

Design of 5',7'-Dihydroxyflavones and β -D-Glucopyranosyl Heterocyclic Derivatives as Glycogen Phosphorylase Inhibitors.

by

Daniel Barr

A thesis submitted in partial fulfilment for the requirements for the degree of
MSc (by Research) at the University of Central Lancashire

May 2018

STUDENT DECLARATION FORM



Concurrent registration for two or more academic awards

I declare that while registered as a candidate for the research degree, I have not been a registered candidate or enrolled student for another award of the University or other academic or professional institution

Material submitted for another award

I declare that no material contained in the thesis has been used in any other submission for an academic award and is solely my own work

Collaboration

Where a candidate's research programme is part of a collaborative project, the thesis must indicate in addition clearly the candidate's individual contribution and the extent of the collaboration. Please state below:

Signature of Candidate _____

Type of Award **Masters By Reseach**

School **School of Physical and Computational Sciences**

Abstract

There has been a significant increase in the number of people diagnosed with diabetes from 1980 to 2016, rising from 108 million to approximately 420 million people worldwide, and is predicted to rise to above 640 million by the year 2040 and type II diabetes is seen in 90-95% of those diagnosed. Many treatments currently exist to treat type II diabetes, although there are considerable adverse health effects associated with these drugs including a risk of hypoglycaemia. Accordingly, there is a swift need for a new, effective treatment that has little to no side effect for those suffering from T2D. β -D-Glucopyranosyl derivatives are known to inhibit glycogen phosphorylase, which is a valid target for controlling hyperglycaemia in type 2 diabetes. Computational methods, such as molecular docking with Glide and GOLD as well as post-docking free energy calculations using MM-GBSA calculations were used to screen a library of β -D-glucopyranosyl analogues and we, for the first time, have derived computational models of MMGBSA for the GPb catalytic site have revealed excellent predictive potential based on a thorough statistical analysis. Using these models, correlations between predicted and experimental inhibitory potential as high as 0.95 – 0.97 were obtained for a training set of ligands. These methods have substantial potential for discovery of new effective compounds in the treatment of T2D as thousands of potential ligands could in the future be screened. Previous computational screening of 5,7-dihydroxyflavone analogues which have been predicted to bind at the caffeine binding site have been performed and a number of these analogues have been synthesized and will undergo kinetic experiments which will give insight into the effectiveness of 5',7'-dihydroxyflavone derivatives so that a wider library of similar compounds could be tested that displays biological activity towards glycogen phosphorylase.

Acknowledgements

I would like to express my deepest gratitude to Dr. Joseph Hayes (DoS) and Dr. Timothy Snape (co-supervisor) for their continued guidance and supervision throughout this research project. I would also like to thank Matthew Davies and Benjamin Chetter for their help in a few aspects of this project and the technician staff at the University of Central Lancashire for their help throughout the year.

List of Abbreviations

4-DMAP	4-Dimethylaminopyridine
AcOH	Acetic Acid
ADIs	α -Glucosidase Inhibitors
AMBER	Assisted Model Building and Energy Refinement
AMPK	AMP-Activated Protein Kinase
ATP	Adenosine Triphosphate
BCCs	Bond Charge Corrections
Beta Cells	β -Cells
BVR	Baker-Venkataraman Rearrangement
CCD	Chemical Component Dictionary
CHARMM	Chemistry at HARvard Molecular Mechanics
CoMFA	Comparative Molecular Field Analysis
CSII	Continuous Subcutaneous Insulin Infusion
DCM	Dichloromethane
DMF	Dimethylformamide
DMSO	Dimethylsulphoxide
DR	Diabetic Retinopathy
EtOAc	Ethyl Acetate
GA	Genetic Algorithm
GD	Gestational Diabetes
Glc-1-P	(Glucose-1-Phosphate)
GOLD	Genetic Optimization For Ligand Docking
GP	Glycogen Phosphorylase
GS	Glidescore
H ₂ SO ₄	Sulphuric Acid
HBr	Hydrobromic Acid
HCl	Hydrochloric Acid
HLGPb	Human Liver GPb
IMP	Inosinic Acid

K ₂ CO ₃	Potassium Carbonate
LBDD	Ligand Based Drug Design
LC-MS	Liquid Chromatography – Mass Spectrometry
LiHMDS	Lithium Bis-(trimethylsilyl)amide
LOO-cv	Leave-One-Out Cross Validation
MgSO ₄	Magnesium Sulphate
MM	Molecular Mechanics
MMFF	Merck Molecular Force Field
MM-GBSA	Molecular Mechanics – Generalized Born Surface Area
MOM	Methoxymethyl Chloride
NADH	Nicotinamide Adenine Dinucleotide
NAG	N-Acetyl-β-D-Glucopyranosyl-Amine
NaH	Sodium Hydride
NaOH	Sodium Hydroxide
NMR	Nuclear Magnetic Resonance
OHA	Oral Hypoglycaemic Agents
OPLS3	Optimized Potential for Liquid Screening 3
PCOS	Polycystic Ovary Syndrome
PDB	Protein Data Bank
PLP	Pyridoxal 5'-Phosphate
QM	Quantum Mechanics
QM-PLD	Quantum Mechanics - Polarized Ligand Docking
QSAR	Quantitative Structure-Activity Relationship
RMGPb	Rabbit Muscle GPb
RMSD	Root mean Square Deviation
RMSE	Root Mean Square Error
SBDD	Structure Based Drug Design
SP	Standard Precision
T1D	Type I Diabetes
T2D	Type II Diabetes

TEA	Triethylamine
THF	Tetrahydrofuran
TLC	Thin Layer Chromatography
TZDs	Thiazolidinediones
UDP-Glucose	Uridine Phosphate Glucose
VdW	Van der Waals
XP	Extra Precision

List of Equations

Equation 1: Phosphorylation of Glycogen.

Equation 2: The Hansch Equation

Equation 3: The Free-Wilson

Equation 4: Potential Energy Function

Equation 5: Bonded and Non-Bonded Terms

Equation 6: ΔG_{bind} Equation

Equation 7: GlideScore Scoring Function

Equation 8: RMSD Equation

Equation 9: ChemPLP Fitness

Equation 10: Prime MM-GBSA Equation.

Equation 11. ΔG_{bind} Equation including Entropic Terms

List of Tables

Table 1.1	Differences of Type I and Type II diabetes.....	3
Table 2.1	The Structure of Most Potent Inhibitors of RMGPb Catalytic Site Discovered to Date.....	16
Table 2.2	The Structure Highlighting the Importance of the α - Substituent in Inhibiting RMGPb.....	17
Table 2.3	Potencies of Dihydropyridine AMP site Inhibitors that Bind at the Allosteric Site.....	18
Table 2.4	IC ₅₀ values (μ M) for Inhibition of GP _a and GP _b of Flavonoids Tested in Concentrations up to 50 μ M...24	
Table 4.1	Set of Training Set Ligands used in Computational Studies and their known Inhibitory Constants.....	42
Table 4.2	Table Showing the Solvation of the Various Receptors using Cluster Waters.....	47
Table 4.3	Ligand RMSDs GeneralClusterProtein.....	51
Table 4.4	Ligand RMSDs ClusterResidueIn.....	52
Table 4.5	Ligand RMSDs ClusterResidueOut.....	53
Table 4.6	Glide SP/XP Results, QPLD Results for GeneralClusterProtein.....	55
Table 4.7	Glide SP/XP Results for ClusterResidueIn.....	56
Table 4.8	Glide SP/XP Results for ClusterResidueOut.....	57
Table 4.9	GOLD Results for GeneralClusterProtein.....	58

Table 4.10	Prime MM-GBSA Results GeneralClusterProtein...60
Table 4.11	Prime MM-GBSA GOLD Results for GeneralClusterProtein.....62
Table 4.12	Prime MM-GBSA QM-PLD Results for GeneralClusterProtein.....63
Table 4.13	Prime MM-GBSA Results for ClusterResidueIn.....64
Table 4.14	Prime MM-GBSA Results for ClusterResidueOut....66
Table 4.15	Tautomeric Corrections.....68
Table 4.16	GeneralClusterWater Models 1-4 Including Entropy and Tautomeric Correction.....69
Table 4.17	Strike Statistics Model 1-4.....70
Table 4.18	Ligands used in models 1-4 along with the codes used.....71
Table 4.19	Predicted Ki values for Unknown Set of Ligands....73
Table 5.1	Demethylation products of 2-Methoxyflavone Derivatives.....88
Table 5.2	Different Catalysts used in the Single Ketone Synthesis of 4-Methoxyflavone.....90
Table 5.3	Different Solvents used in the Single Ketone Synthesis of 4-Methoxyflavone.....92

List of Figures

Figure 1.1	Map Showing the Number of Those Diagnosed with Diabetes in 2015, and the Respective Predicted Amount in 2040.....	1
Figure 1.2	Metformin. The Most common Biguanide.....	6
Figure 1.3	Typical Sulphonylurea Structure.....	7
Figure 1.4	Acarbose, an example of an α -Glucosidase inhibitor.....	8
Figure 1.5	The Structure of Meglitinide.....	9
Figure 2.1	Glycogen Phosphorylase Dimer. The known Binding Sites are Highlighted.....	11
Figure 2.2	The processes of both glycolysis and gluconeogenesis.....	12
Figure 2.3	The structure of Glycogen phosphorylase in the inactive T state and the active R state respectively.....	14
Figure 2.4	The mode of action of PLP as an essential co-factor.....	15
Figure 2.5	The structure of NAG.....	18
Figure 2.6	The different ligand interactions of a dihydropyridine dicarboxylic acid.....	20

Figure 2.7	Binding of ellagic acid at the GP inhibitor site as determined by X-ray crystallography (PDB code: 4YUA).....	21
Figure 2.8	Binding of quercetin at the Quercetin binding site as determined by X-ray crystallography (PDB code: 4MRA).....	22
Figure 2.8	Various different flavonoid structures, all follow the typical flavonoid backbone.....	25
Figure 3.1	Overview of drug discovery based on Structure based drug design (SBDD).....	29
Figure 3.2	A workflow showing the general process followed in docking.....	34
Figure 3.3	Overview of rigid docking ligand ensemble method.....	35
Figure 4,1	The Different Residue Orientations.....	44
Figure 4.2	Correlation of the Aromatic R groups In Ki against glide gscore using SP without H ₂ O's present.....	55
Figure 4.3	Prime (No strain) correlation of Aromatic R groups without waters using input poses from glide-SP.....	60
Figure 4.4	Correlation between predicted and experimental In Ki from Prime (NS) using Model 1.....	71

Figure 4.5	Correlation between predicted and experimental $\ln K_i$ from Prime (NS) using Model 2.....	72
Figure 4.6	Correlation between predicted and experimental $\ln K_i$ from Prime (NS) using Model 3.....	72
Figure 4.7	Correlation between predicted and experimental $\ln K_i$ from Prime (NS) using Model 4.....	73
Figure 4.8	Predicted binding pose of 2,5-Diazole using model 3.....	78
Figure 4.9	Predicted binding pose of 2,4-Diazole using model 3.....	78
Figure 5.1	Binding of the flavone chrysin at the GP inhibitor site.....	81
Figure 5.2	Basic backbone structure of flavone, showing the hydroxyl groups at the 5 and 7 position.....	82
Figure 5.3	2'-hydroxychalcone.....	89
Figure 5.4	1-(2-hydroxyphenyl)-3-phenyl-2-propenone.....	89
Figure 6.1	The glycogenolysis pathway.....	97
Figure 6.2	The basic structure of a flavone. Hydroxyl substituents are shown at the 5 and 7 position.....	98
Figure 6.3	The structure of 3'-hydroxyflavone.....	105
Figure 6.4	The 3'-benzoylflavone structure formed.....	108

Figure 6.5	¹ H-NMR of 3-(5-bromo-2-chloro-benzoyl)-2-(5-bromo-2-chloro-phenyl)-5,7-dihydroxy-chromen-4-one....	109
Figure 6.6	The mixture of products formed in the attempted synthesis of 2-(2-chlorophenyl)-5,7-dihydroxy-chromen-4-one.....	110
Figure 6.7	¹ H-NMR of 3-(2-chlorobenzoyl)-2-(2-chlorophenyl)-5,7-dihydroxy-chromen-4-one.....	111
Figure 6.8	LC-MS showing the 1:1 ratio of O-Cl(a):O-Cl(b)...	113

List of Schemes

Scheme 5.1	The mechanism involved in the Baker-Venkataraman Rearrangement.....	83
Scheme 5.2	The General Method Used.....	84
Scheme 5.3	Synthesis of Tricin.....	85
Scheme 5.4	Microwave Synthesis of Flavones.....	86
Scheme 5.5	Demethylation Rearrangement of 2'-Methoxyflavones.....	87
Scheme 5.6	Demethylation Rearrangement Derivatives...	88
Scheme 5.7	Synthesis of 2'-Hydroxychalcone.....	89
Scheme 5.8	Synthesis of 1-(2-hydroxyphenyl)-3,2-propenone.....	90
Scheme 5.9	Synthesis of 4'-Methoxyflavone using an Iodine Catalyst.....	91
Scheme 5.10	Cyclocondensation of 1-(2-hydroxyphenyl)-3-phenyl-3-propenone.....	92
Scheme 5.11	Synthesis of 8'-Bromochrysin.....	93
Scheme 5.12	Flavone Synthesis with Protecting Groups....	94
Scheme 5.13	Flavone Synthesis with a Palladium Catalyst.....	95
Scheme 6.1	General Method of Synthesising 5',7'-Dihydroxyflavones.....	99

Scheme 6.2	Synthesis of 2-(4-Bromophenyl)-5',7'- Dihydroxy-chromen-4-one.....	100
Scheme 6.3	Synthesis of 2-(3-Fluorophenyl)-5',7'- Dihydroxy-chromen-4-one.....	101
Scheme 6.4	Synthesis of 2-(3-Chloro-4-Fluoro-phenyl)-5',7'- Dihydroxy-chromen-4-one.....	102
Scheme 6.5	Synthesis of 2-(3-Bromo-4-Fluoro-phenyl)-5',7'- dihydroxy-chromen-4-one.....	103
Scheme 6.6	Mechanism of the Formation of 3'- Benzoylflavones.....	106
Scheme 6.7	Synthesis of 2-(5-Bromo-2-Chloro-phenyl)-5',7'- dihydroxy-chromen-4-one.....	107
Scheme 6.8	Hydrolysis of (2-Acetyl-3,5-Dimethoxy-phenyl) 2-Methoxybenzoate.....	114
Scheme 6.9	Synthesis of 5,7-dihydroxy-2-(2- hydroxyphenyl)chromen-4-one.....	115

Table of Contents

Abstract.....	ii
Acknowledgement.....	iii
List of Abbreviations.....	iv
List of Equations.....	vii
List of Tables.....	viii
List of Figures.....	x
List of Schemes.....	xiv
1. Introduction	1
1.1 Introduction to Diabetes	1
1.1.1 Prevalence of Diabetes.....	1
1.1.2 Type I Diabetes.....	2
1.1.3 Type II Diabetes.....	2
1.1.4 Gestational Diabetes.....	3
1.2 Symptoms of Diabetes.....	4
1.3 Long-Term Diabetes Complications.....	4
1.4 Current Treatment.....	5
1.4.1 Lifestyle Changes.....	5

1.4.2	Oral Hypoglycaemic Agents.....	6
1.4.3	Biguanides.....	6
1.4.4	Sulphonylureas.....	7
1.4.5	α -Glucosidase Inhibitors.....	8
1.4.6	Meglitinides.....	8
1.4.7	Thiazolidinediones.....	9
1.4.8	Insulin Therapy.....	9
1.5	Synopsis.....	10
Chapter 2. Glycogen Phosphorylase.....		11
2.1	Introduction.....	11
2.2	Glycogenolysis and Gluconeogenesis.....	11
2.3	Structure of Glycogen Phosphorylase.....	13
2.4	The Catalytic Site.....	14
2.5	Allosteric Sites.....	18
2.6	Inhibitor Site.....	20
2.7	Quercetin Binding Site.....	22
2.8	Flavonoids.....	23
2.9	Conclusion.....	27

Chapter 3. Molecular Modelling.....	28
3.1 Introduction.....	28
3.2 Molecular Mechanics.....	31
3.2.1 Introduction.....	31
3.2.2 Forcefield.....	32
3.3 Molecular Docking.....	33
3.3.1 Introduction.....	33
3.3.2 Search Algorithms.....	34
3.3.3 Scoring Functions.....	35
3.3.4 Glide Docking.....	36
3.3.5 Quantum-Mechanics Polarized Ligand Docking.....	37
3.3.6 GOLD.....	38
3.3.7 Prime MM-GBSA.....	38
3.4 Conclusion.....	39
Chapter 4. In Silico Screening of β -D-Glucopyranosyl Analogues as Inhibitors of Glycogen Phosphorylase.....	40

4.1	Introduction.....	40
4.2	Aim.....	41
4.2.1	Specific Objectives.....	41
4.3	Computational Details.....	41
4.3.1	Ligand Preparation.....	41
4.3.2	Protein Preparation.....	44
4.3.3	Solvation Modelling.....	45
4.3.4	Docking with the Imidazole Receptor With GeneralClusterProtein Water Model.....	47
4.3.5	Docking using Proteins containing Different Residue Orientations.....	48
4.3.6	Quantum-Mechanics Polarized Ligand Docking.....	48
4.3.7	GOLD Docking Details.....	48
4.3.8	Docking Performance Assessment.....	49
4.3.9	Prime MM-GBSA Post-Docking.....	50
4.4	Results and Discussions.....	51
4.4.1	Glide Docking.....	51
4.4.2	Docking – Relative Ranking of Ligands and Scoring Results.....	54
4.4.3	Prime MM-GBSA Results.....	59
4.4.4	Refinement of MM-GBSA Model.....	67

4.4.5	Prediction of Inhibitory Potential for Unknown Compounds.....	75
4.5	Conclusion.....	79
Chapter 5. Synthetic Methods.....		81
5.1	Introduction.....	81
5.2	Baker-Venkataraman Rearrangement.....	83
5.2.1	β -Diketone Cyclization.....	84
5.3	Microwave Synthesis.....	85
5.4	Demethylation Rearrangement.....	87
5.5	Single Ketone Synthesis.....	89
5.6	Metal Catalyst Synthesis.....	95
5.7	Conclusion.....	96
Chapter 6. Synthetic Results.....		97
6.1	Introduction.....	97
6.2	Synthetic Attempt of 2-(4-Bromophenyl)-5',7'-Dihydroxy-chromen-4-one.....	100

6.3	Synthetic Attempt of 2-(3-Fluorophenyl)-5',7'- Dihydroxy-chromen-4-one.....	101
6.4	Synthetic Attempt of 2-(3-Chloro-4-Fluorophenyl) -5',7'-dihydroxy-chromen-4-one.....	102
6.5	Synthetic Attempt of 2-(3-Bromo-4-Fluorophenyl) -5',7'-dihydroxy-chromen-4-one.....	103
6.6	The Method Limitations.....	104
6.7	Synthetic Attempt of 2-(5-Bromo-2-Chlorophenyl) -5',7'-dihydroxy-chromen-4-one.....	107
6.8	Synthetic Attempt of 2-(2-Chlorophenyl)-5',7'- Dihydroxy-chromen-4-one.....	110
6.8.1	First Attempt at Synthesis of 2-(2- chlorophenyl)-5',7'-dihydroxy-chromen-4- one.....	110
6.8.2	Second Attempt at Synthesis of 2-(2- chlorophenyl)-5',7'-dihydroxy-chromen-4- one.....	112
6.8.3	Third Attempt at Synthesis of 2-(2- chlorophenyl)-5',7'-dihydroxy-chromen-4- one.....	112
6.9	Synthetic attempt of 5,7-dihydroxy-2-(2- hydroxyphenyl)chromen-4-one.....	114

6.9.1	Second Attempt of synthesizing (2-acetyl-3,5-dimethoxy-phenyl) 2-methoxybenzoate.....	115
6.10	Conclusion.....	115
6.11	Experimental.....	115
6.11.1	2-(4-Bromophenyl)-5',7'-dihydroxy-chromen-4-one.....	118
6.11.2	2-(3-Fluorophenyl)-5',7'-dihydroxy-chromen-4-one.....	119
6.11.3	2-(2-Chloro-4-fluorophenyl)-5',7'-dihydroxy-chromen-4-one.....	120
6.11.4	2-(3-Bromo-4-fluorophenyl)-5',7'-dihydroxy-chromen-4-one.....	122
6.11.5	3-(5-Bromo-2-chloro-benzoyl)-2-(5-bromo-2-chloro-phenyl)-5,7-dihydroxy-chromen-4-one.....	123
6.11.6	5,7-Dihydroxy-2-(2-hydroxyphenyl)chromen-4-one.....	125
Chapter 7. Conclusion		127
References.....		129
Appendix.....		135

Chapter 1 – Introduction.

1.1 Introduction to Diabetes

1.1.1 Prevalence of Diabetes

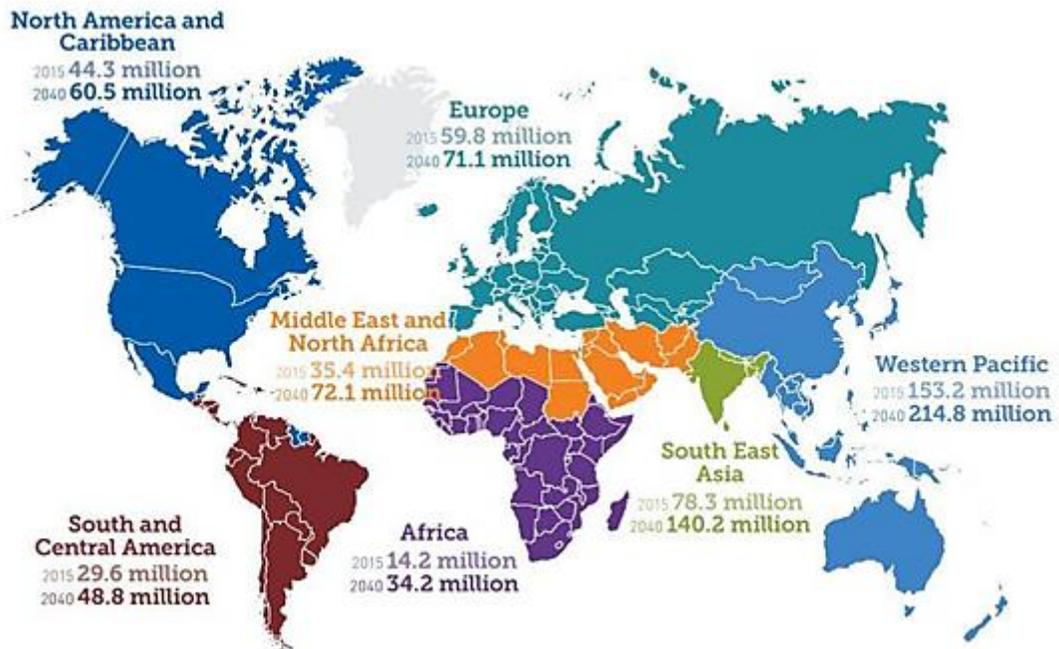


Figure 1.1 - Map showing the number of those diagnosed with diabetes in 2015, and the respective predicted amount in 2040.^[2]

There has been a significant increase in the number of people diagnosed with diabetes from 1980 to 2016, rising from 108 million to approximately 420 million people worldwide, and is predicted to rise to above 640 million by the year 2040 at the current rate. (Figure 1.1) ^[3] Each year, 7 million people develop diabetes and the most dramatic increases in type 2 Diabetes (T2D) have occurred in populations where there have been rapid and major changes in lifestyle, demonstrating the important role played by lifestyle factors with respect to diabetes.^[4] Over the past decade, diabetes has risen substantially in countries at all income levels, mirroring the global increase in the number of people who are overweight or obese, but has risen faster in low- and middle-income countries. The global prevalence of diabetes has grown from 4.7% in 1980 to 8.5% in 2014.^[5]

1.1.2 Type I Diabetes

Type I diabetes (T1D) or diabetes mellitus is characterized by the absence of insulin production within the pancreas. The onset of T1D usually begins in childhood or adolescence and is classified as an autoimmune disease. In T1D, beta cells (β -cells) in the pancreas that produce insulin are gradually destroyed. Eventually, insulin deficiency is absolute. Without insulin (to move glucose into cells) blood glucose levels become excessively high, a condition known as hyperglycaemia. Because the body cannot utilize the sugar, it persists in the blood stream before entering into the urine (glycosuria), where it is lost.^[6] T1D is less common than T2D and consists of only 5-10% of cases of diabetes with both males and females being of equal risk. T1D can occur at any age but usually between infancy and the late 30s. A few risk factors that increase the likelihood of developing the disease are; being ill in early infancy, having a parent with T1D, having an older mother and having other autoimmune disorders for example.^[6] The symptoms associated with T1D can take some time to show as β -cell insulin production can still occur. At the point of no insulin production, the symptoms progress rapidly. Frequent urination, unusual thirst, extreme hunger, weight loss, weakness, fatigue, changes in eyesight are all associated with T1D (Table 1.1.)

1.1.3 Type II Diabetes.

Type II diabetes or non-insulin dependent diabetes mellitus is the most common form of diabetes and is seen in 90-95% of those diagnosed. There are several factors that can increase the risk of developing type 2 diabetes. Obesity, unhealthy diet, physical inactivity, smoking, having a first-degree relative with T2D and being of South Asian and African-Caribbean descent are examples of factors that can increase the likelihood of developing T2D (Table 1.1.)^[7] Too much glucose left to circulate in the blood (hyperglycaemia) can cause tissue damage whereas too little glucose (hypoglycaemia) can lead to tiredness, shaking and loss of consciousness, demonstrating the importance of insulin in blood glucose level regulation. As well as the translocation of glucose into cells,^[8] insulin also regulates the breakdown of glycogen to glucose (glycogenolysis) and also the production of glucose from

pyruvate and other non-carbohydrate substances (gluconeogenesis). T2D is characterised by three factors; (1) some resistance of insulin action on glucose uptake in tissues such as adipocytes, liver tissue and skeletal muscle tissues. (2) A relatively low production of insulin from the β -cells of the pancreas and (3) an impaired insulin action to inhibit the hepatic glucose production. [9] The T2D form of diabetes frequently goes undiagnosed for many years because the hyperglycaemia develops gradually and is less severe at earlier stages.[10]

Table 1.1 – Differences of Type I and Type II diabetes^[11]

	Type I	Type II
Onset	Acute; severe	Mild-severe; often insidious
Insulin Secretion	None	Variable
Insulin Sensitivity	Normal	Decreased
Insulin Dependence	Permanent	Temporary; may occur later
Racial/ethnic groups at increased risk	All (low in Asians)	African Americans, Hispanics, Native Americans, Asian/Pacific Islanders.
Association: Obesity	No	Strong
Autoimmune etiology	Yes	No

1.1.4 Gestational Diabetes.

Gestational diabetes (GD) is a temporary disease that can affect pregnant women and arises typically in the second or third trimester and this usually dissipates after the pregnancy. During the later stages of pregnancy as the foetus develops, there is an increased need for insulin by up to two to three times that of normal levels. If the body cannot make the required amount of insulin, glucose will remain in the blood stream and cause hyperglycaemia.[12] Most women suffering from GD will not

require insulin and instead blood sugar levels can be reduced through changes in diet and exercise. Although the majority of women with GD will experience a normal pregnancy and give birth to healthy babies, GD can cause problems such as: polyhydramnios – an excess of amniotic fluid, which can cause premature labour, the baby growing larger than usual – possibly leading to difficulties during delivery and pre-eclampsia – a condition that causes high blood pressure during pregnancy which can lead to pregnancy complications if not treated. GD will usually disappear after pregnancy but women who have suffered from this are more likely to develop T2D in the future.^[13]

1.2 Symptoms of Diabetes.

The symptoms of diabetes are caused as a result of glucose remaining in the blood rather than entering the cells for glucose metabolism. The symptoms are seen in both T1D and T2D and develop quickly in T1D often over a few weeks but these symptoms can take longer to manifest in T2D and typically develop over a few years - often going undiagnosed. Symptoms such as more frequent urination, particularly at night, increased thirst, tiredness, and unexplained weight loss, slow healing and blurred vision are all associated with diabetes. Hyperglycaemia causes the main symptoms of diabetes and can occur from ineffective diabetes medication, eating too much of the wrong foods or being unwell.^[14]

1.3 Long-Term Diabetes Complications

There are many possible long term complications associated with diabetes usually effecting the eyes (retinopathy), heart (cardiovascular disease), kidneys (nephropathy), and nerves and feet (neuropathy). The eyes are particularly at high risk of damage with those who suffer from diabetes. Diabetes is associated with damage to the tiny blood vessels in the retina leading to diabetic retinopathy (DR). Untreated, DR can cause blood vessels in the retina to leak fluid or haemorrhage (bleed), distorting vision. In its most advanced stage, new abnormal blood vessels proliferate on the surface of the retina, leading to cell loss of the area.^[15] People with diabetes are also at a much greater risk of developing problems with their feet, due

to the damage caused by raised blood sugar relating to sensation and reduced circulation. Untreated, these problems can cause foot ulcers, infections and may possibly lead to amputation.^[16] Those with diabetes are twice as likely as non-sufferers to have heart disease or a stroke, and those same people tend to develop heart disease or have strokes at an earlier age than other people. Over time, high blood glucose levels can lead to increased deposits of fatty materials on blood vessel walls. These deposits may affect blood flow, increasing the chance of clogging and hardening of blood vessels (atherosclerosis).^[17] Kidney disease develops very slowly in people who have diabetes for over a period of about 20 years. About one in three people with diabetes will develop kidney disease, this is decreasing as treatments improve. Hyperglycaemia causes damage to small blood vessels. This damage can cause the kidneys to work less efficiently. Keeping a normal blood glucose levels can greatly reduce the risk of kidney disease.^[16]

1.4 Current Treatment

Each sufferer of diabetes requires individualized treatment. Those with T1D will always require insulin as well as diet and exercise changes. T2D may require insulin or oral hypoglycaemic agents (medication that results in lower blood sugar), if diet and exercise alone fail to lower blood glucose levels.

1.4.1 Lifestyle Changes

It should not be understated that lifestyle changes can cause serious improvements in health of those who suffer from diabetes. Dietary changes can be powerful in preventing and reversing diabetes. Diabetic experts have taken care to limit fats – especially saturated fats that can raise cholesterol levels, and to limit protein for those with impaired kidney function. Fat is a problem for people with diabetes as increased fat in the diet reduces the efficacy of insulin to get glucose into cells. It is advised that patients eat unlimited amounts of grains, fruits and vegetables. Modest amounts of non-fat condiments, alcohol and caffeine containing beverages are fine. High blood pressure is often associated with diabetes and as a result, patients are advised to limit the amount of salt in the diet.^[18] A study found that patients taking

oral medications and patients on insulin were able to get off their medications after some days on a near-vegetarian diet and exercise program.^[19] Regular physical activity helps the body cells take up glucose for metabolism and this lowers blood glucose levels. Regular exercise also helps weight loss as well as controlling blood cholesterol and blood pressure both of which can be caused as a result of diabetes. The benefits of a regular aerobic exercise program in diabetes management include decreased need for insulin, decreased risk of obesity, and decreased risk of heart disease.^[18]

1.4.2 Oral Hypoglycaemic Agents.

As T2D is a progressive disease, lifestyle changes alone may not be enough to control elevated blood glucose levels and a patient may also need to take medication. Oral hypoglycaemic agents (OHA) are an example of such drugs and are all similar in that they reduce blood glucose levels.

1.4.3 Biguanides

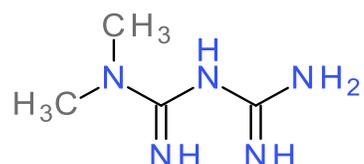


Figure 1.2 Metformin. The Most common Biguanide.

Metformin is the most common OHA prescribed for the treatment of T2D (Figure 1.2), being prescribed to at least 120 million people worldwide. The main effect of this drug from the biguanide family is to acutely decrease hepatic glucose production, mostly through a mild and transient inhibition of the mitochondrial respiratory-chain complex 1. The resulting decrease in hepatic energy activates the AMP-activate protein kinase (AMPK), a cellular metabolic sensor, providing a generally accepted mechanism for metformin action on hepatic gluconeogenesis.^[20] In addition, there is evidence for other beneficial effects, including a reduction in cardiovascular disease^[21] as well as being an effective ovulation induction agent for

non-obese women for polycystic ovary syndrome (PCOS)^[22] Metformin does not have many significant adverse effects but could possibly cause lactic acidosis, causing dizziness, severe drowsiness, muscle pain, tiredness, chills, blue/cold skin and stomach pains. Metformin isn't particularly known for causing hypoglycaemia; however, low blood glucose levels may occur if this drug is used with other anti-diabetic drugs. Hypoglycaemia is more likely to occur with heavy exercise, drinking large amounts of alcohol, or not consuming enough calories from food.^[23]

1.4.4 Sulphonylureas

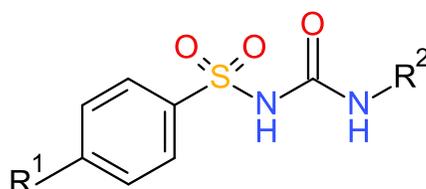


Figure 1.3 - Typical Sulphonylurea structure.

Sulphonylureas can be used to treat T2D as they are able to stimulate insulin secretion from pancreatic β -cells (Figure 1.3). The principle target of these drug is the inhibition of adenosine triphosphate (ATP)-sensitive potassium (K_{ATP}) channel, which plays a major role in controlling the β -cell membrane potential. The K_{ATP} channel consists of an inwardly rectifying K^+ channel and a sulphonylurea receptor. Binding of either ATP or a sulphonylurea to the intracellular domains of the subunit produces channel inhibition.^[24] Inhibition of K_{ATP} channels causes depolarization of the β -cell membrane; in turn, this triggers the opening of the voltage-gates Ca^{2+} channels, eliciting Ca^{2+} influx and a rise in intracellular Ca^{2+} which stimulates the exocytosis of insulin containing secretory granules.^[25] Although Sulphonylureas are usually well tolerated, the most common side effect is hypoglycaemia which is more common with long-acting sulphonylureas such as chlorpropamide and glibenclamide.^[26]

1.4.5 α -Glucosidase inhibitors

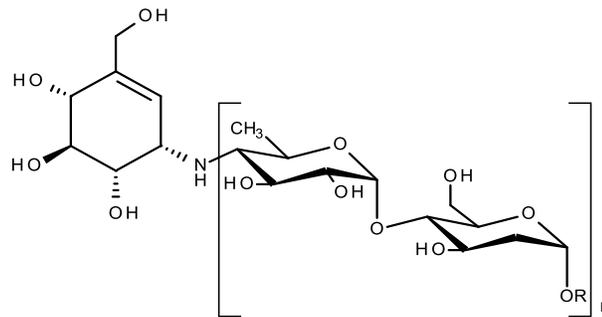


Figure 1.4 - Acarbose, an example of an α -Glucosidase inhibitor.

An effective management of T2D to decrease hyperglycaemia is to retard the absorption of glucose by inhibition of carbohydrate hydrolysing enzymes, such as α -glucosidase (Figure 1.4). α -Glucosidase is the key enzyme in catalysing the final step in the digestive process of carbohydrates. Therefore, α -glucosidase inhibitors (ADI) can retard the liberation of D-glucose from dietary complex carbohydrates and delay glucose absorption.^[27] Examples of ADI's include: acarbose, miglitol and voglibose. Typical side effects of ADI's concern gastrointestinal problems such as: flatulence and diarrhoea.

1.4.6 Meglitinides

Meglitinides work in a similar way to sulphonylureas in that they stimulate insulin secretion from the pancreas (Figure 1.5). Meglitinides bind to a different part of the sulphonylurea receptor than sulphonylurea drugs, which may mean less effect on ischemic preconditioning, but have the same effect on increasing β -cell apoptosis.^[28] Meglitinides have very few side effects and those caused are often caused by wrong dosage causing hypoglycaemia.

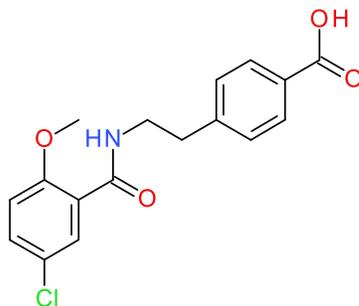


Figure 1.5 - The structure of Meglitinide.

1.4.7 Thiazolidinediones

Thiazolidinediones (TZDs) have various effects that help to control hyperglycaemia, but their primary mechanism of action is to improve insulin sensitivity. A common TZD, pioglitazone, stimulates the peroxisome proliferator-activated receptor gamma, which alters key transcription genes responsible for carbohydrate and lipid metabolism.^[29] The side effects associated with TZDs are more notable than other OHAs and include edema, weight gain, macular edema and heart failure.^[30]

1.4.8 Insulin Therapy

Insulin therapy is recommended for patients with T2D with initial A1C levels (test measuring the glucose blood levels by assessing the amount of glycated haemoglobin) greater than 9 percent or if the diabetes is uncontrolled despite receiving OHAs.^[31] Numerous studies have shown that patients with experienced physicians can achieve excellent glycaemic control with insulin therapy and will experience very few adverse health effects.^[32] Weight gain is very common in patients undergoing insulin therapy as there is reduced glycosuria and an increase of glucose uptake in cells and it is therefore suggested that dietary and exercise modifications would be beneficial.^[33] Another common side effect of insulin therapy is hypoglycaemia, especially if used in combination therapy with other OHAs. There are a few possible ways a patient may administer insulin to the body, the most common of which is insulin injections where the insulin is injected into a muscle. Long term insulin injections can cause localized lipoatrophy at the injection site, considered to be an adverse immunological side effect of insulin therapy.^[34] Another

popular method is to use an insulin pen which use replaceable cartridges containing an insulin reservoir. Continuous Subcutaneous Insulin Infusion (CSII) devices or insulin pumps are a modern form of insulin delivery and are programmed to deliver insulin under the skin. These devices are sometimes fitted with internal blood glucose monitors and are sometimes able to automatically deliver insulin when needed.

1.5 Synopsis

Type II diabetes mellitus is becoming increasingly common due to the changing lifestyles and affects approximately 420 million people worldwide, and is predicted to rise to above 640 million by the year 2040 at the current rate.^[3] This thesis investigates the prospect of targeting glycogen phosphorylase (GP) as a potential treatment of T2D and in particular, *in silico* screening of β -D-glucopyranosyl derivatives with different 5-membered heterocyclic linkers targeting the catalytic site and synthesis of flavonoid analogues as inhibitors of the caffeine binding site. Liver GP is an important target for the treatment of T2D due to the fact that that this enzyme catalyses the breakdown of glycogen to glucose, therefore inhibition can reduce hyperglycaemia.

In this project, by exploiting docking and post-docking computational methods of β -D-glucopyranosyl derivatives using different parameters, effective statistical models can be created and then used to predict potential candidates for synthesis. Synthetic attempts of 5,7-dihydroxyflavone derivatives with the potential to inhibit GP with more efficacy than chrysin was explored, where a synthetic protocol provided by collaborators was optimised for these syntheses by exploiting the Baker-Venkataraman re-arrangement and the flavones produced will eventually undergo kinetic experiments at our collaborators (University of Thessaly, Greece) to determine any inhibition. This data will direct for further lead optimisation efforts from exploiting structure activity relationship methods, with the goal of designing novel, effective compounds which can be used in the treatment of T2D which are safer than medication currently on the market.

Chapter 2 – Glycogen Phosphorylase

2.1 Introduction

Glycogen Phosphorylase (GP (EC number 2.4.1.1) is an essential enzyme in the regulation of blood glucose levels (Figure 2.1). GP catalyses the first step of glycogenolysis through phosphorylation of the terminal α -1,4-glycosidic bond to yield glucose-1-phosphate (Glc-1-P).^[35] The equation for this phosphorylation reaction is;



Due to the importance of GP in glycogen metabolism, this enzyme has become an attractive target for new therapeutic agents with the ability to directly control blood glucose levels in T2D by inhibition of its catalytic site activity.

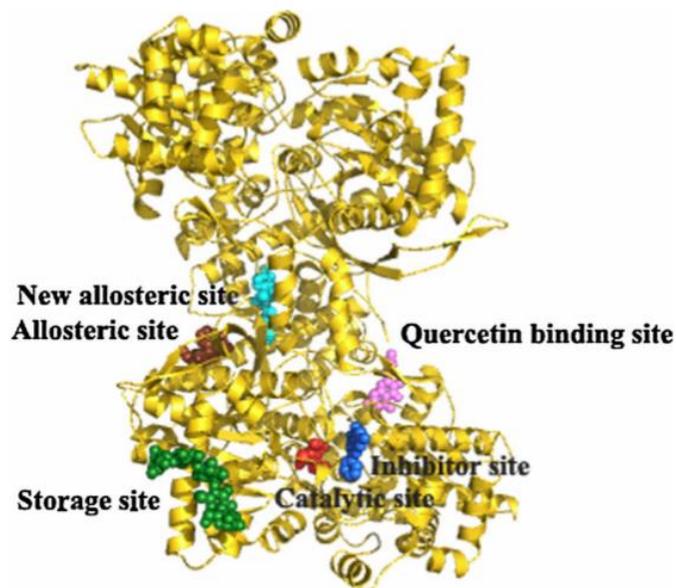


Figure 2.1 - Glycogen phosphorylase dimer. The known binding sites are highlighted.^[36]

2.2 Glycogenolysis and Gluconeogenesis

Generation of glucose in the body arises from the processes of either glycogenolysis or gluconeogenesis (Figure 2.2), both of which are regulated by insulin. Glycogenolysis produces glucose by breaking down glycogen into the smaller

glucose subunits (which can then be used to form pyruvate through glycolysis) whereas gluconeogenesis is the synthesis of glucose by the use of the substrate pyruvate and other non-carbohydrate precursors.^[37] Glycogenolysis occurs in the muscle, brain and liver tissues and is stimulated by hormonal responses. Epinephrine markedly stimulated glycogenolysis in the muscles and, to a lesser

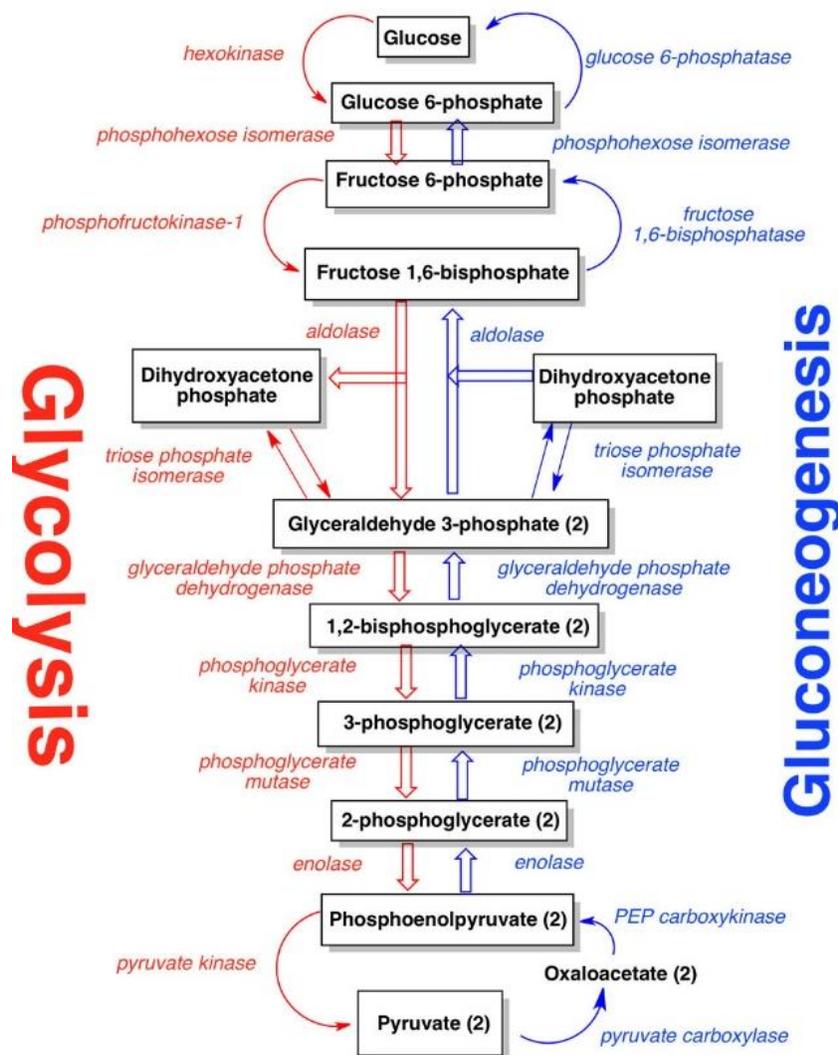


Figure 2.2 - The processes of both glycolysis and gluconeogenesis.^[1]

extent, in the liver. The liver is more responsive to glucagon, a polypeptide hormone secreted by the α -cells of the pancreas.^[38] Gluconeogenesis occurs predominantly in the liver but is also known to occur in the cortex of the kidneys during periods of

fasting, starvation, low-carbohydrate diets, or intense exercise and the pathway requires eleven enzyme-catalysed steps.^[39] The process of glycolysis has a pathway which requires ten enzymes and is the near reversal of gluconeogenesis, altered only by the use of a few differing enzymes.

2.3 Structure of Glycogen Phosphorylase

GP is an enzyme of relative high abundance in muscle, brain and liver that is composed of 842 amino acids which has a mass of 97,434 kDa. The crystal structure of GP is well known with over 215 solved structures published on the protein data bank.^[40] GP is a dimer existing in two interconvertible forms controlled by allosteric interactions and reversible phosphorylation; GP_b is the non-phosphorylated form and is less active. GP_b is converted to GP_a, the more active form, upon the enzyme catalysed phosphorylation of the Serine-14 (Ser-14) residue in each subunit by phosphorylase kinase.^[41] Both forms of the enzyme can also exist in two states, the R state which shows high activity and high substrate binding affinity, and the T state which shows low activity and low substrate binding affinity (figure 2.3.) The equilibrium for GP_a favours the R state whereas the equilibrium for GP_b favours the T state.^[42] The R and T states trigger some structural changes, the greatest occurring in residues 262-290, consisting of tower helices (α 7 helices), α 8 helices and the connecting 280s loop with Asp283-Asn284-Phe285. The angle between the two helices change from +20° in the T state to -80° in the R state. In the T state, the tower helices pack antiparallel with a cluster of hydrogen bonds between Asn270 and Asn274. Tyr262 is in van der Waals contact with Pro281 of the 280s loop of the other subunit. This loop blocks access to the catalytic site in the T state. In the R state, the tower helices change the angle of tilt with respect to each other. The Asn/Asn contacts are broken and there are no contact between Tyr262 and the 280s loop of the other subunit. The 280s loop is disordered and no longer blocks access to the catalytic site.^[43] To date, six different GP ligand binding sites have been identified: the catalytic, allosteric, new allosteric, inhibitor, glycogen storage and a very recently discovered quercetin-binding site (Figure 2.1).

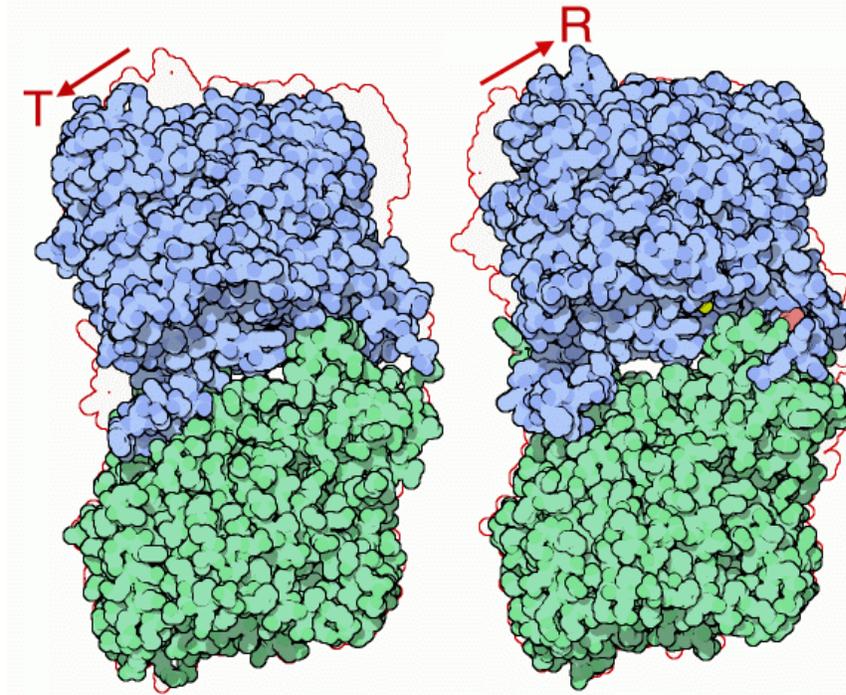


Figure 2.3 - The structure of Glycogen phosphorylase in the inactive T state and the active R state respectively.^[44]

2.4 The Catalytic Site

The catalytic site is the active site of GP, where glycogen is phosphorylated to glucose-1-phosphate (G-1-P) and then shortened by one sub-unit. Accessible to bulk solvent through a 15 Å long channel, the catalytic site is located at the junction of the N and C terminal domains. Each is an α/β structure consisting of a β -sheet core surrounded by a tier of α -helices.^[36] The active site residues are located on flexible loops of polypeptide chain emanating from the domain boundaries. α -D-glucose, the physiological inhibitor of GP ($K_i = 1.7\text{mM}$) participates in at least five well-defined hydrogen bonds with the surrounding residues and presents a complementary molecular surface to the active site at the hydrogen-bonded positions of the ligand. The catalytic site also contains pyridoxal 5'-phosphate (PLP) as an essential cofactor which is located at the center of the GP monomer subunit and is attached by a covalent Schiff base linkage to Lys-680 formed by a reaction of the aldehyde group of the PLP and the ϵ -amino group of the lysine residue in the C-terminal domain.

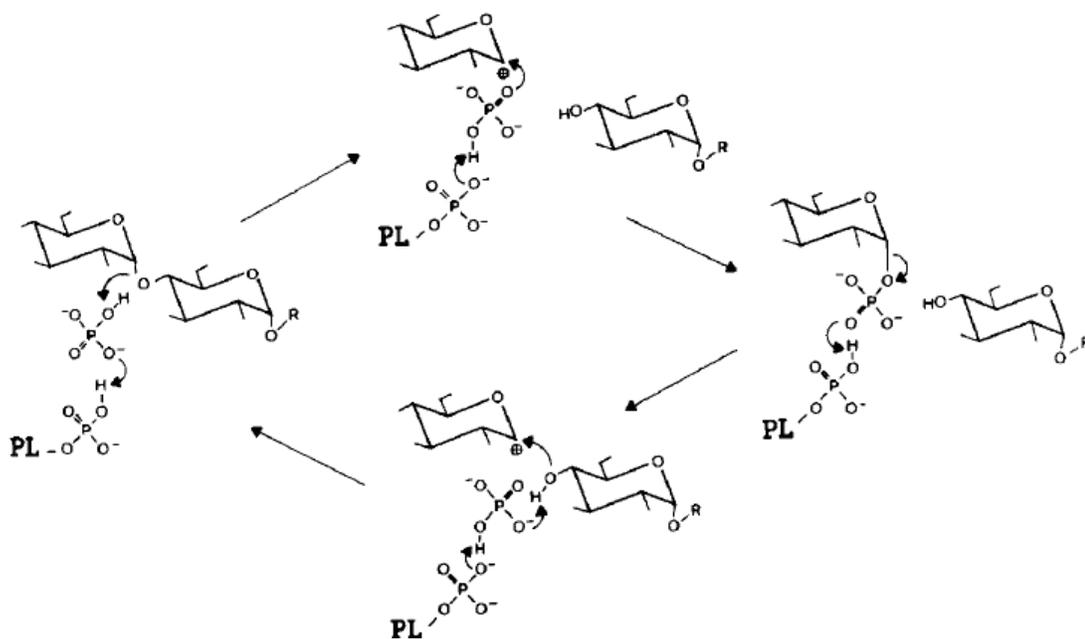
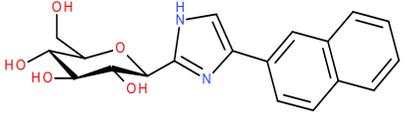
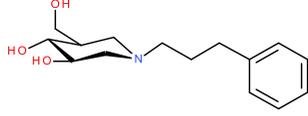
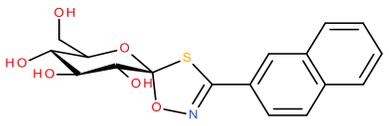
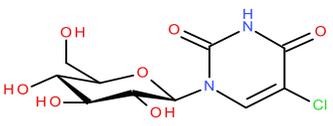
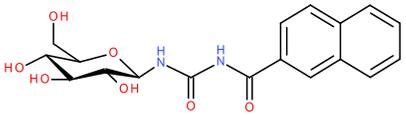
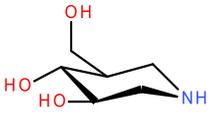
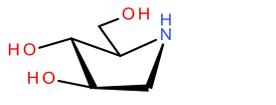
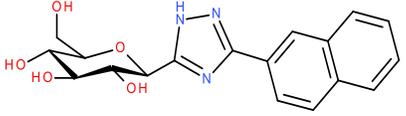
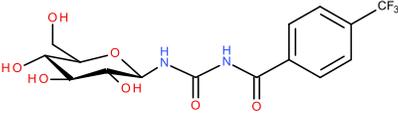
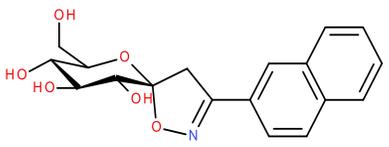
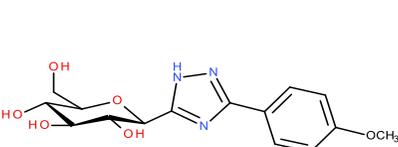
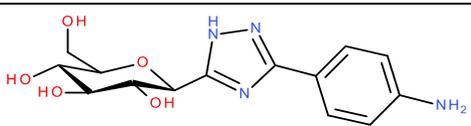
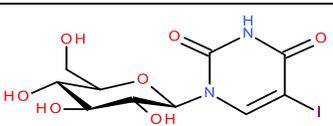
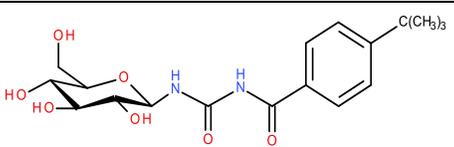


Figure 2.4 – The mode of action of PLP as an essential co-factor.^[45]

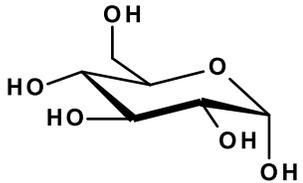
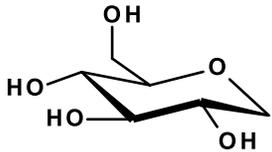
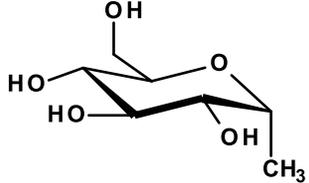
The role of the co-factor PLP (figure 2.4) is to donate a proton from one of its phosphate groups to an inorganic phosphate molecule bound to the terminal glucose of a glycogen chain, allowing the inorganic phosphate to in turn be deprotonated by the oxygen forming the α -1,4 glycosidic linkage, the glycogen chain is separated from the terminal glycogen in an S_N1 mechanism, resulting in the formation of a glucose molecule with a secondary carbocation at the 1 position. The deprotonated inorganic phosphate acts as a nucleophile and bonds with the carbocation, resulting in the formation of G-1-P and a glycogen chain shortened by one glucose molecule.^[46] The phosphorylation site, Ser-14, is contained within a region of polypeptide that binds across the subunit interface (residues 5–22) and is phosphorylated by phosphorylase kinase, and is dephosphorylated by the protein phosphatase, PP1.^[47] When an inhibitor enters and binds to the catalytic site, there is a stabilization of the conformation at the 280s loop (residue 282-287) located between helices α -7 (residues 261-274) and α -8 (residues 289-314), which causes steric hindrance, blocking the substrate (glycogen) reaching the site, promoting the inactive T-state conformation. The most potent inhibitors of the catalytic site have demonstrated lowering of blood glucose levels in vivo, the majority of which are glucose analogues which display low micromolar K_i or IC_{50} values (Table 2.1.)

Table 2.1 - The structures of the most potent inhibitors of RMGPb catalytic site discovered to date.^[36]

Structure	Ki [μM]	Structure	Ki [μM]
	0.031		0.84
	0.16		1.02
	0.35		1.21
	0.39		1.7
	0.41		1.8
	0.63		1.9
	0.67		1.94
	0.7		

Glucose based ligands have been the focus of GP catalytic site drug design with both α and β -substitutions on the anomeric C1 atom. The β -substitutions aim to exploit the catalytic subsite known as the β -cavity, an empty space at the β -1-C configuration lined by both polar and non-polar groups.^[48] The sugar rings make

Table 2.2 – The structures highlighting the importance of the α -substituent in inhibiting RMGPb

α-D-glucose	1-deoxyglucose	α-methyl-1-deoxyglucose
		
$K_i = 1.7\text{mM}$	$K_i = 10.7\text{mM}$	$K_i = 53\text{mM}$

hydrogen bonds to suitable protein groups through each of the peripheral hydroxyl groups including hydrogen bonds via a water bridging molecule to an aspartic acid, Asp283, and a hydrogen bond from 2-OH to Asn284. The importance of the α -1-OH is shown by the poorer inhibition constants for 1-deoxyglucose ($K_i = 10.7\text{mM}$)^[49] and α -methyl-1-deoxyglucose ($K_i = 53\text{mM}$).^[50] A key structural feature of glucose based ligands is the so called linker heterocycle which connects the glucose moiety with an R-group,^[51] this R-group is of importance with certain groups aiding to the greatest inhibition exploiting β -cavity features. In most cases, the most favourable residue to occupy the β -cavity is the 2-naphthyl group owing to its large planar hydrophobic nature.^[52] In most of these glucose analogues displaying low micromolar inhibition, there is an important ligand-protein interaction formed from the linker heterocycle consisting of a hydrogen bond between an N-H to the protein His377 backbone O. *N*-Acetyl- β -D-glucopyranosyl-amine (NAG) is a simple glucose analogue which forms this interaction and inhibits RMGPb ($K_i = 32\mu\text{M}$) which is approximately 50 times more potent than its natural physiological inhibitor, α -D-glucose.^[36] Solvation effects of the catalytic site are important, this has been highlighted through protein crystallographic work by E.D Chrysin et al. where similar glucose analogues were investigated showing nitrogen atoms of the heterocycle interacting with a direct hydrogen bond with Leu136 N and water-mediated hydrogen bonds with Asp 283.^[53]

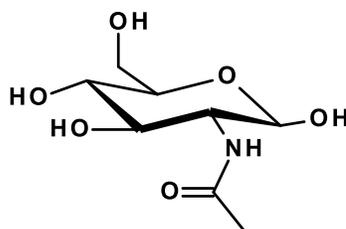


Figure 2.5 – The structure of NAG.

The design of most GP catalytic site inhibitors are targeted at human liver GPb (HLGPb) however, most in vitro studies are performed using rabbit muscle GPb (RMGPb) due to the relative availability as well as its propensity to produce good quality crystals. The active sites of both HLGPb and RMGPb are identical in both amino acid sequence and structural architecture and the overall structure of RMGPb shares an 80% similarity to that of HLGPb. This similarity allows for the use of in vitro studies using RMGPb with a reliability that the conclusions drawn from experimental data gathered from these studies can also be applicable to HLGPb.

2.5 Allosteric Sites

Table 2.3 - Potencies of selected dihydropyridine AMP site inhibitors that bind at the allosteric site.

Entry	R ₂	R ₂	R ₃	R ₄	HLGP(a) K _i (μM)	Cell EC ₅₀ (μM)	
1	Cl	CO ₂ H	(<i>i</i> -Pr)-O	Et	0.039	2.2	
2	H	CO ₂ H	(<i>i</i> -Pr)-O	Et	0.395	6.5	
3	Cl	H	(<i>i</i> -Pr)-O	Et	15.6	NA	
4	Cl	CO ₂ H	(<i>i</i> -Pr)-O	H	0.692	11.0	
5	Cl	CO ₂ H	(<i>i</i> -Pr)-NH	Et	9.1	NA	
6	Cl	CO ₂ H	(<i>i</i> -Pr)-O	CH ₂ Ph	0.011	1.13	
7	Cl	CO ₂ H	(<i>i</i> -Pr)-O		0.002	0.48	
8	Cl	CO ₂ H	(<i>i</i> -Pr)-O		0.004	0.27	

The allosteric site is located roughly 30 Å from the catalytic site and is composed of structural elements from two subunits of the functional dimer located on opposite sides of the enzyme molecule. Two α -helices (residues 47-78, and 289-314, respectively) and four β -strands (residues 153-160, 191-193, 222-232, and 237-247).^[36] This site is partially exposed to solvent and binds the activator AMP and the natural inhibitor Glc-6-P but recognizes a variety of phosphorylated compounds, for instance ATP, nicotinamide adenine dinucleotide (NADH), inosinic acid (IMP), uridine diphosphate glucose (UDP)-glucose, β -glycerophosphate and inorganic phosphate. The equilibrium between the T and R-state can be regulated by certain compounds activating the allosteric site, such as Glc-6-P and adenosine phosphates. Indirect inhibition of substrate binding to this site can arise as a result of the stabilization of the T or T' state (T' referring to the symmetry related monomer subunit.). Inhibition can be a result of direct inhibition of AMP binding.^[48] In resting muscle cells, levels of Glc-6-P are sufficient to retain the enzyme in the T state, and prevent wasteful glycogen degradation. Glc-6-P binds at the allosteric site and promotes conformational changes in the T-state GPb.^[54] Binding interactions between Glc-6-P to the allosteric site are mostly hydrogen-bonding interactions and there are relatively few van der Waals interactions. In contrast, there are examples of interactions from compounds such as dihydropyridines (Table 2.2) that are dominated largely by non-polar van der Waals interactions and ionic interactions. The non-polar contacts appear to be the major source of binding energy that results in an inhibitor with nM affinity compared with the μ M affinity exhibited by Glc-6-P. The different binding modes and contacts observed reveals a remarkable degree of versatility for the allosteric site, which is able to recognize specifically dissimilar compounds by employing the same residues.^[48]

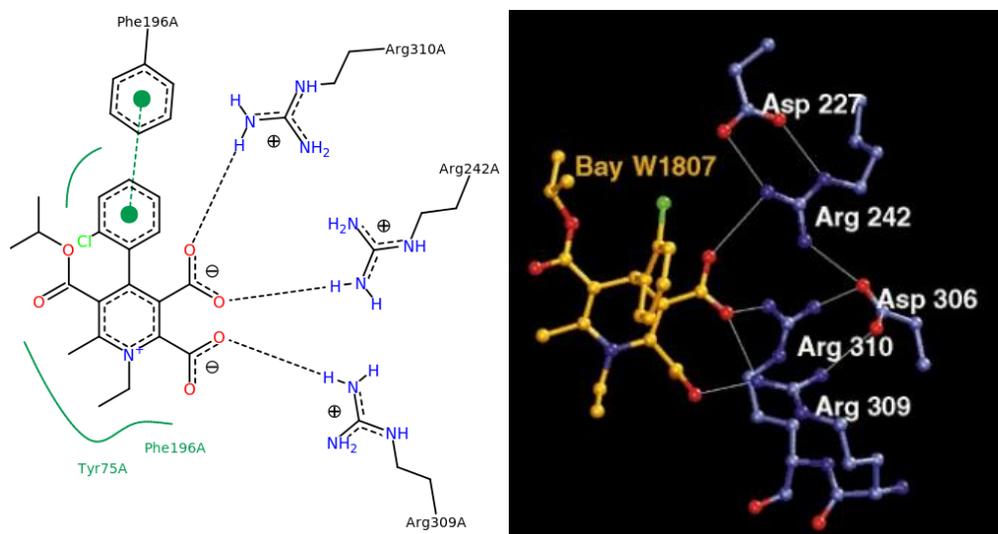


Figure 2.6 – Diagram showing the different ligand interactions of a dihydropyridine dicarboxylic acid. Black dashed lines indicate hydrogen bonds. Green solid line show hydrophobic interactions and green dashed lines show π - π stacking interactions.^[35]

A second allosteric site, the new allosteric site (commonly known as the indole binding site), is located 33 Å from the catalytic site and is 37 Å from the inhibitor site within the central cavity region of the subunit dimerization interface. Crystal structures of ligand binding at this site have validated the environment within the indole site that constitutes for the binding between the receptor-ligand. Binding occurs with only minor distortions of the structure, and upon binding there is a displacement of nine water molecules. The increasing entropy from the displacement of the water in combination the van der Waals, CH/ π , halogen/polar, and the specific polar/polar interactions appear to be the major source of binding energy responsible for inhibitor binding.^[36] Inhibitors of this site are known to stabilise the T-state conformation of the enzyme and hence inhibit GPb activity.^[55]

2.6 Inhibitor Site.

The inhibitor site (commonly known as the caffeine binding site) is located 12 Å from the catalytic site, occupying part of the entrance. The inhibitor site has been probed with purines, nucleosides, nucleotides and heterocycle compounds such as uric acid, riboflavin,^[56-59] and flavopiridol.^[60, 61] In the less active T state, Phe285, belonging to the 280s loop is stacked close to Tyr613, belonging to the α 19 helix, and these aromatic residues form the core of the inhibitor site. The main source of

the binding affinity of ligands which bind at the inhibitor site is from π -stacking interactions with the sidechains of Phe285 and Tyr613 forming sandwich type complexes; they promote the T state conformation of the enzyme through stabilization of the closed position of 280s loop, blocking access of the substrate to the catalytic site. Conversely, on transition from the T- to the R state, the 280s loop becomes disordered and displaced, opening a channel that allows a crucial residue Arg569, to enter the catalytic site in place of Asp283 and to create the recognition site for the substrate phosphate; this transition enables the glycogen substrate to reach the catalytic site.^[62] The most potent natural inhibitor site known to bind at the site is ellagic acid through a comprehensive kinetic and crystallographic study on the binding properties of gallic acid and its dimer ellagic acid to GP;^[63] The K_i values of ellagic acid are 7.5 and 13.4 μM for GPa and GPb respectively whereas gallic acid displays a weaker K_i of 3.9 and 1.7 mM, respectively. As expected, ellagic acid and gallic acid experience the characteristic stacking binding modes between Phe285 and Tyr613. Ellagic acid exploits nearly twice as many van der Waals interactions with GPb compared to gallic acid; it forms one direct hydrogen bond to the Asn 282 sidechain (ND atom) and a water-bridged interaction with Asp283, Ile570, Ala610, Gly612, His614 and Met615.^[63]

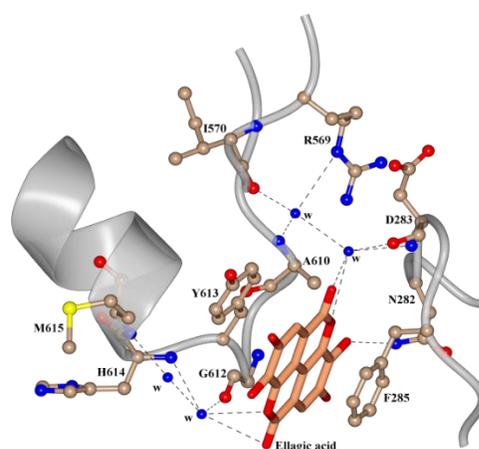


Figure 2.7 - Binding of ellagic acid at the GP inhibitor site as determined by X-ray crystallography (PDB code: 4YUA).^[63]

2.7 Quercetin Binding Site

Recent screening of 13 polyphenolic compounds obtained by extraction of *Vitis vinifera* (common grape vine) vinification by-products displayed significant inhibitory potency for RMGPb *in vitro* (IC₅₀ values in the range of low µg/mL) revealed that the most active ingredient of these extracts is the flavonoid quercetin.^[64] Although unknown at the time, quercetin was found to bind to a novel binding site. The site is 42 Å from the allosteric site, 32 Å from the inhibitor site and 15 Å away from the catalytic site. The site is a shallow groove consisting of Lys544, Arg551, Lys655 and Tyr548 on one side with Glu120 and Glu123 on the other. The binding mode of quercetin to GP is well understood, with hydrogen bonds and van der Waals interactions forming between Arg551 and Lys544. There are also water mediated bridging interactions with Glu121, Lys655, Leu494, Cys495 and Glu654.

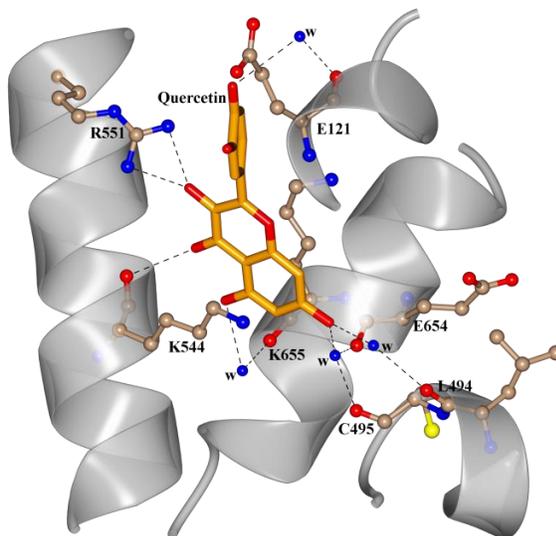


Figure 2.7 – Binding of quercetin at the Quercetin binding site as determined by X-ray crystallography (PDB code: 4MRA)^[64]

2.8 Flavonoids.

Polyphenols are naturally occurring compounds abundant in many vegetables, fruits, cereals and beverages. Fruits like grapes, apples, pears, cherries and berries (such as strawberries and blackberries) contain up to 200-300mg polyphenols per 100 grams of fresh weight. Polyphenols exist as secondary metabolites of plants and are generally involved in defence against Ultra Violet radiation or aggression by pathogens.^[65] Flavonoids are a diverse group of polyphenols with over 10,000 compounds that have been identified until now, with only very few being investigated in detail.^[66] The effects flavonoids have within the body has been the center of many research studies in recent years and the beneficial effects are becoming abundantly clear. Chemically, flavonoids are based upon a fifteen-carbon skeleton consisting of two phenyl rings (A and B as shown in Figure 2.7) linked via a heterocyclic pyrene ring (C).^[67] Flavonoids can be divided into a number of subclasses, flavones (e.g. flavone, chrysin, epigenin), flavonols (e.g. quercetin, kaemferol, myricetin), flavanones e.g. flavanone, hesperetin and naringenin) (Figure 2.8).

Table 2.4 - IC₅₀ values (μM) for inhibition of GP_a and GP_b of flavonoids tested in concentrations up to 50 μM.

Compound	IC ₅₀ GP _a (μM)	IC ₅₀ GP _b (μM)	Compound	IC ₅₀ GP _a (μM)	IC ₅₀ GP _b (μM)
Flavones			Anthocyanidins		
DHF	>50 [67]	>50 [67]	Pelargonidin	43.6 [68]	6.2 [68]
Chrysin	>27.5 [68]	15.3 [68]	Cyanidin	3.0 [68]	9.0 [68]
		19.0 [69]	Delphinidin	3.1 [68]	10.7 [68]
Apigenin	>30 [68]	19.01 (K _i) [59]	Peonidin	25.1 [68]	17.6 [68]
Luteolin	15.6 [68]	No effect up to 30 [68]	Malvidin	>50 [68]	>50 [68]
		29.7 [70]	Catechins		
		31.7 [70]	Catechin	No effect [68]	No effect [68]
Tricin	>50 [68]	No effect [68]	Epicatechin	No effect [68]	No effect [68]
Baicalein	11.2 [68]	10.2 [68]			No effect [70]
Hypolaetin		15.7 [70]	Epigallocatechin	No effect [68]	No effect [68]
6-hydroxyluteolin		11.6 [70]			No effect [68]
Isoscutellarein		46.1 [70]	ECG	12.5 [68]	50 [68]
Flavonols			EGCG	7.7 [68]	27 [71]
Quercetin	4.8 [68]	20.9 [68]			33.9 [68]
		33.5 [70]	EGC		34 [71]
Quercetagetin		9.7 [70]	GCG		>400 [71]
		3.5 (K _i) [70]	CG		6.3 [71]
(-)-Epicatechin (EC)		290 [71]			35 [71]
Flavanones			Other		
Naringenin	>50 [68]	No effect [68]	Ellagic acid	3.2 [68]	12.1 [68]
Hesperitin	No effect [68]	No effect [68]		7.5 (K _i) [69]	13.4(K _i) [69]
Eriodictyol	No effect [68]	No effect [68]	Gallic acid	No effect [68]	No effect [68]
Isoflavones				3900(K _i) [69]	1730(K _i) [69]
Daidzein	No effect [68]	No effect [68]	Caffeine		130 [51]
Genistein	>50 [68]	No effect [68]			

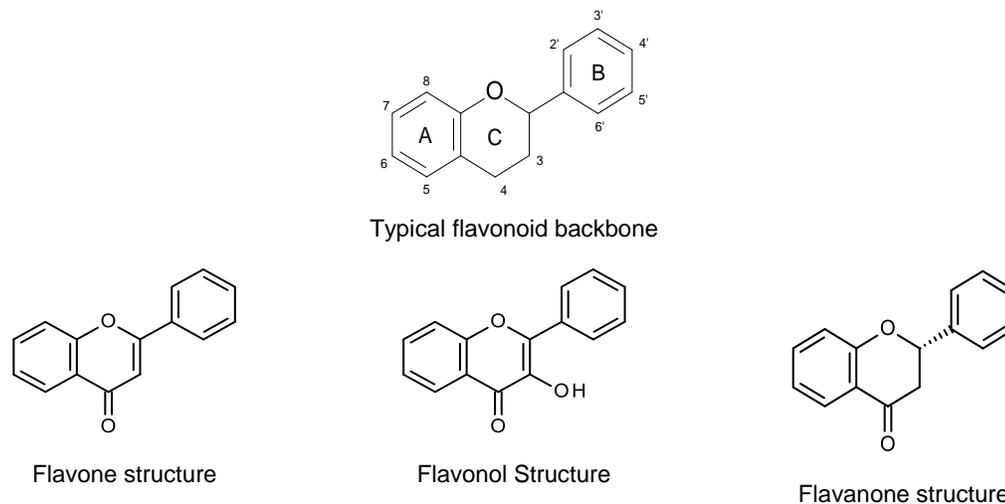


Figure 2.8 – Various different flavonoid structures, all follow the typical flavonoid backbone.

Several observations have suggested that natural flavonoids have inhibitory effects on various types of cancer cells mediated by different molecular targets and acting through diverse metabolic pathways^[72] and also inhibit the enzymes responsible for superoxide anion production, such as xanthine oxidase and protein kinase C. Flavonoids have also been shown to inhibit cyclooxygenase, glutathione S-transferase, mitochondrial succin-oxides, and NADH oxidase, all involved in reactive oxygen species generation.^[73]

Flavonoids have also been shown to have biological activity on GP with natural flavonoid derivatives investigated as potential GP inhibitors^[68, 71, 74] with ongoing experiments of many compounds against RMGP_a and RMGP_b. Jakobs et al found in their study that almost all flavonoids tested inhibited active RMGP_a and un-phosphorylated AMP activated GP_b, however inhibition of GP_a was two to four times stronger than inhibition for RMGP_b^[68] with the leading flavonoid inhibitors of GP on the low μM range (Table 2.3). Inhibition of GP with flavonoids are sensitive to the different substituent groups on them, a number of flavonoids in each structural subclass demonstrate little or no inhibition dependent upon the different substituents. Of those compounds tested, no flavanone or isoflavone shows inhibition based on the few compounds tested. Flavone's are known to bind at the inhibitor site, an investigation into the binding modes of chrysin to GP_b (PDB code: 3EBO) shows that there is the characteristic π - π stacking in between the two

aromatic residues (Phe285 and Tyr613), and when chrysin binds to GPb, it forms polar/polar interactions and exploits a number of van der Waals contacts dominated by the substantial contacts made to Phe285 and Tyr613 by the aromatic rings.^[59] The most potent compound to bind at the inhibitor site discovered to date is flavopirodol^[59, 61] and its analogues.^[75] While flavopirodol has a K_i of 1.2 μM , analogues of flavopirodol with a flattened tetrahydro-pyridine ring and halogen substituted phenyl ring B have revealed potencies ranging from 0.83-2.75 μM . The most favourable substitution in this series is the 3-Cl analogue. Halogen substitution and binding effects through the σ -hole phenomenon are now widely recognized and exploited in drug design.^[76] The crystal structure of flavopirodol in complex with RMGPb PDB codes: 1C8K, 3EBP) has been solved^[59, 61] and the structural features of inhibition are similar to that of chrysin.^[59] The common flavonoid core of flavopirodol and chrysin superimpose well stacked between Phe285 and Tyr613. Flavopirodol makes a few polar/polar interactions with water molecules and has a numerous van der Waals contacts. The chlorophenyl ring B extends towards the hydrophobic pocket lined by Phe771, Tyr573 and Ile380 but the chlorine atom does not form a strong halogen bond with GPb and is mainly involved in van der Waals interactions.

2.9 Conclusion

Glycogen Phosphorylase presents a valid target in the treatment of T2D. Glucose analogues have been demonstrated to be effective in the inhibition of the GP catalytic site (Table 2.1) and this section also highlights the potential antihyperglycaemic effects of natural flavonoids acting as inhibitors of the GP binding site (Table 2.3). The scaffold provided from both of these structures serve in the potential development of more effective pharmaceuticals for T2D treatment as a number of these compounds are low μM inhibitors and a glucose analogue has been discovered which displays nanomolar inhibition of the catalytic site, 2-(β -D-glucopyranosyl)-4(5)-(2-naphthyl)-imidazole ($K_i = 31\text{nM}$).^[36] The structural basis of these compounds relative to potency as currently known has been studied in some detail, which provides information for further structure based drug design (SBDD) efforts.

Chapter 3 – Molecular Modelling

3.1 Introduction

Drug design is a complex process, from the original idea of a new drug to the launch of a finished product can take 12-15 years and cost in excess of \$1 billion.^[77] In the past, traditional medicines were often found through the identification of active compounds by chance discovery. Typically, a large library of chemical compounds would be tested experimentally *in vitro* and *in vivo* with the hope that some biological activity would be displayed in a number of compounds (high throughput screening).^[78] This method is incredibly expensive, time consuming and is impractical. In recent years, increased target identification and validation enables increased knowledge of the relationship between the target and disease.

Knowledge-based screening involves selecting from the chemical library a smaller subset of molecules that are likely to have activity at the target receptor based on knowledge of a target protein. This type of knowledge have given rise to early discovery paradigms using pharmacophores and molecular modelling to conduct virtual screening of compound databases.^[79] A pharmacophore model is a geometrical description of the chemical functionalities required for a ligand to interact favourably with a target and pharmacophore mapping is one of the major elements of drug design in which there is an absence of structural data for the target receptor (known as ligand based drug design (LBDD)). For instance, when a set of active compounds are known for a given target, pharmacophore modelling is used to identify the common structural features as well as the relative distances between them that lead to this activity. Once a pharmacophore is built, a thorough search of available on-line databases such as ZINC^[80] is performed which identifies molecules which share the same features of the parent compounds to be used as potential drug candidates. Another important modelling method is quantitative structure-activity relationship (QSAR). QSAR models relate a set of predictor variables to the potency of the response variable. QSAR models first summarize a supposed relationship between chemical structures and biological activity in a data-set of chemicals and secondly, predict the activity of new chemicals. The Hansch equation

(Equation 3.1) relates the relationship between the biological activity of a series of compounds and their physicochemical properties including hydrophobic ($\log P$), electronic (σ), steric (E_s), and other effects;

$$\text{Log } \frac{1}{C} = k_1(\log P)^2 + k_2 \log P + k_3 \sigma + k_4 E_s + k_s \quad \text{(Equation 3.1)}$$

The Free-Wilson equation (Equation 3.2) relates the contributions that the substituents have to the activity of the parent structure.

$$\text{Activity} = k_1 X_1 + k_2 X_2 + \dots + k_n X_n + Z \quad \text{(Equation 3.2)}$$

X_n is an indicator variable which is indicated by the values 1 and 0 dependent on whether the substituent (n) is present or not. The contribution of each substituent to activity is determined by the value of k_n and Z is a constant representing the parent compound. 3D QSAR has advantages over the Hansch and Free-Wilson approach in that there is no reliance on experimental values, it can be applied to molecules with unusual substituents, and it is not necessarily restricted to molecules of the same class and has a higher predictive capability. Comparative Molecular field Analysis (CoMFA) generates a 3D grid where molecules are described by values of molecular fields calculated on a point within the grid. These molecular fields are usually steric and electrostatic and this grid is able to derive a correlation between the biological activity of a set of molecules and their 3D shape, electrostatic and hydrogen bonding characteristics.

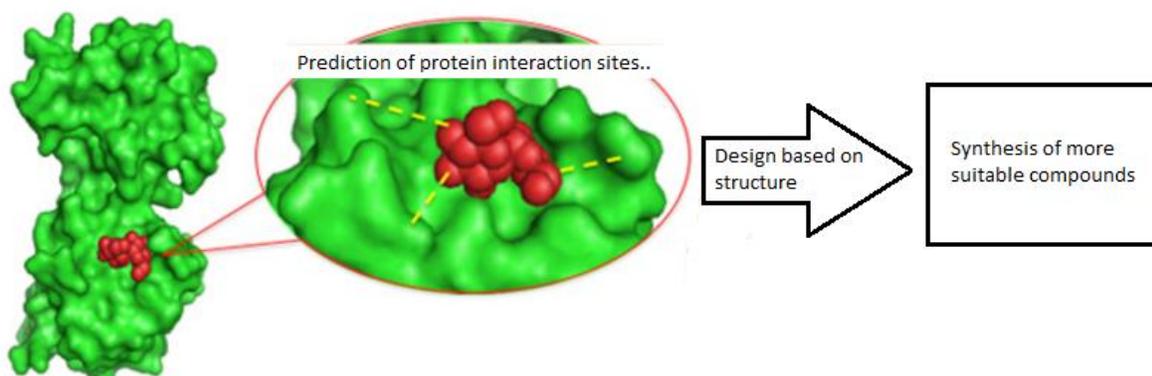


Figure 3.1 – Overview of drug discovery based on Structure based drug design (SBDD).

Structure based drug design (SBDD) is employed when the structure of the target is known. Recent advances in protein X-ray crystallography, nuclear magnetic

resonance (NMR) and electron microscopy have allowed for the precise mapping of atoms that make up the different residues of a protein structure, with over 40500 distinct protein sequences available.^[40] Crystal structures of ligand-protein complexes are considerably valuable in drug design effects as they allow for visual representations of the environment which the ligand occupies within a binding site, and thus prediction of modifications could be added to provide increased potency or selectivity. In SBDD the most common type of drug design is molecular docking, where ligand binding to the receptor is predicted by exploiting the many known biological macromolecular targets solved and deposited in the RCSB Protein Database^[40] These solved targets can be used to screen potentially active compounds and this aids in the ability to find hit compounds. With the advancement of computers and algorithms, docking accuracy has increased exponentially in recent years due to the increasing sophistication of docking software. Docking may also help in lead optimization efforts, suggesting modifications that have the potential to improve activity.

In the current work, the structure of GP is well known and by extension, the core interactions and binding modes of glucose based analogues and the surrounding residues are also well understood. Accurate modelling of these interactions must ensure that constraints are placed on groups whose relative positions and interactions are known. The current chapter will describe the many key computational screening programmes used in virtual docking as well as the role molecular mechanics (MM) has in molecular modelling, which forms the basis of molecular-mechanics-Generalized Born Surface Area (MM-GBSA), the post-docking method used in this study.

3.2 Molecular Mechanics.

3.2.1. Introduction

Molecular Mechanics (MM) forcefields are the method of choice for protein simulations, which are essential in the study of conformational flexibility.^[81] Computational structure-based drug design involves studies based on the three-dimensional structure of the biomolecular target, usually a protein. Whilst quantum mechanical (QM) methods are useful for studying properties of isolated ligands and for limited studies on simple models of protein binding sites, it can be more desirable to perform simulations on the entire protein. Molecular docking studies on entire proteins therefore are performed often exploiting molecular mechanic force fields, which approximate QM energy surfaces with a classic mechanical model, thus decreasing computational cost of simulations on large systems by orders of magnitude. Furthermore, MM potential energy functions allow for relatively accurate representations of dispersion interactions.^[82] Molecular mechanics works on the principle of an additive potential energy function which represent the vast majority of force fields currently used. (Equation 3.3, Equation 3.4)

$$E_{total} = E_{bonded} + E_{nonbonded} \quad \text{(Equation 3.3)}$$

$$E_{bonded} = \sum_{bonds} K_b (b - b_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{dihedral} \sum_{n=1}^6 K_{\theta,n} (1 + \cos(n\theta - \delta_n))$$

$$E_{nonbonded} = \sum_{estat.} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \sum_{vdw} \epsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min,ij}}{r_{ij}} \right)^6 \right] \quad \text{(Equation 3.4)}$$

The total energy of the system is found by addition of the bonded and non-bonded terms. The bonded interactions arise from bond stretching terms, angle bending terms and dihedral or torsion terms whereas the non-bonded terms arise from electrostatic forces as well as van der Waals interactions. The bond and angle terms dominate the local covalent structure around each atom and in this approximation, only 2 parameters are needed for each bond and angle: the reference or equilibrium value (b_0 and θ_0) and the force constants (K_b and K_θ).^[82] Finally, the torsional energy is represented by a sum of cosine functions with multiplicities $n = 1, 2, 3, \dots$ and amplitudes $K_{\theta,n}$. The phases δ_n are usually constrained at 0° or 180° so that the

energy surface of achiral molecules is symmetric and so that enantiomers have the same energetic properties. Electrostatic terms in force fields are handled by Coulomb interactions between fixed point charges q_i and q_j centred on the atoms. Electrostatic interactions are additive because the charges do not affect each other and all the individual atom-atom electrostatic interactions may simply be summed to yield the total electrostatic energy of the system. For the van der Waals interaction component, a classic Lennard-Jones 6-12 potential, defined by the radius $R_{\min,ij}$ and the well depth ϵ_{ij} , is typically used.

3.2.2 Forcefield

Many forcefields have been developed by various groups and are often designed for different purposes and will therefore differ slightly in the types of simulations they may run independent on the parameterization used. The most commonly used forcefields include: Assisted Model Building and Energy Refinement (AMBER)^[83] which is often used for proteins and DNA. Chemistry at HARvard Molecular Mechanics (CHARMM)^[84] which is widely used for both small molecules and macromolecules. Merck Molecular Force Field (MMFF)^[85] which is used for a broad range of molecules, but predominately organic. The forcefield of concern being used in this project is the Optimized Potential for Liquid Screening 3 (OPLS3) used in Schrodinger's maestro software.^[86] All of these forcefields share basic characteristics; they rely on a fixed charged electrostatic model, they employ a standard Lennard-Jones 6-12 van der Waals potential to model electronic repulsion and dispersive nonbonded interactions, and they utilize harmonic stretching and bending terms, and dihedral angle based torsional potentials, to model the valence component of the energy.^[87]

To adequately cover medicinal chemical space, the OPLS3 forcefield employs over an order of magnitude more reference data and associated parameter types relative to other commonly used small molecule force fields (e.g., MMFF and OPLS_2005). As a consequence, OPLS3 achieves a high level of accuracy across performance benchmarks that assess small molecule conformational propensities and solvation. In the OPLS3 forcefield, stretching and bend terms are fit to quantum chemical data.

These terms are insensitive to the other terms in the forcefield and so are determined as the first step in the protocol. The van der Waals (vdw) terms are obtained from liquid state simulations, along with the core charge set used in the liquid simulations. Bond charge corrections (BCC's) for the CM1A-BCC charge model are defined and tested (using, for example, calculations of solvation free energies in water). Finally, in the last step the torsional parameters are fit to *ab initio* quantum chemical data. A number of key assumptions are made with OPLS3 with regard to whether significant predictive power compared to experiment will be created. The assumptions are as follows: (1) Classical treatment of nuclear motion does not systematically distort results compared to experiment. (2) The use of a fixed charge force field provides an adequate description of electrostatic interactions in the condensed phase problem of interest. The use of a gas phase model on water for instance is that there is a drastic underestimation of the cohesiveness of liquid water.^[88] Protein and water contribute the vast majority of atoms in a simulation and it is important that physically reasonable structures and energies emerge given the tuning at average polarization. (3) Torsional terms fit to gas phase potential surface can be transferred to condensed phase with minimal errors. (4) The contribution of fundamentally non-classical quantum mechanical non-bonded interactions, such as oxygen-sulphur interactions, π - π interactions, and non-linear vdW coupling, is negligible.^[87]

3.3 Molecular Docking

3.3.1 Introduction

In silico Ligand-Protein docking is perhaps the most widely used procedure in computer aided structure based drug design. Molecular docking programs are able to predict the most likely binding modes and geometries of a given ligand against a target, referred to as poses, as well as the predicted binding affinities between small molecules and the receptor target. A scoring function is used which predicts the binding energies between a ligand and the receptor for each of the poses generated, which are then ranked based on the energies; the top ranked pose, in theory should be the most favourable conformation of the ligand in the binding site.

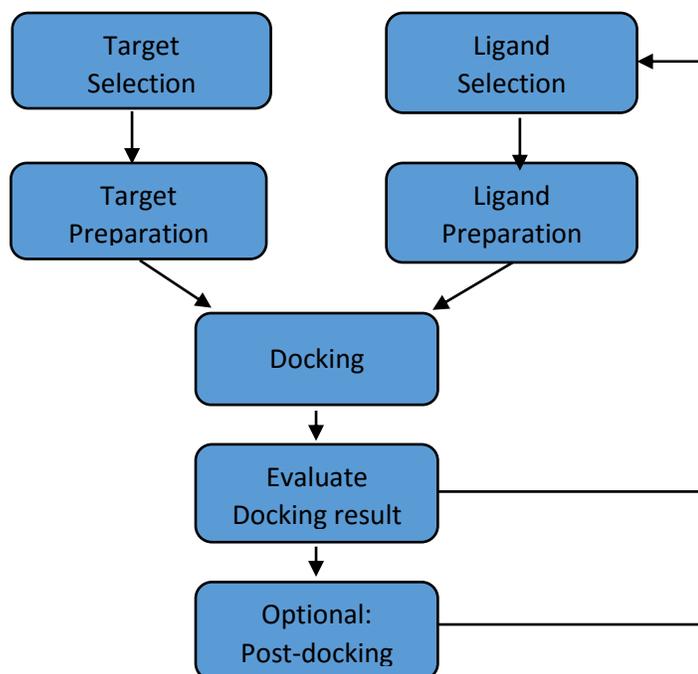


Figure 3.2 – A workflow showing the general process followed in docking.

3.3.2 Search Algorithms

Flexible and rigid docking constitute the two types of docking. Flexible docking is the most common form of docking used and the software generates different conformations of each ligand ‘on-the-fly’ by using a search algorithm.^[89] Flexible docking is much more effective when conformational changes become a bigger factor; an example being that hydrogen bonds and other features might be missed to the lack of flexibility. The vast amount of ligands have a degree of flexibility and may be incorrectly dismissed if they are docked as rigid bodies in the wrong conformation.^[90] In one flexible docking approach, the ligand is divided into rigid fragments and connecting flexible chains. All rigid fragments are docked to all possible places in the cavity independently of each other, yielding either a small or large amount of acceptable combinations of poses. The flexible chains are then fitted to the specific rigid fragment poses that comprise a matching pose-set. The reconstructed solutions then define a rough binding pose and conformation of the ligand. These poses are refined by a local energy minimization in the active site of the receptor, driven by the scoring function

In rigid docking (Figure 3.3), the protein and ligand are treated as rigid, thus restricting the search space. The ligands are built before-hand, and an image of the target binding site is built up from multiple overlapping spheres of varying sizes, which are mapped via the molecular surfaces of the target protein conformation and structure. With this map, the ligand is then superimposed with the center of the spheres. If the conformation of the ligand fits well with the spheres, it is placed into the receptor, minimised, and the position is then scored.

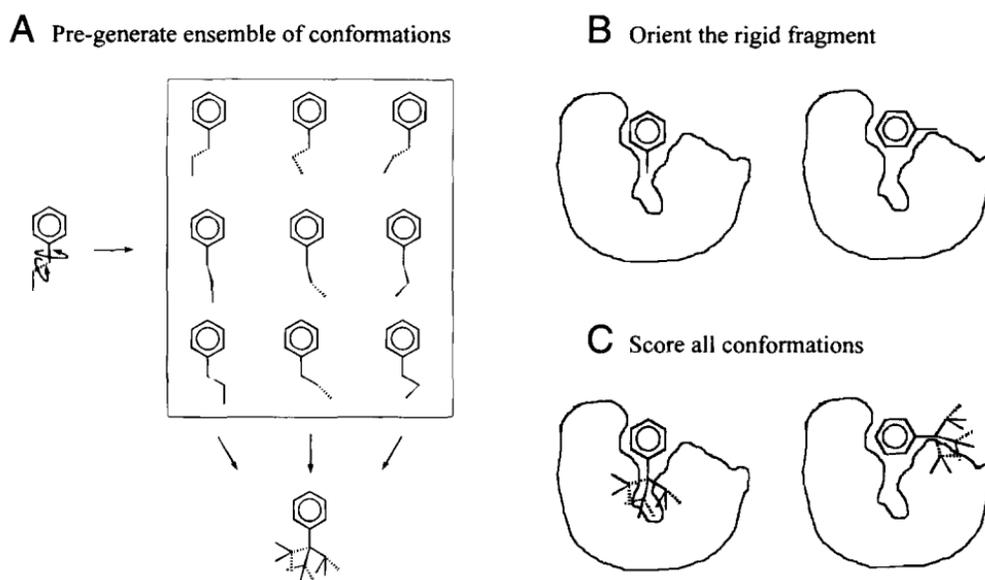


Figure 3.3 - Overview of rigid docking ligand ensemble method. A: rigid atoms are fixed in position and the conformational space of the rest of the molecule is systematically sampled at 60° or 120° increments. B: The rigid fragment is orientated in the target site. C: All n flexible fragments of the molecule are scored in the orientation of the rigid fragment.^[110]

3.3.3 Scoring functions

Scoring functions allow us to discriminate whether ligands are predicted to be active or inactive, by calculating the binding energy of a particular pose. If a ligand has a high binding affinity, the ligand is deemed active and is therefore predicted to exhibit biological activity. Docking scoring functions are often forcefield based.

The sum all of all the various contributions to the binding free energy (ΔG_{bind}) are predicted upon docking as part of the scoring function. The binding free energy is found through the sum of the additive equation, for example the Prime MM-GB/SA function;

$$\Delta G_{bind} = \Delta G_{solvent} + \Delta G_{conf} + \Delta G_{int} + \Delta G_{rot} + \Delta G_{t/r} + \Delta G_{vib} \quad \text{(Equation 3.5)}$$

$\Delta G_{solvent}$ describes the energy contribution by solvation effects and is calculated from the interactions between the solvent and the ligand, protein and intermolecular complex. ΔG_{conf} arises from the conformational changes arising from changes within the protein and/or ligand. Most docking methods assume a rigid receptor and so there is usually no conformational changes from the protein. However, the ligand changes conformation in solution to a dominant conformation in the bound state. ΔG_{int} is the free energy from specific protein-ligand interactions. ΔG_{rot} is the loss of free energy associated with holding internal rotations of the protein and ligand in place. $\Delta G_{t/r}$ is the loss in translational and rotational free energy due to the two bodies (the ligand and the receptor forming a single body (the intermolecular complex)). This is usually assumed to be the same for all ligands and is largely ignored when looking at relative binding strengths of numerous ligands to a certain protein. ΔG_{vib} is the free energy due to changes in vibrational modes. This contribution is hard to calculate and so is generally ignored.

3.3.4 Glide Docking

In this work, Glide, as part of Schrodinger's molecular modelling package,^[91] was used to simulate protein-ligand docking. Glide uses a funnel like system of hierarchical filters to search for possible conformations of the ligands in the active-site of the target receptor. The shape and properties of the receptor are mapped to a grid, which is prepared in the pre-processing steps, by different sets of fields that provide sequentially more accurate scoring of the ligand pose. In the next step, a set of initial ligand conformations are produced. The initial ligand conformations are subjected to initial screening to find promising ligand poses, skipping poses that sterically clash with the receptor. These poses are then taken forward and minimised using the OPLS3 forcefield. Finally, the minimised poses are re-scored with Schrödinger's GlideScore scoring function, which ranks the successful poses. Glidescore is derived from the ChemScore function, as follows:

$$GS = (vdW) + (Coul) + (Lipo) + (H-bonds) + (Metal) + (Rewards) + (RotB) + (Site) \quad \text{(Equation 3.6)}$$

The (vdW) van der Waals energy term takes into account non-bonding interactions. Where (Coul) is the Coulombic term accounting for electron-electron repulsion. The lipophilic (Lipo) term is derived from previous experimental data. The metal (Metal) term represents any metal-ligand interactions may be present. (Rewards) takes into account favourable bindings whilst penalising polar groups in regions of hydrophobicity. The rotatable bond (RotB) term scores lower for non-moving rotatable ligand binding bonds. Finally, (Site) gives a higher score for non-hydrogen bonding areas of hydrophobicity.

Root mean square deviation (RMSD) calculations are standardly used to assess the accuracy of docking calculations, in particular, the poses generated from the calculation. The RMSD is the measure of the average distance between atoms of a superimposed ligand. The poses created by docking software are compared to the native protein-ligand complex obtained from the PDB.^[40] The lower the RMSD value of a particular ligand, the more similar to the native ligand in the complex it is. The RMSD is calculated using the following equation:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2} \quad \text{(Equation 3.7)}$$

where δ_i is the distance between atom i in the predicted and reference structure.

3.3.5 Quantum-Mechanics Polarized Ligand Docking

Quantum-Mechanics Polarized Ligand Docking (QM-PLD or simply QPLD)^[86] aims to improve the partial charges on the ligand atoms during docking by replacing them with charges derived from quantum mechanical calculations on the ligand in the field of the receptor instead of a molecular mechanic forcefield. Polarization of the charges on the ligand by the receptor is accounted for and redocking of the ligands with the new charges can result in improved docking accuracy. In the QM-PLD protocol, ligands are docked with Glide, then charges on the ligand induced by the protein are calculated using QSite, and the ligand poses are then redocked.

3.3.6 GOLD.

Genetic Optimization for Ligand Docking.^[92] (GOLD) is distributed by Cambridge Crystallographic Data Centre. GOLD employs a genetic algorithm (GA) to explore the full range of conformational flexibility of the ligand and also the rational flexibility of selected receptor hydrogens as well as the water molecules. The mechanism for ligand placement is based on fitting points, similar to Glide. Fitting points are added to hydrogen-bonding groups on protein and ligand and the program maps acceptor points in the ligand on donor points in the protein and vice versa. GOLD maps ligand CH groups onto hydrophobic fitting points generated on the protein binding site cavity. GOLD also optimizes flexible ligand dihedrals, ligand ring geometries, dihedrals of protein hydroxyl and amino groups and mappings of the latter. GOLD offers a choice of fitness functions: GoldScore, ChemScore, ASP, CHEMPLP and User Defined Score. CHEMPLP^[93] has been found to give the highest average success rates for both pose prediction and virtual screening experiments against diverse validation test sets and is therefore the default scoring function in GOLD. The piecewise linear potential (f_{PLP}) is used to model the steric complementarity between protein and ligand, whereas CHEMPLP additionally considers the distance- and angle- dependent hydrogen bonding and metal bonding terms ($f_{chem-hb}$, $f_{chem-cho}$, $f_{chem-met}$). The internal score of the ligand consists of the heavy-atom clash potential ($f_{lig-clash}$) as well as the torsion potential used within ChemScore ($f_{lig-tors}$). Both fitness functions are capable of covalent docking ($f_{chem-cov}$), considering side-chains ($f_{chem-prot}$) and explicit water molecules as well as handling constraints (f_{cons}).

$$fitness_{PLP} = -(W_{PLP} \cdot f_{PLP} + W_{lig-clash} \cdot f_{lig-clash} + W_{lig-tors} \cdot f_{lig-tors} + f_{chem-cov} + W_{prot} \cdot f_{chem-prot} + W_{cons} \cdot f_{cons})$$

$$fitness_{CHEMPLP} = fitness_{PLP} - (f_{chem-hb} + f_{chem-cho} + f_{chem-met}) \quad \text{(Equation 3.8)}$$

3.3.7 Prime MM-GBSA

The Prime molecular mechanics-Generalized Born surface area (MM-GBSA)^[94] post docking method is used to calculate the binding-free energy (ΔG_{bind}) of each ligand pose generated from Glide and GOLD. Prime adopts a surface-generalized Born

model using a Gaussian surface instead of a van der Waals surface for better representation of the solvent-accessible surface area.^[95] Prime allows the use of protein flexibility, permitting the residues around the ligand to adopt more specific geometries for individual ligands being screened and should therefore give more accurate results. Poses are minimized using the local optimization feature of Prime and the energies of complex were calculated. The binding free energy (ΔG_{bind}) is then estimated using the equation:

$$G_{\text{bind}} = E_{\text{R:L}} - (E_{\text{R}} + E_{\text{L}}) \quad \text{(Equation 3.9)}$$

Where $E_{\text{R:L}}$ is energy of the complex, $E_{\text{R}} + E_{\text{L}}$ is sum of the energies of the ligand and unliganded receptor, the outcome of the OPLS3 force field.^[96]

3.4 Conclusion

Computational chemistry has given rise to an extremely powerful tool that can aid in the design of potentially new therapeutic drugs. Molecular docking presents itself as a valuable tool that is used in modern day drug discovery throughout the medical field. In this work, its effectiveness will arise from its ability to narrow down a library of β -D-glucopyranosyl derivatives which may possibly display activity on the GP catalytic site, which has already been demonstrated with a number of successful computationally driven studies.^[48, 51] Recent advances in computer technology have allowed for more computationally demanding methods to be used for predicting more accurately the binding energies and a reduction in prediction of false positives.

Chapter 4 – In silico screening of β -D-glucopyranosyl analogues as inhibitors of Glycogen Phosphorylase.

4.1 Introduction

Type-2 Diabetes (T2D) is a heterogeneous disease which is characterised by hyperglycaemia. T2D is characterised by three factors; (1) some resistance of insulin action on glucose uptake in tissues such as adipocytes, liver tissue and skeletal muscle tissues. (2) A relatively low production of insulin from the β -cells of the pancreas and (3) an impaired insulin action to inhibit the hepatic glucose production, 70% of which comes from the breakdown of glycogen by GP via the glycogenolysis pathway.^[9] The inhibition of GP is therefore a promising treatment of T2D. The catalytic site, where glycogen is broken down to G-1-P (and subsequently to glucose) is the main focus in this study, and β -D-glucopyranosyl derivatives, with some known examples showing nanomolar inhibition.^[97] The catalytic site is accessible to bulk solvent through a 15 Å long channel (Figure 2.1), and is located at the junction of the N and C terminal domains. Each is an α/β structure consisting of a β -sheet core surrounded by a tier of α -helices.^[36] Inhibition of the catalytic site promotes the T-state conformation of the enzyme via stabilisation of the 280s loop, blocking access to the site, hindering biological activity. In this chapter, the virtual screening of β -D-glucopyranosyl derivatives with different 5-membered heterocyclic linkers is presented. Initial screening was performed using both Glide-SP and -XP and GOLD. This was followed by MM-GBSA post-docking calculations which were then used to predict binding free energy (ΔG_{bind}). A training set of ligands was first employed so as to optimize the predictive capability of the derived model. This training set consisted of 27 previously studied β -D-glucopyranosyl derivatives with 5-membered heterocyclic linkers as GP inhibitors (Table 4.1). The best MM-GBSA model was then used to predict inhibition constant (K_i) values of a series of unknown inhibitors of this type.

4.2 Aim.

Using the most accurate derived computational screening method for a training set of ligands, to predict the binding affinity of a set of new β -D-glucopyranosyl derivatives with different 5-membered heterocyclic linkers as GP inhibitors.

4.2.1 Specific Objectives.

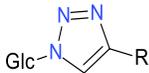
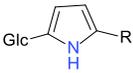
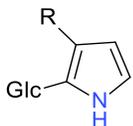
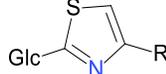
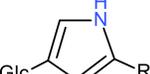
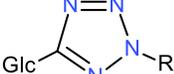
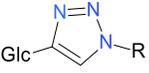
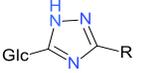
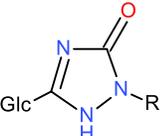
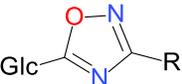
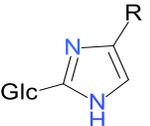
- To explore the effect of different protein conformations from the PDB database on docking results for a training set of ligands.
- To explore the effect of explicit cavity H₂O molecules on the docking and post-docking MM-GBSA results.
- Compare the performance of Glide and GOLD with different scoring functions for ranking the ligand potencies at the GPb catalytic site.
- To identify an effective computational model to screen for new catalytic site inhibitors.
- Analyse modification to standard Prime MM-GBSA so as to improve predictive capabilities.

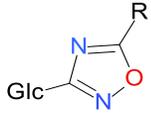
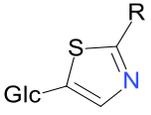
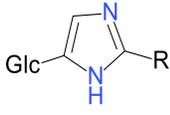
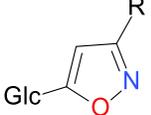
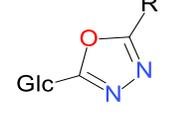
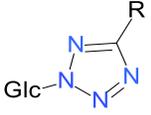
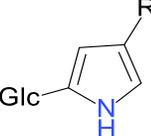
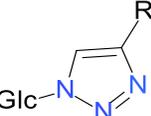
4.3 Computational details.

4.3.1 Ligand Preparation

A set of 80 glucose analogues (Table 4.1) were built using Schrodinger's Maestro programme.^[91] This consisted of 27 training set ligands with known K_i values and a set of 53 ligands whose K_i's were to be predicted. The ligands were then minimized generating energetically correct molecular structures using the LigPrep application with default settings except for the determination of stereoisomers, which was set to determine chiralities from 3D structures. The compounds were prepared for docking in terms of correct protonation, ionization states which were generated at a target pH of 7.0 \pm 2.0, tautomerisation (where 32 tautomers were created) and stereochemistry to produce broad chemical and structural diversity so as to achieve more accurate docking simulations.

Table 4.1 – Set of training set ligands used in computational studies and their known inhibitory constants. Ligands to be predicted are highlighted in red.^a

<p>1,2,3-Triazole_4R</p>  <table border="1" data-bbox="277 443 656 646"> <thead> <tr> <th>R group</th> <th>Ki (μM)</th> </tr> </thead> <tbody> <tr> <td>1-Naphthyl</td> <td>136^[53]</td> </tr> <tr> <td>2-Naphthyl</td> <td>16^[53]</td> </tr> <tr> <td>CH₂OH</td> <td>14^[53]</td> </tr> <tr> <td>Phenyl</td> <td>151^[53]</td> </tr> </tbody> </table>	R group	Ki (μM)	1-Naphthyl	136 ^[53]	2-Naphthyl	16 ^[53]	CH ₂ OH	14 ^[53]	Phenyl	151 ^[53]	<p>2N-Pyrrole_3R</p>  <table border="1" data-bbox="716 443 1068 646"> <thead> <tr> <th>R group</th> <th>Ki (μM)</th> </tr> </thead> <tbody> <tr> <td>1-Naphthyl</td> <td>N/A</td> </tr> <tr> <td>2-Naphthyl</td> <td>N/A</td> </tr> <tr> <td>CH₂OH</td> <td>382</td> </tr> <tr> <td>H</td> <td>N/A</td> </tr> </tbody> </table>	R group	Ki (μM)	1-Naphthyl	N/A	2-Naphthyl	N/A	CH ₂ OH	382	H	N/A	<p>2N-Pyrrole_5R</p>  <table border="1" data-bbox="1105 443 1474 646"> <thead> <tr> <th>R group</th> <th>Ki (μM)</th> </tr> </thead> <tbody> <tr> <td>1-Naphthyl</td> <td>N/A</td> </tr> <tr> <td>2-Naphthyl</td> <td>No inh. At 625</td> </tr> <tr> <td>Phenyl</td> <td>No inh. at 625</td> </tr> </tbody> </table>	R group	Ki (μM)	1-Naphthyl	N/A	2-Naphthyl	No inh. At 625	Phenyl	No inh. at 625		
R group	Ki (μM)																															
1-Naphthyl	136 ^[53]																															
2-Naphthyl	16 ^[53]																															
CH ₂ OH	14 ^[53]																															
Phenyl	151 ^[53]																															
R group	Ki (μM)																															
1-Naphthyl	N/A																															
2-Naphthyl	N/A																															
CH ₂ OH	382																															
H	N/A																															
R group	Ki (μM)																															
1-Naphthyl	N/A																															
2-Naphthyl	No inh. At 625																															
Phenyl	No inh. at 625																															
<p>2S5N-Thiazole_4R</p>  <table border="1" data-bbox="277 842 656 993"> <thead> <tr> <th>R group</th> <th>Ki (μM)</th> </tr> </thead> <tbody> <tr> <td>1-Naphthyl</td> <td>N/A</td> </tr> <tr> <td>2-Naphthyl</td> <td>158</td> </tr> <tr> <td>Phenyl</td> <td>310</td> </tr> </tbody> </table>	R group	Ki (μM)	1-Naphthyl	N/A	2-Naphthyl	158	Phenyl	310	<p>Glucopyrrole_3R</p>  <table border="1" data-bbox="716 842 1068 993"> <thead> <tr> <th>R group</th> <th>Ki (μM)</th> </tr> </thead> <tbody> <tr> <td>1-Naphthyl</td> <td>N/A</td> </tr> <tr> <td>2-Naphthyl</td> <td>N/A</td> </tr> <tr> <td>Phenyl</td> <td>N/A</td> </tr> </tbody> </table>	R group	Ki (μM)	1-Naphthyl	N/A	2-Naphthyl	N/A	Phenyl	N/A	<p>2,3,4,5-Tetrazole_3R</p>  <table border="1" data-bbox="1105 810 1474 993"> <thead> <tr> <th>R group</th> <th>Ki (μM)</th> </tr> </thead> <tbody> <tr> <td>1-Naphthyl</td> <td>N/A</td> </tr> <tr> <td>2-Naphthyl</td> <td>N/A</td> </tr> <tr> <td>Phenyl</td> <td>N/A</td> </tr> <tr> <td>Me</td> <td>N/A</td> </tr> </tbody> </table>	R group	Ki (μM)	1-Naphthyl	N/A	2-Naphthyl	N/A	Phenyl	N/A	Me	N/A				
R group	Ki (μM)																															
1-Naphthyl	N/A																															
2-Naphthyl	158																															
Phenyl	310																															
R group	Ki (μM)																															
1-Naphthyl	N/A																															
2-Naphthyl	N/A																															
Phenyl	N/A																															
R group	Ki (μM)																															
1-Naphthyl	N/A																															
2-Naphthyl	N/A																															
Phenyl	N/A																															
Me	N/A																															
<p>2,3,4-Triazole_4R</p>  <table border="1" data-bbox="277 1276 634 1423"> <thead> <tr> <th>R group</th> <th>Ki (μM)</th> </tr> </thead> <tbody> <tr> <td>1-Naphthyl</td> <td>N/A</td> </tr> <tr> <td>2-Naphthyl</td> <td>N/A</td> </tr> <tr> <td>Phenyl</td> <td>N/A</td> </tr> </tbody> </table>	R group	Ki (μM)	1-Naphthyl	N/A	2-Naphthyl	N/A	Phenyl	N/A	<p>2,3,5-Triazole_4R</p>  <table border="1" data-bbox="716 1140 1068 1413"> <thead> <tr> <th>R group</th> <th>Ki (μM)</th> </tr> </thead> <tbody> <tr> <td>1-Naphthyl</td> <td>4.98^[97]</td> </tr> <tr> <td>2-Naphthyl</td> <td>0.41^[97]</td> </tr> <tr> <td>Phenyl</td> <td>7^[97]</td> </tr> <tr> <td>Me</td> <td>499^[97]</td> </tr> <tr> <td>Ph-Me</td> <td>1.7^[97]</td> </tr> <tr> <td>CH₂OH</td> <td>105^[97]</td> </tr> </tbody> </table>	R group	Ki (μM)	1-Naphthyl	4.98 ^[97]	2-Naphthyl	0.41 ^[97]	Phenyl	7 ^[97]	Me	499 ^[97]	Ph-Me	1.7 ^[97]	CH ₂ OH	105 ^[97]	<p>2,3-Diazole_4R</p>  <table border="1" data-bbox="1105 1276 1474 1423"> <thead> <tr> <th>R group</th> <th>Ki (μM)</th> </tr> </thead> <tbody> <tr> <td>1-Naphthyl</td> <td>N/A</td> </tr> <tr> <td>2-Naphthyl</td> <td>N/A</td> </tr> <tr> <td>Phenyl</td> <td>400</td> </tr> </tbody> </table>	R group	Ki (μM)	1-Naphthyl	N/A	2-Naphthyl	N/A	Phenyl	400
R group	Ki (μM)																															
1-Naphthyl	N/A																															
2-Naphthyl	N/A																															
Phenyl	N/A																															
R group	Ki (μM)																															
1-Naphthyl	4.98 ^[97]																															
2-Naphthyl	0.41 ^[97]																															
Phenyl	7 ^[97]																															
Me	499 ^[97]																															
Ph-Me	1.7 ^[97]																															
CH ₂ OH	105 ^[97]																															
R group	Ki (μM)																															
1-Naphthyl	N/A																															
2-Naphthyl	N/A																															
Phenyl	400																															
<p>2,4,5-Triazole-3-One_4R</p>  <table border="1" data-bbox="277 1644 646 1833"> <thead> <tr> <th>R group</th> <th>Ki (μM)</th> </tr> </thead> <tbody> <tr> <td>1-Naphthyl</td> <td>N/A</td> </tr> <tr> <td>2-Naphthyl</td> <td>80</td> </tr> <tr> <td>Phenyl</td> <td>191</td> </tr> <tr> <td>H</td> <td>No inh. at 625</td> </tr> </tbody> </table>	R group	Ki (μM)	1-Naphthyl	N/A	2-Naphthyl	80	Phenyl	191	H	No inh. at 625	<p>2,4N-5-Oxadiazole_3R</p>  <table border="1" data-bbox="716 1644 1084 1812"> <thead> <tr> <th>R group</th> <th>Ki (μM)</th> </tr> </thead> <tbody> <tr> <td>1-Naphthyl</td> <td>19^[98]</td> </tr> <tr> <td>2-Naphthyl</td> <td>12^[98]</td> </tr> <tr> <td>Phenyl</td> <td>64^[98]</td> </tr> </tbody> </table>	R group	Ki (μM)	1-Naphthyl	19 ^[98]	2-Naphthyl	12 ^[98]	Phenyl	64 ^[98]	<p>2,5-Diazole_4R</p>  <table border="1" data-bbox="1105 1644 1474 1812"> <thead> <tr> <th>R group</th> <th>Ki (μM)</th> </tr> </thead> <tbody> <tr> <td>1-Naphthyl</td> <td>N/A</td> </tr> <tr> <td>2-Naphthyl</td> <td>0.031</td> </tr> <tr> <td>Phenyl</td> <td>0.28</td> </tr> </tbody> </table>	R group	Ki (μM)	1-Naphthyl	N/A	2-Naphthyl	0.031	Phenyl	0.28				
R group	Ki (μM)																															
1-Naphthyl	N/A																															
2-Naphthyl	80																															
Phenyl	191																															
H	No inh. at 625																															
R group	Ki (μM)																															
1-Naphthyl	19 ^[98]																															
2-Naphthyl	12 ^[98]																															
Phenyl	64 ^[98]																															
R group	Ki (μM)																															
1-Naphthyl	N/A																															
2-Naphthyl	0.031																															
Phenyl	0.28																															

<p>2,5N-3-Oxadiazole_4R</p>  <table border="1"> <thead> <tr> <th>R group</th> <th>Ki (μM)</th> </tr> </thead> <tbody> <tr> <td>1-Napthyl</td> <td>No inh. at 625</td> </tr> <tr> <td>2-Napthyl</td> <td>38</td> </tr> <tr> <td>Me</td> <td>N/A</td> </tr> <tr> <td>Phenyl</td> <td>No inh. at 625</td> </tr> </tbody> </table>	R group	Ki (μM)	1-Napthyl	No inh. at 625	2-Napthyl	38	Me	N/A	Phenyl	No inh. at 625	<p>2S4N-Thiazole_3R</p>  <table border="1"> <thead> <tr> <th>R group</th> <th>Ki (μM)</th> </tr> </thead> <tbody> <tr> <td>1-Napthyl</td> <td>N/A</td> </tr> <tr> <td>2-Napthyl</td> <td>N/A</td> </tr> <tr> <td>Phenyl</td> <td>N/A</td> </tr> </tbody> </table>	R group	Ki (μM)	1-Napthyl	N/A	2-Napthyl	N/A	Phenyl	N/A	<p>2,4-Diazole_3R</p>  <table border="1"> <thead> <tr> <th>R group</th> <th>Ki (μM)</th> </tr> </thead> <tbody> <tr> <td>1-Napthyl</td> <td>N/A</td> </tr> <tr> <td>2-Napthyl</td> <td>N/A</td> </tr> <tr> <td>Phenyl</td> <td>N/A</td> </tr> </tbody> </table>	R group	Ki (μM)	1-Napthyl	N/A	2-Napthyl	N/A	Phenyl	N/A		
R group	Ki (μM)																													
1-Napthyl	No inh. at 625																													
2-Napthyl	38																													
Me	N/A																													
Phenyl	No inh. at 625																													
R group	Ki (μM)																													
1-Napthyl	N/A																													
2-Napthyl	N/A																													
Phenyl	N/A																													
R group	Ki (μM)																													
1-Napthyl	N/A																													
2-Napthyl	N/A																													
Phenyl	N/A																													
<p>2N3O-Isoxazole_4R</p>  <table border="1"> <thead> <tr> <th>R group</th> <th>Ki (μM)</th> </tr> </thead> <tbody> <tr> <td>1-Napthyl</td> <td>N/A</td> </tr> <tr> <td>2-Napthyl</td> <td>N/A</td> </tr> <tr> <td>CH₂OH</td> <td>218</td> </tr> <tr> <td>Phenyl</td> <td>No inh. at 625</td> </tr> </tbody> </table>	R group	Ki (μM)	1-Napthyl	N/A	2-Napthyl	N/A	CH ₂ OH	218	Phenyl	No inh. at 625	<p>2O3N-Oxazole_4R</p>  <table border="1"> <thead> <tr> <th>R group</th> <th>Ki (μM)</th> </tr> </thead> <tbody> <tr> <td>1-Napthyl</td> <td>N/A</td> </tr> <tr> <td>2-Napthyl</td> <td>N/A</td> </tr> <tr> <td>Phenyl</td> <td>N/A</td> </tr> </tbody> </table>	R group	Ki (μM)	1-Napthyl	N/A	2-Napthyl	N/A	Phenyl	N/A	<p>2O-4,5N-Oxadiazole_3R</p>  <table border="1"> <thead> <tr> <th>R group</th> <th>Ki (μM)</th> </tr> </thead> <tbody> <tr> <td>1-Napthyl</td> <td>10% inh. at 625</td> </tr> <tr> <td>2-Napthyl</td> <td>10% inh. at 625</td> </tr> <tr> <td>Me</td> <td>145^[99]</td> </tr> <tr> <td>Phenyl</td> <td>10% inh. at 625</td> </tr> </tbody> </table>	R group	Ki (μM)	1-Napthyl	10% inh. at 625	2-Napthyl	10% inh. at 625	Me	145 ^[99]	Phenyl	10% inh. at 625
R group	Ki (μM)																													
1-Napthyl	N/A																													
2-Napthyl	N/A																													
CH ₂ OH	218																													
Phenyl	No inh. at 625																													
R group	Ki (μM)																													
1-Napthyl	N/A																													
2-Napthyl	N/A																													
Phenyl	N/A																													
R group	Ki (μM)																													
1-Napthyl	10% inh. at 625																													
2-Napthyl	10% inh. at 625																													
Me	145 ^[99]																													
Phenyl	10% inh. at 625																													
<p>3S5N-Thiazole_4R</p>  <table border="1"> <thead> <tr> <th>R group</th> <th>Ki (μM)</th> </tr> </thead> <tbody> <tr> <td>1-Napthyl</td> <td>N/A</td> </tr> <tr> <td>2-Napthyl</td> <td>N/A</td> </tr> <tr> <td>Phenyl</td> <td>N/A</td> </tr> </tbody> </table>	R group	Ki (μM)	1-Napthyl	N/A	2-Napthyl	N/A	Phenyl	N/A	<p>Ar-2,3,4,5-Tetrazole_Glc-4R</p>  <table border="1"> <thead> <tr> <th>R group</th> <th>Ki (μM)</th> </tr> </thead> <tbody> <tr> <td>1-Napthyl</td> <td>N/A</td> </tr> <tr> <td>2-Napthyl</td> <td>N/A</td> </tr> <tr> <td>Phenyl</td> <td>N/A</td> </tr> <tr> <td>Me</td> <td>N/A</td> </tr> </tbody> </table>	R group	Ki (μM)	1-Napthyl	N/A	2-Napthyl	N/A	Phenyl	N/A	Me	N/A	<p>Ar-2S4N_Thiazole_Glc-3R</p>  <table border="1"> <thead> <tr> <th>R group</th> <th>Ki (μM)</th> </tr> </thead> <tbody> <tr> <td>1-Napthyl</td> <td>N/A</td> </tr> <tr> <td>2-Napthyl</td> <td>N/A</td> </tr> <tr> <td>Phenyl</td> <td>N/A</td> </tr> </tbody> </table>	R group	Ki (μM)	1-Napthyl	N/A	2-Napthyl	N/A	Phenyl	N/A		
R group	Ki (μM)																													
1-Napthyl	N/A																													
2-Napthyl	N/A																													
Phenyl	N/A																													
R group	Ki (μM)																													
1-Napthyl	N/A																													
2-Napthyl	N/A																													
Phenyl	N/A																													
Me	N/A																													
R group	Ki (μM)																													
1-Napthyl	N/A																													
2-Napthyl	N/A																													
Phenyl	N/A																													
<p>Glucopyrrole_4R</p>  <table border="1"> <thead> <tr> <th>R group</th> <th>Ki (μM)</th> </tr> </thead> <tbody> <tr> <td>1-Napthyl</td> <td>N/A</td> </tr> <tr> <td>2-Napthyl</td> <td>N/A</td> </tr> <tr> <td>Phenyl</td> <td>N/A</td> </tr> </tbody> </table>	R group	Ki (μM)	1-Napthyl	N/A	2-Napthyl	N/A	Phenyl	N/A	<p>Me-2,3,4_Triazole_Glc-4R</p>  <table border="1"> <thead> <tr> <th>R group</th> <th>Ki (μM)</th> </tr> </thead> <tbody> <tr> <td>Me</td> <td>N/A</td> </tr> </tbody> </table>	R group	Ki (μM)	Me	N/A																	
R group	Ki (μM)																													
1-Napthyl	N/A																													
2-Napthyl	N/A																													
Phenyl	N/A																													
R group	Ki (μM)																													
Me	N/A																													

^a K_i = N/A = not available and to be included in test set for prediction.

4.3.2 Protein Preparation

A set of 12 known protein structures containing a glucose analogue with a 5-membered heterocyclic linker bound at the catalytic site was downloaded from the protein database^[40] (PDB Code: 5JTU, 5JTT, 3G2H, 3G2I, 3G2K, 3G2L, 1XL0, 1XL1, 5LRC, 5LRF, 5LRE and 5LRD). Notable differences in structure were recognized for residues Asn282, Asp283, Asn284 and Phe285 from the 280s loop in solved 3G2H (bound ligand 1,2,3-Triazole_4R=phenyl)^[53] and 3G2L (bound ligand 1,2,3-triazole_4R=1-naphthyl)^[100] where these residues had different orientations (residues facing outwards relative to the catalytic site) in contrast to the remaining 10 proteins (residue facing inwards relative to the catalytic site) whose residues superimpose relatively well (Figure 4.1). As these residues are within the catalytic site, protein-ligand binding may be influenced by interactions between the ligand and these residues. Accordingly, two protein structures for calculations were generated (PTO)

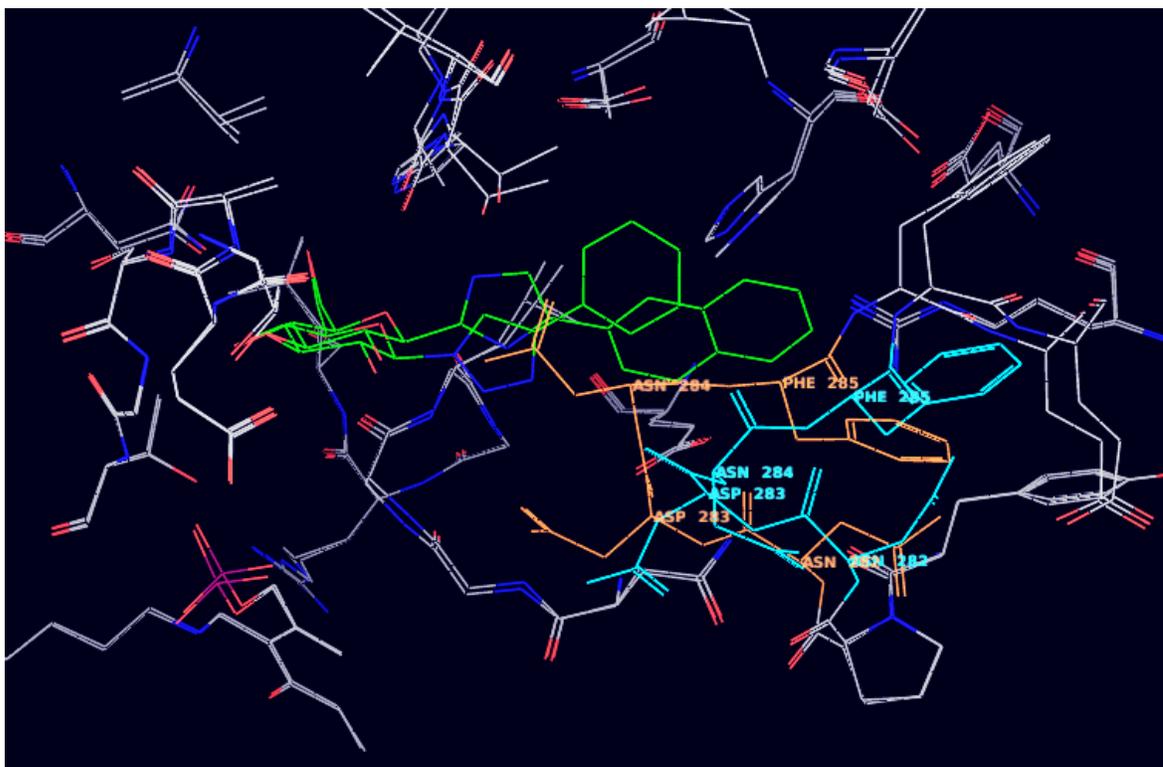


Figure 4.1 – The different orientation of residues Asn282, Asp283, Asn284 and Phe285 using the 5JTU protein with bound ligand 2,5-Diazole_4R=2-Naphthyl (**ClusterResidueIn** shown in brown) and the 3G2H protein with bound ligand 1,2,3-Triazole_4R=Phenyl (**ClusterResidueOut** shown in blue).

whose orientations belong to the two different orientation/geometries observed to explore the effects these residue conformations may have on predicted binding energies and a third general protein was generated (PDB code: 5JTU) to represent the common solvation seen for all 12 proteins (*vide infra*). Protein 5JTU was also chosen to represent the most common orientation seen (residues in), due to the fact this crystal structures native ligand shows the best known inhibition to date, and the R group of 2-naphthyl occupying the β -cavity allows for multiple R groups to be explored and protein PDB code: 3G2H was chosen to represent the less common orientation found (residues out). Protein structures were downloaded from the PDB database (www.rcsb.org) and the crystallographic waters subsequently removed. Water molecules generated from the ClusterWaters algorithm were then superimposed onto the protein, explained in the Solvation Modelling subsection (4.3.3). All protein structures were preprocessed using the default settings in the Import and Process tab of the Protein Preparation Wizard.^[86] This tool assigns bond orders on untemplated residues using the chemical component dictionary (CCD) and automatically adds missing hydrogen atoms to the protein residues and other structures. The protein structures were analysed using the Review and Modify tab, the DMSO molecules were deleted and issues with alternative residue positions were corrected to optimize interactions. The Refine tab was kept at default settings, where orientations of water molecules (if present) were sampled into orientations which maximize interactions. Protonation states of protein residues were found using the PROPKA^[101] program predicting pK_a values at pH 7. Minimization was performed with and without the conserved waters detailed in the solvation subsection to decipher the effects of the docking results in terms of binding geometries and predicted related potencies using the OPLS3 forcefield.

4.3.3 Solvation Modelling.

Solvation effects are usually overlooked when performing VS docking simulations as the presence of explicit water molecules in such systems is one of the most controversial factors in terms of their overall effect on VS efficacy, and the currently accepted view is that such effects are system-dependent to a considerable

degree.^[102] Some authors have reported significant improvement in docking performance when water molecules were included,^[103, 104] whereas other authors found that the inclusion of water molecules had little effect in the quality of their docking calculations.^[105, 106] The GP catalytic site in solved structures contains a number of crystallographic water molecules, some being well ordered and conserved. A number of determined crystal structures of GP in complex with various ligands reveals the importance of solvent for inhibition binding,^[107] some water molecules providing key beneficial interactions often bridging the ligand to certain residues through hydrogen bonding networks. As the residues of the two different protein conformations being explored have two different orientations, the catalytic site water molecules for the proteins will also likely be different due to steric and electrostatic effects. 'Cluster waters' is a statistical analysis algorithm in the form of a python script which allows for the identification of conserved water molecules in a set of protein structures.^[86] The set of proteins are superimposed, and clustering identifies those waters that overlap through a region of space at distances we set at 0.5 Å. Those that met the statistical requirement were retained and those which did not were discarded. The protein used was 5JTU representing the set of all 12 proteins irrespective of 280's loops conformation but with all 12 proteins used to identify common/conserved water molecules. Applying this method, 107 water molecules were retained, 5 of these occupying the catalytic site. The Protein prepared with these explicit water molecules as described in 4.3.2 was labelled **GeneralClusterProtein**. For the less common occurring residue orientation (residue out) clustering of the waters from proteins PDB codes: 3G2H and 3G2L was performed. In total 510 water molecules were found within the two structures, and 172 of these were mutually conserved, 5 of these occupying the catalytic site. This Protein prepared from 3G2H was labelled **ClusterResidueOut**. The set of common water molecules related to the most common orientation (residue in) were generated from a cluster of the other remaining 10 aforementioned proteins, including 5JTU. In total 3034 water molecules were found within the 10 protein structures, and 134 of these were common to all proteins, 7 of these occupying the catalytic site. The protein was labelled **ClusterResidueIn**. The results of the cluster

waters solvation model of each protein is summarised in Table 4.2. The file generated from the cluster waters script contain only the oxygen atoms of the water molecules which were superimposed onto the corresponding protein and then the Protein Preparation process as outlines in 4.3.2 performed.

Table 4.2 - Table showing the solvation of the various receptors using Cluster Waters.

Water Model	PDB model structure used	Number of Proteins used for clustering	Total Number of Waters	Cavity H₂O's
GeneralClusterProtein	5JTU	12	107	5
ClusterResidueIn	5JTU	10	134	7
ClusterResidueOut	3G2H	2	172	5

4.3.4 Docking with the Imidazole receptor (PDB code: 5JTU) with GeneralClusterProtein Water Model.

Docking of all ligands were performed first using the imidazole protein (PDB code: 5JTU) solvated with 107 water molecules. The shape and properties of the catalytic site were mapped onto grids with dimensions of 27 Å x 27 Å x 27 Å centered on the native glucose analogue inhibitor so as to ensure full exploration of the catalytic site and the surrounding residues. Default settings were applied, including van der Waals radius scaling (1.0). As glucose in the GP catalytic site has 5 well defined hydrogen bonds with the surrounding residues, core constraints were placed on the glucose ring. Trial simulations revealed that constraining the glucose ring too tightly causes incorrect pose generation, as a certain degree of flexibility is needed. Core constraints (maximum RMSD of 0.75 Å for core atom positions) were applied to the glucose ring atoms and positional constraints (maximum RMSD of 0.75 Å for atom positions) were placed on the hydrogen atoms of the hydroxyl groups of the glucose ring also to encourage the correct orientation of the hydrogen atoms. Docking calculations were performed using both Glide standard precision (SP) and extra-precision (XP) *with* and *without* water molecules, both of which included post-docking minimization with strain correction. 10 poses per ligand were saved for each docking calculation in order to generate a large number of diverse poses.

4.3.5 Docking using the Proteins Considering Different Residue Orientations.

Docking was also performed again using the imidazole protein structure (PDB code: 5JTU) this time solvated instead with 134 water molecules – **ClusterResidueIn** (Table 4.2). Alternatively, docking with the least common orientation of the protein residues (**ClusterResidueOut**; Table4.2) was performed using the PDB code: 3G2H protein structure which was solvated with 172 water molecules. The shape and properties of both the catalytic sites were mapped onto grids as per 4.3.4 and the same atomic constraints were also used. Docking calculations were performed using both Glide SP and XP *with* and *without* water molecules, and post-docking minimizations were performed.

4.3.6 Quantum-Mechanics Polarized Ligand Docking (QPLD)

The same grid previously created for **GeneralClusterProtein** without waters was used in QPLD calculations. This protein/solvation model was chosen on the basis of its superior performance for the standard Glide and post-docking MM-GBSA calculations (*vide infra*). During the Initial Glide docking, the precision was set to SP and ligand van der Waals scaling was kept at default, as well as the protocol for discarding duplicate poses. The charges for ligands were calculated in the field of the receptor using quantum mechanics (QM) and the Jaguar accurate charge model, and the ligands were then redocked with the updated QM charges using the same default parameters set for the previous Glide docking. Poses, as before, were ranked using GlideScore.

4.3.7 GOLD Docking Details

GOLD (Genetic Optimisation for Ligand Docking)^[92] uses .mol2, .pdb or the .sdf file format. The GOLD wizard allows a protein taken from our previous protein preparation calculations with the Protein Preparation Wizard (4.3.2) without modifications. The protein is selected and the co-crystallised native ligand is removed, but used to define the binding site. The binding site is generated and a residue interaction at a distance set to 10 Å from the native ligand was chosen. The detect cavity option was selected, restricting atom selection to solvent-accessible

surface, and also forces all H-bond donors/acceptors to be treated as solvent accessible. GOLD has many options regarding water molecules, including the toggle state and/or spin state. Three solvation models were explored using the **GeneralClusterProtein** receptor as this outperformed other the receptor models on the basis of Glide docking poses. The first solvation model contained 25 water molecules; GOLD docking only allows for a maximum of 25 water molecules to be used for docking and as these water molecules were in and around the catalytic site at a distance of 8 Å, they were retained and were kept stationary. Two other solvation models were created consisting of two water molecules (HOH 467 and HOH 469 in PDB code: 5JTU) that make favourable binding interactions with the ligand heterocycle moiety within the within the catalytic site. Two solvation modes were considered for the latter which allowed these two water molecules to both toggle and spin whereas another solvation model held the two waters stationary. Lastly, docking was performed without any water molecules present for comparison. The scoring function used was CHEMPLP (piecewise linear potential)^[93] throughout. The search efficiency was kept at the default (100%). Constraints were placed on the glucose ring scaffold as well as the individual hydrogen atoms of the hydroxyl groups of glucose with a constraint weight of 5 for both. Docking was set to generate 10 poses per ligand to achieve a large number of diverse poses.

4.3.8 Docking performance Assessment

Geometry.

For all 12 solved protein-ligand complexes, comparative RMSD calculations were performed for both Glide-SP and -XP with and without water molecules comparing the predicted pose for a ligand with its native conformation. The aim was to see which protein residue orientation (in or out) and solvation model performs the most accurate docking calculation with respect to geometry and whether SP or XP gave similar poses for the same receptor. An RMSD of less than 1.5 Angstroms was considered accurate.^[108]

Scoring.

The highest ranked poses for each training set ligand with Glide and GOLD were selected and the corresponding Glidescores (GSs) for Glide and ChemPLP fitness for GOLD were plotted against the $\ln K_i$ values for each known ligand to give a correlation plot. In the case of ligands with tautomers only the best scoring tautomer was plotted so that each ligand only one receives one rank.

4.3.9 Prime MM-GBSA docking

The Prime MM-GBSA method^[109] was used to calculate the binding-free energy (ΔG_{bind}) of each ligand, using the following equation:

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{solv} \quad \text{(Equation 4.1)}$$

where ΔE_{MM} is the difference in the minimized energies between the receptor-inhibitor complex and the sum of the energies of the free protein and unbound inhibitor. ΔG_{solv} is the difference in the GBSA solvation energies comparing solvation energy of the receptor-ligand complex and the sum of the solvation energies for the unliganded protein and inhibitor. These simulations were performed on the receptor-ligand pose viewer structures obtained from the molecular docking calculations and should in theory give more accurate results than the preliminary docking calculations if the receptor is relaxed using Prime, as the residues around the ligand were in more optimal positions to the specific ligands; additionally a more accurate model for solvation is used. Pose viewer files were taken from the docking simulations performed on all receptors from Glide –SP and -XP and GOLD – ChemPLP, and used in MM-GBSA calculations as implemented in the Prime^[94] module from the Schrodinger Suite using default settings. For this, the ligand poses were minimized using the local optimization feature in Prime, with the OPLS3 force field^[87] and the variable-dielectric generalized Born solvation model, to more accurately account for solvation effects. During the simulation process, the ligand strain energy was also considered so that MM-GBSA calculations were ran with both the receptor rigid and also with residues within 5 Å of the ligand poses are allowed to be flexible for comparison.

4.4 Results and Discussions

4.4.1 Glide Docking

RMSD Calculations.

Exploiting the 12 known/solved crystallographic structures previously mentioned, we explored the potential of our receptor models and effect of conserved waters on pose prediction, as well as the effect of protein conformation. The root mean square deviation (RMSD) was compared for ligand heavy atoms from the top ranked poses for both SP and XP that have solved crystal structures. An RMSD cut off less than 1.5 Å was considered similar. The results of these RMSD calculations as shown in Tables 4.3-4.5

Table 4.3 – Ligand RMSDs (Å) achieved from Glide-SP & -XP docking compared with their crystallographic equivalents for the **GeneralClusterProtein** (PDB Code 5JTU) receptor considered with and without conserved water molecules retained. Ligands above the cut-off RMSD for accuracy (>1.5 Å) are highlighted in orange.

PDB Code	GeneralClusterProtein			
	With Waters		Without Waters	
	SP	XP	SP	XP
5JTU	0.068	0.099	0.107	0.130
5JTT	0.09	0.084	1.031	0.101
3G2H	0.24	0.229	1.028	1.523 ^a
3G2I	0.243	1.146 (3) ^a	0.189	1.147 (3) ^a
3G2K	0.183	1.666	0.184	1.671 ^a
3G2L	1.902 ^b	1.548 (2) ^{ab}	1.680 ^b	1.509 ^{ab}
1LX0	0.188	0.217	0.203	1.098 (3) ^{ab}
5LRC	0.108	0.093	1.032	0.104
5LRD	1.015	1.015	1.014	1.014
5LRE	0.149	0.16	0.166	0.146
5LRF	0.112	0.124	0.128	0.109
Average	0.391	0.580	0.615	0.777
Standard Deviation	0.567	0.631	0.552	0.660

^a The Heterocycle has flipped and has a different orientation to the ligand native to the receptor.

^b The R-Group has flipped and has a different orientation to the ligand native to the receptor.

^{ab} The Heterocycle & R-group has flipped and has a different orientation to the ligand native to the receptor.

(n) Given in parenthesis, the rank of the pose generated that correlates to the correct pose.

Table 4.4 – Ligand RMSDs (Å) achieved from Glide-SP & -XP docking compared with their crystallographic equivalents for the **ClusterResidueIn** (PDB Code 5JTU) receptor considered with and without conserved water molecules retained. Ligands above the cut-off RMSD for accuracy (>1.5 Å) are highlighted in orange.

PDB Code	ClusterResidueIn			
	With Waters		Without Waters	
	SP	XP	SP	XP
5JTU	0.138	0.13	0.131	0.129
5JTT	0.097	0.12	0.120	1.031
3G2H	1.031	1.031	1.298 (3) ^a	1.299
3G2I	1.145 (2) ^a	1.143 (2) ^a	1.166 (2) ^{ab}	1.133 (2) ^{ab}
3G2K	1.662 (2) ^a	1.669 ^a	0.164	1.668 ^a
3G2L	1.837 ^b	1.533 (2) ^a	1.541 (2) ^{ab}	1.474 ^a
1LX0	0.217	0.16	0.166	1.087 (2) ^{ab}
5LRC	0.152	0.146	0.130	0.119
5LRD	1.02	1.022	1.017	1.021
5LRE	0.163	0.189	0.182	0.192
5LRF	0.139	0.164	0.182	0.192
Average	0.691	0.664	0.554	0.850
Standard Deviation	0.667	0.619	0.569	0.581

^a The Heterocycle has flipped and has a different orientation to the ligand native to the receptor.

^b The R-Group has flipped and has a different orientation to the ligand native to the receptor.

^{ab} The Heterocycle & R-group has flipped and has a different orientation to the ligand native to the receptor.

(n) Given in parenthesis, the rank of the pose generated that correlates to the correct pose.

Overall, both Glide –SP and –XP with/without waters give very accurate results with respect to pose generation, but a few consistent outliers exist. All receptors used consistently generate inaccurate poses for the native ligand of 3G2L (1,2,3-Triazole_4R-1-Napthy; Table 4.1) usually due to the imprecise positioning of the 1-naphthyl R group, which are usually 90° perpendicular to the native structures, so this ligand could arguably be classed as an outlier. In general, -SP gives better results with respect to pose accuracy than –XP, both with and without waters. **GeneralClusterProteins** (Table 4.3) gives the most experimental consistent results in most cases whereas **ClusterResidueIn**, on average, gives the highest RMSD values, and therefore, generates the least accurate poses. Docking without waters

Table 4.5 – Ligand RMSDs (Å) achieved from Glide-SP & -XP docking compared with their crystallographic equivalents for the **ClusterResidueOut** (PDB Code 3G2H) receptor considered with and without conserved water molecules retained. Ligands above the cut-off RMSD for accuracy (>1.5 Å) are highlighted in orange.

PDB Code	ClusterResidueOut			
	With Waters		Without Waters	
	SP	XP	SP	XP
5JTU	0.120	0.126	0.142	0.143
5JTT	1.032	1.033	0.108	0.157
3G2H	1.041	0.145	1.046	0.178
3G2I	0.334	1.317 (3) ^a	0.300	1.051(2) ^a
3G2K	0.206	1.602 ^a	0.244	0.332
3G2L	1.879 ^b	1.511 ^a	1.956(3) ^b	1.941 ^b
1LX0	0.114	0.126	0.186	0.186
5LRC	0.117	0.145	0.130	0.109
5LRD	0.106	0.136	0.141	1.018
5LRE	1.507 ^b	0.145	0.167	0.156
5LRF	1.498 ^b	0.128	0.129	0.121
Average	0.723	0.583	0.414	0.490
Standard Deviation	0.682	0.636	0.578	0.595

^a The Heterocycle has flipped and has a different orientation to the ligand native to the receptor.

^b The R-Group has flipped and has a different orientation to the ligand native to the receptor.

^{ab} The Heterocycle & R-group has flipped and has a different orientation to the ligand native to the receptor.
(n) Given in parenthesis. The rank of the pose generated that correlates to the correct pose.

gives the best results for **ClusterResidueOut**, with only the outlying 3G2L protein-ligand above the cut-off with Glide- SP yielding an average RMSD of 0.414 Å and – XP yielding an average RMSD of 0.490 Å but not for the two other receptor conformation. Using Glide-XP, 1,2,3-Triazole glucose analogue poses had a tendency to be imprecise as the heterocycle “flipped” so that position of the nitrogen atoms are roughly 180° relative to the known ligand-protein complex. Usually, if the highest ranked pose generated has the incorrect orientation, a pose is generated which is similar to that of the known ligand in the protein complex and has a glide-score close to that of the highest ranking ligand pose. Glide-SP generated poses with the most accuracy, although Glide–SP (average RMSD = 0.691 Å) using the **ClusterResidueIn** (Table 4.4) receptor generated inaccurate poses similar to that

of Glide-XP (average RMSD = 0.664 Å) in the same receptor when docking with waters. When docking to this receptor with no waters, only two protein-ligands are above the cut-off point, however, overall give relatively poor RMSD values. Docking to **ClusterResidueOut** with waters is the only example of a receptor whose highest scoring pose for the 5LRE and 5LRF native ligands are incorrect. It was the orientation of the R group which was unlike that of the known ligand in the protein complex. This may be due to the environment in which the R-group occupies within this receptor having a different structure making up the β -cavity, therefore forming different interactions contributing to the common orientation seen.

4.4.2 Docking – Relative Ranking of Ligands and Scoring Results

*Glide Results for **GeneralClusterProtein**.*

The aim of docking was to derive an accurate model so as to screen a range of glucose analogues with heterocyclic linkers in order to predict their relative binding affinities to the catalytic site of GPb. 80 ligands (2 with known inhibitory (K_i) constants) were generated which after including tautomers gave 111 structures, which were then docked using Glide -SP and -XP with and without water molecules. Solvation of the **GeneralClusterProtein** was carried out with ClusterWaters considering all 12 proteins. 107 water molecules were retained, 5 of these occupying the catalytic site. The results of these calculations docking *with* and *without* waters are shown in Table 4.6. When docking with these water molecules retained, 440 diverse poses were generated using SP whereas only 147 poses were generated using XP; when docking with the absence of water molecules, fewer poses were generated with SP (409) and slightly more for XP (165). The ligands were ranked by best GlideScore (GS) and the highest ranked pose for each ligand was selected and then plotted against the $\ln K_i$ values. As aromatic R group substituents (Table 4.1) have been found to be the most effective for β -cavity binding we performed separate correlations studies (R^2 , the square of the Pearson correlation coefficient) for these groups (phenyl, 1 and 2-naphthalene) and also considering all ligands. The best correlation was from Glide-SP for hydrophobic aromatic R groups where values of $R^2 = 0.48$ without waters were obtained but which still cannot be called a reasonable

correlation (Figure 4.2). The best correlation received for all ligands was achieved when docking with Glide-XP with waters, obtaining a correlation of $R^2 = 0.38$. Ligands that ranked well usually had an available N-H on the heterocycle to hydrogen bond with the His377 backbone O. Large, hydrophobic R groups extending into the β -cavity of the catalytic site usually show the most favourable binding over small, polar groups.^[36] Comparing the results of the QPLD calculations without waters, we can see that the reparametrized charges did not lead to an improved correlation.

Table 4.6. – Table showing the results of the Glide-SP & -XP and quantum mechanics polarized ligand docking (QPLD) for **GeneralClusterProtein** in terms of number of poses obtained and correlation between predicted (Glidescore) and experimental ($\ln K_i$) binding affinities

SP/XP	Number of Poses with (without) H ₂ O's ^a	Correlation (R^2)			
		With H ₂ O's		Without H ₂ O's	
		All Ligands	Aromatic R groups	All Ligands	Aromatic R groups
SP	440 (409)	0.25	0.24	0.29	0.48
XP	147 (165)	0.38	0.45	0.31	0.26
QPLD	409	-	-	0.24	0.30

^a Without H₂O's are given in parentheses

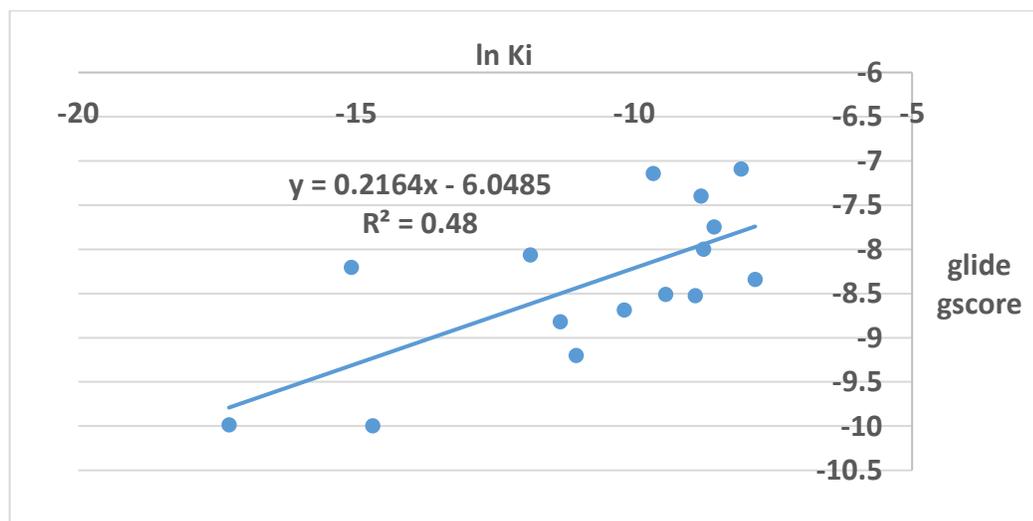


Figure 4.2 – Correlation of the Aromatic R groups $\ln K_i$ against glide gscore using SP without H₂O's present

Glide Results for ClusterResidueIn

The results of these calculations are shown in Table 4.7. Docking to this receptor model with Glide –SP and –XP generated 324 and 134 poses with waters, respectively. Docking without waters generated 378 poses for –SP and 141 poses for –XP. Docking without waters generated an increased amount of poses compared to with waters (as expected) but notably less poses were generated than the **GeneralClusterProtein** with waters. This is likely due to the increased amount of H₂O molecules (2 more) occupying the catalytic site sterically hindering orientations required for pose generation. The correlation achieved for this receptor was in general, slightly higher than for the previous receptor. A best correlation was obtained for hydrophobic aromatic R groups where values of $R^2 = 0.38$ with waters and $R^2 = 0.53$ without waters were achieved but which still cannot be called a reasonable correlation. Glide-XP gave a correlation of $R^2 = 0.42$ which was obtained for all ligands without waters, The increased amount of water molecules in the catalytic site will contribute to an increased amount of binding interactions with surrounding residues, the larger R groups having less space to adopt different conformations in the β -cavity, but in theory more correct poses. Finally correlations with this receptor model are generally slightly better without the H₂O's present.

Table 4.7 – Table showing the correlation achieved from Glide-SP & -XP for the **ClusterResidueIn**.

SP/XP	Number of Poses with (without) H ₂ O's ^a	Correlation (R ²)			
		With H ₂ O's		Without H ₂ O's	
		All Ligands	Aromatic R groups	All Ligands	Aromatic R groups
SP	324 (378)	0.30	0.38	0.36	0.53
XP	134 (141)	0.35	0.36	0.42	0.45

^a Without H₂O's are given in parentheses

*Glide Results for **ClusterResidueOut**.*

Docking to the receptor model **ClusterResidueOut** with Glide –SP and –XP generated 445 and 142 poses respectively with water and 440 and 178 poses respectively without water. The results of these calculations are shown in Table 4.8. Docking to this receptor generated a similar amount of poses to the **GeneraClusterProtein**, and contained the same amount of water molecules (5) occupying the catalytic site, showing that steric hindrance caused by the water molecules may indeed have an effect on pose generation. The best correlation of $R^2 = 0.46$ was obtained with Glide-SP for aromatic R groups and Glide-XP ($R^2 = 0.36$) for ligands with aromatic R groups without waters, values which can be considered modest. This protein conformation performs notably poorer when docking with waters in comparison to the other receptors. It is possible that this protein orientation is unfavourable for binding of most ligands, consistent with only 2 of the 12 PDBs studied adopting this orientation. In the most common orientation of the Asn282, Asp283, Asn284 and Phe285 residues (**ClusterResidueIn**), ligands whose heterocycle is able to act as a H-bond acceptor also benefits from hydrogen bonding the with Asn284 residue NH; in **ClusterResidueout**, Asn284 is facing a different direction in the orientation away from the binding site and cannot therefore achieve this favourable binding interaction, leading to a lower average Glide-Score across all of the ligands.

Table 4.8 – Table showing the correlation achieved from Glide-SP & -XP for the **ClusterResidueOut**.

SP/XP	Number of Poses with (without) H2O's ^a	Correlation (R^2)			
		With H2O's		Without H2O's	
		All Ligands	Aromatic R groups	All Ligands	Aromatic R groups
SP	445 (440)	0.22	0.21	0.23	0.44
XP	142 (178)	0.19	0.25	0.31	0.45

^a Without H2Os are given in parentheses

GOLD Results for GeneralClusterProtein

Table 4.9 – Table showing the correlations achieved from GOLD for **GeneralClusterProtein**

Water Model	Number of poses	Correlation (R ²)	
		All ligands	Aromatic R groups
Without H ₂ O	334	0.37	0.42
25 Static H ₂ O	343	0.34	0.35
2 Static H ₂ O	337	0.36	0.41
2 Spin/Toggle H ₂ O	363	0.35	0.40

Docking with GOLD using the **GeneralClusterProtein** receptor model generated 343 and 334 poses with static water molecules and without water molecules, respectively, as well as 337 and 363 poses with two static waters and when two waters were allowed to toggle and spin, respectively. The results of these calculation are shown in Table 4.9 Both GOLD and Glide were set to generate 10 poses per ligand and docking to this protein with GOLD generated on average less poses than that of Glide. Similar correlations were achieved for all docking models tested: without H₂O and these explicit H₂O models. These ranged from 0.34-0.37 for all ligands and 0.35-0.42 for the aromatic R groups. In comparison with the Glide docking (Table 4.6), the correlations without waters was slightly better for all ligands ($r^2 = 0.37$ vs $r^2 = 0.29$). Overall, for the GOLD water models, slightly improved correlations were observed compared to Glide –SP and –XP, however, these improvements cannot be considered significant, so that none of the docking models tested yielded predictive capability.

Summary of docking results

Overall, docking performance with Glide-SP was better than that of –XP and generated vastly more poses, which were on average more accurate based on the RMSDs. Each receptor displayed an improved correlation when docking was performed with no water molecules and the best correlations were achieved with aromatic R groups and in particular the best correlations were achieved with the **ClusterResidueIn** receptor. Docking to this receptor with Glide-SP and no waters

achieved a correlation of $R^2 = 0.53$ for the aromatic R groups. However, the correlations achieved overall from Glide are relatively modest and docking with GOLD considering different water models did not lead to improved results. More accurate Prime MM-GBSA post-docking calculation should in theory yield better results.

4.4.3 Prime MM-GBSA Results

GeneralClusterProtein Post-docking

Prime post-docking calculations were performed on the docking poses (up to 10 for each structure) and ΔG_{bind} values calculated according to Equation 4.1. These calculations were performed *with* and *without* explicit waters, and also *with* and *without* strain correction. Strain correction corresponds to relaxation of the protein and/or ligand in the free unbound state as compared with its conformation in the bound state. The results of the Prime MM-GBSA post-docking calculations for **GeneralClusterProtein** poses revealed that the poses gathered from both Glide –SP and –XP achieve better correlations compared to docking alone when the predicted ΔG_{bind} values were plotted against $\ln K_i$ values (Table 4.10). The correlations achieved without explicit waters from Glide-SP outperforms all other solvation models with this receptor which suggests a better estimation of the relative ΔG_{bind} values. The Prime-MM-GBSA calculations give better correlations from the –SP poses as opposed to –XP; this is consistent with the poses generated from Glide-SP having the most accurate pose orientations based on the RMSD calculations. Additionally Glide –SP poses performed better without including ligand strain for the rigid receptor but performed better with the protein/ligand strain at a receptor flexibility of 5 Å. Glide –XP poses performed better including strain correction for both a rigid receptor and that of 5 Å flexibility but overall performed better with a rigid receptor, suggesting that Prime calculations for both Glide –SP and –XP are more accurate when keeping the protein rigid. Overall, the best correlation achieved for this receptor was using Glide –SP poses without waters using a rigid receptor and without including ligand strain.

Table 4.10. - Table showing the correlation achieved from Prime MM-GBSA for the GeneralClusterProtein.

Glide-SP/XP	Protein Flexibility	Prime Correlation (R^2) with Strain (No Strain)			
		With H2O's		Without H2O's	
		All ligands ^a	Aromatic R groups ^a	All ligands ^a	Aromatic R groups ^a
SP	0 Å flexibility	0.62 (0.73) [0.25]	0.69 (0.80) [0.24]	0.71 (0.75) [0.29]	0.73 (0.86) [0.48]
	5 Å flexibility	0.50 (0.51)	0.58 (0.65)	0.70 (0.59)	0.81 (0.72)
XP	0 Å flexibility	0.54 (0.52) [0.38]	0.55 (0.51) [0.45]	0.59 (0.58) [0.31]	0.58 (0.51) [0.26]
	5 Å flexibility	0.44 (0.47)	0.43 (0.47)	0.53 (0.49)	0.56 (0.50)

^a Previous Glide docking correlations are given in square parentheses

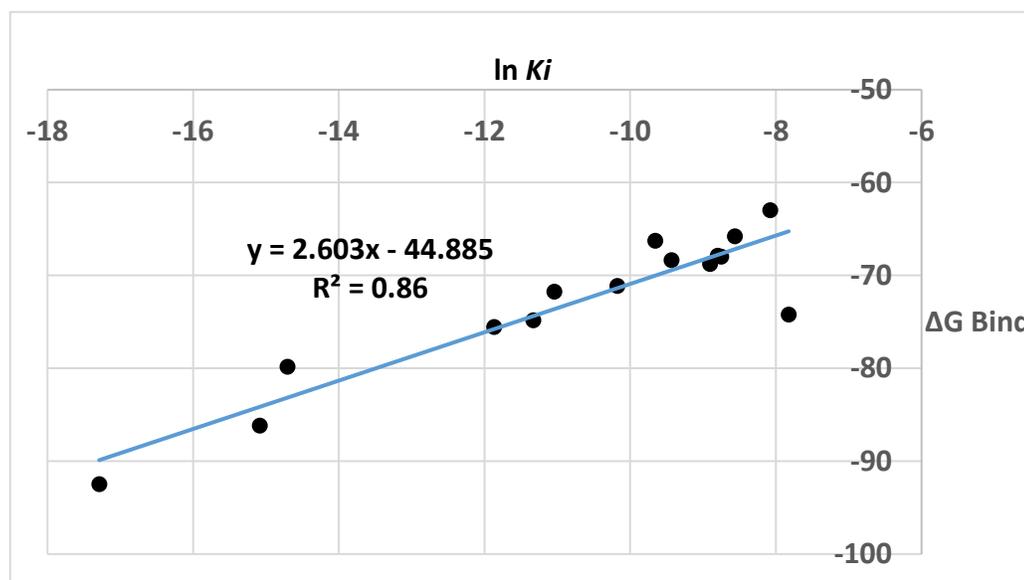


Figure 4.3 – Prime (No strain) correlation of Aromatic R groups without waters using input poses from glide-SP.

With this, R^2 correlations for all ligands and for the aromatic R groups of 0.75 and 0.86 (Figure 4.3), respectively, were obtained. However, similar favourable correlations were also achieved with waters for all ligands ($R^2 = 0.73$) and the aromatic R groups ($R^2 = 0.80$) also using a rigid receptor without ligand strain of particular significance. These correlations are considerable improvements on those gathered from docking with Glide –SP with waters for all ligands ($R^2 = 0.25$) and for the aromatic R groups ($R^2 = 0.24$), and without waters for all ligands ($R^2 = 0.29$) and aromatic R groups ($R^2 = 0.48$).

For comparative purposes, the poses gathered from GOLD were then also used as input for Prime MM-GBSA calculations. Again, the Prime MM-GBSA results consistently gave significantly better correlations than those gathered from the Gold-ChemPLP docking alone. As with the Glide poses, Prime calculation using the GOLD docking poses with a rigid receptor outperformed Prime with 5 Å protein flexibility for all of the solvation models used. The best correlation achieved was from using the poses from the docking with two toggle and spin water molecules model for all ligands ($R^2 = 0.66$) and also for the aromatic R groups ($R^2 = 0.67$) including the ligand strain term. Poses from the same model also gave the best correlation without including the ligand strain term for all ligands ($R^2 = 0.58$) and for the aromatic R groups ($R^2 = 0.62$). Similar correlation were also obtained for the 2 static H₂O's. In contrast to Glide, these particular solvation models with H₂O's produced better correlations than without explicit H₂O's. Significantly, this was one of the few explicit water models that outperformed MM-GBSA without H₂O's included. The best correlations without H₂O's with no strain where values of $R^2 = 0.59$ (all ligands) and $R^2 = 0.62$ (Aromatic R groups) were obtained. The correlations achieved can be considered reasonable, but are poor in comparison to the poses generated from Glide –SP for the same receptor.

Table 4.11 – Table showing the correlation achieved from GOLD using ChemPLP and Prime for GeneralClusterProtein

Protein Flexibility	Prime Correlation (R^2) with Strain (No Strain)							
	Without H ₂ O's		With static H ₂ O's		With two Static H ₂ O's		With two Spin/Toggle H ₂ O's	
	All ligands ^a	Ar R groups ^a	All ligands ^a	Ar R groups ^a	All ligands ^a	Ar R groups ^a	All ligands ^a	Ar R groups ^a
0 Å flexibility	0.59 (0.56)	0.62 (0.55)	0.56 (0.56)	0.59 (0.56)	0.62 (0.58)	0.64 (0.59)	0.66 (0.58)	0.67 (0.62)
	[0.37]	[0.42]	[0.34]	[0.35]	[0.36]	[0.41]	[0.35]	[0.40]
5 Å flexibility	0.45 (0.53)	0.49 (0.57)	0.47 (0.49)	0.48 (0.51)	0.56 (0.56)	0.59 (0.59)	-	-

^a Previous GOLD docking correlations are given in square parentheses

Quantum Mechanics Polarized Ligand Docking Post-docking results

With respect to MM-GBSA results with the QPLD docking poses from Glide, the new quantum mechanics partial charges calculated for the ligands in the field of the receptor achieved good correlations, but does not lead to improvements over the standard Glide results using molecular mechanic forcefield based parametrisation. The correlations are shown in Table 4.12. The best correlation achieved was using a rigid receptor for all ligands ($R^2 = 0.69$) and for the aromatic R groups, a value of $R^2 = 0.81$ was obtained. In comparison, post-docking using the same receptor with MM forcefield parameters achieved correlation for all ligands ($R^2 = 0.75$) and for the aromatic R groups ($R^2 = 0.86$).

Table 4.12. - Table showing the correlation achieved from QM-PLD for the **GeneralClusterProtein**.

	Protein Flexibility	Prime Correlation (R^2) with Strain (No Strain)	
		Without H2O's	
		All ligands	Aromatic R groups
Glide	N/A	0.24	0.30
Prime	0 Å flexibility	0.67 (0.69)	0.75 (0.81)
	5 Å flexibility	0.61 (0.48)	0.68 (0.61)

ClusterResidueIn Post-docking

For this protein conformation, Prime MM-GBSA post-docking calculations revealed that the poses gathered from both Glide –SP and XP achieve similar correlations with and without water when the predicted ΔG_{bind} values were plotted against $\ln K_i$ values (Table 4.13). Similar to the previous receptor, the correlations achieved from Glide–SP without waters slightly outperform all other solvation models used in this receptor which again suggests a better estimation of the relative ΔG_{bind} values. Also, Prime calculations using Glide –SP gave better correlations than that of –XP. In agreement; the RMSD calculations (4.4.1) show again that Glide –SP was more accurate than -XP on average in terms of pose generation, although they were not as accurate as the last receptor. The correlations achieved with Prime from Glide –SP were again best using a rigid receptor with no ligand strain. This again suggests that Prime MM-GBSA calculations are more accurate when keeping the protein rigid. The best correlation achieved for this receptor was gathered from using Glide –SP without waters using a rigid receptor for all ligands ($R^2 = 0.75$) and for the aromatic R groups ($R^2 = 0.81$) without including the ligand strain term. Similar favourable correlations were also achieved with waters for all ligands ($R^2 = 0.75$) and the aromatic R groups ($R^2 = 0.81$) also using a rigid receptor without ligand strain. These

correlations are again significant improvements on those gathered from docking with Glide –SP with waters for all ligands ($R^2 = 0.30$) and for the aromatic R groups ($R^2 = 0.38$) and without waters for all ligands ($R^2 = 0.36$) and for the aromatic R groups ($R^2 = 0.53$). This receptors residue orientation is the same as the previous receptor and differs only in solvation; with the catalytic site containing 2 more water molecules and the overall protein containing 27 more water molecules altogether. As discussed previously, the presence of these waters in the catalytic site likely impacted negatively on the RMSD of the poses generated and therefore, correlations achieved for this receptor are slightly less than that of the previous receptor.

Table 4.13. - Table showing the correlation achieved from Prime MM-GBSA for the ClusterResidueIn.

Glide-SP/ XP	Protein Flexibility	Prime Correlation (R^2) with Strain (No strain)			
		With H2O's		Without H2O's	
		All ligands ^a	Aromatic R groups _a	All ligands ^a	Aromatic R groups _a
SP	0 Å flexibility	0.60 (0.75) [0.30]	0.72 (0.81) [0.38]	0.69 (0.75) [0.36]	0.69 (0.81) [0.53]
	5 Å flexibility	0.54 (0.56)	0.49 (0.55)	0.42 (0.54)	0.52 (0.64)
XP	0 Å flexibility	0.54 (0.49) [0.35]	0.57 (0.54) [0.36]	0.58 (0.52) [0.42]	0.57 (0.49) [0.45]
	5Å flexibility	0.45 (0.48)	0.44 (0.47)	0.55 (0.56)	0.56 (0.59)

^a Previous Glide docking correlations are given in square parentheses.

ClusterResidueOut Post-docking

For this receptor model, unlike the previous receptors, Prime MM-GBSA post docking calculations show that poses gathered from both –SP and –XP achieve better correlations post-docking with explicit water molecules present in the protein when the predicted ΔG_{bind} values were plotted against $\ln K_i$ values (Table 4.14). Prime calculations using Glide –SP poses gave better correlations than those of –XP both with and without water. The poses that were generated from Glide –SP with waters generated the least accurate geometry across all receptors based on the RMSD calculations (4.4.1), however, the most accurate poses without waters. The inaccurate poses generated are often due to the wrong orientations of the R-group substituents and not the heterocyclic linkers themselves. Similar to the previous receptors, the correlations achieved with Prime from the Glide–SP poses are greater than those of Glide-XP, but displayed better correlations than the previous receptors for –XP in general. Interestingly, this receptor revealed greater correlations with 5 Å flexibility and with the strain term included as opposed to holding the receptor rigid. With flexibility, the receptor has the ability to adopt a geometry similar to those in the more common receptors which are more favourable for binding. The best correlation achieved for this receptor was gathered from using Glide –SP with waters and a receptor flexibility of 5 Å for all ligands ($R^2 = 0.71$) and for the aromatic R groups ($R^2 = 0.75$). Favourable correlations were also achieved using Glide –XP with water using a receptor flexibility of 5 Å without the ligand strain for all ligands ($R^2 = 0.63$) and for the aromatic R groups ($R^2 = 0.66$). A good correlation was also achieved using Glide-SP without waters with the strain term included for the aromatic R groups ($R^2 = 0.72$). These correlations are significant improvements on those gathered from docking with Glide –SP with waters for all ligands ($R^2 = 0.22$) and for the aromatic R groups ($R^2 = 0.21$) and without waters for all ligands ($R^2 = 0.23$) and aromatic R groups ($R^2 = 0.44$). Surprisingly, however, the correlations achieved from Prime (0.22-0.29) are worse than those gathered from using Glide –XP without waters for all ligands ($R^2 = 0.31$) and the aromatic R groups ($R^2 = 0.45$). However, with H₂O present, the post-docking Prime MM-GBSA results revealed considerable

improvements ($R^2 = 0.51-0.68$) compared to the docking alone ($R^2 = 0.19$ – all ligands; $R^2 = 0.09$ – Aromatic R groups)

Table 4.14. - Table showing the correlation achieved from Prime MM-GBSA for the ClusterResidueOut.

Glide-SP/ XP	Protein Flexibility	Prime Correlation (R^2) with Strain (No strain)			
		With H2O's		Without H2O's	
		All ligands ^a	Aromatic R groups ^a	All ligands ^a	Aromatic R groups ^a
SP	0 Å flexibility	0.62 (0.66) [0.22]	0.60 (0.67) [0.21]	0.38 (0.35) [0.23]	0.48 (0.43) [0.44]
	5 Å flexibility	0.71 (0.46)	0.75 (0.53)	0.55 (0.46)	0.72 (0.50)
XP	0 Å flexibility	0.53 (0.53) [0.19]	0.51 (0.51) [0.09]	0.29 (0.27) [0.31]	0.25 (0.22) [0.45]
	5Å flexibility	0.58 (0.63)	0.68 (0.66)	0.28 (0.28)	0.29 (0.25)

^a Previous Glide docking correlations are given in square parentheses

Summary

For Prime MM-GBSA post-docking calculations, -SP poses generally scored better than their -XP counterparts; also without waters generally gave slightly better correlations. However, for MM-GBSA using GOLD poses retaining just two waters, both static and in spin/toggle mode, better correlations were obtained compared to without explicit H₂O. Nevertheless, these correlations were still less than those obtained for post-docking with the Glide poses. No strain predominantly produced better results compared to experimental in terms of correlations; likewise retaining a rigid receptor for these calculations. Reparameterization of ligand partial charges

using quantum mechanics also led to good correlations but were not superior to those obtained using standard (Glide) MM forcefield charges. Overall, the best MM-GBSA model obtained was for **GeneralClusterProtein** using Glide-SP poses, without waters, with a rigid receptor and with no strain correction. This model produced a correlation of $R^2 = 0.86$ (aromatic R groups) and $r^2 = 0.75$ (all ligands). On this basis, we took this model forward as the optimal model obtained. However, in the next section, we looked at this model in more detail considering refinements that may improve its predictive capability when applied to the unknown ligands from Table 4.1.

4.4.4 Refinement of MM-GBSA Model.

Inclusion of Entropy Effects

An estimate for the loss of ligand entropy (ΔS_{MM}) on binding was calculated using MM with the OPLS3 forcefield and the Rigid Rotor Harmonic Oscillator (RRHO) approximation. Using this method, the change in vibrational, rotational and translational entropy (ΔS_{MM}) of the ligands on binding was considered. RRHO calculations was performed using MacroModel 11.0.^[86] The ΔG_{bind} values from **GeneralClusterProtein** using Glide-SP poses, without waters and a rigid receptor were corrected with the entropy term;

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{solv} - T\Delta S_{MM} \quad (\text{Equation 4.2})$$

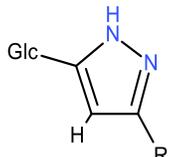
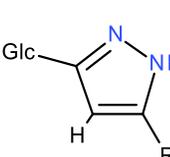
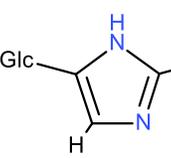
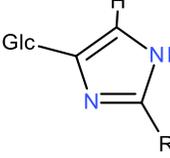
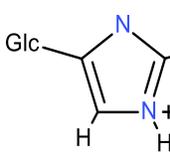
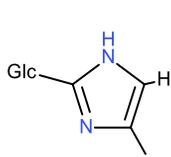
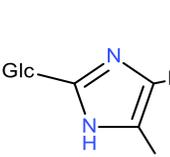
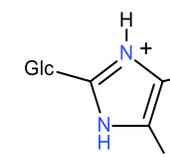
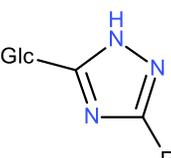
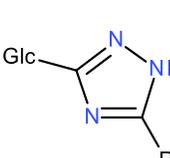
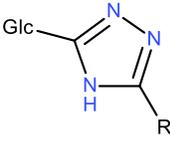
This model produced a correlation of $R^2 = 0.70$ (Aromatic R groups) and $R^2 = 0.69$ (all ligands) including the strain term (Table 4.10). The model also gave a correlation of $R^2 = 0.83$ (Aromatic R groups) and $R^2 = 0.75$ (all ligands) without the strain term. Hence, inclusion of the extra entropy term gave similar correlations as to those obtained with standard Prime (Equation 4.1).

Inclusion of Tautomeric Correction.

Our set of 27 training set ligands (Table 4.1) consists of 4 ligands shown in Table 4.15 (2,3-Diazole_4R, 2,4-Diazole_4R, 2,5-Diazole_4R, 2,3,5-Triazole_4R) with the potential to form tautomers and/or different ionization states. Until now, we have

considered the best ΔG_{bind} as the predicted experiment value of each ligand irrespective of ionization/tautomeric state. However, to consider more accurately the potential of these ligands to bind in these different states, quantum mechanics (QM) calculations were performed by a collaborator^[110] to establish which of the tautomeric states of the ligands are the most favourable in the unbound state. The results of the calculation are also shown in Table 4.15.

Table 4.15. – Relative energies (gas and solution phase (in parentheses)) of the tautomeric forms of ligands considered in this work calculated using quantum mechanics. Relevant ionization states are also shown.

Ligand	Neutral state			Ionized form
	Tautomer 1	Tautomer 2	Tautomer 3	
2,3-Diazole_4R-phenyl	 +3.0 (+2.3)	 0.0 (0.0)	-	-
2,4-Diazole_4R	 +4.8 (+4.6)	 0.0 (0.0)	-	 N/A
2,5-Diazole_4R	 0.0 (0.0)	 +0.61 (+2.09)	-	 N/A
2,3,5-Triazole_4R	 0.0 (0.0)	 +3.18 (+1.61)	 +6.12 (+6.19)	-

The most favouring binding state of the 2,4-diazole_4R and 2,5-diazole_4R ligands were consistently protonated (+1) state. The calculated pKa values of these ligands (R = phenyl), by our collaborator, were 5.9 and 5.5, respectively, meaning that these ligands will be partially protonated in solution, favourable for binding. For 2,3,5-triazole_4R, the most stable tautomer in the free state (tautomer 1) was almost the most favourable in terms of ΔG_{bind} . Hence, there was no tautomeric correction. However, for 2,3-diazole_4R-phenyl, the most stable tautomer in the free state was tautomer 2, but ΔG_{bind} was stronger for its tautomer 1. Taking the ΔG_{bind} value instead for tautomer 2, led to a significant improvement in correlation (Table 4.16, models 3 and 4) leading to R^2 values as large as 0.95-0.97 for the aromatic R group.

Table 4.16. – Correlations of the best MM-GBSA models with and without the entropy term as well as the tautomeric correction.^a

Model Number	MM-GBSA model	Correlation (R^2)	
		All Ligands	Aromatic R groups
		Strain (No Strain)	Strain (No Strain)
1	Best Prime model, Eq. (4.1)	0.71 (0.75)	0.73 (0.86)
2	MM-GBSA model, Eq. (4.2) - including entropy	0.69 (0.75)	0.70 (0.83)
3	Eq. (4.1) with tautomeric correction	0.82 (0.83)	0.84 (0.97)
4	Eq. (4.2) with tautomeric correction	0.82 (0.83)	0.86 (0.95)

^a Calculations were performed using Strike 5.01^[86]

Based on the results in Table 4.16, we considered the models for prediction of heterocyclic linkers with different aromatic R groups without the strain term included, the most useful based on their excellent R^2 values. We then looked at these models in more detail using Strike 5.01^[86] so as to obtain a full set of statistics and make predictions on the unknown compounds in Table 4.1.

Table 4.17. – Results of the Strike statistics performed for models 1-4.

Model	Equation	R ²	LOO Cv Q ²	RMSE	F[df ₁ ,df ₂]	p-value
1	$(\ln K_i)^{\text{pred}} = 0.33701(\pm 0.036963) \Delta G_{\text{bind}}^{\text{pred}} + 13.890(\pm 2.7190)$	0.86	0.84	1.013	83.1 [1,13]	5.2e-7
2	$(\ln K_i)^{\text{pred}} = 0.32118(\pm 0.040879) \Delta G_{\text{bind}}^{\text{pred}} + 6.5651(\pm 2.2290)$	0.83	0.79	1.149	61.7 [1,13]	2.7e-6
3	$(\ln K_i)^{\text{pred}} = 0.33911(\pm 0.17491) \Delta G_{\text{bind}}^{\text{pred}} + 13.795(\pm 1.2746)$	0.97	0.96	0.50	375.9 [1,13]	5.6e-11
4	$(\ln K_i)^{\text{pred}} = 0.32590(\pm 0.020994) \Delta G_{\text{bind}}^{\text{pred}} + 6.5347(\pm 1.1279)$	0.95	0.93	0.62	241.0 [1,13]	9.0e-10

The results of the multiple linear regressions using the four models for the training set of 15 ligands with aromatic R groups and the equations relating to them are shown in Table 4.17. A comprehensive statistical evaluation of the relationship between predicted and experimental activities ($\ln K_i$) was obtained which included the correlation coefficient R^2 , leave-one-out cross validation (LOO-cv) Q^2 , F and p-values and also the root mean square error (RMSE). All models gave consistently excellent R^2 values and LOO-cv Q^2 values (Q^2 values only slightly less than their R^2 counterparts), and all of the models have corresponding p-values $\ll 0.05$ showing statistical significance. These models are shown in Figures 4.4 - 4.7. All of the aforementioned models were then used to predict the K_i values of the set of unknown compounds in Table 4.1. In order to consider a number of factors that are relevant to inhibition; including entropy and tautomeric correction. All models have favourable attributes and give similar excellent statistical values. Accordingly, using these models 1-4, predicted K_i values were produced and given as a range (incorporating RMSE values) which should provide considerable insight into the inhibitory potential of the unknown compounds.

Table 4.18. – Ligands used in models 1-4 along with the codes present on Figure 4.4-4.7.^a

Ligand	Code		
	Phenyl	1-Napthyl	2-Napthyl
1,2,3-Triazole_4R	1a	1b	1c
1-Glc-2S5NThiazole_4R	2a	N/A	2c
2,3,5-Triazole_4R	3a	N/A	3c
2,3-Diazole_3R	4a	N/A	N/A
2,4,5-Triazole3-one_4R	5a	N/A	5c
2,4N-5Oxodiazole_3R	6a	N/A	6c
2,5-Diazole_4R	7a	N/A	7c
2,SN3O-Oxodiazole	N/A	N/A	8c

^a N/A = not available and included in test set for prediction

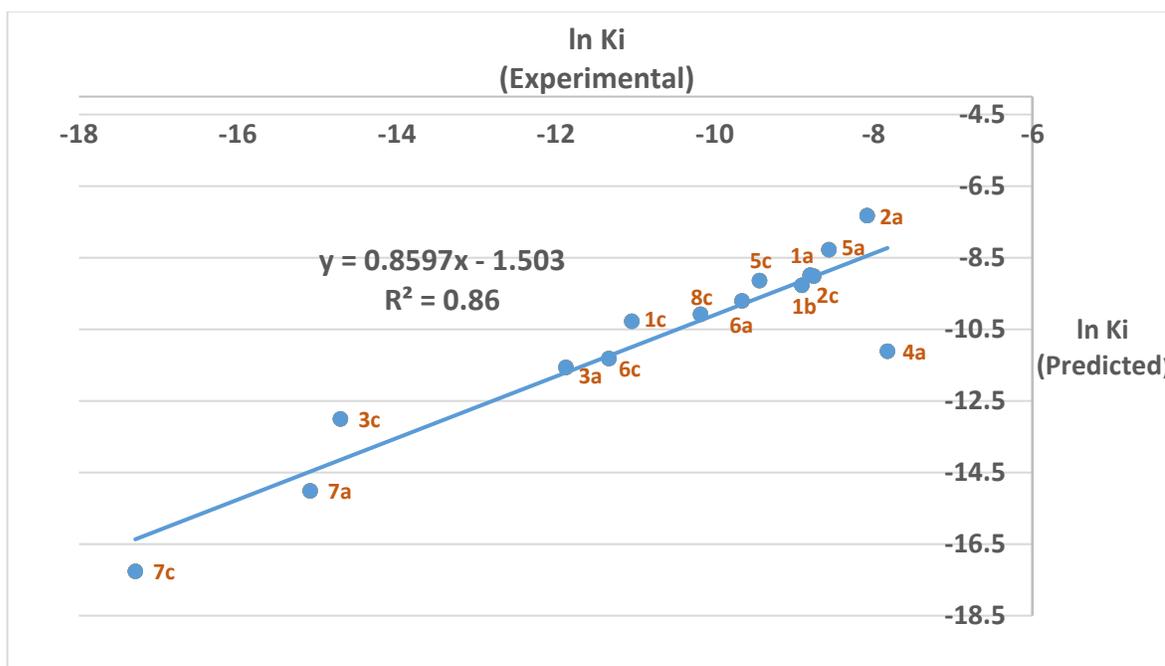


Figure 4.4 – Correlation between predicted and experimental ln Ki from Prime (NS) using Model 1. (Table 4.17)

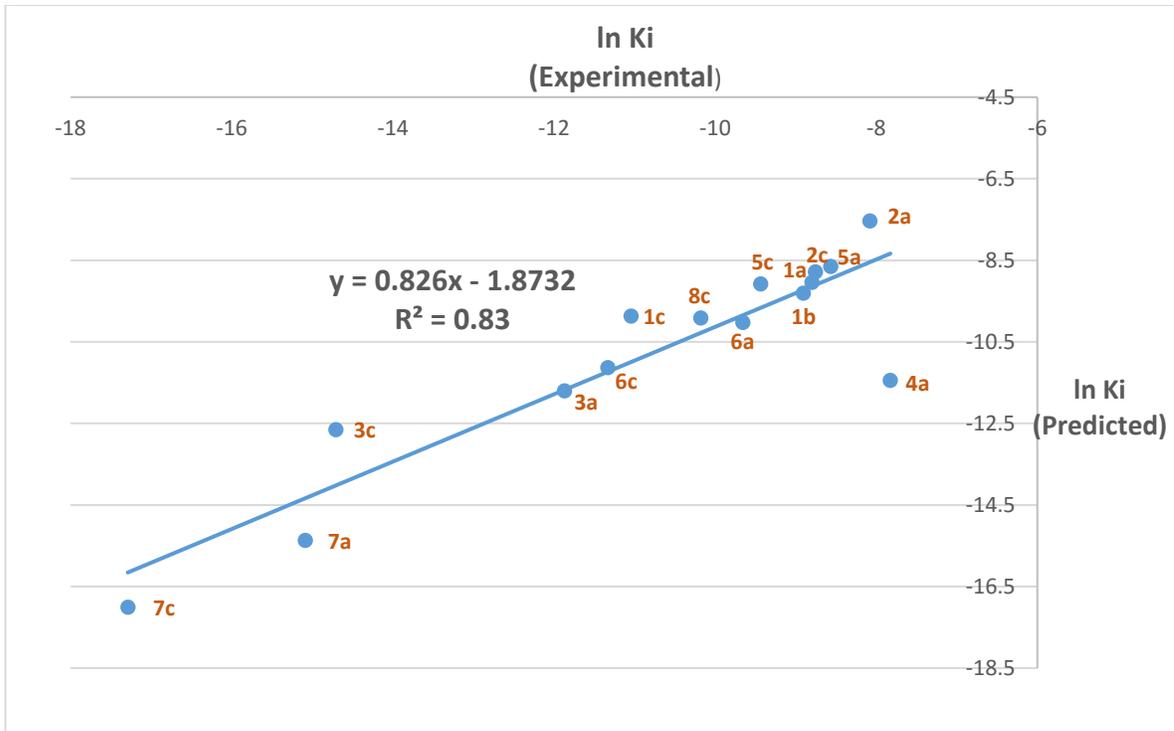


Figure 4.5 – Correlation between predicted and experimental In Ki from Prime (NS) using Model 2. (Table 4.17)

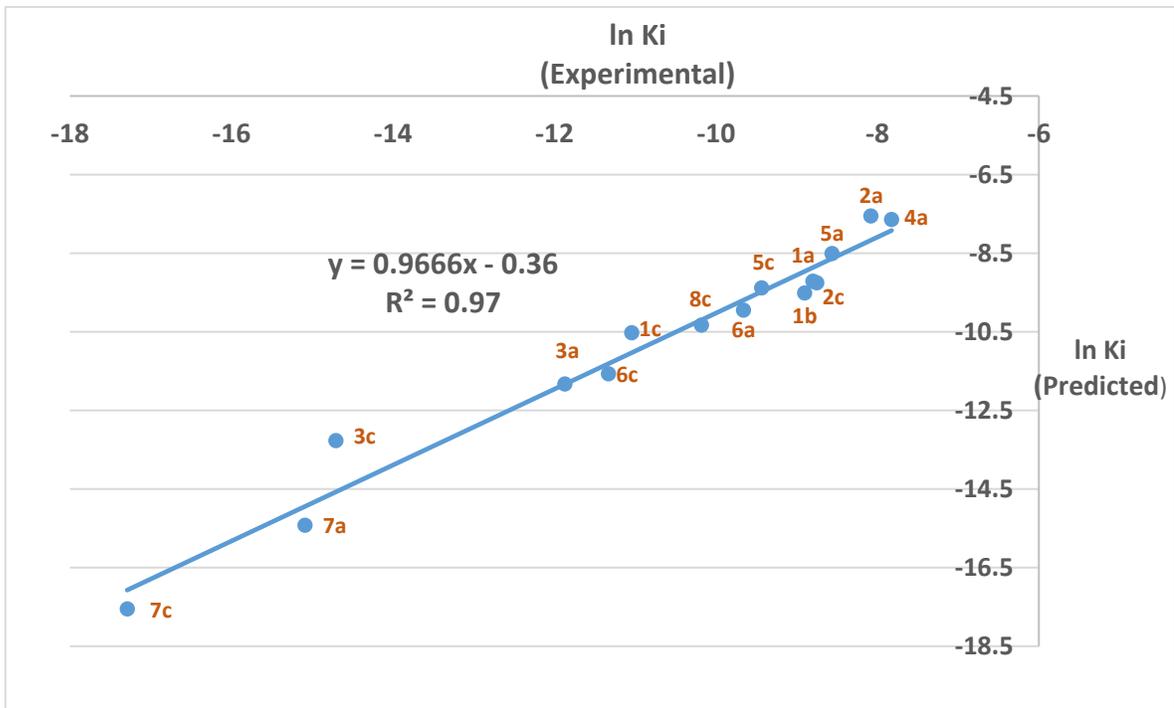


Figure 4.6 – Correlation between predicted and experimental In Ki from Prime (NS) using Model 3. (Table 4.17)

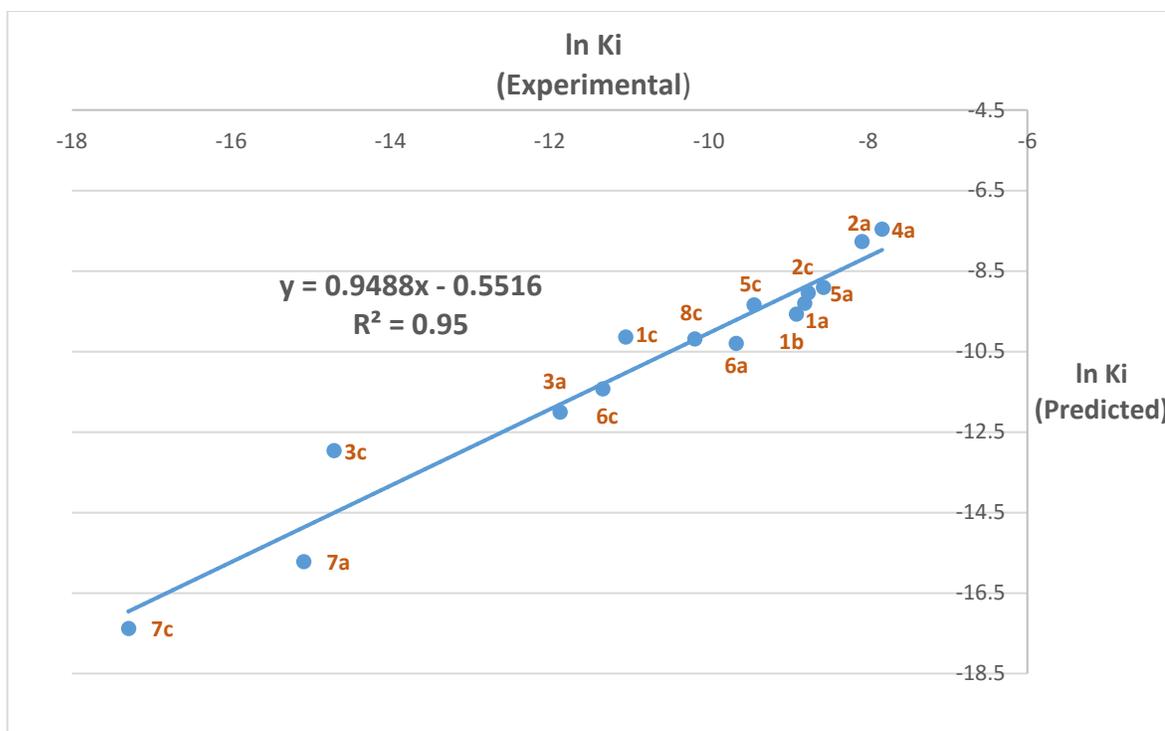
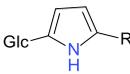
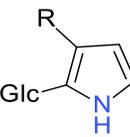


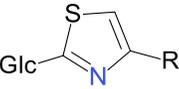
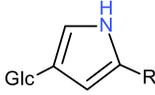
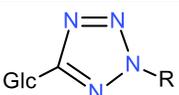
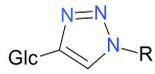
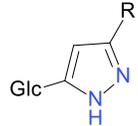
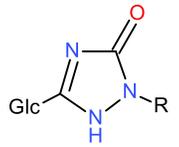
Figure 4.7 – Correlation between predicted and experimental $\ln K_i$ from Prime (NS) using Model 4 (Table 4.17)

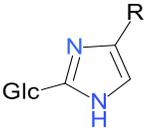
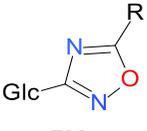
4.4.5 Prediction of Inhibitory Potential for Unknown Compounds.

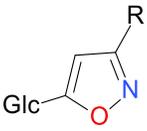
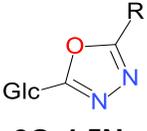
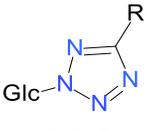
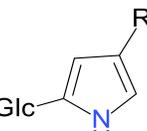
The results of the predictions using models 1-4 (Table 4.17) are shown in Table 4.18.

Table 4.19. - The predicted K_i values of unknown set of ligands in Table 4.1.

Heterocycle	R group	Predicted Activities as $\ln K_i$ (K_i in μM) [K_i range] ^a				Overall K_i Range (μM)
		Model 1	Model 2	Model 3	Model 4	
 2N-Pyrrole_3R	1-Naphthyl	-7.33 (658) [239-1813]	-7.57 (518) [164-1636]	-7.55 (525) [318-865]	-7.80 (409) [219-762]	164-1813
	2-Naphthyl	-7.31 (672) [244-1850]	-7.5 (553) [175-1745]	-7.53 (535) [345 - 882]	-7.74 (436) [234-814]	175-1850
 2N-Pyrrole_5R	Phenyl	NI. ^b	NI. ^b	NI. ^b	NI. ^b	NI. ^b
	1-Naphthyl	-4.41 (12155) [4413-33480]	-4.33 (13207) [4185-41677]	-4.62 (9862) [5982 - 16260]	-4.52 (10911) [5850 - 20349]	4185-41677
	2-Naphthyl	NI. ^b	NI. ^b	NI. ^b	NI. ^b	NI. ^b

 2S5N-Thiazole_4R	1-Naphthyl	-8.78 (154) [56-424]	-8.37 (233) [74-735]	-9.02 (121) [74-200]	-8.62 (181) [97-338]	56-735
 GlucoPyrrole_3R	Phenyl	-5.51 (4042) [1468 - 11133]	-6.24 (1960) [621- 6184]	-5.73 (3257) [1975- 5370]	-6.45 (1576) (845- 2938]	621- 11133
	1-Naphthyl	-7.75 (431) [156- 1186]	-7.78 (418) [133- 1320]	-7.98 (342) [208- 564]	-8.02 (328) [176-613]	133- 1320
	2-Naphthyl	-7.21 (741) [269- 2040]	-7.26 (705) [223- 2223]	-7.43 (591) [358- 974]	-7.49 (558) [299- 1040]	223- 2223
 2,3,4,5-Tetrazole_3R	Phenyl	-8.45 (214) [78-589]	-8.73 (162) [51-511]	-8.68 (169) [103-279]	-8.98 (126) [67-234]	51-589
	1-Naphthyl	-10.93 (18) [7-49]	-10.84 (20) [6-62]	-11.18 (14) [9-23]	-11.12 (15) [8-28]	6-62
	2-Naphthyl	-10.74 (22) [8-60]	-10.46 (29) [9-90]	-10.99 (16.87) [10-28]	-10.74 (22) [12-40]	8-90
 2,3,4-Triazole_4R	Phenyl	-8.89 (138) [50-381]	-9.41 (82) [26-258]	-9.12 (109) [66-180]	-9.68 (63) [34-117]	26-381
	1-Naphthyl	-9.87 (52) [19-142]	-9.95 (48) [15-150]	-10.11 (45) [25-70]	-10.23 (36) [19-68]	15-150
	2-Naphthyl	-10.59 (25) [9-69]	-10.70 (23) [7-71]	-10.84 (20) [20-32]	-10.98 (17) [9-32]	7-71
 2,3-Diazole_4R	1-Naphthyl	-10.91 (18) [7-50]	-10.93 (18) [6-56]	-11.16 (14) [9-23]	-11.22 (13) [7-25]	6-56
	2-Naphthyl	-11.95 (6.5) [2-18]	-11.91 (7) [2-21]	-12.20 (5) [3-8]	-12.22 (5) [3-9]	2-21
 2,4,5-Triazole3-one_4R	1-Naphthyl	-7.53 (536) [194- 1477]	-7.28 (688) [218- 2173]	-7.76 (426) [259- 703]	-7.52 (545) [292- 1016]	194- 2173

 2,5-Diazole_4R	1-Naphthyl	-17.74 (0.02) 0.007- 0.1]	-17.30 (0.03) [0.01-0.1]	-18.03 (0.015) [0.009- 0.024]	-17.68 (0.02) [0.01- 0.04]	0.007- 0.1
 2,5N-3-Oxadiazole	1-Naphthyl	-10.46 (29) [10-79]	-10.39 (31) [10-97]	-10.71 (22) [14-37]	-10.67 (23) [12-43]	10-97
 2S4N-Thiazole_3R	Phenyl	-10.71 (22) [8-62]	-11.15 (14) [5-46]	-10.96 (17) [11-29]	-11.44 (11) [6-20]	5-62
	1-Naphthyl	-11.32 (12) [4-33]	-11.17 (14) [4-45]	-11.57 (9) [6-16]	-11.46 (11) [6-20]	4-45
	2-Naphthyl	-11.81 (7) [3-21]	-11.78 (8) [2-24]	-12.06 (6) [4-10]	-12.07 (6) [3-11]	2-24
 2,4-Diazole_3R	Phenyl	-12.71 (3) [1-8]	-13.22 (2) [1-6]	-12.97 (2) [1-4]	-13.55 (1) [0.7-2]	0.7-8
	1-Naphthyl	-16.06 (0.1) [0.04- 0.3]	-15.95 (0.1) [0.04-0.4]	-16.33 (0.08) [0.05-0.1]	-16.31 (0.08) [0.04- 0.15]	0.04- 0.15
	2-Naphthyl	-15.34 (0.2) [0.1-0.6]	-15.31 (0.2) [0.07-0.7]	-15.62 (0.17) [0.1-0.3]	-15.67 (0.16) [0.08- 0.29]	0.07- 0.7
 2N30-Isoxazole_4R	Phenyl	-9.43 (81) [29-222]	-9.71 (60) [19-191]	-9.67 (63) [39-105]	-9.98 (46) [25-86]	19-222
	1-Naphthyl	-10.18 (38) [14-105]	-10.28 (35) [11-111]	-10.42 (30) [18-49]	-10.54 (27) [14-50]	11-111
	2-Naphthyl	-10.80 (21) [7-56]	-10.62 (25) [8-77]	-11.04 (16) [10-26]	-10.90 (18) [10-34]	7-77

 2O3N-Oxazole_4R	Phenyl	-10.03 (44) [16-122]	-10.49 (28) [9-88]	-10.27 (35) [21-57]	-10.77 (21) [11-39]	9-122
	1-Naphthyl	-10.11 (41) [15-112]	-10.11 (41) [13-129]	-10.35 (32) [19-53]	-10.38 (31) [17-58]	13-129
	2-Naphthyl	-10.97 (17) [6-47]	-9.82 (54) [17-171]	-11.22 (13) [8-22]	-11.30 (12) [7-23]	6-171
 2O-4,5N-Oxadiazole_3R	Phenyl	-7.90 (372) [135-1023]	-8.23 (268) [95-845]	-8.13 (295) [179-487]	-8.47 (209) [112-390]	95-1023
	1-Naphthyl	-9.88 (51) [19-141]	-9.85 (53) [17-167]	-10.13 (40) [24-66]	-10.12 (40) [22-75]	17-167
 3S5N-Thiazole_4R	Phenyl	-7.59 (508) [184-1399]	-7.99 (340) [108-1074]	-7.81 (404) [245-666]	-8.23 (267) [143-497]	108-1399
	1-Naphthyl	-9.89 (51) [18-140]	-9.62 (66) [21-209]	-10.14 (40) [24-65]	-9.89 (51) [27-94]	18-209
	2-Naphthyl	-8.62 (180) [65-496]	-8.81 (149) [47-470]	-8.86 (142) [86-235]	-9.07 (115) [62-215]	47-496
 Ar-2,3,4,5-Tetrazole_Glc-4R	Phenyl	-9.07 (115) [42-318]	-9.47 (77) [24-244]	-9.31 (91) [55-150]	-9.73 (59) [32-110]	24-318
	1-Naphthyl	-10.73 (22) [8-61]	-10.87 (19) [6-60]	-10.98 (17) [10-28]	-11.15 (14) [8-27]	6-61
	2-Naphthyl	-10.52 (27) [10-74]	-10.48 (28) [9-88]	-10.77 (21) [13-35]	-10.77 (21) [11-39]	9-88
 GlucoPyrrole_4R	Phenyl	-9.70 (61) [22-169]	-10.23 (36) [11-114]	-9.94 (48) [29-79]	-10.51 (27) [15-51]	11-169
	1-Naphthyl	-10.58 (26) [9-70]	-10.60 (25) [8-79]	-10.82 (20) [12-40]	-10.88 (19) [10-35]	8-79
	2-Naphthyl	-10.98 (17) [6-47]	-11.07 (16) [5-49]	-11.23 (13) [8-22]	-11.35 (12) [6-22]	5-49

^a Models as per Table 4.17 with the range for K_i values including RMSE given in square parentheses.

^bNI = No inhibition. Poses not generated for these structures.

The four models used gave quite similar $\ln K_i$ predictions and a range incorporating the RMSE values was generated for each ligand. The ligand that was predicted to be the most potent biological inhibitor of GP was 2,5-diazole with a 1-naphthyl substituent (Figure 4.8), which had a predicted K_i range of 0.007-0.1 μM . The position of the NH of this ligand heterocycle has the ability to form a favourable H-bond interaction with the His377 backbone and the heterocycle can also form interactions with Asp283 backbone O. The 2,4-diazole heterocycle was also predicted to be efficient with predicted K_i 's for phenyl of 0.7-8 μM , 1-naphthyl 0.04-0.15 μM and 2-naphthyl of 0.07-0.7 μM (Figure 4.8). This ligand also has the ability to form favourable H-bond interactions with His377 (3.58 Å), but primarily with Asp283 (2.88 Å for Model 3). 2,3-Diazole with the 2-naphthyl substituent (2-21 μM) has been predicted to be a good inhibitor and can again form favourable H-bond interactions with His377. Although thiazoles usually show relatively poor inhibition of GPb (Table 4.1), 2S4N-thiazole with the phenyl (5-62 μM), 1-naphthyl (4-45 μM) and 2-naphthyl substituent (2-24 μM) is predicted to be relatively potent μM inhibitors, further analysis shows the potential of the heterocycle to form a favourable H-bond with Asn284 NH, which could contribute to the low micromolar inhibition. Glucopyrrole with the 1-naphthyl (8-79 μM) and 2-naphthyl substituent (5-49 μM) also shows predicted low micromolar inhibition of GP, this ligand also having the ability to form a H-bond with His377. Other ligands that have been predicted to show low micromolar inhibition of GP include; 2,3,4,5-tetrazole with the 1-naphthyl (6-62 μM) and 2-naphthyl substituents (8-90 μM), 2,3,4-triazole with the 2-naphthyl substituent (7-71 μM), 2,5N-3O-oxodiazole with the 1-naphthyl substituent (10-97 μM), 2N3O-isoxazole with the 2-naphthyl substituent (7-77 μM), 203N-oxazole with the phenyl (9-122 μM) and 2-naphthyl substituent (6-171 μM).

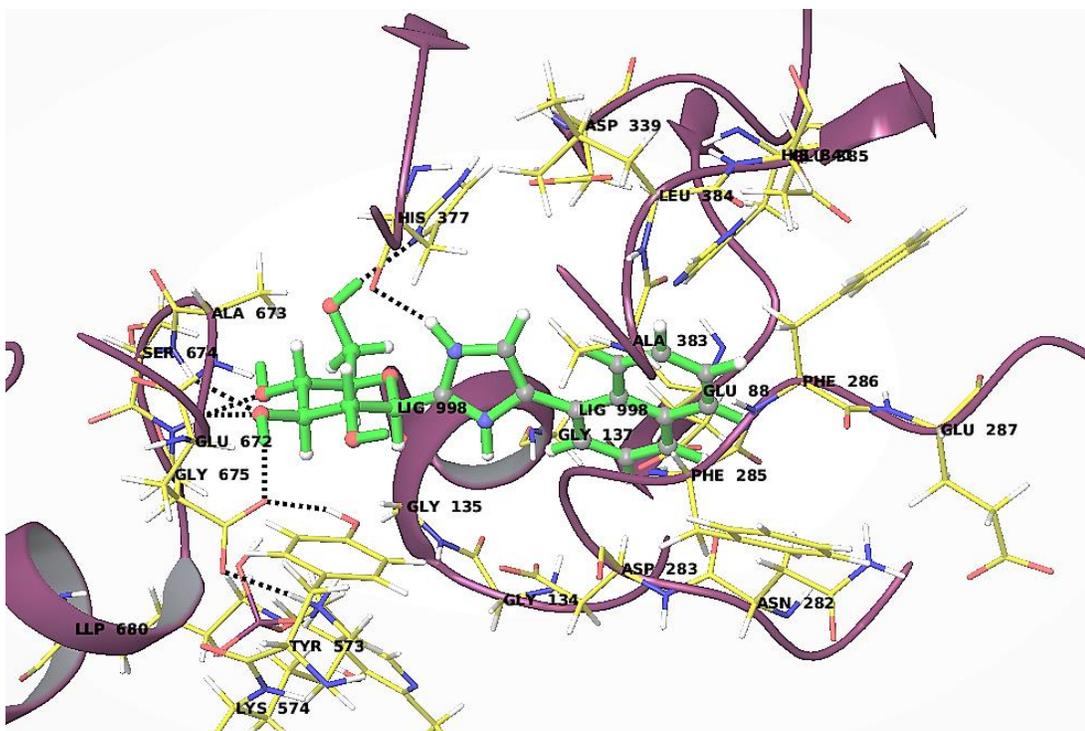


Figure 4.8. - Predicted binding pose of 2,5-Diazole using model 3.

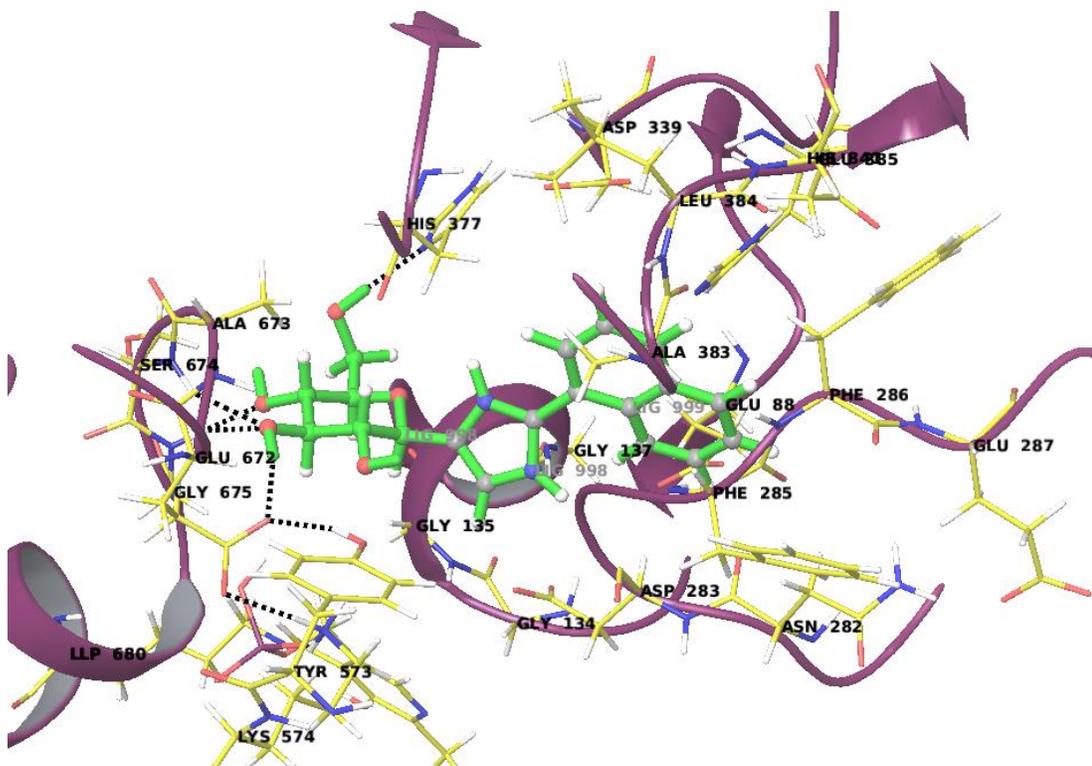


Figure 4.9. - Predicted binding pose of 2,4-Diazole using model 3.

4.5 Conclusion.

Computational drug design has many important and useful applications and gives information regarding the inhibition of GP for the treatment of T2D. By controlling many of the different parameters that effect ligand-protein binding, including exploration of solvation effects, atomic constraints of the glucose core and it's adjoined hydrogen atoms as well as the exploration of the two varying catalytic site structures based on residues Asn282, Asp283, Asn284 and Phe285, we were able to see how these different factors contribute to the binding of potential inhibitors at the GPb catalytic site. Firstly, solvation of the enzyme using the cluster waters algorithm using a cut-off distance of 0.5 Angstroms revealed that over 100 explicit water molecules were found in common across relevant solved GP-inhibitor complexes and 5-7 common waters in the catalytic site. In general, the correlations achieved for both Glide-SP, -XP and post-docking Prime MM-GBSA are better with a receptor containing no water molecules. Overall, the atomic constraints applied to the glucose analogues on the glucose ring itself and the hydrogen atoms on the hydroxyl groups of glucose produced experimentally consistent poses like that of known ligands in the protein complex with only a few exceptions. The different residue orientations displayed interesting features when binding. Docking calculations to the **ClusterResidueOut** receptor performed worse in general than the other receptors, indicating that there may be a crucial interactions for favourable binding with the Asn284 residue, which points away from the catalytic site in this orientation. However, this same protein conformation outperforms the others when the receptor has 5 Å flexibility; here, the receptor has the ability to adopt a geometry similar to those in the more common receptors. The **GeneralClusterProtein** and the **ClusterResidueIn** featured the same residue orientation and correlations achieved for Prime MM-GBSA post-docking are impressive and similar, with the **GeneralClusterProtein** achieving a slightly higher correlation demonstrating that the more common orientation found across the 10 proteins (Code: 5JTU, 5JTT, 3G2I, 3G2K, 1XL0, 1XL1, 5LRC, 5LRF, 5LRE and 5LRD) contribute to a better correlation. A comprehensive statistical evaluation of the relationship between predicted and experimental activities (ln Ki) was performed for all four models

generated from the **GeneralClusterProtein**. The predictive capabilities of these models are significant and give excellent statistical values which provide considerable insight into the inhibitory potential of the unknown compounds. This is the first time that an MM-GBSA model(s) has been demonstrated to show such predictive capability for the GPb catalytic site. Applying the statistical models indicate that some unknown ligands have low micromolar/nanomolar inhibition such as 2,5-diazole with the 1-naphthyl substituent (0.007-0.1 μM) as well as 2,4-diazole the phenyl (0.7-8 μM), 1-naphthyl (0.04-0.15 μM) and 2-naphthyl substituents (0.07-0.7 μM). These compounds are therefore worthy of future synthesis and in vitro kinetics experiments.

Chapter 5 - Synthetic Methods

5.1 Introduction

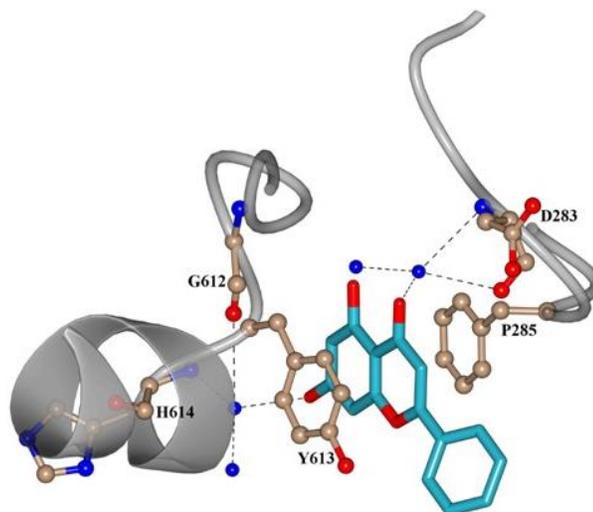


Figure 5.1 – Binding of the flavone chrysin at the GP inhibitor site (PDB code: 3EBO). A sandwich type complex has formed with chrysin in between residues Phe285 and Tyr613 which exploit π - π stacking interactions and a number of water bridging (blue) interactions.^[59]

The flavone nucleus is an important scaffold, since both natural and synthetic derivatives draw a large variety of biological and pharmacological activities; including antitumor, anti-inflammatory, antiviral and antioxidant properties.^[111] Interest in flavonoids as a potential treatment for T2D is growing, with some flavonoids, such as chrysin, inhibiting GP by binding to the caffeine binding site, found at the surface of GP roughly 12 Å from the catalytic site.^[68] Interactions are made through the flavone skeleton as the A & C rings are able to form π - π stacking with the two aromatic residues within the inhibitor site: Phe285 and Tyr613 (Figure 5.1). Binding at this site leads to stabilisation of the T-state conformation (section 2.6), blocking access to the catalytic site. Previous computational screening performed by Chetter, B. et al^[110] has predicted that a subgroup of flavonoids known as flavones can bind to the inhibitor site with some efficacy. The previous computational screening performed ranked a series of 5,7-dihydroxy flavones based on their expected performance, with the only difference being the substituents of the

B ring (Figure 5.2). In this project, attempts will be made to synthesise a number of these flavones which are predicted to bind at the inhibitor site. Future kinetic studies will determine whether these compounds are biologically active.

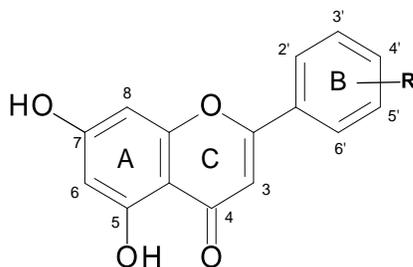
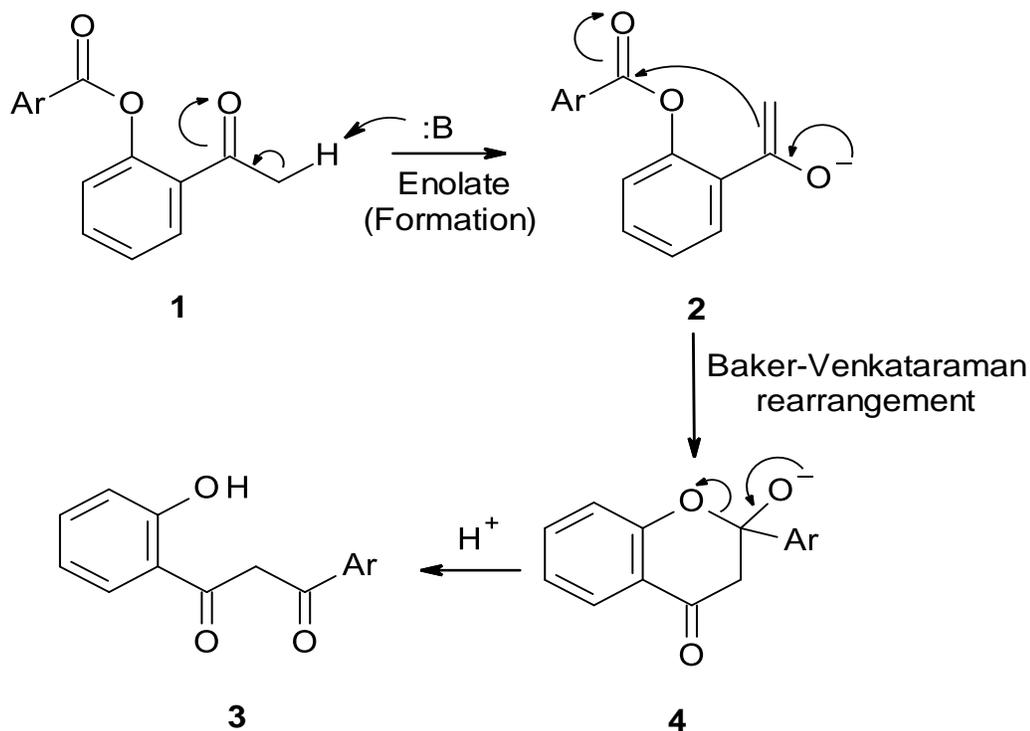


Figure 5.2 - Basic backbone structure of flavone, showing the hydroxyl groups at the 5 and 7 position.

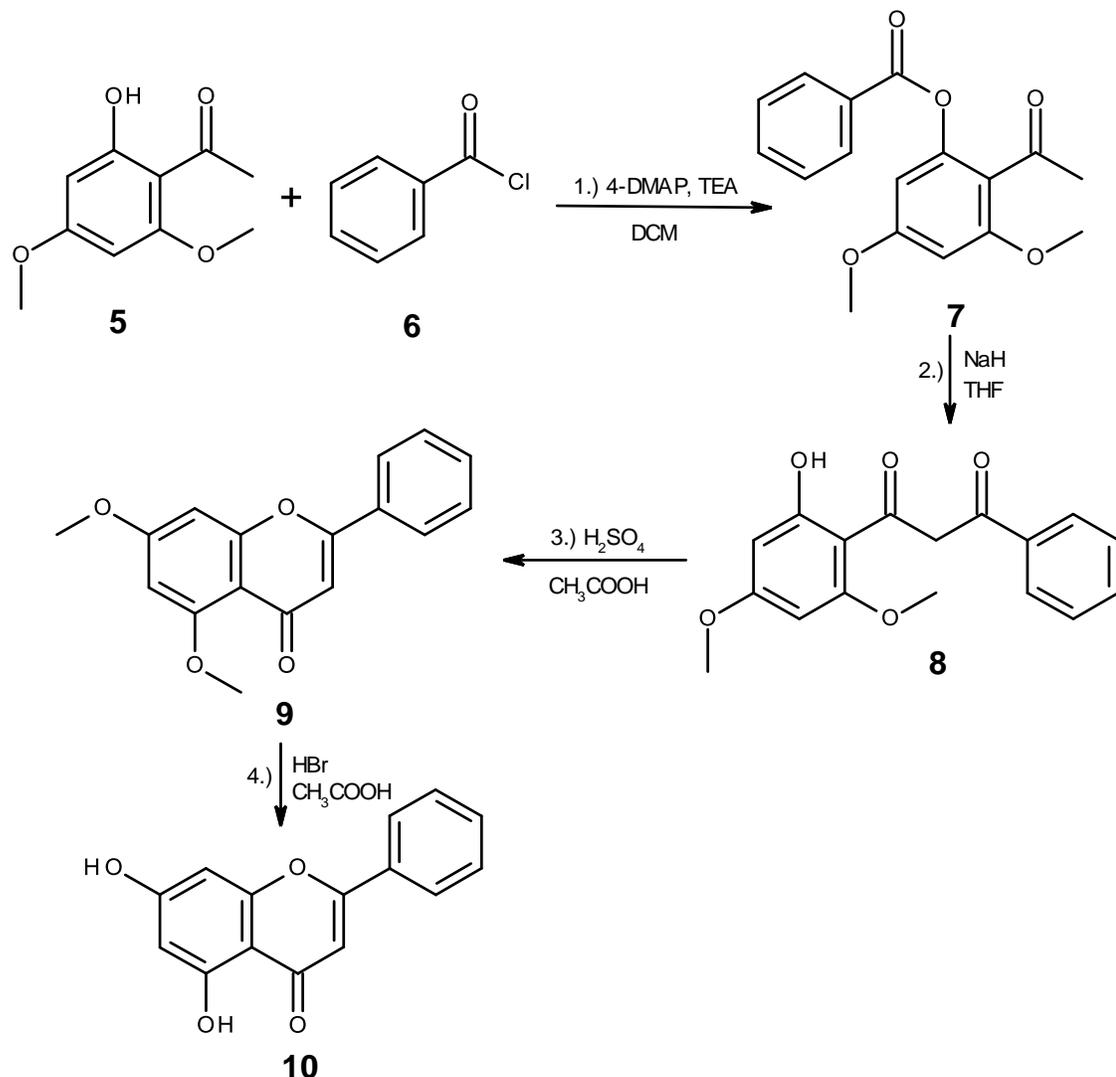
Many synthetic methods exist to synthesize these phenolic structures whilst keeping the hydroxyl group at the 5 and 7 position intact (Figure 5.2). Most procedures start with the A and B ring, joined by a carbon chain, then cyclized into the C ring. This section explores the different methods of synthesizing flavones.

5.2 Baker-Venkataraman Rearrangement.



Scheme 5.1 – The mechanism involved in the Baker-Venkataraman rearrangement.

One of the most commonly used methods to obtain flavones involves the cyclization of 1, 3-diphenylpropane-1,3-diones or 2'-hydroxychalcones which are prepared from 2'-hydroxyacetophenone (**5**) and benzoylating agents (**6**) or benzaldehydes. Exploited numerous times in the preparation of flavones, the Baker-Venkataraman Rearrangement (BVR) has been shown to be effective in the preparation of flavones.^[112, 113] 2'-Hydroxyacetophenones are converted into benzoyl esters, (**7**) and the BVR is used under basic conditions to form 1, 3-diphenylpropane-1, 3-diones (**8**) which, following cyclization with sodium acetate or sulphuric acid in acetic acid, yield the flavones (**9**). Further steps can be added in the synthesis for more diverse compounds; for instance, demethylation (**10**), (Scheme 5.1) oxidation or reduction. The next method of synthesis looks at forming the same β -diketone intermediate structure without exploiting the Baker-Venkataraman rearrangement.

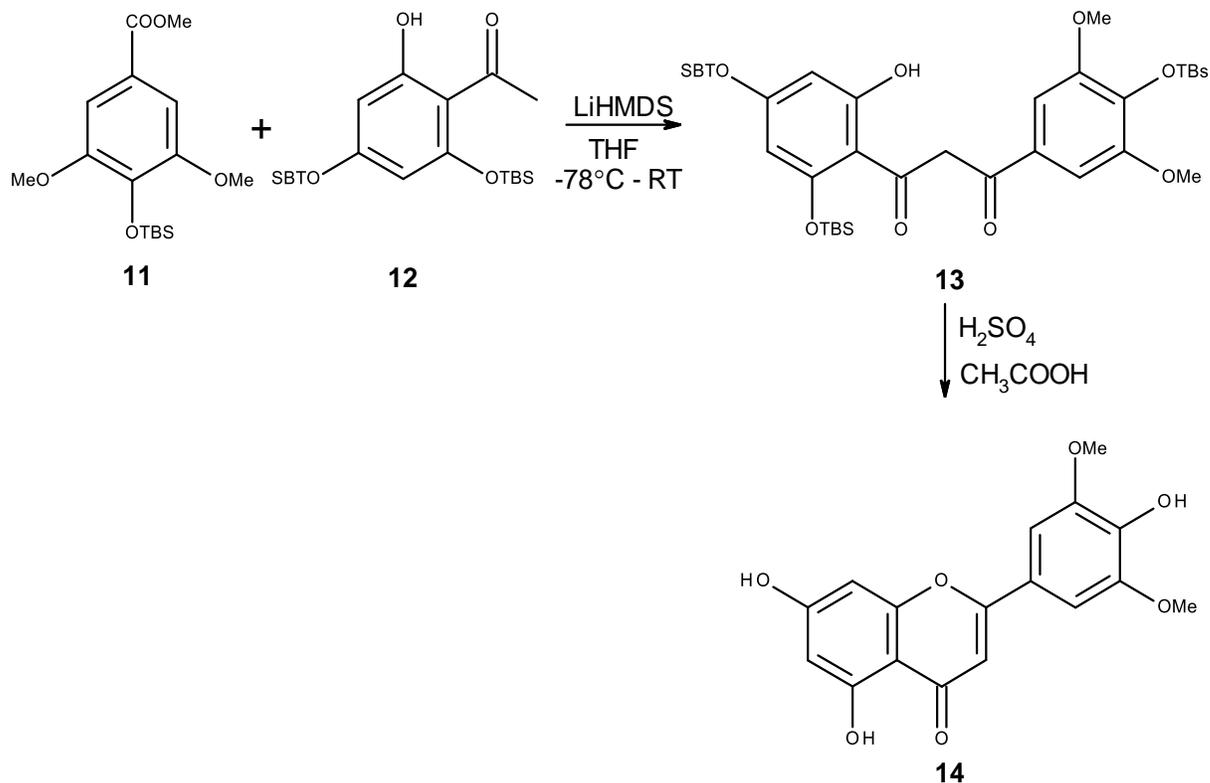


Scheme 5.2 – A general method of synthesizing 5,7-dihydroxyflavone exploiting the Baker-Venkataraman rearrangement.

5.2.1 β -diketone cyclization

Tricin, a 5,7-dihydroxyflavone derivative was prepared by M. Ninomiya, et al^[114] with a reasonable yield of 68%. The synthesis involved the condensation reaction of 4-*O-tert*-butyldimethylsilyl-3,5-dimethoxy-benzoate (**11**) and 1.5 equivalents 2',4'-*O*-bis(*tert*-butyldimethylsilyl)-6'-hydroxyacetophenone (**12**) using 8 equivalents of lithium bis(trimethylsilyl)amide (LiHMDS) in THF at -78°C raised to room temperature over 3 days to form the β -diketone (**13**) as a mixture of tautomers. These were subjected to acid cyclodehydration and deprotection with 0.5% H₂SO₄

in acetic acid at 100°C overnight to yield tricrin (**14**). The next method of synthesis will explore the use of a microwave assisted synthesis, known for an efficient, 'greener' synthesis.



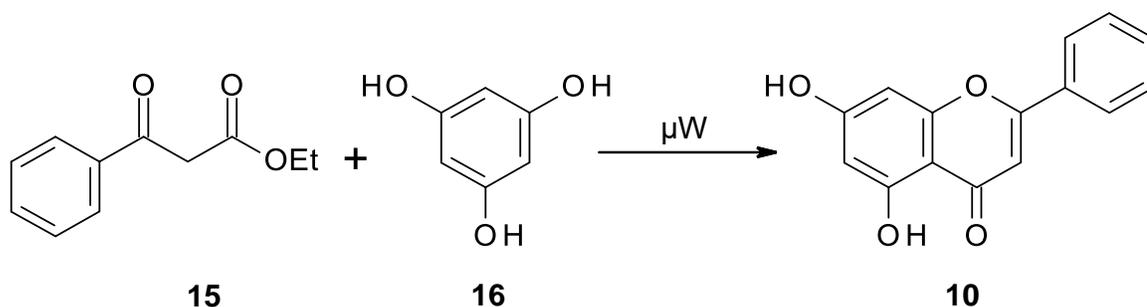
Scheme 5.3 – Synthesis of tricrin (**14**) through using a β -diketone intermediate.

5.3 Microwave Synthesis

Synthesis following a β -diketone pathway is the most commonly used method of synthesis for flavone derivatives, but an extremely efficient synthesis has been found using microwave radiation of similar β -diketones; which has numerous advantages over thermal methods. Microwaves provide high efficiency of heating, low operating costs, purity in final product, reduction in unwanted side reactions and often a reduced use of solvents.^[115]

Kabalka. G.W, et al reported achieving impressive yields of over 90% for numerous flavone derivatives using a microwave-assisted synthesis.^[116] This method of microwave synthesis involves irradiation of a 1,3-diketone to yield the desired

flavone. To the reaction 0.1 mmol CuCl₂ and 1.0 mmol of the 1,3-diketone in 3ml of ethanol was subjected to microwave irradiation at 80°C at 100W for 5 minutes. The product was separated and purified using flash column chromatography. A solvent free microwave method (Scheme 5.3) was investigated by Seijas. J.A, et al achieved by irradiation of phloroglucinol (**16**) and β-ketoesters (**15**),^[117] such as ethyl benzoylacetate as opposed to the benzoyl chloride and o-hydroxyacetophenone. Due to the solvent free synthesis, a substantially more powerful microwave was used, heating to 240°C at 800W with irradiation times ranging from 3-12 minutes.

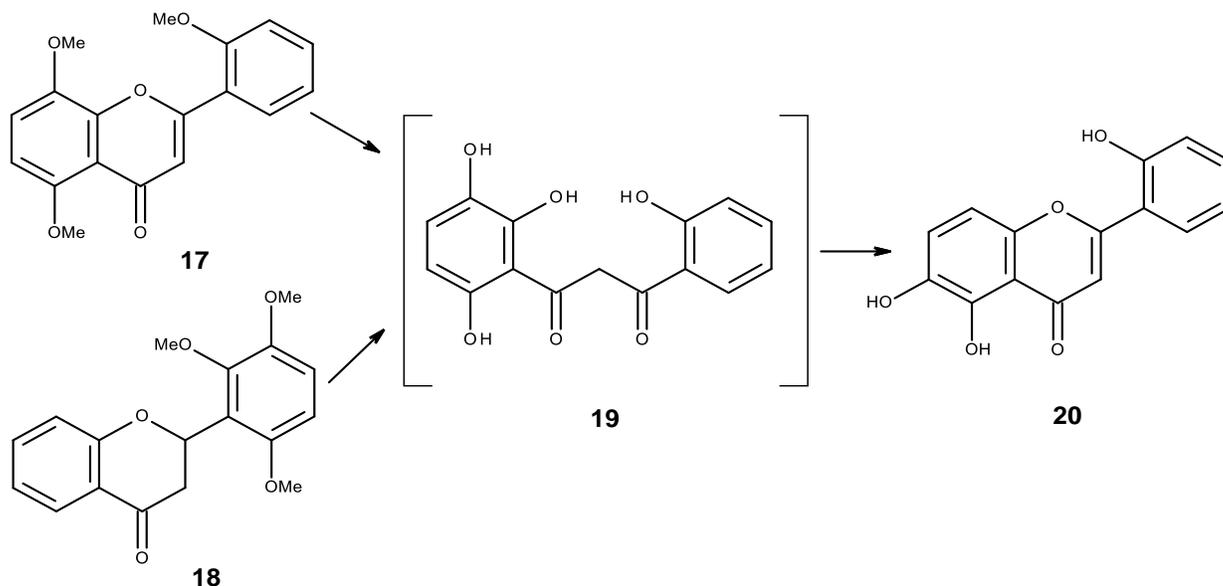


Scheme 5.4 – Flavone synthesis using highly the efficient Microwave synthesis.

Solvent free microwave syntheses are highly efficient as well as safer, as there are no harmful solvents, acids or bases associated with this synthesis method. High yields can be achieved, with Seijas. J.A, et al achieving yields of 96% when synthesising chrysin (**10**).^[117] Microwave synthesis is often a fast process, with some reactions taking several minutes as opposed to some thermal methods taking several hours, making for a more energy efficient process. The next synthetic pathway looks at the rearrangement of 2'-methoxyflavones under drastic conditions to give diverse structures.

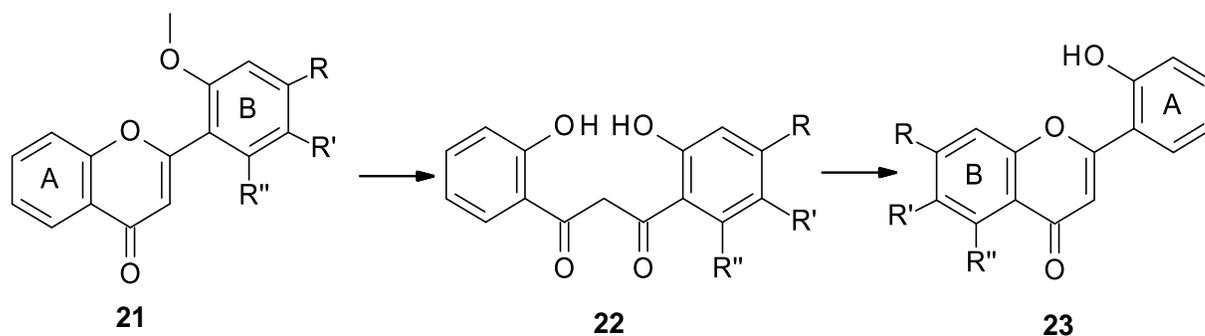
5.4 Demethylation rearrangement

Rearrangement reactions are used extensively in organic chemistry as formation of the desired structural isomer may be difficult to achieve using other synthetic methods.



Scheme 5.5 – Rearrangement of 2'-methoxyflavones upon demethylation.

Gallagher, K, et al found that certain 2'-methoxyflavones are rearranged during demethylation (Scheme 5.5) by hydroiodic acid under sufficiently drastic conditions to give the related 2'-hydroxyflavones in which the 2-phenyl group (B-ring) and the fused aromatic ring (A-ring) of the original flavone are interchanged.^[118] Their suggestion is that an intermediate disalicyloylmethane (**19**) occurs in the rearrangement. Both 5,8,2'-trimethoxyflavone (**17**) and the 2',3',6'-isomer (**18**) give 5,6,2'-trihydroxyflavone (**20**) on treatment with hydroiodic acid.



Scheme 5.6 – The general scheme showing the 2-phenyl group and the fused aromatic ring interchanging upon demethylation ^[119]

Table 5.1 – Table showing the different products forming from demethylation of (21)

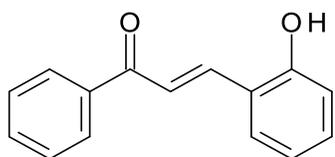
Reactant (21)				Product (22)			
Name	R	R'	R''	Name	R	R'	R''
2':5'- Dimethoxyflavone	H	OMe	H	6':2'- Dihydroxyflavone	H	OH	H
2':4'- Dimethoxyflavone	OMe	H	H	2':4'- Dihydroxyflavone	OH	H	H
2':6'- Dimethoxyflavone	H	H	OMe	2':6'- Dihydroxyflavone	H	H	OH
2':4':6'- Trimethoxyflavone	OMe	H	OMe	2':4':6'- Trihydroxyflavone	OH	H	OH

2-Methoxyflavones corresponding to (21) were prepared for demethylation experiments from the Baker-Venkataraman method from the appropriate o-arylacetophenones with satisfactory yields which were demethylated with hydroiodic acid under suitable conditions to form the different compounds seen in Table 5.1.

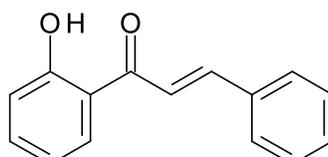
All previous methods explored in this chapter have exploited the use of a β -diketone intermediate, the next method of synthesis involves the cyclisation of the less commonly used single ketone intermediate to yield flavones.

5.5 Single Ketone synthesis

Reaction pathways involving single ketone groups as an intermediate have been observed in the synthesis of flavones. The single ketones are derivatives of 2'-hydroxychalcone (**24**) (Figure 5.3) or 1-(2-hydroxyphenyl)-3-phenyl-2-propenone (**25**) (Figure 5.4) and these structures are able to undergo cyclocondensation, which forms the C ring of the flavone.



24

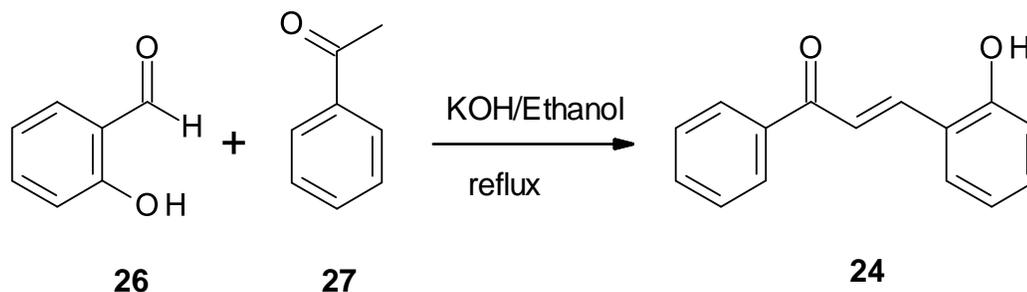


25

Figure 5.3 - 2'-hydroxychalcone

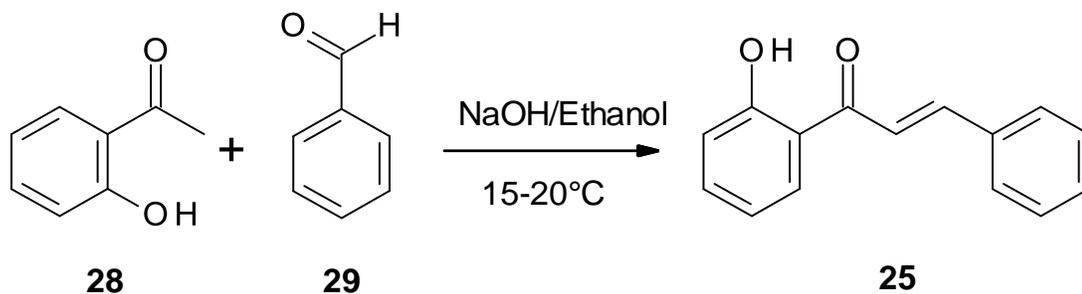
Figure 5.4 - 1-(2-hydroxyphenyl)-3-phenyl-2-propenone

Both of these compounds are usually prepared under basic conditions. The synthesis of 2'-hydroxychalcone (**24**) was achieved by Sashidhara. K.V, et al;^[120] through the reaction of salicylaldehyde (**26**) and acetophenone (**27**) with 10% aq. potassium hydroxide and ethanol at reflux via the Claisen-Schmidt condensation, a variation of the aldol condensation.



Scheme 5.7 – Synthesis of 2'-hydroxychalcone

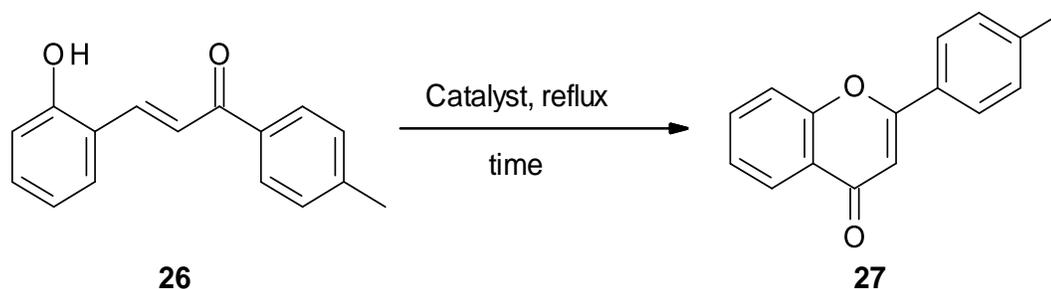
1-(2-hydroxyphenyl)-3-phenyl-2-propenone (**25**) was prepared by Zambare. A.S, et al using sodium hydroxide in ethanol at 15-20°C in an aldol condensation as seen in Scheme 5.8.^[121]



Scheme 5.8 – Synthesis of 1-(2-hydroxyphenyl)-3-phenyl-2-propenone

The cyclocondensation of 2'-hydroxychalcone (**24**) was investigated by Sashidhara. K.V, et al with the use of different catalysts and solvents. Many catalysts were explored, as well as the time under reflux (Table 5.2).^[120]

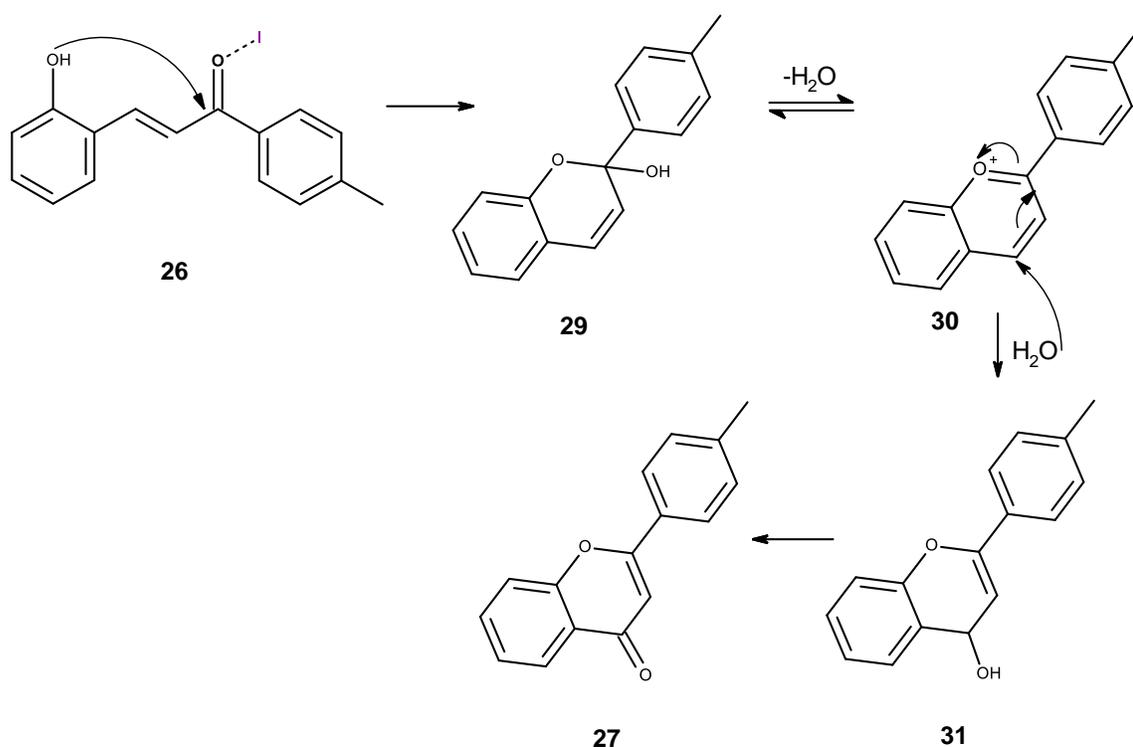
Table 5.2 – Table showing the use of different catalysts including reaction time and the effect this has on the yield in synthesising 4-methoxyflavone from (**26**).^[120]



Entry	Catalyst	Time (h)	Yield (%)
1	Aluminium chloride	2.5	28
2	Zinc chloride	3.0	10
3	Boron trifluoride diethyl etherate	3.0	0
4	Tin (II) chloride dehydrate	2.0	54
5	Mercury (II) chloride	3.0	5
6	Iron (III) chloride	3.0	5
7	Tin (IV) chloride	3.0	0
8	Iodine	1.5	74

It was reported that iodine as a catalyst gave the best results in terms of yield and reaction time for this reaction, reporting that the use of a 10 mol % of iodine was most favourable. Surprisingly, the solvent used had a large impact on the reaction and the absence solvent achieved the most impressive yield of 74% with the use of the iodine catalyst (Table 5.2). However, this is still not the highest yield reported thus far.

Sashidhara. K.V, et al has suggested a mechanism using iodine as a catalyst. Intramolecular cyclisation results in a hemiacetal species (**29**), which gets converted to more reactive flavylum ion (**30**). Water molecules then attack on the more reactive 3-position of (**30**) to form adduct (**31**), that on oxidation in the presence of iodine yields flavone (**27**).^[120]

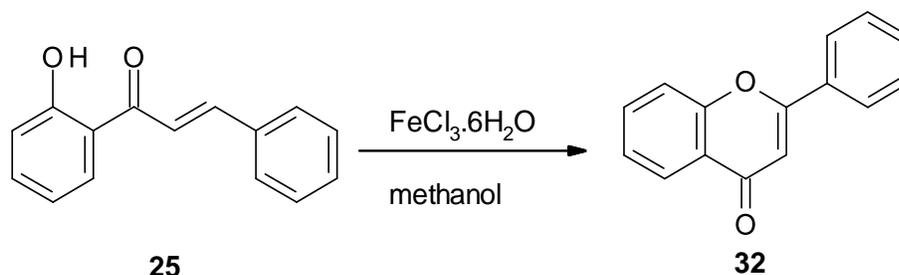


Scheme 5.9 – The synthesis of 4'-methoxyflavone using an Iodine catalyst without the use of a solvent.

Table 5.3 – Table showing the use of different solvents and the effect this has on yield in synthesising 4-methoxyflavone from **(26)**^[120]

Entry	Solvent	Yield (%)
1	1,4-Dioxane	56
2	Acetonitrile	0
3	Ethanol	33
4	Methanol	12
5	Tetrahydrofuran	15
6	Dimethyl sulfoxide	20
7	Neat	74

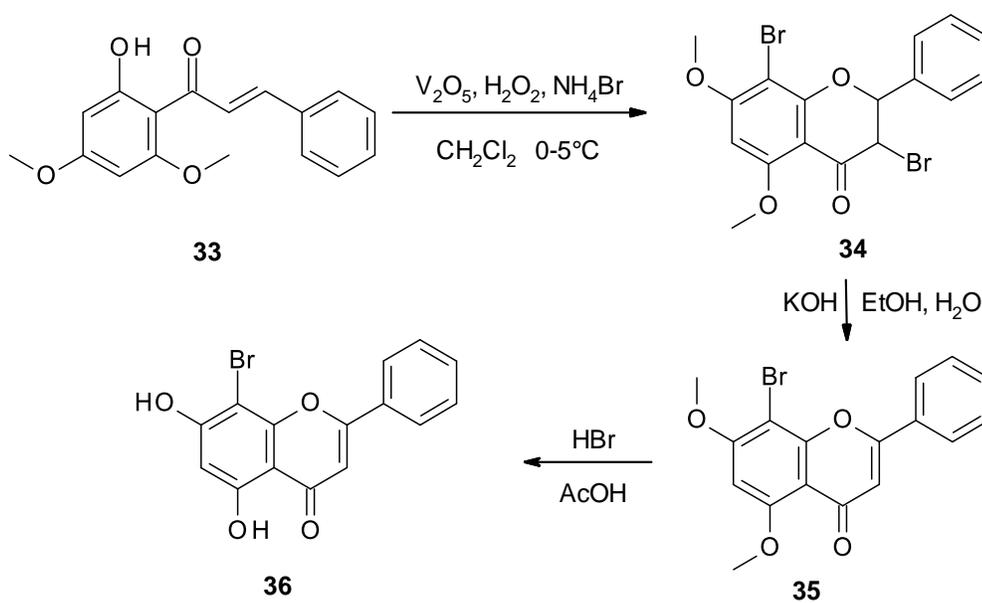
H. Kumar, et al ^[122] found that using iron (III) chloride hexahydrate ($\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$) in methanol at reflux was a viable procedure for the cyclocondensation of 1-(2-hydroxyphenyl)-3-phenyl-2-propenone (**25**) to form a flavone (**32**) (Scheme 5.10), ethanol was also found to be an effective solvent, but this synthesis seems to be unsuccessful in non-alcohol solvents (dimethylformamide (DMF), dimethylsulphoxide (DMSO), tetrahydrofuran (THF), toluene and ethyl acetate). $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$ in methanol acquired a 55% yield, with a reaction time of 10 hours and in comparison, this method is therefore less efficient than others previously discussed.



Scheme 5.10 – cyclocondensation of 1-(2-hydroxyphenyl)-3-phenyl-2-propenone to give **(32)** as a product.

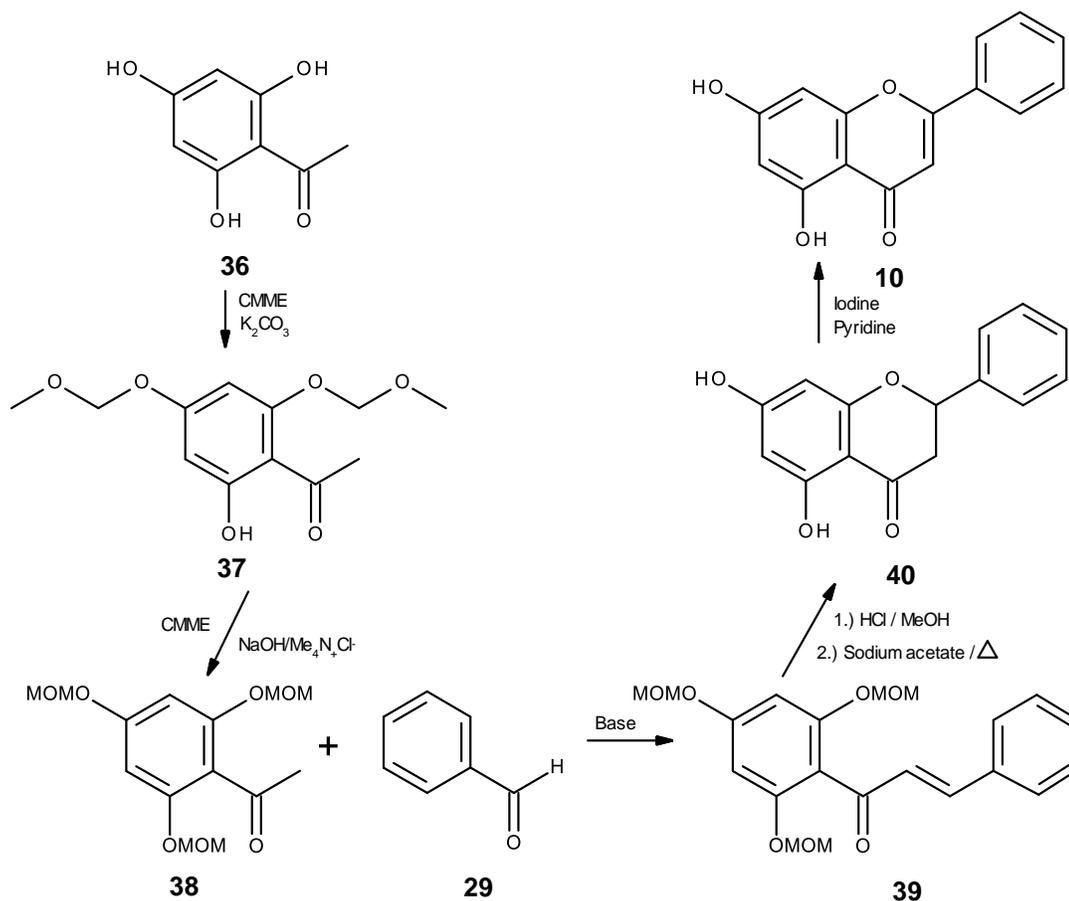
Formation of the C ring using ethanol has been observed repeatedly; Zambare. A, et al^[121] have used 10 mol% oxalic acid in ethanol to give the cyclised flavone (**32**)

with reported yields of 95% over 6 hours. Methanol with 10 mol% oxalic acid was also used to give the cyclised flavone to give slightly less yields of 92% after 6.5 hours, therefore validating the aforementioned solvent choice in this reaction.^[122] Using oxalic acid as a catalyst allowed for a wider range of solvents to be used, unlike the iron (III) chloride hexahydrate route; acetonitrile over 12 hours gave an 85% yield and THF over 10 hours gave an 88% yield. The oxalic acid method is therefore more efficient than that of the iron (III) chloride procedure. Both of these methods are both less favourable than that of the greener microwave assisted methods mentioned earlier.



Scheme 5.11 – The synthesis of 8'-bromochrysin from 2'-hydroxy-4',6'-dimethoxychalcone.

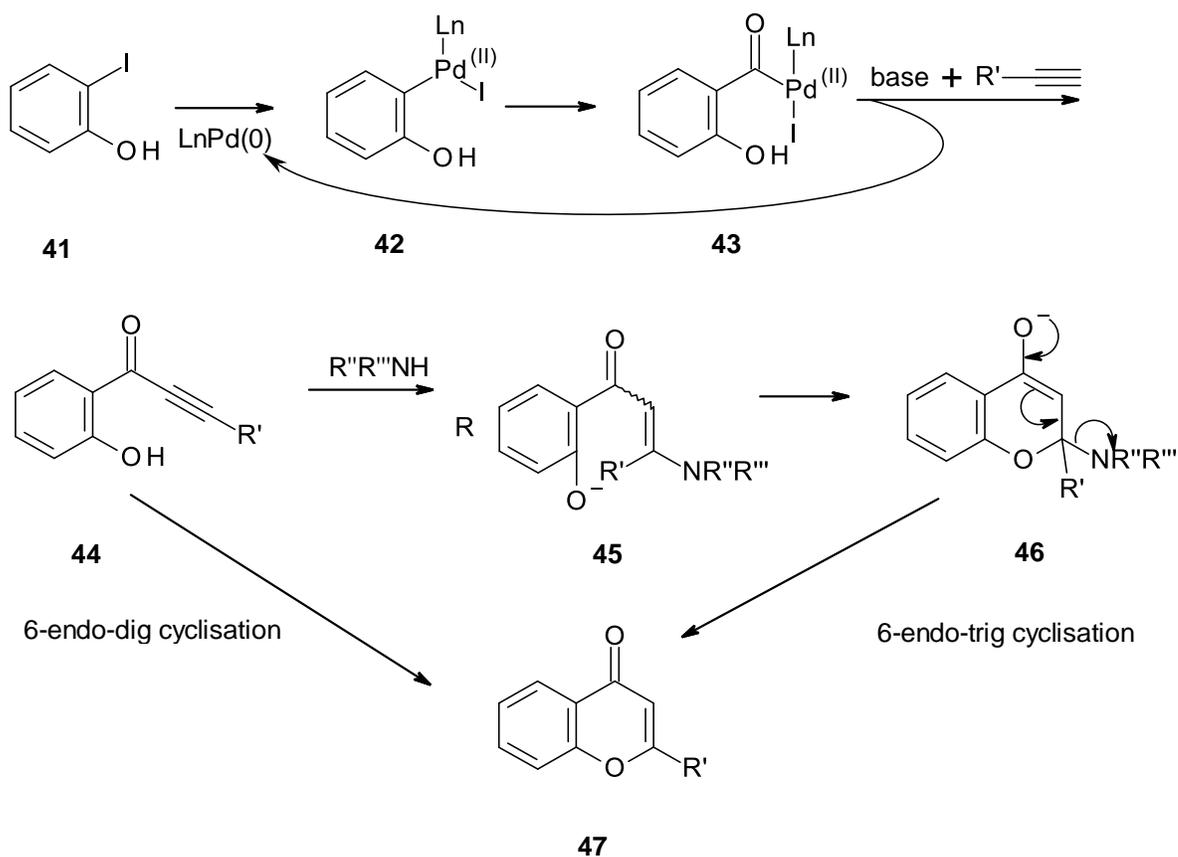
Khan. A, et al observed the synthesis of the cyclised product with the use of ammonium bromide and vanadium pentoxide with hydrogen peroxide. This method achieved a modest yield of 62%; this method is therefore more inefficient in comparison to other aforementioned methods, and also introduces a bromine at the 8-position of the A-ring (34), which required further synthetic manipulation to remove, such as exploiting a Pt(III) surface for hydro-debromination.^[123] Demethylation of the methoxy groups to yield the corresponding hydroxyl groups must also be performed via the use of HBr with AcOH at reflux for 24 hours, as reported by Yenjai. C, et al^[124] whom achieved a 95% yield using such an approach.



Scheme 5.12 – Method of synthesizing flavones with the use of protecting groups.

Mavel, S, et al synthesised a series of 5,7-dihydroxyflavone derivatives via an enone intermediate (Scheme 5.12).^[125] The trihydroxyacetophenone (**36**) we protected in a step-wise manner using methoxymethyl (MOM) chloride and K_2CO_3 , the third addition of the MOM group required the addition of a stronger base, NaOH, and the phase transfer catalyst tetrabutyl ammonium chloride. MOM-ether protected hydroxyacetophenone (**28**) was then transformed into the chalcone derivative (**39**) through a base-catalysed condensation with benzaldehyde (**29**). Hydrolysis of MOM-protected groups were performed with dilute HCl in methanol followed by cyclization (**40**) with solid sodium acetate under reflux. Oxidation of the crude intermediate with iodine in pyridine led to the target flavones (**10**).

5.6 Metal Catalysed Synthesis.



Scheme 5.13 – The synthesis of flavones with the use of a palladium catalyst.

The use of a palladium-based catalyst has been explored to synthesise flavone-type compounds. For example, Miao, H, et al achieved high yields under basic conditions under a carbon monoxide atmosphere.^[126] This catalytic approach starts with iodophenol (**41**), which undergoes an oxidative addition to form the organometallic intermediate complex (**42**) (Scheme 5.13). Carbon monoxide is then inserted to give the carbon monoxide insertion product (**43**); the desired R group is then introduced, with base, attached to an alkyne carbon-carbon triple bond (**44**). For the synthesis of flavones, the R group must be aromatic. This compound can then undergo a 6-endo-dig cyclisation which is a direct path to form the desired flavone (**47**). Alternatively, the former compound can undergo a Michael addition, to apply the diethylamine ($\text{NR}''\text{R}'''$) group, forming alkene (**45**); this compound can then undergo a 6-endo-trig cyclisation (**46**) which then rearranges to form the desired flavone. The

drawback of this method lies with the possibility of compound (**44**) to rearrange to form a palladium compound. This issue was resolved by Torii and Kalinin using an excess of diethylamine to form (**45**). The conditions required for this reaction require heating the reaction to 120°C under a pressure of 20kg/cm², which is impractical with regards to previously the previously aforementioned methods.

5.7 Conclusion

T2D is a widespread illness affecting millions of people worldwide. Flavones present themselves as potentially promising treatment of T2D through inhibition of the GP caffeine binding site. Synthetic methods presented in this chapter all represent viable options available for the synthesis of flavones, and can be used for the synthesis of chrysin and other 5,7-dihydroxy flavone analogues. However, synthesis using the Baker-Venkataraman rearrangement is well understood and can be used to yield diverse flavone structures that have undergone previous computational screening by a collaborator and are predicted to be potentially active against the caffeine binding site. Kinetic experiments can be performed on the synthesised compounds to determine their activity against GP. These kinetic results can then influence future drug design as structure activity relationships can be used to closely refine and then predict future ligands for synthesis which aim for greater biological activity.

Chapter 6 – Synthetic Results

6.1 - Introduction

Type 2 diabetes (T2D) is a chronic illness that is becoming more prevalent, with a significant increase of those diagnosed with diabetes from 1980 to 2016, rising from 108 million to approximately 420 million people worldwide; this is predicted to rise to above 640 million by the year 2040 at the current rate.^[3] Multiple adverse health effects arise from T2D and sufferers are at a higher risk of developing other health complications as a direct result, which could otherwise be avoidable; such as angina, cardiac failure, strokes retinopathy, renal replacement therapy and amputations.^[127] Diabetes mellitus is characterized by high levels of glucose in the bloodstream (hyperglycemia), peripheral insulin resistance and low insulin production in the pancreas.^[128] Glycogen Phosphorylase (GP) has been identified as a key enzyme in T2D, as it catalyses the first step in the intracellular degradation of glycogen to give α -D-glucose-1-phosphate (G-1-P), via the glycogenolysis pathway, shown in Figure 6.1. Inhibition of this enzyme will therefore reduce glucose production, making it a promising target for T2D treatment. Studies in both diabetic and non-diabetic rats have shown that flavonoids have anti-hyperglycemic properties as they inhibit glycogen metabolism and are known to bind at multiple sites on GP (*c.f.* Chapter 2.6-2.8).^[129, 130]

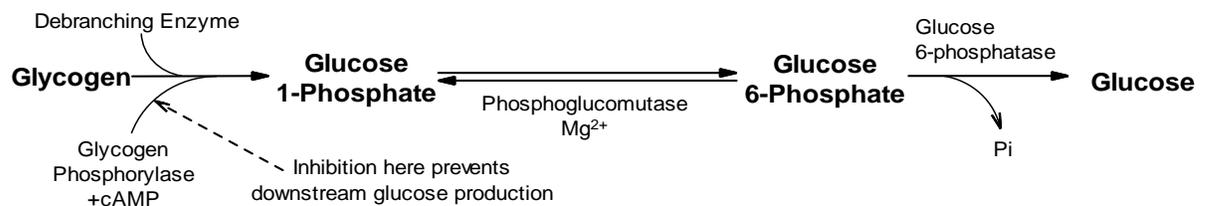


Figure 6.1 – The glycogenolysis pathway.^[131]

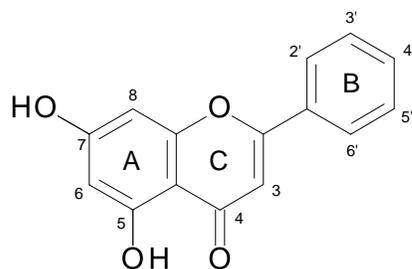
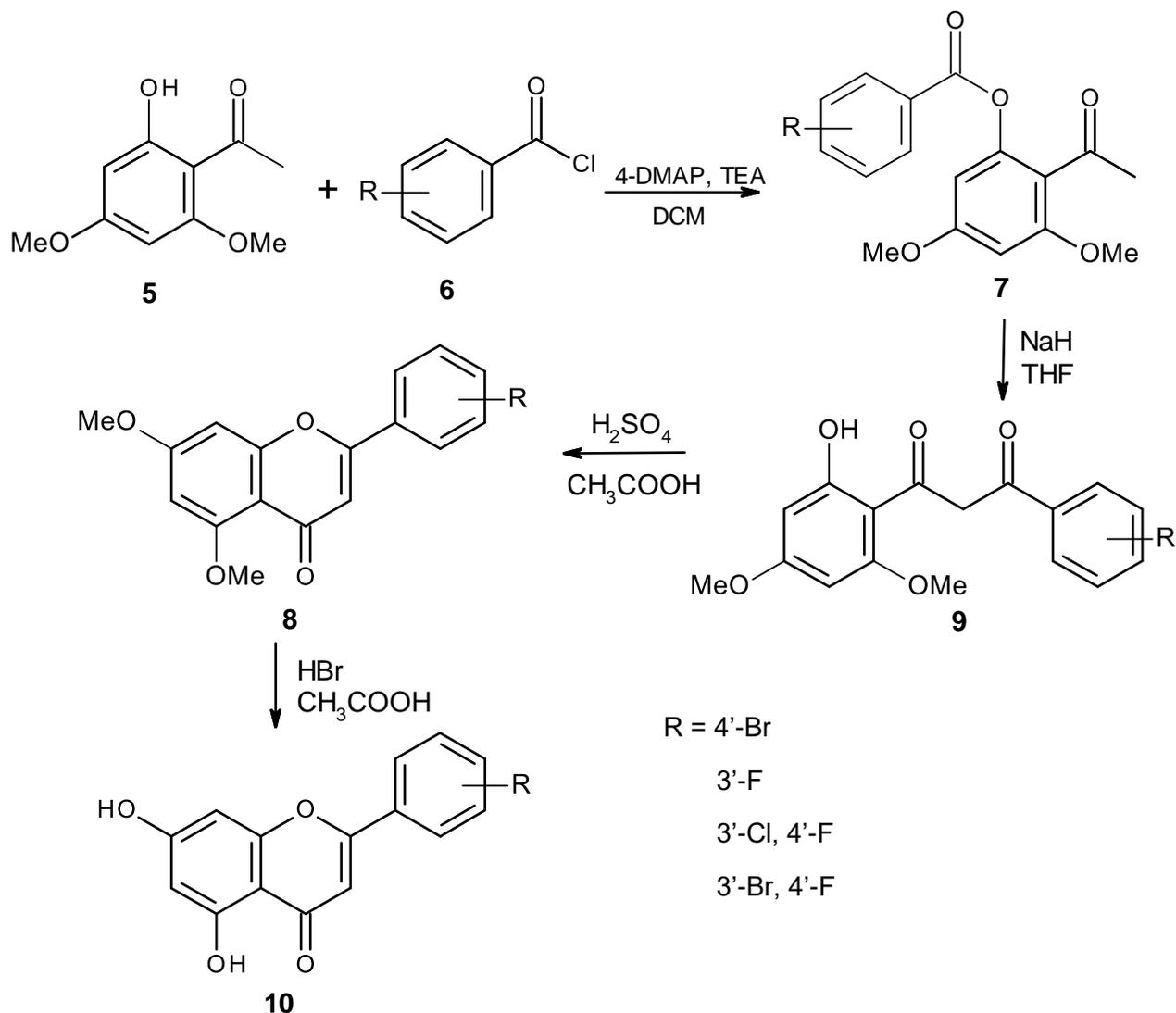


Figure 6.2 – The basic structure of a flavone. Hydroxyl substituents are shown at the 5 and 7 position.

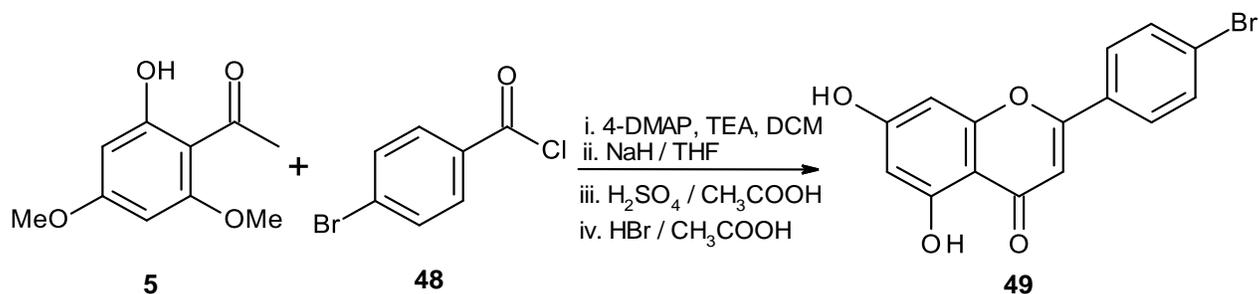
Many synthetic pathways have been reported for the synthesis of flavonoids; The focus on this work are a group of flavonoids known as flavones and the possible biological effect of different substituents on the B ring may have on GP at the caffeine binding site. There are many synthetic routes and methods available which can yield the flavone structure with the desired substituents on the B. For GP inhibition, hydroxyl groups on the 5 and 7 position of the A ring are required to increase bioavailability (Figure 6.2), and as such, synthetic routes must be followed which allow for the protection of the hydroxyl groups in order to stop undesired products from forming. Chetter. B, reported an efficient synthetic method (Scheme 6.1),^[110] capable of producing diverse 5',7-dihydroxy flavones in relatively high yields with the use of 2'-hydroxy-4',6'-dimethoxyacetophenone as a starting material, which allows for reactions of the lone hydroxyl group, whilst leaving the methoxy groups intact, which can later be removed, yielding the di-hydroxyl compound. 2'-Hydroxy-4',6'-dimethoxyacetophenone is reacted with a benzoyl chloride, with varying substituents in an esterification mechanism, followed by the Baker-Venkataraman rearrangement; the resultant product is then cyclised and then the methoxy groups removed to form hydroxyl groups at the 5 and 7 position.



Scheme 6.1 – The method used to synthesise a number of 5',7'-dihydroxy flavones.

The method outlined above was used to synthesise a number of flavone analogues (see R-groups) as these compounds were predicted through previous computational screening to show inhibitory potential against GP (Section 5.1), however, there were limits found with this method when synthesising certain derivatives, and as a result, other synthetic methods were explored further on which are explained later in this chapter.

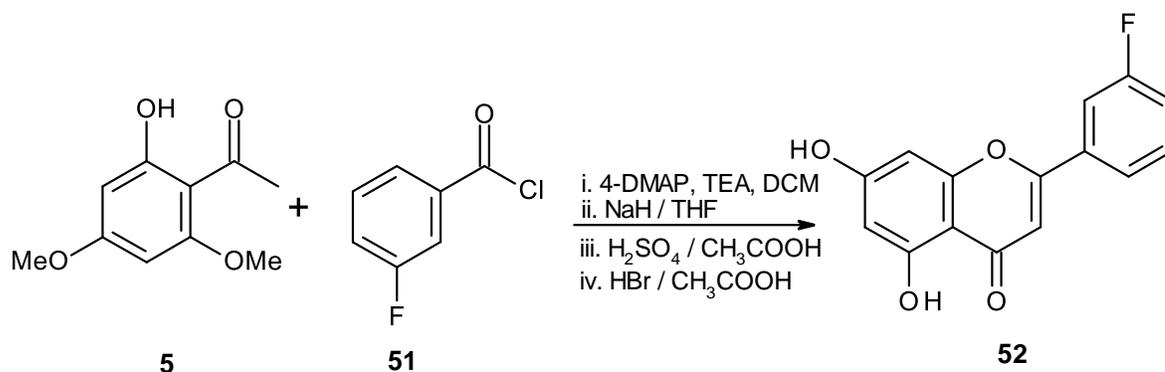
6.2 Attempted Synthesis of 2-(4-bromophenyl)-5',7'-dihydroxy-chromen-4-one



Scheme 6.2– The method used to synthesise 2-(4-bromophenyl)-5',7'-dihydroxy-chromen-4-one

This reaction followed the synthetic pathway mentioned in Scheme 6.1 The ester was formed using 2-hydroxy,-4'6'-dimethoxyacetophenone (**5**) which was added to 1.1 equivalents of 4-bromobenzoyl chloride (**48**) in a mixture of dichloromethane (DCM), 4-dimethylaminopyridine (4-DMAP) and trimethylamine (TEA), this was stirred under a nitrogen atmosphere at room temperature until the reaction was determined complete by TLC. The product was acidified with hydrochloric acid (HCl). The organic layer was washed with distilled water and brine then dried (MgSO₄) and isolated *in vacuo* leaving a pale yellow solid. The Baker-Venkataraman rearrangement was then performed with equivalent amounts of sodium hydride (NaH) in tetrahydrofuran (THF), which was refluxed until deemed complete by TLC. This was extracted with ethyl acetate (EtOAc) and washed with HCl, water and brine then isolated *in vacuo* to yield a yellow solid. This was then cyclised using (sulphuric acid) H₂SO₄ in acetic acid (CH₃COOH), heated at reflux for 15 hours until completion, as determined by TLC. This was poured over an ice/water slurry, and vacuum filtration was used to collect a pale yellow solid. The demethylation was performed with hydrobromic acid (HBr) which was heated at reflux for 40 hours until completion, as determined by TLC. This was poured over an ice/water slurry, and vacuum filtration was used to collect the desired compound as a yellow solid obtaining a 25% yield (**49**), which was washed with water.

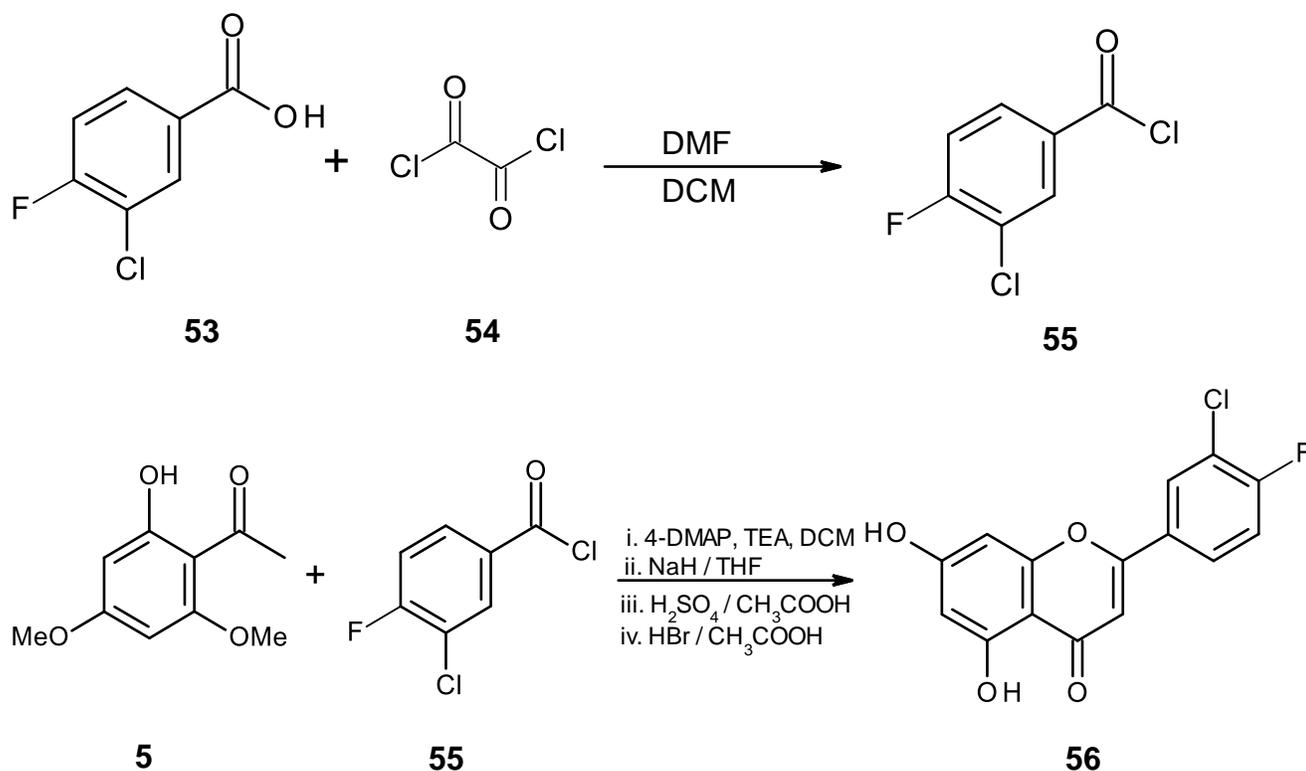
6.3 Attempted Synthesis of 2-(3-fluorophenyl)-5',7'-dihydroxy-chromen-4-one.



Scheme 6.3 – The method used to synthesise 2-(3-fluorophenyl)-5',7'-dihydroxy-chromen-4-one.

The same pathway that was highlighted in Scheme 6.1 was also used to synthesise the meta-para flavone analogue, using the same reagents with the exception of the benzoyl chloride; where 1.1 equivalents of 3-fluorobenzoyl chloride (**51**) was used in the esterification step instead. The Baker-Venkataraman step was performed with an equivalent amount of NaH in anhydrous THF heated at reflux until completion, as determined by TLC. This was then acidified with HCl and extracted with EtOAc, washed with water and brine and then dried (MgSO₄). This was then isolated *in vacuo* to yield a brown crystalline solid. The cyclisation step performed with H₂SO₄ and CH₃COOH yielded a crude product which was recrystallized from hot ethanol to yield the desired compound. The demethylation step using HBr and CH₃COOH also required less time to reach completion, and after 20 hours, the reaction was determined complete by TLC. The reaction was neutralised with NaOH and extracted with EtOAc, washed with water and brine, and isolated *in vacuo* to yield a brown crystalline solid as pure compound with an overall yield of 22%. (**52**).

6.4 Attempted Synthesis of 2-(3-chloro-4-fluoro-phenyl)-5',7'-dihydroxy-chromen-4-one

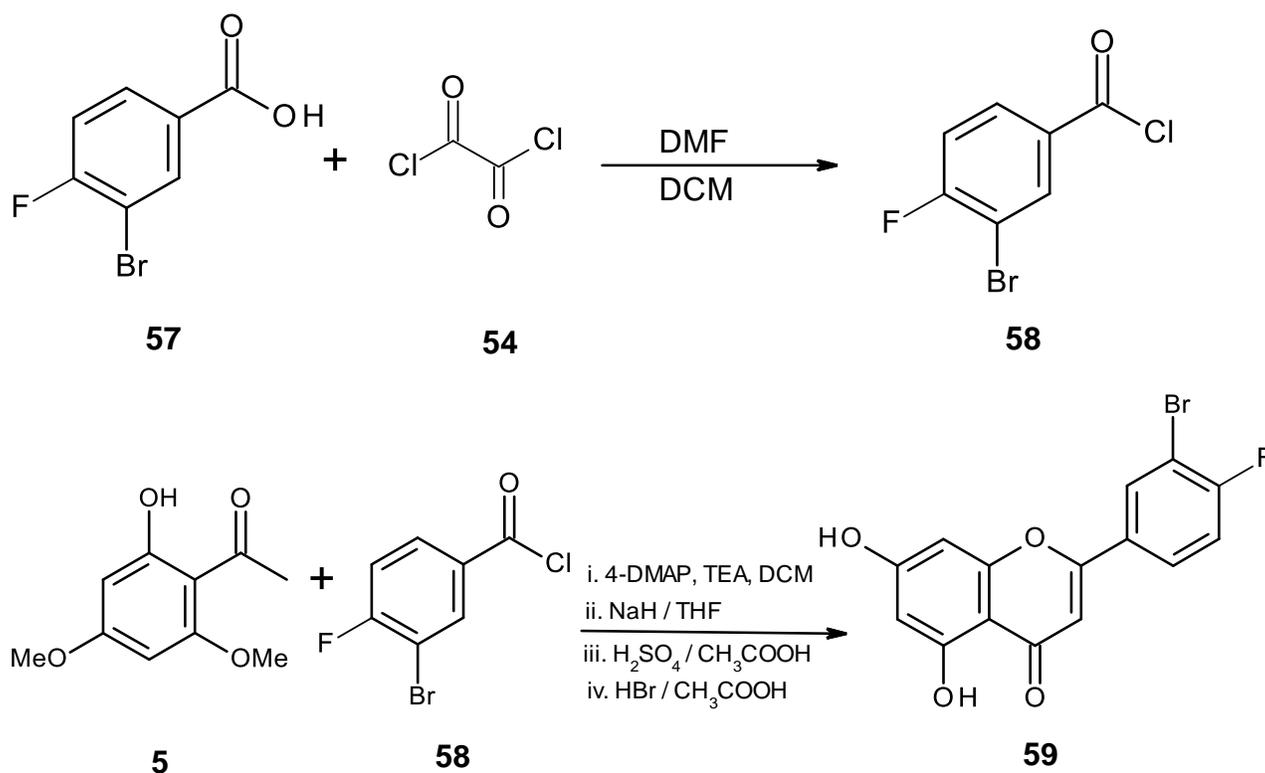


Scheme 6.4 – The method used to synthesise 2-(3-chloro-4-fluoro-phenyl)-5',7'-dihydroxy-chromen-4-one

The method used to synthesise the 3'-chloro,4'-fluoro analogue (**56**) was the same as mentioned previously in (Scheme 6.1) with the exception of the first step, where the acid (**53**) had to be transformed into the corresponding acyl chloride (**55**) using 1.1 equivalents of oxalyl chloride (**54**) and 2 drops of dimethylformamide (DMF) in DCM until completion, as determined by TLC. The product was isolated *in vacuo* to yield a viscous yellow oil (**55**). The esterification step was performed with 1.1 equivalents of the benzoyl chloride, and after 24 hours, there was still presence of (**5**) and so flash column chromatography was to isolate the ester as a pale brown crystalline solid. The Baker-Venkataraman step was again performed with equivalent amounts of NaH in anhydrous THF and heated at reflux until completion, as determined by TLC. The solution was acidified using HCl, and vacuum filtration

was used to collect an orange crystalline solid. The cyclisation step was performed using an excess of H_2SO_4 in CH_3COOH heated at reflux until completion, and poured over an ice/water slurry, then vacuum filtration was used to collect a pale brown crystalline solid which was purified using flash column chromatography. Demethylation was achieved using an excess of HBr in CH_3COOH heated at reflux for 30 hours until completion, as determined by TLC. This was poured over an ice/water slurry and then vacuum filtration was used to collect a brown crystalline solid as pure product with an overall yield of 6%. (**56**).

6.5 Attempted Synthesis of 2-(3-bromo-4-fluoro-phenyl)-5',7'-dihydroxy-chromen-4-one



Scheme 6.5 – The method used to synthesise 2-(3-bromo-4-fluoro-phenyl)-5',7'-dihydroxy-chromen-4-one.

The method used to synthesise the 3'-bromo,4'-fluoro analogue (**59**) was the same as mentioned previously in Scheme 6.1, however, the acid (**57**) again had to be transformed into the corresponding acyl chloride (**58**) with the use of 1.1 equivalents

of oxalyl chloride (**54**), along with 2 drops of DMF in DCM, which were allowed to react until completion, as determined by TLC. The esterification was achieved with the use of 1.1 equivalents of the benzoyl chloride, as well as 4-DMAP, TEA and DCM which were allowed to react until completion and then purified using flash column chromatography, yielding pale yellow crystals. The Baker-Venkataraman rearrangement was achieved with equivalent amounts of NaH in anhydrous THF, heated at reflux for 20 hours until completion, as determined by TLC. The solution was acidified with HCl and then collected using vacuum filtration to yield a yellow solid. The cyclisation step was performed with excess H₂SO₄ in CH₃COOH, which was heated at reflux until completion, as determined by TLC. This was poured over an ice/water slurry and vacuum filtration was used to collect a pale brown crystalline solid. Demethylation was performed using an HBr in CH₃COOH heated at reflux for 40 hours until completion. This was poured over an ice/water slurry and then vacuum filtration was used to collect an orange crystalline solid with an overall yield of 8.5% (**59**).

6.6 The method limitations

Despite the successful synthesis of a number of flavone analogues using the described method (Scheme 6.1), there was a significant limitation encountered that had not been observed before, but which has subsequently been identified in a 2011 publication after a careful literature search on the proposed structure.^[132] When an *o*-chloro (or possibly any ortho-halogen) benzoyl chloride substituent was used to form the B-ring under the conditions of the general procedure, spectroscopic data such as ¹H-NMR spectroscopy of the products formed from these reactions revealed peaks which are characteristic of extra aromatic hydrogens, as well as C-NMR showing peaks characteristic of extra aromatic carbons and an extra carbonyl peak.

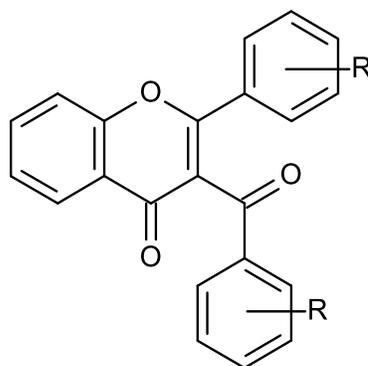
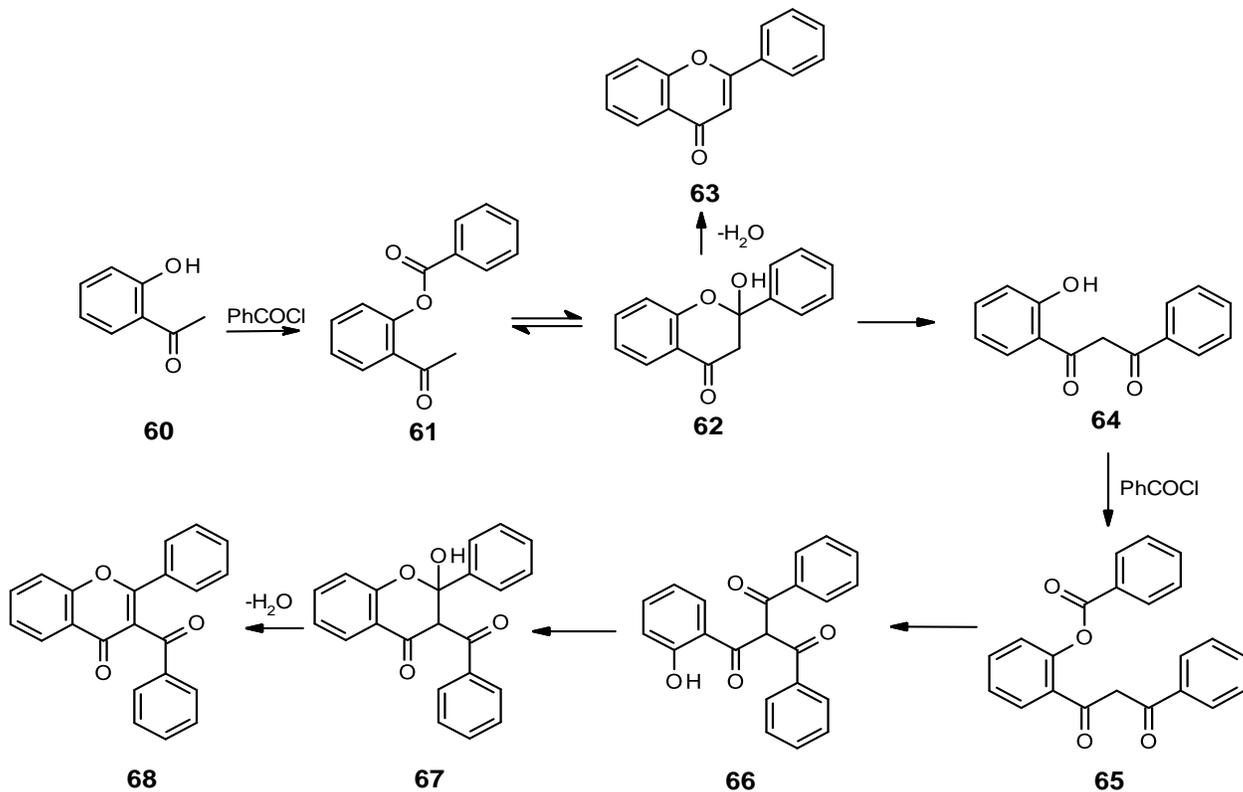


Figure 6.3 – The structure of 3'-hydroxyflavone

Chee, C.F, et al^[132] observed the formation of 3-benzoyl flavones (Figure 6.3) during a one-pot synthesis which can occur during either the esterification or the Baker-Venkataraman rearrangement in the presence of excess remaining benzoyl chloride. A mechanism was proposed by Chee, C.F, et al (Scheme 6.6); when 2'-hydroxyacetophenone (**60**) was treated with benzoyl chloride and base, addition of the first equivalent of benzoyl chloride produces 2'-benzoylacetophenone (**61**). In the presence of a base, the enolate of the acetyl group is formed and attacks the carbonyl of the ester to give the hemiacetal (**62**), which presumably undergoes ring-opening to give the β -diketone (**64**), the Baker-Venkataraman rearrangement product. In excess of benzoyl chloride, the phenolic group of the β -diketone undergoes esterification with another equivalent of benzoyl chloride to form benzoyloxydiketone (**65**). Rearrangement of the benzoyldiketone gives a triketone intermediate (**66**), which finally dehydrates upon cyclisation to produce the 3-benzoyl flavone (**68**) via a hemiacetal (**67**).^[132]

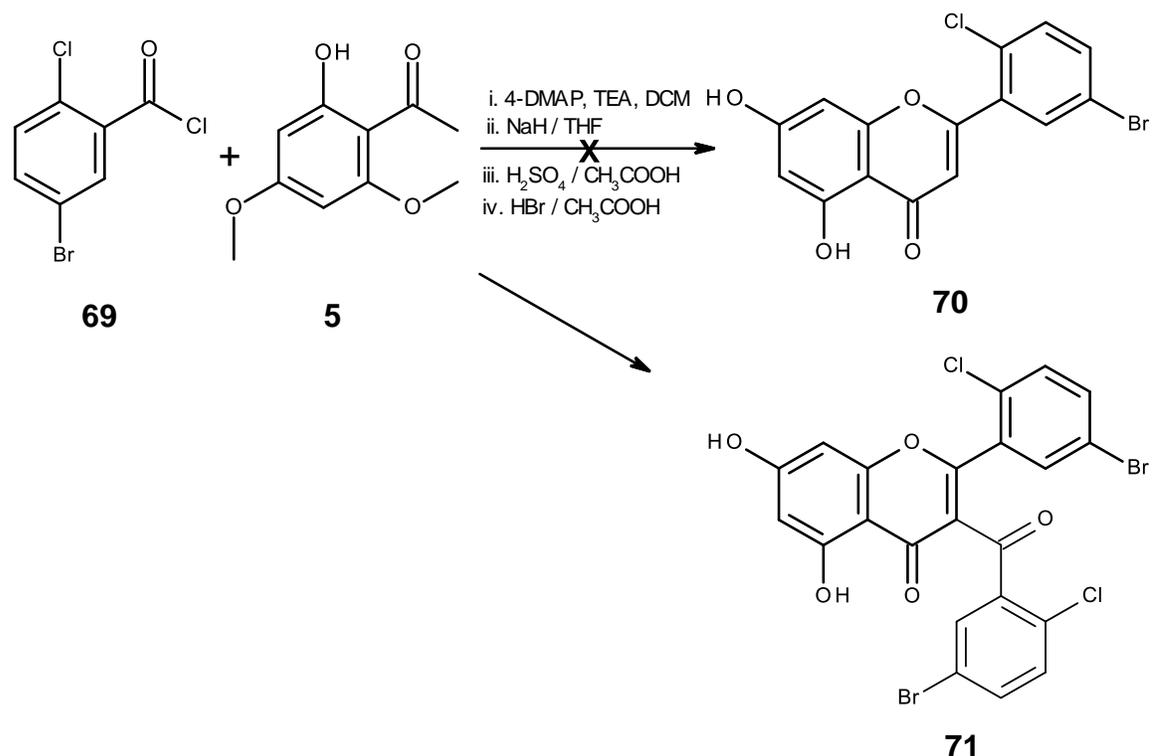


Scheme 6.6 – The proposed mechanism of the formation of 3'-benzoylflavone

Once the 3-benzoylflavone has formed along with the desired flavone, they are difficult to separate using recrystallisation or column chromatography as the polarities of the compounds are near identical. The formation of 3-benzoyl flavones were only observed when using *o*-chloro-benzoyl chlorides, showing that these substituents may have a higher affinity than other benzoyl chlorides in forming 3-benzoyl flavones, which could be due to the electronic or steric effects of the 2-chloro substituent. Although the reaction pathways followed in this project are different to those performed by Chee, C.F et al, remaining benzoyl chloride that hasn't been removed from the reaction may still undergo an esterification of the remaining alcohol group at the Baker-Venkataraman stage, and then rearrangement of the benzoyldiketone would ultimately lead to the formation of the 3-benzoyl flavone. Most synthetic attempts involving an ortho-chloro substituent don't lead to the formation of the pure 3-benzoyl flavone but instead form a mixture of the desired flavone and the unwanted structure. Formation of pure 3-benzoyl flavone are

produced in relatively low yields, as it is with excess benzoyl chloride that these products can form.

6.7 Attempted synthesis of 2-(5-bromo-2-chloro-phenyl)-5',7'-dihydroxy-chromen-4-one



Scheme 6.7 – The method used to synthesise 2-(5-bromo-2-chloro-phenyl)-5',7'-dihydroxy-chromen-4-one

The synthesis of the 2'-chloro,5'-bromo flavone compound was adapted since, once again, the starting 2'-chloro-5'-bromo benzoyl chloride is only commercially available as a carboxylic acid and therefore must be activated as its acid chloride (**69**). 2'-Chloro-5'-bromo benzoic acid was dissolved in DCM and stoichiometric amounts of oxalyl chloride was added followed by 2 drops of DMF. The reaction was performed at room temperature under a nitrogen atmosphere. Upon completion, the product was isolated *in vacuo* to yield a viscous yellow oil (**69**). 0.9 equivalents 2'-hydroxy-4',6'-methoxyacetophenone (**5**) was added in a mixture of DCM, 4-DMAP and TEA. The reaction was left until completion, as determined by TLC however a spot remained with similar *rf* values to 2'-chloro-5'-bromobenzoyl chloride which was

thought to be removable by the work-up. The organic layer was separated and washed, then concentrated under vacuum to leave a golden oil. LC-MS confirmed the presence of unreacted 2'-hydroxy-4',6'-methoxyacetophenone (**5**) and so aqueous NaOH was added to wash impurities from the product and no further purification was performed. Equivalent amounts of NaH was added along with anhydrous THF as a solvent. The reaction was heated to 120°C and allowed to stir overnight. After 20 hours, the solution was acidified and vacuum filtration was used to obtain a pale yellow crystalline solid. An NMR of the crude product were performed, and peaks characteristic of β -diketone enol tautomerisation was observed. Cyclization was performed with an excess of H₂SO₄ was added along with CH₃COOH. This was refluxed at reflux for 6 hours and poured over an ice/water to yield pale brown coloured crystals formed, which were collected through vacuum filtration. Demethylation was performed with excess HBr in CH₃COOH, and this was allowed to react for 20 hours until completion, as determined by TLC. This was poured over an ice/water slurry to yield brown crystals. The first NMR performed was indicative of two compounds formed, and this was primarily thought to be the presence of partially demethylated product. A TLC was performed using a different solvent system (2:8 acetone:petroleum ether) and the presence of two compounds were confirmed. The demethylation step was repeated for a further 20 hours and a TLC using the same solvent system showed the presence of one compound. This was poured over an ice/water slurry to obtain a small amount of pale brown crystals with an overall yield of 11.5% (**71**).

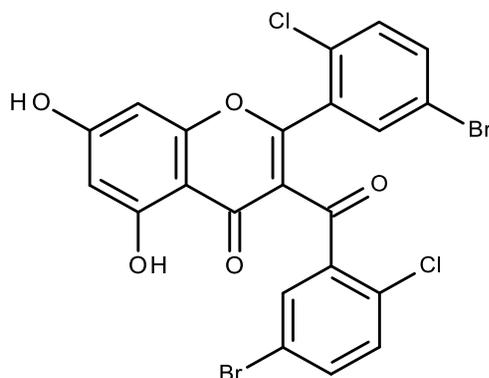


Figure 6.4 – The 3'-benzoylflavone structure formed from this reaction (**71**).

The H-NMR performed on this compound has an integration of 8H when the desired compound should integrate to 6H in the aromatic area. Identification of the desired flavone can be confirmed by a diagnostic H-NMR peak that typically appears at around 7ppm which indicates a hydrogen atom in the 3-position (Figure 6.4) adjacent to the ketone group; and this peak is no longer present when the 3'-benzoyl flavone forms. The C-NMR shows 22 carbon atoms, and the presence of two carbonyl groups, whereas the 5',7'-dihydroxyflavone only contains 15 carbons. There are no methyl peaks present on both the carbon and the proton NMR and the presence of hydroxyl groups integrate to 1H showing that the product had fully demethylated. An LC-MS was performed which showed a mw of 582, indicative of the compound as well as the isotope pattern of 2Cl and 2Br atoms in a singular molecule owing to the different isotope ratios of $^{35}\text{Cl}:$ ^{37}Cl and $^{79}\text{Br}:$ ^{81}Br

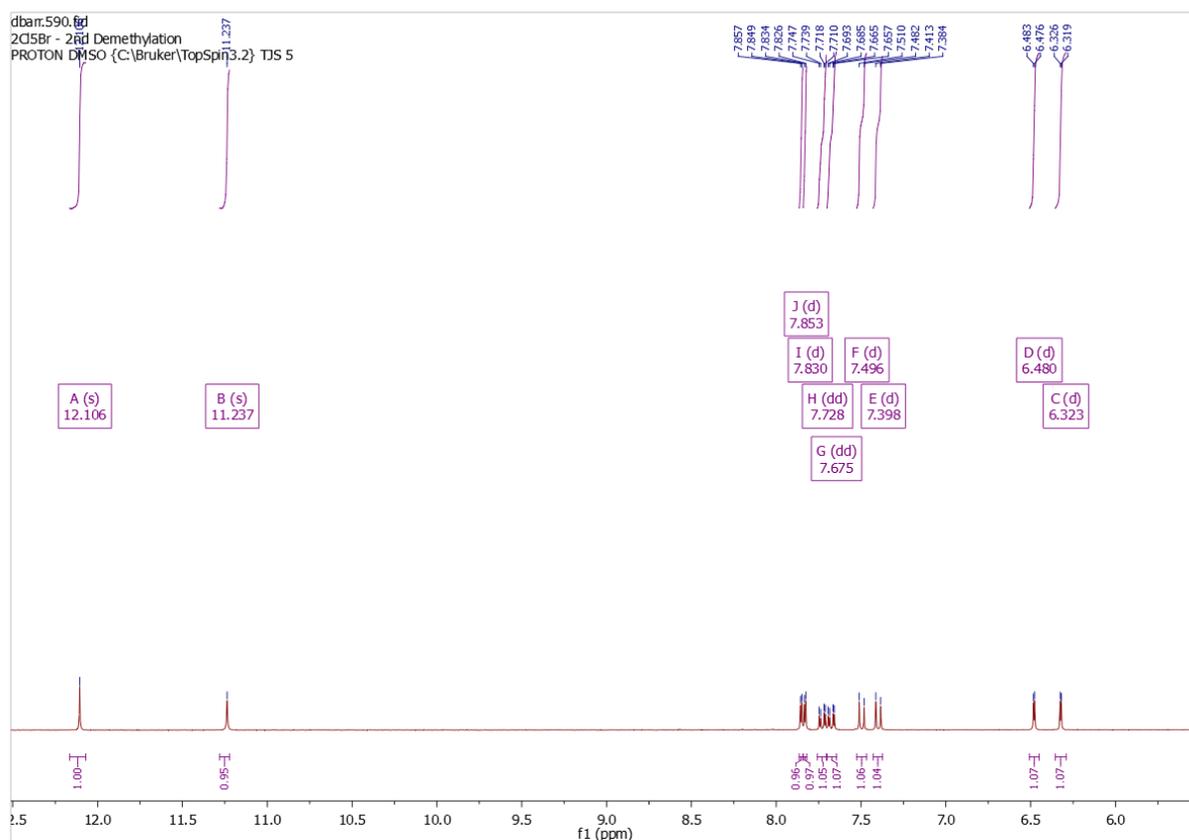


Figure 6.5 – $^1\text{H-NMR}$ of 3-(5-bromo-2-chloro-benzoyl)-2-(5-bromo-2-chloro-phenyl)-5,7-dihydroxy-chromen-4-one

6.8 Attempted synthesis of 2-(2-chlorophenyl)-5',7'-dihydroxy-chromen-4-one

Three separate attempts were made to synthesize 2-chloroflavone, with two synthetic attempts forming the unwanted 3-benzoyl flavone in differing amounts when reacted under different conditions; including the base used in the esterification and the Baker-Venkataraman rearrangement, temperature and reaction time.

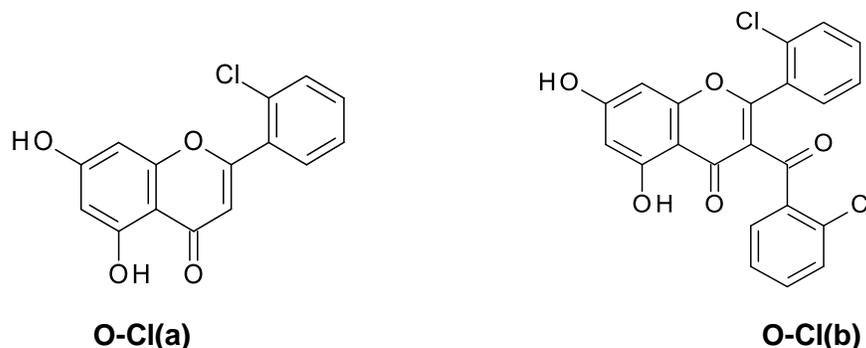


Figure 6.6 – The mixture of products formed in the attempted synthesis of 2-(2-chlorophenyl)-5,7-dihydroxy-chromen-4-one

Table 6.1 – Table showing the different conditions used in an attempt to synthesise 2-(2-chlorophenyl)-5,7-dihydroxy-chromen-4-one

Attempt	Esterification		Baker-Venkataraman rearrangement		Ratio O-Cl(a) : O-Cl(b)
	Base	Temperature (°C)	Base	Temperature (°C)	
1	4-DMAP/TEA	RT	NaH/THF	120	1.5:8.5
2	K ₂ CO ₃ H ₂ O/Toluene	60	C ₅ H ₅ N/KOH	120	-
3	K ₂ CO ₃ H ₂ O/Toluene	60	NaH/THF	110	4:1

6.8.1 First Attempt at Synthesis of 2-(2-chlorophenyl)-5',7'-dihydroxy-chromen-4-one

Attempt one of the synthesis followed the general procedure (Scheme 6.1) with 2'-hydroxy-4',6'-methoxyacetophenone and 2'-chlorobenzoyl chloride. This attempt involved the use of 4-DMAP and TEA as a base, at room temperature, for the

esterification with 1.5 equivalents of the benzoyl chloride. These were allowed to react over 20 hours, forming an orange solution. TLC showed the formation of a new compound, but a second compound also persisted which was thought to be unreacted benzoyl chloride. The organic layer was washed and separated, then added to an equivalent amount of NaH and THF which was heated at reflux for 20 hours; the mixture was acidified, then washed and separated forming a crystalline brown solid. The crystalline compound was then boiled in glacial acetic acid and sulphuric acid and refluxed for 19 hours, then poured onto an ice/water slurry to form a pale brown solid. TLC showed the presence of three compounds and so recrystallization from ethanol was performed leaving only a small amount of product. The product was then refluxed in hydrobromic acid and glacial acetic acid for 35 hours, then poured onto an ice/water slurry to achieve a small yield of brown crystals. Analysis of this product showed that only the pure 3-benzoylflavone remained.

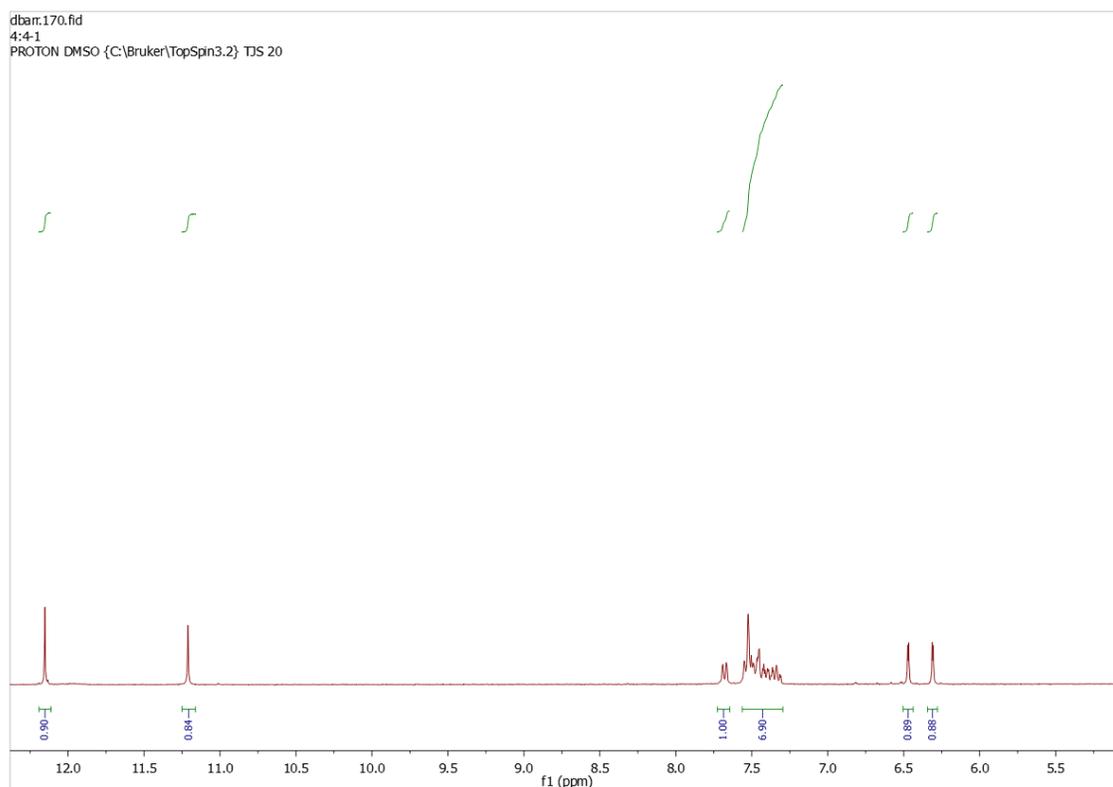


Figure 6.7 – $^1\text{H-NMR}$ of 3-(2-chlorobenzoyl)-2-(2-chlorophenyl)-5,7-dihydroxy-chromen-4-one
 $^1\text{H NMR}$ (300 MHz, $\text{DMSO-}d_6$) δ 12.15 (s, 1H), 11.21 (s, 1H), 7.68 (dd, $J = 1.6$ Hz, 1H), 7.56 – 7.30 (m, 7H), 6.47 (d, $J = 2.1$ Hz, 1H), 6.31 (d, $J = 2.1$ Hz, 1H).

6.8.2 Second Attempt at Synthesis of 2-(2-chlorophenyl)-5',7'-dihydroxy-chromen-4-one

The second attempt followed the same general procedure (Scheme 6.1), with slight modifications on the reagents used using the same amount of 2'-hydroxy-4',6'-methoxyacetophenone with only 1.1 equivalents of 2'-chlorobenzoyl chloride using potassium carbonate in H₂O/toluene as a base for the esterification which was heated at 60°C and allowed to react until completion, as determined by TLC. The organic layer was separated, washed and isolated *in vacuo*. This was then added to an equivalent amount of pyridine and potassium hydroxide which was heated at reflux for 20 hours. TLC showed the presence of two compounds of which seemed to be 2'-chlorobenzoic acid and the 2'-hydroxy-4',6'-methoxyacetophenone, which was likely due to hydrolysis of the ester. The solution was acidified, then washed and separated forming a crystalline brown solid. The presence of starting material was then confirmed using LC-MS and NMR.

6.8.3 Third Attempt at Synthesis of 2-(2-chlorophenyl)-5',7'-dihydroxy-chromen-4-one

As the esterification of the second attempt was a success, the same conditions were kept in the third synthetic attempt keeping the same moles of 2'-hydroxy-4',6'-methoxyacetophenone with 1.1 equivalents of 2'-chlorobenzoyl chloride; modifications were made to the BVR step whereby the aforementioned conditions and base used in attempt one were replicated. 2'-Hydroxy-4',6'-methoxyacetophenone and 2'-chlorobenzoyl chloride were added to potassium carbonate in H₂O/toluene which was heated at 60°C and allowed to react until completion, as determined by TLC. The organic layer was separated, washed and isolated *in vacuo* forming a yellow oil; which solidified into pale yellow crystals when left in a desiccator. Equivalent amounts of NaH and THF were then added and heated at reflux for 20 hours to form an amber solution. TLC showed the formation of the new product as well as a faint spot indicative of a small amount of another compound which persisted. The organic layer was acidified, then washed and separated forming a crystalline brown solid. Crude NMR analysis confirmed production of the β -diketone. The compound was then heated at reflux in glacial

acetic acid and sulphuric acid for 4 hours, then poured onto an ice/water slurry to form a pale brown solid. TLC showed the formation of two compounds with r.f values on silica gel TLC plates near identical.

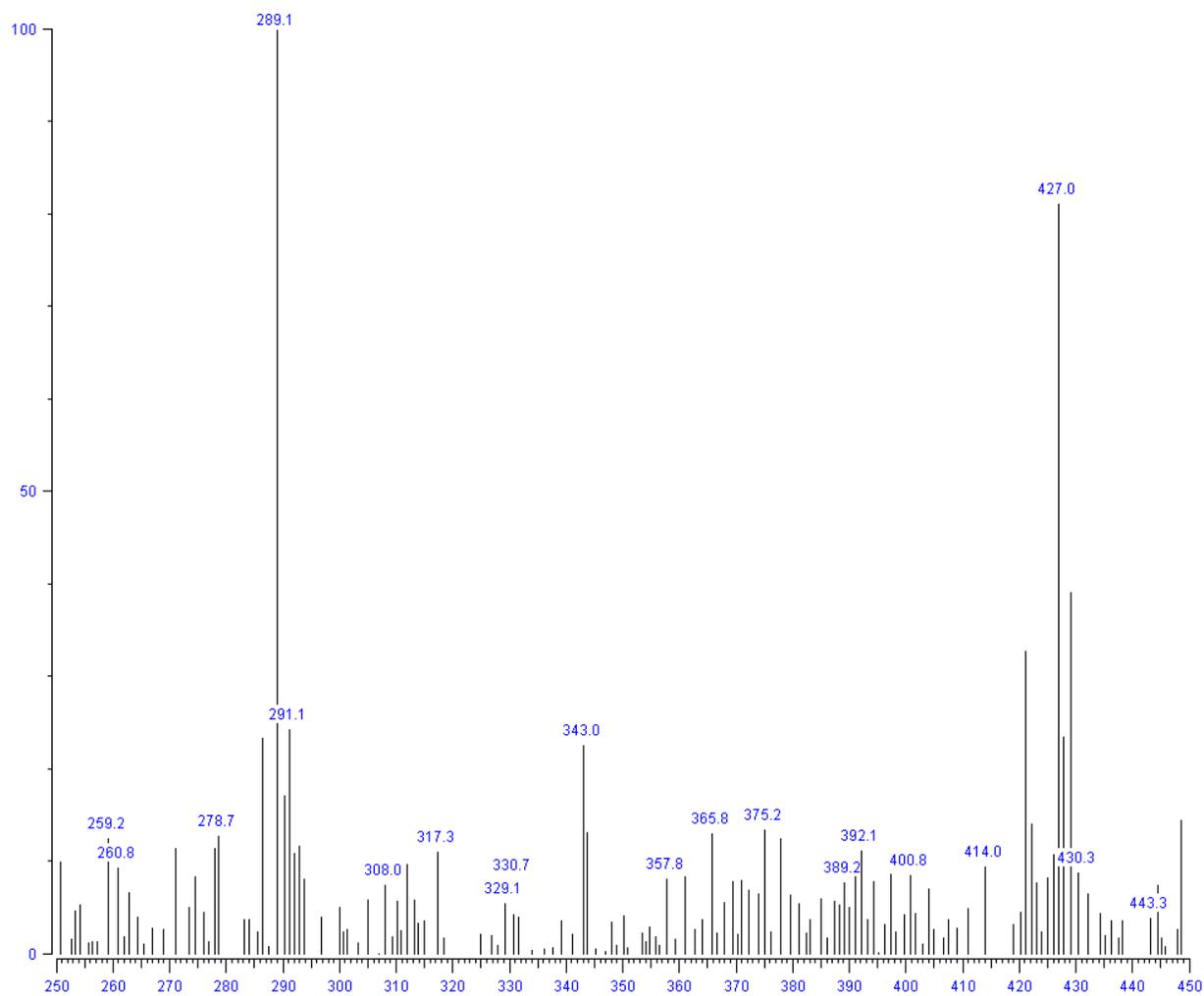
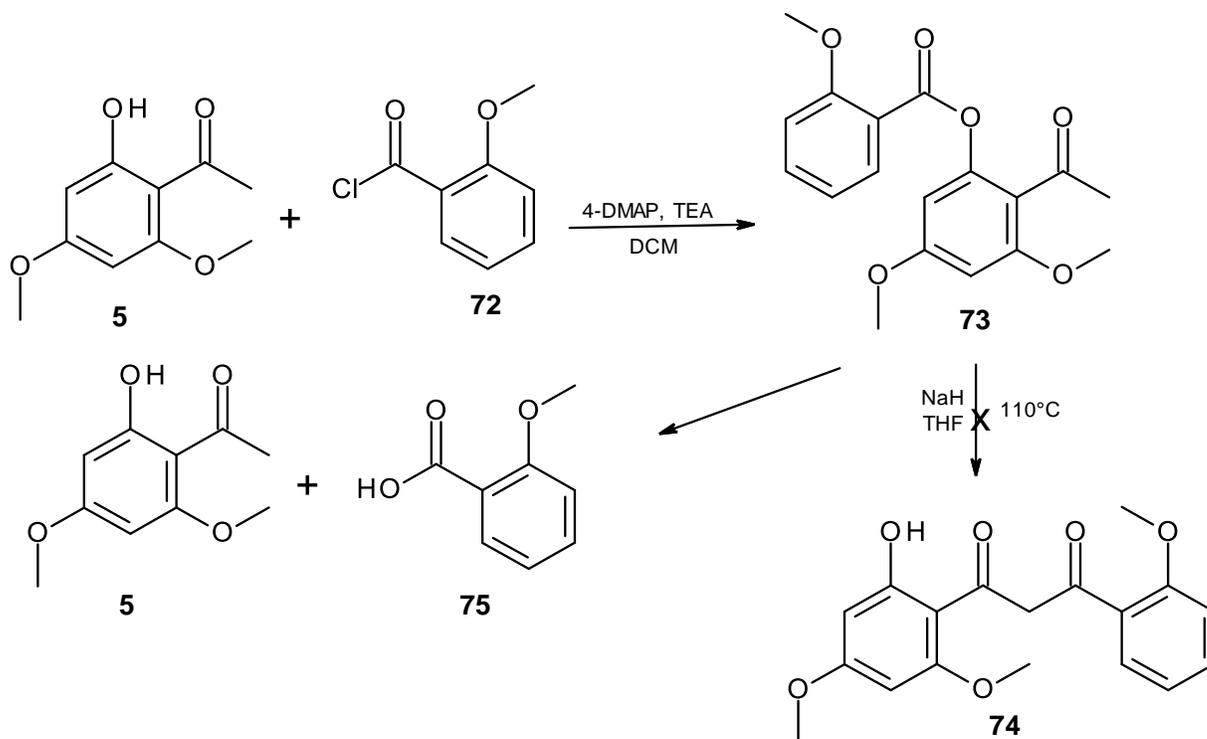


Figure 6.8 – LC-MS showing the 1:1 ratio of OCl(a):OCl(b)

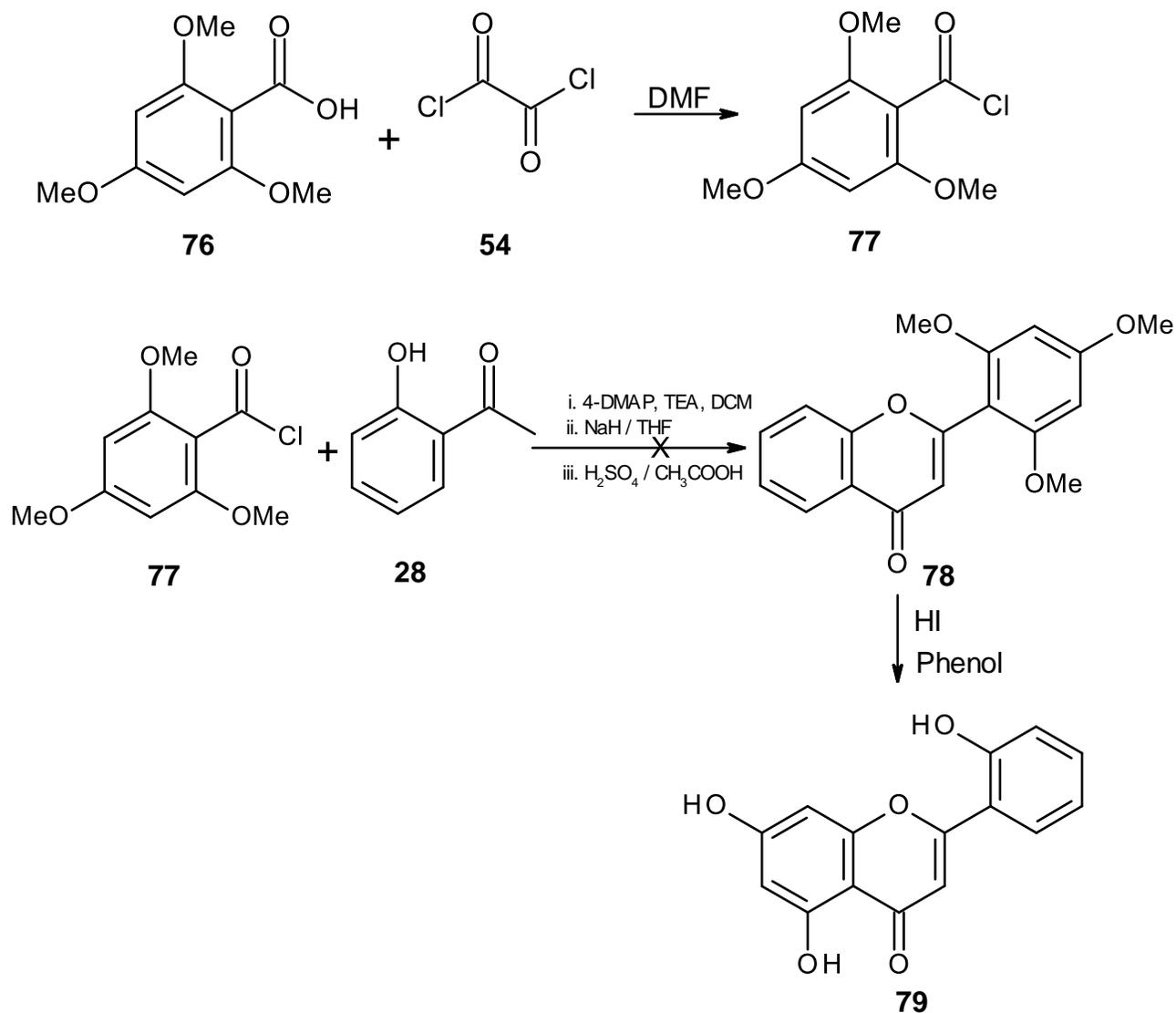
6.9 Attempted Synthesis of 5,7-dihydroxy-2-(2-hydroxyphenyl)chromen-4-one



Scheme 6.8 – The hydrolysis of (2-acetyl-3,5-dimethoxyphenyl) 2-methoxybenzoate reforming starting materials.

An attempt to synthesize 5,7-dihydroxy-2-(2-hydroxyphenyl)chromen-4-one was performed using the Baker-Venkataraman rearrangement route highlighted in Scheme 6.1. 2'-Hydroxy-4',6'-dimethoxyacetophenone (**5**) was added to 1.1 equivalents 2-methoxybenzoyl chloride (**72**) in a mixture of dichloromethane, triethylamine and 4-dimethylaminopyridine and stirred forming a slight yellow solution. The reaction was left until completion, as determined by TLC. The organic layer was separated and washed, then an equivalent amount of sodium hydride was added along with tetrahydrofuran and stirred at reflux in an attempt to form (**74**). Confusingly, TLC showed the formation of two products and further analysis showed to be hydrolysis products of the ester (**5** & **75**). Due to the lack of success with this synthetic attempt, a new synthetic route was approached and the use of anhydrous THF was used to minimize ester hydrolysis.

6.9.1 Second Attempted synthesis of (2-acetyl-3,5-dimethoxy-phenyl) 2-methoxybenzoate



Scheme 6.9 – The method used to synthesise (2-acetyl-3,5-dimethoxy-phenyl) 2-methoxybenzoate. The second attempt of synthesising 5,7-dihydroxy-2-(2-hydroxyphenyl)chromen-4-one was performed using a method observed by Gallagher, K.M, et al,^[119] who found that certain 2'-methoxyflavones rearrange during demethylation by hydroiodic acid under sufficiently drastic conditions, to give the related 2'-hydroxyflavones in which the B and the A ring of the original flavone are interchanged. The starting materials were adapted in order to form a compound which demethylated in order to form (79).

The reaction pathway followed for this reaction is similar to the one described in Scheme 6.1 with the exception of the demethylation step, whereby the reagents used are modified. 2',4',6'-Trimethoxybenzoic acid (**76**) was transformed into the corresponding benzoyl chloride (**77**) by the introduction of 1.1 equivalents of oxalyl chloride (**54**) in DCM followed by one drop of DMF and reacted until completion to form the corresponding benzoyl chloride. TLC showed the presence of two compounds, likely from unreacted acid (**76**) and the newly formed 2',4',6'-trimethoxybenzoyl chloride (**77**). The esterification was performed with 1.1 equivalents of 2'-hydroxyacetophenone (**28**) in a mixture of 4-DMAP, TEA and DCM and allowed to react for 20 hours, forming a yellow solution. TLC showed the presence of multiple compounds and so the organic layer was washed with aqueous NaOH to remove any of the unreacted acidic materials, followed by water and brine and no further purification was performed. The organic layer was then separated and isolated *in vacuo* leaving a viscous yellow oil (**78**). The ester was added to an equivalent amount of NaH in THF and heated at reflux for 20 hours forming an amber solution; this was acidified with HCl, forming a yellow crystalline powder which was collected using vacuum filtration. Flash column chromatography was performed on silica gel to isolate the product. This was then added to glacial acetic acid (CH₃COOH) and sulphuric acid (H₂SO₄) and refluxed for 7 hours before being poured over an ice/water slurry, yielding a small amount of white crystalline material which was collected using vacuum filtration. This was then refluxed in a mixture of hydroiodic acid and phenol. This modification in the demethylation step forms the intermediate disalicyloylmethane (Figure 5.6), which undergoes a ring closure, with the A and B-rings interchanged. After 2 hours this was then poured onto an ice/water slurry, forming a small amount of white crystals (**79**), which were collected using vacuum filtration.

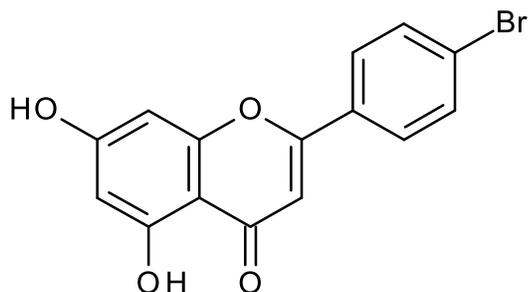
6.10 Conclusion

The Baker-Venkatarman rearrangement has shown to be effective in the synthesis of multiple 5,7-dihydroxyflavones. The BVR route successfully lead to the synthesis of five flavone analogues, four of these containing halogen groups and one of these containing a hydroxyl group. Method limitations were discovered, however, when attempts were made to functionalise the B ring with a 2'-chloro substituent. This limitation lead to the formation of 3'-benzoylflavones with the desired flavone usually synthesised alongside. Two near pure 3'-benzoyl flavones have been synthesised in this study and confirmed using spectroscopic data. The synthesised compounds will be taken forward and undergo kinetic studies by our collaborators in Thessaly, Greece, to compare the previous computational results to the experimental results. The data gathered from these kinetic experiments can be used to refine a search for a more ideal ligand and more potent compounds could be found.

6.11 Experimental

Reactions were followed by analytical thin layer chromatography (TLC) using plastic-backed TLC plates coated in silica G/UV254, run in a variety of solvent systems and visualized with a UV light at 254 nm, p-anisaldehyde stain and/or potassium permanganate stain. Commercially available solvents and reagents were purchased from Fisher, Sigma Aldrich, TCI and Fluorochem and were used without further purification unless specified in the syntheses. Flash column chromatography was carried out on DAVISIL silica 60 Å (40 -63µm) under bellows pressure. ¹H and ¹³C NMR were carried out on a Bruker Fourier 300 (300 MHz). GCs were ran with DSQ II, Trace GC ultra, Triplus, trace1800 GC & ITQ700. LC-MS were ran with the Finnigan LCW Advantage Max, Finnigan PDA Plus detector. Infrared spectra were recorded on a solid sample using a Thermo Nicolet IR-200 FT-IR.

6.11.1 - 2-(4-Bromophenyl)-5',7'-dihydroxy-chromen-4-one

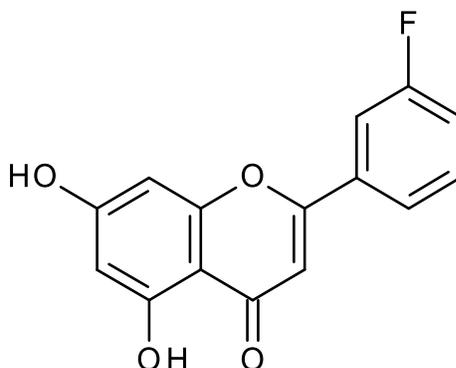


The general procedure was followed using 2'-Hydroxy-4',6'-methoxyacetophenone (**5**) (1.5g, 7.6mmol) which was added to 4'-bromobenzoyl chloride (**48**) (1.85g, 8.41mmol) dissolved in DCM (55mL), 4-DMAP (0.065g) and TEA (1.75mL) was added and the reaction was allowed to stir for 24 hours under a nitrogen atmosphere. The product was acidified with HCl (30mL) and the organic layer was washed with distilled water (2 x 25mL), brine (30mL) and then dried (MgSO₄) The solution was evaporated *in vacuo* leaving a pale yellow solid behind (3g, 7.9 mmol). NaH (1.0 eq, 0.31g) was added along with dry THF (75mL) and the reaction was heated to 66 °C for 20 hours. The solution was washed with HCl (30mL) and EtOAc (70mL) was added to extract. The organic layer was washed with distilled water (2 x 25mL), brine (30mL) and then dried (MgSO₄). The solution was evaporated *in vacuo* leaving a pale brown crystalline solid behind (2.6g, 6.9mmol). Conc. H₂SO₄ (5mL) was added along with CH₃COOH (50mL) and this was allowed to stir at 118 °C for 15 hours. After allowing to cool, the warm solution was poured onto an ice/water slurry (300mL) forming a pale yellow crystalline solid (1.2g, 3.3mmol) which was collected using vacuum filtration. HBr (35mL) and CH₃COOH (20mL) was added and heated at 122 °C for 40 hours. The warm solution was poured onto an ice/water slurry (300mL) yielding the title compound as a pure yellow crystalline product (**49**) (0.62 g, 1.9mmol).

2-(4-Bromophenyl)-5',7'-dihydroxy-chromen-4-one (0.62g, 25%) Chromatography solvent 1:5 ethyl acetate:petroleum ether. Mp 278-280°C. ¹H NMR (300 MHz, DMSO-*d*₆) δ 12.78 (s, 1H, OH), 10.97 (s, 1H, OH), 8.03 (d, *J* = 8.5 Hz, 2H, Ar), 7.78 (d, *J* = 8.5 Hz, 2H, Ar), 7.04 (s, 1H, Ar), 6.53 (d, *J* = 2.0 Hz, 1H, Ar), 6.22 (d, *J* = 2.0 Hz, 1H, Ar).

^{13}C NMR (100.6 MHz, DMSO): δ_{C} (181.75, $\underline{\text{C}}=\text{O}$), (164.44, $\underline{\text{Ar}}\text{-OH}$), (162.01, Ar-C), (161.37, $\underline{\text{Ar}}\text{-OH}$), (157.32, Ar-C), (132.09, Ar-C), (129.89, Ar-C), (128.32, Ar-C), (125.74, Ar-Br), (105.47, Ar-C), (103.92, Ar-C), (99.02, Ar-C), (94.11, Ar-C). $m/z = 333$ [M^+] IR (neat, cm^{-1}) $\nu = 3342$ (O-H), 1651 (C=O), 1508 (Ar C=C), 1267 (O-H), 1185 (C-O-C). 632 (C-Br),

6.11.2 - 2-(3-fluorophenyl)-5',7'-dihydroxy-chromen-4-one.

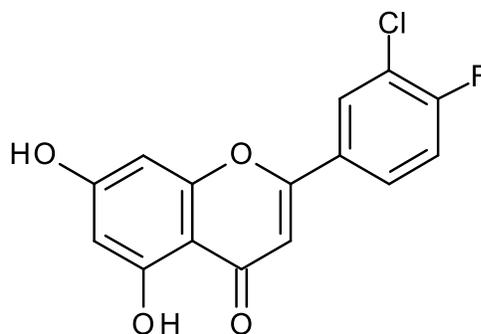


The general procedure was followed using 2'-hydroxy-4',6'-Methoxyacetophenone (**5**) (1.0g, 5mmol) added to 3'-fluorobenzoyl chloride (**51**) (0.7ml, 5.8mmol) dissolved in DCM (50ml), 4-DMAP (0.069g) and TEA (1.1mL) was added and the reaction was allowed to stir for 24 hours under a nitrogen atmosphere until presumed to be complete by TLC. The product was washed with HCl (25mL). The organic layer was washed with distilled water (2 x 25mL), brine (30mL) and then dried (MgSO_4). The solution was evaporated *in vacuo* leaving a white crystalline solid behind (1.57g, 4.9mmol). NaH (1.0 eq, 0.21g) was added along with dry THF (50mL) and the reaction was heated to 66 °C for 20 hours. The solution was acidified with HCl (45mL) and EtOAc (70mL) was added. The organic layer was washed with distilled water (2 x 25mL), brine (30mL) and then dried (MgSO_4). The solution was evaporated *in vacuo* leaving a pale brown crystalline solid behind (1.2g, 3.7mmol). Conc. H_2SO_4 (5mL) was added along with CH_3COOH (50mL) and this was allowed to stir at 118 °C for 20 hours. The warm solution was poured onto an ice/water slurry (350mL) forming a crude yellow crystalline solid (1.0g, 3mmol) which was collected using vacuum filtration. This was recrystallized from ethanol yielding a white crystalline solid (0.6g, 2mmol). HBr (25mL) and CH_3COOH (25mL) was added and

heated at 122 °C for 20 hours. The solution was neutralised with NaOH (2 x 25mL) and EtOAc (60mL) was added. The organic layer was washed with distilled water (2x30mL), brine (50mL) and then dried (MgSO₄). The solution was evaporated *in vacuo* leaving a pale brown crystalline solid behind as pure product (**52**) (0.31g, 1.1mM).

2-(3-fluorophenyl)-5',7'-dihydroxy-chromen-4-one (0.31g, 22%) Mp 269-271°C. ¹H NMR (300 MHz, DMSO-*d*₆) δ 12.76 (s, 1H, OH), 10.98 (s, 1H, OH), 8.02 – 7.91 (m, 2H, Ar-H), 7.63 (dt, *J* = 8.5, 6.0 Hz, 1H, Ar-H), 7.47 (td, *J* = 8.0 Hz, 2.5 Hz, 1H, Ar-H), 7.08 (s, 1H, Ar-H), 6.56 (d, *J* = 2.0 Hz, 1H, Ar-H), 6.23 (d, *J* = 2.0 Hz, 1H, Ar-H). ¹³C NMR (100.6 MHz, DMSO): δ_c (181.78, C=O), (164.49, Ar-OH), (161.54, Ar-C), (161.33, Ar-OH), (157.32, Ar-F), (132.24, Ar-C), (122.50, Ar-C), (118.62, Ar-C), (118.01, Ar-C), (113.37, Ar-C), (113.13, Ar-C), (105.91, Ar-C), (103.96, Ar-C), (99.03, Ar-C), (94.18, Ar-C). *m/z* = 273 [M⁺]. IR (neat, cm⁻¹) ν = 3144 (O-H), 1652 (C=O), 1507 (Ar C=C), 1250 (C-F), 1286 (O-H), 1180 (C-O-C).

6.11.3 - 2-(3-chloro-4-fluoro-phenyl)-5,7-dihydroxy-chromen-4-one.

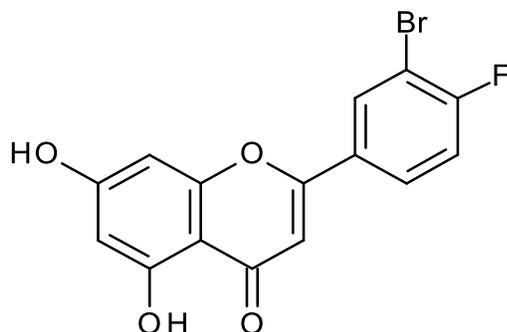


3-Chloro-4-fluorobenzoic acid (**53**) (1.2g 5.5mmol) was added to oxalyl chloride (**54**) (0.51mL, 6.1mmol) in 60mL DCM followed by 2 drops of DMF. This was allowed to react for 1 hour and was isolated *in vacuo* to yield a viscous yellow oil (**55**). The general procedure (Scheme 6.1) could then be followed using 2'-hydroxy-4',6'-methoxyacetophenone (**5**) (1g, 5mM) which was added to 3'-chloro-4'-fluorobenzoyl chloride (**55**) (1.3g, 5.5mM) dissolved in DCM (55ml), 4-DMAP (0.065g) was added and the reaction flask was purged under nitrogen for 10 minutes. TEA (1.1mL) was added and the reaction was allowed to react for 20 hours and the progress of the

reaction was monitored by TLC, the presence of starting material was still present and therefore flash column chromatography was used to obtain product as pale yellow crystals (1.03g, 3mmol). NaH (1.0 eq, 0.12g) was added along with anhydrous THF (65mL) and the reaction was heated to 66 °C for 20 hours. The solution was acidified with HCl (200mL) and the precipitate formed was collected using vacuum filtration leaving an orange solid (0.7g, 2mmol). H₂SO₄ (5mL) was added along with CH₃COOH (50mL) and this was allowed to stir at 118 °C for 20 hours. The warm solution was poured onto an ice/water slurry (400mL) forming a brown solid (0.35g, 1mM) which was collected using vacuum filtration. HBr (25mL) and CH₃COOH (25mL) was added and heated at 122 °C for 20 hours. The warm solution was poured onto an ice/water slurry (450mL) yielding the title compound as a pure brown crystalline product (**56**) (0.09g, 0.3mmol).

2-(3-chloro-4-fluoro-phenyl)-5,7-dihydroxy-chromen-4-one (0.09g, 6%) Mp 268-270°C. ¹H NMR (300 MHz, DMSO-*d*₆) δ 12.74 (s, 1H), 10.99 (s, 1H), 8.35 (dd, *J* = 7.1, 2.4, 0.9 Hz, 1H), 8.18 – 8.05 (m, 1H), 7.62 (t, *J* = 9.0, 1.0 Hz, 1H), 7.07 (s, 1H), 6.56 (d, *J* = 2.1, 1.0 Hz, 1H), 6.21 (d, *J* = 2.1, 1.0 Hz, 1H). ¹³C NMR (100.6 MHz, DMSO): δ_c (181.65, C=O), (164.46, Ar-OH), (161.30, Ar-C), (160.62, Ar-OH), (157.82, Ar-F), (157.25, Ar-C), (128.73, Ar-C), (127.50, Ar-C), (127.42, Ar-C), (120.76, Ar-Cl), (117.69, Ar-C), (105.73, Ar-C), (103.80, Ar-C), (99.02, Ar-C), (94.21, Ar-C). *m/z* = 305 [M⁺]. IR (neat, cm⁻¹) ν = 3221 (O-H), 1654 (C=O), 1501 (Ar C=C), 1248 (O-H), 1116 (C-O-C), 840 (C-F), 632 (C-Cl).

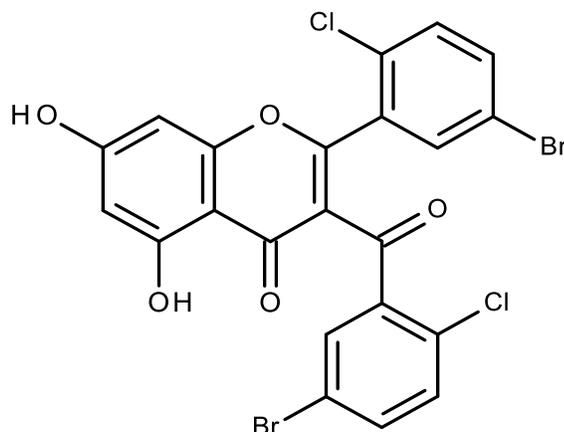
6.11.4 - Synthesis of 2-(3-bromo-4-fluoro-phenyl)-5,7-dihydroxy-chromen-4-one.



3-Bromo-4-fluorobenzoic acid (**57**) (1.2g, 5.5mmol) was added to oxalyl chloride (**54**) (0.55mL, 6.5mmol) in 60mL DCM followed by 2 drops of DMF. This was allowed to react for 1 hour and was isolated *in vacuo* to yield a viscous yellow oil (1.4g, 6mmol) (**58**). The general procedure was then followed using 2-hydroxy-4',6'-methoxyacetophenone (**5**) (1.25g, 5.9mmol) and 3'-chloro-4'-fluorobenzoyl chloride (**58**) (1.4g, 6mmol) and this dissolved in DCM (55ml), 4-DMAP (0.065g) was added and the reaction flask was purged under nitrogen for 10 minutes. TEA (1.1mL) was added and the reaction was allowed to react for 20 hours and the progress of the reaction was monitored by TLC, the presence of starting material (**5**) was still present and therefore flash column chromatography was used to obtain product as pale yellow crystals (1.15g 3mmol). NaH (1.0 eq, 0.12g) was added along with dry THF (65mL) and the reaction was heated to 66 °C for 20 hours. The solution was acidified with HCl (200mL) and the precipitate formed was collected using vacuum filtration, leaving a pale brown crystalline solid behind (0.95g, 2.4mmol). H₂SO₄ (5mL) was added along with CH₃COOH (50mL) and this was allowed to stir at 118 °C for 20 hours. The warm solution was poured onto an ice/water slurry (400mL) forming a brown solid (0.6g, 2.4x10⁻⁴mmol) which was collected using vacuum filtration. HBr (25mL) and CH₃COOH (25mL) was added and heated at 122 °C for 20 hours. The warm solution was poured onto an ice/water slurry (450mL) yielding the title compound as a pure brown crystalline product (0.16, 0.5mmol).

2-(3-bromo-4-fluoro-phenyl)-5,7-dihydroxy-chromen-4-one (0.16g (8.5%) Mp 269-271°C. ¹H NMR (300 MHz, DMSO-*d*₆) δ 12.71 (s, 1H, O-H), 10.94 (s, 1H, O-H), 8.37 (dd, *J* = 7.0, 2.0 Hz, 1H, Ar-H), 8.09 (ddd, *J* = 8.7, 4.5, 2.0 Hz, 1H, Ar-H), 7.53 (t, *J* = 8.7 Hz, 1H, Ar-H), 7.01 (s, 1H, Ar-H), 6.51 (d, *J* = 2.0 Hz, 1H, Ar-H), 6.18 (d, *J* = 2.0 Hz, 1H, Ar-H). ¹³C NMR (100.6 MHz, DMSO): δ_c (181.13, C=O), (164.92, Ar-OH), (161.80, Ar-C), (161.05, Ar-OH), (159.41, Ar-F), (157.73, Ar-C), (131.98, Ar-C), (129.27, Ar-C), (128.60, Ar-C), (117.82, Ar-C), (109.65, Ar-Br), (106.20, Ar-C), (104.30, Ar-C), (99.50, Ar-C), (94.68, Ar-C). *m/z* = 351 [M⁺]. IR (neat, cm⁻¹) ν = 3285 (O-H), 1652 (C=O), 1480 (Ar C=C), 1250 (O-H), 1164 (C-O-C), 840 (C-F), 632 (C-Br).

6.11.5 - 3-(5-bromo-2-chloro-benzoyl)-2-(5-bromo-2-chloro-phenyl)-5,7-dihydroxy-chromen-4-one

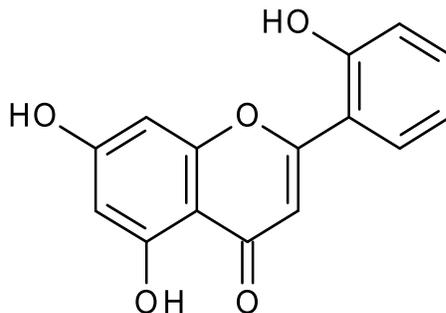


2'-Chloro-5'-bromobenzoic acid (2g, 8.5mmol) was dissolved in DCM (80mL), the reaction flask was purged with nitrogen for 10 minutes and then oxalyl chloride (0.80mL, 9.3mmol) was added along with 2 drops of DMF. This was allowed to react and the progress was monitored by TLC over the period of an hour until completion. The product was isolated *in vacuo* to yield a viscous yellow oil (2g, 7.9mmol). 2'-Hydroxy-4',6'-dimethoxyacetophenone (1.5g, 7.6mmol) and 2'-chloro,5-bromobenzoyl chloride (2g, 7.9mmol) was dissolved in DCM (50mL) along with 4-DMAP (0.075g) and TEA (1.75mL) was added and the reaction was allowed to progress for 20 hours at room temperature until completion, however, a spot persisted with an *rf* value similar to that of 2'-chloro-5'-benzoyl chloride which persisted. The organic layer was separated and washed with distilled water

(2x25mL) and brine (30mL) and then dried (MgSO₄). The remaining solution was isolated *in vacuo* leaving a pale brown crystalline solid behind (2.7g, 6.9mmol). NaH (0.28g, 6.9mmol) was added along with anhydrous THF (75mL) and this was heated to 66 °C for 23 hours. The reaction was then allowed to cool to room temperature and HCl (200mL) was added in 15mL aliquots, and the precipitate that had formed was collected through vacuum filtration (0.87g, 2.1mmol). The crude product was then dissolved in CH₃COOH (60mL) and of H₂SO₄ (5 drops). This was heated to reflux at 118 °C for 6 hours. The flask was taken off heat and allowed to cool before being poured over an ice/water slurry (500mL) to yield a cream/brown precipitate which was collected through vacuum filtration (0.94g, 2.4mmol). HBr (50mL) was added along with CH₃COOH (60mL) and this was heated to reflux at 122 °C for 30 hours. The flask was then allowed to cool and then poured over an ice/water slurry (500mL) forming a pale brown precipitate to form, which was collected using vacuum filtration (0.51g, 0.9mmol).

2-(5-bromo-2-chloro-phenyl)-5,7-dihydroxy-chromen-4-one (0.51g, 11.5%) Mp 273-277°C. ¹H NMR (300 MHz, DMSO-*d*₆) δ 12.11 (s, 1H, O-H), 11.24 (s, 1H, O-H), 7.85 (d, *J* = 2.5 Hz, 1H, Ar-H), 7.83 (d, *J* = 2.5 Hz, 1H, Ar-H), 7.73 (dd, *J* = 8.5, 2.5 Hz, 1H, Ar-H), 7.68 (dd, *J* = 8.5, 2.5 Hz, 1H, Ar-H), 7.50 (d, *J* = 8.7 Hz, 1H, Ar-H), 7.40 (d, *J* = 8.7 Hz, 1H, Ar-H), 6.48 (d, *J* = 2.0 Hz, 1H, Ar-H), 6.32 (d, *J* = 2.0 Hz, 1H, Ar-H). ¹³C NMR (100.6 MHz, DMSO): δ_C (187.63, C=O), (178.52, C=O), (165.38, Ar-OH), (162.95, Ar-OH), (161.70, Ar-C), (157.21, Ar-C), (138.80, Ar-C), (135.68, Ar-C), (135.38, Ar-Cl), (132.54, Ar-C), (132.30, Ar-C), (132.05, Ar-C), (131.43, Ar-Cl), (130.11, Ar-C), (120.05, Ar-Br), (120.00, Ar-Br), (119.93, Ar-C), (103.71, Ar-C), (99.97, Ar-C), (94.62, Ar-C) *m/z* = 584 [M⁺]. IR (neat, cm⁻¹) ν = 3265 (O-H), 1610 (C=O), 1416 (Ar C=C), 1278 (O-H), 1161 (C-O-C), 816 (C-F), 761 (C-Br).

6.11.6 Synthesis of 5,7-dihydroxy-2-(2-hydroxyphenyl)chromen-4-one



2,4,6-Trimethoxybenzoic acid (1.1g, 5.2mmol), was dissolved in DCM (65mL), the reaction flask was purged with nitrogen for 10 minutes and then oxalyl chloride (0.50mL, 5.7 mmol) was added along with 2 drops of DMF. This was allowed to react and the progress was monitored by TLC over the period of an hour until completion. The product was isolated *in vacuo* to yield a viscous yellow oil (1.4g, 6mmol). 2'-Hydroxyacetophenone (0.66mL, 5.5mmol) was added and dissolved in DCM (50mL), along with TEA (1.1mL) and 4-DMAP (0.070g) and the reaction was allowed to progress for 23 hours at room temperature, the presence of multiple spots on the TLC made the identification of the compound difficult and therefore the organic layer was extracted, washed with NaOH (2 x 25mL), followed by water (2 x 25mL), brine (30mL), this was then isolated *in vacuo* to yield a viscous yellow oil (1.8g, 5.4mmol). The flask was purged with nitrogen for ten minutes followed by the addition of NaH (0.22g, 5.4mmol) along with anhydrous THF (50mL) and this was heated at 66 °C for 20 hours. The solution was then acidified with HCl (250mL) to yield a pale yellow crystalline solid, which was collected using vacuum filtration (1.12g, 3.4mmol). Flash column chromatography was performed to isolate the desired product (0.8g (2.5mmol). H₂SO₄ (5ml) was added along with CH₃COOH (60mL) and this was heated to reflux at 118 °C for 6 hours until completion, as determined by TLC. The solution was then poured over an ice/water slurry to yield cream like crystals (0.2g 0.64mmol). Hydroiodic acid (5mL) was added along with phenol (3.21g) and this was heated to 122 °C for 2 hours. The solution was then poured over an ice/water slurry (300mL) to yield pale a pale-cream like crystalline powder (0.02g 0.07mmol).

5,7-dihydroxy-2-(2-hydroxyphenyl)chromen-4-one (0.02g, 1.3%) Mp 260-262°C.
¹H NMR (300 MHz, DMSO-*d*₆) δ 9.48 (m, 3H O-H), 8.04 (dd, 1.5 Hz, 1H, Ar-H), 7.77 (ddt, *J* = 15, 7.5, 1.0 Hz, 1H Ar-H), 7.57 (d, *J* = 8.5 Hz, 1H, Ar-H), 7.53 – 7.43 (t, *J* = 15.0, 7.2 Hz, 1H, Ar-H), 6.23 (s, 1H, Ar-H), 5.90 (d, *J* = 1.0 Hz, 2H, Ar-H). *m/z* = 271 [M⁺]. IR (neat, cm⁻¹) ν = 3102 (O-H), 1604 (C=O), 1458 (Ar C=C), 1266 (O-H), 1137 (C-O-C), 870 (C-F), 743 (C-Br).

Chapter 7 – Conclusion

Type II diabetes (T2D) mellitus is the most common form of diabetes and is seen in 90-95% of those diagnosed. Current methods for treatment of T2D are inefficient, and lifestyle changes are often the first approach taken to manage the disease as serious improvements in health can be achieved. Regular physical activity helps the body cells take up glucose for metabolism and this in turn lowers hyperglycaemia. [18] There are currently many oral hypoglycaemic drugs used to control the disease; biguanides, such as metformin, sulphonylureas, such as chlorpropamide and Meglitinide. However, there are substantial side-effects associated with the current drugs including a risk of hypoglycaemia. Accordingly, there is a swift need for a new, effective treatment that has little to no side effect for those suffering from T2D. Inhibition of both the catalytic site and the caffeine binding site of glycogen phosphorylase are seen as an extremely attractive target for an effective treatment for T2D as this enzyme is responsible for the first step of the glycogenolysis pathway.

Computational methods such as docking and post docking calculations were exploited in this project on a group of β -D-glucopyranosyl derivatives with different heterocyclic linkers which have potential to bind at the catalytic site of GP. Exploiting known co-crystallised complexes from the PDB database, two protein structures for calculations were generated whose orientations belong to the two different orientation/geometries that were observed in residues Asn282, Asp283, Asn284 and Phe285, additionally, three different solvation models were explored. Two of these solvation models belonged to the solved crystal structures of the two different orientation observed and another was generated which took into account all solved crystal structures irrespective of residue orientation. Computational simulations have yielded a number of ligands predicted to have low micromolar inhibition of GP (potentially nanomolar) and can now be taken forward for kinetic experiments. We, for the first time, have derived computational models of MMGBSA for the GPb catalytic site which have revealed excellent predictive potential based on a thorough statistical analysis. Using these models, correlations between predicted and experimental inhibitory potential as high as 0.95 – 0.97 were obtained for a training

set of ligands. These methods have substantial potential for discovery of new effective compounds in the treatment of T2D as thousands of potential ligands could in the future be screened.

The second component of this thesis involved the synthesis of a number of 5',7'-dihydroxy flavone derivatives with different substituents on the B-ring which was achieved using a generalised procedure involving the Baker-Venkataraman rearrangement followed by cyclisation and finally a de-protection step to yield the desired compound. Five compounds were synthesised with different substituents on the B-ring, para-bromo (4'-Br), meta-fluoro, (3'-F), meta-chloro, para-fluoro (3'-Cl,4'-F), meta bromo, para fluoro (3'-bromo,4'-fluoro) and ortho-hydroxyl (2'-OH). However, method limitations were found for a small number of compounds containing o-chloro substituents where 3'-benzoylflavones were synthesised. The 5',7'-dihydroxy flavones synthesised have undertaken previous computational screening and have been predicted to show biological activity against GP. Further kinetic experiments of the compounds listed will give insight into the effectiveness of 5',7'-dihydroxyflavones derivatives and which substituent combination is most ideal in its ability to inhibit GP.

While this thesis has demonstrated the potential of efficiently predicting the inhibition constants of a group of compounds, many opportunities for extending the scope of this thesis remain; the effects of solvation modelling is controversial in computational drug design and a more detailed analysis of this using different cut-off points of the cluster-waters script would elaborate the effects that including more or less explicit water molecules may have on predicted binding energies. Method development of synthetic limitations could be explored in order to produce ortho-chloro 5',7'-dihydroxyflavone derivatives so that a wider library of compounds could be tested with kinetic experiments; adaptations to the current procedure, the use of a different synthetic procedure and other parameters could be then explored in an attempt to efficiently produce these compounds. This thesis has laid the groundwork for the future exploration of these areas.

References

1. Berg. J.M, et al., *Gluconeogenesis and Glycolysis Are Reciprocally Regulated*, in *Biochemistry. 5th Edition*. 2002, W H Freeman: New York. Section 16.4.
2. *Insulin Algorithm*. Available from: <http://www.insulinalgorithms.com> [cited 2017 03/01].
3. IDF. *IDF Diabetes Atlas - 7TH EDITION*. 2015. Available from: <http://www.diabetesatlas.org/> [cited 2017 15/01].
4. Tabish. S.A, *Is Diabetes Becoming the Biggest Epidemic of the Twenty-first Century?*, in *Int J Health Sci (Qassim)*. 2007. p. V-viii.
5. Organisation, W.H. *Global Report on Diabetes*. Global burden of Diabetes, 2016. Available from: who.int/ [cited 2017 15/01]
6. Harvey. S. *Diabetes - type 1*. 2012. Available from: <https://www.umm.edu/health> [cited 2016 29/12].
7. Diabetes.co.uk. *Type 2 Diabetes*. 2017. Available from: <https://www.diabetes.co.uk/type2-diabetes.html> [cited 2017 03/01].
8. Saltiel. A.R, *Insulin signalling and the regulation of glucose and lipid metabolism*. *Nature*, 2011. **13**(414): p. 799-806.
9. Kahn. B.B, *Type 2 Diabetes: When Insulin Secretion Fails to Compensate for Insulin Resistance*. *Cell*, 1998. **92**(5): p. 593-596.
10. American Diabetes Association, *Diagnosis and Classification of Diabetes Mellitus*, in *Diabetes Care*. 2010. p. 62-9.
11. Lawrence. S.R, et al., *Adolescent Health Services: Missing Opportunities*. 2009, The National Academies Press. p. 67-71.
12. Diabetes.co.uk. *Gestational Diabetes*. 2017. Available from: <https://www.diabetes.co.uk/gestational-diabetes.html> [cited 2017 03/01].
13. nhs.uk. *Gestational diabetes*. 2017 [cited 2017 03/01].
14. NHS. *Type 2 diabetes - Symptoms*. 2017. Available from: www.nhs.co.uk [cited 2017 03/01].
15. NEI. *National Eye Institute. Facts About Diabetic Eye Disease*. 2015. Available from: <https://nei.nih.gov/health/diabetic/retinopathy> [cited 2017 04/01].
16. Diabetes.org.uk. *Diabetes Complications*. 2017. Available from: <https://www.diabetes.org.uk/guide-to-diabetes/complications> [cited 2017 04/01].
17. NIDDK. *National Institute of Diabetes and Digestive and Kidney Diseases. Diabetes, Heart Disease, and Stroke*. 2013. p. 1-12. Available from: <https://www.niddk.nih.gov> [cited 2017 04/01]
18. Asif. M, *The prevention and control the type-2 diabetes by changing lifestyle and dietary pattern*, in *J Educ Health Promot*. 2014: India.
19. Pérez-Jiménez. F, et al., *A Mediterranean and a high-carbohydrate diet improve glucose metabolism in healthy young persons*. *Diabetologia*, 2001. **44**(11): p. 2038-2043.
20. Viollet. B, *Cellular and molecular mechanisms of metformin: an overview*. 2012. **122**(6): p. 253-70.
21. Rena. G, et al., *Molecular mechanism of action of metformin: old or new insights?*, in *Diabetologia*. 2013: Berlin/Heidelberg. p. 1898-906.
22. Johnson. N.P, *Metformin use in women with polycystic ovary syndrome*, in *Ann Transl Med*. 2014.
23. Nasri. H, et al., *Metformin: Current knowledge*, in *J Res Med Sci*. 2014: India. p. 658-64.
24. Teramoto. N, *Physiological roles of ATP-sensitive K(+) channels in smooth muscle*. *The Journal of Physiology*, 2006. **572**(Pt 3): p. 617-624.

25. Proks. P, et al., *Sulfonylurea Stimulation of Insulin Secretion*. Diabetes, 2002. **51**(suppl 3): p. 368.
26. Daniele. S, et al., *Sulphonylureas and their use in clinical practice*. 2015. p. 840-848.
27. Kumar. S, et al., *α -glucosidase inhibitors from plants: A natural approach to treat diabetes*. 2011, Pharmacogn Rev. p. 19-29.
28. Nyenwe. E.A, et al., *Management of type 2 diabetes: evolving strategies for the treatment of patients with type 2 diabetes*. Metabolism, 2011. **60**(1): p. 1-23.
29. Irons. B.K, *Drug treatment of type 2 diabetes mellitus in patients for whom metformin is contraindicated*, in *Diabetes Metab Syndr Obes*. 2014. p. 15-24.
30. Rizos. C.V, et al., *How safe is the use of thiazolidinediones in clinical practice?* Expert Opinion on Drug Safety, 2009. **8**(1): p. 15-32.
31. Petznick. A., *Insulin Management of Type 2 Diabetes Mellitus*. 2011, *Am Fam Physician*. p. 183-190.
32. DeFronzo. A.R, et al., *Pharmacologic Therapy for Type 2 Diabetes Mellitus*. 1999, American College of Physicians–American Society of Internal Medicine. p. 281-303.
33. Wheatley. T, *Insulin Oedema and its Clinical Significance: Metabolic Studies in Three Cases*. Diabetic Medicine, 1985. **2**(5): p. pg 400-404.
34. Mokta. J.K, et al., *Insulin lipodystrophy and lipohypertrophy*, in *Indian J Endocrinol Metab*. 2013: India. p. 773-4.
35. Zographos. S.E, et al., *The structure of glycogen phosphorylase b with an alkyldihydropyridine-dicarboxylic acid compound, a novel and potent inhibitor*. Structure. **5**(11): p. 1413-1425.
36. Hayes. J.M, et al., *Natural products and their derivatives as inhibitors of glycogen phosphorylase: potential treatment for type 2 diabetes*. Phytochemistry Reviews, 2014. **13**(2): p. 471-498.
37. Berg. J.M, et al., *Glucose Can Be Synthesized from Noncarbohydrate Precursors.*, in *Biochemistry. 5th edition.*, W.H.F.a. Company., Editor. 2002. Section 16.3.
38. Berg. J.M, et al., *Epinephrine and Glucagon Signal the Need for Glycogen Breakdown.*, in *Biochemistry. 5th edition*. 2002., W. H. Freeman and Company: New York. Section 21.3.
39. Widmaier. E.P, et al., *Vander's human physiology. The mechanisms of body function*. 2008, Boston: McGraw-Hill Higher Education.
40. PDB. *Protein Data Bank*. 2017. Available from: <https://www.rcsb.org/> [cited 2017 05/02].
41. Leonidas. D.D, et al., *Kinetic properties of tetrameric glycogen phosphorylase b in solution and in the crystalline state*. Protein Science : A Publication of the Protein Society, 1992. **1**(9): p. 1123-1132.
42. Berg. J.M, et al., *Phosphorylase Is Regulated by Allosteric Interactions and Reversible Phosphorylation*, in *Biochemistry. 5th Edition.*, W.H.F.a. Company, Editor. 2002: New York.
43. Tsai. C.S, *BIOMACROMOLECULES. Introduction to Structure, Function and Informatics*. 2007, New Jersey: John Wiley & Sons.
44. Goodsell. D.S, et al., *The RCSB PDB "Molecule of the Month": Inspiring a Molecular View of Biology*. 2001 [cited 2017 06/02].
45. Livanova. N.B, et al., *Pyridoxal 5'-phosphate as a catalytic and conformational cofactor of muscle glycogen phosphorylase B*. Biochemistry (Mosc), 2002. **67**(10): p. 1089-98.
46. Palm. D, et al., *The role of pyridoxal 5'-phosphate in glycogen phosphorylase catalysis*. Biochemistry, 1990. **29**(5): p. 1099-107.
47. Rath. V.L, et al., *Activation of human liver glycogen phosphorylase by alteration of the secondary structure and packing of the catalytic core*. Mol Cell, 2000. **6**(1): p. 139-48.

48. Hayes, J.M, et al., *Computation as a tool for glycogen phosphorylase inhibitor design*. Mini Rev Med Chem, 2010. **10**(12): p. 1156-74.
49. Street. I.P, *Hydrogen bonding and specificity. Fluorodeoxy sugars as probes of hydrogen bonding in the glycogen phosphorylase-glucose complex*. Biochemistry, 1986. **25**(20): p. 6021-6027.
50. Martin. J.L, et al., *Glucose analog inhibitors of glycogen phosphorylase: the design of potential drugs for diabetes*. Biochemistry, 1991. **30**(42): p. 10101-10116.
51. Parmenopoulou. V, et al., *Structure based inhibitor design targeting glycogen phosphorylase B. Virtual screening, synthesis, biochemical and biological assessment of novel N-acyl-beta-d-glucopyranosylamines*. Bioorg Med Chem, 2014. **22**(17): p. 4810-25.
52. Praly. J.P, *Inhibition of glycogen phosphorylase in the context of type 2 diabetes, with focus on recent inhibitors bound at the active site*. Mini Rev Med Chem, 2010. **10**(12): p. 1102-26.
53. Chrysin. E.D, et al., *Amide-1,2,3-triazole bioisosterism: the glycogen phosphorylase case*. Tetrahedron: Asymmetry, 2009. **Volume 20**(6-8): p. 733-740.
54. Johnson. L.N, et al., *Crystallographic binding studies on the allosteric inhibitor glucose-6-phosphate to T state glycogen phosphorylase b*. J Mol Biol, 1993. **232**(1): p. 253-67.
55. Kemp. D.M., *Type 2 Diabetes: Disease Overview, in New Therapeutic Strategies for Type 2 Diabetes*, . 2012: Royal Society of Chemistry.
56. Chebotareva, N.A, et al., *Regulation of muscle glycogen phosphorylase by physiological effectors*. Biotechnol Genet Eng Rev, 2001. **18**: p. 265-97.
57. Oikonomakos. N.G, *Glycogen phosphorylase as a molecular target for type 2 diabetes therapy*. Curr Protein Pept Sci, 2002. **3**(6): p. 561-86.
58. Ekstrom. J.L, et al., *Structure-activity analysis of the purine binding site of human liver glycogen phosphorylase*. Chem Biol, 2002. **9**(8): p. 915-24.
59. Tsitsanou. K.E, et al., *Sourcing the affinity of flavonoids for the glycogen phosphorylase inhibitor site via crystallography, kinetics and QM/MM-PBSA binding studies: comparison of chrysin and flavopiridol*. Food Chem Toxicol, 2013. **61**: p. 14-27.
60. Kaiser. A, et al., *The cyclin-dependent kinase (CDK) inhibitor flavopiridol inhibits glycogen phosphorylase*. Arch Biochem Biophys, 2001. **386**(2): p. 179-87.
61. Oikonomakos. N.G, et al., *Flavopiridol inhibits glycogen phosphorylase by binding at the inhibitor site*. J Biol Chem, 2000. **275**(44): p. 34566-73.
62. Johnson. L.N, *Glycogen phosphorylase: control by phosphorylation and allosteric effectors*. Faseb j, 1992. **6**(6): p. 2274-82.
63. Kasvinsky. P.J, et al., *The regulation of glycogen phosphorylase alpha by nucleotide derivatives. Kinetic and x-ray crystallographic studies*. J Biol Chem, 1978. **253**(9): p. 3343-51.
64. Kantsadi. A.L, et al., *Biochemical and biological assessment of the inhibitory potency of extracts from vinification byproducts of Vitis vinifera extracts against glycogen phosphorylase*. Food Chem Toxicol, 2014. **67**: p. 35-43.
65. Pandey. K.B, et al., *Plant polyphenols as dietary antioxidants in human health and disease*. Oxid Med Cell Longev, 2009. **2**(5): p. 270-8.
66. Kozłowska. A, *Flavonoids--food sources and health benefits*. Rocznik Panstw Zakl Hig, 2014. **65**(2): p. 79-85.
67. Kumar. S, et al., *Chemistry and Biological Activities of Flavonoids: An Overview*. TheScientificWorldJournal ed. 2013.
68. Jakobs. S, et al., *Natural flavonoids are potent inhibitors of glycogen phosphorylase*. Mol Nutr Food Res, 2006. **50**(1): p. 52-7.

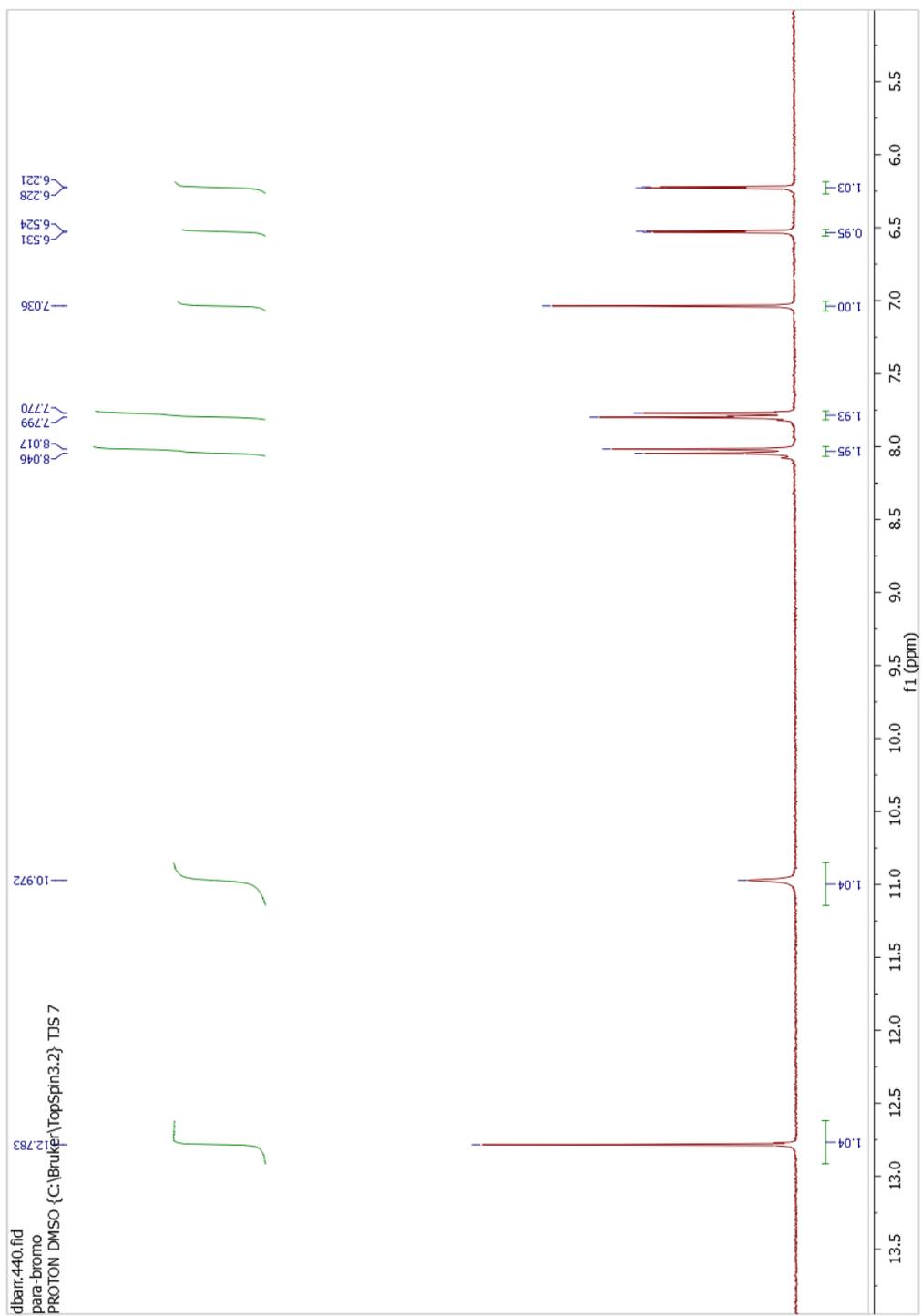
69. Kyriakis. E, et al., *Natural flavonoids as antidiabetic agents. The binding of gallic and ellagic acids to glycogen phosphorylase b*. FEBS Lett, 2015. **589**(15): p. 1787-94.
70. Kato. A, et al., *Structure–Activity Relationships of Flavonoids as Potential Inhibitors of Glycogen Phosphorylase*. Journal of Agricultural and Food Chemistry, 2008. **56**(12): p. 4469-4473.
71. Kamiyama. O, et al., *In vitro inhibition of α -glucosidases and glycogen phosphorylase by catechin gallates in green tea*. Food Chemistry, 2010. **122**(4): p. 1061-1066.
72. Sak. K, *Cytotoxicity of dietary flavonoids on different human cancer types*. Pharmacogn Rev, 2014. **8**(16): p. 122-46.
73. Pietta. P.G, *Flavonoids as Antioxidants*. Journal of Natural Products, 2000. **63**(7): p. 1035-1042.
74. Kato. A, et al., *Protective Effects of Dietary Chamomile Tea on Diabetic Complications*. Journal of Agricultural and Food Chemistry, 2008. **56**(17): p. 8206-8211.
75. Hampson. L.J, et al., *Bioactivity of glycogen phosphorylase inhibitors that bind to the purine nucleoside site*. Bioorganic & Medicinal Chemistry, 2006. **14**(23): p. 7835-7845.
76. Wilcken. R, et al., *Principles and Applications of Halogen Bonding in Medicinal Chemistry and Chemical Biology*. Journal of Medicinal Chemistry, 2013. **56**(4): p. 1363-1388.
77. Hughes. J.P, et al., *Principles of early drug discovery*. Br J Pharmacol, 2011. **162**(6): p. 1239-49.
78. Ekins. S, *In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling*. British Journal of Pharmacology, 2007. **152**(1): p. 9-20.
79. McInnes. C, *Virtual screening strategies in drug discovery*. Curr Opin Chem Biol, 2007. **11**(5): p. 494-502.
80. Irwin. J.J, et al., *ZINC--a free database of commercially available compounds for virtual screening*. J Chem Inf Model, 2005. **45**(1): p. 177-82.
81. Vanommeslaeghe. K, *Molecular Mechanics*. Current pharmaceutical design, 2014. **20**(20): p. 3281-3292.
82. Hobza. P, et al., *Structure, energetics, and dynamics of the nucleic Acid base pairs: nonempirical ab initio calculations*. Chem Rev, 1999. **99**(11): p. 3247-76.
83. Case. D.A, et al., *AMBER*. 2017: University of California, San Francisco.
84. Brooks. B.R, et al., *CHARMM: The Biomolecular Simulation Program*. Journal of computational chemistry, 2009. **30**(10): p. 1545-1614.
85. Halgren, T.A, et al., *Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94*. 1996: Merck Research Laboratories. p. 490–519.
86. **Schrödinger Release 2017-2**. 2017, LLC: New York, NY.
87. Harder. E, et al., *OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins*. Journal of Chemical Theory and Computation, 2016. **12**(1): p. 281-296.
88. Berendsen. H.J.C, et al., *The missing term in effective pair potentials*. The Journal of Physical Chemistry, 1987. **91**(24): p. 6269-6271.
89. Dias. R, et al., *Molecular docking algorithms*. Curr Drug Targets, 2008. **9**(12): p. 1040-7.
90. Lorber. D.M, *Flexible ligand docking using conformational ensembles*. Protein Sci, 1998. **7**(4): p. 938-50.
91. **Schrödinger Release 2017-2 : GlideSchrödinger Release 2017-2**. 2017, LLC: New York, NY.
92. Jones. G, et al., *Development and validation of a genetic algorithm for flexible docking*. Journal of Molecular Biology, 1997. **267**(3): p. 727-748.
93. Korb. O, *Empirical scoring functions for advanced protein-ligand docking with PLANTS*. J Chem Inf Model, 2009. **49**(1): p. 84-96.

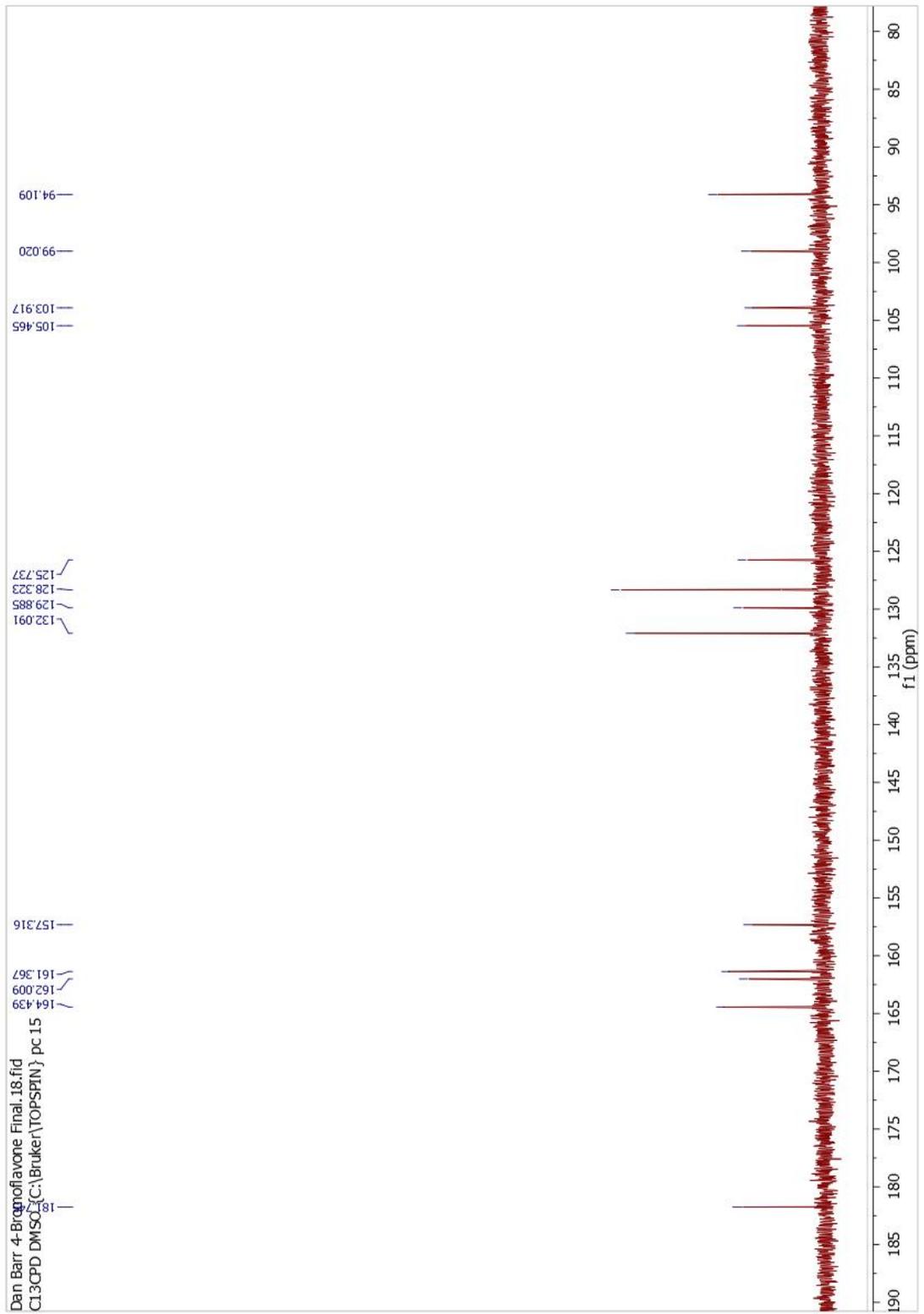
94. Schrodinger, *Prime*. 2012, LCC: New York, NY.
95. Yu. Z, et al., *What role do surfaces play in GB models? A new-generation of surface-generalized born model based on a novel gaussian surface for biomolecules*. *J Comput Chem*, 2006. **27**(1): p. 72-89.
96. Srivastava. M, et al., *Molecular modeling evaluation of the antimalarial activity of artemisinin analogues: molecular docking and rescoring using Prime/MM-GBSA Approach*. 2010.
97. Kun. S, et al., *New synthesis of 3-(beta-D-glucopyranosyl)-5-substituted-1,2,4-triazoles, nanomolar inhibitors of glycogen phosphorylase*, in *Eur J Med Chem*. 2014, 2014 Elsevier Masson SAS: France. p. 567-79.
98. Tóth. M, et al., *Synthesis and structure–activity relationships of C-glycosylated oxadiazoles as inhibitors of glycogen phosphorylase*. *Bioorganic & Medicinal Chemistry*, 2009. **17**(13): p. 4773-4785.
99. Chrysin. E.D, et al., *Kinetic and crystallographic studies on 2-(β-D-glucopyranosyl)-5-methyl-1, 3, 4-oxadiazole, -benzothiazole, and -benzimidazole, inhibitors of muscle glycogen phosphorylase b. Evidence for a new binding site*.
100. Chrysin. E.D, et al., *Amide-1,2,3-triazole bioisosterism: the glycogen phosphorylase case*. *Tetrahedron: Asymmetry*, 2009. **Volume 20**(6-8): p. 733-740.
101. Olsson. M.H.M, et al., *PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions*. *Journal of Chemical Theory and Computation*, 2011. **7**(2): p. 525-537.
102. Pottel. J, et al., *Docking Ligands into Flexible and Solvated Macromolecules. 6. Development and Application to the Docking of HDACs and other Zinc Metalloenzymes Inhibitors*. *Journal of Chemical Information and Modeling*, 2014. **54**((1)): p. 254-265.
103. Yang. J.M, et al., *GEMDOCK: a generic evolutionary method for molecular docking*. *Proteins*, 2004. **55**(2): p. 288-304.
104. Pospisil. P, et al., *Methodology and Problems of Protein-Ligand Docking: Case Study of Dihydroorotate Dehydrogenase, Thymidine Kinase, and Phosphodiesterase 4*. *Journal of Receptors and Signal Transduction*, 2002. **22**(1-4).
105. Birch. L, et al., *Sensitivity of molecular docking to induced fit effects in influenza virus neuraminidase*. *Computer-Aided molecule design*, 2002. **16**: p. 855-869.
106. de Graaf. C, et al., *Binding mode prediction of cytochrome p450 and thymidine kinase protein-ligand complexes by consideration of water and rescoring in automated docking*. *J Med Chem*, 2005. **48**(7): p. 2308-18.
107. Mavrokefalos. N, et al., *Discovery of the Glycogen Phosphorylase-Modulating Activity of a Resveratrol Glucoside by Using a Virtual Screening Protocol Optimized for Solvation Effects*. *Planta Med*, 2015. **81**(06): p. 507-516.
108. Hevener. K.E, et al., *Validation of Molecular Docking Programs for Virtual Screening against Dihydropteroate Synthase*. *Journal of chemical information and modeling*, 2009. **49**(2): p. 444-460.
109. Hayes. J, et al., *MM-GB(PB)SA Calculations of Protein-Ligand Binding Free Energies*. 2012.
110. Chetter. B, *Computer-aided design and synthesis of 5'7-dihydroxyflavone derivatives as glycogen phosphorylase inhibitors*. 2016 University of Central Lancashire.
111. Thomas. N, *Pharmacological activities of chromene derivatives: An overview*. Vol. 6. 2013. p. 11-15.
112. Hu. K, et al., *Synthesis and cytotoxicity of novel chrysin derivatives*. *Medicinal Chemistry Research*, 2011. **20**(7): p. 838-846.

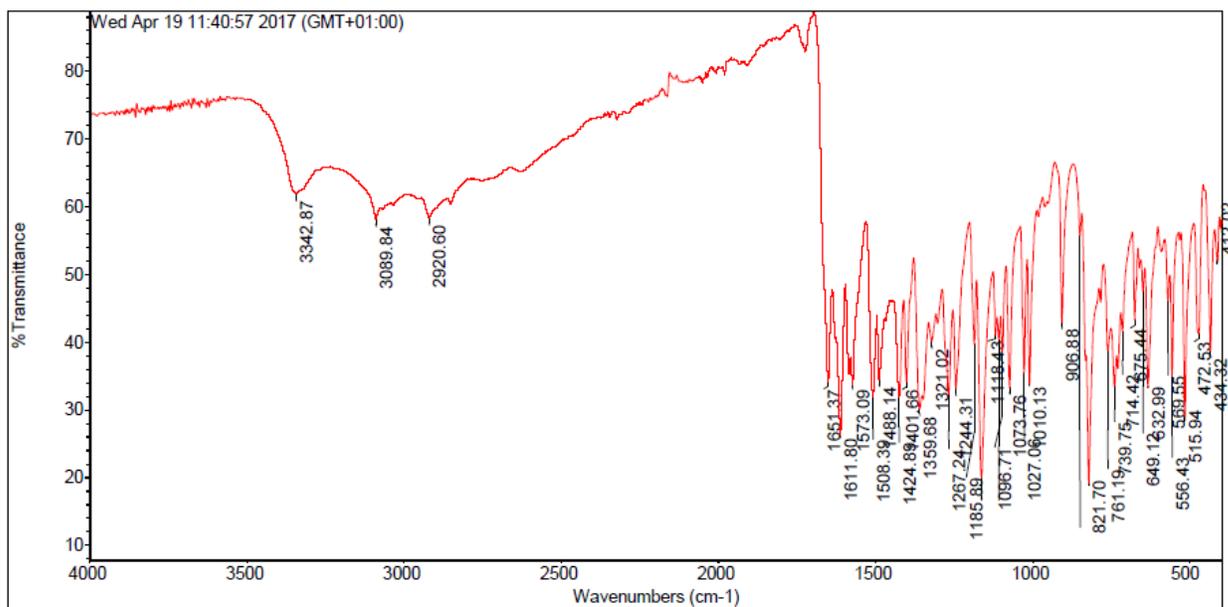
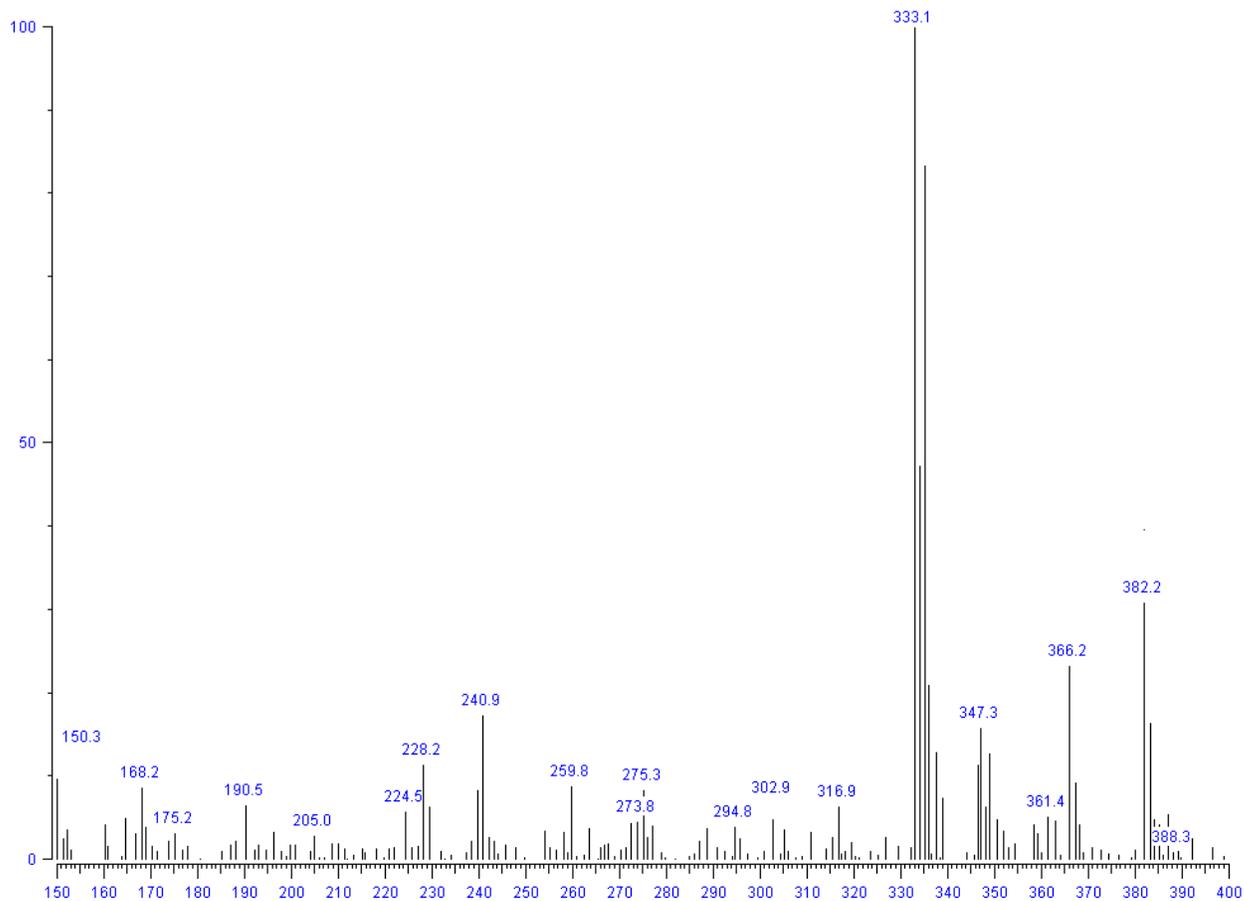
113. Ameen. D, *Mechanism and Application of Baker–Venkataraman O→C Acyl Migration Reactions*. Vol. 47. 2014.
114. Ninomiya. M, et al., *Increased bioavailability of tricin-amino acid derivatives via a prodrug approach*. J Med Chem, 2011. **54**(5): p. 1529-36.
115. Varma. R.S, *Green Chemistry green chemistry with Microwave Energy green chemistry with microwave energy*, in *Encyclopedia of Sustainability Science and Technology*, R.A. Meyers, Editor. 2012, Springer New York: New York, NY. p. 4642-4673.
116. Kabalka. G.W, *Microwave-assisted synthesis of functionalized flavones and chromones*. Tetrahedron Letters, 2005. **46**(37): p. 6315-6317.
117. Seijas. J.A, et al., *Solvent-Free Synthesis of Functionalized Flavones under Microwave Irradiation*. The Journal of Organic Chemistry, 2005. **70**(7): p. 2855-2858.
118. Doporto. M.L, et al., *Rearrangement in the demethylation of 2[prime or minute]-methoxyflavones. Part II. Further experiments and the determination of the composition of Iotoflavin*. Journal of the Chemical Society (Resumed), 1955(0): p. 4249-4256.
119. Gallagher. K.M, et al., *771. Rearrangement in the demethylation of 2[prime or minute]-methoxyflavones*. Journal of the Chemical Society (Resumed), 1953(0): p. 3770-3777.
120. Sashidhara. K.V, et al., *A novel route to synthesis of flavones from salicylaldehyde and acetophenone derivatives*. Tetrahedron Letters, 2012. **53**(18): p. 2355-2359.
121. Zambare. A.S, et al., *Development of mild and efficient method for synthesis of substituted flavones using oxalic acid catalyst*. Chinese Chemical Letters, 2009. **20**(2): p. 171-174.
122. Kumar. K.H, et al., *A novel one-pot oxidative cyclization of 2'-amino and 2'-hydroxychalcones employing FeCl₃ center dot 6H(2)O-methanol. Synthesis of 4-alkoxy-2-aryl-quinolines and flavones*. Tetrahedron, 2007. **63**(38): p. 9531-9535.
123. Lee. A.F, et al., *Hydrodebromination of Bromobenzene over Pt(111)*. The Journal of Physical Chemistry C, 2007. **111**(28): p. 10455-10460.
124. Yenjai. C, et al., *Structural modification of 5,7-dimethoxyflavone from Kaempferia parviflora and biological activities*. Arch Pharm Res, 2009. **32**(9): p. 1179-84.
125. Mavel. S, et al., *Synthesis and biological evaluation of a series of flavone derivatives as potential radioligands for imaging the multidrug resistance-associated protein 1 (ABCC1/MRP1)*. Bioorg Med Chem, 2006. **14**(5): p. 1599-607.
126. Miao. H, *Regiospecific Carbonylative Annulation of Iodophenol Acetates and Acetylenes To Construct the Flavones by a New Catalyst of Palladium–Thiourea–dppp Complex*. Organic Letters, 2000. **2**(12): p. 1765-1768.
127. HSCIC UK., *National Diabetes audit. Report 2: Complications and Mortality*. 2011 - 2012. Available from www.digital.nhs.uk/nda p. 5. [cited 05/01]
128. Kahn. B.B, *Type 2 diabetes: when insulin secretion fails to compensate for insulin resistance*. Cell, 1998. **92**(5): p. 593-6.
129. Ong. K.C, *Effects of myricetin on glycemia and glycogen metabolism in diabetic rats*. Life Sci, 2000. **67**(14): p. 1695-705.
130. Anila. L, et al., *Beneficial effects of flavonoids from Sesamum indicum, Emblica officinalis and Momordica charantia*. Phytother Res, 2000. **14**(8): p. 592-5.
131. Szablewski. L, *Glucose Homeostasis and Insulin Resistance*. 2011: Bentham Books.
132. Chee, C.F, et al., *An efficient one-pot synthesis of flavones*. Tetrahedron Letters, 2011. **52**(24): p. 3120-3123.

Appendix

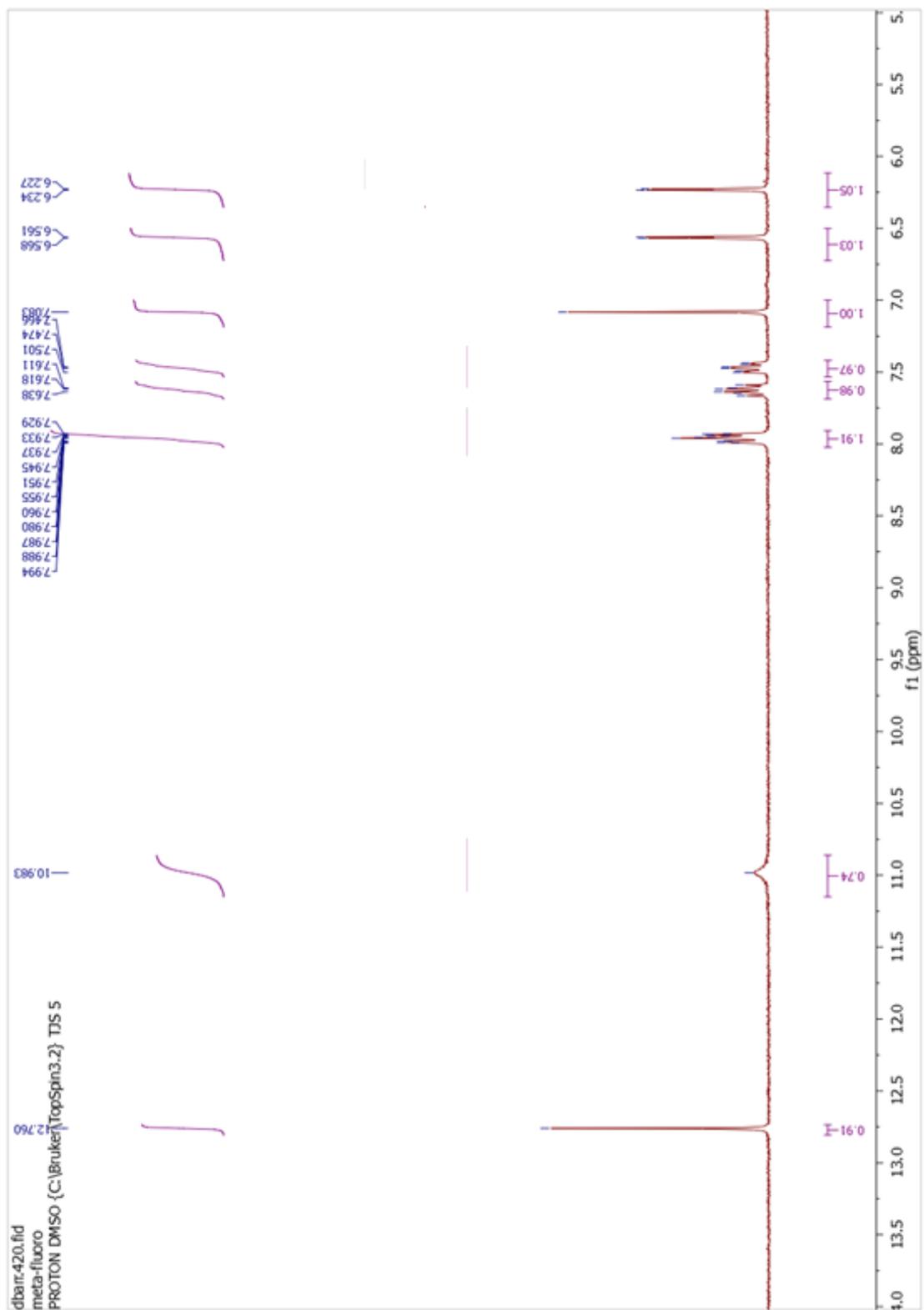
2-(4-bromophenyl)-5,7-dihydroxy-chromen-4-one

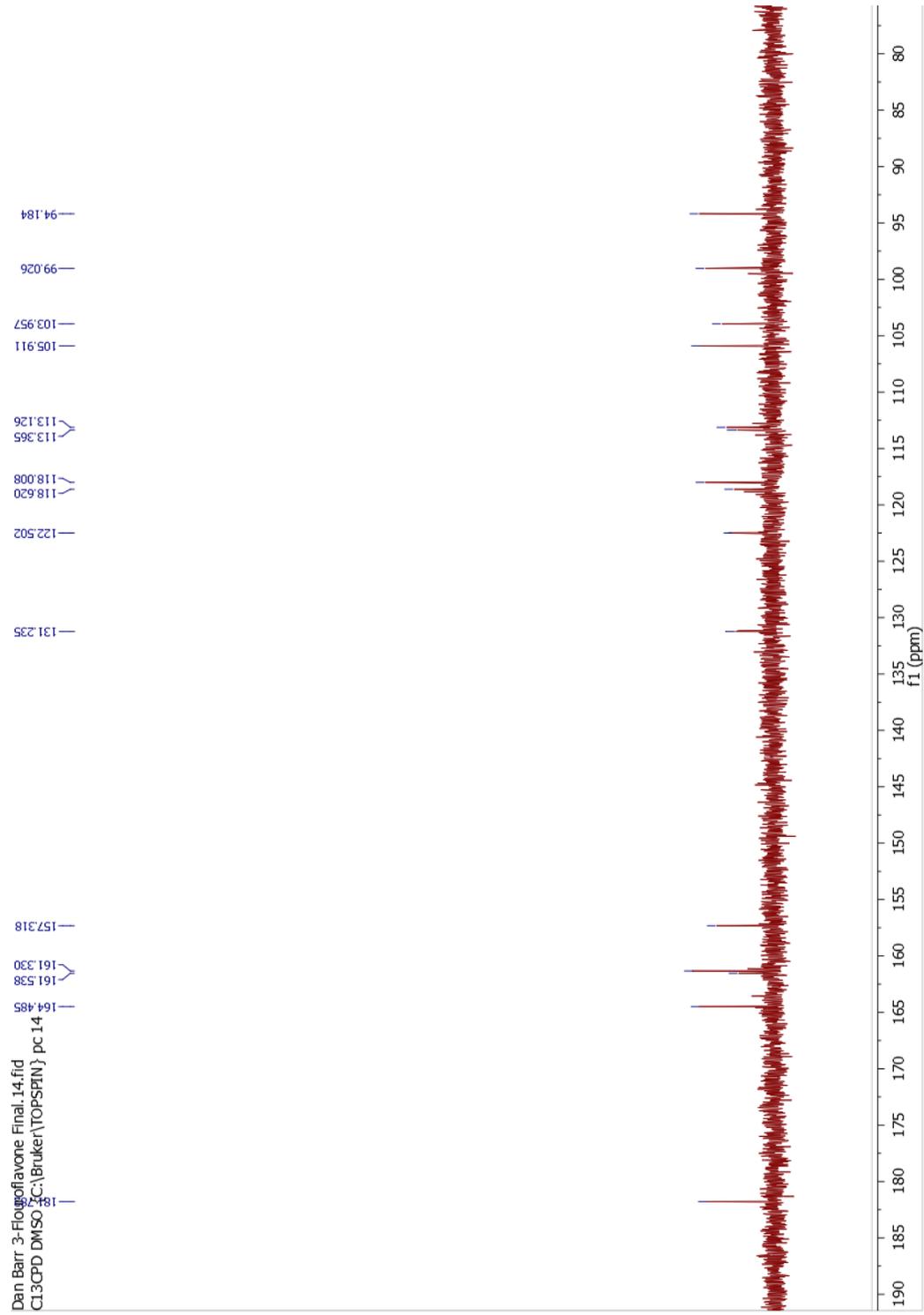


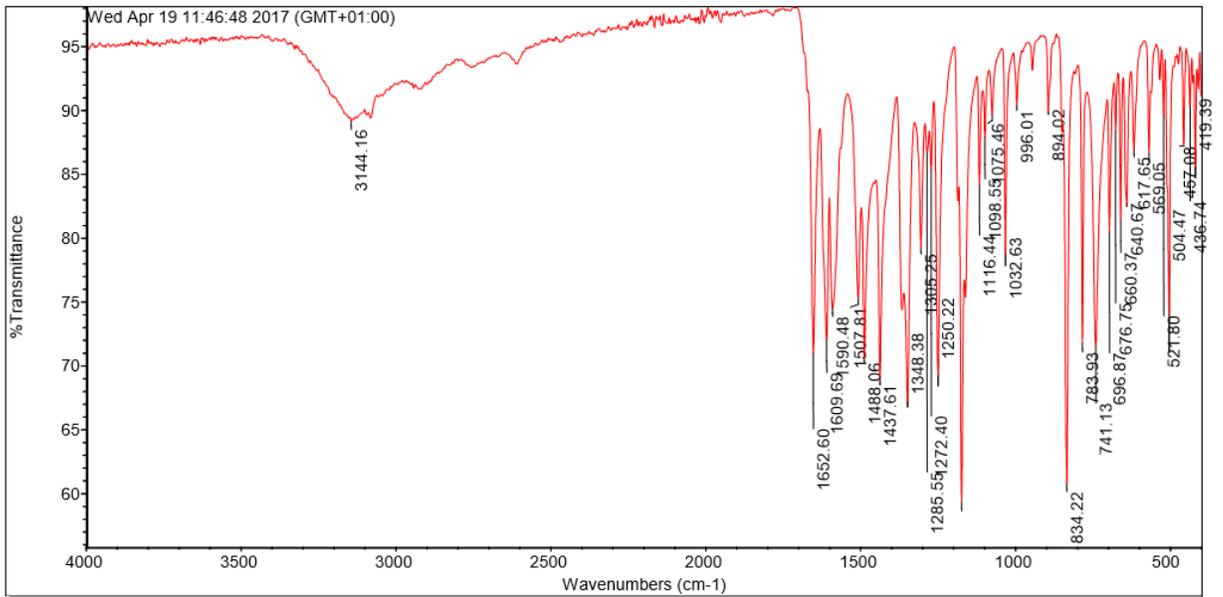
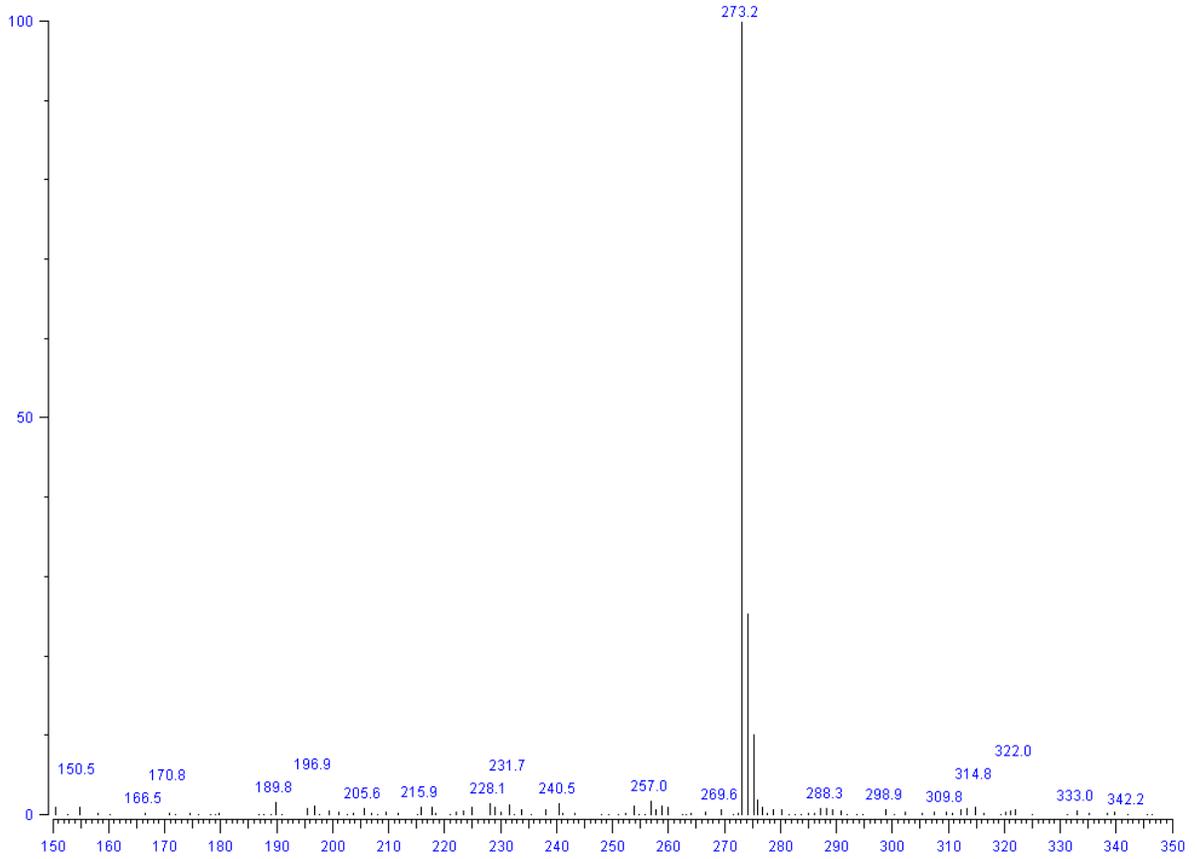




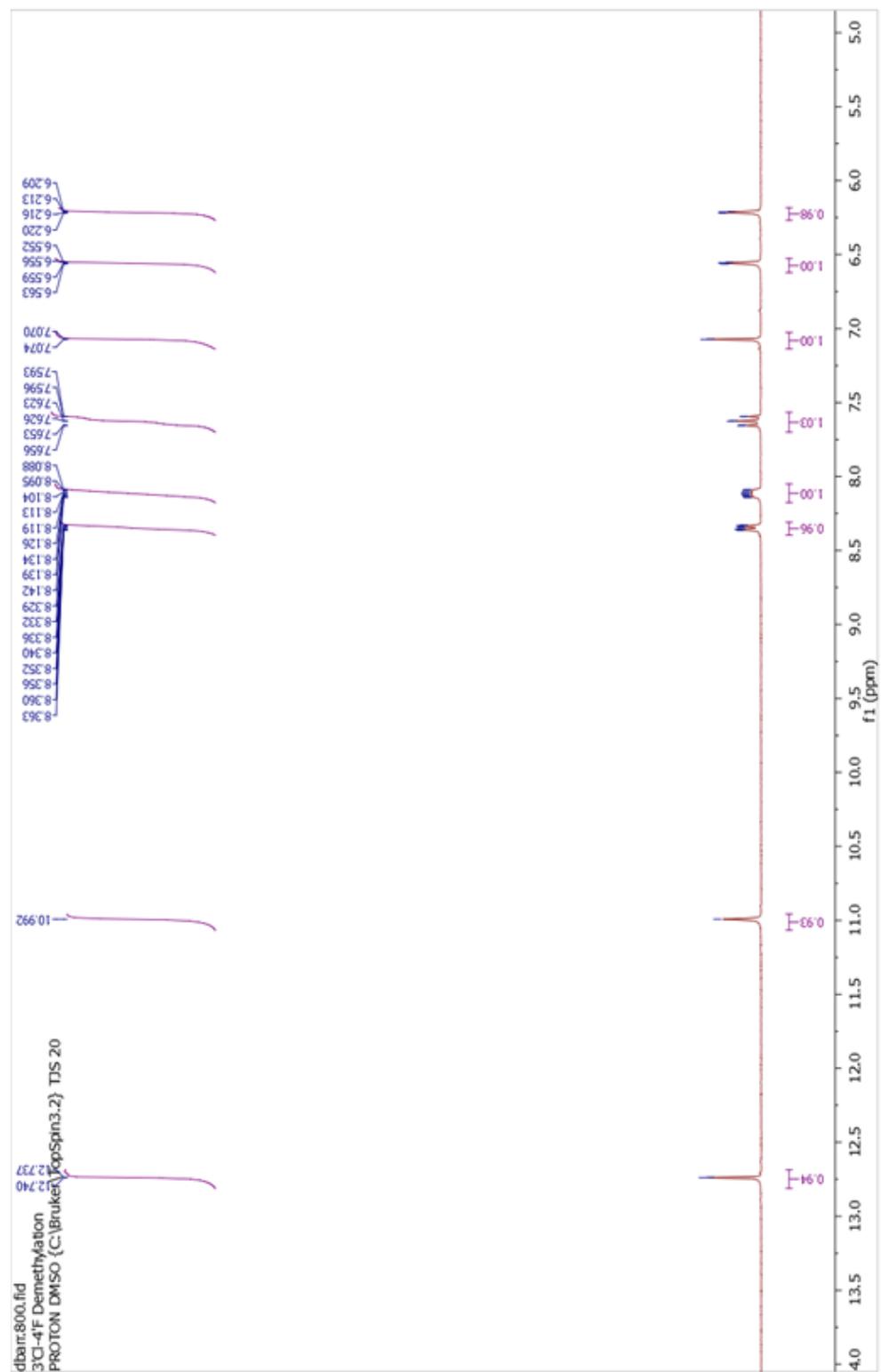
2-(3-fluorophenyl)-5,7-dihydroxy-chromen-4-one

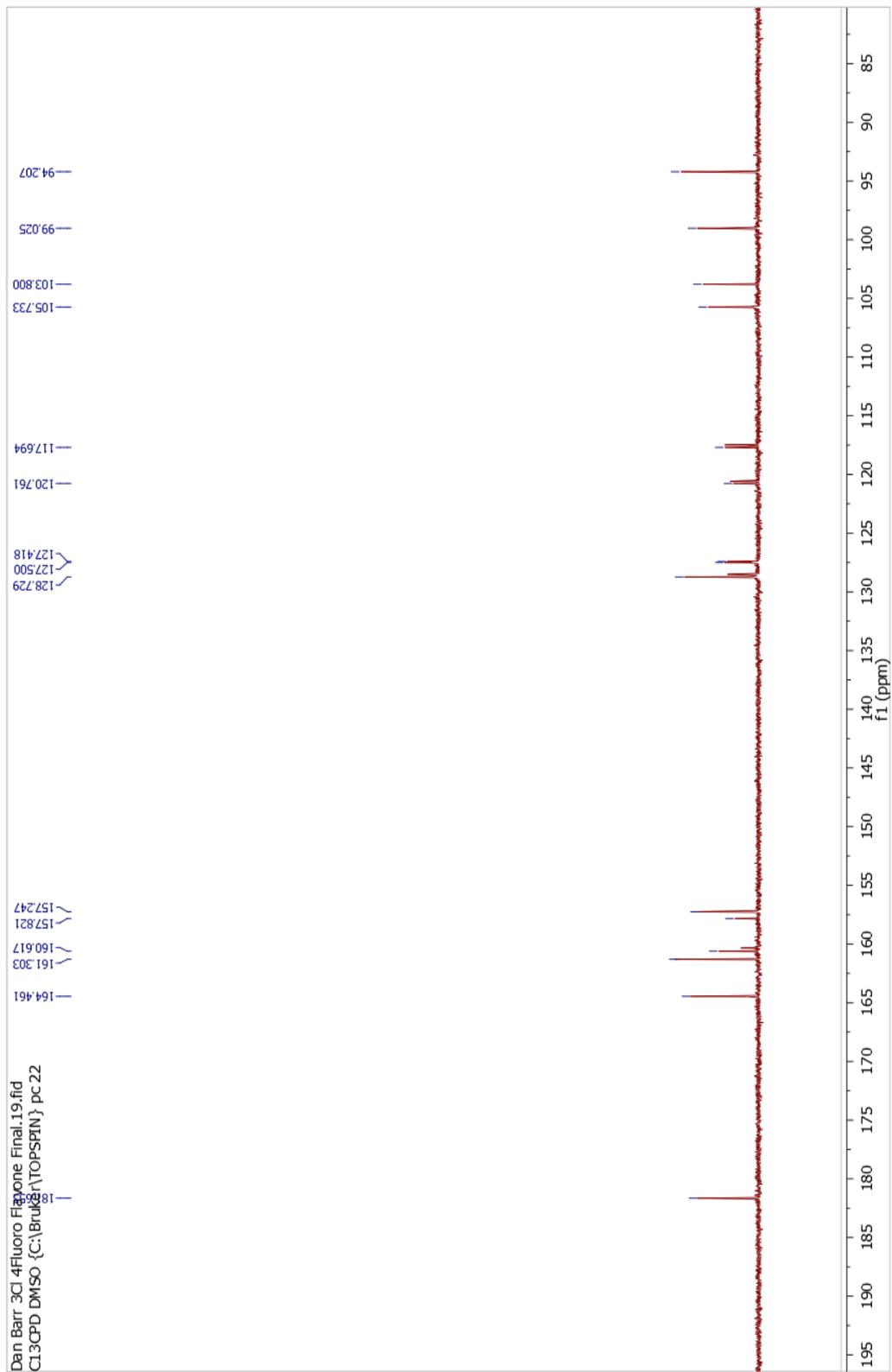


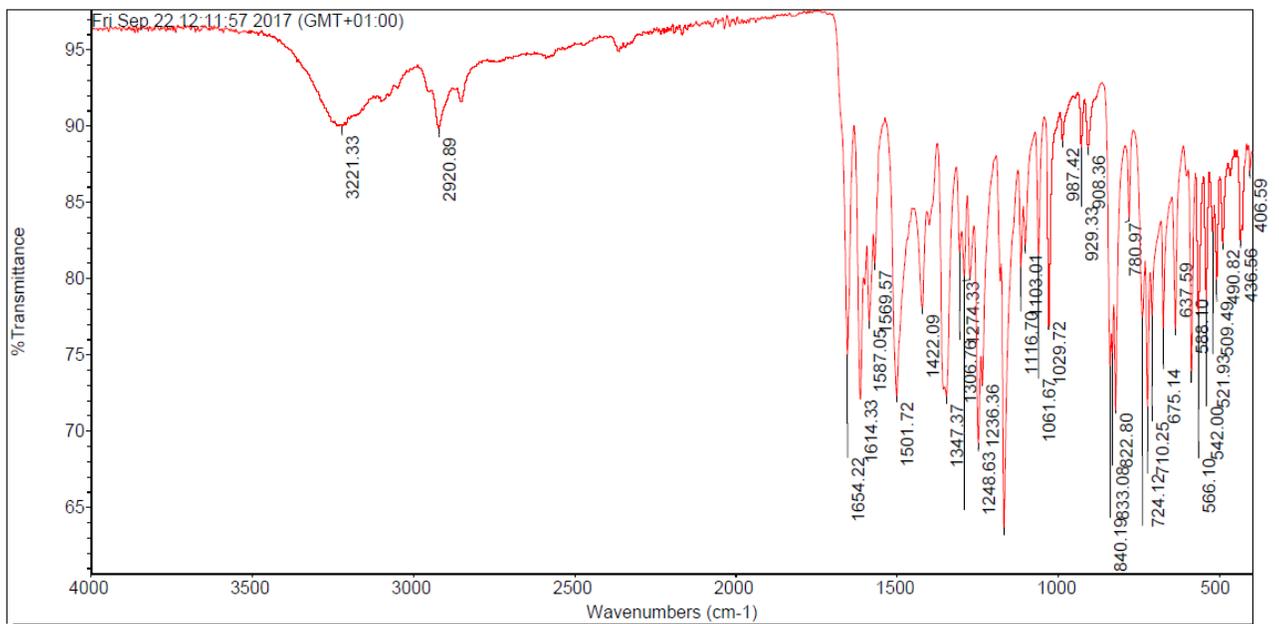
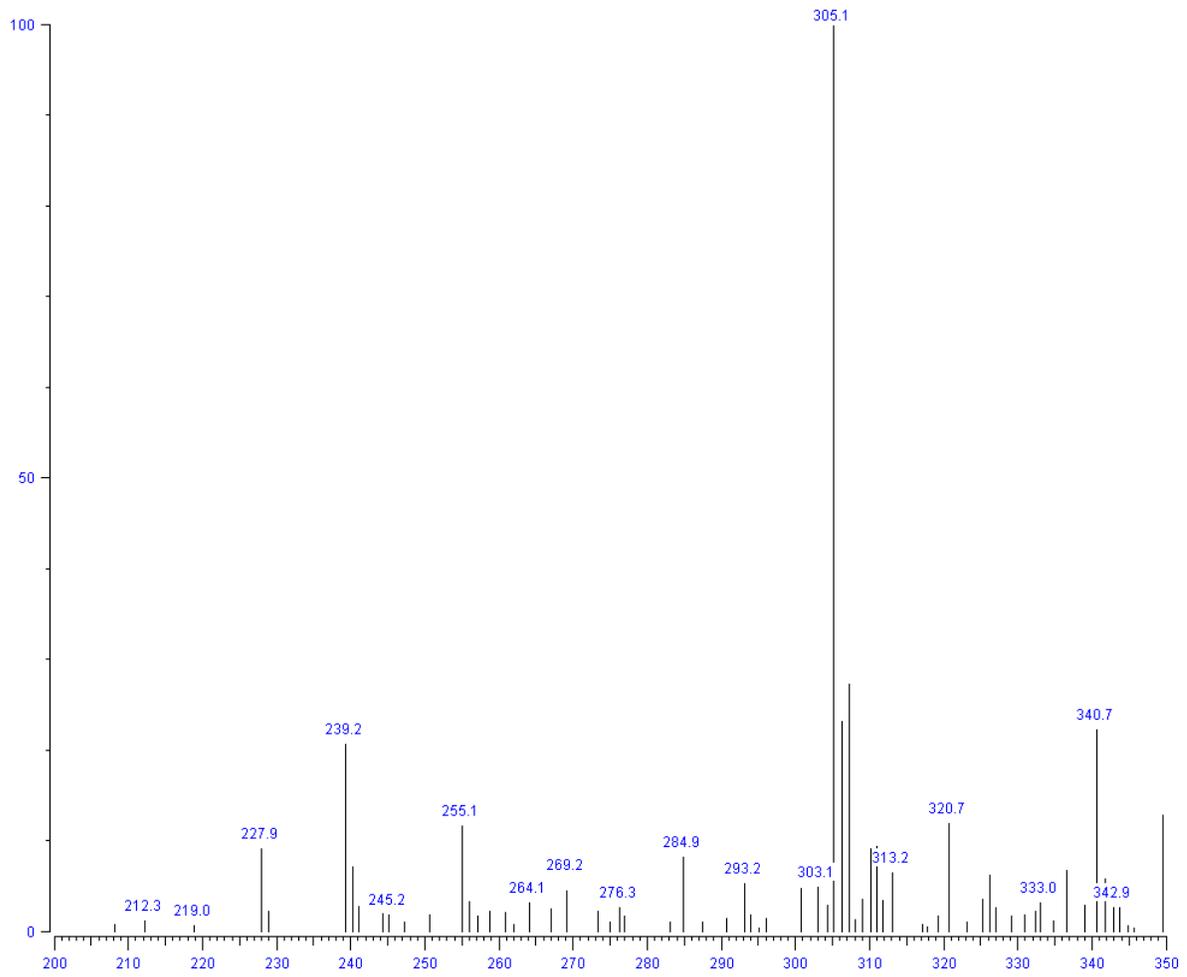




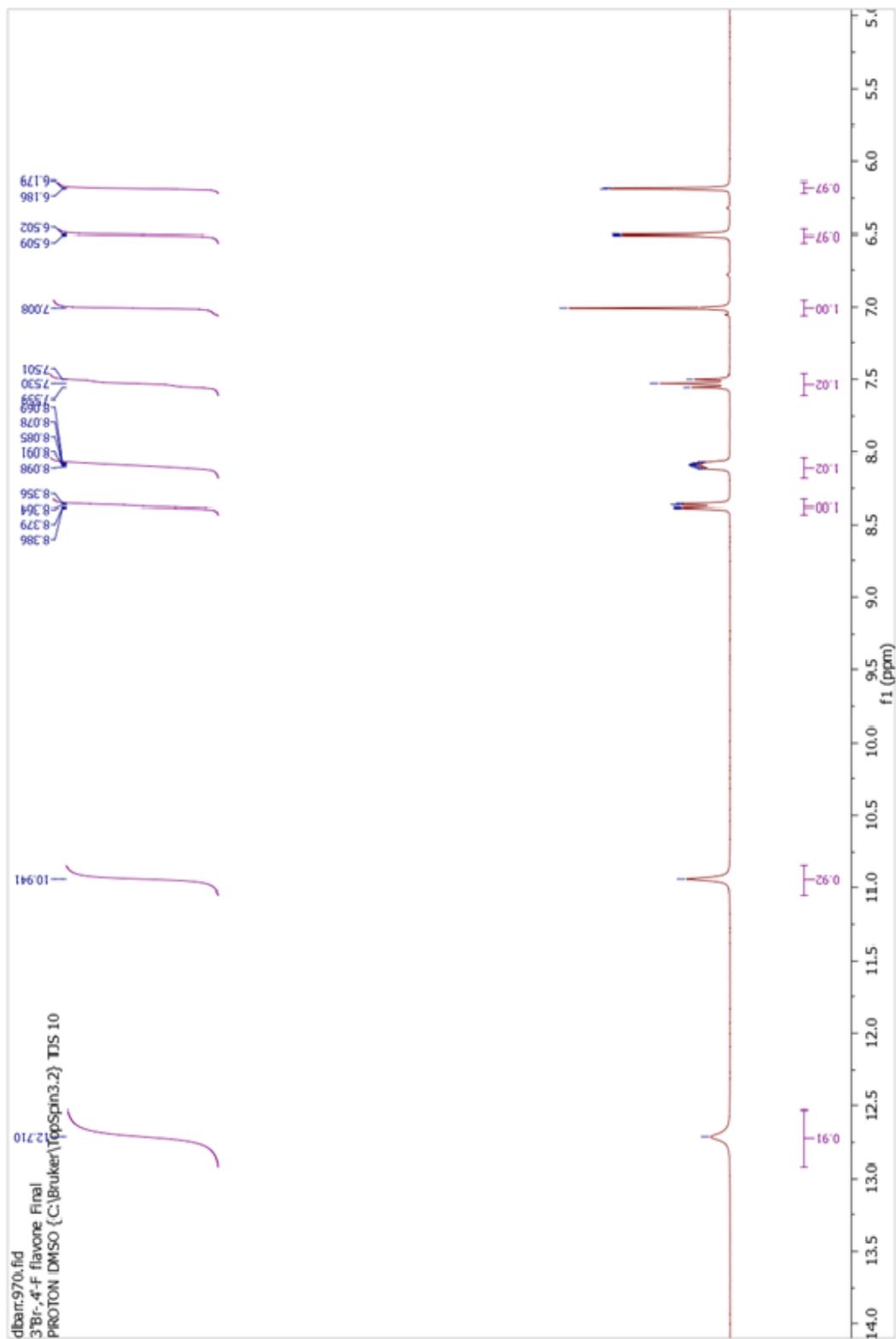
2-(3-chloro-4-fluoro-phenyl)-5,7-dihydroxy-chromen-4-one

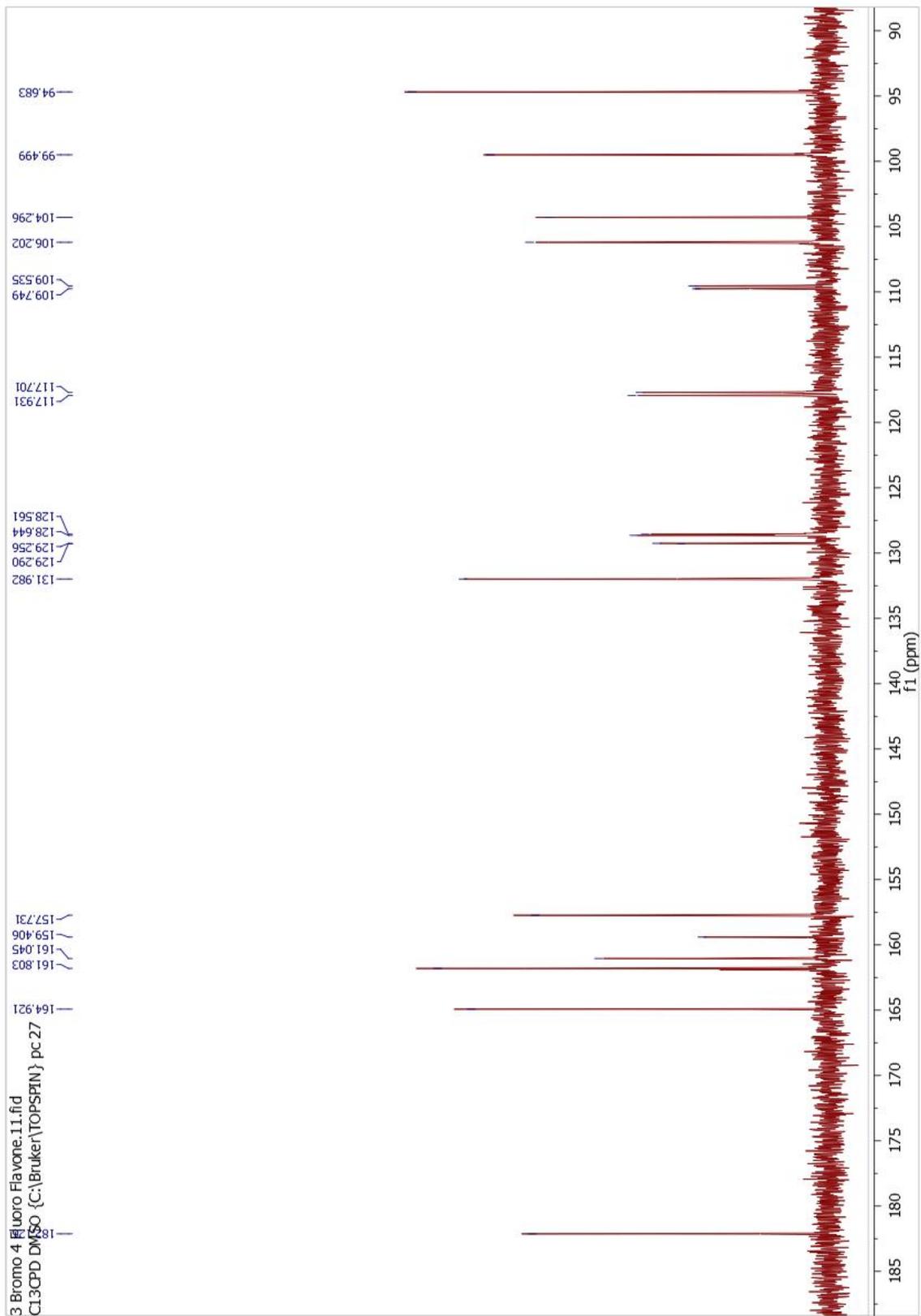


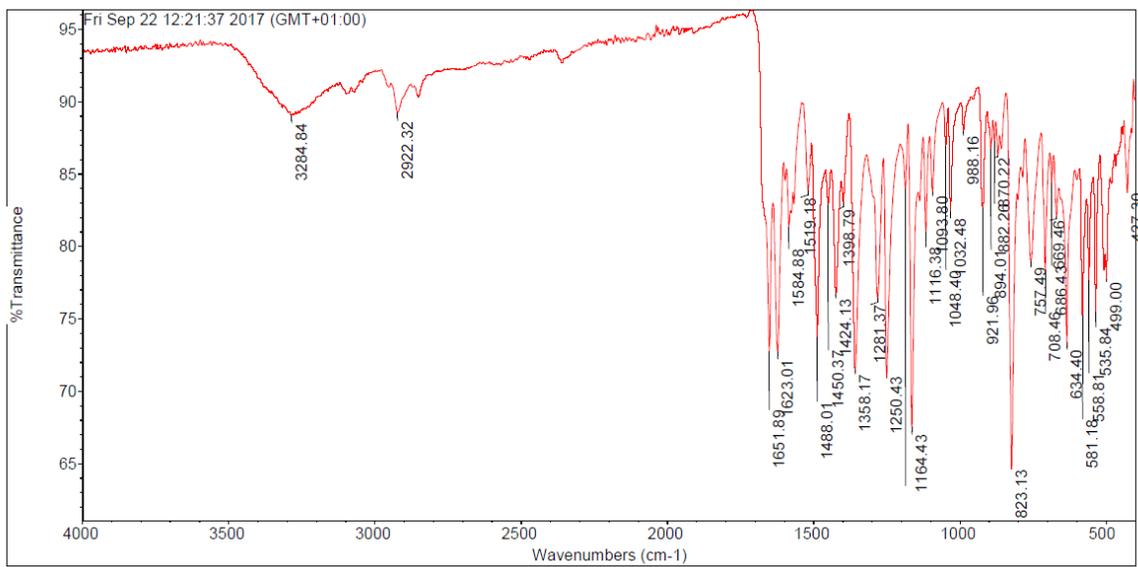
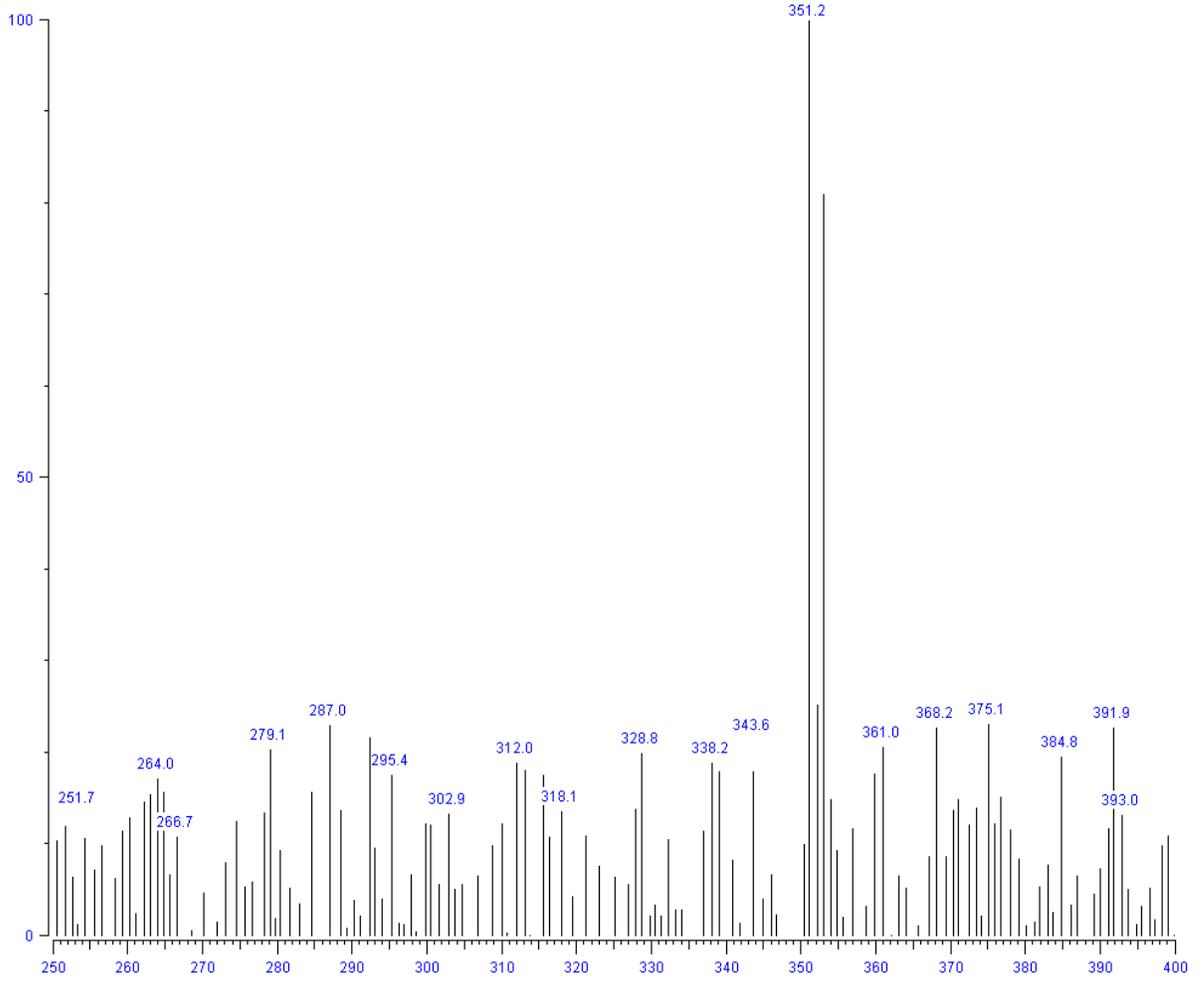




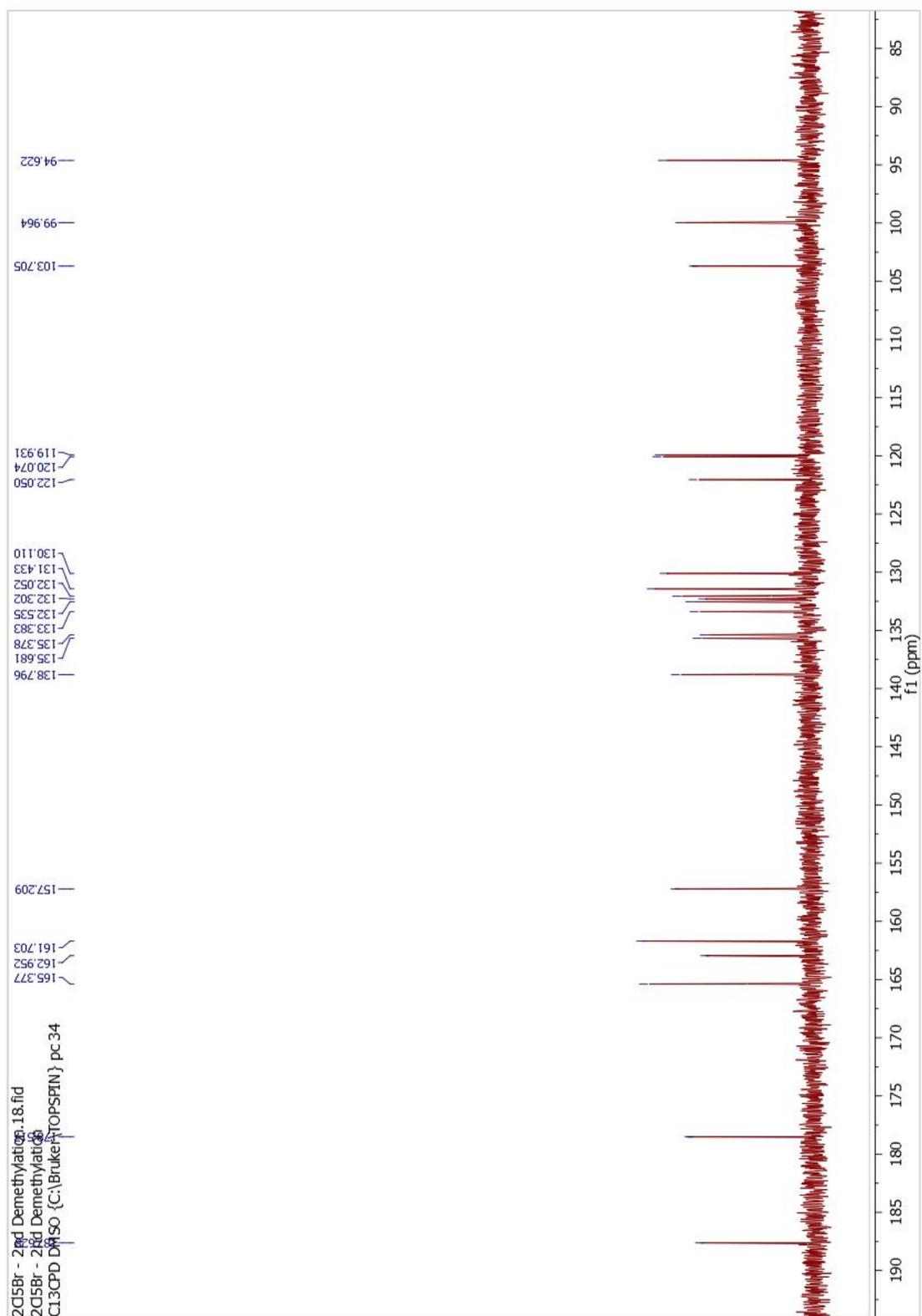
2-(3-bromo-4-fluoro-phenyl)-5,7-dihydroxy-chromen-4-one.

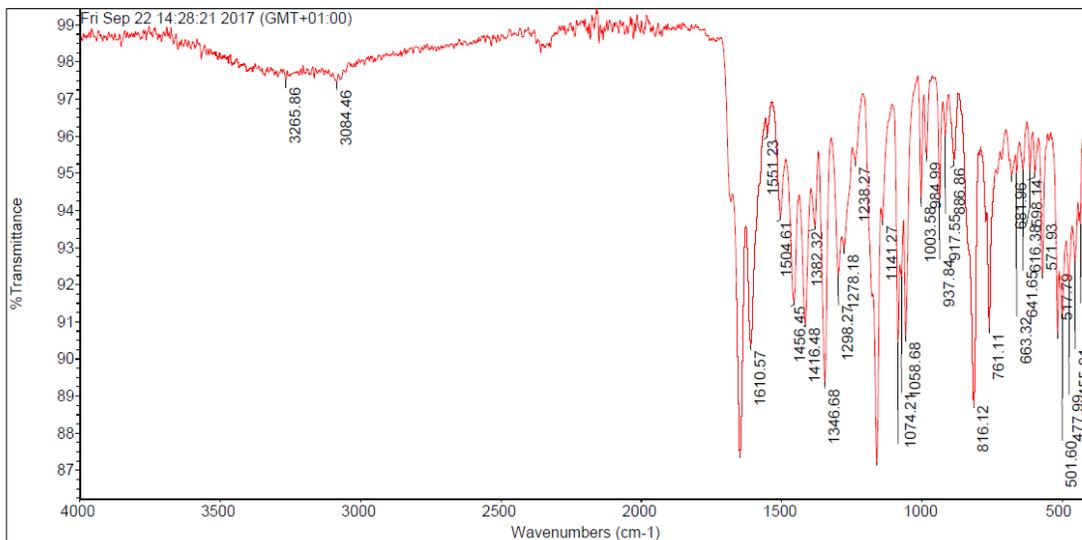
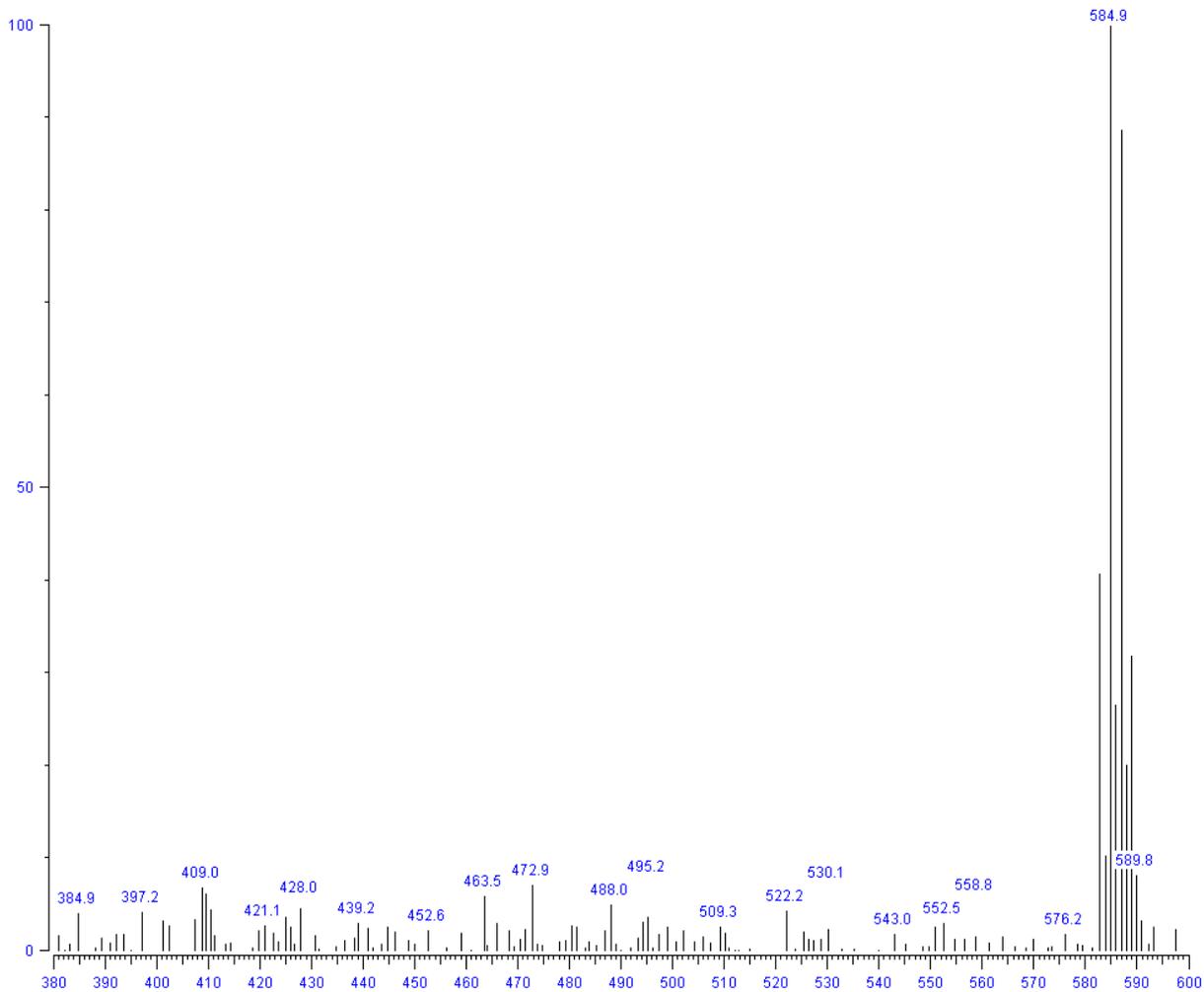




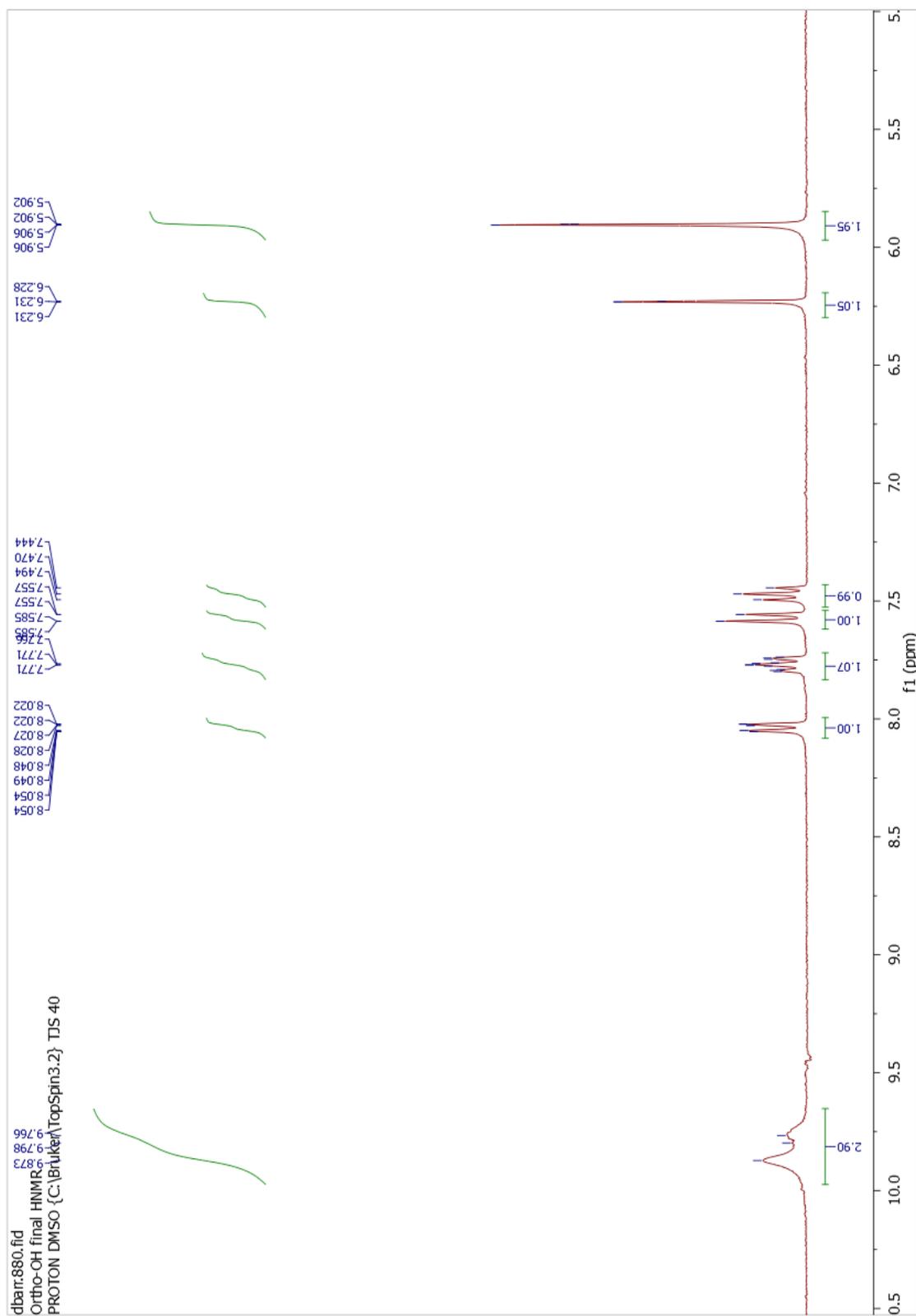


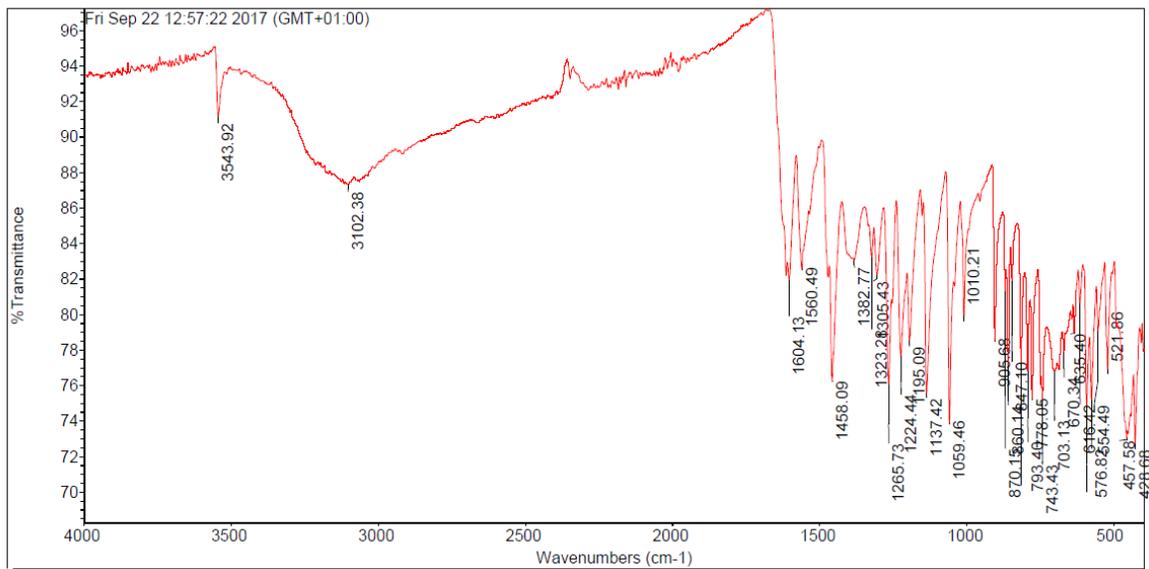
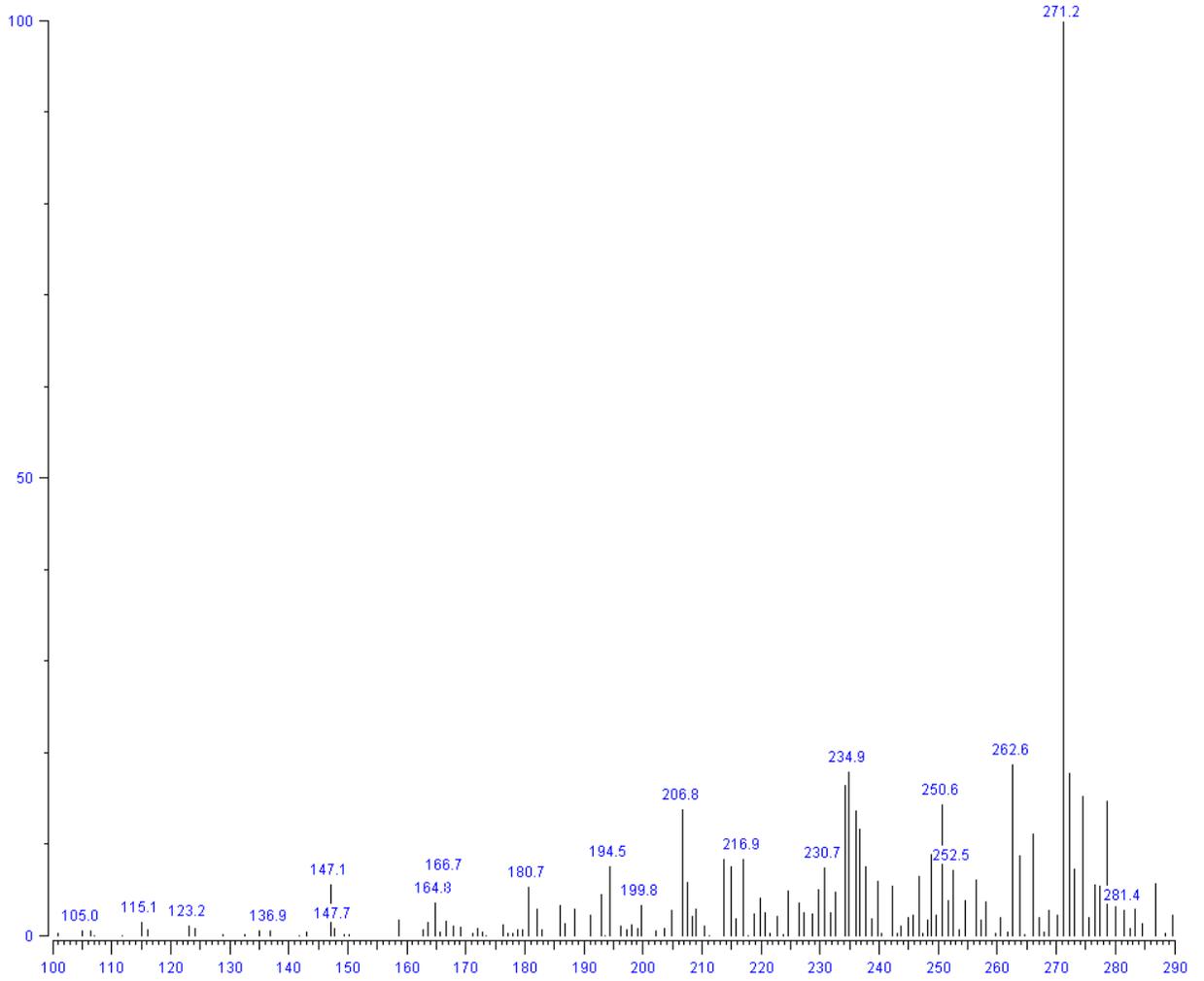
2-(5-bromo-2-chloro-phenyl)-5,7-dihydroxy-chromen-4-one





5,7-dihydroxy-2-(2-hydroxyphenyl)chromen-4-one





Scripts.

Protocol for running Prime-MMGBSA using multiple protein conformations generated from GOLD docking poses.

Aim: Docking with GOLD generates...

Generating Complexes from GOLD.

```
gold_utils -write_complexes -conf <conf_filename> -o <output directory> -format  
<output_format>
```

Splitting of the complex into receptor and ligand pose viewer

```
$SCHRODINGER/run pv_convert.py -mode split_pv <filename>
```

Deleting/Renaming files in a directory.

```
rename 's/gold_complex_LigandsForGOLD_//' gold_complex_LigandsForGOLD_*
```

To replace text

(to replace lig for instance)

```
$SCHRODINGER/prime_mmgsa -prime_opt OPLS_VERSION/OPLS3 -out_type COMPLEX - WAIT
```

Making a command executable

```
Chmod u+x
```

(to run the command) ./file > file.out

Running Prime

```
$SCHRODINGER/prime_mmgsa -prime_opt OPLS_VERSION/OPLS3 -out_type COMPLEX -WAIT  
<out_pv.mae>
```

Deleting text from files

```
:%s/<name>/ /g
```

Protocol for Generating Solvation of different residue orientations.

Cluster Waters

(with 0.5 Å cutoff distance)

```
$SCHRODINGER/run cluster_waters.py -c 0.5 <protein1> <protein2> <output_file.pdb>
```

(to keep only waters that superimpose across 100% of all proteins)

```
grep 1.00100.00 <output_file.pdb> > <name.pdb>
```