Doctoral   Science

2022

Development of novel anticancer agents targeting G protein coupled receptor: GPR120

Ajay Pal
Technological University Dublin, ajaypinfo@gmail.com

Development of novel anticancer agents targeting G protein coupled receptor: GPR120

Ajay Pal

Thesis submitted to Technological University Dublin

in fulfilment of requirement for the degree of

"Doctor of Philosophy"

School of Food Science & Environmental Health,

Environmental Sustainability and Health Institute (ESHI)

College of Sciences and Health

Technological University Dublin, Ireland

Supervised by:

Dr Gemma K. Kinsella

Prof. James F. Curtin

March 2022

Abstract

The G-protein coupled receptor, GPR120, has ubiquitous expression and multifaceted roles in modulating metabolic and anti-inflammatory processes. GPR120 - also known as Free Fatty Acid Receptor 4 (FFAR4) is classified as a free fatty acid receptor of the Class A GPCR family. GPR120 has recently been implicated as a novel target for cancer management. GPR120 gene knockdown in breast cancer studies revealed a role of GPR120-induced chemoresistance in epirubicin and cisplatin-induced DNA damage in tumour cells. Higher expression and activation levels of GPR120 is also reported to promote tumour angiogenesis and cell migration in colorectal cancer. A number of agonists targeting GPR120 have been reported, such as TUG891 and Compound39, but to date development of small-molecule inhibitors of GPR120 is limited.

This research applied a rational drug discovery approach to discover and design novel anticancer agents targeting the GPR120 receptor. A homology model of GPR120 (short isoform) was generated to identify potential anticancer compounds using a combined *in silico* docking-based virtual screening (DBVS), molecular dynamics (MD) assisted pharmacophore screenings, structure–activity relationships (SAR) and *in vitro* screening approach. A pharmacophore hypothesis was derived from analysis of 300 ns all-atomic MD simulations on apo, TUG891-bound and Compound39-bound GPR120 (short isoform) receptor models and was used to screen for ligands interacting with Trp277 and Asn313 of GPR120. Comparative analysis of 100 ns all-atomic MD simulations of 9 selected compounds predicted the effects of ligand binding on the stability of the "ionic lock" – a characteristic of Class A GPCRs activation and inactivation. The "ionic lock" between TM3(Arg136) and TM6(Asp) is known to prevent G-protein recruitment while GPCR agonist binding is coupled to outward movement of TM6 breaking the "ionic lock" which facilitates G-protein recruitment. The MD-assisted pharmacophore hypothesis predicted **Cpd 9**, (*2-hydroxy-N-{4-[(6-hydroxy-2-methylpyrimidin-4-yl) amino] phenyl} benzamide*) to act as a GPR120S antagonist which can be evaluated and characterised in future studies.

Additionally, DBVS of a small molecule database (~350,000 synthetic chemical compounds) against the developed GPR120 (short isoform) model led to selection of the 13 hit molecules which were then tested *in vitro* to evaluate their cytotoxic, colony forming and cell migration activities against SW480 – human CRC cell line expressing GPR120.

Two of the DBVS hit molecules showed significant (> 90%) inhibitory effects on cell growth with micromolar affinities (at 100 µM) - AK-968/12713190 (*dihydrospiro(benzo[h]quinazoline-5,1'-cyclopentane)-4(3H)-one*) and AG-690/40104520 (*fluoren-9-one*). SAR analysis of these two test compounds led to the identification of more active compounds in cell-based cytotoxicity assays – AL-281/36997031 (IC_{50} = 5.89–6.715 µM), AL-281/36997034 (IC_{50} = 6.789 to 7.502 µM) and AP-845/40876799 (IC_{50} = 14.16-18.02 µM). In addition, AL-281/36997031 and AP-845/40876799 were found to be significantly target-specific during comparative cytotoxicity profiling in GPR120-silenced and GPR120-expressing SW480 cells. In wound healing assays, AL-281/36997031 was found to be the most active at 3 µM (IC_{25}) and prevented cell migration. As well as in the assessment of the proliferation ability of a single cell to survive and form colonies through clonogenic assays, AL-281/36997031 was found to be the most potent of all three test compounds with the survival rate of ~ 30% at 3 µM.

The inter-disciplinary approach applied in this work identified potential chemical scaffolds – spiral benzo-quinazoline and fluorenone, targeting GPR120 which can be further explored for designing anti-cancer drug development studies.

Declaration

I certify that this thesis which I now submit for examination for the award of Doctor of Philosophy (PhD), is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work.

This thesis was prepared according to the regulations for graduate study by research of the Technological University Dublin and has not been submitted in whole or in part for another award in any other third level institution.

The work reported on in this thesis conforms to the principles and requirements of the TU Dublin's guidelines for ethics in research.

TU Dublin has permission to keep, lend or copy this thesis in whole or in part, on condition that any such use of the material of the thesis be duly acknowledged

Ajay Pal

Acknowledgements

List of Abbreviations

αLA	α-linolenic acid
AC	Adenyl cyclase
ADME-tox	Absorption, distribution, metabolism, excretion, and toxicity
AI	Artificial intelligence
AJCC	American Joint Committee on Cancer
ANGPT	Angiopoietin
BRAF	Proto-oncogene B-Raf
cAMP	Cyclic adenosine monophosphate
CADD	Computer aided drug design
CIN	Chromosomal instability
CIMP	CpG (5'-C-phosphate-G-3') island methylator phenotype
CRC	Colorectal cancer
DAG	Diacylglycerol
DHA	Docosahexaenoic acid
EGFR	Epidermal growth factor receptor
ECL	Extracellular loop
EPA	Eicosapentaenoic acid
FDA	Food and Drug Administration
FFA	Free fatty acids
FFAR	Free fatty acid receptor
FGF	Fibroblast growth factor
GDP	Guanosine diphosphate
GLP-1	Glucagon like peptide
GRK	G protein-coupled receptor kinases
GTP	Guanosine triphosphate
GPCR	G protein coupled receptor
HTS	High-throughput screening
IARC	International Agency for Research on Cancer
ICL	Intracellular loop
IP3	Inositol triphosphate
JAG1	Jagged1 signalling

KRAS	Ki-ras2 sarcoma viral oncogene
LCFA	Long-chain fatty acids
LPS	Lipopolysaccharide
mAb	Monoclonal Antibody
mCRC	Metastatic colorectal cancer
MD	Molecular dynamics
MLP	Molecular Libraries Program
MMP	Matrix metalloproteases
MSA	Multiple sequence alignment
MSI	Microsatellite instability pathway
NAMD	Negative allosteric modulating drug
NIH	National Institute of Health
NTRK	Neurotrophic tyrosine receptor kinase
Notch/DLL4	Notch/delta-like 4
PD-1	Programmed death-1
PD-L1	Programmed death-ligand-1
PDGF	Platelet-derived growth factor
PIP2	Phosphatidylinositol biphosphate
PKA	Protein kinase A
PLC	Phospholipase C
PUFA	Poly unsaturated fatty acids
RMSD	Root mean square deviation
SAR	Structure activity relationship
TM	Transmembrane
TNT	Total neoadjuvant therapy
VEGF	Vascular endothelial growth factor
VS	Virtual screening
WHO	World Health Organisation

Table of Contents

List of Figures

Figure 5.3: a) Percentage relative fold gene expression (by $\Delta\Delta C_T$ method) of GPR120 in GPR120-siRNA transfected SW480 cells within 72-hours of transfection. Results show mean

List of Tables

Chapter 1
Background and literature review

According to World Health Organization (WHO), cancer – also called neoplasm and malignant tumour, is a term used to define a large group of diseases affecting any part of the body. The defining feature of cancer is an abnormal and uncontrolled growth of cells that can be altered by various internal and external environmental factors (WHO, 2022). While a tumour is defined as mass of tissue due to abnormal proliferation of cells (Jiang, Puntis and Hallett, 1994). Cells produce signals to control how much and how often the cells divide. If there is a shift in the control mechanisms due to changes within the genes of a cell or a group of cells that govern cell proliferation and differentiation, cells may start to grow and multiply at an abnormal rate and form a group called a tumour (Figure 1.1) (Jiang, Puntis and Hallett, 1994; Luther and Chan, 2016).
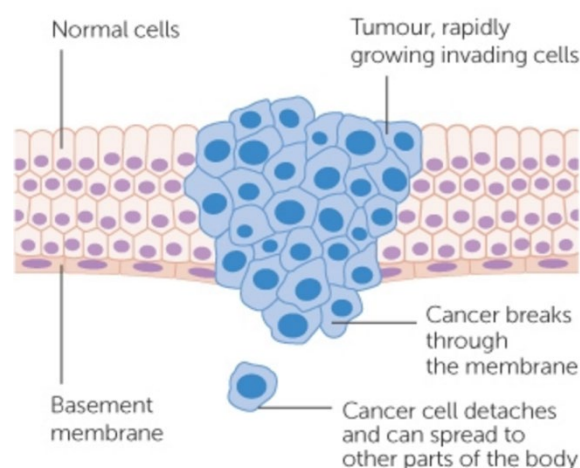
Figure 1.1: Graphical diagram of tumour growth through the basement membrane of epithelium / mesothelium / endothelium tissue; with detached cancer cells starting metastasis. (Image Copyright of Cancer Research UK)

A small subpopulation of cells within the tumour can be described as tumour stem cells, which retain the ability to undergo repeated cycles of proliferation as well as to migrate to distant sites in the body to colonize various organs in a process called metastasis (Figure 1.1). Tumour stem cells often have chromosome abnormalities reflecting their genetic instability, which leads to progressive selection of subclones that can survive more readily in the multicellular environment of the host (Luther and Chan, 2016; Palucka and Coussens, 2016).

Quantitative abnormalities in various metabolic pathways and cellular components accompany the tumour progression. The invasive and metastatic processes as well as a series of metabolic abnormalities resulting from the cancer cause illness and eventual death of the patient unless the neoplasm can be eradicated with treatment (Jiang, Puntis and Hallett, 1994; Luther and Chan, 2016).

The International Agency for Research on Cancer (IARC) in GLOBACAN cancer statistics survey for 2020 estimated 19.3 million new cases of cancer and ~10 million deaths from cancer worldwide (Ferlay et al., 2021). Lung cancer, liver cancer and stomach cancer were the top three causes of cancer death. Worldwide, lung, liver and stomach cancers were the top three causes of cancer death. In Europe, breast cancer, prostate cancer, colorectal cancer, and lung cancer are the top four diagnosed cancers accounting for 50% of all diagnosed cancers in both sexes (Bray et al. 2018; Ferlay et al. 2021). Colorectal cancer was the third leading cause of cancer deaths following prostate and lung cancer in males, and breast and lung cancer in females in 2020 (Ferlay et al., 2021). With an ageing population worldwide, colorectal cancer (CRC) is emerging as a formidable health problem (Arnold et al., 2016). According to the IARC GLOBOCAN reports from 2018 and 2020 (Figure 1.2), CRC remains the third most common cancer in both sexes (1,931,590 cases, 10% of the total in 2020) and the second in causing deaths due to cancer following lung cancer (935,173 cases, 9.4% of the total) worldwide with 55% of the cases occurring in more developed regions (Bray et al., 2018; Ferlay et al., 2021; Arnold et al., 2016).
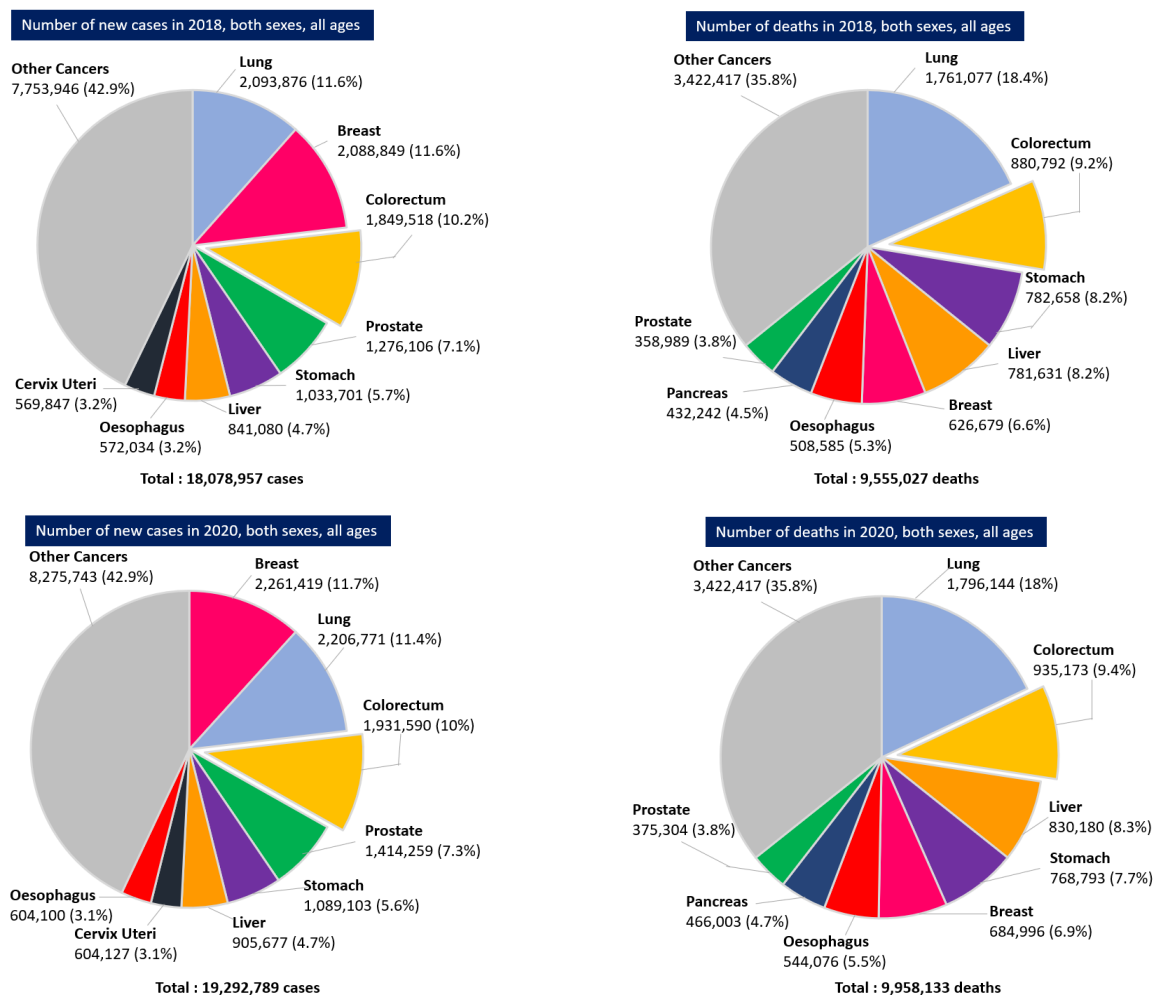
Figure 1.2: Graphical overview of incidence and mortality rates of different cancers worldwide in 2018 and 2020. (Bray et al., 2018; Ferlay et al., 2021)

1.1 Colorectal cancer

CRC is a cancer of the lower bowels in which tumour growth occurs in the tissues of the colon and/or rectum. As colon cancer and rectum cancer have many common features, they are mostly referred to as colorectal cancer (CRC). Although the prevalence of CRC has slowly increased since the 1960s worldwide, the mortality rate has decreased over the past decades due to improved treatments (Arnold et al., 2016). The rates of CRC incidence and mortality vary widely worldwide. Global patterns suggest that CRC incidence and mortality are rising rapidly in many middle-income developing countries linked to rapid westernisation of diet and lifestyle. Countries, such as Japan, have observed a substantial increase in the number of new cases of CRC whereas the African nations have the lowest recorded rates of CRC incidence and mortality (Arnold et al., 2016; Bray et al., 2018). Such trends in CRC incidence and mortality have been observed in Ireland as well (Figure 1.3, Table 1.1). The

number of new CRC cases recorded increased from 2,426 (in 2010) to 3178 (in 2020) which is corresponding to an increase in deaths from 942 (in 2010) to 1,282 (in 2020) in the Irish population (Cancer Incidence, Survival and Mortality Data - HSE.ie, 2019; Cancer today, 2020).
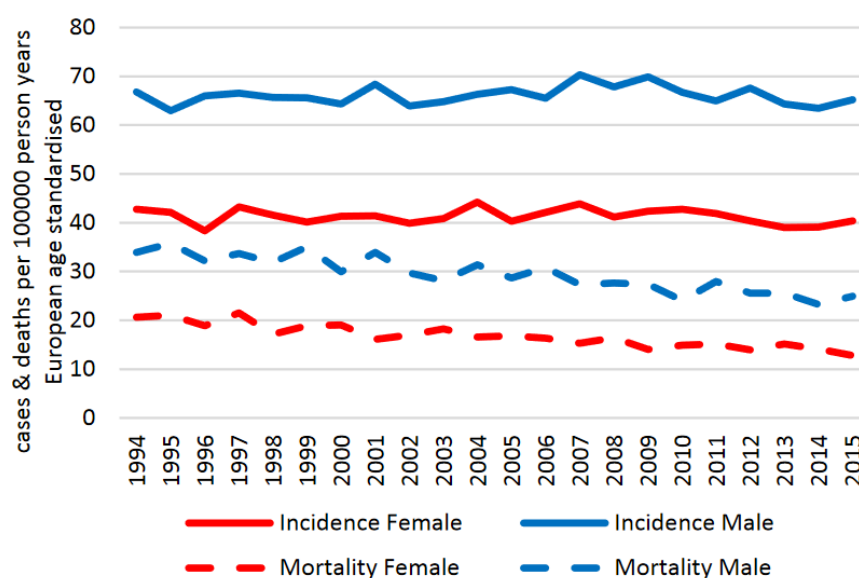
Figure 1.3: Trends in incidence and mortality of CRC cases in Ireland. Annual percentage changes from 1994 to 2015. (NCRI.ie, 2021)

In Ireland, CRC is the fourth most common, newly diagnosed cancer, among men and women (Table 1.1). The number of new cases is expected to increase significantly over the next 10 years, due mainly to an increasing and ageing population (Arnold et al. 2016). Cancer incidence projections 2020-2045 report of Ireland predicted that the annual number of CRC cases to increase in males from 1,021 in 2015 to 2,196 in 2045 (+115%) and in females from 776 in 2015 to 1,617 in 2045 (+108%) – an increase to 3,813 overall (+112%) (NCRI.ie, 2021). CRC is currently the second most common cause of cancer death in Ireland (up from the third in 2018) and about 40 percent of colorectal cancer patients die from the disease. With increasing projections of CRC incidence, CRC related deaths will likely also increase. (Arnold et al. 2016; Cancer incidence projections for Ireland 2020-2045, 2021). Relatively few CRCs occur in the population below 40 years of age. However, rates increase rapidly with age, more markedly for colorectal cancer (Keum and Giovannucci, 2019; Arnold et al. 2016).

Table 1.1: Top five cancers sited from estimated number of new cases and deaths in 2020 recorded in Ireland both sexes and individually for male and female population – IARC GLOBOCAN 2020 (Cancer today, 2020)

Table 1.1A: Top 5 cancer sites by number of cases and deaths in both sexes			
Incidence case count		**Mortality case count**	
Prostate	4503	Lung	2232
Breast	3433	Colorectum	1282
Lung	3271	Breast	745
Colorectum	3178	Pancreas	618
Skin	1316	Prostate	569
Table 1.1B: Top 5 cancer sites by number of cases and deaths in male population			
Incidence case count		**Mortality case count**	
Prostate	4503	Lung	1194
Colorectum	1856	Colorectum	749
Lung	1643	Prostate	569
Bladder	729	Pancreas	358
Kidney	673	Oesophagus	347
Table 1.1C: Top 5 cancer sites by number of cases and deaths in female population			
Incidence case count		**Mortality case count**	
Breast	3433	Lung	1038
Lung	1628	Breast	745
Colorectum	1322	Colorectum	533
Corpus uteri	808	Ovary	297
Skin	772	Pancreas	260

1.1.1 Aetiology of CRC

CRC results from the progressive accumulation of genetic mutations and epigenetic alterations that lead to the transformation of normal colonic epithelium to colon adenocarcinoma. Epigenetic alterations are the inheritable changes in gene activity and expression that occur without an alteration in DNA sequence (Keum and Giovannucci, 2019).

The molecular genesis of CRC is based on a multistep event that may take several years to develop full scale malignant tumours (Fearon and Vogelstein, 1990; Keum and Giovannucci, 2019; Lengauer, Kinzler and Vogelstein, 1998). These can include:

1) Genetic and epigenetic (non-genetic influences on gene expression) alterations that underlie and promote the colon tumour formation process because they provide a clonal growth advantage to the cells that acquire them. (Hong, 2018; Jung et al., 2020)

2) The loss of genomic stability from point mutations to chromosomal rearrangements which leads to tumour development.

3) The hereditary cancer syndromes which frequently correspond to germ line forms of key genetic defects whose somatic occurrences drive the emergence of sporadic colon cancers.

Tumour development in sporadic CRC is known to arise from different pathways such as: chromosomal instability (CIN), microsatellite instability pathway (MSI) and CpG (5'-C-phosphate-G-3') island methylator phenotype (CIMP) (De Rosa et al., 2015).

1.1.1.1 Chromosomal instability (CIN)

CIN is characterised by an accumulation of mutations in tumour suppressor genes (such as APC, MADR2, and p53 genes) and oncogenes (such as KRAS, ErbB2, PI3KCA, and CCND1 genes), which lead to transformation of normal colonic epithelium to colon adenocarcinoma (Dunican et al., 2002). CIN occurs in 60-70% of all CRC cases, and characteristically are aneuploid (the presence of an abnormal number of chromosomes in a cell – one less or one more); highly differentiated, with no lymphocytic infiltration; have a poor prognosis, and no specific tumour site (colon or rectal) predominance (Dunican et al., 2002; Arriba et al., 2015).

1.1.1.2 Microsatellite instability (MSI)

MSI refers to repetitive DNA expansions and contractions in the cell. Molecular causes of the MSI phenotype are DNA replication and repair defects with 15% occurrence in CRC, (Dunican et al., 2002) and is predominantly caused by hyper methylation of the promoter region of MLH1 (encode for MutL homolog 1), resulting in transcriptional silencing (Koyuncuer and Ozkan, 2020). MSI results in inactivation of the mismatch repair system, and thereby failure to correct nucleotide mismatches, which causes a high frequency of mutations in coding and non-coding regions of repetitive sequences throughout the cancer genome. Typical of MSI tumours are frameshift mutations in specific genes such as β-catenin, transforming growth factor β receptor II (TGFβRII), epidermal growth factor receptor (EGFR) or Bcl-2-associated X protein (BAX) (Koyuncuer and Ozkan, 2020). MSI tumours are also associated with larger tumour size and are more frequently observed in women (Koyuncuer and Ozkan, 2020). Rectal tumours exhibiting MSI are rare, in terms of occurrence; they are often associated with Lynch syndrome – a hereditary disorder caused by

a mutation in a mismatch repair gene in which affected individuals have a higher-than-normal chance of developing CRC, endometrial cancer, etc. (Razvi, Giardiello and Law, 2017).

1.1.1.3 CpG island methylator phenotype (CIMP)

CpG sites or CG sites are regions of DNA where a phosphate links a cytosine and a guanine in the linear sequence of bases (5'—C—phosphate—G—3'), that is, cytosine and guanine separated by only one phosphate. CpG islands are part of the promoter of ~50% of all genes (Tapial et al., 2019). Promoter CpG island hyper methylation results in inactivation of tumour suppressor and DNA repair genes causing transcriptional silencing, which reflects an epigenetic change (Bae et al., 2016). The MLH1 gene is frequently inactivated, which is reported in most CRCs and has also been identified in adenomas (Bae et al., 2016; Tapial et al., 2019). CIMP tumours are associated with proximal colon localisation, older age, MSI, high frequency of BRAF (proto-oncogene B-Raf) and KRAS (Ki-ras2 sarcoma viral oncogene) mutations and poor differentiation (Bae et al., 2016).

1.1.2 Pathology and staging of CRC

Colon and rectum walls are made up of several layers which are (from inner to outside layer): mucosa, submucosa, muscularis propria and serosa (serosa is not found on most of the rectum). CRC starts in the innermost layer (the mucosa) in the form of non-cancerous polyps (Figure 1.4) and its roots can grow through some or all the other layers. When cancer cells are in the wall; they can then grow into blood vessels or lymph vessels (tiny channels that carry away waste and fluid). From there, they can travel to nearby lymph nodes or to distant parts of the body. The extent of spread of CRC depends on its depth into the intestinal wall and if it has spread outside the colon or rectum (Kay Washington, 2008).
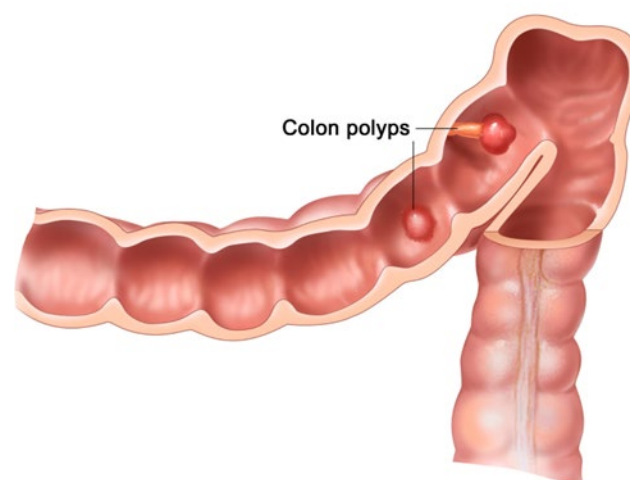
Figure 1.4: Graphical view of colonic polyps.
(Image Copyright of Cancer Research UK)

Polyp is a non-specific term for the growth on the inner surface of the colon or rectum which may eventually develop into cancer. The two-common type of polyps found in colorectal regions are: 1) Hyperplastic and inflammatory polyps (these are generally non-cancerous); 2) Adenomas / adenomatous polyps. Adenomas are pre-malignant lesions most likely to become cancers, about 96% of CRCs are adenocarcinomas (cancer developed from adenomas), which evolve from glandular tissue (Uyar, 2020; Stewart et al., 2006). These start in the glandular cells that make mucus to lubricate the inside of the colon and rectum (Kay Washington, 2008; Dubé et al., 2017). Adenomas are sub classified as tubular, tubulovillous and villous based on their architectural features (Dubé et al., 2017).

1.1.2.1 Cancer staging systems

Staging defines the severity and extent of cancer. The American Joint Committee on Cancer (AJCC) established the tumour/node/metastasis (TNM) system, which is most commonly used for cancer staging (AJCC Cancer Staging Manual, 7th Edition).   The TNM system assigns a number based on:

- Degree of invasion of intestinal wall – "T"

    - TX : Primary tumour which cannot be evaluated.

    - T0 : Primary tumour is not evident.

    - Tis : Early cancer stage limited to specific tissue / area.

    - T1-T4 : Size and/or extent of tumour.

- Degree of lymphatic node involvement – "N"

    - NX: Regional lymph nodes cannot be evaluated.

    - N0: Cancer is not present in lymph nodes.

    - N1-N3: Number/extent of lymph nodes involved.

- Degree of metastasis – "M"

    - M0: Cancer cells have not metastasised.

    - M1: Cancer cells metastasised to other tissues/organs.

Different cancers have different classification systems in which the letters and numbers may not have the same meaning. The broader stage of a cancer is defined by I, II, III, IV derived from the TNM values grouped together. The TNM staging of CRC in humans is described in table 1.2 (Tong et al., 2018).

Table 1.2: TNM staging of CRC in prognostic groups (Tong et al., 2018)

AJCC stages	TNM stages	Prognosis
Stage 0	Tis N0 M0	Tis: Tumour confined to mucosa; cancer-in-situ
Stage I	T1 N0 M0	T1: Tumour invades submucosa
Stage I	T2 N0 M0	T2: Tumour invades muscularis propria
Stage II-A	T3 N0 M0	T3: Tumour invades sub serosa or beyond (without other organs involved)
Stage II-B	T4 N0 M0	T4: Tumour invades adjacent organs or perforates the visceral peritoneum (the inner lining of abdominal cavity)
Stage III-A	T1-2 N1 M0	N1: Metastasis to 1 to 3 regional lymph nodes. T1 or T2.
Stage III-B	T3-4 N1 M0	N1: Metastasis to 1 to 3 regional lymph nodes. T3 or T4.
Stage III-C	any T, N2 M0	N2: Metastasis to 4 or more regional lymph nodes. Any T.
Stage IV	any T, any N, M1	M1: Distant metastases present. Any T, any N.

CRC stage progression, tumour development and metastasis are the result of essential cell physiology alterations such as: insensitivity to growth inhibitory signals; self-sufficiency in growth factors; evasion of apoptosis; unchecked replicative potential; reprogrammed energy metabolism; evasion of immune destruction and tumour angiogenesis (Tong et al., 2018; Mathonnet, 2014; Salem, Puccini and Tie, 2020). The formation of new vasculature (angiogenesis) is reported to be actively involved in the metastasis of CRC, which can turn even benign tumours to malignant tumours (Salem, Puccini and Tie, 2020).

1.2 Tumour angiogenesis in CRC

Angiogenesis is the growth of new blood vessels from existing vessels and involves the migration, growth, and differentiation of endothelial cells to line the inside wall of blood vessels. The term comes from the Greek words: *angio*, meaning blood vessel, and *genesis*, meaning beginning (Folkman, 2007).

The process of angiogenesis is controlled by chemical mediators (illustrated in Figure 1.5) in the body. These mediators can stimulate both the repair of damaged blood vessels and the formation of new blood vessels (angiogenesis). Other chemical factors, called angiogenesis inhibitors, interfere with blood vessel formation. Normally, the stimulating and inhibiting effects of these chemical factors are in equilibrium so that blood vessels form only when and where they are required (Adams and Alitalo, 2007; Folkman, 2007; Yin et al., 2020). In healthy adults, a balance of growth factor signalling maintains endothelial cells in a quiescent or resting state.

Blood vessels are comprised of an inner lining of closely assembled endothelial cells sheathed by pericytes, (the basement membrane) embedded in the stromal compartment (various stromal cells and extracellular matrix) (Yin et al., 2020). But conditions such as hypoxia or other endogenous signals activate these cells and induce the release of signalling factors, such as vascular endothelial growth factor (VEGF), fibroblast growth factor (FGF) and chemokines. VEGF is secreted by cells (e.g., tumour cells) and binds to its receptor (VEGF receptor 2) and neuropilin-I receptor on endothelial cells (Figure 1.5). VEGF is the most common of at least six other pro-angiogenic proteins released from tumours (Adams and Alitalo, 2007; Folkman, 2007).

For capillary formation (Figure 1.5), a tumour cell stimulated / influenced by Neuropilin, VEGF/VEGFR, notch/delta-like 4 (Notch/DLL4) and Jagged1 (JAG1) signalling and releases matrix metalloproteases (MMP) to degrade the basement membrane and remodel the extracellular matrix. MMPs, released from tumour cells, mobilize pro-angiogenic proteins from stroma (Adams and Alitalo, 2007).

Tumour cells secrete angiopoietin 2 (ANGPT2), which is pro-angiogenic and competes with ANGPT1 (housekeeping protein / anti-angiogenic) for binding to the endothelial Tyrosine-Protein Kinase Receptor Tie-2 receptor (TEK). ANGPT2 increases the degradation of the vascular basement membrane and migration of endothelial cells, therefore facilitating capillary / sprout formation (Adams and Alitalo, 2007; Folkman, 2007) (Figure 1.5). Tumour

cells are polarized and extend numerous filopodia to guide sprout migration (via semaphorins, ephrins, and integrins guidance signals) towards angiogenic stimuli (VEGF gradient) (Adams and Alitalo 2007).

Platelet-derived growth factor (PDGF) and basic fibroblast growth factor (bFGF; also known as FGF2) are angiogenic proteins secreted by some tumour cells (Folkman, 2007). Endothelial cells also contain integrins, which carry signals both upstream and downstream to facilitate endothelial cell binding to extracellular membranes, helping the cells to maintain viability and responsiveness to growth regulatory proteins (Folkman, 2007).

Fusion of neighbouring branches occurs when tumour cells encounter each other, establish endothelial cell junctions, and form a continuous lumen of blood vessel. An extracellular matrix is deposited to establish a new basement membrane; endothelial cell proliferation ceases, and pericytes are recruited to stabilize the new vessel. Once blood flow is established, the perfusion of oxygen and nutrient reduces angiogenic stimuli (VEGF expression) and inactivates endothelial cell oxygen sensors, re-establishing the quiescent state of the blood vessel (Adams and Alitalo, 2007).

The growth and proliferation of cancer cells, particularly in metastatic colorectal cancer (mCRC), depends essentially on tumour angiogenesis through various signalling pathways. Tumour angiogenesis generates neovascularization (formation of new blood vessels) in response to cellular need for nutrients and oxygen. During this process, the angiogenic switch (such as VEGF and EGFR) is activated leading to new vessel sprouting that sustain the expanding tumour growth (Adams and Alitalo, 2007; Folkman 2007). The process of cell growth, cell death, cell migration and alterations in vascular matrix is a complex system of vascular remodelling that is controlled by manifold physical, biochemical, and genetic components. Molecular triggers commonly derived from tumour cells imbalance the process of vascular remodelling leading to tumour progression (Adams and Alitalo, 2007; Yin et al., 2020).
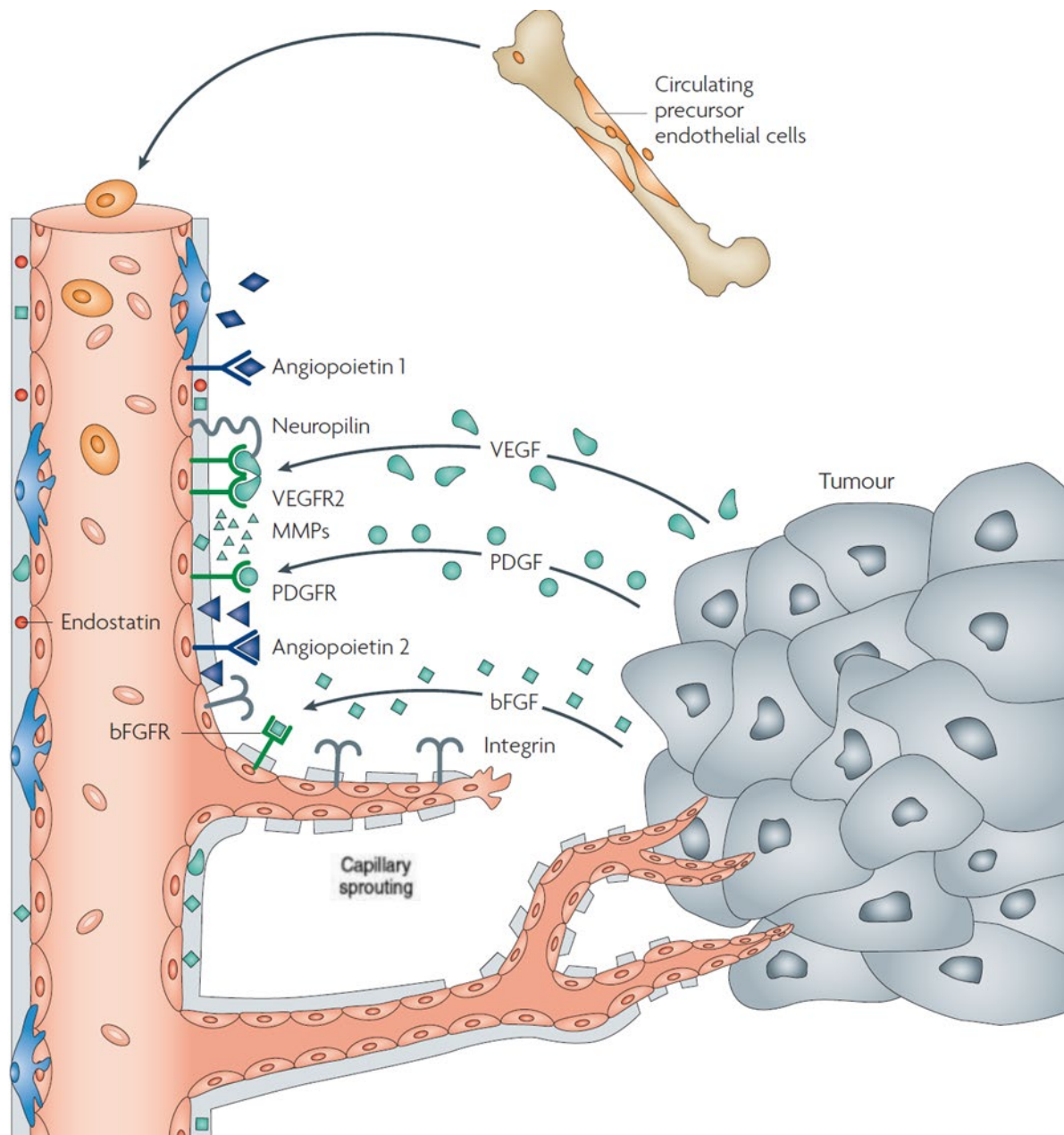
Figure 1.5: Key factors involved in the angiogenic switch in tumour angiogenesis. The angiogenic switch is the interactional initiation between the tumour and its microenvironment. Tumour-derived chemoattractants (ANGPT1, ANGPT2, MMP) promote the secretion of multiple angiogenic growth factors and proteinases (VEGF, bFGF, PDGF). Reproduced with copyright permission from Nature Publishing Group (Folkman 2007).

1.3 Treatment of CRC

In developed countries widespread screening programmes – especially for risk populations are the primary mode of early detection leading to earlier diagnosis and treatment have been successful in preventing late-stage CRC development. The treatment of CRC is individualised and based on the CRC staging to provide the patient with the best quality of life after survival. For patients with localised and potentially curable CRC, surgery is the primary treatment. The localised stage-dependent relative five-year survival rate (~64%) has improved in CRC patients (SEER, 2021; Xie, Chen and Fang, 2020) due to a combination of curative resection and evolved therapy implementation. Radiation and chemotherapy with recent targeted therapies have played an important role in the advancement of CRC treatment. Preoperative and postoperative radiotherapy studies have shown decreased risk of CRC recurrence (Sauer et al., 2012). A new approach termed total neoadjuvant therapy (TNT) is being practised for advanced stage CRC patients where chemotherapy and chemoradiation therapy are given six months before the surgery resulting in increased complete pathologic response and prolonged disease-free survival rate of ~39% (Cercek et al., 2018; Petrelli et al., 2020). In brief, these therapeutics-based CRC treatments can be divided into chemotherapy, targeted therapy, and immunotherapy (Xie, Chen and Fang, 2020; Petrelli et al., 2020).

1.3.1 Chemotherapy

A regimen of chemotherapy consists of a specific number of cycles over a set period during which the patient may receive a combination of different anticancer drugs based on the therapist's discretion. The commonly used chemotherapy regimens are summarised in Table 1.3 (Biller and Schrag, 2021) along with the associated toxicities and adverse effects.

Table 1.3: Commonly used chemotherapy regimens for CRC treatment. (Biller and Schrag, 2021)

Regimen	Component drugs	Mechanism of action / drug target	Toxicities and adverse effects	First-line use
FOLFOX: the most commonly used adjuvant regimen.	fluorouracil, leucovorin, and oxaliplatin	Inhibition of DNA synthesis in cancer cells: **fluorouracil** inhibits the formation of thymidylate - the precursor of thymidine triphosphate, which is essential for DNA synthesis. **Leucovorin** enhances the activity of fluorouracil by stabilizing the interactions of its active metabolite to the enzyme thymidylate synthetase. **Oxaliplatin** binds to the guanine and cytosine moieties of DNA, leading to cross-linking of DNA	Neuropathy, hypersensitivity, and pancytopenia (lower than normal number of red and white blood cells and platelets in the blood)	Yes
Irinotecan: patients with uridine diphosphate glucuronosyltransferase 1A1 (UGT1A) polymorphism	Irinotecan	Inhibition of DNA synthesis by blocking topoisomerase I-DNA complex	Severe diarrhea and neutropenia	No
FOLFIRI: Not used in adjuvant regimens.	fluorouracil, leucovorin, and irinotecan	Inhibition of DNA synthesis	Pancytopenia, diarrhea	Yes
Fluorouracil and leucovorin: for frail patients with major co-morbidities	fluorouracil and leucovorin	Inhibition of DNA synthesis	Pancytopenia, mucositis	Yes
Capecitabine: preferred if no plans to	capecitabine	A prodrug that is selectively	Pancytopenia, hand/foot	Yes

		tumour-activated to its cytotoxic – fluorouracil	syndrome	
CAPOX / XELOX: most common; substitutes oral capecitabine for intravenous fluorouracil.	capecitabine and oxaliplatin	Inhibition of DNA synthesis	Pancytopenia, diarrhea, hand/foot-syndrome, neuropathy, and hypersensitivity	Yes
CAPIRI /XELIRI: Substitutes oral capecitabine for intravenous fluorouracil	capecitabine and irinotecan	Inhibition of DNA synthesis	Pancytopenia, diarrhea, hand/foot-syndrome	Yes
FOLFOXIRI: Intensive regimen used for patients who are fit for surgical resection.	fluorouracil, leucovorin, oxaliplatin, and irinotecan	Inhibition of DNA synthesis	Pancytopenia, diarrhea, neuropathy, and hypersensitivity	Yes
FOLFOX + BRAF/KRAS inhibitors: For tumors that are KRAS/NRAS/BRAF wild type; ineffective for tumors with sequence variations in these genes.	fluorouracil, leucovorin, oxaliplatin and cetuximab or panitumumab	**Cetuximab** and **panitumumab** binds to EGFR and competitively inhibits the binding of epidermal growth factor resulting in inhibition of cell growth, induction of apoptosis.	Pancytopenia, diarrhea, hand/foot-syndrome, hypomagnesemia, hypersensitivity reactions, neuropathy, skin toxicity	Yes
IROX: Nonstandard regimen for patient's intolerant of fluorouracil due to severe dihydropyrimidine dehydrogenase deficiency or coronary vasospasm	irinotecan and oxaliplatin	Inhibition of DNA synthesis	Diarrhea and neuropathy	Rare

1.3.2 Targeted therapy

As the name suggests, targeted therapy targets the CRC specific genes, proteins, or the tissue environment contributing to the tumour growth. Targeted therapies include monoclonal antibodies (mAb) and small molecules with a molecular weight of <900 Dalton, which can directly inhibit cell proliferation, differentiation, or migration. The targeted inhibitors are also included in some chemotherapy regimens (Table 1.3). For CRC, the following are the targeted therapy options:

EGFR inhibitors such as Cetuximab (Garrett and Eng, 2011; Price et al., 2014) and Panitumumab (Price et al., 2014) are used in the treatment of mCRC. These inhibitors bind to EGFR and competitively inhibit the binding of epidermal growth factor resulting in inhibition of cell growth, induction of apoptosis. EGFR mAb in combination with chemotherapy has shown improvement in progression-free survival in CRC patients (Chan et al., 2017).

Anti-angiogenesis therapy inhibits tumour angiogenesis and prevents the tumour growth. Preclinical and clinical evidence suggest the benefit of anti-angiogenic agents such as bevacizumab (Avastin - Genentech, Inc., 2021; Cao et al., 2019), aflibercept (Sun and Patel, 2013) and regorafenib (Papadimitriou and Papadimitriou, 2021) in the treatment of mCRC.

Combined targeted therapies: In ~8 % of CRC patients, a specific mutation called BRAF V600E was detected (AACR Project GENIE: Powering Precision Medicine through an International Consortium, 2017). The negative charge of glutamic acid (E) due to the V600E mutation initiates phosphorylation of nearby serine and threonine residues and therefore functions to activate BRAF. This loss of inhibition of BRAF leads to an increase of basal oncogenic activity (Liang, Khorana and Kalady, 2015). Patients with BRAF V600E mutation markers are treated with BRAF inhibitors such as encorafenib and cetuximab (Liang, Khorana and Kalady, 2015).

Tumour-agnostic therapy: This approach targets specific genomic alterations or molecular features irrespective of tumour sites. In mCRC patients, neurotrophic tyrosine receptor kinase (NTRK) gene fusion is another biomarker where surgery is not possible due to extensive spread of cancer and other treatments have been reported to fail (Cocco, Scaltriti and Drilon, 2018). In this genetic alteration, two genes of the NTRK family fuse together and produce an altered tropomyosin receptor kinase (TRK) protein, which leads to uncontrolled cell growth (Lange and Lo, 2018). Larotrectinib (Burki, 2018) and entrectinib (Drilon et al., 2017) are the TRK inhibitors used for mCRC treatment where NTRK alterations are present.

1.3.3 Immunotherapy

Immunotherapy in CRC treatments explores the concept of T-cell mediated immunosuppression. Tumour cells can upregulate immune checkpoint molecules – programmed death (PD-1), programmed death-ligand 1 (PD-L1), and CD152, that exhaust T cells and inhibit apoptosis of malignant cells. PD-1 binds to PD-L1 to prevent binding of T cells and escape cellular apoptosis. Inhibitors of either PD-1 (Pembrolizumab, Nivolumab, Cemplimab) or PD-L1 (Atezolizumab, Avelumab, Durvalumab) (Tumeh et al., 2014; Golshani and Zhang, 2020) are used for mCRC treatment.

The improved aggressive surgical resections and development of novel chemotherapy regimens has increased the survival rate in cancer patients but with less focus on the quality of life of cancer patients and survivors (Buiting and Olthuis, 2020). Various surveys and statistical studies have reported CRC patients and survivors, especially elderly (aged > 80), consider quality of life as important as survival both during the treatment and post-treatment (Yucel, 2015; McCombie, Frampton and Frizelle, 2021). As the number of toxicities and adverse effects of the present mCRC and CRC chemotherapy regimens (enlisted in Table 1.3) are very high, the quality of life of patients is reduced to a great extent. About one-third of the CRC patients have been reported to terminate their chemotherapy regimens prematurely (known as early discontinuation - EDChemo) due to strong adverse effects (Boakye et al., 2021).

Over the last two decades, with despite ~70 new cancer therapies approved (from 2002 to 2014) for various cancer types, the overall cancer survival rate has barely improved - by ~2 months, (Fojo, Mailankody and Lo, 2014) compared to the cost of drug development and treatment. More recently, anti-cancer research has shifted focus from cytotoxic drugs to target-based drugs to reduce the toxicities and adverse effects. The known drug-targets are being repurposed for anticancer therapies against chemoresistance, tumour angiogenesis, etc. (Spugnini and Fais, 2019; Huang, Zhao, Liu and Liu, 2018; Houthuijzen et al., 2017; Wu et al., 2013). The combination of these target-based drugs as adjuvants and/or neo-adjuvants with lower-dose chemotherapy regimens (Spugnini and Fais, 2019; Boakye et al., 2021), which help reducing the adverse treatment effects and improve patient compliance to the regimen. One of the analytical studies (Munker et al., 2018) has inferred that dosage reduction with longer chemotherapy regimen in mCRC therapy reduces the treatment adverse effects without compromising the survival rate. With extensive data linking G protein coupled receptors (GPCRs) and CRC, GPCRs have emerged as attractive targets to develop

novel strategies targeting tumour progression and metastasis in CRC (Kumari, Reabroi and North, 2021; Usman et al., 2020; Liu et al., 2017).

1.4 G protein coupled receptors – novel drug targets for CRC therapy.

New drug discovery is impeded by the limited number of novel and validated disease targets (Chandra, 2011; Gilbert, 2013). Target-based drug discovery focuses on specific known protein targets and can explore the possible importance of functionally lesser understood proteins that may be involved in disease mechanism. Hence, the approach provides an effective means of moving beyond well understood targets to discover novel drug targets (Kue et al., 2016).

The Molecular Libraries Program (MLP) by the National Institute of Health (NIH) performed large-scale screenings to identify new chemical entities to explore novel drug targets (Austin, 2004). These screens identified G protein coupled receptors (GPCRs) as one of the most targeted protein family involved in high-throughput screening (HTS) in the MLP (Przydzial et al., 2013; Insel et al., 2019). From a 2011-2019 sales report, drugs targeting GPCRs were estimated to hold ~ 30-34 % of marketed therapeutics sales (Hauser et al., 2017; Hauser et al., 2018; Zhou and Wild, 2019; Kooistra et al., 2020).

GPCRs are found only in eukaryotes - including animals, plants, fungi, and protozoa (Salon, Lodowski and Palczewski, 2011). In humans, GPCRs have about 800 members known according to phylogenetic analysis studies (Kooistra et al., 2020)**.** However, a-limited number of known GPCRs (~110) out of the ~800 human GPCRs were considered as drug targets (Salon, Lodowski and Palczewski, 2011; Qu, Wang and Wu, 2020) while the remaining were regarded as orphan receptors as neither their endogenous ligand nor their physical function has been elucidated. The identification and quantification of previously unrecognised GPCRs further increased the number of viable GPCR drug targets to 134 (Sriram and Insel, 2018; Insel et al., 2019). Thus, GPCRs represent a rich source of drug targets (34% of marketed drugs), comprising the largest (4% of human genome) and most diverse group of membrane receptors in the human genome (Salon, Lodowski and Palczewski, 2011; Kooistra et al., 2020).

The signalling pathways of GPCRs have been reported to play crucial roles in blood vessel development, such as lysophosphatidic acid/LPA4, sphingosine 1-phosphate/Edg-1, and lysophospholipid / GPR4 (Liu et al., 2000; Yang et al., 2007; Parker, Parker, Sah and Sallee,

2005). Crosstalk between GPCR downstream signalling and VEGFR2 pathways has been reported but not profoundly understood during development of angiogenesis. It has been shown that $G\alpha_q/G\alpha_{11}$ and $G\alpha_{13}$ proteins regulated VEGFR2 tyrosine phosphorylation through interaction with VEGFR2 (Zeng et al., 2003). The disruption of the gene for the heteromeric G protein subunit $G\alpha_{13}$ disable endothelial cells from developing into an organised vascular system. Thrombin protein (binds to protease activated receptors) has also been shown to directly stimulate VEGFR1 and VEGFR2 expression (Richard, Vouret-Craviari and Pouysségur, 2001), suggesting that the transcription of VEGF/VEGFR2 and tumour angiogenesis could be regulated by GPCRs.

Mutations and alterations in some GPCRs have been linked to tumorigenicity where tumour cells overexpress and aberrantly activate GPCRs by releasing an excess of GPCR agonists such as bioactive lipids, peptides, chemokines, hormones, etc (Liang et al., 2020; Usman et al., 2020). A number of GPCR targeted drugs against various cancers have been approved by the Food and Drug Administration (FDA) but none for CRC yet (enlisted in Table 1.4).

Table 1.4: FDA approved GPCR target-based drugs for anticancer therapy since 1996 (Usman et al., 2020).

Drugs	GPCR Receptor	Cancer	Approval year
Cabergoline	Dopamine receptor D1	Neuroendocrine tumours, pituitary tumours	1996 (Lin, Zhang, Zhang and Wu, 2019)
Lanreotide	Somatostatin receptor	Pancreatic cancer	2007 (Godara, Siddiqui, Byrne and Saif, 2018)
Degarelix	Gonadotropin releasing factor hormone receptor	Prostate cancer	2008 (Olsson et al., 2017)
Plerixafor	C-X-C chemokine receptor 4	Multiple myeloma	2008 (Wang, Tannous, Poznansky and Chen, 2020)
Vismodegib	Smoothened receptor	Metastatic basal cell carcinoma	2012 (Bánvölgyi et al., 2019)
Raloxifene	Estrogen receptor	Breast cancer	2014 (Ağardan et al., 2020; Mirzapur, Khazaei, Moradi and Khazaei, 2018)
Sonidegib	Smoothened receptor	Metastatic basal cell carcinoma	2015 (Chen et al., 2018)
Mogamulizumab	C–C Chemokine receptor 4	T cell lymphoma	2018 (Moore et al., 2019)

Overexpression of a variety of GPCRs is associated with CRC development. Overexpression of Formylpeptide receptor 2 (GPCR) is linked to abnormal migration and proliferation in mammalian colon cells (Chen et al., 2013; Liang et al., 2020; Tian et al., 2020). Similarly, for phospholipid GPCRs – activation of Lysophosphatidic acid receptor 1 and 2 has been reported to stimulate cell proliferation and cell invasion activity in CRC cells (Yun, 2019; Kitamura et al., 2019; Ishimoto et al., 2020). Another GPCR - GPR120 also called free fatty acid receptor 4 (FFAR4), has been identified as a promotor of tumour angiogenesis in human colorectal carcinoma (Wu et al., 2013).

1.4.1 Introduction to GPCRs

GPCRs, also known as seven transmembrane (TM) domain receptors, hepta-helical receptors, serpentine receptors, and G protein-linked receptors (GPLR), sense external stimuli molecules outside the cell and transmit cellular signals to the cytoplasm and eventually to the nucleus (Salon, Lodowski and Palczewski, 2011).

GPCRs have a common structural feature of seven helical domain which transverse the plasma membrane forming the serpentine structure as shown in Figure 1.6. The GPCR structure can be split into three segments: the seven TM helical bundle (which forms the core of the GPCR) and often with a perpendicular helix (H8) in the intracellular domain; three extracellular loops (ECL), which connect the helices at the extracellular end and modulate ligand access; and three intracellular loops (ICL), which connect the helices at the intracellular end and interacts with effector signalling proteins. The bundle-like core conformation of the seven TM domains is stabilized by an essential disulfide linkage formed by two cysteine residues, one on the TM3/ECL1 interface and one in ECL2 (Figure 1.6) in all the GPCRs. The N-terminus of a GPCR is at the extracellular side, and the C-terminus is located intracellularly. The various domains of GPCRs vary in length in different types of GPCRs providing specific structural and functional properties (Grisshammer, 2017; Insel et al., 2019).
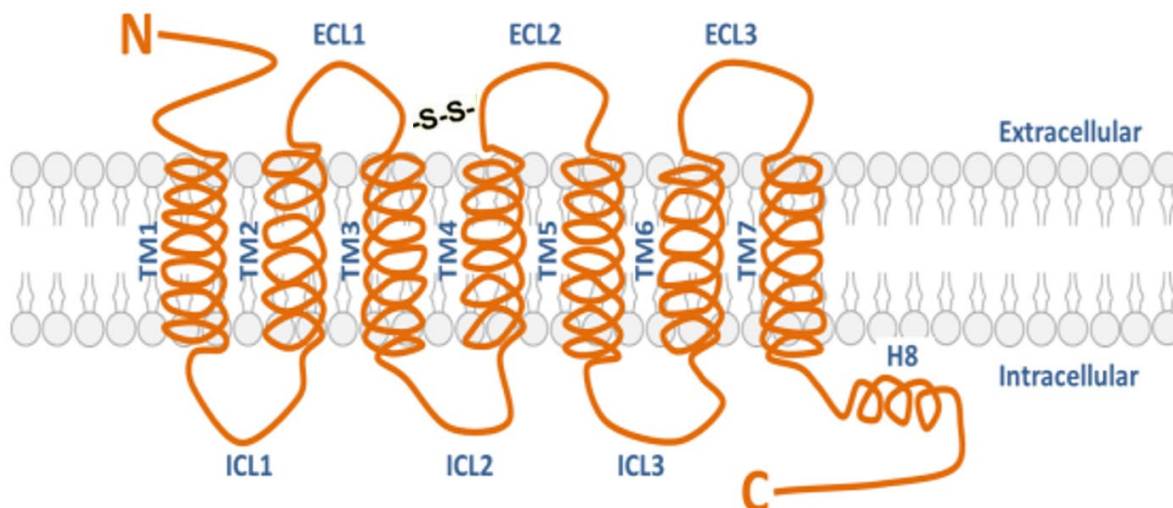
Figure 1.6: Schematic view of the general structure of a G protein-coupled receptor (GPCR) embedded in the cell membrane illustrating the conserved S-S (disulfide linkage) essential for structural stability. Figure adapted from "Virtual Background – Cell Membrane", by BioRender.com (2021).

1.4.2 Classification of GPCRs

Based on sequence homology, the Nomenclature Committee of the International Union of Basic and Clinical Pharmacology (NC-IUPHAR) classified six GPCR classes (A to F) (Kolakowski, 1994; Attwood and Findlay, 1994), which includes GPCRs from all species even classes that do not appear in humans (such as Class D and Class E).

 Class A-F classification system is represented as:

    Class A: Rhodopsin-like GPCR

    Class B: Secretin-like GPCR

    Class C: Metabotropic glutamate-like GPCR

    Class D: Fungal mating pheromone GPCR

    Class E: Cyclic AMP GPCR

    Class F: Frizzled/smoothened GPCR

More recently another popular system was presented by Fredriksson known as the GRAFS classification system (Fredriksson, Lagerström, Lundin and Schiöth, 2003) based on phylogenetic sequence relations: Glutamate (Class C), Rhodopsin (Class A), Adhesion, Frizzled and Secretin (Class B) which is summarised in Table 1.5. Both the classification schemes have shortcomings and inconsistencies that cannot assign new GPCR sequences

from invertebrate species to any existing family or subfamily (Scholz, Langenhan and Schöneberg, 2019).

Table 1.5: GPCR Glutamate, Rhodopsin, Adhesion, Frizzled, Secretin (GRAFS) Classification with characteristics, biological roles, and representative examples for each group. (Fredriksson, Lagerström, Lundin and Schiöth, 2003; Kooistra et al., 2020; Byrne et al., 2021)

GRAFS / Number of known human Sequences	Characteristics	Biological roles	Representative examples
Glutamate (Class C) 22	Presence of 9 conserved cysteine residues known as cysteine rich domain (CRD) or NCD3G in the N-terminal region	Neurobiological roles, gustatory roles (sweet and sour tastes)	Metabotropic glutamate receptors, GABA receptors, taste-1 receptors
Rhodopsin (Class A) 719	Presence of D/ERY in TM3, NPxxY in TM7 and disulfide bond by cysteines (C) in ELC1 and ELC2	Olfactory and vision stimuli, neurotransmitter signalling, cardiovascular and immunological functions, etc.	Adrenergic receptors, opioid receptors, chemokine receptors
Adhesion (Class B2) 33	Long N-termini containing multiple functional domains and having numerous sites for glycosylation	Developmental biology of Central Nervous System, immunological functions, etc.	Latrophilin receptors, Brain-specific angiogenesis inhibitor proteins
Frizzled (Class F) 11	Characterized by presence of 10 conserved cysteine residues known as FZ_CRD domain or FZ domain between N-terminal and TM regions	Developmental biology, cancer development, perception of bitter taste, also known as receptors for Wnt proteins	Smoothened receptor, taste-2 receptors
Secretin (Class B1) 15	Evolved from Adhesion receptor family	Endocrine and metabolic disorders	Glucagon receptor, GLP-1 receptor, growth hormone releasing hormone receptor

1.4.3 Class A - Rhodopsin-like GPCRs signalling

The present research has focused on the Rhodopsin-like receptors, which play a crucial role in many physiological processes, for instance, the transmission of the light and odorant signal, the mediation of neurotransmission and hormonal action, cell growth and immune defence (Ballesteros, Shi and Javitch, 2001; Kinoshita and Okada, 2015). These GPCRs initiate downstream signalling networks by coupling with heterotrimeric guanine-nucleotide-binding regulatory proteins (G-proteins) and β–arrestin resulting in a broad range of physiological and pathological processes (Kinoshita and Okada, 2015).

GPCRs can possess some degree of constitutive activity (a basal level of activity in the absence of any agonist) (Salon, Lodowski and Palczewski, 2011; Kinoshita and Okada, 2015). The basal activity of a receptor can increase or decrease depending upon the type of ligand binding in an orthosteric cavity located in the TM-bundle core towards the extracellular end of the receptor. The binding ligands can be characterised according to their pharmacological effects on the target receptor (Figure 1.7) as:

1) Full agonist: induce the maximal level of activation possible.
2) Partial agonist: activate the receptor above basal levels but not maximally.
3) Neutral antagonist: bind to the receptor but maintains basal levels by neither stimulating nor inhibiting the receptor.
4) Inverse agonist: decrease the level of receptor activation below basal levels
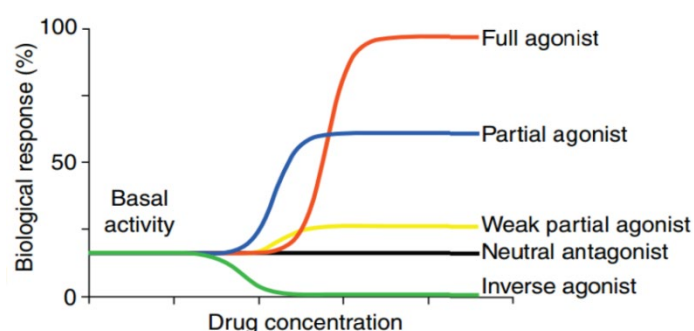
Figure 1.7: Pharmacological effect of agonists, partial agonists, neutral antagonists, and inverse agonists (Neubig, Spedding, Kenakin and Christopoulos, 2003)

The receptor activation typically starts with agonist binding and triggers a conformational re-arrangement of TM and loop regions, which allows the coupling of a G protein present in the cytosol leading to downstream signalling (Figure 1.8 a, b). During activation the GPCRs transforms from basal (ground) state to an active state mediated by small conformational

changes of the ligand binding (agonist) at the orthosteric site. These small structural changes are amplified into larger structural changes at the intracellular domain (known as the "active state" - Figure 1.8a) of the GPCR to facilitate G-protein coupling or arrestin interaction (Latorraca, Venkatakrishnan and Dror, 2016). Where the ligand binding at the orthosteric site is an antagonist or inverse agonist, the receptor retains the basal state or attains "inactive / closed" state (Figure 1.8a), which cannot facilitate the G-protein coupling or arrestin interaction at the intracellular domain (Latorraca, Venkatakrishnan and Dror, 2016; Nygaard et al., 2013).

The heterotrimeric G protein consists of one α-subunit (there are 16 encoding genes expressing 20 different Gα) bound to guanosine diphosphate (GDP); a β subunit (5 genes); and a γ subunit (12 genes) (Kolakowski, 1994; Nygaard et al., 2013). The G protein remains in an "inactive" state when its α subunit is bound to GDP and associated with its respective βγ subunit. When the agonist binds to the inactive GPCR, the α subunit of the G protein catalyses the GDP-GTP exchange and this promotes the dissociation of the βγ subunit (Figure 1.8). Subsequently, the free GTP-α (s, i, q, 12) and βγ subunits can transfer the signal to their intracellular effectors, such as enzymes and ions channels through second messenger molecules (such as cAMP, DAG and IP3) which lead to effector functions (Table 1.6) (Ballesteros, Shi and Javitch, 2001; Kinoshita and Okada, 2015; Nygaard et al., 2013).

As Figure 1.8b illustrates two of the signalling pathways – Gαs and Gαq. Activated Gαs stimulates membrane associated enzyme adenylyl cyclase (AC), which increases ATP to cyclic AMP (cAMP) conversion. cAMP acts as a second messenger to activate protein kinase A (PKA) which can phosphorylate multiple downstream targets; whereas the Gαi pathway inhibits AC. Activated Gαq stimulates the membrane-bound phospholipase C (PLC) to cleave phosphatidylinositol biphosphate (PIP2) into second messenger's inositol triphosphate (IP3) and diacylglycerol (DAG). IP3 increases intracellular calcium concentrations (Ca^{2+}) while membrane-bound DAG activates protein kinase C (PKC) by translocating it from cytosol to plasma membrane.
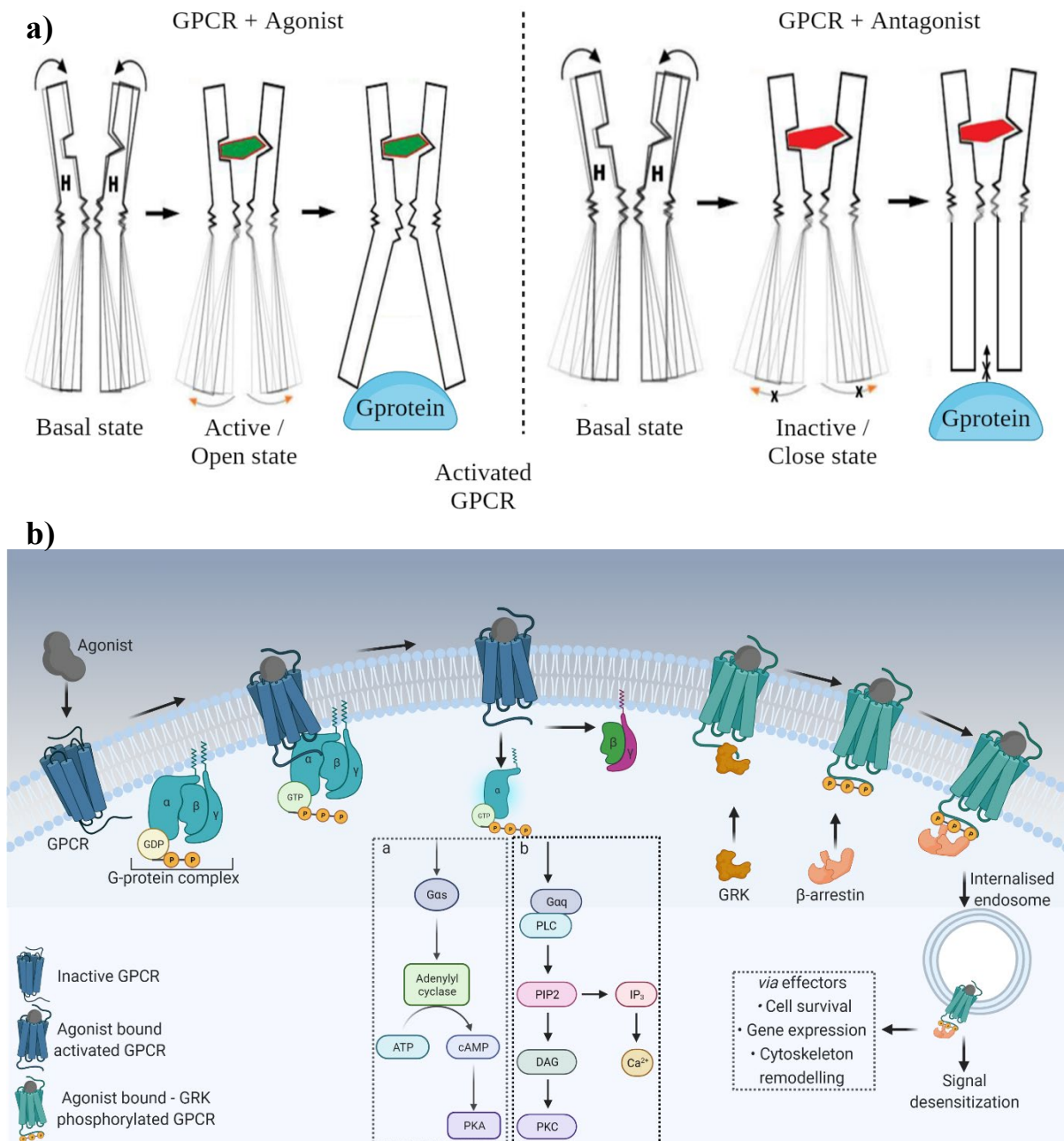
Figure 1.8: a) General flow of transformation of GPCR from basal state to active state when an agonist (green) binds or inactive state when an antagonist / inverse agonist (red) binds at the orthosteric binding site. H – helical transmembrane domains (Adapted from Nygaard et al., 2013); b) Schematic diagram of G-Protein Coupled receptors diverse signaling pathways upon activation of G-protein subunits [α, β and γ] or β-arrestin binding; Figure adapted from "Virtual Background – Cell Membrane", by BioRender.com (2021). Retrieved from https://app.biorender.com/biorender-templates. (Byrne et al., 2021)

Activated GPCRs also produce G protein independent biochemical responses through *β*-arrestin molecules (versatile adapter proteins) that form complexes with GPCRs following agonist binding and phosphorylation of receptors by GPCR kinase (GRK). GRK phosphorylates G-Protein independent ligand-bound GPCRs to initiate the recruitment of β-arrestin and blocks G-Protein coupling. GPCR-β–arrestin complex promotes endocytosis, trafficking ligand-GPCRs to sorting endosomes for either recycling to plasma membrane or signaling and regulation of various cellular processes (Figure 1.8b). They play a central role in the interrelated processes of homologous desensitization and GPCR internalization, which lead to the termination of G protein-GPCR signal transduction (Kinoshita and Okada, 2015; Nygaard et al., 2013). Similarly, several GPCRs have been shown to complex with small GTP-binding proteins such as Ras, Rab, and Rho, leading to activation of phospholipase D (Kinoshita and Okada, 2015).

Table 1.6: Classification of G proteins (Kolakowski, 1994)

Class	α subtype	Effector functions
G_s	α_s, α_{olf}	Adenyl cyclase (AC) stimulation, Ca^{2+} channel activation
G_i	α_o, α_{i1-3}, α_t, α_{gust}, α_z	AC inhibition, regulation of K^+ and Ca^{2+} channels, activates cyclic guanosine monophosphate (cGMP) phosphodiesterase
$G_{q/11}$	α_q, α_{11}, α_{14-16}	Phospholipase C-β activation
G_{12}	α_{12}, α_{13}	Na^+/ K^+ exchange; stimulates Bruton's tyrosine kinase and ras-GTPase activating protein.

1.5 GPR120 – discovery and therapeutics

GPR120 was originally described as an orphan GPCR in 2003 (Fredriksson, Lagerström, Lundin and Schiöth, 2003). An orphan receptor is a protein that has a similar structure to other identified receptors but are activated by binding to unknown signalling molecules. The orphan receptors are said to be deorphanized when their endogenous / signalling molecules are identified by various experimental assays describing their pharmacological activity (Yasi et al., 2019; Laschet, Dupuis and Hanson, 2018). GPR120 was later deorphanized and found to be a Rhodopsin-like GPCR based on the presence of canonical NPxxY and D/ERY (ERM in GPR120) conserved motifs indicative of rhodopsin-like GPCRs (Hirasawa et al., 2004). Subsequently, based on the nomenclature of the Free fatty acid receptor (FFAR) family, GPR120 was named systematically as the Free fatty acid receptor 4 (FFAR4) - as its endogenous ligands, poly-unsaturated long-chain free fatty acids (FFA), were known (Davenport et al., 2013; Fredriksson, Lagerström, Lundin and Schiöth, 2003; Hirasawa et al., 2008). However, the GPR120 receptor, unexpectedly, is only distantly related in terms of sequence identity and location to the other FFA-responsive GPCRs such as FFAR1-3 (Table 1.7). The distant relatedness of GPR120 to the other members of FFAR family has been reported in various other species as well (Mus musculus, Rattus norvegicus, Bos taurus, Gorilla gorilla, Rhesus monkey, etc.) at NCBI Gene database (https://www.ncbi.nlm.nih.gov/gene/).

Table 1.7: Summary of characteristic profiles and biological activity of the members of Free fatty acid receptor sub-family (Milligan, Shimpukade, Ulven and Hudson, 2016; Grundmann, Bender, Schamberger and Eitner, 2021; Carullo et al., 2021).

Nomenclature	GPR40/FFAR1	GPR43/FFAR2	GPR41/FFAR3	GPR120/FFAR4
Coupling protein	Gαq/11 Gαi/o Gαs Gα12/13 β-arrestin	Gαq/11 Gαi/o Gα12/13 β-arrestin	Gαi/o β-arrestin	Gαq/11 Gαi/o β-arrestin
Location (Homo sapiens)	19q13.12	19q13.12	19q13.12	10q23.33
Location (Mus musculus)	7; 7 B1	7; 7B1	7; 7B1	19; 19 C2
Location (Rhesus monkey)	19	19	19	9
GPR120 Sequence identity	20.66%	18.64%	23.43%	100%
Endogenous Agonist (FFA)	Medium-long chain, C6-C12	Short chain, C3~C4~C2	Short chain, C3>C4>>C2	Medium-long chain, C14-18
Expression sites	Pancreas (β-cells) Intestine (L, K, I cells) Bone Central nervous system Immune cells (Monocytes)	PMNs (Neutrophils, Eosinophils) Lymphocytes Monocytes Pancreas (β-cells) Intestine (L cells, IECs) White adipose tissue	Peripheral nervous system Pancreas (β-cell) Intestine (L, K cells) Immune tissue (DCs, thymus)	Adipose tissue Macrophages Lung Intestine (L, K, I cells) Bone
Major functions	Insulin secretion, Gut hormone secretion, Bone remodelling, Pain perception, Macrophage M2 differentiation	Immune cell activation, Cytokine secretion, Insulin release, Gut hormone secretion, Immune-modulatory, Reduction in lipolysis, lipid accumulation, and insulin resistance	Increase in heart rate, energy expenditure, reduction of gut motility, Inhibition of insulin secretion, Gut hormone release, Decrease Th2 response	Differentiation, browning, Anti-inflammatory, Epithelial repair, Gut hormone release, Bone formation

1.5.1 GPR120 expression and functions

During deorphanization, expression of GPR120 mRNA (human) in intestine, lung, spleen, adrenal glands, and thymus was found to be abundant (Hirasawa et al., 2004). Later it was found to be highly expressed in adipose tissues especially, during adipocyte differentiation (Gotoh et al., 2007; Wang, Xie, Zhang and Leung, 2019). The detailed expression profiling of GPR120 in intestine indicated that GPR120 is expressed in each of L, K, and I cells of intestine (Iwasaki et al., 2015). These cells of the intestine secrete gastrointestinal hormones or peptides which help digestion of ingested foods and absorption of released nutrients. The global gene expression analysis of GPR120 by various groups showed that GPR120 is also expressed in the taste buds and pancreatic islets (Matsumura et al., 2007; Wang, Xie, Zhang and Leung, 2019), with gene co-expression and protein-protein interactions associated with insulin secreting islets and haemoglobin glycosylation (Kaur et al., 2021). Many studies have now clearly demonstrated that GPR120 plays a significant role in pathophysiology of various metabolic disorders such as diabetes, obesity, and inflammatory disorders (Kaur et al., 2021).

The Human Protein Atlas (HPA) analysed RNA-seq data from 37 human tissues samples to estimate protein expression of GPR120 (Figure 1.9) (Uhlen et al., 2017) indicating possible expression of GPR120 on the somatic cells or possibly expressed by a subpopulation of cells in the corresponding tissue. The HPA RNA-seq analysis reported the highest TPM (transcripts per million) of GPR120 in rectum and colon tissues. The expansive expression pattern and diverse physiological roles of GPR120 (Table 1.5) has made GPR120 receptor a potential target of therapeutic interventions as well as lead to a major challenge of off-target effects.

The high expression of GPR120 mRNA in subcutaneous, epididymal and mesenteric adipose tissue in HFD-fed (high fat diet) mice suggests that GPR120 has important roles in adipocyte differentiation and maturation (Ichimura et al., 2012) as GPR120 deficient HFD-fed mice are reported to develop obesity, glucose intolerance and fatty liver with decreased adipocyte differentiation, lipogenesis, and enhanced hepatic lipogenesis (Ichimura et al., 2012). Adipocyte differentiation from bone marrow mesenchymal stem cells has also been suggested to be promoted by GPR120 (Gao et al., 2015).
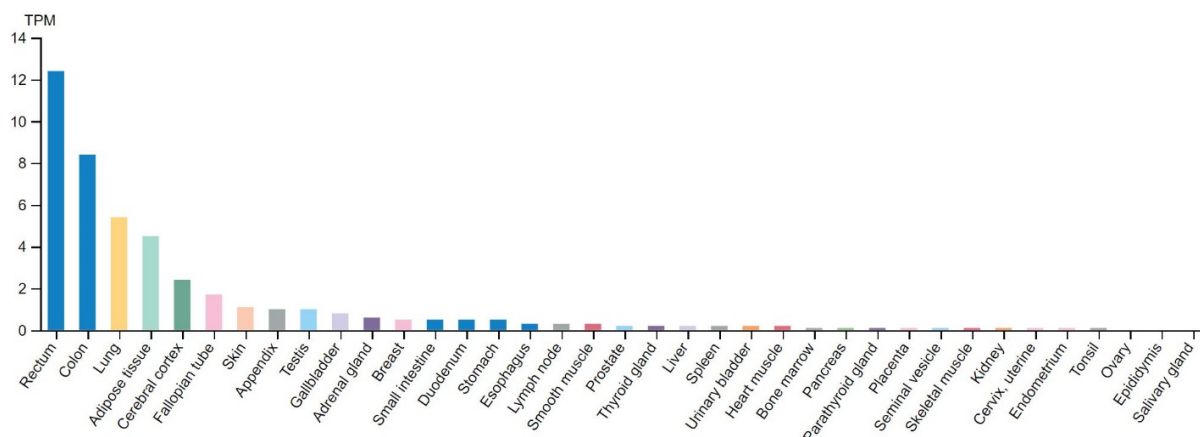
Figure 1.9: GPR120 mRNA expression levels across 37 tissue samples from Human Protein Atlas RNA-seq dataset. RNA-seq tissue data reported as mean TPM (transcripts per million), color-coding is based on tissue groups, each consisting of tissues with functional features in common (Uhlen et al., 2017). (Figure obtained from HPA - https://www.proteinatlas.org/)

GPR120 is responsive to long-chain fatty acids (LCFAs), and in particular, LCFAs stimulate secretion of the incretin GLP-1 from the STC-1 mouse enteroendocrine cell line (Hirasawa et al., 2004) and α-LA increases plasma GLP-1 levels in a rat model (Tanaka et al., 2008). The long-term administration of LCFAs can upregulate the proliferation of pancreatic β-cells. The resulting increased secretion and plasma levels of GLP-1 has been shown to enhance the glucose stimulated insulin secretion (Tanaka et al., 2008) playing a significant role in type II diabetes (Milligan, Shimpukade, Ulven and Hudson, 2016)

Eicosapentaenoic acid (EPA), an endogenous agonist of GPR120 (Figure 1.10), activated GPR120 induced autophagy in mesenchymal stem cells derived from murine bone marrow (Gao et al., 2015; Gao et al., 2016). GPR120 has also been reported to work as a dual acting factor in bone metabolism as it increases the osteoblastic bone formation as well as decreasing osteoclastic bone resorption (Ahn et al., 2016). Thus, GPR120 works as an important target for modulating bone production and degradation.

GPR120 expressed in macrophages has been shown to have an anti-inflammatory role. Docosahexaenoic acid (DHA), an endogenous agonist of GPR120 (Figure 1.10), activated GPR120 and significantly reduced the lipopolysaccharide (LPS) stimulated secretion of cytokines TNF-α (tumour necrosis factor) and IL-6 (interleukins) plasma inflammatory markers (Alvarez-Curto and Milligan, 2016).

α-linolenic acid (αLA)

Eicosapentaenoic acid (EPA)

Docosahexaenoic acid (DHA)
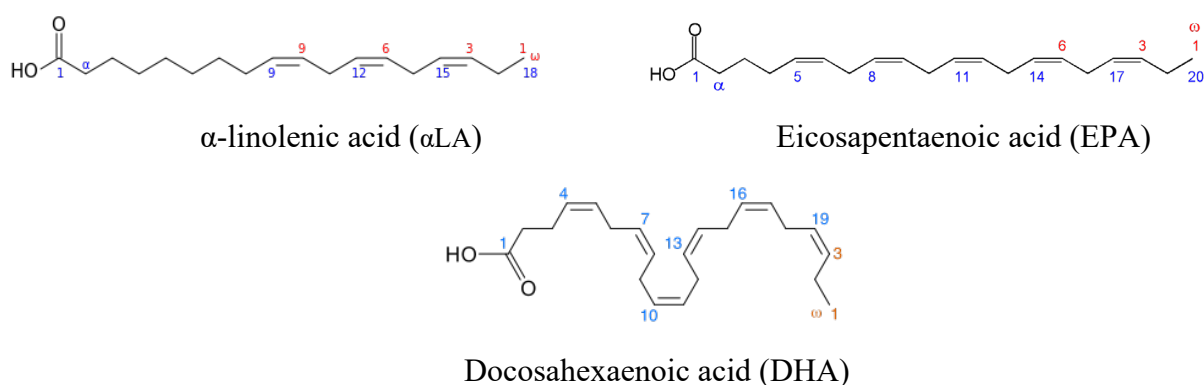
Figure 1.10: Endogenous long chain fatty acid activators of GPR120.

1.5.2 GPR120 – isoforms and signalling

In humans, GPR120 is present in two splice variants: the short isoform (361 amino acid long – GPR120S) and long isoform (377 amino acid long – GPR120L) (Figure 1.11) (Hirasawa et al., 2004; Watson, Brown and Holliday, 2012; Hudson, Shimpukade, Milligan and Ulven, 2014). Notably, due to alternative splicing GPR120L contains an additional 16 amino acid sequence within the third intracellular loop (ICL3) (Figure 1.11) (Watson, Brown and Holliday, 2012), a GPCR domain that is typically involved in protein interactions, downstream signalling, and desensitization (Davenport et al. 2013). In humans the ubiquitously expressed short isoform couples effectively to both Gαi/o, Gα $_{q/11}$ and β-arrestin2 pathways, while the long isoform is known to only bind β-arrestin and relays signalling to ERK1/2 (Senatorov et al., 2020; Carullo et al., 2021; Watterson et al., 2017). The short isoform is present in all parts of the human gastrointestinal tract while the long isoform has been detected only in human colon or colon epithelial cell lines being suggested as a distinct human gene (Senatorov et al., 2020). The GPR120 long isoform has not been detected in rodents, canines, bovines or nonhuman primates such as cynomolgus monkeys (Kim et al., 2015; Senatorov et al., 2020).

GPR120 has been reported to couple with both Gαq/11 and Gαi/o when activated by LCFAs in different tissues (Engelstoft et al., 2013; Stone et al., 2014). The Gαq/11 coupling induces mobilisation of intracellular calcium (Hirasawa et al., 2004; Hudson et al., 2013; Carullo et al., 2021). Gαq/11-coupled GPR120 has been reported to function in release of GLP-1 from enteroendocrine cells in the colon region (Tanaka et al., 2008). While Gαi/o coupled GPR120 inhibits adenylate cyclase activity which in turn inhibits ghrelin secretion from gastric mucosal cells (Engelstoft et al., 2013). Alvarez-Curto reported that Gαq/11 knockout / null cells could not produce mobilisation of intracellular calcium upon agonist exposure,

confirming Gαq/11 mediated signalling at GPR120 (Alvarez-Curto and Milligan, 2016). GPR120 activation by LCFAs and PUFAs also resulted in β-arrestin2 recruitment (Hudson et al., 2013). The anti-inflammatory actions of GPR120 are induced through β-arrestin2 / TAB1, independent of G protein - Gαq/11-coupled pathway (Engelstoft et al., 2013).
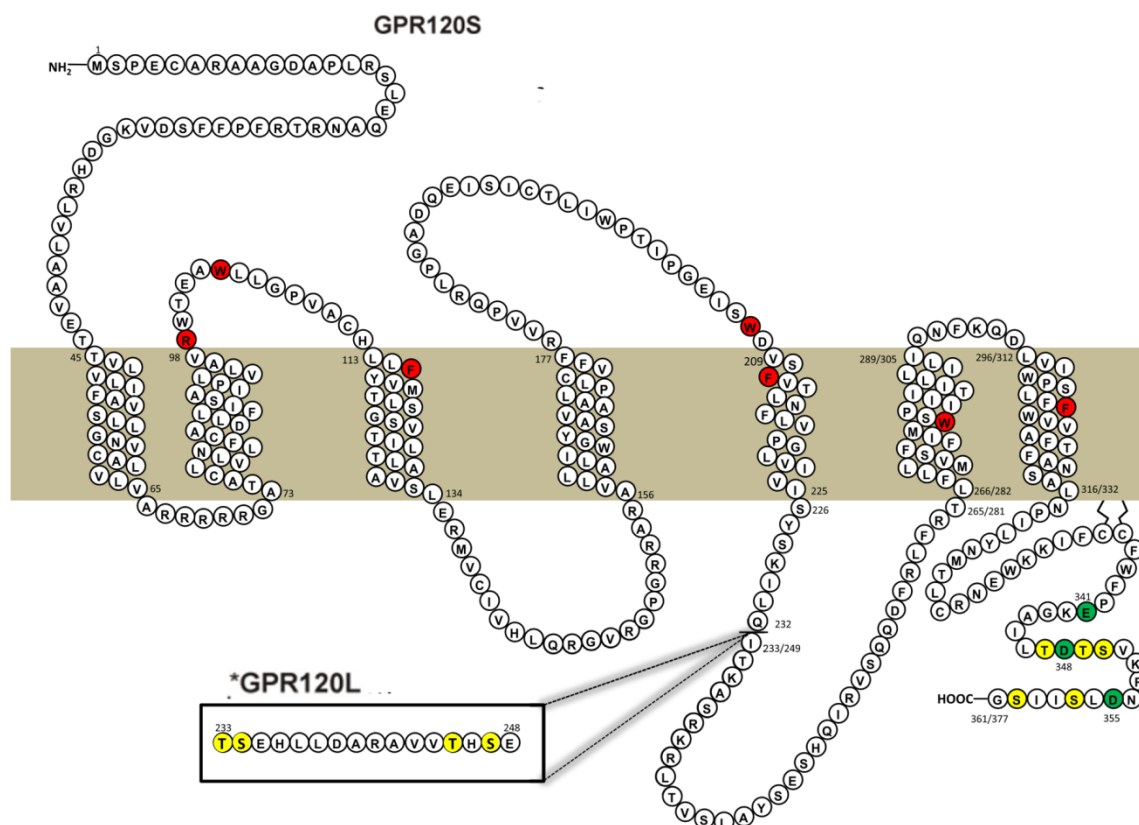
Figure 1.11: Human GPR120S structure: Snake plot. The insert shows the additional 16 amino acids gap introduced to GPR120L. The amino acid residues shaded in red are implicated as residues involved in the orthosteric binding pocket (Arg99(R), Trp104(W), Phe115(F), Trp207(W), Phe211(F), Trp277(W), Phe304(F)). The amino acids shaded in yellow are known phosphorylation sites (Thr347(T), Thr349(T), Ser350(S), Ser357(S), Ser361(S)) of GPR120S and GPR120L, which along with the noted acidic residues (Glu341(E), Asp348(D), Asp355(D)) shaded in green, create the β-arrestin phosphosensors (Hudson, Shimpukade, Milligan and Ulven, 2014). Figure was modified from Senatorov et al., 2020.

The physiological consequences of the two GPR120 receptor isoforms in humans is unknown although GPR120L contains four additional phosphor-labile residues in ICL3 (threonine and serine) compared to GPR120S (Kim et al., 2015). Agonism with FFAs lead to the phosphorylation of both isoforms at the same time with no significant difference but the basal

phosphorylation of GPR120S is two-fold higher than that of GPR120L in the absence of any FFA (Senatorov et al., 2020). The difference in the basal phosphorylation might be due to the blocking of phosphorylation sites by the 16 additional amino acids of GPR120L. Despite such studies, the functional and physiological importance of the long isoform remains uncertain (Watson, Brown and Holliday, 2012; Milligan, Shimpukade, Ulven and Hudson, 2016; Senatorov et al., 2020).

1.5.3 GPR120 ligands

Both the GPR120 isoforms are activated by specific long-chain fatty acids (LCFAs), both saturated and unsaturated, and by the poly unsaturated fatty acids (PUFAs) such as α-linolenic acid (αLA), EPA, DHA (shown in Figure 1.10) (Hirasawa et al., 2008) as the orthosteric binding residues are identical in both isoforms. A single arginine in TM2 (Arg99) has been reported to form a critical interaction between receptor and -COOH group (carboxylate) of its ligands (Hudson, Shimpukade, Milligan and Ulven, 2014). The site-specific mutation study identified six other specific residues essential for agonist binding and activation of GPR120 which are: Trp104 (ECL1), Phe115 (TM3), Trp207, Phe211 (TM5), Trp277 (TM6) and Phe304 (TM7) (Hudson, Shimpukade, Milligan and Ulven, 2014) (Figure 1.11, residues in red).

There is a marked overlap in activation of GPR120 and FFAR1 receptors by fatty acid receptor ligands even though they share less than 10% homology in amino acid sequence (Senatorov et al., 2020). The residues involved in recognition of endogenous ligands of the FFA1 receptor such as pair of Arg residues in TM5(Arg183) and TM7(Arg258) as well as Asp residue in TM6 (Asn244) are not conserved in GPR120 (Milligan et al., 2015). This elevates the need to develop receptor selective and specific ligands for GPR120 to achieve targeted effect of receptor activation / inhibition. GPR120 can be activated by both saturated and unsaturated medium to long chain fatty acids (Figure 1.10). Being a potential therapeutic target with diverse physiological functions, several labs (such as Graeme Milligan Lab, University of Glasgow; Trond Ulven-Lab, University of Southern Denmark) are still developing GPR120 ligands (Son, Kim and Im, 2021; Carullo et al., 2021).
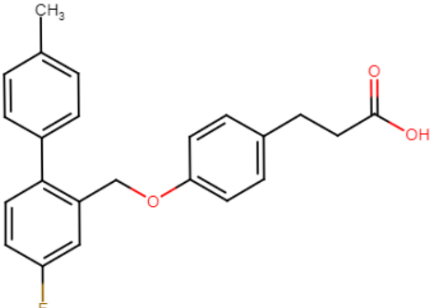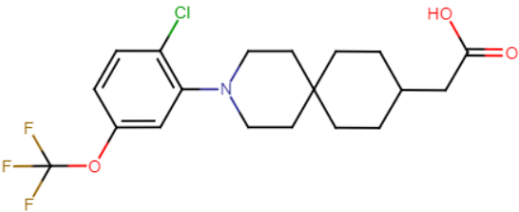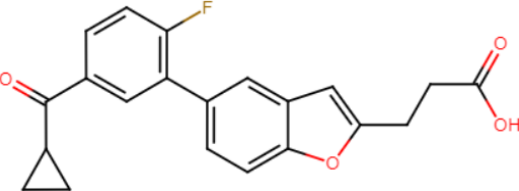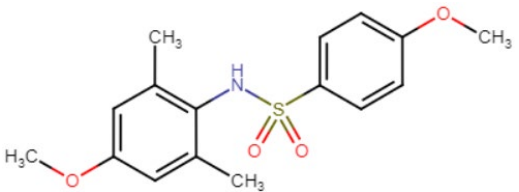
GW9508 (see Table 1.8) was the first synthetic ligand identified to bind to GPR120, which was originally developed as a FFAR1 agonist (Briscoe et al., 2006). Later, another non-selective ligand, NCG21, was developed as a GPR120 agonist using structure-activity relationship (SAR) analysis in combination with computer aided drug design (CADD) (Sun et
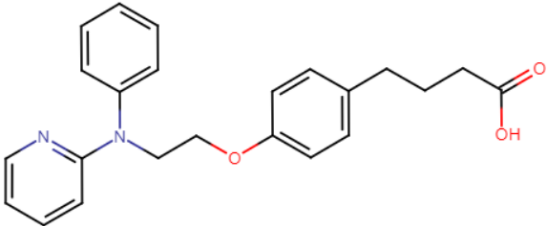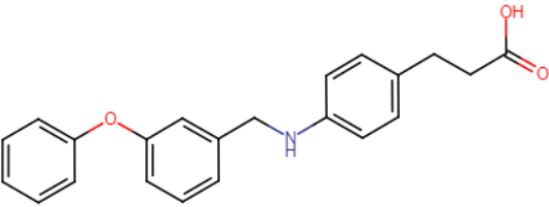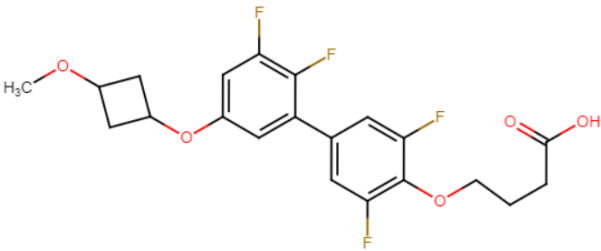
al., 2010). Finally, this drive led to the development of the first GPR120 specific ligand TUG891, which acts as a receptor agonist (Hudson et al., 2013). Over the years, a few GPR120 ligands have been developed (Table 1.8)– all of which act as receptor agonists and to date no antagonists have been developed for the GPR120 receptor. Lombardo and group discovered a series of benzofuran propanoic acid derivatives as potent and selective GPR120 agonists. Compound39 of this series demonstrated acute mechanism-based pharmacodynamic effects with high selectivity for GPR120 over FFAR1 in human and mouse models (Lombardo et al., 2016).
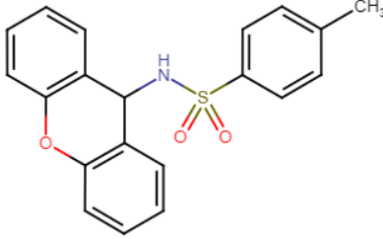
Compound AH7614 was developed by GlaxoSmith Kline (Sparks et al., 2014) and was postulated to be an orthosteric antagonist of GPR120 but later it was found to be a probe-dependent (screened with endogenous agonists) negative allosteric modulating drug (NAMD) of GPR120 (Watterson et al., 2017). As a NAMD, AH7614 does not bind to the orthosteric binding pocket of receptor, instead it binds to another pocket on the receptor, which results in conformational changes in the orthosteric binding pocket and reduces the binding affinity of other ligands (Watterson et al., 2017; Wild, Cunningham and Zhou, 2014).

Table 1.8: Human GPR120 ligands – representations of different chemical scaffolds are presented according to IC_{50} / EC_{50} values (Carullo et al., 2021; Son, Kim and Im, 2021).

Ligand	Ligand / Type	GPR120 mediated functions	Bioassays with drug potency data
 CHEMBL3952043	Agonist; Chromane propionic acid analogues	Improvement of glucose tolerance.	Agonist activity at human GPR120S in CHOK1 cells; β-arrestin binding assay => EC_{50} = 5 nM; IP1 accumulation – HTRF assay => EC_{50} = 14 nM; (Adams et al., 2016)
 CHEMBL3973101	Agonist; Aryloxybutanoic acid analogues	Improvement of glucose tolerance.	Agonist activity at human GPR120S in CHOK1 cells; IP1 accumulation – HTRF assay => EC_{50} = 21 nM; (Lombardo et al., 2016)
 CHEMBL3910333	Agonist; Non-acidic analogues	Improvement of glucose tolerance.	Agonist activity at human GPR120 receptor in Flp-In T-Rex293 cells; Calcium ions mobilization assay by Fura2-AM dye => EC_{50} = 38.9 nM (Azevedo et al., 2016)

TUG-891 / CHEMBL2058533	Agonist; Biphenyl analogues	Improve inflammation and insulin resistance; Increase of fat oxidation and reduction of fat mass; Increase of GLP-1 secretion; Decrease of circulating LDL; Repair of acute kidney injury; Inhibition of motility and phagocytosis in alveolar macrophages; Induction of adipogenic differentiation. (Murtaza et al., 2020)	Agonist activity at human GPR120 receptor in HEK293 cells; β-arrestin recruitment assay by BRET assay => EC_{50} = 43.65 nM. (Shimpukade et al., 2012)
CpdA / CHEMBL3919973	Agonist; Spirocyclic	Anti-inflammation in macrophages; Improvement of glucose tolerance; Decrease of hyperinsulinemia; Increase of insulin sensitivity; Decrease of hepatic steatosis; Improve atopic dermatitis. (Oh et al., 2014, Son et al., 2020)	Agonist activity at human GPR120S in CHOK1 cells; β-arrestin recruitment assay by luminescence assay => EC_{50} = 66 nM. (Cox et al., 2016)
Compound39 / CHEMBL3927879	Agonist; Benzofuran propanoic acid analogues	Improvement of glucose tolerance.	Agonist activity at human GPR120 in CHOK1 cells; IP1 accumulation – HTRF assay => EC_{50} = 97 nM; (Lombardo et al., 2016)
GSK137647 / CHEMBL3311198	Agonist; Diarylsulfonamides	Increase in mineralization of differentiated osteoblasts; Suppression of adipogenic differentiation of mesenchymal stem cells; Protection of pancreatic β cell dysfunction; Inhibition of islet inflammation. (Hasan et al., 2017)	Agonist activity at human GPR120 in human bone osteosarcoma epithelial cells (U2OS); Calcium ions mobilization assay by FLIPR => EC_{50} = 398.11 nM. (Sparks et al., 2014)

NCG21 / CHEMBL463785	Agonist; Acidic analogues	Increase of GLP-1 secretion. (Sun et al., 2010)	Agonist activity at human GPR120 in HEK 293 cells; β-arrestin recruitment assay by BRET assay => EC_{50} = 2398.23 nM (Shimpukade et al., 2012)
GW9508 / CHEMBL207881	Partial Agonist; Acidic analogues	Increase glucose-stimulated insulin release; Inhibition of inflammatory responses; Decrease of ghrelin secretion; Anti-osteogenesis; Induction of InterLeukin-4 secretion from eosinophils; Promotion of angiogenesis and motility in colorectal carcinoma.	Agonist activity at human GPR120 in HEK 293 cells; β-arrestin recruitment assay by BRET assay => EC_{50} = 1412.54 nM. GPR120 non-selective; higher selectivity towards FFAR1 (Shimpukade et al., 2012)
SR13	Agonist; Biphenyl analogues	Improvement of glucose tolerance.	Agonist activity in Chinese hamster ovary (CHO) cells expressing human GPR120; Calcium influx activity assay => EC_{50} = 93 nM High selectivity over FFAR1 (Sheng et al., 2018; Zhang, Sun, Wen and Yuan, 2019)

AH7614 / CHEMBL3311302	Negative allosteric modulator (Watterson et al. 2017); Diarylsulfonamides	Inhibit endogenous signalling of GPR120; Reduce DHA induced inhibition of lipid accumulation. (Kang et al., 2018)	Block α-linolenic acid induced intracellular Ca^{2+} response in U2OS cells; IC_{50} = 79.43 nM. Recently characterised as a NAMD (Watterson et al. 2017; Sparks et al., 2014).

1.6 GPR120 signalling in CRC and angiogenesis.

In addition to prior research, colorectal gene expression profiling by various groups (Kheirelseid, Miller and Kerin, 2013) enlisted GPCR expressing genes that are consistently represented as anticancer drug targets (Table 1.4**;** Usman et al., 2020). Dysregulation of GPR120 (a rhodopsin-like GPCR) expression and signalling has been recognized as one of the markers in cancers, such as breast cancer, renal cancer, prostate cancer, hepatic cancer, pancreatic cancer, and CRC (Fukushima et al., 2015; Houthuijzen et al., 2017; Senatorov and Moniri, 2018). GPR120's role in tumorigenesis, migration and metastasis is becoming evident with increasing reports.

GPR120 has been reported to play a pro-oncogenic role in human colon cancer by enhancing angiogenesis and cell migration (Wu et al., 2013), implying GPR120 antagonists could have a therapeutic effect. Regarding signal transduction of GPR120, direct effects of GW9508 (a partial agonist - Table 1.8) were examined in CRC cell lines which demonstrated activation of PI3K/Akt and nuclear factor kB. Regarding the effects of GPR120 agonists on the proliferation and migration, the GPR120 agonist EPA was shown to activate ribosomal protein S6 kinase β-1 (p70S6K1) in the CRC Caco2 cell line (Hopkins, Liu and Meier, 2014). S6K1 is a serine/threonine kinase that acts downstream of PIP3 and phosphoinositide-dependent kinase-1 in the PI3 kinase pathway. The kinase activity of p70S6K1 leads to an increase in protein synthesis and cell proliferation.

The GPR120 partial agonist, GW9508 is reported to be mitogenic for CRC cells (Wu et al., 2013), but not for prostate cancer cells (Hopkins, Liu and Meier, 2014). GW9508 (at 10 μM dose) stimulates migration of colon cancer cells whereas TUG-891 (GPR120 agonist) inhibits migration of prostate cancer cells (Wu et al., 2013; Hopkins, Liu and Meier, 2014). TUG-891 and omega-3 fatty acids stimulating GPR120 were shown to inhibit proliferation of DU145 prostate cancer cells (Hopkins and Meier, 2015). However, the knockdown of GPR120 prevented TUG-891-induced inhibition of growth and migration in these cells, confirming the key role of GPR120 (Hopkins and Meier, 2015). The contrasting reports from such studies further confuse the role GPR120 in cell proliferation and migration requiring further study.

Omega-3 PUFAs (such as DHA and EPA, see Figure 1.10) bind to a variety of known GPCRs such as lysophospholipid receptors, prostanoid receptors, free fatty acid receptros (FFAR1, GPR120), etc., (Im, 2012; Hopkins, Liu and Meier, 2014; Gao et al., 2016) as well

as various nuclear receptors (peroxisome proliferator-activated receptor α, γ; retinoic acid receptor RXR- α, β, γ), ion channels, transcriptase, reductase, kinase receptors, matrix metalloproteinases, etc. DHA has been reported to inhibit VEGF-induced phosphorylation of cell migration (translation of cells from one location to another), ERK1/2 and endothelial nitric oxide synthase 3 (eNOS) (Chao et al., 2014). eNOS is primarily responsible for the generation of NO in the vascular endothelium which plays crucial roles in regulating vascular tone, cellular proliferation, leukocyte adhesion, and platelet aggregation (Fish and Marsden 2006). There is a crosstalk relationship between these pathways. DHA shows the inhibitory actions via binding to GPR120 and induction of protein phosphatase 2 (PP2A, a tumour suppressor) enzyme activity (Chao et al., 2014).

Whereas GPR120 activation by EPA is reported to upregulate VEGF-A by stimulating protein kinase C (PKC), which subsequently activates PI3K in HEK293 cells (Hasan et al., 2015). Silencing of the GPR120 gene diminished the EPA-induced upregulation of VEGF-A release (Hasan et al., 2015). In the downstream cascade, a specific peroxisome proliferator-activated receptor γ (PPARγ) antagonist GW9662 annulled EPA-induced (eicosapentaenoic acid) release of VEGF-A. The transfection experiments in HEK293 cells confirmed that GPR120 activation with EPA specifically enhanced the binding of peroxisome proliferator-activated receptor γ (PPARγ) to the PPAR-response element (PPRE) in the VEGF-A promoter region. Intracellular EPA can also directly activate PPARγ, which also enhances VEGF-A release from adipocytes (Hasan et al., 2015).

In summary, GPR120 receptor activation has been reported to promote angiogenesis in CRC cells by stimulating release of VEGF, IL-8 and COX-2-derived PGE2 (Wu et al., 2013). GPR120 has also been shown to regulate tumour growth and migration of various cancer types, including melanoma and prostate cancers (Hopkins and Meier, 2016; Houthuijzen, 2016). GPR120 activation results in a signalling cascade that ultimately induces chemoresistance suggesting that antagonism of GPR120 can limit the development of chemotherapy resistance (Houthuijzen et al., 2017; Wang et al., 2019). Taken together, these results support contrasting roles of DHA and EPA through activation of GPR120 in cell migration and VEGF induced angiogenesis, which are implicated in GPR120's anticancer therapeutic functions. Several conflicting studies showing pro and antitumour effects pointed out a contribution for GPR120 in CRC progression and prognosis. Being a novel target in

CRC therapy, exploration studies for GPR120 ligands may yield better therapies and help the next generation of CRC patients (Ungaro, D'Alessio and Danese, 2021).

1.7 GPCRs and drug development

As mentioned previously, GPCRs are amongst the most frequently investigated drug targets accounting for ~34% of the prescription drugs (Kooistra et al., 2020). However, successful drug development strategies are assisted by the target structure determination. Structure determination of GPCRs have had a high failure rate as it is challenging to crystalise a highly unstable protein structure. GPCRs are transmembrane signalling proteins with flexible, multi-conformational states as a functional requirement (O'Brien et al., 2020). Being membrane proteins, GPCRs have large hydrophobic surface with low polar surface areas which reduce the crystal formation (Broecker, Eger and Ernst, 2017). Furthermore, GPCRs become thermolabile and unstable outside of the cell membrane for which short-chain detergents or fused mAb are required to stabilise the GPCR proteins in crystallising solution (Zimmermann et al., 2018).

1.7.1 Protein structure prediction

Over the last two decades, with evolving crystallisation techniques there has been an exponential increase in GPCR 3D structure determination. Since the publication of the first GPCR crystal structure of bovine rhodopsin in 2000, a total of ~110 unique (~600 in total) GPCR structures has been solved by X-ray crystal or cryo-EM methods in different conformations – Inactive state: ~30; Active state: 64; Intermediate states: 16 (https://gpcrdb.org/structure/statistics; Qu, Wang and Wu, 2020). More recently, the availability of such an expansive range of GPCR structures has boosted GPCR targeted structure-based drug discovery (SBDD). However, comparing the total number of GPCRs determined structures (~110) against the total number of GPCRs known in human genome (~400 druggable GPCRs) (Qu, Wang and Wu, 2020), highlights the fact that solving GPCR structures is still a tedious, time consuming and expensive strategy. However, the knowledgebase built from the solved GPCR structures has been successfully employed to predict the 3D structures of homologous GPCR members (homology/comparative modelling) and study protein dynamics using *in silico* / computational methodologies (Gusach et al., 2020) and possibly lead to CADD.

The principle behind protein structure prediction is that protein structures are more conserved than protein sequences during evolution in response to mutations (Illergård, Ardell and

Elofsson, 2009; Pearce and Zhang, 2021; Tiss et al., 2021). If the two proteins are evolutionarily related and share similarity in the primary sequence, the term "homology modelling" is used or synonymously "comparative modelling" – where a 3D structure of a protein can be predicted computationally if its sequence is comparable to an experimentally determined structure of a protein. The experimentally determined protein is used as a template and the sequence of the template is aligned to the sequence of the query protein with unknown structure. Using a computational algorithm, such as MODELLER (Webb and Sali, 2016), the 3D coordinates of the query structure are optimised according to the template structure and sequence alignment (Sailapathi et al., 2021).

Recently, a neural network-based 3D structure predictor of proteins – AlphaFold, demonstrated prediction accuracy comparable to experimental structures at the Critical Assessment of protein Structure Prediction – CASP14 challenge (Jumper et al., 2021). The latest application of artificial intelligence (such as AlphaFold2 from DeepMind) for *in silico* protein structure predictions has reported accuracy of free-modelling predictions improved to ~ 90% (Jumper et al., 2021). With the advancement of the computational 3D structure predictors, the bottlenecked structural elucidation of hard to crystallise proteins like GPCRs is becoming more precise and accurate (Pearce and Zhang, 2021).

1.7.2 Molecular docking

The experimentally determined GPCR structure or a homology model can be used to analyse GPCR-ligand binding patterns by molecular docking. Molecular docking predicts the binding pose of a small molecule or peptide in the binding pocket of the target protein. The docking algorithm predicts the binding pose by exploring the conformational landscape of the small molecule defined by the degree of freedom (number of rotatable bonds) and then ranks the candidate poses by summing the electrostatic and van der Waal forces (Pagadala, Syed and Tuszynski, 2017). The ranking of the candidate poses is performed by scoring functions which approximate the summed-up forces for greater computational efficiency. Large databases or computational chemical libraries can be screened by molecular docking – termed as virtual screening (VS). The molecular docking and scoring algorithms can be streamlined using the unique interaction fingerprints determined by GPCR-ligand co-crystal structures or site-specific mutagenic studies (Mitro et al., 2012; Hudson, Shimpukade, Milligan and Ulven, 2014).

1.7.3 Molecular dynamics simulations

Crystal structures can reveal important atomic level details of GPCR ligand binding and the activation mechanism but as a single snapshot of the protein-ligand complex. Molecular dynamics (MD) simulations employ Newtonian mechanics to study the interactions in the protein-ligand complex over a time period. Using molecular mechanics, the acting forces are divided into bonded (connected atoms) and non-bonded (disconnected atoms, separated by more than 3 bonds) terms. The bonded term can be separated into 3 components: bonds, angles, and dihedrals, while the non-bonded terms can be divided into Coulomb's law for electrostatic forces and Lennard-Jones potential for van der Waals forces (Swegat, Schlitter, Krüger and Wollmer, 2003).

While performing the MD simulation on the protein-ligand system, a random initial velocity is provided to the initial 3D protein-ligand structure. Forces acting on every particle is then calculated as a function of time (Swegat, Schlitter, Krüger and Wollmer, 2003). At every set timestep, a 3D snapshot of the system is saved along with all the forces. At the end of the simulation, the time-averaged energies from an ensemble of all the snapshots, thermodynamic and kinetic properties of the protein-ligand can be measured.

Employing SBDD in combination with homology modelling and MD simulations have previously resulted in several successful drug discovery cases such as Dorzolamide – anti-glaucoma agent; Imatinib – anticancer tyrosine kinase inhibitor; and Vemurafenib – anticancer BRAF inhibitor (Sancineto, Massari, Iraci and Tabarrini, 2013; Jacquemard and Kellenberger, 2019; Wang et al., 2020; Bagchi, 2020).

1.7.4 Pharmacophore screening

A pharmacophore can be defined as a three-dimensional spatial arrangement of chemical features in a drug molecule required for biological activity. These chemical features include H-bond donors, H-bond acceptors, aromatic and lipophilic regions as well as positive and negative ionizable ions (Wermuth, Ganellin, Lindberg and Mitscher, 1998; Voet et al., 2014, Voet et al., 2014a). Structure-based pharmacophore models (SBPMs) are generated by probing the known or predicted protein-ligand interactions. The co-crystalised protein-ligand structures solved by X-ray or cryo-electron microscopy assist the development of verified SBPMs along with knowledge obtained from biological assays such as site-specific mutation or alanine scanning studies (Dhasmana et al., 2019). The predictive SBPMs are ones where the essential protein-ligand interactions are not known and predictive computational tools

such as molecular docking and MD simulations of protein-ligand complexes are performed to predict the essential protein-ligand interactions (Voet et al., 2014). SBPMs overcome the limitations of ligand-based pharmacophore and molecular docking as they are based on the protein-ligand interaction residues and spatial conformation of the target active site (Joseph-McCarthy et al., 2003; Steindl, Schuster, Laggner and Langer, 2006; Voet et al., 2014a). Several studies employing SBPMs have identified various chemical scaffolds proposed as potential leads for development of novel anticancer agents (Crisan, Avram and Pacureanu, 2017; (khtar, Jabeen, Jalal and Antilla, 2018).

1.7.5 Pharmacokinetics compliance

Pharmacokinetics represents the study of the dynamic movements of drug molecules during their passage through the patient's body and encompassing the kinetics of ADME-tox. Estimation of absorption, distribution, metabolism, excretion, and toxicity (ADME-tox) is of vital importance during early drug discovery as it helps reducing the pharmacokinetics-related failures in the clinical phases (Hay et al., 2014). When a drug is absorbed orally it gets distributed via the portal circulation to liver where hepatic metabolism occurs. Typically, the drug metabolism pathway consists of oxidation followed by conjugation of oxidised moiety with polar molecules like cysteine, methionine, glucose, glutathione, or glucose. Both oxidation and conjugation phases are catalysed by various enzymes. The important drug properties such as metabolic stability, drug-drug interactions and drug toxicity are determined during drug metabolism (Bhhatarai et al., 2019). Drug toxicity remains the most significant and unpredictable property to date as it is most difficult property to predict or screen. Drug toxicity could involve multiple host factors and could be organ specific or species specific which are difficult to model adequately at early-stage drug discovery experiments. Drug toxicity is the leading cause of restricted use or withdrawal of drug from market even after lead candidates were reported to be safe in preclinical or early clinical trials (Bhhatarai et al., 2019; Tetko et al., 2006; Sohel et al., 2022).

Pharmacokinetics parameters are used as filters in virtual screenings to evaluate lead-like or drug-like molecules. Some of the filters being used can be listed as (Daina, Michielin and Zoete, 2017):

- Lipinski's filter includes molecular weight $\leq$ 500, logP (lipophilicity) $\leq$ 4.15, hydrogen bond acceptors $\leq$ 10, and hydrogen bond donors $\leq$ 5.

- Ghose's filter includes $160 \leq$ molecular weight ≤ 480, $-0.4 \leq$ logP ≤ 5.6, $40 \leq$ the molar refractivity ≤ 130, and $20 \leq$ number of atoms ≤ 70.

- Veber's filter includes the number of rotatable bonds ≤ 10 and the total polar surface area (TPSA) ≤ 140.

- Egan's filter includes logP ≤ 5.88 and TPSA ≤ 131.

- Muegge's filter includes $200 \leq$ molecular weight ≤ 600, $-2 \leq$ logP ≤ 5, TPSA ≤ 150, the number of rings ≤ 7, the number of carbons > 4, the number of heteroatoms > 1, the number of rotatable bonds ≤ 15, the hydrogen bond acceptors ≤ 10, and the hydrogen bond donors ≤ 5.

1.8 Cytotoxicity screening of small molecules

For decades, monolayer cell cultures – growing cells on flat surfaces of laboratory vessels, have been used extensively as the cellular *in vitro* cancer model for early-stage drug discovery and development (Stock et al., 2016). Cell-based *in vitro* cancer models for drug discovery provide a commercially viable and robust methodology to screen potentials lead compounds for their cytotoxicity, anti-proliferative and anti-migratory properties (Voskoglou-Nomikos, Pater and Seymour, 2003; Stock et al., 2016). For drug discovery screening, a cell line with higher expression of target receptor and lower expression of receptors sharing same ligands of the target receptor are recommended (Stock et al., 2016).

SW480 – human CRC cell line, was selected for *in vitro* screening of compounds derived by SBDD targeting GPR120 receptor. SW480 are the epithelial cells of large intestine established from a primary adenocarcinoma of the colon of Dukes' type B CRC patient (ATCC-SW480, 2021). A gene expression study reported that GPR120 is overexpressed (3.17 folds) in SW480 cell lines as well as did not detect GPR40 (also known as FFAR1) expression in SW480 (Wu et al., 2013). GPR40 is another member of FFAR family which shares orthosteric ligands with GPR120 (Wu et al., 2013; Grundmann, Bender, Schamberger and Eitner, 2021). The absence of GPR40 expression in SW480 enabled GPR120 selective *in vitro* screenings and reducing the probability of false positives.

Fluorometric assays of cytotoxicity and cell viability are recommended over traditional dye exclusion and colorimetric assays for being more sensitive and non-toxic to cells under treatment (Lindhagen, Nygren and Larsson, 2008; Aslantürk, 2018). These assays can be performed with ease using a fluorometer, fluorescence microplate reader or flow cytometer. Alamar Blue (resazurin salt) assay is a reduction assay based on the conversion of the blue

non-fluorescent resazurin dye to the pink fluorescent resorufin by mitochondrial and other cellular enzymes (Aslantürk, 2018). It is a nontoxic and cell permeable indicator of cell viability. As the dye enters the viable cells, it is reduced to fluorescent compound and increases the overall fluorescence of the cell culture medium. CFDA-AM assay (5-carboxyfluorescein diacetate, acetoxymethyl ester) and protease viability marker assay (glycylphenylalanyl-aminofluorocoumarin - GF-AFC) are other fluorometric dye assays that are used for cytotoxicity quantification (Aslantürk, 2018).

For target identification of test compounds, the process of regulating the target gene expression *in vitro* and quantifying the drug response in regulated and unregulated environment can be useful. Small interfering RNA (siRNA) also known as silencing RNA provides a simple and inexpensive *in vitro* toolkit to confirm target specificity of a potential test compound (Neumeier and Meister, 2021). siRNA transfection modulates the mRNA and temporarily supresses the gene-product of interest. In mammalian cells, the siRNA duplexes can be introduced directly by transfection to generate transient cell lines with silenced gene expression. SW480 has been reported a suitable transfection host (ATCC-SW480, 2021).

1.9 Overview of thesis

This introduction has demonstrated that target-based drugs when used as adjuvants and/or neo-adjuvants in combination with lower-dose chemotherapy regimens (Spugnini and Fais, 2019; Boakye et al., 2021) might fulfil the need for CRC chemotherapy regimens that could reduce the adverse treatment effects. GPR120 has tremendous potential of being such a novel therapeutic target for inhibiting tumour progression and tumour angiogenesis especially in colorectal cancer due to high expression level of GPR120 in colorectal tissue (Figure 1.9). In addition, the development of GPR120-targeted anticancer drugs might help to improve CRC and mCRC prognosis with reduced adverse effects. Although the pharmacology of GPR120 has been well documented, the understanding of its signal transduction is still at an early stage and despite its therapeutic potential no drugs have yet been approved by the FDA for GPR120.

The present study aims to answer the research question 'Can a combination of *in silico* and *in vitro* approaches lead to identification of chemical scaffolds targeting GPR120, which can be optimised by structural activity relationship profiling to design potential anti-cancer lead compounds?'.

To address these questions following research objectives were identified:

1. A structural model of the human GPR120 receptor will be developed and validated to facilitate *in silico* ligand design.

2. Molecular docking of reference (known) ligands to validate the binding mode and stability of ligand binding and protein-ligand interactions using MD simulations.

3. *In silico* (docking and pharmacophore) and *in vitro* (cytotoxicity) screening for new small molecule ligands based on the better understanding of the ligand binding site from the developed 3D model of the GPR120 receptor.

4. Optimisation of the hit compound(s) through structure-activity relationship (SAR) analysis, and computational physicochemical profiling.

5. Advanced *in vitro* assays on optimised hits, e.g., wound healing and clonogenicity to further evaluate the anticancer and potential anti-angiogenic properties of the preferred compounds

6. siRNA-based functional target validation: Comparative profiling of hit compounds in siRNA-mediated GPR120 silenced CRC cell line (SW480) and wild-type CRC cell line (SW480) by i*n vitro* cell proliferation assays.

Chapter 2 details and describes the methods and protocols used for the computational and biological experimental work.

Chapter 3 first details the results of GPR120 homology model generation and validation of the 3D model of GPR120 (short isoform). The validated 3D model of GPR120 was used to determine the docking poses of reference ligands (TUG891 and Compound39) followed by preparation of GPR120-ligand complexes for MD simulations.

Chapter 4 details the analysis of MD simulations of TUG891 and Compound39 bound GPR120 systems and development of an Asn313 interaction hypothesis to design GPR120 antagonists. The potential hits derived from the Asn313 hypothesis were further analysed by GPR120-ligand MD simulations.

Chapter 5 enlists the results from virtual screening of chemical libraries followed by *in vitro* screenings in SW480 CRC cell lines and SAR analysis of the initial hit compounds. The best hits from the SAR analysis were further tested by wound healing and colony formation assays. The target specificity of the best hits was tested by siRNA-GPR120 transfected SW480 cell lines.

Chapter 6 summarises the work presented with critical discussion and projects on the future work that could be performed.

Chapter 2
Methods and Protocols

2.1 Computational methodology

In this section, the computational methodologies used are outlined along with the structural and sequence retrieval information. The visual analysis and homology model building were carried out on an in-house 8 node (Intel® Core™ i7-4790 CPU @ 3.60GHz × 8) Linux cluster. The molecular dynamics simulations were performed on the Irish Centre for High-End Computing (ICHEC – www.ichec.ie) cluster. Class B (Project name: nmlif042b) and Class C (Project name: dtlif001) projects were approved by the Irish Centre for High End Computing (ICHEC), which provided 60,000 CPU core units for computational processing and 1000 gigabits (GB) of data storage space.

2.1.1 Protein data retrieval and analysis

The primary protein sequences used for various bioinformatics analysis were retrieved from the Universal Protein Resource (UniProt) database (Uniprot, 2016) that contains ~ 550,000 manually annotated and reviewed protein entries. The short (S) and long (L) isoform (Q5NUL3-2 and Q5NUL3) sequences of human GPR120 were used. Class A GPCRs for which crystal structures were known (in January 2017), were retrieved from the RCSB protein databank (Deshpande et al., 2005) to be used as a template database for homology modelling. The X-ray crystal structures were manually cleaned using PyMol (DeLano, 2018) and screened to include only templates in the inactive state i.e., co-complexed with an antagonist or inverse agonist (Table 3.1).

The sequence alignments of screened templates were performed against human GPR120 (S and L) sequences on the Clustal Omega server (Sievers and Higgins, 2017) and the sequence alignments were visually inspected for the alignment of conserved motifs in the TM helices of the sequences. Three different secondary structure prediction algorithms - UCD-Porter, JPred, TMHMM server (Mirabello and Pollastri, 2013; Drozdetskiy et al., 2015; Krogh et al., 2001) were used to predict the range of α-helical TM regions and the topologies of the loop regions in the extra- and intracellular regions of the plasma membrane. The phylogenetic analysis of the screened templates with human GPR120 (S and L isoforms) was performed in SeaView (Gouy, Guindon and Gascuel, 2009) using the PhyML algorithm at

100 bootstrap values (assigns measures of accuracy to sample estimates) to trace evolutionary sequence conservation.

The 3D crystal structures of selected homologous proteins were visually analysed and screened for the presence of non-protein molecules (such as water molecules, lipids, heteroatoms, etc.) using PyMol Open-source version 2.1.0 (DeLano, 2018). The PDB files of 3D structures were cleaned manually to keep a single chain of protein in PyMol (DeLano, 2018).

2.1.2 Homology modelling and validation

The homology modelling of the human GPR120 (S and L isoforms) receptor was performed using the MODELLER v9.14 homology modelling tool (Webb and Sali, 2016). The procedure for modelling involved python scripts run through command line for advanced homology modelling based on multiple templates. The multiple templates-based HM in MODELLER started from multiple sequence alignment (MSA) (Appendix Table IIa), template structures alignment, and 3D model generation to model evaluation by discrete optimized protein energy (DOPE) based scoring function. Each step was processed by a separate python script supplied with the MODELLER installation suite. The automated sequence alignment file generated by MODELLER was replaced by a manually annotated sequence alignment file generated by Clustal Omega (Sievers and Higgins, 2017). The 3D model generation script was set to generate 100 models for each template group as multiple template combinations were used for HM development (detailed in section 3.4.2).

As mentioned, different combinations of selected templates (Appendix Table IIa) were used for 3D model generation of human GPR120 S and L isoforms – making the process repetitive and time consuming. A python-based pipeline was developed in KNIME (Konstanz Information Miner; Berthold et al., 2008), GITHUB (https://github.com/jay4pal/KNIME_GPCRs) which read the user-defined configuration file to automatically make template combinations to build and evaluate the homology models (Figure 2.1). KNIME is a data analytics platform which allows the integration of configurable nodes and community-built nodes in different scripting languages. The nodes developed for the HM pipeline using MODELLER were configured in the Python 3.0 scripting language. The output from the pipeline consists of sub-subdirectories named after each template combination containing respective models generated and a table of scores of the models.
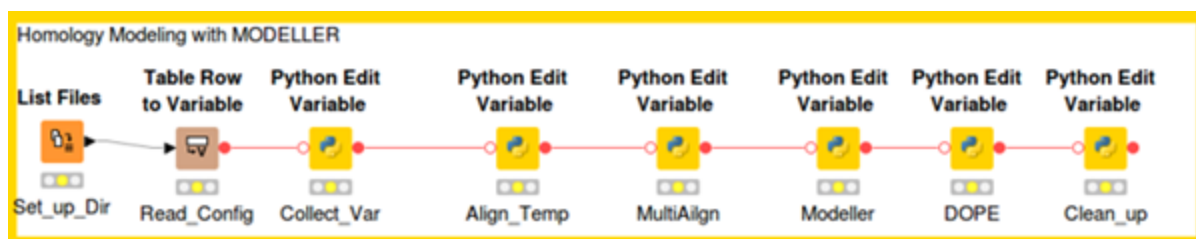
Figure 2.1: KNIME workflow to automate the homology modelling using MODELLER (moving from left to right). "Set_up_Dir" node collects the configuration file from working directory which is read by the "Read_config" node into machine language and then parsed by the "Collect_Var" node. The next four nodes align the templates, generate 3D models, and evaluate them by the scoring function respectively. The last node removes the temporary files created during the processing.

The top ten scoring models from the selected template combination were assessed for stereochemical properties using the MolProbity webserver (Chen et al. 2009). The side chains of the top ten selected models were optimized by SCWRL4 (Krivov, Shapovalov and Dunbrack, 2009) followed by a second stereochemical assessment using the MolProbity webserver (Chen et al., 2009). The difference in stereochemical properties between before and after structures were compared to determine if side-chain optimization by SCWRL4 improved the structural stability. Finally, the structural conformations and stereochemistry of the transmembrane (TM), intra-cellular loop (ICL) and extra-cellular loop (ECL) regions of the predicted model were assessed through Ramachandran (ϕ-Ψ) plots using the MolProbity webserver (Chen et al., 2009) and structural quality by ProSA (Wiederstein and Sippl, 2007) and ERRAT (Colovos and Yeates, 1993) webservers were used.

2.1.3 Molecular docking protocol

SMINA – scoring and minimization in AutoDock VINA – was the docking program used for molecular docking (Koes, Baumgartner and Camacho, 2013; Trott and Olson, 2009) (selected after benchmark studies of four different docking algorithms - Appendix III). SMINA is available as an open-source program, and it can treat both the ligand and the protein as flexible structures. Prior to molecular docking, the protein structure was prepared in AutoDock Tools (ADT) (Morris et al., 2009). The protein was read in from the PDB file to remove non-polar hydrogens, add polar hydrogens and charges are applied. Specific residues in the binding pocket were selected as residues (Hudson et al., 2014) with flexible sidechains and the output file were saved in as PDBQT format with default charges and torsions. For the

next step, the search space – the binding site grid box (Table 2.1) was visually defined by employing AutoDock tools' Grid setting feature, based on the site-specific mutation study by Hudson et al., 2014 to include residues – Arg99(TM2), Trp104 (ECL1), Phe115 (TM3), Trp207 (TM5), Phe211 (TM5), Trp277 (TM6) and Phe304 (TM7), deemed essential for ligand binding.

Table 2.1: Binding pocket co-ordinates and box size used in docking experiments with the GPR120 homology model.

center_x	61.622
center_y	59.75
center_z	46.597
size_x	60
size_y	60
size_z	60
Spacing	0.375 Å

The docking experiment was run through the command line which defines the ligand or an ensemble file of multiple ligands, target receptor to dock into, the binding pocket space (Table 2.1) and the exhaustiveness set at 8 (it can be set between 1 and 8 – where 8 is the highest level of exhaustive search with a net increase on the computational time). Each docking experiment was run at random seed to keep the experiment random and unbiased.

For VS of large chemical libraries, the chemical libraries were retrieved and processed in Biovia Discovery Studio's Pipeline Pilot v9.1 from Dassault Systèmes (Dassault Systèmes 2017) to remove compounds with molecular weight (greater than) > 650 and (less than) < 250 Dalton. The stereoisomers of all the ligands used for docking experiments were built in the Pilot Pipeline tool available with BIOVIA suite (by Dr Gemma Kinsella). The filtered chemical database was energy minimised using MMFF94 forcefield (steepest descent) in the open-source OpenBabel software package (http://openbabel.org/) and saved in sdf format. The docking experiments with single ligands were performed on the in-house 8 core linux cluster while docking experiments with large ligand ensemble databases (such as the decoy database, virtual screening databases, etc.) were performed on the ICHEC cluster. For chemical database libraries ICHEC cluster's taskfarm module was used, which read in the shell script containing command line arguments. The taskfarm module allocates one command per thread for execution on ICHEC cluster.

Later, in collaboration with Dr Anthony J. Chubb (Royal College of Surgeons, Ireland), the KNIME pipeline for HM was extended to include the molecular docking pipeline and increase the flexibility to give more options to the user (detailed in Appendix IV).

2.1.3.1 Enrichment study

In the enrichment studies, a set of inactive compounds was used to perform the enrichment test. A group of 28 active ligands of GPR120S was selected based on their known activity and potency from the literature (Appendix table IIb). The active ligands were used as the active set (i.e., known binders), and for the decoys, a set of 1,150 (unique) druglike decoys obtained from DUD.E (Database of Useful Decoys: Enhanced) (Mysinger, Carchia, Irwin and Shoichet, 2012). Tautomer and stereoisomers were not generated for the decoy databases. The decoys were energy minimized using Avogadro software and docked with the GPR120S model using the docking and binding protocols (section 2.1.3) in SMINA. The docked molecules were then ranked according to the docking score obtained from SMINA. Receiver Operating Characteristic (ROC) curves were then plotted and their corresponding AUCs calculated using R-Studio's ROCR package with gplots (RStudio Team 2015). The ROC graphically relates the sensibility and the specificity of the selected model.

2.1.3.2 Scoring function for VS of chemical libraries

Default scoring function of AutoDock SMINA was used for all docking experiments unless mentioned otherwise. The docked poses generated by AutoDock SMINA by VS chemical libraries (Chapter 5) were rescored with three different scoring functions, AutoDock Vina (Trott and Olson, 2009), NNScore 2.0 (neural network-based scoring function) (Durrant and McCammon, 2011) and DLSCORE (Deep learning based scoring function) https://github.com/sirimullalab/DLSCORE), to calculate the consensus binding affinity score (C_{score}) using *Equation I*. AutoDock VINA and SMINA predicts the binding affinities as approximations of interaction energies (kcal/mol) such as Gaussian, repulsion, hydrogen bonding, hydrophobic, non-hydrophobic contact and Lennard-Jones 4–8 van der Waals interactions (Koes and Camacho, 2012), while NNScore 2.0 and DLSCORE predicts the binding affinities as pK_d values (Brunner, 2004).

Equation I:

$$C_{score} = \left. V_{score} + S_{score} + (-NN_{score}) + (-DL_{score}) \middle/ Number\ of\ scoring\ functions \right.$$

where: C_{score} is consensus docking score.

V_{score} is docking using AutoDock Vina's default scoring function.

S_{score} is docking score using SMINA's default scoring function.

NN_{score} is docking score using NNScore 2.0 function.

DL_{score} is docking score using DLScore function.

2.1.4 Energy minimization (EM) and molecular dynamics (MD)

As mentioned in section 1.5, GPR120 is transmembrane receptor, which spans through the thickness of the plasma membrane (~35 Å) (Calmet et al., 2016). As such the predicted models of GPR120S and GPR120L were inserted in a lipid bilayer to mimic the *in-situ* environment and relax the predicted receptor models. The lipid bilayer of 512 molecules of 1-palmitoyl-2oleoyl-sn-glycerophosphocholine (POPC) was generated from the Berger lipid parameters for POPC molecules (gro and itp files), obtained from the ATB repository (Koziara et al., 2014), using GROMACS v5.1.4 (Abraham et al., 2015). The protein structure file of the predicted structure was converted to a GROMACS readable file using the gmx_pdb2gmx function with the GROMOS 54a7 forcefield. Then the protein was inserted into the generated lipid bilayer using Lambda and InflateGro2 program (Schmidt and Kandt., 2012) and solvated with water molecules using the single point charge water (SPC) model (spc216.gro) parameter file using gmx_solvate function of GROMACS. The charge on the solvated protein-lipid system was neutralized through addition of the required number of Na^+ and Cl^- counter-ions with gmx_genion. As the system prepared did not contain any bound ligand in the protein – it was termed the Apo protein system. Similarly, ligand-bound protein models of GPR120S were prepared (using the molecular docking protocol) and inserted into the lipid bilayer. The general workflow of GROMACS based MD system preparation, equilibration, production run and analysis are shown in Figure 2.2.

As GROMACS cannot prepare the topologies of heteroatoms (ligands), the topology and GROMOS force field parameter for the ligands were obtained from the ATB repository server (Koziara et al., 2014). The prepared protein system was subjected to 50,000 steps of the steepest descent EM algorithm with maximum force set at 1,000 kJ/mol/nm, which was further consecutively reduced to 10 kJ/mol/nm using a conjugate gradient EM. The ligand-bound systems were energy minimized using position restraints on the ligand topologies to keep the ligands bound to the protein. The ligand position restraints were used during EM and equilibration steps in all the MD simulations and removed in MD production runs unless mentioned otherwise.
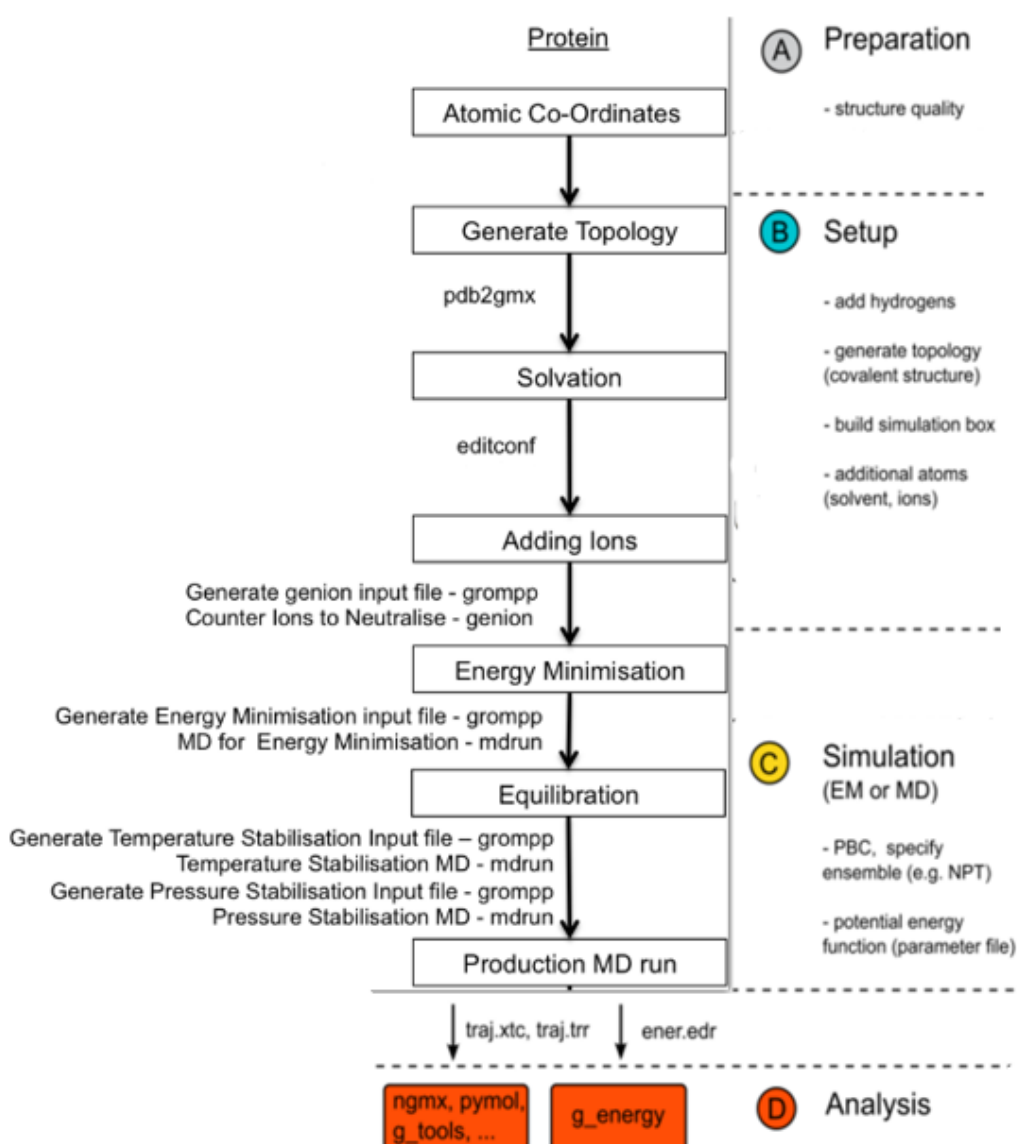
Figure 2.2: GROMACS Molecular Dynamics Simulation Workflow. (Image adapted from http://www2.mpibpc.mpg.de/groups/grubmueller/Lugano_Tutorial/)

For the MD equilibration runs, the energy minimized systems were subjected to a 500 ps position restraining simulation with 1,000 kJ mol^{-1} harmonic constraint to relieve the close contacts with POPC and water under NVT (constant Number of particles, Volume, and Temperature) ensemble conditions, with a Vrescale (modified Berendsen) temperature coupler (Bussi, Donadio and Parrinello., 2007). This was followed by another 5 ns equilibration run under NPT (constant Number of particles, Pressure, and Temperature) ensemble conditions, before a final production run of 300 ns on each system. The systems were run at 300 K, i.e., above the phase transition temperature of pure POPC, to ensure that the lipids maintained their proper density, and 1 bar pressure under isothermal-isobaric ensemble with ligand constraints. Nosé-Hoover (which is used widely for membrane NPT and MD simulations) temperature and Parrinello-Rahman pressure couplers were used to maintain the temperature and pressure values with the protein, ligands, lipids and water (plus ions) molecules coupled separately with a coupling constant of τ_t=0.1 ps. Semi-isotropic pressure coupling was set with τ_p=2 ps, allowing the bilayer to deform in the x–y plane independently of the z-axis. A time-step of 2 fs was used throughout with periodic boundary conditions. LINCS constraint algorithm was used to maintain the geometry of the molecules (Hess, 2008). Long-range electrostatic interactions were calculated using the particle-mesh Ewald (PME) method. Van der Waal's interactions and Coulomb interactions were cut off at 12 Å with updates every five steps. The checkpoint on the MD production run were saved every 50 ps which recorded the conformational 3D coordinates, system velocities and energy parameters.

The trajectory and energetics of the protein systems were analysed by in-built GROMACS tools and visualized in PyMol (DeLano 2018) and XMGRACE (http://plasma-gate.weizmann.ac.il/Grace/). The overall stability of the simulated systems was also checked with respect to temperature, pressure, and potential energy of the systems to check thermodynamic equilibrium during the production simulation runs, confirming the convergence of individual trajectories. The webserver WADDAICA was used to compute the protein-ligand binding free energies of protein-ligand snapshots extracted from the MD trajectory every 5 ns using the Binding Affinity by AI module of WADDAICA (Bai et al., 2021). The protein-ligand interaction plots from the MD trajectory were generated using Molecular-dynamics-Interaction-plot. (https://github.com/tavolivos/Molecular-dynamics-Interaction-plot).

2.1.5 Structure-based pharmacophore screening

A six-feature pharmacophore model was generated by manual curation of docked reference structures (TUG-891 and compound **39**) using the ZINCPharmer web interface (Sterling and Irwin, 2015; Koes and Camacho, 2012) (Table 2.3), which was then screened against the ZINC15 database containing ~230 million purchasable lead-like 3D molecules (Sterling and Irwin, 2015). The pharmacophore screening results from ZINCPharmer were downloaded as structures data files (sdf format). The 3D molecules were then prepared and docked against the GPR120S receptor model following the set protocol in section 2.1.3.

2.1.6 Pharmacokinetic profiling

The physicochemical and drug-likeness profiling of the screened hits was performed using the online webserver SwissADME (Daina, Michielin and Zoete, 2017). The selected compounds were prepared for 100 ns MD simulations following the protocol described in section 2.1.3 and 2.1.4. Multiple linear regression (MLR) was used to correlate the physicochemical descriptors (obtained from SwissADME) with the binding affinity predictions from WADDAICA (Bai et al., 2021). The ChemMine tools webserver was used to develop a structural similarity clustering scatterplot (Backman, Cao and Girke, 2011). ChemMine clustering tool converts the Tanimoto similarity matrix into distance matrix to cluster the molecules.

2.2 Biochemical assay protocols

2.2.1 Cell culture and materials

SW480, a human CRC cell line was obtained from Dr Alan Casey's Nanolab (TU Dublin) and was cultured in RPMI1640 growth media (R8758) supplemented with 10% (v/v) Fetal Bovine Serum (cat no. F7524) and incubated at 37 °C ± 1 °C in a humidified atmosphere with 5% CO_2. Trypsin-EDTA 0.25% sterile-filtered (cat no. T4049) solution was used for detaching adherent cells from the culture flask surface. The autoclaved sterile phosphate buffered saline (PBS) was used for washing purposes. The PBS tablets (cat no. P4417) were dissolved in deionised water as per the manufacturer's instructions to yield 140mM NaCl, 10mM phosphate buffer and, 3mM KCl solution at pH 7.4. The cell culture media and reagents were procured from Merck-Sigma, unless otherwise mentioned. TUG891 (SML1914), AH7614 (SML2025), Adapalene (A7486), Azilsartan (SML0432), Gliquidone (CDS021537) and Lapatinib (SML2259) were procured from Sigma-Aldrich (Merck), and Sonidegib (M4841) was procured from AbMole (abmole.com). The test compounds for *in*

vitro screening assays mentioned in section 5.3 were procured from SPECS (www.specs.net). A stock solution of test compounds (20 mM) was prepared in 100% (w/v) DMSO and stored at −20 °C.

2.2.2 SW480 *in vitro* screenings by cell-based cytotoxicity assay

Alamar blue based cytotoxicity assays were performed to determine inhibitory effects of the test compounds. SW480 cells were seeded at 10,000, 5,000 and 2,500 cells per well in 96-well plates for 24, 48 and/or 72 h drug treatment, respectively. Required dilutions of test compounds were freshly prepared in 0.5% (v/v) DMSO growth media for cytotoxicity assays. After the treatment period, drug concentrations were replaced with 6% (v/v) alamar blue dye solution and the cells were incubated for 3 h under cell incubation conditions. Finally, the fluorescence signal was read using 560 nm excitation and 590 nm emission filters with Varioskan LUX Multimode Microplate Reader from ThermoFisher Scientific.

2.2.3 SW480 siRNA transfection

SW480 cells were transfected with 50 nM GPR120 (human) − 27mer siRNA duplex (cat no. SR317391) using siTran 2.0 siRNA transfection reagent (OriGene – www.origene.com) according to manufacturer's instructions after overnight seeding. Universal scrambled siRNA duplex (cat no. SR30004) was used as negative control in transfection experiments. The 27mer GPR120-siRNA duplex was provided as two different samples labelled –

siRNA-1 (rArGrArArArUrGrArCrUrUrGrUrCrGrArUrUrArUrUrUrCTG) and

siRNA-2 (rGrGrArUrGrCrArArGrArGrCrUrGrUrCrGrUrGrArCrUrCAC).

siRNA-1 was selected for GPR120 silencing in *in vitro* assays based on the results obtained from relative quantification studies (section 2.2.4 and 5.2.2; Figure 5.3a).

2.2.4 RT-qPCR protocol

Real-time quantitative polymerase chain reaction (RT-qPCR) was used for detection and measurement of GPR120 expression in control and siRNA-treated SW480 cell lines. Fast 96-well PCR reaction (cat no. 7321161) plates covered with optical adhesive films (cat no. 4360954 from Thermo Fisher Scientific) were used in Applied Biosystems 7500 Fast Real-time PCR system for the RT-qPCR run. RT-qPCR run was performed according to manufacturer's instructions.

2.2.4.1 RNA extraction

RNA was extracted from GPR120-siRNA treated SW480 cells using TRI reagent (cat no. 93289) obtained from Merck-Sigma and following the protocol provided by the manufacturer. After 24 hours of siRNA transfection treatment cells were collected from 6 well plates by directly lysing the cellular monolayer using 1 ml per well TRI solution in a sterile fume hood. After addition of TRI solution, the cell lysate was repeatedly pipetted up and down to form a homogenous lysate mix. The homogenous lysate mix was transferred to 1.5 ml sterile microcentrifuge tubes and allowed to stand for 5 minutes at room temperature. Then for each 1ml of TRI solution 0.2 ml chloroform (HPLC grade) was added and mixed thoroughly by gently inverting the tubes repeatedly for 15 seconds and then allowed to stand for another 5 minutes. The resting tubes were transferred to the centrifuge pre-calibrated at 4°C, 12,000 x g for 15 minutes. At this stage, the centrifugated mixture was separated into 3 phases- colourless top aqueous phase containing RNA, a turbid interphase containing DNA and the pink-coloured organic phase at bottom containing protein materials.

The colourless aqueous phase was transferred to a fresh tube followed by addition of 0.5 ml 2-propanol (HPLC grade) per ml of TRI reagent used. The sample tubes were allowed to stand for 5 min at room temperature and then centrifuged at 12,000 x g for 10 minutes at 4°C. After centrifugation, the RNA precipitate formed a colourless pellet at the bottom of the tube. The supernatant was removed, and the pellet was washed with 1ml of 75% ethanol per 1ml of TRI solution used. The sample was vortexed and then centrifuged at 7,500 x g for 5 minutes at 4°C. The excess of ethanol supernatant was removed, and the pellet was left for air drying in the PCR laminar hood to remove the residual ethanol. The dried RNA pellet was re-suspended in 20 μL of RNase/DNase-free-water.

The RNA yield of the collected sample was calculated using the μDrop plate in Multimode Microplate Reader from ThermoFisher Scientific at 230, 260, 280 nm wavelengths with RNase/DNase-free-water as blank (Appendix Table IId).

2.2.4.2 cDNA synthesis and RT-qPCR run

Total RNA extracted from siRNA treated and wild SW480 cells (control) extracted was transcribed into complementary DNA (cDNA) following the instructions provided with the qScript cDNA SuperMix (cat no. 95048-025) obtained from Quantabio. SuperMix, extracted RNA samples (from siRNA-1 treated, siRNA-2 treated and untreated control cells) and RNase/DNase-free water were thawed on ice and centrifuged for 5-10 seconds to collect

residual liquid from sides of the tube. For each sample, 1 µg of RNA template was mixed with 4 µL of qScript cDNA SuperMix and make up the volume to 20 µL with RNase/DNase-free water in 0.2mL micro-tubes sitting on ice. A control sample without RNA template was prepared as well. After sealing each reaction tube, the mix was gently vortexed and then centrifuged to collect component at the bottom of tube.

Appropriately labelled reaction tubes were placed in the Thermocycler with following cycling steps:

Cycle 1:	25°C for 5 min
Cycle 2:	42°C for 30 min
Cycle 3:	85°C for 5 min
Cycle End:	4°C hold

If RT-qPCR was not performed on the same day, the cDNA mix tubes were stored at -20°C, else the cDNA mix tubes were place on ice and steps for RT-qPCR run were followed.

The second step of RT-qPCR follows the protocol provided with PerfeCTa SYBR Green FastMix – Low ROX (cat no 95074-025 from Qunatbio). The primers for GPR120 were designed using Primer3 online webserver (Ye et al., 2012) with NM_181745.4 (Homo sapiens free fatty acid receptor 4, transcript variant 1, mRNA) as input and optimum primer melting temperature (T_m) set at 60 °C. The primer sequence for GAPDH (housekeeping gene as control) was obtained from literature (Wang et al., 2019). The primers (Table 2.2) were obtained from Merck Sigma in lyophilised form and reconstituted with RNase/DNase-free water to prepare 100µM solutions.

Table 2.2: Sequence of primers used for RT-qPCR of GPR120 and GAPDH.

GPR120 Forward primer	5′GGATGCAAGAGCTGTCGTGA3′	Tm = 60.39°C
GPR120 Reverse Primer	5′TTACCGACGCTGTGGATGTC3′	Tm = 60.11°C
GAPDH Forward primer	5′GGATTTGGTCGTATTGGG3′	Tm = 53.5°C
GAPDH Reverse Primer	5′GGAAGATGGTGATGGGATT3′	Tm = 54.3°C

The final reaction cocktail for each sample was prepared by mixing 10 µL of PerfeCTa SYBR Green FastMix – Low ROX with 5 µL of cDNA and 1 µL of each forward and reverse

primer (10μM) and make up the volume to 20 μL with RNase/DNase-free water in each well of 96-well Fast reaction plate. The plate layout is illustrated in Table 2.3. To prevent pipetting error of small volumes, 80 μL of reaction mix for each sample was mixed well in 1 well and then dispensed 20 μL to each well of the corresponding column. The optical adhesive film was used to seal the plate and the plate was centrifuged briefly to collect the residual volumes from sides of wells at the bottom. Finally, the plate was placed in the Applied Biosystems 7500 Fast Real-time PCR system with following cycling steps:

1) Initial denaturing (holding stage):   95°C   30 sec
2) PCR cycling (45 cycles):   95°C   05 sec
3) Collect data at end of extension step: 60°C   30 sec

Table 2.3: RT-qPCR 96-well plate layout followed for the experiments.

	Untreated Control	siRNA-1 treated	siRNA-2 treated	Blank (No cDNA)
GPR120	Repeat 1	Repeat 1	Repeat 1	Repeat 1
	Repeat 2	Repeat 2	Repeat 2	Repeat 2
	Repeat 3	Repeat 3	Repeat 3	Repeat 3
GAPDH	Repeat 1	Repeat 1	Repeat 1	Repeat 1
	Repeat 2	Repeat 2	Repeat 2	Repeat 2
	Repeat 3	Repeat 3	Repeat 3	Repeat 3

At the end of RT-qPCR run, the result sheet was exported as an excel file for final calculations.

2.2.5 Wound healing / Scratch assay

SW480 cells were seeded at 3x 10^6 cells per well in 6 well-plates 24 hours prior to scratch to obtain a confluent monolayer. The monolayer was scratched using a sterile 200 μL pipette tip held at 45° angle in the laminar hood. Each well was scratched both horizontally and vertically to obtain at cross point at centre of the well. The cross point was used as a reference point while taking snapshots of wells. The old media was discarded, and cells was gently washed with PBS solution without disturbing or dislodging the cells in the well. After washing, the vehicle control and subtoxic (IC_{25}) concentrations of test compounds were added. Snapshots of scratched surfaces were taken at 10X zoom using the Tucsen camera (ISH500) mounted Optika XDS-2 trinocular inverse microscope and plates were incubated at 37°C with 5% CO_2 for 24 hours. After incubation, plates were removed turn by turn and snapshots were taken at 10X zoom following the same protocol. Image processing and

relative scratch area quantification was performed using Fiji (Schindelin et al., 2012) with the Wound Healing tool plugin. Three independent experiments were performed for each test compound in triplicates.

2.2.6 Clonogenic assay

SW480 cells (5 cell/well) were seeded on 96-well plates and maintained in a humidified chamber atmosphere comprising 95% air and 5% CO_2 at 37°C overnight. After which, media was replaced with serial dilutions of test compounds 5, 7 and 2X prepared in 0.5% DMSO / RPMI1640 media and incubated for 2 weeks at 5% CO_2 at 37°C. Following treatment, cells were fixed with 0.25 % glutaraldehyde at room temperature for 20 min and stained using crystal violet solution (0.1% crystal violet and 1X PBS) at room temperature for 30 mins. Stained cells were washed with water by dropping gently, and air dried at room temperature. The number of colonies were quantified using the Tucsen camera (ISH500) mounted Optika XDS-2 trinocular inverse microscope and Fiji program (version 1.6.0; Schindelin et al., 2012). Survival rate of the colony formation was calculated using the equation II:

Equation II:

$$Survival\ rate = \frac{No.\ of\ colonies\ formed\ after\ survival}{No.\ of\ cells\ seeded\ X\ 100}$$

2.3 Statistical analysis

All data was presented as mean ± standard error of the mean (SEM). For the analysis and graphical representation of biological experimental IC_{50} values were calculated (GraphPad Prism 6 Software, La Jolla CA) using dose–response curves for the compounds. For plotting the dose–response curve, a non-linear regression curve fitting method was used where the mean positive control (50% DMSO in growth media) was defined as 0% and the mean vehicle control (0.5% DMSO in growth media) was defined 100%. For significance analysis, two-way analysis of variance (ANOVA) followed by Tukey's multiple comparison test was performed. $P < 0.05$ was considered statistically significant.

Chapter 3

Homology Modelling of G-Protein Coupled Receptor: GPR120

GPCRs are the largest family of transmembrane proteins in the human body comprising ~800 distinct receptors. In 2000, the first high resolution crystal structure of a mammalian GPCR - bovine rhodopsin – PDB-ID: 1F88 (Palczewski, 2000) was resolved. Later in 2007, a structure of an inverse-agonist bound human β_2 adrenoceptor was crystallized in a lipid environment at 3.4/3.7 Å resolution (Rasmussen et al., 2007). These structures provided significant insights into the GPCR function and stability of the inactive state upon binding of antagonist or inverse agonist. Since then, crystal structures of 110 unique GPCRs – 88 Class A GPCRs, and ~320 ligand-receptor complexes have been elucidated (https://gpcrdb.org/structure/statistics - January 2022; Kooistra et al., 2020). However, about 80% of GPCRs have not yet been crystallised including FFA2, FFA3 and GPR120. The process of crystalising GPCRs remains slow and challenging (Grisshammer, 2017) due to low expression levels and difficulties regarding the crystallization process itself (Ghosh, Kumari, Jaiman and Shukla, 2015). The available experimentally (NMR or X-ray) determined GPCR structures have been used in comparative homology modelling and *in silico* / virtual screening (VS) campaigns of early-stage drug discovery (Chahal, Nirwan and Kakkar, 2019; Lin, Li and Lin, 2020; Muhammed and Aki-Yalcin, 2018). Subsequently, millions of compounds can be computationally screened through VS to discover novel drug-like compounds (Dailey et al., 2009).

Based upon the observation of the experimentally determined 3D protein structures, the central dogma of structural biology of proteins is stated as the tertiary (3D) structures of homologous proteins are evolutionarily more conserved than their primary structure (amino acid sequence) (Kaczanowski and Zielenkiewicz, 2009). Therefore, if two proteins have sufficient similarity in sequence (more than 30 %) (Waterhouse et al., 2018), the structural information of one protein (the template or known 3D structure) can be used as a scaffold for generating a 3D model of the other protein (target – unknown 3D structure). The amino acid sequence of the protein of interest and additional information from the protein database websites are used to identify different goals such as phylogeny of the protein, the prediction of secondary and tertiary structure of the protein (homology modelling - HM) and the

identification of functional properties (Basith et al., 2018; Kaczanowski and Zielenkiewicz, 2009).

HM has been used to predict the structures of various GPCRs (Sailapathi et al., 2021) and predicts the 3D-structure of an unknown protein based on the known structures of a similar / homologous protein from its amino acid sequence with an accuracy that is comparable to the best experimental results such as NMR or X-ray crystal structures. The HM methods primarily consists of four different steps: template selection, target-template alignment, model building and model assessment. The steps of HM are briefly described below as in (Webb and Sali, 2016). (Figure 3.1).
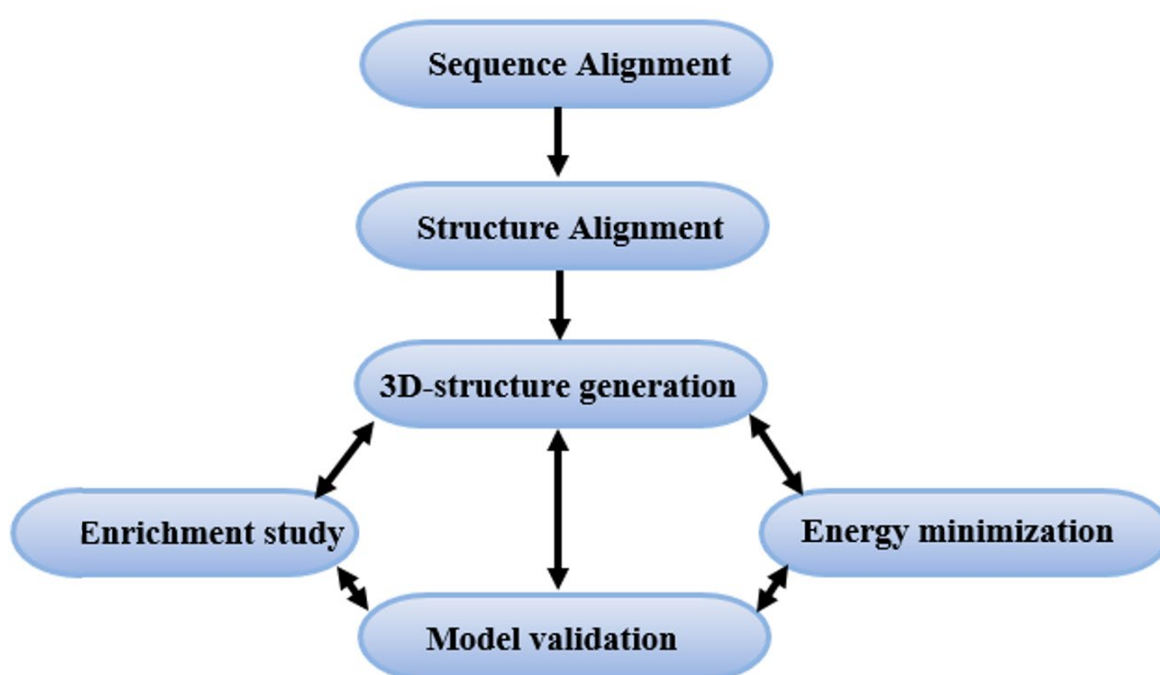
Figure 3.1: General schematic of homology modelling methodology followed.

When the experimental 3D structure of a target protein is known, protein-ligand interactions and binding site information can be used to generate an ensemble of binding modes of protein-ligand configurations. Different ligands can then be ranked with respect to their binding modes in the binding pocket of the receptor (Beuming et al., 2015). In cases where, the crystal structure of the target protein is unknown, and the 3D structure has been generated by HM, it may contain inaccuracies and errors that may interfere with ligand binding in the binding site of the receptor. Enrichment studies can be performed to measure the selectivity of the receptor towards the active known binders through molecular docking studies. The enrichment factor demonstrates the ability of the generated model to enrich the known / active ligands of the receptor in the top ranks of the screening from among a wide database of

molecules. The screening databases for enrichments are prepared to contain actives / known ligands and drug-like decoys which resemble the active ligands in geometry but are chemically distinct from actives. The enrichment factor is evaluated by plotting the corresponding Area Under the Receiver Operating Characteristic (ROC) curves (AUC) (Lenselink et al., 2014; Mysinger, Carchia, Irwin and Shoichet, 2012).

The overall objective of this chapter is the generation and validation of a structural model of human GPR120 receptor as well as confirm the basis of recognition of known GPR120 agonist - TUG891 by molecular docking analysis.

3.1 Template selection and sequence alignment

The target protein sequence is used as the query to identify the experimental 3D structures of homologous proteins in the databases of protein structures such as RCSB-PDB (Deshpande et al., 2005), GPCRdb (Munk et al., 2016), SCOP (Andreeva et al., 2004), DALI (Dietmann et al., 2001) and CATH (Pearl et al., 2005). The sequence comparison methods - BLAST and FASTA, are utilised to detect sequence identity / similarity and quantify the results in terms of sequence identity or statistical measures such as an E-value or z-score (Altschul et al., 1990). For more distantly related homologs, sensitive searching methods including profile matching and Hidden Markov Models (HMM) (Gribskov, McLachlan, and Eisenberg, 1987; Krogh et al., 1994), position specific scoring matrix (PSSM) (Altschul et al., 1997) are utilized. In general, the heuristic method, BLAST search – Basic Local Alignment Search Tool, is a reliable approach that identifies hits with sufficiently low E-value signifying close evolutionary relatedness. A template with a very poor E-value is not recommended as it can lead to the generation of a misguided model (Altschul et al., 1990).

There are factors other than sequence similarity that need to be considered to select a template from identified templates by the search method. The environmental condition of the target must be considered with respect to the template's native environment such as solvent, pH, and ligands complexed in the crystal structure, quaternary interactions, and the location of the target in the cell (Loo et al., 2018). Another important factor is the experimental quality of the template. The accuracy of a crystallographic structure depends on the resolution and R-factor while for a NMR structure, the number of restraints per residue is the indicative factor for the structure's accuracy (Webb and Sali, 2014; Loo et al., 2018).

The human GPR120S and GPR120L sequences were used unless mentioned otherwise. The sequences of short and long isoforms of GPR120 - human GPR120S (Q5NUL3-2), and

human GPR120L (Q5NUL3) were retrieved from UniProt (Uniprot, 2016) (section 2.1.1). The sequences of other members of the FFAR family - human FFAR1 (O14842), human FFAR2 (O15552) and human FFAR3 (O14843) – were also retrieved from UniProt. As GPR120 belongs to the Class A (Rhodopsin family) of the GPCR gene superfamily, all Class A GPCRs (Kargman et al., 1999; Fredriksson et al., 2003) for which crystal structures have been submitted were retrieved from the RCSB PDB (Table 3.1). The retrieved PDB structure files were screened based on several criteria. Firstly, each template had to include an antagonist or inverse agonist and with an overall sequence identity above 20% with hGPR120 (an exception was for FFAR family members). Secondly, the templates should not be missing more than two loop regions. All the selected templates (shown in Table 3.1) were pre-processed manually to remove any additional structures (such as water molecules, lipids, heteroatoms, ligands, ions, etc.) from the PDB file using Pymol (DeLano, 2018).

To determine the most suitable template(s) for building GPR120 models, the amino acid sequence of each selected GPCR (Table 3.1) was individually aligned against the GPR120S (short isoform) as well as performing MSA using Clustal Omega (Sievers and Higgins, 2017) followed by phylogenetic analysis using SeaView (Gouy, Guindon, and Gascuel, 2009) which read in FASTA sequences.

The crystal structure of human FFAR1 (PDB ID: 4PHU) was elucidated in 2014 (Srivastava et al., 2014). As FFAR1 and GPR120 belong to the same subfamily of FFA receptors of Class A GPCRs, FFAR1 was selected as one of the templates even though the X-ray crystal structure was co-crystallised with an allosteric ligand. The phylogenetic analysis of the FFA receptor subfamily shows that FFAR1, FFAR2 and FFAR3 have more than 30% sequence identity and are phylogenetically closer to nucleotide receptors in the phylogenetic tree of the Class A family of GPCRs (Figure 3.2). Whereas GPR120 has TM sequence identity of 15% with FFAR1 and less than 20% with other members of the subfamily and is clustered with orphan receptors. The PhyML method based phylogenetic analysis in SeaView of the selected template GPCR sequences showed that endothelin, orexins, and opioids were most closely related to human GPR120S in the evolutionary tree (Figure 3.2). Consequently, Orexin receptors (OX1R and OX2R), Opioid Kappa-type receptor (Kappa-type), Opioid delta-like receptor (Delta-type) and Endothelin type-B receptor (Endothelin) were selected for generating human GPR120S homology models.

Table 3.1: Crystal structures of Class A GPCRs in inactive state available in the PDB (Data retrieved: January 2017).

G Protein Coupled Receptor	PDB-ID	Resolution
Crystal Structure of the CCR5 Chemokine Receptor	4MBS	2.71 Å
The structure of the CXCR4 chemokine receptor in complex with small molecule antagonist IT1t	3ODU	2.5 Å
Structure of N/OFQ Opioid Receptor in Complex with Peptide Mimetic	4EA3	3.0 Å
Structure of the human kappa opioid receptor in complex with JDTic	4DJH	2.9 Å
Structure of the human delta opioid 7TM receptor	4N6H	1.8 Å
Solution conformation of substance P in water complexed with NK1R (NMR structure)	2KS9	NMR
Crystal structure of Human GPR40 bound to allosteric agonist TAK-875	4PHU	2.3 Å
Crystal Structure of a Lipid G Protein-Coupled Receptor	3V2Y	2.8 Å
Crystal Structure of Bovine Rhodopsin	1U19	2.2 Å
Crystal structure of the chimeric protein of 5-HT2B-BRIL in complex with ergotamine	4IB4	2.7 Å
Structure of human dopamine D3 receptor in complex with eticlopride	3PBL	2.8 Å
Structure of the human histamine H1 receptor in complex with doxepin	3RZE	3.1 Å
The 2.6 A Crystal Structure of a Human A2A Adenosine Receptor bound to ZM241385	3EML	2.6 Å
Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist	3UON	3.0 Å
Crystal structure of the chimeric protein of 5-HT1B-BRIL in complex with ergotamine	4IAR	2.7 Å
XFEL structure of human Type 1 Angiotensin-II Receptor	4YAY	2.9 Å
Human Orexin 2 receptor bound to Suvorexant (insomnia drug)	4S0V	2.5 Å
Human endothelin receptor type-B in ligand-free form	5GLI	2.5 Å

A multiple sequence alignment employing the Clustal Omega software provided insights about highly and less conserved areas of the target protein (Figure 3.3). Even though all the Class A specific (D/ERY in TM3; WxP in TM6; NPxxY in TM7 – highlighted in Figure 3.3, 3.4) motifs are conserved throughout the FFA receptor family with few exceptions, large gaps and non-identical residues can be observed between the aligned GPR120 isoforms and other members of the FFA receptor family. Interestingly, the highly conserved GPCR D/ERY motif of GPR120 shows a difference where Y (tyrosine) is mutated to M (methionine). There are a few examples of variations at this position in other Class A GPCRs such as for the Oxytocin receptor which has a DRC motif; Cysteinyl-leukotriene type 1 receptor has FRC.

As the sequence alignments for HM are guided by the highly conserved amino acid residues (motifs that are shared by the members of this family) between the template and the target, FFAR1 (the only FFA receptor with an elucidated crystal structure) was not used as it has the alterations in the conserved motifs compared to GPR120 such as ERM to GRY in TM3 – negatively charged residues replaced by small uncharged Glycine which favours turns, WxP to VxP in TM6 and NPxxY to NPxxT of TM7/8 – large aromatic residues replaced by non-aromatic residues.
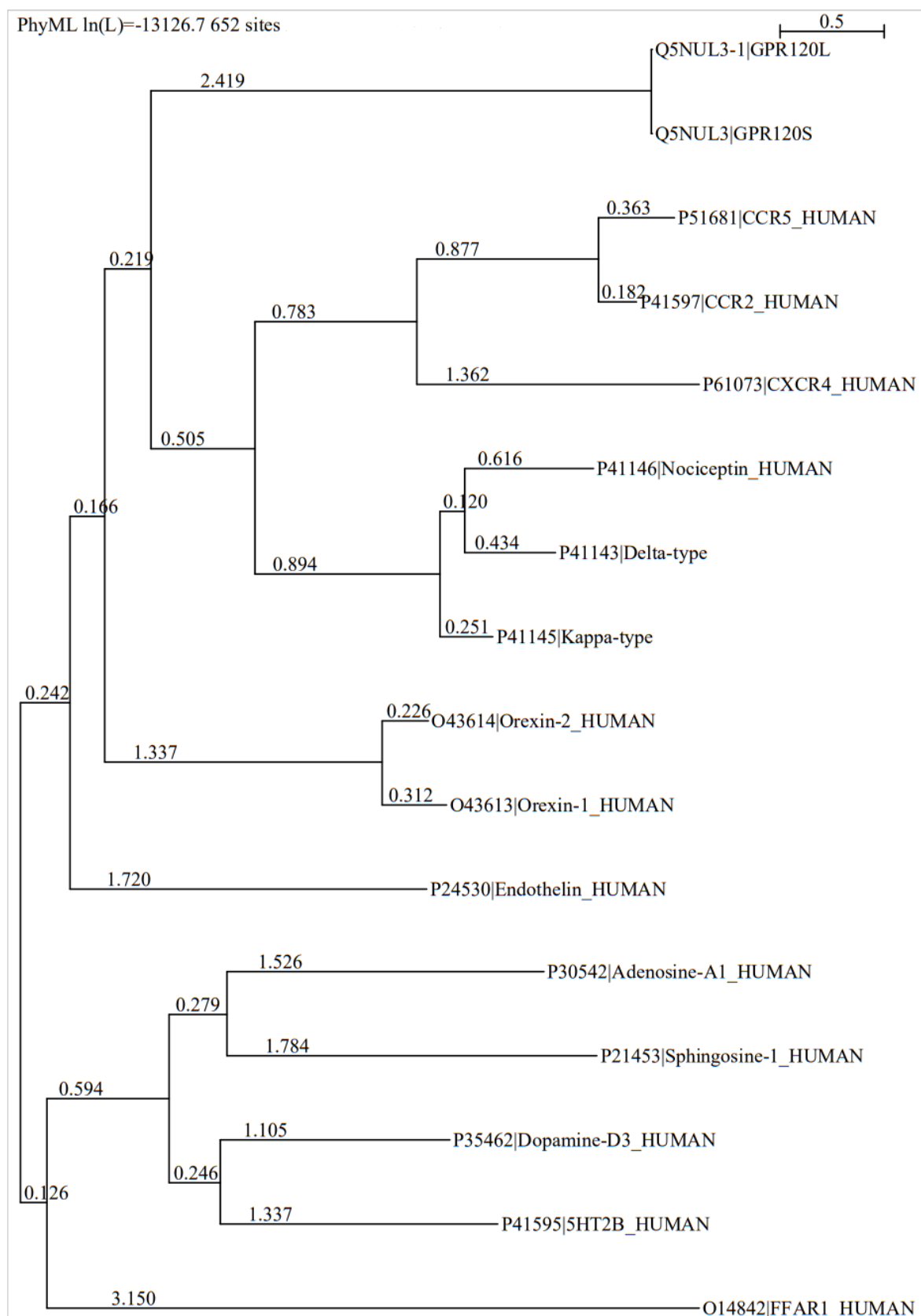
Figure 3.2: Phylogenetic tree of GPR120L and GPR120S with selected Class A GPCRs built in SeaView (Gouy, Guindon, and Gascuel 2009) using the PhyML algorithm at 100 bootstrap values – branch length is directly proportional to the number of substitutions per site.

Figure 3.3: Multiple sequence alignment of the FFA receptor family generated using Clustal Omega. Conserved residues are shown in square boxes; * indicates a fully conserved residue; : indicates conservation between groups of strongly similar properties; . indicates conservation between groups of weakly similar properties. The conserved motifs of Rhodopsin-like family are highlighted in black boxes and disulfide bond forming cysteines are highlighted in yellow boxes.

The sequence identity of the seven transmembrane regions between the GPR120 receptor and the selected GPCRs with elucidated crystal structures (listed in Table 3.1) with a cut off ≥ 20 percent sequence identity is detailed in Table 3.2. Only antagonist or inverse agonist-bound crystal structures were used for templates as they are likely to be in open / inactive conformation (Figure 1.8). The open / inactive conformation of receptor implies that the orthosteric binding pocket of the receptor has not closed on (i.e., trapped) the ligand and ECL2 has not covered the entrance (acting as a lid) of the binding pocket to trap the agonist inside the pocket. These conformational changes are usually observed in the agonist-bound Class A GPCRs (activation detailed in section 1.4.3; Figure 1.8). As the GPR120 models were generated to be used for VS and molecular docking studies, the model was required to be in an open / inactive conformation.

Table 3.2: Sequence identity / similarity scores from pairwise alignment between the selected Class A GPCR templates with %identity ≥ 20 % and GPR120S target sequence for the TM regions from GPCRdb (GPCRdb.org).

GPR120S pairwise aligned against	%Similarity	%Identity	Resolution
Orexin OX2 receptor [4S0V]	45	24	2.5 Å
Opioid Kappa-type receptor [4DJH]	43	22	2.9 Å
Orexin OX1 receptor [4ZJ8]	41	24	2.75 Å
Adenosine A1 receptor [5UEN]	40	22	3.2 Å
Chemokine C-X-C type 4 [3ODU]	39	22	2.5 Å
Endothelin type B receptor [5GLI]	39	21	2.5 Å
Dopamine D3 receptor [3PBL]	39	21	2.89 Å
Chemokine C-C type 5 [4MBS]	38	21	2.71 Å
Opioid delta-like receptor [4N6H]	38	21	1.8 Å
5-hydroxytryptamine receptor 2B [4IB4]	38	21	2.7 Å
Chemokine C-C type 2 [5T1A]	38	21	2.81 Å
Sphingosine 1-phosphate receptor 1 [5A86]	38	21	2.25 Å
Nociceptin/orphanin FQ receptor [4EA3]	37	20	3.01 Å

Furthermore, the templates were examined for their crystal structure resolutions (Table 3.2). The higher resolution (lower values) crystal structures were selected as this indicates a more refined crystal structure useful to investigate receptor-ligand interactions. The crystal structures of the selected templates were also manually screened for the presence of a disulfide bond between TM3 and ECL2 forming a lid to trap ligands as well as alignment of conserved motifs, such as the D/ERY motif in TM3, WxP motif in TM6 and NPxxY motif in TM7 (Figure 3.4).

Figure 3.4: Conserved motifs of Class A GPCRs selected as templates for GPR120 homology model. Sequence alignments were generated in GPCRdb (Munk et al., 2016). The position of the characteristic motifs in the sequence is specified in the Ballesteros-Weinstein numbering scheme (Ballesteros and Weinstein, 1995).

3.2 GPR120 topology prediction and homology model generation.

The secondary structure prediction for each amino acid is "three-state" as each residue can be in one of three "states": α, β, or other (O) and secondary structure prediction algorithms and databases can be used to predict whether an amino acid is in a α−helix, a β−sheet, or a loop region. The secondary structure information of GPR120 isoforms available on Uniprot (Uniprot, 2016) when compared with other databases such as GPCRdb (Munk et al., 2016) and the results from various secondary structure prediction algorithms (UCD-Porter, JPred, TMHMM) (Mirabello and Pollastri, 2013; Drozdetskiy et al., 2015; Krogh et al., 2001) showed different ranges of TM region (Appendix V a-c To have a singular range of TM regions, a consensus method from three different secondary structure prediction algorithms (UCD – Porter, JPred, TMHMM) was applied and an overlapping topology range of residues was selected (Table 3.3).

Table 3.3: Predicted topology of the TM regions of GPR120 short and long isoform based on the consensus from the topology prediction algorithms.

Helix	Amino acid sequence	Range GPR120S	Range GPR120L
TM1	*LVLAAVETTVLVLIFAVSLLGNVCALVLVA*	37-65	37-65
TM2	*ATACLVLNLFCADLLFISAIPLVLAVRWT*	73-101	73-101
TM3	*PVACHLLFYVMTLSGSVTILTLAAVSLERMVCIV*	107-141	107-141
TM4	*RRARAVLLALIWGYSAVAALPLCVF*	152-175	152-175
TM5	*EISWDVSFVTLNFLVPGLVIVISYSKILQIT*	204-233	204-233
TM6	*IRVSQQDFRLFRTLFLLMVSFFIMWSPIIITILLILI*	252-289	268-305
TM7	*VIWPSLFFWVVAFTFANSALNPILYMTLCRNEWK*	296-324	312-347
TM8	*CRNEWKKIFCCFW*	326-339	349-362

MODELLER v9.14 (Webb and Sali, 2016) was incorporated into an in-house KNIME pipeline for HM (Figure 2.1) and used for generating GPR120 homology models using the protocol explained in section 2.1.2. For each template combination (shown in Appendix Table IIa), 100 homology models were generated. Initially, models were generated using the

complete sequence of both GPR120S (361 residues) and GPR120L (377 residues). As the protein sequence contains a long N-terminal chain (35 residues) and ICL3 of 17 residues in GPR120S and 33 residues in GPR120L, the initial models were generated with disordered structures as shown in Figure 3.5. The N-terminal chain of GPR120L was bending towards the intracellular region, implying that it was entering the membrane bilayer while in GPR120S, the N-terminal chain was wrapped around the ECL2; and ICL3 (Cyan – in GPR120L - long isoform only) was packed into the intracellular pocket, which would result in steric clashes with trimeric subunits of G-protein (detailed in section 1.4.3; Figure 1.8). A manual screening was applied on such GPR120 models in which ECL and ICL as well as N and C-termini enter the membrane as MODELLER protocols (section 2.1.2) do not include any information about the location of a protein with respect to the membrane.
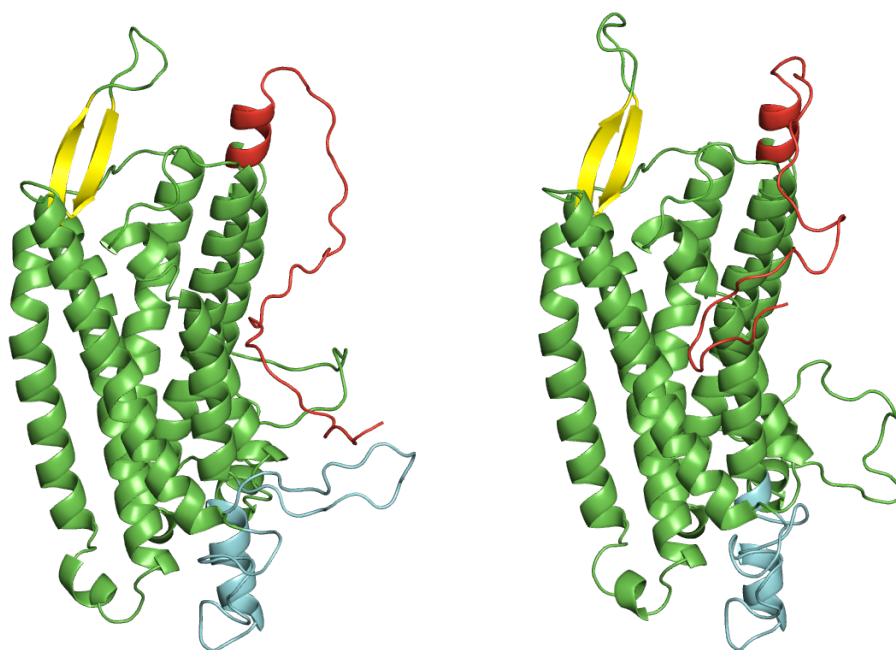
Figure 3.5: Top DOPE scored 3D model of GPR120L (left) and GPR120S (right) generated using the 4S0V-4N6H-4DJH-5GLI-4JZ8 template combination; the figure illustrates the disordered conformation of the N-terminal region (shown in red) and ICL3 (shown in cyan). Visualised in PyMol (DeLano 2018).

Previous literature suggested that the N-terminal of GPCRs plays an important role in peptide recognition and binding of bulkier ligands (ligands above 600 Dalton) (Rutz, Klein and Schülein, 2015). Class C GPCRs are distinguished by their large N-terminals, which contain a ligand-binding domain, such ligand binding domain are not common in class A

GPCRs (Rutz, Klein and Schülein, 2015) (explained in section 1.3.1). As the scope of the research was to screen only small drug like compounds using the 3D models of GPR120, the N-terminal (1-35 residues) of GPR120S and GPR120L, and C-terminal (339 -361 residues) of GPR120S were cleaved off to refine the modelled structure (Figure 3.6). If required, the N-terminal and C-terminal can be threaded in the future to the refined seven TM domains of GPR120.

For the next iteration of HM generation, the protein sequence of GPR120S (36-339 residues) was utilized (Figure 3.6) and GPR120L (36-355 residues), which has an extra 16 residues compared to the short isoform with a total of 33 residues in ICL3.

The DOPE scores of the predicted models from all the 25 template combinations (Table 2.1) were compared and models generated from the **4S0V-4N6H** template combination were found to have the best DOPE scores. The model with the lowest DOPE score signifies the most stable protein model. MODELLER generated the MSA (shown in Figure 3.6) of target sequences (4S0V and 4N6H) extracted from the PDB (3D coordinate) files of experimental structures and the GPR120S sequence to build the 3D models of GPR120S sequence. The cleavage of N and C terminals from GPR120S sequence improved the percent identity of 4S0V from 24 to 28 % and 4N6H from 21 to 24 % against the query sequence (Table 3.4). The models generated by 4S0V-4N6H templates were evaluated to confirm the conformations and stereochemistry of TM and loop regions.

Table 3.4: Clustal Omega Percent Identity Matrix of templates and target sequence used for the generating the homology model of GPR120S.

1:	GPR120S_trimmed (35-339)	100.00	24.21	28.16
2:	4N6H (Delta-like opioid receptor)	24.21	100.00	30.00
3:	4S0V (Orexin type 2 receptor)	28.16	30.00	100.00

```
                 ▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟     ▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟
GPR120S    ----------LVLAAVETTVLVLIFAVSLLGNVCALVLVAR-RRRRGATACLVLNLFCAD    84
4N6H       GSPGARSASSLALAIAITALYSAVCAVGLLGNVLVMFGIVRYTKMKTATNIYIFNLALAD    60
4S0V       ---------PKEYEWVLIAGYIIVFVVALIGNVLVCVAVWKNHHMRTVTNYFIVNLSLAD    51
                     .    :     :  .*.*:*** . . : :   :* .:.** . **
           ▟▟▟▟▟ ▟▟▟▟ ▟▟▟▟▟▟▟▟▟▟     ▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟  ▟▟▟
GPR120S    LLFIS-AIPLVLAVRWTEAWLLGPVACHLLFYVMTLSGSVTILTLAAVSLERMVCIVHLQ    143
4N6H       ALATS-TLPFQSAKYLMETWPFGELLCKAVLSIDYYNMFTSIFTLTMMSVDRYIAVCHPV    119
4S0V       VLVTITCLPATLVVDITETWFFGQSLCKVIPYLQTVSVSVSVLTLSCIALDRWYAICHPS    111
              *     :*    .    *:* :* ☆: : :  .::**: ::::* .:  *
           ▟▟▟▟▟▟▟▟ ▟▟▟▟▟     ▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟      ▶▶▶▶▶▶▶      ▶▶▶▶  ▟▟▟
GPR120S    RGVRGPGRRARAVLLALIWGYSAVAALPLCVFFRVVPQRLPGADQEISICTLIWPTIPG-    202
4N6H       KALDFRTPAKAKLINICIWVLASGVGVPIMVMA-VTR---PR--DGAVVCMLQFPSPSWY    173
4S0V       TAKRARNSIV------IIWIVSCIIMIPQAIVMECST---VFKTTLFTVCDERWG--GEI    160
              .            **  :.    :*   :.         ☆   :
           ▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟     ◣    ◣   ▟▟▟▟▟▟▟▟▟▟▟▟▟
GPR120S    -EISWDVSFVTLNFLVPGLVIVISYSKILQITKASRKRLTVSLAYSESHQIRVSQQDFRL    261
4N6H       WDTVTKICVFLFAFVVPILIITVCYGLMLLRLRSVRLL------SGSKEKDRSL---RRI    224
4S0V       YPKMYHICFFLVTYMAPLCLMVLAYLQIF------RKL------WCRQKQIRAR---RKT    205
             .:...  . ::.*   ::.:.*  ::        *      ..:  *       :
           ▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟     ▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟▟
GPR120S    FRTLFLLMVSFFIMWSPIIITILLI-LIQNFKQDLVIWPSLFFWVVAFTFAWSALNPILY    320
4N6H       TRMVLVVVGAFVVCWAPIHIFVIVWTLVDIDRRDP-LVVAALHLCIALGYAWSSLNPVLY    283
4S0V       ARMLMVVLLVFAICYLPISILNVLKRVFGMFAHDRETVYAWFTFSHWLVYAWSAANPIIY    265
             *  ::::::  *  : :  ** *   ::  :.    :*    :  :   : :***: **::*
           ▟▟▟▟ ▟▟▟▟▟▟▟▟▟▟▟▟
GPR120S    NMTL--CRNEWKKIFCCFWI--                                         339
4N6H       AFLDENFKRCFRQL--CRKPCG                                         303
4S0V       NFLSGKFREEFKAAFSCGIDCS                                         287
              :      :. ::      *
```

Figure 3.6: Sequence alignment of GPR120S and templates (4N6H and 4S0V) used for homology model generation; Conserved residues shown in square boxes. Multiple sequence alignment generated using Clustal Omega. '*' indicates fully conserved residue; ':' indicates conservation between groups of strongly similar properties; '.' indicates conservation between groups of weakly similar properties. The disulfide bond forming Cysteines are indicated by ☆. The conserved motifs of the Rhodopsin-like family are bolded. Red indicates helical / transmembrane regions; Green - β-strands; yellow – loop regions.

3.3 Side-chain optimization

From the 4S0V-4N6H model series, the top 10 models were selected (Table 3.5) and were further screened by their stereochemical properties by assessing their Ramachandran ϕ-Ψ (phi-psi) plots using MolProbity webserver (Chen et al., 2009) and quality of the predicted structure by ProSA webserver (Wiederstein and Sippl, 2007).

Table 3.5: DOPE score of top 10 models of GPR120S derived from templates 4S0V-4N6H

	3D Model	DOPE score		3D Model	Dope Score
1	GPR120S_Model_30	-41156.738	6	GPR120S_Model_31	-40959.097
2	GPR120S_Model_63	-41070.296	7	GPR120S_Model_24	-40953.445
3	GPR120S_Model_97	-41042.429	8	GPR120S_Model_77	-40917.972
4	GPR120S_Model_18	-41029.484	9	GPR120S_Model_33	-40915.375
5	GPR120S_Model_45	-40976.371	10	GPR120S_Model_32	-40899.261

The comparative analysis of the stereochemical parameters of the top ten DOPE scoring models (Table 3.5) of GPR120S by the Molprobity webserver showed high plausibility of the generated models with respect to the protein stereochemistry. The Ramachandran ϕ-Ψ evaluation showed that there were no residues (0%) in outlier region and 98-99% residues were in favoured region. But the models had some poor rotamers / sidechain rotations above the permitted range of 0.3%, which were removed by sidechain optimization using SCWRL4 (Krivov, Shapovalov and Dunbrack, 2009). GPR120S_model30 with the best DOPE score was selected for further optimisation. From this point forward, GPR120S_model30 is referred to as GPR120S.

SCWRL4 contains a backbone-dependent rotamer library based on kernel density estimates, and it search for the average sidechain rotamer over samples of conformations about the position of residue in the rotamer library. It consists of a fast-anisotropic hydrogen bonding functions and using the short-range, soft van der Waals atom-atom interaction potential, it detects sidechain-sidechain and sidechain-backbone collisions. Finally, it generates an optimized model from all parameters by determining the interaction graph within the crystal environment using symmetry operators of the crystallographic space group (Krivov, Shapovalov and Dunbrack, 2009; Ryu and Kim, 2012).

SCWRL4 read in the 3D coordinate file of the predicted model backbone from which sidechains were removed. The algorithm searched for the sidechain rotamers over the rotamer conformational library to generate the complete model with sidechains. Sidechains of all the selected models (Table 3.5) were optimised using SCWRL4 and the validation parameters of initial and sidechain optimised models were compared to evaluate and select the best predicted model (Table 3.6).

The sidechain optimization by SCWRL4 increased the stereochemical quality of the generated 3D model (Table 3.6). The number of poor sidechain rotamers decreased from five to one, bringing the model closer to the permitted range (~ 0.3%) of allowed poor rotamers from ~2% in 3D models.

Table 3.6: Stereochemical parameters of top side chain optimised GPR120S model generated by HM (Molprobity webserver).

Model (303 residues)	GPR120S		Sidechain optimised GPR120S	
Poor rotamers (Goal: <0.3%)	5	1.88%	1	0.38%
Favoured rotamers (Goal: >98%)	244	91.73%	264	99.25%
Ramachandran outliers (Goal: <0.05%)	0	0.00%	0	0.00%
Ramachandran favoured (Goal: >98%)	297	98.67%	297	98.67%
Bad bonds (Goal: 0%)	0 / 2473	0.00%	7/ 2473	0.28%
Bad angles (Goal: <0.1%)	33 / 3379	0.98%	6/ 3379	0.18%

3.4 Validation of GPR120S homology model

Ramachandran plot analysis (MolProbity - Chen et al., 2009) of the sidechain optimised model (Table3.6) reveals that ~98% (297/301 residues) of residues are in favourable regions while only one glycine residue is in the disallowed region (Figure 3.7 a). As glycine do not have side chains and is the least sterically hindered amino acid and therefore often adopts conformations that are forbidden to other residues. Glycine can adopt ɸ-Ψ (phi-psi) angles in all four quadrants of the Ramachandran plot covering a large area on the plot and appear in the unfavourable or disallowed regions (Ho and Brasseur, 2005).
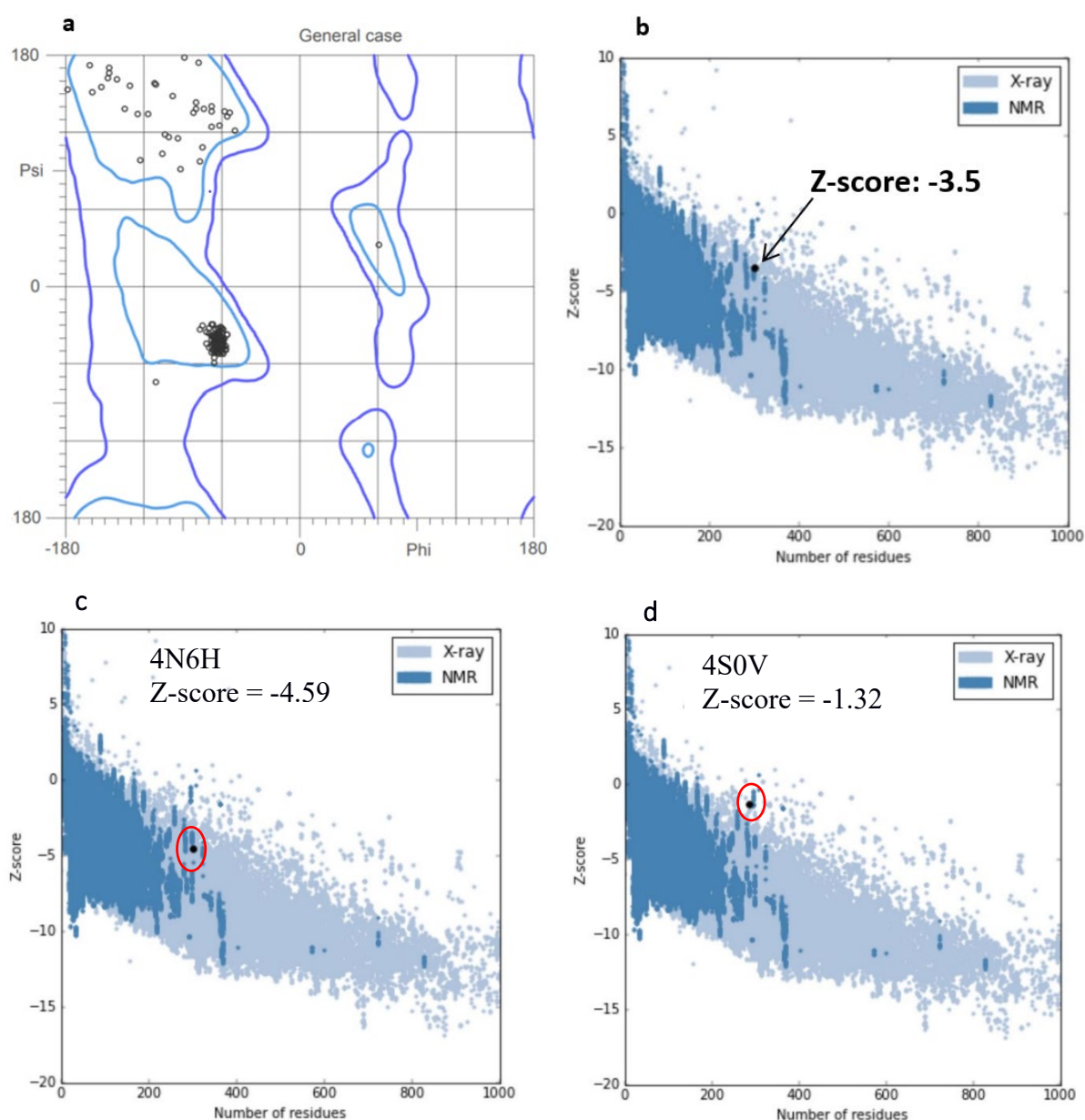
Figure 3.7: Ramachandran plot (a) and Z-score (b-d) of the top GPR120S homology model and homology templates 4N6H and 4S0V. The most favoured regions in Ramachandran plot are enclosed by blue boundaries, additional allowed, generously allowed regions are enclosed in purple boundaries and all areas outside of purple lines are disallowed regions / outliers. No residues have a bad conformation / are outlier, with only two residues in allowed regions.

Protein structure analysis (ProSA) gives the quality of the overall model in terms of the deviation of the experimental result from the most probable result in the form of a graph (Wiederstein and Sippl 2007). The number of residues is plotted versus Z-score for known structures determined by X-Ray crystallography and NMR. Z-score is a statistical measure that quantifies the distance in standard deviations of a data point from the mean of a

data set. The Z-score of the selected SCWRL optimised model (-3.5) indicated that overall quality of the protein is comparative to the Z-score of the templates used (4N6H: -4.59; 4S0V: -1.32) and lies in the favourable range (Figure 3.7 b-d). In general, positive values correspond to poor quality of the input structure. The black dot represents the predicted protein, and the analysis shows that the modelled protein is similar to proteins which have been characterized using NMR/X-ray. These values indicate that the structural average for the generated quality control values is within a normal range. Therefore, the final refined model passes the formal evaluation and need to be considered as suitable for further studies.

The ERRAT analysis of 3D model of GPR120S (Figure 3.8) resulted in an overall quality factor of 92.6 % where high-resolution structures produce values around 95% or higher while lower resolutions (2.5 to 3 Å) produce average overall quality factor of ~91%. The highest error values were recorded in the intracellular loop (ICL) regions – ICL2 (142-151) and ICL3 (234-251) which are not involved in the protein-ligand interactions. ERRAT quality factor represented clear agreement that the generated 3D model of GPR120S is of comparative quality with X-ray crystal structures and can be used for further screening studies.
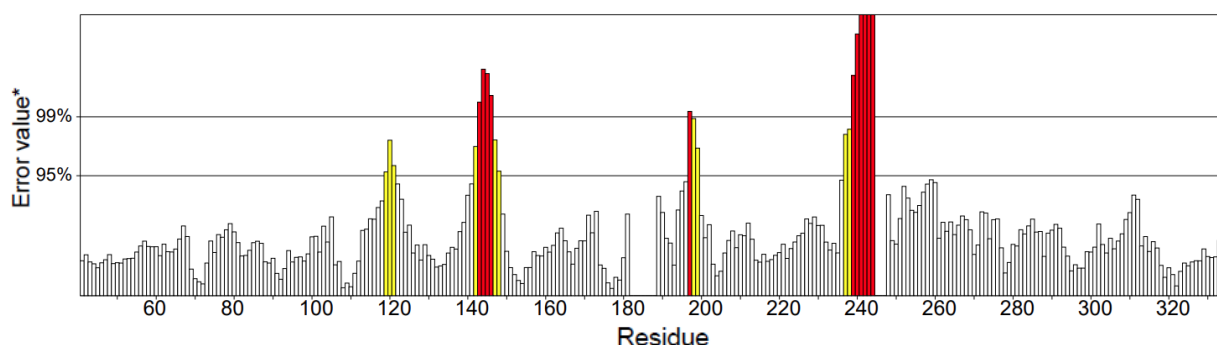
Figure 3.8: ERRAT statistical representation of overall quality of sidechain optimised GPR120S model in comparison to refined X-ray crystal structures (Colovos and Yeates, 1993). The highest error values were recorded in the intracellular loop (ICL) regions – ICL2 (142-151) and ICL3 (234-251).

Finally, the quantitative analysis of structural features of the generated model was performed. The X-ray crystallised 3D structures of templates (4N6H and 4S0V) when superimposed to the GPR120S model, the backbone (Cα, C and N atoms) root mean square deviation (RMSD) of selected model against 4N6H and 4S0V was found to be 1.386 Å and 0.652 Å, respectively. RMSD measures the difference between C-alpha atom positions between two proteins. The sidechain optimised homology model GPR120S was selected for

further EM and investigational studies as it has the lowest mean RMSD (1.019 Å) and highest structural similarity when compared to other models as well as it was most optimised model on various parameters such as DOPE scores, Z-score, number of bad angles and ϕ-Ψ evaluation.

Initially, the GPR120S model was generated and validated and later GPR120L was modelled from the validated GPR120S 3D model to keep the 3D structure of GPR120L ordered as the ICL3 (the long isoform – with 16 extra amino acids) of GPR120L has very low similarity with other GPCRs (Appendix Vd). The range of transmembrane domain for GPR120L was maintained as that for GPR120S with topology as shown in Table 3.3. Although a model for GPR120L was also generated the focus was on GPR120S for compound screening as both the short and long isoforms have identical orthosteric binding pocket sequentially as well as structurally (Figure 1.11). As the predicted 3D models of GPR120S and GPR120L were generated to screen hit compounds by VS of chemical databases, only GPR120S was used for VS of chemical databases and further studies.

3.5 Energy minimization of GPR120S model in a phospholipid bilayer

The aim of the initial energy minimization was to achieve the net force on each atom in the GPR120S model closest to zero. EM explores different protein motions and conformations to find optimised free energy states of the protein obtained during evolution of protein structure towards the global minimum of free energy. Although attaining global minimum of free energy during the short energy minimisation steps is almost impossible, the optimised protein structures usually attain local minimum of free energy. The optimised state proteins have low free energy due to the least number of steric clashes and abnormal geometry (Abraham et al., 2015). GROMACS v5.1.4 uses semi empirical force fields to run the molecular energetics of the protein model. Interatomic interactions such as the bond lengths, angles, and torsions as well as the van der Waals and electrostatic interactions, of the protein model are approximated to compute the energy of different conformations to find the most accurate ground state conformation possible.

Being a transmembrane protein, for EM the GPR120S model must be embedded in the phospholipid bilayer solvated by water under periodic boundary conditions to provide the optimum environment. This environment mimics the natural lipid bilayer and conserves the functional properties of the receptor. For example, in silico MD studies have shown that gradual inactivation of the β_2 adrenergic receptor occurred in the neutral lipid membrane of *1-*

palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylcholine (POPC) molecules (Dror et al., 2011) as well as activation of rhodopsin receptor induced changes in membrane (Salas-Estrada, Leioatts, Romo and Grossfield, 2018). Prior to EM the GPR120S was embedded in an explicit bilayer membrane consisting of POPC molecules (Figure 3.9 a, b), solvated with water molecules and charge neutralized by addition of sodium and chloride ions following the set protocol (as explained in section 2.1.4). EM was carried out for an apo receptor model to ensure that the protein system has no steric clashes or abnormal geometry.

The coordinate file of a hydrated, equilibrated 128 lipid POPC (Figure 3.10) bilayer along with lipid parameters for the GROMOS 54a7 force field were obtained from the ATB repository (Koziara et al., 2014). The coordinate file was then resized using Inflategro2 (Schmidt and Kandt, 2012) - modified to comply with GROMACS v5.1.4, to produce a fully hydrated, 512 POPC lipid bilayer (Figure 3.10c). First, InflateGro2 incorporates LAMBADA (Schmidt and Kandt, 2012) which determine the protein's LSmin (hydrophilicity profile) configuration using a recursive optimization to test different protein orientations thus aligning membrane and protein. Then it automatically embeds the membrane protein into lipid bilayer patches and removes the clashing lipids. Inflategro2 then stretches the model in the plane of the lipid bilayer and energy minimizes the system again on contraction allowing optimization of lipid/protein interface over 1000 EM steps for 20 iterative cycles, using a scaling factor of 0.5 (Kandt, Ash and Tieleman, 2007).

Embedding of the GPR120S receptor resulted in the removal of 33 lipid molecules. The final system contained a 303 amino acid long protein, 477 POPC molecules, 25,529 water molecules and 14 chloride ions (shown in Figure 3.9d). Each system was charge neutralised by adding the required number of sodium or chloride ions. The setup of such large systems (104,505 atoms) invariably leads to close contacts in the initial positions of the atoms. The close contacts can result in extremely high repulsive forces and cause large displacements by accelerating the atoms and distort the overall spatial conformation of atoms. EM can relieve the close contacts and stabilize the system (Kandt, Ash and Tieleman, 2007).

The analysis of 50,000 EM steps showed that after ~14,000 EM steps, the apo protein system reached a local minimum (U = -2,025,501 kJ/mol/nm) nearest to the starting system and attained a probable relaxed state (Figure 3.10a). As the EM converged the system to acceptable potential energy levels before all iterations were completed. The model with the

lowest energy state from the energy minimized ensemble was selected and subjected to model validation through enrichment studies.

Chemical structure of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylcholine (POPC) molecule
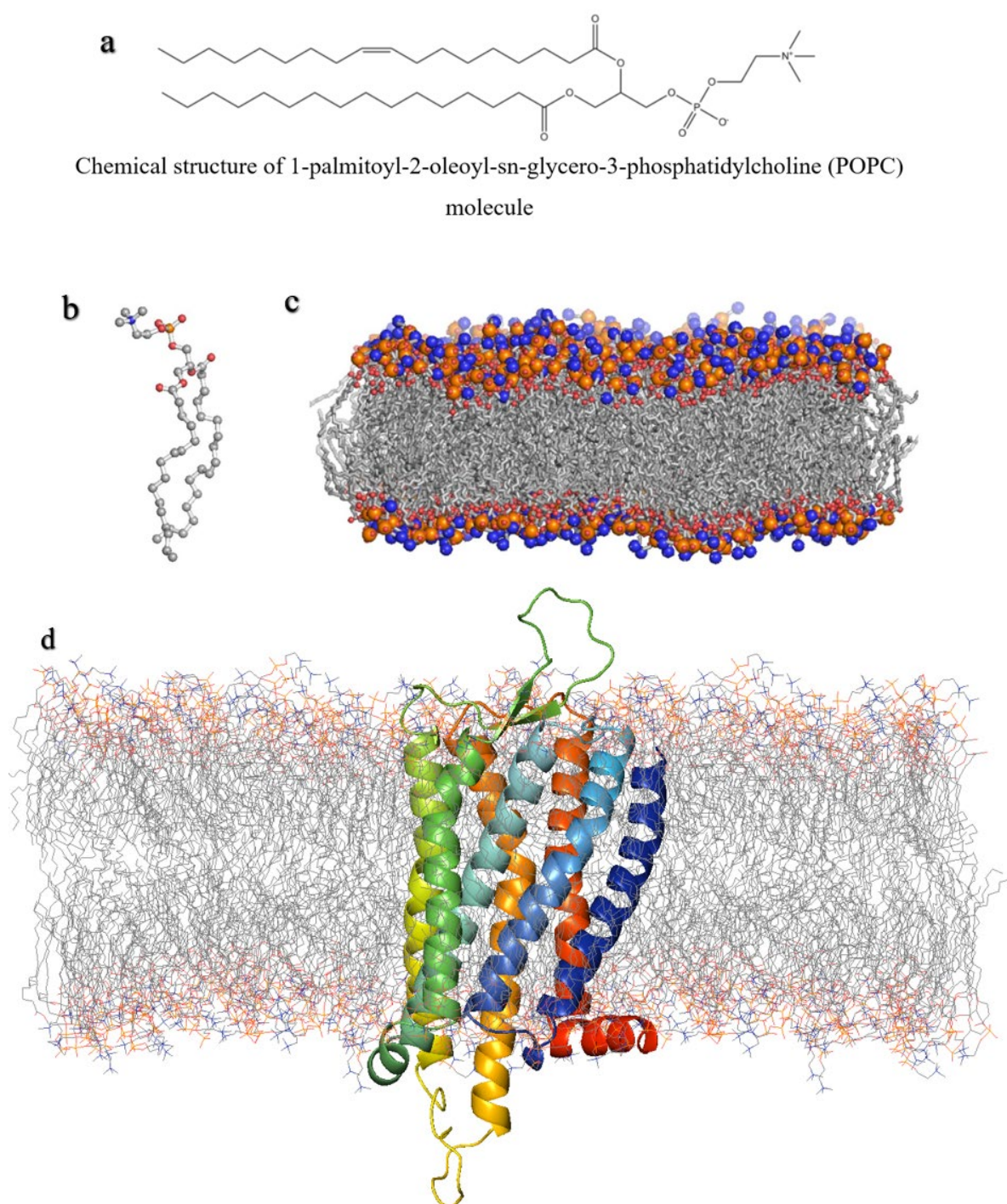
Figure 3.9: a) Chemical structure of POPC molecule; b) single POPC cartoon view; c) full minimized bilayer of 512 POPC molecules; d) GPR120S receptor embedded in POPC bilayer; water molecules removed for clear visualization. Images were rendered and visualized in PyMol (DeLano, 2018)

The residue root mean square fluctuation (RMSF) is a measure of the flexibility of a residue. It is typically calculated for the Cα atom of each residue. RMSF of the apo protein system obtained from the EM run is shown below in Figure 3.10b, representing the measure of conformational variance analysis. It highlights that the loop portions especially ECL2 (177-203) and ICL3 (236-252) of the protein structure, fluctuated from their mean structure the most and the TM helices showed the least fluctuation (below 0.2 nm). The larger loop regions have more disordered structures and are known to fluctuate the most (Baştuğ, and Kuyucak, 2012). The minimised model of GPR120S was used for further studies.
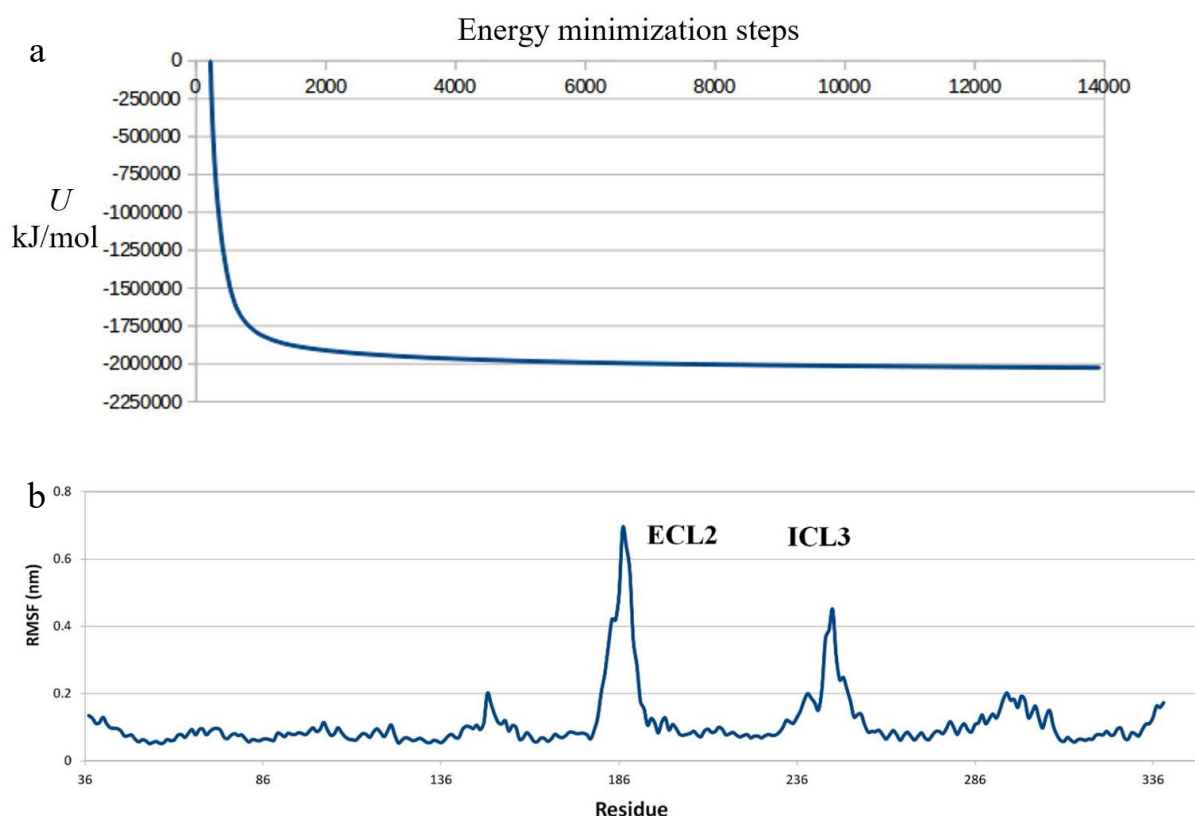
Figure 3.10: a) Potential energy (U) curve of the Apo protein energy minimization. The EM plot demonstrates that the system steadily converged until ~ 4,500 EM steps and afterwards attained steady state potential; b) Root mean square fluctuation (RMSF) of Cα atoms of each residue during the EM. The ECL2 and ICL3 regions showed the maximum fluctuation.

3.6 Molecular docking and enrichment studies for affinity prediction

Previous molecular docking and site-specific mutation studies of GPR120 have revealed that a single arginine residue in TM2 (Arg99) has a critical interaction between the receptor and the –COOH (carboxylate) of its ligands (Watson, Brown and Holliday, 2012; Hudson et al., 2013; Hudson et al., 2014). Six other residues were defined essential for TUG891 binding

and interaction with GPR120 were: Trp104 (ECL1), Phe115 (TM3), Trp207 (TM5), Phe211 (TM5), Trp277 (TM6) and Phe304 (TM7) (Hudson et al. 2014) (shown in Figure 3.11). These seven residues were selected as the main criterion for defining the orthosteric binding pocket as well as protein-ligand interactions for binding pose prediction of GPR120 ligands. To date, while some GPR120 agonists have been developed no orthosteric antagonist of GPR120 is available. AH7614 (Table 1.8) was postulated as a GPR120S antagonist in 2014 (Sparks et al. 2014) but has recently been reported to act as a negative allosteric modulator (NAMD) of GPR120 (Watterson et al. 2017).
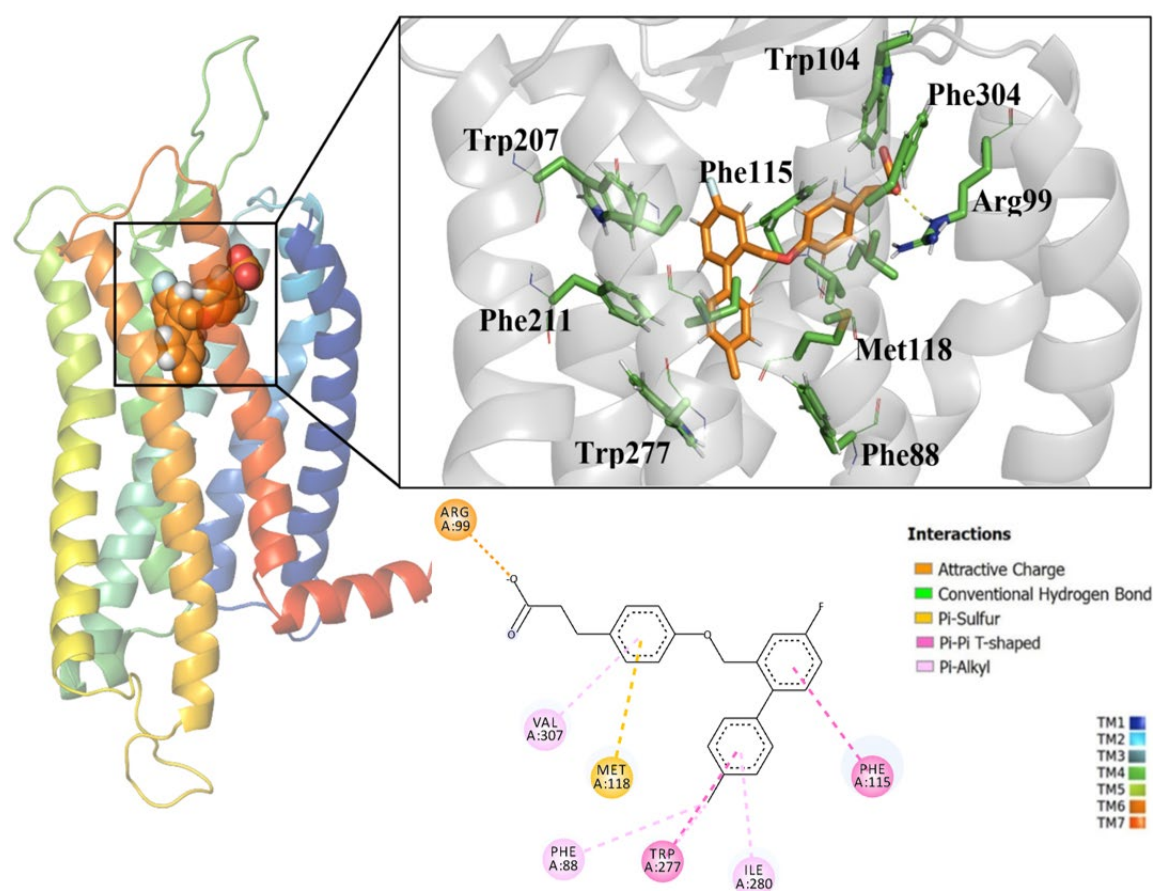
Figure 3.11: Selected docked pose of TUG891 (binding score -9.875) illustrating hydrogen bond interactions with Arg99 of GPR120S and a 2D interaction map of TUG891 in the orthosteric pocket binding pocket. The 3D images were visualized and rendered in PyMol v2.1.0. The 2D interaction maps were generated in BIOVIA DS Client visualizer v19.1 2019.

The *in-silico* investigation of receptor activation commenced with semi-flexible docking (rigid protein vs flexible ligand) of a known agonist to the receptor model in the inactive state (GPR120S homology model). The best docked pose for TUG891 yielded a binding score -

9.875 kcal/mol (free energy of binding calculated by AutoDock SMINA). The carboxylate of TUG891 forms a salt bridge with Arg99 and a strong T-type (perpendicular) pi-stacking interaction between Phe115 and the cyclic aromatic core structure of TUG891 stabilize the ligand into the pocket formed between TM3, TM6 and TM7 (Fig 3.11). A molecular docking study conducted by Hudson and team (Hudson et al., 2014) used the HM of human GPR120S (based on a nanobody-stabilized active state β2 adrenoreceptor template; PDB id: 3P0G) to dock TUG891. An extensive overlap is observed between the binding pocket residues interacting with the agonist in both studies. Other equitable hydrophobic interactions were also observed to be stabilizing the docked TUG891 in the orthosteric binding pocket (Table 3.7).

Table 3.7: List of residues forming contacts with TUG891 docked to the orthosteric binding pocket of GPR120S; obtained from PLIP (Salentin et al. 2015; https://plip-tool.biotec.tu-dresden.de/plip-web/plip/index)

Salt-bridge / Hydrogen bond interactions		
Residue	Distance Å	Ligand Group
Arg99	4.32	Carboxylate
π-Stacking interactions		
Residue	Distance Å	Angle (Degree)	Type
Phe115	5.04	72.46	T
Hydrophobic interactions		
Residue	Distance (Å)	
Phe88	3.84	
Val95	3.99	
Trp104	3.84	
Leu114	3.84	
Phe115	3.59	
Phe115	3.53	
Thr199	3.67	
Leu173	3.99	
Trp207	3.19	
Phe211	3.77	
Ile280	3.47	
Ile284	3.61	
Phe303	3.65	
Phe303	3.56	
Val307	3.93	
Val307	3.46	

Ideally, for the purpose of binding mode prediction, a homology model is validated via the reproduction of the experimentally determined binding mode of another ligand (Beuming et al., 2015; Dhasmana et al., 2019). In the absence of a receptor crystal structure, the energy minimised GPR120S model was validated using enrichment studies that demonstrated the ability of the modelled GPR120S to predict the binding affinities of known ligands and distinguish between known ligands and decoys (Kim and Skolnick, 2008; Park et al., 2009). The selected model was assessed for the ability to distinguish between known binders of GPR120 and a set of drug-like decoys (obtained from DUD.E; see Appendix Table IIb) in a virtual screening exercise, which was evaluated by plotting their corresponding ROC curves and calculating the AUC (Beuming et al., 2015; Lenselink et al., 2014; Mysinger, Carchia, Irwin and Shoichet, 2012). The docked pose of TUG891 (Figure 3.11) was used as a reference pose for docking the active and decoy sets to perform enrichment studies.

The best **AUC** obtained for the ROC curve for the GPR120S receptor model was 0.89 (Figure 3.12), which indicates a significant ability to distinguish between the active compounds and the decoys. The AUC obtained for the GPR120S receptor (< 0.9), was expected considering the binding affinity prediction and ranking power of the docking algorithms in the absence of high-resolution protein-ligand crystal structure. An analysis of the top-ranking decoys which contributed significantly to the false positive rate at the initial part of the test revealed them to be mostly compounds of high logP (above 5) and molecular weight (more than 450 Daltons). The Enrichment factor (EF) of the top 1% of the data set is 9.375 (top 20% is 4.23) indicating that the modelled structure bound with the active ligands in the top 1% of the ranked database. The EF of 9.3 indicates that ~9 times more active compounds in the top 1% of the screening would appear. Based on the ROC curves obtained, the enrichment studies provided additional support to the validation of selected model as a reasonable representation of the GPR120S receptor. Using the same set of actives and decoys, further enrichment studies were performed using SMINA docking algorithm and four different scoring algorithms to re-score the top docked pose of SMINA to benchmark different scoring functions – SMINA, VINA, NNScore, DLScore and CScore (section 2.1.3.2; Appendix Table IIc). SMINA scoring function was found to be the best scoring function. The validated GPR120S model with an EF (1%) of 9.375 was used for further MD studies.
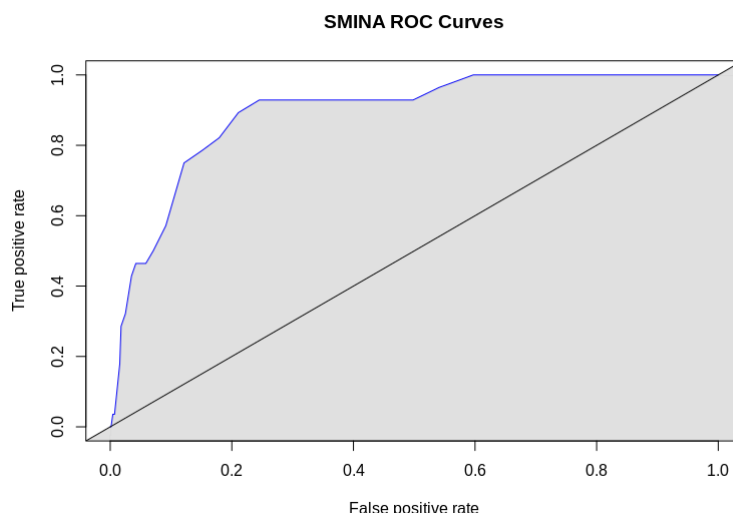
Figure 3.12: Area under ROC plot of decoys and actives docked to GPR120S shown in linear scale. The line of identity (black, diagonal) represents the expected line in completely random selection, which would result in an AUC of 0.5. Performed in R using in-house R-script.

3.7 Comparative analysis of GPR120 model predicted by DeepMind - artificial intelligence.

DeepMind, a knowledge-based artificial intelligence (AI) operation, developed a 3D model predictor of protein folding – AlphaFold. It uses a distribution over pairwise distances between residues corresponding to a statistical potential function to predict the protein folding (Wei, 2019; Jumper et al., 2021). The prediction algorithm improves the accuracy by training the AI nodes based on the evolutionary, physical, and geometric constraints of elucidated protein structures. The 3D model of human GPR120 short isoform (Q5NUL3 - 361 amino acids) predicted by AlphaFold was published on AlphFold DB in July 2021 (https://alphafold.ebi.ac.uk/entry/Q5NUL3 Jumper et al., 2021). The AI predicted model of GPR120S (Figure 3.14) was compared to the validated GPR120S model generated for this study in January 2017. The N-terminal (1-36) and C-terminal (340-361) domains, which were removed from the GPR120S model to improve the structural quality (section 3.2; Figure 3.5) due to their low structural similarity to templates were reported as the unstructured regions with very low prediction scores (per-residue confidence score < 50) in the AI model as well (Figure 3.13). With the exception of TM5, all the TM domains and Helix 8 of AI model have the same sequence range (Table 3.3) as used for the GPR120S homology model (Figure 3.13). When the GPR120S homology model was superimposed to the AI predicted model, the backbone (Cα, C and N atoms) RMSD was found to be 1.351 Å. With such low RMSD

values, the accuracy of the GPR120S homology model generated by our study is comparable to the AI predicted model of GPR120 short isoform.
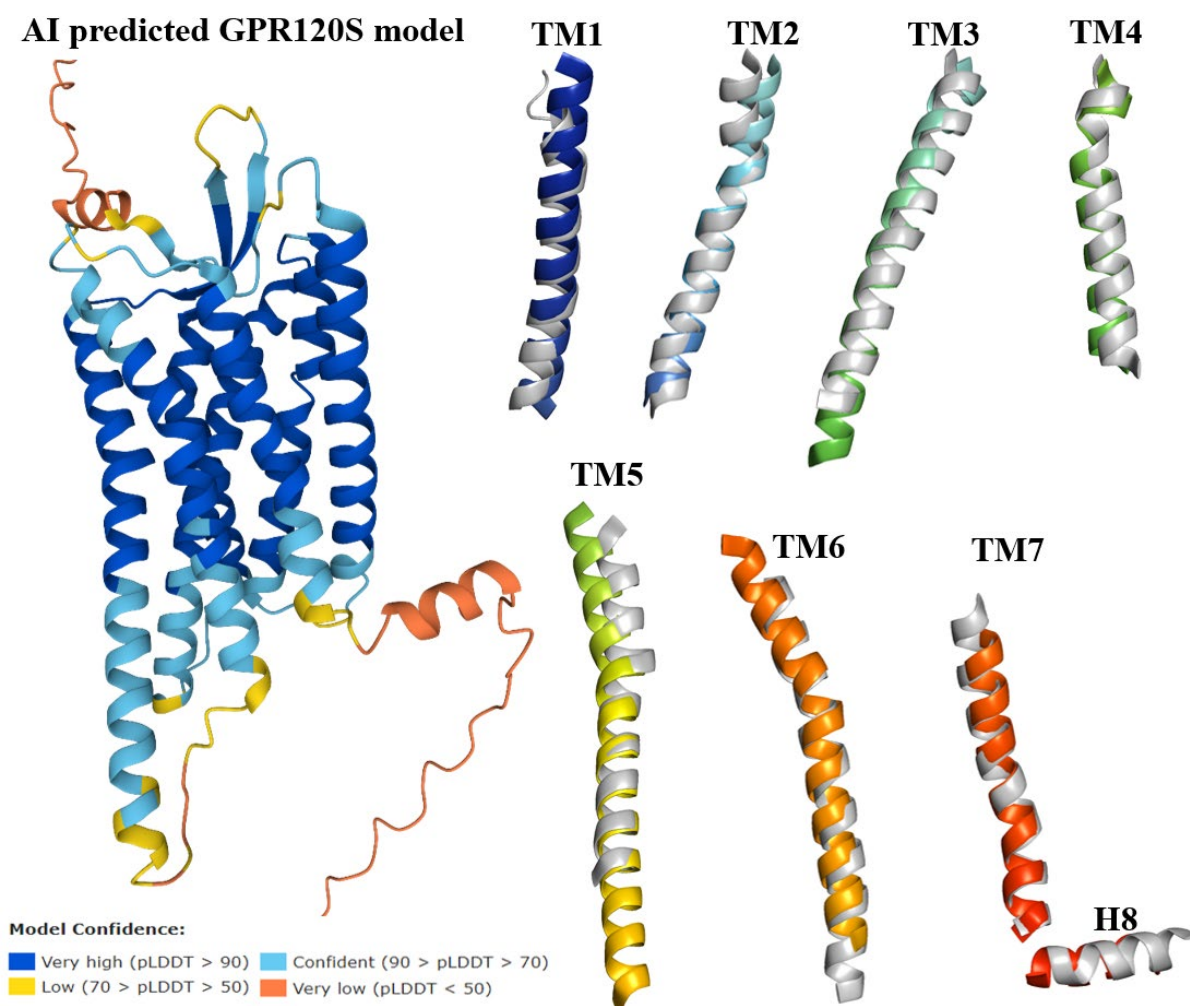
Figure 3.13: 3D structure of human GPR120S predicted by AlphaFold artificial intelligence (AI) program (Jumper et al., 2021). GPR120S seven TM domains and Helix - H8 of AlphaFold predicted model (spectrum colour) superimposed on validated GPR120S homology model (grey). AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100.

3.8 Discussion

The initial focus was to build a 3D structural model of GPR120S through *in silico* methodologies. Homology models as well as molecular docking are useful predictions and not the ultimate solutions to determine the structures of macromolecules. Computational approaches have gone through major developments in recent years and the available software

packages are powerful and reliable to predict comparative naturally resembling models (Malathi and Ramaiah, 2018). However, the quality of the predictions completely relies on the prior knowledgebase available for target proteins. The models produced by combination of a strong knowledgebase and reliable computational methods are of use in providing guidance to experimental biologists to minimize unsuccessful experiment setup designs (Malathi and Ramaiah 2018; Jumper et al., 2021).

Homology modelling of GPCRs has been employed effectively in various structure-based drug discovery studies of GPCRs (Bissantz, Bernard, Hibert and Rognan, 2002; Jaiteh, Rodríguez-Espigares, Selent and Carlsson, 2020). Homology modelling based on single template structure for low sequence identity target proteins often generate less accurate protein models (Larsson, Wallner, Lindahl and Elofsson, 2008) with GPCRs having sequence identity in the range of 20-30% are best modelled using multiple template homology modelling (Baker, 2001; Bender, Marlow and Meiler, 2020).

The homology model of GPR120S was developed, based on the high-resolution crystal structures of human delta-like opioid and human orexin type-2 receptors, which are closer to GPR120S in phylogenetic evolution of the available templates (performed in January 2017). To the best of our knowledge, six different studies of GPR120 generated homology models have been published which used single template-based modelling such as photoactivated bovine rhodopsin (Sun et al., 2010; Hara et al., 2011; Takeuchi et al., 2013), active state β2 adrenoreceptor (Hudson, Shimpukade, Milligan and Ulven, 2014), inactive state delta-like opioid receptor (Chinthakunta et al., 2018) and activated turkey β1 adrenoreceptor (Zhang, Sun, Wen and Yuan, 2019).

The recently elucidated crystal structure of orexin OX2 receptor (in 2015) was found to be a suitable template for GPR120S as it covers around 87% of the GPR120S (first 200 residues) with 26% identity. It was used in combination with the delta-like opioid receptor which covers around 84% of GPR120S (full sequence) with 26% identity. The MSA of orexin OX2 and delta-like opioid receptor over GPR120S improved the sequence alignment as well as the quality of the 3D model generated. The current study used an antagonist-bound human delta-like opioid receptor and orexin 2 receptors as templates for GPR120 model prediction available in January 2017. However, the latest (December 2021) protein BLAST search (Altschul et al., 1990) conducted for GPR120S templates against the Protein Data Bank (Deshpande et al., 2005) enlisted the same Orexin 2 receptor and delta-like opioid receptors

as the best templates for homology model predictions covering 87% and 84% of the full query sequence, respectively. While the sequence coverage improved to 93% and 96% respectively for GPR120S sequence N and C terminals removed.

One limitation in the GPR120S model generation that can be argued is the ECL and ICL domains (loop) modeling, especially ECL2 which plays a significant role in ligand binding in Class A GPCRs. Side-chain optimization using SCWRL4 was performed to reduce this bias to some extent and improve ligand binding predictions. Side-chain optimization may also improve the stereochemistry of the model as during model generation limited rotamers for each sidechain at fixed C-alpha of the protein backbone are explored. As there can be exhaustive rotamer conformations for each residue and depending on the protein sequence length, these enumerations can become astronomical. The side-chain optimization algorithm, SCWRL4, was used to explore the bigger library of rotamers after homology modelling to find combination of rotamers representing the lowest-energy conformations. The molecular docking study against β-catenin by Low et al (Low et al., 2021) reported the application of SCRWL-based sidechain sampling in iterative fashion to optimise homology models with improved ligand enrichment performance. Considering the sequence identity of <30% and the flexibility of loop regions, an initial 50,000 step EM and further molecular dynamics simulations (Chapter 4) were performed in apo and ligand-bound systems to attain stable protein ensemble conformations.

The prediction of ligand binding by molecular docking experiments is sensitive to side-chain conformations of residues forming the binding pocket. The template selection criteria for the GPR120 homology model were based on the antagonist-bound crystal structures. The potential of binding pocket sidechain conformations of homology models biased towards the chemical scaffolds of the templates used particularly for the ligand-bound templates could result in false positives and pseudo binding predictions.

Over the course of combined energy minimization and semi-flexible molecular docking studies, the model has been investigated. The reported hydrogen bonding pattern, lipophilic interactions, and binding energy of the selective ligand – TUG891 has been examined in the structural refinement. The amino acid residues involved in the putative binding site – Arg99 (TM2), Trp104 (ECL1), Phe115 (TM3), Trp207, Phe211 (TM5), Trp277 (TM6) and Phe304 (TM7) (Hudson et al. 2014) are facing towards the pocket in the validated model and have shown reliable interactions with the GPR120S selective agonist TUG891. A molecular

docking study conducted by Hudson (Hudson et al. 2014) used the HM of human GPR120S (based on β_2 adrenoreceptor template; PDB id: 3P0G) to dock TUG891. The predicted binding poses of TUG891 (Hudson et al. 2014) illustrates the residues in proximity (distance unknown) of the agonist. There was an extensive overlap between the binding pocket residues interacting with the agonist in both studies (Table 3.6; Figure3.11).

Enrichment studies are used to measure the selectivity of the receptor towards the active known binders. The enrichment factor demonstrates the ability of the generated model to enrich the known / active ligands of the receptor in the top ranks of the screening from among a wide database of molecules. The screening databases for enrichments are prepared to contain actives / known ligands and drug-like decoys which resemble the active ligands in geometry but are chemically distinct from actives. The enrichment factor is evaluated by plotting the corresponding AUC plots (Lenselink et al. 2014; Mysinger, Carchia, Irwin and Shoichet, 2012).

A ROC curve is the graphical plot that illustrates the diagnostic ability of the evaluation test. The true positive rate/known binders (Sensitivity) are plotted in function of the false positive rate/drug-like decoys (Specificity) for different cut-off points. Each point on the ROC plot represents a sensitivity/specificity pair corresponding to a particular decision threshold. A test with perfect discrimination (no overlap in the two distributions) has a ROC plot that passes through the upper left corner (100% sensitivity, 100% specificity) (Mysinger, Carchia, Irwin and Shoichet, 2012). Therefore, the closer the ROC plot is to the upper left corner, the higher the overall accuracy of the test.

AUC is considered as an effective measure of inherent validity of the predicted model to rank randomly chosen known binders higher than a randomly chosen drug-like decoy. This curve is useful in (i) evaluating the discriminatory ability of the predicted model to correctly pick up decoys and known binders; (ii) finding optimal cut-off point to least misclassify decoys and known binders (Mysinger, Carchia, Irwin and Shoichet, 2012). As mentioned in section 3.6 that benchmarking of different scoring functions – SMINA, VINA, NNScore, DLScore and CScore (Appendix Table IIc) reported SMINA as the best scoring function. It should be noted that due to the small number of actives (28) molecules against a large set of decoys (1400) molecules, the results can be biased. Similarly, the worst performing scoring functions – NNScore and DLScore, were trained and optimised using a small training set (14 active molecules; remaining active molecules used as test set) which could have skewed the data

projections. A consensus score, Cscore, was calculated as an average score from the four scoring functions which resulted in AUC of 0.77 (Appendix IIc) and it can be proposed as an unbiased score which overcomes the overcorrected SMINA and VINA functions and undertrained neural networks of NNScore and DLScore. CScore function was used only for VS of large databases (Chapter 5) where pose selection by protein-ligand interactions was difficult otherwise SMINA docking and scoring function was used.

The generated model GPR120S attained EF score of 9.3 and was used for structure-based pharmacophore design (chapter 4; Appendix I Paper 2; Pal, Curtin and Kinsella, 2021a) as well as for structure-based virtual screening (see chapter 5; Appendix I Paper 1; Pal, Curtin and Kinsella, 2021). The limiting factor for the enrichment study was the number and chemotype of active ligands used for decoy database generation, as all the ligands were either agonists or partial agonists except AH-7614 (NAMD). Unfortunately, the six different studies where GPR120 homology models were generated to perform docking and VS experiments did not published data related to enrichment studies for comparison (Sun et al., 2010; Hara et al., 2011; Takeuchi et al., 2013; Chinthakunta et al., 2018; Zhang, Sun, Wen and Yuan, 2019).

Nevertheless, the results from the present methodologies provided valuable information concerning the optimal GPR120S model requirements for agonist selectivity recognition. In order to design new GPR120S selective ligands, both agonist and antagonist – the non-conserved residues (with respect to FFAR family) involved in the binding pocket need to be targeted. These residues are identical for GPR120S and GPR120L and share proximity to the ligands. The binding pattern of the designed ligands through CADD methodologies and MD simulations of protein-ligand complexes will help to confirm and optimize the validated GPR120S model.

To address the inherent limitations such as low-sequence identity of templates, disordered loop regions, binding site plasticity, lower enrichment factors, etc., which can lead to errors / inaccuracies in the structure-based virtual screening experiments; molecular dynamics simulations of 300 ns were next conducted (Chapter 4). The resultant refined models would minimise these limitations for latter SBDD studies.

Chapter 4

Structure based prediction of a novel GPR120 antagonist based on pharmacophore screening and molecular dynamics simulations.

Active state (agonist bound) GPCR structures are known to represent a specific conformation, which is recognized by the heterotrimeric G proteins at intracellular domains of GPCR (Kling, Clark and Gmeiner, 2016; Weis and Kobilka, 2018; Zhou et al., 2019). The conformational changes are transferred from the ligand binding pocket at the extracellular domain to the intracellular domain. There is a common mechanism of activation of GPCRs (especially in Class A – explained in section 1.4.3) which starts with agonist binding (step1), inducing inward motions of the extracellular domains of TM3, TM5, TM6, and TM7 (step 2). This inward movement is accompanied by an outward movement of the intracellular domain of TM5, TM6 and TM7 and inward movement of the intracellular domain of TM3 (step 3), allowing the G protein to bind (step 4) and become activated to relay downstream signalling (Weis and Kobilka, 2018; Zhou et al., 2019).

During this activation process, some prominent conformational changes have been observed in Class A GPCRs (Figure 4.1) such as breakage of the "ionic lock" between TM3 (E/DRY motif: Arg136 and Asp259) and TM6 at the cytoplasmic end; reorganization / rotation of the "activation switch" residue in TM6 (WxP motif: Trp277) and breakage of an electrostatic interaction in TM7 (NPxxY motif: Asn317, Pro318 and Tyr321) (Weis and Kobilka, 2018; Zhou et al., 2019). The conserved $P^{(TM5)}$ $I^{(TM3)}$ $F^{(TM6)}$ motif (Pro219, Ile126 and Phe274) residues in the TM regions and Proline residues closer to the WxP motif (TM6) serve as hinges to transfer the conformational motion from extracellular to intracellular domains (Figure 4.1) (Weis and Kobilka, 2018; Zhou et al., 2019).
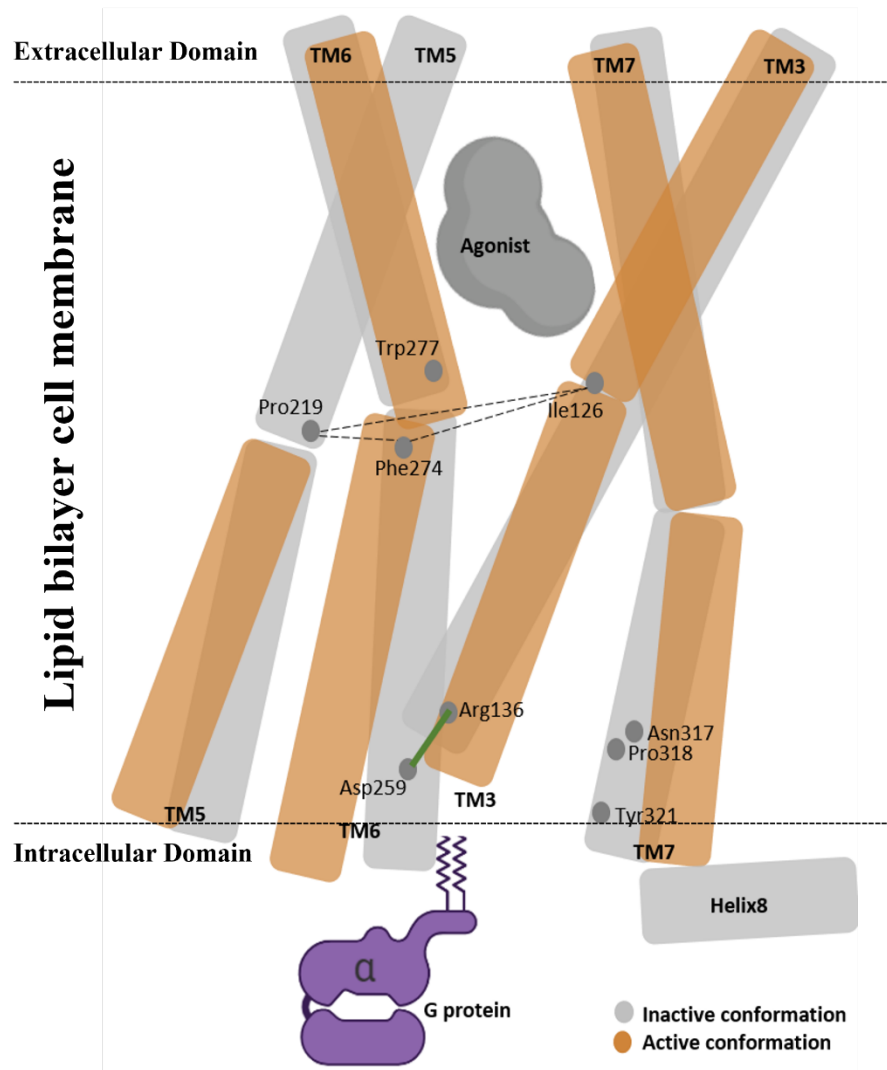
Figure 4.1: General schematic of Class A GPCRs activation with GPR120S specific residues, Inactive state (Grey) versus Active state (Orange); - illustrate the bound agonist in the orthosteric pocket activates the microswitches on TM3, TM6 and TM7. The inward motions of the extracellular domains of TM3, TM5, TM6, and TM7 is followed by the outward movement of the intracellular domains. The G protein binds at the intracellular domain and relays the downstream signalling. (Zhou et al., 2019)

As stated, (section 3.1; Figure 3.4), Class A GPCRs share several conserved motifs. The template structures, delta-like opioid and orexin O_2, show the characteristic packing of TM domains and interactions corresponding to the inactive state of Class A GPCRs, so the generated model GPR120S should also exhibit specific characteristics such as:

- Ionic lock formed between TM3 (D/ERY – Arg136) and TM6 (Asp259) (Figure 4.1, 4.2a)

- Disulfide bridge between TM3 (Cys111) and ECL2 (Cys194) (Figure 4.2b)

- H-bond network mediating interactions between TM7 (NPxxY – Asn317, Ser314), TM1 (Ser54, Asn58), TM2 (Asp85) at cytoplasmic end (Figure 4.1, 4.2c)

- Conserved interface between TM5, TM3 and TM6 by triad PIF motif - TM5 (Pro219), TM3 (Ile126) and TM6 (Phe274) near the base of the ligand binding pocket (Figure 4.1, 4.2d).

As the templates for homology modelling were antagonist bound / in the inactive state, these structural features and interactions were found to be present and stabilize the inactive conformation of the GPR120S model as reported in other inactive Class A GPCR studies (Wacker et al., 2013; Kato et al., 2019; Zhou et al., 2019).

Here, with an objective of investigating the structural and conformational changes of GPR120S from a computational perspective - MD simulations were performed, with a particular emphasis on the early events of its possible activation mechanism (Zhou et al., 2019; Perkins et al., 2014). MD simulations for an apo receptor; and two agonist-bound receptor complexes were performed to study the ligand-induced process of conformational change that leads to a possibly active receptor conformation (Wang and Chan, 2017; Weis and Kobilka, 2018; Zhou et al., 2019).

For a comparative study along with TUG891 (Figure 3.11) another agonist bound system was prepared with Compound39 – a benzofuran propanoic acid analogue (Table 1.8) (Lombardo et al., 2016) employing the set docking protocols (Figure 4.3). MD simulations were performed on three systems in a lipid bilayer: (1) Apo-GPR120S, (2) GPR120S-TUG891 (agonist bound; EC_{50} = 43.65 nM) (Figure 3.11) and (3) GPR120S-Compound39 (agonist bound; EC_{50} = 97 nM) (Figure 4.3) (Shimpukade et al., 2012; Lombardo et al., 2016). The apo protein system (model generated using antagonist-bound templates) was assumed to bias towards the features characteristics of inactive state whereas simulations of the agonist-bound systems should bias to express the ligand-induced conformational changes leading to an active state receptor conformation (Zhou et al., 2019).
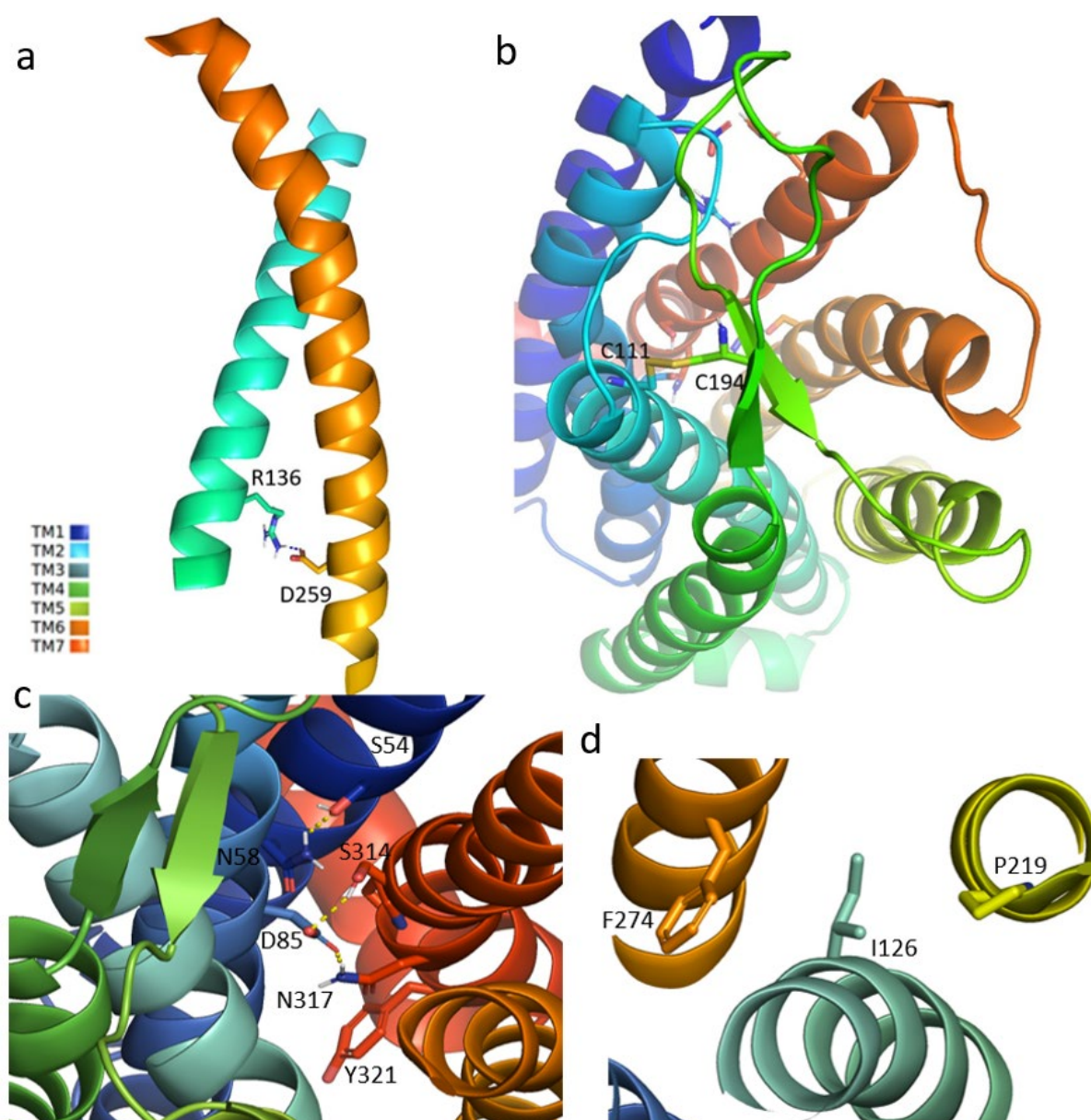
Figure 4.2: Characteristic features of Class A GPCRs observed in the GPR120S receptor model; a) Ionic lock formed between TM3 and TM6 with a H-bond distance of ~2 Å; b) Disulfide link between TM3 (Cys111) and ECL2 (Cys194); c) H-bonding (2-3 Å) network at cytoplasmic end by $N^{317}P^{318}xxY^{321}$ motif; and d) conserved $P^{219}\ I^{126}\ F^{274}$ triad forming an interface between TM 5, 3 and 6. The images were visualized and rendered in PyMol (DeLano 2018).

4.1 Molecular docking analysis of Compound39-bound GPR120S

The docked pose of Compound39 showed favourable hydrophobic interactions and formed hydrogen bonds with (Arg99, Trp277) two out of the seven experimentally validated residues (Figure 1.11) as well as two hydrogen bonds (Thr125, Asn313) with other residues in the orthosteric binding pocket. Compared to the docked pose of TUG891 (Figure 3.11, 3.13), instead of the carboxylate (-COOH) tail, the carboxyl (C=O) group of Compound39 formed a hydrogen bond with Arg99 (Table 4.1; Figure 4.3). It might be attributed to the reverse docking – where the ligand gets flipped in the pocket during docking process due to the presence of carbonyl and carboxylic groups at opposite ends. To confirm if the docking pose was not just an artifact / anomaly, the ligand was docked using different seeds / input conformations of the ligand. The consensus from ten different docking runs showed the obtained docking pose of Compound39, was the most preferred docking pose.
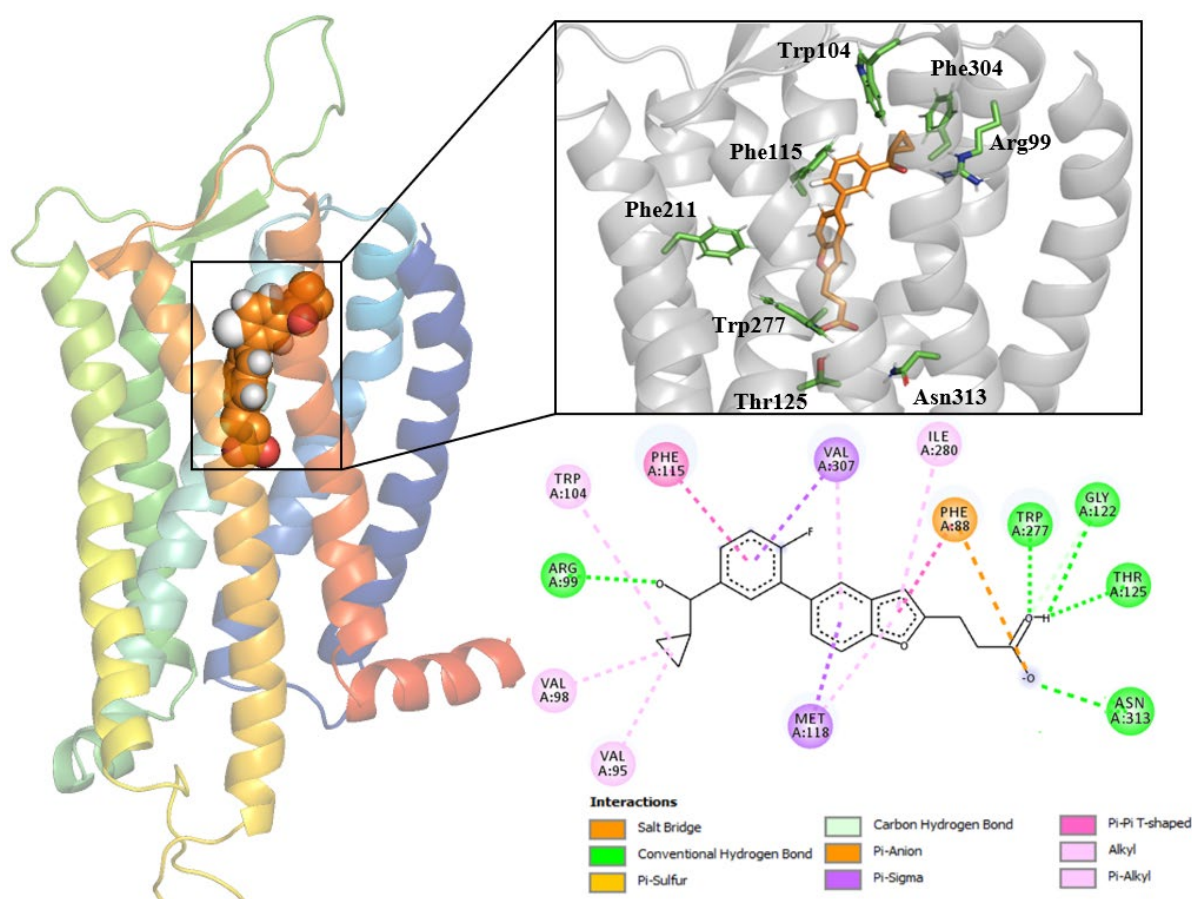
Figure 4.3: Selected docked pose of Compound39 (binding score -9.828 kcal/mol) illustrating hydrogen bond interactions with Arg99 of GPR120S and a 2D interaction map of Compound39 in the orthosteric pocket binding pocket. The 3D images were visualized and

rendered in PyMol v2.1.0. The 2D interaction maps were generated in BIOVIA DS Client visualizer v19.1 2019.

Table 4.1: List of residues forming contacts with Compound39 docked to the orthosteric binding pocket of GPR120S.

Hydrophobic interactions

Residue	Distance (Å)
Trp104	3.88
Met118	3.52
Ile280	3.98
Phe303	3.95
Phe303	3.77
Val307	3.47
Thr310	3.76

H-bonding interactions

Residue	Distance (Å) H-A	Distance (Å) D-A	Donor Angle (degree)
Arg99	1.90	2.83	151.40
Arg99	3.39	3.97	118.58
Gly122	2.81	3.65	149.66
Thr125	1.80	2.75	157.09
Trp277	1.80	2.79	171.41
Asn313	1.77	2.75	165.25

4.2 Conformational analysis of equilibrated GPR120S models in ligand-bound and apo form.

To generate the conformational ensemble, the equilibrated systems were given random velocities and subjected to further MD production runs of 300 ns where a stable protein backbone RMSD was achieved (Figure 4.4). MD trajectory of the systems at checkpoints every 25 ns was analysed to record the plateauing of protein backbone RMSD as well as potential energy of the systems, signifying the convergence achieved by the MD simulations of protein systems. The protein backbone of starting structures in all the three systems (apo; TUG891 bound; Compound39 bound) reached stability and after 200 ns no significant changes in the RMSD values were observed. The difference in the RMSD values suggested that binding of the ligands resulted in different but relatively stable conformations of the GPR120S model systems during the 300 ns MD production run.
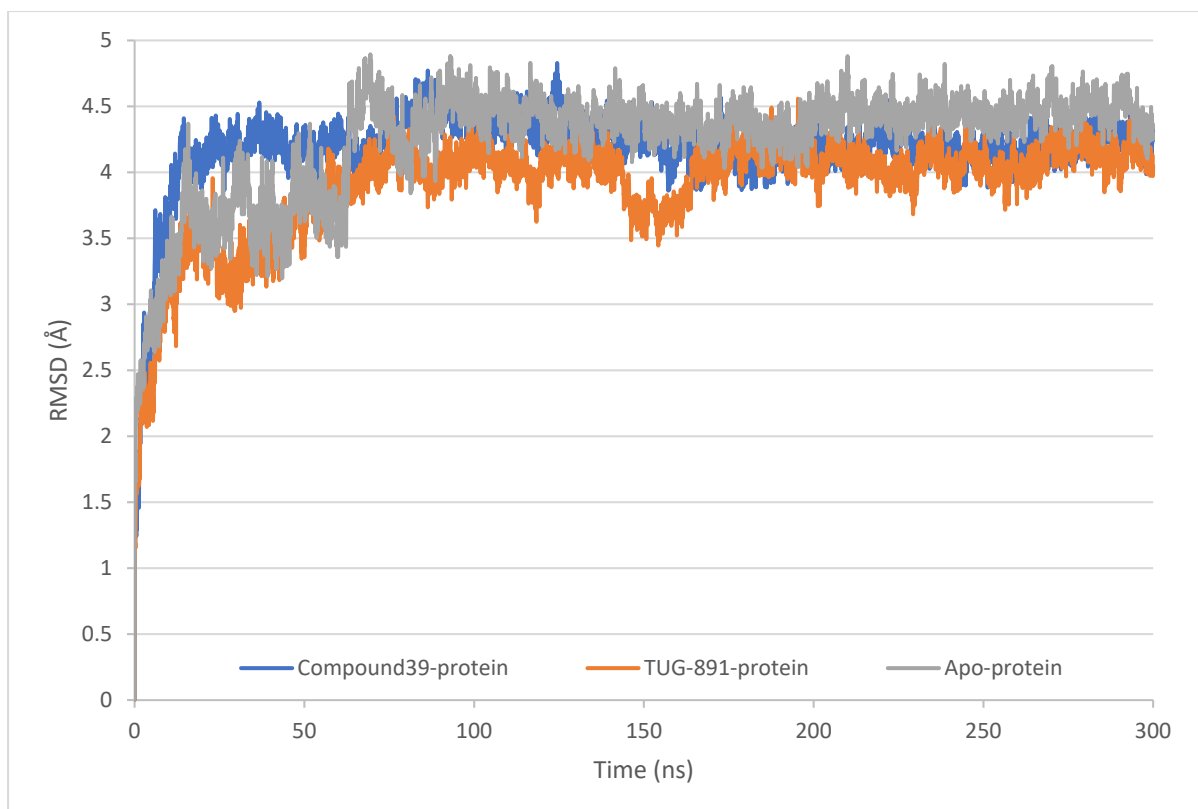
Figure 4.4: RMSD (Å) of the backbone atoms recorded during the 300 ns MD production run of GPR12S models without bound ligands – Apo (Grey) and with bound agonists – Compound39 (Blue) and TUG891 (Orange) plotted versus time (ns).

The RMSF (Root Mean Square Fluctuation) analysis of the protein backbones over the period of 300 ns MD production runs illustrate the highest range of fluctuation in the loop regions especially the ECL2 (177-203) and ICL3 (236-252) domains of the ligand bound protein models compared to the apo protein model (Figure 4.5a). The agonist-bound systems showed fluctuations in all the loop regions in comparison to apo-protein systems. The Compound39-bound system recorded a marked difference in the ICL1 (65-71) domain as well compared to the TUG891-bound system suggesting that both the agonists might be inducing conformational changes by two different mechanisms. The fluctuations in the ECL domains might be due to the presence of ligands, which due to interactions with the essential binding pocket residues (Arg99 and Trp277 specifically) pull or push the residues and transmit conformational changes to the ICL domains as well via TM regions to bring the protein to active conformation. The low range of fluctuations in the apo protein system suggest that the generated protein model was in stable inactive state and remained in the inactive state as it might be in a global minimum due to the absence of ligand induced conformational changes during MD simulation run.

The continuous disulfide linkage between two cysteine residues (Cys111 and Cys194) was observed throughout the 300 ns MD production runs (Figure 4.5b). During MD production runs, no special bond constraints were applied to keep the disulphide bond intact. The Cysteine bridge between TM3 and ECL2 is highly conserved in most of rhodopsin-like receptors and used as anchor points in modelling the individual backbone course of hydrophilic loops (Fredriksson et al., 2003; Kinoshita and Okada, 2015; Zhou et al., 2019). The analysis confirms that the disulfide bridge is conserved in GPR120S models and essential for the packing and stabilization of a restricted number of conformations of the seven TM domains.
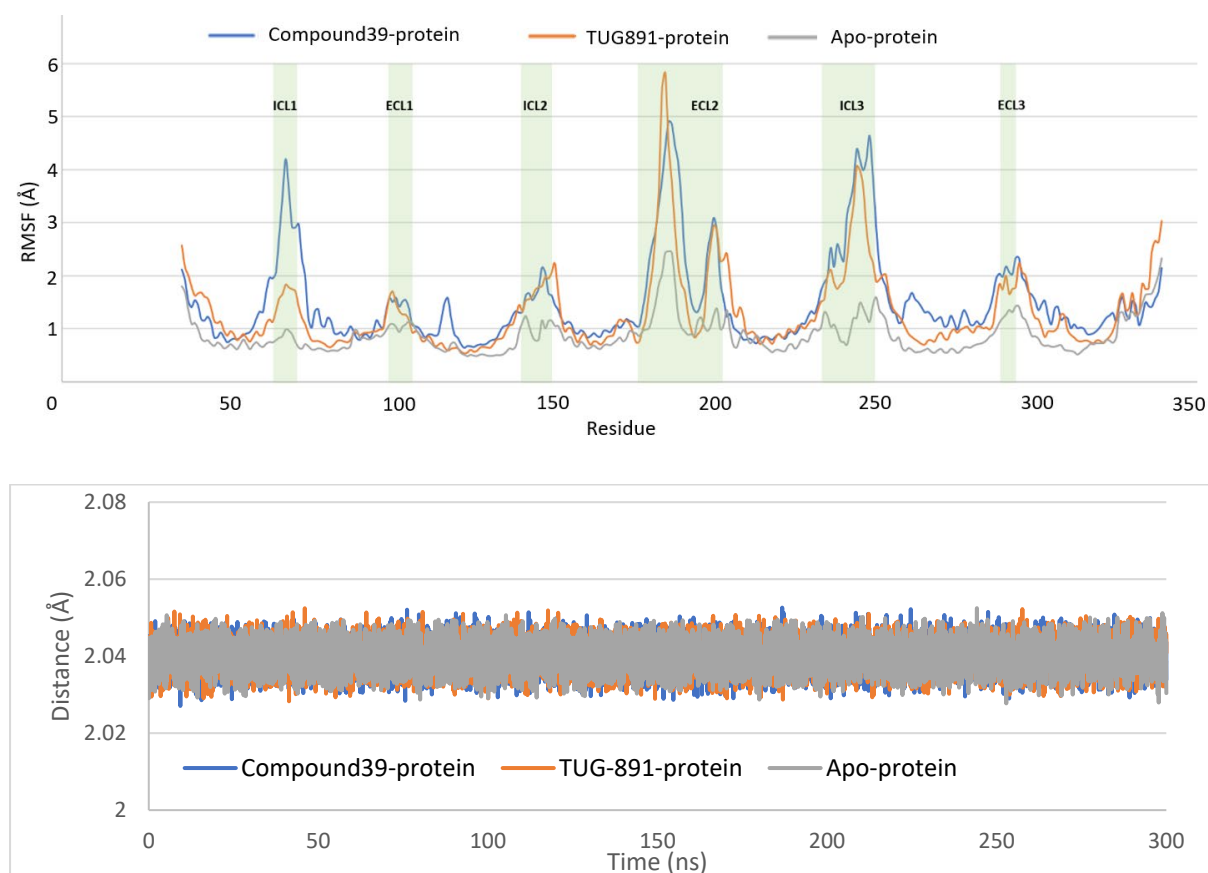
Figure 4.5: a) RMSF (Å) of the backbone atoms recorded during the 300 ns MD production run of GPR120S models; b) Distance (Å) between the centre of mass of the "cysteine bridge" residues (C111-C194) recorded during the 300 ns MD production run of GPR12S models without bound ligands – Apo (Grey) and with bound agonists – Compound39 (Blue) and TUG891 (Orange) plotted versus time (ns).

4.3 Conformational changes in the "ionic lock" [D/E]RY motif

The comparative analysis of the putatively active state models (both agonist-bound) with the inactive (apo-model) state model was performed and conformational changes in the residues involved in the ionic lock – Arg136 of TM3 and Asp259 of TM6 were analysed over the period of the MD runs (Figure 4.6a). The inactive state model (Apo) kept the conformation with the ionic lock formed between the intracellular ends of TM3 and TM6. The centre of mass of both the residues stayed within the range of 2-4 Å – appropriate to form and sustain the salt-bridge, making the inactive state of the model stable (Figure 4.6b). As expected, the active state model of TUG891-bound GPR120S showed the two residues drifting apart from each other over the period of the 300 ns MD production runs. The analysis demonstrates that the two residues were at the distance of 10 Å at the end of production run disrupting the salt bridge at the very start of the simulation.

However, the Compound39-bound protein model demonstrated unexpected behaviour of staying close to the inactive conformation. The average distance between the two residues remained ~ 5 Å throughout the MD production run. Although the residues were not close enough for salt bridge formation but with these residues being this close suggested that the intracellular cavity of the receptor would not be able to accommodate the heteromeric G protein subunits. Two possible inferences can be proposed from the Compound39-bound protein simulation: 1) The protein model might be in a local minimum of the energy landscape at that specific conformation (Goedecker, 2004); 2) The interaction network pattern of Compound39 carboxylate tail with Thr125, Trp277 and Asn313 (Figure 4.3), not observed in TUG891 docking interactions, led the MD trajectory to different conformational changes at the intracellular domain especially at TM3 – ER_{136}M motif ([D/E]RY) involved in the "ionic lock" formation.

The MD production simulation was rerun for 300 ns on the same starting conformation of Compound39-bound protein system with random initial velocities verified that the variance was reproducible. The comparative analysis of the distance between the residues (Arg136 and Asp259 – involved in formation of "ionic lock") from the 300 ns MD rerun reported an increase with a higher range of fluctuations compared to the first 300 ns MD run but were significantly less than those of the TUG891-bound system. The average distance (~8 Å) was greater than 4 Å limiting the hydrogen bond interactions for the salt bridge formation (Figure 4.6c).
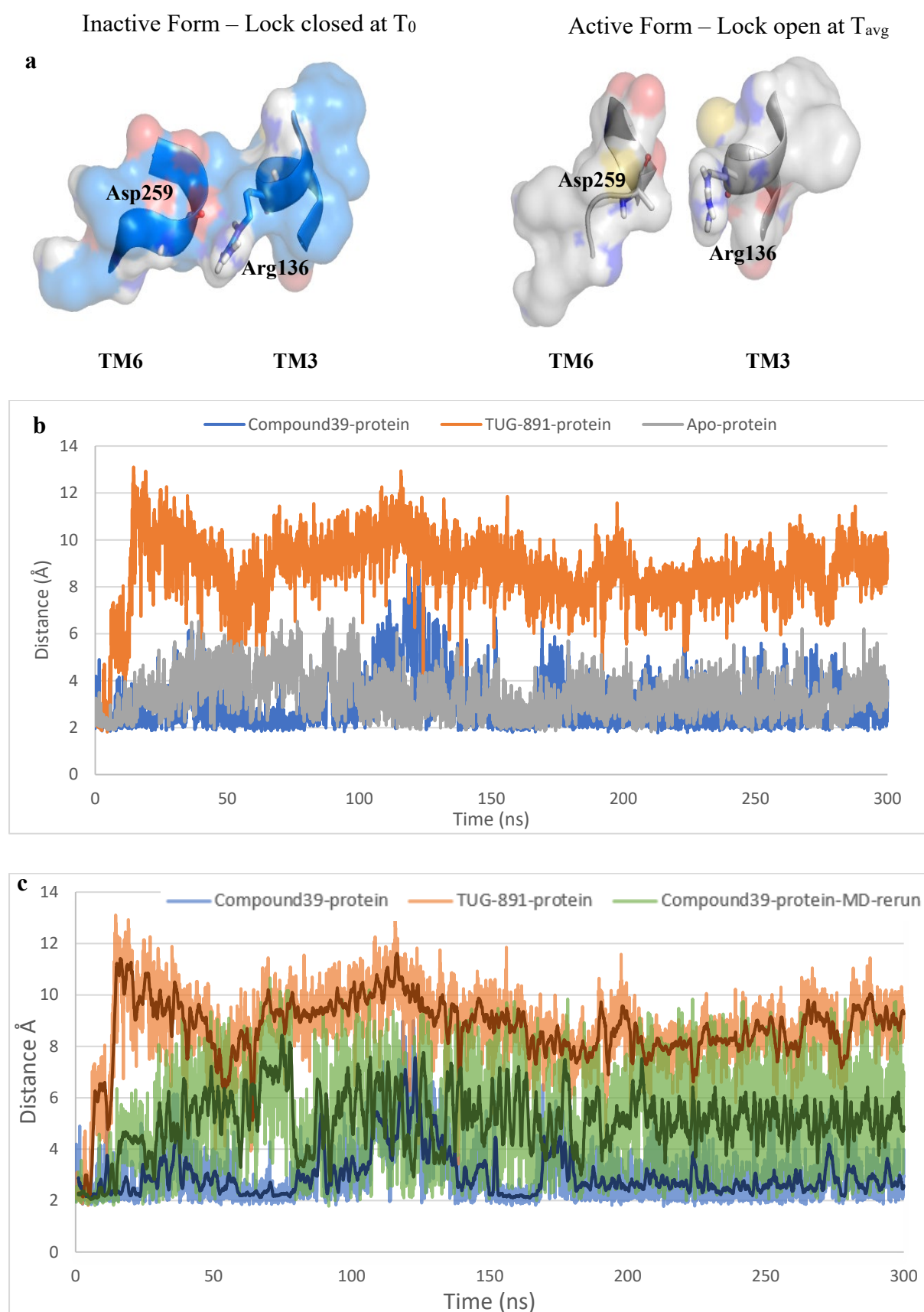
Figure 4.6: a) Graphical picture of ionic lock closed at simulation time T_0 and open conformation at T_{avg}; representing the average conformation from the 300 ns MD run of the

TUG-891 bound protein model; b) Distance between centre of mass of Arg136(TM3) and Asp259(TM6) recorded during 300 ns MD production run predicting the effect of bound ligands at the structural conformation of ionic lock residues; c) Distance between centre of mass of Arg136(TM3) and Asp259(TM6) recorded during 300 ns MD production re-run. Running average of distance every 1 ns is shown for clarity.

To the best of our knowledge, these are the first simulations of GPR120S complexed with ligands (TUG891 and Compound39) at a 300 ns timescale. A recently published study investigated protein-ligand stability from a 200 ns MD simulation of SR13 - a chromane propionic acid analogue, derived from the same Merck & Co. patented series as Compound39, (Table 1.8) docked to homology model of GPR120 (Uniprot ID: Q5NUL3-2) (Zhang, Sun, Wen and Yuan, 2019; Adams et al., 2016). This study highlighted the conformational changes adopted by SR13 to enter the binding pocket (Zhang, Sun, Wen and Yuan, 2019). However, to the best of our knowledge, these are the first simulations of GPR120S complexed with ligands (TUG-891 and Compound39) at a 300 ns timescale.

As the human GPR120S receptor can bind to the flexible FFAs (PUFAs) as well as a diverse set of rigid compounds such as TUG891 suggested the existence of different binding conformations in the orthosteric binding pocket of the receptor leading to a cascade effect inducing protein activation. The difference in binding pattern of TUG891 and Compound39 highlighted residues (Thr125, Trp277 and Asn313) in the orthosteric binding pocket for further examination. Residues Thr125 and Asn313 were not analyzed in the previous site-specific mutagenic study while the Trp277Ala mutation resulted in a loss of receptor activity (Hudson et al., 2014). The hydrogen bonding analysis of Compound39-bound protein simulations showed an average of ~60% H-bond occupancy between Compound39 and the Trp277 sidechain during the 300 ns production run accompanied with ~35% and ~10% occupancy for Asn313 and Thr125 sidechains, respectively. The interaction network of Trp277 and / or Asn313 with Compound39 might be affecting the conformational changes as observed in TUG891-bound protein simulations.

4.3.2 Conformational changes in the PIF and NPxxY motif

As shown in Figure 4.2d, the $P^{(TM5)} I^{(TM3)} F^{(TM6)}$ triad of residues form an interface between TM 5, 3 and 6 near the activation switch - 'WxP' (TM6) and inner core of the orthosteric binding pocket in the GPR120S model which is consistent with the Class A GPCRs (Wacker et al., 2013; Rasmussen et al., 2007; Ballesteros and Weinstein, 1995). From the

conformational analysis of TUG891 (Figure 4.7a, c) and Compound39 (Figure 4.7c, c) bound MD trajectories with respect to the apo-GPR120S MD conformations, a large shift (movement) in the position of the Phe274 side chain was observed coupled with a rotameric switch in Ile126. Both these changes corresponded to the down-outward movement of the intracellular domain of TM6 (Figure 4.7 d, e) with sideways movement in TM5.
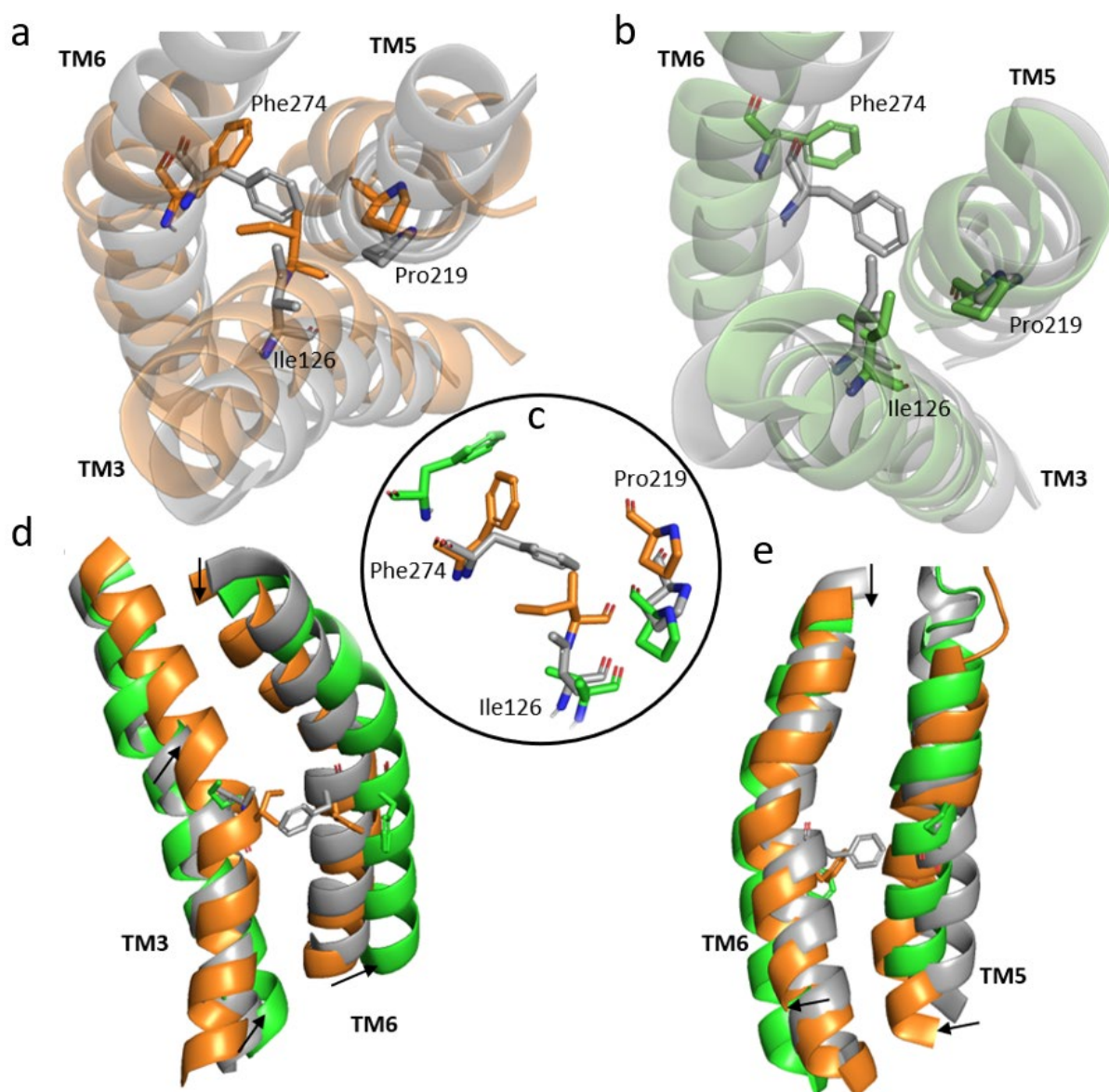
Figure 4.7 Conserved P-I-F triad motif in Apo (Grey); a) TUG891-bound (Orange) and b) Compound39-bound (Green) GPR120S models. A rotameric shift of Ile126 (c) resulted in rotation of Phe274 in TUG891 bound (orange) and movement of Phe274 in Compound39 bound (Green) GPR120S. The Phe274 (TM6) rotation and / or shift resulted in displacement of TM3 (d) and TM5 (e).

Although TUG891 and Compound39 bound systems showed Phe274 sidechain movement in opposite directions, in both systems the shift resulted in the spatial rearrangement of intracellular domains of TM helices of different magnitudes of 2-4 Å. Again, the slight differences in spatial rearrangement in the Compound39 bound system might be attributed to Thr125 and Asn313 interactions with the ligand. These rearrangements of TM regions were inferred to open the helical bundle at the intracellular domain and facilitate the G-protein binding, thus mimicking the translation of receptor confirmation from inactive state (apo) to active state (agonist-bound) (Rasmussen et al., 2007).

The conserved NPxxY motif at the cytoplasmic end of TM7 is another activation switch of class A GPCRs and the interplay between residues of D/ERY and NPxxY motifs has been defined as a crucial determinant during transition of Class A GPCRs from the inactive to active state (Fritze et al., 2003). In the classical example of β_2 adrenergic receptors the intracellular end of TM7 moves towards the TM bundle core accompanied by the Tyr (of NPxxY motif) side chain rotation, which moves the side chain further into the TM bundle (Katritch, Cherezov and Stevens, 2013; Rasmussen et al., 2007). MD trajectory analysis of TUG891-bound and Compound39-bound GPR120S systems reported sharp rotation of Tyr321 sidechain during the 300 ns simulation compared to the apo-protein system (Figure 4.8) with a difference of small outward displacement of TM7 in both agonist-bound systems. Asn317 – the first residue of NPxxY motif, reported a highly conserved H-bond interaction with Asp85 of TM2, which is a characteristic feature known in the inactive state of Class A GPCRs (Rasmussen et al., 2007). The interaction was not reported in either agonist bound GPR120S MD systems (Figure 4.8 c) due to displacement of intracellular domain of TM7 by 2-3 Å (Figure 4.8 a, b).

As the human GPR120S receptor can bind to the flexible FFAs (PUFAs) as well as a diverse set of rigid compounds such as TUG891 suggested the existence of different binding conformations in the orthosteric binding pocket of the receptor leading to a cascade effect inducing protein activation. The difference in binding pattern of TUG891 and Compound39 highlighted residues (Thr125, Trp277 and Asn313) in the orthosteric binding pocket for further examination. Residues Thr125 and Asn313 were not analyzed in the previous site-specific mutagenic study while the Trp277Ala mutation resulted in a loss of receptor activity (Hudson et al., 2014). The hydrogen bonding analysis of Compound39-bound protein simulations showed an average of ~60% H-bond occupancy between Compound39 and the Trp277 sidechain during the 300 ns production run accompanied with ~35% and ~10%

occupancy for Asn313 and Thr125 sidechains, respectively. The interaction network of Trp277 and / or Asn313 with Compound39 might be affecting the conformational changes as observed in TUG891-bound protein simulations.

Figure 4.8: Conserved NPxxY motif in Apo (Grey); a) TUG891-bound (Orange) and b) Compound39-bound (Green) GPR120S models. H-bond interaction between TM7 and TM2 were not observed in agonist bound protein models (c). A sharp rotation of Tyr321 (d) side chain was observed in agonist bound protein models with respect to apo model.

The above stated results from the conformational analysis of the MD trajectories of apo and agonist bound GPR120S receptor models suggested that the binding of agonists – especially TUG891, enabled the transition of GPR120S receptor modelled using inactive templates towards the active state conformation. The specific reasons for larger shifts observed in Compound39-bound model such as Cα of Phe274 and rotation of Tyr321 sidechain compared to TUG891-bound model from the present MD simulation production runs remain to be fully determined. The conformational changes and coupling interactions between conserved motifs were consistent with the Class A GPCRs and validated the docking predictions of TUG891 as

its binding decreases the probability of inactive-like states for PIF, D/ERY, WxP, NPxxY and ionic lock motifs. By definition an antagonist is a drug molecule which upon binding to the receptor prevents the conformational transition of the receptor from inactive state to active state and inhibits the physiological action (Figure 1.7). These results also further established the hypothesis that the Compound39 interactions with Trp277 and / or Asn313 due to reverse docking prediction might be affecting the conformational transition of protein from inactive to active state compared to TUG891-bound protein simulations. A small molecule with stable binding interactions with Trp277 and / or Asn313 might act as a GPR120 antagonist and stabilizes GPR120 receptor in the basal activity state.

4.4 Trp277 and Asn313-based pharmacophore screening

Based on the inferred significance of the Trp277 and Asn313 interaction network with the carboxylate chain of Compound39, a single structure-based pharmacophore model was generated by enumerating the 3D conformation of functional features present in the receptor binding pocket. The docked conformations of TUG-891 and Compound39 were superimposed to generate the pharmacophore model using the ZINCPharmer package (Koes and Camacho, 2012), which resulted in a six-featured hypothesis consisting of two HBA (Hydrogen bond acceptor), two Ar (Aromatic ring systems) and two Hb (hydrophobic) with preferred chemical features (Figure 4.9, Appendix Table IIe). The selected hypothesis was generated to focus screening on ligands interacting with the Trp277 and Asn313 (Figure 4.9b) residues with scaffold features of TUG-891 and Compound39 being added to attain rigidity and better anchorage in the binding pocket by interacting with essential binding residues (Hudson et al., 2014; Hudson et al., 2013) such as Arg99, Trp104, Phe115, Trp207, Phe211 and Phe304.
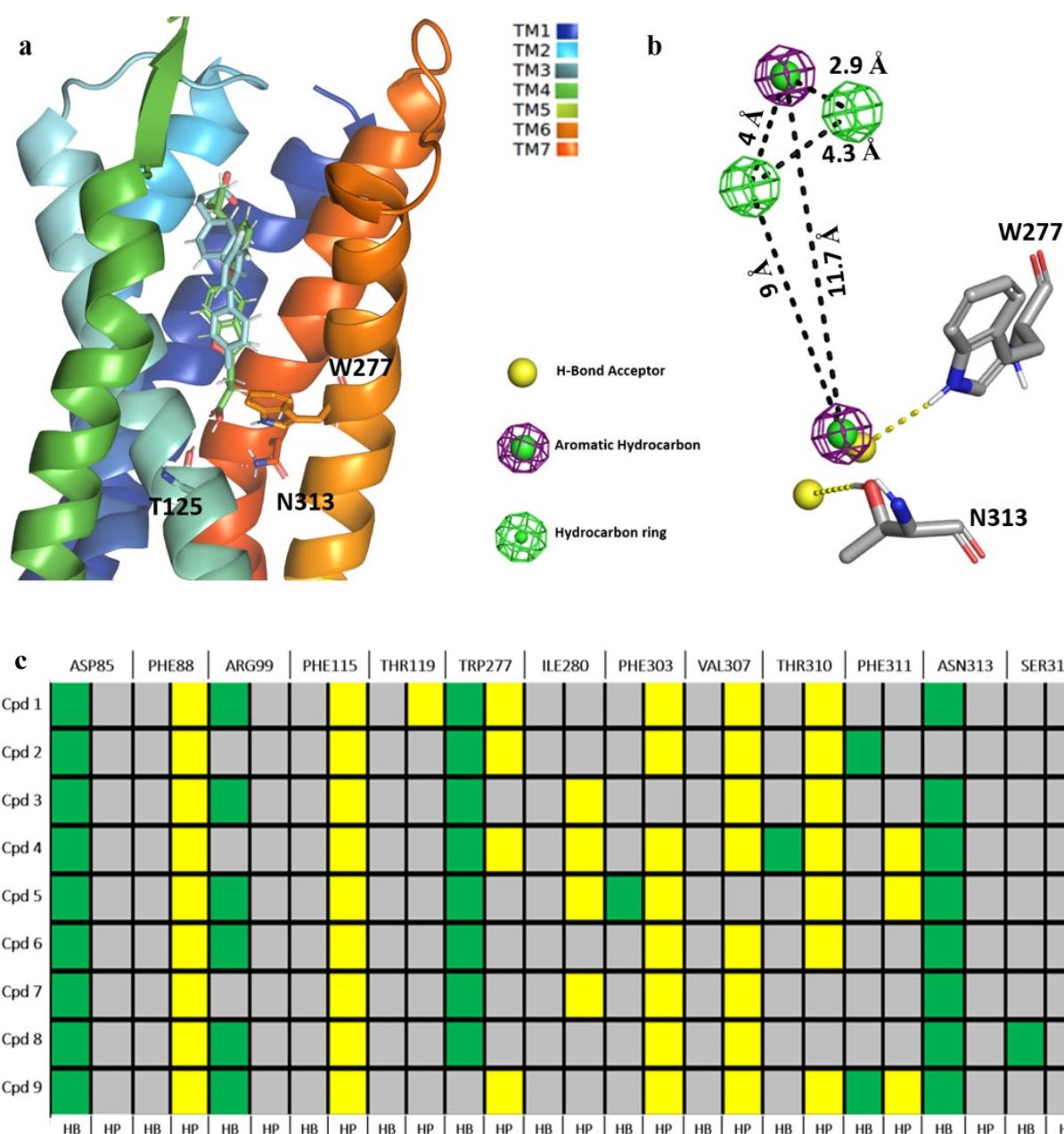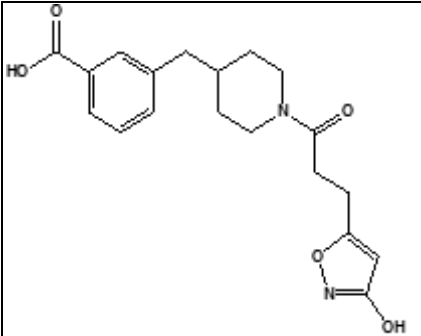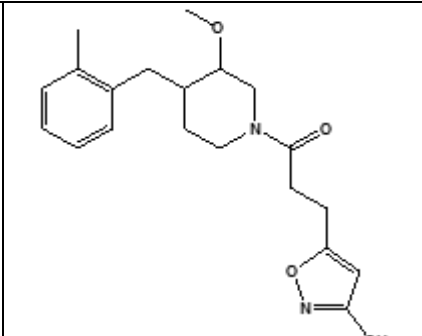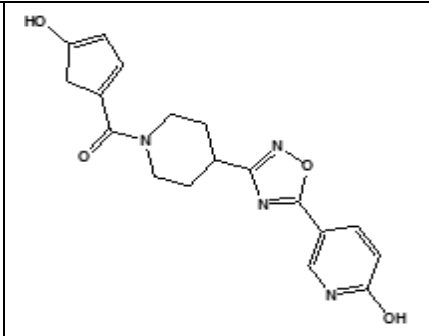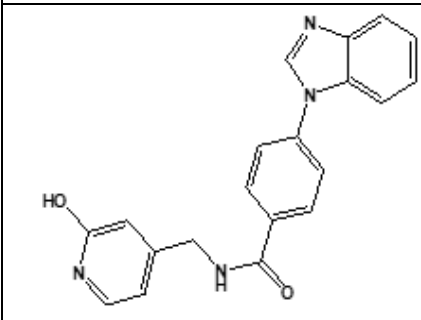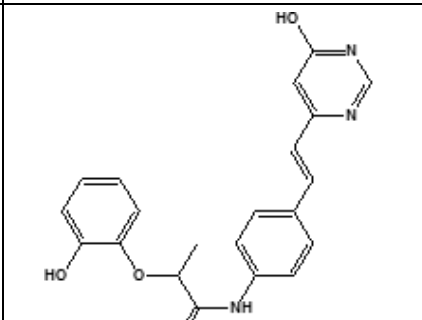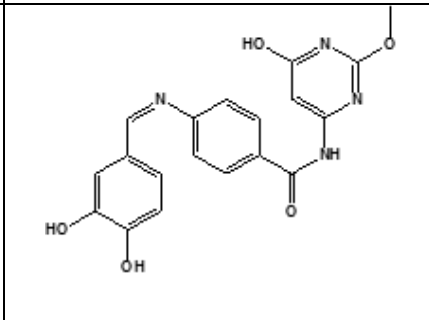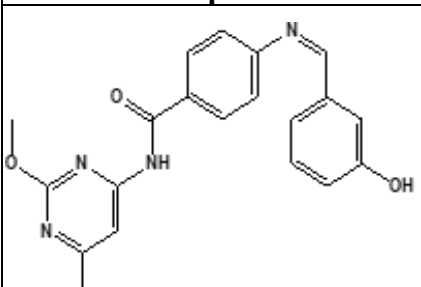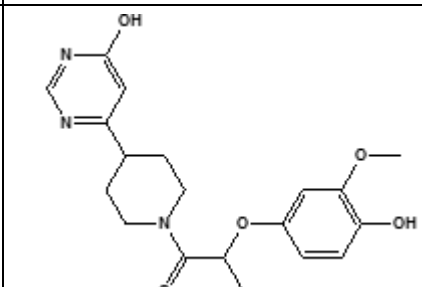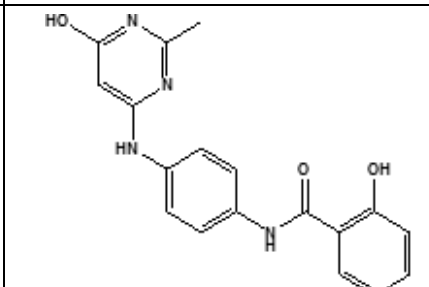
Figure 4.9: a) Superimposed docked poses of TUG-891 (Cyan stick model) and Compound39 (Green stick model) with GPR120S. TM5 is hidden in the image for a clearer view; b) Developed pharmacophore model with its corresponding chemical features used for screening the ZINC database (Sterling and Irwin, 2015; Koes and Camacho, 2012); c) Protein-ligand interaction fingerprint map of the 9 docked compounds with GPR120S; Green – (HB) Hydrogen bond; Yellow – (HP) Hydrophobic interactions; Grey – No interactions.

The pharmacophore-based virtual screening (VS) resulted in 63 unique chemical hits identified from the ZINC15 commercial database of ~230 million compounds. Further, selection of these 63 compounds for structure-based VS was performed by physicochemical

profiling using SwissADME (Daina, Michielin and Zoete, 2017) to predict the druglike and / or lead-like nature of the compounds. In combination with analysis of predicted physicochemical descriptors, an *in cerebro* assessment was applied to select compounds with diverse and synthesizable scaffolds. The final screening resulted in 9 best-hits (shown in Table 4.2, 4.3 and 4.4) based on the SwissADME predictions over compounds' physicochemical and pharmacokinetic properties.

Table 4.2: Assigned nomenclature and 2D structures of selected 9 best hit compounds from SBPM.

Cpd1	**Cpd2**	**Cpd3**
Cpd4	**Cpd5**	**Cpd6**
Cpd7	**Cpd8**	**Cpd9**

The physicochemical and pharmacokinetic parameters of the selected candidates were following the drug-like molecule rules such as - Lipinski violations, Ghose violations, Veber violations, Egan violations and Muegge violations (section 1.7.5; Table 4.3, 4.4). Molecular weights (MW) of all the 9 best hits were in the range of 300-400 Daltons and the impact of the MW being in the ideal range can be observed on the GI absorption predictions (Table 4.4) of these compounds. Similarly, lipophilicity (iLOG Po/w) of 9 hit compounds was predicted to be lower than 3.5 by SwissADME which can be a determinant factor in several ADMETox parameters as well as potency. Solubility and metabolism of the drug candidates are known to be reduced at high lipophilicity (~5) as well as blood brain barrier permeability of the test candidates is high which can result in the adverse effects. SwissADME predictions reported that the hit compounds (except Cpd 2) are not BBB permeant and are soluble in aqueous phase (Table 4.4). The number of hydrogen bonding donors (<5) and acceptors (<10) present in the drug-like candidate are related to the molecule's polarity and permeability of oral drugs. As TPSA is related to hydrogen bonding donors and acceptors, these parameters dictate the predictions of drug permeability and oral bioavailability (Daina, Michielin and Zoete, 2017). As lower TPSA values (~80 Å) are known to favor permeability and oral bioavailability, Cpd 6 was reported to have highest (137 Å) TPSA which can be attributed to 8 H-bond acceptors in the molecular structure.

Cytochrome P450 monooxygenase (CYP) enzymes play a pivotal role in drug metabolism and elimination (Daina, Michielin and Zoete, 2017) and 50-90% of the drug candidates have been estimated to act as CYP enzyme substrates and result in adverse effects. Cpd 4 was predicted to be a substrate of the five major CYP enzymes (Table 4.4) and should be reported clearly if Cpd 4 proceeds to later stages of drug discovery. Identification of unfavorable fragments which can lead to toxicity was reported as PAINS – pan-assay interference compounds and Brenk alerts. Cpd 6 and Cpd 7 raised the alert that they contain fragments which could either interfere with the assay readouts or cause covalent protein/DNA modifications and subsequently relay adverse effects (Daina, Michielin and Zoete, 2017).

Such an informed manual selection of ligands after virtual screenings have previously been reported to refine the screening performance (Voet, Kumar, Berenger and Zhang, 2014; Voet et al., 2011). These hit compounds were then prepared and docked into the receptor binding pocket following the set protocol (section 2.1.3). The protein-ligand interactions of these 9 compounds were analyzed (Figure 4.9c, Appendix Figure IIf) to confirm Trp277 / Asn313 H-

bond interaction(s) before preparing the protein-ligand complexes for 100 ns MD production runs following the set protocol (section 2.1.4).

Table 4.3: Physicochemical parameters and docked binding scores (kcal/mol) from SMINA of the selected pharmacophore-based VS hits obtained SwissADME drug-likeness profiling enlisting molecular weight (MW), hydrogen bond acceptors (HBA), hydrogen bond donors (HBD), total polar surface area (TPSA) and predicted oil/water partition co-efficient (iLOG Po/w) (Daina, Michielin and Zoete, 2017).

Molecule	MW	H-bond acceptors	H-bond donors	TPSA (Å)	iLOG Po/w	SMINA Score (kcal/mol)
Cpd 1	357.38	6	1	106.7	2.54	-10.9
Cpd 2	358.43	5	1	75.8	3.06	-10.8
Cpd 3	354.36	7	2	112.58	2.75	-9.8
Cpd 4	344.37	4	2	80.04	2.63	-9.5
Cpd 5	377.39	6	3	104.57	2.86	-9.3
Cpd 6	380.35	8	4	137.16	1.7	-9.0
Cpd 7	364.35	7	3	116.93	2.36	-9.0
Cpd 8	373.4	7	2	105.01	3.24	-8.7
Cpd 9	336.34	5	4	107.37	1.87	-8.7

Table 4.4: Pharmacokinetics parameters of the selected pharmacophore-based VS hits obtained from SwissADME drug-likeness profiling (Daina, Michielin and Zoete, 2017). Color code – Green : Highly favourable; White : Favourable; Red : Unfavourable.

	Cpd 1	Cpd 2	Cpd 3	Cpd 4	Cpd 5	Cpd 6	Cpd 7	Cpd 8	Cpd 9
GI absorption	High	High	High	High	High	High	High	High	High
BBB permeant	No	Yes	No	No	No	No	No	No	No
P-glycoprotein interaction	No	No	Yes	Yes	No	No	No	Yes	No
CYP1A2 inhibitor	No	No	No	Yes	No	Yes	Yes	No	Yes
CYP2C19 inhibitor	Yes	No	No	Yes	No	No	No	No	No
CYP2C9 inhibitor	No	No	No	Yes	Yes	No	Yes	No	No
CYP2D6 inhibitor	No	Yes	No	Yes	Yes	No	No	No	Yes
CYP3A4 inhibitor	No	Yes	No	Yes	No	No	No	No	Yes
Skin permeant log Kp (cm/s)	-6.85	-6.57	-7.41	-6.43	-6.18	-7.03	-6.68	-7.18	-5.26
Lipinski violations	0	0	0	0	0	0	0	0	0
Ghose violations	0	0	0	0	0	0	0	0	0
Veber violations	0	0	0	0	0	0	0	0	0
Egan violations	0	0	0	0	0	1	0	0	0
Muegge violations	0	0	0	0	0	0	0	0	0
PAINS alerts	0	0	0	0	0	1	0	0	0
Brenk alerts	0	0	0	0	0	2	1	0	0
Lead likeness violations	1	1	1	0	1	1	1	1	1
Synthetic Accessibility	3.47	3.99	4.06	2.25	3.37	3.09	3.03	3.58	2.55
Estimated Aqueous Solubility	Soluble	Soluble	Soluble	Soluble	Moderately soluble	Soluble	Soluble	Soluble	Moderately soluble

4.4.1 MD simulations of Trp277 and Asn313 interacting compounds.

Comparative MD simulations of 100 ns duration for each of the 9 complexes were carried out to evaluate the significance of the Asn313 interaction in terms of conformational changes to the protein. A MD time scale of 100 ns was chosen based on observations from previous MD simulation of TUG891 and Compound39 (Figure 4.6) as well as other studies where 100 ns MD simulations were shown to be sufficient to illustrate the dynamics of protein-ligand fingerprinting and induced conformational changes in GPCR molecular studies (Chan, Filipek and Yuan, 2016; Rahman et al., 2020; Zhang, Sun, Wen and Yuan, 2019).

The protein-ligand binding affinity of the selected 9 compounds over the duration of 100 ns MD simulations predicts the strength of the binding interaction between the receptor and the compounds (Figure 4.10a). Cpd 2 and Cpd 8 were predicted to show a continuous decreasing gradient in the binding affinity while Cpd 9 was predicted to show a continuous increase in the binding affinity over time during the MD simulation. Other compounds did not show a consistent pattern of decreasing or increasing binding affinity predictions (Figure 4.10a).

The binding affinity predictions from the 100 ns MD studies gave insights to the binding stability of docked hits and non-covalent interactions with residues in the binding pocket. Amongst all the compounds, Cpd 1, 7 and 9 were predicted to conserve Trp277 and Asn313 H-bond interactions during the 100 ns MD simulations (Figure 4.10b). As the binding affinities predicted for this study used only the protein-ligand complex (without solvent) snapshots extracted from the MD simulations, it should be cautioned that solvent (water) molecules also play an important role in non-covalent interactions – forming bridge interactions between ligand and protein (Venkatakrishnan et al., 2019).

Most of the interactions with Cpd9 (electrostatic or hydrophobic) were observed to be continuous during the simulation. Similarly, the protein-ligand binding affinity prediction over the 100 ns MD simulation suggested a stable binding of Cpd9 in the orthosteric binding pocket of GPR120 (Figure 4.10a). In contrast, Cpd 2 showed the lowest binding affinity in the binding pocket compared to the other ligands and Cpd 2 was the only ligand which did not form hydrophobic and /or H-bond interactions with Arg99 as well as Asn313 (Figure 4.9c and 4.10a) at the starting conformation. Cpd 7, which had a Asn313 H-bond interaction (Figure 4.9c) at the initial (0 ns) conformation was predicted to have a high binding affinity of -10.8 kcal/mol. However, the binding affinity of Cpd 7 reduced to -10.25 kcal/mol over the MD simulation as the number of Asn313 H-bond interactions of Cpd 7 reduced (Figure

4.10b). The correlative binding affinity predictions and Asn313 interactions could infer the important role of Asn313 residue in the ligand binding stability.
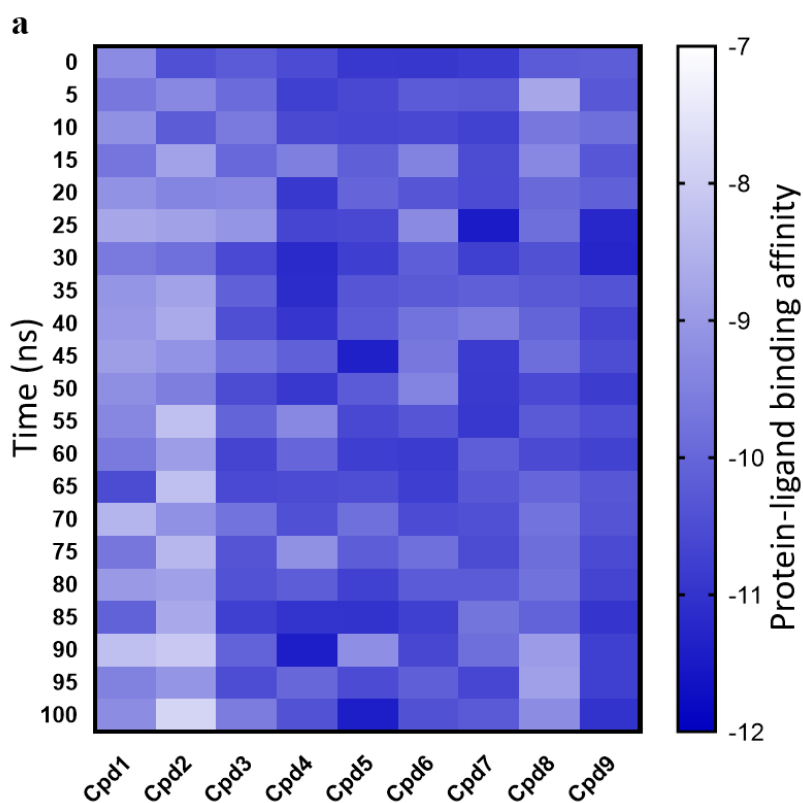
Figure 4.10: a) Heatmap of protein-ligand binding affinities (kcal/mol) of snapshots extracted from the 100 ns MD simulation trajectory and scored by webserver WADDAICA (Bai et al., 2021) (continued)
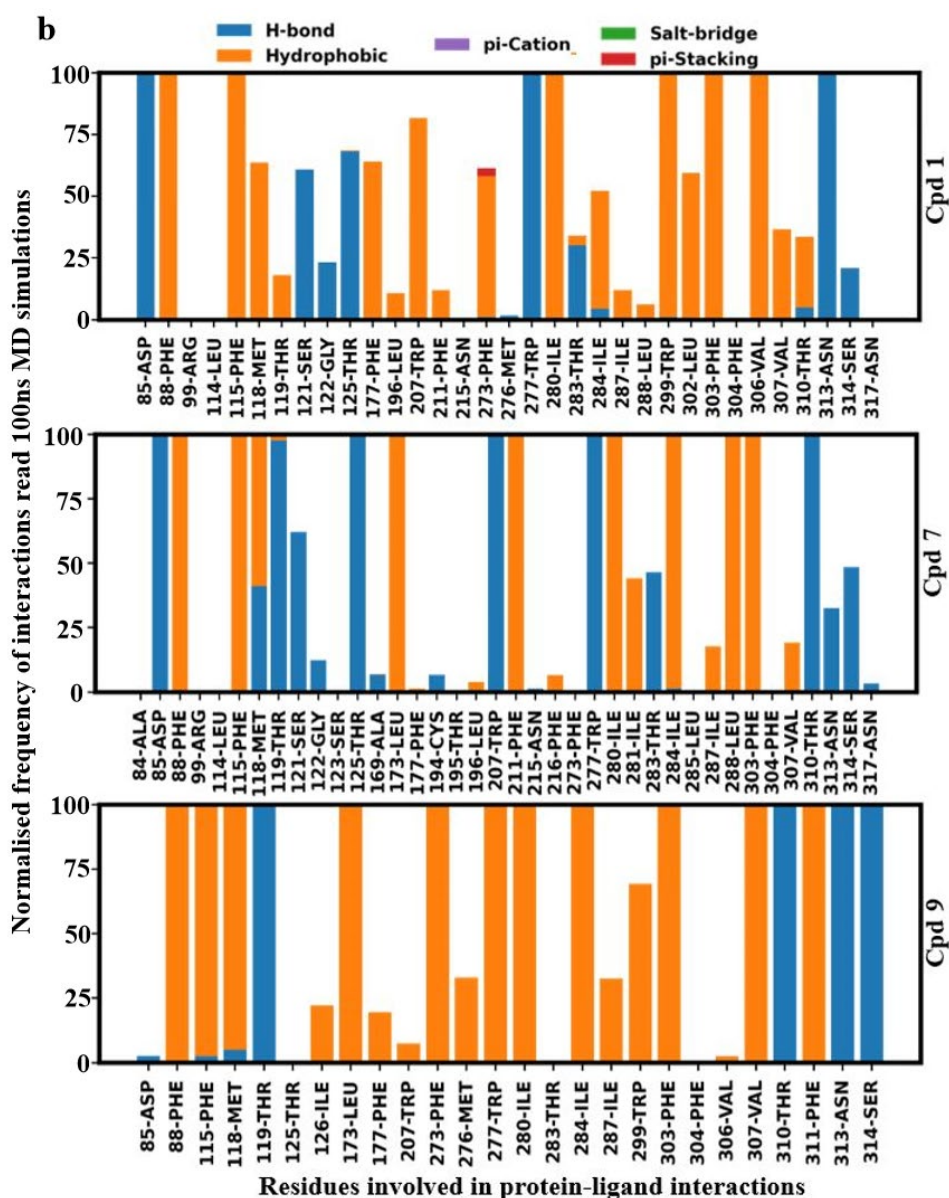
Figure 4.10: b) Protein-ligand interaction fingerprint map (Available at GitHub link: https://github.com/tavolivos/Molecular-dynamics-Interaction-plot) plotting the normalized frequency of interactions of the residue with the ligands, shows compounds 1, 7 and 9 with conserved Trp277 and Asn313 H-bond interactions over the period of 100 ns MD production runs.

As GPR120S was modelled from templates in the inactive form, the binding of ligands with antagonistic activity should keep the receptor in the inactive state without causing major conformational changes at the intracellular domain of the receptor. The protein's structural stability evaluation by RMSF analysis of 100 ns trajectory showed (Figure 4.11a; Additional Figures in Appendix VII a-e) that Cpd 9 and Cpd 1 stabilized the protein backbone as well as

reducing the fluctuations of ECL2 (177-203) and ICL3 (236-252) regions observed in Apo, TUG-891 and Compound39 bound protein systems while Cpd 7 bound protein system has recorded the highest range of fluctuations (~12 Å) in the ICL3 region. Binding of other ligand molecules have an overall similar effect on the protein backbone fluctuations, but higher than that of Cpd 9.

Focusing on the ligands themselves Cpd 9 and Cpd 7, they were also found to be the most stable in the GPR120S orthosteric binding pocket with RMSD values below ~1.5 Å and 2 Å respectively throughout the 100 ns MD trajectory (Figure 4.11b). Such low ligand RMSD values suggest that these two ligands are tightly bound to the orthosteric binding pocket without major changes in their initial docked orientations. It is important to mention that both ligands lead to contrasting effects on the protein-backbone RMSD values – Cpd 9 stabilized the protein-backbone in its initial inactive form whereas Cpd 7 deviated the protein-backbone away from its initial conformation. Cpd 9 stabilized the protein within the timespan of the first 20 ns, keeping the average RMSD of protein model below 4 Å. While the binding of Cpd 7 to GPR120S leads to the highest RMSD values (~ 8 Å). Ligand RMSD analysis of Cpd 1 also presented a range of fluctuations with protein backbone RMSD reaching above ~5 Å (Figure 4.11b; Additional Figures in Appendix VII a-e).

During the 100 ns production run the "ionic lock" remained closed, inhibiting the coupling between receptor and G-protein, and hence keeping the receptor in the inactive state. Compared with other ligands, the Cpd 9 bound protein model predicted the least movement of the distance between Arg136 and Asp259 (involved in the "Ionic lock") at the intracellular domain of GPR120S (Figure 4.11c) that is the site specific for G-protein coupling. The study published by Provasi et al. (Provasi et al., 2011) used inactive and active crystal structures of GPCRs with ligands eliciting different pharmacological actions performing 20 ns MD simulations. The study reported that depending on their pharmacological activity, the ligand bound to the receptor can shift the conformational equilibrium towards active or inactive state of the receptor. A similar conformational shift (Cpd7 – from inactive to active state) as well as stability (Cpd9 – stabilized inactive state) in receptor state was observed in our study. Here, during the 100 ns MD production run of Cpd9 the "ionic lock" remained closed (Figure 4.11c, Figure in Appendix VII d), inhibiting the coupling between receptor and G-protein, and hence keeping the receptor in the inactive state.

a
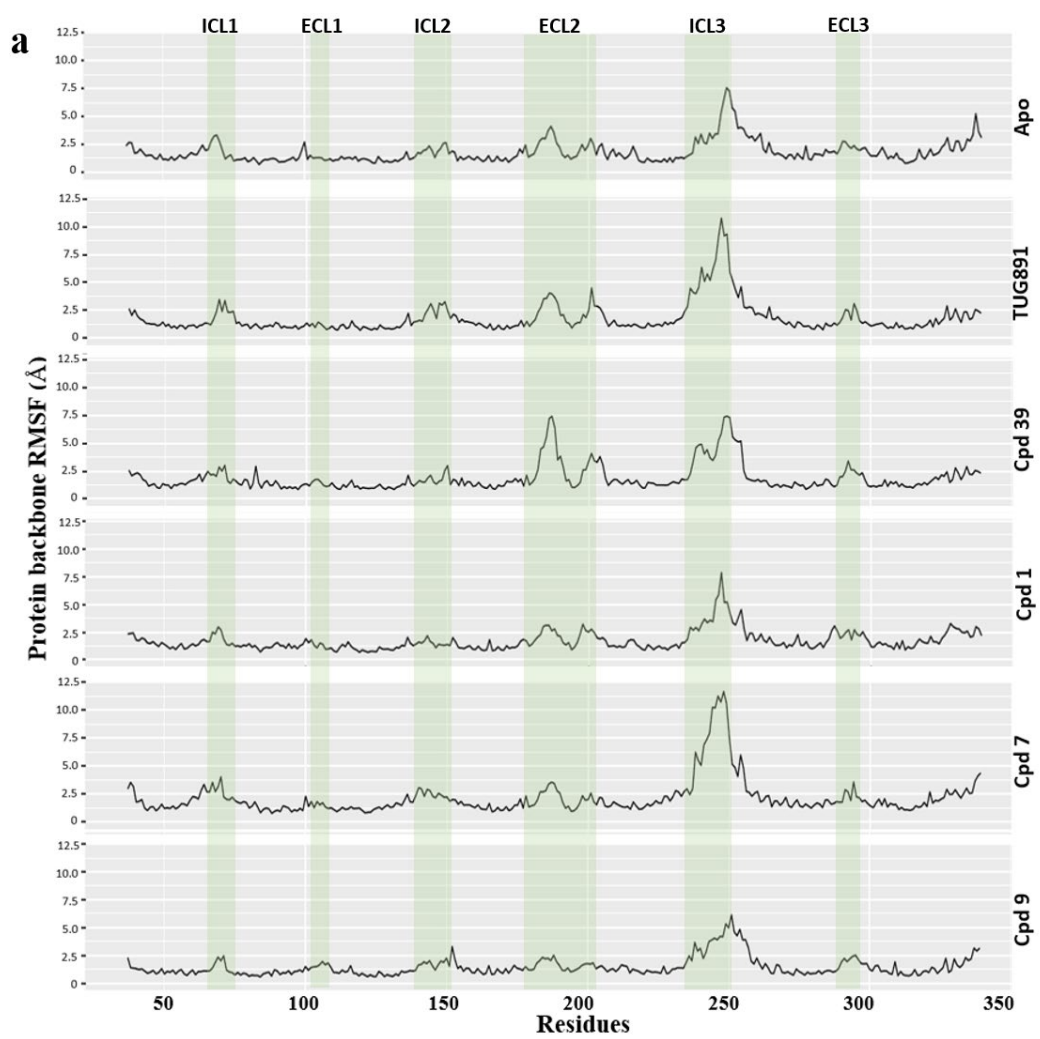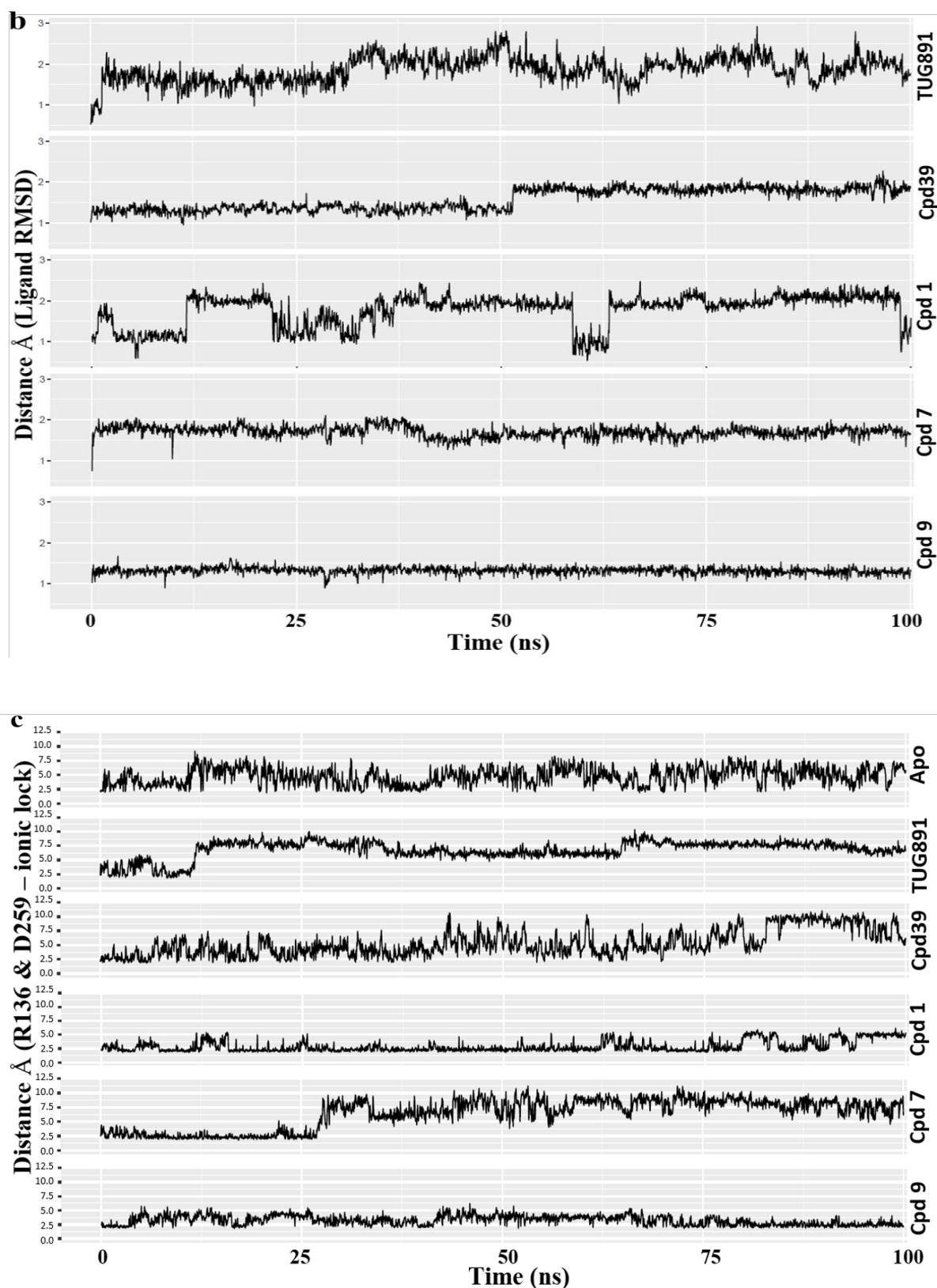
Figure 4.11: Comparative MD simulation analysis of best hits, compound 1, 7 and 9 with respect to Apo, TUG-891 and Compound39 bound proteins over 100 ns timescale; a) RMSF plot of the protein backbone. The intracellular loop (ICL) and extracellular loop

(ECL) regions are highlighted as shaded regions. The transmembrane regions in the 3D structure are - TM1: 36-65; TM2: 73-101; TM3: 107-141; TM4: 152-175; TM5: 204-233; TM6: 252-289; TM7: 296-324; b) RMSD plot of the ligand atoms; c) Distance plot between the center of mass of residues Arg136(TM3) and Asp259(TM6) involved in "ionic-lock" conformation.

4.4.2 Multiple linear regression model and structural diversity analysis of the selected compounds

For further analysis, MLR was applied to predict the contribution of physicochemical descriptors (independent variable) such as molecular weight, H-bond donors, H-bond acceptors, logP and topological polar surface area (TPSA) on binding affinity (dependent variable) (Funar-Timofei, Borota and Crisan, 2017; da Silva Costa et al., 2018) (Figure 4.12). The average binding affinity (obtained from WADDAICA webserver) of the last 20 ns of MD snapshots (from 80 to 100 ns) was used for the MLR model. The performance of the MLR model is expressed in terms of R^2, which was found to be 0.799 signifying that ~ 80% of the data fit the regression model. Cpd 9 with the highest binding affinity value from the WADDAICA server (Bai et al., 2021), was also predicted to have the highest binding affinity by the MLR model. Indeed, the MLR model agreed with the binding affinity predictions of the WADDAICA webserver, but it should be noted that the majority of the variables used in the MLR model were obtained from other prediction algorithms – such as binding affinity, logP and TPSA.
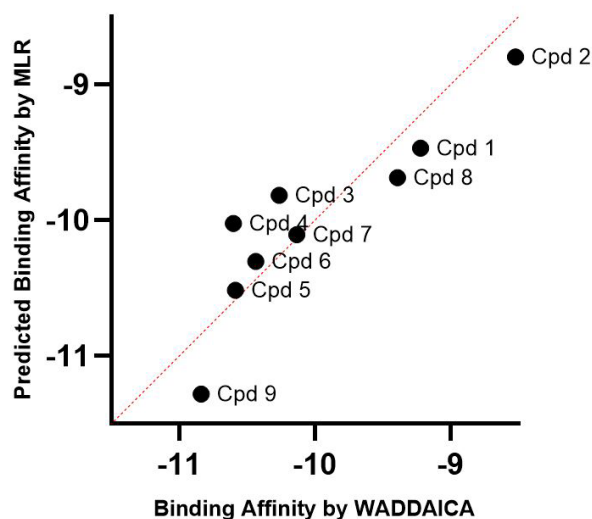
Figure 4.12: Scatter plot of a multiple linear regression model of the contribution of physicochemical descriptors –molecular weight, number of H-bond acceptors and H-bond donors, TPSA and logP from SwissADME (Daina, Michielin and Zoete, 2017) over the WADDAICA (Bai et al., 2021) binding affinity for the 9 compounds.

The selected compounds Cpd 1-9 were also clustered with the co-crystalized ligands of templates (4N6H-EJ4 and 4S0V-SUV) used for homology model generation and reference ligands used for generation of pharmacophore (TUG-891 and Compound39) to verify the unbiased nature of the screened compounds. The two-dimensional scaling cluster (Figure 4.13; Table 4.5) confirms that screening of Cpd9 and Cpd7 was not biased towards the templates or reference ligands used in the study.

Table 4.5: Clustering of structures based on distance matrix analysis of 2D structural similarity. (Backman, Cao and Girke, 2011) V1 and V2 denotes distance values derived from all-against-all comparisons of compounds using atom pair similarity scores.

Molecules	cluster	V1	V2
Cpd_1	1	-0.085	0.3574
Cpd_2	1	-0.057	0.3855
Cpd_3	3	-0.122	0.243
Cpd_4	4	-0.225	-0.188
Cpd_5	5	-0.222	-0.205
Cpd_6	5	-0.264	-0.327
Cpd_7	5	-0.265	-0.34
Cpd_8	8	-0.093	0.2649
Cpd_9	5	-0.218	-0.272
4S0V-SUV	10	0.6775	-0.191
4N6H-EJ4	11	-0.024	0.2465
Compound39	10	0.6775	-0.191
TUG-891	13	0.2192	0.215
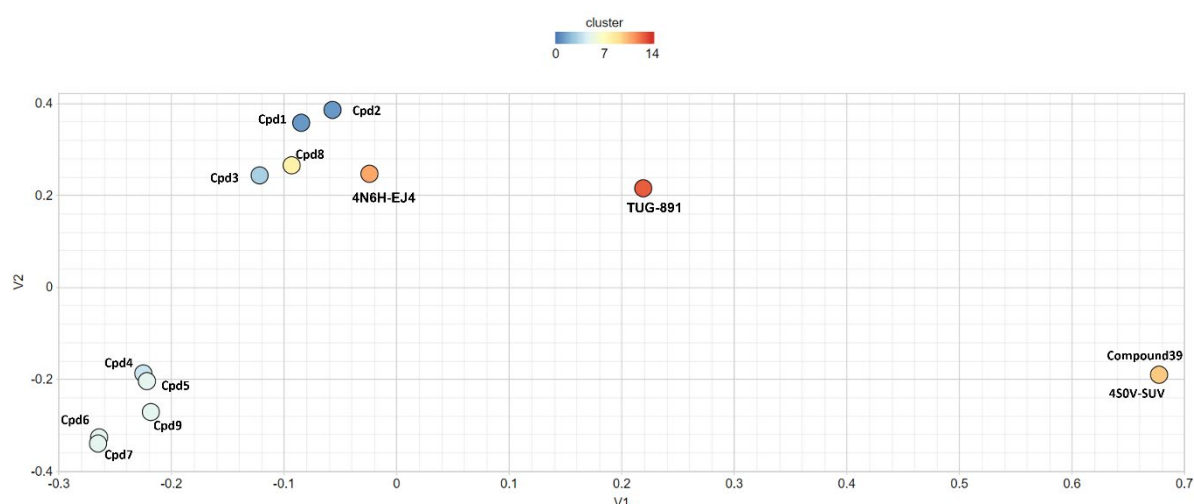
Figure 4.13: Clustering of template ligands (PDBs: 4S0V-SUV and 4N6H-EJ4); agonists (TUG-891 and Compound 39) and selected compounds 1-9 by Two Dimensional similarities with a default similarity cut-off of 0.4 (Backman, Cao and Girke, 2011).

Finally, with the design and discovery of novel scaffolds by *in silico* methods, the pharmacokinetic profile and synthesizability of these compounds can be a limiting factor. The proposed antagonist Cpd9 has a promising predicted ADME profile indicating leadlikeness but with a moderate predicted solubility (see supplementary material). Furthermore, a feasible retrosynthesis scheme obtained from CAS SciFinder (https://scifinder.cas.org) confirms the ease to synthesis of this compound (Figure 4.14).

Step1: Synthesis of substituted salicylanilides under microwave irradiation ; 7 min, rt → 210 °C

Scheme 1 (1 Reaction)
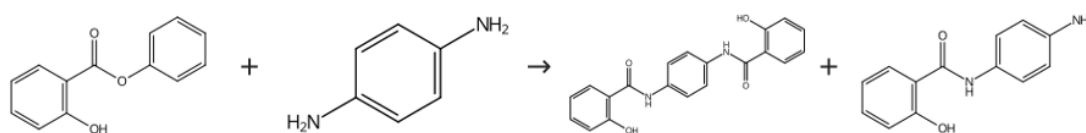
Step2: Pyrimidine derivatives synthesis with Boron tribromide in solvents: Dichloromethane; 0 °C; 0 °C → rt; overnight, rt

Scheme 1 (1 Reaction)

Step3: Preparation of naphthyridinone derivatives with Sodium bicarbonate in Methanol, Water; overnight, 60 °C

Scheme 1 (1 Reaction)
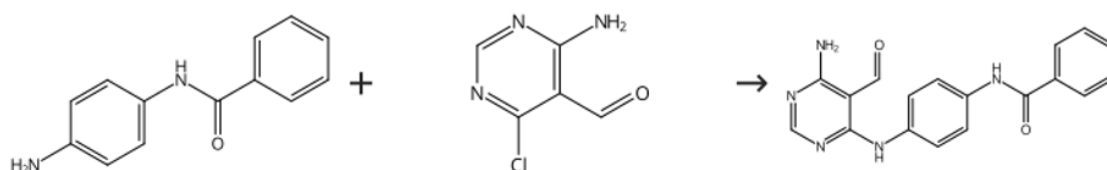
Figure 4.14: Retrosynthesis plan for Cpd9 predicted using CAS SciFinder. (https://scifinder.cas.org)

To summarize the results (Figure 4.15), virtual screening of ZINC database against SBPM targeting Trp277 and Asn313 enlisted 63 hits. Furthering screening analysis by SwissADME and protein-ligand interaction fingerprinting, nine compounds were selected for 100 ns MD simulations. MD analysis predicted compounds 1, 7 and 9 as candidates of interest to be acting as GPR120 agonists (Cpd1) and GPR120 antagonists (Cpd7 and Cpd9).
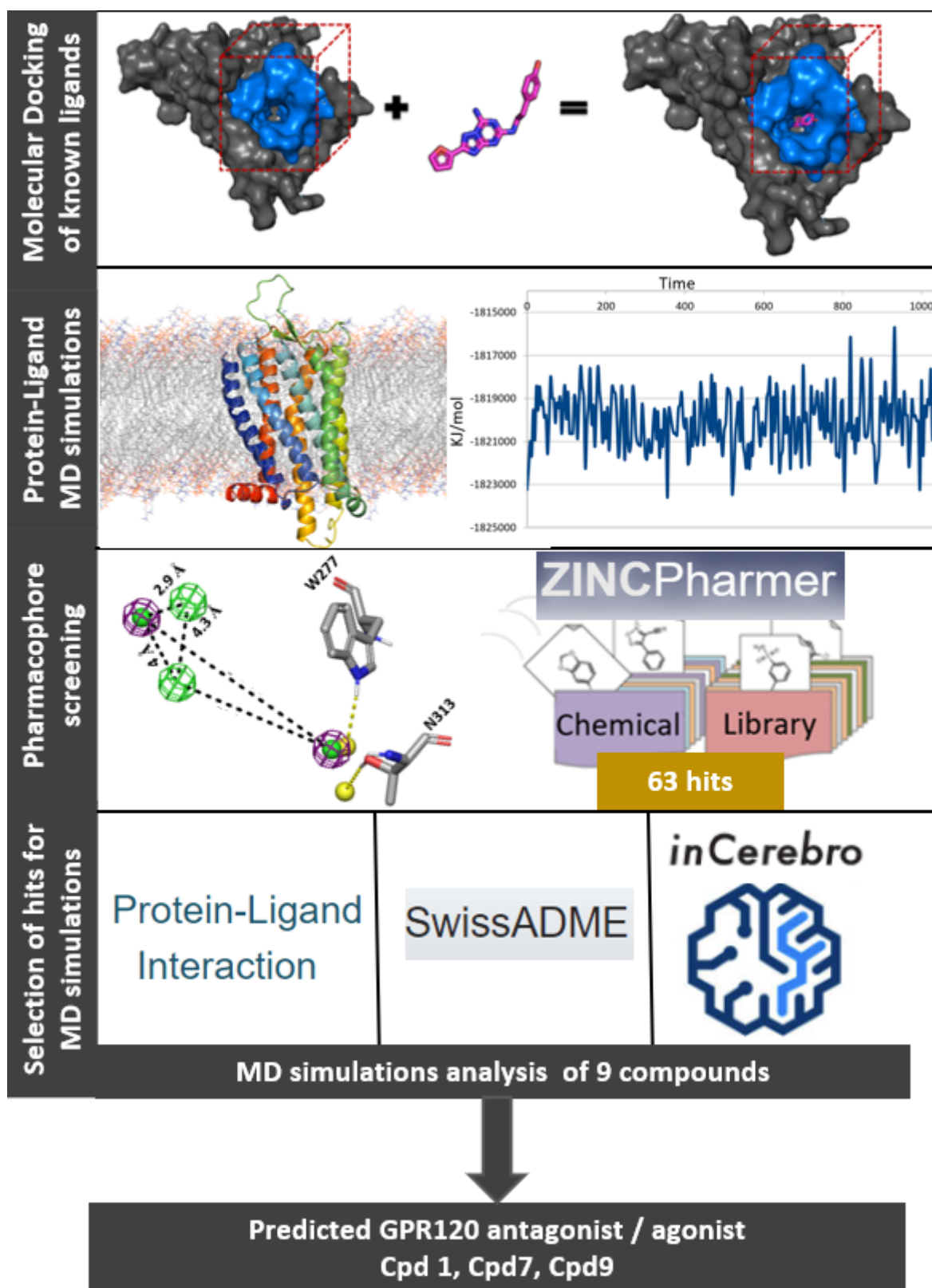
Figure 4.15: Summary of the methodology and results from pharmacophore-based screening of ZINC database.

4.5 Discussion

The conformational sampling by MD simulations enhanced the study of receptor activation using mathematical algorithms and high-performance computers. MD simulations can be set biased to favor the exploration of conformational energy landscape of protein systems and the selection of parameters can be challenging as well as crucial for achieving sufficient sampling at the most efficient computational cost.

The embedding of the protein model in a phospholipid lipid bilayer to mimic the *in-situ* environment of the protein during MD production run is of functional importance (Lee, 2011). The lipid molecules modify the ways that TM helices pack into bundles. The hydrophobic residues orientate towards the lipids of the bilayer and the hydrophilic residues move towards the core of the TM bundle. The lipid molecules penetrate between the helices and bind into clefts between the helices to modulate the activity of a membrane protein (Liang, Adamian and Jackups, 2005; Lee, 2011; Stansfeld and Sansom, 2011; Marrink et al., 2019). The GPR120S model was embedded in a POPC bilayer to relax the TM bundle in the lipid environment to mimic the lipid raft paradigm (Marrink et al., 2019; Lingwood and Simons, 2009). The lipid raft paradigm states that the membrane protein and the bilayer lipids form nanoscale functional units and when the protein (G protein subunits) match with lipid molecules at specific locations they both can contribute to protein activation (Lingwood and Simons, 2009). For future work, GPR120S and G-protein complex can be analyzed for the lipid raft paradigm at microscale MD simulations.

For the best conformational sampling of inactive and active states of GPCRs by MD simulations, comparative studies between agonist bound and antagonist bound protein systems is one of the accepted criteria. As there were no orthosteric antagonists of GPR120 known at the time of this research, the apo protein system was set as the reference point for inactive protein conformation state. Although all the conserved motifs such as [E/D]RY, WxP, NPxxY and PIF, cys-cys disulfide bond between TM3 and ECL2 as well as "ionic lock" interaction of TM3-TM6 observed in the GPR120S apo protein model were consistent with Class A GPCRs inactive state (Figure 4.1, 4.2), it could be a limiting factor in sampling the conformational space of the inactive state. On the other hand, focusing on the primary objective of our research to perform *in silico* screening of chemical databases to discover a series of small molecules which can bind into the orthosteric binding pocket and show pharmacological activity in *in vitro* screening assays (Chapter 5). As mentioned in Chapter 3,

the orthosteric binding pocket of GPR120S model in an inactive or basal activity state (apo model) will be able to facilitate the ligand binding which can either transit the receptor from an inactive to active state or stabilize the receptor in the inactive state.

A recent steered and supervised MD study of lipid receptors (cannabinoid receptor 1, sphingosine-1-phosphate receptor and lysophosphatidic receptor) which employed tabu-like / "forbidden" search algorithm (Jakowiecki et al., 2020) suggested that the large N-terminal domain plugs the ligand binding channel located between extracellular domain of TM1 and TM7 and plays a significant role in the ligand entry and exit into the orthosteric binding site from the extracellular domain. Similarly, the protein-water H-bond networks of GPCRs has been suggested to form an extensive flexible H-bond network that spans the inner core of GPCR bundle and bridge all conserved motifs leading to activation mechanism of GPCRs (Bertalan, Lešnik, Bren and Bondar, 2020).

In the present study, the GPR120S protein was modelled and simulated without N and C termini and the MD simulations performed for the screening of potential of GPR120 ligands were to only validate the docked agonist models and predict conformational changes in conserved motifs induced by screened ligands. They were not steered or supervised to study either entry or exit of the ligands into the orthosteric binding pocket. Further studies can be proposed with models of both short and long isoforms of GPR120 containing N and C termini along with sodium ion bound in the allosteric pocket (Zhou et al., 2019) to improve the pharmacodynamics of the proposed ligands and refine the prediction of activation mechanism. To consider protein-water H-bond for prediction of receptor activation mechanism and virtual screening of potential ligands high resolution crystal structures of GPR120 are preferred to verify the number of internal water molecules. As the templates used for GPR120S homology modelling has <25% sequence identity, it is not recommended as a reliable methodology to superimpose the templates to the energy minimised water solvated GPR120S model to locate the conserved water clusters for MD simulations and predictions.

The study of binding mechanisms of drugs and induced activation pathway of GPCRs require long-time scale all-atom MD simulations to perform conformational state sampling. The enormous system size and energy barriers between various ligand-bound and ligand-free states are difficult to cross and limit the conformation sampling by conventional MD simulations. Enhanced sampling techniques such as accelerated MD (aMD) and Gaussian

accelerated MD (GaMD) (Miao, Nichols and McCammon, 2014), can be applied to overcome this limitation. The aMD sampling technique, as the name suggests, accelerate the conformational sampling by providing a boost potential to the system and decrease the energy barriers allowing easy transition of system from one energy well to another. The aMD simulations can be used to verify the discrepancies observed in Compound39-bound systems behavior from TUG891-bound as if the system was stuck in local minima or the effects were ligand specific. As the aMD simulations generate large energetic noise which can overshadow the results. GaMD follows a near-Gaussian distribution to improve the aMD method (Miao, Nichols and McCammon, 2014) and can be used for further sampling of Compound39 and TUG891 bound systems.

The MD analysis of 300 ns production runs of agonist bound GPR120S models led to the generation of a pharmacophore hypothesis targeting Trp277 and Asn313 residues of GPR120 receptor to discover potential candidates as GPR120 antagonists. The pharmacophore generated on the basis of the Trp277 and Asn313 hypothesis was used to screen ZINC database for potential GPR120 antagonists. The present hypothesis was validated by running MD simulations of pharmacophore identified 9 hits over a period of 100 ns suggesting that H-bond interactions of Cpd 9 (*2-hydroxy-N-{4-[(6-hydroxy-2-methylpyrimidin-4-yl)amino]phenyl}benzamide*) with Trp277 and Asn313 stabilized the occupancy of the ligand in the orthosteric pocket and kept the protein in the inactive form. While the interactions of a phenylimino-phenol analogue - Cpd 7 phased the protein from inactive to active form due to breakage of the ionic lock which could lead to G-protein coupling at the intracellular domain. Cpd 1 warrants further investigation as although it stabilized the "ionic lock" with slight fluctuations at the end of MD run (Figure 4.11c), a large range of fluctuations in the ligand RMSD with respect to protein backbone (> 5 Å) were observed (Figure 4.11b) suggesting weak protein-ligand interactions.

Further *in silico* site-specific mutation studies or *in vitro* alanine scanning studies targeting Asn313 alone as well as in combination with other binding pocket residues could confirm the importance of Asn313 interactions in GPR120 antagonist design. Therefore, the insights from the present study can potentially be employed to enhance the selectivity of GPR120S ligands and target interactions with key residues (Trp277 and Asn313) to develop novel agonists and antagonists.

As the conformational analysis of the 300 ns long MD trajectories of apo and TUG891 bound GPR120S receptor models validated the GPR120S receptor modelled using inactive templates - the energy minimized apo model was selected to screen chemical libraries (Chapter 5) to discover and design GPR120 antagonist or inverse agonist for CRC management.

Chapter 5

In silico and *in vitro* screening of chemical databases for potential anti-cancer candidates targeting GPR120.

Virtual screening (VS) is a*n in-silico* drug discovery tool to predict novel hit compounds by evaluating their high potency and selectivity towards the target protein in a computational model. The strategy of applying 'in silico' screening is to bring a more focused approach to the wet-lab experiments using pharmacophore searches of 3D databases, homology searching and docking (Zhu et al., 2013; Ul-Haq, Uddin and Gul, 2011). Some important points to be considered for virtual screening are - the availability of the compounds to be screened against the receptor; the knowledge about the structure of the receptor and the receptor-ligand interactions; and the knowledge about required drugs and their characteristics (Zhu et al., 2013).

VS can be performed by two approaches: the first is "structure-based drug discovery (SBDD)", which requires knowledge of the 3D structure of the target protein and binding site to increase the success rate (Figure 5.1); and the second is "ligand-based drug discovery (LBDD)", where no information on the protein is necessary. Instead, one or more compounds that are known to bind to the protein are used as a structural query to predict hit compounds for subsequent experimental validation (Zhu et al., 2013).

VS of chemical libraries by SBPMs is focused on specific 3D arrangement of chemical moieties with respect to selective protein-ligand interaction (as applied in Chapter 4). The SBPM model targeting Trp277 and Asn313 residues previously used for screening ZINC database was built on a prediction inferred from MD simulations. Similarly, LBDD (includes ligand-based pharmacophore model) is focussed on physical and chemical similarities to known drug candidates previously proposed to interact with the protein of interest. Both these SBPMs and LBDD approaches can limit the chemical scaffold diversity of the libraries and miss out potential hit compounds (Dhasmana et al., 2019). On the contrary, SBDD is a robust approach to screen the available chemical libraries to predict the best interaction modes between the ligand and the residues of the defined binding pocket of the protein target.
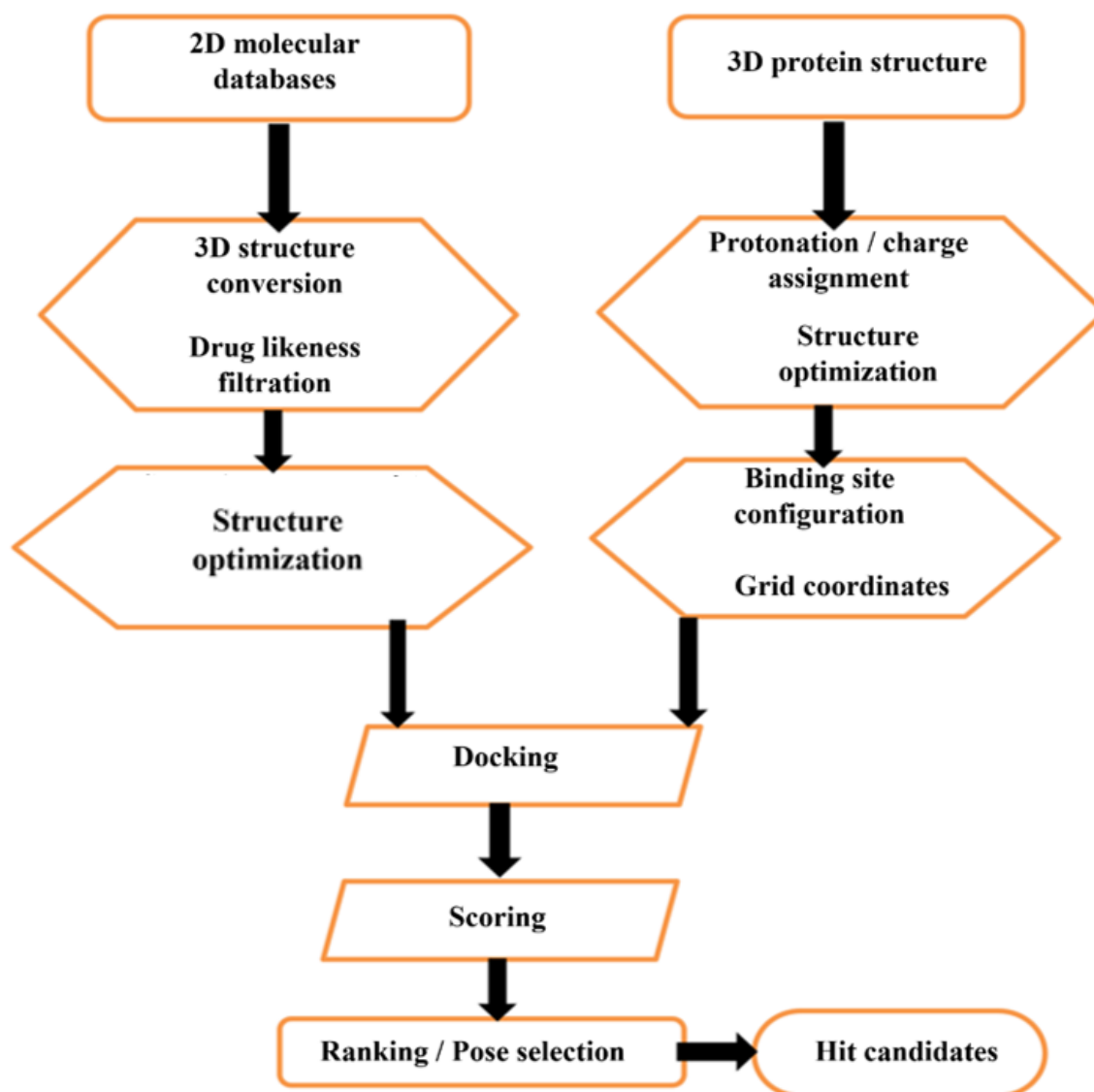
Figure 5.1: General schematic of VS applied for the search of human GPR120S receptor modulators.

The overall objective of this chapter is to apply SBDD methodology to screen potential GPR120 binding ligands and validate the *in-silico* hits. The protocol developed for identifying GPR120S hit compounds was based on the optimised 3D homology model of GPR120S (as described in chapter 3) and the docking protocol (section 2.1.3) summarised in Figure 5.1. Subsequently, identified hits will be evaluated by *in vitro* functional assays such as fluorometric cell viability assay using Alamar Blue, cell viability assay in GPR120-siRNA transfected cells for target identification, wound healing assay and clonogenic assay.

5.1 Chemical databases and drug-likeness

In VS, large chemical databases / libraries such as DrugBank (Wishart, 2006), ZINC (Irwin et al., 2012), ASINEX (www.asinex.com), SPECS (www.specs.net), contain small molecules which are screened for hit compounds by SBDD, LBDD or both. These computational databases require pre-screening filtration to remove unwanted compounds with undesirable or toxic physicochemical properties. The filtered databases should have compounds with favourable characteristics (termed drug-likeness) such as solubility, stability, and the absence of toxic moieties (reactive functional groups / toxicophores) (detailed in section 1.7.4). The filtration process also reduces the library size making VS computationally less expensive as well as increasing the chances of finding new ligands. Some databases (e.g., ZINC) pre-screen compounds for drug-likeness using several rules such as "Lipinski's Rule of Five" (Lipinski, Lombardo, Dominy and Feeney, 2012), which states that drug-like compounds for oral bioavailability should have molecular weight (m.wt.) lower than 500 Daltons, lipophilicity (logP) lower than 5, less than five hydrogen bond donors, and less than 10 hydrogen bond acceptors.

For VS of the GPR120S receptor, the DrugBank and SPECS databases were selected which are freely available online databases. DrugBank is a knowledgebase for drugs and drug targets with comprehensive drug action information, first released in 2006. The latest release of DrugBank (version 5 at the time of study, 2017) (Wishart, 2006; Wishart et al., 2017) was used which contained 2,627 FDA approved small molecule drugs to potentially repurpose the approved drugs as GPR120 targeting therapeutics. The SPECS database (www.specs.net) is another online database of commercially (purchasable) available compounds used, which contains ~350,000 compounds. All the databases were pre-screened for Lipinski's Rule of Five using BIOVIA Pipeline Pilot (Dassault Systèmes 2017) to increase the VS efficiency as the orthosteric binding pocket of GPR120S (Class A GPCRs) cannot accommodate molecules bigger than ~600 Dalton (m.wt.) (Jaiteh, Rodríguez-Espigares, Selent and Carlsson, 2020; Beuming et al., 2015).

5.2 Structure based drug discovery / VS

In this study, the SBDD approach was used to identify the binding pose of each small molecule in a test library (by docking), and from that identify the predicted binding affinity of that molecule (scoring). The set of hit compounds are then predicted by sorting all compounds in the test library by their binding affinity score and deciding on a threshold score. Compounds scoring better than the threshold are regarded as hits and evaluated further.

This approach is analogous to experimental high throughput screening - HTS, where the percentage inhibition obtained from HTS serve the same role as the score in structure-based VS (Ruiz-Torres et al., 2017; Chan and Zhang, 2020). The virtual screening method is fast and economical to predict active compounds.

The validated GPR120S receptor homology model was used as the structural basis for the following database search. As mentioned earlier (section 1.5.2), the essential binding motifs for GPR120S known ligands involves hydrogen bonding interactions between Arg99 (TM2) and Trp277 (TM6), aromatic stacking interactions and hydrophobic interactions with: Trp104 (ECL1), Phe115 (TM3), Trp207, Phe211 (TM5) and Phe304 (TM7). In addition to these core interactions, ligands might be able to extend their interaction network in binding pocket. The set of pre-screened databases with all the possible stereoisomers were screened against the GPR120S target using the docking protocol (section 2.1.3).

For molecular docking, a set of 28 active ligands (Appendix IIb) of human GPR120S was prepared for which the experimental data (EC_{50}) was available to validate the VS protocol (section 3.6; Figure 3.12). The experimental data obtained from previous published studies was not uniform as the type of bioassays varied in all the studies (Lombardo et al., 2016; Sparks et al., 2014; Azevedo et al., 2016; Hudson, Shimpukade, Milligan and Ulven, 2014). It should be noted that the ability of this docking model to successfully predict correlation between binding affinities and experimental activity is a complex function governed by several factors, such as the bioavailability of drug in the experimental models, physicochemical parameters of these compounds and the experimental values (Ki / EC_{50}) used. In this regard, the set docking protocol with a validated consensus scoring (CScore) function (section 2.1.3; section 3.6, 3.8 and Appendix IIc) was used as the pose selection by protein-ligand interactions was difficult for the VS of large chemical databases.

5.2.1 SBDD of the DrugBank database

From screening of the prepared DrugBank database (Wishart, 2006), 73 top scoring poses (cut off of -9.0 with respect to binding score of TUG891 of -9.8) of small molecule approved drugs were obtained. These top scoring poses of drugs were then analysed manually for the interactions with essential binding residues - Arg99, Trp104, Phe115, Trp207, Phe211, Trp277, and Phe304 (section 3.6) as well as based on the known targets of the candidates. The selected 24 compounds (enlisted in Table 5.1) were further characterised based on their

mechanism of actions – if they are known binders of transmembrane receptors and main chemical entity (scaffold) for selection of compounds for further *in vitro* screening assays.

Table 5.1: Selected hits from the top scoring molecules from VS of DrugBank database with GPR120S receptor model and their indication.

Drug – DrugBank ID	SMINA Score	Treatment indications	Mechanism of Action	Reference
Dutasteride (DB01126) Steroidal compound	-13.085	Benign prostatic hyperplasia	5-alpha reductase inhibitor	Keam and Scott, 2008
Lapatinib (DB-1259) Anilino-qiunozoline compound	-12.249	Solid tumours in breast cancer and lung cancer	Tyrosine kinase inhibitor for epidermal growth factor receptor type 2 (HER2/ ERBB2) and epidermal growth factor receptor (HER1/EGFR/EGBB1)	Nelson and Dolder, 2006
Eluxadoline (DB09272) Benzoic acid derivative	-12.180	Diarrhea and abdominal pain; inflammatory bowel disesease	Mu- and kappa-opioid receptor agonist; delta-opioid receptor antagonist	Garnock-Jones, 2015
Lumacaftor (DB09280) Benzoic acid derivative	-11.728	Cystic fibrosis	Acts as a protein-folding chaperone preventing misfolding of CFTR ion channels destruction	Boyle et al., 2014
Sonidegib / LDE225 (DB09143) Biphenyl carboxamide compound	-11.613	Anticancer agent; Hedgehog signalling pathway inhibitor	Smoothened receptor inhibtor – GPCR (Class F)	Zollinger et al., 2014
Adapalene (DB00210) Naphthoic acid derivative	-11.182	Topical retinoid; acne vulgaris; anti-inflammatory	Binds to retinoid X receptors; normalize differentiation of	Kolli et al., 2019

			follicular epithelial cells	
Amrubicin (DB06263) synthetic anthracycline derivative	-11.094	Bladder and gastric carcinoma; lung cancer	Inhibit topoisomerase II; acts as cytotoxic and antimitotic agent	Katou et al., 2008; Maesaka et al., 2019
Nandrolone (DB13169) Androgenic steroid	-10.916	Anabolcic steroids for catabolic states such as burn injuries.	binds to the androgen receptor (agonist)	Ghizoni, Bertelli, Grala and da Silva, 2012
Tasosartan (DB01349) Pyrimidinone compound	-10.908	essential hypertension;	long-acting angiotensin II (AngII) receptor blocker	Elokdah et al., 2002
Dasabuvir (DB09183) Pyrimidinyl-naphthalenyl	-10.871	Antiviral; chronic Hepatitis C	NS5B inhibitor- terminating RNA polymerization	Gentile, Buonomo and Borgia, 2014
Azilsartan medoxomil (DB08822) Benzodiazole compound	-10.844	mild to moderate essential hypertension	Angiotensin II receptor antagonist	Sica et al., 2011; Sood, Bajaj and Bajaj, 2018
Posaconazole (DB01263) Triazole compound	-10.761	Antifungal drug used in invasive Candida and Aspergillus infections.	Impaire the functions of membrane-bound CYP-450 enzyme systems.	Greer, 2007
Gliquidone (DB01252) Sulfonylurea compounds	-10.684	Anti-diabetic drug – Type 2 diabetes mellitus	ATP-dependent potassium (KATP) channel blocker	Furman, 2016
Abiraterone (DB05812) Steroidal progesterone	-10.602	Hormone refractory prostate cancer	Selective and irreversible inhibitor of 17 alpha-hydroxylase	O'Donnell et al., 2004

			(CYP17A1)	
Bazedoxifene (DB06401) indole derivatives	-10.570	Post-menopausal osteoporosis; decreases bone resorption;	oestrogen-receptor modulator (agonist and/or antagonist)	Komm and Chines, 2011
Darifenacin (DB00496) Pyrrolidineacetamide	-10.499	Treatment of urinary incontinence	M3 muscarinic acetylcholine receptor blocker	Steers, 2006
Apixaban (DB07828) Pyridine-carboxamide	-10.439	Thromboembolic diseases - reduce the risk of stroke and systemic embolism	Factor Xa (FXa) inhibitor (clotting factor)	Deshpande, 2012
Ezetimibe (DB00973) Hydroxyphenyl-azetidinone	-10.145	Anti-hyperlipidemic; decreasing cholesterol absorption in the small intestine	Blocks Niemann-Pick C1-Like 1 (NPC1L1) protein on GIT epithelial cells and hepatocytes	Hammersley and Signy, 2016
Canagliflozin (DB08907) Oxane-triol moiety	-10.083	Anti-diabetic	Inhibit the sodium-glucose transport protein 2 (SGLT2)	Jakher, Chang, Tan and Mahaffey, 2019
Levocabastine (DB01106) Piperidine carboxylic acid	-10.129	Selective second-generation Histamine H1-receptor antagonist	allergic conjunctivitis	Dechant and Goa, 1991
Deferasirox (DB01609) Triazolyl benzoic acid	-9.899	Iron chelator; treatment of chronic iron overload due to long-term blood transfusion	Binds to trivalent ferric ion and forms a stable complex	Yang, Keam and Keating, 2007
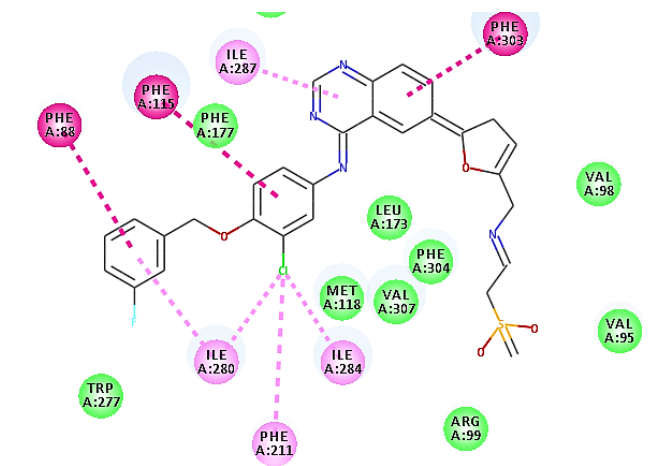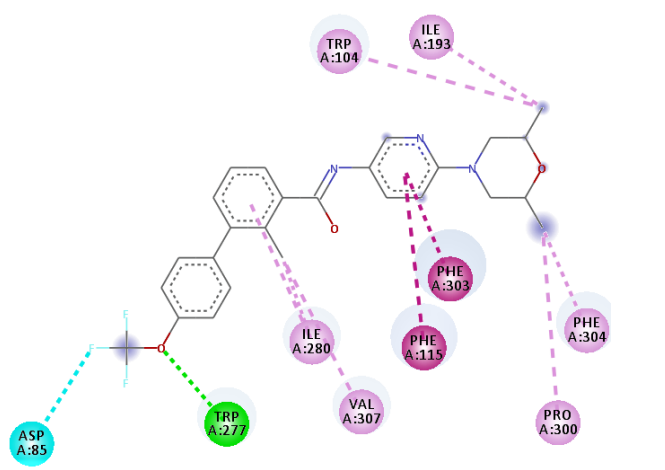
Estradiol (DB00783) Steroid	-9.639	Urogenital symptoms associated with post-menopausal atrophy	Estrogen receptor agonist	Rossouw, 2002
Pipotiazine (DB01621) Phenothiazine	-9.548	CNS depressants and anaesthetics	Antagonist of dopaminergic-receptors, serotonergic-receptors, histaminergic-receptors, M1/M2-receptors	Mustafa, 2016
Enzalutamide (DB08899) Benzamide	-9.533	Treatment of castration-resistant prostate cancer	Androgen receptor inhibitor	Beer et al., 2014

In addition, Table 5.2 shows the predicted GPR120S binding poses of candidate compounds representing different chemical scaffolds – adapalene, azilsartan, gliquidone, lapatinib and LDE225, which were selected for an initial *in vitro* screening assay based on structural diversity, cost, and availability at the time of procurement. Adapalene is a retinoid X receptor used for topical skin applications (Kolli et al., 2019). The retinoid X receptors are nuclear receptors which interact with PPARγ receptors. As mentioned in chapter 1, there is a proposed theory of interaction between PPARγ and GPR120 through EPA (Figure 1.12), screening this compound for GPR120 specific might lead to positive results. Azilsartan is a mild hypertension drug which targets Angiotensin II receptor acting as its antagonist (Sica et al., 2011; Sood, Bajaj and Bajaj, 2018). As Angiotensin II and GPR120 both are rhodopsin-like receptors, which share common features both in TM and orthosteric binding domains, Azilsartan was selected for *in vitro* screening assays. The docked pose of Azilsartan (Table 5.2) was found to have H-bond interactions with Arg99 and Trp277 as well as Asn313 as hypothesized for SBPM for GPR120 antagonist. As the functional assays performed in the present study cannot verify if the ligand is acting as an agonist or an antagonist, Azilsartan can be a proposed as a potential candidate for future studies.

Past studies available throughout the literature suggest the significant involvement of GPR120 isoforms in anti-diabetic therapies (Azevedo et al., 2016). Gliquidone, an anti-diabetic drug (Furman, 2016) was selected for *in vitro* screening from top scoring poses as anti-diabetic compounds are probable binders of GPR120S and are of significant interest as dual therapeutics. Lapatinib is a tyrosine kinase inhibitor for epidermal growth factor receptor type 2 (HER2/ ERBB2) and epidermal growth factor receptor (HER1/EGFR/EGBB1) and plays a significant role in breast cancer tumour therapy (Nelson and Dolder, 2006). Another anticancer agent- LDE225 - Hedgehog signalling pathway inhibitor which acts by inhibiting the smoothened GPCR receptor (Class F) was top scoring ligand in VS (Zollinger et al., 2014). Both the anticancer ligands were also selected for initial *in vitro* screenings.

Table 5.2: Predicted binding modes of the 5 hit compounds selected for *in vitro* screening assays from DrugBank. The 2D interaction maps were generated in BIOVIA DS Client visualizer v19.1.

Test Compound	2D interaction map
Adapalene Docking SMINA Score: -11.182 kcal/mol	
Azilsartan Docking SMINA Score: -10.844 kcal/mol	
Gliquidone Docking SMINA Score: -10.684 kcal/mol	

Lapatinib Docking SMINA Score: -12.249 kcal/mol	
Sonidegib / LDE225 Docking SMINA Score: -11.613 kcal/mol	

5.2.2 I*n vitro* screening of VS hits from the DrugBank database

The stock solutions of procured (section 2.2.1) test drugs were prepared at 10 µM concentrations in 0.05% DMSO / RPMI-1640 culture media. Serial dilutions from 10 to 0.039 µM were tested in SW480 cells according to the set protocol for 24-hour alamar blue cytotoxicity assay (section 2.2.2; Figure 5.2). With the exception of lapatinib (IC$_{50}$ 1.16 to 1.77 µM), all test compounds failed to show strong cytotoxic activity against the SW480 cells. As mentioned earlier, lapatinib is actively used for breast cancer and tumour therapy, such results were expected for lapatinib as well as LDE225. But LDE225 showed negligible cytotoxic effects with IC$_{50}$ values of 21.6 to 68.2 µM, which might be related to the expression of smoothened GPCR receptor in SW480 cells. The expression levels of smoothened GPCR receptor in SW480 are currently unknown.
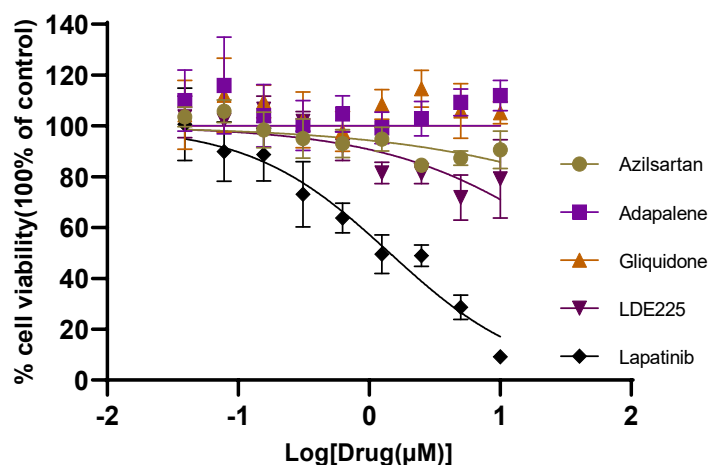
Figure 5.2: Cytotoxicity profile of DrugBank compounds using SW480 cells for 24-hour treatment. Results show mean and standard error of 5 replica samples. Where no error bars are visible, they are obscured by the symbol.

Validated 27mer GPR120 siRNA duplex (from Origene) was used to silence the GPR120 expression in SW480 cell-line. Successful downregulation of the GPR120 was validated by the data obtained from RT-qPCR analysis of GPR120 mRNA levels from siRNA transfected SW480 cells (Figure 5.3a). GPR120 siRNA duplex kit was provided with two different samples of GPR120-siRNA. GPR120 expression was found to be ~36% in siRNNA-2 treated SW480 cells and ~59% in siRNA-1 treated SW480 cells in comparison to untreated (control) SW480 cell line (Figure 5.3a). As the siRNA-2 transfected SW480 cell lines reported the lowest expression (~36%) of GPR120, inferring that siRNA-2 is a more potent GPR120 gene silencer than siRNA-1. siRNA-2 was used to silence GPR120 expression in SW480 cell cultures for further experiments, if not mentioned otherwise. Detailed data and relative fold gene expression calculations delta-delta Ct method for RT-qPCR are provided in appendix VIII.

As the cytotoxic effect of lapatinib observed in the above experiment might be solely due to tyrosine kinase inhibition in HER/EGFR receptors, lapatinib was tested in GPR120-siRNA transfected SW480 cells (Figure 5.3b). The comparative results between GPR120-silenced (IC$_{50}$ 0.61 to 1.21 μM) and control (IC$_{50}$ 1.16 to 1.77 μM) experiments presented no significant change in the cytotoxic activity of lapatinib confirming that the lapatinib is not a GPR120 binder and the observed cytotoxic activity could be due to other protein targets.
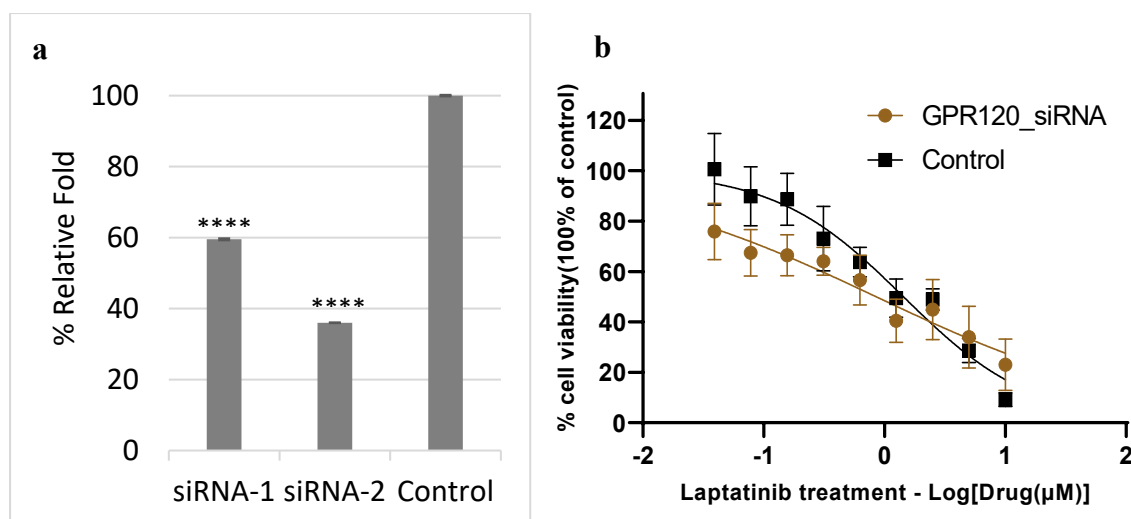
Figure 5.3: a) Percentage relative fold gene expression (by $\Delta\Delta C_T$ method) of GPR120 in GPR120-siRNA transfected SW480 cells within 72-hours of transfection. Results show mean and standard error of 3 replica samples from 2 independent experiments. Where error bars are not visible, they are negligible; b) siRNA-mediated silencing of GPR120 in SW480 cells – Lapatinib was screened against GPR120-siRNA transfected and non-transfected (control) SW480 cells for 24-hour treatment. Results from five replicates are expressed as the mean ± standard error, Percent relative cell viability for all treatments were quantified and normalised to the maximal response induced by vehicle control. Data was analysed by two-way ANOVA followed by Tukey's multiple comparison and asterisk values denote significance (****$P <$ 0.0001).

5.2.3 Design and I*n vitro* screening of Deferasirox and derivatives

Deferasirox, an iron chelator (Yang, Keam and Keating, 2007), was also predicted as one of the VS hits from the DrugBank database. It is a tridentate binder of iron (III), the phenols from two molecules of deferasirox trap one iron atom. Replacing either of the two phenols by a phenyl (removal of a hydroxy group) might result in reduced or negligible iron chelation properties of deferasirox (Figure 5.4). The 3D conformation of deferasirox was found to be similar to AH7614 – a negative allosteric modulator of GPR120 (Watterson et al., 2017; Lombardo et al., 2016). It was hypothesized that removal of a single hydroxyl from one of the phenols of deferasirox might strip the molecule's iron chelation property and act as a novel GPR120 allosteric modulator or binding competitor of AH7614. It was proposed that the flexible di-phenyl triazol moiety of the analogue compared to the xanthene moiety of AH7614 would enable the molecule to adapt to the 3D space in the allosteric binding pocket (location not currently elucidated) and act as a strong binder.
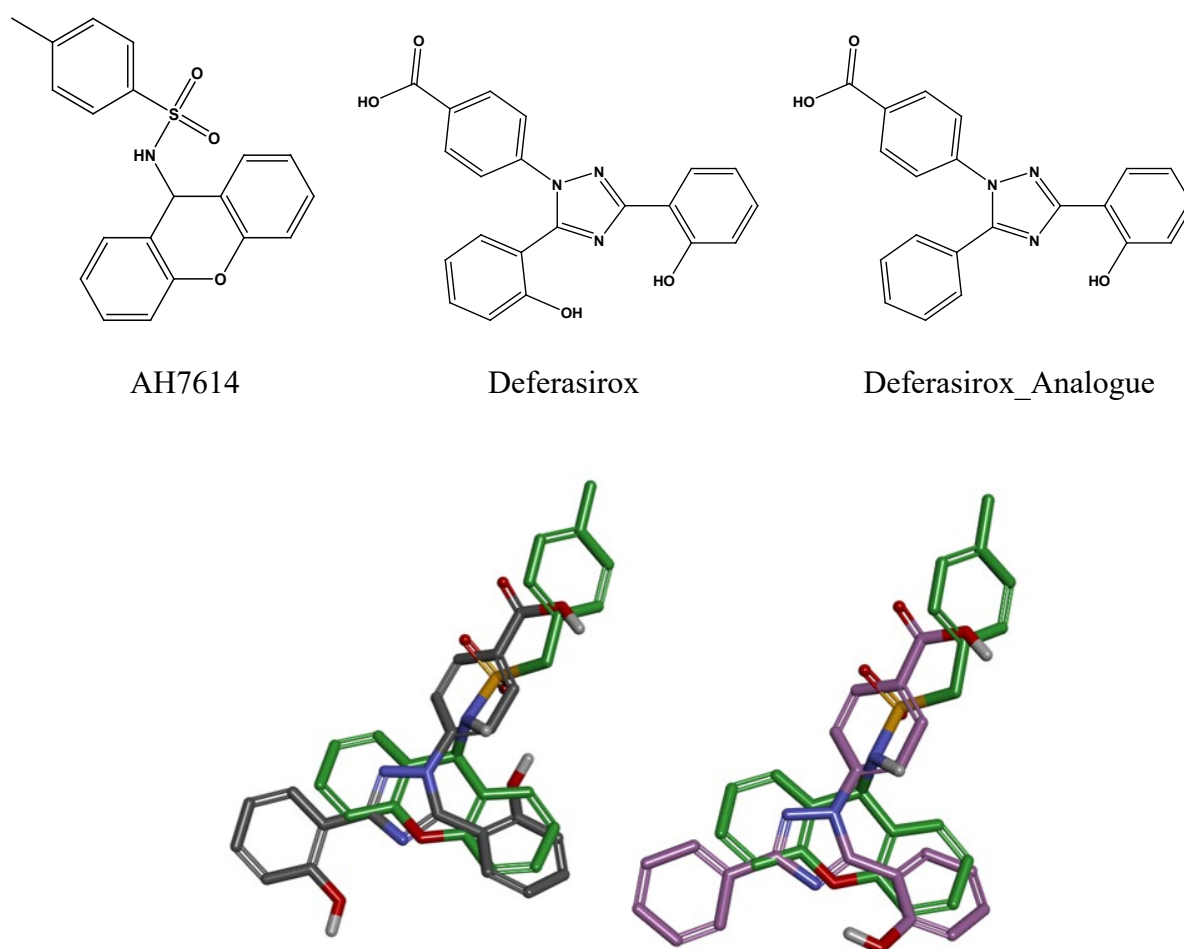
Figure 5.4: Structures of AH7614, deferasirox and deferasirox_analogue, a proposed model of a novel GPR120 allosteric modulator. 3D molecular overlay of AH7614 (Green) against Deferasirox (Grey) and Deferasirox_analogue (Pink) was generated using consensus method with flexible alignment of rotatable bonds in BIOVIA DS Client visualizer v19.1 2019.

The proposed deferasirox_analogue was synthesized by Dr Gráinne Hargaden (Hargaden Chemistry Lab, TU Dublin). TUG891, AH7614, deferasirox and deferasirox_analogue were tested *in vitro* in GPR120-siRNA transfected and non-transfected SW480 cells by alamar blue cytotoxicity assay (Figure 5.5). The cytotoxic effects of deferasirox (50 µM) were reduced from 80% to ~30% (deferasirox_analogue) by replacing the phenol with a phenyl group. It was notable that deferasirox (50 µM) exhibited ~40-50% cytotoxicity even in the GPR120_siRNA transfected cells, while the deferasirox_analogue did not show any cytotoxic activity in GPR120_siRNA transfected cells showing that the deferasirox_analogue might be active through the GPR120 receptor. Additional experiments are required to confirm whether the effects are due to orthosteric or allosteric binding at GPR120 receptor.

AH7614 and TUG891 did not show significant changes in the cytotoxicity or cell proliferation rates. TUG891 treated non-transfected SW480 cells registered a slight (~30%) increase in the proliferation rate, while no such increase in proliferation rate was observed in GPR120_siRNA transfected cells suggesting that TUG891 induced GPR120 activation can result in increased rate of CRC cell proliferation. A similar study in bovine granulosa cells (Maillard et al., 2018) reported 2-fold increase in cell proliferation when treated with TUG891 (50 µM). On the contrary, another study in prostate cancer cells (Liu et al., 2014) showed that TUG891 as the most potent inhibitor of proliferation induced by endogenous GPR120 ligands. The ambiguous role of TUG891 induced stimulation of GPR120 in cell proliferation with respect to various cell types is still not clearly defined.

The above experiments and observations show that none of the selected compounds from the DrugBank database were able to induce cytotoxic effects through GPR120 in SW480 cells. While deferasirox (iron chelator) and the analogue of deferasirox modified to mimic and compete against the GPR120 NAMD AH7614 was found to be a significant cytotoxic agent acting through GPR120 receptor.
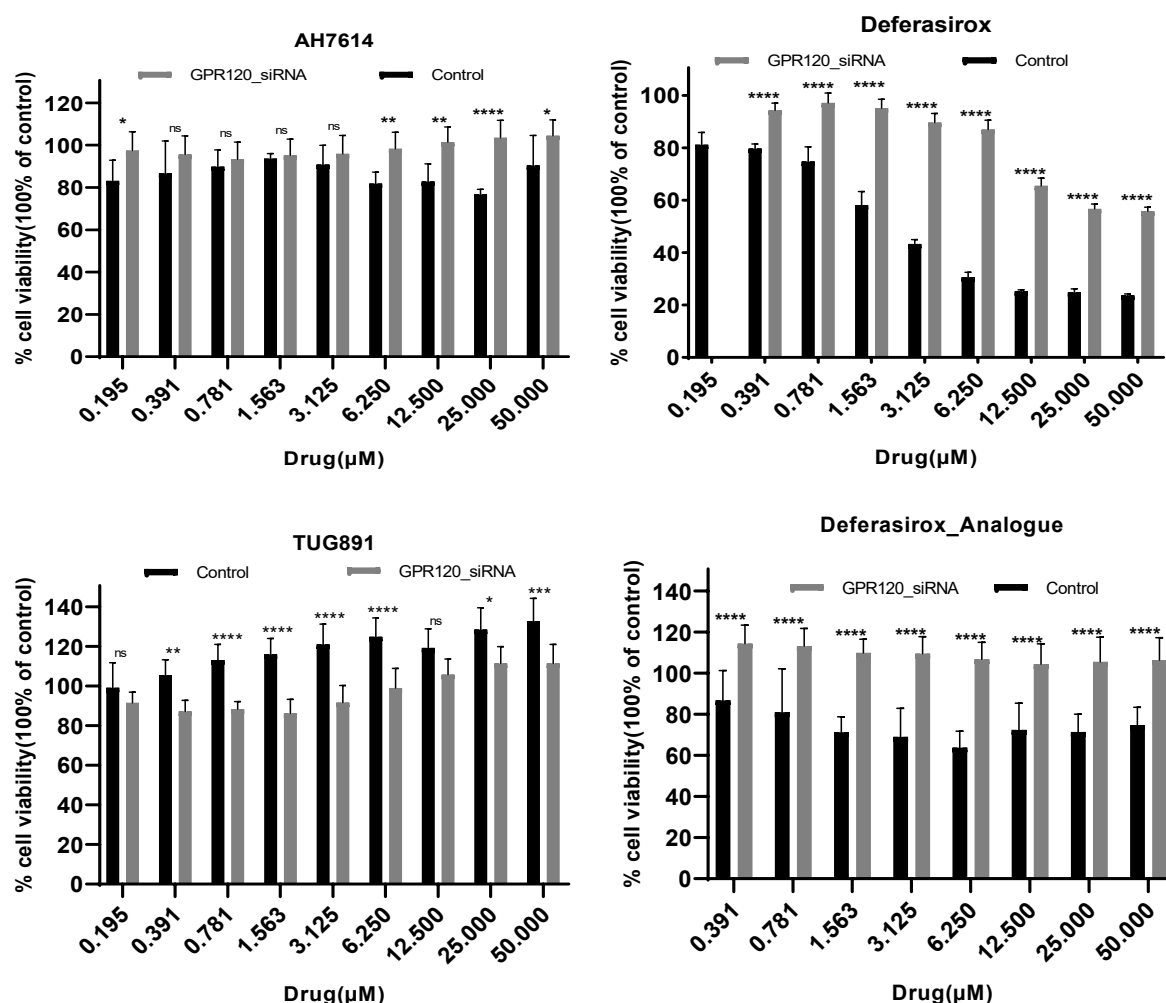
Figure 5.5: Cytotoxicity profile of AH7614, Deferasirox, TUG891 and Deferasirox_analogue using GPR120-siRNA transfected and non-transfected (control) SW480 cells for 24-hour treatment. Results show mean and standard error of 5 replica samples from two independent experiments. Where no error bars are visible, they are obscured by the symbol; Results from replicates are expressed as the mean ± standard error, Percent relative cell viability for all treatments were quantified and normalised to the maximal response induced by vehicle control. Data was analysed by two-way ANOVA followed by Tukey's multiple comparison and asterisk values denote significance (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$).

5.3 SBDD of the SPECS database

For virtual screening, the SPECS database (www.specs.net) containing ~350,000 commercially available, well-characterised and drug-like molecules was screened against the GPR120S homology model. AutoDock SMINA was used as the molecular docking algorithm and the docked poses were rescored using an in-house consensus scoring function from

Equation I (see methods). The best scoring poses docked into the receptor were manually evaluated by comparison with the docked pose of TUG891.

The docking-based virtual screening with a C_{score} cut off was set lower than -9 (comparable to the reference ligand TUG891 of -9.8) resulted in ~66,000 compounds. These were further analysed manually using PyMol Open-source version 2.1.0 (DeLano, 2018) to enlist molecules for phase I of the *in vitro* screening through Alamar Blue cytotoxicity assay (section 2.2.2). The manual docking-pose analysis resulted in 13 compounds (Table 5.3) based on their diverse scaffold chemistry. As per the docking evaluation, these compounds are predicted to bind to the orthosteric binding pocket of GPR120S (Figure 5.6) as well as having similar hydrophobic and / or electrostatic interactions with one or more of the residues reported essential for the pharmacological activity of the receptor.

Table 5.3: Virtual-HTS hit compounds for GPR120S identified from the SPECS database with scores from each scoring function and the consensus score (Cscore).

SPECS_ID	VINA	NNScore	DLScore	SMINA	Cscore
AN-970/40920574	-14.079	10.616	7.580	-14.082	-11.589
AK-968/41925665	-14.150	8.477	8.173	-14.154	-11.238
AO-299/41877474	-13.332	9.334	8.266	-13.334	-11.066
AE-848/32608035	-13.030	9.560	8.375	-13.030	-10.999
AN-970/40920575	-12.821	9.985	7.643	-12.825	-10.818
AG-690/40104520	-13.103	8.887	8.141	-13.103	-10.809
AJ-292/40857565	-13.176	8.324	8.034	-13.178	-10.678
AN-758/14707017	-12.559	9.557	7.362	-12.558	-10.509
AK-968/15252756	-12.586	8.380	8.211	-12.583	-10.441
AO-081/14456496	-11.786	9.224	7.783	-12.258	-10.263
AK-968/12713190	-12.337	8.540	7.689	-12.337	-10.226
AB-131/42301549	-12.271	8.472	6.879	-12.852	-10.119
AG-690/12137150	-12.267	6.848	7.074	-12.268	-9.614
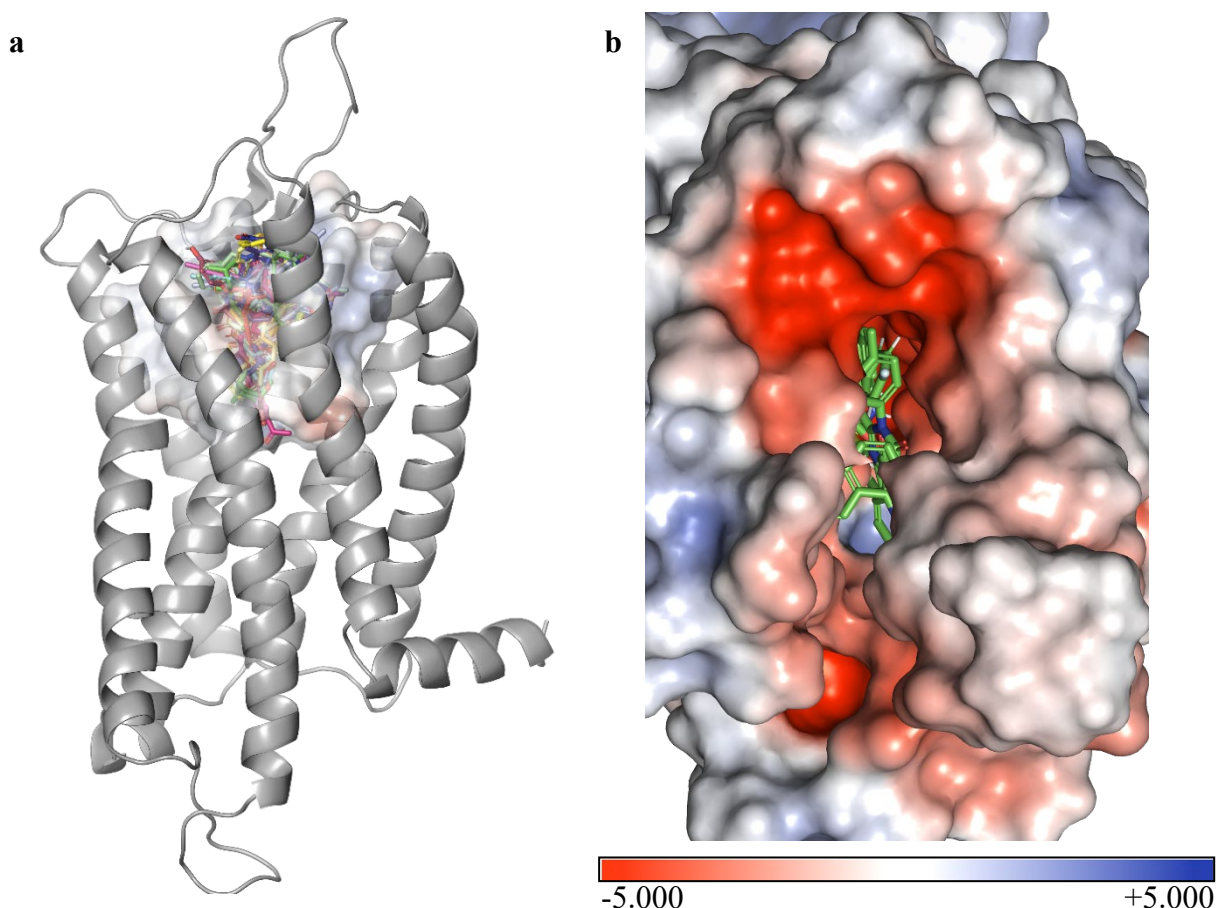
Figure 5.6: Docking simulation of 13 test molecules with surface topology of the GPR120S orthosteric binding pocket (a). Zoomed in view of the electrostatic potential molecular surface of the orthosteric binding pocket as viewed from above (b) was calculated with APBS (Adaptive Poisson-Boltzmann Solver) plugin in PyMOL. Blue denotes a positively charged surface; red denotes a negatively charged surface. The bound test molecules are shown as green stick models. The 3D images were visualized and rendered in PyMol v2.1.0.

5.3.1 *in vitro* screening of VS hits from the SPECS database

The selected 13 compounds from the *in silico* VS were evaluated for their potential anticancer activity by an alamar blue cytotoxicity assay (section 2.2.2) using SW480 cells expressing GPR120 (Wu et al., 2013). For initial screening, the SW480 cell-line was treated with three concentrations (100, 10 and 1 µM) of each test compound for 72 hours to confirm if extended treatment at lower concentrations (1 µM) result in significant cytostatic effects (Kummar, Gutierrez, Doroshow and Murgo, 2006). As illustrated in Figure 5.7a, most of the compounds (11 out of 13) displayed null to negligible (~30%) cytotoxic or cytostatic effects against SW480 cells at the highest tested concentration of 100 µM. However, two of the test

compounds showed significant (> 90%) inhibitory effects on cell growth with micromolar affinities (at 100 μM), reported in Figure 5.7a.
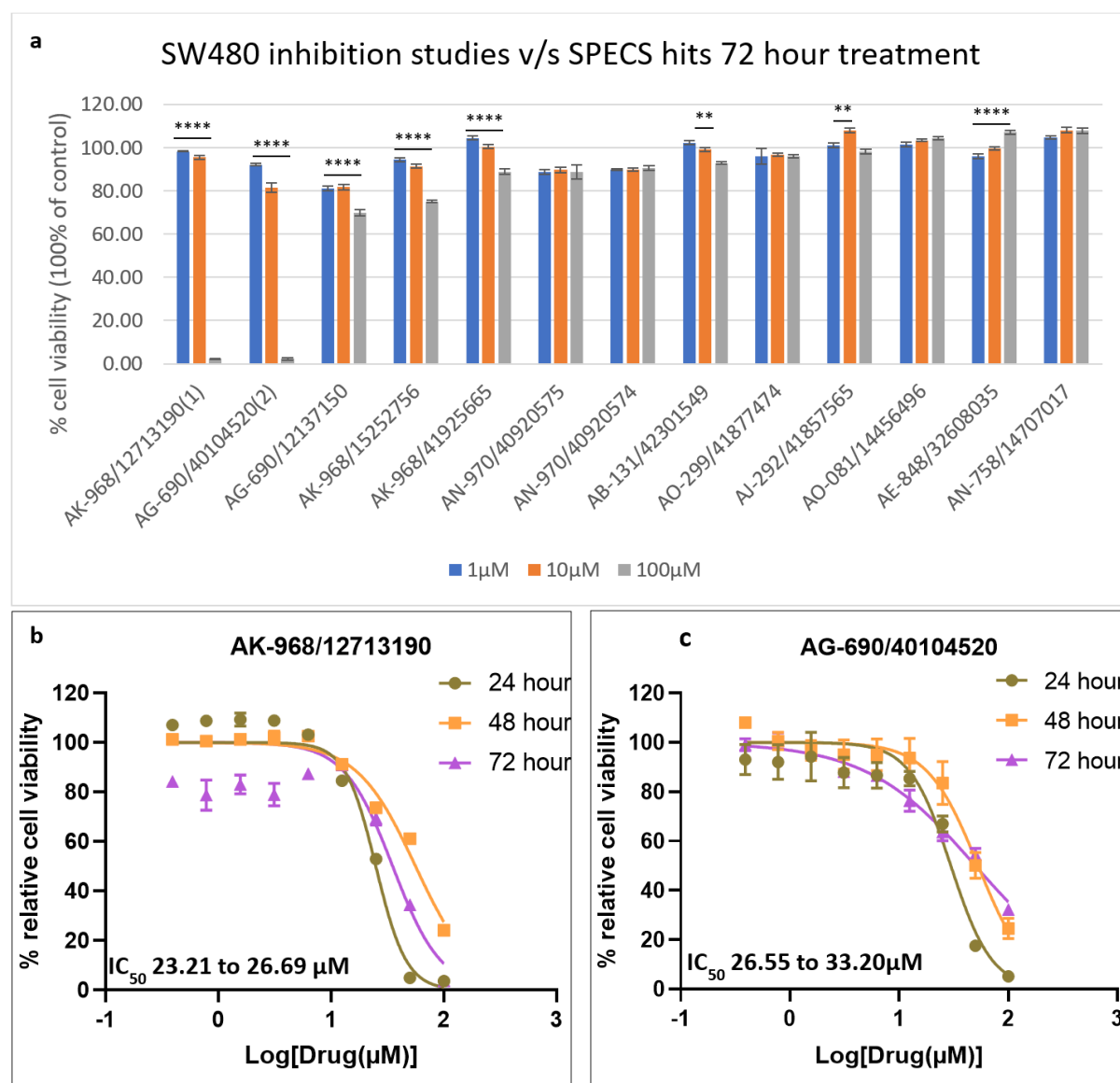
Figure 5.7: a) Cytotoxicity assay of test compounds in SW480 cells which express GPR120 at three concentrations 100, 10 and 1 μM. Results from six replicates are expressed as the mean ± Standard error. The cytotoxicity of b) AK-968/12713190 (experimental 24-hour IC$_{50}$ 23.21 to 26.69 μM) and c) AG-690/40104520 (experimental 24-hour IC$_{50}$ 26.55 to 33.2 μM) was assayed by using SW480 cells using 9 serial dilutions from 100 μM to 0.39 μM at three different treatment time periods. Results show mean and standard error of 5 replica samples. Results are representative of three individual experiments. Where no error bars are visible, they are obscured by the symbol. Percent relative cell viability for all treatments were quantified and normalised to the maximal response induced by vehicle control. Data for (a) was analysed by two-way ANOVA followed by Tukey's multiple comparison and asterisk

values denote significant differences between 100, 10 and 1 µM treatment for each compound (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$).

The two active compounds - AK-968/12713190 and AG-690/40104520, were further tested at a wider concentration range to construct a dose-response curve and determine their IC$_{50}$ values (see Figure 5.7 b, c). Both the test compounds were active in cell line measurements with modest inhibitory activity at different treatment times. The 24-hour drug treatments indicated IC$_{50}$ values of 23.21- 26.69 µM for AK-968/12713190 and 26.55 - 33.2 µM for AG-690/40104520.

As the Dose-response-time (DRT) can highlight the dose-response patterns over time in pharmacological studies (Gabrielsson, Andersson, Jirstrand and Hjorth, 2018), these two test compounds were tested over a longer treatment time of 48 hours and 72 hours for their cytotoxicity activity (see Figure 5.6b, c). After 48 hours, the IC$_{50}$ of both compounds drops to ~50-60 µM. Similar effects over time have been reported in another oncogenic study (Zakharia et al., 2017) suggesting that the cells might have developed acquired resistance to test compounds at lower concentration after 48 hours of exposure. This acquired resistance might enable them to escape the cytostatic state and start cell proliferation which can be traced back to the augmented chemoresistance in breast cancer treatment through GPR120 overexpression (Wang et al., 2019). Also, the increased metabolic activity of cancer cells can be related to anticancer drug metabolism responsible for the resistance to cytotoxic agents (Cree, 2011; Iyanagi, 2007), hence reducing the cytotoxicity of these two compounds over time. While the 72-hour experiments showed a slight decrease of 10 µM in IC$_{50}$ values of both compounds, it should be noted that the drug concentrations were not replaced over the treatment time intervals. The decreased cell growth or increased cytotoxic effects of these two compounds at 72 hours might be the result of a lack of nutrients and increased metabolic waste in the culture solution (Ackermann and Tardito, 2019).

5.3.2 SAR / similarity search and *in vitro* screening of SAR compounds

The top-scoring docked poses of AK-968/12713190 and AG-690/40104520 (Figure 5.8) predicted that the two molecules interact with several residues reported significant for protein-ligand binding by Hudson et al. 2014 and which also interacted with the selected docked pose of TUG891 (Figure 3.11) such as Ile280, Ile284, Val307. AK-968/12713190 consists of a benzo-quinazoline ring structure as the chemical scaffold with smaller benzyl-methyl and benzyl substituents. The phenylalanine residue at TM3 (Phe115) shows strong π-

π interactions with the main scaffold as well as π-sulfur interactions with the sulfanyl linker. While AG-690/40104520 consists of a 9-fluorenone as the chemical scaffold with symmetric dimeric naphthalene groups at both ends linked by an aminosulfonyl. The presence of dimeric naphthyl substituents showing strong hydrophobic π-π stacked interactions at one end and simple π-Sigma interactions at the other suggesting strong binding interactions in the binding pocket. The chemical scaffold of AG-690/40104520 - (PubChem CID: 10241) is actively used in preparation of antimalarial drugs, functional polymers, and dyes (9-Fluorenone, 2021).
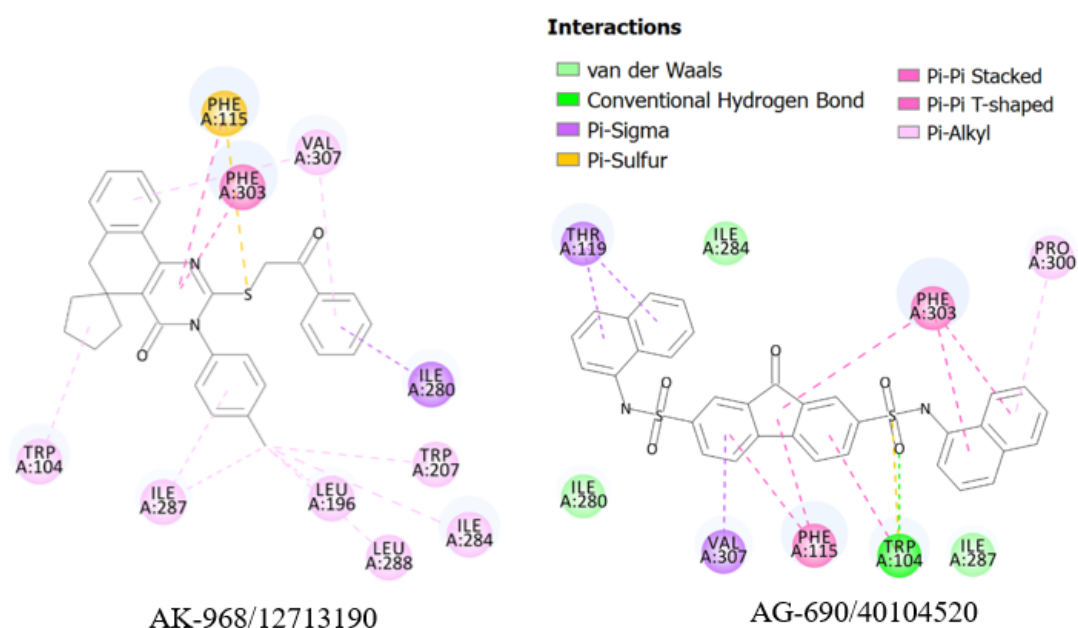
Figure 5.8: 2D interaction maps of docked poses of compounds AK-968/12713190 and AG-690/40104520 with hGPR120S. The 2D interaction maps were generated in BIOVIA DS Client visualizer v19.1 2019.
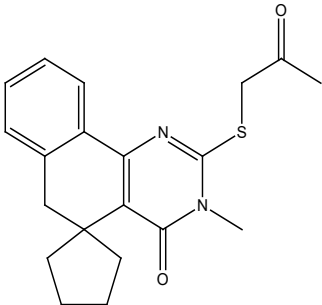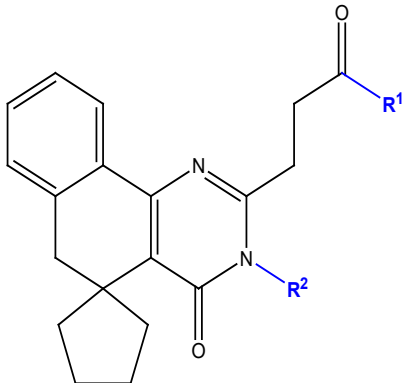
These two compounds were selected for SAR studies based on the strong docking predictions with the GPR120S model, their novelty with respect to the literature and the micromolar cytotoxic activity in CRC cell line. For phase II of *in vitro* screening, the chemical scaffold of the two most cytotoxic test compounds (from phase I) was used for a substructure search from the SPECS database docked pool using Discovery Studio's Pipeline Pilot from Dassault Systèmes 2017.

5.3.2.1 SAR profiling of AK-968/12713190

To expand the SAR profile, the generic chemical structure of AK-968/12713190, *dihydrospiro(benzo[h]quinazoline-5,1'-cyclopentane)-4(3H)-one*, in combination with sulfanyl acetone tail was used as a query for a substructure search against the pre-processed SPECS database using Biovia pipeline pilot (Table 5.4).

The substructure search of AK-968/12713190 resulted in 16 compounds from the pre-screened SPECS database. This set of compounds explored R-groups in position R^1 and R^2 of the scaffold (Table 5.2) in combination with *in silico* ADME profiling using SwissADME (http://www.swissadme.ch/) (Daina, Michielin and Zoete, 2017) to procure the selective compounds for *in vitro* screening. The analogues were selected exploring the alkyl to aryl substitutions at the R^2 position and simple halobenzene substitutions at the R^1 position connected by a sulfanyl acetone linker keeping *dihydrospiro(benzo[h]quinazoline-5,1'-cyclopentane)-4(3H)-one* scaffold structure intact. Based on the docking analysis, manual screening, and the availability of compounds at SPECS, seven compounds were tested for *in vitro* cytotoxicity assay in SW480 cell line following the set protocol (Figure 5.9).

Table 5.4: SAR profile of AK-968/12713190 with modified groups to determine functional potency. Cytotoxic activity (experimental 24-hour IC_{50}) of SAR compounds in SW480 cell line measured by Alamar Blue assay. Lipinski filters provided by SPECS database.

dihydrospiro(benzo[h]quinazoline-5,1'-cyclopentane)-4(3H)-one

Generic SAR structure

SPECS Compound ID	-R¹	-R²	Docking Cscore	Experimental IC_{50} (µM)	Lipinski's Violations
AL-281/36997030	1,4-C_6H_4Cl	-CH_3	-9.339	22.92 to 27.58	0
AJ-292/12930007	1,4-C_6H_4Cl	-C_5H_9	-9.741	24.26 to 26.95	2 (MW 505; logP 4.57)
AL-281/36997031	1,4-C_6H_4Cl	-C_6H_5	-10.706	5.890 to 6.715	2 (MW 513; logP 4.03)
AG-690/12134207	1,4-C_6H_4Cl	-C_2H_4-C_6H_5	-10.213	68.92 to 84.21	2 (MW 541; logP 4.32)
AL-281/36997034	1,4-C_6H_4Cl	1,4-C_6H_4-O- CH_3	-9.668	6.789 to 7.502	2 (MW 543; logP 4.38)
AN-512/12673388	1,4-C_6H_4Br	1,4-C_6H_4-CH_3	-10.704	25.54 to 28.87	2 (MW 571; logP 4.77)
AN-512/12674229	1,4-C_6H_4Br	2,2-(CH_3)₂-C_5H_7O	-9.988	N/A	2 (MW 593; logP 4.5)
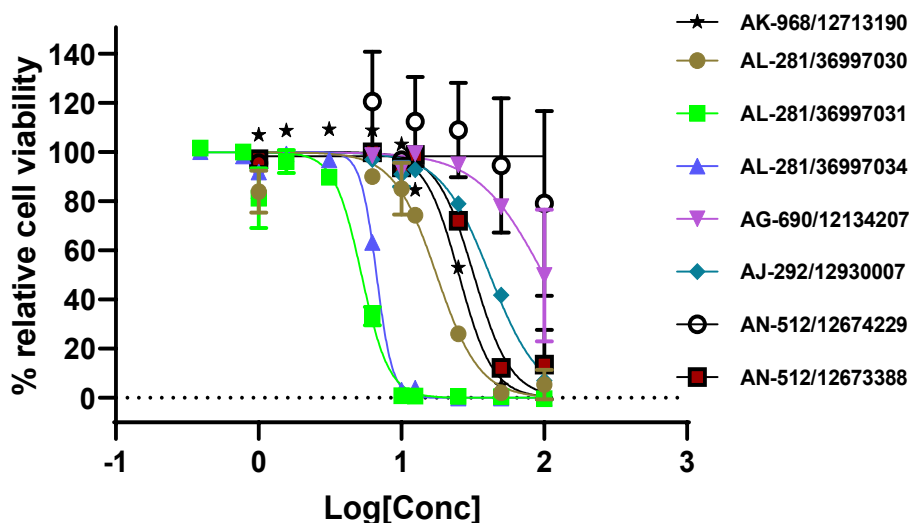
Figure 5.9: Cytotoxicity profile of compound AK-968/12713190 SAR analogues using SW480 cells for 24-hour treatment. Results show mean and standard error of 4 replica samples. Where no error bars are visible, they are obscured by the symbol; Percent relative cell viability for all treatments were quantified and normalised to the maximal response induced by vehicle control.

Based on the SAR study, the IC_{50} values of the new hits enable an initial identification of the essential pharmacophore features required in the *dihydrospiro(benzo[h]quinazoline-5,1'-cyclopentane)-4(3H)-one* scaffold. The presence of an aromatic halogen at the sulfanyl acetone tail exhibited an increase in potency (AL-281/36997030, AL-281/36997031 and AL-281/36997034). However, the substitution of chloride in comparison to bromide seems to be more effective for pharmacological activity. The drop in potency of AN-512/12673388 with respect to the parent compound (AK-968/12713190) and AL-281/36997031 can be related to the larger atomic size of bromine atom compared to chlorine which can result in decreased solubility and hence lowering the bioavailability of the compounds. Although not predicted in molecular docking which showed similar binding scores −10.22, −10.706 and −10.704 (compounds AK-968/12713190, AL-281/36997031 and AN-512/12673388 respectively), the greater size of bromine might be responsible for steric clashes with neighbouring residues in the binding pocket which can be further analysed by future MD studies. Fluorine atom is smaller in size and exhibit slight electronegativity which has been reported to increase the electrostatic bonding affinity of the compounds (Khosravan, Marani and Sadeghi Googheri,

2017). Incorporation of fluoroaromatics at the sulfanyl acetone tail may increase the solubility and hence bioavailability of these SAR analogues.

Substitution of the benzo-methyl at the R^2 position in the parent structure and AN-512/12673388 by a smaller methyl group (AL-281/36997030) or an aromatic six-membered group (AL-281/36997031) resulted in a significant increase in potency of the analogues. The substitution of a non-aromatic cyclic group (AJ-292/12930007) resulted in reduced activity of the parent compound (AK-968/12713190). The addition of a methoxy group to this aromatic ring at R^2 position in AL-281/36997034 resulted in the second most active compound of the SAR profiling. The quinazoline ring linked to five or six membered aromatic ring structures at R^2 position by a single C-C bond length seems to be the optimum as when the linker length in AL-281/36997031 is increased (-C-C_2H_4-C-), the cytotoxic activity of AG-690/12134207 registered a drastic decrease from ~ 6 µM to ~ 80 µM. The total inactivity of AN-512/12674229 may confirm the above inferences as it contains bromo-aromatic group at R^1 position and non-aromatic cyclic ring with a longer linker at the R^2 position.

5.3.2.2 SAR profiling of AG-690/40104520
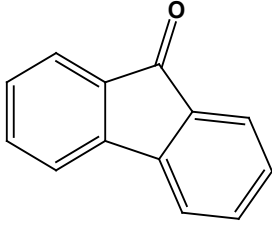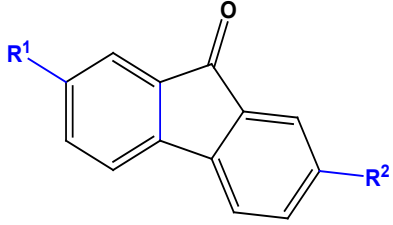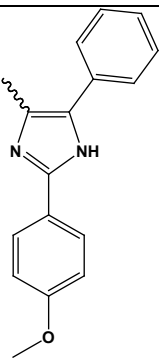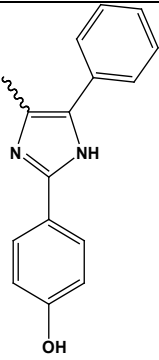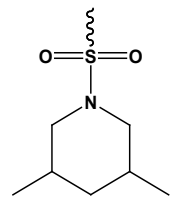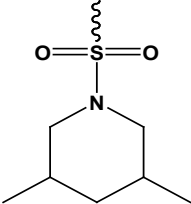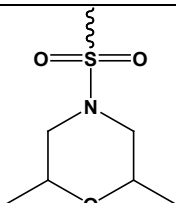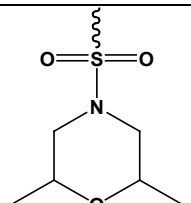
Similarly, the generic chemical structure of AG-690/40104520, *fluoren-9-one*, was used as a query for a substructure search. The substructure search of *fluoren-9-one* from the pre-screened SPECS database resulted in 28 hits, out of which 8 compounds (Table 5.5) were procured for *in vitro* cytotoxicity screenings based on the docking analysis and *in silico* ADME profiling using SwissADME (Daina, Michielin and Zoete, 2017). Substitutions at the R^1 and R^2 positions of the *fluoren-9-one* substructure were explored to build a SAR profile (Figure 5.10). The IC$_{50}$ (µM) values obtained from the cytotoxicity assay (Table 5.5) suggested that the presence of the sulfonamide linker and/or identical substitutions at the R^1 and R^2 positions did not exhibit significant changes in potency of SAR compounds. Since the only two active compounds in the SAR study were AG-205/11944202 (non-identical substitutions without the sulfonamide linker) and AP-845/40876799 (identical substitutions with sulfonamide linkers).
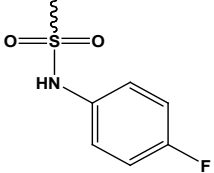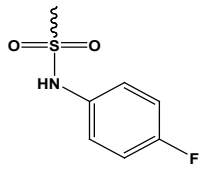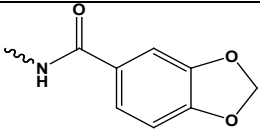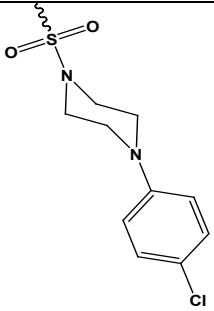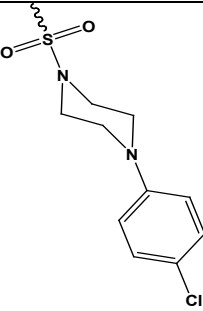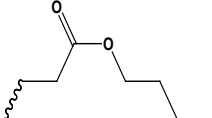
The substitution of methoxy-phenyl by phenol group in the imidazole R^2 groups of compounds AG-205/11945004 and AG-205/11944202 increased the cytotoxicity from negligible (>1446 µM) to ~40 µM. The drastic increase in cytotoxicity by replacing the -OCH$_3$ (methoxy phenyl) with -OH (phenol) suggested that the bigger hydrophobic -OCH$_3$ group might be having strong steric clashes which impacted the binding or entry of AG-

205/11945004 while the hydrophilic -OH group eased the binding of AG-205/11944202 in the orthosteric pocket of GPR120 receptor. Further removal of the second phenyl group from the imidazole ring of AG-205/11944202 could be proposed to reduce the hydrophobicity as well as size at the R^2 substitution which might increase the cytotoxicity profile of R^2 imidazole substitution analogues.

Amongst symmetrical compounds with identical R^1 and R^2 substitutions in the presence of the sulfonamide linker, except AP-845/40876799 no other compounds (AP-845/40876779, AG-690/11665662 and AG-690/11665659) exhibit cytotoxicity activity. AP-845/40876799 was found to be the best hit compound compared to the parent compound (AG-690/40104520) with improved IC_{50} values from ~32 μM to 18 μM. The inactivity of compound AG-690/11665659 can be attributed to the bulky phenyl-piperazine groups preventing the ligand-receptor binding or the strong lipophilic character (logP 6.58) might have resulted in the ligand getting trapped (dissolving) in the plasma membrane.

Table 5.5: SAR of AG-690/40104520 with modified groups to determine functional potency. Cytotoxic activity (experimental 24-hour IC$_{50}$) of SAR compounds in SW480 cell line measured by Alamar Blue assay. Lipinski filters provided by SPECS database.

fluoren-9-one			Generic SAR structure		
SPECS Compound ID	**-R^1**	**-R^2**	**Docking Cscore**	**Experimental IC$_{50}$ (µM)**	**Lipinski's Violations**
AG-205/11945004	-H		-11.533	>1446	1 (logP 6.99)
AG-205/11944202	-H		-11.467	38.73-44.8	1 (logP 6.43)
AP-845/40876779			-11.218	N/A	2 (MW 530; logP 6.25)
AG-690/11665662			-11.125	N/A	1 (MW 534)

AP-845/40876799			-11.102	**14.16-18.02**	2 (MW 526; logP 5.43)
AO-080/43378433	-H		-10.053	>480	0
AG-690/11665659			-9.389	N/A	2 (MW 697; logP 6.58)
AG-219/37040030	-H		-8.802	>571.6	1 (logP 5.39)
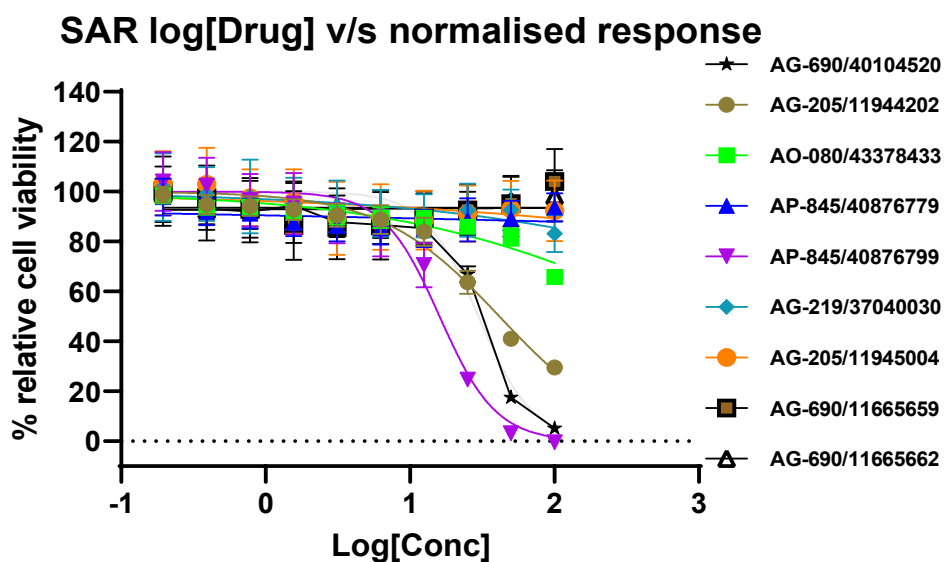
Figure 5.10: Cytotoxicity profile of compound AG-690/40104520 SAR analogues using SW480 cells for 24-hour treatment. Results show mean and standard error of 4 replica samples. Where no error bars are visible, they are obscured by the symbol; Percent relative cell viability for all treatments were quantified and normalised to the maximal response induced by vehicle control.

Compounds AP-845/40876779 and AG-690/11665662 are approximately of the same size as compound AP-845/40876799 (Figure 5.11), with a difference of aromatic ring structure and a rotatable -N-C- bond between fluorophenyl and sulfonamide. Absence of both the aromatic ring structure linked by a rotatable linker might be the contributing factors for the inactivity of AP-845/40876779 and AG-690/11665662, where the strong π–π (or weaker Sigma-π) interactions between the aromatic phenyl ring and the aromatic residues in the receptor binding pocket might be stabilising the protein-ligand binding. While compound AO-080/43378433 with aromatic benzo-dioxole at R^2 position linked by an amide bond showed negligible cytotoxicity (> 480 μM). It is a well-studied fact that rotation is not permitted about amide bonds but with allowed torsion about –(C=O) – (benzo-dioxole) as well as (*fluoren-9-one*)–NH- bonds (Fischer, 2000). The rigidity of the amide bond in AO-080/43378433 might be the limiting factor, therefore, suggesting that the presence of an aromatic ring attached to the main scaffold by a rotatable linker might be essential for substitution groups. While the smallest and the most flexible compound AG-219/37040030 also showed negligible (>571.6 μM) cytotoxic activity as it lacks the aromatic groups at R^2 position in comparison to AO-080/43378433.

The present SAR study explored some of the chemical space around the two parent compounds AK-968/12713190 and AG-690/40104520 and resulted in identification of two analogues of AK-968/12713190 (AL-281/36997031 and AL-281/36997034) and one analogue of AG-690/40104520 (AP-845/40876799) with improved cytotoxicity activities (Figure 5.11). The pharmacokinetics profiling of these five test compounds (Table 5.6) showed that they are poorly soluble in the aqueous phase which could result in low gastrointestinal absorption. The lower gastrointestinal absorption of these compounds is predicted to reduce the oral bioavailability of these compounds.

The profiling also enlisted violations of different rules set for the compounds to have drug-likeness. Recent drug discovery has shown marketed drugs breaking some of these rules which has resulted in many extensions to the Lipinski's Rule of Five (Congreve, Carr, Murray and Jhoti, 2003; Jhoti, Williams, Rees and Murray, 2013). The newer drug-likeness rules such as Ghose's, Veber's, Muegge's filters, etc., (Table 5.6; enlisted in section 1.7.4) choose lower parameter values for drug-like and fragment-like compounds because during ligand / lead optimisation the values of the parameters such as molecular weight, hydrophobicity, rotatable bonds, etc. increase inevitably (Brogi, 2019). As the SAR analysis of the above compounds also laid the blueprint for scaffold design as well as selection of new

compounds for *in vitro* screenings, further optimisation by SAR profiling of these compounds is required to improve their pharmacokinetics. The improved pharmacokinetics might result in more potent and selective cytotoxic agents targeting GPR120 receptor.
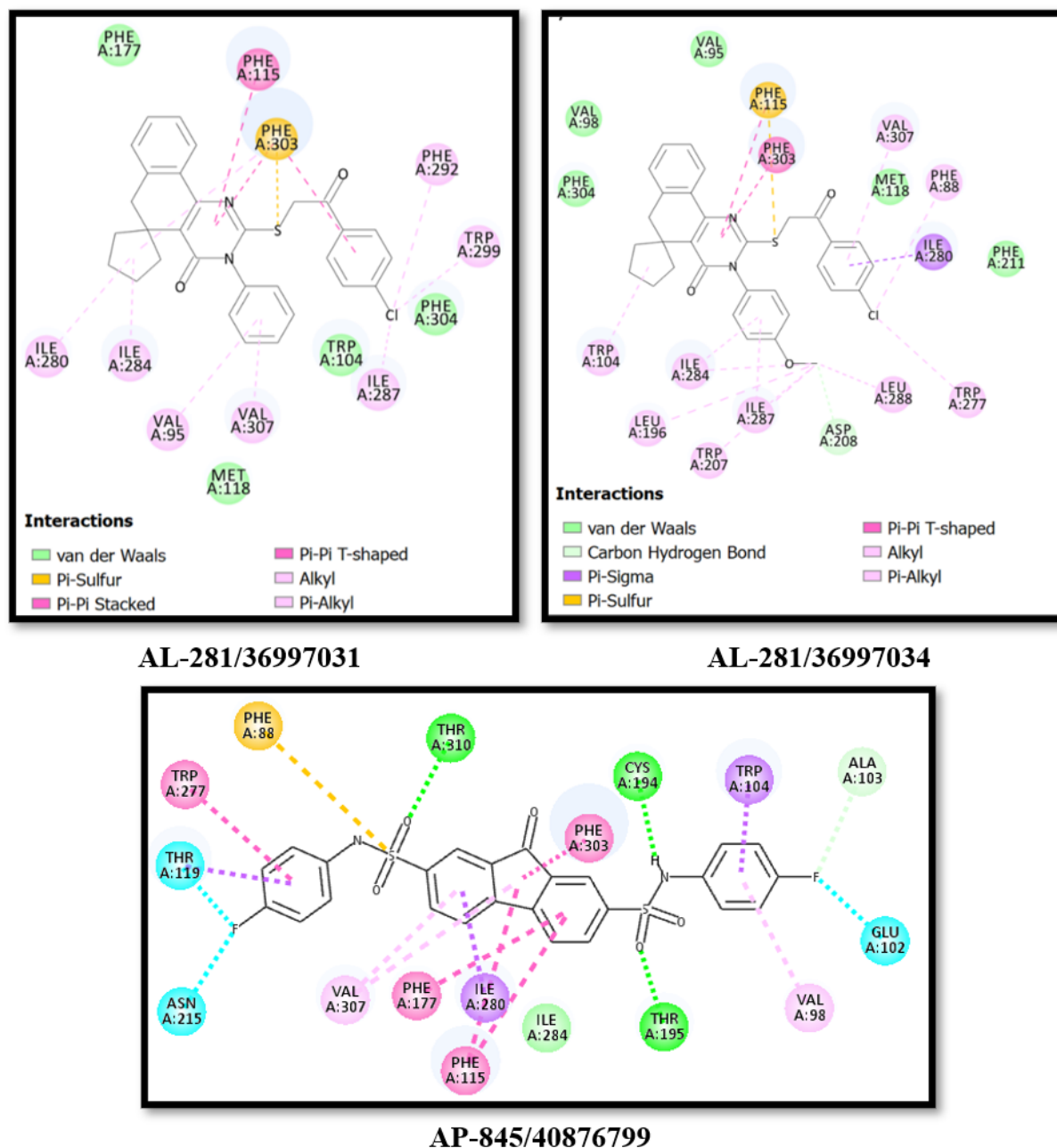
Figure 5.11: 2D interaction maps of docked poses of compounds AL-281/36997031 and AL-281/36997034 and AP-845/40876799 with hGPR120S. The 2D interaction maps were generated in BIOVIA DS Client visualizer v19.1 2019.

Table 5.6: Physicochemical and pharmacokinetics parameters of the potential anticancer two parent compounds and their SAR hits obtained from SwissADME drug-likeness profiling (Daina, Michielin and Zoete, 2017). Color code – Green : Highly favourable; White : Favourable; Red : Unfavourable;

Physicochemical properties	AK-968-12713190 (Parent1)	AL-281-36997031 (SAR hit)	AL-281-36997034 (SAR hit)	AG-690-40104520 (Parent2)	AP-845-40876799 (SAR hit)
Molecular Weight	492.63	513.05	543.08	590.67	526.53
Rotatable bonds	5	5	6	6	6
H-bond acceptors	3	3	4	5	7
H-bond donors	0	0	0	2	2
TPSA	77.26	77.26	86.49	126.17	126.17
iLOG Po/w	4.44	4.03	4.38	2.92	2.17
Pharmacokinetic properties					
GI absorption	Low	Low	Low	Low	Low
BBB permeant	No	No	No	No	No
P-glycoprotein interaction	Yes	Yes	Yes	No	No
CYP1A2 inhibitor	No	No	No	Yes	No
CYP2C19 inhibitor	No	Yes	Yes	No	Yes
CYP2C9 inhibitor	Yes	Yes	Yes	No	Yes
CYP2D6 inhibitor	No	No	No	No	No
CYP3A4 inhibitor	No	No	No	Yes	Yes
Skin permeant log Kp (cm/s)	-4.04	-3.98	-4.18	-5.21	-6.45
Lipinski violations	1	2	2	1	1
Ghose violations	3	3	3	3	2
Veber violations	0	0	0	0	0
Egan violations	1	1	1	1	1
Muegge violations	1	1	1	1	0
PAINS alerts	0	0	0	0	0
Brenk alerts	0	0	0	0	0
Lead likeness violations	2	2	2	2	2
Synthetic Accessibility	4.62	4.5	4.57	3.82	3.31
Estimated Aqueous Solubility	Poorly soluble	Poorly soluble	Poorly soluble	Poorly soluble	Moderately soluble

5.3.3 Target validation of hit compounds by siRNA cell transfection.

As cancer therapies need to be target specific to limit their general toxicity to healthy cells and prevent obstruction to normal cellular homeostasis (Padma, 2015). The next step after identification of three potential hit compounds (AL-281/36997031, AL-281/36997034 and AP-845/40876799) (Figure 5.11) was to confirm that the cytotoxic compounds were exhibiting their effects by targeting the intended GPR120 receptor, hence showing target specificity. Scientific literature and public databases such as DrugBank, ChEMBL, ZINC, etc., (Wishart, 2006; Papadatos and Overington, 2014; Irwin et al., 2012) can be used to identify drug targets in cases where the potential test compounds are well researched or already known. For novel test compounds, the process of regulating the target expression *in vitro* and quantifying the drug response in regulated and unregulated environment can be used for target identification.

The three test compounds - AL-281/36997031 and AL-281/36997034 were tested in GPR120 silenced SW480 cells at approximate experimental IC_{50} (5 µM) and AP-845/40876799 was tested at approximate experimental IC_{50} (15 µM) (Figure 5.12). The cytotoxicity of AL-281/36997031 and AP-845/40876799 was observed to be significantly suppressed in GPR120-siRNA transfected SW480 cells compared to non-transfected SW480 cell line. AL-281/36997034 presented a slight change in its cytotoxicity properties in GPR120-siRNA transfected SW480 cells. The comparative results between GPR120-silenced and control experiments suggested that AL-281/36997031 and AP-845/40876799 were GPR120 specific as they exhibit their cytotoxic effects via GPR120 while AL-281/36997034 might be exhibiting cytotoxic activity through multiple targets including GPR120. TUG891 (10 µM) presented neither cytotoxic nor proliferative activity and was equivalent to the vehicle. Also, no significant change was observed with TUG891 in control GPR120_siRNA transfected cell lines.
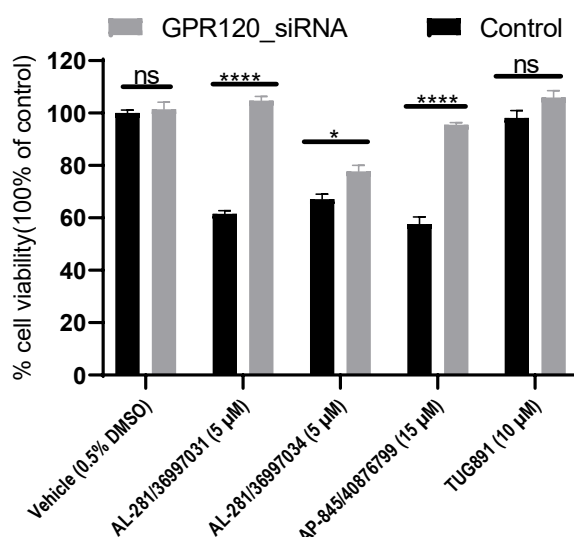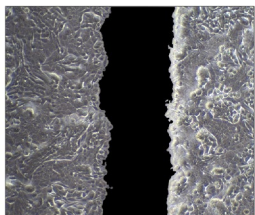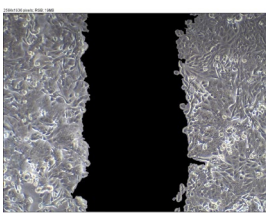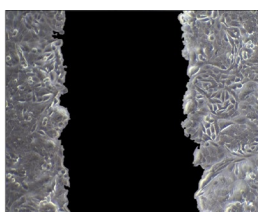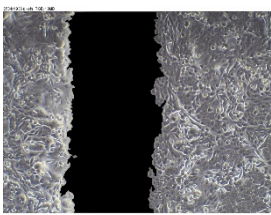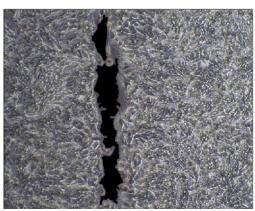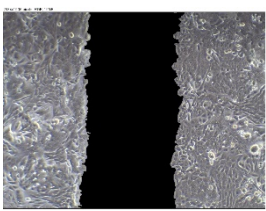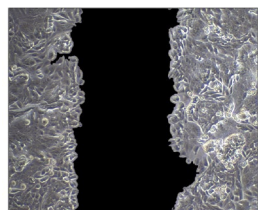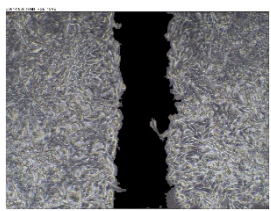
Figure 5.12: siRNA-mediated silencing of GPR120 in SW480 cells – Compounds AL-281/36997031, AL-281/36997034 and AP-845/40876799 were screened against GPR120-siRNA and scrambled-siRNA (control) transfected SW480 cells. Results from five replicates are expressed as the mean ± standard error, ****$p<0.0001$, ***$p=0.0003$, *$p=0.0420$, ns => not significant as indicated (from two-way ANOVA, Sidak's multiple comparisons test). Percent relative cell viability for all treatments were quantified and normalised to the maximal response induced by vehicle control.

5.3.4 Wound healing assay of test compounds

Cell migration is a hallmark of wound repair, cancer invasion, metastasis, angiogenesis, etc (Arnold, Opdenaker, Flynn and Sims-Mourtada, 2015). Wound healing is a complex cellular process involving dynamic interactions and crosstalk between intra- and extra-cellular matrix molecules (Arwert, Hoste and Watt, 2012). Analysis of cell migration *in vitro* is useful to quantify alterations in cell migration in response to various factors or treatments. Cancer cell-lines (SW480) can migrate at a higher rate compared to normal cell-lines and promote tumour invasion, tumour angiogenesis and metastasis (He et al., 2018). Wound healing or *in vitro* scratch assay is a well-developed inexpensive, simple, and versatile methodology to quantify the migration capacity of the cells. It can be real-time image monitoring or time-interval image monitoring of cell migration to heal the wound i.e., scratched cellular monolayer (He et al., 2018; Somchai et al., 2020). The percentage scratch area recovered overtime in the presence of cytotoxic drug treatments per experiment measures the migration rate. The image analysis of scratch area by image capturing and analysis via software tools like Fiji with

Wound healing tool automate the process (Schindelin et al., 2012) and reduce the human error.

In this study, the three cytotoxic hits (AL-281/36997031, AL-281/36997034 and AP-845/40876799) identified from previous results were tested for their ability to inhibit cell migration in SW480 cell-line at subtoxic concentrations (from approximated IC_{50} values – Table 5.4 and 5.5) obtained from their dose-response curves. Representative images of mechanical scratch wound in the absence or presence from timepoints 0 and 16 hours are illustrated in Figure 5.13a. The experimental data (Figure 5.13b) illustrated that treatment with AL-281/36997031 and AL-281/36997034 (scaffold: *dihydrospiro(benzo[h]quinazoline-5,1'-cyclopentane)-4(3H)-one*) caused significant inhibition of cell migration in comparison to the compound AP-845/40876799 and control. The time interval of 16 hours was selected based on the trial experiments of control to find the optimum range where mechanical wound scratch would not be completely recovered. The inhibitory effects of test compounds agreed with their cytotoxicity profiling results where AL-281/36997031 was found to be the most active of the three test compounds.

a) Vehicle Control 0.5% DMSO / RPMI+10%FBS	AL-281/36997031 Subtoxic Conc. (IC$_{25}$) 3 μM (IC$_{50}$ 5.890 - 6.715 μM)	AL-281/36997034 Subtoxic Conc. (IC$_{25}$) 3.5 μM (IC$_{50}$ 6.789 - 7.502 μM)	AP-845/40876799 Subtoxic Conc. (IC$_{25}$) 7.5 μM (IC$_{50}$ 14.24 to 18.02 μM)
%RSA = 82.23%	%RSA = 2.38%	%RSA = 8.13%	%RSA = 41.48%
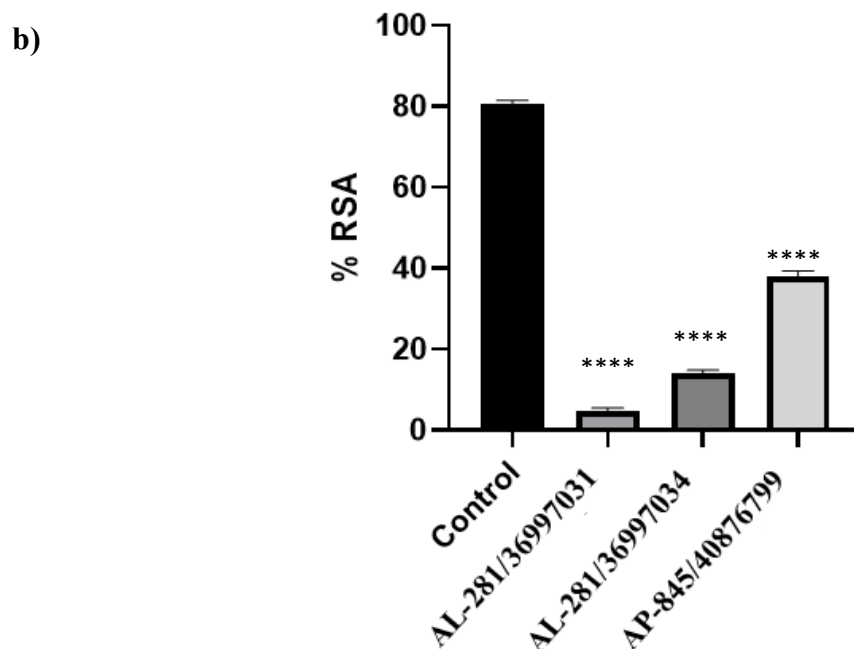
Figure 5.13: a) Representative images of scratched SW480 monolayer captured at time 0 (Top) and 16 (bottom) hours against Control (0.5% DMSO / RPMI+10%FBS) and treatment of three test compounds AL-281/36997031, AL-281/36997034 and AP-845/40876799; b) Bar plot illustrating scratch assay results from 3 independent experiments performed in triplicates

with standard error of mean. Rate of migration of SW480 cells in presence of drug treatments calculated as percentage recovered surface area (%RSA). Data was analysed by two-way ANOVA followed by Sidak's multiple comparisons test and asterisk values denote significance (****$P < 0.0001$).

5.3.5 Colony formation or clonogenic assay

Colony formation / clonogenic assays were developed to determine the ability of a single cell to form a colony of cells (He et al., 2018; Franken et al., 2006), especially stem cells can form colonies. In tumours, only a few cells retain this property and after migration and metastasis initiate tumorigenesis (He et al., 2018). A clonogenic assay was used to screen the three test compounds AL-281/36997031, AL-281/36997034 and AP-845/40876799 in a dose-dependent manner. The survival fraction (Figure 5.14) calculated from the clonogenic assay revealed that treatment with test compounds decreased the colony forming ability of SW480 cells in comparison to the control (0.5% DMSO/RPMI1640). Due to the lower number of cells per well (5 cells/well) the survival rate at concentrations higher than IC_{50} was negligible and colony formation was observed only below IC_{50} values in each drug treatment (Figure 5.14a). AL-281/36997031 was found to be the most potent of all three test compounds keeping the survival rate ~ 30% at 3 µM. The results indicated that` the selected test compounds possess cytotoxic function and can be developed as potential anticancer therapeutic.
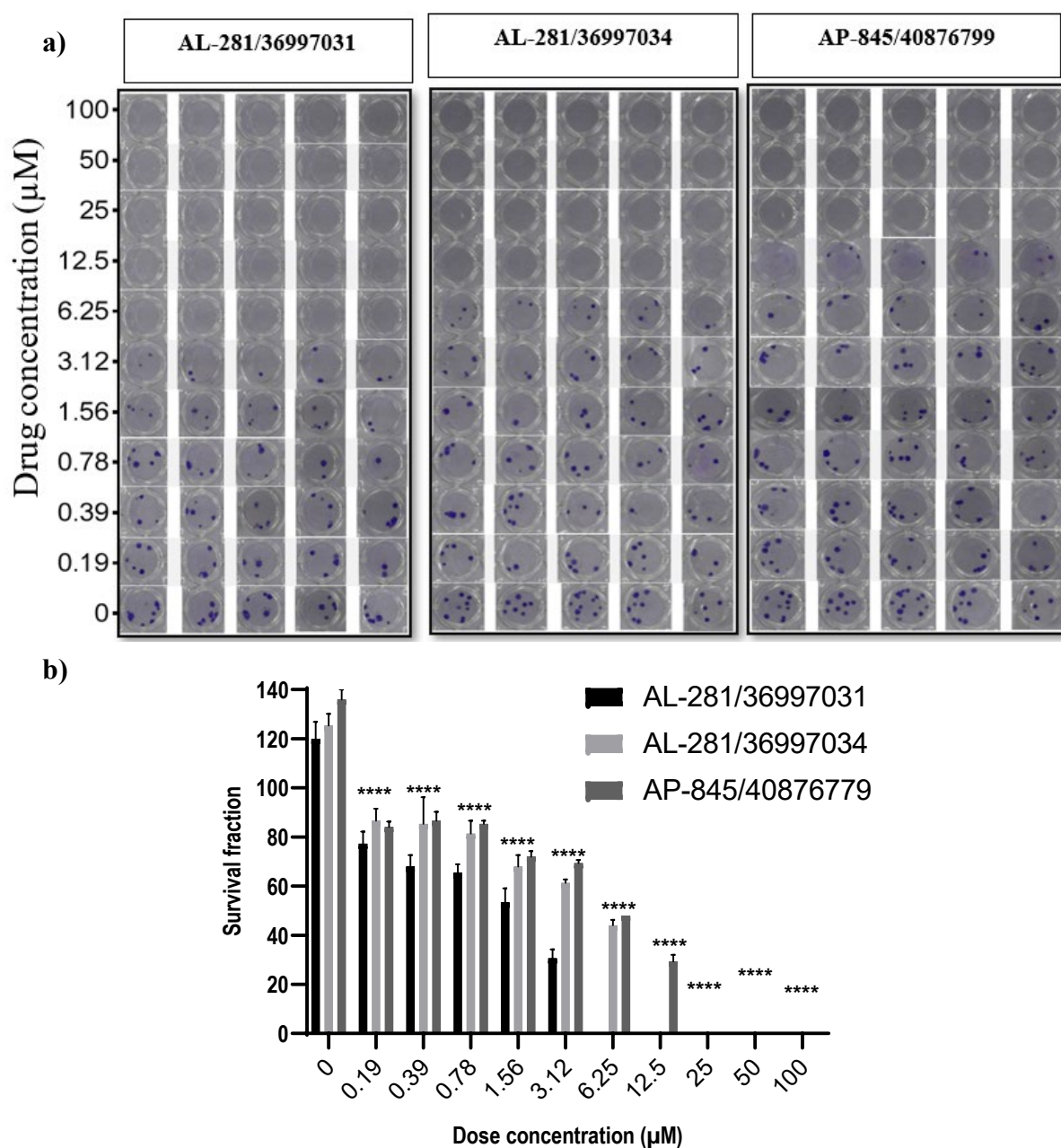
Figure 5.14: a) Representative images of SW480 colonies captured after two weeks of incubation against treatment of three test compounds AL-281/36997031, AL-281/36997034 and AP-845/40876799; b) Bar plot illustrating dose-dependent survival rate of SW480 cells with results from 3 independent experiments performed in triplicates with standard error of mean; ****$p < 0.0001$ as indicated (from two-way ANOVA, Sidak's multiple comparisons test). Percent relative cell viability for all treatments were quantified and normalised to the maximal response induced by vehicle control.

To summarise the findings from *in silico* and *in vitro* screenings (Table 5.7), the lead optimization of AK-968/12713190 and AG-690/40104520 with structural modifications using SAR analysis resulted in better hit compounds. Based on the identification and validation using *in vitro* assays, the cytotoxic properties of the three test compounds AL-281/36997031, AL-281/36997034 and AP-845/40876799 in a CRC cell line was confirmed.

Table 5.7: Summary of results obtained from VS of screening of Chemical libraries – DrugBank (left) and SPECS (right) databases against GPR120S model and *in vitro* screening assays in SW480 cells.

DrugBank database – VS of 2627 molecules	SPECS database – VS of ~350,000	
73 top scoring	66,27 top scoring	
↓	↓	
24 selected for *in vitro* screenings	13 compounds procured for in vitro screening	
↓	↓	
5 compounds procured and tested *in vitro*	2 hit compounds	
↓	↓	
Lapatinib (IC$_{50}$ 1.16 to 1.77 µM) showed cytotoxic activity in SW480	**AK-968/12713190** (IC$_{50}$ 23.21 to 26.69 µM)	**AG-690/40104520** (IC$_{50}$ 26.55 to 33.2 µM)
↓	↓	↓
Cytotoxicity of Lapatinib was not GPR120 specific when tested against GPR120-siRNA transfected SW480	Substructure search 16 hits	Substructure search 28 hits
	↓	↓
Similar spatial geometry of AH7614 and Deferasirox	7 compounds procured and tested *in vitro*	8 compounds procured and tested *in vitro*
↓	↓	↓
Deferasirox_analogue synthesized and tested *in vitro* with AH7614	2 hit compounds **AL-281/36997031** (IC$_{50}$ 5.890 to 6.715 µM)	1 hit compound **AP-845/40876799** (IC$_{50}$ 14.16 to 18.02 µM)
↓	**AL-281/36997034** (IC$_{50}$ 6.789 to 7.502 µM)	
Deferasirox_analogue reported reduced cytotoxicity compared to Deferasirox		

5.4 Discussion

The research aim was to identify novel lead molecules for selective binding to the human GPR120S receptor. The predicted and validated model of GPR120S was used for virtual screening of commercially available databases to find hit compounds having better binding scores using the molecular docking program SMINA (Koes, Baumgartner and Camacho 2013). Docking-based VS are growing in the number of success cases reported (Villoutreix et al., 2009; Rognan, 2017). The development of an accurate empirical scoring functions to predict protein–ligand binding affinities is a key aspect in SBDD.

Random forest-based scoring functions has been reported as best performing scoring functions (Wójcikowski, Ballester and Siedlecki 2017). The random forest is an ensemble approach that acts as a nearest neighbour predictor. The working principle of this method is that a group of "weak learners" (training set with limited parameters) can come together to form a "strong learner". For developing a random forest-based scoring function large number (minimum 300) of known ligands of GPR120S were required (Wójcikowski, Ballester and Siedlecki 2017). As the number of active ligands with uniform experimental data on GPR120S is low, preparing the random-forest based scoring function would not be useful in scoring and ranking the protein-ligand interactions. In this regard, consensus-based scoring function using rescoring programs which employ diverse types of algorithms was used to predict the best docking poses.

The iterative combination of *in silico* and *in vitro* methods employed resulting in discovery of three potential cytotoxic compounds (Table 5.6). From 73 top scoring virtual hits of DrugBank only 24 were selected for *in vitro* screening but only 5 compounds were procured and tested *in vitro*. The VS of DrugBank database might be useful in drug repurposing once the remaining virtual hits from DrugBank are screening *in vitro* (Sahragardjoonegani, Beall, Kesselheim and Hollis, 2021). While the initial *in vitro* screening of SPECS virtual hits reported AK-968/12713190 and AG-690/40104520 as lead cytotoxic compounds. SAR profiling of these two cytotoxic compounds resulted in three potential hits - AL-281/36997031, AL-281/36997034 and AP-845/40876799. When these compounds were tested at concentrations below experimental IC$_{50}$ values in GPR120-silenced SW480 cells, this cytotoxic effect of AL-281/36997031 and AP-845/40876799 was significantly suppressed in GPR120-siRNA transfected SW480 cells while AL-281/36997034 showed ~ 10% higher cytotoxicity levels in siRNA transfected cells. The comparative study between

GPR120-silenced and control experiments suggested that AL-281/36997031 exhibited cytotoxic effects through GPR120 binding while cytotoxic activity of AL-281/36997034 might be either through multiple targets including GPR120 or through another target. However, further *in silico* and *in vitro* validation is required to confirm their anti-cancer potential targeting GPR120. As literature suggests that high levels of GPR120 expression in CRC cell lines increases the cell proliferation rate and reduces apoptosis, it can be hypothesized that compounds AL-281/36997031, AL-281/36997034 and AP-845/40876799 inhibit GPR120 and hence increase the apoptosis rate.

A comparative study of these test compounds against a competitive antagonist would be useful, but as mentioned earlier no GPR120 antagonists are available to date. AH7614 (4-Methyl-*N*-9*H*-xanthen-9-yl-benzenesulfonamide) was first reported as a GPR120 selective antagonist by GlaxoSmithKline in 2014 but its mechanism of antagonism was not known. Later collaborative research by the Ulven and Milligan labs in 2017 reported that AH7614 was a negative allosteric modulator of GPR120 (Watterson et al., 2017). As it does not bind at the orthosteric binding pocket of the receptor, hence AH7614 is not a competitive antagonist of GPR120. The deferasirox_analogue designed as a novel GPR120 ligand showed significant activity and further functional assays are required to validate it as an allosteric or orthosteric ligand of GPR120.

As mentioned previously cell migration and colony formation are the characteristic hallmarks of cancer cells – a cytotoxic compound capable of preventing cell migration and colony formation can be characterised as an anti-cancer drug molecule (Arnold, Opdenaker, Flynn and Sims-Mourtada, 2015; Arwert, Hoste and Watt, 2012; He et al., 2018). Treatment with compound AP-845/40876799 (7.5 µM) showed lower inhibitory effects on the cell migration compared to AL-281/36997031 (3 µM) and AL-281/36997034 (3.5 µM), even though it was used at double concentration (Figure 5.13). While AP-845/40876799 reported higher significance in target identification when tested with SW480 cells with downregulated GPR120 expression (Figure 5.12). Similar results were observed in the clonogenic assay (Figure 5.14). The lower inhibitory effects of AP-845/40876799 can be related to its lower binding affinity towards the GPR120 orthosteric binding pocket. Further design optimisation of AP-845/40876799 by SAR analysis might improve the potency of new compounds from micro-molar to nano molars.

Chapter 6
Overall Discussion and Conclusion with Future Prospects

Being key sensors of extracellular signalling and regulation of various physiological processes, GPCR structural biology has been flooded with atomic-level information to design GPCR-targeted drugs (Zhu, Wu, Huang and An, 2021). Insights into GPCR-ligand binding and resulting conformational changes in GPCR structure upon activation are being studied. During this period, CADD – specifically SBDD has increasingly become popular for identifying novel therapeutics against GPCR families (Kosciolek, Mordalski and Bojarski, 2011; Yuan and Xu, 2018; Ferruz et al., 2018; Ibrahim and Clark, 2019). Recent studies have profiled GPR120 as a target of interest for developing anticancer therapeutics as it was found to play a significant role in chemoresistance in breast cancer tumour cells (Houthuijzen et al., 2017; Wang et al., 2019; Senatorov and Moniri, 2018). GPR120 is a Class A GPCR which has been reported to play a pro-oncogenic role in CRC by enhancing tumour angiogenesis and cell migration (Wu et al., 2013), and multifaceted roles in other cancer management (Senatorov and Moniri, 2018).

The present work has focused on studying the SBDD against GPR120 for development of anticancer drugs in CRC (Wu et al., 2013; Kumari, Reabroi and North, 2021). To attain the research goal, a spectrum of computational and biological methods to discover potential anticancer ligands from large chemical libraries, VS hits to potent lead optimisation, and molecular dynamics of ligand binding and activation were explored.

In the absence of experimentally elucidated 3D structure of GPR120, knowledge-based predictive computational methodologies such as homology modelling, molecular docking and MD simulations were applied to guide the *in vitro* studies towards the design and discovery of novel GPR120 ligands. Our first step was to build a 3D structural model of GPR120S through *in silico* methodologies. It is worthwhile reporting the inherent limitations in homology modelling of GPCRs with templates with low sequence identity (below 40%) (Tiss et al., 2021).

The previous GPR120 homology models reported in literature were generated using only a single template (Sun et al., 2010; Hara et al., 2011; Takeuchi et al., 2013; Chinthakunta et al., 2018; Zhang, Sun, Wen and Yuan, 2019). While the data regarding template selection for the

human GPR120 short isoform (Q5NUL3 - 361 amino acids) model prediction by AlphaFold2 is not available.

In this work, the highest sequence identity of templates available for GPR120 homology modelling was ~25% - Delta opioid and orexin 2 receptor, with 80 % query sequence coverage. Only recently, (January 2021), new crystal structures with higher sequence identity have been deposited in PDB database such as Neuropeptide Y2 receptor (~30%) with only 50% sequence coverage (Tang et al., 2021). An improved GPR120 homology model can be built using these higher sequence identity templates for future work as the recently developed software packages are powerful and reliable to predict comparative naturally resembling models such as AlhpaFold2 – AI predictive algorithm for protein folding (Malathi and Ramaiah 2018; Jumper et al., 2021).

To date and to the best of our knowledge, our GPR120S model was the first to be generated using pairwise template alignment using antagonist-bound human delta-like opioid receptor and orexin 2 receptors as templates covering 87% and 84% of the full query sequence, respectively. The homology model of GPR120S (short isoform) generated by combination of the two templates for the present study (generated in January 2017) provided apt guidance to focus our *in vitro* evaluation assays on the potential GPR120 ligands and enhanced the success rate of experimental design setup for lead screenings. Furthermore, none of the listed studies published data related to enrichment studies as well which prevented us from performing comparative analysis between these models and the generated model.

Another major knowledge gap missing is the link between active and inactive conformations of GPCRs obtained from X-ray crystallography or cryo-electron microscopic models. As mentioned earlier, ligand binding results in conformational changes leading to receptor activation. These conformational insights can lead to designing ligands with specific efficacy and activity profiles. In order to predict conformational changes upon activation, the GPR120S homology model in this study was built using templates in an inactive state (bound to an antagonist). The inactive state of generated GPR120S model was confirmed by structural analysis of characteristic Class A GPCR motifs such as [D/E]RY motif – "ionic lock" between TM3 and TM6 (Figure 4.3) (Katritch, Cherezov and Stevens, 2013; Rasmussen et al., 2007).

Using *in silico* methodology of molecular docking, GPR120 agonists (TUG891 and Compound39) were docked into the orthosteric binding pocket of the generated GPR120S model. The protein-ligand interaction profiling of the selected docked poses showed that the ligands are interacting with Arg99 of TM3 and all the residues essential for biological activity (Sun et al., 2010; Hudson, Shimpukade, Milligan and Ulven, 2014) are in the proximity of the ligands (~ 4 Å). The ligand bound GPR120S receptor models obtained by molecular docking were energy minimized to bring models to local minima (stable energy state) in the protein conformational energy landscape using steepest descent before MD assisted structural investigation of the ligand bound models. The site-specific mutation studies by Hudson et al reported – Arg99 (TM2), Trp104 (ECL1), Phe115 (TM3), Trp207, Phe211 (TM5), Trp277 (TM6) and Phe304 (TM7) (Hudson et al 2014) as essential amino acid residues required for significant binding interactions with the GPR120 agonists.

Our study performed 300 ns long all atomic MD simulations using the agonist-bound GPR120S models (TUG891 and Compound39) to mimic / predict and understand the ligand-driven modulation of equilibrium between different states of receptor. The topology (secondary structure) of the transmembrane receptors, which dictates the functionality of the receptor, is established by their physiological environment. As the effects and importance of immediate environment of transmembrane proteins on stabilising their folded protein conformations is well understood (Liang, Adamian and Jackups, 2005; Lee, 2011; Stansfeld and Sansom, 2011; Sandoval-Perez, Pluhackova and Böckmann, 2017; Marrink et al., 2019), we embedded the GPR120S models in the lipid bilayer followed by solvation with SPC water molecules and neutralisation by sodium and chloride ions to run MD simulations.

In the absence of a known GPR120 antagonist, one apo-GPR120S protein system and two agonist-bound GPR120S protein systems were built to run MD simulations (sections 4.1, 4.2) to get comparative results for structural analysis. MD analysis of the apo-GPR120S system (GPR120S model without a docked ligand) confirmed our two hypothesizes that - (i) the inactive state GPR120S homology model was generated by using templates of inactive receptors and (ii) an inactive GPR120S model will remain in an inactive state if not stimulated by agonist binding. The lowest energy conformation of apo-GPR120S model was extracted from the 300 ns MD trajectory and used for structure-based pharmacophore screening and VS of SPECS chemical database detailed in chapters 4 and 5 respectively.

While the MD analysis of the agonist-bound GPR120S model (TUG891) confirmed that agonist binding stimulated the GPR120S receptor and steered the protein conformation from an inactive state to an active state (Figure 4.6). The molecular docking analysis from consensus docking runs of Compound39 (Figure 4.3 and Table 4.1) illustrated that in addition to H-bond interactions shown by TUG891 (Arg99 and Trp277), Compound39 formed H-bonding interactions with Thr125 and Asn313 as well. As the comparative analysis of various GPCR Class A conserved motifs such as PIF, D/ERY and NPxxY motifs in the apo and the agonist-bound models characterised residues Asn313 as a molecular switch for antagonist binding. Further site-specific mutation studies targeting Asn313 alone as well as in combination with other binding pocket residues could be employed to confirm the importance of Asn313 interactions in GPR120 antagonist design.

Based on the inferred results, a structure-based pharmacophore hypothesis focused on Trp277 and Asn313 was generated (Figure 4.9b) to screen ZINC chemical library for small molecules which can conserve the Class A GPCR "ionic lock" and stabilise the protein in inactive state, preventing receptor activation. Structure-based pharmacophore screening resulted in screening of 63 compounds which were further analysed for interacting with residues essential for biological activity (Hudson, Shimpukade, Milligan and Ulven, 2014) as well as Asn313 (section 4.3, Table 4.2). MD analysis from the 100 ns all-atomic simulations of the final nine compounds reported significant differences in the conformational changes stimulated by docked ligands – Cpd1, Cpd7 and Cpd9 (Figure 4.11). The protein-ligand interaction fingerprint mapping of these three compounds over the 100 ns simulation period showed conserved interactions with Trp277 and Asn313. PLIF results were seconded by ligand RMSD studies which showed that Cpd1, Cpd7 and Cpd9 were the most stable in the orthosteric binding pocket of the GPR120S model (Appendix VII). Further analysis of the MD trajectory focusing on the "ionic lock" (Arg136-Asp259) conservation predicted that docking of Cpd7 could induce a conformational shift in receptor from inactive state towards active state, thus acting as a potential agonist of GPR120 receptor. While Cpd9 was predicted to stabilise the inactive state conformation of the GPR120S model thus acting as a potential GPR120 antagonist. The MD results obtained from our study illustrate the importance of understanding the conformational changes to design ligands with specific activities.

As mentioned in section 4.5, *in silico* investigations can be performed using extended ensemble MD sampling in presence of allosteric ions, protein-water H-bond networks, as well as water bridges between ligand and protein (Bertalan, Lešnik, Bren and Bondar, 2020). *In silico* site-specific mutations studies, especially for Asn313 as well as in combination with Trp277 of GPR120 (for short and long isoforms) can be further explored to confirm their predicted effects on ligand binding and receptor activation. Further the predictions from the MD derived structure-based pharmacophore screenings require validation by *in vitro* GPCR binding assays such as surface plasmon resonance (SPR) GPCR-ligand binding assay, radioligand or fluorescent ligand binding assays to study ligand bias, agonism and antagonistic profiling and G-protein dependent assays (cAMP assay, calcium influx assay, etc.) as well as G-protein independent assays (GPCR internalisation assay, label-free whole cell assay, β-Arrestin recruitment assay, etc.) (Zhang and Xie, 2012).

Recent review studies (Congreve, de Graaf, Swain and Tate, 2020; Zhu, Wu, Huang and An, 2021; Sanjeevi et al., 2022) elaborate on the impact of structural elucidation of GPCRs on development and increasing number of SBDD studies. GPCR-targeted SBDD methodology has been emplyed by successfully by well known CADD industrial players like DESRES (D. E. Shaw Research https://www.deshawresearch.com/), Sosei Heptares (https://soseiheptares.com/ - previously known as HEPTARES Therapeutics) as well as by newly emerging biotech companies like leadXpro (https://leadxpro.ch/), Confo Therpaeutics (http://www.confotherapeutics.com/), are some of the several which are working on SBDD. With a primary objective of GPCR-targeted SBDD, the energy refined model of GPR120S was used for VS experiments in chapter 5.

As specified in section 3.6 (Figure 3.11), the binding pocket defined for the docking algorithm (SMINA) was based on the site-specific mutation studies (Hudson et al. 2014). For future prospects, improvements in the homology model based on the available templates with higher sequence identity and homology together with increasing computational power will enable the evaluation of large set of analogs to improve hits from VS. As mentioned in section 3.8 and 5.4, due to low number of known ligands of GPR120 the AUC scores for Cscore from enrichment studies of the homology model of GPR120S was found to be lower than accepted (>0.9). The haystack used for enrichment studies can be refined using the actives and in-actives obtained from the present study to future work. An alternative approach could be employed to reduce the chemotype bias of the binding pocket is to generate an

ensemble of models and perform ensemble enrichment studies by evaluating the docked poses and protein-ligand interactions of the known ligands among the multiple homology models. Scoring or ranking of docked poses by scoring functions remains the principal bottleneck in SBDD. Different studies have published contrasting results related to consensus scoring for molecular docking experiments. Teramoto and Fukunishi proposed that supervised consensus scoring can improve the predictive power of VS by compensating for the deficiencies of each scoring function and taking into account protein-ligand interactions (Teramoto and Fukunishi, 2007). While one of the latest studies on evaluation of consensus scoring functions for AutoDock Vina and SMINA reported no performance gain in ranking docking poses and proposed that default scoring function of SMINA is the best approach for researchers (Masters, Eagon and Heying, 2020).

DrugBank and SPECS are the two chemical libraries screened against GPR120S model to discover lead candidates binding to GPR120 receptor (Table 5.7). For future studies, the present methods can be combined with improved computational approaches and specifc biological assays to further explore chemical databases like ZINC, ASINEX, etc,. As observed in various studies, the screening of large chemical databases can be improved by applying pan-assay interference (PAINs) analysis as well as selectivity screenings pre-docking which can filter out the chemical conpounds containing chemical moeties known to impart non-specificity, toxicity and potency related adverse effects (Baell and Holloway, 2010). In the present study, PAINs were performed through SwissADME only for hit candidates obtained after screening experiments (Table 4.4 and Table 5.6). For future propspects, these filters can be applied pre-docking a large chemical database like ZINC.

VS of DrugBank against GPR120S model was performed aiming to repurpose the commercially available FDA approved drugs which have been reported in various studies (Crisan, Avram and Pacureanu, 2017; Sahragardjoonegani, Beall, Kesselheim and Hollis, 2021). Only five hit candidates from VS of DrugBank were screened *in vitro*, out of which only Lapatinib - a known tyrosine kinase inhibitor of HER/EGFR receptors, showed strong cytotoxicity (IC$_{50}$ 1.16 to 1.77 µM) against SW480 cells. Lapatinib failed to show selectivity when tested against GPR120_siRNA treated cells confirming that the observed cytotoxic activity was not through GPR120 binding.

Visual analysis of Deferasirox – FDA approved iron chelating drug, and further 3D molecular overlay of Deferasirox against AH7614 (Figure 5.4) showed significant structural similarity.

AH7614 is a negative allosteric modulator of GPR120 (Watterson et al. 2017). Since the residues composing the allosteric site of GPR120 where AH7614 binds have not been confirmed by previous studies, design of Deferasirox_analogue (Figure 5.4) can be interpreted as ligand-based drug design. It can be argued that the DrugBank database was screened against GPR120S model focusing on the orthosteric binding pocket of GPR120 receptor (Hudson, Shimpukade, Milligan and Ulven, 2014). Elucidation of AH7614 bound GPR120 receptor by NMR, X-ray, or cryo-EM methods or site-specific mutation studies are required to debate this argument in detail. Both Deferasirox and the analogue when tested for GPR120 selectivity in GPR120-siRNA (Figure 5.5) treated were found to be inactive confirming their selectivity towards GPR120 providing a novel purpose for the Deferasirox scaffold (1,3,5-triphenyl-1H-[1,2,4] triazole). The present 2D cell experimental setup for this study did not provide significant results for proliferative or anti-proliferative effects of TUG891 (GPR120 agonist), due to which co-treatment of TUG891 and AH7614 versus TUG891 and Deferasirox_analogue was not performed for comparative analysis of allosteric activity of Deferasirox and the analogue. To the best of our knowledge, most of the published TUG891 studies assayed Ca^{2+} mobilization, β-arrestin recruitment, ERK phosphorylation, etc., (Son, Kim and Im, 2021), no studies have been published regarding proliferative or anti-proliferative effects of TUG891.

As academic research projects have limited funds, procurement of hit candidates obtained from VS of chemical databases often becomes a bottleneck due to high costs of custom synthesis or overseas shipments. To prevent such procurement limitations, our study focused VS of chemical libraries available in SPECS database (https://www.specs.net/), as the library inlcues only test compounds which are available for procurement. The hits from VS of SPECS were screened *in vitro* to test cytotoxic effects of the hit compounds (Pal, 2021). The combination of iterative optimization and testing of hit compounds from the initial *in silico* and *in vitro* screenings through physicochemical profiling and SAR analysis resulted in the discovery of three potential anticancer drugs. The three compounds - AL-281/36997031, AL-281/36997034 and AP-845/40876799 were confirmed as cytotoxic agents by wound healing and clonogenicity assays.

The SAR analysis of two test compounds AK-968/12713190 and AG-690/40104520, from initial screenings lead to the discovery of three compounds AL-281/36997031, AL-281/36997034 and AP-845/40876799 with improved IC_{50} values (section 5.3.2). SAR

performed in the present study was limited to the peripheral substitutions and replacements to the identified substructure and within the scope of SAR compounds available for procurement from the SPECS database. SAR profiling can be expanded beyond the SPECS database to screen a larger set of compounds to discover potential anti-cancer agents targeting GPR120. SAR profiling of substructures of AK-968/12713190 and AG-690/40104520 (Table 5.2; 5.3) may also lead to discovery of novel scaffolds exhibiting anticancer properties.

While GPR120-siRNA transfected cell lines showed that only compounds AL-281/36997031 and AP-845/40876799 are exhibiting their cytotoxic activity through GPR120. As mentioned, to confirm these compounds as GPR120-targeted therapeutics further binding assays such as calcium influx assay, BRET assay, etc., are required to be performed (Zhang and Xie, 2012; Shimpukade et al., 2012; Lombardo et al., 2016; Watterson et al., 2017). As the project was inspired by Wu et al., 2013 study, which identified GPR120 as angiogenesis inducing and tumour-promoting receptor, we aimed to discover and design GPR120-targeted novel anticancer therapeutics for CRC – the proposed future studies could focus on measurement of angiogenic phenotype in 2D or 3D cell models, if possible, animal models such as zebrafish models could be of great interest. Quantification of alterations in the angiogenic phenotype by delivery of screened candidates could confirm if the results are in agreement with Wu's lab (Wu et al., 2013).

Finally, since GPR120 is a hot target for therapeutics in metabolic disorders, escpecialy type 2 diabetes mellitus and obesity (Azevedo et al., 2016; Houthuijzen, 2016; Houthuijzen et al., 2017) – repurposing of the designed ligands such as Cpd1, Cpd7, Cpd9, and Deferasirox_analogue as well as compounds AL-281/36997031, AL-281/36997034 and AP-845/40876799 can be studied for various metabolic disorders.

To conclude, the present research has shown SBDD approaches and rigorous MD studies can focus the *in vitro* studies towards identification and optimisation of GPCR ligands. GPCR-based therapeutics continue to be developed as anticancer agents as they have the potential for use as a component of novel target selective agents to regulate tumour growth and metastasis. This work presents an example of the necessity and novelty of the convergence between computational and biological approaches to reap new grounds in molecular biology.

Bibliography

Ackermann, T. and Tardito, S., 2019. Cell Culture Medium Formulation and Its Implications in Cancer Metabolism. Trends in Cancer, 5(6), pp.329-332.

Adams, G., Velazquez, F., Jayne, C., Shah, U., Miao, S., Ashley, E., Madeira, M., Akiyama, T., Di Salvo, J., Suzuki, T., Wang, N., Truong, Q., Gilbert, E., Zhou, D., Verras, A., Kirkland, M., Pachanski, M., Powles, M., Yin, W., Ujjainwalla, F., Venkatraman, S. and Edmondson, S., 2016. Discovery of Chromane Propionic Acid Analogues as Selective Agonists of GPR120 with in Vivo Activity in Rodents. ACS Medicinal Chemistry Letters, 8(1), pp.96-101.

Adams, R. and Alitalo, K., 2007. Molecular regulation of angiogenesis and lymphangiogenesis. Nature Reviews Molecular Cell Biology, 8(6), pp.464-478.

Ağardan, N., Değim, Z., Yılmaz, Ş., Altıntaş, L. and Topal, T., 2020. Tamoxifen/raloxifene loaded liposomes for oral treatment of breast cancer. Journal of Drug Delivery Science and Technology, 57, p.101612.

Ahn, S., Park, S., Baek, J., Lee, S., Baek, W., Lee, S., Lee, Y., Yoo, H., Kim, H., Lee, S., Im, D., Lee, S., Kim, B. and Koh, J., 2016. Free Fatty Acid Receptor 4 (GPR120) Stimulates Bone Formation and Suppresses Bone Resorption in the Presence of Elevated n-3 Fatty Acid Levels. Endocrinology, 157(7), pp.2621-2635.

Akhtar, N., Jabeen, I., Jalal, N. and Antilla, J., 2018. Structure-based pharmacophore models to probe anticancer activity of inhibitors of protein kinase B-beta (PKB β). Chemical Biology & Drug Design, 93(3), pp.325-336.

Alvarez-Curto, E. and Milligan, G., 2016. Metabolism meets immunity: The role of free fatty acid receptors in the immune system. Biochemical Pharmacology, 114, pp.3-13.

Arnold, K., Opdenaker, L., Flynn, D. and Sims-Mourtada, J., 2015. Wound Healing and Cancer Stem Cells: Inflammation as a Driver of Treatment Resistance in Breast Cancer. Cancer Growth and Metastasis, 8, p.CGM.S11286.

Arnold, M., Sierra, M., Laversanne, M., Soerjomataram, I., Jemal, A. and Bray, F., 2016. Global patterns and trends in colorectal cancer incidence and mortality. Gut, 66(4), pp.683-691.

Arriba, M., García, J., Inglada-Pérez, L., Rueda, D., Osorio, I., Rodríguez, Y., Álvaro, E., Sánchez, R., Fernández, T., Pérez, J., Hernández, J., Benítez, J., González-Sarmiento, R., Urioste, M. and Perea, J., 2015. DNA copy number profiling reveals different patterns of

chromosomal instability within colorectal cancer according to the age of onset. Molecular Carcinogenesis, 55(5), pp.705-716.

Arwert, E., Hoste, E. and Watt, F., 2012. Epithelial stem cells, wound healing and cancer. Nature Reviews Cancer, 12(3), pp.170-180.

Aslantürk, Ö., 2018. In Vitro Cytotoxicity and Cell Viability Assays: Principles, Advantages, and Disadvantages. Genotoxicity - A Predictable Risk to Our Actual World, http://dx.doi.org/10.5772/intechopen.71923

ATCC-SW480, 2021. SW480 [SW-480] | ATCC. [online] Atcc.org. Available at: <https://www.atcc.org/products/ccl-228#detailed-product-information> [Accessed 4 March 2018].

Attwood, T. and Findlay, J., 1994. Fingerprinting G-protein-coupled receptors. "Protein Engineering, Design and Selection", 7(2), pp.195-203.

Austin, C., 2004. MOLECULAR BIOLOGY: NIH Molecular Libraries Initiative. Science, 306(5699), pp.1138-1139.

Azevedo, C., Watterson, K., Wargent, E., Hansen, S., Hudson, B., Kępczyńska, M., Dunlop, J., Shimpukade, B., Christiansen, E., Milligan, G., Stocker, C. and Ulven, T., 2016. Non-Acidic Free Fatty Acid Receptor 4 Agonists with Antidiabetic Activity. Journal of Medicinal Chemistry, 59(19), pp.8868-8878.

Backman, T., Cao, Y. and Girke, T., 2011. ChemMine tools: an online service for analyzing and clustering small molecules. Nucleic Acids Research, 39(suppl), pp.W486-W491.

Bae, J., Rhee, Y., Kim, K., Wen, X., Song, Y., Cho, N., Kim, J. and Kang, G., 2016. Are clinicopathological features of colorectal cancers with methylation in half of CpG island methylator phenotype panel markers different from those of CpG island methylator phenotype–high colorectal cancers? Human Pathology, 47(1), pp.85-94.

Baell, J. and Holloway, G., 2010. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. Journal of Medicinal Chemistry, 53(7), pp.2719-2740.

Bagchi, A., 2020. Latest trends in structure-based drug design with protein targets. Advances in Protein Chemistry and Structural Biology, pp.1-23.

Bai, Q., Ma, J., Liu, S., Xu, T., Banegas-Luna, A., Pérez-Sánchez, H., Tian, Y., Huang, J., Liu, H. and Yao, X., 2021. WADDAICA: A webserver for aiding protein drug design by artificial intelligence and classical algorithm. Computational and Structural Biotechnology Journal, 19, pp.3573-3579.

Baker, D., 2001. Protein Structure Prediction and Structural Genomics. Science, 294(5540), pp.93-96.

Ballesteros, J. and Weinstein, H., 1995. [19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. Methods in Neurosciences, pp.366-428.

Ballesteros, J., Shi, L. and Javitch, J., 2001. Structural Mimicry in G Protein-Coupled Receptors: Implications of the High-Resolution Structure of Rhodopsin for Structure-Function Analysis of Rhodopsin-Like Receptors. Molecular Pharmacology, 60(1), pp.1-19.

Bánvölgyi, A., Anker, P., Lőrincz, K., Kiss, N., Márton, D., Fésűs, L., Gyöngyösi, N. and Wikonkál, N., 2019. Smoothened receptor inhibitor vismodegib for the treatment of basal cell carcinoma: a retrospective analysis of efficacy and side effects. Journal of Dermatological Treatment, 31(4), pp.387-398.

Basith, S., Cui, M., Macalino, S., Park, J., Clavio, N., Kang, S. and Choi, S., 2018. Exploring G Protein-Coupled Receptors (GPCRs) Ligand Space via Cheminformatics Approaches: Impact on Rational Drug Design. Frontiers in Pharmacology, 9, ISSN:1663-9812.

Beer, T., Armstrong, A., Rathkopf, D., Loriot, Y., Sternberg, C., Higano, C., Iversen, P., Bhattacharya, S., Carles, J., Chowdhury, S., Davis, I., de Bono, J., Evans, C., Fizazi, K., Joshua, A., Kim, C., Kimura, G., Mainwaring, P., Mansbach, H., Miller, K., Noonberg, S., Perabo, F., Phung, D., Saad, F., Scher, H., Taplin, M., Venner, P. and Tombal, B., 2014. Enzalutamide in Metastatic Prostate Cancer before Chemotherapy. New England Journal of Medicine, 371(5), pp.424-433.

Bender, B., Marlow, B. and Meiler, J., 2020. Improving homology modeling from low-sequence identity templates in Rosetta: A case study in GPCRs. PLOS Computational Biology, 16(10), p.e1007597.

Bertalan, É., Lešnik, S., Bren, U. and Bondar, A., 2020. Protein-water hydrogen-bond networks of G protein-coupled receptors: Graph-based analyses of static structures and molecular dynamics. Journal of Structural Biology, 212(3), p.107634.

Beuming, T., Lenselink, B., Pala, D., McRobb, F., Repasky, M. and Sherman, W., 2015. Docking and Virtual Screening Strategies for GPCR Drug Discovery. Methods in Molecular Biology, pp.251-276.

Bhhatarai, B., Walters, W., Hop, C., Lanza, G. and Ekins, S., 2019. Opportunities and challenges using artificial intelligence in ADME/Tox. Nature Materials, 18(5), pp.418-422.

Biller, L. and Schrag, D., 2021. Diagnosis and Treatment of Metastatic Colorectal Cancer. JAMA, 325(7), p.669.

Bissantz, C., Bernard, P., Hibert, M. and Rognan, D., 2002. Protein-based virtual screening of chemical databases. II. Are homology models of g-protein coupled receptors suitable targets?. Proteins: Structure, Function, and Bioinformatics, 50(1), pp.5-25.

Boakye, D., Jansen, L., Halama, N., Chang-Claude, J., Hoffmeister, M. and Brenner, H., 2021. Early discontinuation and dose reduction of adjuvant chemotherapy in stage III colon cancer patients. Therapeutic Advances in Medical Oncology, 13, p.175883592110063.

Boned, R., van Gunsteren, W. and Daura, X., 2008. Estimating the Temperature Dependence of Peptide Folding Entropies and Free Enthalpies from Total Energies in Molecular Dynamics Simulations. Chemistry - A European Journal, 14(16), pp.5039-5046.

Boyle, M., Bell, S., Konstan, M., McColley, S., Rowe, S., Rietschel, E., Huang, X., Waltz, D., Patel, N. and Rodman, D., 2014. A CFTR corrector (lumacaftor) and a CFTR potentiator (ivacaftor) for treatment of patients with cystic fibrosis who have a phe508del CFTR mutation: a phase 2 randomised controlled trial. The Lancet Respiratory Medicine, 2(7), pp.527-538.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R., Torre, L. and Jemal, A., 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 68(6), pp.394-424.

Briscoe, C., Peat, A., McKeown, S., Corbett, D., Goetz, A., Littleton, T., McCoy, D., Kenakin, T., Andrews, J., Ammala, C., Fornwald, J., Ignar, D. and Jenkinson, S., 2006. Pharmacological regulation of insulin secretion in MIN6 cells through the fatty acid receptor GPR40: identification of agonist and antagonist small molecules. British Journal of Pharmacology, 148(5), pp.619-628.

Broecker, J., Eger, B. and Ernst, O., 2017. Crystallogenesis of Membrane Proteins Mediated by Polymer-Bounded Lipid Nanodiscs. Structure, 25(2), pp.384-392.

Brogi, 2019. Computational Approaches for Drug Discovery. Molecules, 24(17), p.3061.

Buiting, H. and Olthuis, G., 2020. Importance of Quality-of-Life Measurement Throughout the Disease Course. JAMA Network Open, 3(3), p.e200388.

Burki, T., 2018. Larotrectinib in TRK fusion-positive cancers. The Lancet Oncology, 19(4), p.e187.

Byrne, K., Pal, A., Curtin, J., Stephens, J. and Kinsella, G., 2021. G-protein-coupled receptors as therapeutic targets for glioblastoma. Drug Discovery Today;26(12), pp.2858-2870.

Cancer Discovery, 2017. AACR Project GENIE: Powering Precision Medicine through an International Consortium. 7(8), pp.818-831.

Cao, D., Zheng, Y., Xu, H., Ge, W. and Xu, X., 2019. Bevacizumab improves survival in metastatic colorectal cancer patients with primary tumor resection: A meta-analysis. Scientific Reports, 9(20326).

Carullo, G., Mazzotta, S., Vega-Holm, M., Iglesias-Guerra, F., Vega-Pérez, J., Aiello, F. and Brizzi, A., 2021. GPR120/FFAR4 Pharmacology: Focus on Agonists in Type 2 Diabetes Mellitus Drug Discovery. Journal of Medicinal Chemistry, 64(8), pp.4312-4332.

Cercek, A., Roxburgh, C., Strombom, P., Smith, J., Temple, L., Nash, G., Guillem, J., Paty, P., Yaeger, R., Stadler, Z., Seier, K., Gonen, M., Segal, N., Reidy, D., Varghese, A., Shia, J., Vakiani, E., Wu, A., Crane, C., Gollub, M., Garcia-Aguilar, J., Saltz, L. and Weiser, M., 2018. Adoption of Total Neoadjuvant Therapy for Locally Advanced Rectal Cancer. JAMA Oncology, 4(6), p.e180071.

Chahal, V., Nirwan, S. and Kakkar, R., 2019. Combined approach of homology modeling, molecular dynamics, and docking: computer-aided drug discovery. Physical Sciences Reviews, 4(10).

Chan, D., Segelov, E., Wong, R., Smith, A., Herbertson, R., Li, B., Tebbutt, N., Price, T. and Pavlakis, N., 2017. Epidermal growth factor receptor (EGFR) inhibitors for metastatic colorectal cancer. Cochrane Database of Systematic Reviews.

Chan, H., Filipek, S. and Yuan, S., 2016. The Principles of Ligand Specificity on beta-2-adrenergic receptor. Scientific Reports, 6(34736).

Chan, W. and Zhang, Y., 2020. Virtual Screening of Human Class-A GPCRs Using Ligand Profiles Built on Multiple Ligand–Receptor Interactions. Journal of Molecular Biology, 432(17), pp.4872-4890.

Chandra, N., 2011. Computational approaches for drug target identification in pathogenic diseases. Expert Opinion on Drug Discovery, 6(10), pp.975-979.

Chao, C., Lii, C., Ye, S., Li, C., Lu, C., Lin, A., Liu, K. and Chen, H., 2014. Docosahexaenoic Acid Inhibits Vascular Endothelial Growth Factor (VEGF)-Induced Cell Migration via the GPR120/PP2A/ERK1/2/eNOS Signaling Pathway in Human Umbilical Vein Endothelial Cells. Journal of Agricultural and Food Chemistry, 62(18), pp.4152-4158.

Chen, K., Liu, M., Liu, Y., Yoshimura, T., Shen, W., Le, Y., Durum, S., Gong, W., Wang, C., Gao, J., Murphy, P. and Wang, J., 2013. Formylpeptide receptor-2 contributes to colonic epithelial homeostasis, inflammation, and tumorigenesis. Journal of Clinical Investigation, 123(4), pp.1694-1704.

Chen, L., Aria, A., Silapunt, S., Lee, H. and Migden, M., 2018. Treatment of advanced basal cell carcinoma with sonidegib: perspective from the 30-month update of the BOLT trial. Future Oncology, 14(6), pp.515-525.

Chen, V., Arendall, W., Headd, J., Keedy, D., Immormino, R., Kapral, G., Murray, L., Richardson, J. and Richardson, D., 2009. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallographica Section D Biological Crystallography, 66(1), pp.12-21.

Chinthakunta, N., Cheemanapalli, S., Chinthakunta, S., Anuradha, C. and Chitta, S., 2018. A new insight into identification of in silico analysis of natural compounds targeting GPR120. Network Modeling Analysis in Health Informatics and Bioinformatics, 7.pp1-21.

Cocco, E., Scaltriti, M. and Drilon, A., 2018. NTRK fusion-positive cancers and TRK inhibitor therapy. Nature Reviews Clinical Oncology, 15(12), pp.731-747.

Colovos, C. and Yeates, T., 1993. Verification of protein structures: Patterns of nonbonded atomic interactions. Protein Science, 2(9), pp.1511-1519.

Congreve, M., Carr, R., Murray, C. and Jhoti, H., 2003. A 'Rule of Three' for fragment-based lead discovery? Drug Discovery Today, 8(19), pp.876-877.

Congreve, M., de Graaf, C., Swain, N. and Tate, C., 2020. Impact of GPCR Structures on Drug Discovery. Cell, 181(1), pp.81-91.

Cox, J., Chu, H., Chelliah, M., Debenham, J., Eagen, K., Lan, P., Lombardo, M., London, C., Plotkin, M., Shah, U., Sun, Z., Vaccaro, H., Venkatraman, S., Suzuki, T., Wang, N., Ashley, E., Crespo, A., Madeira, M., Leung, D., Alleyne, C., Ogawa, A., Souza, S., Thomas-Fowlkes, B., Di Salvo, J., Weinglass, A., Kirkland, M., Pachanski, M., Powles, M., Tozzo, E., Akiyama, T., Ujjainwalla, F., Tata, J. and Sinz, C., 2016. Design, Synthesis, and Evaluation of Novel and Selective G-protein Coupled Receptor 120 (GPR120) Spirocyclic Agonists. ACS Medicinal Chemistry Letters, 8(1), pp.49-54.

Cree, I., 2011. Cancer Biology. Methods in Molecular Biology, pp.1-11.

Crisan, L., Avram, S. and Pacureanu, L., 2017. Pharmacophore-based screening and drug repurposing exemplified on glycogen synthase kinase-3 inhibitors. Molecular Diversity, 21(2), pp.385-405.

da Silva Costa, J., da Silva Lopes Costa, K., Cruz, J., da Silva Ramos, R., Silva, L., Do Socorro Barros Brasil, D., de Paula da Silva, C., dos Santos, C. and da Cruz Macedo, W., 2018. Virtual Screening and Statistical Analysis in the Design of New Caffeine Analogues Molecules with Potential Epithelial Anticancer Activity. Current Pharmaceutical Design, 24(5), pp.576-594.

Dailey, M., Hait, C., Holt, P., Maguire, J., Meier, J., Miller, M., Petraccone, L. and Trent, J., 2009. Structure-based drug design: From nucleic acid to membrane protein targets. Experimental and Molecular Pathology, 86(3), pp.141-150.

Daina, A., Michielin, O. and Zoete, V., 2017. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Scientific Reports, 7(42717).

Davenport, A., Alexander, S., Sharman, J., Pawson, A., Benson, H., Monaghan, A., Liew, W., Mpamhanga, C., Bonner, T., Neubig, R., Pin, J., Spedding, M. and Harmar, A., 2013. International Union of Basic and Clinical Pharmacology. LXXXVIII. G Protein-Coupled Receptor List: Recommendations for New Pairings with Cognate Ligands. Pharmacological Reviews, 65(3), pp.967-986.

De Rosa, M., Pace U., Rega, D., Costabile, V., Duraturo, F., Izzo, P. and Delrio, P., 2015. Genetics, diagnosis and management of colorectal cancer (Review). Oncology Reports, 34(3), pp.1087-1096.

Dechant, K. and Goa, K., 1991. Levocabastine. Drugs, 41(2), pp.202-224.

Deshpande, N., 2012. Apixaban versus warfarin in patients with atrial fibrillation. Indian Heart Journal, 64(1), p.111.

Dhasmana, A., Raza, S., Jahan, R., Lohani, M. and Arif, J., 2019. High-Throughput Virtual Screening (HTVS) of Natural Compounds and Exploration of Their Biomolecular Mechanisms. New Look to Phytomedicine, pp.523-548.

Drilon, A., Siena, S., Ou, S., Patel, M., Ahn, M., Lee, J., Bauer, T., Farago, A., Wheler, J., Liu, S., Doebele, R., Giannetta, L., Cerea, G., Marrapese, G., Schirru, M., Amatu, A., Bencardino, K., Palmeri, L., Sartore-Bianchi, A., Vanzulli, A., Cresta, S., Damian, S., Duca, M., Ardini, E., Li, G., Christiansen, J., Kowalski, K., Johnson, A., Patel, R., Luo, D., Chow-Maneval, E., Hornby, Z., Multani, P., Shaw, A. and De Braud, F., 2017. Safety and Antitumor Activity of the Multitargeted Pan-TRK, ROS1, and ALK Inhibitor

Entrectinib: Combined Results from Two Phase I Trials (ALKA-372-001 and STARTRK-1). Cancer Discovery, 7(4), pp.400-409.

Dror, R., Arlow, D., Maragakis, P., Mildorf, T., Pan, A., Xu, H., Borhani, D. and Shaw, D., 2011. Activation mechanism of the β 2-adrenergic receptor. Proceedings of the National Academy of Sciences, 108(46), pp.18684-18689.

Dubé, C., Yakubu, M., McCurdy, B., Lischka, A., Koné, A., Walker, M., Peirson, L. and Tinmouth, J., 2017. Risk of Advanced Adenoma, Colorectal Cancer, and Colorectal Cancer Mortality in People With Low-Risk Adenomas at Baseline Colonoscopy: A Systematic Review and Meta-Analysis. American Journal of Gastroenterology, 112(12), pp.1790-1801.

Dunican, D., McWilliam, P., Tighe, O., Parle-McDermott, A. and Croke, D., 2002. Gene expression differences between the microsatellite instability (MIN) and chromosomal instability (CIN) phenotypes in colorectal cancer revealed by high-density cDNA array hybridization. Oncogene, 21(20), pp.3253-3257.

Elokdah, H., Friedrichs, G., Chai, S., Harrison, B., Primeau, J., Chlenov, M. and Crandall, D., 2002. Novel human metabolites of the angiotensin-II antagonist tasosartan and their pharmacological effects. Bioorganic & Medicinal Chemistry Letters, 12(15), pp.1967-1971.

Engelstoft, M., Park, W., Sakata, I., Kristensen, L., Husted, A., Osborne-Lawrence, S., Piper, P., Walker, A., Pedersen, M., Nøhr, M., Pan, J., Sinz, C., Carrington, P., Akiyama, T., Jones, R., Tang, C., Ahmed, K., Offermanns, S., Egerod, K., Zigman, J. and Schwartz, T., 2013. Seven transmembrane G protein-coupled receptor repertoire of gastric ghrelin cells. Molecular Metabolism, 2(4), pp.376-392.

Fearon, E. and Vogelstein, B., 1990. A genetic model for colorectal tumorigenesis. Cell, 61(5), pp.759-767.

Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D., Piñeros, M., Znaor, A. and Bray, F., 2021. Cancer statistics for the year 2020: An overview. International Journal of Cancer, 149(4), pp.778-789.

Ferruz, N., Doerr, S., Vanase-Frawley, M., Zou, Y., Chen, X., Marr, E., Nelson, R., Kormos, B., Wager, T., Hou, X., Villalobos, A., Sciabola, S. and De Fabritiis, G., 2018. Dopamine D3 receptor antagonist reveals a cryptic pocket in aminergic GPCRs. Scientific Reports, 8(897).

Fischer, G., 2000. Chemical aspects of peptide bond isomerisation. Chemical Society Reviews, 29(2), pp.119-127.

Fojo, T., Mailankody, S. and Lo, A., 2014. Unintended Consequences of Expensive Cancer Therapeutics—The Pursuit of Marginal Indications and a Me-Too Mentality That Stifles Innovation and Creativity. JAMA Otolaryngology–Head & Neck Surgery, 140(12), p.1225.

Folkman, J., 2007. Angiogenesis: an organizing principle for drug discovery?. Nature Reviews Drug Discovery, 6(4), pp.273-286.

Franken, N., Rodermond, H., Stap, J., Haveman, J. and van Bree, C., 2006. Clonogenic assay of cells in vitro. Nature Protocols, 1(5), pp.2315-2319.

Fredriksson, R., Höglund, P., Gloriam, D., Lagerström, M. and Schiöth, H., 2003. Seven evolutionarily conserved human rhodopsin G protein-coupled receptors lacking close relatives. FEBS Letters, 554(3), pp.381-388.

Fredriksson, R., Lagerström, M., Lundin, L. and Schiöth, H., 2003. The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. Molecular Pharmacology, 63(6), pp.1256-1272.

Fritze, O., Filipek, S., Kuksa, V., Palczewski, K., Hofmann, K. and Ernst, O., 2003. Role of the conserved NPxxY(x)5,6F motif in the rhodopsin ground state and during activation. Proceedings of the National Academy of Sciences, 100(5), pp.2290-2295.

Fukushima, K., Yamasaki, E., Ishii, S., Tomimatsu, A., Takahashi, K., Hirane, M., Fukushima, N., Honoki, K. and Tsujiuchi, T., 2015. Different roles of GPR120 and GPR40 in the acquisition of malignant properties in pancreatic cancer cells. Biochemical and Biophysical Research Communications, 465(3), pp.512-515.

Funar-Timofei, S., Borota, A. and Crisan, L., 2017. Combined molecular docking and QSAR study of fused heterocyclic herbicide inhibitors of D1 protein in photosystem II of plants. Molecular Diversity, 21(2), pp.437-454.

Furman, B., 2016. Gliquidone. Reference Module in Biomedical Sciences.

Gabrielsson, J., Andersson, R., Jirstrand, M. and Hjorth, S., 2018. Dose-Response-Time Data Analysis: An Underexploited Trinity. Pharmacological Reviews, 71(1), pp.89-122.

Gao, B., Han, Y., Wang, L., Lin, Y., Sun, Z., Lu, W., Hu, Y., Li, J., Lin, X., Liu, B., Jie, Q., Yang, L. and Luo, Z., 2016. Eicosapentaenoic acid attenuates dexamethasome-induced apoptosis by inducing adaptive autophagy via GPR120 in murine bone marrow-derived mesenchymal stem cells. Cell Death & Disease, 7(5), pp.e2235-e2235.

Gao, B., Huang, Q., Jie, Q., Lu, W., Wang, L., Li, X., Sun, Z., Hu, Y., Chen, L., Liu, B., Liu, J., Yang, L. and Luo, Z., 2015. GPR120: A bi-potential mediator to modulate the osteogenic and adipogenic differentiation of BMMSCs. Scientific Reports, 5(14080).

Garnock-Jones, K., 2015. Eluxadoline: First Global Approval. Drugs, 75(11), pp.1305-1310.

Garrett, C. and Eng, C., 2011. Cetuximab in the treatment of patients with colorectal cancer. Expert Opinion on Biological Therapy, 11(7), pp.937-949.

Gco.iarc.fr. 2021. Cancer today. [online] Available at: <https://gco.iarc.fr/today/online-analysis-table?v=2020&mode=cancer&mode_population=continents&population=900&populations=372&key=asr&sex=0&cancer=39&type=0&statistic=5&prevalence=0&population_group=0&ages_group%5B%5D=0&ages_group%5B%5D=17&group_cancer=1&include_nmsc=1&include_nmsc_other=1#collapse-group-0-4> [Accessed 5 July 2021].

Gene.com. 2021. Avastin - Genentech, Inc.. [online] Available at: <https://www.gene.com/download/pdf/avastin_prescribing.pdf> [Accessed 5 July 2018].

Gentile, I., Buonomo, A. and Borgia, G., 2014. Dasabuvir: A Non-Nucleoside Inhibitor of NS5B for the Treatment of Hepatitis C Virus Infection. Reviews on Recent Clinical Trials, 9(2), pp.115-123.

Ghizoni, M., Bertelli, J., Grala, C. and da Silva, R., 2012. The Anabolic Steroid Nandrolone Enhances Motor and Sensory Functional Recovery in Rat Median Nerve Repair With Long Interpositional Nerve Grafts. Neurorehabilitation and Neural Repair, 27(3), pp.269-276.

Ghosh, E., Kumari, P., Jaiman, D. and Shukla, A., 2015. Methodological advances: the unsung heroes of the GPCR structural revolution. Nature Reviews Molecular Cell Biology, 16(2), pp.69-81.

Gilbert, I., 2013. Drug Discovery for Neglected Diseases: Molecular Target-Based and Phenotypic Approaches. Journal of Medicinal Chemistry, 56(20), pp.7719-7726.

Godara, A., Siddiqui, N., Byrne, M. and Saif, M., 2018. The safety of lanreotide for neuroendocrine tumor. Expert Opinion on Drug Safety, 18(1), pp.1-10.

Goedecker, S., 2004. Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems. The Journal of Chemical Physics, 120(21), pp.9911-9917.

Golshani, G. and Zhang, Y., 2020. Advances in immunotherapy for colorectal cancer: a review. Therapeutic Advances in Gastroenterology, 13, p.175628482091752.

Gotoh, C., Hong, Y., Iga, T., Hishikawa, D., Suzuki, Y., Song, S., Choi, K., Adachi, T., Hirasawa, A., Tsujimoto, G., Sasaki, S. and Roh, S., 2007. The regulation of adipogenesis through GPR120. Biochemical and Biophysical Research Communications, 354(2), pp.591-597.

Greer, N., 2007. Posaconazole (Noxafil): A New Triazole Antifungal Agent. Baylor University Medical Center Proceedings, 20(2), pp.188-196.

Grisshammer, R., 2017. New approaches towards the understanding of integral membrane proteins: A structural perspective on G protein-coupled receptors. Protein Science, 26(8), pp.1493-1504.

Grundmann, M., Bender, E., Schamberger, J. and Eitner, F., 2021. Pharmacology of Free Fatty Acid Receptors and Their Allosteric Modulators. International Journal of Molecular Sciences, 22(4), p.1763.

Gusach, A., Maslov, I., Luginina, A., Borshchevskiy, V., Mishin, A. and Cherezov, V., 2020. Beyond structure: emerging approaches to study GPCR dynamics. Current Opinion in Structural Biology, 63, pp.18-25.

Hammersley, D. and Signy, M., 2016. Ezetimibe: an update on its clinical usefulness in specific patient groups. Therapeutic Advances in Chronic Disease, 8(1), pp.4-11.

Hara, T., Hirasawa, A., Ichimura, A., Kimura, I. and Tsujimoto, G., 2011. Free Fatty Acid Receptors FFAR1 and GPR120 as Novel Therapeutic Targets for Metabolic Disorders. Journal of Pharmaceutical Sciences, 100(9), pp.3594-3601.

Hasan, A., Ohmori, K., Konishi, K., Igarashi, J., Hashimoto, T., Kamitori, K., Yamaguchi, F., Tsukamoto, I., Uyama, T., Ishihara, Y., Noma, T., Tokuda, M. and Kohno, M., 2015. Eicosapentaenoic acid upregulates VEGF-A through both GPR120 and PPARγ mediated pathways in 3T3-L1 adipocytes. Molecular and Cellular Endocrinology, 406, pp.10-18.

Hauser, A., Attwood, M., Rask-Andersen, M., Schiöth, H. and Gloriam, D., 2017. Trends in GPCR drug discovery: new agents, targets and indications. Nature Reviews Drug Discovery, 16(12), pp.829-842.

Hauser, A., Chavali, S., Masuho, I., Jahn, L., Martemyanov, K., Gloriam, D. and Babu, M., 2018. Pharmacogenomics of GPCR Drug Targets. Cell, 172(1-2), pp.41-54.e19.

Hay, M., Thomas, D., Craighead, J., Economides, C. and Rosenthal, J., 2014. Clinical development success rates for investigational drugs. Nature Biotechnology, 32(1), pp.40-51.

He, J., Li, Y., Han, Z., Zhou, J., Chen, W., Lv, Y., He, M., Zuo, J. and Zheng, L., 2018. The CircRNA-ACAP2/Hsa-miR-21-5p/ Tiam1 Regulatory Feedback Circuit Affects the Proliferation, Migration, and Invasion of Colon Cancer SW480 Cells. Cellular Physiology and Biochemistry, 49(4), pp.1539-1550.

Hirasawa, A., Hara, T., Katsuma, S., Adachi, T. and Tsujimoto, G., 2008. Free Fatty Acid Receptors and Drug Discovery. Biological and Pharmaceutical Bulletin, 31(10), pp.1847-1851.

Hirasawa, A., Tsumaya, K., Awaji, T., Katsuma, S., Adachi, T., Yamada, M., Sugimoto, Y., Miyazaki, S. and Tsujimoto, G., 2004. Free fatty acids regulate gut incretin glucagon-like peptide-1 secretion through GPR120. Nature Medicine, 11(1), pp.90-94.

Ho, B. and Brasseur, R., 2005. The Ramachandran plots of glycine and pre-proline. BMC Structural Biology, 5(14).

Hopkins, M. and Meier, K., 2015. Free Fatty Acid Receptor 4 (FFA4) Mediates the Inhibitory Effects of Omega-3 Fatty Acids on Growth Factor-mediated Proliferation and Migration in Human Prostate Cancer Cells. The FASEB Journal, 29(S1).

Hopkins, M. and Meier, K., 2016. Free Fatty Acid Receptors and Cancer: From Nutrition to Pharmacology. Free Fatty Acid Receptors, pp.233-251.

Hopkins, M., Liu, Z. and Meier, K., 2014. Effects of eicosapentaenoic acid and the GPR120 agonist TUG-891 on lysophosphatidic acid signaling in human prostate cancer cells (1066.10). The FASEB Journal, 28(S1).

Houthuijzen, J., 2016. For Better or Worse: FFAR1 and FFAR4 Signaling in Cancer and Diabetes. Molecular Pharmacology, 90(6), pp.738-743.

Houthuijzen, J., Oosterom, I., Hudson, B., Hirasawa, A., Daenen, L., McLean, C., Hansen, S., Jaarsveld, M., Peeper, D., Sadatmand, S., Roodhart, J., Lest, C., Ulven, T., Ishihara, K., Milligan, G. and Voest, E., 2017. Fatty acid 16:4(n-3) stimulates a GPR120-induced signaling cascade in splenic macrophages to promote chemotherapy resistance. The FASEB Journal, 31(5), pp.2195-2209.

HSE.ie. 2019. Cancer Incidence, Survival and Mortality Data - HSE.ie. [online] Available at: <https://www.hse.ie/eng/services/list/5/cancer/pubs/intelligence/registrydata.html> [Accessed 2 July 2019].

Huang, J., Zhao, D., Liu, Z. and Liu, F., 2018. Repurposing psychiatric drugs as anti-cancer agents. Cancer Letters, 419, pp.257-265.

Hudson, B., Shimpukade, B., Mackenzie, A., Butcher, A., Pediani, J., Christiansen, E., Heathcote, H., Tobin, A., Ulven, T. and Milligan, G., 2013. The Pharmacology of TUG-891, a Potent and Selective Agonist of the Free Fatty Acid Receptor 4 (FFA4/GPR120), Demonstrates Both Potential Opportunity and Possible Challenges to Therapeutic Agonism. Molecular Pharmacology, 84(5), pp.710-725.

Hudson, B., Shimpukade, B., Milligan, G. and Ulven, T., 2014. The Molecular Basis of Ligand Interaction at Free Fatty Acid Receptor 4 (FFA4/GPR120). Journal of Biological Chemistry, 289(29), pp.20345-20358.

Ibrahim, P. and Clark, T., 2019. Metadynamics simulations of ligand binding to GPCRs. Current Opinion in Structural Biology, 55, pp.129-137.

Ichimura, A., Hirasawa, A., Poulain-Godefroy, O., Bonnefond, A., Hara, T., Yengo, L., Kimura, I., Leloire, A., Liu, N., Iida, K., Choquet, H., Besnard, P., Lecoeur, C., Vivequin, S., Ayukawa, K., Takeuchi, M., Ozawa, K., Tauber, M., Maffeis, C., Morandi, A., Buzzetti, R., Elliott, P., Pouta, A., Jarvelin, M., Körner, A., Kiess, W., Pigeyre, M., Caiazzo, R., Van Hul, W., Van Gaal, L., Horber, F., Balkau, B., Lévy-Marchal, C., Rouskas, K., Kouvatsi, A., Hebebrand, J., Hinney, A., Scherag, A., Pattou, F., Meyre, D., Koshimizu, T., Wolowczuk, I., Tsujimoto, G. and Froguel, P., 2012. Dysfunction of lipid sensor GPR120 leads to obesity in both mouse and human. Nature, 483(7389), pp.350-354.

Illergård, K., Ardell, D. and Elofsson, A., 2009. Structure is three to ten times more conserved than sequence-A study of structural response in protein cores. Proteins: Structure, Function, and Bioinformatics, 77(3), pp.499-508.

Insel, P., Sriram, K., Gorr, M., Wiley, S., Michkov, A., Salmerón, C. and Chinn, A., 2019. GPCRomics: An Approach to Discover GPCR Drug Targets. Trends in Pharmacological Sciences, 40(6), pp.378-387.

Irwin, J., Sterling, T., Mysinger, M., Bolstad, E. and Coleman, R., 2012. ZINC: A Free Tool to Discover Chemistry for Biology. Journal of Chemical Information and Modeling, 52(7), pp.1757-1768.

Ishimoto, K., Minami, A., Minami, K., Ueda, N. and Tsujiuchi, T., 2020. Different effects of lysophosphatidic acid receptor-2 (LPA2) and LPA5 on the regulation of chemoresistance in colon cancer cells. Journal of Receptors and Signal Transduction, 41(1), pp.93-98.

Iwasaki, K., Harada, N., Sasaki, K., Yamane, S., Iida, K., Suzuki, K., Hamasaki, A., Nasteska, D., Shibue, K., Joo, E., Harada, T., Hashimoto, T., Asakawa, Y., Hirasawa, A. and Inagaki, N., 2015. Free Fatty Acid Receptor GPR120 Is Highly Expressed in Enteroendocrine K Cells of the Upper Small Intestine and Has a Critical Role in GIP Secretion After Fat Ingestion. Endocrinology, 156(3), pp.837-846.

Iyanagi, T., 2007. Molecular Mechanism of Phase I and Phase II Drug-Metabolizing Enzymes: Implications for Detoxification. International Review of Cytology, pp.35-112.

Jacquemard, C. and Kellenberger, E., 2019. A bright future for fragment-based drug discovery: what does it hold?. Expert Opinion on Drug Discovery, 14(5), pp.413-416.

Jaiteh, M., Rodríguez-Espigares, I., Selent, J. and Carlsson, J., 2020. Performance of virtual screening against GPCR homology models: Impact of template selection and treatment of binding site plasticity. PLOS Computational Biology, 16(3), p.e1007680.

Jakher, H., Chang, T., Tan, M. and Mahaffey, K., 2019. Canagliflozin review - safety and efficacy profile in patients with T2DM. Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy, Volume 12, pp.209-215.

Jakowiecki, J., Orzeł, U., Chawananon, S., Miszta, P. and Filipek, S., 2020. The Hydrophobic Ligands Entry and Exit from the GPCR Binding Site-SMD and SuMD Simulations. Molecules, 25(8), p.1930.

Jhoti, H., Williams, G., Rees, D. and Murray, C., 2013. The 'rule of three' for fragment-based drug discovery: where are we now?. Nature Reviews Drug Discovery, 12(8), pp.644-644.

Jiang, W., Puntis, M. and Hallett, M., 1994. Molecular and cellular basis of cancer invasion and metastasis: Implications for treatment. British Journal of Surgery, 81(11), pp.1576-1590.

Joseph-McCarthy, D., Thomas, B., Belmarsh, M., Moustakas, D. and Alvarez, J., 2003. Pharmacophore-based molecular docking to account for ligand flexibility. Proteins: Structure, Function, and Genetics, 51(2), pp.172-188.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S., Ballard, A., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A., Kavukcuoglu, K., Kohli, P. and Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature,

Kaczanowski, S. and Zielenkiewicz, P., 2009. Why similar protein sequences encode similar three-dimensional structures? Theoretical Chemistry Accounts, 125(3-6), pp.643-650.

Kang, S., Huang, J., Lee, B., Jung, Y., Im, E., Koh, J. and Im, D., 2018. Omega-3 polyunsaturated fatty acids protect human hepatoma cells from developing steatosis through FFA4 (GPR120). Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids, 1863(2), pp.105-116.

Kato, H., Zhang, Y., Hu, H., Suomivuori, C., Kadji, F., Aoki, J., Krishna Kumar, K., Fonseca, R., Hilger, D., Huang, W., Latorraca, N., Inoue, A., Dror, R., Kobilka, B. and Skiniotis, G., 2019. Conformational transitions of a neurotensin receptor 1–Gi1 complex. Nature, 572(7767), pp.80-85.

Katou, M., Soga, N., Onishi, T., Arima, K. and Sugimura, Y., 2008. Small cell carcinoma of the prostate treated with amrubicin. International Journal of Clinical Oncology, 13(2), pp.169-172.

Katritch, V., Cherezov, V. and Stevens, R., 2013. Structure-Function of the G Protein–Coupled Receptor Superfamily. Annual Review of Pharmacology and Toxicology, 53(1), pp.531-556.

Kaur, S., Mirza, A., Overgaard, A., Pociot, F. and Størling, J., 2021. A Dual Systems Genetics Approach Identifies Common Genes, Networks, and Pathways for Type 1 and 2 Diabetes in Human Islets. Frontiers in Genetics, 12.

Kay Washington, M., 2008. Colorectal Carcinoma: Selected Issues in Pathologic Examination and Staging and Determination of Prognostic Factors. Archives of Pathology & Laboratory Medicine, 132(10), pp.1600-1607.

Keam, S. and Scott, L., 2008. Dutasteride. Drugs, 68(4), pp.463-485.

Keum, N. and Giovannucci, E., 2019. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. Nature Reviews Gastroenterology & Hepatology, 16(12), pp.713-732.

Kheirelseid, E., Miller, N. and Kerin, M., 2013. Clinical applications of molecular profiling in colorectal cancer: Review of the literature. American Journal of Molecular Biology, 03(02), pp.131-138.

Khosravan, A., Marani, S. and Sadeghi Googheri, M., 2017. The effects of fluorine substitution on the chemical properties and inhibitory capacity of Donepezil anti-Alzheimer drug; density functional theory and molecular docking calculations. Journal of Molecular Graphics and Modelling, 71, pp.124-134.

Kim, J., Lee, K., Park, S., Kang, S., Huang, J., Lee, J., Sato, K., Chung, H., Okajima, F. and Im, D., 2015. Omega-3 fatty acids induce Ca2+ mobilization responses in human colon epithelial cell lines endogenously expressing FFA4. Acta Pharmacologica Sinica, 36(7), pp.813-820.

Kinoshita, M. and Okada, T., 2015. Erratum: Structural conservation among the rhodopsin-like and other G protein-coupled receptors. Scientific Reports, 5(12253).

Kitamura, C., Sonoda, H., Nozawa, H., Kano, K., Emoto, S., Murono, K., Kaneko, M., Hiyoshi, M., Sasaki, K., Nishikawa, T., Shuno, Y., Tanaka, T., Hata, K., Kawai, K., Aoki, J. and Ishihara, S., 2019. The component changes of lysophospholipid mediators in colorectal cancer. Tumor Biology, 41(5), p.101042831984861.

Kling, R., Clark, T. and Gmeiner, P., 2016. Comparative MD Simulations Indicate a Dual Role for Arg1323.50 in Dopamine-Dependent D2R Activation. PLOS ONE, 11(1), p.e0146612.

Koes, D. and Camacho, C., 2012. ZINCPharmer: pharmacophore search of the ZINC database. Nucleic Acids Research, 40(W1), pp.W409-W414.

Kolakowski Jr, L. F., 1994. GCRDb: A G protein-Coupled Receptor Database. Receptors Channels, 2, 1.

Kolli, S., Pecone, D., Pona, A., Cline, A. and Feldman, S., 2019. Topical Retinoids in Acne Vulgaris: A Systematic Review. American Journal of Clinical Dermatology, 20(3), pp.345-365.

Komm, B. and Chines, A., 2011. Bazedoxifene: the evolving role of third-generation selective estrogen-receptor modulators in the management of postmenopausal osteoporosis. Therapeutic Advances in Musculoskeletal Disease, 4(1), pp.21-34.

Kooistra, A., Mordalski, S., Pándy-Szekeres, G., Esguerra, M., Mamyrbekov, A., Munk, C., Keserű, G. and Gloriam, D., 2020. GPCRdb in 2021: integrating GPCR sequence, structure and function. Nucleic Acids Research, 49(D1), pp.D335-D343.

Kosciolek, T., Mordalski, S. and Bojarski, A., 2011. Rapid binding site analysis by means of structural interaction fingerprint patterns – an implication to GPCR-targeted CADD. Journal of Cheminformatics, 3(S1).

Koyuncuer, A. and Ozkan, H., 2020. Clinicopathological and molecular features of sporadic colorectal cancers with DNA mismatch repair deficiency: A single center experience. Medicine Science | International Medical Journal, 9(4), p.1014.

Kue, C., Kamkaew, A., Burgess, K., Kiew, L., Chung, L. and Lee, H., 2016. Small Molecules for Active Targeting in Cancer. Medicinal Research Reviews, 36(3), pp.494-575.

Lange, A. and Lo, H., 2018. Inhibiting TRK Proteins in Clinical Cancer Therapy. Cancers, 10(4), p.105.

Larsson, P., Wallner, B., Lindahl, E. and Elofsson, A., 2008. Using multiple templates to improve quality of homology models in automated homology modeling. Protein Science, 17(6), pp.990-1002.

Latorraca, N., Venkatakrishnan, A. and Dror, R., 2016. GPCR Dynamics: Structures in Motion. Chemical Reviews, 117(1), pp.139-155.

Lee, A., 2011. Biological membranes: the importance of molecular detail. Trends in Biochemical Sciences, 36(9), pp.493-500.

Lengauer, C., Kinzler, K. and Vogelstein, B., 1998. Genetic instabilities in human cancers. Nature, 396(6712), pp.643-649.

Lenselink, E., Beuming, T., Sherman, W., van Vlijmen, H. and IJzerman, A., 2014. Selecting an Optimal Number of Binding Site Waters To Improve Virtual Screening Enrichments Against the Adenosine A2A Receptor. Journal of Chemical Information and Modeling, 54(6), pp.1737-1746.

Liang, J., Adamian, L. and Jackups, R., 2005. The membrane–water interface region of membrane proteins: structural bias and the anti-snorkeling effect. Trends in Biochemical Sciences, 30(7), pp.355-357.

Liang, J., Khorana, A. and Kalady, M., 2015. BRAF-mutated colorectal cancer: clinical implications for a distinct subset of the disease. Colorectal Cancer, 4(3), pp.125-133.

Liang, W., Chen, K., Gong, W., Yoshimura, T., Le, Y., Wang, Y. and Wang, J., 2020. The Contribution of Chemoattractant GPCRs, Formylpeptide Receptors, to Inflammation and Cancer. Frontiers in Endocrinology, 11.

Lin, S., Zhang, A., Zhang, X. and Wu, Z., 2019. Treatment of Pituitary and Other Tumours with Cabergoline: New Mechanisms and Potential Broader Applications. Neuroendocrinology, 110(6), pp.477-488.

Lin, X., Li, X. and Lin, X., 2020. A Review on Applications of Computational Methods in Drug Screening and Design. Molecules, 25(6), p.1375.

Lindhagen, E., Nygren, P. and Larsson, R., 2008. The fluorometric microculture cytotoxicity assay. Nature Protocols, 3(8), pp.1364-1369.

Lingwood, D. and Simons, K., 2009. Lipid Rafts As a Membrane-Organizing Principle. Science, 327(5961), pp.46-50.

Lipinski, C., Lombardo, F., Dominy, B. and Feeney, P., 2012. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews, 64, pp.4-17.

Liu, Y., Wada, R., Yamashita, T., Mi, Y., Deng, C., Hobson, J., Rosenfeldt, H., Nava, V., Chae, S., Lee, M., Liu, C., Hla, T., Spiegel, S. and Proia, R., 2000. Edg-1, the G protein–coupled receptor for sphingosine-1-phosphate, is essential for vascular maturation. Journal of Clinical Investigation, 106(8), pp.951-961.

Liu, Z., Hopkins, M., Zhang, Z., Quisenberry, C., Fix, L., Galvan, B. and Meier, K., 2014. Omega-3 Fatty Acids and Other FFA4 Agonists Inhibit Growth Factor Signaling in Human Prostate Cancer Cells. Journal of Pharmacology and Experimental Therapeutics, 352(2), pp.380-394.

Lombardo, M., Bender, K., London, C., Plotkin, M., Kirkland, M., Mane, J., Pachanski, M., Geissler, W., Cummings, J., Habulihaz, B., Akiyama, T., Di Salvo, J., Madeira, M., Pols, J., Powles, M., Finley, M., Johnson, E., Roussel, T., Uebele, V., Crespo, A., Leung, D., Alleyne, C., Trusca, D., Lei, Y., Howard, A., Ujjainwalla, F., Tata, J. and Sinz, C., 2016. Discovery of benzofuran propanoic acid GPR120 agonists: From uHTS hit to mechanism-based pharmacodynamic effects. Bioorganic & Medicinal Chemistry Letters, 26(23), pp.5724-5728.

Low, J., Du, W., Gocha, T., Oguz, G., Zhang, X., Chen, M., Masirevic, S., Yim, D., Tan, I., Ramasamy, A., Fan, H. and DasGupta, R., 2021. Molecular docking-aided identification of small molecule inhibitors targeting β-catenin-TCF4 interaction. iScience, 24(6), p.102544.

Luther, J. and Chan, A., 2016. Malignant Tumors of the Colon. Yamada's Atlas of Gastroenterology, pp.238-245.

Maesaka, F., Nakai, Y., Tomizawa, M., Owari, T., Miyake, M., Inoue, T., Anai, S., Tanaka, N. and Fujimoto, K., 2019. Amrubicin is effective against small cell carcinoma of the prostate as a second-line chemotherapeutic agent: A case report. IJU Case Reports, 2(3), pp.133-136.

Maillard, V., Desmarchais, A., Durcin, M., Uzbekova, S. and Elis, S., 2018. Docosahexaenoic acid (DHA) effects on proliferation and steroidogenesis of bovine granulosa cells. Reproductive Biology and Endocrinology, 16(40).

Malathi, K. and Ramaiah, S., 2018. Bioinformatics approaches for new drug discovery: a review. Biotechnology and Genetic Engineering Reviews, 34(2), pp.243-260.

Marrink, S., Corradi, V., Souza, P., Ingólfsson, H., Tieleman, D. and Sansom, M., 2019. Computational Modeling of Realistic Cell Membranes. Chemical Reviews, 119(9), pp.6184-6226.

Masters, L., Eagon, S. and Heying, M., 2020. Evaluation of consensus scoring methods for AutoDock Vina, smina and idock. Journal of Molecular Graphics and Modelling, 96, p.107532.

Mathonnet, M., 2014. Hallmarks in colorectal cancer: Angiogenesis and cancer stem-like cells. World Journal of Gastroenterology, 20(15), p.4189.

Matsumura, S., Mizushige, T., Yoneda, T., Iwanaga, T., Tsuzuki, S., Inoue, K. and Fushiki, T., 2007. GPR expression in the rat taste bud relating to fatty acid sensing. Biomedical Research, 28(1), pp.49-55.

McCombie, A., Frampton, C. and Frizelle, F., 2021. Quality of life preferences in colorectal cancer patients aged 80 and over. ANZ Journal of Surgery.

Miao, Y., Nichols, S. and McCammon, J., 2014. Free energy landscape of G-protein coupled receptors, explored by accelerated molecular dynamics. Physical Chemistry Chemical Physics, 16(14), p.6398.

Milligan, G., Alvarez-Curto, E., Watterson, K., Ulven, T. and Hudson, B., 2015. Characterizing pharmacological ligands to study the long-chain fatty acid receptors GPR40/FFA1 and GPR120/FFA4. British Journal of Pharmacology, 172(13), pp.3254-3265.

Milligan, G., Shimpukade, B., Ulven, T. and Hudson, B., 2016. Complex Pharmacology of Free Fatty Acid Receptors. Chemical Reviews, 117(1), pp.67-110.

Mirzapur, P., Khazaei, M., Moradi, M. and Khazaei, M., 2018. Apoptosis induction in human breast cancer cell lines by synergic effect of raloxifene and resveratrol through increasing proapoptotic genes. Life Sciences, 205, pp.45-53.

Mitro, N., Gilardi, F., Giudici, M., Godio, C., Scotti, E. and Crestani, M., 2012. Site-Directed Mutagenesis to Study the Role of Specific Amino Acids in the Ligand Binding Domain of PPARs. Peroxisome Proliferator-Activated Receptors (PPARs), pp.137-144.

Mokhtari, R., Homayouni, T., Baluch, N., Morgatskaya, E., Kumar, S., Das, B. and Yeger, H., 2017. Combination therapy in combating cancer. Oncotarget, 8(23), pp.38022-38043.

Moore, D., Elmes, J., Shibu, P., Larck, C. and Park, S., 2019. Mogamulizumab: An Anti-CC Chemokine Receptor 4 Antibody for T-Cell Lymphomas. Annals of Pharmacotherapy, 54(4), pp.371-379.

Muhammed, M. and Aki-Yalcin, E., 2018. Homology modeling in drug discovery: Overview, current applications, and future perspectives. Chemical Biology & Drug Design, 93(1), pp.12-20.

Munker, S., Gerken, M., Fest, P., Ott, C., Schnoy, E., Fichtner-Feigl, S., Wiggermann, P., Vogelhuber, M., Herr, W., Stroszczynski, C., Schlitt, H., Evert, M., Reng, M., Klinkhammer-Schalke, M. and Teufel, A., 2018. Chemotherapy for metastatic colon

cancer: No effect on survival when the dose is reduced due to side effects. BMC Cancer, 18(455).

Murtaza, B., Hichami, A., Khan, A., Shimpukade, B., Ulven, T., Ozdener, M. and Khan, N., 2020. Novel GPR120 agonist TUG891 modulates fat taste perception and preference and activates tongue-brain-gut axis in mice. Journal of Lipid Research, 61(2), pp.133-142.

Mustafa, F., 2016. Switching away from pipotiazine palmitate: a naturalistic study. Therapeutic Advances in Psychopharmacology, 7(1), pp.25-28.

Mysinger, M., Carchia, M., Irwin, J. and Shoichet, B., 2012. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. Journal of Medicinal Chemistry, 55(14), pp.6582-6594.

Ncri.ie. 2021. [online] Available at: <https://www.ncri.ie/sites/ncri/files/pubs/annualreport2018_26112018.pdf> [Accessed 5 July 2021].

Ncri.ie. 2021. [online] Available at: <https://www.ncri.ie/sites/ncri/files/pubs/CancerIncidenceProjections_NCRI_fullreport_09042019_final.pdf> [Accessed 5 July 2021].

Nelson, M. and Dolder, C., 2006. Lapatinib: A Novel Dual Tyrosine Kinase Inhibitor with Activity in Solid Tumors. Annals of Pharmacotherapy, 40(2), pp.261-269.

Neubig, R., Spedding, M., Kenakin, T. and Christopoulos, A., 2003. International Union of Pharmacology Committee on Receptor Nomenclature and Drug Classification. XXXVIII. Update on Terms and Symbols in Quantitative Pharmacology. Pharmacological Reviews, 55(4), pp.597-606.

Neumeier, J. and Meister, G., 2021. siRNA Specificity: RNAi Mechanisms and Strategies to Reduce Off-Target Effects. Frontiers in Plant Science, 11.

Nygaard, R., Zou, Y., Dror, R., Mildorf, T., Arlow, D., Manglik, A., Pan, A., Liu, C., Fung, J., Bokoch, M., Thian, F., Kobilka, T., Shaw, D., Mueller, L., Prosser, R. and Kobilka, B., 2013. The Dynamic Process of β2-Adrenergic Receptor Activation. Cell, 152(3), pp.532-542.

O'Brien, E., Fuglestad, B., Lessen, H., Stetz, M., Lin, D., Marques, B., Gupta, K., Fleming, K. and Wand, A., 2020. Membrane Proteins Have Distinct Fast Internal Motion and Residual Conformational Entropy. Angewandte Chemie International Edition, 59(27), pp.11108-11114.

O'Donnell, A., Judson, I., Dowsett, M., Raynaud, F., Dearnaley, D., Mason, M., Harland, S., Robbins, A., Halbert, G., Nutley, B. and Jarman, M., 2004. Hormonal impact of the 17α-

hydroxylase/C17,20-lyase inhibitor abiraterone acetate (CB7630) in patients with prostate cancer. British Journal of Cancer, 90(12), pp.2317-2325.

Oh, D., Walenta, E., Akiyama, T., Lagakos, W., Lackey, D., Pessentheiner, A., Sasik, R., Hah, N., Chi, T., Cox, J., Powels, M., Di Salvo, J., Sinz, C., Watkins, S., Armando, A., Chung, H., Evans, R., Quehenberger, O., McNelis, J., Bogner-Strauss, J. and Olefsky, J., 2014. A Gpr120-selective agonist improves insulin resistance and chronic inflammation in obese mice. Nature Medicine, 20(8), pp.942-947.

Olsson, H., Petri, N., Erichsen, L., Malmberg, A. and Grundemar, L., 2017. Effect of Degarelix, a Gonadotropin-Releasing Hormone Receptor Antagonist for the Treatment of Prostate Cancer, on Cardiac Repolarisation in a Randomised, Placebo and Active Comparator Controlled Thorough QT/QTc Trial in Healthy Men. Clinical Drug Investigation, 37(9), pp.873-879.

Padma, V., 2015. An overview of targeted cancer therapy. BioMedicine, 5(4).

Pagadala, N., Syed, K. and Tuszynski, J., 2017. Software for molecular docking: a review. Biophysical Reviews, 9(2), pp.91-102.

Pal, A., Curtin, J. and Kinsella, G., 2021. In silico and in vitro screening for potential anticancer candidates targeting GPR120. Bioorganic & Medicinal Chemistry Letters, 31, p.127672.

Pal, A., Curtin, J. and Kinsella, G., 2021a. Structure based prediction of a novel GPR120 antagonist based on pharmacophore screening and molecular dynamics simulations. Computational and Structural Biotechnology Journal, 19, pp.6050-6063.

Palczewski, K., 2000. Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor. Science, [online] 289(5480), pp.739-745. Available at: <http://DOI: 10.1126/science.289.5480.739> [Accessed 21 April 2021].

Palucka, A. and Coussens, L., 2016. The Basis of Oncoimmunology. Cell, 164(6), pp.1233-1247.

Papadatos, G. and Overington, J., 2014. The ChEMBL database: a taster for medicinal chemists. Future Medicinal Chemistry, 6(4), pp.361-364.

Papadimitriou, M. and Papadimitriou, C., 2021. Antiangiogenic Tyrosine Kinase Inhibitors in Metastatic Colorectal Cancer: Focusing on Regorafenib. Anticancer Research, 41(2), pp.567-582.

Parker, S., Parker, M., Sah, R. and Sallee, F., 2005. Angiogenesis and rhodopsin-like receptors: A role for N-terminal acidic residues?. Biochemical and Biophysical Research Communications, 335(4), pp.983-992.

Pearce, R. and Zhang, Y., 2021. Toward the solution of the protein structure prediction problem. Journal of Biological Chemistry, 297(1), p.100870.

Perkins, A., Phillips, J., Kerkvliet, N., Tanguay, R., Perdew, G., Kolluri, S. and Bisson, W., 2014. A Structural Switch between Agonist and Antagonist Bound Conformations for a Ligand-Optimized Model of the Human Aryl Hydrocarbon Receptor Ligand Binding Domain. Biology, 3(4), pp.645-669.

Petrelli, F., Trevisan, F., Cabiddu, M., Sgroi, G., Bruschieri, L., Rausa, E., Ghidini, M. and Turati, L., 2020. Total Neoadjuvant Therapy in Rectal Cancer. Annals of Surgery, 271(3), pp.440-448.

Price, T., Peeters, M., Kim, T., Li, J., Cascinu, S., Ruff, P., Suresh, A., Thomas, A., Tjulandin, S., Zhang, K., Murugappan, S. and Sidhu, R., 2014. Panitumumab versus cetuximab in patients with chemotherapy-refractory wild-type KRAS exon 2 metastatic colorectal cancer (ASPECCT): a randomised, multicentre, open-label, non-inferiority phase 3 study. The Lancet Oncology, 15(6), pp.569-579.

Provasi, D., Artacho, M., Negri, A., Mobarec, J. and Filizola, M., 2011. Ligand-Induced Modulation of the Free-Energy Landscape of G Protein-Coupled Receptors Explored by Adaptive Biasing Techniques. PLoS Computational Biology, 7(10), p.e1002193.

Przydzial, M., Bhhatarai, B., Koleti, A., Vempati, U. and Schürer, S., 2013. GPCR ontology: development and application of a G protein-coupled receptor pharmacology knowledge framework. Bioinformatics, 29(24), pp.3211-3219.

Pubchem.ncbi.nlm.nih.gov. 2021. 9-Fluorenone. [online] Available at: <https://pubchem.ncbi.nlm.nih.gov/compound/9-Fluorenone> [Accessed 1 June 2021].

Qu, X., Wang, D. and Wu, B., 2020. Progress in GPCR structure determination. GPCRs, pp.3-22.

Rahman, M., Saha, T., Islam, K., Suman, R., Biswas, S., Rahat, E., Hossen, M., Islam, R., Hossain, M., Mamun, A., Khan, M., Ali, M. and Halim, M., 2020. Virtual screening, molecular dynamics and structure–activity relationship studies to identify potent approved drugs for Covid-19 treatment. Journal of Biomolecular Structure and Dynamics, pp.1-11.

Rasmussen, S., DeVree, B., Zou, Y., Kruse, A., Chung, K., Kobilka, T., Thian, F., Chae, P., Pardon, E., Calinski, D., Mathiesen, J., Shah, S., Lyons, J., Caffrey, M., Gellman, S., Steyaert, J., Skiniotis, G., Weis, W., Sunahara, R. and Kobilka, B., 2007. Crystal structure of the β2 adrenergic receptor–Gs protein complex. Nature, 477(7366), pp.549-555.

Razvi, M., Giardiello, F. and Law, J., 2017. DNA Mismatch Repair and Lynch Syndrome. Current Colorectal Cancer Reports, 13(3), pp.212-219.

Richard, D., Vouret-Craviari, V. and Pouysségur, J., 2001. Angiogenesis and G-protein-coupled receptors: signals that bridge the gap. Oncogene, 20(13), pp.1556-1562.

Rossouw, J., 2002. Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women: Principal Results From the Women's Health Initiative Randomized Controlled Trial. JAMA: The Journal of the American Medical Association, 288(3), pp.321-333.

Rubbino, F., Garlatti, V., Garzarelli, V., Massimino, L., Spanò, S., Iadarola, P., Cagnone, M., Giera, M., Heijink, M., Guglielmetti, S., Arena, V., Malesci, A., Laghi, L., Danese, S. and Vetrano, S., 2021. GPR120 Prevents Colorectal Adenocarcinoma Progression by Sustaining The Mucosal Barrier Integrity.

Ruiz-Torres, V., Encinar, J., Herranz-López, M., Pérez-Sánchez, A., Galiano, V., Barrajón-Catalán, E. and Micol, V., 2017. An Updated Review on Marine Anticancer Compounds: The Use of Virtual Screening for the Discovery of Small-Molecule Cancer Drugs. Molecules, 22(7), p.1037.

Ryu, J. and Kim, D., 2012. Protein structure optimization by side-chain positioning via beta-complex. Journal of Global Optimization, 57(1), pp.217-250.

Sailapathi, A., Gunalan, S., Somarathinam, K., Kothandan, G. and Kumar, D., 2021. Importance of Homology Modeling for Predicting the Structures of GPCRs. Homology Molecular Modeling - Perspectives and Applications.

Sahragardjoonegani, B., Beall, R., Kesselheim, A. and Hollis, A., 2021. Repurposing existing drugs for new uses: a cohort study of the frequency of FDA-granted new indication exclusivities since 1997. Journal of Pharmaceutical Policy and Practice, 14(3).

Salas-Estrada, L., Leioatts, N., Romo, T. and Grossfield, A., 2018. Lipids Alter Rhodopsin Function via Ligand-like and Solvent-like Interactions. Biophysical Journal, 114(2), pp.355-367.

Salem, M., Puccini, A. and Tie, J., 2020. Redefining Colorectal Cancer by Tumor Biology. American Society of Clinical Oncology Educational Book, (40), pp.147-159.

Salon, J., Lodowski, D. and Palczewski, K., 2011. The Significance of G Protein-Coupled Receptor Crystallography for Drug Discovery. Pharmacological Reviews, 63(4), pp.901-937.

Sancineto, L., Massari, S., Iraci, N. and Tabarrini, O., 2013. From Small to Powerful: The Fragments Universe and it's; Chem-Appeal;. Current Medicinal Chemistry, 20(11), pp.1355-1381.

Sandoval-Perez, A., Pluhackova, K. and Böckmann, R., 2017. Critical Comparison of Biomembrane Force Fields: Protein–Lipid Interactions at the Membrane Interface. Journal of Chemical Theory and Computation, 13(5), pp.2310-2321.

Sanjeevi, M., Hebbar, P., Aiswarya, N., Rashmi, S., Rahul, C., Mohan, A., Jeyakanthan, J. and Sekar, K., 2022. Methods and applications of machine learning in structure-based drug discovery. Advances in Protein Molecular and Structural Biology Methods, pp.405-437.

Sauer, R., Liersch, T., Merkel, S., Fietkau, R., Hohenberger, W., Hess, C., Becker, H., Raab, H., Villanueva, M., Witzigmann, H., Wittekind, C., Beissbarth, T. and Rödel, C., 2012. Preoperative Versus Postoperative Chemoradiotherapy for Locally Advanced Rectal Cancer: Results of the German CAO/ARO/AIO-94 Randomized Phase III Trial After a Median Follow-Up of 11 Years. Journal of Clinical Oncology, 30(16), pp.1926-1933.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J., White, D., Hartenstein, V., Eliceiri, K., Tomancak, P. and Cardona, A., 2012. Fiji: an open-source platform for biological-image analysis. Nature Methods, 9(7), pp.676-682.

Scholz, N., Langenhan, T. and Schöneberg, T., 2019. Revisiting the classification of adhesion GPCRs. Annals of the New York Academy of Sciences, 1456(1), pp.80-95.

SEER. 2021. Cancer of the Colon and Rectum - Cancer Stat Facts. [online] Available at: <https://seer.cancer.gov/statfacts/html/colorect.html> [Accessed 24 April 2021].

Senatorov, I. and Moniri, N., 2018. The role of free-fatty acid receptor-4 (FFA4) in human cancers and cancer cell lines. Biochemical Pharmacology, 150, pp.170-180.

Senatorov, I., Cheshmehkani, A., Burns, R., Singh, K. and Moniri, N., 2020. Carboxy-Terminal Phosphoregulation of the Long Splice Isoform of Free-Fatty Acid Receptor-4 Mediates β-Arrestin Recruitment and Signaling to ERK1/2. Molecular Pharmacology, 97(5), pp.304-313.

Shen, J., Li, W., Liu, G., Tang, Y. and Jiang, H., 2009. Computational Insights into the Mechanism of Ligand Unbinding and Selectivity of Estrogen Receptors. The Journal of Physical Chemistry B, 113(30), pp.10436-10444.

Sheng, R., Yang, L., Zhang, Y., Xing, E., Shi, R., Wen, X., Wang, H. and Sun, H., 2018. Discovery of novel selective GPR120 agonists with potent anti-diabetic activity by hybrid design. Bioorganic & Medicinal Chemistry Letters, 28(15), pp.2599-2604.

Shimpukade, B., Hudson, B., Hovgaard, C., Milligan, G. and Ulven, T., 2012. Discovery of a Potent and Selective GPR120 Agonist. Journal of Medicinal Chemistry, 55(9), pp.4511-4515.

Sica, D., White, W., Weber, M., Bakris, G., Perez, A., Cao, C., Handley, A. and Kupfer, S., 2011. Comparison of the Novel Angiotensin II Receptor Blocker Azilsartan Medoxomil vs Valsartan by Ambulatory Blood Pressure Monitoring. The Journal of Clinical Hypertension, 13(7), pp.467-472.

Sohel, M., Sultana, H., Sultana, T., Al Amin, M., Aktar, S., Ali, M., Rahim, Z., Hossain, M., Al Mamun, A., Amin, M. and Dash, R., 2022. Chemotherapeutic potential of hesperetin for cancer treatment, with mechanistic insights: A comprehensive review. Heliyon, p.e08815.

Somchai, P., Phongkitkarun, K., Kueanjinda, P., Jamnongsong, S., Vaeteewoottacharn, K., Luvira, V., Okada, S., Jirawatnotai, S. and Sampattavanich, S., 2020. Novel Analytical Platform For Robust Identification of Cell Migration Inhibitors. Scientific Reports, 10(931).

Son, S., Kim, N. and Im, D., 2021. Development of Free Fatty Acid Receptor 4 (FFA4/GPR120) Agonists in Health Science. Biomolecules & Therapeutics, 29(1), pp.22-30.

Son, S., Park, S., Koh, J. and Im, D., 2020. Free fatty acid receptor 4 (FFA4) activation ameliorates 2,4-dinitrochlorobenzene-induced atopic dermatitis by increasing regulatory T cells in mice. Acta Pharmacologica Sinica, 41(10), pp.1337-1347.

Sood, M., Bajaj, J. and Bajaj, J., 2018. Azilsartan Medoxomil: Angiotensin Receptor Blocker in the Treatment of Hypertension. International Journal of Medical and Dental Sciences, 2(1), p.100.

Sparks, S., Chen, G., Collins, J., Danger, D., Dock, S., Jayawickreme, C., Jenkinson, S., Laudeman, C., Leesnitzer, M., Liang, X., Maloney, P., McCoy, D., Moncol, D., Rash, V., Rimele, T., Vulimiri, P., Way, J. and Ross, S., 2014. Identification of diarylsulfonamides as agonists of the free fatty acid receptor 4 (FFA4/GPR120). Bioorganic & Medicinal Chemistry Letters, 24(14), pp.3100-3103.

Spugnini, E. and Fais, S., 2019. Drug repurposing for anticancer therapies. A lesson from proton pump inhibitors. Expert Opinion on Therapeutic Patents, 30(1), pp.15-25.

Sriram, K. and Insel, P., 2018. G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs?. Molecular Pharmacology, 93(4), pp.251-258.

Stansfeld, P. and Sansom, M., 2011. Molecular Simulation Approaches to Membrane Proteins. Structure, 19(11), pp.1562-1572.

Steers, W., 2006. Darifenacin: Pharmacology and Clinical Usage. Urologic Clinics of North America, 33(4), pp.475-482.

Steindl, T., Schuster, D., Laggner, C. and Langer, T., 2006. Parallel Screening: A Novel Concept in Pharmacophore Modeling and Virtual Screening. Journal of Chemical Information and Modeling, 46(5), pp.2146-2157.

Sterling, T. and Irwin, J., 2015. ZINC 15 – Ligand Discovery for Everyone. Journal of Chemical Information and Modeling, 55(11), pp.2324-2337.

Stock, K., Estrada, M., Vidic, S., Gjerde, K., Rudisch, A., Santo, V., Barbier, M., Blom, S., Arundkar, S., Selvam, I., Osswald, A., Stein, Y., Gruenewald, S., Brito, C., van Weerden, W., Rotter, V., Boghaert, E., Oren, M., Sommergruber, W., Chong, Y., de Hoogt, R. and Graeser, R., 2016. Capturing tumor complexity in vitro: Comparative analysis of 2D and 3D tumor models for drug discovery. Scientific Reports, 6(28951).

Stone, V., Dhayal, S., Brocklehurst, K., Lenaghan, C., Sörhede Winzell, M., Hammar, M., Xu, X., Smith, D. and Morgan, N., 2014. GPR120 (FFAR4) is preferentially expressed in pancreatic delta cells and regulates somatostatin secretion from murine islets of Langerhans. Diabetologia, 57(6), pp.1182-1191.

Sun, Q., Hirasawa, A., Hara, T., Kimura, I., Adachi, T., Awaji, T., Ishiguro, M., Suzuki, T., Miyata, N. and Tsujimoto, G., 2010. Structure-Activity Relationships of GPR120 Agonists Based on a Docking Simulation. Molecular Pharmacology, 78(5), pp.804-810.

Sun, W. and Patel, A., 2013. Ziv-aflibercept in metastatic colorectal cancer. Biologics: Targets and Therapy, p.13.

Swegat, W., Schlitter, J., Krüger, P. and Wollmer, A., 2003. MD Simulation of Protein-Ligand Interaction: Formation and Dissociation of an Insulin-Phenol Complex. Biophysical Journal, 84(3), pp.1493-1506.

Takeuchi, M., Hirasawa, A., Hara, T., Kimura, I., Hirano, T., Suzuki, T., Miyata, N., Awaji, T., Ishiguro, M. and Tsujimoto, G., 2013. FFA1-selective agonistic activity based on docking simulation using FFA1 and GPR120 homology models. British Journal of Pharmacology, 168(7), pp.1570-1583.

Tanaka, T., Yano, T., Adachi, T., Koshimizu, T., Hirasawa, A. and Tsujimoto, G., 2008. Cloning and characterization of the rat free fatty acid receptor GPR120: in vivo effect of the natural ligand on GLP-1 secretion and proliferation of pancreatic β cells. Naunyn-Schmiedeberg's Archives of Pharmacology, 377(4-6), pp.515-522.

Tang, T., Hartig, C., Chen, Q., Zhao, W., Kaiser, A., Zhang, X., Zhang, H., Qu, H., Yi, C., Ma, L., Han, S., Zhao, Q., Beck-Sickinger, A. and Wu, B., 2021. Structural basis for ligand recognition of the neuropeptide Y Y2 receptor. Nature Communications, 12(732).

Tapial, S., Olmedillas-López, S., Rueda, D., Arriba, M., García, J., Vivas, A., Pérez, J., Pena-Couso, L., Olivera, R., Rodríguez, Y., García-Arranz, M., García-Olmo, D., González-Sarmiento, R., Urioste, M., Goel, A. and Perea, J., 2019. Cimp-Positive Status is More Representative in Multiple Colorectal Cancers than in Unique Primary Colorectal Cancers. Scientific Reports, 9(10516).

Teramoto, R. and Fukunishi, H., 2007. Supervised Consensus Scoring for Docking and Virtual Screening. Journal of Chemical Information and Modeling, 47(2), pp.526-534.

Tetko, I., Bruneau, P., Mewes, H., Rohrer, D. and Poda, G., 2006. Can we estimate the accuracy of ADME–Tox predictions?. Drug Discovery Today, 11(15-16), pp.700-707.

Tian, C., Chen, K., Gong, W., Yoshimura, T., Huang, J. and Wang, J., 2020. The G-Protein Coupled Formyl Peptide Receptors and Their Role in the Progression of Digestive Tract Cancer. Technology in Cancer Research & Treatment, 19, p.153303382097328.

Tiss, A., Ben Boubaker, R., Henrion, D., Guissouma, H. and Chabbert, M., 2021. Homology Modeling of Class A G-Protein-Coupled Receptors in the Age of the Structure Boom. Methods in Molecular Biology, pp.73-97.

Tong, G., Zhang, G., Liu, J., Zheng, Z., Chen, Y., Niu, P. and Xu, X., 2018. Comparison of the eighth version of the American Joint Committee on Cancer manual to the seventh version for colorectal cancer: A retrospective review of our data. World Journal of Clinical Oncology, 9(7), pp.148-161.

Tumeh, P., Harview, C., Yearley, J., Shintaku, I., Taylor, E., Robert, L., Chmielowski, B., Spasic, M., Henry, G., Ciobanu, V., West, A., Carmona, M., Kivork, C., Seja, E., Cherry, G., Gutierrez, A., Grogan, T., Mateus, C., Tomasic, G., Glaspy, J., Emerson, R., Robins, H., Pierce, R., Elashoff, D., Robert, C. and Ribas, A., 2014. PD-1 blockade induces responses by inhibiting adaptive immune resistance. Nature, 515(7528), pp.568-571.

Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhori, G., Benfeitas, R., Arif, M., Liu, Z., Edfors, F., Sanli, K., von Feilitzen, K., Oksvold, P., Lundberg, E., Hober, S.,

Nilsson, P., Mattsson, J., Schwenk, J., Brunnström, H., Glimelius, B., Sjöblom, T., Edqvist, P., Djureinovic, D., Micke, P., Lindskog, C., Mardinoglu, A. and Ponten, F., 2017. A pathology atlas of the human cancer transcriptome. Science, 357(6352), p.2507.

Ul-Haq, Z., Uddin, R. and Gul, S., 2011. Optimization of Structure Based Virtual Screening Protocols Against Thymidine Monophosphate Kinase Inhibitors as Antitubercular Agents. Molecular Informatics, 30(10), pp.851-862.

Ungaro, F., D'Alessio, S. and Danese, S., 2020. The Role of Pro-Resolving Lipid Mediators in Colorectal Cancer-Associated Inflammation: Implications for Therapeutic Strategies. Cancers, 12(8), p.2060.

Usman, S., Khawer, M., Rafique, S., Naz, Z. and Saleem, K., 2020. The current status of anti-GPCR drugs against different cancers. Journal of Pharmaceutical Analysis, 10(6), pp.517-521.

Venkatakrishnan, A., Ma, A., Fonseca, R., Latorraca, N., Kelly, B., Betz, R., Asawa, C., Kobilka, B. and Dror, R., 2019. Diverse GPCRs exhibit conserved water networks for stabilization and activation. Proceedings of the National Academy of Sciences, 116(8), pp.3288-3293.

Voet, A., Callewaert, L., Ulens, T., Vanderkelen, L., Vanherreweghe, J., Michiels, C. and De Maeyer, M., 2011. Structure based discovery of small molecule suppressors targeting bacterial lysozyme inhibitors. Biochemical and Biophysical Research Communications, 405(4), pp.527-532.

Voet, A., Kumar, A., Berenger, F. and Zhang, K., 2014. Combining in silico and in cerebro approaches for virtual screening and pose prediction in SAMPL4. Journal of Computer-Aided Molecular Design, 28(4), pp.363-373.

Voet, A., Qing, X., Lee, X., De Raeymaecker, J., Tame, J., Zhang, K. and De Maeyer, M., 2014a. Pharmacophore modeling: advances, limitations, and current utility in drug discovery, 7, pp.81-92.

Voskoglou-Nomikos, T., Pater, J. and Seymour, L., 2003. Clinical predictive value of the in vitro cell line, human xenograft, and mouse allograft preclinical cancer models. Clin. Cancer Res., 9(11), pp.4227–4239.

Wacker, D., Wang, C., Katritch, V., Han, G., Huang, X., Vardy, E., McCorvy, J., Jiang, Y., Chu, M., Siu, F., Liu, W., Xu, H., Cherezov, V., Roth, B. and Stevens, R., 2013. Structural Features for Functional Selectivity at Serotonin Receptors. Science, 340(6132), pp.615-619.

Wang, J., Tannous, B., Poznansky, M. and Chen, H., 2020. CXCR4 antagonist AMD3100 (plerixafor): From an impurity to a therapeutic agent. Pharmacological Research, 159, p.105010.

Wang, X., He, S., Gu, Y., Wang, Q., Chu, X., Jin, M., Xu, L., Wu, Q., Zhou, Q., Wang, B., Zhang, Y., Wang, H. and Zheng, L., 2019. Fatty acid receptor GPR120 promotes breast cancer chemoresistance by upregulating ABC transporters expression and fatty acid synthesis. EBioMedicine, 40, pp.251-262.

Wang, Y. and Chan, Y., 2017. Understanding the molecular basis of agonist/antagonist mechanism of human mu opioid receptor through gaussian accelerated molecular dynamics method. Scientific Reports, 7(7828).

Wang, Y., Xie, T., Zhang, D. and Leung, P., 2019. GPR120 protects lipotoxicity-induced pancreatic β-cell dysfunction through regulation of PDX1 expression and inhibition of islet inflammation. Clinical Science, 133(1), pp.101-116.

Wang, Z., Sun, H., Shen, C., Hu, X., Gao, J., Li, D., Cao, D. and Hou, T., 2020. Combined strategies in structure-based virtual screening. Physical Chemistry Chemical Physics, 22(6), pp.3149-3159.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F., de Beer, T., Rempfer, C., Bordoli, L., Lepore, R. and Schwede, T., 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Research, 46(W1), pp.W296-W303.

Watson, S., Brown, A. and Holliday, N., 2012. Differential Signaling by Splice Variants of the Human Free Fatty Acid Receptor GPR120. Molecular Pharmacology, 81(5), pp.631-642.

Watterson, K., Hansen, S., Hudson, B., Alvarez-Curto, E., Raihan, S., Azevedo, C., Martin, G., Dunlop, J., Yarwood, S., Ulven, T. and Milligan, G., 2017. Probe-Dependent Negative Allosteric Modulators of the Long-Chain Free Fatty Acid Receptor FFA4. Molecular Pharmacology, 91(6), pp.630-641.

Webb, B. and Sali, A., 2016. Comparative Protein Structure Modeling Using MODELLER. Current Protocols in Protein Science, 86(1), Chapter 2: Unit2.9.

Wei, G., 2019. Protein structure prediction beyond AlphaFold. Nature Machine Intelligence, 1(8), pp.336-337.

Weis, W. and Kobilka, B., 2018. The Molecular Basis of G Protein–Coupled Receptor Activation. Annual Review of Biochemistry, 87(1), pp.897-919.

Wermuth, C., Ganellin, C., Lindberg, P. and Mitscher, L., 1998. Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). Pure and Applied Chemistry, 70(5), pp.1129-1143.

WHO, 2022. Cancer. [online] Who.int. Available at: <https://www.who.int/news-room/fact-sheets/detail/cancer> [Accessed 5 May 2022].

Wiederstein, M. and Sippl, M., 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research, 35(Web Server), pp.W407-W410.

Wild, C., Cunningham, KA., Zhou, J., 2014. Allosteric Modulation of G Protein-Coupled Receptors: An Emerging Approach of Drug Discovery. Austin J Pharmacol Ther., 2 (1): 1011. ISSN: 2373-6208.

Wishart, D., 2006. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Research, 34(90001), pp.D668-D672.

Wishart, D., Feunang, Y., Guo, A., Lo, E., Marcu, A., Grant, J., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., Pon, A., Knox, C. and Wilson, M., 2017. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Research, 46(D1), pp.D1074-D1082.

Wong, C., 2018. Steered molecular dynamics simulations for uncovering the molecular mechanisms of drug dissociation and for drug screening: A test on the focal adhesion kinase. Journal of Computational Chemistry, 39(19), pp.1307-1318.

Wu, Q., Wang, H., Zhao, X., Shi, Y., Jin, M., Wan, B., Xu, H., Cheng, Y., Ge, H. and Zhang, Y., 2013. Identification of G-protein-coupled receptor 120 as a tumor-promoting receptor that induces angiogenesis and migration in human colorectal carcinoma. Oncogene, 32(49), pp.5541-5550.

Xie, Y., Chen, Y. and Fang, J., 2020. Comprehensive review of targeted therapy for colorectal cancer. Signal Transduction and Targeted Therapy, 5(1), SN.2059-3635.

Yang, L., Keam, S. and Keating, G., 2007. Deferasirox. Drugs, 67(15), pp.2211-2230.

Yang, L., Radu, C., Roy, M., Lee, S., McLaughlin, J., Teitell, M., Iruela-Arispe, M. and Witte, O., 2007. Vascular Abnormalities in Mice Deficient for the G Protein-Coupled Receptor GPR4 That Functions as a pH Sensor. Molecular and Cellular Biology, 27(4), pp.1334-1347.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. and Madden, T., 2012. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics, 13(134).

Yin, L., Li, J., Ma, D., Li, D. and Sun, Y., 2020. Angiogenesis in primary colorectal cancer and matched metastatic tissues: Biological and clinical implications for anti-angiogenic therapies. Oncology Letters, 19(5), pp.3558-3566.

Yuan, X. and Xu, Y., 2018. Recent Trends and Applications of Molecular Modeling in GPCR–Ligand Recognition and Structure-Based Drug Design. International Journal of Molecular Sciences, 19(7), p.2105.

Yucel, I., 2015. Quality of life in colorectal cancer patients during chemotherapy in the era of monoclonal antibody therapies. Journal of Clinical Oncology, 33(15_suppl), pp.e14525-e14525.

Yun, C., 2019. Lysophosphatidic Acid and Autotaxin-associated Effects on the Initiation and Progression of Colorectal Cancer. Cancers, 11(7), p.958.

Zakharia, Y., Monga, V., Swami, U., Bossler, A., Freesmeier, M., Frees, M., Khan, M., Frydenlund, N., Srikantha, R., Vanneste, M., Henry, M. and Milhem, M., 2017. Targeting epigenetics for treatment of BRAF mutated metastatic melanoma with decitabine in combination with vemurafenib: A phase lb study. Oncotarget, 8(51), pp.89182-89193.

Zeng, H., Zhao, D., Yang, S., Datta, K. and Mukhopadhyay, D., 2003. Heterotrimeric Gαq/Gα11 Proteins Function Upstream of Vascular Endothelial Growth Factor (VEGF) Receptor-2 (KDR) Phosphorylation in Vascular Permeability Factor/VEGF Signaling. Journal of Biological Chemistry, 278(23), pp.20738-20745.

Zhang, R. and Xie, X., 2012. Tools for GPCR drug discovery. Acta Pharmacologica Sinica, 33(3), pp.372-384.

Zhang, X., Sun, H., Wen, X. and Yuan, H., 2019. A Selectivity Study of FFAR4/FFAR1 Agonists by Molecular Modelling. Journal of Chemical Information and Modelling, 59(10), pp.4467-4474.

Zhou, J. and Wild, C., 2019. GPCR Drug Discovery: Emerging Targets, Novel Approaches and Future Trends. Current Topics in Medicinal Chemistry, 19(16), pp.1363-1364.

Zhou, Q., Yang, D., Wu, M., Guo, Y., Guo, W., Zhong, L., Cai, X., Dai, A., Jang, W., Shakhnovich, E., Liu, Z., Stevens, R., Lambert, N., Babu, M., Wang, M. and Zhao, S., 2019. Common activation mechanism of class A GPCRs. eLife, 8.

Zhu, S., Wu, M., Huang, Z. and An, J., 2021. Trends in application of advancing computational approaches in GPCR ligand discovery. Experimental Biology and Medicine, 246(9), pp.1011-1024.

Zhu, T., Cao, S., Su, P., Patel, R., Shah, D., Chokshi, H., Szukala, R., Johnson, M. and Hevener, K., 2013. Hit Identification and Optimization in Virtual Screening: Practical Recommendations Based on a Critical Literature Analysis. Journal of Medicinal Chemistry, 56(17), pp.6560-6572.

Zimmermann, I., Egloff, P., Hutter, C., Arnold, F., Stohler, P., Bocquet, N., Hug, M., Huber, S., Siegrist, M., Hetemann, L., Gera, J., Gmür, S., Spies, P., Gygax, D., Geertsma, E., Dawson, R. and Seeger, M., 2018. Synthetic single domain antibodies for the conformational trapping of membrane proteins. eLife, 7.

Zollinger, M., Lozac'h, F., Hurh, E., Emotte, C., Bauly, H. and Swart, P., 2014. Absorption, distribution, metabolism, and excretion (ADME) of 14C-sonidegib (LDE225) in healthy volunteers. Cancer Chemotherapy and Pharmacology, 74(1), pp.63-75.

Zhu, S., Wu, M., Huang, Z. and An, J., 2021. Trends in application of advancing computational approaches in GPCR ligand discovery. Experimental Biology and Medicine, 246(9), pp.1011-1024.

Appendix I: Publications and disseminations

Paper 1:

The G-protein coupled receptor - GPR120 has recently been implicated as a novel target for colorectal cancer (CRC) and other cancer managements. In this study, a homology model of GPR120S (short isoform) was generated to identify potential anti-cancer compounds targeting the GPR120 receptor using a combined *in silico* docking-based virtual screening (DBVS), structure–activity relationships (SAR) and *in vitro* screening approach. SPECS database of synthetic chemical compounds (~350,000) was screened using the developed GPR120S model to identify molecules binding to the orthosteric binding pocket followed by an AutoDock SMINA rigid-flexible docking protocol.

The best 13 hit molecules were then tested *in vitro* to evaluate their cytotoxic activity against SW480 – human CRC cell line expressing GPR120. The test compound **1** (3-(4-methylphenyl)-2-[(2-oxo-2-phenylethyl)sulfanyl]-5,6-dihydrospiro(benzo[h]quinazoline-5,1′-cyclopentane)-4(3H)-one) showed ~ 90% inhibitory effects on cell growth with micromolar affinities (IC$_{50}$ = 23.21–26.69 μM). Finally, SAR analysis of compound **1** led to the identification of a more active compound from the SPECS database showing better efficacy

during cell-based cytotoxicity assay –**5** (IC$_{50}$ = 5.89–6.715 µM), while a significant reduction in cytotoxic effects of **5** was observed in GPR120-siRNA pre-treated SW480 cells.

The GPR120S homology model generated, and SAR analysis conducted by this work discovered a potential chemical scaffold, *dihydrospiro(benzo[h]quinazoline-5,1'-cyclopentane)-4(3H)-one*, which will aid future research on anti-cancer drug development for CRC management.

Paper 2:

Structure based prediction of a novel GPR120 antagonist based on pharmacophore screening and molecular dynamics simulations.

Ajay Pal; James F. Curtin; Gemma K. Kinsella
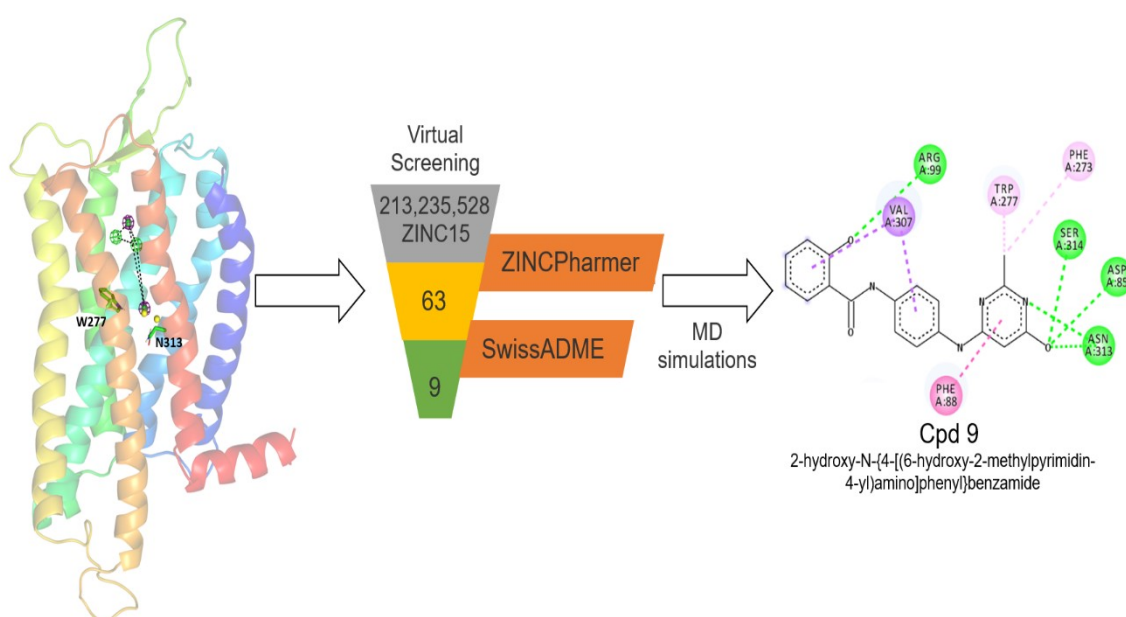
The G-protein coupled receptor, GPR120, has ubiquitous expression and multifaceted roles in modulating metabolic and anti-inflammatory processes. Recent implications of its role in cancer progression have presented GPR120 as an attractive oncogenic drug target. GPR120 gene knockdown in breast cancer studies revealed a role of GPR120-induced chemoresistance in epirubicin and cisplatin-induced DNA damage in tumour cells. Higher expression and activation levels of GPR120 is also reported to promote tumour angiogenesis and cell migration in colorectal cancer. Some agonists targeting GPR120 have been reported,

such as TUG891 and Compound39, but to date development of small-molecule inhibitors of GPR120 is limited.

Herein, following homology modelling of the receptor a pharmacophore hypothesis was derived from 300 ns all-atomic molecular dynamics (MD) simulations on apo, TUG891-bound and Compound39-bound GPR120S (short isoform) receptor models embedded in a water solvated lipid bilayer system. We performed comparative MD analysis on protein-ligand interactions between the two agonist and apo simulations on the stability of the "ionic lock" – a Class A GPCRs characteristic of receptor activation and inactivation. The detailed analysis predicted that ligand interactions with W277 and N313 are critical to conserve the "ionic-lock" conformation (R136 of Helix 3) and prevent GPR120S receptor activation. The results led to generation of a W277 and N313 focused pharmacophore hypothesis and the screening of the ZINC15 database using ZINCPharmer through the structure-based pharmacophore. 100 ns all-atomic molecular dynamics (MD) simulations were performed on 9 small molecules identified and Cpd **9**, (2-hydroxy-N-{4-[(6-hydroxy-2-methylpyrimidin-4-yl) amino] phenyl} benzamide) was predicted to be a small-molecule GPR120S antagonist. The conformational results from the collective all-atomic MD analysis provided structural information for further identification and optimisation of novel druggable inhibitors of GPR120S using this rational design approach, which could have future potential for anti-cancer drug development studies.

Paper 3:

G-protein-coupled receptors as therapeutic targets for glioblastoma.

Kate F. Byrne; Ajay Pal; James F. Curtin; John C. Stephens; Gemma K. Kinsella
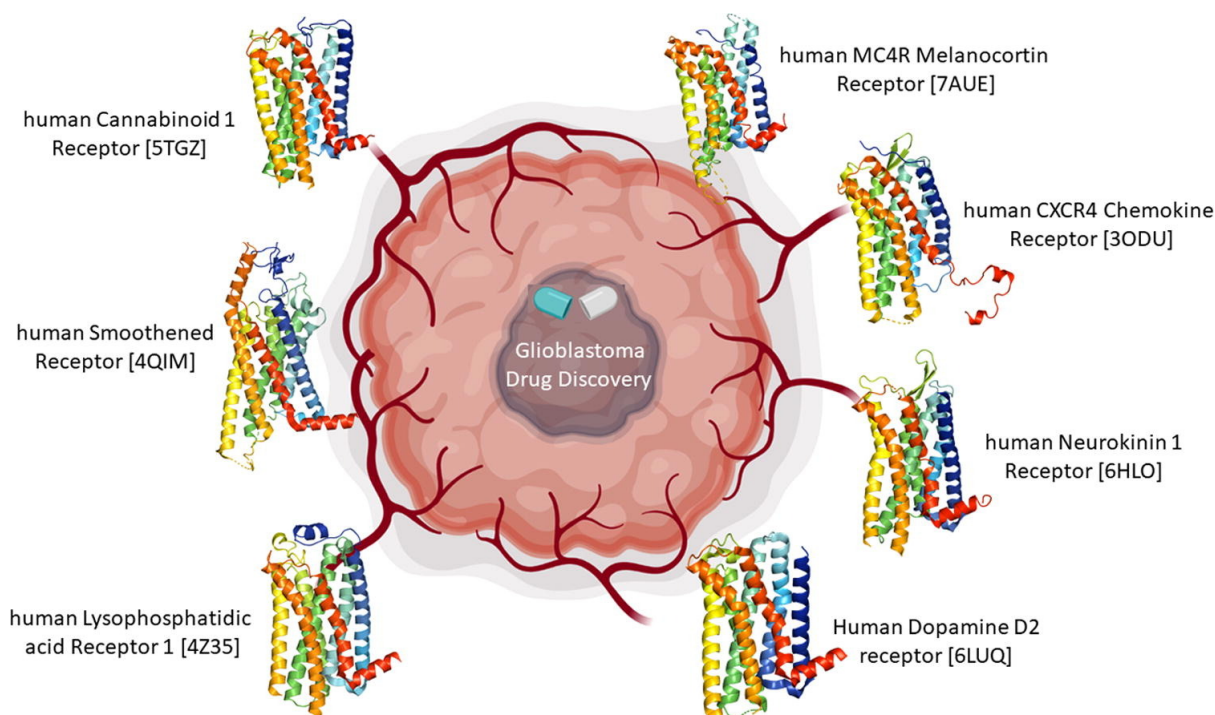
Glioblastoma multiforme (GBM) is the most common and aggressive primary brain tumour in adults. Treatments include surgical resection, radiotherapy, and chemotherapy. Despite this, the prognosis remains poor, with an impacted quality of life during treatment coupled with brain tumour recurrence; thus, new treatments are desperately needed. In this review, we focus on recent advances in G-protein-coupled receptor (GPCR) targets. To date, the most promising targets are the chemokine, cannabinoid, and dopamine receptors, but future work should further examine the melanocortin receptor-4 (MC4R), adhesion, lysophosphatidic acid (LPA) and smoothened (Smo) receptors to initiate new drug-screening strategies and targeted delivery of safe and effective GBM therapies.

List of posters and presentations	
Poster + Flash Presentation: (February 2021)	MGMS Young Modellers' Forum 2020/21: In silico structural & in vitro functional analysis of GPR120 receptor to screen potential anticancer candidates.

Poster Presentation: (May 2018)	UK-QSAR and MGMS: Structure activity Relationships, Cardiff: G-Protein Coupled Receptor-KNIME Automated Molecular Modelling Platform: GPCR_KAMP
Poster Presentation: (April 2018)	CADD symposium and workshop: linking Design, Biology, Chemistry and Medicine, Trinity College Dublin: G-Protein Coupled Receptor-KNIME Automated Molecular Modelling Platform: GPCR_KAMP

Appendix II: Supplementary tables and figures

Appendix Table IIa: Combinations of multiple templates (PDB codes) used to build GPR120 homology models. 4S0V (OX2 orexin receptor); 4N6H (Delta-like opioid receptor); 4DJH (kappa-opioid receptor); 5GLI (Endothelin type-B receptor); 4EA3 (Nociceptin/Orphanin FQ opioid receptor).
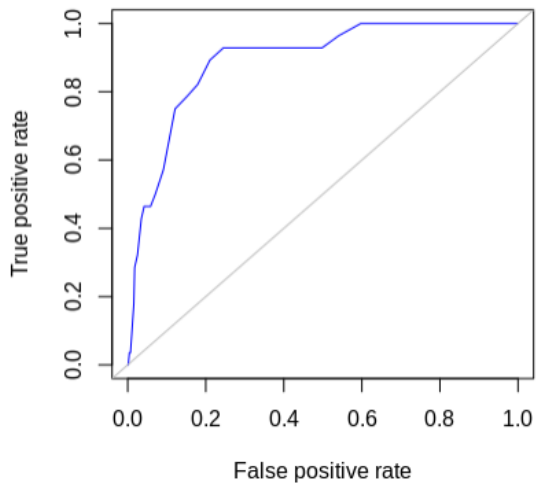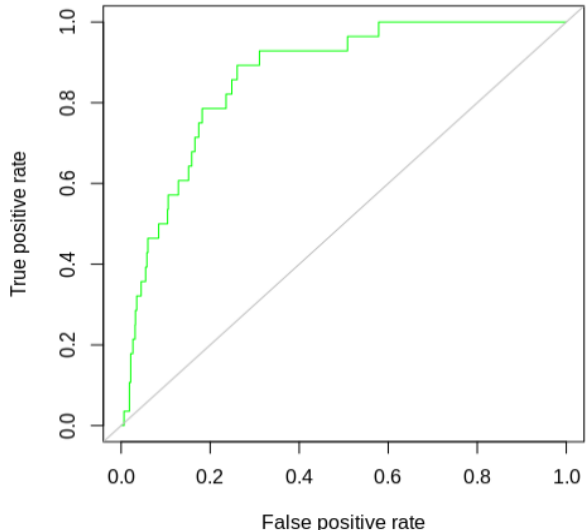
4S0V	4S0V-4N6H	4DJH-5GLI	4S0V-4N6H-4DJH	4DJH-4EA3-5GLI
4N6H	4S0V-4DJH	4DJH-4EA3	4S0V-4N6H-5GLI	4S0V-4N6H-4DJH-5GLI
4DJH	4S0V-5GLI	4N6H-4DJH	4S0V-4N6H-4EA3	4S0V-4N6H-4DJH-4EA3
5GLI	4S0V-4EA3	4N6H-5GLI	4N6H-4DJH-5GLI	4N6H-4DJH-5GLI-4EA3
4EA3	5GLI-4EA3	4N6H-4EA3	4N6H-4DJH-4EA3	4S0V-4N6H-4DJH-5GLI-4EA3

Appendix IIb: GPR120 actives used for Decoy generation (Lombardo et al., 2016; Sparks et al., 2014; Azevedo et al., 2016; Hudson, Shimpukade, Milligan and Ulven, 2014)

Compound name	Binding Score (kcal/mol)	EC$_{50}$ (nM)	SMILES representation
Compound48	-9.9143	83	c1cc2c(cc1c3cc(ccc3F)OC(F)(F)F)cc(o2)CCC(=O)[O-]
TUG891	-9.87591	43	Cc1ccc(cc1)c2ccc(cc2COc3ccc(cc3)CCC(=O)[O-])F
Compound40	-9.87491	102	c1cc2c(cc1c3cc(ccc3Cl)OC(F)(F)F)cc(o2)CCC(=O)[O-]
PartialA1	-9.82831	NA	CC#C[C@@H](CC(=O)[O-])c1ccc(s1)OCc2cccc(c2)c3c(cc(cc3C)OCCCS(=O)(=O)C)C
Compound39	-9.82688	97	c1cc2c(cc1c3cc(ccc3F)C(=O)C4CC4)cc(o2)CCC(=O)[O-]
Compound18	-9.7935	88	c1cc(cnc1)Oc2ccc(c(c2)c3ccc(cc3)OCCCC(=O)[O-])Cl
Compound10	-9.7395	100	c1ccc(cc1)Oc2ccc(c(c2)c3ccc(cc3)OCCCC(=O)[O-])F
Compound41	-9.5415	68	Cc1cc(cc(c1)OC2CCC2)c3ccc4c(c3)cc(o4)CCC(=O)[O-]
Compound5	-9.4734	170	c1ccc(cc1)Oc2cccc(c2)c3ccc(cc3)OCCCC(=O)[O-]
TUG1197	-9.44996	128	c1ccnc(c1)Oc2cc(cc(c2)F)N3C[C@@H]4C=CC=C[C@@H]4S3(=O)=O
TUG1506	-9.2101	NA	Cc1ccc(cc1)S(=O)(=O)C[C@H]2c3ccccc3S[C@@H]4[C@H]2C=CC=C4
Compound37	-9.1646	63	c1cc2c(cc1c3cc(ccc3F)OC(F)(F)F)cc(o2)CCC(=O)[O-]
Compound27	-9.11864	149	Cc1cc(cc(c1OCCCC(=O)[O-])C)c2cc(ccc2F)OC(F)(F)F
Compound6	-9.0261	57	c1cc(ccc1c2cc(ccc2Cl)OC(F)(F)F)OCCCC(=O)[O-]
CpdA	-9.02491	66	c1cc(c(cc1OC(F)(F)F)N2CCC3(CCC(CC3)CC(=O)[O-])CC2)Cl
Compound19	-8.92716	60	c1cc(cc(c1)F)Oc2ccc(c(c2)c3ccc(cc3)OCCCC(=O)[O-])F
AH7614	-8.90068	79.43	Cc1ccc(cc1)S(=O)(=O)N[C@H]2c3c(cccc3)Oc4c2cccc4
Compound20	-8.73304	182	Cc1ccc(cc1)Oc2cccc(c2)c3ccc(cc3)OCCCC(=O)[O-]
Compound28	-8.59616	20	c1cc(c(cc1OC(F)(F)F)c2cc(c(c(c2)F)OCCCC(=O)[O-])F)F
PartialA2_full	-8.56813	NA	Cc1cc(cc(c1c2cccc(c2)COc3ccc4c(c3)OC[C@H]4CC(=O)[O-])C)OCCCS(=O)(=O)C
Compound23	-8.30635	94	Cc1cc(cc(c1OCCCC(=O)[O-])C)c2cc(ccc2Cl)OC(F)(F)F
Compound8	-8.24819	290	CCOc1ccc(c(c1)c2ccc(cc2)OCCCC(=O)[O-])F
Compound24	-7.91255	185	Cc1cc(ccc1OCCCC(=O)[O-])c2cc(ccc2Cl)OC(F)(F)F
Compound1	-7.87396	474	c1ccc2c(c1)c(no2)c3ccc(c(c3Cl)Cl)OCCCC(=O)[O-]
GSK39	-7.60468	79	Cc1ccc(cc1)S(=O)(=O)[N-]c2cccc(c2C)C(=O)OC

Compound17	-7.57706	198	Cc1ccc(cc1c2ccc(cc2)OCCCC(=O)[O-])Oc3cccc(c3)F
GSK137647A	-7.46444	398.11	Cc1cc(c(c(c1)C)NS(=O)(=O)c2ccc(cc2)OC)C
Compound7	-7.40539	121	CCOc1ccc(c(c1)c2ccc(cc2)OCCCC(=O)[O-])Cl

Appendix IIc: ROC and enrichment studies to benchmark different scoring functions.

Docking Algorithm: SMINA Scoring Algorithm: SMINA AUC obtained: 0.892295 Top 1% EF score: 9.735	
Docking Algorithm: SMINA Scoring Algorithm: VINA AUC obtained: 0.8633163 Top 1% EF score: 3.33	

Docking Algorithm: SMINA Scoring Algorithm: NNScore AUC obtained: 0.546352 Top 1% EF score: 0	 **NNScore ROC Curves**
Docking Algorithm: SMINA Scoring Algorithm: DLScore AUC obtained: 0.6107143 Top 1% EF score: 0	 **DLScore ROC Curves**
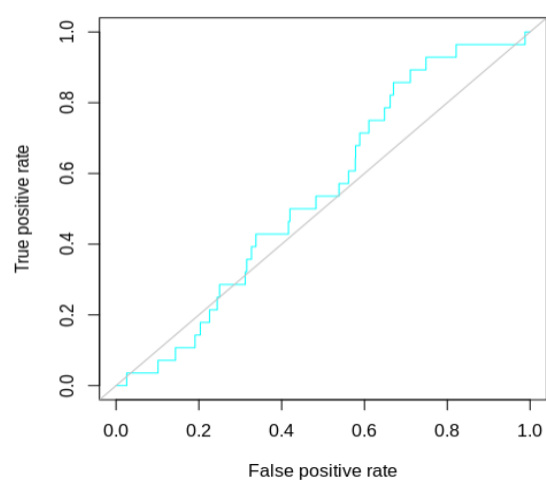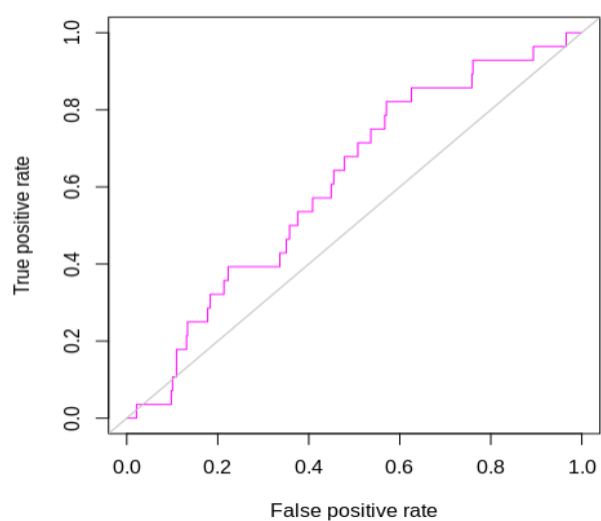
Scoring Algorithm: CScore (obtained by Equation I) AUC obtained: 0.7114976 Top 1% EF score: 3.33	 **C_Score ROC Curves**
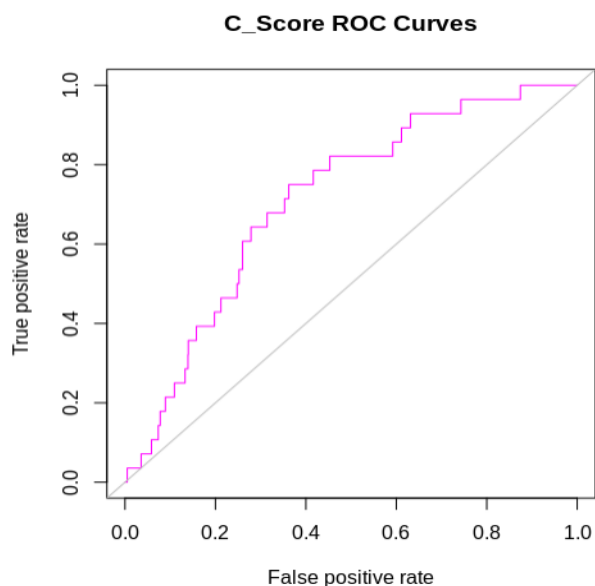
```
########### ROC Script written in R to find ROC – AUC and Enrichment factors####################
library("ROCR")
lig <- unique(Actives$Ligands)
dec <- unique(Decoy$Compounds)
#merging active and decoy and  #chnaging colnames to identicals
colnames(Decoy) =c("Ligands","SMINA","VINA","Nnscore","Dlscore")
ROC$IsActive=as.numeric(ROC$Ligands %in% lig)
predSminaScore = prediction((ROC$SMINA*-1), ROC$IsActive)
perfSmina = performance(predSminaScore, 'tpr','fpr')
jpeg("test_ROC.jpg")
plot(perfSmina,main="SMINA ROC Curves",col="blue")
abline(0,1,col="grey")
dev.off()
#AUC
auc_SMINA=ROCR::performance(predSminaScore,"auc")
auc.area_SMINA= slot(auc_SMINA,"y.values")[[1]]
cat(auc.area_SMINA)
#Enrichment
EF_Smina=perfSmina@y.values[[1]]/perfSmina@x.values[[1]]
EF_Smina1=EF_Smina[which(perfSmina@x.values[[1]]>0.01)[1]]
EF_Smina20=EF_Smina[which(perfSmina@x.values[[1]]>0.2)[1]]
cat(EF_Smina1)
cat(EF_Smina20)
#ROC for VINA, NNScore, DLScore, Cscore can be calulcauted using the scripted blocks
#################################################################################
```

Appendix IId: RNA yield of the collected sample calculated using the µDrop.

	230nm	260nm	280nm	260/280	RNA Yield (ng/µL) = $(260)^{nm}*50*20$
Blank	0.0615	0.0791	0.0724	1.091851	-
GPR120_siRNA-1	0.7213	1.4750	0.7415	1.989122	1475.0
GPR120_siRNA-2	0.8813	1.7918	0.8895	2.01439	1791.8
Control	0.9113	1.9577	0.9733	2.011508	1957.7

Appendix IIe: List of pharmacophore features used to screen ZINC database using ZincPharmer

Pharmacophore feature	X-coordinate	Y-coordinate	Z-coordinate	Radii
Aromatic	61.42	60.36	40.99	1.1
Aromatic	62.41	60.4	52.66	1.1
Hydrogen Acceptor	62.66	61.45	53.17	0.5
Hydrogen Acceptor	64.23	60.23	54.29	0.5
Hydrophobic	61.42	60.36	40.99	1
Hydrophobic	62.41	60.4	52.66	1
Hydrophobic	61.02	57.04	44.38	1
Hydrophobic	59.04	60.5	42.67	1

Appendix IIf: Docked pose analysis of Cpd1-9 by 2D interaction map generated in BIOVIA DS Client visualizer v19.1 2019.

Some of the protein-ligands interactions are missing with respect to Figure 4.9c due to different calculation algorithm used by PLIP and DS client visualizer.

Cpd6

Cpd7

Cpd8

Cpd9
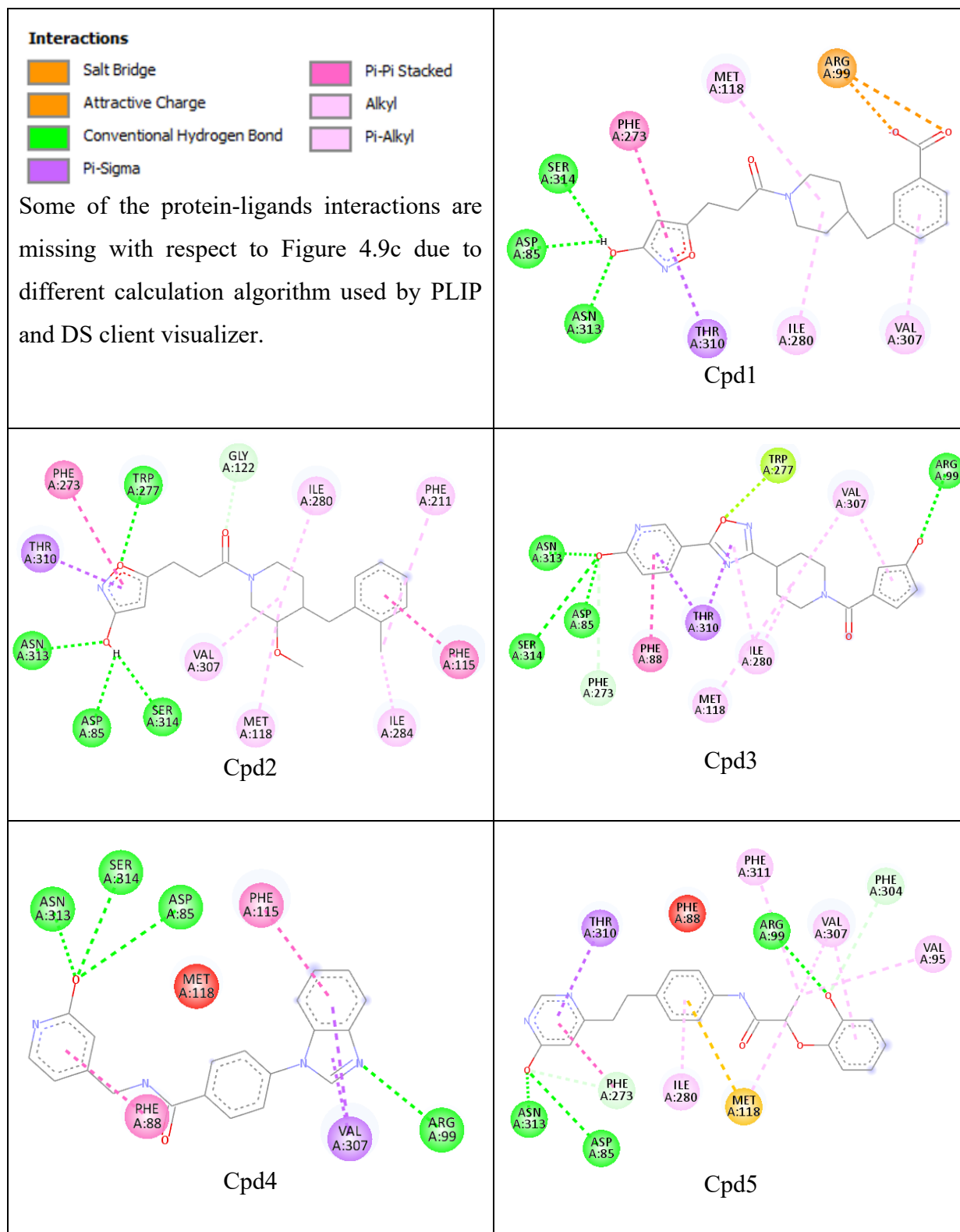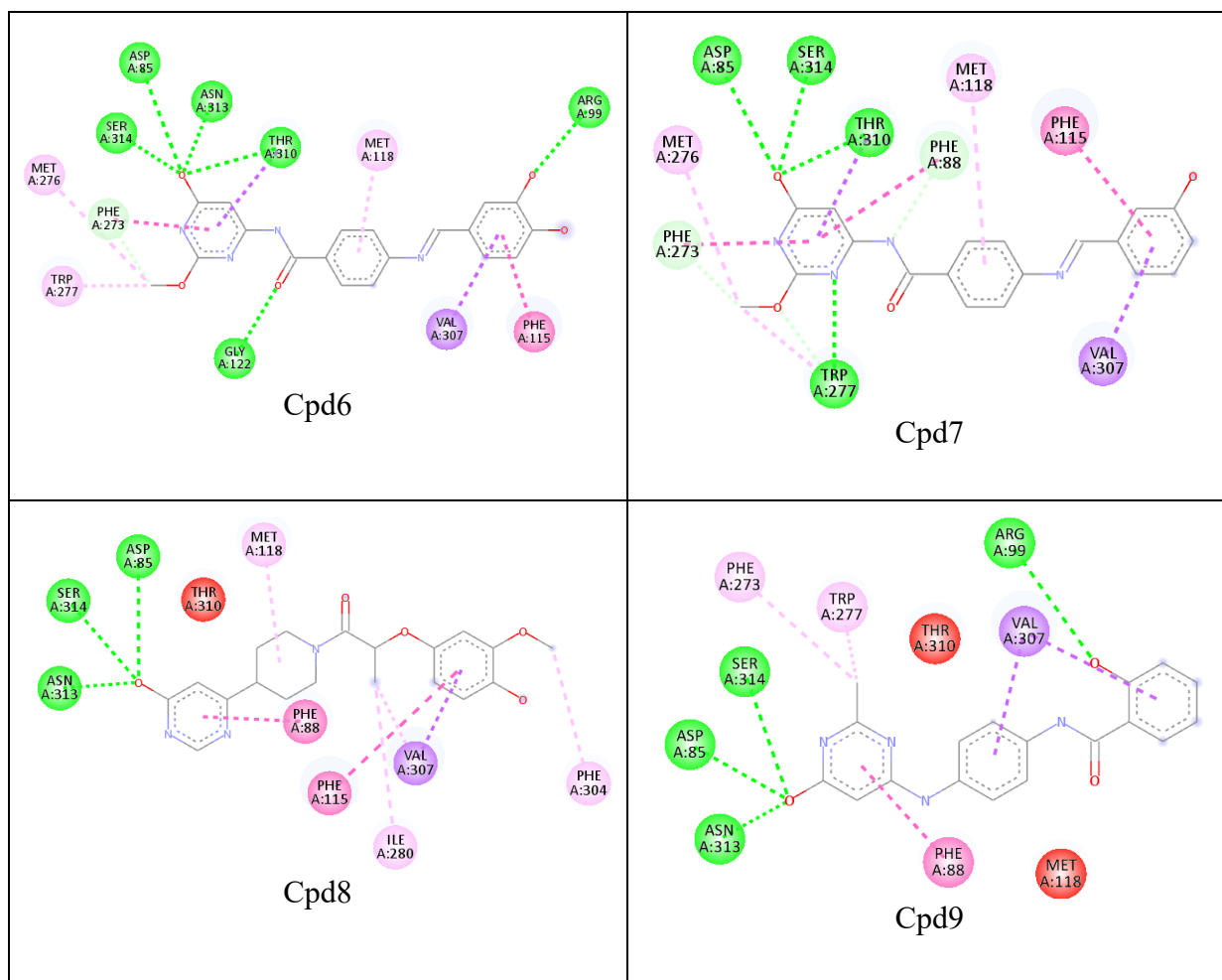
Appendix III: Benchmarking of docking algorithms

Four docking algorithms (AutoDock SMINA (Trott and Olson 2009), Dock6 (Allen et al. 2015), AutoDock4 (Morris et al. 2009) and ROSIE (Moretti et al. 2017)) were validated using high-resolution X-ray co-crystal structures of five Class A GPCRs (PDB IDs: 3EML [AdenosineA2A]; 2VT4 [β_1-Adr]; 3NY8 [β_2-Adr]; 3PBL [Dopa D3]; 3ODU [CXCR4]) to select a docking algorithm to be used. The structures of protein and ligand were prepared according to the input requirements for each docking algorithm. Then the docking algorithms were applied to reproduce the bound conformation of a ligand in crystal structure. The docking protocols were compared in reproducing the crystallographic pose of ligand inside the receptor binding pocket and the one with the lowest RMSD value, the lowest mean RMSD value and the highest number of poses with RMSD value < 2 Å was selected.

AutoDock SMINA demonstrated the best performance in comparison to the other ligand-docking algorithms considered for G-protein coupled receptors. SMINA is a fork of the AutoDock VINA software – which uses the same docking and scoring algorithms with additional flexibility to use user-defined custom scoring function, was used for molecular docking experiments in the project instead of VINA (Koes, Baumgartner and Camacho 2013). Also, SMINA overcomes the limitation of VINA to enlist only one ligand in the ligand file set for docking. SMINA allows the user to set a ligand file comprising different conformations of different molecules for the docking experiment.

Table IIIa: Evaluation of docking algorithms by RMSD (Å) between co-crystalised and the re-docked co-crystalised ligand into the GPCR orthosteric binding pocket.

GPCR X-ray Crystals	SMINA	Dock6	Autodock4	ROSIE
3EML[AdenosineA2A]	0.82	0.55	0.93	0.96
2VT4[β_1-Adr]	0.49	0.89	2.93	0.51
3NY8[β_2-Adr]	1.02	0.71	1.22	1.00
3PBL[Dopa D3]	0.99	0.41	0.48	0.98
3ODU[CXCR4]	0.79	3.98	2.63	0.79
RMSD Avg. (Å)	*0.82*	**1.31**	**1.64**	**0.85**
RMSD Std. Deviation	*0.21*	**1.51**	**1.08**	**0.21**

Appendix IV: Automated KNIME pipeline for GPCR homology modelling and molecular docking.
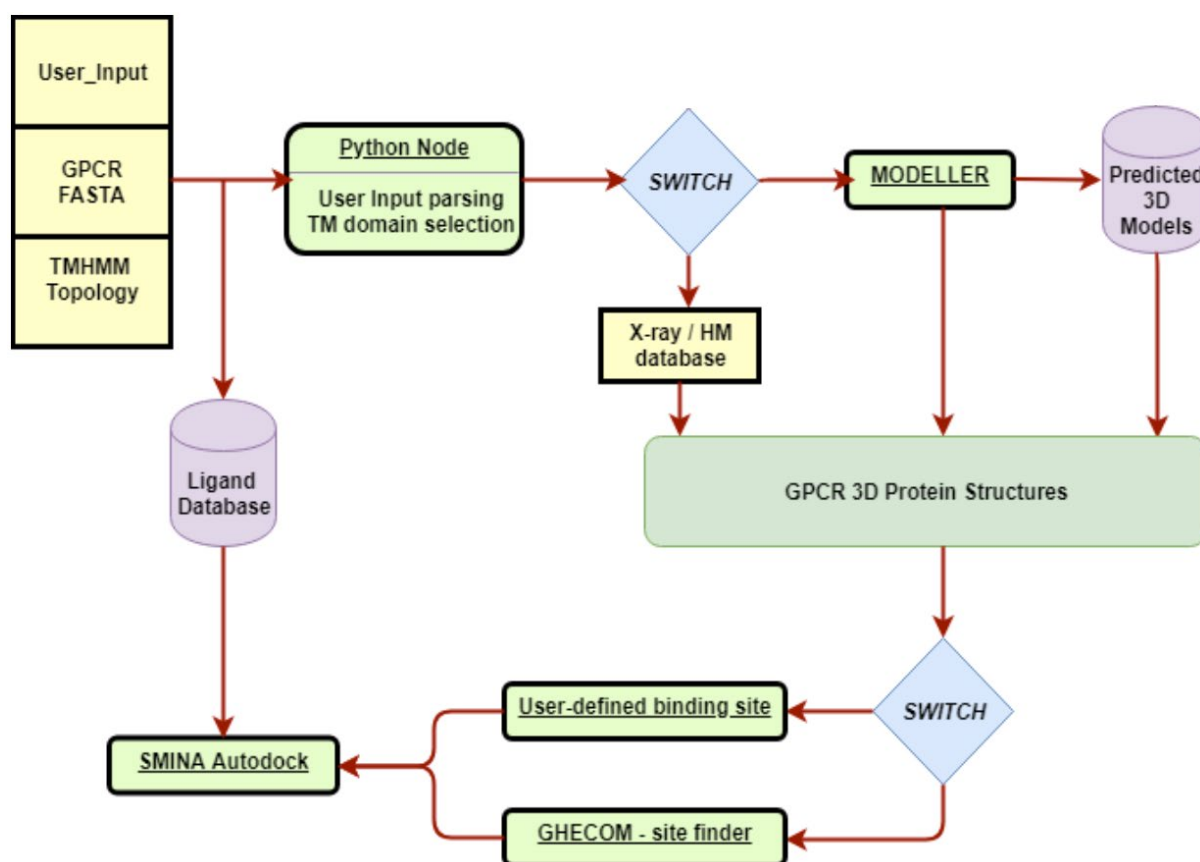
Figure IV.1: Schematic of KNIME workflow for GPCR homology modelling and molecular docking for virtual screening.

The KNIME workflow was designed to incorporate the open-source tools to serve as a backbone of the desktop CADD infrastructure targeting GPCRs on the linux platforms (specifically Ubuntu distributions). The secondary structure topology provided by the user from the TMHMM server along with GPCR protein sequence files was used to predict the 3D model of desired proteins. The user can choose between HM tool - MODELLER (open-source academic license) to predict the 3D models or define the database of GPCR 3D models. The advanced user has the option to develop a pipeline to include different HM algorithms of their choice and incorporate it into the workflow.

The predicted models were combined in a temporary virtual database which can be used in virtual screening if the user decides to proceed with workflow. The user can skip the HM pipeline if they already have a pre-processed database of 3D protein models / structures for VS through molecular docking using SMINA. The next pre-requisite for VS is the defined

binding pocket of each protein target to be used for screening in the docking experiments. At this point, the user has an option to define the configuration file with binding pockets for each protein or GHECOM site finder tool embedded in the pipeline can automatically detect the protein cavities. The python script embedded in pipeline chooses the largest pocket (from the GHECOM predictions) at the extracellular end of the GPCRs as the orthosteric binding pocket for the docking experiment.

The ligand databases to be screened against the protein structures were prepared and processed by the OpenBabel tool available as an open-source software package. The pipeline creates sub-directories for all the docking results with respect to the protein target in the working directory. The KNIME pipeline is designed to provide preconfigured starting points for the new users introduced to the field of CADD-SBDD. The advanced users can adapt the flexible workflow according to their needs using the various chemoinformatics research nodes available in the KNIME repository. The workflow developed will be available as open source via GitHub web portal (https://github.com/jay4pal/KNIME_GPCRs).

Appendix V: Topology prediction of human GPR120 and homology modelling

Appendix Va: Topology Prediction of GPR120L by TMHMM:

#TMHMM ------ WEBSEQUENCE Length: 377

WEBSEQUENCE Number of predicted TMHs: 7

WEBSEQUENCE POSSIBLE N-term signal sequence

WEBSEQUENCE	TMHMM2.0	outside	1	43
WEBSEQUENCE	TMHMM2.0	TMhelix	44	66
WEBSEQUENCE	TMHMM2.0	inside	67	74
WEBSEQUENCE	TMHMM2.0	TMhelix	75	97
WEBSEQUENCE	TMHMM2.0	outside	98	111
WEBSEQUENCE	TMHMM2.0	TMhelix	112	134
WEBSEQUENCE	TMHMM2.0	inside	135	154
WEBSEQUENCE	TMHMM2.0	TMhelix	155	177
WEBSEQUENCE	TMHMM2.0	outside	178	208
WEBSEQUENCE	TMHMM2.0	TMhelix	209	231
WEBSEQUENCE	TMHMM2.0	inside	232	282
WEBSEQUENCE	TMHMM2.0	TMhelix	283	305
WEBSEQUENCE	TMHMM2.0	outside	306	314
WEBSEQUENCE	TMHMM2.0	TMhelix	315	337
WEBSEQUENCE	TMHMM2.0	inside	338	377

Appendix Vb: GPR120 Topology Prediction of GPR120S by UCD-Porter: C- coil; H-helix (Green); E-beta sheet (Yellow).

M	S	P	E	C	A	R	A	A	G	D	A	P	L	R	S	L	E	Q	A	N	R	T	R	F	P	F	F	S	D	V	K	G	D	H	R	L	V	L	A	A	V	E	T	T	V	L	V	L	I	F	A	V	S	L	L	G	N	V	C
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	H	H	H	C	C	C	C	C	C	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H

TM1 (37-66)

A	L	V	L	V	A	R	R	R	R	R	G	A	T	A	C	L	V	N	L	F	C	A	D	L	L	F	I	S	A	I	P	L	V	L	A	V	R	W	T	E	A	W	L	L	G	P	V	A	C	H	L	L	F	Y	V	M	T	L
H	H	H	H	H	H	C	C	C	C	C	C	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C	C	H	H	H	H	H	H	H	C	C	C	C	C	C	H	H	H	H	H	H	H	H	H	H	H	H	H	H

TM1   TM2 (73-101)   TM3 (108-141)

S	G	S	V	T	I	L	T	L	A	A	V	S	L	E	R	M	V	C	I	V	H	L	Q	R	G	V	R	G	P	G	R	R	A	R	A	V	L	L	A	L	I	W	G	Y	S	A	V	A	L	P	L	C	V	F	R	V	V
H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C	C	C	C	C	C	C	C	C	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C	C	H	H	H	H	C	E	E	E	E

TM3   TM4 (152-174)

P	Q	R	L	P	G	A	D	Q	E	I	S	I	C	T	L	I	W	P	T	I	P	G	E	I	S	W	D	V	S	F	V	T	L	N	F	L	V	P	G	L	V	I	V	I	S	Y	S	K	I	L	Q	T	S	E	H	L	L	D	A
E	E	C	C	C	C	E	E	E	E	C	C	C	C	C	C	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C	C	C	C	C	C	C	C

TM5 (198-232)

R	A	V	V	T	H	S	E	I	T	K	A	S	R	K	R	L	T	V	S	L	A	Y	S	E	S	H	Q	I	R	V	S	Q	Q	D	F	R	L	F	R	T	L	F	L	L	M	V	S	F	F	I	M	W	S	P	I	I	I	T	I
C	C	C	C	C	C	C	C	C	H	H	H	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H

TM6 (253-289)

L	L	I	L	I	Q	N	F	K	Q	D	L	V	I	W	P	S	L	F	F	W	V	V	A	F	T	F	A	N	S	A	L	N	P	I	L	Y	N	M	T	L	C	R	N	E	W	K	K	I	F	C	C	F	W	F	P	E	K	G	A
H	H	H	H	C	C	C	C	C	C	C	C	C	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C	C	C	C	C	C	C	C	C	C

TM6   TM7 (298-334)

I	L	T	D	T	S	V	K	R	N	D	L	S	I	I	S	G
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C

Appendix Vc: Topology Prediction of GPR120L by JPred Secondary Structure Prediction; (Red: Helix; Green-Beta sheet; Black- coil/loop)

jp_UabS4Mz/1-377 PGEISWDVSFVTLNFLVPGLVIVISYSKILQTSEHLLDARAVVTHSEITKASRKRLTVSLAYSESHQIRVSQQDFRLFRTLFLLMVSFFIMWSPIIITILLILIQNFKQDL

Lupas_21 . . . .

Lupas_14 . . . .

Lupas_28 . . . .

jnetpred

JNETCONF

4289999998753112223554304533333331100277777777101111111017777777777612232301143322333416889999999999999999882357 8 8 8 7

JNETSOL25 BBBBBBBBBBB- BB- BBB- BBBBBBBB- B- - - B- - BBBBBBBBB- BBBBBBBBBBBBBBB- BBBBB- - B- - - B- - - - - - - - - - - - - BBB- BB- BB- BBB- BBBBBBBBBBBBB- B- - BB

JNETSOL5 - - - - - BB- BB- - - - - - - - BBBBB- - - - - - - - B- B- - BB- - BB- - B- B- - - B- - - - -

JNETSOL0 - B- - - - - - - - - -

JNETHMM

JNETPSSM

JNETJURY * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
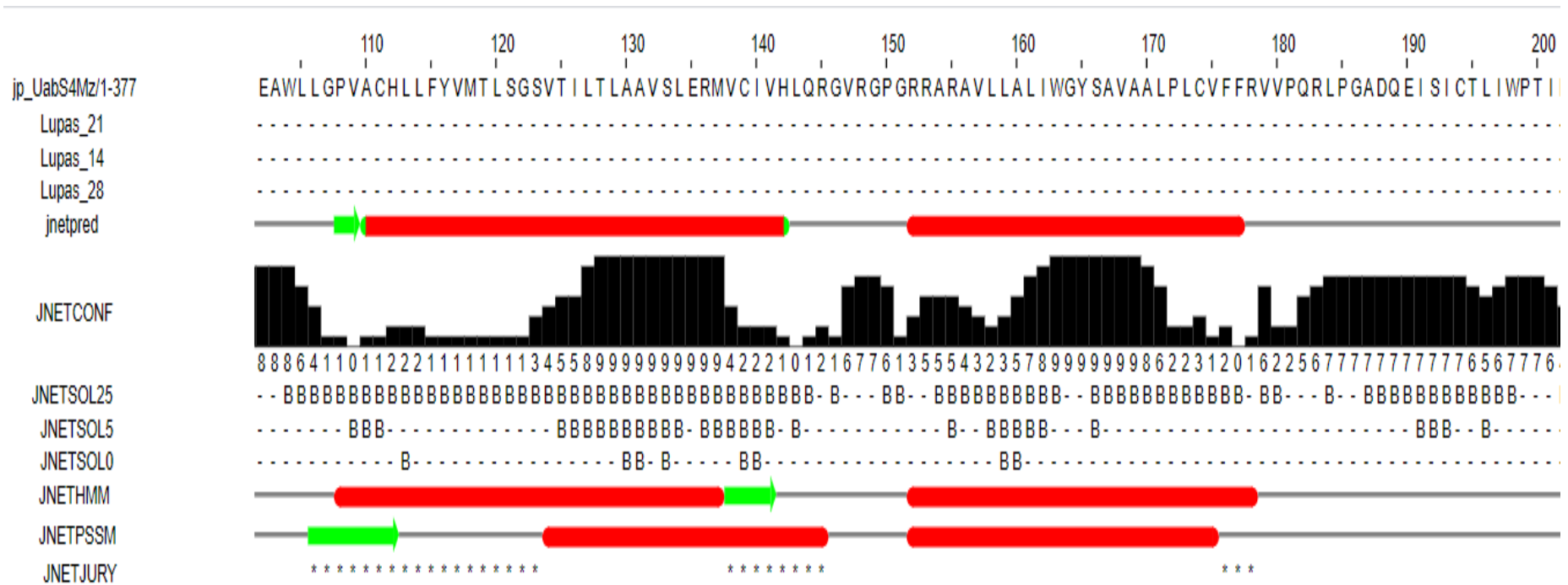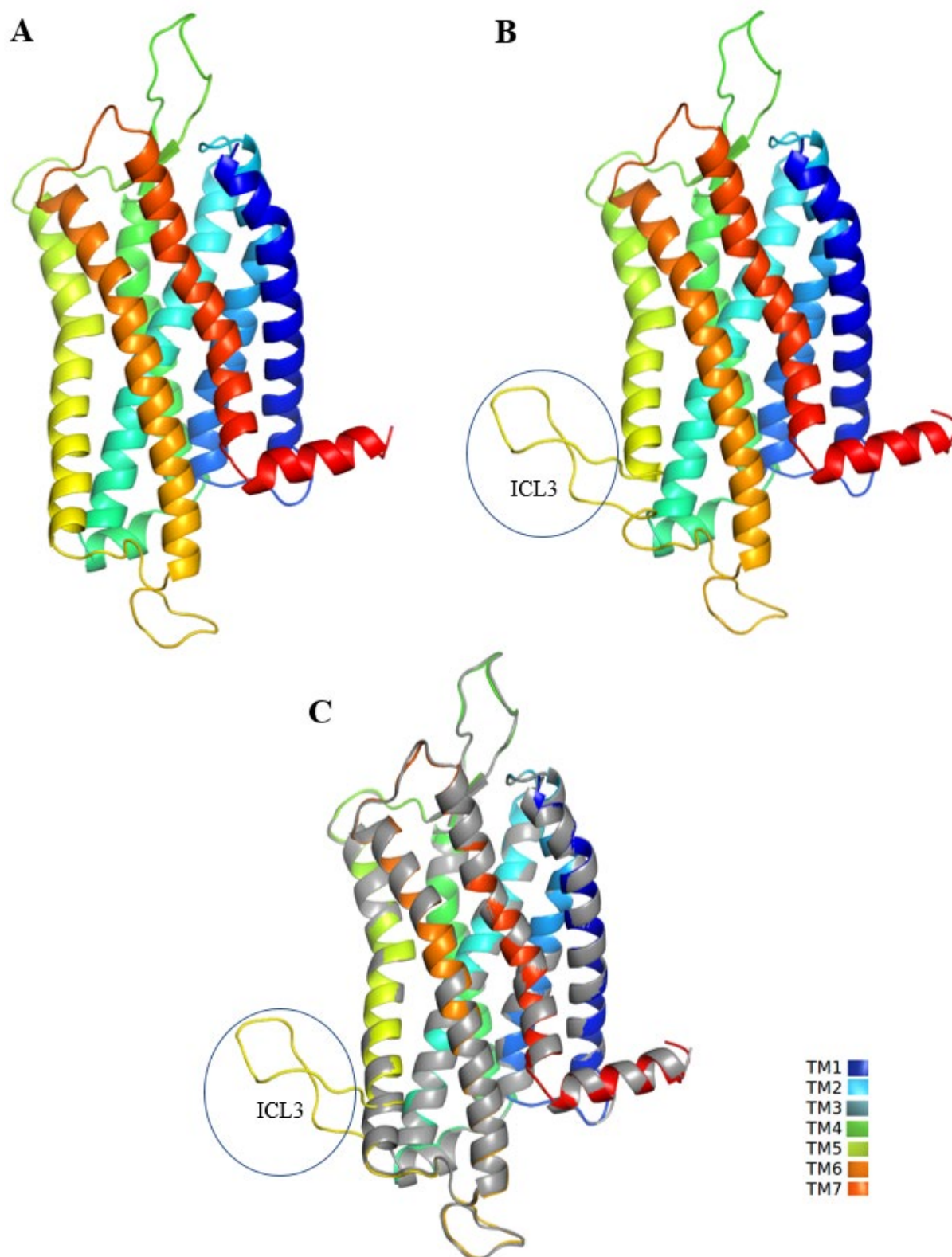
Appendix Vd: Homology models of GPR120L

Figure Vd: Validated model of GPR120S (A) used as template for predicting GPR120L (B) structure. (C) Superimposed structures of GPR120S (Grey) and GPR120L (Spectrum colour)

Appendix VI: Equilibration of GPR120S receptor

Equilibration of Apo GPR120S receptor

The 500 ps NVT equilibration run of the Apo GPR120S system showed that both the total energy and potential energy of the system plateaued after ~400 ps characterised by the decrease in total energy from -1,432,239 (2 ps) to -1,513,052 kJ/mol (372 ps) (Figure VIa). The visual inspection of the system ensemble after NVT run showed that the phospholipid bilayer leaflets started separating after 20 ps of NVT run (Figure VIb). The leaflets stayed separated throughout the 500 ps NVT run. This leaflet separation phenomenon has been observed frequently in MD with large heterogeneous systems as the POPC molecules were trying to orient themselves with respect to both water and protein.
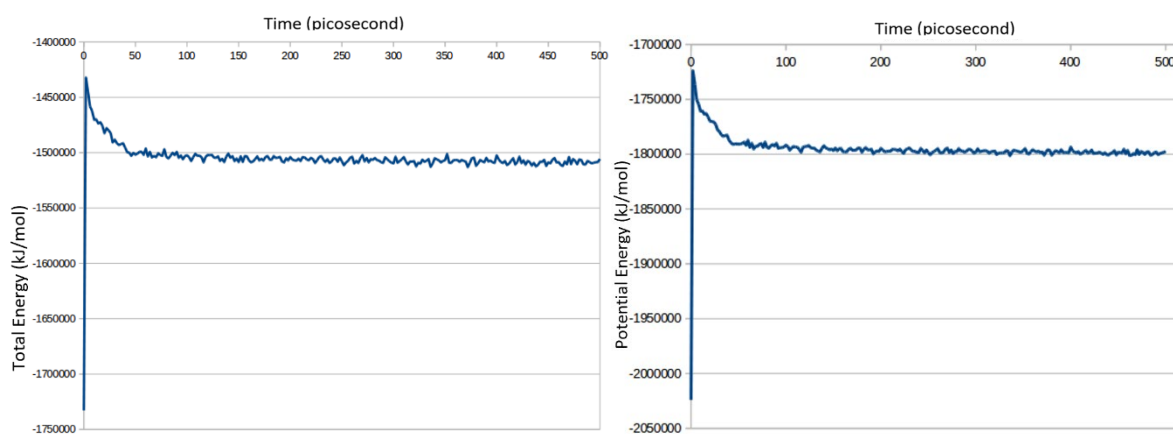

Figure VIa: Total energy (left) and Potential energy (right) of Apo GPR120S during NVT equilibration run of 500ps.
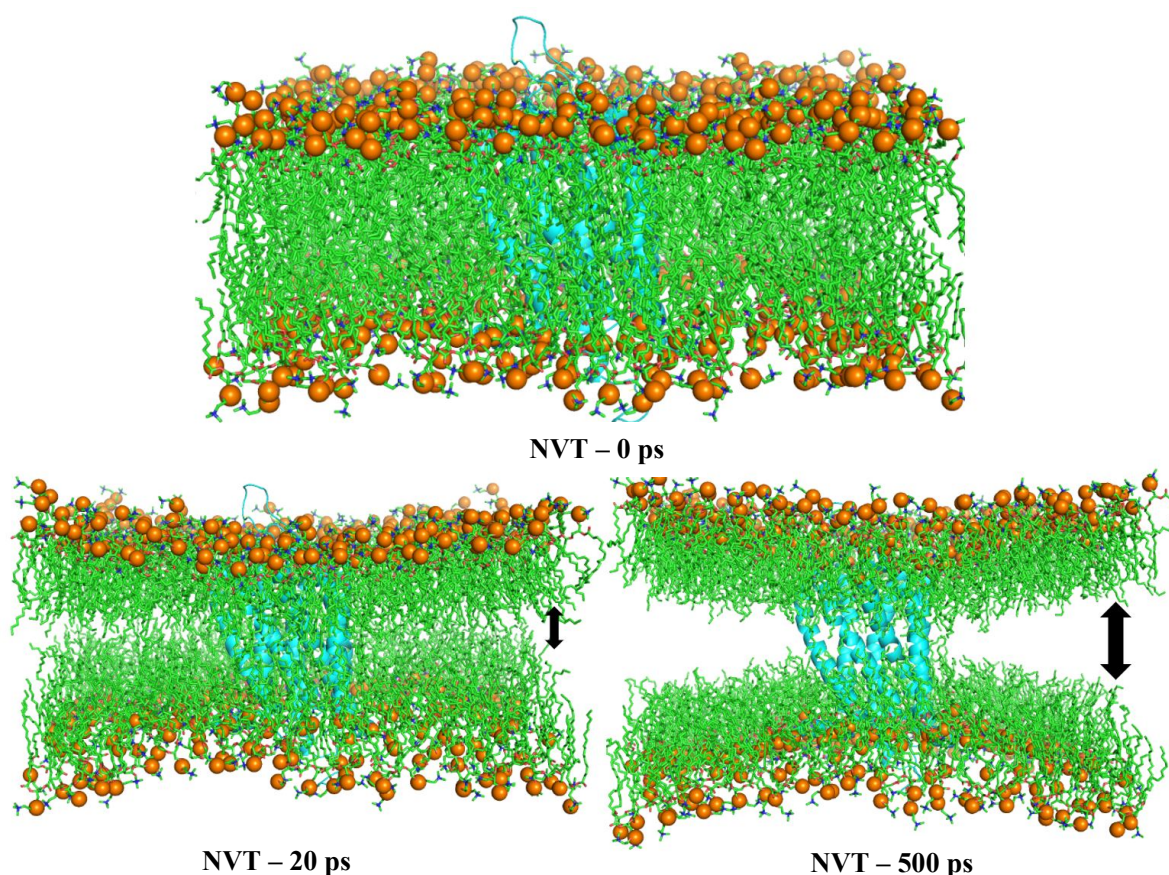
NVT – 0 ps

NVT – 20 ps

NVT – 500 ps

Figure VIb: Snapshots of the phospholipid bilayer leaflets in the Apo GPR120S system. The leaflets separated during 500 ps NVT equilibration step; Phosphate heads are shown as orange spheres and lipid tails as green sticks with the GPR120S protein (cyan) embedded. The water box is not shown in the image for clarity. The image was visualized and rendered in PyMol (DeLano 2018).

The extended NPT run with consecutively decreasing position restraints on POPC phosphate heads was simulated to bring the leaflets closer without stretching the lipid bonds. The position restraints of 1000, 500, 100, 50 and 10 kJ/mol/nm^2 were applied for five consecutive NPT runs of 1000 ps (1 ns) each to harmonically reduce stress and relax the system. The total energy as well as the potential energy of the system were observed to stabilise around 4000 ps with lowest potential energy of -1,823,318.5 kJ/mol at 4118 ps. The biggest decline in total energy as well as potential energy of the system occurred during the first 3 ns (Figure VIc) which can be attributed to the relaxation in position restraints. Afterwards, with a gradual decrease the total energy of the system plateaued during the last 500 ps of the equilibration simulations. The temperature of the system portrayed a uniform trend (300 K

+/- 2) throughout the 5 ns equilibration simulation (Figure VId-Left). As pressure is a macroscopic property, the system showed large fluctuations from one simulation step to another. But the average pressure overtime stabilized at ~1 bar after equilibration (Figure VId-Right). The visual analysis of NPT ensemble also showed that the leaflet has returned to their normal locations and relaxed with the system (Figure VIe). The system conformation at 4,118 ps (with lowest potential energy of -1,823,318.5 kJ/mol) was used for running MD production runs.
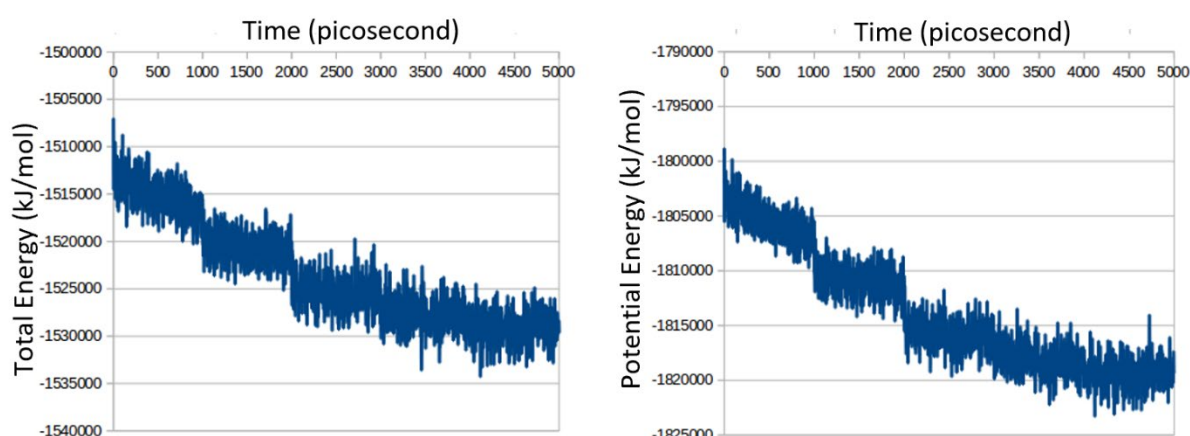

Figure VIc: Total energy (left) and Potential energy (right) of Apo GPR120S during NPT equilibration run of 5 ns with position constraint on phosphate heads of POPC.
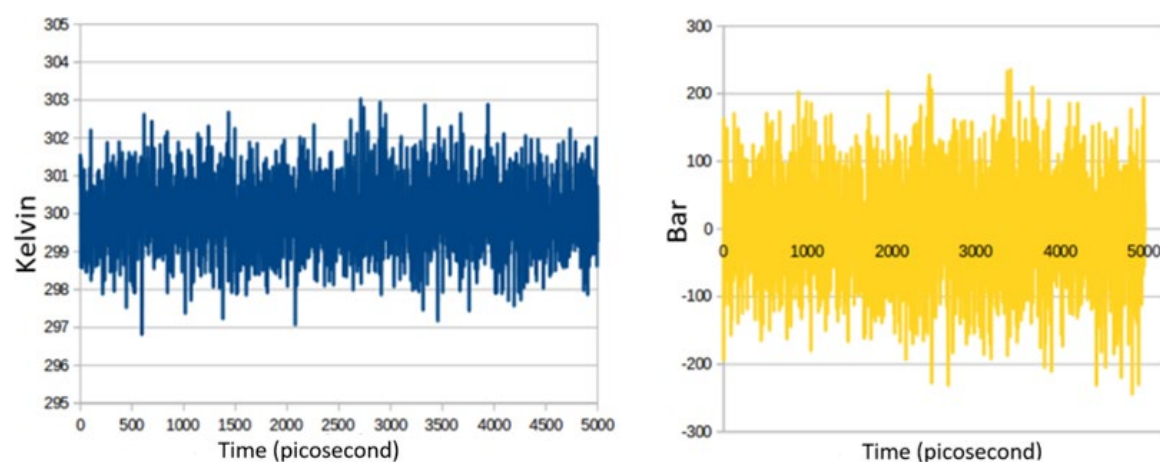

Figure VId: Temperature (left) and pressure (right) of Apo GPR120S during NPT equilibration run of 5 ns with position constraint on phosphate heads of POPC.
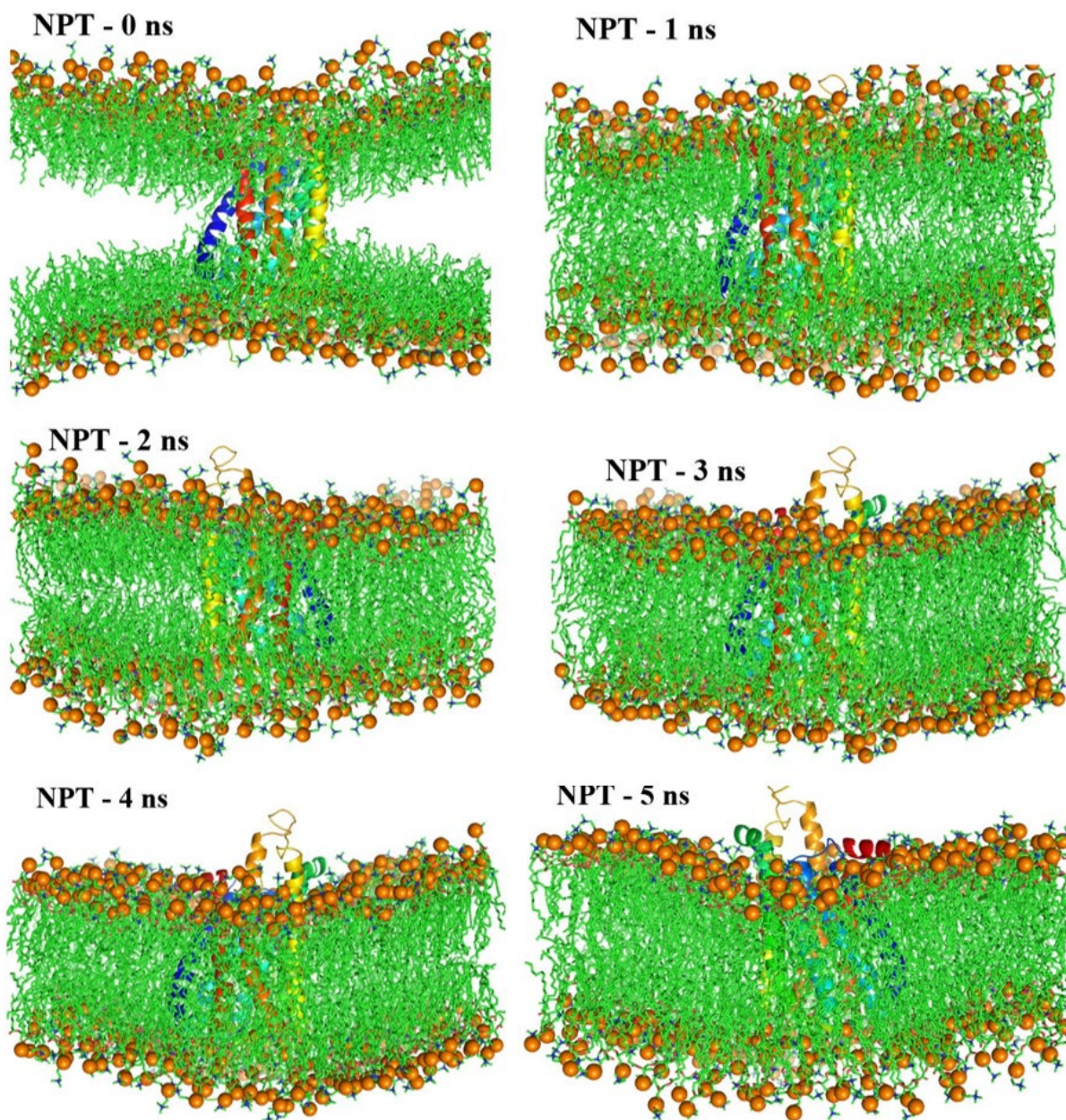
Figure VIe: Snapshots of the phospholipid bilayer leaflets in the Apo GPR120S system. The leaflets retained "sandwich" conformation after 5 ns NPT equilibration step; Phosphate heads are shown as orange spheres and lipid tails as green sticks with the GPR120S protein (cartoon) embedded. The water box is not shown in the image. The image was visualized and rendered in PyMol (DeLano 2018).

EQUILIBRATION OF THE ANTAGONIST AND AGONIST-BOUND SYSTEM

The equilibration of the prepared systems with position constraints on ligands (TUG891 and Compound39) successfully kept the ligands inside the binding pocket throughout the simulations. But a similar problem of leaflet separation was observed in agonist-bound systems during NVT equilibration. As the project was aimed to look for differences in the conformational behaviour of the GPR120S model in the presence of an agonist from the apo system, first equilibration of the receptor model in the bilayer had to be ensured. The same protocol of Apo system - NVT (500 ps) and five consecutive NPT (1 ns each) with decreasing position constraints for phosphate heads of phospholipid bilayer to prevent leaflet separation was followed. The total energies of both the agonist-bound systems were stable by the end of the equilibration process (Figure VIf).

At this point, the AH7614-bound system was dropped from further MD simulation studies as it was declared NAMD of GPR120 receptor isoforms (Watterson et al. 2017).
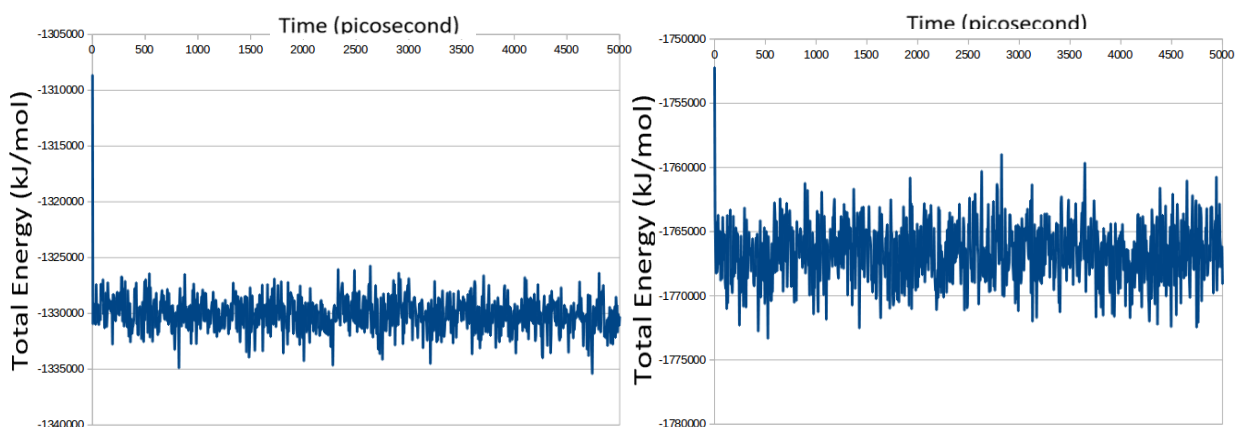

Figure VIf: Total energy of TUG891-bound GPR120S model (Left) and Compound39-bound GPR120S model (right) during NPT equilibration run of 5 ns with position constraint on phosphate heads of POPC

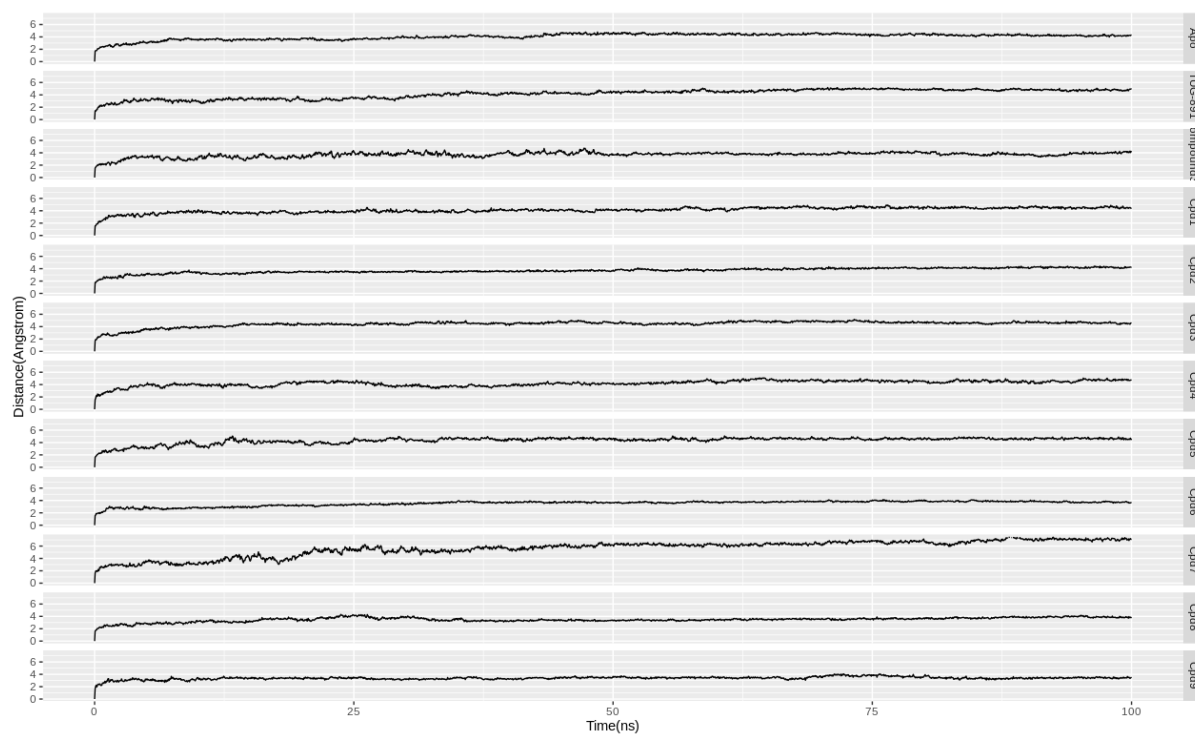# Appendix VII: Analysis of 100 ns MD simulations of Compounds Cpd1-9

Figure VIIa: Root Mean Square Deviation (RMSD) plot of Cα of the protein backbone (Å) recorded during 100 ns MD simulation run of GPR120S in unbound (Apo) and bound form with TUG891, Compound39 and compounds 1-9.
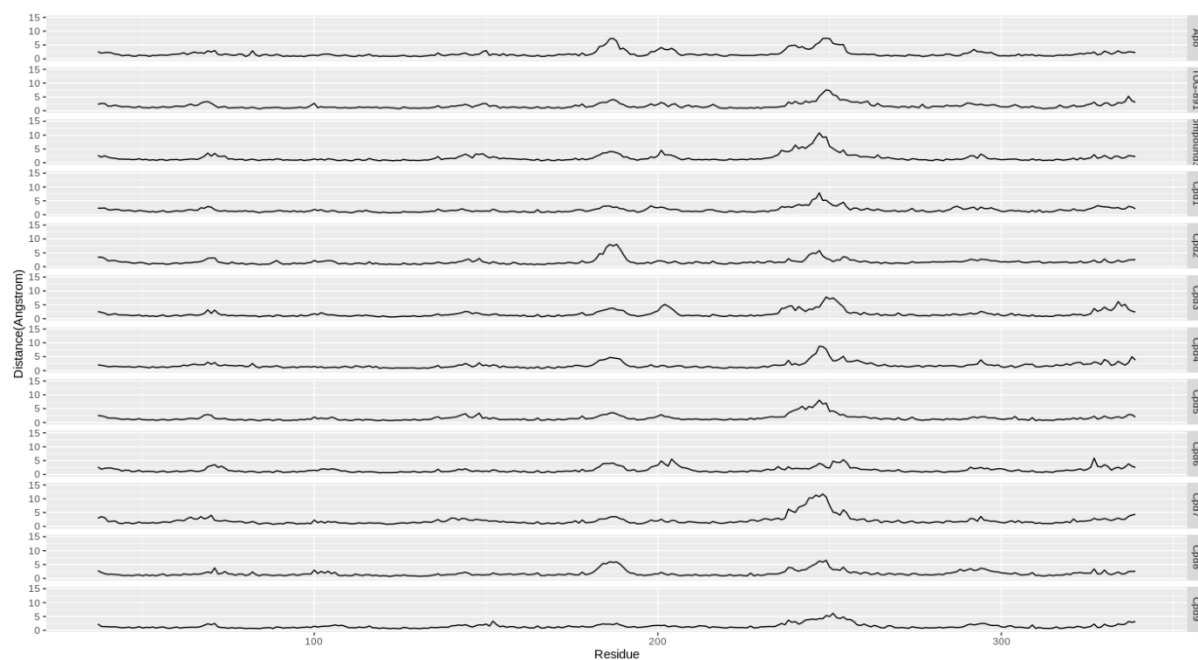
Figure VIIb: Root Mean Square Fluctuation (RMSF) plot of the protein backbone (Å) recorded during 100 ns MD simulation run of GPR120S in unbound (Apo) and bound form with TUG891, Compound39 and compounds 1-9.
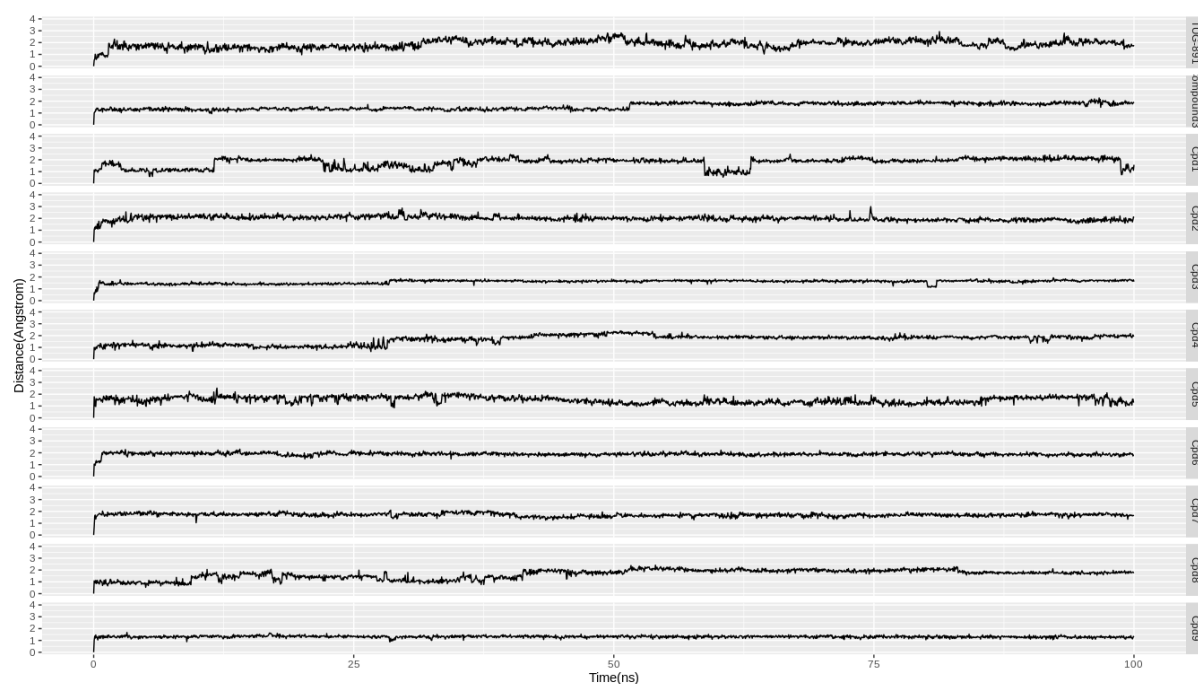

Figure VIIc: Root Mean Square Deviation (RMSD) plot of ligands bound to protein (Å) recorded during 100 ns MD simulation run of GPR120S in bound forms with TUG891, Compound39 and compounds 1-9.
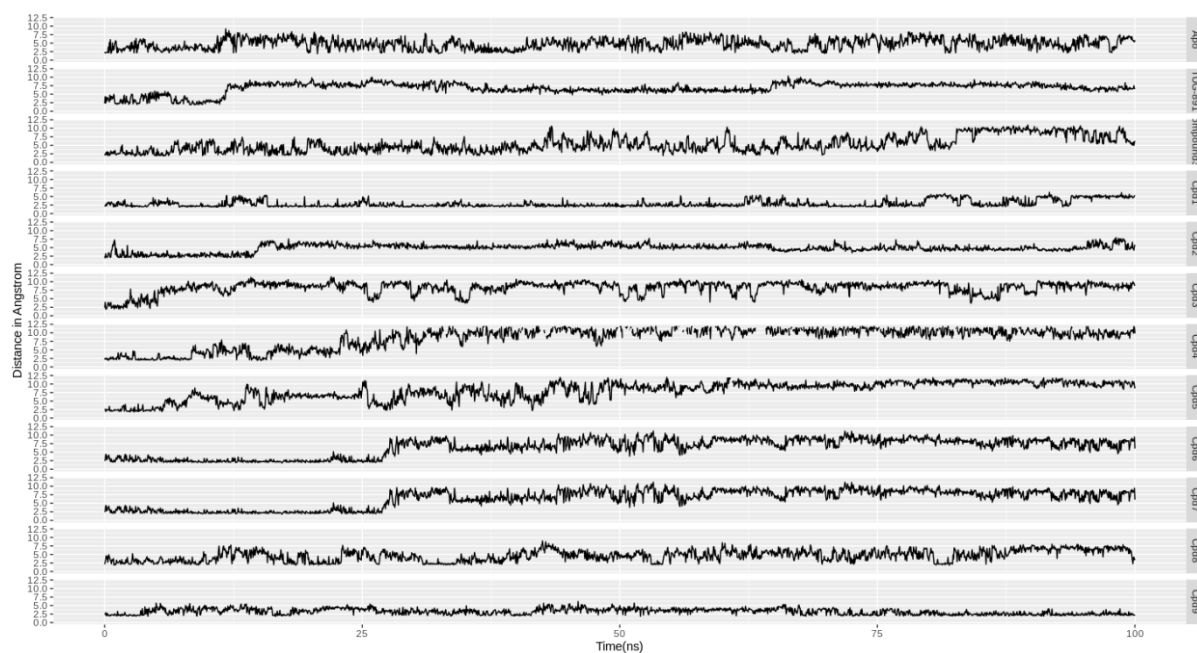
Figure VIId: Distance (Å) plot between the center of mass of residues Arg136(TM3) and Asp259(TM6) involved in "ionic-lock" conformation recorded during 100 ns MD simulation run of GPR120S in bound forms with TUG891, Compound39 and compounds 1-9
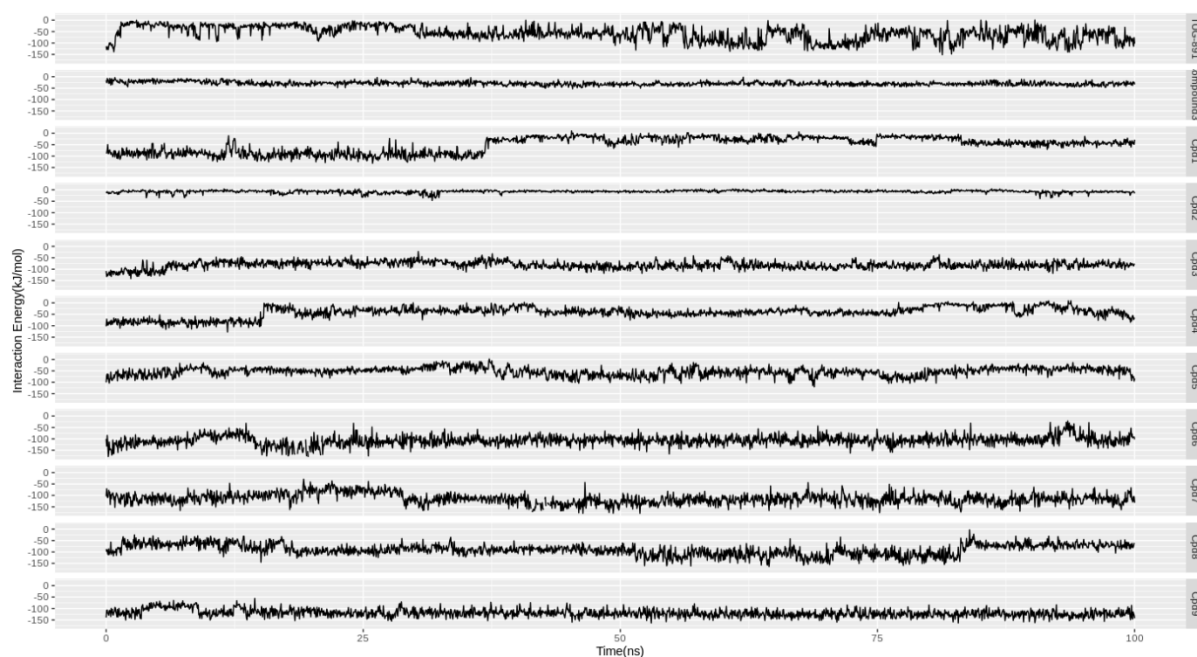

Figure VIIe: Interaction energy of (Coulombic interactions) of ligand bound protein systems recorded during 100 ns MD simulation run of GPR120S in bound forms with TUG891, Compound39 and compounds 1-9

Appendix VIII: RT-qPCR readings of GPR120 siRNA expression

	Housekeeping GAPDH						
Sample	C_T^1	C_T^2	C_T^3	C_T^4	C_T^5	C_T^6	C_T^{mean}
siRNA-1	8.192475	8.708578	9.434944	8.442725	8.778606	8.722689	8.713336
siRNA-2	7.732936	7.338948	7.806002	7.550343	8.087684	7.664701	7.696769
Control	8.513725	8.449002	9.202267	8.873232	9.143499	9.024377	8.867684

	Target gene GPR120										
Sample	C_T^1	C_T^2	C_T^3	C_T^4	C_T^5	C_T^6	C_T^{mean}	ΔC_T	$\Delta\Delta C_T$	2^-ΔΔCt	%Relative Fold
siRNA-1	30.20934	31.91719	32.2821	31.89163	32.14431	31.98971	31.73905	23.02571	0.748621	0.595172	59.517
siRNA-2	31.09578	31.43562	31.66361	31.85941	31.09859	31.53338	31.44773	23.75096	1.473875	0.360014	36.001
Control	30.42481	31.7924	30.86534	31.71351	30.54261	31.52996	31.14477	22.27709	0	1	100

Where:

C_T^{mean} = Average of C_T values

$\Delta C_T = C_T^{mean}$ (Target gene GPR120) - C_T^{mean} (Housekeeping gene GAPDH)

Calibrator / reference sample $\Delta C_T = \Delta C_T$ (Control of target gene) => 22.27709

$\Delta\Delta C_T = \Delta C_T$ (treated sample) - ΔC_T (Calibrator)

Relative fold gene expression = $2^{-\Delta\Delta C_T}$

Appendix IX: List of Employability Skill and Discipline Specific Skills Training

Module Title	Institution	ECTs awarded	Date
Discipline Specific Skills			
BIOL 9222 Biological Basis of Disease	TU Dublin	10	January 2017
GRSO 1001 Research Methods	TU Dublin	5	October 2017
GRSO 1005 Introduction to Statistics	TU Dublin	5	March 2018
Employability Skills			
GRSO 1010 Introduction to Pedagogy	TU Dublin	5	January 2017
Scientific Programming Concepts	ICHEC	5	February 2018
RESM 1953 Research Integrity	TU Dublin	5	November 2018
Techniques and Strategies in Molecular Medicine	CRDI - TCD	5	December 2018