. Wallace and A. Watts, *Nature communications*, 2018, **9**, 1710-1710.
86. W. K. Kroeze, D. J. Sheffler and B. L. Roth, *Journal of Cell Science*, 2003, **116**, 4867.
87. R. Fredriksson, M. C. Lagerström, L.-G. Lundin and H. B. Schiöth, *Molecular Pharmacology*, 2003, **63**, 1256.
88. H. B. Schiöth and R. Fredriksson, *General and Comparative Endocrinology*, 2005, **142**, 94-101.
89. E. Ghosh, P. Kumari, D. Jaiman and A. K. Shukla, *Nature Reviews Molecular Cell Biology*, 2015, **16**, 69-81.
90. A. S. Hauser, M. M. Attwood, M. Rask-Andersen, H. B. Schiöth and D. E. Gloriam, *Nature reviews. Drug discovery*, 2017, **16**, 829-842.
91. X.-l. Tang, Y. Wang, D.-l. Li, J. Luo and M.-y. Liu, *Acta Pharmacologica Sinica*, 2012, **33**, 363-371.
92. H. R. Kim, N. M. Duc and K. Y. Chung, *Biomolecules & therapeutics*, 2018, **26**, 101-108.
93. K. Sriram and P. A. Insel, *Molecular pharmacology*, 2018, **93**, 251-258.
94. P. A. Insel, K. Sriram, M. W. Gorr, S. Z. Wiley, A. Michkov, C. Salmerón and A. M. Chinn, *Trends in Pharmacological Sciences*, 2019, **40**, 378-387.
95. K. A. Jacobson, S. Costanzi and S. Paoletta, *Trends in pharmacological sciences*, 2014, **35**, 658-663.
96. S. Basith, M. Cui, S. J. Y. Macalino, J. Park, N. A. B. Clavio, S. Kang and S. Choi, *Frontiers in Pharmacology*, 2018, **9**.
97. A. Heifetz, G. F. X. Schertler, R. Seifert, C. G. Tate, P. M. Sexton, V. V. Gurevich, D. Fourmy, V. Cherezov, F. H. Marshall, R. I. Storer, I. Moraes, I. G. Tikhonova, C. S. Tautermann, P. Hunt, T. Ceska, S. Hodgson, M. J. Bodkin, S. Singh, R. J. Law and P. C. Biggin, *Naunyn-Schmiedeberg's archives of pharmacology*, 2015, **388**, 883-903.
98. X. Yuan and Y. Xu, *International journal of molecular sciences*, 2018, **19**, 2105.

99. M. Esguerra, A. Siretskiy, X. Bello, J. Sallander and H. Gutiérrez-de-Terán, *Nucleic Acids Research*, 2016, **44**, W455-W462.
100. V. V. Gurevich and E. V. Gurevich, *International journal of molecular sciences*, 2017, **18**, 2519.
101. P. S. H. Park, *Current medicinal chemistry*, 2012, **19**, 1146-1154.
102. D. Provasi, M. C. Artacho, A. Negri, J. C. Mobarec and M. Filizola, *PLOS Computational Biology*, 2011, **7**, e1002193.
103. P. Samama, S. Cotecchia, T. Costa and R. J. Lefkowitz, *Journal of Biological Chemistry*, 1993, **268**, 4625-4636.
104. S. M. de Munnik, M. J. Smit, R. Leurs and H. F. Vischer, *Frontiers in Pharmacology*, 2015, **6**.
105. B. K. Kobilka, *Biochimica et biophysica acta*, 2007, **1768**, 794-807.
106. J. Bockaert and J. Philippe Pin, *The EMBO Journal*, 1999, **18**, 1723-1729.
107. M. N. Davies, A. Secker, A. A. Freitas, M. Mendao, J. Timmis and D. R. Flower, *Bioinformatics*, 2007, **23**, 3113-3118.
108. F. Horn, E. Bettler, L. Oliveira, F. Campagne, F. E. Cohen and G. Vriend, *Nucleic acids research*, 2003, **31**, 294-297.
109. F. Horn, J. Weare, M. W. Beukers, S. Hörsch, A. Bairoch, W. Chen, Ø. Edvardsen, F. Campagne and G. Vriend, *Nucleic Acids Research*, 1998, **26**, 275-279.
110. G.-M. Hu, T.-L. Mai and C.-M. Chen, *Scientific reports*, 2017, **7**, 15495-15495.
111. G. Pándy-Szekeres, C. Munk, T. M. Tsonkov, S. Mordalski, K. Harpsøe, A. S. Hauser, A. J. Bojarski and D. E. Gloriam, *Nucleic acids research*, 2018, **46**, D440-D446.
112. A. J. Harmar, *Genome Biol*, 2001, **2**, Reviews3013.
113. D. M. Rosenbaum, S. G. F. Rasmussen and B. K. Kobilka, *Nature*, 2009, **459**, 356-363.
114. S. B. Gacasan, D. L. Baker and A. L. Parrill, *AIMS biophysics*, 2017, **4**, 491-527.
115. K. Palczewski, *Annual review of biochemistry*, 2006, **75**, 743-767.
116. M. C. Lagerström and H. B. Schiöth, *Nature Reviews Drug Discovery*, 2008, **7**, 339-357.
117. D. C. Teller, T. Okada, C. A. Behnke, K. Palczewski and R. E. Stenkamp, *Biochemistry*, 2001, **40**, 7761-7772.
118. K. Sasaki, S. Dockerill, D. A. Adamiak, I. J. Tickle and T. Blundell, *Nature*, 1975, **257**, 751-757.
119. C. Parthier, S. Reedtz-Runge, R. Rudolph and M. T. Stubbs, *Trends in Biochemical Sciences*, 2009, **34**, 303-310.
120. M. Wheatley, D. Wootten, M. T. Conner, J. Simms, R. Kendrick, R. T. Logan, D. R. Poyner and J. Barwell, *British journal of pharmacology*, 2012, **165**, 1688-1703.
121. C. R. R. Grace, M. H. Perrin, M. R. DiGruccio, C. L. Miller, J. E. Rivier, W. W. Vale and R. Riek, *Proceedings of the National Academy of Sciences of the United States of America*, 2004, **101**, 12836-12841.
122. V. Karageorgos, M. Venihaki, S. Sakellaris, M. Pardalos, G. Kontakis, M.-T. Matsoukas, A. Gravanis, A. Margioris and G. Liapakis, *Hormones*, 2018, **17**, 45-59.
123. C. de Graaf, G. Song, C. Cao, Q. Zhao, M.-W. Wang, B. Wu and R. C. Stevens, *Trends in Biochemical Sciences*, 2017, **42**, 946-960.

124. F. Wu, L. Yang, K. Hang, M. Laursen, L. Wu, G. W. Han, Q. Ren, N. K. Roed, G. Lin, M. A. Hanson, H. Jiang, M.-W. Wang, S. Reedtz-Runge, G. Song and R. C. Stevens, *Nature Communications*, 2020, **11**, 1272.
125. K. Hollenstein, C. de Graaf, A. Bortolato, M.-W. Wang, F. H. Marshall and R. C. Stevens, *Trends in pharmacological sciences*, 2014, **35**, 12-22.
126. L. Chun, W.-h. Zhang and J.-f. Liu, *Acta pharmacologica Sinica*, 2012, **33**, 312-323.
127. P. Rondard, C. Goudet, J. Kniazeff, J.-P. Pin and L. Prézeau, *Neuropharmacology*, 2011, **60**, 82-92.
128. S. D. Hellyer, S. Albold, T. Wang, A. N. Y. Chen, L. T. May, K. Leach and K. J. Gregory, *Molecular Pharmacology*, 2018, **93**, 504.
129. B.-O. Hans and A. J. Petrine Wellendorph and Anders, *Current Drug Targets*, 2007, **8**, 169-184.
130. C. S. Tautermann, *Bioorganic & Medicinal Chemistry Letters*, 2014, **24**, 4073-4079.
131. B. L. Roth, *Nature Structural & Molecular Biology*, 2019, **26**, 535-544.
132. X. C. Zhang, J. Liu and D. Jiang, *Protein & cell*, 2014, **5**, 492-495.
133. T. C. Møller, D. Moreno-Delgado, J.-P. Pin and J. Kniazeff, *Biophysics Reports*, 2017, **3**, 57-63.
134. D. Zhang, Q. Zhao and B. Wu, *Molecules and cells*, 2015, **38**, 836-842.
135. M. Dong, C. Koole, D. Wootten, P. M. Sexton and L. J. Miller, *British Journal of Pharmacology*, 2014, **171**, 1085-1101.
136. M. Orel, E. Padrós and J. Manyosa, *The FEBS Journal*, 2012, **279**, 2357-2367.
137. A. Bortolato, A. S. Doré, K. Hollenstein, B. G. Tehan, J. S. Mason and F. H. Marshall, *British journal of pharmacology*, 2014, **171**, 3132-3145.
138. K. J. Culhane, Y. Liu, Y. Cai and E. C. Y. Yan, *Frontiers in Pharmacology*, 2015, **6**.
139. P. Tewatia, N. Agrawal, M. Gaur and S. Sahi, *Biochimie*, 2014, **101**, 168-182.
140. B. Trzaskowski, D. Latek, S. Yuan, U. Ghoshdastider, A. Debinski and S. Filipek, *Current medicinal chemistry*, 2012, **19**, 1090-1109.
141. W. I. Weis and B. K. Kobilka, *Annual review of biochemistry*, 2018, **87**, 897-919.
142. D. Wacker, R. C. Stevens and B. L. Roth, *Cell*, 2017, **170**, 414-427.
143. P. R. Gentry, P. M. Sexton and A. Christopoulos, *Journal of Biological Chemistry*, 2015, **290**, 19478-19488.
144. H. N. Motlagh, J. O. Wrabl, J. Li and V. J. Hilser, *Nature*, 2014, **508**, 331-339.
145. T. Ruth Nussinov and Chung-Jung, *Current Pharmaceutical Design*, 2012, **18**, 1311-1316.
146. N. Tuteja, *Plant signaling & behavior*, 2009, **4**, 942-947.
147. C. D. Hanlon and D. J. Andrew, *Journal of cell science*, 2015, **128**, 3533-3542.
148. P. J. Conn, A. Christopoulos and C. W. Lindsley, *Nature reviews. Drug discovery*, 2009, **8**, 41-54.
149. E. C. Tibaduiza, C. Chen and M. Beinborn, *Journal of Biological Chemistry*, 2001, **276**, 37787-37793.
150. Y.-W. Li, L. Fitzgerald, H. Wong, S. Lelas, G. Zhang, M. D. Lindner, T. Wallace, J. McElroy, N. J. Lodge, P. Gilligan and R. Zaczek, *CNS Drug Reviews*, 2005, **11**, 21-52.
151. T. Z. Baram, D. T. Chalmers, C. Chen, Y. Koutsoukos and E. B. De Souza, *Brain research*, 1997, **770**, 89-95.

152. J. R. Williamson, *Nature Chemical Biology*, 2008, **4**, 458-465.
153. T. Lenaerts, J. Ferkinghoff-Borg, J. Schymkowitz and F. Rousseau, *BMC systems biology*, 2009, **3**, 9-9.
154. I. G. Denisov and S. G. Sligar, *Archives of biochemistry and biophysics*, 2012, **519**, 91-102.
155. Y.-J. I. Jong, S. K. Harmon and K. L. O'Malley, *British journal of pharmacology*, 2018, **175**, 4026-4035.
156. V. Zachariou, R. S. Duman and E. J. Nestler, in *Basic Neurochemistry (Eighth Edition)*, eds. S. T. Brady, G. J. Siegel, R. W. Albers and D. L. Price, Academic Press, New York, 2012, pp. 411-422.
157. H. Schulman, in *From Molecules to Networks (Third Edition)*, eds. J. H. Byrne, R. Heidelberg and M. N. Waxham, Academic Press, Boston, 2014, pp. 119-148.
158. J. Doijen, T. Van Loy, B. Landuyt, W. Luyten, D. Schols and L. Schoofs, *Biosensors and Bioelectronics*, 2019, **137**, 33-44.
159. G. J. Augustine, eds. D. Purves, G. Augustine, D. Fitzpatrick, L. Katz, A.-S. LaMantia, J. McNamara and M. Williams, Sinauer Associates, Sunderland MA, Third edn., 2004.
160. A. Warshel and A. Papazyan, *Current Opinion in Structural Biology*, 1998, **8**, 211-217.
161. S. E. Braslavsky, *Pure and Applied Chemistry*, 2007, **79**, 293-465.
162. M. Amin and J. Küpper, *ChemistryOpen*, 2020, **9**, 691-694.
163. M. Amin and J. Küpper, *arXiv e-prints*, 2020, arXiv:2001.07053.
164. C. N. Schutz and A. Warshel, *Proteins: Structure, Function, and Bioinformatics*, 2001, **44**, 400-417.
165. A. S. Alshami, J. Tang and B. Rasco, *Food and Bioprocess Technology*, 2017, **10**, 1548-1561.
166. A. Warshel and J. Aqvist, *Annual Review of Biophysics and Biophysical Chemistry*, 1991, **20**, 267-298.
167. L. Li, C. Li, Z. Zhang and E. Alexov, *Journal of Chemical Theory and Computation*, 2013, **9**, 2126-2136.
168. M. Jaiteh, I. Rodríguez-Espigares, J. Selent and J. Carlsson, *PLOS Computational Biology*, 2020, **16**, e1007680.
169. D. Hilger, M. Masureel and B. K. Kobilka, *Nature Structural & Molecular Biology*, 2018, **25**, 4-12.
170. P. Nakliang, R. Lazim, H. Chang and S. Choi, *Biomolecules*, 2020, **10**, 631.
171. M. Congreve, C. de Graaf, N. A. Swain and C. G. Tate, *Cell*, 2020, **181**, 81-91.
172. C. de Graaf, C. Rein, D. Piwnica, F. Giordanetto and D. Rognan, *ChemMedChem*, 2011, **6**, 2159-2169.
173. A. Ciancetta and K. A. Jacobson, in *Computational Methods for GPCR Drug Discovery*, ed. A. Heifetz, Springer New York, New York, NY, 2018, pp. 45-72.
174. F. Ballante, A. Rudling, A. Zeifman, A. Luttens, D. D. Vo, J. J. Irwin, J. Kihlberg, J. Brea, M. I. Loza and J. Carlsson, *Journal of Medicinal Chemistry*, 2020, **63**, 613-620.
175. D. Bassani, M. Pavan, M. Sturlese and S. Moro, *Journal*, 2022, **15**.
176. J. Zhang, S. Gu, X. Sun, W. Li, Y. Tang and G. Liu, *RSC Advances*, 2016, **6**, 13490-13497.

177. A. M. Bodnaruc, D. Prud'homme, R. Blanchet and I. Giroux, *Nutrition & metabolism*, 2016, **13**, 92-92.
178. R. Sharma, T. S. McDonald, H. Eng, C. Limberakis, B. D. Stevens, S. Patel and A. S. Kalgutkar, *Drug Metabolism and Disposition*, 2013, **41**, 2148.
179. J. Li, J. Zheng, S. Wang, H. K. Lau, A. Fathi and Q. Wang, *Frontiers in Physiology*, 2017, **8**.
180. M.-W. Taing, F. J. Rose and J. P. Whitehead, *Drug design, development and therapy*, 2014, **8**, 677-688.
181. E. Tomas and J. F. Habener, *Trends in endocrinology and metabolism: TEM*, 2010, **21**, 59-67.
182. J. J. Holst, *Physiological Reviews*, 2007, **87**, 1409-1439.
183. D. Donnelly, *British journal of pharmacology*, 2012, **166**, 27-41.
184. D. Athauda and T. Foltynie, *Drug Discovery Today*, 2016, **21**, 802-818.
185. B. Ahrén, *Experimental cell research*, 2011, **317**, 1239-1245.
186. G. Song, D. Yang, Y. Wang, C. de Graaf, Q. Zhou, S. Jiang, K. Liu, X. Cai, A. Dai, G. Lin, D. Liu, F. Wu, Y. Wu, S. Zhao, L. Ye, G. W. Han, J. Lau, B. Wu, M. A. Hanson, Z.-J. Liu, M.-W. Wang and R. C. Stevens, *Nature*, 2017, **546**, 312-315.
187. Y. Zhang, B. Sun, D. Feng, H. Hu, M. Chu, Q. Qu, J. T. Tarrasch, S. Li, T. Sun Kobilka, B. K. Kobilka and G. Skiniotis, *Nature*, 2017, **546**, 248-253.
188. C. d. Graaf, D. Donnelly, D. Wootten, J. Lau, P. M. Sexton, L. J. Miller, J.-M. Ahn, J. Liao, M. M. Fletcher, D. Yang, A. J. H. Brown, C. Zhou, J. Deng and M.-W. Wang, *Pharmacological reviews*, 2016, **68**, 954-1013.
189. D. Wootten, C. A. Reynolds, K. J. Smith, J. C. Mobarec, C. Koole, E. E. Savage, K. Pabreja, J. Simms, R. Sridhar, S. G. B. Furness, M. Liu, P. E. Thompson, L. J. Miller, A. Christopoulos and P. M. Sexton, *Cell*, 2016, **165**, 1632-1643.
190. A. B. Bueno, A. D. Showalter, D. B. Wainscott, C. Stutsman, A. Marín, J. Ficorilli, O. Cabrera, F. S. Willard and K. W. Sloop, *The Journal of biological chemistry*, 2016, **291**, 10700-10715.
191. Y.-L. Liang, M. Khoshouei, A. Glukhova, S. G. B. Furness, P. Zhao, L. Clydesdale, C. Koole, T. T. Truong, D. M. Thal, S. Lei, M. Radjainia, R. Danev, W. Baumeister, M.-W. Wang, L. J. Miller, A. Christopoulos, P. M. Sexton and D. Wootten, *Nature*, 2018, **555**, 121-125.
192. Y. Xu, Y. Wang, K. Liu, Y. Peng, D. Yao, H. Tao, H. Liu and G. Song, *IUCrJ*, 2019, **6**, 996-1006.
193. A. B. Bueno, B. Sun, F. S. Willard, D. Feng, J. D. Ho, D. B. Wainscott, A. D. Showalter, M. Vieth, Q. Chen, C. Stutsman, B. Chau, J. Ficorilli, F. J. Agejas, G. R. Cumming, A. Jiménez, I. Rojo, T. S. Kobilka, B. K. Kobilka and K. W. Sloop, *Nature Chemical Biology*, 2020, **16**, 1105-1110.
194. T. Kawai, B. Sun, H. Yoshino, D. Feng, Y. Suzuki, M. Fukazawa, S. Nagao, D. B. Wainscott, A. D. Showalter, B. A. Droz, T. S. Kobilka, M. P. Coghlan, F. S. Willard, Y. Kawabe, B. K. Kobilka and K. W. Sloop, *Proceedings of the National Academy of Sciences*, 2020, **117**, 29959.
195. A. Vangone, J. Schaarschmidt, P. Koukos, C. Geng, N. Citro, M. E. Trellet, L. C. Xue and A. M. J. J. Bonvin, *Bioinformatics*, 2019, **35**, 1585-1587.

196. Z. Kurkcuoglu, P. I. Koukos, N. Citro, M. E. Trellet, J. Rodrigues, I. S. Moreira, J. Roel-Touris, A. S. J. Melquiond, C. Geng and J. Schaarschmidt, *Journal of computer-aided molecular design*, 2018, **32**, 175-185.
197. T. Kar, U. Narsaria, S. Basak, D. Deb, F. Castiglione, D. M. Mueller and A. P. Srivastava, *Scientific reports*, 2020, **10**, 1-24.
198. S. A. Hinke, A. M. Cieniewicz, T. Kirchner, K. D'Aquino, R. Nanjunda, J. Aligo, R. Perkinson, P. Cooper, K. Boayke, M. L. Chiu, S. Jarantow, E. R. Lacy, Y. Liang, D. L. Johnson, J. M. Whaley, R. B. Lingham and A. J. Kihm, *Molecular metabolism*, 2018, **10**, 87-99.
199. S. Milo, R. A. Heylen, J. Glancy, G. T. Williams, B. L. Patenall, H. J. Hathaway, N. T. Thet, S. L. Allinson, M. Laabei and A. T. A. Jenkins, *Scientific Reports*, 2021, **11**, 3726.
200. Y. E. Ryzhkova, M. N. Elinson, A. N. Vereshchagin, K. A. Karpenko, F. V. Ryzhkov, I. E. Ushakov and M. P. Egorov, *Chemistry*, 2022, **4**.
201. D. Rodríguez, Z. G. Gao, S. M. Moss, K. A. Jacobson and J. Carlsson, *J Chem Inf Model*, 2015, **55**, 550-563.
202. W. K. B. Chan and Y. Zhang, *J Mol Biol*, 2020, **432**, 4872-4890.
203. H. Chen, W. Fu, Z. Wang, X. Wang, T. Lei, F. Zhu, D. Li, S. Chang, L. Xu and T. Hou, *ACS Chemical Neuroscience*, 2019, **10**, 677-689.
204. D. R. Weiss, J. Karpiak, X.-P. Huang, M. F. Sassano, J. Lyu, B. L. Roth and B. K. Shoichet, *Journal of Medicinal Chemistry*, 2018, **61**, 6830-6845.
205. J. S. E. Loo, A. L. Emtage, K. W. Ng, A. S. J. Yong and S. W. Doughty, *Journal of Molecular Graphics and Modelling*, 2018, **80**, 38-47.
206. Y.-D. Gong, H.-G. Cheon, T. Lee and N. S. Kang, 2010, **31**.
207. T. Akabli, F. Lamchouri, S. Senhaji and H. Toufik, *Structural Chemistry*, 2019, **30**, 1495-1504.
208. D. F. Veber, S. R. Johnson, H. Y. Cheng, B. R. Smith, K. W. Ward and K. D. Kopple, *J Med Chem*, 2002, **45**, 2615-2623.
209. H. Jhoti, G. Williams, D. C. Rees and C. W. Murray, *Nat Rev Drug Discov*, 2013, **12**, 644-645.
210. H. van de Waterbeemd, D. A. Smith and B. C. Jones, *Journal of Computer-Aided Molecular Design*, 2001, **15**, 273-286.
211. C. Wang, A. Avdeef, W. Zhang and K. Y. Tam, *Bio-Medical Materials and Engineering*, 2014, **24**, 3849-3854.
212. U. Rashid, S. F. Hassan, S. Nazir, A. Wadood, M. Waseem and F. L. Ansari, *Medicinal Chemistry Research*, 2015, **24**, 304-315.
213. S. F. Hassan, U. Rashid, F. L. Ansari and Z. Ul-Haq, *Journal of Molecular Graphics and Modelling*, 2013, **45**, 202-210.
214. K. Beaumont, S. M. Cole, K. Gibson and J. R. Gosset, in *Metabolism, Pharmacokinetics and Toxicity of Functional Groups: Impact of Chemical Building Blocks on ADMET*, The Royal Society of Chemistry, 2010, pp. 61-98.
215. J. Fernandes and C. R. Gattass, *Journal of Medicinal Chemistry*, 2009, **52**, 1214-1218.
216. V. R. Vukic, D. M. Loncar, D. V. Vukic, L. R. Jevric, G. Benedekovic, J. Francuz, V. Kojic, M. Z. Karadzic Banjac and V. Popsavin, *Computational Biology and Chemistry*, 2019, **83**, 107112.

217. M. A. Dorato and L. A. Buckley, *Curr Protoc Toxicol*, 2007, **Chapter 19**, Unit19 11.
218. S. Parasuraman, *J Pharmacol Pharmacother*, 2011, **2**, 74-79.
219. F. Pognan, M. Beilmann, H. C. M. Boonen, A. Czich, G. Dear, P. Hewitt, T. Mow, T. Oinonen, A. Roth, T. Steger-Hartmann, J.-P. Valentin, F. Van Goethem, R. J. Weaver and P. Newham, *Nature Reviews Drug Discovery*, 2023, **22**, 317-335.
220. L. Prasad-Reddy and D. Isaacs, *Drugs in context*, 2015, **4**, 212283-212283.
221. R. Gentilella, V. Pechtner, A. Corcos and A. Consoli, *Diabetes Metab Res Rev*, 2019, **35**, e3070.
222. C. S. Odoemelam, E. Hunter, J. Simms, Z. Ahmad, M.-W. Chang, B. Percival, I. H. Williams, M. Molinari, S. C. Kamerlin and P. B. Wilson, *Applied Biosciences*, 2022, **1**, 143-162.
223. D. R. Weiss, A. Bortolato, B. Tehan and J. S. Mason, *Journal of Chemical Information and Modeling*, 2016, **56**, 642-651.
224. L. B. Knudsen, D. Kiel, M. Teng, C. Behrens, D. Bhumralkar, J. T. Kodra, J. J. Holst, C. B. Jeppesen, M. D. Johnson, J. C. de Jong, A. S. Jorgensen, T. Kercher, J. Kostrowicki, P. Madsen, P. H. Olesen, J. S. Petersen, F. Poulsen, U. G. Sidelmann, J. Sturis, L. Truesdale, J. May and J. Lau, *Proceedings of the National Academy of Sciences of the United States of America*, 2007, **104**, 937-942.
225. F. Li, H. Wang, Y. Wang, S. Feng, B. Hu, X. Zhang, J. Wang, W. Li and M. Cheng, *Journal of Biomolecular Structure and Dynamics*, 2019, 1-18.
226. W. Dong, B. Han, K. Shao, Z. Yang, Y. Peng, Y. Yang and W. Liu, *J Mater Sci Mater Med*, 2012, **23**, 2945-2952.
227. M. Maximo da Silva, M. Comin, T. Santos Duarte, A. M. Foglio, E. J. De Carvalho, M. Do Carmo Vieira and S. A. Nazari Formaggio, *Molecules*, 2015, **20**.
228. M. L. Landry and J. J. Crawford, *ACS Med Chem Lett*, 2020, **11**, 72-76.
229. M. J. Waring, *Expert Opin Drug Discov*, 2010, **5**, 235-248.
230. C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, in *Adv Drug Deliv Rev*, Netherlands, 2001, vol. 46, pp. 3-26.
231. M. J. Waring, J. Arrowsmith, A. R. Leach, P. D. Leeson, S. Mandrell, R. M. Owen, G. Pairaudeau, W. D. Pennie, S. D. Pickett, J. Wang, O. Wallace and A. Weir, in *Nat Rev Drug Discov*, England, 2015, vol. 14, pp. 475-486.
232. Y. Gao, C. Gesenberg and W. Zheng, in *Developing Solid Oral Dosage Forms (Second Edition)*, eds. Y. Qiu, Y. Chen, G. G. Z. Zhang, L. Yu and R. V. Mantri, Academic Press, Boston, 2017, pp. 455-495.
233. M. Mohammadi, H. Mohammadiarani, V. S. Shaw, R. R. Neubig and H. Vashisth, *Proteins: Structure, Function, and Bioinformatics*, 2019, **87**, 146-156.
234. K. D. Schuster, M. Mohammadi, K. B. Cahill, S. L. Matte, A. D. Maillet, H. Vashisth and R. H. Cote, *PLOS ONE*, 2019, **14**, e0214554.
235. P. Renault, M. Louet, J. Marie, G. Labesse and N. Floquet, *Scientific Reports*, 2019, **9**, 5495.
236. G. Vettoretti, E. Moroni, S. Sattin, J. Tao, D. A. Agard, A. Bernardi and G. Colombo, *Scientific Reports*, 2016, **6**, 23830.
237. P. Urban, T. Lautier, D. Pompon and G. Truan, *Int J Mol Sci*, 2018, **19**.
238. P. Manikandan and S. Nagini, *Curr Drug Targets*, 2018, **19**, 38-54.

239. R. Santes-Palacios, A. L. Marroquín-Pérez, S. L. Hernández-Ojeda, R. Camacho-Carranza, T. Govezensky and J. J. Espinosa-Aguirre, *Toxicology in Vitro*, 2020, **62**, 104681.
240. Y. S. Sun, Z. Zhao, Z. N. Yang, F. Xu, H. J. Lu, Z. Y. Zhu, W. Shi, J. Jiang, P. P. Yao and H. P. Zhu, *Int J Biol Sci*, 2017, **13**, 1387-1397.
241. A. G. Waks and E. P. Winer, *JAMA*, 2019, **321**, 288-300.
242. R. J. Santen, H. Brodie, E. R. Simpson, P. K. Siiteri and A. Brodie, *Endocrine Reviews*, 2009, **30**, 343-375.
243. O. Artigalás, T. Vanni, M. H. Hutz, P. Ashton-Prolla and I. V. Schwartz, *BMC Medicine*, 2015, **13**, 139.
244. J. Sgrignani, A. Cavalli, G. Colombo and A. Magistrato, *Mini Reviews in Medicinal Chemistry*, 2016, **16**, 1112-1124.
245. A. A. Walsh, G. D. Szklarz and E. E. Scott, *The Journal of biological chemistry*, 2013, **288**, 12932-12943.
246. Z. Dutkiewicz and R. Mikstacka, *Bioinorganic chemistry and applications*, 2018, **2018**, 3924608-3924608.
247. J. McGraw, in *Handbook of Pharmacogenomics and Stratified Medicine*, ed. S. Padmanabhan, Academic Press, San Diego, 2014, pp. 323-340.
248. K. J. Aaftaab Sethi, K. Sasikala and Mallika Alvala, in *Drug Discovery and Development - New Advances*, ed. P. K. a. A. T. Vishwanath Gaitonde, IntechOpen, 2019.
249. F. N. Novikov, V. S. Stroylov, A. A. Zeifman, O. V. Stroganov, V. Kulkov and G. G. Chilov, *Journal of Computer-Aided Molecular Design*, 2012, **26**, 725-735.
250. O. V. Stroganov, F. N. Novikov, V. S. Stroylov, V. Kulkov and G. G. Chilov, *Journal of Chemical Information and Modeling*, 2008, **48**, 2371-2385.
251. S. Sehgal, A. Tahir, S. Shafique, M. Hassan and S. Rashid, *theoretical and computational science*, 2014, **1**.
252. M. Kubo, K. Yamamoto and T. Itoh, *Bioorg Med Chem*, 2019, **27**, 285-304.
253. R. Santes-Palacios, D. Ornelas-Ayala, N. Cabañas, A. Marroquín-Pérez, A. Hernández-Magaña, S. Del Rosario Olgún-Reyes, R. Camacho-Carranza and J. J. Espinosa-Aguirre, *Biomed Res Int*, 2016, **2016**, 5341081.
254. L. Scotti, F. Mendonça Junior, D. Moreira, M. Silva, I. Pitta and M. Scotti, *Current topics in medicinal chemistry*, 2013, **12**.
255. S. J. Teague, A. M. Davis, P. D. Leeson and T. Oprea, *Angewandte Chemie International Edition*, 1999, **38**, 3743-3748.
256. S. A. Kumar and B. L. Bhaskar, *Heliyon*, 2019, **5**, e02420.
257. S. Rampogu, M. Son, C. Park, H. H. Kim, J. K. Suh and K. W. Lee, *Biomed Res Int*, 2017, **2017**, 2105610.
258. N. Suvannang, C. Nantasenamat, C. Isarankura-Na-Ayudhya and V. Prachayasittikul, *Molecules*, 2011, **16**, 3597-3617.
259. C. Parks, Z. Gaieb and R. E. Amaro, *Frontiers in Molecular Biosciences*, 2020, **7**.
260. K. C. Hsu, Y. F. Chen and J. M. Yang, 2008.
261. M. Prabakaran, in *Characterisation and Design of Tissue Scaffolds*, ed. P. Tomlins, Woodhead Publishing, 2016, pp. 149-168.

262. G. Di Nardo, C. Zhang, A. G. Marcelli and G. Gilardi, *Journal*, 2021, **22**.
263. Y. Hong, H. Li, Y. C. Yuan and S. Chen, *Ann N Y Acad Sci*, 2009, **1155**, 112-120.
264. H. J. Chan, K. Petrossian and S. Chen, *J Steroid Biochem Mol Biol*, 2016, **161**, 73-83.
265. D. Ghosh, C. Egbuta and J. Lo, *The Journal of Steroid Biochemistry and Molecular Biology*, 2018, **181**, 11-19.
266. M. Fantacuzzi, B. De Filippis, M. Gallorini, A. Ammazalorso, L. Giampietro, C. Maccallini, Z. Aturki, E. Donati, R. S. Ibrahim, E. Shawky, A. Cataldi and R. Amoroso, *European Journal of Medicinal Chemistry*, 2020, **185**, 111815.
267. I. Tarvainen, T. Zimmermann, P. Heinonen, M. H. Jäntti, J. Yli-Kauhaluoma, V. Talman, H. Franzyk, R. K. Tuominen and S. B. Christensen, *ACS Medicinal Chemistry Letters*, 2020, **11**, 671-677.
268. J.-Y. Tsai, D. Rédei, P. Forgo, Y. Li, A. Vasas, J. Hohmann and C.-C. Wu, *Journal of Natural Products*, 2016, **79**, 2658-2666.
269. J. Neves Cruz, K. S. da Costa, T. A. A. de Carvalho and N. A. N. de Alencar, *Journal of Biomolecular Structure and Dynamics*, 2020, **38**, 1425-1434.
270. A. Fatima, J. K. Holien, C. Tiwari, M. W. Parker, R. J. Rodgers and L. L. Martin, *Biology of Reproduction*, 2020, **102**, 1261-1269.
271. A. Filler, *Nature Precedings*, 2009.
272. R. W. Darbeau, *Applied Spectroscopy Reviews*, 2006, **41**, 401-425.
273. W. R. Nitz, in *Medizintechnik: Verfahren – Systeme – Informationsverarbeitung*, ed. R. Kramme, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 339-356.
274. M. Balci, in *Basic ^{1H}- and ^{13C}-NMR Spectroscopy*, ed. M. Balci, Elsevier Science, Amsterdam, 2005, pp. 9-24.
275. K. Zia, T. Siddiqui, S. Ali, I. Farooq, M. S. Zafar and Z. Khurshid, *European journal of dentistry*, 2019, **13**, 124-128.
276. J. A. Tossell, in *Nuclear Magnetic Shieldings and Molecular Structure*, ed. J. A. Tossell, Springer Netherlands, Dordrecht, 1993, pp. 279-296.
277. L. B. Krivdin, *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2019, **112-113**, 103-156.
278. L. B. Casabianca, *Magnetic Resonance in Chemistry*, 2020, **58**, 611-624.
279. L. B. Casabianca and A. C. de Dios, *J Chem Phys*, 2008, **128**, 052201.
280. L. B. Krivdin, *Magnetic Resonance in Chemistry*, 2020, **58**, 478-499.
281. K. Jackowski and M. Jaszuński, *Concepts in Magnetic Resonance Part A*, 2007, **30A**, 246-260.
282. J. C. Facelli, *Concepts in Magnetic Resonance Part A*, 2004, **20A**, 42-69.
283. P. B. Wilson, M. Grootveld and S. C. L. Kamerlin, *Magnetic Resonance in Chemistry*, 2020, **58**, 116-117.
284. A. N. Bootsma and S. Wheeler, 2019.
285. S. E. Wheeler and K. N. Houk, *Journal of Chemical Theory and Computation*, 2010, **6**, 395-404.
286. N. Mardirossian and M. Head-Gordon, *Journal of Chemical Theory and Computation*, 2016, **12**, 4303-4325.
287. S. Lemonick, Density functional theory error discovered, (accessed 18/06/2022, 2022).

288. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. G. Fox, Inc., Wallingford CT,, *Journal*, 2016.
289. S. Nerli, A. C. McShan and N. G. Sgourakis, *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2018, **106-107**, 1-25.
290. A. Jazayeri, M. Rappas, A. J. H. Brown, J. Kean, J. C. Errey, N. J. Robertson, C. Fiez-Vandal, S. P. Andrews, M. Congreve, A. Bortolato, J. S. Mason, A. H. Baig, I. Teobald, A. S. Doré, M. Weir, R. M. Cooke and F. H. Marshall, *Nature*, 2017, **546**, 254-258.
291. A. Evers, M. Bossart, S. Pfeiffer-Marek, R. Elvert, H. Schreuder, M. Kurz, S. Stengel, M. Lorenz, A. Herling, A. Konkar, U. Lukasczyk, A. Pfenninger, K. Lorenz, T. Haack, D. Kadereit and M. Wagner, *Journal of Medicinal Chemistry*, 2018, **61**, 5580-5593.
292. P. Zhao, Y.-L. Liang, M. J. Belousoff, G. Deganutti, M. M. Fletcher, F. S. Willard, M. G. Bell, M. E. Christie, K. W. Sloop, A. Inoue, T. T. Truong, L. Clydesdale, S. G. B. Furness, A. Christopoulos, M.-W. Wang, L. J. Miller, C. A. Reynolds, R. Danev, P. M. Sexton and D. Wootten, *Nature*, 2020, **577**, 432-436.
293. X. Zhang, M. J. Belousoff, P. Zhao, A. J. Kooistra, T. T. Truong, S. Y. Ang, C. R. Underwood, T. Egebjerg, P. Šenel, G. D. Stewart, Y.-L. Liang, A. Glukhova, H. Venugopal, A. Christopoulos, S. G. B. Furness, L. J. Miller, S. Reedtz-Runge, C. J. Langmead, D. E. Gloriam, R. Danev, P. M. Sexton and D. Wootten, *Molecular Cell*, 2020, **80**, 485-500.e487.
294. H. Ma, W. Huang, X. Wang, L. Zhao, Y. Jiang, F. Liu, W. Guo, X. Sun, W. Zhong, D. Yuan and H. E. Xu, *Cell Research*, 2020, **30**, 1140-1142.
295. Z. Cong, L.-N. Chen, H. Ma, Q. Zhou, X. Zou, C. Ye, A. Dai, Q. Liu, W. Huang, X. Sun, X. Wang, P. Xu, L. Zhao, T. Xia, W. Zhong, D. Yang, H. Eric Xu, Y. Zhang and M.-W. Wang, *Nature Communications*, 2021, **12**, 3763.
296. X. Zhang, M. J. Belousoff, Y.-L. Liang, R. Danev, P. M. Sexton and D. Wootten, *Cell Reports*, 2021, **36**.
297. X. Zhang, R. M. Johnson, I. Drulyte, L. Yu, A. Kotecha, R. Danev, D. Wootten, P. M. Sexton and M. J. Belousoff, *Structure*, 2021, **29**, 963-974.e966.
298. R. M. Johnson, X. Zhang, S. J. Piper, T. J. Nettleton, T. H. Vandekolk, C. J. Langmead, R. Danev, P. M. Sexton and D. Wootten, *Biochemical and Biophysical Research Communications*, 2021, **578**, 84-90.

Appendix

Chapter 4 - Supplementary Information

Appendix 1

Table S1: Table showing the GPCR crystal structures resolved between 2017 and 2021

PDB ID	Date	Ligand Bound	Crystallisation Method	Resolution	Reference
5NX2	2017	Peptide 1	X-ray diffraction	3.7 Å	290
6KJV	2019	PF-0637222	X-ray diffraction	2.8 Å	192
6KK1	2019	PF-0637222	X-ray diffraction	2.8 Å	192
6GB1	2018	Peptide 11	X-ray diffraction	2.73 Å	291
6KK7	2019	PF-0637222	X-ray diffraction	3.1 Å	192
6LN2	2020	FAB7F38, PF-06372222	X-ray diffraction	3.2 Å	124
6ORV	2020	TT-OAD2	Electron microscopy	3.0 Å	292
6VCB	2020	LSN3160440	Electron microscopy	3.3 Å	193
6X18	2020	Glp-1	Electron microscopy	2.1 Å	293
6X19	2020	CHU-128	Electron microscopy	2.1 Å	293
6X1A	2020	PF-06882961	Electron microscopy	2.5 Å	293
6XOX	2020	LY3502970	Electron microscopy	3.1 Å	194
7C2E	2020	RGT1383	Electron microscopy	4.2 Å	294
7DUQ	2021	Compound 2	Electron microscopy	2.5 Å	295
7DUR	2021	Compound 2	Electron microscopy	3.3 Å	295

7E14	2021	Compound 2	Electron microscopy	2.9 Å	295
7EVM	2021	Compound 2	Electron microscopy	2.5 Å	295
7KI0	2021	Semaglutide	Electron microscopy	2.5 Å	296
7KI1	2021	Taspoglutide	Electron microscopy	2.5 Å	296
7LCI	2021	Pf-06882961	Electron microscopy	2.9 Å	297
7LCJ	2021	Pf-06882961	Electron microscopy	2.82 Å	297
7LCK	2021	Pf-06882961	Electron microscopy	3.24 Å	297
7RTB	2021	Peptide 19	Electron microscopy	2.14 Å	298

Appendix 2

Table S2. The binding energy of the ligands upon docking onto the crystal structures studied

Pose	Ligand	Structure	Binding Energy (Kcal/mol)
1	1	5VAI	-6.46
2	1	5VAI	-6.34
3	1	5VAI	-6.25
1	2	5VAI	-7.55
2	2	5VAI	-7.45
3	2	5VAI	-7.37
1	3	5VAI	-6.44
2	3	5VAI	-6.31
3	3	5VAI	-6.28
1	4	5VAI	-7.20
2	4	5VAI	-7.159
3	4	5VAI	-7.157
1	1	6B3J	-6.70
2	1	6B3J	-6.59
3	1	6B3J	-6.43
1	2	6B3J	-7.67
2	2	6B3J	-7.59

3	2	6B3J	-7.45
1	3	6B3J	-7.27
2	3	6B3J	-7.12
3	3	6B3J	-6.92
1	4	6B3J	-7.95
2	4	6B3J	-7.77
3	4	6B3J	-7.73
1	1	5VEW	-8.50
2	1	5VEW	-8.18
3	1	5VEW	-6.26
1	2	5VEW	-7.189
2	2	5VEW	-7.066
3	2	5VEW	-6.873
1	3	5VEW	-7.877
2	3	5VEW	-6.866
3	3	5VEW	-6.827
1	4	5VEW	-9.48
2	4	5VEW	-9.27
3	4	5VEW	-8.22

Appendix 3

Table S3. Table showing amino acids present in the binding sites for compound 1 Pose 2

S/N	5VAI	5VEW	6B3J
1	VAL 331	TYR 402	LEU 354
2	THR 355	SER 352	PHE 390
3	LEU 401	ASN 406	HIS 363
4	TYR 402	LEU 401	PHE 393
5	LEU 356	VAL 405	PRO 358
6	SER 352	LYS 351	MET 397

7	THR 353	ASN 407	
8	LEU 349	ARG 348	
9		LEU 349	
10		ILE 345	
11		GLU 408	
12		HIS 180	
13		ARG 176	

Appendix 4

Table S4. Table showing amino acids present in the binding sites for compound 1 Pose 3

S/N	5VAI	5VEW	6B3J
1	LEU 401	SER 389	LEU 359
2	PHE 393	PHE 390	LEU 401
3	GLN 394	GLN 394	ASN 406
4	PHE 390	MET 397	TYR 402
5	MET 397	LEU 396	LEU 356
6	PRO 358	LEU 359	
7	ILE 357	PHE 393	
8	THR 355		
9	LEU 354		
10	LYS 351		

Appendix 5

Table S5. Table showing amino acids present in the binding sites for compound 2 Pose 2

S/N	5VAI	5VEW	6B3J
1	PHE 324	THR 355	LEU 354
2	PHE 321	PRO 358	ILE 357
3	VAL 365	PHE 324	ILE 366

4	ILE 357	LEU 354	PRO 358
5	PHE 390	ILE 328	PHE 390
6	HIS 363	VAL 331	
7	GLU 387		
8	ILE 366		
9	ILE 325		

Appendix 6

Table S6. Table showing amino acids present in the binding sites for compound 2 Pose 3

S/N	5VAI	5VEW	6B3J
1	LEU 401	LEU 359	MET 397
2	LEU 354	THR 355	HIS 363
3	HIS 363	LEU 354	PHE 390
4	PHE 390	VAL 331	PRO 358
5		PHE 347	VAL 405
6		PHE 324	LEU 401
7		PRO 358	ILE 357
8			LEU 354

Appendix 7

Table S7. Table showing amino acids present in the binding sites for compound 3 Pose 2

S/N	5VAI	5VEW	6B3J
1	SER 352	ARG 348	LEU 356
2	VAL 331	LEU 349	LEU 359
3	LEU 356	LYS 351	GLU 408
4		LEU 401	VAL 405
5		VAL 405	LEU 401
6			TYR 402

Appendix 8

Table S8. Table showing amino acids present in the binding sites for compound 3 Pose 3

S/N	5VAI	5VEW	6B3J
1	ARG 176	ILE 345	MET 397
2	HIS 180	VAL 405	PRO 358
3	TYR 402	LYS 351	PHE 390
4	SER 352	LEU 401	PHE 393
5	THR 353	SER 352	PHE 385
6		ARG 348	
7		ASN 406	

Appendix 9

Table S9. Table showing amino acids present in the binding sites for compound 4 Pose 2

S/N	5VAI	5VEW	6B3J
1	PHE 324	HIS 363	LYS 351
2	PHE 321	LEU 359	ILE 357
3	VAL 365	THR 391	VAL 405
4	ILE 357	GLN 394	THR 355
5	PHE 390	LEU 360	MET 397
6	ILE 325	MET 397	LEU 401
7		LEU 396	PRO 358
8		PHE 393	HIS 363
9		ILE 400	PHE 390

Appendix 10

Table S10. Table showing amino acids present in the binding sites for compound 4 Pose 3

S/N	5VAI	5VEW	6B3J
-----	------	------	------

1	LEU 251	LYS 351	ILE 357
2	LEU 356	LEU 401	LYS 351
3	THR 355	VAL 405	THR 355
4	LYS 351	ARG 348	LEU 354
5		LEU 349	LEU 401
6			MET 397
7			PRO 358
8			HIS 363
			PHE 390

Chapter 9 - Supplementary Information

Table S11 Relative DFT computed NMR shielding of optimised 4,4'-dimethyl-1,1'-biphenyl pruned using (50,194) grid with rotations on the x axis

Ato		10	20	30	40	50	60	70	80
m	0	degrees							
16-		0.9999	0.9999	1.0011	0.9999	1.0011	0.9999	0.9999	0.9999
H	1	708	7497	6369	5412	6786	5412	708	7497
20-		0.9999	0.9999	0.9999	1.0011	0.9999	1.0011	0.9999	0.9999
H	1	708	708	5412	6786	5412	6369	7497	7497
13-		0.9999	0.9999	0.9986	0.9998	0.9986	0.9998	0.9999	0.9999
H	1	7085	6669	4662	1261	4246	1261	7085	7085
17-		0.9999	0.9999	0.9998	0.9986	0.9998	0.9986	0.9999	0.9999
H	1	7085	7085	1261	4246	1261	4246	6669	7085
15-		0.9999	0.9999	0.9976	1.0000	1.0024	1.0000	0.9999	0.9999
H	1	9174	9174	6291	3718	2916	3718	9174	9587
19-		0.9999	0.9999	1.0000	1.0024	1.0000	1.0024	0.9999	0.9999
H	1	9174	9174	3718	2503	3718	2503	9174	9174
14-		1.0000	1.0000	0.9976	0.9999	0.9976	0.9999	1.0000	1.0000
H	1	0412	0412	6291	5878	6291	5878	0824	0412

18-H	1	1.0000 0412	1.0000 0824	0.9999 5878	0.9976 6291	0.9999 5878	0.9976 6291	1.0000 0412	1.0000 0412
22-H	1	1	1	1.0006 2999	0.9999 8967	1.0006 2999	0.9999 9311	1	1
28-H	1	0.9999 7262	1	0.9999 8967	1.0006 2654	0.9999 8967	1.0006 2999	1	1
23-H	1	0.9999 7947	0.9999 7947	1.0093 4883	1.0000 3422	1.0093 6252	1.0000 3764	0.9999 692	0.9999 8289
26-H	1	1.0000 2037	0.9999 692	1.0000 3764	1.0093 5568	1.0000 3422	1.0093 5225	0.9999 7947	0.9999 7947
24-H	1	1.0000 2037	1.0000 2037	0.9901 2315	0.9999 7963	0.9901 0957	0.9999 7623	1.0000 3056	1.0000 2037
27-H	1	1.0000 2716	1.0000 3056	0.9999 7623	0.9901 1636	0.9999 7963	0.9901 2315	1.0000 2037	1.0000 2037
Ato m	0	10 degrees	20 degrees	30 degrees	40 degrees	50 degrees	60 degrees	70 degrees	80 degrees
1-C	1	1.0002 0322	1.0002 0559	1.0006 4512	1.0007 231	1.0006 5457	1.0007 231	1.0002 0086	1.0001 9613
7-C	1	1.0002 0086	1.0002 0086	1.0007 231	1.0006 4275	1.0007 3019	1.0006 4748	1.0002 0559	1.0001 9141
4-C	1	0.9999 8354	0.9999 5298	1.0004 3959	0.9998 6366	1.0004 584	0.9998 7541	0.9999 5769	0.9999 5534
10-C	1	0.9999 5769	0.9999 5769	0.9998 7306	1.0004 4665	0.9998 8246	1.0004 3724	0.9999 5298	0.9999 6709
3-C	1	0.9999 0049	0.9999 2302	0.9997 7283	1.0000 6383	0.9997 6907	1.0000 4881	0.9999 1927	0.9999 2866
9-C	1	0.9999 2866	0.9999 1927	1.0000 4694	0.9997 916	1.0000 5445	0.9997 7283	0.9999 2302	0.9999 2115
5-C	1	0.9999 5498	0.9999 5873	0.9997 9366	0.9997 5614	0.9997 9178	0.9997 5239	0.9999 5873	0.9999 5873

11- C	1	0.9999 6061	0.9999 5686	0.9997 5051	0.9997 8428	0.9997 5426	0.9997 9366	0.9999 5873	0.9999 5498
6- C	1	1.0000 9431	1.0000 8897	1.0024 3431	0.9998 4519	1.0024 432	0.9998 4519	1.0000 9431	1.0000 8719
12- C	1	1.0000 9431	1.0000 9431	0.9998 4519	1.0024 3253	0.9998 523	1.0024 3609	1.0000 8897	1.0000 8541
2- C	1	1.0000 6755	1.0000 64	0.9984 3204	1.0007 8398	0.9984 3915	1.0007 7865	1.0000 6044	1.0000 5867
8- C	1	1.0000 5689	1.0000 5867	1.0007 7687	0.9984 4271	1.0007 8398	0.9984 3204	1.0000 6578	1.0000 5867
21- C	1	1.0000 0879	1.0000 0314	1.0000 5086	1.0000 0377	1.0000 5149	1.0000 0565	1.0000 0377	1.0000 0314
25- C	1	1.0000 0377	1.0000 0377	1.0000 0565	1.0000 5024	1.0000 044	1.0000 5024	1.0000 0314	1.0000 0377

Table S12 Relative DFT computed NMR shielding of optimised 4,4'-dimethyl-1,1'-biphenyl pruned using (75,302) grid with rotations on the x axis

Ato m	0	10 degrees	20 degrees	30 degrees	40 degrees	50 degrees	60 degrees	70 degrees	80 degrees
16- H	1	1	1.0000 1253	0.9999 9582	1	1	1	1.0000 2088	1
20- H	1	1	1.0000 2088	1	1	1	0.9999 9582	1.0000 1253	1
13- H	1	0.9999 9583	1.0000 668	0.9999 9166	0.9999 9583	0.9999 9583	0.9999 9166	1.0000 0834	0.9999 9583
17- H	1	0.9999 9583	1.0000 0834	0.9999 9166	1	1	0.9999 9166	1.0000 1668	0.9999 9583
15- H	1	1.0000 0413	0.9999 9587	1.0000 0413	1	1	1.0000 0826	1.0000 0826	1.0000 0413

19-		1.0000	1.0000	1.0000			1.0000	0.9999	1.0000
H	1	0413	0826	0826	1	1	0413	9587	0413
14-			1.0000	0.9999			0.9999	0.9999	0.9999
H	1	1	0412	9588	1	1	9176	9588	9588
18-		0.9999	0.9999	0.9999			0.9999	1.0000	0.9999
H	1	9588	9176	9176	1	1	9588	0412	9588
22-		1.0000		1.0000			1.0000	1.0000	1.0000
H	1	0688	1	0344	1	1	0344	0344	0344
28-		1.0000	1.0000	1.0000	1.0000		1.0000		1.0000
H	1	0344	0344	0344	0344	1	0344	1	0688
23-		1.0000	0.9999	1.0000			1.0000	1.0000	1.0000
H	1	1026	7948	1368	1	1	2394	2394	1026
26-		1.0000	1.0000	1.0000			1.0000	0.9999	1.0000
H	1	1026	2394	2394	1	1	1368	7948	1026
24-		0.9999	1.0000	0.9999	1.0000	1.0000	0.9999	0.9999	0.9999
H	1	966	2378	8981	0679	034	7962	7622	9321
27-		0.9999	0.9999	0.9999	1.0000	1.0000	0.9999	1.0000	0.9999
H	1	9321	7622	7622	034	0679	8981	2378	966
Ato		10	20	30	40	50	60	70	80
m	0	degrees							
4-		1.0000	1.0000	0.9999	1.0000	1.0000		1.0000	1.0000
C	1	1882	3058	9765	1411	0941	1	3293	1647
10-		1.0000	1.0000		1.0000	1.0000		1.0000	1.0000
C	1	1647	3293	1	1176	1411	1	2823	2117
1-		1.0000	0.9999	1.0000	1.0000	1.0000	1.0000	0.9999	1.0000
C	1	2112	249	2816	1643	1877	2816	2021	2347
7-		1.0000	0.9999	1.0000	1.0000	1.0000	1.0000	0.9999	1.0000
C	1	1877	2021	2816	1643	1643	2816	249	2112
3-		0.9999	1.0000	0.9999	0.9999	0.9999	0.9999	1.0000	0.9999
C	1	9437	2441	8686	9061	9061	8686	2066	9624

9-		0.9999	1.0000	0.9999	0.9999	0.9999	0.9999	1.0000	0.9999
C	1	9624	2066	8686	9061	9249	8686	2253	9249
5-		1.0000	1.0000	0.9999	1.0000	1.0000		1.0000	1.0000
C	1	0375	1689	9625	0563	0375	1	2815	0188
11-		1.0000	1.0000		1.0000	1.0000	0.9999	1.0000	1.0000
C	1	0375	2815	1	0375	0563	9437	1501	0375
6-		1.0000	0.9999	1.0000	0.9468	1.0000	0.9999	0.9999	1.0000
C	1	2838	7694	2661	023	2306	8226	8226	2483
12-		1.0000	0.9999	1.0000	1.0000	1.0000	0.9999	0.9999	1.0000
C	1	2483	8226	2661	2306	1774	8226	7871	2661
2-		0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
C	1	9646	6102	9469	9823	9291	9291	5394	9469
8-		0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
C	1	9469	5394	9114	9291	9823	9469	6102	9469
21-		1.0000	0.9999	1.0000	1.0000	1.0000	1.0000	0.9999	1.0000
C	1	0565	9937	0188	0188	0126	0126	9937	0251
25-		1.0000	0.9999	0.9999	1.0000	1.0000	1.0000	0.9999	1.0000
C	1	0188	9937	9498	0126	0126	0188	9937	0565

Table S13 Relative DFT computed NMR shielding of optimised 4,4'-dimethyl-1,1'-biphenyl pruned using (99,590) grid with rotations on the x axis

Ato		10	20	30	40	50	60	70	80
m	0	degrees							
16-		1.0000		1.0000		1.0000	1.0000		
H	1	042	1	042	1	042	042	1	1
20-				1.0000	1.0000		1.0000		1.0000
H	1	1	1	042	042	1	042	1	042
13-		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
H	1	042	042	042	042	042	042	042	042

17-H	1	1.0000 042							
15-H	1	1.0000 083	1.0000 041	1.0000 041	1.0000 041	1.0000 041	1.0000 041	1	1
19-H	1	1	1	1.0000 041	1.0000 041	1.0000 041	1.0000 041	1.0000 041	1.0000 083
14-H	1	0.9999 917	0.9999 959						
18-H	1	0.9999 959	0.9999 917						
22-H	1	1.0000 034							
28-H	1	1	1.0000 034						
23-H	1	1.0000 239	1.0000 102	1.0000 17	1.0000 102	1.0000 17	1.0000 17	1.0000 068	1.0000 102
26-H	1	1.0000 102	1.0000 068	1.0000 17	1.0000 17	1.0000 102	1.0000 17	1.0000 102	1.0000 239
24-H	1	0.9999 727	0.9999 898	0.9999 83	0.9999 898	0.9999 83	0.9999 796	0.9999 898	0.9999 898
27-H	1	0.9999 898	0.9999 898	0.9999 796	0.9999 796	0.9999 864	0.9999 83	0.9999 898	0.9999 727
Ato m	0	10 degrees	20 degrees	30 degrees	40 degrees	50 degrees	60 degrees	70 degrees	80 degrees
4-C	1	1.0000 024	1	1.0000 047	1.0000 071	1.0000 024	1.0000 047	1.0000 047	1.0000 024
10-C	1	1.0000 024	1.0000 047	1.0000 047	1.0000 024	1.0000 071	1.0000 047	1	1.0000 024
1-C	1	0.9999 789	0.9999 789	0.9999 789	0.9999 789	0.9999 766	0.9999 766	0.9999 789	0.9999 813

7-		0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
C	1	813	789	766	766	789	766	789	789
3-		0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
C	1	981	019	038	019	056	056	019	1
9-			1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
C	1	1	019	056	056	019	038	019	981
5-		0.9999	0.9999		0.9999	1.0000	1.0000	0.9999	0.9999
C	1	944	981	1	925	019	038	925	887
11-		0.9999	0.9999	1.0000	1.0000	0.9999		0.9999	0.9999
C	1	887	925	019	019	925	1	981	944
6-		0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
C	1	894	858	841	858	858	841	841	858
12-		0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
C	1	858	982	823	858	858	841	858	894
2-		0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
C	1	858	947	947	929	929	929	947	929
8-		0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
C	1	929	947	965	929	929	947	947	911
21-		1.0000			1.0000	0.9999	0.9999	1.0000	1.0000
C	1	006	1	1	013	994	994	013	019
25-		1.0000	1.0000	0.9999	0.9999	1.0000	0.9999		1.0000
C	1	013	013	994	994	006	994	1	006

Table S14 Relative DFT computed NMR shielding of optimised 4,4'-dimethyl-1,1'-biphenyl pruned using (175,974) grid with rotations on the x axis

Ato		10	20	30	40	50	60	70	80
m	0	degrees							
16-		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
H	1	042	042	042	042	042	042	042	042

20- H	1	1.0000 042							
13- H	1	1	1	1	1	1	1	1	1
17- H	1	1	1	1	1	1	1	1	1
15- H	1	1.0000 041	1	1.0000 041	1	1	1	1	1
19- H	1	1	1	1	1	1	1	1	1.0000 041
14- H	1	0.9999 959							
18- H	1	0.9999 959							
22- H	1	1	1	1	1	1	1	1	1
28- H	1	1	1	1	1	1	1	1	1
23- H	1	1.0000 17	1.0000 136						
26- H	1	1.0000 136	1.0000 17						
24- H	1	0.9999 796	0.9999 83						
27- H	1	0.9999 83	0.9999 83	0.9999 83	0.9999 864	0.9999 83	0.9999 83	0.9999 83	0.9999 796
Ato m	0	10 degrees	20 degrees	30 degrees	40 degrees	50 degrees	60 degrees	70 degrees	80 degrees
4- C	1	1.0000 024	1.0000 024	1	1.0000 024	1.0000 024	1.0000 024	1.0000 024	1

10- C	1	1	1.0000 024	1.0000 047	1.0000 047	0.9999 976	1.0000 024	1.0000 024	1.0000 024
1- C	1	0.9999 883	0.9999 883	0.9999 859	0.9999 859	0.9999 859	0.9999 859	0.9999 859	0.9999 859
7- C	1	0.9999 859	0.9999 859	0.9999 883	0.9999 977	0.9999 859	0.9999 883	0.9999 883	0.9999 883
3- C	1	1.0000 038	1	1.0000 019	1	1.0000 038	1.0000 019	1.0000 019	1.0000 019
9- C	1	1.0000 019	1.0000 019	1.0000 038	1	1.0000 019	1	1	1.0000 019
5- C	1	1	1.0000 019	1.0000 019	1.0000 038	1	1	1.0000 019	0.9999 981
11- C	1	0.9999 981	1.0000 019	1	0.9999 981	1.0000 056	1.0000 038	1.0000 019	1
6- C	1	0.9999 965	0.9999 982	0.9999 965	1.0000 018	0.9999 965	0.9999 947	0.9999 965	0.9999 965
12- C	1	0.9999 965	0.9999 965	0.9999 947	1.0000 018	0.9999 982	0.9999 965	0.9999 982	0.9999 965
2- C	1	0.9999 894	0.9999 876	0.9999 894	1.0000 018	0.9999 894	0.9999 911	0.9999 894	0.9999 911
8- C	1	0.9999 894	0.9999 894	0.9999 894	1	0.9999 894	0.9999 894	0.9999 876	0.9999 894
21- C	1	1.0000 006	1.0000 006	1.0000 006	1	1	1.0000 006	1.0000 006	1.0000 006
25- C	1	1.0000 006	1.0000 006	1.0000 006	1	1	1	1.0000 006	1.0000 006

Table S15 Relative DFT computed NMR shielding of transition state 4,4'-dimethyl-1,1'-biphenyl pruned using (50,194) grid with rotations on the x axis

Ato m	0	10 degrees	20 degrees	30 degrees	40 degrees	50 degrees	60 degrees	70 degrees	80 degrees
16- H	1	0.9996 305	0.9994 819	0.9995 286	0.9991 676	1.0007 899		1.0004 035	1.0001 317
20- H	1	1.0001 317	1.0004 035	1.0004 459	1.0007 899	0.9991 676		0.9994 819	0.9996 305
13- H	1	1.0001 317	1.0004 035	1.0004 459	1.0007 899	0.9991 676		0.9994 819	0.9996 305
17- H	1	0.9996 305	0.9994 819	0.9995 286	0.9991 676	1.0007 899		1.0004 035	1.0001 317
15- H	1	0.9993 735	0.9994 27	0.9990 313	0.9985 202	1.0018 302	0.9999 918	1.0007 172	1.0008 904
19- H	1	1.0008 904	1.0007 172	1.0012 531	1.0018 302	0.9985 202	0.9999 918	0.9994 27	0.9993 735
14- H	1	1.0008 904	1.0007 172	1.0012 531	1.0018 302	0.9985 202	0.9999 918	0.9994 27	0.9993 735
18- H	1	0.9993 735	0.9994 27	0.9990 313	0.9985 202	1.0018 302	0.9999 918	1.0007 172	1.0008 904
22- H	1	0.9997 629	0.9998 488	1.0128 841	0.9998 866	0.9998 866	1.0088 608	0.9998 488	0.9997 629
28- H	1	0.9997 629	0.9998 488	1.0047 104	0.9998 866	0.9998 866	1.0088 608	0.9998 488	0.9997 629
23- H	1	0.9967 749	0.9970 712	0.9958 861	0.9928 823	1.0068 043	0.9912 102	1.0026 802	1.0030 412
26- H	1	1.0030 412	1.0026 802	1.0039 879	1.0068 043	0.9928 823		0.9970 712	0.9967 749
24- H	1	1.0030 412	1.0026 802	0.9908 39	1.0068 043	0.9928 823		0.9970 712	0.9967 749
27- H	1	0.9967 749	0.9970 712	0.9908 39	0.9928 823	1.0068 043	0.9912 102	1.0026 802	1.0030 412

Ato m	0	10 degrees	20 degrees	30 degrees	40 degrees	50 degrees	60 degrees	70 degrees	80 degrees
4- C	1	0.9996 216	1.0006 52	1.0005 354	1.0000 262	1.0000 262	1.0000 381	1.0006 52	0.9996 216
10- C	1	0.9996 216	1.0006 52	1.0005 354	1.0000 262	1.0000 262	1.0000 381	1.0006 52	0.9996 216
1- C	1	0.9999 937	1.0000 897	1.0013 354	0.9999 124	0.9999 124	1.0000 23	1.0000 897	0.9999 937
7- C	1	0.9999 937	1.0000 897	1.0013 354	0.9999 124	0.9999 124	1.0000 23	1.0000 897	0.9999 937
3- C	1	1.0007 062	0.9996 893	0.9987 853	1.0006 045	0.9984 181	0.9999 812	1.0003 352	1.0000 264
9- C	1	1.0000 264	1.0003 371	0.9999 906	0.9984 181	1.0006 045	0.9999 793	0.9996 893	1.0007 062
5- C	1	1.0000 264	1.0003 371	0.9999 906	0.9984 181	1.0006 045	0.9999 793	0.9996 893	1.0007 062
11- C	1	1.0007 062	0.9996 893	0.9987 853	1.0006 045	0.9984 181	0.9999 831	1.0003 352	1.0000 264
6- C	1	1.0008 575	0.9997 985	0.9993 407	1.0006 527	1.0012 388	1.0000 05	1.0009 158	1.0003 73
12- C	1	1.0003 73	1.0009 158	1.0009 89	1.0012 388	1.0006 527	1.0000 067	0.9997 985	1.0008 575
2- C	1	1.0003 73	1.0009 158	1.0009 89	1.0012 388	1.0006 527	1.0000 067	0.9997 985	1.0008 575
8- C	1	1.0008 575	0.9997 985	0.9993 407	1.0006 527	1.0012 388	1.0000 05	1.0009 158	1.0003 73
21- C	1	1.0001 958	1.0003 383	1.0002 008	1.0002 736	1.0002 736	1.0000 27	1.0003 383	1.0001 958
25- C	1	1.0001 958	1.0003 383	1.0002 008	1.0002 736	1.0002 736	1.0000 264	1.0003 383	1.0001 958

Table S16 Relative DFT computed NMR shielding of transition state 4,4'-dimethyl-1,1'-biphenyl pruned using (75,302) grid with rotations on the x axis

Ato		10	20	30	40	50	60	70	80
m	0	degrees							
16-		0.9999	0.9999	0.9999	0.9999	0.9997	1.0000	0.9994	0.9999
H	1	405	066	915	958	07	042	352	873
20-		0.9998	0.9998	1.0000	0.9999	1.0000	0.9999	1.0003	1.0000
H	1	769	132	042	958	425	915	822	085
13-		0.9998	0.9998	1.0000	0.9999	1.0000	0.9999	1.0003	1.0000
H	1	769	132	042	958	425	915	822	042
17-		0.9999	0.9999	0.9999	0.9999	0.9997	1.0000	0.9994	0.9999
H	1	405	066	915	958	07	042	352	873
15-		1.0000	1.0000	0.9999	0.9999	0.9996	1.0000	0.9991	0.9999
H	1	618	453	876	959	95	124	098	753
19-		0.9999	0.9999	1.0000		1.0002	0.9999	1.0008	1.0000
H	1	382	712	124	1	967	876	985	082
14-		0.9999	0.9999	1.0000		1.0002	0.9999	1.0008	1.0000
H	1	382	712	124	1	967	876	985	041
18-		1.0000	1.0000	0.9999	0.9999	0.9996	1.0000	0.9991	0.9999
H	1	618	453	876	959	95	124	098	67
22-		1.0087	1.0092		1.0000	1.0077		1.0053	0.9999
H	1	898	365	1	034	486	1	742	588
28-		1.0087	1.0000			1.0077		1.0053	1.0089
H	1	898	103	1	1	486	1	742	066
23-		1.0002	0.9911	0.9999	0.9999	1.0012	1.0000	1.0035	1.0000
H	1	146	147	455	898	09	477	725	749
26-		0.9910	1.0002	1.0000	1.0000	0.9911	0.9999	0.9910	0.9910
H	1	909	588	477	034	351	455	159	636
24-		0.9910	0.9997	1.0000	1.0000	0.9911	0.9999	0.9910	0.9999
H	1	909	514	477	068	351	455	159	523

27- H	1	1.0002 146	1.0002 588	0.9999 455	1.0000 034	1.0012 09	1.0000 477	1.0035 725	1.0000 954
Ato m	0	10 degrees	20 degrees	30 degrees	40 degrees	50 degrees	60 degrees	70 degrees	80 degrees
4- C	1	1.0001 333	1.0002 071	0.9999 548	0.9999 691	1.0000 5	0.9999 548	0.9998 762	0.9999 691
10- C	1	1.0001 333	1.0002 071	0.9999 548	0.9999 691	1.0000 5	0.9999 548	0.9998 762	0.9999 714
1- C	1	0.9997 016	1.0001 127	0.9999 666	0.9999 75	0.9996 035	0.9999 666	0.9995 806	1.0000 605
7- C	1	0.9997 016	1.0001 127	0.9999 666	0.9999 75	0.9996 035	0.9999 666	0.9995 806	1.0000 605
3- C	1	1.0000 264	1.0001 186	0.9999 831	0.9999 831	0.9999 925	0.9999 868	0.9994 803	0.9999 548
9- C	1	0.9999 812	1.0000 546	0.9999 868	0.9999 944	1.0004 444	0.9999 831	0.9998 87	0.9999 605
5- C	1	0.9999 812	1.0000 546	0.9999 868	0.9999 944	1.0004 444	0.9999 831	0.9998 87	0.9999 642
11- C	1	1.0000 264	1.0001 186	0.9999 831	0.9999 831	0.9999 925	0.9999 868	0.9994 803	0.9999 567
6- C	1	0.9999 051	0.9998 951	0.9999 75	0.9999 85	0.9996 971	0.9999 834	0.9993 725	0.9999 767
12- C	1	0.9997 553	0.9996 588	0.9999 834	0.9999 85	0.9999 334	0.9999 75	1.0004 145	0.9999 95
2- C	1	0.9997 553	0.9996 588	0.9999 834	0.9999 85	0.9999 334	0.9999 75	1.0004 145	0.9999 917
8- C	1	0.9999 051	0.9998 951	0.9999 75	0.9999 834	0.9996 971	0.9999 834	0.9993 725	0.9999 75
21- C	1	1.0001 242	1.0001 017	0.9999 969	0.9999 994	1.0000 446	0.9999 969	1.0001 136	1.0000 195

25-		1.0001	1.0001	0.9999	0.9999	1.0000	0.9999	1.0001	1.0000
C	1	242	017	969	994	446	969	136	195

Table S17 Relative DFT computed NMR shielding of transition state 4,4'-dimethyl-1,1'-biphenyl pruned using (99,590) grid with rotations on the x axis

Ato	m	0	10 degrees	20 degrees	30 degrees	40 degrees	50 degrees	60 degrees	70 degrees	80 degrees
16-			0.9999	0.9997		0.9998	0.9998	0.9999	0.9999	0.9997
H	1		703	877	1	853	132	915	958	41
20-			0.9999	0.9998	0.9999	0.9998	0.9998	1.0000	1.0000	0.9999
H	1		703	811	958	132	853	085	042	49
13-			0.9999	0.9998	0.9999	0.9998	0.9998	1.0000		0.9999
H	1		618	811	958	132	853	085	1	49
17-			0.9999	0.9997		0.9998	0.9998	0.9999	0.9999	0.9997
H	1		618	877	1	853	132	915	958	41
15-			0.9999	0.9999	1.0000	1.0000	0.9999	0.9999	0.9999	0.9997
H	1		876	011	041	33	423	876	959	774
19-			0.9999	1.0000	0.9999	0.9999	1.0000	1.0000	1.0000	1.0001
H	1		876	824	959	423	33	124	082	855
14-			0.9999	1.0000	0.9999	0.9999	1.0000	1.0000	1.0000	1.0001
H	1		67	824	959	423	33	124	082	855
18-			0.9999	0.9999	1.0000	1.0000	0.9999	0.9999	0.9999	0.9997
H	1		67	011	041	33	423	876	959	774
22-			0.9999	1.0086	1.0000	1.0000	1.0000	1.0000	1.0089	1.0097
H	1		45	11	034	206	206	034	512	827
28-			1.0090	1.0086	1.0000	1.0000	1.0000	1.0089	1.0000	1.0000
H	1		199	11	034	206	206	203	034	069
23-			1.0000	1.0003	1.0000	1.0001	0.9998	0.9999	1.0000	0.9911
H	1		477	372	136	873	059	455	17	114

26-		0.9910	0.9911	0.9999	0.9998	1.0001	0.9911	1.0000	1.0008
H	1	501	318	83	059	873	08	204	003
24-		0.9999	0.9911	0.9999	0.9998	1.0001	1.0000	0.9911	0.9991
H	1	591	318	796	059	873	511	08	827
27-		0.9999	1.0003	1.0000	1.0001	0.9998	1.0000	0.9999	0.9991
H	1	591	372	136	873	059	511	762	827
Ato		10	20	30	40	50	60	70	80
m	0	degrees							
4-		1.0000	0.9999	0.9999	1.0000	1.0000	0.9999	0.9999	0.9999
C	1	095	714	905	452	452	929	905	334
10-		1.0000	0.9999	0.9999	1.0000	1.0000	0.9999	0.9999	0.9999
C	1	071	714	905	452	452	929	929	334
1-		1.0001	0.9997	0.9999	0.9997	0.9997	0.9999	0.9999	0.9997
C	1	857	83	896	329	35	896	896	913
7-		1.0001	0.9997	0.9999	0.9997	0.9997	0.9999	0.9999	0.9997
C	1	878	83	896	329	35	896	896	913
3-		0.9998	1.0000	1.0000	1.0001	0.9999		1.0000	0.9999
C	1	889	734	038	694	887	1	038	322
9-		0.9998	1.0002	1.0000	0.9999	1.0001	1.0000	1.0000	1.0000
C	1	87	014	019	887	694	056	019	64
5-		0.9998	1.0002	1.0000	0.9999	1.0001	1.0000	1.0000	1.0000
C	1	757	014	019	887	694	056	019	64
11-		0.9998	1.0000	1.0000	1.0001	0.9999		1.0000	0.9999
C	1	739	734	038	694	887	1	038	322
6-		1.0000	0.9998	1.0000	0.9999	0.9998	0.9999	0.9999	0.9998
C	1	899	602	017	617	052	933	967	335
12-		1.0000	0.9999	0.9999	0.9998	0.9999	1.0000	1.0000	1.0000
C	1	916	917	983	069	617	05	017	083
2-		1.0000	0.9999	0.9999	0.9998	0.9999	1.0000	1.0000	1.0000
C	1	849	917	983	069	617	05	017	083

8-C	1	1.0000 849	0.9998 602	1.0000 017	0.9999 617	0.9998 052	0.9999 933	0.9999 95	0.9998 335
21-C	1	0.9999 504	1.0000 176	0.9999 962	1.0000 157	1.0000 151	0.9999 962	0.9999 962	1.0000 125
25-C	1	0.9999 504	1.0000 176	0.9999 969	1.0000 157	1.0000 151	0.9999 962	0.9999 969	1.0000 125

Table S18 Relative DFT computed NMR shielding of transition state 4,4'-dimethyl-1,1'-biphenyl pruned using (175,974) grid with rotations on the x axis

Ato	m	0	10 degrees	20 degrees	30 degrees	40 degrees	50 degrees	60 degrees	70 degrees	80 degrees
16-H	1	0.9999 788	0.9998 344	1.0000 042	0.9998 981	0.9999 958	1	1	0.9997 919	
20-H	1	0.9997 919	0.9998 811	1	0.9998 599	1.0000 042	1.0000 042	1	0.9999 788	
13-H	1	0.9997 919	0.9998 811	1	0.9998 599	1.0000 042	1.0000 042	1	0.9999 788	
17-H	1	0.9999 788	0.9998 344	1.0000 042	0.9998 981	0.9999 958	1	1	0.9997 919	
15-H	1	1.0001 896	0.9999 505	1.0000 041	1.0000 165	0.9999 794	0.9999 959	1	0.9997 898	
19-H	1	0.9997 898	1.0000 247	0.9999 959	0.9999 629	1.0000 041	1.0000 041	1	1.0001 896	
14-H	1	0.9997 898	1.0000 247	0.9999 959	0.9999 629	1.0000 041	1.0000 041	1	1.0001 896	
18-H	1	1.0001 896	0.9999 505	1.0000 041	1.0000 165	0.9999 959	0.9999 959	1	0.9997 898	
22-H	1	1.0000 206	1.0000 137	1	1.0088 687	1.0000 034	1	1	1.0000 206	

28-		1.0000	1.0000		1.0090	1.0089		1.0000	1.0000
H	1	206	137	1	783	924	1	034	206
23-		1.0007	0.9998	1.0000	1.0000	1.0000	0.9999		0.9992
H	1	731	331	102	988	136	83	1	133
26-		0.9992	1.0001	0.9999	0.9998	0.9911	1.0000	0.9999	1.0007
H	1	133	396	83	91	046	102	966	731
24-		0.9992	1.0001	0.9999	0.9911	0.9999	1.0000	0.9999	1.0007
H	1	133	396	83	251	796	102	932	731
27-		1.0007	0.9998	1.0000	0.9911	0.9999	0.9999		0.9992
H	1	731	331	102	251	796	83	1	133
Ato		10	20	30	40	50	60	70	80
m	0	degrees							
4-		0.9999	0.9999	0.9999	1.0001	0.9999	0.9999	0.9999	0.9999
C	1	453	5	952	047	952	952	976	453
10-		0.9999	0.9999	0.9999	1.0001	0.9999	0.9999	0.9999	0.9999
C	1	453	5	952	047	952	952	976	453
1-		0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9998
C	1	748	729	875	207	875	875	875	748
7-		0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9998
C	1	748	729	875	207	875	875	875	748
3-		1.0000	0.9998	1.0000	1.0001	1.0000	1.0000	1.0000	1.0000
C	1	772	946	075	205	075	094	094	038
9-		1.0000	1.0001	1.0000	0.9999	1.0000	1.0000	1.0000	1.0000
C	1	038	318	094	925	075	075	056	772
5-		1.0000	1.0001	1.0000	0.9999	1.0000	1.0000	1.0000	1.0000
C	1	038	318	094	925	075	075	056	772
11-		1.0000	0.9998	1.0000	1.0001	1.0000	1.0000	1.0000	1.0000
C	1	772	946	075	205	075	094	094	038
6-		1.0000	0.9999	1.0000	0.9999	0.9999	0.9999		0.9998
C	1	583	018	033	367	983	983	1	835

12-		0.9998	1.0000	0.9999	0.9998	1.0000	1.0000	1.0000	1.0000
C	1	835	366	983	335	017	033	017	583
2-		0.9998	1.0000	0.9999	0.9998	1.0000	1.0000	1.0000	1.0000
C	1	835	366	983	335	017	033	017	583
8-		1.0000	0.9999	1.0000	0.9999	0.9999	0.9999	0.9999	0.9998
C	1	583	018	033	367	983	983	983	835
21-		1.0000	1.0000	0.9999	1.0000	0.9999	0.9999	0.9999	1.0000
C	1	307	213	962	238	969	956	962	307
25-		1.0000	1.0000	0.9999	1.0000	0.9999	0.9999	0.9999	1.0000
C	1	307	213	962	238	969	956	962	307