Development of modulators targeting G protein-coupled receptors using a combination of conventional drug discovery methodologies and artificial intelligence techniques

A Thesis Presented to the Faculty of Graduate Studies of Lakehead University by

Wenjie Liu

In partial fulfilment of the requirements for the degree of Doctor of Philosophy in Chemistry and

Material Science

Supervised by

Dr. Jinqiang Hou

Lakehead University, Thunder Bay, Ontario, Canada, 2023

Abstract

Lysophosphatidic acid receptor 1 (LPA_1) is one of six G protein-coupled receptors (GPCRs) in the LPA receptor family (LPA_{1-6}) and is activated by lysophosphatidic acid (LPA). LPA_1 is involved in a wide range of cellular activities, including the regulation of cell proliferation, migration, invasion, and apoptosis. Numerous studies have reported its implication in the initiation, progression, and metastasis of breast cancer. Bioinformatic analysis has also revealed the role of LPA_1 in breast cancer. However, the current repertoire of LPA_1 ligands targeting cancer remains highly limited, and only one LPA_1 radioligand for positron emission tomography (PET) imaging of pulmonary fibrosis has been reported so far. Therefore, the objective of this study is to develop novel LPA_1 antagonists and LPA_1 radioligands for breast cancer therapeutics and diagnosis.

Chapter 1 provides an overview of the current development status of ligands targeting LPA_{1-6}. Chapter 2 presents the detailed synthesis and biological evaluation of carbamate-derived and urea-derived ligands in three parts. In part 1, the first series of carbamate-derived LPA_1 antagonists were synthesized and assessed. Among these compounds, 1-(4'-(5-(((1-(3-fluorophenyl)ethoxy)carbonyl)amino)-4-methyl-1H-1,2,3-triazol-1-yl)-[1,1'-biphenyl]-4-yl)cyclopropane-1-carboxylic acid exhibited the highest potency and selectivity, effectively inhibiting cell survival, migration, and invasion induced by LPA in the breast cancer cell line MDA-MB-231. In part 2, the second series of LPA_1 antagonists, derived from urea, were synthesized and evaluated. However, this series demonstrated overall lower activity compared to the first series of compounds. Lastly, in part 3 of this chapter, one of the most potent and selective LPA_1 antagonists, 1-(4'-(5-((((3-ethylbenzyl)oxy)carbonyl)amino)-4-methyl-1H-1,2,3-triazol-1-yl)-[1,1'-biphenyl]-4-yl)cyclopropane-1-carboxylic acid, was chosen for radiolabeling in order to be used as a radiotracer for PET imaging, facilitating the diagnosis of breast cancer. The radiosynthesis and evaluation of this compound are discussed in detail in this section.

The field of medicinal chemistry has entered into the 'big data era' due to the emergence of large-scale and high-dimensional data. Traditional machine learning and deep learning, as subsets of artificial intelligence (AI), provide a versatile framework that enables the identification of intricate patterns from complex datasets, surpassing the limitations of traditional methods in drug discovery. Motivated by the rapid emergence of AI in medicinal chemistry, the objective was to

apply these cutting-edge techniques to drug discovery, as discussed in chapter 3. In part 1, several machine learning models were constructed and trained using a diverse dataset consisting of peptides and small molecules, with the goal of predicting small molecule binding. The models developed in the study exhibited high accuracy and precision in their predictions, making them valuable for guiding drug design. In part 2, the exploration of AI was expanded to the field of deep learning. A local augmentation mechanism was incorporated into graph neural networks, resulting in the construction of local augmented graph convolution networks and local augmented graph attention networks. These neural network models were employed to predict the ligand type of GPCRs. The models achieved state-of-the-art performance and demonstrated accurate prediction capabilities. All the models developed in this chapter hold potential for guiding drug design in their respective fields.

Acknowledgements

I did not realize how quickly the four years of my PhD study have passed until now. Although these four years of my life were filled with copious amounts of hard work, sleepless nights, anxieties, and occasional all-nighters, they have also been incredibly rewarding and enjoyable experiences. The lessons I've learned, the friendships I've made, and the personal growth I've experienced will undoubtedly influence my future endeavors. I would like to express my deepest gratitude to my supervisor, Dr. Hou, whose unwavering support, invaluable guidance, and constant encouragement have played a pivotal role in shaping my research journey. I am grateful to him for providing me with this excellent platform and funding to perform my research work, through which I developed a true passion for the world of science. His expertise and mentorship have provided me with the necessary foundation to excel in my work. I would also like to express my immense gratitude to my co-supervisor, Dr. Jiang, and my committee members, Dr. Campbell and Dr. Gottardo, for their input, evaluation, and suggestions. Their expertise and insights have greatly contributed to the success of my research endeavors.

I am grateful to my co-supervisor Dr. Lees for generously providing me the opportunity to conduct a crucial part of my project in his lab at NOSM University. I would also like to express my appreciation to Dr. Qin for his support in granting me access to his lab, where I conducted plasmid-related experiments essential to my project's success. Additionally, I extend my heartfelt thanks to our collaborators, Dr. Thu and Dr. Ramachandran, for their contributions to the biological studies, as well as to Dr. Du for their assistance in resolving coding-related challenges. These collaborations have significantly broadened the scope and impact of my research.

I would like to express my gratitude to Dr. Rotstein from the University of Ottawa for graciously accepting the role of the external examiner. I am sincerely thankful for his valuable insights and comments on my thesis.

I would like to extend my gratitude to our program facilitator Brenda, TBRHRI staff members Tanya, Shannon, and Carmen, and biosafety officer Dr. Dayer for their assistance, responses, and coordination. Their support has facilitated the smooth progress of my research.

I express my heartfelt gratitude to my co-worker for their camaraderie, support, and discussions that have enriched my research experience.

Lastly, I want to offer my special thanks to my parents for their unwavering love, encouragement, and belief in my abilities throughout this journey. Their support has been the pillar of my strength and motivation.

I am truly grateful to each and every individual mentioned above, as well as those who have supported me in other ways, for their invaluable contributions to my academic and personal growth.

Thank you all!

Table of contents

List of Abbreviations

ACC	Accuracy
AAE	Adversarial autoencoder
ANOVA	Analysis of Variance
AUC	Area under the receiver operating characteristic curve
Arg	Arginine
AI	Artificial intelligence
Asp	Aspartic acid
AE	Autoencoder
ATX	Autotaxin
BSA	Bovine serum albumin
CB_1	Cannabinoid 1 receptor
CHO	Chinese hamster ovary
CVAE	Conditional variational autoencoder
CXCR4	C-X-C chemokine receptor type 4
DGL	Deep Graph Library
DNN	Deep neural network
DCM	Dichloromethane
DMSO	Dimethyl sulfoxide
DMEM	Dulbecco's Modified Eagle Medium
EDG	Endothelial differentiation gene
EGFR	Epidermal growth factor receptor
ER	Estrogen receptor
EtOH	Ethanol
EtOAc	Ethyl acetate
ECFP4	Extended-connectivity fingerprint
XGBoost	eXtrem Gradient Boosting
FDA	Food and Drug Administration
FN	False negatives
FPR	False positive rate

FP	False positives
FBS	Fetal bovine serum
GPCR	G protein-coupled receptors
GRU	Gated recurrent unit
GAN	Generative adversarial network
GR	Ghrelin receptor
Gln	Glutamine
Gly	Glycine
GAT	Graph Attention Network
GCN	Graph Convolutional Network
GNN	Graph neural network
GFP	Green fluorescent protein
HR-MS	High resolution mass spectrometry
HPLC	High-performance liquid chromatography
H_3R	Histamine H_3 receptor
HER2	Human epidermal growth factor receptor 2
IBMX	Isobutyl-1-methylxanthine
F-12K	Kaighn's Modification of Ham's F-12 Medium
LC–MS	Liquid chromatography–mass spectrometry
LAGAT	Local Augmented Graph Attention Network
LAGCN	Local Augmented Graph Convolutional Network
LSTM	Long-short-term memory
LPA_1	Lysophosphatidic acid receptor 1
ML	Machine learning
MCC	Matthews correlation coefficient
MeOH	Methanol
MACCS	Molecular ACCess System keys
MW	Molecular weight
NMR	Nuclear Magnetic Resonance
NAR	Number of aromatic rings
NHA	Number of hydrogen bond acceptors

NHD	Number of hydrogen bond donors
NRB	Number of rotatable bonds
Phe	Phenylalanine
PBS	Phosphate buffer saline
PET	Positron emission tomography
Prec	Precision
PR	Progesterone receptor
RF	Random forest
Rec	Recall
ROC	Receiver operating characteristic
RNN	Recurrent neural network
SHAP	SHapley Additive exPlanations
SMILES	Simplified molecular-input line-entry system
SEM	Standard error of mean
SAR	Structure–activity relationship
SVM	Support vector machine
t-SNE	t-Distributed stochastic neighbor embedding
TB	Terrific broth
TLC	Thin layer chromatography
Thr	Threonine
TPSA	Topological polar surface area
TN	True negatives
TPR	True positive rate
TP	True positives
VAE	Variational autoencoder

Chapter 1

The development of modulators for lysophosphatidic acid receptors: a comprehensive review

The development of modulators for lysophosphatidic acid receptors: a comprehensive review

Abstract:

Lysophosphatidic acids (LPAs) are bioactive phospholipids implicated in a wide range of cellular activities that regulate a diverse array of biological functions. They recognize two types of G protein-coupled receptors: LPA_{1-3} receptors and LPA_{4-6} receptors that belong to the endothelial differentiation gene (EDG) family and non-EDG family, respectively. In recent years, the LPA signaling pathway has captured an increasing amount of attention because of its involvement in various diseases, such as idiopathic pulmonary fibrosis, cancers, cardiovascular diseases and neuropathic pain, making it a promising target for drug development. While no drugs targeting LPA receptors have been approved by the FDA thus far, at least three antagonists have entered phase II clinical trials for idiopathic pulmonary fibrosis (**BMS-986020** and **BMS-986278**) and systemic sclerosis (**SAR100842**), and one radioligand **(BMT-136088/^{18}F-BMS-986327**) has entered phase I clinical trials for positron emission tomography (PET) imaging of idiopathic pulmonary fibrosis. This article provides an extensive review on the current status of ligand development targeting LPA receptors to modulate LPA signaling and their therapeutic potential in various diseases.

Graphical abstract:
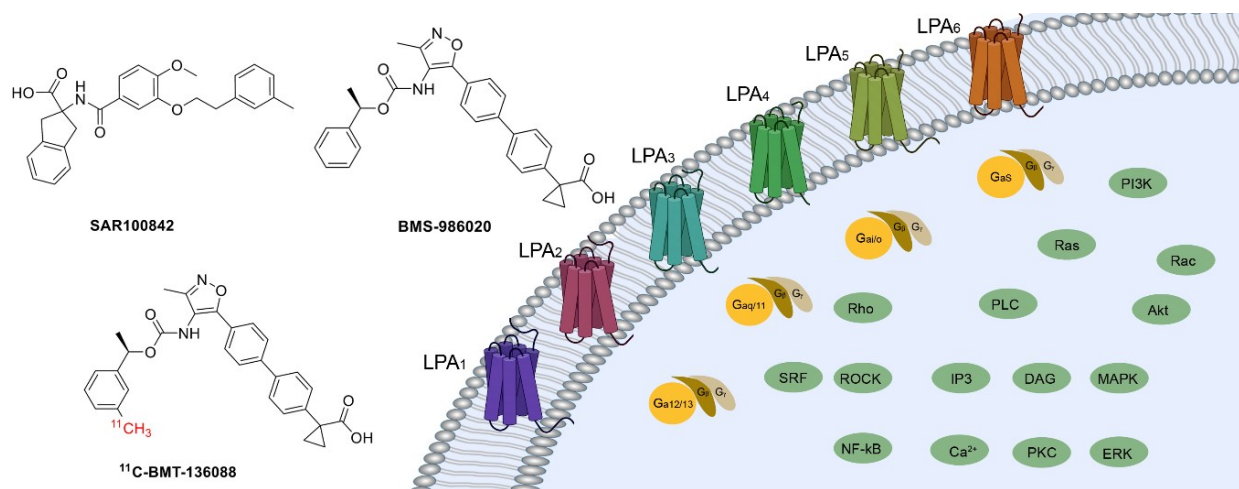
SAR100842

BMS-986020

^{11}C-BMT-136088

LPA$_1$ LPA$_2$ LPA$_3$ LPA$_4$ LPA$_5$ LPA$_6$

GaS Gi Gγ

Gai/o Gβ Gγ

Gaq/11 Gβ Gγ

Ga12/13 Gβ Gγ

PI3K

Ras

Rac

Akt

Rho PLC

SRF ROCK IP3 DAG MAPK

NF-kB Ca^{2+} PKC ERK

1. Introduction

Lysophosphatidic acid (LPA) is a bioactive phospholipid, consisting of a phosphate headgroup and a glycerol backbone with a single saturated or unsaturated fatty acid chain.[1] In serum and plasma, LPA was detected in various forms, such as 1-acyl- and 2-acyl-LPA, but when discussing LPA in the context of a signaling molecule, the term LPA usually refers to 1-acyl-2-hydroxy-*sn*-glycero-3-phosphate.[2-3] The hydrocarbon chain length and degree of saturation vary greatly and are dependent on the precursor phospholipid. The most abundant forms of LPA in human plasma are as follows: **16:0 LPA** (1-palmitoyl-2-hydroxy-*sn*-glycero-3-phosphate) with 16 carbon atoms and 0 double bonds, **18:2 LPA** (1-linoleoyl-2-hydroxy-*sn*-glycero-3-phosphate) with 18 carbon atoms and 2 double bonds, and **18:1 LPA** (1-oleoyl-2-hydroxy-*sn*-glycero-3-phosphate) with 18 carbon atoms and 1 double bond.[4] LPA is produced under physiological and pathophysiological conditions in both cells and extracellular fluids. It mediates multiple cell responses and activities, including cell proliferation, migration, invasion,[5-7] cytokine production,[8-9] reactive oxygen species (ROS) generation,[10] and macrophage formation.[11] Therefore, dysregulation of LPA is linked to certain kinds of diseases, including cancer,[12] atherosclerosis,[9, 13] fibrosis,[14] and immune system disfunction.[15-16] So far, LPA has been detected in many human fluids, including plasma,[17] serum, cerebrospinal fluid,[18] saliva, gingival crevicular fluid,[19] and pleural effusions.[20]

LPA can be produced from membrane phospholipids, including phosphatidylcholine (PC), phosphatidylethanolamine (PE), and phosphatidylserine (PS) through at least two enzymatic pathways.[2, 21] In plasma and serum, LPA is mainly generated by a lysophospholipase D (lysoPLD) named autotaxin (ATX). Plasma levels of LPA in ATX heterozygous mice have been shown to decrease by 50% compared with wild type controls.[22-23] In this pathway (Fig 1.1A), phospholipids are firstly hydrolyzed into lysophospholipids, including lysophosphatidylcholine (LPC), lysophosphatidylethanolamine (LPE), and lysophosphatidylserine (LPS), by phospholipase A enzymes (PLA$_1$ and PLA$_2$), followed by the cleavage of these lysophospholipids by ATX to produce LPA. Aside from ATX, LPA in serum and plasma is also produced from lecithin: cholesterol acyltransferase (LCAT). After binding to a lipoprotein, LCAT cleaves the *sn*-2 fatty acid chain of PC to produce 1-acyl-LPC, and then 1-acyl-LPC is converted into 1-acyl-LPA by ATX.[24-25] In platelets and some malignant cells, LPA is mainly

produced from phosphatidic acid (PA). In this pathway (Fig 1.1B), phospholipids are converted into PA by phospholipase D (PLD), and PA is subsequently deacylated by PLA_1 and PLA_2 to produce LPA.

In circulation, LPA has a half-life of around 0.5-3 minutes, depending on the LPA species.[26-27] The turnover of LPA is regulated by the enzymes that are involved in LPA production and degradation. Three pathways mediate the degradation of LPA.[28] In the first pathway, LPA is dephosphorylated to monoacylglycerol (MAG) through lipid phosphate phosphatases (LPPs). In LPP null mice, the plasma level of LPA is significantly increased and the clearance rate of intravenously injected LPA is four times slower compared with wild type.[29] The second pathway involves LPA acyltransferase, which catalyzes the transfer of the acyl group to LPA to produce PA. In the third pathway, LPA is converted into glycerol-3-phosphate (G3P) by phospholipases.
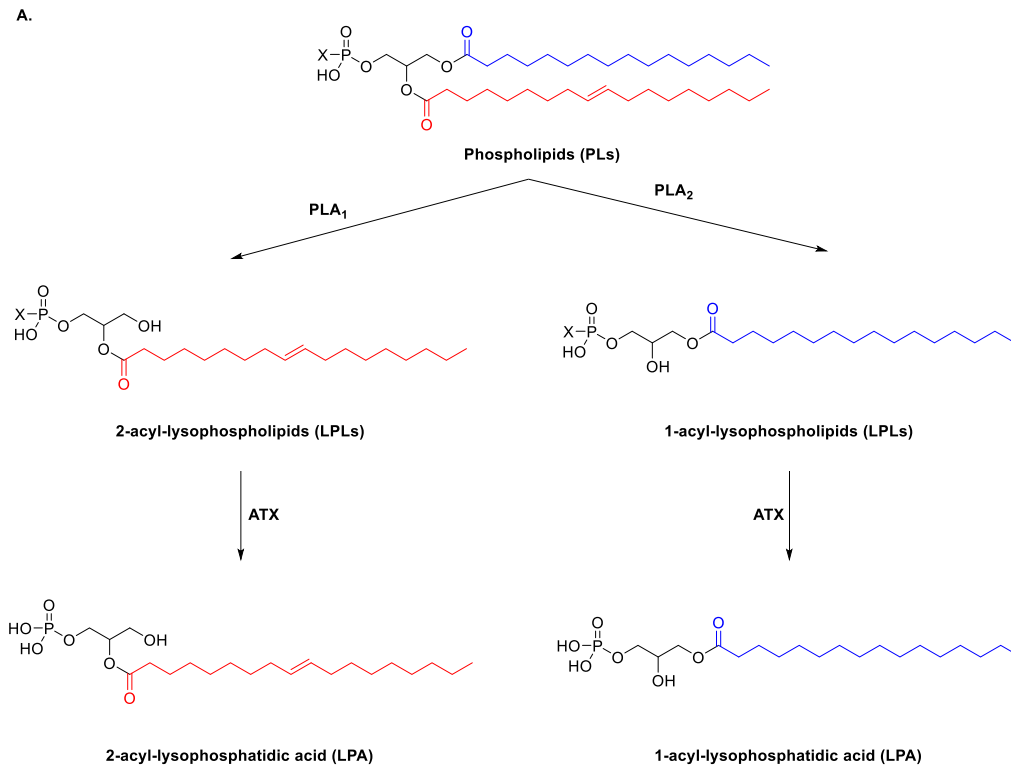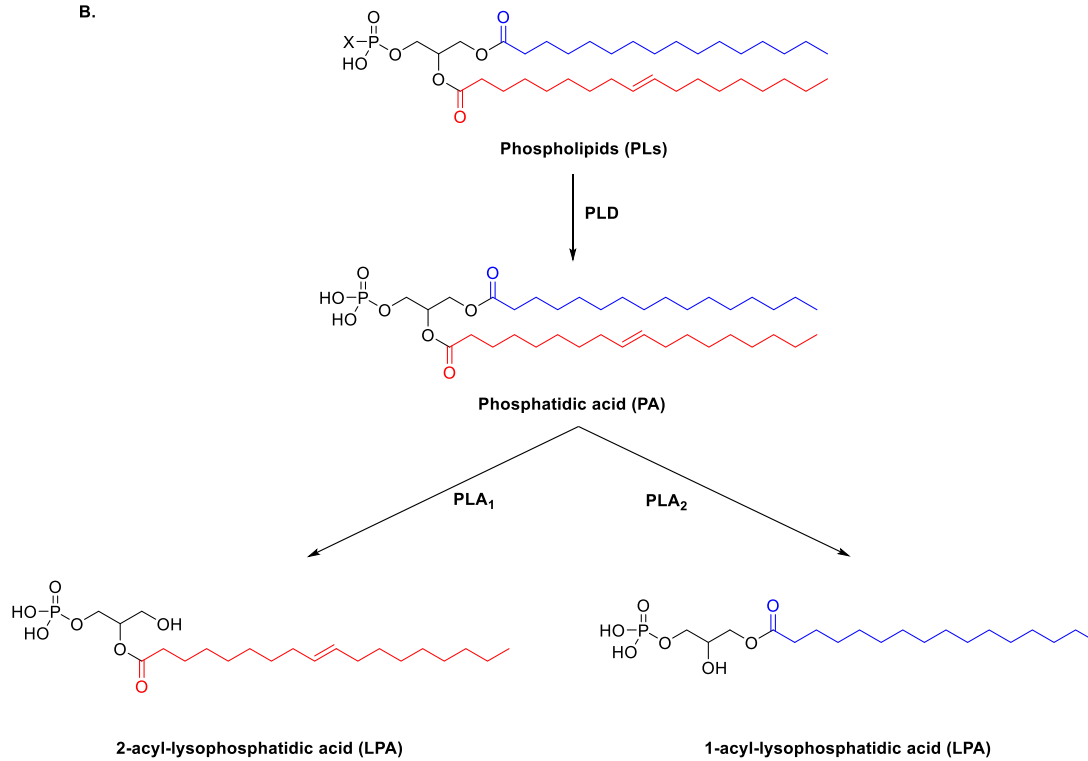
Figure 1.1. Major pathways of LPA production. X: hydrophilic head group.

LPA functions mainly by acting on a family of G protein-coupled receptors (GPCRs) known as the LPA receptors.[12] These LPA receptors then activate Rho (via $G_{12/13}$ protein), PLC (via G_q protein), RAS-MAPK or PI3-Akt (via $G_{i/o}$ protein), and AC (via G_s protein)-mediated pathways through several distinct class of heterotrimeric G proteins,[4, 12, 30-32] leading to a diverse array of downstream signaling and cell activities (Fig 1.2).

In mammals, six LPA receptors (LPA_{1-6}) have been discovered to date (Fig 1.2), and they are divided into two LPA receptor families according to phylogeny: the endothelial differentiation gene (EDG) family and the non-EDG family. All of these receptors are widely distributed in diverse human tissues. The EDG family, the LPA_{1-3} receptors, share 45-56% similarity in amino acid sequence with each other, while the LPA_{4-6} receptors form the non-EDG family and share 35-55% similarity in amino acid sequence.[33]
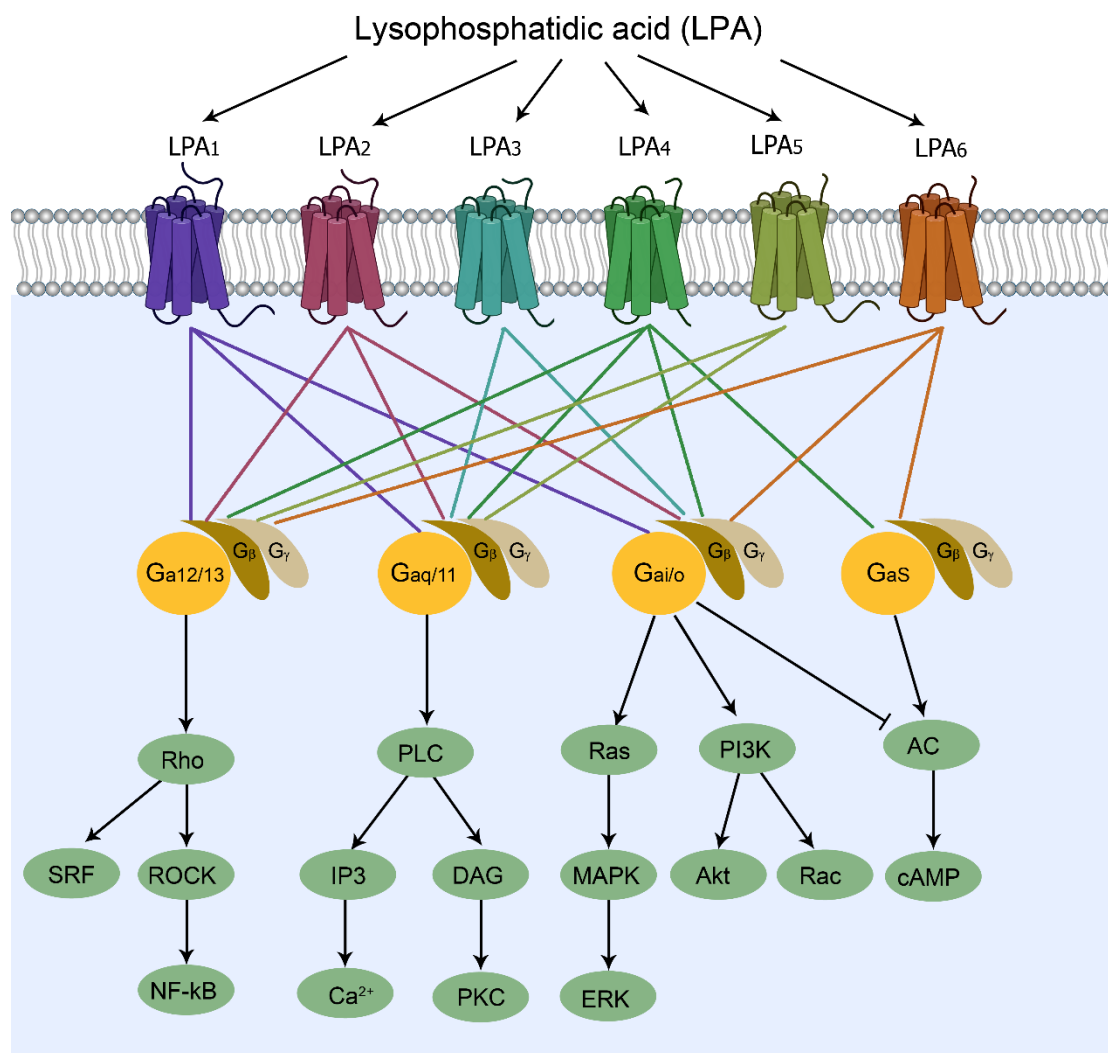
Figure 1.2. LPA downstream signaling.

As the largest family of cell transmembrane molecules, GPCRs are widely distributed in various cell types and play important roles in the transduction of a wide range of cell signals. Because of their significant roles in cell signaling, they have gained special interest for drug discovery. Currently, approximately 27% of marketed drugs are GPCR agonists or antagonists.[34] LPA receptor-mediated signaling cascades are important pathways involved in many processes, therefore drug discovery based on LPA is an attractive platform for researchers to explore. Because of the heterogeneity of LPA receptor subtypes, tissue distribution and expression profiles, blocking LPA production globally by inhibiting ATX can lead to a myriad of undesirable side effects. Hence, LPA-based drug development comes down to the discovery of

novel, potent LPA ligands targeting a specific receptor subtype without negating the physiological roles of other subtypes.

In recent years, there have been numerous studies on the synthesis and biological evaluation of agonists and antagonists that target LPA signaling, with variable efficacy and selectivity. The initial studies on ligands targeting LPA receptors were LPA-like compounds with a phosphate moiety and a long fatty acid chain. Based on the structure of **18:1 LPA**, several aspects of this endogenous ligand have been modified in structure-activity relationship studies to identify new drugs. These include modifications to the fatty acid chain length and its saturation, conversion of the phosphate group to isosteres, and substitution at the *sn*-2 position.[35-36] However, these compounds demonstrate undesirable bioavailability due to their hydrophobic nature. The first non-lipid LPA$_1$ ligands, isoxazole and thiazole derivatives, were reported in 2001,[37] followed by the development of the isoxazole-derived LPA receptor antagonist **Ki16425**. The phenylethoxy carbamoyl moiety of **Ki16425** inspired many other LPA ligands discovery with different five membered rings to achieve subtype activity and selectivity.[38-40] Among ligands targeting LPA receptors, LPA$_1$- and LPA$_1$/LPA$_3$-targeting ligands have been the most studied. The aim of this chapter is to review recent developments in drug discovery that target LPA receptors, as well as LPA ligands under clinical trials for diagnostic and therapeutic purposes.

2. LPA$_1$ ligands

2.1 Agonists

N-acyl ethanolamine phosphoric acid (**NAEPA**) (Fig 1.3A) was the first agonist based on LPA.[41] It was synthesized by replacing the glycerol on LPA with ethanolamine and characterized as an LPA mimetic with an EC$_{50}$ of around 40 nM in human platelet aggregation. Later, this compound was confirmed as an LPA$_1$/LPA$_2$ dual agonist which is equipotent with **18:1 LPA** at inducing calcium release *in vitro*.[35]

By using **NAEPA** as a lead compound, several modifications were made to achieve higher activity and selectivity. Substituting the α-carbon of its phosphate head group with an electron-withdrawing α-keto group (**1**) (Fig 1.3A) resulted in a selective LPA$_1$ agonist (EC$_{50}$ = 221 nM) with minimal LPA$_2$ activity and no LPA$_3$ activity from GTP[γ^{35}S] assays (data not available for LPA$_{4-6}$ receptors).[42] Furthermore, isosteric substitution of the phosphate with thiophosphate (**2**) (Fig 1.3A) resulted in a highly selective compound with LPA$_1$ agonistic activity (EC$_{50}$ = 318 nM)

without any LPA$_{2-3}$ agonism (Table 1.1).[42] However, α-monofluoromethylene- (**3**) (Fig 1.3A) and α-difluoromethylene- (**4**) (Fig 1.3A) substituted alkyl LPA analogues were found to be weak LPA$_{1-3}$ agonists in inducing cell migration.[43]
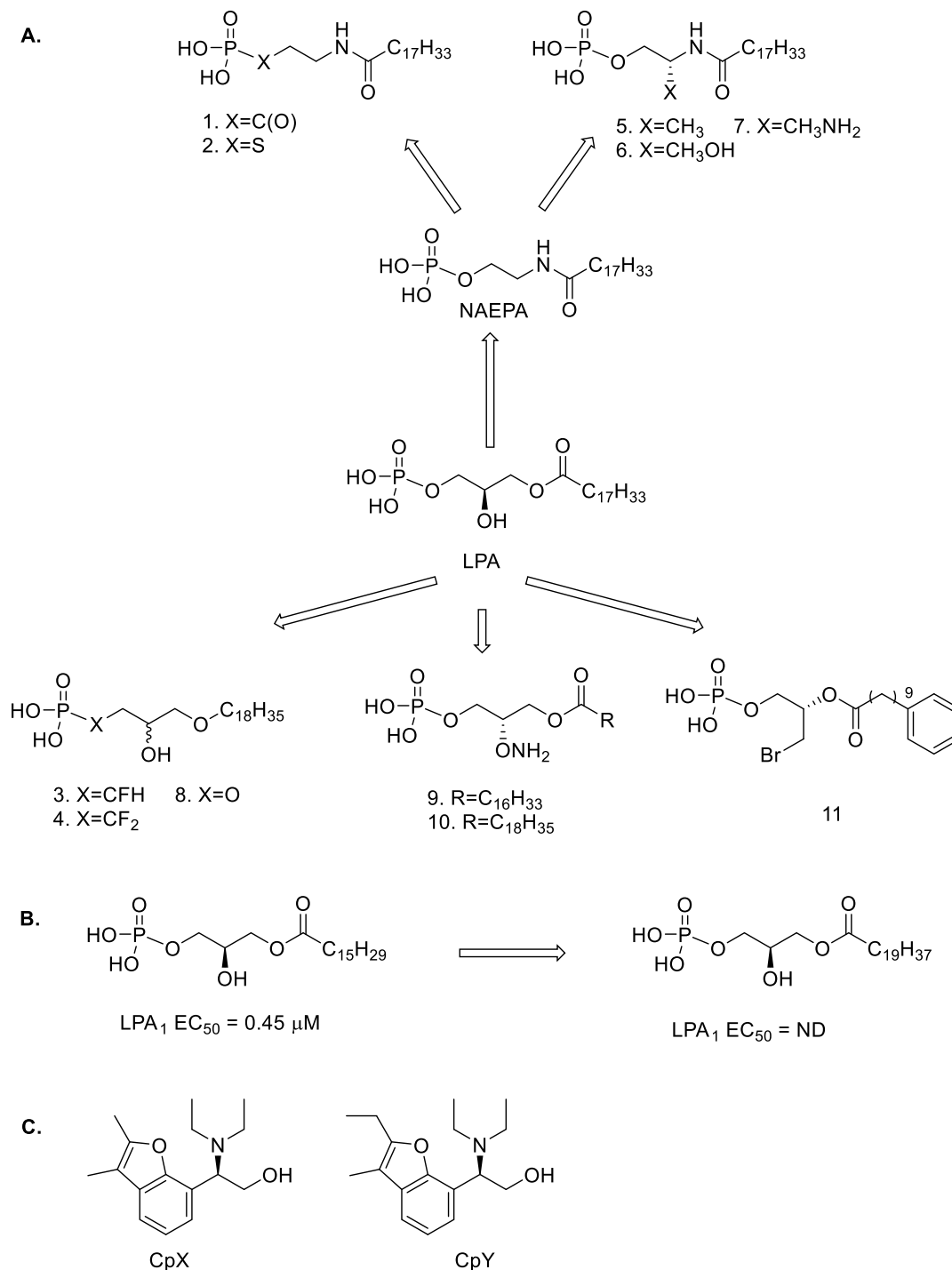

Figure 1.3. Lipid-based (A and B) and non-lipid (C) LPA$_1$ agonists. NE: No effect was observed.

Based on **NAEPA**, substitutions were made at the second carbon to obtain a series of 2-substituted **NAEPA** analogues (**5, 6, 7**) (Fig 1.3A). It was found that the agonistic activity decreased as the substituent bulk increased. Smaller substituents showed greater potency, with the methyl (**5**), methylene hydroxy (**6**), and methylene amino substituents (**7**) being more potent at the LPA$_1$ receptor than **18:1 LPA** from the GTP[γ^{35}S] binding assay. Strong stereoselectivity was also observed in this series of ligands, with the (*R*)-configured ligands being more potent.[44]

Based on the structure of **18:1 LPA**, Xu *et al.* replaced the ester by exchanging this functional group with an ether, resulting in (*R*) and (*S*) enantiomeric alkyl LPA analogues (**8**) (Fig 1.3A). These two alkyl LPA analogues were equipotent at inducing cell migration as **18:1 LPA** in three cancer cell lines, and were confirmed as non-enantioselective, non-selective agonists for the LPA$_{1-3}$ receptors.[43]

Modifications were also made at the *sn*-2 position of LPA to obtain compounds in racemic or optically pure forms. Gajewiak *et al.* introduced nucleophilic aminooxy (AO) functionality at the *sn*-2 position in a stereo-controlled way to produce enantiomerically pure *sn*-2 AO-LPA analogues with a palmitoyl (**9**) or oleoyl (**10**) acyl chain (Fig 1.3A). These AO-LPA analogues showed LPA$_1$ (EC$_{50}$ > 4630 nM), LPA$_2$ (EC$_{50}$ = 382-2450 nM), and LPA$_4$ (EC$_{50}$ ~1880 nM) receptor agonistic activity, with the LPA$_2$ receptor activity being the most potent. Antagonistic activity was observed towards the LPA$_3$ receptor (IC$_{50}$ = 56-2670 nM). The oleoyl ligand was the most potent LPA$_2$ agonist among this series of compounds with an EC$_{50}$ of 382 nM for LPA$_2$ (LPA$_1$ EC$_{50}$ > 7440 nM, LPA$_4$ EC$_{50}$ = 1880 nM).[45]

The length and structure of the alkyl chain of LPA also influence the activity. In an attempt to discover novel compounds with specific LPA$_1$ agonistic activity, González-Gil *et al.* started with the structure of **18:1 LPA** and made comprehensive modifications on the alkyl chain to obtain a series of compounds. The activities of these compounds were evaluated by the calcium mobilization assay. It was found that small changes in the number of methylene units exerted significant influence on EC$_{50}$ values. Cutting two methylene units resulted in a twofold improvement in LPA$_1$ agonistic activity, while adding two more methylene units resulted in a completely inactive compound (Fig 1.3B).[46] By replacing the hydrophobic chain with a phenyl moiety, the most potent and selective LPA$_1$ agonist to date was identified (Table 1.1), with

enantioselectivity towards the (*S*) configuration (**11**) (Fig 1.3A). This compound had an EC_{50} of 0.24 μM for the LPA_1 receptor without any activity on LPA_{2-6}.[46]

In 2020, two non-lipid benzofuran ethanolamine derivatives were reported as LPA_{1-3} agonists which were initially characterized as orphan receptor smooth muscle contracting agents. These agonists, **CpX** and **CpY** (Fig 1.3C), showed lower potency in the calcium assay compared with **18:1 LPA**, but higher *in vivo* efficacy because of better plasma stability. In a rat model, these two LPA agonists could induce intraurethral pressure, which was dose-dependently antagonized by the LPA antagonist **Ki16425** (Fig 1.5).[47]

2.2 Antagonists

In some cases, agonists and antagonists for the same target can share similar structures; a good agonist can be used as an inspiration for antagonist design. The LPA_1/LPA_3 dual antagonist **VPC12249** (Fig 1.4) was discovered by changing substituents at the second carbon of the LPA_1 agonist **NAEPA**. In this series of **NAEPA** analogues, compounds with small 2-substituted moieties were potent LPA agonists. However, this agonism was changed into antagonism with the introduction of a bulky group containing two aromatic rings (Fig 1.4). This benzyl-4-oxybenzyl-substituted **NAEPA** analogue, **VPC12249**, was found to be devoid of any agonist activity. It turned out to be an LPA_1/LPA_3 antagonist with an enantiomeric preference for the (*S*) configuration, having K_i values of 137 nM and 428 nM for the LPA_1 and LPA_3 receptors, respectively.[44]

Based on the structure of **VPC12249**, the acyl chain and linker regions were modified and a series of **VPC12249** analogues were synthesized. In this new series of **VPC12249** analogues, a 2-pyridyl derivative (**12**) (Fig 1.4) was found to be a potent dual LPA_1/LPA_3 antagonist, with a K_i value of 18 nM at LPA_1 in the GTPγS binding assay.[48]
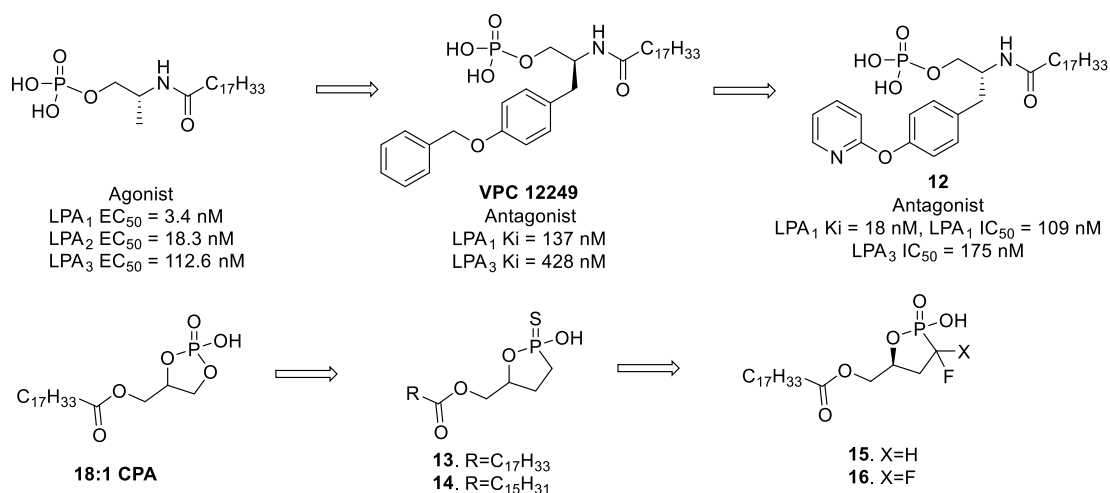
Figure 1.4. Lipid-based LPA$_1$ antagonists.

Cyclic phosphatidic acid (**CPA**) is a naturally occurring analogue of LPA in which an oxaphospholane is formed by the *sn*-2 hydroxy group and the *sn*-3 phosphate group (Fig 1.4). Based on the structure of **CPA**, Xu *et al.* replaced the phosphate group with its bioactive mimetics and found that phosphonothioate (**13, 14**) (LPA$_1$ IC$_{50}$ = 799-941 nM, LPA$_3$ IC$_{50}$ = 1270-2340 nM) and monofluoromethylene phosphonate (**15**) (LPA$_1$ IC$_{50}$ = 106 nM, LPA$_3$ IC$_{50}$ = 7720 nM) analogues were LPA$_1$/LPA$_3$ dual antagonists with no activity on the LPA$_2$ receptor. However, the antagonism was lost and weak agonism was observed with the introduction of a second fluorine atom (**16**) (Fig 1.4).[49]

The non-lipid low-molecular-weight compound **Ki16425** (Fig 1.5) was discovered by screening from a library established by the Kirin Brewery Co. Ltd, and its pharmacological properties were tested *in vitro*.[50] **Ki16425** showed effective antagonism against the LPA$_1$ and LPA$_3$ receptors in several cell types in both the calcium assay and the GTPγS binding assay. As well, this compound markedly inhibited **18:1 LPA**-induced cell proliferation and migration in 3T3 fibroblasts.[50] By using **Ki16425** as a lead compound, Yamamoto *et al.* converted the 3-methyl-isoxazole ring of **Ki16425** to 3- and 5-unsubstituted-isoxazolyl derivatives and a chlorocyclopentene isoxazole derivative (**17**) (Fig 1.5), resulting in 5 times higher LPA$_1$ antagonistic activity (IC$_{50}$ = 0.13 µM).[51] Later, enantiomers of **Ki16425** were synthesized using a 1,3-dipolar cycloaddition strategy. **(R)-Ki16425** showed greater potency and inhibitory activity against cell migration than racemic and **(S)-Ki16425**.[52]
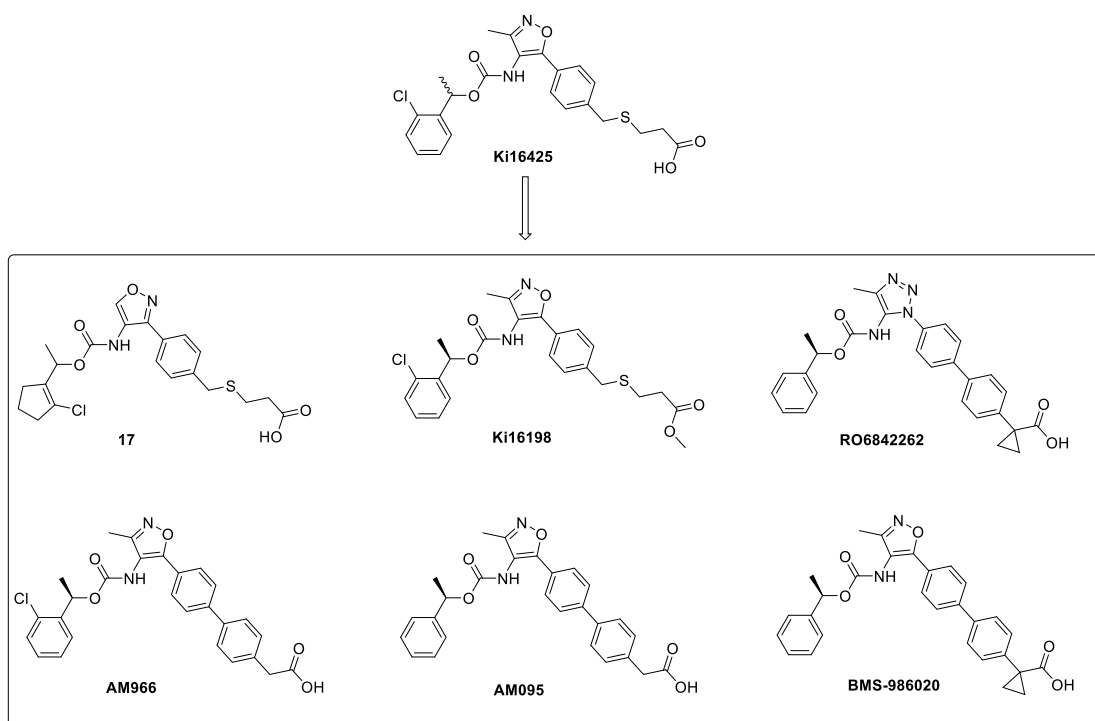
Figure 1.5. **Ki16425**-derived non-lipid LPA$_1$ antagonists.

Based on the structure of **(R)-Ki16425**, Komachi *et al.* prepared a methyl ester derivative of **Ki16425** to improve its oral activity. This ester product, **Ki16198** (Fig 1.5), showed similar potency and selectivity as **Ki16425**. **Ki16198** prevented pancreatic cancer cell migration and invasion induced by **18:1 LPA** *in vitro* and, when orally administered in a mouse model of pancreatic cancer, ameliorated metastasis.[53]

Based on the structure of **Ki16425**, Amira Pharmaceuticals designed and synthesized a series of biphenyl-substituted isoxazole analogues. Among them, **AM966**, **AM095**, and **BMS-986020** (previously **AM152/BMS-986202**) (Fig 1.5) stand out as promising LPA ligands. **AM966** is an orally active LPA$_1$ antagonist with significantly greater potency and selectivity than **Ki16425**. The IC$_{50}$ of **AM966** is 17 nM at the LPA$_1$ receptor as determined by the calcium assay in CHO cells, with negligible activity on the LPA$_2$, LPA$_3$, LPA$_4$ and LPA$_5$ receptors (IC$_{50}$ = 1700 nM, 1600 nM, 7700 nM, and 8600 nM, respectively). In human melanoma cells, lung fibroblasts, and CHO cells, **AM966** inhibited chemotaxis induced by **18:1 LPA**. Administration of **AM966** in a lung-injured mouse model ameliorated markers of injury, vascular leakage, fibrosis, and inflammation.[54]

AM095 has a similar structure compared to **AM966**, but has an unsubstituted phenyl group rather than a chlorine-substituted phenyl group (Fig 1.5). This gives a similar potency as **AM966** but dramatically higher selectivity towards LPA_1 (LPA_1 IC_{50} = 25 nM, LPA_{2-5} IC_{50} > 8000 nM) (Table 1.1). In a mouse model of scleroderma, pharmacological inhibition of LPA_1 with **AM095** attenuated dermal fibrosis.[55] Furthermore, **AM095** also inhibited histamine release stimulated by **18:1 LPA** in a mouse model and attenuated lung injury by reducing lung collagen, vascular leakage, and inflammation. The antifibrotic properties of **AM095** were further observed in the kidney of a mouse model of unilateral ureteral obstruction[56] and diabetic nephropathy.[57] In recent years, the therapeutic potential of **AM095** was also shown in mouse models of cerebral ischemia by antagonizing both LPA-induced brain damage and proinflammatory responses through the LPA_1 receptor.[58]

In 2012, Qian *et al.* discovered the triazole-derived carboxylic acid compound **RO6842262** (Fig 1.5),[59-60] which had great potency and high LPA_1 selectivity (LPA_1 IC_{50} = 25 nM, LPA_3 IC_{50} > 30 µM) (Table 1.1) in the calcium release assay. This compound decreased **18:1 LPA**-induced proliferation and contraction in human lung fibroblasts. Oral dosing of this compound in experimental mice showed attenuated plasma histamine release induced by LPA in a dose-dependent manner,[60] showing great potential for the treatment of lung fibrosis. Two years later, the same group reported a novel chemical class of non-carboxylic acid 5-amino-4-cyanopyrazole analogues with antagonism against the LPA_1 receptor. Among this group of compounds, three ligands (**18, 19, 20**) (Fig 1.6) showed high potency and selectivity towards the LPA_1 receptor by the FLIPR assay (LPA_1 IC_{50} = 74-111 nM, LPA_3 IC_{50} > 30 µM) (data not available for $LPA_{2, 4-6}$ receptors), and these novel compounds showed an inhibitory effect towards lung fibroblast proliferation and contraction in response to **18:1 LPA**.[61]

Over the past two decades, it has been discovered that some GPCR ligands can preferentially target specific pathways while blocking other pathways of a receptor. These compounds turned out to be 'biased' ligands that can achieve functional selectivity. The development of biased ligands is now an active area of research, as selectively activating or blocking specific signaling pathways results in increased drug efficacy with reduced side effects.[62]

In order to achieve functional selectivity, Shimizu *et al.* used high-throughput screening to discover five LPA_1 negative allosteric modulators with biased antagonism that preferentially

inhibit LPA-induced activation of G_q intracellular downstream signals. These five ligands showed comparable inhibitory activity in the calcium mobilization assay with **Ki16425**, with compound **LQ4** (Fig 1.6) being the most potent (IC_{50} = 17 nM). None of these ligands showed any activity in cAMP accumulation or β-arrestin recruitment assays,[63] which suggests that these ligands induced biased signaling.
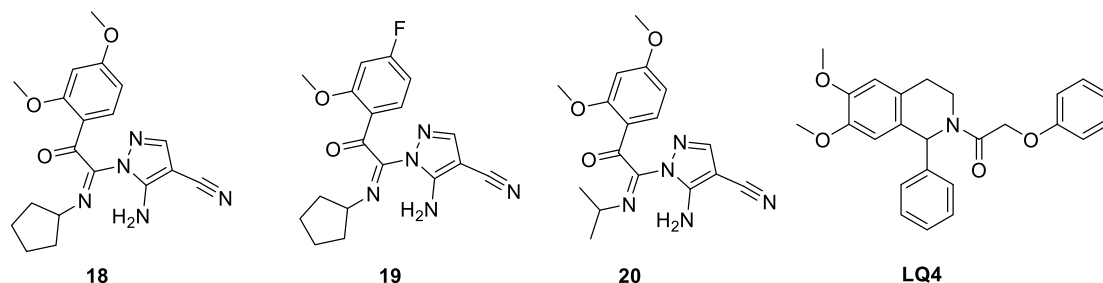
Figure 1.6. LPA$_1$ antagonists with novel chemical scaffolds.

By optimizing a hit compound from high-throughput screening, the benzamide derivative with high antagonistic activity against LPA$_1$ (LPA$_1$ IC_{50} = 160 nM, LPA$_2$ IC_{50} = 8600 nM, LPA$_3$ IC_{50} > 10000 nM), **ONO-7300243** (Fig 1.7) (Table 1.1), was discovered.[64] This compound showed strong therapeutic effects *in vivo* in a rat intraurethral pressure model induced by LPA dose-dependently, highlighting the potential for the treatment of benign prostatic hyperplasia. Furthermore, in the lung cancer cell line A549, treatment with **ONO-7300243** completely inhibited cell migration, proliferation and colony formation induced by **18:1 LPA**.[65] By using **ONO-7300243** as a lead compound, the same research group converted the amide to a secondary alcohol group by scaffold hopping and synthesized a series of compounds to attain better *in vivo* efficacy. The best LPA$_1$ antagonist was obtained by incorporating an indane and pyrrole structure into the benzene rings. This ligand, named **ONO-0300302** (Fig 1.7), showed the best reduction of intraurethral pressure in rat and dog models of prostatic hyperplasia, with an *in vitro* efficacy of 0.16 nM as shown by the calcium assay after 24h incubation.[66]
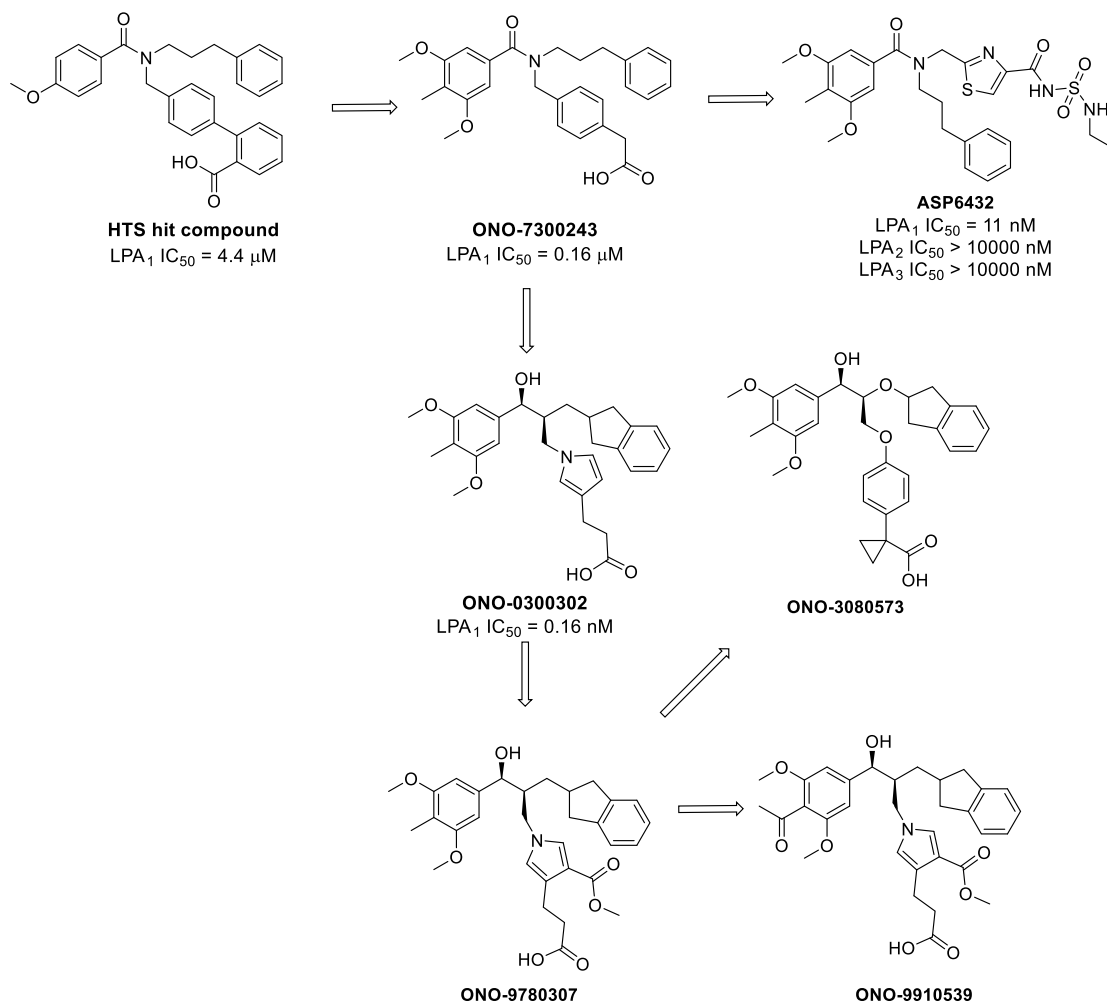
Figure 1.7. **ONO-7300243**-derived LPA$_1$ antagonists.

ASP6432 is an **ONO-7300243** analogue with a sulfamoyl thiazole moiety developed by Astellas Pharma Inc (Fig 1.7). **ASP6432** has an IC$_{50}$ of 11 nM in human LPA$_1$ and 30 nM in rat LPA$_1$ with high selectivity for the LPA$_1$ receptor (LPA$_2$ IC$_{50}$ > 10000 nM, LPA$_3$ IC$_{50}$ > 10000 nM, LPA$_4$ IC$_{50}$ = 114 nM, and LPA$_5$ IC$_{50}$ > 30000 nM; LPA$_6$ not tested). In a rat model, administration of **ASP6432** inhibited LPA-induced urethral and prostate strips contractions and intraurethral pressure elevation. In cultured human prostate stromal cells, **ASP6432** suppressed **18:1 LPA**-induced cell proliferation.[67] Furthermore, **ASP6432** decreased urethral perfusion pressure and improved voiding dysfunction in a rat model, showing potential for the treatment of prostate hyperplasia and other lower urinary tract diseases.[68]

In 2015, three crystal structures of human LPA_1 bound with antagonists (PDB: 4Z34, 4Z35, 4Z36) were reported.[69] Three antagonists, **ONO-3080573**, **ONO-9780307**, and **ONO-9910539** (Fig 1.7) were designed to be co-crystalized with the LPA_1 receptor. **ONO-9780307** was selected as the initial structure based on the stability assay data in order to stabilize the LPA_1 receptor in a single conformation. Based on **ONO-9780307**, **ONO-9910539** and **ONO-3080573** were designed with ideal stability and lipophilicity to improve interactions with the LPA_1 receptor and reduce torsional strain. These three antagonists, **ONO-3080573**, **ONO-9780307**, and **ONO-9910539**, have respective IC_{50} values of 11 nM, 27 nM, and 22 nM for the LPA_1 receptor by the calcium assay.

3. LPA_2 ligands

3.1 Agonists

Based on computational models, Virag *et al.* selected fatty alcohol phosphate (**FAP**) as the lead scaffold to synthesize a series of compounds with a phosphate headgroup attached to a hydrocarbon chain with varying lengths, leading to the discovery of the first LPA_2-selective agonist with weak antagonistic activities for the LPA_1 and LPA_3 receptors, the dodecyl fatty alcohol phosphate compound **21** (EC_{50} = 700 nM) (Fig 1.8).[70] This finding suggested that **FAP**, which does not contain a glycerol backbone, represents the minimal pharmacophore that can recognize LPA receptors. Based on the structure of these **FAP-LPA** analogues, the phosphate was replaced by thiophosphate and an oleoyl-thiophosphate compound was produced (**22**) (Fig 1.8) that turned out to be a full LPA_2 agonist (EC_{50} = 244 nM) with partial agonistic activity towards LPA_1 and LPA_3.[71]
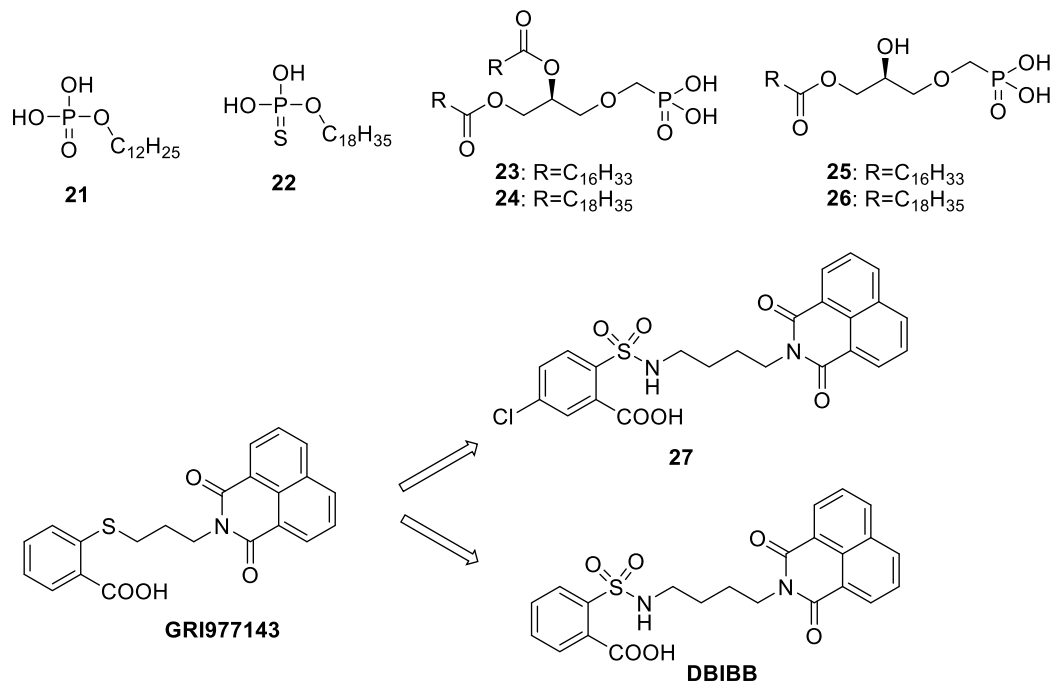
Figure 1.8. LPA$_2$ agonists.

Based on the structures of PA and LPA, a methylene group was inserted between the glycerol and the phosphate head to develop bis-acylated PA analogues (**23, 24**) (Fig 1.8) and *sn*-2-OH alkoxymethylene phosphonate LPA analogues (**25, 26**) (Fig 1.8). These four ligands were tested in HT-29 colon cancer cell lines, in which the LPA$_2$ receptor is exclusively expressed. In HT-29 cells, it was found that these four ligands were more potent than **18:1 LPA** in activating downstream kinase of LPA signaling pathways, suggesting that these compounds are LPA$_2$ agonists.[72]

In search of non-lipid LPA$_2$ agonists, structure-based virtual screening was performed, and four non-lipid compounds stood out as LPA$_2$ agonists. Among these LPA$_2$ agonists, **GRI977143** (Fig 1.8), which has a specific agonistic activity towards LPA$_2$ (EC$_{50}$ = 3.3 μM), was selected for further evaluation. In several cell types, **GRI977143** could mimic the effect of LPA on cell growth and cell invasion. Furthermore, **GRI977143** played a pro-survival role in several cell types through the LPA$_2$ receptor.[73] The therapeutic potential of **GRI977143** was also proved in an autoimmune encephalomyelitis disease model, which could counterbalance the loss of endogenous LPA.[74]

By using **GRI977143** as a scaffold, a series of sulfamoyl benzoic acid analogues were developed by using an LPA$_2$ homology model, and these compounds were tested for LPA$_{1-5}$ activity. In doing this, the first non-lipid agonist of LPA$_2$ with picomolar activity was discovered (**27**) (Fig 1.8); furthermore, this ligand was able to specifically target LPA$_2$ without any activity on other LPA receptor subtypes (LPA$_2$ EC$_{50}$ = 5.1 pM, LPA$_1$ EC$_{50}$ > 10000 nM, LPA$_3$ EC$_{50}$ > 10000 nM, LPA$_4$ EC$_{50}$ > 10000 nM, and LPA$_5$ EC$_{50}$ > 10000 nM) (Table 1.1).[75] **DBIBB** (2-[[[4-(1,3-dioxo-1H-benz[de]isoquinolin-2(3H)-yl)butyl]amino]sulfonyl]-benzoic acid) (Fig 1.8) is another compound in this series with an EC$_{50}$ of 1.42 μM towards the LPA$_2$ receptor and no effect on other LPA receptor subtypes (data not available for LPA$_6$ receptor). In a mouse model of asthma, administration of **DBIBB** reduced respiratory inflammation and cytokines.[76] Potential for the therapeutic use of **DBIBB** in treating acute radiation syndrome caused by γ-radiation was also observed.[77]

3.2 Antagonists

The first reported LPA$_2$ antagonist was a tetradecyl phosphonate **FAP** analogue (**28**) (Fig 1.9) with an IC$_{50}$ of 5.5 μM for LPA$_2$ (LPA$_1$ IC$_{50}$ > 30 μM, LPA$_3$ IC$_{50}$ = 3.1 μM), and this compound reduced LPA-induced responses in RH7777 cells.[71]

The first non-lipid LPA$_2$ antagonist with high potency and selectivity was discovered by high-throughput screening. The HTS hit compound was modified to afford two different chemotypes of compounds that are potent LPA$_2$ antagonists with high selectivity: a thienopyrimidine tolyl derivative (**29**) (Fig 1.9) (LPA$_2$ IC$_{50}$ = 0.26 μM, LPA$_1$ IC$_{50}$ > 50 μM, LPA$_3$ IC$_{50}$ > 50 μM) (Table 1.1), and a thienopyrimidine sulfonyl derivative (**30**) (Fig 1.9) (LPA$_2$ IC$_{50}$ = 0.017 μM, LPA$_1$ IC$_{50}$ > 50 μM, LPA$_3$ IC$_{50}$ > 50 μM) (Table 1.1). The sulfonyl derivative was further tested *in vitro* and showed an inhibitory effect on ERK phosphorylation and cell proliferation induced by LPA in the HCT-116 colon cancer cell line.[78]
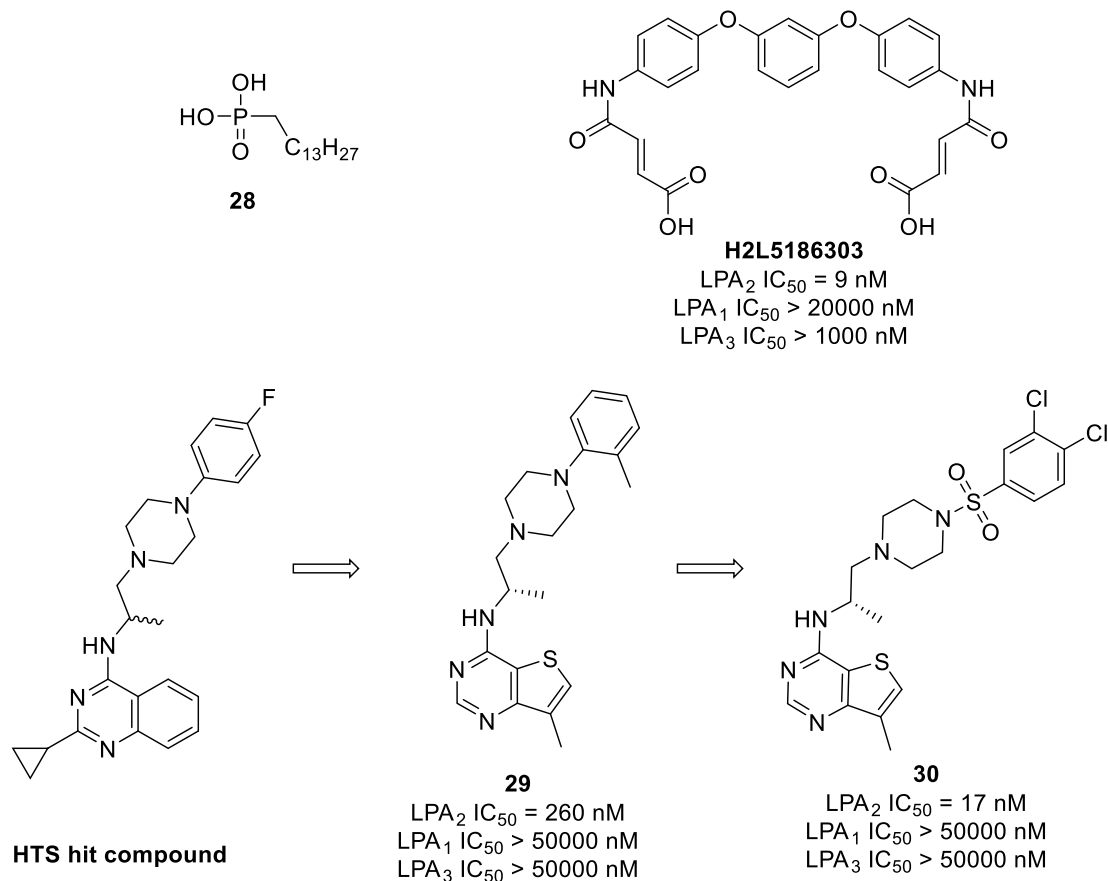
Figure 1.9. LPA$_2$ antagonists.

The most potent LPA$_2$ antagonist, **H2L5186303** (Fig 1.9), was discovered by structure-based virtual screening using a three-point pharmacophore. **H2L5186303** showed the highest potency and selectivity for LPA$_2$ (LPA$_2$ IC$_{50}$ = 9 nM, LPA$_1$ IC$_{50}$ = 27354 nM, LPA$_3$ IC$_{50}$ = 1230 nM) (Table 1.1) among over 250 000 compounds screened.[79] In a mouse model of allergic inflammation, treatment with **H2L5186303** reduced airway hyperresponsiveness by inhibiting inflammation and decreasing the production of cytokines induced by LPA.[80]

4. LPA$_3$ ligands

4.1 Agonists

Initial structural-relationship studies on LPA ligands suggested that the phosphate headgroup on LPA is critical for LPA function.[81] Inspired by this, **OMPT** (1-oleoyl-2-*O*-methyl-*rac*-glycerophosphothionate) was discovered by converting the **LPA 18:1** hydroxy and phosphate groups to *O*-methoxy and thiophosphate groups (Fig 1.10), respectively. Indeed, **OMPT** conferred great potency and selectivity towards LPA$_3$.[82] **OMPT** induced downstream ERK

phosphorylation equipotently as **18:1 LPA** with a stronger proliferative effect. It induced these responses through the LPA_3 receptor specifically while the LPA_1 and LPA_2 receptors were not involved, suggesting high selectivity towards the LPA_3 receptor.[82] Later, enantiomerically pure **OMPT** was synthesized. It was reported that the unnatural configuration **(2S)-OMPT** was 5- to 20-fold more potent than **(2R)-OMPT** at activating the LPA_3 receptor and its downstream signaling due to less steric interference compared with **(2R)-OMPT**.[83] In OVCAR3 ovarian cancer cells, **(2S)-OMPT** had an EC_{50} value of 9.0 nM as ascertained by calcium release, while the EC_{50} of **(2R)-OMPT** was 46.5 nM.
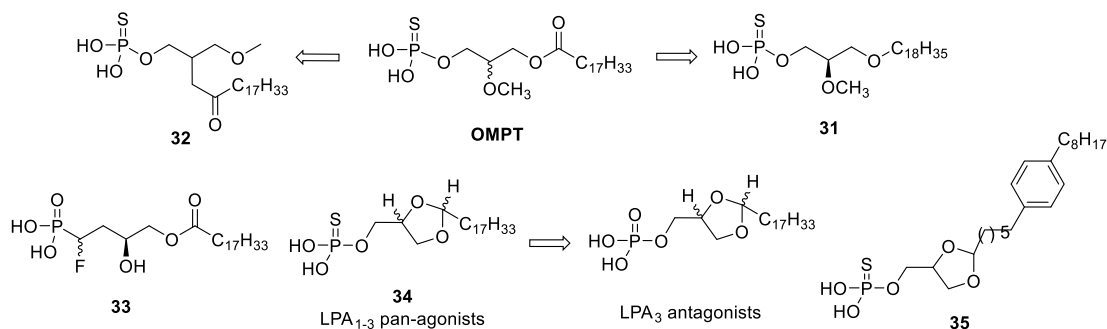
Figure 1.10. LPA_3 agonists.

Based on the structure of **OMPT**, the *sn*-1 *O*-acyl group was converted to an *O*-alkyl ether, and a series of enantiospecific alkyl phosphorothioate LPA analogues was obtained. These alkyl **OMPT** analogues were also found to be LPA_3 agonists, with (*R*)-alkyl **OMPT** (**31**) (Fig 1.10) being the most potent (EC_{50} of 62 nM in RH7777 cells). This LPA_3 agonist stimulated cell growth dose-dependently.[84]

Further replacement of the *sn*-2 methoxy group on **OMPT** by a methoxyethane group was carried out by Jiang *et al.* The resulting *sn*-2 radyl-**OMPT** analogues showed similar agonism towards the LPA_{1-6} receptors as their parent compound **OMPT**. An 18:1 acyl phosphorothioate analogue (**32**) (Fig 1.10) proved to be the most potent LPA_3 agonist, with an EC_{50} of 0.21 nM by the TGFα shedding assay.[85]

Based on the structure of *sn*-1-/*sn*-2-acyl-LPA, Xu *et al.* made isosteric substitutions on the phosphate group with fluorine-containing groups and consequently found a potent acyl α-fluoromethylene phosphate analogue (**33**) (Fig 1.10) which displayed an EC_{50} of 0.5 nM for the LPA_3 receptor. This compound was more potent than **18:1 LPA** at inducing calcium

mobilization and could activate downstream MAPK/AKT signaling more effectively.[86-87] However, changing the acyl chain of the fluoromethylene phosphate analogues to an alkyl chain resulted in the loss of LPA agonistic activity.[87]

Darmstoff analogues are LPA-like ligands with dioxolane incorporated into the hydrocarbon chain. Inspired by the structure of Darmstoff analogues, the phosphate group was replaced by thiophosphate and a series of optically pure Darmstoff analogues were developed. These pure stereoisomers with a thiophosphate group (**34**) (Fig 1.10) were pan-agonists of LPA_{1-3}, with the greatest potency for the LPA_3 receptor (LPA_3 EC_{50} = 127-265 nM). However, their corresponding stereoisomers with a phosphate group turned out to be LPA_3 antagonists (LPA_3 IC_{50} = 136-484 nM). Introduction of an aromatic ring in the hydrocarbon chain (**35**) (Fig 1.10) resulted in a selective LPA_3 agonist (EC_{50} = 692 nM) without any activity on the LPA_1 and LPA_2 receptors (Table 1.1).[88]

4.2 Antagonists

Diacylglycerol pyrophosphate (**DGPP**) and dioctyl-phosphatidic acid (**PA**) are naturally occurring phospholipids with an ionic phosphate group and a hydrophobic chain identified in microorganisms. It was reported that short chain **8:0 DGPP** and **8:0 PA** (Fig 1.11) are competitive, selective LPA_3 receptor antagonists with a weak antagonistic effect on LPA_1 and no activity at LPA_2. They inhibited LPA-induced responses in several cell types, with an IC_{50} value of 250 nM for LPA_3 and an IC_{50} value of 3 μM for LPA_1.[89] **8:0 DGPP** and **8:0 PA** are also involved in atherosclerotic processes by inhibiting platelet activities induced by acyl- and alkyl-LPA.[90] Based on the structure of **8:0 PA**, enantiomerically pure **PA** analogues with thiophosphate replacing the phosphate headgroup were synthesized, and these analogues antagonized the LPA_1/LPA_3 receptors with enantioselectivity. The (2R)-dioctanoyl PA analogue (**36**) (Fig 1.11) was the most potent LPA_3 antagonist with an IC_{50} value of 11 nM, while its (2S)-isomer was found to be an LPA_3 agonist with an EC_{50} value of 115 nM. The (2R)-dioctyl **PA** analogue (**37**) (Fig 1.11) was the most potent LPA_3 agonist with an EC_{50} value of 3 nM, while its (2S)-isomer was found to be an LPA_1/LPA_3 dual antagonist.[91]
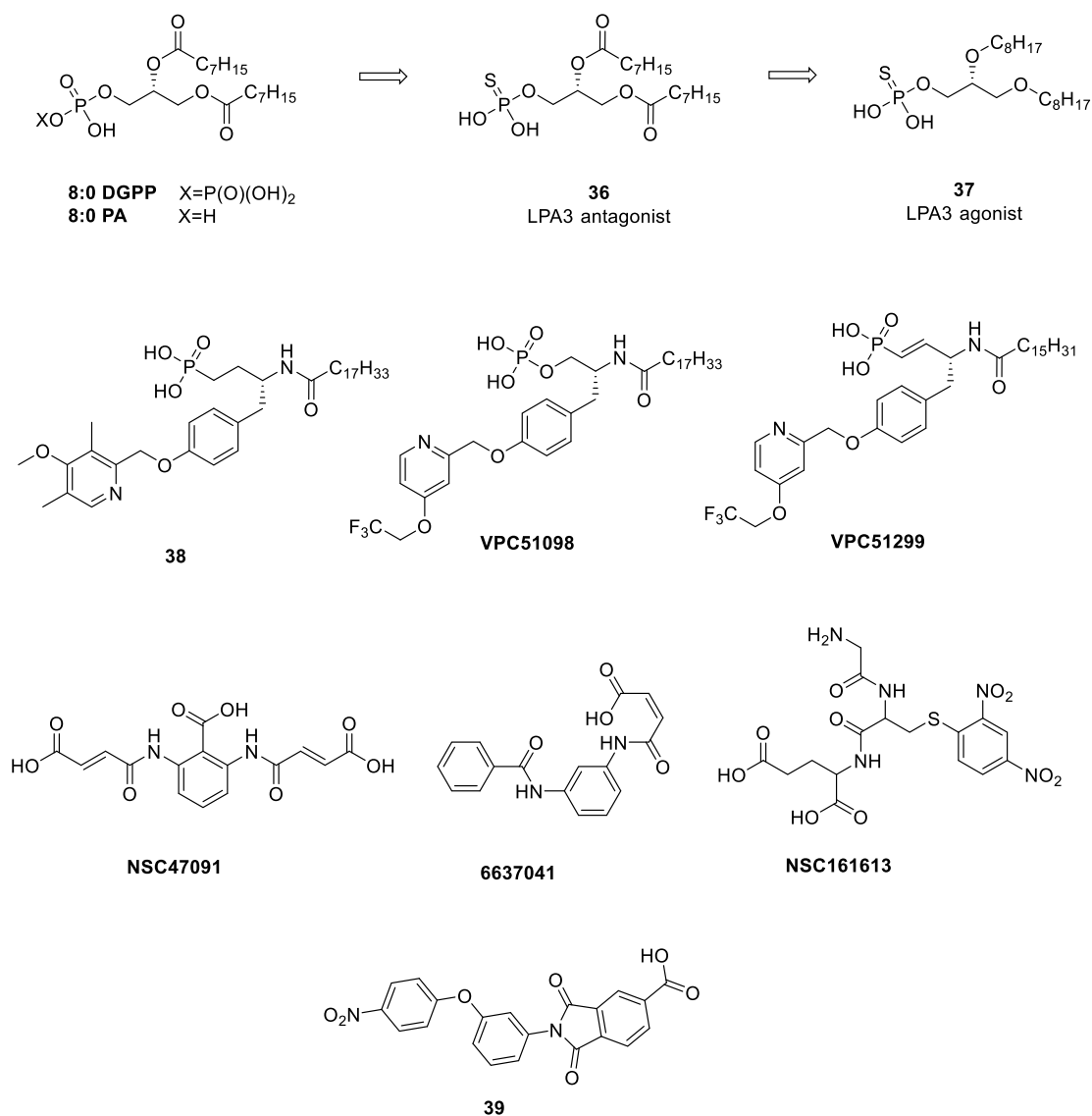
Figure 1.11. LPA$_3$ antagonists.

The first metabolically stable LPA$_3$-selective antagonist was developed in 2004. In order to achieve better stability, nonhydrolyzable phosphate head group mimetics were incorporated into the LPA$_1$/LPA$_3$ dual antagonist **VPC12249**, and a novel series of compounds with 2-pyridyl substituents was developed. Among them, a compound with a methoxy dimethylpyridine substituent (**38**) (Fig 1.11) was found to be a highly subtype-selective LPA$_3$ antagonist (IC$_{50}$ = 150 nM) without any activity on the LPA$_1$ receptor (Table 1.1). Another compound with a trifluoroethoxy substituent, **VPC51098** (Fig 1.11), was found to be a potent LPA$_1$/LPA$_3$ dual

antagonist (LPA$_1$ IC$_{50}$ = 84 nM, LPA$_3$ IC$_{50}$ = 48 nM) with greater potency at the LPA$_3$ receptor than **VPC12249**.[92]

Based on the structure of **VPC51098**, the labile phosphate head group was converted to vinyl phosphonate to achieve better stability. This phosphatase-resistant compound **VPC51299** (Fig 1.11) was validated as an LPA$_1$/LPA$_3$ dual antagonist with an indistinguishable potency from **Ki16425**. It blocked LPA agonist-induced calcium release dose-dependently with a half-life of 90 minutes in rats by intravenous dosing.[93]

In order to identify novel non-lipid LPA$_3$ antagonists with subtype selectivity, a pharmacophore was built for virtual screening. Two compounds had stood out as promising LPA$_3$ antagonists in this study, **NSC47091** and **NSC161613** (Fig 1.11). **NSC47091** was the first reported LPA$_2$/LPA$_3$ dual antagonist, with an IC$_{50}$ of 355 nM and 30 nM for the LPA$_2$ and LPA$_3$ receptors, respectively. **NSC161613** showed high potency and subtype selectivity towards the LPA$_3$ receptor with an IC$_{50}$ of 24 nM while it did not show any detectable activity on other LPA receptor subtypes (Table 1.1), making it the first reported non-lipid selective antagonist for the LPA$_3$ receptor. In an attempt to find more LPA$_3$ antagonists, a similarity search was performed by using **NSC47091** as the search target and five more LPA$_3$ antagonists were screened out. Among them, compound **6637041** (Fig 1.11) exhibited the greatest potency for LPA$_3$ by having an IC$_{50}$ of 347 nM.[94] By using the weak LPA$_3$ antagonist **H2L5747876** as the target, another similarity search was performed by the same research group and 10 novel antagonists were identified. The most potent compound was an LPA$_1$/LPA$_3$/LPA$_5$ antagonist (**39**) (Fig 1.11) with an IC$_{50}$ of 94 nM for LPA$_1$, 752 nM for LPA$_3$, and 463 nM for LPA$_5$.[79]

5. LPA$_4$ ligands

5.1 Agonists

LPA$_4$ is the first of the non-EDG family of LPA receptors. There has been significantly less success in the discovery of selective agonists and antagonists for LPA$_{4-6}$. Nevertheless, some ligands have still been reported, although mostly as partial or pan-agonists and antagonists (Fig 1.12A).
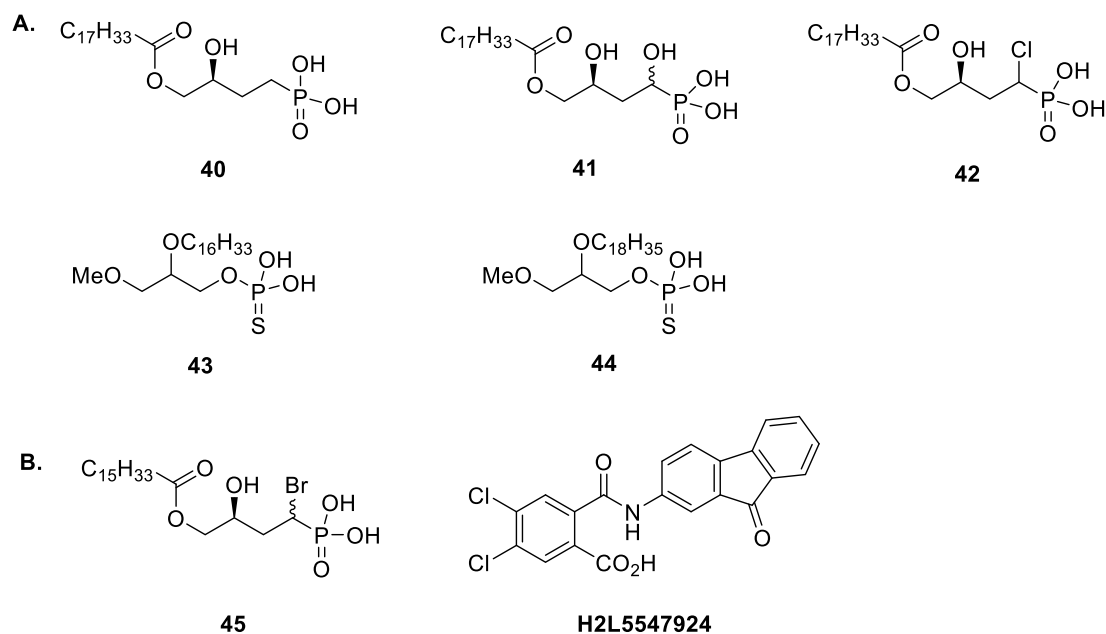
Figure 1.12. LPA$_4$ agonists (A) and antagonists (B).

Based on **18:1 LPA**, the phosphate group was modified to a methylene or substituted methylene group to achieve better stability, and several LPA$_4$ agonists were discovered.[95] The α-methylene phosphonate analogue (**40**) (Fig 1.12A) was seen to be weakly agonistic towards LPA$_4$, displaying an EC$_{50}$ of 3900 nM.[95] However, it was seen to have a stronger agonistic effect towards the LPA$_{1-3}$ receptors, with respective EC$_{50}$ values of > 2520 nM, > 281 nM, and > 1710 nM for these receptor isoforms. The α-hydroxymethylene phosphonate analogue (**41**) (Fig 1.12A) was discovered to be a potent and selective LPA$_3$ agonist (IC$_{50}$ value of 393 nM) with weak agonistic activity towards LPA$_1$, LPA$_2$, and LPA$_4$, exhibiting an IC$_{50}$ of > 1150 nM for LPA$_4$.[95] The α-chloromethylene phosphonate analogue (**42**) (Fig 1.12A) was shown to have a mixed agonist-antagonist profile.[95] This ligand displayed agonistic activity towards LPA$_4$ as well as LPA$_1$ and LPA$_3$, with respective EC$_{50}$ values of 3480 nM, 528 nM, and > 2670 nM, but had an IC$_{50}$ of 1690 nM for the LPA$_2$ receptor. This group employed calcium mobilization assays for the determination of ligand properties, utilizing RH7777 cells for LPA$_{1-3}$ assays and CHO cells for LPA$_4$ assays.[95]

In 2013, Jiang *et al*. reported on the synthesis and bioactivities of several phosphorothioate LPA analogues on LPA$_{1-6}$. These were measured by using the TGFα shedding assay in HEK293 cells, which endogenously express all members of the LPA receptor family. For the *sn*-2 alkyl

phosphorothioate analogues **43** and **44** (Fig 1.12A), good LPA$_4$ agonistic activity was seen, having EC$_{50}$ values of 3.8 nM and 4.5 nM, respectively. As well, an EC$_{50}$ of 2.4 nM was observed for the racemic acyl **OPMT** analogue **32** (Fig 1.10). However, these compounds were not shown to be specific for LPA$_4$, as they demonstrated a higher potency for both the LPA$_3$ and LPA$_5$ receptors.[85]

5.2 Antagonists

In 2007, Jiang et al. reported the first antagonists for LPA$_4$. Their palmitoyl α-bromo analogue of LPA (**45**) (Fig 1.12B) was determined to have an IC$_{50}$ of 266 nM for LPA$_4$. This ligand was observed to be a pan-antagonist of the LPA receptor family, displaying weak antagonistic activity towards LPA$_1$, LPA$_2$, and LPA$_3$, with respective IC$_{50}$ values of 1500 nM, 1420 nM, and 1160 nM.[95] In a virtual screening that was done to identify LPA$_2$ agonists, a ligand named **H2L5547924** (Fig 1.12B) was reported to demonstrate moderate antagonist activity towards LPA$_4$, with an IC$_{50}$ of 1.3 μM. However, it also demonstrated similar levels of antagonist activity towards LPA$_1$ and LPA$_3$, as well as a similar agonist potency for LPA$_2$, making it highly nonspecific.[73]

6. LPA$_5$ ligands

6.1 Agonists

Several LPA$_5$ agonists have been reported in the recent literature. Williams *et al.* succinctly summarized the effects of 14 previously reported lipid-based LPA receptor agonists on LPA$_5$, performing calcium mobilization assays with RH7777 and CHO cells. Among these ligands, **18:1 LPA** and **18:1 AGP** (Fig 1.13) had agonistic activity for the LPA$_5$ receptor, with the EC$_{50}$ values being 15 nM and 2 nM, respectively.[96] Oh *et al.* reported the potencies for small lipid-based molecules in 2008.[97] They studied several ligands, and their EC$_{50}$ values were measured by the serum response element (SRE)-luciferase assay in CV-1 cells infected with the LPA$_5$ receptor. Among these ligands, farnesyl pyrophosphate (**FPP**) (Fig 1.13) had the highest agonistic activity for the LPA$_5$ receptor, displaying an EC$_{50}$ of 260 nM. In Jiang *et al.*'s report on phosphorothioate analogues of LPA, the most important finding concerning LPA$_5$ was that compound **43** (Fig 1.12A), an *sn*-2 alkyl phosphorothioate analogue, was the most potent for this receptor, showing an EC$_{50}$ of 0.26 nM in the TGFα shedding assay using HEK293 cells.[85]

AGP 18:1

Farnesyl pyrophosphate (FPP)
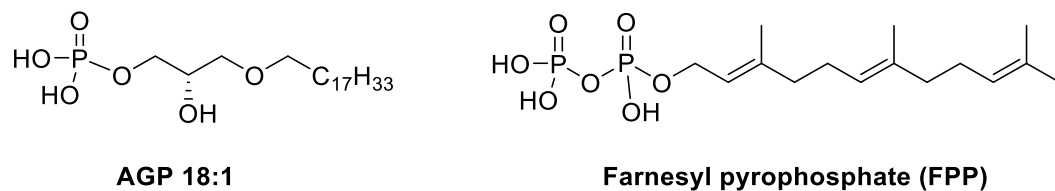
Figure 1.13. LPA$_5$ agonists.

6.2 Antagonists

Based on computational methods and previous ligands, a pharmacophore of the LPA$_5$ receptor was established and used for *in silico* screening, and two novel LPA$_5$ antagonists with non-lipid structures, **H2L5987411** and **H2L5765834** (Fig 1.14), were discovered. Their respective IC$_{50}$ values were 3.5 μM and 463 nM for LPA$_5$ in the calcium assay. **H2L5987411** also displayed an IC$_{50}$ of 1.4 μM for LPA$_4$ (no effect on LPA$_{1-3}$), while **H2L5765834** had respective IC$_{50}$ values of 94 nM and 752 nM for LPA$_1$ and LPA$_3$ (no effect on LPA$_2$ and LPA$_4$). Each of the two antagonists play an inhibitory role in LPA-induced platelet shape change.[96]

In 2012, Kozian *et al*. identified a diphenyl pyrazole carboxylic acid modulator of LPA$_5$ activity via high-throughput screening at Sanofi-Aventis Deutschland GmbH. This compound (**47**) (Fig 1.14) was seen to have an IC$_{50}$ of 2.2 μM on LPA$_5$ receptor-mediated platelet aggregation. When this compound was tested as an inhibitor in an RH7777 cell line stably expressing human LPA$_5$, it was determined to have an IC$_{50}$ of 0.8 μM.[98]

In 2013, Nazaré *et al*. from Sanofi-Aventis reported on the activities of exemplary compounds from a series of pyridine derivatives, a series of pyrazole derivatives, and a series of benzo[1,3]dioxine derivatives, as well as their uses as LPA$_5$ antagonists. Within each series, three assays were performed: a platelet aggregation assay using human washed platelets, a FLIPR assay using the human mast cell line HMC-1, and a second FLIPR assay using the murine microglia cell line BV-2. The most potent compound among the series of pyridine derivatives was compound **48** (Fig 1.14), which had IC$_{50}$ values of 3.6 μM, 0.13 μM, and 3.2 μM with respect to the assays listed above. Among the series of pyrazole derivatives, compound **49** (Fig 1.14) was the most potent, showing IC$_{50}$ values of 0.03 μM for the HMC-1 FLIPR assay and 0.2 μM for the BV-2 FLIPR assay. The most potent compound in the platelet aggregation assay for this series was compound **50** (Fig 1.14), with an IC$_{50}$ of 2.9 μM. Lastly, among the series of

benzo[1,3]dioxine derivatives, compounds **51** (Fig 1.14) and **52** (Fig 1.14) were seen to be the most potent. Compound **51** had respective IC$_{50}$ values of 1.1 μM, 3.2 μM, and 4.1 μM, whereas compound **52** had an IC$_{50}$ of 1.5 μM in both FLIPR assays.[99-101]
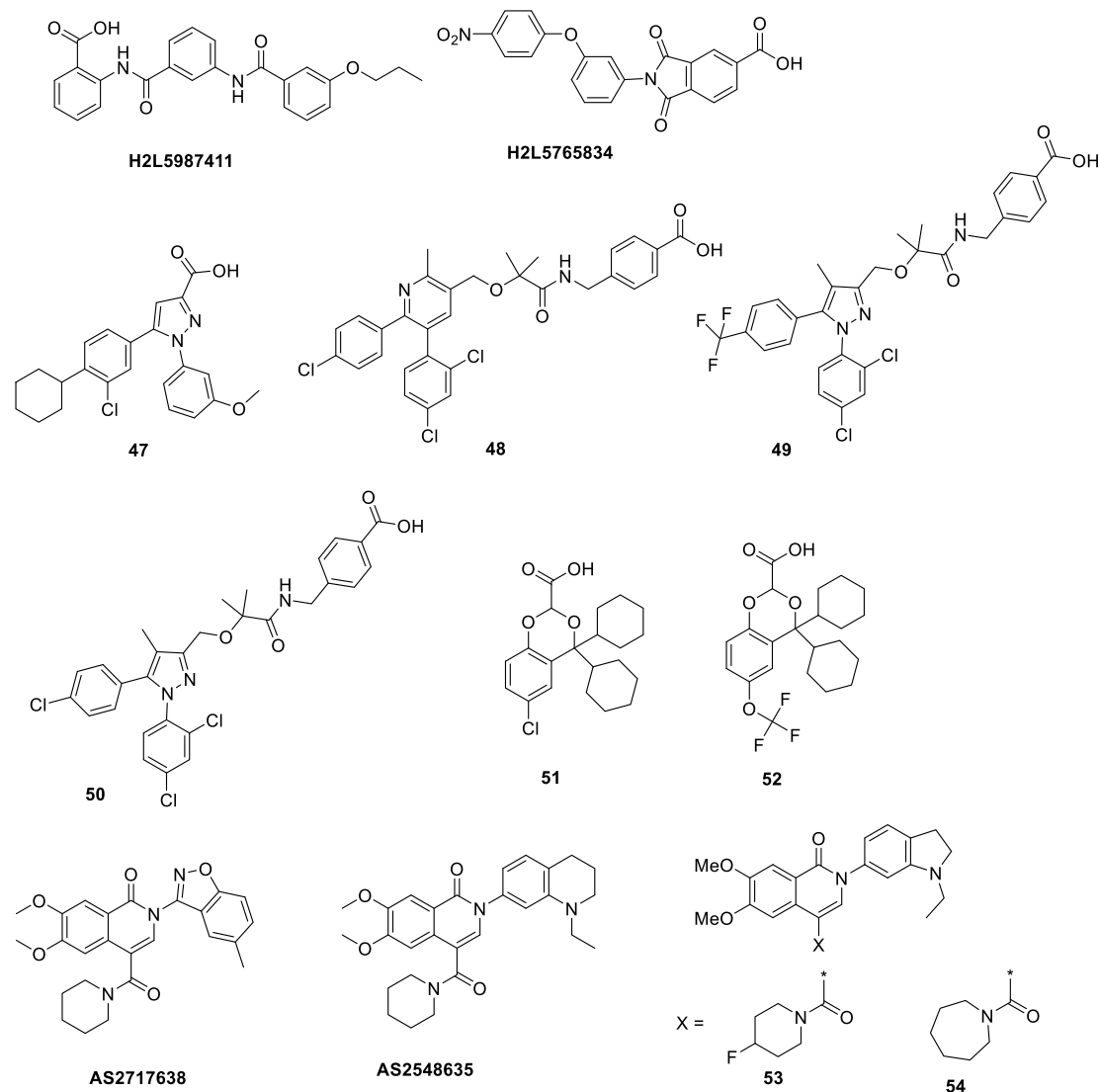
Figure 1.14. LPA$_5$ antagonists.

In 2017, researchers at Astellas Pharma Inc. examined the analgesic effects of two novel LPA$_5$ antagonists, **AS2717638** and **AS2548635** (Fig 1.14), in mouse and rat models. In CHO cells, these novel compounds were able to potently inhibit cAMP production, exhibiting respective IC$_{50}$ values of 0.038 μM and 0.042 μM. These values are about 3-90 times more potent than previously reported LPA$_5$ antagonists. Both compounds showed significantly worse activity

against LPA$_{1-3}$, having IC$_{50}$ values of ≥ 9.9 μM in LPA-induced calcium mobilization assays, meaning that based on these results, both compounds are potent LPA$_5$ antagonists that are highly selective against LPA$_{1-3}$ (LPA$_4$ and LPA$_6$ receptors were not tested) (Table 1.1). The therapeutic effects of **AS2717638** were further tested in experimental animals. It was found that at an oral dosage of 10 mg/kg, **18:1 LPA**-induced allodynia and thermal hyperalgesia were ameliorated significantly, indicating that antagonism of the LPA$_5$ receptor is a promising new method for pain therapy.[102] Furthermore, **AS2717638** was able to inhibit LPA-mediated pro-inflammatory transcription factor phosphorylation and attenuate LPA-induced cyto-/chemokine secretion by murine BV-2 microglia.[103]

In a similar study in 2018, Kawamoto *et al*. performed a high-throughput screening in order to identify novel LPA$_5$ antagonists that possessed the ability to pass through the blood-brain barrier (BBB). Other than **AS2717638** and **AS2548635**, most of the other LPA$_5$ antagonists that had been reported until that point were either phosphorothioate LPA analogues or contained a carboxylic acid moiety, neither of which demonstrate a high BBB permeability. A *2H*-isoquinolin-1-one derivative was identified to have moderate antagonist activity towards the LPA$_5$ receptor. Modifications on this hit compound were made in order to maximize its pharmacokinetic properties, resulting in the generation of several new compounds, including compound **53** (Fig 1.14). Structurally, this compound used indoline as a substituent at the 2-position and fluoropiperidinamide at the 4-position of the core scaffold. Compound **53** displayed good potency for the LPA$_5$ receptor in the cAMP accumulation assay with an IC$_{50}$ of 0.12 μM. Another compound (**54**) (Fig 1.14) that was created from the SAR study displayed a slightly improved IC$_{50}$ of 0.11 μM, but was found to have worse CNS druglikeness.[104]

7. LPA$_6$ ligands

7.1 Agonists

Some of the first ligands that showed agonist activity towards LPA$_6$ were reported by Lee *et al*. in 2009. This group reported that at a concentration of 5 μM, both **FPP**, which had recently been identified as an LPA$_5$ agonist along with its known role as an LPA$_{1-3}$ antagonist, as well as the compound geranylgeranyl diphosphate (**GGPP**) (Fig 1.15), were able to induce SRE reporter activity in LPA$_6$-transfected hBRIE 380i cells.[105] This suggests that both of these ligands are able to activate LPA signaling pathways through the LPA$_6$ receptor.

Geranylgeranyl diphosphate (GGPP)
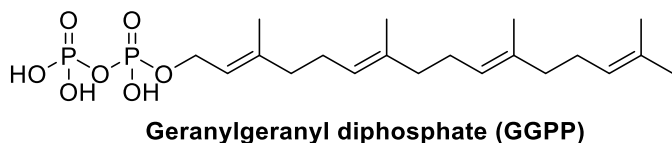
Figure 1.15. LPA$_6$ agonists.

7.2 Antagonists

To the best of existing knowledge, no LPA$_6$ antagonists have been reported in the literature thus far. However, Taniguchi *et al.* were able to determine the crystal structure of zebrafish LPA$_6$ (PDB: 5XSZ) in the absence of any ligand,[106] which can facilitate future drug development targeting the LPA$_6$ receptor. It is anticipated that novel antagonists for the LPA$_6$ receptor will be discovered in the near future, likely assisted by structure-based design using the reported crystal structure.

8. Conclusion

The LPA family of GPCRs has been implicated in the progression of numerous diseases including cancer and fibrosis. Much of the interest in targeting LPA receptors for therapy is focused on cancer, as LPA receptor signaling has been shown to be involved in every aspect of tumour development[12, 107-112] for many cancer types, including ovarian, breast, brain, and colon cancer.[113-116] In recent years, significant research efforts have been made in search of small, novel, non-lipid ligands that can target this family of receptors in order to exploit them for both diagnostic and therapeutic effects. Most of the modulators reported so far have cross-activities on more than one LPA receptors, potentially leading to complicated therapeutic outcomes and thus hindering the clinical transformation of targeting LPA receptors using the modulators. Modulators with high subtype selectivity are highly desirable due to the complicated and probably contradictory cellular activities mediated by different LPA receptors. For example, in colon cancer cells, LPA$_1$ promotes colony formation while LPA$_6$ inhibits colony formation.[117] For a proof-of-concept study to treat colon cancer, a ligand with activity for LPA$_1$ but little or no activity for LPA$_6$, or a ligand with antagonist activity for LPA$_1$ but agonist activity for LPA$_6$ will be more beneficial. Another challenge in targeting LPA receptors is their rather complex G protein coupling and subsequent downstream signaling pathways. Most of the LPA receptors couple more than two G proteins, resulting in multifaceted outcomes.[118] For example, LPA$_4$ couples four trimeric G proteins (G$_{a12/13}$, G$_{aq/11}$, G$_{ai/o}$, and G$_{as}$) (Figure 1.2). However, the

G_i/RAS/MAPK pathway is associated with tumour-promoting activities, while the G_s/AC/cAMP pathway plays an anticancer role. Potent and selective ligands targeting the EDG family of LPA receptors have been reported, including **RO6842262** for LPA_1, **27** for LPA_2, and **35** for LPA_3, among other ligands (Table 1.1). As well, many ligands that target both LPA_1 and LPA_3 have been discovered due to their high homologies. However, for the non-EDG family of LPA receptors, less progress has been made. Though ligands have been discovered that can target LPA_4, LPA_5, and LPA_6 with high potency, many of these ligands are dual- or pan-agonists or antagonists, such as phosphorothioate LPA analogues. One bright spot regarding the non-EDG LPA receptors is that several potent and selective LPA_5 antagonists have been discovered, such as those reported by Astellas Pharma Inc. in 2017 (Table 1.1). Needless to say, further research will be required in order to develop more selective ligands for LPA_{4-6}, and especially LPA_6, considering that no LPA_6 antagonists have been reported as of yet.

Although LPA receptors are widely expressed in many normal tissues, their expression levels are low. High expression of LPA receptors has been found in many types of cancers, such as ovarian,[119] breast,[120] liver,[121] colon,[122] and thyroid[123] cancer, naturally making them potential targets for the development of PET radioligands for diagnostic applications. More research will need to be performed regarding the synthesis and evaluation of LPA receptor-targeting radioligands, as only one has been reported. The discovery and development of LPA receptor-targeting ligands has proven to be a difficult task, but considering the potential clinical benefits of exploiting these receptors, such research is crucial in the advancement of the medical toolbox.

Table 1.1. LPA ligands with high subtype selectivity.

Ligands	IC_{50} (nM)						Refs
	LPA_1	LPA_2	LPA_3	LPA_4	LPA_5	LPA_6	
2 [a]	318	>5000	> 5000	-[2]	-	-	[42]
11 [a]	240	>10000	> 10000	-	-	-	[46]
27 [a]	>10000	0.005	> 10000	>10000	>10000	-	[75]
35 [a]	>30000	>30000	692	-	-	-	[88]
29 [b]	>50000	260	> 50000	-	-	-	[78]
30 [b]	>50000	17	>50000	-	-	-	[78]

38 [b]	>10000	-	150	-	-	-	92
AM095 [b]	25	>10000	>10000	8500	>10000	-	55
AS2548635 [b]	9900	>10000	> 10000	-	42	-	124
AS2717638 [b]	>10000	>10000	> 10000	-	38	-	124
H2L5186303 [b]	27354	9	1230	NE[3]	NE	-	79
NSC161613 [b]	NE	NE	24	NE	NE	-	94
ONO-7300243 [b]	160	8600	> 10000	-	-	-	64
RO6842262 [b]	25	-	> 30000	-	-	-	60

[1] [a] : agonist; [b] : antagonist;

[2] "-" data is not available for a given receptor;

[3] no effect was reported.

9. References

1. Blaho, V. A.; Hla, T., Regulation of mammalian physiology, development, and disease by the sphingosine 1-phosphate and lysophosphatidic acid receptors. *Chem Rev* **2011,** *111* (10), 6299-6320.

2. Aoki, J., Mechanisms of lysophosphatidic acid production. *Semin Cell Dev Biol* **2004,** *15* (5), 477-489.

3. Sugiura, T.; Nakane, S.; Kishimoto, S.; Waku, K.; Yoshioka, Y.; b, A. T.; Hanahan, D. J., Occurrence of lysophosphatidic acid and its alkyl ether-linked analog in rat brain and comparison of their biological activities toward cultured neural cells. *Biochimica et Biophysica Acta* **1999,** *1440*, 194-204.

4. Choi, J. W.; Herr, D. R.; Noguchi, K.; Yung, Y. C.; Lee, C. W.; Mutoh, T.; Lin, M. E.; Teo, S. T.; Park, K. E.; Mosley, A. N.; Chun, J., LPA receptors: subtypes and biological actions. *Annu Rev Pharmacol Toxicol* **2010,** *50*, 157-186.

5. Hopkins, M. M.; Zhang, Z.; Liu, Z.; Meier, K. E., Eicosopentaneoic acid and other free fatty acid receptor agonists inhibit lysophosphatidic acid- and epidermal growth factor-induced proliferation of human breast cancer cells. *J Clin Med* **2016,** *5* (2), 16.

6. Boucharaba, A.; Serre, C. M.; Gres, S.; Saulnier-Blache, J. S.; Bordet, J. C.; Guglielmi, J.; Clezardin, P.; Peyruchaud, O., Platelet-derived lysophosphatidic acid supports the progression of osteolytic bone metastases in breast cancer. *J Clin Invest* **2004,** *114* (12), 1714-1725.

7. Jeong, G. O.; Shin, S. H.; Seo, E. J.; Kwon, Y. W.; Heo, S. C.; Kim, K. H.; Yoon, M. S.; Suh, D. S.; Kim, J. H., TAZ mediates lysophosphatidic acid-induced migration and proliferation of epithelial ovarian cancer cells. *Cell Physiol Biochem* **2013,** *32* (2), 253-263.

8. Li, H. Y.; Oh, Y. S.; Choi, J. W.; Jung, J. Y.; Jun, H. S., Blocking lysophosphatidic acid receptor 1 signaling inhibits diabetic nephropathy in db/db mice. *Kidney Int* **2017,** *91* (6), 1362-1373.

9. Zhou, Z.; Subramanian, P.; Sevilmis, G.; Globke, B.; Soehnlein, O.; Karshovska, E.; Megens, R.; Heyll, K.; Chun, J.; Saulnier-Blache, J. S.; Reinholz, M.; van Zandvoort, M.; Weber, C.; Schober, A., Lipoprotein-derived lysophosphatidic acid promotes

atherosclerosis by releasing CXCL1 from the endothelium. *Cell Metab* **2011,** *13* (5), 592-600.

10. Chen, W. M.; Chiang, J. C.; Lin, Y. C.; Lin, Y. N.; Chuang, P. Y.; Chang, Y. C.; Chen, C. C.; Wu, K. Y.; Hsieh, J. C.; Chen, S. K.; Huang, W. P.; Chen, B. P. C.; Lee, H., Lysophosphatidic acid receptor LPA3 prevents oxidative stress and cellular senescence in Hutchinson-Gilford progeria syndrome. *Aging Cell* **2020,** *19* (1), e13064.

11. Ray, R.; Rai, V., Lysophosphatidic acid converts monocytes into macrophages in both mice and humans. *Blood* **2017,** *129* (9), 1177-1183.

12. Mills, G. B.; Moolenaar, W. H., The emerging role of lysophosphatidic acid in cancer. *Nat Rev Cancer* **2003,** *3* (8), 582-591.

13. Yang, L.; Kraemer, M.; Fang, X. F.; Angel, P. M.; Drake, R. R.; Morris, A. J.; Smyth, S. S., LPA receptor 4 deficiency attenuates experimental atherosclerosis. *J Lipid Res* **2019,** *60* (5), 972-980.

14. Tager, A. M.; LaCamera, P.; Shea, B. S.; Campanella, G. S.; Selman, M.; Zhao, Z.; Polosukhin, V.; Wain, J.; Karimi-Shah, B. A.; Kim, N. D.; Hart, W. K.; Pardo, A.; Blackwell, T. S.; Xu, Y.; Chun, J.; Luster, A. D., The lysophosphatidic acid receptor LPA1 links pulmonary fibrosis to lung injury by mediating fibroblast recruitment and vascular leak. *Nat Med* **2008,** *14* (1), 45-54.

15. Zhang, D.; Zhang, Y.; Zhao, C.; Zhang, W.; Shao, G.; Zhang, H., Effect of lysophosphatidic acid on the immune inflammatory response and the connexin 43 protein in myocardial infarction. *Exp Ther Med* **2016,** *11* (5), 1617-1624.

16. Mathew, D.; Torres, R. M., Lysophosphatidic acid is an inflammatory lipid exploited by cancers for immune evasion via mechanisms similar and distinct from CTLA-4 and PD-1. *Front Immunol* **2020,** *11*, 531910.

17. Xu, Y.; Shen, Z.; Wiper, D. W.; Wu, M.; Morton, R. E.; Elson, P.; Kennedy, A. W.; Belinson, J.; Markman, M.; Casey, G., Lysophosphatidic acid as a potential biomarker for ovarian and other gynecologic cancers. *JAMA* **1998,** *280*, 719-723.

18. Jiang, D.; Ju, W.; Wu, X.; Zhan, X., Elevated lysophosphatidic acid levels in the serum and cerebrospinal fluid in patients with multiple sclerosis: therapeutic response and clinical implication. *Neurol Res* **2018,** *40* (5), 335-339.

19. Bathena, S. P.; Huang, J.; Nunn, M. E.; Miyamoto, T.; Parrish, L. C.; Lang, M. S.; McVaney, T. P.; Toews, M. L.; Cerutis, D. R.; Alnouti, Y., Quantitative determination of lysophosphatidic acids (LPAs) in human saliva and gingival crevicular fluid (GCF) by LC-MS/MS. *J Pharm Biomed Anal* **2011,** *56* (2), 402-407.

20. Bai, C. Q.; Yao, Y. W.; Liu, C. H.; Zhang, H.; Xu, X. B.; Zeng, J. L.; Liang, W. J.; Yang, W.; Song, Y., Diagnostic and prognostic significance of lysophosphatidic acid in malignant pleural effusions. *J Thorac Dis* **2014,** *6* (5), 483-490.

21. Aoki, J.; Inoue, A.; Okudaira, S., Two pathways for lysophosphatidic acid production. *Biochim Biophys Acta* **2008,** *1781* (9), 513-518.

22. Tanaka, M.; Okudaira, S.; Kishi, Y.; Ohkawa, R.; Iseki, S.; Ota, M.; Noji, S.; Yatomi, Y.; Aoki, J.; Arai, H., Autotaxin stabilizes blood vessels and is required for embryonic vasculature by producing lysophosphatidic acid. *J Biol Chem* **2006,** *281* (35), 25822-25830.

23. van Meeteren, L. A.; Ruurs, P.; Stortelers, C.; Bouwman, P.; van Rooijen, M. A.; Pradere, J. P.; Pettit, T. R.; Wakelam, M. J.; Saulnier-Blache, J. S.; Mummery, C. L.; Moolenaar, W. H.; Jonkers, J., Autotaxin, a secreted lysophospholipase D, is essential for blood vessel formation during development. *Mol Cell Biol* **2006,** *26* (13), 5015-5022.

24. Yamashita, A.; Hayashi, Y.; Nemoto-Sasaki, Y.; Ito, M.; Oka, S.; Tanikawa, T.; Waku, K.; Sugiura, T., Acyltransferases and transacylases that determine the fatty acid composition of glycerolipids and the metabolism of bioactive lipid mediators in mammalian cells and model organisms. *Prog Lipid Res* **2014,** *53*, 18-81.

25. Zhai, G.; Pelletier, J.-P.; Liu, M.; Aitken, D.; Randell, E.; Rahman, P.; Jones, G.; Martel-Pelletier, J., Activation of the phosphatidylcholine to lysophosphatidylcholine pathway is associated with osteoarthritis knee cartilage volume loss over time. *Sci Rep* **2019,** *9* (1), 9648.

26. Albers, H. M. H. G.; Dong, A.; Meeteren, L. A. v.; Egan, D. A.; Sunkara, M.; Tilburg, E. W. v.; Schuurman, K.; Tellingen, O. v.; Morris, A. J.; Smyth, S. S.; Moolenaar, W. H.; Ovaa, H., Boronic acid-based inhibitor of autotaxin reveals rapid turnover of LPA in the circulation. *PNAS* **2010,** *107* (16), 7257-7262.

27. Salous, A. K.; Panchatcharam, M.; Sunkara, M.; Mueller, P.; Dong, A.; Wang, Y.; Graf, G. A.; Smyth, S. S.; Morris, A. J., Mechanism of rapid elimination of lysophosphatidic acid and related lipids from the circulation of mice. *J Lipid Res* **2013,** *54* (10), 2775-2784.

28. Yung, Y. C.; Stoddard, N. C.; Chun, J., LPA receptor signaling: pharmacology, physiology, and pathophysiology. *J Lipid Res* **2014,** *55* (7), 1192-1214.

29. Tomsig, J. L.; Snyder, A. H.; Berdyshev, E. V.; Skobeleva, A.; Mataya, C.; Natarajan, V.; Brindley, D. N.; Lynch, K. R., Lipid phosphate phosphohydrolase type 1 (LPP1) degrades extracellular lysophosphatidic acid in vivo. *Biochem J* **2009,** *419* (3), 611-618.

30. Geraldo, L. H. M.; Spohr, T. C. L. d. S.; Amaral, R. F. d.; Fonseca, A. C. C. d.; Garcia, C.; Mendes, F. d. A.; Freitas, C.; dosSantos, M. F.; Lima, F. R. S., Role of lysophosphatidic acid and its receptors in health and disease: novel therapeutic strategies. *Signal Transduct Target Ther* **2021,** *6* (1), 45.

31. Jeong, K. J.; Park, S. Y.; Cho, K. H.; Sohn, J. S.; Lee, J.; Kim, Y. K.; Kang, J.; Park, C. G.; Han, J. W.; Lee, H. Y., The Rho/ROCK pathway for lysophosphatidic acid-induced proteolytic enzyme expression and ovarian cancer cell invasion. *Oncogene* **2012,** *31*, 4279–4289.

32. Harrington, B. S.; Annunziata, C. M., NF-κB signaling in ovarian cancer. *Cancers* **2019,** *11* (8), 1182.

33. Xiang, H.; Lu, Y.; Shao, M.; Wu, T., Lysophosphatidic acid receptors: biochemical and clinical implications in different diseases. *J Cancer* **2020,** *11* (12), 3519-3535.

34. Bassilana, F.; Nash, M.; Ludwig, M. G., Adhesion G protein-coupled receptors: opportunities for drug discovery. *Nat Rev Drug Discov* **2019,** *18* (11), 869-884.

35. Lynch, K. R.; Macdonald, T. L., Structure–activity relationships of lysophosphatidic acid analogs. *Biochimica et Biophysica Acta* **2002,** *1582*, 289–294.

36. Hopper, D. W.; Ragan, S. P.; Hooks, S. B.; Lynch, K. R.; Macdonald, T. L., Structure-activity relationships of lysophosphatidic acid: conformationally restricted backbone mimetics. *J Med Chem* **1999,** *42*, 963-970.

37. Llona-Minguez, S.; Ghassemian, A.; Helleday, T., Lysophosphatidic acid receptor (LPAR) modulators: the current pharmacological toolbox. *Prog Lipid Res* **2015,** *58*, 51-75.

38.   Parr, T. A.; Roppe, J. R.; Seiders, T. J. 3- or 5-bi phenyl-4-ylisoxazole-based compounds useful for the treatment of fibrosis, pain, cancer and respiratory, allergic, nervous system or cardiovascular disorders. WO2012138797A1. 2012.

39.   Buckman, B. O.; Nicholas, J. B.; Emayan, K.; Seiwert, S. D. Lysophosphatidic acid receptor antagonists. WO2013025733A1. 2013.

40.   Buckman, B. O.; Nicholas, J. B.; Emayan, K.; Seiwert, S. D.; Yuan, S. Lysophosphatidic acid receptor antagonists. WO2014113485A1. 2014.

41.   Sugiura, T.; Tokumura, A.; Gregory, L.; Nouchi, T.; Weintraub, S. T.; Hanahan, D. J., Biochemical characterization of the interaction of lipid phosphoric acids with human platelets: comparison with platelet activating factor. *Arch Biochem Biophys* **1994,** *311*, 358-368.

42.   Santos, W. L.; Heasley, B. H.; Jarosz, R.; Carter, K. M.; Lynchb, K. R.; Macdonald, T. L., Synthesis and biological evaluation of phosphonic and thiophosphoric acid derivatives of lysophosphatidic acid. *Bioorganic Med Chem Lett* **2004,** *14*, 3473–3476.

43.   Xu, Y.; Tanaka, M.; Arai, H.; Aoki, J.; Prestwich, G. D., Alkyl lysophosphatidic acid and fluoromethylene phosphonate analogs as metabolically-stabilized agonists for LPA receptors. *Bioorg Med Chem Lett* **2004,** *14* (21), 5323-5328.

44.   Heise, C. E.; Santos, W. L.; Schreihofer, A. M.; Heasley, B. H.; Mukhin, Y. V.; Macdonald, T. L.; Lynch, K. R., Activity of 2-substituted lysophosphatidic acid (LPA) analogs at LPA receptors: discovery of a LPA1/LPA3 receptor antagonist. *Mol Pharmacol* **2001,** *60*, 1173–1180.

45.   Gajewiak, J.; Tsukahara, R.; Fujiwara, Y.; Tigyi, G.; Prestwich, G. D., Synthesis, pharmacology, and cell biology of sn-2-aminooxy analogues of lysophosphatidic acid. *Org Lett* **2008,** *10*, 1111-1114.

46.   González-Gil, I.; Zian, D.; Vázquez-Villa, H.; Hernández-Torres, G.; Martínez, R. F.; Khiar-Fernández, N.; Rivera, R.; Kihara, Y.; Devesa, I.; Mathivanan, S.; Valle, C. R. d.; Zambrana-Infantes, E.; Puigdomenech, M.; Cincilla, G.; Sanchez-Martinez, M.; Fonseca, F. R. d.; Ferrer-Montiel, A. V.; Chun, J.; RubénLópez-Vales; Ortega-Gutiérrez, S., A novel agonist of the type 1 lysophosphatidic acid receptor (LPA1), UCM-05194, shows efficacy in neuropathic pain amelioration. *J Med Chem* **2020,** *63* (5), 2372-2390.

47. Guillot, E.; Bail, J. C. L.; Paul, P.; Fourgous, V.; Briand, P.; Partiseti, M.; Cornet, B.; Janiak, P.; Philippo, C., Lysophosphatidic acid receptor agonism: discovery of potent nonlipid benzofuran ethanolamine structures. *J Pharmacol Exp Ther* **2020,** *374*, 283–294.

48. Heasley, B. H.; Jarosz, R.; Lynch, K. R.; Macdonald, T. L., Initial structure-activity relationships of lysophosphatidic acid receptor antagonists: discovery of a high-affinity LPA1/LPA3 receptor antagonist. *Bioorg Med Chem Lett* **2004,** *14* (11), 2735-2740.

49. Xu, Y.; Jiang, G.; Tsukahara, R.; Fujiwara, Y.; Tigyi, G.; Prestwich, G. D., Phosphonothioate and fluoromethylene phosphonate analogues of cyclic phosphatidic acid: novel antagonists of lysophosphatidic acid receptors. *J Med Chem* **2006,** *49*, 5309-5315.

50. Ohta, H.; Sato, K.; Murata, N.; Damirin, A.; Malchinkhuu, E.; Kon, J.; Kimura, T.; Tobo, M.; Yamazaki, Y.; Watanabe, T.; Yagi, M.; Sato, M.; Suzuki, R.; Murooka, H.; Sakai, T.; Nishitoba, T.; Im, D.-S.; Nochi, H.; Tamoto, K.; Tomura, H.; Okajima, F., Ki16425, a subtype-selective antagonist for EDG-family lysophosphatidic acid receptors. *Mol Pharmacol* **2003,** *64*, 994–1005.

51. Yamamoto, T.; Fujita, K.; Asari, S.; Chiba, A.; Kataba, Y.; Ohsumi, K.; Ohmuta, N.; Iida, Y.; Ijichi, C.; Iwayama, S.; Fukuchi, N.; Shoji, M., Synthesis and evaluation of isoxazole derivatives as lysophosphatidic acid (LPA) antagonists. *Bioorg Med Chem Lett* **2007,** *17*, 3736–3740.

52. Sato, T.; Sugimoto, K.; Inoue, A.; Okudaira, S.; Aoki, J.; Tokuyama, H., Synthesis and biological evaluation of optically active Ki16425. *Bioorg. Med Chem Lett* **2012,** *22*, 4323–4326.

53. Komachi, M.; Sato, K.; Tobo, M.; Mogi, C.; Yamada, T.; Ohta, H.; Tomura, H.; Kimura, T.; Im, D.-S.; Yanagida, K.; Ishii, S.; Takeyoshi, I.; Okajima, F., Orally active lysophosphatidic acid receptor antagonist attenuates pancreatic cancer invasion and metastasis in vivo. *Cancer Sci* **2012,** *103* (6), 1099-1104.

54. Swaney, J. S.; Chapman, C.; Correa, L. D.; Stebbins, K. J.; Bundey, R. A.; Prodanovich, P. C.; Fagan, P.; Baccei, C. S.; Santini, A. M.; Hutchinson, J. H.; Seiders, T. J.; Parr, T. A.; Prasit, P.; Evans, J. F.; Lorrain, D. S., A novel, orally active LPA(1) receptor antagonist inhibits lung fibrosis in the mouse bleomycin model. *Br J Pharmacol* **2010,** *160* (7), 1699-1713.

55. Castelino, F. V.; Seiders, J.; Bain, G.; Brooks, S. F.; King, C. D.; Swaney, J. S.; Lorrain, D. S.; Chun, J.; Luster, A. D.; Tager, A. M., Amelioration of dermal fibrosis by genetic deletion or pharmacologic antagonism of lysophosphatidic acid receptor 1 in a mouse model of scleroderma. *Arthritis Rheum* **2011**, *63* (5), 1405-1415.

56. Swaney, J. S.; Chapman, C.; Correa, L. D.; Stebbins, K. J.; Broadhead, A. R.; Bain, G.; Santini, A. M.; Darlington, J.; King, C. D.; Baccei, C. S.; Lee, C.; Parr, T. A.; Roppe, J. R.; Seiders, T. J.; Ziff, J.; Prasit, P.; Hutchinson, J. H.; Evans, J. F.; Lorrain, D. S., Pharmacokinetic and pharmacodynamic characterization of an oral lysophosphatidic acid type 1 receptor-selective antagonist. *J Pharmacol Exp Ther* **2011**, *336* (3), 693-700.

57. Lee, J. H.; Sarker, M. K.; Choi, H.; Shin, D.; Kim, D.; Jun, H.-S., Lysophosphatidic acid receptor 1 inhibitor, AM095, attenuates diabetic nephropathy in mice by downregulation of TLR4/NF-κB signaling and NADPH oxidase. *Biochim Biophys Acta Mol Basis Dis* **2019**, *1865* (6), 1332-1340.

58. Gaire, B. P.; Sapkota, A.; Song, M. R.; Choi, J. W., Lysophosphatidic acid receptor 1 (LPA1) plays critical roles in microglial activation and brain damage after transient focal cerebral ischemia. *J Neuroinflammation* **2019**, *16* (1), 170.

59. Gabriel, S. D.; Hamilton, M. M.; Qian, Y.; Sidduri, A. N-aryltriazole compounds as lpar antagonists. WO2013189865A1. 2013.

60. Qian, Y.; Hamilton, M.; Sidduri, A.; Gabriel, S.; Ren, Y.; Peng, R.; Kondru, R.; Narayanan, A.; Truitt, T.; Hamid, R.; Chen, Y.; Zhang, L.; Fretland, A. J.; Sanchez, R. A.; Chang, K. C.; Lucas, M.; Schoenfeld, R. C.; Laine, D.; Fuentes, M. E.; Stevenson, C. S.; Budd, D. C., Discovery of highly selective and orally active lysophosphatidic acid receptor-1 antagonists with potent activity on human lung fibroblasts. *J Med Chem* **2012,** *55* (17), 7920-7939.

61. Sidduri, A.; Budd, D. C.; Fuentes, M. E.; Lambros, T.; Ren, Y.; Roongta, V.; Schoenfeld, R. C.; Gillespie, P.; Stevenson, C. S.; Truitt, T.; Qian, Y., Discovery of novel non-carboxylic acid 5-amino-4-cyanopyrazole derivatives as potent and highly selective LPA1R antagonists. *Bioorg Med Chem Lett* **2014,** *24* (18), 4450-4454.

62. Smith, J. S.; Lefkowitz, R. J.; Rajagopal, S., Biased signalling: from simple switches to allosteric microprocessors. *Nat Rev Drug Discov* **2018**, *17* (4), 243-260.

63. Shimizu, Y.; Nakayama, M., Discovery of novel Gq-biased LPA1 negative allosteric modulators. *SLAS Discov* **2017,** *22* (7), 859-866.

64. Terakado, M.; Suzuki, H.; Hashimura, K.; Tanaka, M.; Ueda, H.; Kohno, H.; Fujimoto, T.; Saga, H.; Nakade, S.; Habashita, H.; Takaoka, Y.; Seko, T., Discovery of ONO-7300243 from a novel class of lysophosphatidic acid receptor 1 antagonists: from hit to lead. *ACS Med Chem Lett* **2016,** *7* (10), 913-918.

65. Zhao, P. F.; Wu, S.; Li, Y.; Bao, G.; Pei, J. Y.; Wang, Y. W.; Ma, Q.; Sun, H. J.; Damirin, A., LPA receptor1 antagonists as anticancer agents suppress human lung tumours. *Eur J Pharmacol* **2020,** *868*, 172886.

66. Terakado, M.; Suzuki, H.; Hashimura, K.; Tanaka, M.; Ueda, H.; Hirai, K.; Asada, M.; Ikura, M.; Matsunaga, N.; Saga, H.; Shinozaki, K.; Karakawa, N.; Takada, Y.; Minami, M.; Egashira, H.; Sugiura, Y.; Yamada, M.; Nakade, S.; Takaoka, Y., Discovery of a slow tight binding LPA1 antagonist (ONO-0300302) for the treatment of benign prostatic hyperplasia. *ACS Med Chem Lett* **2017,** *8* (12), 1281-1286.

67. Sakamoto, K.; Noguchi, Y.; Ueshima, K.; Yamakuni, H.; Ohtake, A.; Sato, S.; Ishizu, K.; Hosogai, N.; Kawaminami, E.; Takeda, M.; Masuda, N., Effect of ASP6432, a novel type 1 lysophosphatidic acid receptor antagonist, on urethral function and prostate cell proliferation. *J Pharmacol Exp Ther* **2018,** *366* (2), 390-396.

68. Sakamoto, K.; Noguchi, Y.; Imazumi, K.; Ueshima, K.; Ohtake, A.; Takeda, M.; Masuda, N., ASP6432, a type 1 lysophosphatidic acid receptor antagonist, reduces urethral function during urine voiding and improves voiding dysfunction. *Eur J Pharmacol* **2019,** *847*, 83-90.

69. Chrencik, J. E.; Roth, C. B.; Terakado, M.; Kurata, H.; Omi, R.; Kihara, Y.; Warshaviak, D.; Nakade, S.; Asmar-Rovira, G.; Mileni, M.; Mizuno, H.; Griffith, M. T.; Rodgers, C.; Han, G. W.; Velasquez, J.; Chun, J.; Stevens, R. C.; Hanson, M. A., Crystal structure of antagonist bound human lysophosphatidic acid receptor 1. *Cell* **2015,** *161* (7), 1633-1643.

70. Virag, T.; Elrod, D. B.; Liliom, K.; Sardar, V. M.; Parrill, A. L.; Yokoyama, K.; Durgam, G.; Deng, W.; Miller, D. D.; Tigyi, G., Fatty alcohol phosphates are subtype-selective agonists and antagonists of lysophosphatidic acid receptors. *Mol Pharmacol* **2003,** *63*, 1032–1042.

71.   Durgam, G. G.; Virag, T.; Walker, M. D.; Tsukahara, R.; Yasuda, S.; Liliom, K.; Meeteren, L. A. v.; Moolenaar, W. H.; Wilke, N.; Siess, W.; Tigyi, G.; Miller, D. D., Synthesis, structure-activity relationships, and biological evaluation of fatty alcohol phosphates as lysophosphatidic acid receptor ligands, activators of PPARgamma, and inhibitors of autotaxin. *J Med Chem* **2005,** *48* (15), 4919-4930.

72.   Gajewiak, J.; Tsukahara, R.; Tsukahara, T.; Fujiwara, Y.; Yu, S.; Lu, Y.; Murph, M.; Mills, G. B.; Tigyi, G.; Prestwich, G. D., Alkoxymethylenephosphonate analogues of (Lyso) phosphatidic acid stimulate signaling networks coupled to the LPA2 receptor. *ChemMedChem* **2007,** *2* (12), 1789-1798.

73.   Kiss, G. N.; Fells, J. I.; Gupte, R.; Lee, S. C.; Liu, J.; Nusser, N.; Lim, K. G.; Ray, R. M.; Lin, F. T.; Parrill, A. L.; Sümegi, B.; Miller, D. D.; Tigyi, G., Virtual screening for LPA2-specific agonists identifies a nonlipid compound with antiapoptotic actions. *Mol Pharmacol* **2012,** *82* (6), 1162-1173.

74.   Schmitz, K.; Brunkhorst, R.; de Bruin, N.; Mayer, C. A.; Haussler, A.; Ferreiros, N.; Schiffmann, S.; Parnham, M. J.; Tunaru, S.; Chun, J.; Offermanns, S.; Foerch, C.; Scholich, K.; Vogt, J.; Wicker, S.; Lotsch, J.; Geisslinger, G.; Tegeder, I., Dysregulation of lysophosphatidic acids in multiple sclerosis and autoimmune encephalomyelitis. *Acta Neuropathol Commun* **2017,** *5* (1), 42.

75.   Patil, R.; Fells, J. I.; Szabo, E.; Lim, K. G.; Norman, D. D.; Balogh, A.; Patil, S.; Strobos, J.; Miller, D. D.; Tigyi, G. J., Design and synthesis of sulfamoyl benzoic acid analogues with subnanomolar agonist activity specific to the LPA2 receptor. *J Med Chem* **2014,** *57* (16), 7136-7140.

76.   Knowlden, S. A.; Hillman, S. E.; Chapman, T. J.; Patil, R.; Miller, D. D.; Tigyi, G.; Georas, S. N., Novel inhibitory effect of a lysophosphatidic acid 2 agonist on allergen-driven airway inflammation. *Am J Respir Cell Mol Biol* **2016,** *54* (3), 402-409.

77.   Patil, R.; Szabo, E.; Fells, J. I.; Balogh, A.; Lim, K. G.; Fujiwara, Y.; Norman, D. D.; Lee, S. C.; Balazs, L.; Thomas, F.; Patil, S.; Emmons-Thompson, K.; Boler, A.; Strobos, J.; McCool, S. W.; Yates, C. R.; Stabenow, J.; Byrne, G. I.; Miller, D. D.; Tigyi, G. J., Combined mitigation of the gastrointestinal and hematopoietic acute radiation syndromes by an LPA2 receptor-specific nonlipid agonist. *Chem Biol* **2015,** *22* (2), 206-216.

78. Beck, H. P.; Kohn, T.; Rubenstein, S.; Hedberg, C.; Schwandner, R.; Hasslinger, K.; Dai, K.; Li, C.; Liang, L.; Wesche, H.; Frank, B.; An, S.; Wickramasinghe, D.; Jaen, J.; Medina, J.; Hungate, R.; Shen, W., Discovery of potent LPA2 (EDG4) antagonists as potential anticancer agents. *Bioorg Med Chem Lett* **2008,** *18* (3), 1037-1041.

79. Fells, J. I.; Tsukahara, R.; Liu, J.; Tigyi, G.; Parrill, A. L., Structure-based drug design identifies novel LPA3 antagonists. *Bioorg Med Chem* **2009,** *17* (21), 7457-7464.

80. Kondo, M.; Tezuka, T.; Ogawa, H.; Koyama, K.; Bando, H.; Azuma, M.; Nishioka, Y., Lysophosphatidic acid regulates the differentiation of Th2 cells and its antagonist suppresses allergic airway inflammation. *Int Arch Allergy Immunol* **2021,** *182* (1), 1-13.

81. Jalink, K.; Hengeveld, T.; Mulder, S.; Postma, F. R.; Simon, M. F.; Chap, H.; Marel, G. A. v. d.; Boom, J. H. v.; Blitterswijk, W. J. v.; Moolenaar, W. H., Lysophosphatidic acid-induced Ca2+ mobilization in human A431 cells: structure-activity analysis. *Biochem J* **1995,** *307*, 609-616.

82. Hasegawa, Y.; Erickson, J. R.; Goddard, G. J.; Yu, S.; Liu, S.; Cheng, K. W.; Eder, A.; Bandoh, K.; Aoki, J.; Jarosz, R.; Schrier, A. D.; Lynch, K. R.; Mills, G. B.; Fang, X., Identification of a phosphothionate analogue of lysophosphatidic acid (LPA) as a selective agonist of the LPA3 receptor. *J Biol Chem* **2003,** *278* (14), 11962-11969.

83. Qian, L.; Xu, Y.; Hasegawa, Y.; Aoki, J.; Mills, G. B.; Prestwich, G. D., Enantioselective responses to a phosphorothioate analogue of lysophosphatidic acid with LPA3 receptor-selective agonist activity. *J Med Chem* **2003,** *46*, 5575-5578.

84. Qian, L.; Xu, Y.; Simper, T.; Jiang, G.; Aoki, J.; Umezu-Goto, M.; Arai, H.; Yu, S.; Mills, G. B.; Tsukahara, R.; Makarova, N.; Fujiwara, Y.; Tigyi, G.; Prestwich, G. D., Phosphorothioate analogues of alkyl lysophosphatidic acid as LPA3 receptor-selective agonists. *ChemMedChem* **2006,** *1* (3), 376-383.

85. Jiang, G.; Inoue, A.; Aoki, J.; Prestwich, G. D., Phosphorothioate analogs of sn-2 radyl lysophosphatidic acid (LPA): metabolically stabilized LPA receptor agonists. *Bioorg Med Chem Lett* **2013,** *23* (6), 1865-1869.

86. Xu, Y.; Aoki, J.; Shimizu, K.; Umezu-Goto, M.; Hama, K.; Takanezawa, Y.; Yu, S.; Mills, G. B.; Arai, H.; Qian, L.; Prestwich, G. D., Structure-activity relationships of fluorinated lysophosphatidic acid analogues. *J Med Chem* **2005,** *48* (9), 3319-3327.

87.   Xu, Y.; Qian, L.; Prestwich, G. D., Synthesis of monofluorinated analogues of lysophosphatidic acid. *J Org Chem* **2003,** *68* (13), 5320-5330.

88.   Gududuru, V.; Zeng, K.; Tsukahara, R.; Makarova, N.; Fujiwara, Y.; Pigg, K. R.; Baker, D. L.; Tigyi, G.; Miller, D. D., Identification of Darmstoff analogs as selective agonists and antagonists of lysophosphatidic acid receptors. *Bioorg Med Chem Lett* **2006,** *16* (2), 451-456.

89.   Fischer, D. J.; Nusser, N.; Virag, T.; Yokoyama, K.; Da, W.; Baker, D. L.; Bautista, D.; Parrill, A. L.; Tigyi, G., Short-chain phosphatidates are subtype-selective antagonists of lysophosphatidic acid receptors. *Mol Pharmacol* **2001,** *60* (4), 776-784.

90.   Rother, E.; Brandl, R.; Baker, D. L.; Goyal, P.; Gebhard, H.; Tigyi, G.; Siess, W., Subtype-selective antagonists of lysophosphatidic Acid receptors inhibit platelet activation triggered by the lipid core of atherosclerotic plaques. *Circulation* **2003,** *108* (6), 741-747.

91.   Durgam, G. G.; Tsukahara, R.; Makarova, N.; Walker, M. D.; Fujiwara, Y.; Pigg, K. R.; Baker, D. L.; Sardar, V. M.; Parrill, A. L.; Tigyi, G.; Miller, D. D., Synthesis and pharmacological evaluation of second-generation phosphatidic acid derivatives as lysophosphatidic acid receptor ligands. *Bioorg Med Chem Lett* **2006,** *16* (3), 633-640.

92.   Heasley, B. H.; Jarosz, R.; Carte, K. M.; Van, S. J.; Lynch, K. R.; Macdonald, T. L., A novel series of 2-pyridyl-containing compounds as lysophosphatidic acid receptor antagonists: development of a nonhydrolyzable LPA3 receptor-selective antagonist. *Bioorg Med Chem Lett* **2004,** *14* (15), 4069-4074.

93.   East, J. E.; Carter, K. M.; Kennedy, P. C.; Schulte, N. A.; Toews, M. L.; Lynch, K. R.; Macdonald, T. L., Development of a phosphatase-resistant, L-tyrosine derived LPA1/LPA3 dual antagonist. *Medchemcomm* **2011,** *2* (4), 325-330.

94.   Fells, J. I.; Tsukahara, R.; Fujiwara, Y.; Liu, J.; Perygin, D. H.; Osborne, D. A.; Tigyi, G.; Parrill, A. L., Identification of non-lipid LPA3 antagonists by virtual screening. *Bioorg Med Chem* **2008,** *16* (11), 6207–6217.

95.   Jiang, G.; Xu, Y.; Fujiwara, Y.; Tsukahara, T.; Tsukahara, R.; Gajewiak, J.; Tigyi, G.; Prestwich, G. D., Alpha-substituted phosphonate analogues of lysophosphatidic acid (LPA) selectively inhibit production and action of LPA. *ChemMedChem* **2007,** *2* (5), 679-690.

96.   Williams, J. R.; Khandoga, A. L.; Goyal, P.; Fells, J. I.; Perygin, D. H.; Siess, W.; Parrill, A. L.; Tigyi, G.; Fujiwara, Y., Unique ligand selectivity of the GPR92/LPA5 lysophosphatidate receptor indicates role in human platelet activation. *J Biol Chem* **2009,** *284* (25), 17304-17319.

97.   Oh, D. Y.; Yoon, J. M.; Moon, M. J.; Hwang, J. I.; Choe, H.; Lee, J. Y.; Kim, J. I.; Kim, S.; Rhim, H.; O'Dell, D. K.; Walker, J. M.; Na, H. S.; Lee, M. G.; Kwon, H. B.; Kim, K.; Seong, J. Y., Identification of farnesyl pyrophosphate and N-arachidonylglycine as endogenous ligands for GPR92. *J Biol Chem* **2008,** *283* (30), 21054-21064.

98.   Kozian, D. H.; Evers, A.; Florian, P.; Wonerow, P.; Joho, S.; Nazaré, M., Selective non-lipid modulator of LPA5 activity in human platelets. *Bioorg Med Chem Lett* **2012,** *22* (16), 5239-5243.

99.   Nazaré, M.; Kozian, D.; Evers, A.; Czechtizky, W. Pyridine derivatives and their use in the treatment of conditions associated with pathological thrombus formation. WO2013171316A1. 2013.

100.   Nazaré, M.; Kozian, D.; Evers, A.; Czechtizky, W. Pyrazole derivatives and their use as LPAR5 antagonists. WO2013171317A1. 2013.

101.   Nazaré, M.; Kozian, D.; Bossar, M.; Czechtizky, t.; Evers, A. Benzo[1,3]dioxine derivatives and their use as lpar5 antagonists. WO2013171318A1. 2013.

102.   Murai, N.; Hiyama, H.; Kiso, T.; Sekizawa, T.; Watabiki, T.; Oka, H.; Aoki, T., Analgesic effects of novel lysophosphatidic acid receptor 5 antagonist AS2717638 in rodents. *Neuropharmacology* **2017,** *126*, 97-107.

103.   Plastira, I.; Joshi, L.; Bernhart, E.; Schoene, J.; Specker, E.; Nazare, M.; Sattler, W., Small-molecule lysophosphatidic acid receptor 5 (LPAR5) antagonists: versatile pharmacological tools to regulate inflammatory signaling in BV-2 microglia cells. *Front Cell Neurosci* **2019,** *13*, 531.

104.   Kawamoto, Y.; Seo, R.; Murai, N.; Hiyama, H.; Oka, H., Identification of potent lysophosphatidic acid receptor 5 (LPA5) antagonists as potential analgesic agents. *Bioorg Med Chem* **2018,** *26* (1), 257-265.

105.   Lee, M.; Choi, S.; Hallden, G.; Yo, S. J.; Schichnes, D.; Aponte, G. W., P2Y5 is a G(alpha)i, G(alpha)12/13 G protein-coupled receptor activated by lysophosphatidic acid

that reduces intestinal cell adhesion. *Am J Physiol Gastrointest Liver Physiol* **2009,** *297* (4), G641-654.

106.   Taniguchi, R.; Inoue, A.; Sayama, M.; Uwamizu, A.; Yamashita, K.; Hirata, K.; Yoshida, M.; Tanaka, Y.; Kato, H. E.; Nakada-Nakura, Y.; Otani, Y.; Nishizawa, T.; Doi, T.; Ohwada, T.; Ishitani, R.; Aoki, J.; Nureki, O., Structural insights into ligand recognition by the lysophosphatidic acid receptor LPA6. *Nature* **2017,** *548* (7667), 356-360.

107.   Hanahan, D.; Weinberg, R. A., The hallmarks of cancer. *Cell* **2000,** *100* (1), 57-70.

108.   Hanahan, D.; Weinberg, R. A., Hallmarks of cancer: the next generation. *Cell 144* (5), 646-674.

109.   Sengupta, S.; Wang, Z.; Tipps, R.; Xu, Y., Biology of LPA in health and disease. *Semin Cell Dev Biol* **2004,** *15* (5), 503-512.

110.   Leblanc, R.; Peyruchaud, O., New insights into the autotaxin/LPA axis in cancer development and metastasis. *Exp Cell Res* **2015,** *333* (2), 183-189.

111.   Jesionowska, A.; Cecerska-Heryć, E.; Marczuk, N.; Safranow, K.; Dołęgowska, B., Lysophosphatidic acid and malignant neoplasms. *Postepy Biochem* **2015,** *61* (4), 381-387.

112.   Xu, Y., Lysophospholipid signaling in the epithelial ovarian cancer tumor microenvironment. *Cancers (Basel)* **2018,** *10* (7), 227.

113.   Hama, K.; Aoki, J.; Fukaya, M.; Kishi, Y.; Sakai, T.; Suzuki, R.; Ohta, H.; Yamori, T.; Watanabe, M.; Chun, J.; Arai, H., Lysophosphatidic acid and autotaxin stimulate cell motility of neoplastic and non-neoplastic cells through LPA1. *J Biol Chem* **2004,** *279* (17), 17634-17639.

114.   Mills, G. B.; Eder, A.; Fang, X.; Hasegawa, Y.; Mao, M.; Lu, Y.; Tanyi, J.; Tabassam, F. H.; Wiener, J.; Lapushin, R.; Yu, S.; Parrott, J. A.; Compton, T.; Tribley, W.; Fishman, D.; Stack, M. S.; Gaudette, D.; Jaffe, R.; Furui, T.; Aoki, J.; Erickson, J. R., Critical role of lysophospholipids in the pathophysiology, diagnosis, and management of ovarian cancer. *Cancer Treat Res* **2002,** *107*, 259-283.

115.   Hayashi, M.; Okabe, K.; Kato, K.; Okumura, M.; Fukui, R.; Fukushima, N.; Tsujiuchi, T., Differential function of lysophosphatidic acid receptors in cell proliferation and migration of neuroblastoma cells. *Cancer Lett* **2012,** *316* (1), 91-96.

116. Hoelzinger, D. B.; Nakada, M.; Demuth, T.; Rosensteel, T.; Reavie, L. B.; Berens, M. E., Autotaxin: a secreted autocrine/paracrine factor that promotes glioma invasion. *J Neurooncol* **2008,** *86* (3), 297-309.

117. Takahashi, K.; Fukushima, K.; Otagaki, S.; Ishimoto, K.; Minami, K.; Fukushima, N.; Honoki, K.; Tsujiuchi, T., Effects of LPA1 and LPA6 on the regulation of colony formation activity in colon cancer cells treated with anticancer drugs. *J Recept Signal Transduct Res* **2018,** *38* (1), 71-75.

118. Xu, Y., Targeting lysophosphatidic acid in cancer: the issues in moving from bench to bedside. *Cancers (Basel)* **2019,** *11* (10), 1523.

119. Si, J.; Su, Y.; Wang, Y.; Yan, Y.; Tang, Y., Expressions of lysophosphatidic acid receptors in the development of human ovarian carcinoma. *Int J Clin Exp Med* **2015,** *8* (10), 17880-17890.

120. Chen, M.; Towers, L. N.; O'Connor, K. L., LPA2 (EDG4) mediates Rho-dependent chemotaxis with lower efficacy than LPA1 (EDG2) in breast carcinoma cells. *Am J Physiol Cell Physiol* **2007,** *292* (5), C1927-1933.

121. Zuckerman, V.; Sokolov, E.; Swet, J. H.; Ahrens, W. A.; Showlater, V.; Iannitti, D. A.; Mckillop, I. H., Expression and function of lysophosphatidic acid receptors (LPARs) 1 and 3 in human hepatic cancer progenitor cells. *Oncotarget* **2016,** *7* (3), 2951-2967.

122. Shida, D.; Watanabe, T.; Aoki, J.; Hama, K.; Kitayama, J.; Sonoda, H.; Kishi, Y.; Yamaguchi, H.; Sasaki, S.; Sako, A.; Konishi, T.; Arai, H.; Nagawa, H., Aberrant expression of lysophosphatidic acid (LPA) receptors in human colorectal cancer. *Lab Invest* **2004,** *84* (10), 1352-1362.

123. Schulte, K. M.; Beyer, A.; Köhrer, K.; Oberhäuser, S.; Röher, H. D., Lysophosphatidic acid, a novel lipid growth factor for human thyroid cells: over-expression of the high-affinity receptor edg4 in differentiated thyroid cancer. *Int J Cancer* **2001,** *92* (2), 249-256.

124. Yanagida, K.; Kurikawa, Y.; Shimizu, T.; Ishii, S., Current progress in non-Edg family LPA receptor research. *Biochim Biophys Acta* **2013,** *1831* (1), 33-41.

Chapter 2

Design, synthesis and biological evaluation of fluorine-containing triazole derivatives targeting lysophosphatidic acid receptor 1 for the treatment and diagnosis of breast cancer

Part one

Design, synthesis and biological evaluation of fluorine-containing carbamate compounds targeting lysophosphatidic acid receptor 1 for the treatment of breast cancer

Abstract:

Lysophosphatidic acid receptor 1 (LPA$_1$) belongs to the G protein-coupled receptor (GPCR) family and is widely expressed in several types of tissues and cells. Activation of LPA$_1$ by its ligand is involved in a diverse array of cellular activities. Aberrant expression of LPA$_1$ is associated with breast cancer development and progression. LPA$_1$ overexpression enhanced tumor growth, while silencing or pharmacological inhibition of LPA$_1$ significantly reduces the tumor size and blocked metastases in breast cancer mouse models. The bioinformatics analysis has also suggested that LPA$_1$ is widely involved in cancer cell activities. Higher expression of LPA$_1$ is associated with lower survival in breast cancer patients. The aim of this study was to develop ligands targeting LPA$_1$ as a novel therapy for breast cancer. The triazole derivatives initially developed by Roche for the treatment of idiopathic pulmonary fibrosis were used as the starting point. Based on the most potent and selective compound (**RO6842262**) discovered in their work, fluorine was incorporated in different regions of the compound guided. A total of nine novel compounds were synthesized. Among them, compound (**12f**) exhibited the highest potency against LPA$_1$, with an IC$_{50}$ of 16.0 nM in cAMP assay. The *in vitro* studies showed that **12f** blocked LPA-induced cell survival, migration, and invasion of the breast cancer cell line MDA-MB-231. The LPA$_1$ antagonists developed in this study have the potential to create a novel therapy for breast cancer. In addition, these fluorine-containing compounds could potentially be used as positron emission tomography (PET) imaging agents for early diagnosis of breast cancer.
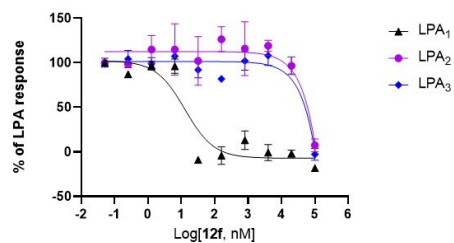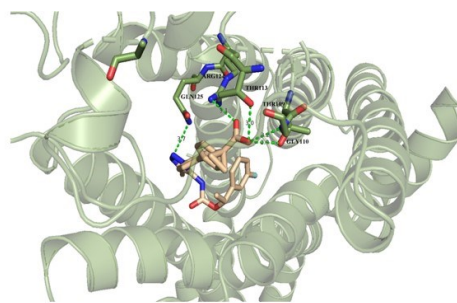
Graphical abstract:
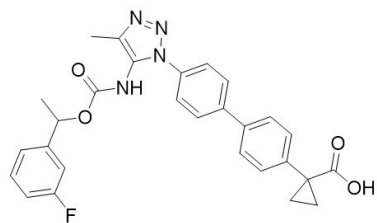
Compound 12f

LPA$_1$ IC$_{50}$ = 16.0 nM
LPA$_2$ IC$_{50}$ > 10.0 μM
LPA$_3$ IC$_{50}$ > 10.0 μM

➢ Antiproliferative activity

➢ Antimigration activity

➢ Anti-invasion activity

1. Introduction

Lysophosphatidic acid receptor 1 (LPA$_1$) is one of six GPCRs in the LPA receptor family (LPA$_{1-6}$), and it is activated by the bioactive phospholipid, lysophosphatidic acid (LPA). LPA$_1$ is implicated in a diverse array of cellular pathways that regulate cell proliferation, migration, invasion, and apoptosis.[1-2] It has been reported that the mRNA expression of LPA$_1$ is elevated in advanced stages of breast cancer compared with early stages. In several breast cancer cell lines, the expression of LPA$_1$ is significantly higher compared with non-tumorigenic cell lines.[3] In several breast cancer cell lines, activation of LPA$_1$ stimulated cell migration and invasion, while LPA antagonists inhibited the effects of LPA-induced proliferation and migration.[4-5] In breast cancer mouse models, overexpression of LPA$_1$ enhanced tumor growth and promoted breast cancer metastasis to the bone, while the silencing or pharmacological inhibition of LPA$_1$ significantly reduced the tumor size and blocked metastasis.[5-6] In metastasis mouse models, both administration of LPA$_1$ inhibitor and knockdown of LPA$_1$ expression blocked the metastasis of breast cancer cells to liver and lungs. Mechanistic studies showed that treatment with LPA$_1$ inhibitor reduced the expression level of the proliferation markers Ki67 and pERK on the metastatic sites, while increased the phosphorylation level of the antitumorigenic marker p38.[7]

In addition to the evidence above, a bioinformatics analysis was conducted using various integrative analyzing tools and websites to further understand the association between LPA$_1$ and diseases. LPA$_1$ receptor-related diseases were first explored on OpenTarget, and the constructed bubble graph (Figure 2.1.1 A) showed that LPA$_1$ is associated with several types of cancer, including breast cancer and ovarian cancer, among others. To investigate the functional roles of LPA$_1$ and its interactive or co-expressed proteins, functional enrichment studies were performed. Ten molecules showed significant interaction with LPA$_1$, including CXCL8, EGFR, GNA13, GNAI1, LPA$_5$, LPA$_4$, CD97, LPA$_6$, ENPP2, and GNA12 (Figure 2.1.1 B). KEGG pathway analysis showed that these proteins were mainly enriched in cancer pathways, followed by the Rap1 signaling pathway. This analysis further proved that LPA$_1$ is widely involved in cancer disease activities (Figure 2.1.1 C). To investigate the relationship between LPA$_1$ and breast cancer, the prognostic value of LPA$_1$ in breast cancer was analysed (Figure 2.1.1 D). The result suggested that higher expression of LPA$_1$ is significantly associated with lower overall survival rate in breast cancer patients.
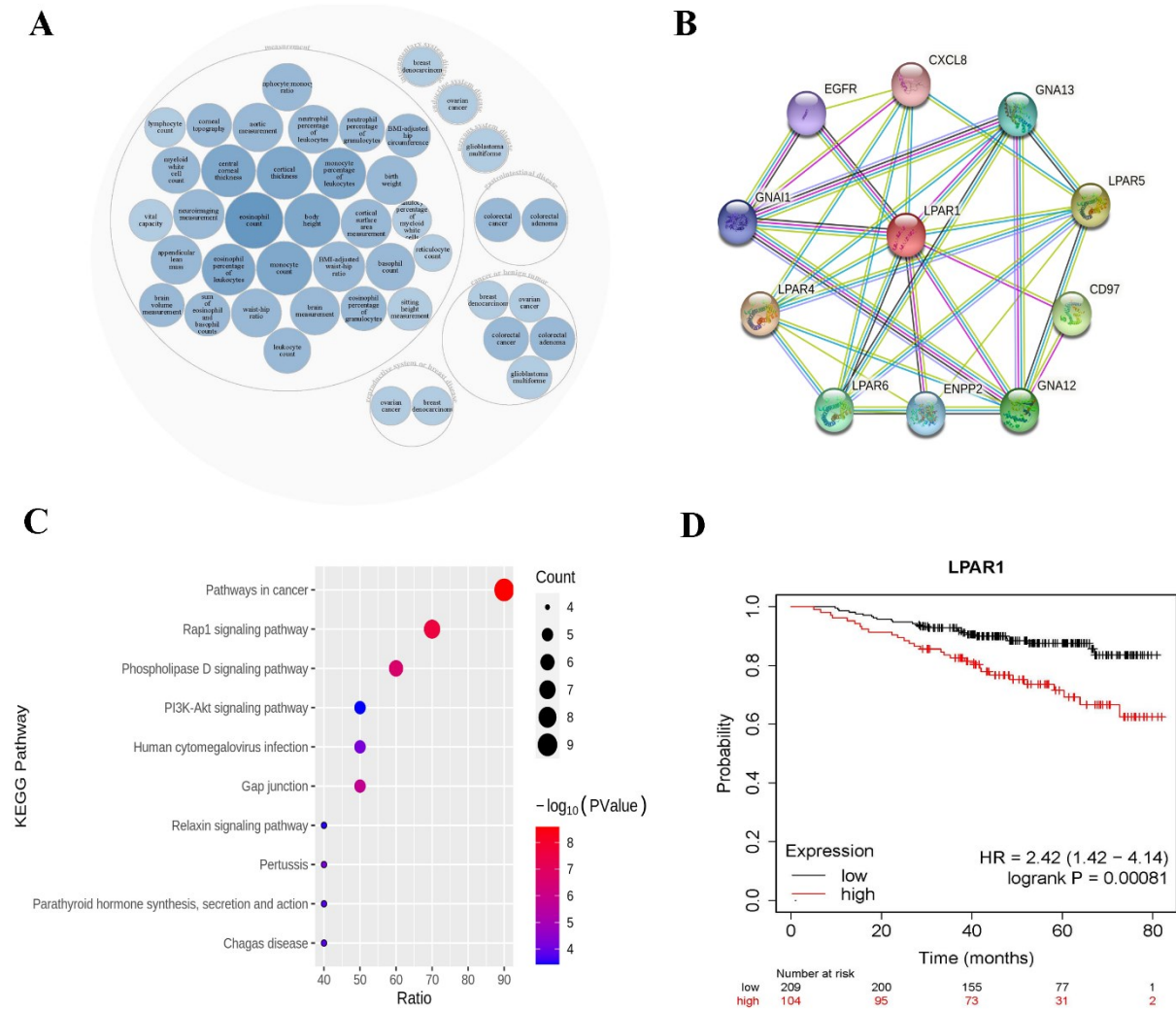
Figure 2.1.1 Bioinformatics analysis. A. The diseases associated with LPA$_1$ were analyzed using the OpenTarget web tool. B. Protein-protein interaction network of LPA$_1$ and LPA$_1$-related genes. C. KEGG pathway analysis of LPA$_1$ and LPA$_1$-related genes. D. LPA$_1$ was identified as a prognostic gene in breast cancer.

These pieces of evidence strongly indicate that LPA$_1$ plays a significant role in both the initiation and progression of breast cancer. Consequently, it holds substantial potential as a therapeutic target, particularly in cases involving breast cancer metastasis. However, the current repertoire of LPA$_1$ ligands targeting cancer remains remarkably limited.

The initial investigations into ligands targeting LPA$_1$ primarily focused on lipid-like molecules featuring lengthy fatty acid chains. These compounds were designed based on the structure of the

native LPA ligand (Figure 2.1.2). However, these molecules often exhibited suboptimal drug-like properties, mainly due to the hydrophobic nature of the moiety. The first non-lipid compound **Ki16425** was reported to be a LPA$_1$/LPA$_3$ dual antagonist.[8] Based on the isoxazole scaffold of **Ki16425** (Figure 2.1.2), other potent and selective LPA$_1$ antagonists were discovered, for example, **AM095**, **BMS-986020** and **[11]C-BMT-136088**. **BMS-986020** (Figure 2.1.2) are currently under phase II clinical trials for the treatment of idiopathic pulmonary fibrosis.[9] **[11]C-BMT-136088** is the only LPA$_1$ radiotracer reported, and it is at phase I of clinical trials.[10-11] **RO6842262** is another **Ki16425**-derived LPA$_1$ antagonist for the treatment of idiopathic pulmonary fibrosis. By substituting the isoxazole to the triazole scaffold, **RO6842262** (Figure 2.1.2) was found to be a potent and selective LPA$_1$ antagonist with an IC$_{50}$ value of 25 nM (LPA$_3$ IC$_{50}$ > 30 µM) in calcium release assay.[12] **RO684226** inhibited lung fibroblast proliferation and contraction induced by LPA *in vitro*, and ameliorated histamine release induced by LPA *in vivo*, demonstrating great therapeutic potential for lung fibrosis.[12]
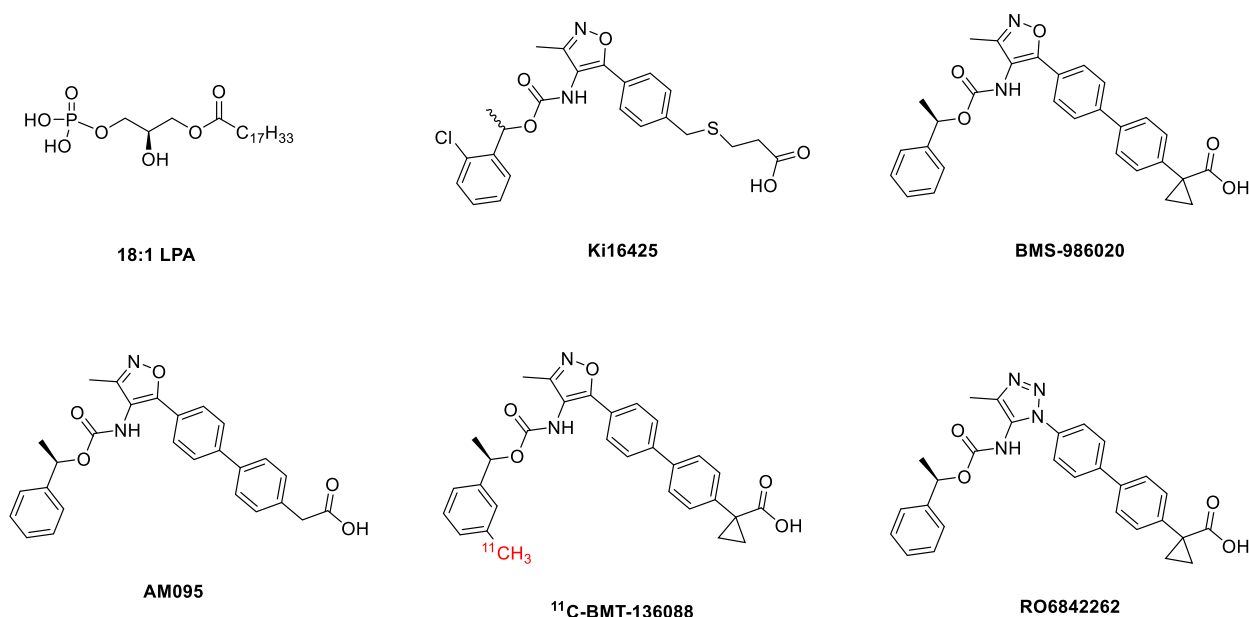
Figure 2.1.2 The structures of the endogenous LPA ligand (**18:1 LPA**) and known LPA$_1$ antagonists.

A key element in drug design and development, fluorine is the most electronegative element, and it is used as a common substitution to modulate molecular properties, such as conformation, pKa, potency, permeability, and pharmacokinetics.[13] Systematic fluorine substitution of ligands is a

promising strategy that could enhance protein-ligand binding.[14] Additionally, fluorine is also one of the most prominent elements used in positron emission tomography (PET) in the form of fluorine-18. To investigate the potential use of LPA_1 antagonist in breast cancer, the structure of **RO6842262** was used as a starting point, and fluorine was incorporated into this compound to develop novel and potent LPA_1 antagonists for the treatment of breast cancer and the potential usage as PET imaging agents. Several compounds with high potency and selectivity were discovered. Particularly, compound **12f** was the most potent in this study with its IC_{50} being 16.0 nM in cAMP assay. It inhibited survival, migration, and invasion in the breast cancer cell line MDA-MB-231 in a dose-dependent manner.

2. Materials and methods

2.1 Materials

2.1 1. Cells

Chinese hamster ovary (CHO)-K1 cell line was a kind gift from Dr. Rithwik Ramachandran (University of Western Ontario). The human breast cancer cell line MDA-MB-231 and cervical cancer cell line HeLa were kind gifts from Dr. Simon Lees (NOSM University). The human ovarian cancer cell line OVCAR-8 was a kind gift from Dr. Leonard Luyt (University of Western Ontario).

2.1.2. Reagents and consumables

2.1.2.1 Cell culture

Dulbecco's Modified Eagle Medium (DMEM)	WISENT Bioproducts
Kaighn's Modification of Ham's F-12 Medium (F-12K)	WISENT Bioproducts
Fetal bovine serum	WISENT Bioproducts
Trypsin	WISENT Bioproducts
Phosphate buffer saline (PBS)	WISENT Bioproducts
Dimethyl sulfoxide (DMSO)	Corning

2.1.2.2 Bacterial transformation

DH5α competent *E. coli*	Dr. Wensheng Qin's lab
Terrific broth (TB)	Bio Basic
LB Agar Plates with 100 µg/mL Ampicillin	Dr. Wensheng Qin's lab
PureLink™ HiPure Plasmid DNA Purification Kits	Invitrogen

2.1.2.3 Transfection

X-tremeGENE 9 DNA Transfection Reagent	Roche
pCMV6-AC-GFP-LPAR$_1$ plasmid	OriGene Technologies

pCMV6-AC-GFP-LPAR$_2$ plasmid	OriGene Technologies
pCMV6-AC-GFP-LPAR$_3$ plasmid	OriGene Technologies

2.1.2.4 cAMP assay

18:1 Lyso PA	Avanti Polar Lipids
Bovine serum albumin	WISENT Bioproducts
cAMP-Glo™ Assay kit	Promega
Isobutyl-1-methylxanthine (IBMX)	Thermo scientific
Ro 20-1724	MilliporeSigma
Forskolin	MilliporeSigma

2.1.2.5 Colony formation assay

Paraformaldehyde	Thermo scientific
Crystal violet	Thermo scientific

2.1.2.6 Transwell cell migration and invasion assays

Basement membrane extract	R&D Systems

2.1.2.7 Consumerbles

100-1000 µL pipette tips	Bio Basic
200 µL pipette tips	Bio Basic
0.1-10 µL pipette tips	Bio Basic
8.0 µm cell culture insert	Corning
0.2 µm filter	Whatman
15 mL graduated centrifuge tube	Bio Basic
50 mL graduated centrifuge tube	Bio Basic
75 cm^2 cell culture flask	Thermo scientific
25 cm^2 cell culture flask	Thermo scientific

6-well cell culture plate	Corning
24-well cell culture plate	Corning
96-well cell culture plate	Corning
96-well white, clear-bottom plate	Corning
2 mL Cryogenic Vial	Fisherbrand
5 mL serological pipette	Bio Basic
1.5 mL microcentrifuge tube	Corning

2.1.3. Instruments and software

Eppendorf Pipette	Eppendorf
Centrifuge	Thermo scientific
IX51 Inverted Microscope	Olympus
Inverted Fluorescence Microscope	EVOS FL
Cell culture incubator	Thermo scientific
Biosafety cabinet	Labconco
Incubating shaker	VWR
NanoDrop spectrophotometer	Thermo scientific
Multi-mode microplate reader	BioTek Synergy HTX
ImageJ	NIH
GraphPad Prism	GraphPad Software
AutoDock Vina	Scripps Research
Discovery Studio Visualizer	Dassault Systemes
500 MHz Avance Nuclear Magnetic Resonance (NMR) Spectrometer	Bruker
Liquid chromatography–mass spectrometry (LC–MS)	Waters
Freeze dryer	Labconco

2.2 Methods

2.2.1. Cell culture

Chinese hamster ovary (CHO)-K1 cell line was cultured in F-12K medium supplemented with 10% fetal bovine serum and 1% streptomycin/penicillin at 37°C, 5% CO_2. The MDA-MB-231, HeLa, and OVCAR-8 was cultured in DMEM supplemented with 10% fetal bovine serum and 1% streptomycin/penicillin at 37 °C, 5% CO_2 (Figure 2.1.3).
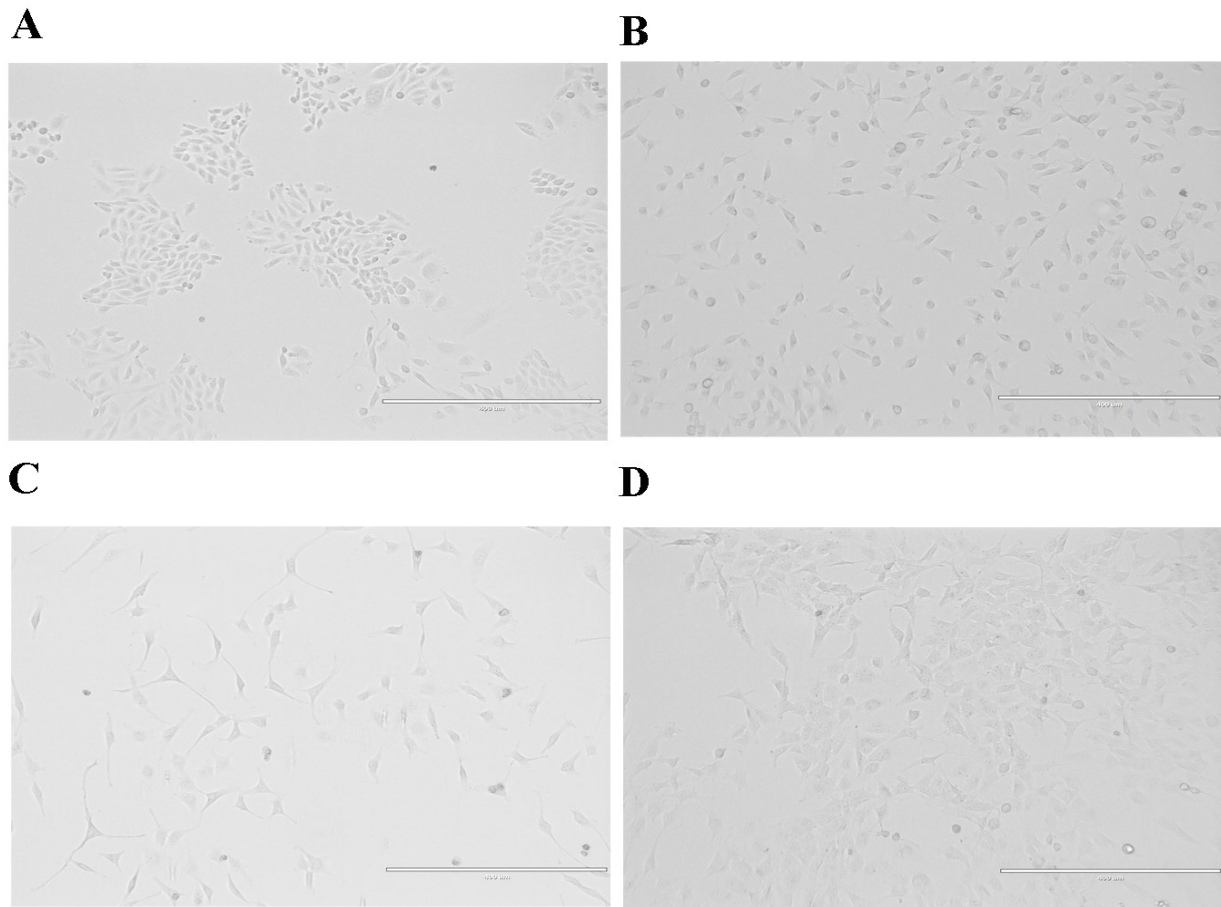
Figure 2.1.3. The CHO-K1 (A), MDA-MB-231 (B), HeLa (C), and OVCAR-8 (D) cell lines under inverted microscope.

2.2.2. Transformation Procedure

The following protocol was followed to conduct the experiment:

1) Thaw DH5α competent *E. coli* cells on wet ice. Place required number of 1.5mL Eppendorf tubes on ice.

2) Gently mix cells, then aliquot 100 µL of DH5α cells into chilled tubes.

3) Refreeze any unused DH5α cells in a dry ice bath for 5 minutes before returning them to the -80°C freezer. Do not use liquid nitrogen.

4) Take 1 µL LPA plasmid, dilute it 10-fold in 10 mM Tris-HCl (pH 7.5) and 1 mM EDTA to a final concentration of 10 ng/µL. Add 1 µL of the dilution to the DH5α cells, moving the pipette through the cells while dispensing. Gently tap tubes to mix.

5) Incubate DH5α cells on ice for 30 minutes.

6) Heat-shock DH5α cells 45 seconds in a 42°C water bath; do not shake.

7) Place the tube on ice for 2 minutes.

8) Add 0.9 mL of room temperature terrific broth (TB) and shake at 225 rpm (37°C) for 1 hour.

9) Spread 100 µL of this dilution on LB plates with 100 µg/mL ampicillin and 50 µg/mL X-gal.

10) Incubate plates overnight at 37°C.

11) Use a 10 µL tip to one clone into a culture tube containing 1.5 mL LB with ampicillin, and culture the clone in a shaker at 37°C with shaking at 220 rpm overnight.

12) Next day, use the PureLink™ HiPure Plasmid Purification Kit to isolate plasmid DNA.

2.2.3. Plasmid isolation

The following protocol was followed to conduct the experiment:

Equilibrate the column: Apply 10 mL Equilibration Buffer (EQ1) to the column. Allow the solution in the column to drain by gravity flow. Proceed to Prepare cell lysate, while the column is equilibrating.

Prepare cell lysate:

1) Use 1-3 mL of an overnight LB culture in a disposable 15 mL conical tube.

2) Harvest the cells by centrifuging the overnight LB culture at 4000 × g for 10 minutes. Remove all medium.

3) Add 0.4 mL Resuspension Buffer (R3) with RNase A to the cell pellet and resuspend the cells until homogeneous.

4) Add 0.4 mL Lysis Buffer (L7). Mix gently by inverting the capped tube until the lysate mixture is thoroughly homogenous. Do not vortex. Incubate at room temperature for 5 minutes. Note: Do not allow lysis to proceed for more than 5 minutes.

5) Add 0.4 mL Precipitation Buffer (N3) and mix immediately by inverting the capped tube until the mixture is thoroughly homogeneous. Do not vortex.

6) Centrifuge the mixture at >12,000 × g for 10 minutes at room temperature.

7) Proceed to Bind and wash DNA.

Bind and wash DNA:

1) Load the supernatant from Prepare cell lysate, step 6 onto the equilibrated column. Allow the solution in the column to drain by gravity flow.

2) Wash the column twice with 2.5 mL Wash Buffer (W8). Allow the solution in the column to drain by gravity flow after each wash. Discard the flow-through.

3) Proceed to Elute and precipitate DNA.

Elute and precipitate DNA:

1) Place a sterile 15 mL centrifuge tube (elution tube) under the column.

2) Add 0.9 mL Elution Buffer (E4) to the column to elute the DNA. Allow the solution to drain by gravity flow. Do not force out any remaining solution.

3) Add 0.63 mL isopropanol to the elution tube. Mix well.

4) Centrifuge the tube at >12,000 × g for 30 minutes at 4°C. Carefully remove and discard the supernatant.

5) Resuspend the pellet in 1 mL 70% ethanol.

6) Centrifuge the tube at >12,000 × g for 5 minutes at 4°C. Carefully remove and discard the supernatant.

7) Air-dry the pellet for 10 minutes.

8) Resuspend the DNA pellet in 50 μL TE Buffer (TE).

9) Measure the concentration of the plasmid using a NanoDrop spectrophotometer.

10) To avoid repeated freezing and thawing of DNA, store the purified DNA at 4°C for immediate use or aliquot the DNA and store at –20°C for long-term storage.

The sequence of the plasmids was verified by Sanger sequencing/whole plasmid sequencing performed by Eurofins Genomics (Appendix E).

2.2.4. Transfection of pCMV6-AC-GFP-LPAR$_1$/pCMV6-AC-GFP-LPAR$_2$/pCMV6-AC-GFP-LPAR$_3$ plasmid DNA into CHO-K1 cells in a 96-well cell culture plate

The day before transfection, the cells were trypsinized and counted. 8×10^3 cells per well were plated in 100 μL of complete growth medium. The following protocol was used to conduct the experiment:

1) Allow X-tremeGENE 9 DNA Transfection Reagent, DNA, and diluent to equilibrate.
2) Dilute X-tremeGENE 9 DNA Transfection Reagent with serum-free medium to a concentration of 3 μL reagent/100 μL medium for a ratio of 3:1, using a sterile tube and gentle mixing.
3) Add 0, 0.1, 0.2, 0.4, 0.5, 0.6, 0.7, 0.8, and 1 μg of DNA to 100 μL of diluted X-tremeGENE 9 DNA Transfection Reagent, respectively; mix gently.
4) Incubate the transfection reagent: DNA complex for 15 minutes at room temperature.
5) Remove the culture vessel from the incubator; removal of growth medium is not necessary. Add 5 μL of the transfection complex to the cells in a dropwise manner in a 96-well plate.
6) Use a fluorescence microscope to observe green fluorescent protein (GFP) and calculate transfection efficiency after 24-48 hours.

Based on the transfection results, 0.8 μg of pCMV6-AC-GFP-LPAR$_1$/ pCMV6-AC-GFP-LPAR$_2$/ pCMV6-AC-GFP-LPAR$_3$ plasmid in 100 μL of diluted transfection reagent was used in the following experiments (Figure 2.1.4).
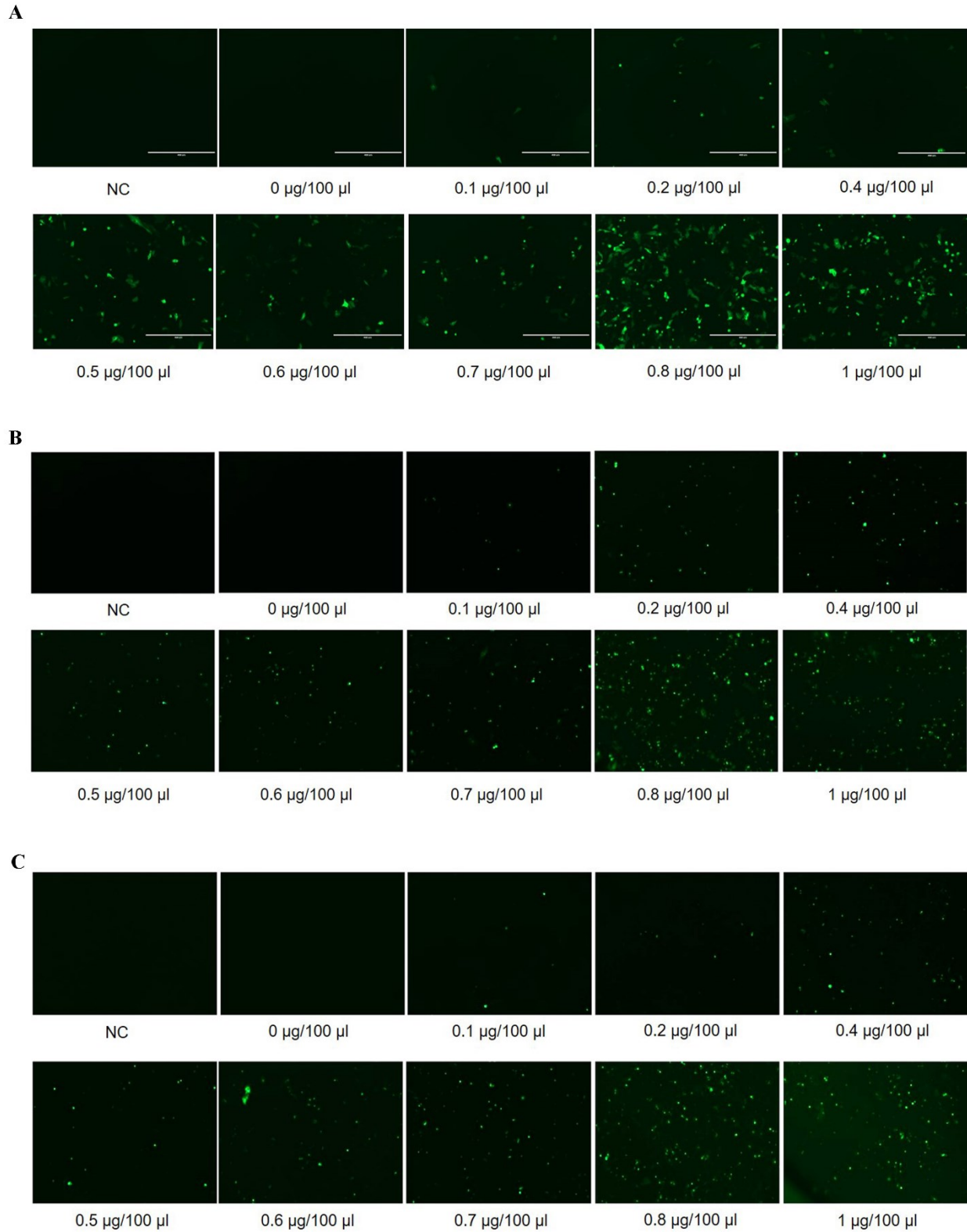
Figure 2.1.4. Determination of transfection efficiencies by GFP visualization. The CHO-K1 cell line was transfected with different concentration of pCMV6-AC-GFP-LPAR$_1$ (A), pCMV6-AC-

GFP-LPAR$_2$ (B), and pCMV6-AC-GFP-LPAR$_3$ (C) plasmid and visualized 48 hours post-transfection under a fluorescence microscopy.

2.2.5. cAMP assay

2.2.5.1 Determining EC$_{90}$ values of **18:1 LPA**

The CHO-K1 cells overexpressing LPAR$_1$/LPAR$_2$/LPAR$_3$ were prepared in a white, clear-bottom 96-well plate. Prior to performing the cAMP assay, a serial 5-fold dilution of **18:1 LPA** with induction buffer containing 100 nM forskolin was performed. The protocol outlined below was followed:

1) Remove the medium and transfer 20 µL of the various concentrations of **18:1 LPA** to the assay plate. Incubate the plate at room temperature for 15 minutes to allow cells and compound to interact.

2) Add 20 µL of Lysis Buffer to all wells. Incubate the plate at room temperature for 15 minutes.

3) Add 2.5 µL Protein Kinase A to 1.0 mL of cAMP-Glo Reaction Buffer to prepare cAMP Detection Solution. Add 40 µL of cAMP Detection Solution to all wells. Mix the plate by shaking for 30-60 seconds, and incubate the plate at room temperature for 20 minutes.

4) Add 80 µL of room-temperature Kinase-Glo Reagent to all wells. Mix the plate by shaking for 30-60 seconds, and incubate the plate at room temperature for 10 minutes.

5) Measure luminescence in each well using a microplate reader.

Relative light units (RLU) were plotted against the log of **18:1 LPA** concentrations. Data were fit to a three-parameter logistic curve to generate the EC$_{90}$ value of **18:1 LPA**. It was determined that **18:1 LPA** has an EC$_{90}$ value of 968.4 nM for LPA$_1$, 421.56 nM for LPA$_2$, and 2071.8 nM for LPA$_3$, respectively (Figure 2.1.5). Based on these EC$_{90}$ values, the cAMP assay was conducted using a final concentration of 1 µM **18:1 LPA** for LPA$_1$, 500 nM for LPA$_2$, and 2.1 µM for LPA$_3$.
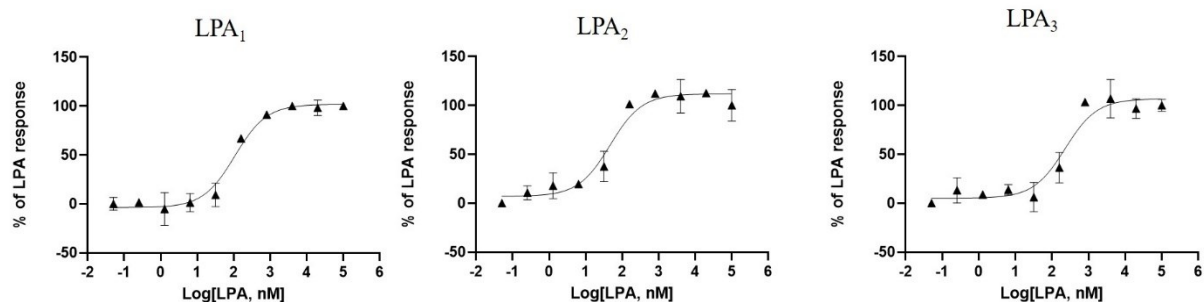
Figure 2.1.5 Dose-response curves of **18:1 LPA** towards the LPA_1, LPA_2, and LPA_3 receptors.

2.2.5.2 Determining IC_{50} values of the test compounds

The CHO-K1 cells overexpressing $LPAR_1$/$LPAR_2$/$LPAR_3$ were prepared in a white, clear-bottom 96-well plate. Prior to performing the cAMP assay, a serial 5-fold dilution of test compounds with induction buffer containing 100 nM forskolin and 1 μM **18:1 LPA** (500 nM for LPA_2, and 2.1 μM for LPA_3) were performed. The following protocol was followed:

1) Remove the medium and transfer 20 μL of the various concentrations of test compounds to the assay plate. Incubate the plate at room temperature for 15 minutes to allow cells and compound to interact.

2) Add 20 μL of Lysis Buffer to all wells. Incubate the plate at room temperature for 15 minutes.

3) Add 2.5 μL Protein Kinase A to 1.0 mL of cAMP-Glo Reaction Buffer to prepare cAMP Detection Solution. Add 40 μL of cAMP Detection Solution to all wells. Mix the plate by shaking for 30-60 seconds, and incubate the plate at room temperature for 20 minutes.

4) Add 80 μL of room-temperature Kinase-Glo Reagent to all wells. Mix the plate by shaking for 30-60 seconds, and incubate the plate at room temperature for 10 minutes.

5) Measure luminescence in each well using a microplate reader.

Relative light units (RLU) were plotted against the log of the test compounds concentrations. Data were fit to a three-parameter logistic curve to generate the IC_{50} values of the test compounds.

2.2.6. Wound healing assay

MDA-MB-231, OVCAR-8, and HeLa cells were seeded in a six-well plate at a density of 3×10^5 cells/well. When the cells formed a tight cell monolayer, a 200-μL plastic pipette tip was used to make a scratch. Cells were washed with PBS to remove cell debris, cultured in serum-free DMEM medium and exposed to **18:1 LPA** or **18:1 LPA** with the LPA_1 antagonists (1, 5, 10 or 20 μM) for a duration of 48 hours. To conduct serum-induced migration assays, cells were cultured in DMEM supplemented with 10% FBS and exposed to varying concentrations of LPA_1 antagonists (1, 5, 10 or 20 μM) for a duration of 48 hours. Wound photographs were recorded at the indicated times and analyzed by ImageJ software. All assays were conducted three times in this study.

2.2.7. Colony formation assay

MDA-MB-231, OVCAR-8, and HeLa cells were trypsinized and plated in 6-well plate at a density of 1,000 cells/well and then cultured in 10% FBS DMEM. Cells were allowed to attach overnight and then exposed to **18:1 LPA** or **18:1 LPA** with the LPA_1 antagonists (1, 5, 10 or 20 μM). Forty-eight hours after chemical treatment, the media was replaced with fresh media, and the plates were incubated at 37°C. Ten days later, the cells were fixed and stained with 4% paraformaldehyde in 0.1% crystal violet. The number of colonies, defined as > 50 cells/colony were counted, and the number of colonies were calculated using ImageJ.

2.2.8. Transwell migration and invasion assays

Transwell assays were conducted in 24-well transwell plates (pore size: 8 μm) to assess the migratory and invasive capacities of MDA-MB-231 cells. For migration assays, 2×10^4 cells were placed in 200 μL of serum-free DMEM in the upper chamber and then 500 μL of DMEM was added to the lower chamber. For the invasion assays, the chamber inserts were precoated with 50 μL of 1:9 mixture of basement membrane extract and DMEM overnight in a 37 °C incubator, then 6×10^4 cells were seeded in the upper chamber. Cells were treated with **18:1 LPA** or **18:1 LPA** with the LPA_1 antagonists (1, 5, 10 or 20 μM) for 48 h. 4% paraformaldehyde was used to fix the cells that had migrated or invaded to the lower surface of the membrane. Crystal violet (0.1%) was applied for staining the fixed cells for 15 min. Five random 100× microscopic fields were selected to count the stained cells.

2.2.9. Molecular docking study

The crystal structure of LPA$_1$ (PBD ID:4z36)[15] was used to perform molecular docking studies. The 3D structure of **RO6842262** and the synthesized compounds in this study were generated using DS Viewer 3.5. Molecular docking was performed using AutoDock Vina.[16] The dimensions of the grid box were set large enough to encompass the whole binding pocket of the LPA$_1$ structure. All other parameters were left as default.

2.2.10. Statistical analysis

The experimental results are expressed as mean ± standard error of mean (SEM). GraphPad Prism 9.0.0 was used for data processing. Student's *t*-test or One-way ANOVA were used to analyze the statistical *P* values between different groups. A *P* value of < 0.05 (two-tailed) was considered as statistically significant.

2.2.11. Synthesis

All solvents and chemicals were reagent grade and purchased from commercial manufacturers. Purity and characterization of compounds were determined by a combination of HPLC, TLC, high-resolution mass spectrometry (HRMS), ^{1}H NMR and ^{13}C NMR analyses. Spectral data was compared with reference data from the literature where possible; for **RO6842262**, ^{1}H NMR was compared.[12] Experimental data and reported data compared favourably in all cases.[12,17-18]

*1-(4-Bromophenyl) cyclopropane-1-carbonitrile (**1**).*

Sodium hydroxide (3.05 g, 76.5 mmol) was dissolved in H$_2$O (5 mL) and toluene (12 mL). 4-bromophenylacetonitrile (1.5 g, 7.65 mmol) and tetrabutylammonium bromide (0.123 g, 0.38 mmol) were added, followed by dibromoethane (0.975 mL, 11.4 mmol) dropwise. The reaction was stirred at 85 °C for 5 hours. The reaction mixture was partitioned between DCM and water. The organic layer was extracted with DCM and washed with 1N HCl and brine, dried over Na$_2$SO$_4$, filtered, and concentrated. The residue was purified by column chromatography (Hexane: EtOAc = 20:1) to give compound **1** (523.33 mg, yield: 30.85%). ^{1}H NMR (500 MHz, Chloroform-*d*) δ 7.47 (d, J = 8.4 Hz, 2H), 7.16 (d, J = 8.3 Hz, 2H), 1.74 (dd, J = 8.0, 5.1 Hz, 2H), 1.38 (dd, J = 7.5, 5.1 Hz, 2H).

*l-(4-Bromophenyl) cyclopropane-1-carboxylic acid (**2**).*

Compound **1** (1.5 g, 6.75 mmol) and sodium hydroxide (1.07 g, 26.80 mmol) were dissolved in ethylene glycol (21 mL), and the reaction was stirred at 180°C for 5 hours. The mixture was added with water and acidified with concentrated HCl. The resulting mixture was then filtered under reduced pressure to give compound **2** (1.53 g, yield: 98.00%). This compound was used in the next step without further purification.

l-(4-Bromophenyl) cyclopropane-1-carboxylic acid ethyl ester (3).

Compound **2** (1.53 g, 5.20 mmol) was dissolved in anhydrous EtOH (20 mL) and concentrated H_2SO_4 (1 mL) was added. The reaction was stirred at 80°C for 9 hours. The mixture was diluted with EtOAc, washed with brine, dried over Na_2SO_4, filtered, and concentrated to give compound **3** (1.10 g, yield: 80.88%). ^1H NMR (500 MHz, Chloroform-d) δ 7.42 (d, J = 8.2 Hz, 2H), 7.21 (d, J = 8.3 Hz, 2H), 4.09 (q, J = 7.1 Hz, 2H), 1.60 (dd, J = 7.0, 4.0 Hz, 2H), 1.20 – 1.10 (m, 5H).

Ethyl 1-(4-(4, 4, 5, 5-tetramethyl-1, 3, 2-dioxaborolan-2-yl) phenyl) cyclopropane-1-carboxylate (4).

The mixture of compound **3** (200 mg, 0.75 mmol), bis(pinacolato)diboron (208.45 mg, 0.825 mmol), and potassium acetate (184.03 mg, 1.875 mmol) were dissolved in dry dioxane (2 mL), and $PdCl_2$(dppf) (54.88 mg, 0.075 mmol) was added. The mixture was bubbled with argon and then stirred at 100°C under argon for 6 hours. The mixture was filtered through diatomaceous earth and partitioned between EtOAc and water. The organic phase was extracted with EtOAc, washed with brine, dried over Na_2SO_4, filtered, and concentrated. The residue was purified by column chromatography (Hexane: EtOAc = 20:1) to give compound **4** (149 mg, yield: 63.15%). ^1H NMR (500 MHz, Chloroform-d) δ 7.76 (d, J = 7.6 Hz, 2H), 7.34 (d, J = 7.6 Hz, 2H), 4.07 (q, J = 7.1 Hz, 2H), 1.59 (dd, J = 7.0, 4.0 Hz, 2H), 1.33 (s, J = 1.7 Hz, 12H), 1.20 – 1.09 (m, 5H). ESI-MS m/z calcd $C_{18}H_{25}BO_4$ $[M + H]^+$ 317.19, found 317.3.

Ethyl 2-(4-(4, 4, 5, 5-tetramethyl-1, 3, 2-dioxaborolan-2-yl) phenyl) acetate (6).

The mixture of compound **5** (1 g, 4.11 mmol), bis(pinacolato)diboron (1.149 mg, 4.52 mmol), and potassium acetate (1.008 g, 10.28 mmol) were dissolved in dry dioxane (10 mL), and $PdCl_2$(dppf) (0.301 g, 0.411 mmol) was added. The mixture was bubbled with argon and then stirred at 100°C under argon for 6 hours. The mixture was filtered through diatomaceous earth and partitioned between EtOAc and water. The organic phase was extracted with EtOAc, washed

with brine, dried over Na$_2$SO$_4$, filtered, and concentrated. The residue was purified by column chromatography (Hexane: EtOAc = 20:1) to give compound **6** (828 mg, yield: 69.00%). ^{1}H NMR (500 MHz, Chloroform-*d*) δ 7.77 (d, J = 8.1 Hz, 2H), 7.29 (d, J = 8.2 Hz, 2H), 4.13 (q, J = 7.1 Hz, 2H), 3.62 (s, 2H), 1.33 (s, 12H), 1.23 (t, J = 7.1 Hz, 3H).

1-Azido-4-bromobenzene (7).

4-Bromoaniline (2 g, 11.63 mmol) was dissolved in 4N HCl (15 mL) and was cooled to 0°C and stirred. To this suspension was added sodium nitrite (0.883 g, 12.79 mmol) dissolved in water (2.33 mL) dropwise. After 30 minutes, sodium azide (0.907 g, 13.95 mmol) was added and the reaction mixture was slowly warmed to rt and stirred for 1 hour. The reaction mixture was stored at 4°C overnight and the resulting residue was filtered to give compound **7** (2.06 g, yield: 89.57%). ^{1}H NMR (500 MHz, Chloroform-d) δ 7.46 (d, J = 8.4 Hz, 2H), 6.90 (d, J = 8.4 Hz, 2H).

Ethyl 1-(4-bromophenyl)-4-methyl-1H-1, 2, 3-triazole-5-carboxylate (8).

To a solution of compound **7** (200 mg, 1.01 mmol) in toluene (1 mL) was added but-2-ynoic acid ethyl ester (0.14 mL, 1.21 mmol). The reaction mixture was heated to reflux at 110°C under argon for 15 hours. The reaction mixture was concentrated and then purified by column chromatography (Hexane: EtOAc = 10:1) to give compound **8** (85 mg, yield: 27.16%). ^{1}H NMR (500 MHz, Chloroform-*d*) δ 7.65 (d, J = 8.8 Hz, 2H), 7.32 (d, J = 8.8 Hz, 2H), 4.29 (q, J = 7.1 Hz, 2H), 2.63 (s, 3H), 1.27 (t, J = 7.1 Hz, 3H).

Ethyl 1-(4-bromophenyl)-4-methyl-1H-1, 2, 3-triazole-5-carboxylic acid (9).

To a solution of compound **8** (80 mg, 0.26 mmol) in MeOH/THF/H$_2$O (0.6/0.6/0.6 mL) was added sodium hydroxide (51.72 mg, 1.29 mmol). The reaction mixture was stirred at rt overnight. The mixture was cooled down to 0°C and neutralized to pH = 4.0 with 3N HCl. The mixture was extracted with EtOAc and the organic phase was dried over Na$_2$SO$_4$, filtered and concentrated to give crude compound **9** (70 mg, yield: 96.21%). ESI-MS m/z calcd C$_{10}$H$_8$BrN$_3$O$_2$ [M + H]$^+$ 281.99, found 282.0.

*2-Phenylpropan-2-yl (1-(4-bromophenyl)-4-methyl-1H-1, 2, 3-triazol-5-yl) carbamate (**10**).*

To a solution of compound **9** (70 mg, 0.248 mmol) in anhydrous toluene (1.75 mL) was added (4-fluorophenyl)methanol (40.36 mg, 0.32 mmol), triethylamine (50 mg), and DPPA (89.25 mg,

0.32 mmol). The reaction mixture was heated to 120°C for 5 hours. The mixture was diluted with EtOAc, washed with brine, dried over Na_2SO_4, filtered, and concentrated. The residue was purified by column chromatography (Hexane: EtOAc = 5:1) to give compound **10** (50 mg, yield: 49.8%). ^{1}H NMR (500 MHz, Chloroform-d) δ 7.58 (d, J = 8.1 Hz, 2H), 7.33 (d, J = 8.3 Hz, 4H), 7.05 (s, 2H), 5.09 (s, 2H), 2.31 (s, 3H).

Ethyl 1-(4'-(5-((((4-fluorobenzyl)oxy)carbonyl)amino)-4-methyl-1H-1,2,3-triazol-1-yl)-[1,1'-biphenyl]-4-yl)cyclopropane-1-carboxylate (11).

To a stirred mixture of compound **10** (100 mg, 0.25 mmol), compound **4** (93.6 mg, 0.30 mmol), and sodium carbonate (55.92 mg, 0.53 mmol) in dioxane (6 mL) and water (2 mL) was added $PdCl_2$(dppf) (18.27 mg, 0.025 mmol). The mixture was bubbled with argon and stirred at 80°C under argon for 6 hours. The mixture was filtered through diatomaceous earth and partitioned between EtOAc and water. The organic phase was extracted with EtOAc, washed with brine, dried over Na_2SO_4, filtered, and concentrated. The residue was purified by column chromatography (DCM: MeOH = 50:1) to give compound **11** (102 mg, yield: 65.88%). ^{1}H NMR (500 MHz, Chloroform-d) δ 7.63 (d, J = 8.2 Hz, 2H), 7.52 (d, J = 8.1 Hz, 2H), 7.48 (d, J = 7.9 Hz, 2H), 7.44 (d, J = 8.1 Hz, 2H), 7.29 (s, 2H), 7.00 (s, 2H), 5.09 (s, 2H), 4.12 (q, J = 7.1 Hz, 2H), 2.28 (s, 3H), 1.65 (dd, J = 7.0, 4.0 Hz, 2H), 1.23 (dd, J = 6.8, 4.0 Hz, 2H), 1.19 (t, J = 7.1 Hz, 3H).

1-(4'-(5-((((4-fluorobenzyl)oxy)carbonyl)amino)-4-methyl-1H-1,2,3-triazol-1-yl)-[1,1'-biphenyl]-4-yl)cyclopropane-1-carboxylic acid (12b).

To a stirred solution of compound **11** (100 mg, 0.19 mmol) in THF (9.9 mL) and water (3.3 mL) was added lithium hydroxide (23.95 mg, 0.97 mmol). The reaction mixture was stirred at rt for 24 hours. The reaction mixture was cooled down to 0°C and neutralized with conc. HCl to pH = 2. The mixture was extracted with EtOAc. The combined organic phase was dried over Na_2SO_4, filtered, and concentrated. The residue was purified by column chromatography (Hexane: EtOAc = 1:1) to give compound **12b** (80 mg, yield: 84.61%). ^{1}H NMR (500 MHz, Methanol-d4) δ 7.79 (d, J = 8.2 Hz, 2H), 7.65 (d, J = 8.3 Hz, 2H), 7.57 (d, J = 8.0 Hz, 2H), 7.51 (d, J = 8.3 Hz, 2H), 7.36 (s, 2H), 7.05 (s, 2H), 5.13 (s, 2H), 2.29 (s, 3H), 1.64 (dd, J = 7.0, 4.0 Hz, 2H), 1.27 (dd, J = 6.9, 3.9 Hz, 2H). ^{13}C NMR (126 MHz, Methanol-d4) δ 170.31, 165.06, 163.11, 143.57, 141.59, 139.45, 135.92, 132.27, 131.54, 131.47, 128.94, 127.84, 125.70, 125.68, 116.37, 116.20, 67.88,

29.65, 17.02, 9.64. HRMS (ESI) m/z: [M + Na]$^+$ Calcd for $C_{27}H_{23}FN_4NaO_4$ 509.1596; Found 509.1600.

1-(4'-(5-((((3-fluorobenzyl)oxy)carbonyl)amino)-4-methyl-1H-1,2,3-triazol-1-yl)-[1,1'-biphenyl]-4-yl)cyclopropane-1-carboxylic acid (12c)

Compound **12c** was prepared in a similar way as compound **12b**, except that (3-fluorophenyl)methanol was used from step 4 in scheme 1 (yield: 83.21%). ^{1}H NMR (500 MHz, Methanol-d4) δ 7.77 (d, J = 7.8 Hz, 2H), 7.63 (d, J = 7.7 Hz, 2H), 7.56 (d, J = 7.7 Hz, 2H), 7.49 (d, J = 7.7 Hz, 2H), 7.30 (s, 1H), 7.04 (m, 3H), 5.14 (s, 2H), 2.28 (s, 3H), 1.62 (br s, 2H), 1.25 (br s, 2H). ^{13}C NMR (126 MHz, DMSO-d6) δ 175.27, 163.07, 161.13, 154.00, 140.83, 140.02, 139.12, 138.35, 137.29, 134.53, 131.01, 130.56, 130.49, 129.57, 127.69, 126.46, 124.08, 123.62, 123.60, 114.95, 114.78, 114.40, 114.23, 65.76, 28.29, 15.84, 9.62. HRMS (ESI) m/z: [M + Na]$^+$ Calcd for $C_{27}H_{23}FN_4NaO_4$ 509.1596; Found 509.1598.

1-(4'-(5-((((2-fluorobenzyl)oxy)carbonyl)amino)-4-methyl-1H-1,2,3-triazol-1-yl)-[1,1'-biphenyl]-4-yl)cyclopropane-1-carboxylic acid (12d)

Compound **12d** was prepared in a similar way as compound **12b**, except that (2-fluorophenyl)methanol was used from step 4 in scheme 1 (yield: 80.56%). ^{1}H NMR (500 MHz, Methanol-d4) δ 7.76 (d, J = 7.9 Hz, 2H), 7.63 (d, J = 7.8 Hz, 2H), 7.55 (d, J = 6.8 Hz, 2H), 7.49 (d, J = 7.7 Hz, 2H), 7.31 (d, J = 6.0 Hz, 2H), 7.08 (s, 2H), 5.20 (s, 2H), 2.27 (s, 3H), 1.63 (br s, 2H), 1.25 (br s, 2H). ^{13}C NMR (126 MHz, Methanol-d4) δ 178.24, 163.33, 161.36, 143.52, 141.42, 139.55, 135.89, 132.25, 131.75, 131.68, 131.21, 128.95, 127.87, 125.60, 125.41, 125.38, 116.43, 116.26, 62.49, 29.56, 17.07, 9.63. HRMS (ESI) m/z: [M + Na]$^+$ Calcd for $C_{27}H_{23}FN_4NaO_4$ 509.1596; Found 509.1588.

1-(4'-(5-(((1-(4-fluorophenyl)ethoxy)carbonyl)amino)-4-methyl-1H-1,2,3-triazol-1-yl)-[1,1'-biphenyl]-4-yl)cyclopropane-1-carboxylic acid (12e)

Compound **12e** was prepared in a similar way as compound **12b**, except that 1-(4-fluorophenyl)ethan-1-ol was used from step 4 in scheme 1 (yield: 79.98%). ^{1}H NMR (500 MHz, Methanol-d4) δ 7.74 (d, J = 6.8 Hz, 2H), 7.61 (d, J = 7.7 Hz, 2H), 7.53 (s, 2H), 7.48 (d, J = 7.7 Hz, 2H), 7.34 (s, 2H), 7.02 (s, 2H), 5.72 (s, 1H), 2.24 (s, 3H), 1.62 (br s, 2H), 1.50 (s, 3H), 1.25 (br s, 2H). ^{13}C NMR (126 MHz, Methanol-d4) δ 178.23, 164.72, 162.77, 143.51, 141.43, 140.16,

139.48, 135.92, 132.26, 131.23, 129.01, 128.94, 127.84, 125.54, 116.29, 116.12, 74.98, 29.55, 22.59, 17.07, 9.61. HRMS (ESI) m/z: $[M + Na]^+$ Calcd for $C_{28}H_{25}FN_4NaO_4$ 523.1752; Found 523.1768.

1-(4'-(5-(((1-(3-fluorophenyl)ethoxy)carbonyl)amino)-4-methyl-1H-1,2,3-triazol-1-yl)-[1,1'-biphenyl]-4-yl)cyclopropane-1-carboxylic acid (12f)

Compound **12f** was prepared in a similar way as compound **12b**, except that 1-(3-fluorophenyl)ethan-1-ol was used from step 4 in scheme 1 (yield: 75.65%). ^{1}H NMR (500 MHz, Methanol-d4) δ 7.75 (d, J = 8.1 Hz, 2H), 7.62 (d, J = 8.2 Hz, 2H), 7.51 (dd, J = 19.9, 8.0 Hz, 4H), 7.29 (s, 1H), 7.19 – 6.68 (m, 3H), 5.73 (s, 1H), 2.26 (s, 3H), 1.63 (dd, J = 7.1, 3.9 Hz, 2H), 1.51 (s, 3H), 1.25 (dd, J = 6.8, 4.0 Hz, 2H). ^{13}C NMR (126 MHz, Methanol-d4) δ 178.29, 165.21, 163.26, 143.55, 141.46, 140.20, 139.52, 132.24, 131.46, 131.39, 131.17, 128.94, 127.85, 125.55, 122.63, 122.61, 115.70, 115.53, 74.83, 29.59, 22.70, 17.06, 9.60. HRMS (ESI) m/z: $[M + Na]^+$ Calcd for $C_{28}H_{25}FN_4NaO_4$ 523.1752; Found 523.1748.

1-(4'-(5-(((1-(2-fluorophenyl)ethoxy)carbonyl)amino)-4-methyl-1H-1,2,3-triazol-1-yl)-[1,1'-biphenyl]-4-yl)cyclopropane-1-carboxylic acid (12g)

Compound **12g** was prepared in a similar way as compound **12b**, except that 1-(2-fluorophenyl)ethan-1-ol was used from step 4 in scheme 1 (yield: 73.81%). ^{1}H NMR (500 MHz, Methanol-d4) δ 7.75 (d, J = 8.1 Hz, 2H), 7.62 (d, J = 8.3 Hz, 2H), 7.53 (d, J = 8.1 Hz, 2H), 7.49 (d, J = 8.2 Hz, 2H), 7.37 (s, 1H), 7.25 (q, J = 6.3 Hz, 1H), 7.15 – 6.84 (m, 2H), 5.98 (s, 1H), 2.26 (s, 3H), 1.63 (dd, J = 7.0, 3.9 Hz, 2H), 1.53 (s, 3H), 1.25 (dd, J = 6.9, 4.0 Hz, 2H). ^{13}C NMR (126 MHz, Methanol-d4) δ 178.27, 161.92, 159.96, 143.52, 141.46, 140.21, 139.54, 135.90, 132.26, 131.18, 130.83, 130.77, 128.96, 127.87, 125.57, 125.54, 116.52, 116.35, 69.79, 29.58, 21.51, 17.06, 9.60. HRMS (ESI) m/z: $[M + Na]^+$ Calcd for $C_{28}H_{25}FN_4NaO_4$ 523.1752; Found 523.1759.

1-(4'-(5-((((3-ethylbenzyl)oxy)carbonyl)amino)-4-methyl-1H-1,2,3-triazol-1-yl)-[1,1'-biphenyl]-4-yl)cyclopropane-1-carboxylic acid (12h)

Compound **12h** was prepared in a similar way as compound **12b**, except that (3-ethylphenyl)methanol was used from step 4 in scheme 1 (yield: 70.09%). ^{1}H NMR (500 MHz, Methanol-d4) δ 7.74 (d, J = 8.2 Hz, 2H), 7.61 (d, J = 8.0 Hz, 2H), 7.54 (d, J = 8.0 Hz, 2H), 7.48

(d, J = 7.9 Hz, 2H), 7.40 – 6.91 (m, 4H), 5.13 (s, 2H), 4.55 (s, 2H), 2.27 (s, 3H), 1.61 (dd, J = 6.9, 4.0 Hz, 2H), 1.23 (dd, J = 6.8, 4.0 Hz, 2H). ^{13}C NMR (126 MHz, Methanol-d4) δ 178.53, 156.27, 143.48, 143.17, 141.64, 140.20, 139.39, 137.61, 135.85, 132.22, 131.26, 129.60, 128.93, 127.91, 127.83, 127.58, 125.58, 68.58, 64.89, 29.72, 16.99, 9.69. HRMS (ESI) m/z: [M + Na]$^+$ Calcd for $C_{28}H_{26}N_4NaO_5$ 521.1795; Found 521.1826.

2-(4'-(5-((((4-fluorobenzyl)oxy)carbonyl)amino)-4-methyl-1H-1,2,3-triazol-1-yl)-[1,1'-biphenyl]-4-yl)acetic acid (12i)

Compound **12i** was prepared in a similar way as compound **12b**, except that (4-ethylphenyl)methanol was used from step 4 in scheme 1 (yield: 86.75%). ^{1}H NMR (500 MHz, DMSO-d6) δ 9.68 (s, 1H), 7.83 (d, J = 8.5 Hz, 2H), 7.66 (d, J = 8.3 Hz, 2H), 7.58 (d, J = 8.1 Hz, 2H), 7.46 (d, J = 8.3 Hz, 2H), 7.28 (s, 4H), 5.09 (s, 2H), 4.46 (s, 2H), 2.19 (s, 3H), 1.48 (dd, J = 6.8, 3.8 Hz, 2H), 1.18 (dd, J = 6.7, 3.9 Hz, 2H). ^{13}C NMR (126 MHz, DMSO-d6) δ 175.81, 154.66, 143.01, 141.25, 140.76, 138.76, 137.61, 134.97, 131.45, 130.13, 128.19, 128.11, 126.89, 124.56, 67.00, 63.02, 28.86, 16.17, 10.08. HRMS (ESI) m/z: [M + Na]$^+$ Calcd for $C_{28}H_{26}N_4NaO_5$ 521.1795; Found 521.1790.

2-(4'-(5-((((4-fluorobenzyl)oxy)carbonyl)amino)-4-methyl-1H-1,2,3-triazol-1-yl)-[1,1'-biphenyl]-4-yl)acetic acid (14a)

Compound **14a** was prepared in a similar way as compound **12b**, except that compound **6** was used from step 5 in scheme 1 (yield: 87.72%). ^{1}H NMR (500 MHz, Methanol-d4) δ 7.77 (d, J = 8.1 Hz, 2H), 7.65 (d, J = 7.7 Hz, 2H), 7.55 (d, J = 8.1 Hz, 2H), 7.43 (d, J = 7.7 Hz, 2H), 7.34 (s, 2H), 7.03 (s, 2H), 5.11 (s, 2H), 3.68 (s, 2H), 2.27 (s, 3H). ^{13}C NMR (126 MHz, Methanol-d4) δ 175.42, 165.04, 163.09, 143.46, 140.16, 139.37, 136.22, 135.92, 131.48, 131.43, 131.17, 128.90, 128.17, 125.66, 116.36, 116.19, 67.87, 41.58, 9.64. HRMS (ESI) m/z: [M + Na]$^+$ Calcd for $C_{25}H_{21}FN_4NaO_4$ 483.1439; Found 483.1453.

3. Results and discussion

The synthesis of fluorine-containing **RO6842262** analogues **12b-12i** and **14a** was outlined in Scheme 1. 1-(4-Bromophenyl)cyclopropane-1-carbonitrile (**1**) was obtained in moderate yield following a similar approach reported previously.[17] Compound **1** was then reacted with ethylene glycol to get compound **2**, followed by esterification to get compound **3**. Compound **3** was reacted with bis(pinacolato)diboron to get compound **4** by Suzuki coupling. Similarly, compound **6** was synthesized by the reaction of commercially available ethyl 2-(4-bromophenyl)acetate (**5**) and bis(pinacolato)diboron (Scheme 1).

1-Azido-4-bromobenzene (**7**) was obtained in high yield by the reaction of 4-bromoaniline and sodium nitrite. Compound **7** was then treated with but-2-ynoic acid ethyl ester to form compound **8**, which was further reacted with sodium hydroxide to form the corresponding carboxylic acid **9**. Curtius rearrangement reaction was then performed with a benzyl alcohol building block to form the intermediate compound **10**, followed by the Suzuki-coupling reaction to afford compound **11**. Lastly, the ester on compound **11** was hydrolysed to afford the final carboxylic acid compound **12b-12i**. Compound **14a** was synthesized in a similar way, except that compound **6** was used from step 5 of Suzuki coupling (Scheme 1).

LPA$_1$ receptor activation results in the inhibition of cAMP production, and antagonizing LPA$_1$ activation induces a change in cAMP production that can be quantified by measuring luminescence. Compounds prepared in Scheme 1 were tested in an *in vitro* cAMP assay using the Chinese hamster ovary (CHO) cell line overexpressing the LPA$_1$ receptor. The endogenous ligand **18:1 LPA** was used to activate the LPA$_1$ receptor. The LPA$_1$ antagonist activity is presented as IC$_{50}$, which is the concentration needed to reverse 50% of the inhibitory effect of **18:1 LPA** on forskolin-mediated cAMP production.
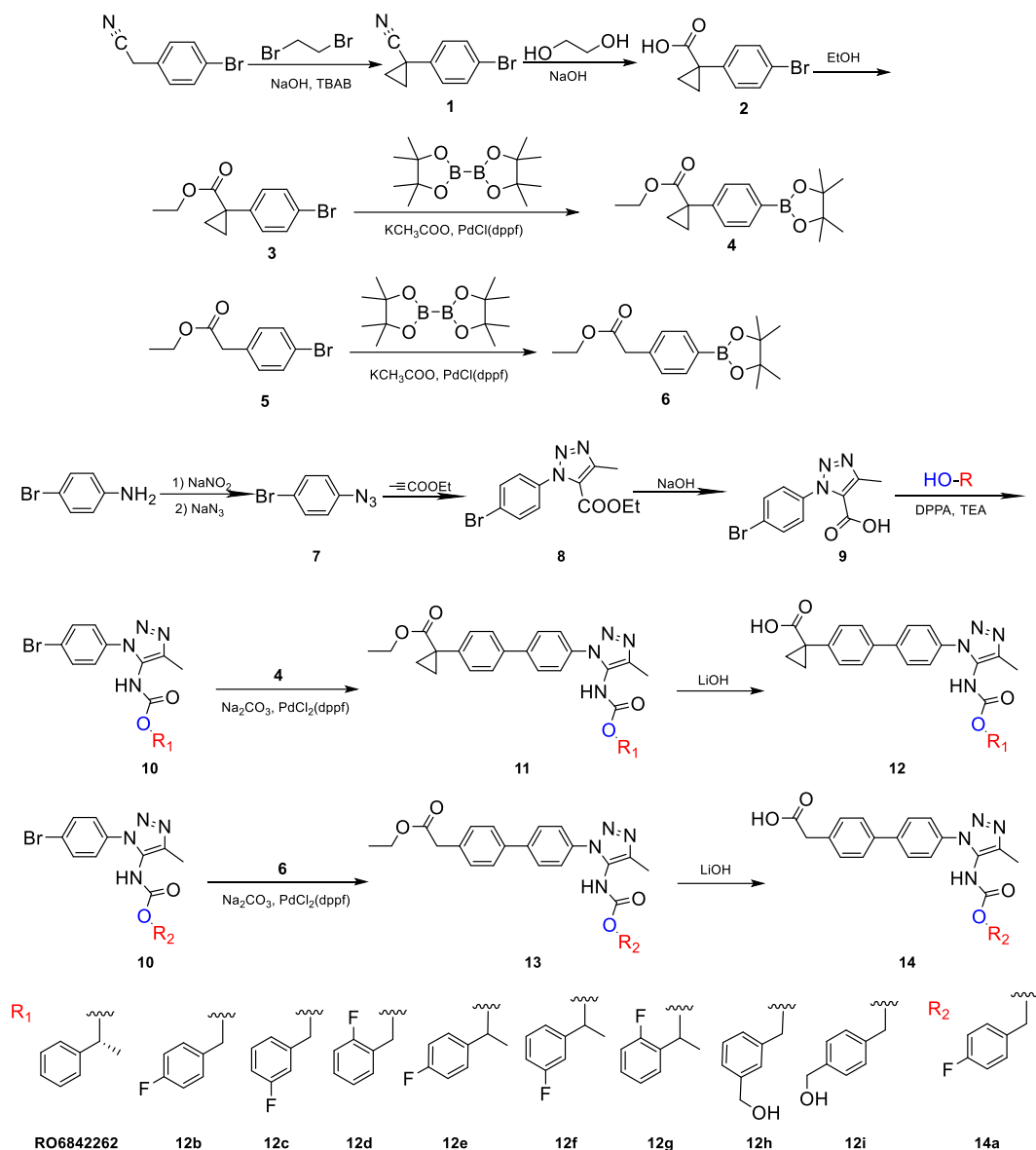
Scheme 1. Synthesis of compounds **12b-12i** and **14a**.

Firstly, three fluorine-containing compounds without the methyl group at the alpha position of the carbamate were synthesized (**12b**, **12c**, and **12d**). The results suggested that the removal of the methyl substitution resulted in a dramatic decrease in the activity (Table 2.1.1). In an attempt to extend hydrogen-bond networks, a hydroxymethyl group was introduced to replace the fluorine (**12h** and **12i**), producing a small series of benzyl alcohols. Interestingly, substituting the fluorine (**12b** and **12c**) with a hydroxymethyl group (**12i** and **12h**) resulted in a drastic increase in activity. For isoxazole-derived LPA$_1$ antagonists, the absence of the cyclopropane at the alpha position of the carboxylic acid did not affect the potency, as was reflected with **AM095**. To

explore the effect of the cyclopropane for triazole-derived LPA_1 antagonists, compound **14a** was synthesized. The result showed that removal of the cyclopropane at the alpha position of the carboxylic acid resulted in a total loss of activity.

Since the removal of the alpha methyl substituent decreased the potency, compounds with the methyl substitution were synthesized. Compounds with the methyl group (**12e**, **12f**, and **12g**) demonstrate drastically higher activity compared with those without. As shown in Table 2.1.1, among this series of compounds (**12e**, **12f**, and **12g**), the compound with meta substitution, compound **12f**, is the most potent with a IC_{50} of 16.0 nM, followed by para (IC_{50} = 93.3 nM) and ortho (IC_{50} = 192.7 nM) substitutions.

Table 2.1.1. SAR of carbamate-derived LPA_1 antagonists **12b-12i** and **14a**.
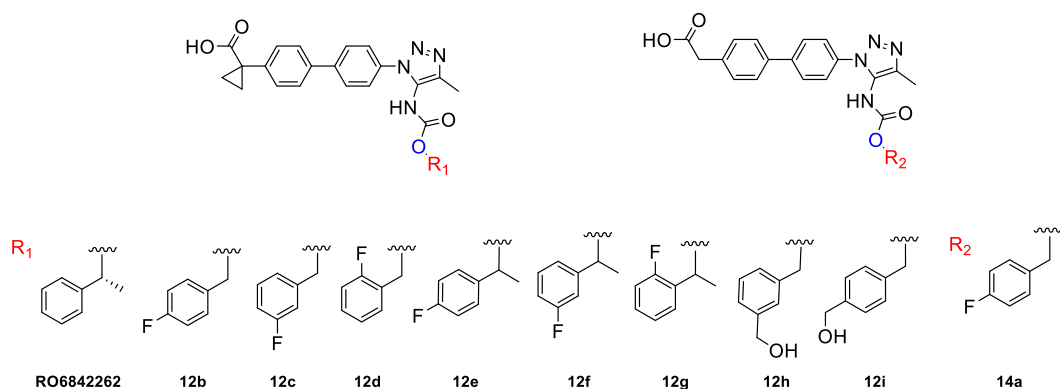
Compounds	IC_{50} (nM)	pIC_{50} [a]	% inhibition [b]
RO6842262	18.3	7.74 ± 0.05	111.60 ± 2.84
12b	3337	5.49 ± 0.10	58.43 ± 13.03
12c	2146	5.67 ± 0.06	77.08 ± 10.44
12d	1488	6.01 ± 0.42	81.92 ± 10.15
12e	93.3	7.11 ± 0.27	102.1 ± 10.95
12f	16.0	7.91 ± 0.22	109.90 ± 2.15
12g	192.7	6.73 ± 0.09	96.41 ± 5.09
12h	91.3	7.04 ± 0.05	115.80 ± 11.81
12i	624.5	6.21 ± 0.02	107.40 ± 6.48
14a	NE	NE [c]	NE

a: Values are the mean pIC_{50} ± SEM of at least two independent experiments performed in duplicate.

b: % maximal inhibition of the response to 1 μM 18:1 LPA.

c: No effect was observed at the highest concentration tested.

Molecular docking studies were performed using the X-ray crystal structure of LPA₁ (PDB ID: 4z36)[15] to investigate the binding mode of the most potent compound discovered in this study, compound **12f**, as well as **RO6842262**. Both compounds were found to bind to the orthosteric binding pocket of LPA₁ in a similar fashion, with the carboxylate oxygen forming hydrogen bonds with Gly110, Thr109, and Thr113. However, additional interactions were observed between **12f** and LPA₁. The deprotonated oxygen from the carboxylic acid of **12f** interacted with the nitrogen from Arg124 through hydrogen bonding, while the nitrogen on the triazole moiety formed hydrogen bonds with Gln125 (Figure 2.1.6). These findings could possibly explain the greater potency of **1f** compared to **RO6842262**.
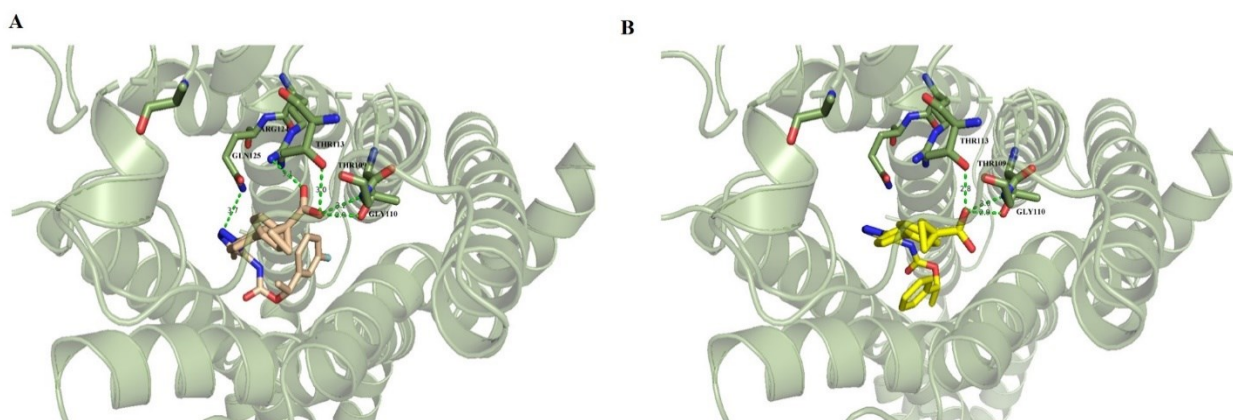
Figure 2.1.6. The binding modes of compound **12f** (A) and **RO6842262** (B) with LPA₁.

The six LPA receptor subtypes are divided into two LPA receptor families according to phylogeny: the endothelial differentiation gene (EDG) family and the non-EDG family. The LPA₁₋₃ receptors belong to the EDG family with 50–60% similarity in amino acid sequence.[19] Studies have shown that LPA₂ and LPA₃ signaling are involved in cell function and survival. Both LPA₂ and LPA₃ signaling maintain vascular homeostasis and improve heart function after ischemic injuries.[20-21] It is important to develop potent and highly selective LPA₁ antagonists to reduce the potential off target effects. As two of the most potent compounds from the series, compounds **12f** and **12h** were selected for further testing for their selectivity against the LPA₂₋₃ receptors. As shown in Figure 2.1.7, both of the compounds showed no significant activity for the LPA₂₋₃ receptors, suggesting decent LPA₁ selectivity (Table 2.1.2).
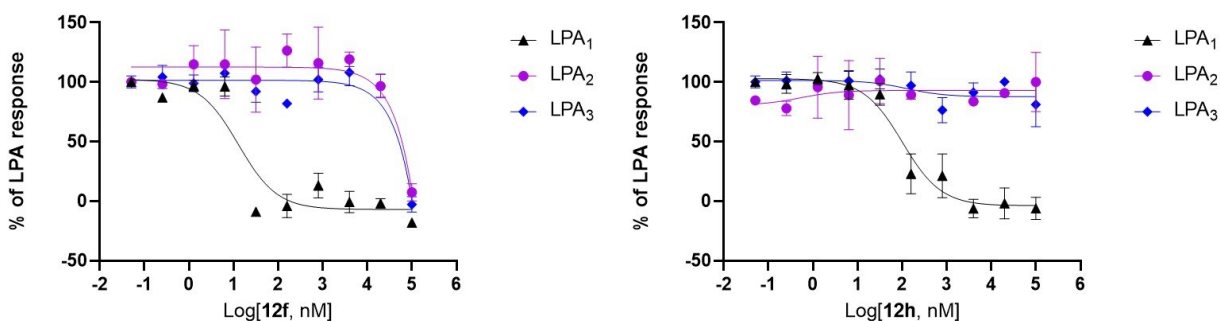
Figure 2.1.7. Dose-response curves of selected compounds **12f** (left) and **12h** (right) towards the LPA$_1$, LPA$_2$, and LPA$_3$ receptors.

Table 2.1.2. Activity of compound **12f** and **12h** towards the LPA$_1$, LPA$_2$, and LPA$_3$ receptors.

Compounds	LPA$_1$ [a] (nM)	LPA$_2$ [a] (nM)	LPA$_3$ [a] (nM)
12f	16.0	> 10000	> 10000
12h	91.3	NE [b]	NE

a: Values are the mean IC$_{50}$ of at least two independent experiments performed in duplicate.

b: No effect was observed at the highest concentration tested.

The most potent compound (**12f**) in this series was then further studied to determine its therapeutic effects in breast cancer. MDA-MB-231 is a commonly used breast cancer cell line and has the highest expression of LPA$_1$ compared with other breast cancer cell lines,[3] so MDA-MB-231 was selected in this study. In order to evaluate the effect of LPA and the compound **12f** on the survival, migration, and invasion of MDA-MB-231, several *in vitro* assays were performed.

Wound healing assay is one of the most common methods to study cell migration, which is based on the observation of cell migration into a wound created on a cell monolayer.[22] The MDA-MB-231 cells were treated with various concentrations of **18:1 LPA** to evaluate its effect on breast cancer cell migration using the wound healing assay. This study showed that stimulation with LPA enhanced cell migratory ability dose-dependently (Figure 2.1.8 A). LPA was able to induce cell migration at concentrations as low as 1.25 µM, with the maximum effect at 20 µM. Following 10 µM LPA stimulation, LPA$_1$ antagonist **12f** impeded cell migration in a dose-dependent manner as shown in Figure 2.1.8 B. At a concentration as low as 1 µM, compound **12f**

exhibited the capability to decrease LPA-induced migration by 27.67%. The highest concentration tested, 20 µM, demonstrated the largest inhibitory effect. When 20 µM of compound **12f** was added, it resulted in a remarkable decrease of LPA-induced migration by 56.39%.

Cancer cells can be stimulated to proliferate and migrate by serum. The anti-migratory effect of compound **12f** was further investigated using serum. The findings revealed that serum was more effective in promoting cell migration compared to LPA alone. The addition of 10% serum resulted in complete wound closure within 48 hours (Figure 2.1.8 C). This outcome was expected, as serum contains multiple growth factors and mitogens in addition to LPA, contributing to enhanced migration.

The impact of compound **12f** on serum-induced cell migration was examined next (Figure 2.1.8 C). The results showed that compound **12f** exerted an inhibitory effect on serum-induced cell migration. Notably, this inhibitory effect was observed at a concentration of 5 µM, resulting in a 29.08% inhibition of serum-induced migration. Moreover, the maximum effect was observed at a concentration of 20 µM of compound **12f**, leading to a substantial inhibition of serum-induced migration by 49.45%. These findings highlight the dose-dependent anti-migratory effect of compound **12f** in the context of serum-induced cell migration.
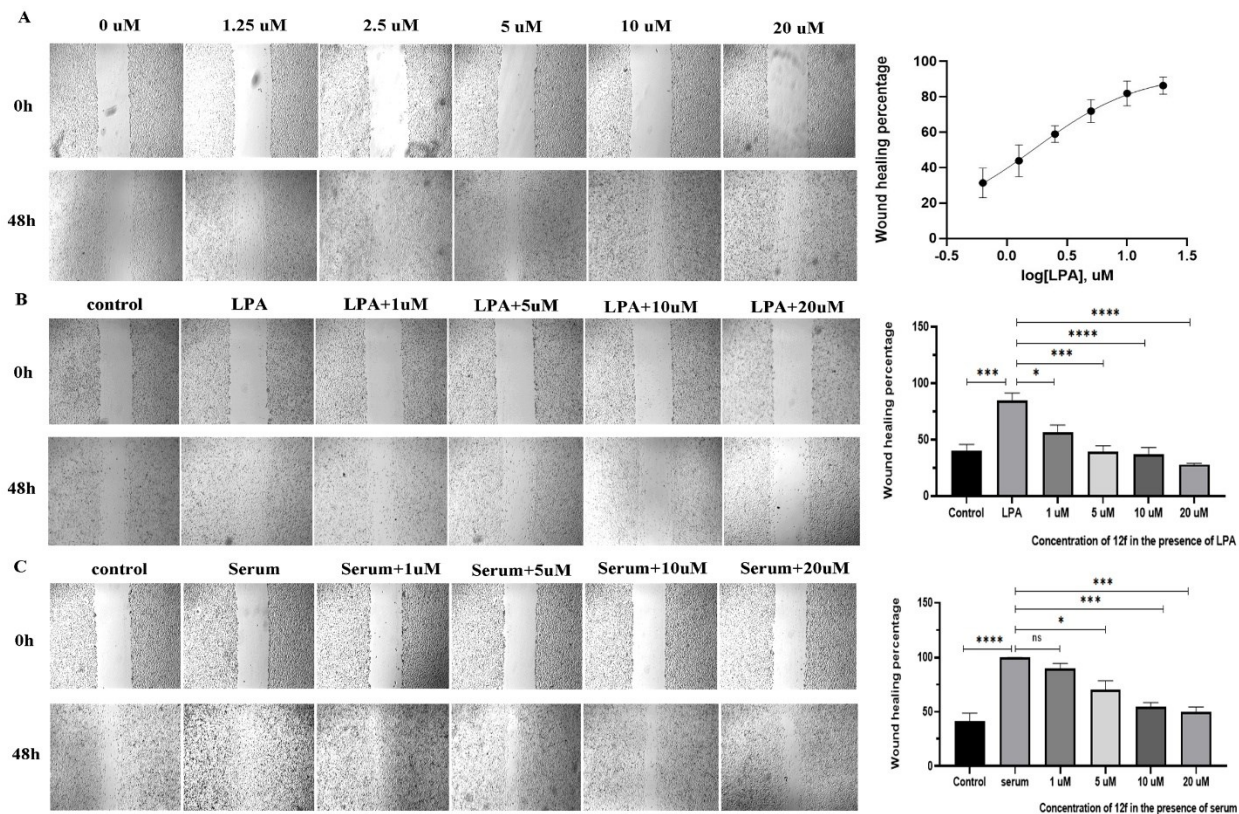
Figure 2.1.8. Effects of LPA, serum and **12f** on cell migration in wound healing assays using MDA-MB-231 cells. A. Effects of various concentrations of LPA on cell migration (left) and its quantification (right). B. Effects of 10 μM LPA and various concentrations of **12f** on cell migration (left) and its quantification (right). C. Effects of 10% serum and various concentrations of **12f** on cell migration (left) and its quantification (right). * $P < 0.05$, *** $P < 0.001$, **** $P < 0.0001$, ns $P > 0.05$.

Transwell assay is based on a two-chamber system separated by a membrane pore through which cells migrate in vertical direction. Transwell migration and invasion assays were introduced in this study to further confirm the effect of LPA and **12f** on cell migratory abilities. As shown in Figure 2.1.9, it was suggested that LPA induced both cell migration and invasion in MDA-MB-231 cells, which was hindered by **12f** dose-dependently. Taken together, these results suggested that LPA promoted breast cancer cell migration, while the LPA antagonist **12f** inhibited cell migration induced by LPA. In addition, **12f** inhibited serum-induced breast cancer cell migration in the absence of LPA.
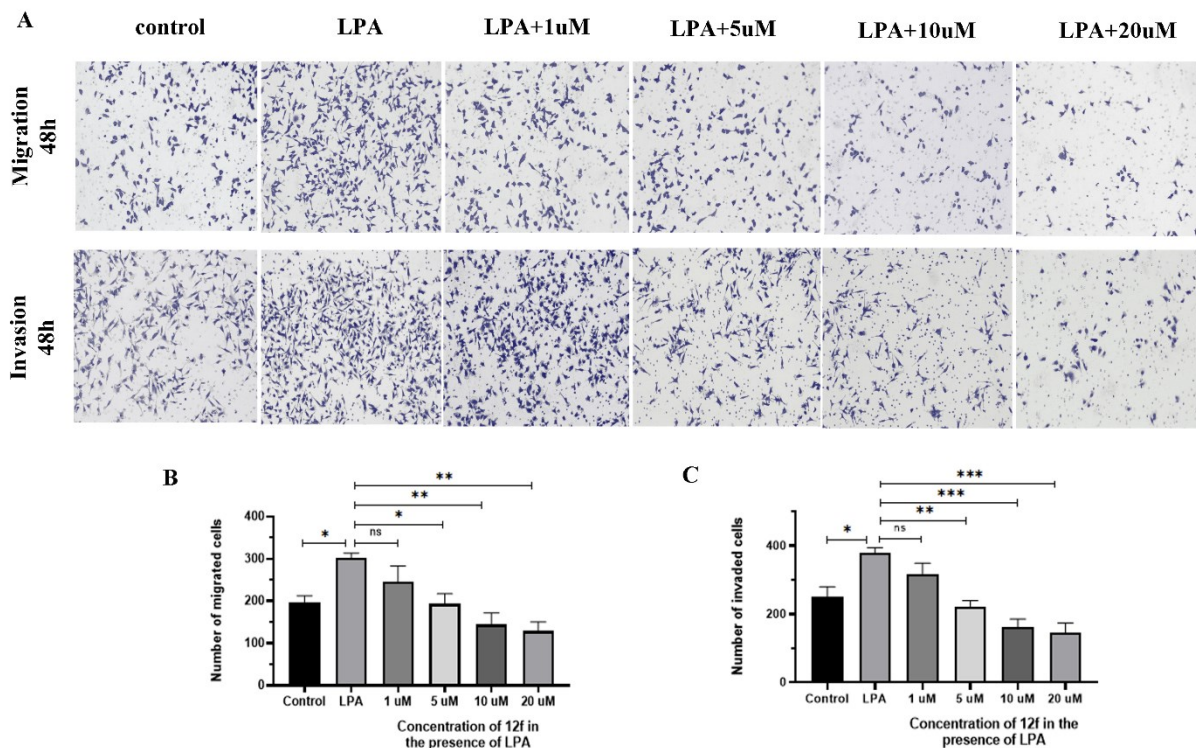
Figure 2.1.9. Effects of 10 µM LPA and various concentrations of **12f** on cell migration and invasion in MDA-MB-231 cells. A. Effects of LPA and **12f** on cell migration and invasion by Transwell assays in MDA-MB-231 cells. B-C. Quantification of Transwell migration (B) and Transwell invasion (C) assays. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$, ns $P > 0.05$.

Colony formation assay is a cell survival assay based on the ability of a single cell to grow into a colony. It was performed to evaluate cell survival.[23] The experimental findings revealed that the presence of 10 µM LPA resulted in enhanced survival of MDA-MB-231 cells. However, when compound **12f** was introduced, it displayed a dose-dependent inhibitory effect on LPA-induced cell survival. Notably, the minimum effective concentration of **12f** was found to be 5 µM, indicating a significant reduction in LPA-induced MDA-MB-231 cell survival (Figure 2.1.10).
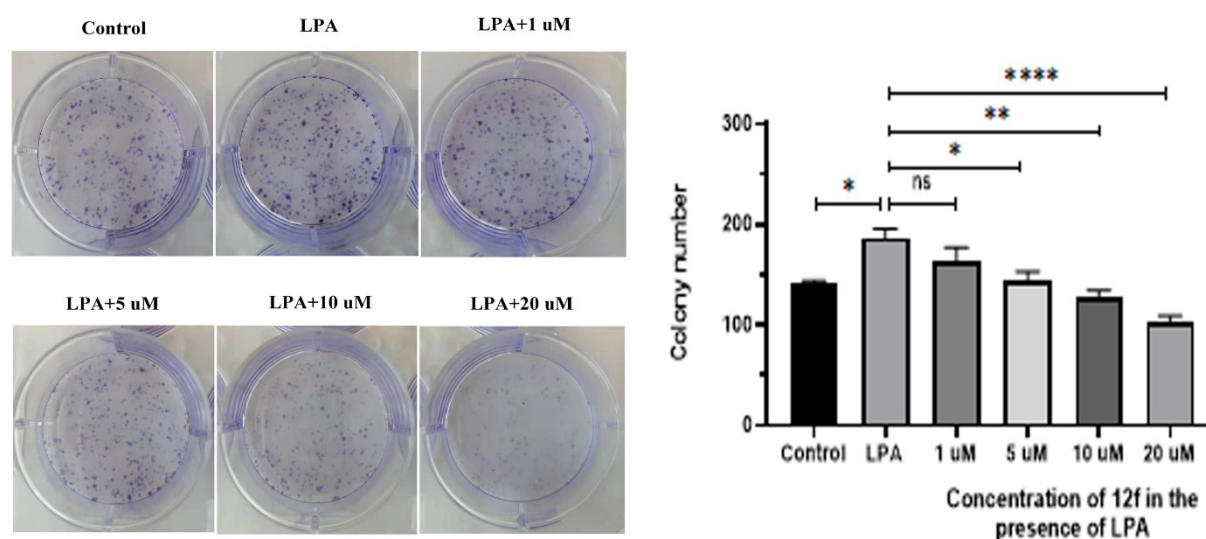
Figure 2.1.10. Effects of 10 µM LPA and various concentrations of **12f** on cell survival in colony formation assays using MDA-MB-231 cells. * $P < 0.05$, ** $P < 0.01$, **** $P < 0.0001$, ns $P > 0.05$.

LPA_1 has been implicated in the initiation and progression of various cancer types, such as renal cancer,[24] prostate cancer,[25] and gastrointestinal cancers.[26] It exerts its effects by enhancing tumor-promoting activities and interacting with other receptors, including the epidermal growth factor receptor (EGFR) and other relevant targets.[2] In addition to investigating the role of LPA_1 in MDA-MB-231, the analysis was extended to several other cancer cell lines, including the human cervical cancer cell line HeLa and the human ovarian cancer cell line OVCAR8, aiming to explore the potential involvement of LPA_1 in different cancer types. The wound healing assay and colony formation assay were conducted. The results suggested that LPA is not involved in cell survival and migration in either HeLa or OVCAR8. Additionally, adding the compound **12f** after LPA stimulation did not have any impact on cell survival or migration (Figure 2.1.11).
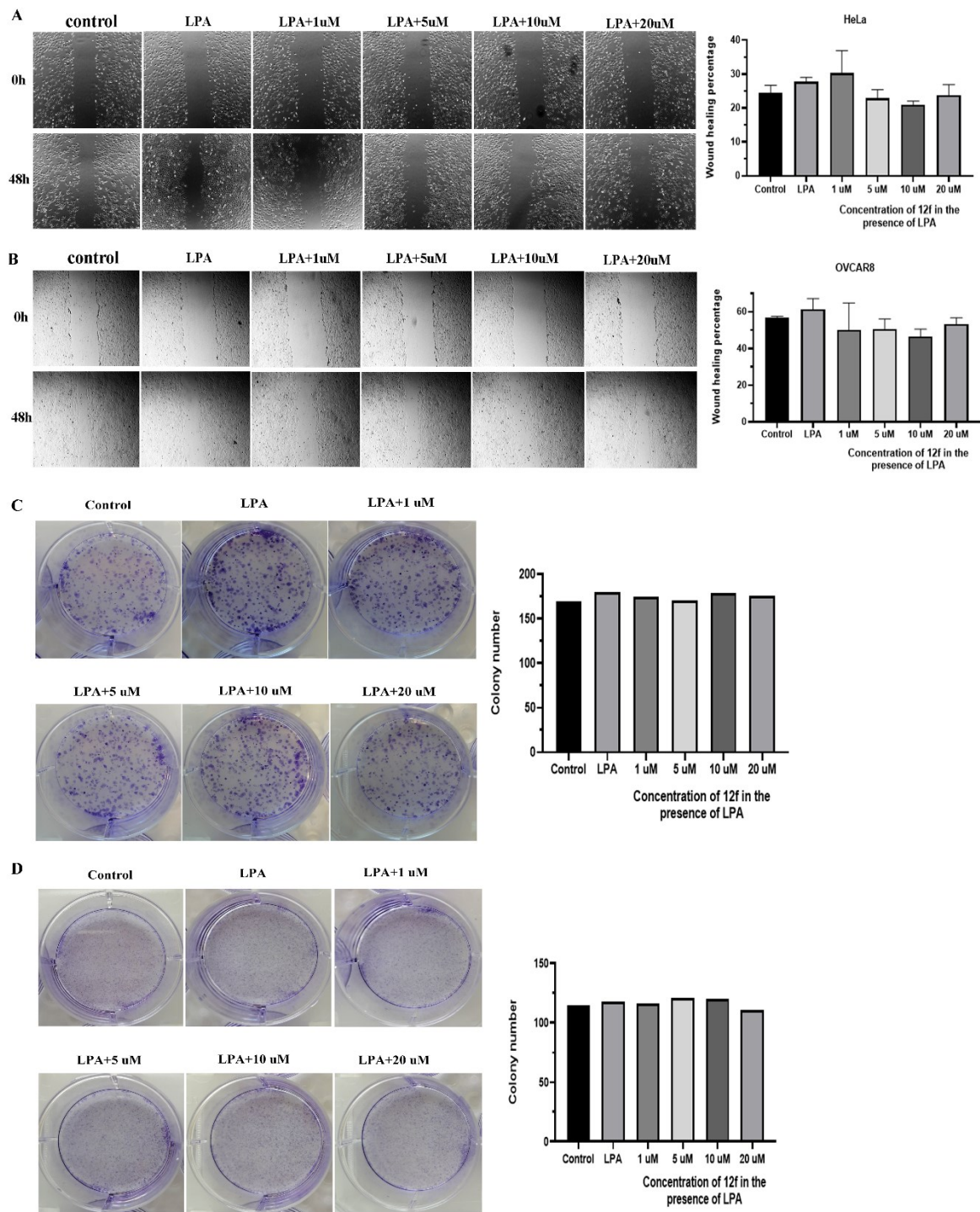
Figure 2.1.11. Wound healing (A and B) and colony formation assays (C and D) using the HeLa (A and C) and OVCAR8 (B and D) cell lines. LPA and LPA$_1$ antagonist **12f** did not affect HeLa and OVCAR8 cell survival and migration.

Cancer is a leading cause of death and a significant obstacle to increasing life expectancy worldwide. In 2020, there were proximately 19.3 million new cases of cancer and 10.0 million deaths attributed to the disease. Among these cases, breast cancer stands out as the most commonly diagnosed cancer in the female population.[27] Current breast cancer chemotherapies aim to target multiple receptors such as estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2). However, these treatments often come with adverse effects, and the emergence of drug resistance remains a significant concern.[28] By identifying alternative targets, the development of more effective and personalized treatment strategies that address the limitations of the current therapies becomes possible. Studies suggest that targeting LPA_1 for the treatment of breast cancer holds potential, given its involvement in the initiation and progress of breast cancer.[29-30] The findings of this study are in line with previously reported research,[3-7] demonstrating that the activation of the LPA_1 receptor promotes cell migration, and invasion in breast cancer. Through the utilization of the most potent LPA_1 antagonist discovered in this study, the pro-proliferative and pro-migratory effects of LPA in breast cancer can be effectively inhibited. In addition, several other cancer cell lines were tested, and the results indicated that neither LPA or the LPA_1 antagonist played a role in their cell survival and migration These findings imply that the LPA_1 receptor does not function as a universal marker across all cancer types; rather, its involvement appears to be specific to certain cancer types.

The current study broadened the collection of triazole-derived LPA_1 antagonists and validated their function across several cancer cell lines. Moreover, the successful integration of fluorine into these compounds without compromising their activity opens the door for their potential development into PET tracers. A noteworthy limitation of this study is the racemic nature of these compounds. Despite their decent activity, evaluating both enantiomers of these racemic compounds would prove valuable. Future studies will be essential to investigate the enantiomeric SAR trend.

4. Conclusion

LPA_1 plays an important role in breast caner. The present study examined SAR on a series of fluorine-containing triazole derivatives and evaluated their potential as potent and selective LPA_1 antagonists. Among them, compound **12f** demonstrated the highest LPA_1 antagonistic activity

and demonstrated excellent inhibitory activity of LPA-induced cell survival, migration, and invasion of a breast cancer cell line. The LPA$_1$ antagonists synthesized in this study have potential to be developed into novel therapies for breast cancer. Moreover, these fluorine-containing compounds could be further developed into ^{18}F-labeled PET imaging agents for use as radiotracers to facilitate breast cancer diagnosis and staging.

5. Reference

1.   Liu, W.; Hopkins, A. M.; Hou, J., The development of modulators for lysophosphatidic acid receptors: A comprehensive review. *Bioorg Chem* **2021,** *117*, 105386.

2.   Xu, Y., Targeting lysophosphatidic acid in cancer: the issues in moving from bench to bedside. *Cancers (Basel)* **2019,** *11* (10), 1523.

3.   Li, T. T.; Alemayehu, M.; Aziziyeh, A. I.; Pape, C.; Pampillo, M.; Postovit, L.-M.; Mills, G. B.; Babwah, A. V.; Bhattacharya, M., Beta-arrestin/Ral signaling regulates lysophosphatidic acid-mediated migration and invasion of human breast tumor cells. *Mol Cancer Res* **2009,** *7* (7), 1064-77.

4.   Hopkins, M. M.; Zhang, Z.; Liu, Z.; Meier, K. E., Eicosopentaneoic Acid and Other Free Fatty Acid Receptor Agonists Inhibit Lysophosphatidic Acid- and Epidermal Growth Factor-Induced Proliferation of Human Breast Cancer Cells. *J Clin Med* **2016,** *5* (2), 16.

5.   Boucharaba, A.; Serre, C.-M.; Guglielmi, J.; Bordet, J.-C.; Clézardin, P.; Peyruchaud, O., The type 1 lysophosphatidic acid receptor is a target for therapy in bone metastases. *Proc Natl Acad Sci U S A* **2006,** *103* (25), 9643-8.

6.   Boucharaba, A.; Serre, C.-M.; Grès, S.; Saulnier-Blache, J. S.; Bordet, J.-C.; Guglielmi, J.; Clézardin, P.; Peyruchaud, O., Platelet-derived lysophosphatidic acid supports the progression of osteolytic bone metastases in breast cancer. *J Clin Invest* **2004,** *114* (12), 1714-1725.

7.   Marshall, J.-C. A.; Collins, J. W.; Nakayama, J.; Horak, C. E.; Liewehr, D. J.; Steinberg, S. M.; Albaugh, M.; Vidal-Vanaclocha, F.; Palmieri, D.; Barbier, M.; Murone, M.; Steeg, P. S., Effect of inhibition of the lysophosphatidic acid receptor 1 on metastasis and metastatic dormancy in breast cancer. *J Natl Cancer Inst* **2012,** `04* (17), 1306-19.

8.   Ohta, H.; Sato, K.; Murata, N.; Damirin, A.; Malchinkhuu, E.; Kon, J.; Kimura, T.; Tobo, M.; Yamazaki, Y.; Watanabe, T.; Yagi, M.; Sato, M.; Suzuki, R.; Murooka, H.; Sakai, T.; Nishitoba, T.; Im, D.-S.; Nochi, H.; Tamoto, K.; Tomura, H.; Okajima, F., Ki16425, a subtype-selective antagonist for EDG-family lysophosphatidic acid receptors. *Mol Pharmacol* **2003,** *64*, 994–1005.

9.   ClinicalTrials.gov:NCT01766817. Https://Clinicaltrials.Gov/Ct2/Show/NCT01766817.

10.   Gallezot, J. D.; Nabulsi, N. B.; Holden, D.; Lin, S. F.; Labaree, D.; Ropchan, J.; Najafzadeh, S.; Donnelly, D. J.; Cao, K.; Bonacorsi, S.; Seiders, J.; Roppe, J.; Hayes, W.;

Huang, Y.; Du, S.; Carson, R. E., Evaluation of the lysophosphatidic acid receptor type 1 radioligand (11)C-BMT-136088 for lung imaging in rhesus monkeys. *J Nucl Med* **2018,** *59* (2), 327-333.

11. ClinicalTrials.gov:NCT02017730. Https://Clinicaltrials.Gov/Ct2/Show/NCT02017730.

12. Qian, Y.; Hamilton, M.; Sidduri, A.; Gabriel, S.; Ren, Y.; Peng, R.; Kondru, R.; Narayanan, A.; Truitt, T.; Hamid, R.; Chen, Y.; Zhang, L.; Fretland, A. J.; Sanchez, R. A.; Chang, K. C.; Lucas, M.; Schoenfeld, R. C.; Laine, D.; Fuentes, M. E.; Stevenson, C. S.; Budd, D. C., Discovery of highly selective and orally active lysophosphatidic acid receptor-1 antagonists with potent activity on human lung fibroblasts. *J Med Chem* **2012,** *55* (17), 7920-7939.

13. Gillis, E. P.; Eastman, K. J.; Hill, M. D.; Donnelly, D. J.; Meanwell, N. A., Applications of Fluorine in Medicinal Chemistry. *J Med Chem* **2015,** *58* (21), 8315-59.

14. Müller, K.; Faeh, C.; Diederich, F., Fluorine in Pharmaceuticals: Looking Beyond Intuition. *Science* **2007,** *317* (5846), 1881-6.

15. Chrencik, J. E.; Roth, C. B.; Terakado, M.; Kurata, H.; Omi, R.; Kihara, Y.; Warshaviak, D.; Nakade, S.; Asmar-Rovira, G.; Mileni, M.; Mizuno, H.; Griffith, M. T.; Rodgers, C.; Han, G. W.; Velasquez, J.; Chun, J.; Stevens, R. C.; Hanson, M. A., Crystal structure of antagonist bound human lysophosphatidic acid receptor 1. *Cell* **2015,** *161* (7), 1633-1643.

16. Forli, S.; Huey, R.; Pique, M. E.; Sanner, M. F.; Goodsell, D. S.; Olson, A. J., Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nature Protocols* **2016,** *11*, 905-919.

17. Buckman, B. O.; Nicholas, J. B.; Emayan, K.; Seiwert, S. D. Lysophosphatidic acid receptor antagonists. *WO2013025733A1* **2013**.

18. Buckman, B. O.; Nicholas, J. B.; Emayan, K.; Seiwert, S. D.; Yuan, S. Lysophosphatidic acid receptor antagonists. *WO2014113485A1* **2014**.

19. Yung, Y. C.; Stoddard, N. C.; Chun, J., LPA receptor signaling: pharmacology, physiology, and pathophysiology. *J Lipid Res* **2014,** *55* (7), 1192-1214.

20. Pei, J.; Cai, L.; Wang, F.; Xu, C.; Pei, S.; Guo, H.; Sun, X.; Chun, J.; Cong, X.; Zhu, W.; Zheng, Z.; Chen, X., LPA2 Contributes to Vascular Endothelium Homeostasis and Cardiac Remodeling After Myocardial Infarction. *Circ Res* **2022,** *131* (5), 388-403.

21. Wang, F.; Liu, S.; Pei, J.; Cai, L.; Liu, N.; Liang, T.; Dong, X.; Cong, X.; Chun, J.; Chen, J.; Hu, S.; Chen, X., LPA3-mediated lysophosphatidic acid signaling promotes postnatal heart regeneration in mice. *Theranostics* **2020,** *10* (24), 10892-10907.

22. Rodriguez, L. G.; Wu, X.; Guan, J.-L., Wound-healing assay. *Methods Mol Biol* **2005,** *294*, 23-9.

23. Franken, N. A. P.; Rodermond, H. M.; Stap, J.; Haveman, J.; Bree, C. v., Clonogenic assay of cells in vitro. *Nat Protoc* **2006,** *1* (5), 2315-9.

24. Hashimoto, S.; Mikami, S.; Sugino, H.; Yoshikawa, A.; Hashimoto, A.; Onodera, Y.; Furukawa, S.; Handa, H.; Oikawa, T.; Okada, Y.; Oya, M.; Sabe, H., Lysophosphatidic acid activates Arf6 to promote the mesenchymal malignancy of renal cancer. *Nat Commun* **2016,** *7*, 10656.

25. Hopkins, M. M.; Liu, Z.; Meier, K. E., Positive and Negative Cross-Talk between Lysophosphatidic Acid Receptor 1, Free Fatty Acid Receptor 4, and Epidermal Growth Factor Receptor in Human Prostate Cancer Cells. *J Pharmacol Exp Ther* **2016,** *359* (1), 124-133.

26. Tveteraas, I. H.; Aasrum, M.; Brusevold, I. J.; Ødegård, J.; Christoffersen, T.; Sandnes, D., Lysophosphatidic acid induces both EGFR-dependent and EGFR-independent effects on DNA synthesis and migration in pancreatic and colorectal carcinoma cells. *Tumour Biol* **2016,** *37* (2), 2519-2526.

27. Sung, H.; Ferlay, J.; Siegel, R. L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F., Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **2021,** *71* (3), 209-249.

28. Burguin, A.; Diorio, C.; Durocher, F., Breast Cancer Treatments: Updates and New Challenges. *J Pers Med* **2021,** *11* (8), 808.

29. Geraldo, L. H. M.; Spohr, T. C. L. d. S.; Amaral, R. F. d.; Fonseca, A. C. C. d.; Garcia, C.; Mendes, F. d. A.; Freitas, C.; dosSantos, M. F.; Lima, F. R. S., Role of lysophosphatidic acid and its receptors in health and disease: novel therapeutic strategies. *Signal Transduct Target Ther* **2021,** *6* (1), 45.

30. Mills, G. B.; Moolenaar, W. H., The emerging role of lysophosphatidic acid in cancer. *Nat Rev Cancer* **2003,** *3* (8), 582-591.

Part two

Design, synthesis and biological evaluation of urea-derived small molecules targeting lysophosphatidic acid receptor 1 for the treatment of breast cancer

Abstract:

The discovery of fluorine-containing carbamate-derived small molecules as potent and selective LPA_1 antagonists was previously reported. To further expand the library of LPA_1 antagonists and potentially enhance their activity, a urea moiety was introduced in replacement of the carbamate group and a second series of LPA_1 antagonists based on a urea scaffold were synthesized and evaluated. Within this series, several compounds exhibited potent LPA_1 antagonism. Notably, compound **17f** emerged as the most active, with an IC_{50} of 215.2 nM in the cAMP assay. Biologically, compound **17f** demonstrated the ability to block LPA-induced cell migration and invasion in the breast cancer cell line MDA-MB-231. The development of these urea-derived LPA_1 antagonists in this study has expanded the repertoire of LPA_1 antagonists and holds promising potential for the development of a novel therapy for breast cancer.

Graphical abstract:

Compound 17f

LPA$_1$ IC$_{50}$ = 215.2 nM
LPA$_2$ IC$_{50}$ > 10.0 µM
LPA$_3$ IC$_{50}$ > 10.0 µM
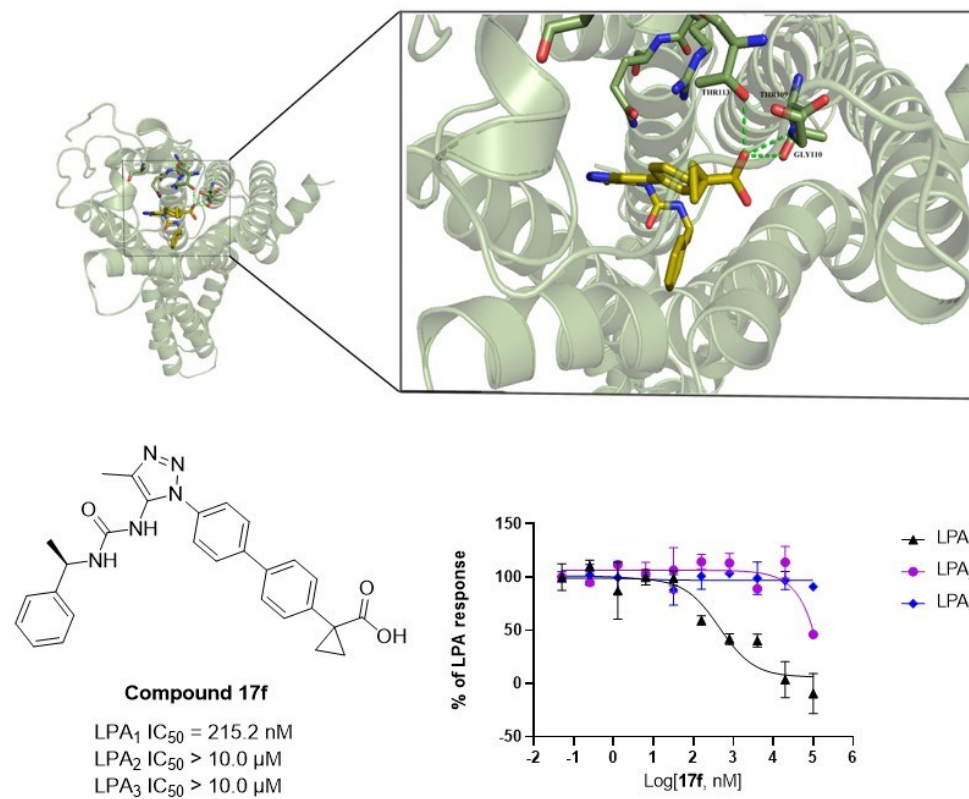
Key words: lysophosphatidic acid receptor 1, breast cancer, urea

1. Introduction

Lysophosphatidic acid receptor 1, known as LPA$_1$, belongs to the family of six G protein-coupled receptors (LPA$_{1-6}$). LPA$_1$ is activated by its endogenous ligand, lysophosphatidic acid (LPA), leading to a diverse range of cellular responses, such as cell proliferation, migration, and survival in various cell types.[1] In recent years, LPA$_1$ has emerged as a promising therapeutic target for breast cancer. Aberrant mRNA expression of LPA$_1$ has been detected in several breast cancer cell lines.[2] In breast cancer mouse models, overexpression of LPA$_1$ promoted the formation of bone metastasis,[3] while pharmacological inhibition of LPA$_1$ using an LPA$_1$ inhibitor blocked the progression of bone metastasis.[4] Furthermore, in the 4T1 spontaneous metastasis mouse model, inhibition of LPA$_1$ had no effect on the growth of breast primary tumors but effectively suppressed spontaneous metastasis.[5, 6] These studies strongly suggest that LPA$_1$ is closely associated with breast cancer metastasis. However, the current availability of LPA$_1$ antagonists developed for the treatment of cancer remains very limited.

In order to address the aforementioned challenges, the approach was initiated by leveraging a potent and selective LPA$_1$ antagonist known as **RO6842262** (Figure 2.2.1), originally identified by Roche for the treatment of lung fibrosis.[7] The carbamate within this compound was converted into a urea, leading to the discovery of a range of compounds with a urea moiety (Figure 2.2.1). The core focus of the endeavor was to evaluate the potential enhancement in functional activity achieved by substituting the hydrogen bond acceptor (oxygen, O) with a hydrogen bond donor (nitrogen, NH). This motivation stemmed from the recognition that urea compounds have the capacity to establish multiple stable hydrogen bonds with receptor targets. These intricate ligand-receptor interactions significantly influence specific biological activities, drug actions, and overall drug characteristics.[8, 9] Within this study, incorporation of a urea moiety was readily achieved through the transformation of the benzyl alcohol building block into its corresponding benzylamine counterpart (Scheme 2). As a result, the substitution of O with NH was performed in an endeavor to enhance the ligand-receptor interaction. Furthermore, urea is generally considered to be more stable than carbamate.

A total of eight urea-derived LPA$_1$ antagonists were synthesized and evaluated. Several compounds with high potency and selectivity were discovered in this study. Particularly, compound **17f** was the most potent in this study with its IC$_{50}$ being 215.2 nM in the cAMP assay.

It inhibited LPA-induced migration, and invasion in the breast cancer cell line MDA-MB-231 in a dose-dependent manner.
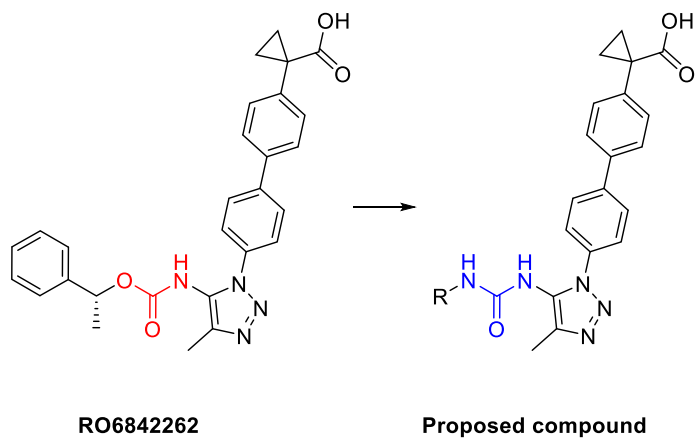
Figure 2.2.1. Structural modification of **RO6842262** leading to urea-derived LPA$_1$ antagonists.

2. Materials and methods

2.1 Synthesis

All solvents and chemicals were of reagent grade and purchased from commercial manufacturers. The purity and characterization of compounds were ascertained through a combination of techniques including HPLC, TLC, high-resolution mass spectrometry (HRMS), as well as ^{1}H NMR and ^{13}C NMR analyses. Spectral data was compared with reference data from the literature where possible; for **RO6842262**, ^{1}H NMR was compared.[7] Experimental data and reported data compared favourably in all cases.[7,10-11]

*1-(1-(4-bromophenyl)-4-methyl-1H-1,2,3-triazol-5-yl)-3-(pyridin-2-ylmethyl)urea (**15**).*

To a solution of compound **9** (250 mg, 0.89 mmol) in anhydrous toluene (6.25 mL) was added pyridin-2-ylmethanamine (143.91 mg, 1.33 mmol), triethylamine (150 mg), and diphenyl phosphoryl azide (316.5 mg, 1.15 mmol). The reaction mixture was heated to 120°C for 10 hours. The mixture was diluted with EtOAc, washed with brine, dried over Na_2SO_4, filtered, and concentrated. The residue was purified by column chromatography (Hexane: EtOAc = 2:1) to give compound **15** (150 mg, yield: 43.00%). ^{1}H NMR (500 MHz, Chloroform-d) δ 8.46 (d, J = 4.7 Hz, 1H), 7.69 (t, J = 7.7 Hz, 1H), 7.60 (d, J = 7.9 Hz, 2H), 7.50 (s, 1H), 7.40 (d, J = 7.9 Hz, 2H), 7.24 (d, J = 8.2 Hz, 2H), 4.65 (d, J = 4.5 Hz, 2H), 2.63 (s, 3H). ^{13}C NMR (126 MHz, Chloroform-d) δ 158.38, 154.66, 148.83, 144.26, 136.93, 135.55, 132.38, 129.25, 126.37, 123.65, 122.67, 122.03, 44.20, 11.50.

*ethyl 1-(4'-(4-methyl-5-(3-(pyridin-2-ylmethyl)ureido)-1H-1,2,3-triazol-1-yl)-[1,1'-biphenyl]-4-yl)cyclopropane-1-carboxylate (**16**).*

To a stirred mixture of compound **15** (150 mg, 0.39 mmol), compound **4** (146.9 mg, 0.46 mmol), and sodium carbonate (86.56 mg, 0.82 mmol) in dioxane (6 mL) and water (2 mL) was added $PdCl_2$(dppf) (28.32 mg, 0.039 mmol). The mixture was purged with argon and stirred at 80 °C under argon for 6 hours. Following this, the mixture was filtered through diatomaceous earth and partitioned between EtOAc and water. The organic phase was extracted with EtOAc, washed with brine, dried over Na_2SO_4, filtered, and concentrated. The residue was purified by column chromatography (DCM: MeOH = 30:1) to give compound **16** (125 mg, yield: 65.00%). ^{1}H NMR (500 MHz, Chloroform-d) δ 8.37 (d, J = 4.7 Hz, 1H), 7.64 (t, J = 9.6 Hz, 4H), 7.56 (d, J = 7.8 Hz,

2H), 7.52 (d, J = 7.7 Hz, 2H), 7.44 (d, J = 7.6 Hz, 2H), 7.22 (d, J = 7.8 Hz, 1H), 7.15 (t, J = 6.3 Hz, 1H), 4.64 (d, J = 4.7 Hz, 2H), 4.12 (q, J = 7.1 Hz, 2H), 2.61 (s, 3H), 1.64 (br s, 2H), 1.25 – 1.15 (m, 5H). ^{13}C NMR (126 MHz, Chloroform-d) δ 174.33, 158.63, 154.76, 148.84, 144.49, 142.27, 139.54, 138.36, 136.85, 135.59, 131.03, 129.21, 127.83, 126.88, 125.19, 122.57, 121.97, 61.06, 44.27, 28.80, 16.50, 14.18, 11.59.

1-(4'-(4-methyl-5-(3-(pyridin-2-ylmethyl)ureido)-1H-1,2,3-triazol-1-yl)-[1,1'-biphenyl]-4-yl)cyclopropane-1-carboxylic acid (17a)

To a stirred solution of compound **16** (125 mg, 0.25 mmol) in THF (9.9 mL) and water (3.3 mL) was added lithium hydroxide (30.81 mg, 1.28 mmol). The reaction mixture was stirred at rt for 24 hours. The reaction mixture was cooled down to 0°C and neutralized with conc. HCl to pH = 2. The mixture was extracted with EtOAc. The combined organic phase was dried over Na$_2$SO$_4$, filtered, and concentrated. The residue was purified by column chromatography (Hexane: EtOAc = 1:1) to give compound **17a** (97 mg, yield: 78.00%). ^{1}H NMR (500 MHz, Methanol-d4) δ 8.50 (d, J = 4.7 Hz, 1H), 7.77 (dd, J = 18.3, 8.1 Hz, 3H), 7.64 (t, J = 7.0 Hz, 4H), 7.51 (d, J = 7.6 Hz, 2H), 7.35 – 7.26 (m, 2H), 4.64 (s, 2H), 2.53 (s, 3H), 1.64 (br s, 2H), 1.27 (br s, 2H). ^{13}C NMR (126 MHz, Methanol-d4) δ 178.33, 161.47, 158.17, 150.03, 144.40, 143.72, 141.54, 139.47, 138.75, 136.80, 132.28, 131.58, 128.88, 127.85, 125.80, 124.02, 123.29, 45.86, 29.61, 17.04, 10.65. ESI-MS m/z calcd C$_{26}$H$_{24}$N$_6$O$_3$ [M + H]$^+$ 469.20, found 468.63.

1-(4'-(5-(3-(4-fluorobenzyl)ureido)-4-methyl-1H-1,2,3-triazol-1-yl)-[1,1'-biphenyl]-4-yl)cyclopropane-1-carboxylic acid (17b)

Compound **17b** was prepared in a similar way as compound **17a**, except that (4-fluorophenyl)methanamine was used from step 4 in scheme 1 (yield: 73.53%). ^{1}H NMR (500 MHz, Methanol-d4) δ 7.78 (d, J = 8.6 Hz, 2H), 7.61 (dd, J = 10.8, 8.4 Hz, 4H), 7.48 (d, J = 8.4 Hz, 2H), 7.17 (dd, J = 8.5, 5.6 Hz, 2H), 6.94 (t, J = 8.8 Hz, 2H), 4.26 (s, 2H), 2.28 (s, 3H), 1.60 (dd, J = 7.0, 3.9 Hz, 2H), 1.23 (dd, J = 6.9, 4.0 Hz, 2H). ^{13}C NMR (126 MHz, Methanol-d4) δ 178.58, 164.31, 162.37, 157.97, 143.49, 141.76, 140.60, 139.43, 136.82, 136.07, 132.27, 131.90, 130.13, 130.06, 128.89, 127.81, 125.78, 116.16, 115.99, 44.00, 29.77, 16.96, 9.76. HRMS (ESI) m/z: [M + Na]$^+$ Calcd for C$_{27}$H$_{24}$FN$_5$NaO$_3$ 508.1755; Found 508.1770.

1-(4'-(5-(3-(3-fluorobenzyl)ureido)-4-methyl-1H-1,2,3-triazol-1-yl)-[1,1'-biphenyl]-4-yl)cyclopropane-1-carboxylic acid (17c)

Compound **17c** was prepared in a similar way as compound **17a**, except that (3-fluorophenyl)methanamine was used from step 4 in scheme 1 (yield: 72.65%). ^{1}H NMR (500 MHz, Methanol-d4) δ 7.77 (d, J = 8.1 Hz, 2H), 7.61 (dd, J = 8.1, 5.3 Hz, 4H), 7.48 (d, J = 7.9 Hz, 2H), 7.22 (dd, J = 8.0, 5.9 Hz, 1H), 7.03 – 6.84 (m, 3H), 4.30 (s, 2H), 2.28 (s, 3H), 1.61 (dd, J = 7.0, 4.0 Hz, 2H), 1.24 (dd, J = 6.9, 4.0 Hz, 2H). ^{13}C NMR (126 MHz, Methanol-d4) δ 178.41, 165.31, 163.37, 158.05, 143.52, 141.54, 140.69, 139.52, 136.05, 132.23, 131.83, 131.27, 131.20, 128.91, 127.83, 125.72, 123.86, 123.84, 114.81, 114.64, 44.15, 29.66, 17.03, 9.75. HRMS (ESI) m/z: [M + Na]$^{+}$ Calcd for $C_{27}H_{24}FN_5NaO_3$ 508.1755; Found 508.1776.

1-(4'-(5-(3-(2-fluorobenzyl)ureido)-4-methyl-1H-1,2,3-triazol-1-yl)-[1,1'-biphenyl]-4-yl)cyclopropane-1-carboxylic acid (17d)

Compound **17d** was prepared in a similar way as compound **17a**, except that (2-fluorophenyl)methanamine was used from step 4 in scheme 1 (yield: 78.90%). ^{1}H NMR (500 MHz, Methanol-d4) δ 7.76 (d, J = 8.5 Hz, 2H), 7.60 (dd, J = 10.6, 8.4 Hz, 4H), 7.48 (d, J = 8.3 Hz, 2H), 7.24 – 7.12 (m, 2H), 7.06 – 6.96 (m, 2H), 4.35 (s, 2H), 2.28 (s, 3H), 1.61 (dd, J = 7.0, 3.9 Hz, 2H), 1.24 (dd, J = 6.9, 4.0 Hz, 2H). ^{13}C NMR (126 MHz, Methanol-d4) δ 178.37, 163.03, 161.08, 157.93, 143.46, 141.53, 140.63, 139.55, 136.06, 132.26, 131.86, 130.37, 130.34, 130.14, 130.08, 128.91, 127.84, 127.46, 127.34, 125.76, 125.30, 125.27, 116.18, 116.01, 38.64, 29.63, 17.04, 9.75. HRMS (ESI) m/z: [M + Na]$^{+}$ Calcd for $C_{27}H_{24}FN_5NaO_3$ 508.1755; Found 508.1770.

(R)-1-(4'-(5-(3-(1-(4-fluorophenyl)ethyl)ureido)-4-methyl-1H-1,2,3-triazol-1-yl)-[1,1'-biphenyl]-4-yl)cyclopropane-1-carboxylic acid (17e)

Compound **17e** was prepared in a similar way as compound **17a**, except that (R)-1-(4-fluorophenyl)ethan-1-amine was used from step 4 in scheme 1 (yield: 70.57%). ^{1}H NMR (500 MHz, Methanol-d4) δ 7.72 (d, J = 8.2 Hz, 2H), 7.59 (d, J = 8.0 Hz, 2H), 7.55 (d, J = 8.2 Hz, 2H), 7.47 (d, J = 7.9 Hz, 2H), 7.24 (dd, J = 8.5, 5.3 Hz, 2H), 6.97 (t, J = 8.6 Hz, 2H), 4.80 (q, J = 6.9 Hz, 1H), 2.24 (s, 3H), 1.60 (dd, J = 7.0, 3.9 Hz, 2H), 1.38 (d, J = 7.1 Hz, 3H), 1.22 (dd, J = 6.8, 3.8 Hz, 2H). ^{13}C NMR (126 MHz, Methanol-d4) δ 178.48, 164.17, 162.23, 157.10, 143.43, 141.70, 141.67, 141.64, 140.53, 139.46, 136.06, 132.24, 131.93, 128.85, 128.75, 128.68, 127.80,

125.71, 116.14, 115.97, 50.47, 29.71, 22.77, 16.99, 9.77. HRMS (ESI) m/z: [M + Na]$^+$ Calcd for C$_{28}$H$_{26}$FN$_5$NaO$_3$ 522.1912; Found 522.1909.

(R)-1-(4'-(4-methyl-5-(3-(1-phenylethyl)ureido)-1H-1,2,3-triazol-1-yl)-[1,1'-biphenyl]-4-yl)cyclopropane-1-carboxylic acid (17f)

Compound **17f** was prepared in a similar way as compound **17a**, except that (*R*)-1-phenylethan-1-amine was used from step 4 in scheme 1 (yield: 80.32%). ^{1}H NMR (500 MHz, Methanol-d4) δ 7.73 (d, J = 8.6 Hz, 2H), 7.60 (d, J = 8.3 Hz, 2H), 7.55 (d, J = 8.6 Hz, 2H), 7.48 (d, J = 8.3 Hz, 2H), 7.30 – 7.22 (m, 4H), 7.22 – 7.15 (m, 1H), 4.82 (q, J = 7.0 Hz, 1H), 2.25 (s, 3H), 1.59 (dd, J = 6.9, 3.8 Hz, 2H), 1.41 (d, J = 7.0 Hz, 3H), 1.19 (dd, J = 6.8, 3.9 Hz, 2H). ^{13}C NMR (126 MHz, Methanol-d4) δ 179.24, 157.22, 145.68, 143.50, 142.35, 140.57, 139.28, 135.98, 132.19, 131.98, 129.53, 128.85, 128.05, 127.75, 126.83, 125.69, 51.10, 30.14, 22.88, 16.73, 9.78. HRMS (ESI) m/z: [M + Na]$^+$ Calcd for C$_{28}$H$_{27}$N$_5$NaO$_3$ 504.2006; Found 504.2002.

(S)-1-(4'-(4-methyl-5-(3-(1-phenylethyl)ureido)-1H-1,2,3-triazol-1-yl)-[1,1'-biphenyl]-4-yl)cyclopropane-1-carboxylic acid (17g)

Compound **17g** was prepared in a similar way as compound **17a**, except that (*S*)-1-phenylethan-1-amine was used from step 4 in scheme 1 (yield: 79.05%). ^{1}H NMR (500 MHz, Methanol-d4) δ 7.73 (d, J = 8.5 Hz, 2H), 7.61 (d, J = 8.2 Hz, 2H), 7.56 (d, J = 8.6 Hz, 2H), 7.49 (d, J = 8.3 Hz, 2H), 7.31 – 7.22 (m, 4H), 7.22 – 7.16 (m, 1H), 4.83 (q, J = 7.0 Hz, 1H), 2.25 (s, 3H), 1.63 (dd, J = 7.0, 3.9 Hz, 2H), 1.41 (d, J = 7.0 Hz, 3H), 1.25 (dd, J = 7.0, 4.0 Hz, 2H). ^{13}C NMR (126 MHz, Methanol-d4) δ 178.30, 157.18, 145.64, 143.39, 141.44, 140.56, 139.57, 136.04, 132.25, 131.94, 129.53, 128.89, 128.06, 127.83, 126.83, 125.69, 51.09, 29.59, 22.86, 17.06, 9.77. HRMS (ESI) m/z: [M + Na]$^+$ Calcd for C$_{28}$H$_{27}$N$_5$NaO$_3$ 504.2006; Found 504.2034.

1-(4'-(4-methyl-5-(3-(1-phenylethyl)ureido)-1H-1,2,3-triazol-1-yl)-[1,1'-biphenyl]-4-yl)cyclopropane-1-carboxylic acid (17h)

Compound **17h** was prepared in a similar way as compound **17a**, except that 1-phenylethan-1-amine was used from step 4 in scheme 1 (yield: 71.01%). ^{1}H NMR (500 MHz, Methanol-d4) δ 7.73 (d, J = 8.5 Hz, 2H), 7.60 (d, J = 8.3 Hz, 2H), 7.55 (d, J = 8.5 Hz, 2H), 7.48 (d, J = 8.3 Hz, 2H), 7.30 – 7.22 (m, 4H), 7.22 – 7.15 (m, 1H), 4.83 (q, J = 7.0 Hz, 1H), 2.25 (s, 3H), 1.60 (dd, J = 6.9, 3.9 Hz, 2H), 1.41 (d, J = 7.0 Hz, 3H), 1.21 (dd, J = 6.9, 3.9 Hz, 2H). ^{13}C NMR (126 MHz,

Methanol-d4) δ 179.00, 157.21, 145.67, 143.46, 142.06, 140.56, 139.37, 136.00, 132.21, 131.97, 129.53, 128.86, 128.05, 127.78, 126.83, 125.69, 51.10, 30.02, 22.88, 16.83, 9.78. HRMS (ESI) m/z: [M + Na]$^+$ Calcd for $C_{28}H_{27}N_5NaO_3$ 504.2006; Found 504.2001.

2.2 Cell culture

All media and cell culture reagents were purchased from WISENT Bioproducts (Quebec, Canada). The Chinese hamster ovary (CHO)-K1 cell line was cultured in F-12K medium supplemented with 10% fetal bovine serum and 1% streptomycin/penicillin at 37°C, 5% CO_2. The MDA-MB-231 cell line was cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum and 1% streptomycin/penicillin at 37°C, 5% CO_2.

2.3 Biological evaluation

The biological studies were conducted in accordance with the methods outlined in the first part of this chapter.

2.4 Molecular docking study

The crystal structure of LPA_1 (PBD ID:4z36)[12] was used to perform molecular docking studies. The 3D structures of **RO6842262** and the compounds synthesized in this study were generated using DS Viewer 3.5. The actual docking was executed using AutoDock Vina,[13] with the grid box dimensions adjusted to adequately cover the entire binding pocket within the LPA_1 structure. The remaining parameters were maintained at their default values.

2.5 Statistical analysis

The experimental outcomes are presented as the mean ± standard error of the mean (SEM). Data processing was carried out using GraphPad Prism 9.0.0. One-way ANOVA was used to analyze the statistical P values between different groups. A P value of < 0.05 (two-tailed) was considered as statistically significant.

3. Results and discussion

The synthetic procedure is illustrated in Scheme 2, depicting the synthesis of urea-derived analogues of **RO6842262** (**17a-17h**). Initially, 2-(4-bromophenyl)acetonitrile underwent a reaction with 1,2-dibromoethane, yielding 1-(4-bromophenyl)cyclopropane-1-carbonitrile (**1**). Subsequent steps involved the conversion of compound **1** to 1-(4-bromophenyl)cyclopropane-1-carboxylic acid (**2**) through a reaction with ethylene glycol, followed by esterification to obtain compound **3**. Compound **3** was then reacted in a Suzuki coupling reaction with bis(pinacolato)diboron, resulting in the formation of ethyl 1-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)cyclopropane-1-carboxylate (**4**).

1-Azido-4-bromobenzene (**7**) was obtained by reacting 4-bromoaniline with sodium nitrite. Subsequently, compound **7** underwent a reaction with but-2-ynoic acid ethyl ester, yielding ethyl 1-(4-bromophenyl)-4-methyl-1H-1,2,3-triazole-5-carboxylate (**8**). Further reaction with sodium hydroxide led to the formation of the corresponding carboxylic acid **9**. Curtius rearrangement reaction was then performed with an benzylamine building block, leading to the formation of the urea compound **15**. Following this, a Suzuki-coupling reaction was performed to generated compound **16**. Lastly, hydrolysis of the ester in compound **16** yielded the final carboxylic acid compounds **17a-17h**, as depicted in Scheme 2.
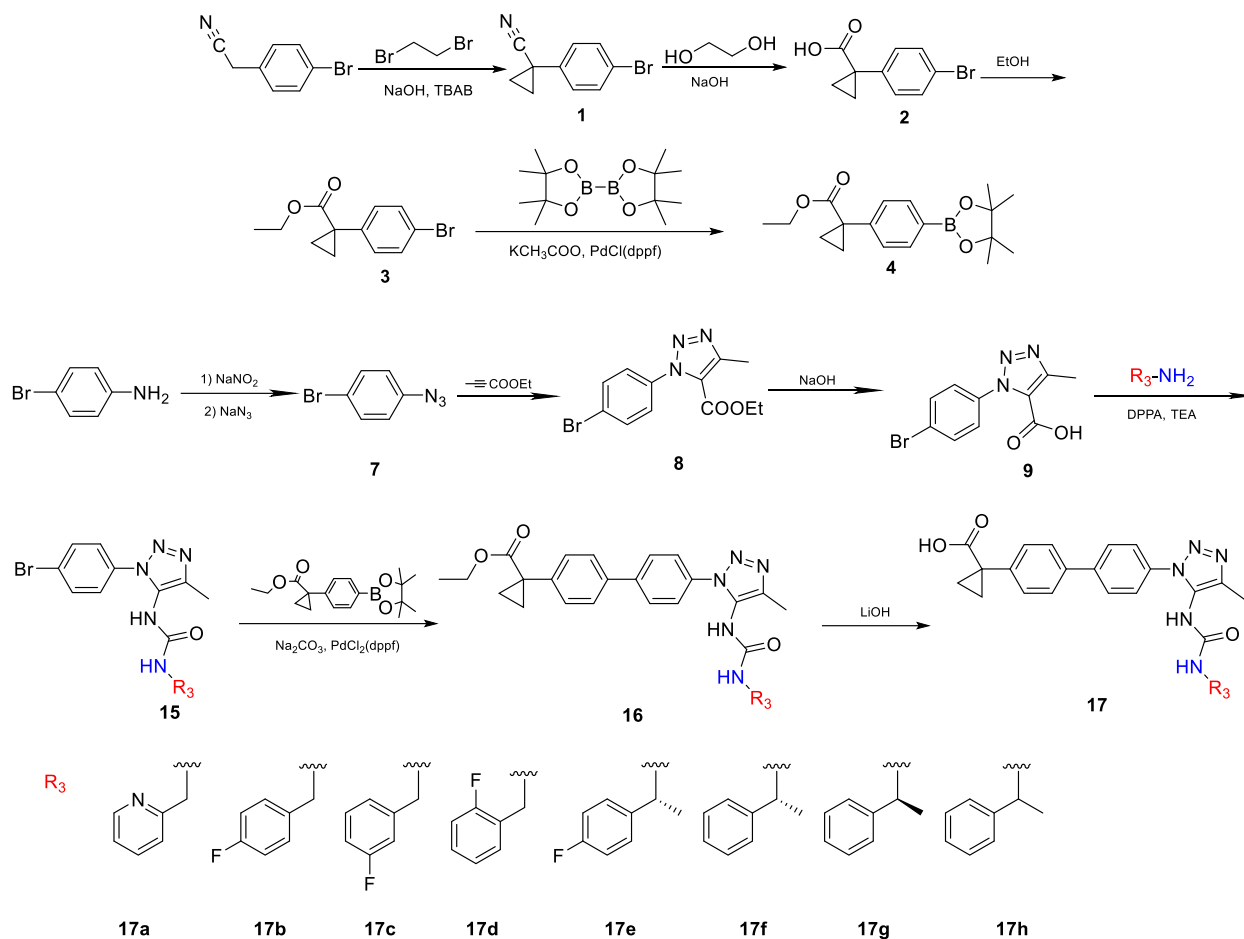
Scheme 2. Synthesis of compounds **17a-17h**.

The compounds prepared in Scheme 2 were tested in the cAMP assay using the Chinese hamster ovary (CHO) cell line overexpressing the LPA$_1$ receptor, following the procedure detailed in part one of this chapter. The endogenous ligand **18:1 LPA** was used for activating the LPA$_1$ receptor. The LPA$_1$ antagonist activity is presented as IC$_{50}$, representing the concentration required to reverse 50% of the inhibitory effect caused by **18:1 LPA** on forskolin-mediated cAMP production.

To investigate the significance of methyl substitution at the alpha position of the urea for compound activity, three fluorine-containing compounds without the methyl group were synthesized (**17b**, **17c**, and **17d**). The results suggested that urea-derived compounds without the methyl substitution showed weak LPA$_1$ antagonistic activity (Table 2.2.1). As a scaffold that is commonly used in medicinal chemistry, the heterocyclic compound pyridine was also introduced to replace the benzene ring (**17a**), resulting in a slight increase in activity.

Since compounds without the alpha methyl substitution showed weak activity, compounds with the methyl substitution were synthesized. Those with the methyl group (**17f**, **17g,** and **17h**) exhibited significantly higher activity when compared to those without it, as shown in Table 2.2.1. Among them, the (*R*)-configuration (**17f**) exhibited the highest activity (IC$_{50}$ = 215.2), followed by the racemate (**17h**) (IC$_{50}$ = 239.7). Given that the (*R*)-configuration is more active than the (*S*)-configuration, a fluorine atom was introduced into the aromatic ring, resulting in the synthesis of compound **17e**. However, the addition of fluorine led to a decrease in activity.
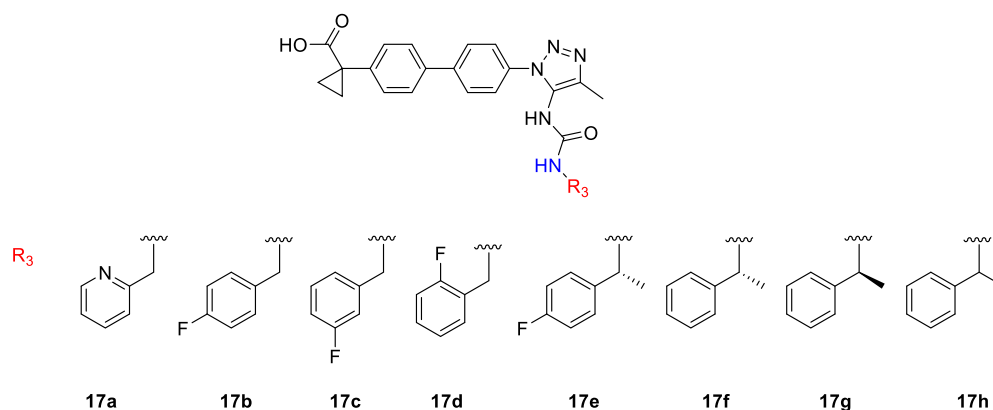
Table 2.2.1. SAR of urea-derived LPA$_1$ antagonists **17a-17h** using the cAMP assay.

Compounds	IC$_{50}$ (nM)	pIC$_{50}$ [a]	% inhibition [b]
RO6842262	18.3	7.74 ± 0.05	111.6 ± 2.8
17a	1005	6.55 ± 0.83	41.84 ± 1.22
17b	3152	5.59 ± 0.28	40.48 ± 20.99
17c	1519	5.83 ± 0.10	88.09 ± 20.46
17d	4469	5.51 ± 0.39	49.52 ± 33.15
17e	780.2	6.11 ± 0.04	102.00 ± 17.70
17f	215.2	6.83 ± 0.28	97.64 ± 5.02
17g	242.4	6.66 ± 0.19	108.10 ± 0.26
17h	239.7	6.81 ± 0.43	83.75 ± 6.24

a: Values are the mean pIC$_{50}$ $\pm$ SEM of at least two independent experiments performed in duplicate.

b: % maximal inhibition of the response to 1 µM **18:1 LPA** presented as mean $\pm$ SEM.

As one of the most potent compounds from this series, compounds **17f** was selected for further testing for its selectivity against the LPA_{2-3} receptors. As shown in Figure 2.2.2, **17f** showed no significant activity for the LPA_{2-3} receptors, suggesting great LPA_1 selectivity (Table 2.2.3).
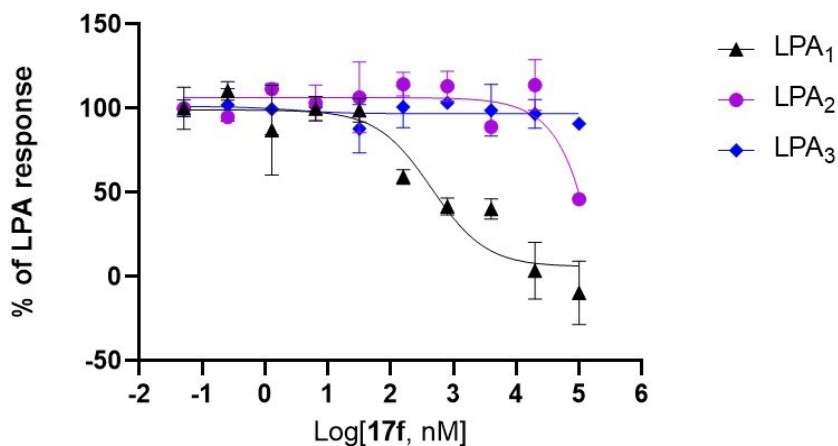
Figure 2.2.2. Dose-response curves of compound **17f** towards the LPA_1, LPA_2, and LPA_3 receptors.

Table 2.2.3. Selectivity of compound **17f**.

Compounds	LPA_1 [a] (nM)	LPA_2 [a] (nM)	LPA_3 [a] (nM)
17f	215.2	> 10000	NE [b]

a: Values are the mean IC_{50} of at least two independent experiments performed in duplicate.

b: No effect was observed at the highest concentration tested.

Molecular docking studies were conducted using the X-ray crystal structure of LPA_1 (PDB ID: 4z36)[10] to explore the possible binding mode of the best compound identified in this study, compound **17f**. The compound exhibited binding to the orthosteric binding pocket of LPA_1 in a similar fashion as **RO6842262**, with the carboxylate oxygen forming hydrogen bonds with Gly110, Thr109, and Thr113 (Figure 2.2.3).
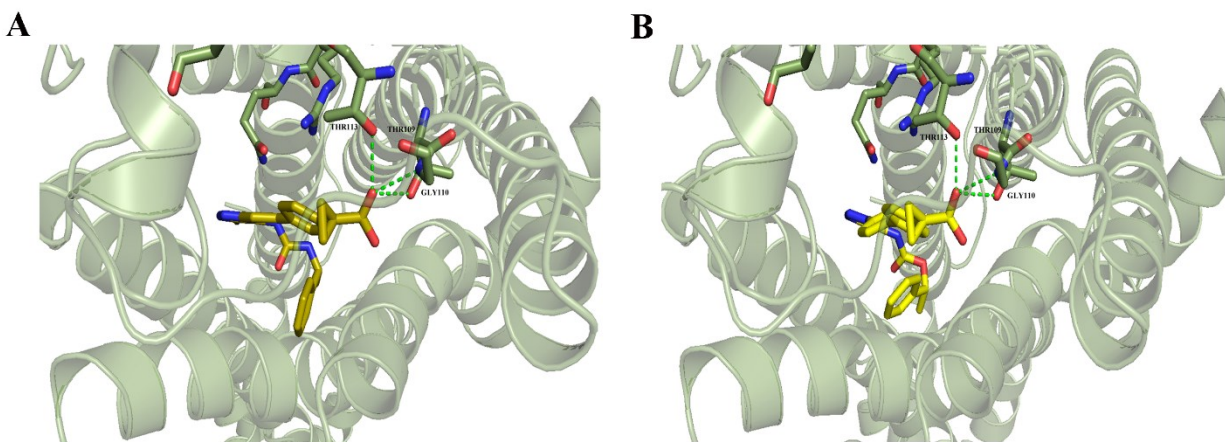
Figure 2.2.3. Binding modes of compound **17f** (A) and **RO6842262** (B) with the LPA$_1$ receptor. Green dashed lines represent hydrogen bond interactions.

Subsequently, the most potent compound (**17f**) within this series was subjected to further investigation to assess its therapeutic impact on metastatic breast cancer using the breast cancer cell line MDA-MB-231. The wound healing assay was firstly performed to investigate the effect of LPA on the migration of MDA-MB-231 cells. When exposed to 10 µM **18:1 LPA**, a notable enhancement in cell migration rate was observed, with the migration rate increasing from the baseline of 41.3% to 93.5% in comparison to control cells. Following LPA stimulation, the introduction of the LPA$_1$ antagonist **17f** resulted in a dose-dependent inhibition of cell migration, as illustrated in Figure 2.2.4A. The antimigration effect of compound **17f** became noticeable at a concentration of 5 µM. The addition of 10 µM and 20 µM of compound **17f** led to reductions in LPA-induced migration by 31.16% and 40.52%, respectively.

The impact of compound **17f** on cell migration was also assessed in the presence of serum. As depicted in Figure 2.2.4B, **17f** exhibited a modest inhibitory effect on serum-induced cell migration, displaying efficacy at a minimum concentration of 20 µM. The introduction of 20 µM of compound **17f** led to a reduction of 19.55% in serum-induced migration. However, concentrations of compound **17f** below 20 µM did not yield any notable impact on cell migration.

The antimigratory effect of **17f** is comparatively weaker in serum-induced migration compared to LPA-induced migration. This is unsurprising given that serum contains a wide range of bioactive molecules aside from LPA. In summary, compound **17f** independently hindered serum-induced cell migration in MDA-MB-231 cells, even in the absence of LPA.
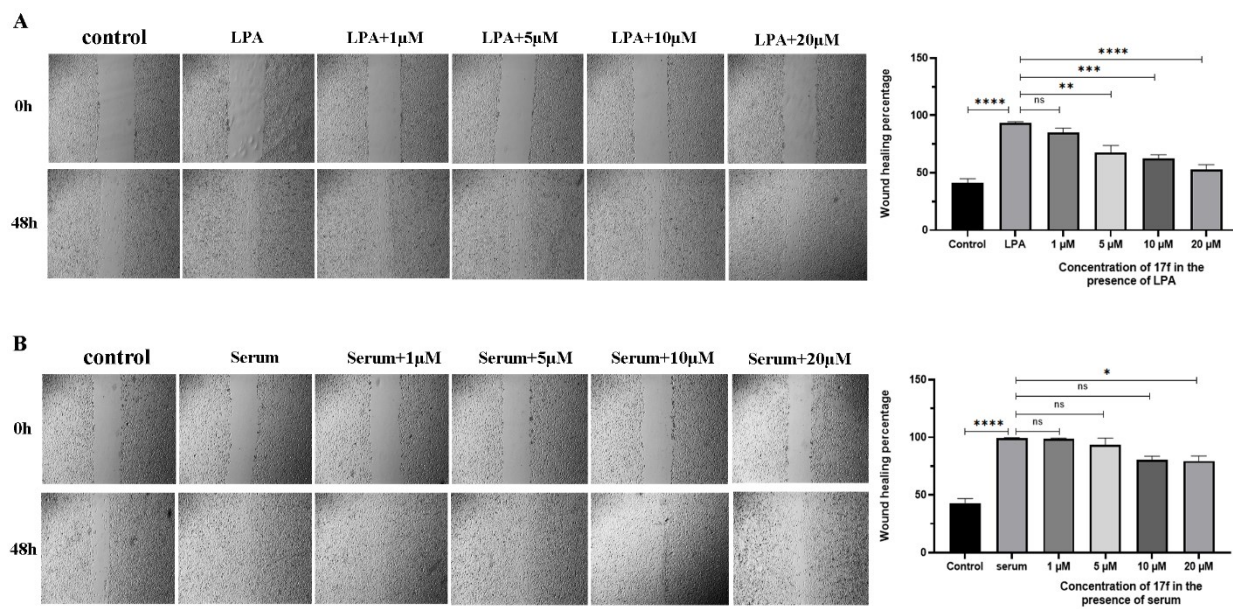
Figure 2.2.4. Effects of LPA, serum and **17f** on cell migration in wound healing assays using MDA-MB-231 cells. A. Effects of 10 µM **18:1 LPA** and various concentrations of **17f** on cell migration (left) and its quantification (right). B. Effects of serum and various concentrations of **17f** on cell migration (left) and its quantification (right). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$, ns $P > 0.05$.

The impact of LPA and **17f** on MDA-MB-231 cell migration was further substantiated through Transwell migration/invasion assays. As depicted in Figure 2.2.5, the exposure to 10 µM **18:1 LPA** led to noticeable cell migration and invasion in MDA-MB-231. However, this effect was progressively impeded by **17f** in a dose-dependent manner. These findings collectively indicate that LPA promotes migration and invasion of breast cancer cells, while the LPA antagonist **17f** effectively counteracts the migratory and invasive effects induced by LPA. Furthermore, **17f** demonstrated the ability to inhibit serum-induced migration of MDA-MB-231 even without the presence of LPA.
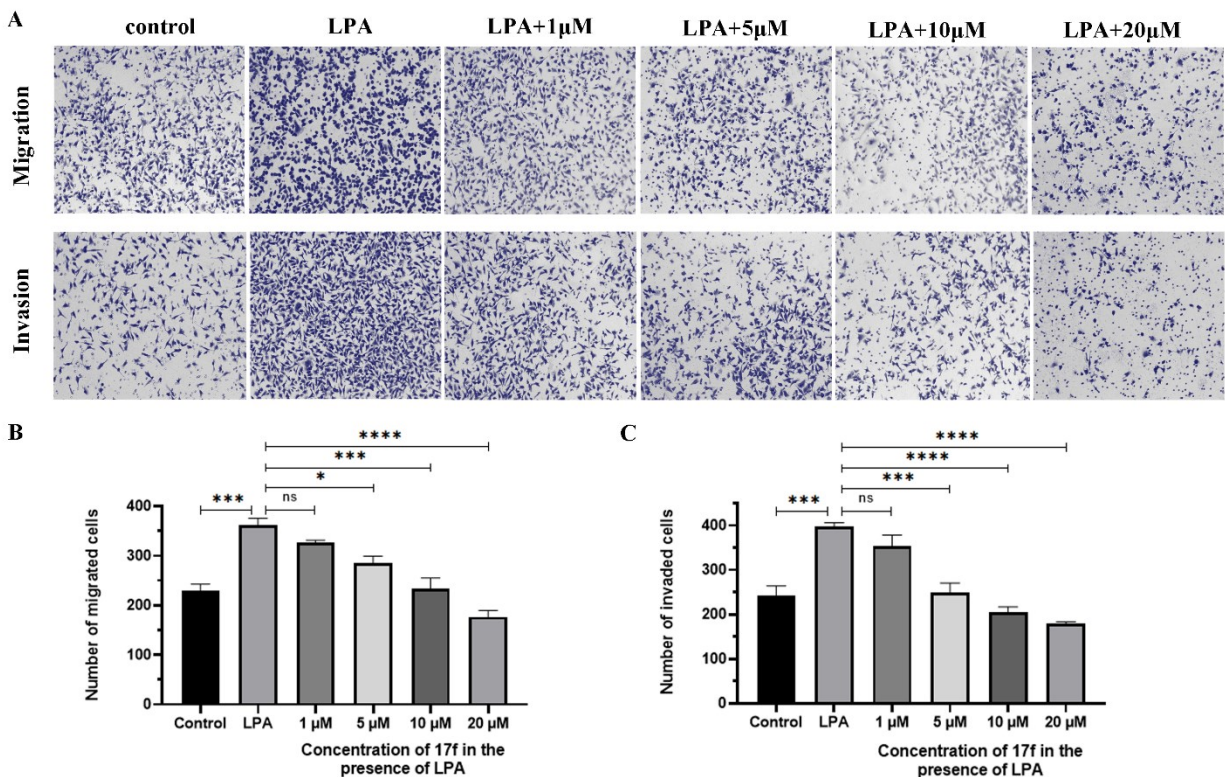
Figure 2.2.5. Effects of LPA and various concentrations of **17f** on cell migration/invasion in Transwell migration/invasion assays using MDA-MB-231 cells. A. Effects of 10 μM **18:1 LPA** and various concentrations of **17f** on cell migration (top) and invasion (bottom). B. Quantification of Transwell migration assay. C. Quantification of Transwell invasion assay. * $P < 0.05$, *** $P < 0.001$, **** $P < 0.0001$, ns $P > 0.05$.

Metastasis is the main cause of death of cancer patients, accounting for more than 90% of tumor mortality.[14] In breast cancer patients, most deaths are caused by the metastatic behavior instead of the primary tumor growth. Approximately 6% to 10% of breast cancer patients are diagnosed with metastatic disease at the initial diagnosis, and approximately 30% of patients initially diagnosed with early-stage breast cancer eventually experience recurrence or metastatic disease.[15] Consequently, addressing the factors responsible for breast cancer metastasis formation poses a significant challenge. This study revealed a connection between LPA_1 and breast cancer migration, offering a potentially promising avenue for addressing the issue of metastatic breast cancer.

LPA_1 is linked to several intracellular signaling pathways that regulate cell behavior. LPA is recognized for instigating the migration and invasion of breast cancer cells by mobilizing LPA receptors and triggering downstream beta-arrestin/Ral signaling pathway.[2] Among LPA receptor subtypes, LPA_1 emerged as the mediator of LPA-induced migration across multiple cancer cell lines.[16] The outcomes of this study are in agreement with previous findings,[3-6] confirming LPA_1's role in breast cancer metastasis. Hence, directing efforts toward LPA_1 for drug development holds significant promise.

While the urea-derived compounds identified in this study demonstrated generally lower activity in comparison to the carbamate-derived compounds detailed in the first part of this chapter, there is potential for them to exhibit improved drug properties due to the inherent stability of urea in contrast to carbamate. Moreover, the most potent compound discovered in this study offers the prospect of further modifications to enhance its activity in future studies.

4. Conclusion

In conclusion, LPA_1 was identified as a novel marker for metastatic breast cancer, leading to the synthesis and assessment of a series of urea-derived LPA_1 antagonists. Notably, compound **17f** demonstrated remarkable LPA_1 antagonistic activity in the cAMP assay. Moreover, it displayed exceptional inhibitory effects on LPA-induced cell migration and invasion in a breast cancer cell line. The LPA_1 antagonists developed herein hold the potential to introduce a new therapeutic avenue for breast cancer. The most potent compound discovered may serve as the starting point for future investigations, facilitating further refinement and optimization.

5. Reference

1. Liu, W.; Hopkins, A. M.; Hou, J., The development of modulators for lysophosphatidic acid receptors: A comprehensive review. *Bioorg Chem* **2021,** *117*, 105386.

2. Li, T. T.; Alemayehu, M.; Aziziyeh, A. I.; Pape, C.; Pampillo, M.; Postovit, L.-M.; Mills, G. B.; Babwah, A. V.; Bhattacharya, M., Beta-arrestin/Ral signaling regulates lysophosphatidic acid-mediated migration and invasion of human breast tumor cells. *Mol Cancer Res* **2009,** *7* (7), 1064-77.

3. Boucharaba, A.; Serre, C.-M.; Grès, S.; Saulnier-Blache, J. S.; Bordet, J.-C.; Guglielmi, J.; Clézardin, P.; Peyruchaud, O., Platelet-derived lysophosphatidic acid supports the progression of osteolytic bone metastases in breast cancer. *J Clin Invest* **2004,** *114* (12), 1714-1725.

4. Boucharaba, A.; Serre, C.-M.; Guglielmi, J.; Bordet, J.-C.; Clézardin, P.; Peyruchaud, O., The type 1 lysophosphatidic acid receptor is a target for therapy in bone metastases. *Proc Natl Acad Sci U S A* **2006,** *103* (25), 9643-8.

5. David, M.; Ribeiro, J.; Descotes, F.; Serre, C.-M.; Barbier, M.; Murone, M.; Clézardin, P.; Peyruchaud, O., Targeting lysophosphatidic acid receptor type 1 with Debio 0719 inhibits spontaneous metastasis dissemination of breast cancer cells independently of cell proliferation and angiogenesis. *Int J Oncol* **2012,** *40* (4), 1133-41.

6. Marshall, J.-C. A.; Collins, J. W.; Nakayama, J.; Horak, C. E.; Liewehr, D. J.; Steinberg, S. M.; Albaugh, M.; Vidal-Vanaclocha, F.; Palmieri, D.; Barbier, M.; Murone, M.; Steeg, P. S., Effect of inhibition of the lysophosphatidic acid receptor 1 on metastasis and metastatic dormancy in breast cancer. *J Natl Cancer Inst* **2012,** `*04* (17), 1306-19.

7. Qian, Y.; Hamilton, M.; Sidduri, A.; Gabriel, S.; Ren, Y.; Peng, R.; Kondru, R.; Narayanan, A.; Truitt, T.; Hamid, R.; Chen, Y.; Zhang, L.; Fretland, A. J.; Sanchez, R. A.; Chang, K. C.; Lucas, M.; Schoenfeld, R. C.; Laine, D.; Fuentes, M. E.; Stevenson, C. S.; Budd, D. C., Discovery of highly selective and orally active lysophosphatidic acid receptor-1 antagonists with potent activity on human lung fibroblasts. *J Med Chem* **2012,** *55* (17), 7920-7939.

8. Ghosh, A. K.; Brindisi, M., Urea Derivatives in Modern Drug Discovery and Medicinal Chemistry. *J Med Chem* **2020,** *63* (6), 2751–2788.

9.   Ronchetti, R.; Moroni, G.; Carotti, A.; Gioiello, A.; Camaioni, E., Recent advances in urea- and thiourea-containing compounds: focus on innovative approaches in medicinal chemistry and organic synthesis. *RSC Med Chem* **2021,** *12* (7), 1046–1064.

10.   Buckman, B. O.; Nicholas, J. B.; Emayan, K.; Seiwert, S. D. Lysophosphatidic acid receptor antagonists. *WO2013025733A1* **2013**.

11.   Buckman, B. O.; Nicholas, J. B.; Emayan, K.; Seiwert, S. D.; Yuan, S. Lysophosphatidic acid receptor antagonists. *WO2014113485A1* **2014**.

12.   Chrencik, J. E.; Roth, C. B.; Terakado, M.; Kurata, H.; Omi, R.; Kihara, Y.; Warshaviak, D.; Nakade, S.; Asmar-Rovira, G.; Mileni, M.; Mizuno, H.; Griffith, M. T.; Rodgers, C.; Han, G. W.; Velasquez, J.; Chun, J.; Stevens, R. C.; Hanson, M. A., Crystal structure of antagonist bound human lysophosphatidic acid receptor 1. *Cell* **2015,** *161* (7), 1633-1643.

13.   Forli, S.; Huey, R.; Pique, M. E.; Sanner, M. F.; Goodsell, D. S.; Olson, A. J., Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nature Protocols* **2016,** *11*, 905-919.

14.   Chaffer, C. L.; Weinberg, R. A., A perspective on cancer cell metastasis. *Science* **2011,** *331* (6024), 1559-64.

15.   Jones, S. E., Metastatic breast cancer: the treatment challenge. *Clin Breast Cancer* **2008,** *8* (3), 224-233.

16.   Hama, K.; Aoki, J.; Fukaya, M.; Kishi, Y.; Sakai, T.; Suzuki, R.; Ohta, H.; Yamori, T.; Watanabe, M.; Chun, J.; Arai, H., Lysophosphatidic acid and autotaxin stimulate cell motility of neoplastic and non-neoplastic cells through LPA1. *J Biol Chem* **2004**, 279 (17), 17634-17639.

Part three

Synthesis and evaluation of a lysophosphatidic acid receptor 1 radioligand for positron emission tomography imaging

Abstract:

Lysophosphatidic acid receptor 1 (LPA_1) belongs to the G-protein coupled receptor (GPCR) family activated by lysophosphatidic acids. The previous studies have shown that LPA_1 mediates cell migration, and invasion in triple negative breast cancer, which can be inhibited by LPA_1 antagonists. In the field of oncology, positron emission tomography (PET) is a valuable tool for early cancer diagnosis, accurate staging, and detection of recurrence. However, as of now, there is no reported radio tracer specifically designed to target LPA_1 for cancer diagnosis. Therefore, the development of a PET radioligand that targets LPA_1 could be beneficial for studying triple negative breast cancer, facilitating early diagnosis, staging, and enabling prognostic evaluations. In earlier studies, a potent and selective LPA_1 antagonist, compound **12h**, was successfully synthesized. This compound was selected for radiolabeling by conducting tosylation followed by fluorination. The resulting fluorinated cold compound **12k** exhibited an IC_{50} of 432.5 nM in the cAMP assay for LPA_1 and showed no activity towards LPA_2 and LPA_3, demonstrating promising potentials as a novel LPA_1 radioligand. The next step involves replacing the fluorine atom with 18-fluorine and evaluating its performance *in vivo* using mouse models of triple negative breast cancer.

Graphical abstract:

Compound **12k**
LPA$_1$ IC$_{50}$ = 432.5 nM
LPA$_2$ IC$_{50}$ > 10 µM
LPA$_3$ IC$_{50}$ > 10 µM
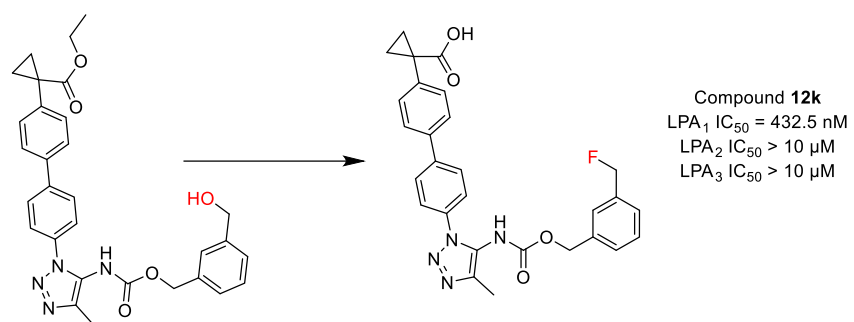
Key words: LPA$_1$, positron emission tomography, radiolabeling

1. **Introduction**

As one of the most commonly used cancer imaging techniques, positron emission tomography (PET) enables the visualization of metabolic and functional processes within the body. In the field of breast cancer, the FDA has approved two radiotracers for clinical use in PET imaging. The primary FDA-approved radiotracer for PET imaging of breast cancer is 2-deoxy-2-[^{18}F]fluoro-D-glucose (FDG). FDG-PET surpasses conventional imaging methods in its effectiveness for detecting distant recurrence and metastases in advanced-stage breast cancer, thereby allowing early detection and accurate staging of the disease.[1] However, there are concerns regarding the accuracy and sensitivity of FDG-PET in detecting breast cancer.[2] This is because FDG uptake itself lacks tumor-specificity, making it challenging to differentiate between malignant and benign breast cells, especially when dealing with instances of breast hypermetabolism.[3] Another PET imaging agent, [^{18}F]-Fluoroestradiol (FES), has recently gained FDA approval for breast cancer. This agent is specifically intended for patients with recurrent or metastatic breast cancer. FES functions as an estrogen analogue, with its uptake directly linked to the concentration of estrogen receptors (ER).[4]

In addition to these two FDA-approved PET tracers for breast cancer, there are ongoing developments of other radioligands for breast cancer PET imaging that target either the progesterone receptor (PR) or the human epidermal growth factor receptor type 2 (HER2).[5] However, there exists a subset of breast cancer patients, approximately 12-17%, who lack the expression of these three receptors. This subtype is known as triple-negative breast cancer (TNBC).[6] TNBC represents the most aggressive form of breast cancer, characterized by early metastasis and a poor prognosis. Due to its stem cell-like characteristics and resistance to targeted therapies, TNBC remains a significant challenge in clinical practice.[7]

Hence, the development of novel radiotracers is imperative to address this issue and fully leverage the potential of PET in the early detection and staging of breast cancer, particularly in instances of TNBC where metastases and recurrences are involved.

Based on the prior research, it has been established that LPA$_1$ exerts a detrimental impact on TNBC.[8-10] In several TNBC cell lines, the expression of LPA$_1$ is notably elevated in contrast to non-tumorigenic cells,[11] rendering it a promising target for the development of radiotracers for PET imaging. The primary goal of this study is to develop radioligands that specifically target

LPA$_1$ for PET imaging in TNBC, with the aim of improving both the sensitivity and accuracy in comparison to FDG. Among the LPA$_1$ antagonists investigated in the previous study, compound **12h** demonstrated exceptional potency and selectivity, making it an ideal candidate for radiolabeling. The hydroxyl group located on the aromatic ring of compound **12h** can be fluorinated via tosylation, thereby serving as a starting point for further exploration.

2. Materials and methods

2.1 Synthesis

All solvents and chemicals were of reagent grade and purchased from commercial manufacturers. The purity and characterization of compounds were ascertained through a combination of techniques including HPLC, TLC, high-resolution mass spectrometry (HRMS), as well as ^{1}H NMR and ^{13}C NMR analyses.

ethyl 1-(4'-(4-methyl-5-(((((3-((tosyloxy)methyl)benzyl)oxy)carbonyl)amino)-1H-1,2,3-triazol-1-yl)-[1,1'-biphenyl]-4-yl)cyclopropane-1-carboxylate (18)

To a stirred mixture of compound **11h** (90 mg, 0.17 mmol) and *p*-toluenesulfonic anhydride (111.63 mg, 0.34 mmol) in dry DCM (2 mL) was added Ytterbium(III) trifluoromethanesulfonate (106.06 mg, 0.17 mmol). The mixture was bubbled with argon and then stirred at 40°C under argon for 6 hours. The mixture was filtered and partitioned between DCM and water. The organic phase was extracted with DCM, washed with brine, dried over Na$_2$SO$_4$, filtered, and concentrated. The residue was purified by column chromatography (Hexane: EtoAc = 2:1) to give compound **18** (50 mg, yield: 43.00%). ESI-MS m/z calcd C$_{37}$H$_{36}$N$_4$O$_7$S [M + H]+ 681.24, found 681.40.

1-(4'-(5-(((((3-(fluoromethyl)benzyl)oxy)carbonyl)amino)-4-methyl-1H-1,2,3-triazol-1-yl)-[1,1'-biphenyl]-4-yl)cyclopropane-1-carboxylic acid (12k)

To a stirred mixture of compound **18** (5 mg, 0.0073 mmol) and tetra-n-butylammonium fluoride (2.3 mg, 0.0088 mmol) in THF was added LiOH (1.76 mg, 0.073 mmol). The mixture was bubbled with argon and then stirred at 90°C under argon for 2 hours. The reaction mixture was cooled down to 0°C and neutralized with conc. HCl to pH = 2. The mixture was extracted with EtOAc. The combined organic phase was dried over Na$_2$SO$_4$, filtered, and concentrated. The residue was purified by column chromatography (DCM: MeOH = 30:1) to give compound **12k** (1.9 mg, yield: 25.00%). ^{1}H NMR (500 MHz, Methanol-d4) δ 7.75 (d, J = 8.1 Hz, 2H), 7.61 (d, J = 8.1 Hz, 2H), 7.54 (d, J = 7.8 Hz, 2H), 7.49 (d, J = 7.9 Hz, 2H), 7.44 – 7.14 (m, 4H), 5.35 (s, 1H), 5.26 (s, 1H), 5.16 (s, 2H), 2.27 (s, 3H), 1.60 (br s, 2H), 1.20 (br s, 2H). ^{13}C NMR (126 MHz, DMSO-d6) δ 170.87, 154.59, 141.33, 138.77, 137.35, 137.07, 137.04, 136.97, 136.84, 134.91, 131.41, 130.07, 129.20, 128.62, 128.60, 128.07, 127.98, 127.94, 127.58, 127.53, 126.74, 124.52,

85.11, 83.82, 66.84, 55.39, 15.86, 10.08. ESI-MS m/z calcd $C_{28}H_{25}FN_4O_4$ [M + H]+ 501.19, found 501.36.

2.2 Biological evaluation

The biological studies were conducted in accordance with the methods outlined in the first part of this chapter.

3. Results and discussion

The synthesis of both the fluorinated cold compound **12k** and its radiolabeled counterpart were detailed in Scheme 3. Starting from the ester form (**11h**) of compound **12h**, the reaction with p-toluenesulfonic anhydride yielded the tosylated compound **18**. Subsequently, compound **18** was subjected to reactions with tetra-n-butylammonium fluoride (TBAF) and LiOH to obtain the final compound **12k**. Likewise, for the proposed radiolabeling, compound **18** will undergo a reaction with $^{18}F^-$ and Kryptofix 222 to generate **^{18}F-12k**.

Proposed synthesis of cold compound
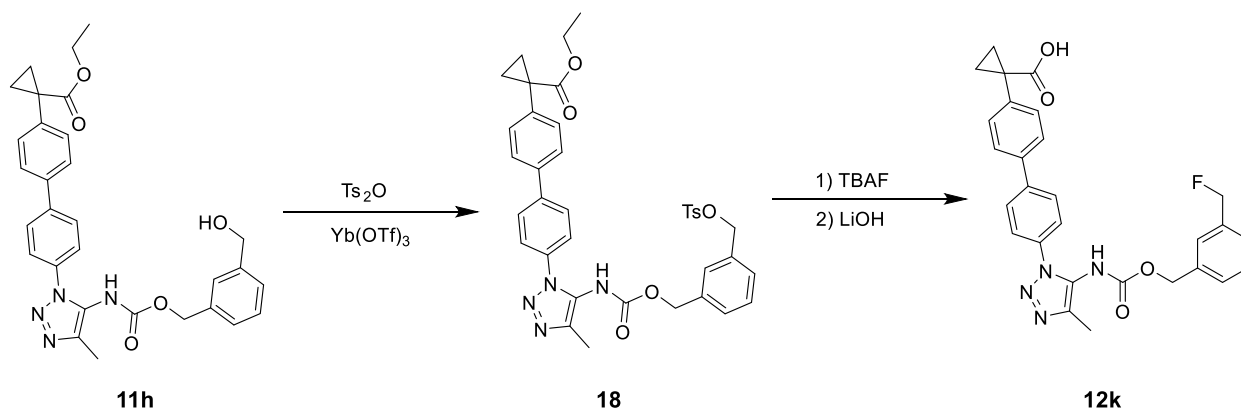
Proposed radiolabelling
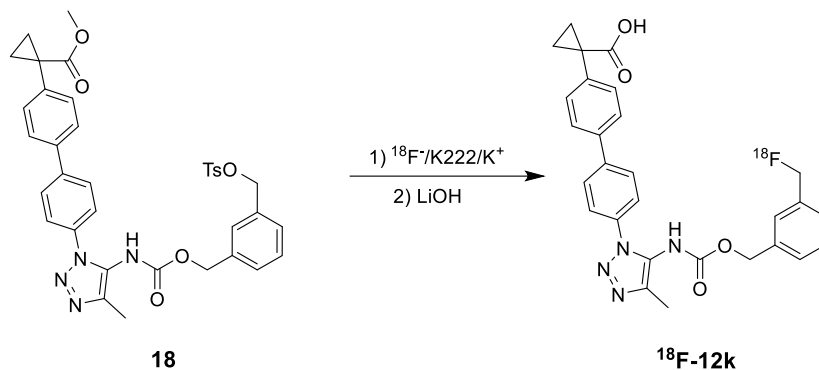
Scheme 3. Synthesis of compound **12k**: from cold to hot.

The synthesis of the cold compound **12k** was successfully accomplished and its activity towards LPA₁, LPA₂, and LPA₃ was evaluated using the CHO cell line, following the previously described cAMP assay. The results of these experiments are presented in Figure 2.3.1 and Table 2.3.1.

Table 2.3.1 Activity of compound **12k** towards LPA₁, LPA₂, and LPA₃.

Targets	pIC₅₀ [a]	IC₅₀ (nM)	% of inhibition [b]
LPA₁	6.37 ± 0.04	432.5	104.60 ± 0.73
LPA₂	< 5	> 10,000	-
LPA₃	< 5	> 10,000	-

a: Values are the mean pIC₅₀ ± SEM of at least two independent experiments performed in duplicate.

b: % maximal inhibition of the response to 1 μM **18:1 LPA**.
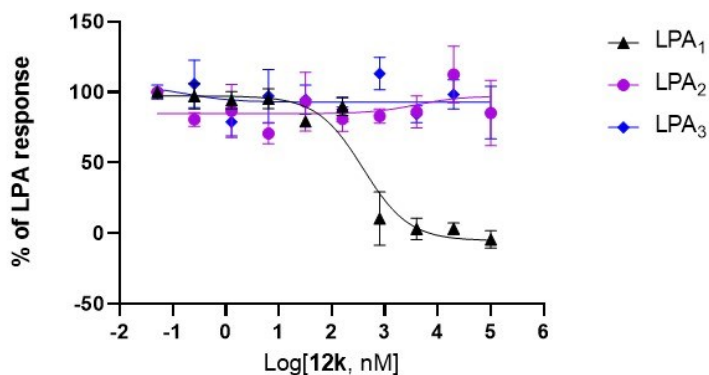
Figure 2.3.1 Dose-response curves of compounds **12k** towards the LPA₁, LPA₂, and LPA₃ receptors.

Replacing the hydroxyl group with fluorine on the aromatic ring resulted in a decrease in ligand activity from 91.3 nM to 432.5 nM (**12h** vs. **12k**). However, this modification did not affect the selectivity of the compound. Despite the reduced activity, compound **12k** remains a promising candidate as a PET radioligand targeting LPA₁ as its highly selective towards LPA₁.

The successful synthesis and evaluation of a fluorinated analogue, **12k**, have been achieved by replacing the hydroxyl group with fluorine on compound **12h**. Compound **12k** exhibits favorable activity and high selectivity towards LPA₁, indicating its significant potential for radiolabeling. Moving forward, the corresponding radiolabeling process will be conducted to acquire 18F-**12k**.

The binding affinity of this radiolabeled compound for a triple-negative breast cancer cell line, MDA-MB-231, will be measured using a cell uptake assay. This will be succeeded by comprehensive evaluations of **18F-12k** using mouse models exhibiting TNBC characteristics.

4. Reference

1. Fowler, A. M.; Strigel, R. M., Clinical advances in PET-MRI for breast cancer. *Lancet Oncol* **2022,** *23* (1), e32-e43.

2. Guo, R.; Lu, G.; Qin, B.; Fei, B., Ultrasound Imaging Technologies for Breast Cancer Detection and Management: A Review. *Ultrasound Med Biol* **2018,** *44* (1), 37-70.

3. T.H. Oude Munnink; W.B. Nagengast; A.H. Brouwers; C.P. Schroder; G.A. Hospers; Hooge, M. N. L.-d.; Wall, E. v. d.; Diest, P. J. v.; Vries, E. G. E. d., Molecular imaging of breast cancer. *The Breast* **2009,** *18*, S66-S73.

4. Hunter, N.; Peterson, L.; Mankoff, D. A.; Muzi, M.; Chen, D. L.; Vinayak, S.; Gwin, W. R.; Specht, J. M.; Linden, H. M., Matched FES and FDG PET imaging in patients with hormone receptor-positive, HER2+ advanced breast cancer. *J Clin Oncol* **2022,** *40* (16_suppl), 1042-1042.

5. Lebron, L.; Greenspan, D.; Pandit-Taskar, N., PET Imaging of Breast Cancer: Role in Patient Management. *PET Clin* **2015,** *10* (2), 159-195.

6. Foulkes, W. D.; Smith, I. E.; Reis-Filho, J. S., Triple-negative breast cancer. *N Engl J Med* **2010,** *363* (20), 1938-48.

7. Carey, L.; Winer, E.; Viale, G.; Cameron, D.; Gianni, L., Triple-negative breast cancer: disease entity or title of convenience? *Nat Rev Clin Oncol* **2010,** *7* (12), 683-692.

8. David, M.; Ribeiro, J.; Descotes, F.; Serre, C.-M.; Barbier, M.; Murone, M.; Clézardin, P.; Peyruchaud, O., Targeting lysophosphatidic acid receptor type 1 with Debio 0719 inhibits spontaneous metastasis dissemination of breast cancer cells independently of cell proliferation and angiogenesis. *Int J Oncol* **2012**, *40* (4), 1133-41.

9. Boucharaba, A.; Serre, C.-M.; Grès, S.; Saulnier-Blache, J. S.; Bordet, J.-C.; Guglielmi, J.; Clézardin, P.; Peyruchaud, O., Platelet-derived lysophosphatidic acid supports the progression of osteolytic bone metastases in breast cancer. *J Clin Invest* **2004**, *114* (12), 1714-1725.

10. Boucharaba, A.; Serre, C.-M.; Guglielmi, J.; Bordet, J.-C.; Clézardin, P.; Peyruchaud, O., The type 1 lysophosphatidic acid receptor is a target for therapy in bone metastases. *Proc Natl Acad Sci U S A* **2006**, *103* (25), 9643-8.

11. Li, T. T.; Alemayehu, M.; Aziziyeh, A. I.; Pape, C.; Pampillo, M.; Postovit, L.-M.; Mills, G. B.; Babwah, A. V.; Bhattacharya, M., Beta-arrestin/Ral signaling regulates

lysophosphatidic acid-mediated migration and invasion of human breast tumor cells. *Mol Cancer Res* **2009**, *7* (7), 1064-77.

Chapter 3

Catalyzing Drug Discovery: Harnessing the Power of Artificial Intelligence in Lead Discovery

Part one

Can Machine Learning 'Transform' Peptides/Peptidomimetics into Small Molecules? A Case Study with Ghrelin Receptor Ligands

Abstract:

There has been considerable interest in transforming peptides into small molecules as peptide-based molecules often present poorer bioavailability and lower metabolic stability. This study looked into building machine learning (ML) models to investigate if ML is able to identify the 'bioactive' features of peptides and use the features to accurately discriminate between binding and non-binding small molecules. The Ghrelin receptor (GR), a receptor that is implicated in various diseases, was used as an example to demonstrate whether ML models derived from a peptide library can be used to predict small molecule binders. ML models based on three different algorithms, namely random forest, support vector machine, and extreme gradient boosting, were built based on a carefully curated dataset of peptide/peptidomimetic and small molecule GR ligands. The results indicated that ML models trained with a dataset exclusively composed of peptides/peptidomimetics provide limited predictive power for small molecules, but that ML models trained with a diverse dataset composed of an array of both peptides/peptidomimetics and small molecules displayed exceptional results in terms of accuracy and false rates. The diversified models can accurately differentiate the binding small molecules from non-binding small molecules using an external validation set with new small molecules that were synthesized previously. Structural features that are the most critical contributors to binding activity were extracted and are remarkably consistent with the crystallography and mutagenesis studies.

Graphical abstract:

Peptides/peptidomimetics
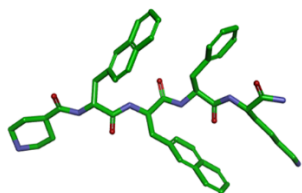
Small molecules
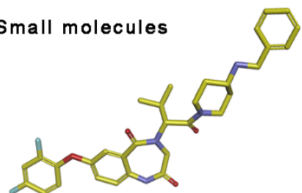
One-hot encoding matrix

0	1	0	1	. . .	0	0	1	0	1	0	0	1
0	0	0	1	. . .	0	1	1	0	1	1	0	1
1	1	0	1	. . .	1	0	1	0	1	0	0	1
0	1	0	0	. . .	0	0	0	1	1	0	1	1
0	1	0	0	. . .	0	0	1	0	1	0	0	1
0	1	1	1	. . .	0	0	1	0	1	0	0	1
. . . .
1	0	0	1	. . .	0	0	1	0	1	0	0	1

Fingerprints:
Morgan, RDKit, MACCS, and PubChem

Machine Learning

New small molecules

Key words: Machine Learning, Drug Design, Ghrelin Receptor, Peptide

120

1. Introduction

In recent years, machine learning (ML) has captured a high degree of attention because of its wide applicability in all branches of research fields. The aim of ML is to develop and apply different algorithms that perform specific tasks, including classification, regression, and clustering, by learning from input data.[1-2] In medicinal chemistry, ML is usually employed to predict categories, such as active/inactive or binding/non-binding, or values, such as IC_{50} or EC_{50}, for test compounds with a specific target by learning from a large collection of training data. Chemical properties of the compounds are transformed into molecular descriptors, which are numerical representations of quantitative physiochemical information for a molecule, in order to be recognised by ML models.[3-4] To date, an increasing number of molecular descriptors have been defined based on different ways of encoding chemical structures. One of the most commonly used types of molecule descriptors in cheminformatics is the fingerprint descriptor, which represents compound information through a chain or vector of bits with varying lengths. Each binary bit vector encodes a specific substructure defined by the fingerprint itself: the presence of the substructure is denoted by 1, and the absence is denoted by 0.[1]

ML models can be categorized into several types based on their algorithms. The most widely used types in medicinal chemistry include decision trees, random forest (RF), support vector machine (SVM), gradient boosting, and different types of neural networks.[1-2] With the increasingly large amount of experimental data available and the development of different ML algorithms, ML models are able to obtain useful information from existing data by characterizing drug molecules based on molecular descriptors. Moreover, successful applications of ML models using different kinds of fingerprints in the field of drug discovery have been reported. For example, Nav1.7 sodium channel inhibitors were accurately identified and optimized based on the prediction from an RF ML model developed with the Chemistry Development Kit (CDK) fingerprint.[5] A deep generative ML model was implemented to facilitate the lead optimization process based on a pyrazolo[3,4-d]pyridazinone scaffold, leading to the successful discovery of potent and selective inhibitors towards discoidin domain receptor 1 (DDR1).[6] Three classification models, namely RF, SVM, and deep neural network (DNN), were built with the extended-connectivity fingerprint (ECFP4) and Molecular ACCess System keys (MACCS) fingerprints to distinguish kinase inhibitors with different binding modes, and with these models,

different kinase inhibitor classes can be predicted with high accuracy.[7] A rapid and effective *de novo* drug design method based on generative tensorial reinforcement learning with the Morgan fingerprint was established for the successful discovery of inhibitors of DDR1.[8] ML approaches have achieved state-of-the-art performance in medicinal chemistry, and predictions from ML models have been able to direct drug design, enabling faster generation of drug candidates and reducing cost and time in the process of drug discovery. The rationale for ML approaches can be applied in all stages of drug development, speeding up the drug discovery process and reducing failure rates.

Ghrelin receptor (GR), previously known as growth hormone secretagogue receptor 1a (GHS-R1a), belongs to the G-protein coupled receptor (GPCR) protein family and is predominantly expressed in the pituitary gland and hypothalamus.[9] Chemical modification of met-enkephalin, an endogenous pentapeptide, led to the discovery of ghrelin as the endogenous ligand for GR, a peptide hormone with 28 amino acids.[10] This ligand is involved in growth hormone metabolism and glucose homeostasis.[11] It has been reported that the expression of GR is significantly increased in several types of cancer, including prostate,[12] breast,[13] and ovarian cancer.[14] Given the wide range of functions that GR displays, the development of drug candidates targeting GR is of great importance both therapeutically and diagnostically. GR ligands with high potency and selectivity have been described in various studies,[15-16] and radiolabelled compounds with nanomolar binding affinity towards GR have also been reported.[17-18] Based on their chemical structures, GR ligands can be divided into two categories: peptides/peptidomimetics and small molecules. Such structurally different types of molecules make the GR ligand pool incredibly diverse as peptides/peptidomimetics and small molecules inherently have large differences in physical and chemical properties.

Peptides have been used as pharmaceuticals for approximately 100 years. Since the discovery of insulin in the 1920s, much research has been performed for the development of other peptide-based drugs.[19] However, as of 2017, peptide drugs only make up a small fraction of drugs available on the market, with around 60 peptide drugs having been approved.[19] This is in stark contrast with the many thousands of drugs that have been approved across major global markets, many of which are small molecule drugs. Although peptides have been shown to generally have higher specificity and lower toxicity than small molecules, there are several drawbacks that limit

the viability of peptides as drug candidates.[20] The most prominent limiting factors for peptide-based drugs are that they have poor stability within the body due to proteases within the gastrointestinal tract, and furthermore, that they generally have difficulties passing through the intestinal mucosal barrier and into the bloodstream.[21] These factors, among others, make it difficult to develop a viable orally administered peptide drug. However, although very few peptide drugs have successfully entered the market, it is believed that they may yet be able to guide future drug development. As previously mentioned, peptides regularly display exceptional specificity for their target. This is largely due to their large size relative to typical small molecule drugs; by having relatively long and varied side chains, peptides can often achieve a very high level of binding with their target by interacting with binding pockets with which small molecules are unable to interact.[21] Ideally, a drug candidate would display a binding affinity similar to that of a peptide while also having the higher stability of small molecules; by combining the key attributes of both peptides and small molecules, it is possible that new drugs with unprecedented levels of both specificity and stability could be developed.

Although some potent peptide/peptidomimetic ligands targeting GR have been discovered,[22-23] they have low stability and incur high levels of degradation in the gastrointestinal tract.[24] It is hoped that ML may be able to determine the key features of both peptides and small molecules and apply them to aid in the design of such drugs. To the best of existing knowledge, no studies have used peptides/peptidomimetics as input data to predict the activity of small molecules with ML. Whether a ML model trained with a compound pool consisting of peptides/peptidomimetics could accurately identify small molecule binders remains unknown.

To address this gap, in this study, different ML models trained with a dataset of peptide/peptidomimetic and small molecule GR ligands split in different strategies were built for the identification of small molecules with binding affinity towards GR. Firstly, binding data from the ChEMBL database was retrieved, compounds with different molecular fingerprints were characterized, and ML classification models were generated based on different splitting strategies. The performance of these models was then evaluated. Models with the best performance were selected to analyse important features that contribute to classification prediction. Key substructures associated with compound binding affinity were identified. The rationality of the models was further confirmed by external validation data.

2. Methods

2.1 Data collection and preparation

Molecules with binding affinity towards GR were collected from the ChEMBL database based on the following criteria: 1) Compounds with a target ChEMBL ID of CHEMBL4616. 2) Compounds with a molecular weight <2000. 3) Compounds with the standard unit being 'nM'. 4) Compounds with the assay type being 'binding'. Duplicates in the dataset were deleted, and the mean value was used for compounds with multiple values from different binding affinity tests. Compounds with $IC_{50}/K_i/K_b$ <1000 nM were considered as ligands with binding affinity, and those with an $IC_{50}/K_i/K_b$ >1000 nM were considered as molecules without binding affinity towards GR. Using this method, a total of 1444 compounds with GR binding data were obtained, among which 1080 compounds were molecules with binding affinity towards GR, while 364 compounds did not display binding affinity towards GR (Table 3.1.1). In order to have a well-balanced dataset to avoid biased prediction, the coverage of non-binding molecules was expanded by randomly selecting 1297 molecules from the ChEMBL database with molecular weights between 320 and 2000 (Table 3.1.1).

Table 3.1.1. Summary of the dataset.

	No. of peptides/ peptidomimetics	No. of small molecules	Total
Binding compounds	312	768	1080
Non-binding compounds	182	182	364
Random compounds	54	1243	1297
Total	548	2193	2741

In order to explore the common scaffold shared by the GR ligands, the binding molecules were clustered based on their Tanimoto coefficients by using the py4cytoscape[25] package in Python (Figure 3.1.1). Tanimoto correlation is the most commonly used algorithm to calculate similarities among molecules by comparing their molecular fingerprints. The Tanimoto value

ranges from 0, when the molecules have nothing in common, to 1, when the molecules are identical. Compounds clustered in groups 2 and 5 (Figure 3.1.1) were defined as peptides/peptidomimetics, and the rest of the compounds were defined as small molecules. The non-binding compounds were defined in the same way (Table 3.1.1).
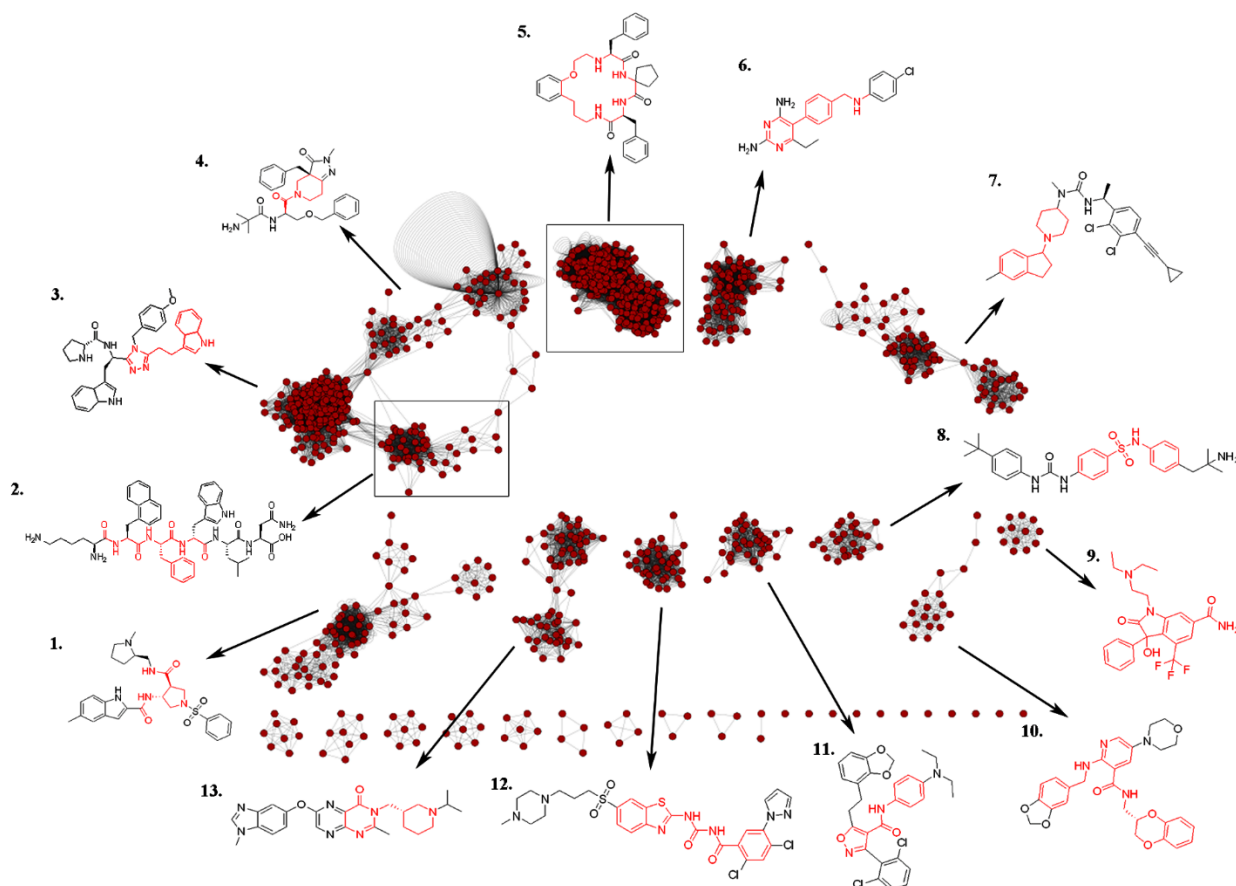
Figure 3.1.1. Binding compounds in the dataset were clustered based on their Tanimoto similarities. The common scaffold in each cluster was highlighted in red. Those in the peptides/peptidomimetics class are highlighted in the rectangular boxes.

To investigate whether ML model derived from different types of compound libraries can be used to accurately identify small molecule binders, four strategies were used to divide the dataset into different training and test sets. For strategy 1, all of the 2741 compounds were randomly split into training, test, and validation sets with a respective ratio of 2:1:1. Therefore, it does not separate peptides/peptidomimetics from small molecules. For strategy 2, the whole dataset was split by compound type. The 548 peptide/peptidomimetic compounds were used as a training set

and the rest of the 2193 small molecules were separated into a test set and a validation set in 1:1 ratio. This strategy was used to see if the ML model derived from the peptide/peptidomimetic library is useful for identifying small molecule binders. For strategy 3, only small molecules were included. Each of the 2193 small molecules were randomly assigned into training, test and validation sets with a ratio of 2:1:1. For strategy 4, all of the 548 peptide/peptidomimetic molecules and 822 randomly chosen small molecules were assigned to a training set, and the rest of the small molecules were evenly distributed into test and validation sets (Table 3.1.2).

All of the data were collected and processed by using the "chembl_webresource_client" package (https://github.com/chembl/chembl_webresource_client)[26] in Python.

Table 3.1.2. Summary of data grouping.

	Training set	Test set	Validation set	Sum
Strategy 1	1370 (mixture)	685 (mixture)	686 (mixture)	2741 (mixture)
Strategy 2	548 (peptides/peptidomimetics)	1096 (small molecules)	1097 (small molecules)	2741 (mixture)
Strategy 3	1096 (small molecules)	548 (small molecules)	549 (small molecules)	2193 (small molecules)
Strategy 4	1370 (mixture)	685 (small molecules)	686 (small molecules)	2741 (mixture)

2.2 Molecular representations

The Morgan, RDKit, MACCS, and PubChem fingerprints were employed due to their excellent performance in previous studies.[7, 27-29] The Morgan fingerprint is a commonly used circular topological descriptor that encodes fragments for a compound with a hashing function based on its atomic neighborhood.[30-31] A radius of 2 and bit size of 1024 were selected for the Morgan fingerprint in this study. The MACCS fingerprint, originally designed for substructure searching, is a fragment fingerprint containing 166 bit of information, each of which encodes the presence or absence of a particular chemical substructure.[32] The RDKit fingerprint is an RDKit-specific topological fingerprint. Inspired by the Daylight fingerprint, this fingerprint generates a bit ID by hashing subgraphs in a compound using bond types and each bond's neighbor count.[33] The bit size was set as 1024 for the RDKit fingerprint. The PubChem fingerprint encodes chemical fragments defined by PubChem with 881 binary bits. The Morgan, MACCS, and RDKit fingerprints were generated by using the open-source cheminformatics tool kit 'RDKit' (version 2021.09.5) in Python.[34] The PubChem fingerprint was generated by using PaDEL-Descriptor.[35] The key information for each fingerprint is summarized in Table 3.1.S1.

To visualize the chemical space of the binding and non-binding compounds in the dataset, t-distributed stochastic neighbor embedding (t-SNE) analyses[36] were performed by using the individual bits as features of each fingerprint. t-SNE is an unsupervised, non-linear technique primarily used for the visualization of high-dimensional data by casting the data into a two- or three-dimensional representation. This technique was widely used in previous medicinal chemistry studies for chemical space analysis.[37-38] The scikit-learn[39] and matplotlib packages[40] in Python were employed for t-SNE analysis and visualization.

2.3 ML model development

For each data splitting strategy, a total of 12 ML classification models were built with a combination of RF, SVM, and eXtrem Gradient Boosting (XGBoost) algorithms and the Morgan, RDKit, MACCS, and PubChem fingerprints.

Random forest. RF is constructed from an ensemble of decision trees built from a bootstrapped sample of training data. For RF models, node splitting is based on a random subset of features for individual trees.[41] The hyperparameters were tuned based on the number of trees in the forest

(n_estimators) and the quality of the split (criterion). The best hyperparameters were determined by setting 'n_estimators' with candidate values of 10, 50, 80, 100, 150, 200, 250, and 280 and by setting the 'criterion' as either gini or entropy. The rest of other parameters were left as default.

Support vector machine. SVM is one of the most commonly used classification algorithms in ML due to its good performance. It constructs a linear or non-linear hyper-plane that separates two classes based on the position and feature values of the data.[42] The SVM kernel was tuned to obtain the best hyperplane by setting the kernel value as either linear, rbf, poly, or sigmoid. The regularization parameter 'C' was optimized using values of 0.5, 0.8, 1.0, 1.5, 1.8, 2.0, and 2.2.

eXtrem Gradient Boosting. XGBoost is based on a gradient boosting decision tree algorithm that is akin to the one used in RF models. Each new model is sequentially created through the improvement of the prior model and then ensembled together to make the final prediction.[43] Similar to RF models, the number of trees (n_estimators) were optimized using candidate values of 10, 50, 100, 150, 200, 250, and 300. The learning rate was tuned from 0.001 to 0.1. The rest of the other parameters were left as default.

The RF and SVM models were built and trained using the scikit-learn package[39] in Python, and the XGBoost model was built and trained using the XGBoost package[43] in Python. Hyperparameters for each model were optimized by validation data using a grid search strategy.

2.4 Performance evaluation

In order to compare the performance of the models and select the best models for further analysis, several metrics were used for model evaluation, including accuracy (ACC), F1 score, Matthews correlation coefficient (MCC), and area under the receiver operating characteristic curve (AUC).

Simply put, accuracy is the fraction of classifications that were correctly predicted by a particular model. F1 score can be regarded as an average of both 'recall', the fraction of true positives that were correctly predicted, and 'precision', the fraction of true positives out of all positive predictions. However, F1 score does not consider false negatives, so the Matthews correlation coefficient was also used. The MCC score is a balanced metric that takes into account true and false positives and negatives, and the score will only be high if both positives and negatives are correctly predicted. AUC score is obtained from integrating the area under a receiver operating

characteristic (ROC) curve and is indicative of a model's ability to distinguish between positives and negatives. The equations that define these metrics are shown below:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

$$F1 = 2 \times \frac{TP}{2TP + FP + FN}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

$$AUC = \int_{x=0}^{1} TPR(FPR - 1(x))dx$$

where TP represents true positives; TN, true negatives; FP, false positives; FN, false negatives; TPR, true positive rate; FPR, false positive rate.

2.5 Feature importance

Based on a game theory approach, SHapley Additive exPlanations (SHAP) values are measurements of feature importance calculated to explain the output of ML models, where the success/total score of a team is considered as the output of a model and the contributions made by each player are considered as the features of the model.[44] This approach has been successfully employed in clinical and drug discovery studies for the identification of crucial features that contribute to ML outputs.[45-46]

To recognise important structural patterns that affect the prediction of binding and non-binding molecules, SHAP values were calculated for each fingerprint bit as a feature value to identify substructures that determine the final prediction of the models.[44] The SHAP values for the RF and XGBoost models were calculated using the TreeExplainer module, and the SHAP values for the SVM models were calculated using the KernelExplainer module in the SHAP package in

Python. The fingerprint bit information was visualized using the DrawMorganBit module in RDKit.

2.6 Molecular docking study

The crystal structure of GR (PDB ID: 6KO5)[47] was used to perform molecular docking studies. The 3D structure of compound **17** was generated using DS Viewer 3.5. Molecular docking was performed using AutoDock Vina.[48] The dimensions of the grid box were set large enough to encompass the whole binding pocket of the GR structure. All other parameters were left as default.

3. Results

3.1 Dataset overview

The general chemical profiles of the whole dataset were overviewed before the data was split and before the models were evaluated. A flowchart for the entire study is shown in Figure 3.1.2. To explore the chemical properties of the compounds in this dataset, several parameters were calculated and compared between binding and non-binding peptides/peptidomimetics and small molecules. These parameters include molecular properties associated with molecule size (molecular weight (MW), and topological polar surface area (TPSA)), hydrophilicity (lipid–water partition coefficient (ALogP)), hydrogen bonding ability (number of hydrogen bond acceptors (NHA) and donors (NHD), number of rotatable bonds (NRB), and number of NHs or OHs), and complexity (fraction of SP^3 hybridized carbons and number of aromatic rings (NAR)). These properties have great impact on the binding affinity of a molecule and were widely used in previous studies to compare compounds of different categories.[5, 49] Furthermore, the relationships between MW, ALogP, and TPSA were paired into scatter plots (Figure 3.1.3A).
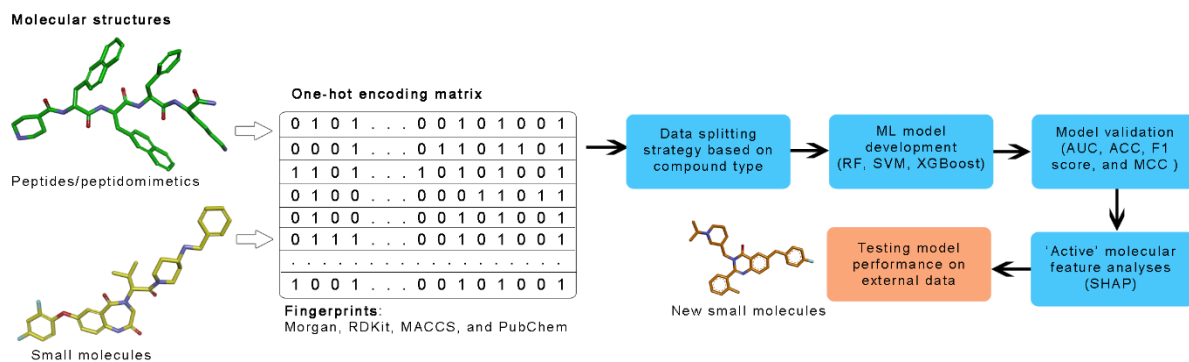
Figure 3.1.2. Schematic summary of the workflow illustrating the course of the study.

The *P* values were calculated by using the Kruskal-Wallis H test[50] to assess the statistical significance among the four groups of molecules. Statistical significance was observed between peptides/peptidomimetics and small molecules, and between binding and non-binding molecules for all these parameters, except for the fraction of sp^3 hybridized carbons, where binding and non-binding molecules showed no difference for both peptides/peptidomimetic and small molecules (Figure 3.1.3B). Although statistical differences were observed for some of the molecular properties between the binding and non-binding compounds, none of these properties could possibly be a crucial factor for the determination of compound binding affinity. It is necessary to build sophisticated ML models to recognize important patterns that differentiate between binding and non-binding compounds.

Figure 3.1.3. Chemical space analysis. A. Scatter plot of MW vs ALogP vs TPSA for binding small molecule (red), non-binding small molecule (blue), binding peptide/peptidomimetic (green), and non-binding peptide/peptidomimetic (purple) compounds in the dataset. B. Violin plots showing the distributions of molecular properties for binding small molecule (red), non-binding small molecule (blue), binding peptide/peptidomimetic (green), and non-binding peptide/peptidomimetic (purple) compounds, including molecular weight (MW), lipid–water partition coefficient (ALogP), fraction of sp^3 hybridized carbons (sp^3), topological polar surface area (TPSA), number of hydrogen bond acceptors (NHA) and donors (NHD), number of rotatable bonds (NRB), number of NHs or OHs (NHOH), and number of aromatic rings (NAR).

3.2 Model performance

For each data splitting strategy, twelve classification ML models were constructed using a combination of RF, SVM, and XGBoost algorithms and four molecular fingerprints. The RF, SVM, and XGBoost methods were selected because they are widely used in the field of medicinal chemistry, and classification and regression tasks performed by these three methods were reported to have high accuracy in previous studies.[45, 51-53]

To obtain models with the best performance, the two most important hyperparameters were tuned by using each of the three validation sets for each model. The results of hyperparameter optimization for each model are shown in Figure 3.1.S1, 3.1.S2, 3.1.S3 and 3.1.S4. The hyperparameter combinations that provided the greatest ACC scores were selected to build each respective model. In SVM models, the choice of kernel had a huge influence on model performance, with either 'poly' or 'rbf' being the best. However, the variation of C values only slightly affected the outcome. In XGBoost models, the scores increased as the number of estimators and learning rate increased. As shown in Figure 3.1.S1, the ACC scores gradually increased from the lower left to upper right.

Overall, good performances for the models based on strategy 1, strategy 3 and strategy 4 were consistently observed in terms of their AUC values (Figure 3.1.4), which range from 0.93 to 0.96, 0.96 to 0.97, and 0.88 to 0.96, respectively. Models based on strategy 3 provided slightly superior performances compared with strategy 1 and strategy 4 in regard to their AUC scores. Models built with strategy 2 provided the worst results, with AUC values ranging from 0.44 to 0.68. For each strategy, the models generated with the Morgan fingerprint were marginally better than other fingerprints, with the AUC values consistently being 0.96 and 0.97 from strategy 1/strategy 4 and strategy 3, respectively (Figure 3.1.4).
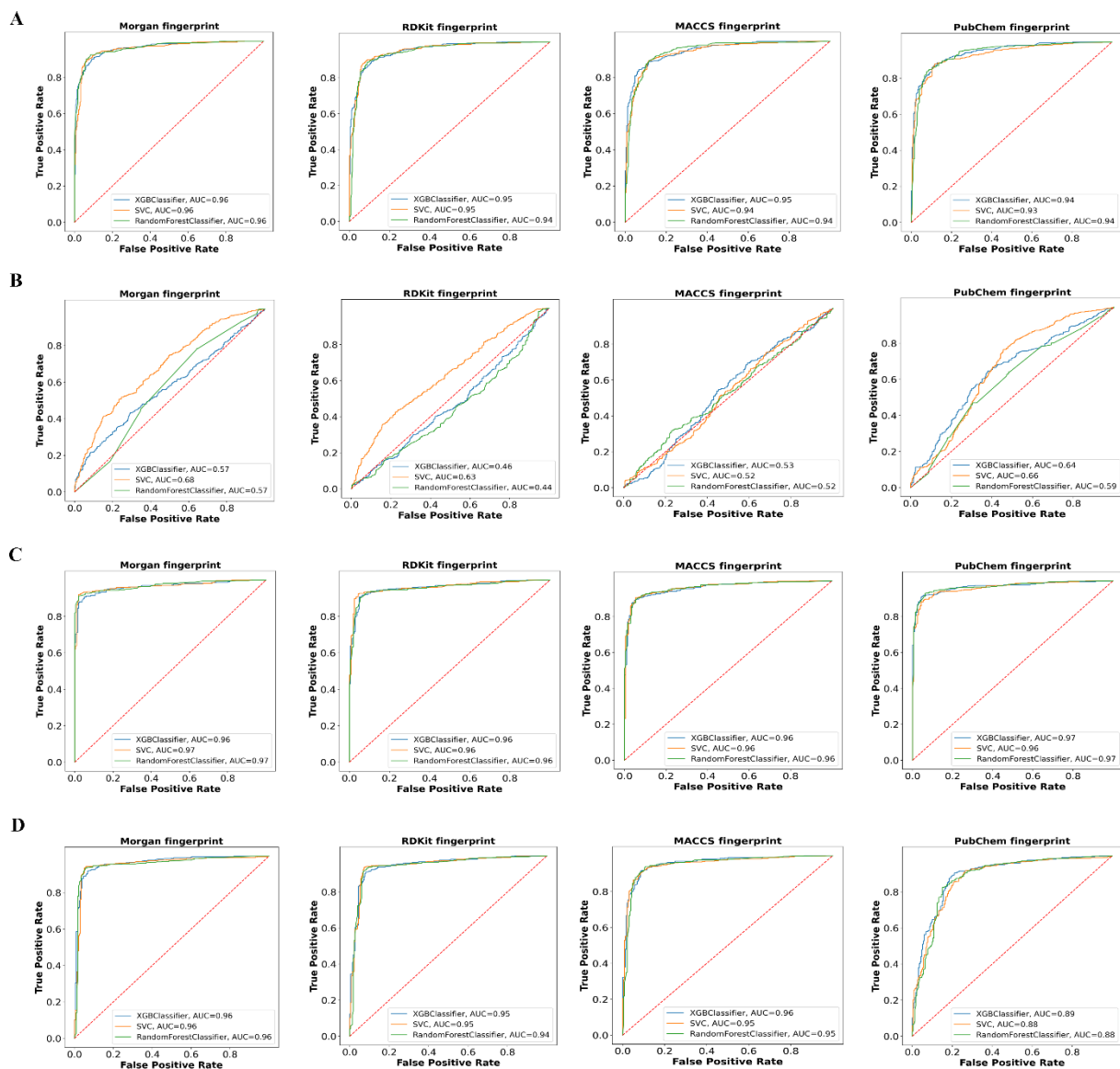
Figure 3.1.4. ROC curves for the models generated with the Morgan, PubChem, RDKit, and MACCS fingerprints using the test sets from strategy 1 (A), strategy 2 (B), strategy 3 (C), and strategy 4 (D). Each graph includes three curves representative of the RF (green), SVM (orange), and XGBoost (blue) classification models.

Given the nearly indistinguishable AUC values from all the models generated with different fingerprint types and ML techniques from each strategy, other widely used metrics, including F1 score and MCC, were also employed to extensively evaluate and compare each model. The metric scores for each model are shown in Table 3.1.3.

Using strategy 1, overall good performances were observed for each model. In particular, the ACC scores were between 0.86 to 0.92, the F1 scores were between 0.83 to 0.93, and the MCC scores were between 0.71 to 0.82. The RF and SVM models built with the same fingerprint demonstrated comparable results in terms of ACC, F1, and MCC scores, whereas the XGBoost models showed slightly worse results. The ACC, F1, and MCC scores of models developed with the Morgan fingerprint were consistently higher than the rest of other models, with the best performing models being RF and SVM.

Using strategy 2, all the metrics were less satisfying. The ACC scores ranged from 0.62 to 0.70, the F1 scores from 0.73 to 0.80, and the MCC scores from 0.01 to 0.31. Although the metrics for strategy 2 are not high compared with those from the other strategies in this study, some of the models may still be used for the identification of small molecule binders. For example, the SVM models built with the Morgan and PubChem fingerprints achieved acceptable results in this series of models with the respective AUC values being 0.68 and 0.66.

Using strategy 3, the scores from each model were consistently high, with the ACC scores ranging from 0.90 to 0.93, F1 scores from 0.91 to 0.95, and MCC scores from 0.79 to 0.86. The three models built with the Morgan fingerprint performed the best compared with models built with the other fingerprints.

Using strategy 4, all the models showed comparable performance with the models from strategy 3. The ACC scores were between 0.83 to 0.93, the F1 scores were between 0.87 to 0.95, and the MCC scores were between 0.63 to 0.84. The models built with the same fingerprint showed nearly equivalent performance. However, the choice of the fingerprint had a bigger influence on the results. Among the four fingerprints, models built with the Morgan fingerprint performed the best, followed by the RDKit and MACCS fingerprint. The PubChem fingerprint afforded the worst performing models.

Table 3.1.3. Summary of model performance using the test sets from strategy 1, strategy 2, strategy 3, and strategy 4.

Strategy	Fingerprint	Model	ACC	F1	MCC	AUC
1	Morgan	RF	0.92	0.93	0.82	0.96
		XGBoost	0.89	0.86	0.76	0.96
		SVM	0.91	0.93	0.82	0.96
	PubChem	RF	0.88	0.90	0.74	0.94
		XGBoost	0.86	0.83	0.71	0.94
		SVM	0.88	0.90	0.75	0.93
	RDKit	RF	0.89	0.91	0.77	0.94
		XGBoost	0.89	0.86	0.77	0.95
		SVM	0.90	0.91	0.79	0.95
	MACCS	RF	0.89	0.91	0.76	0.94
		XGBoost	0.86	0.83	0.71	0.93
		SVM	0.88	0.90	0.76	0.94
2	Morgan	RF	0.66	0.79	0.05	0.57
		XGBoost	0.66	0.80	0.04	0.50
		SVM	0.68	0.80	0.16	0.68
	PubChem	RF	0.63	0.74	0.11	0.59
		XGBoost	0.63	0.73	0.14	0.60
		SVM	0.70	0.78	0.31	0.66
	RDKit	RF	0.66	0.80	0.07	0.44

		XGBoost	0.66	0.79	0.05	0.51
		SVM	0.67	0.80	0.14	0.63
	MACCS	RF	0.62	0.76	0.01	0.52
		XGBoost	0.64	0.77	0.02	0.55
		SVM	0.65	0.78	0.04	0.52
3	Morgan	RF	0.92	0.94	0.84	0.97
		XGBoost	0.93	0.91	0.85	0.97
		SVM	0.93	0.95	0.86	0.97
	PubChem	RF	0.92	0.93	0.82	0.97
		XGBoost	0.91	0.93	0.82	0.97
		SVM	0.91	0.93	0.81	0.96
	RDKit	RF	0.91	0.93	0.81	0.96
		XGBoost	0.92	0.94	0.83	0.96
		SVM	0.93	0.94	0.84	0.96
	MACCS	RF	0.91	0.93	0.81	0.96
		XGBoost	0.90	0.92	0.79	0.96
		SVM	0.92	0.93	0.82	0.96
4	Morgan	RF	0.93	0.94	0.84	0.96
		XGBoost	0.91	0.87	0.80	0.96
		SVM	0.93	0.95	0.84	0.96
	PubChem	RF	0.85	0.89	0.65	0.88
		XGBoost	0.83	0.87	0.63	0.88

	SVM	0.84	0.88	0.64	0.88
RDKit	RF	0.90	0.93	0.78	0.94
	XGBoost	0.91	0.93	0.80	0.95
	SVM	0.92	0.94	0.83	0.95
MACCS	RF	0.92	0.94	0.82	0.95
	XGBoost	0.91	0.93	0.79	0.96
	SVM	0.91	0.93	0.80	0.95

From the four data splitting strategies, comparably high levels of accuracy were consistently achieved by models that employed strategy 1, strategy 3 and strategy 4, with scores from strategy 3 being marginally higher. The lowest scores were observed with strategy 2, where 548 peptides/peptidomimetics were used to train models for the prediction of a large set of small molecules. Such low scores were achieved partially due to the fact that the training set for this strategy was composed of a small number of molecules. Based on such a limited size of training set, the models were not able to learn sufficient information to produce powerful predictions. Another possible reason may be that the structural diversity of the training set was fairly limited since only peptide/peptidomimetic compounds were used to train the models with strategy 2. It has been reported that increasing the number of structurally diverse compounds resulted in increased model performance.[27] Therefore, models developed with the homogenous peptide/peptidomimetic training set of strategy 2 yielded poor results compared with models developed with the more structurally diverse training sets used for strategy 1, strategy 3 and strategy 4. However, when the models were trained with a combination of peptides/peptidomimetics and small molecules, all the metric scores were significantly improved, as was shown with the use of strategy 4. These results further suggest that the size of available data in hand and structural diversity of the dataset influence model performance.

Taken together, a ML model trained solely with a limited number of peptides/peptidomimetics may not be able to identify small molecule binders. However, a ML model trained with a diverse

dataset composed of both peptides/peptidomimetics and small molecules has the ability to accurately identify small molecule binders.

The choice of molecular fingerprint is another factor that affects model performance. The RF, XGBoost, and SVM models provided comparable results within the same fingerprint type, suggesting that these three ML models perform equally well, irrespective of their different algorithms. However, different molecular fingerprints yielded different model performances. Among the four fingerprint types, the models built with the Morgan fingerprint provided superior performance in terms of AUC, F1, ACC, and MCC values compared with the rest of the other fingerprints, followed by the RDKit fingerprint. Models developed with the PubChem and MACCS fingerprints provided slightly lower values with respect to ACC, F1, and MCC scores compared with the Morgan and RDKit fingerprints.

To further analyze the dissimilar performances resulting from different fingerprint types, the four fingerprints were explored and visualized using t-SNE.[36] The t-SNE plots in Figure 3.1.5 suggest that the binding molecules could be better discriminated from the non-binding molecules by using the Morgan and RDKit fingerprints, whereas the binding and non-binding compounds are less distinguishable by using the MACCS and PubChem fingerprints. Additionally, small molecules and peptides/peptidomimetics were distributed in different clusters from these t-SNE plots, suggesting that these two types of molecules show great structural dissimilarity. Such structural dissimilarity may be another reason for the poor performance observed from the models using strategy 2.

Altogether, the Morgan fingerprint produced the best performing models regardless of the ML algorithms. Importantly, it has been shown that ML models trained with a mixture of peptides/peptidomimetics and small molecules are able to provide accurate prediction of the binding category of small molecules. Therefore, the important features of the Morgan fingerprint were further assessed using strategy 4.
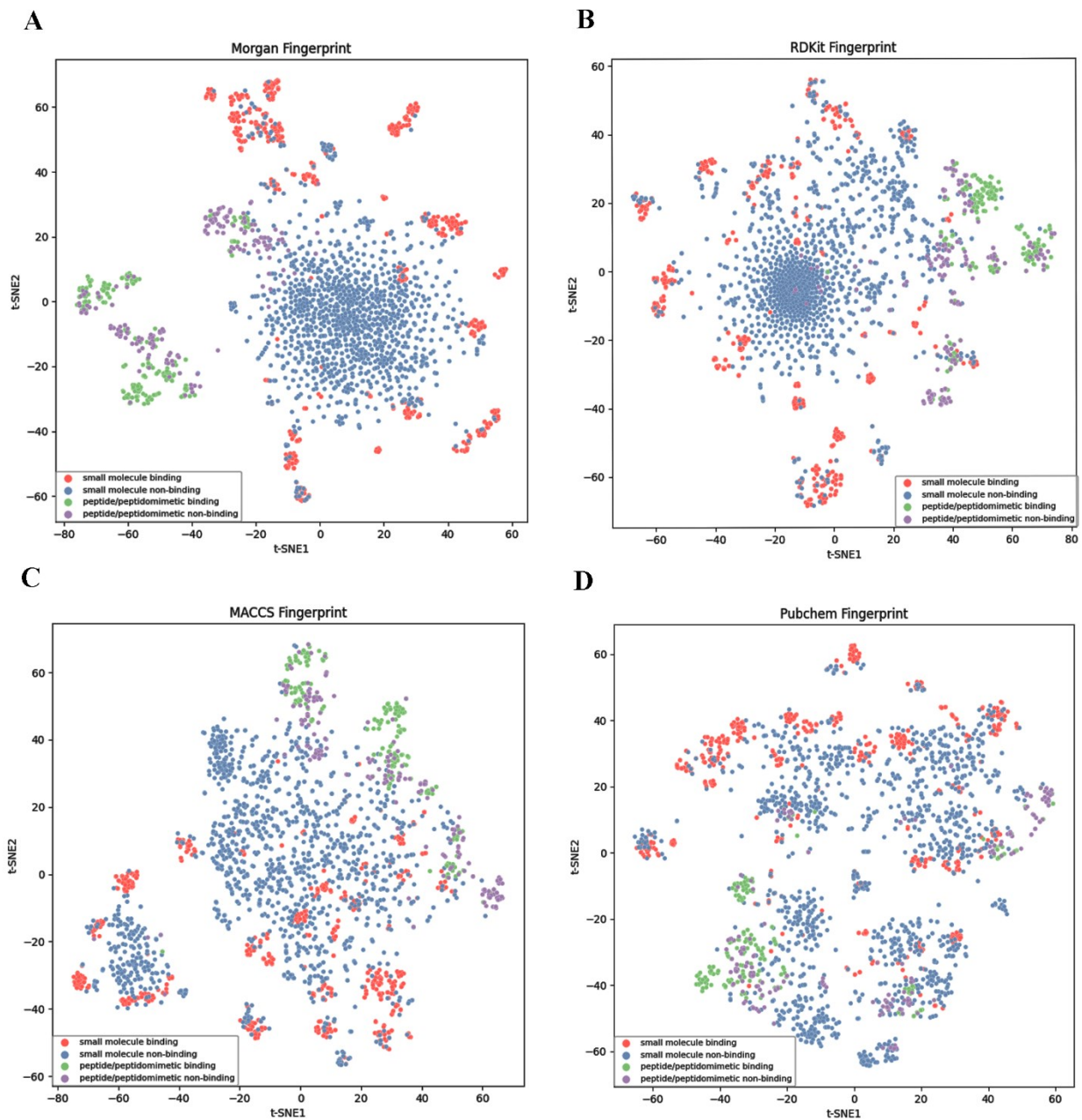
Figure 3.1.5. The t-SNE analysis of the compounds using the Morgan (A), RDKit (B), MACCS (C), and PubChem (D) fingerprints. The red dots indicate binding small molecule, blue: non-binding small molecule; green: binding peptide/peptidomimetic; and purple: non-binding peptide/peptidomimetic compounds.

3.3 Important substructures

In order to interpret the binding predictions from ML models, the important features from models built with the Morgan fingerprint were identified and analysed to present the structural patterns that contribute to the predicted probability of binding category. SHAP[44] was introduced to explain the predictions made by the models by calculating the contribution of each feature.

The Morgan fingerprint bit information was translated to chemical structure features that determine binding category predictions. The top 20 features for the three classification models built with the Morgan fingerprint using strategy 1 (Figure 3.1.S5A), strategy 2 (Figure 3.1.S5B), strategy 3 (Figure 3.1.S5C), and strategy 4 (Figure 3.1.6A, Table 3.1.S2) are shown. Furthermore, the features that are present in all of the three models in each strategy were further extracted. Venn diagrams of the 20 most important features for the models from strategy 1 (Figure 3.1.S6A), strategy 2 (Figure 3.1.S6B), strategy 3 (Figure 3.1.S6C), and strategy 4 (Figure 3.1.6B) are shown. The overlapping sets include 4, 2, 7, and 5 common features from strategy 1, strategy 2, strategy 3, and strategy 4, respectively. As shown in Figure 3.1.6A, almost all of the top 20 features are positive Morgan features that are positively associated with the binding affinity of the GR ligands.

The top-ranked features from strategy 4 can be divided into several categories based on structural similarities. The most prevalent category contains N-bearing conjugated systems, which includes bits 90, 104, 119, 121, 189, 328, 364, 476, 758, 802, 806, 831, 855, and 913. The second most prevalent category contains amine groups, which includes primary, secondary, and tertiary amines. Bits 117, 128, 285, 366, 751, 881, and 896 belong to this category. The next category is the oxygen-containing functional group, which includes bits 169, 320, 650, and 695. The last category contains aliphatic and aromatic carbons, including bits 4, 33, 212, 291, 305, 310, 319, 325, 341, 366, 389, 648, 730, 777 and 878. Conjugated ring systems, amine groups, and oxygen-containing groups are usually active moieties involved in ligand-receptor interactions which contribute to ligand binding affinity. Aliphatic and aromatic carbons, on the other hand, provide hydrophobic environments that interact with the nonpolar parts of GR.

The feature weights from the RF, XGBoost, and SVM models from strategy 4 for one exemplary compound is shown using the SHAP force plot (Figure 3.1.6C). The positive and negative feature contributions are drawn with respective sequential red and blue arrows. The size of each

arrow is proportionate to the SHAP value for the corresponding feature. The top-ranked topological features were mapped onto the displayed compound. The majority of the important features for this compound are patterns that belong to N-containing conjugated ring systems and amine groups.
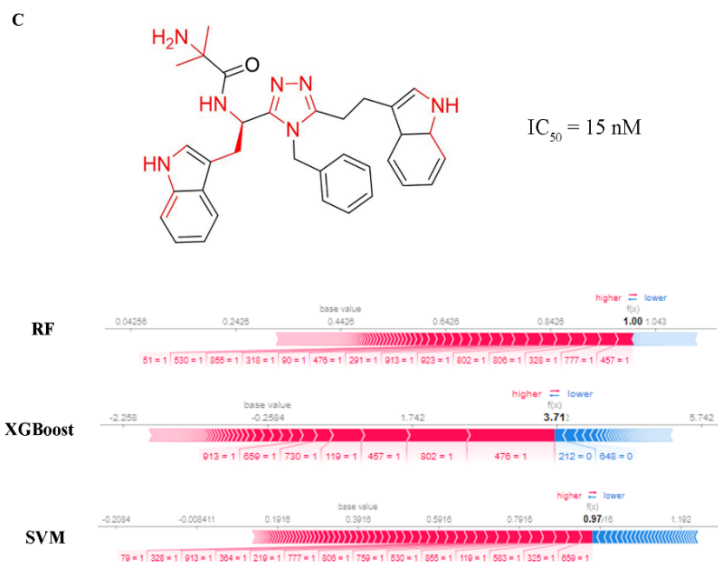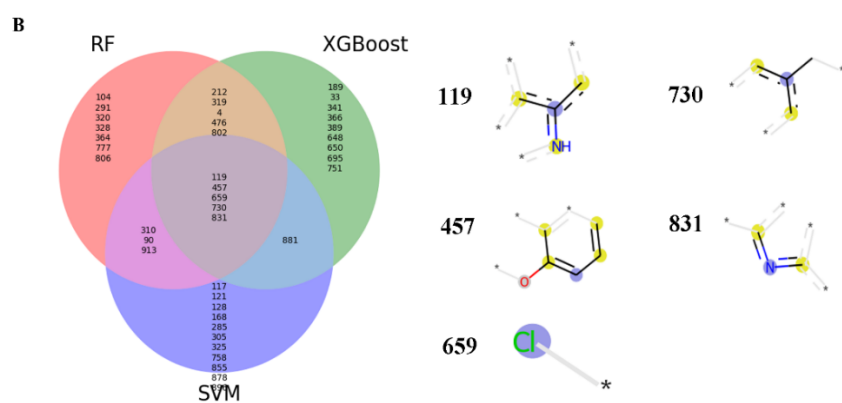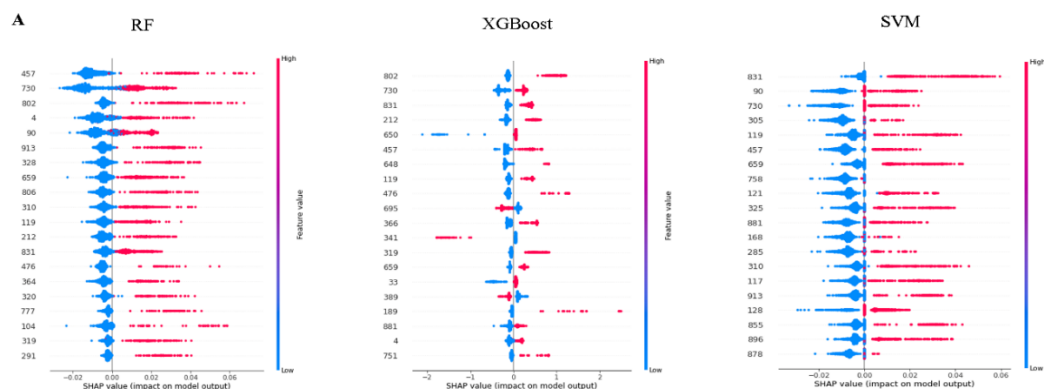
Figure 3.1.6. A. The top 20 important features ranked by SHAP for the RF, XGBoost, and SVM models using strategy 4. B. Venn diagrams demonstrating the features of bit information that are

commonly present among the RF, XGBoost, and SVM models using strategy 4 and their corresponding substructures (the central atom is highlighted in blue, and aromatic atoms are highlighted in yellow. Atoms/bonds in light gray indicate pieces of the structure that affect the connectivity invariants of the atoms, but they are not directly part of the fingerprint).[33] C. Exemplary compound demonstrating top-ranked features for the Morgan fingerprint. The highly weighted features are highlighted in red in the structure of the compound.

3.4 External validation

Recently, a series of fluorine-bearing quinazolinone small molecule derivatives as GR antagonists with high binding affinity towards GR was reported.[17-18] This series of compounds has not yet been included in the ChEMBL database, so these compounds were not used for ML model development. Instead, these compounds were used as external data to additionally investigate the performance of the best model from strategy 2, strategy 3, and strategy 4 in order to test if these models could be used to identify small molecules with binding affinity in a real-world situation. Since only one non-binding GR ligand was found in this study, a total of 32 random small molecules were added to the external validation dataset as non-binding ligands to offset any bias. The ACC, F1 score, and MCC were determined for the external dataset, and important features were mapped onto one compound (**17**).

The results of the metrics using the external data were shown in Table 3.1.4. Excellent performances were observed from SVM models using strategy 3 and strategy 4. Using strategy 2, the binding category for only 33 out of 62 compounds was correctly predicted using SVM. However, the binding category for 54 out of 62 compounds was correctly predicted from strategy 3, demonstrating SVM model possessed very high accuracy. The highest accuracy was achieved using strategy 4, where 57 out of 62 compounds were correctly predicted (Table 3.1.S3). Table 3.1.4 also shows that the SVM model using strategy 2 demonstrated less satisfactory performance. The model trained with strategy 4 in this study performed the best in terms of ACC, F1, and MCC scores. The model trained with strategy 3 showed marginally lower scores compared with that using strategy 4. These results are similar to the ones obtained with the test dataset (Table 3.1.3), which further suggests that a ML model trained with a large number of peptides/peptidomimetics and small molecules is able to yield trustworthy predictions for the binding category of small molecules, but a ML model trained with a limited number of

peptides/peptidomimetics provides unsatisfactory prediction power. Overall, the external validation further proved that models built with strategy 3 and strategy 4 are robust, and predictions from the models are reliable and capable of predicting true positive and negative molecules with high accuracy.

Table 3.1.4. Model performance using the external data.

Strategy	Model	ACC	F1	MCC
2	SVM_Morgan	0.53	0	0
3	SVM_Morgan	0.87	0.85	0.75
4	SVM_Morgan	0.92	0.93	0.84

The top-ranked features for a representative compound (**17**) that were correctly predicted by the SVM model from both strategy 3 and strategy 4 are shown using the SHAP force plot, and all of these top-ranked features are from the piperidine and quinazolinone scaffold of the compound (Figure 3.1.7A). To further confirm the important features found in compound **17** by ML, a molecular docking study was performed. Recently, the crystal structure of GR bound to a ligand with a quinazolinone scaffold was revealed.[47] This co-crystalized quinazolinone derivative compound (**21**) in complex with GR has a very similar structure to compound **17**. Therefore, this crystal structure (PDB ID: 6KO5)[47] was used to perform molecular docking studies as shown in Figure 3.1.7B. The crucial interactions revealed by the crystal structure of compound **21**-bound GR include 1) a π–cation interaction between the 5-azaquinazolinone moiety of compound **21** and ARG102; 2) hydrophobic interactions between the isopropylpiperidine moiety and residues PHE279, PHE309, and PHE312; and 3) a salt bridge interaction between the positively charged amine in the isopropylpiperidine moiety due to protonation under physiological conditions and ASP99 (Figure 3.1.7B). Further mutagenesis studies have showed that alanine mutations of ASP99 and ARG102 abolished the receptor activity, confirming the critical roles of the 5-azaquinazolinone moiety and the isopropylpiperidine moiety. The docking studies showed that compound **17** is able to bind to the orthosteric pocket of GR in an identical fashion as compound **21** (Figure 3.1.7C). All the crucial interactions involving the quinazolinone moiety and the

isopropylpiperidine moiety are preserved. The results show that the ML models derived from strategy 3 and 4 are able to precisely identify the most crucial substructural features responsible for receptor binding.
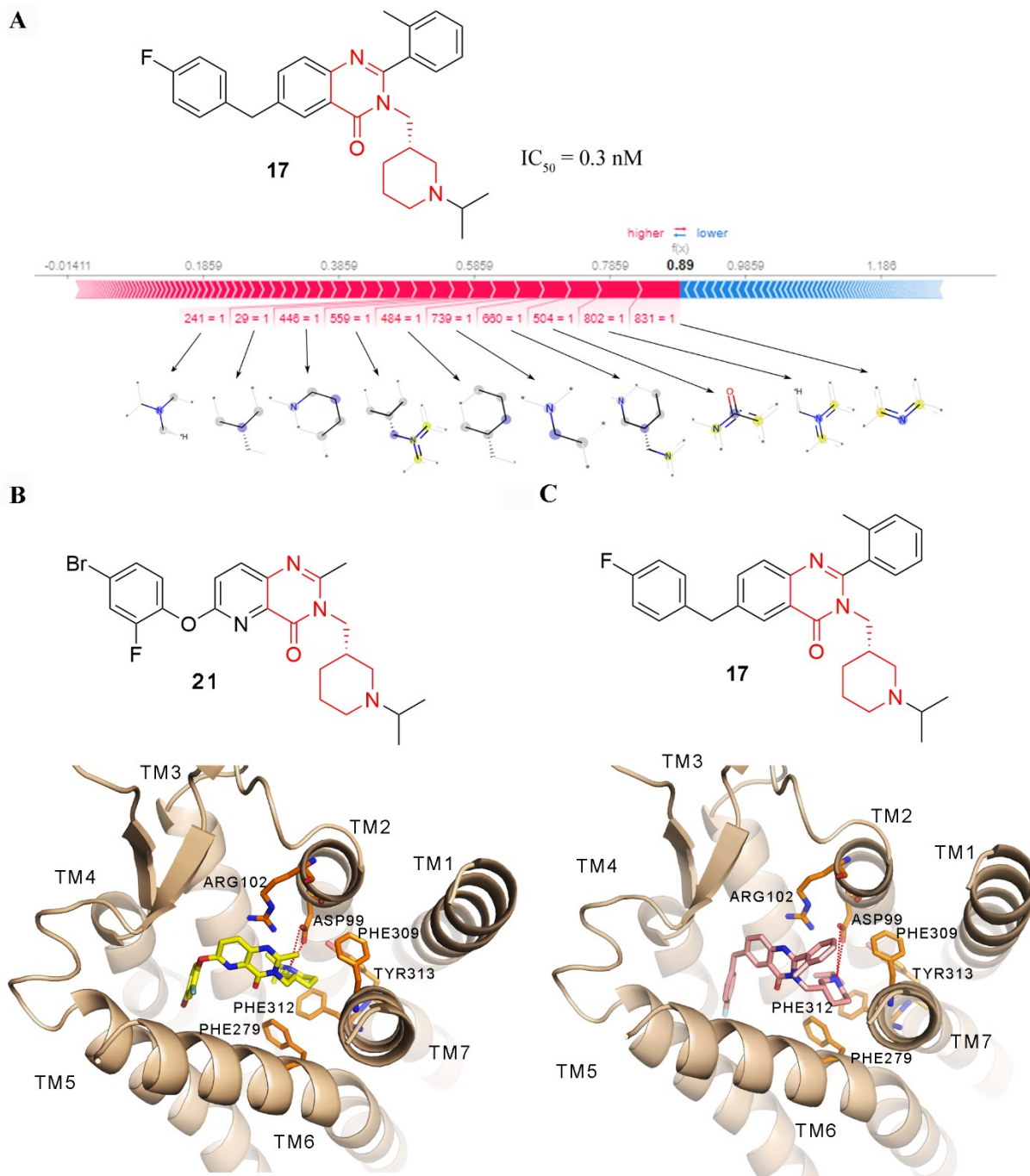
Figure 3.1.7. A. Highly weighted features for one representative compound (**17**) in the validation set, whose binding affinity of 0.3 nM was determined using a radio-ligand binding assay. The binding category of this compound was correctly predicted by the SVM model using strategy 3 and strategy 4. The substructures of the most highly weighted features are shown below the force plot (the central atom is highlighted in blue, and aromatic atoms are highlighted in yellow. Atoms/bonds in light gray indicate pieces of the structure that affect the connectivity invariants of the atoms, but they are not directly part of the fingerprint). The highly weighted features are highlighted in red in the structure of the compound. B. The crystal structure of compound **21** in complex with GR. The ligand and key residues are shown in yellow and orange sticks, respectively. C. The binding mode of compound **17** from the molecular docking study. The ligand and key residues are shown in pink and orange sticks, respectively.

4. Conclusions

In this study, different ML methods including RF, XGBoost, and SVM were developed with training sets of distinct size and chemical composition in an attempt to distinguish GR binding small molecules from those without any binding affinity. To the best of existing knowledge, this is the first study on the development of ML models based on a dataset of compounds that could be split into peptides/peptidomimetics and small molecules. Robust models were obtained, as demonstrated by different performance metrics. Highly weighted features contributing to positive predictions were identified and mapped onto representative compounds. The correct identification of positive features provided a rationale for successful classification tasks. Furthermore, the best performing models in this study were further successfully validated by using 62 new small molecules, and the major substructures contributing to the receptor-ligand interactions were further supported by the crystal structure of GR in complex with a small molecule that has similar structure with the molecules in the validation set.

It was found that the size of the dataset, the structural diversity of the compounds, and the choice of molecular fingerprint influence the model performance the most. ML models trained with a combination of peptides/peptidomimetics and small molecules can be used to identify small molecule binders. However, whether ML models trained solely with peptides/peptidomimetics are able to accurately perform these predictions remains to be seen. More available peptides/peptidomimetics need to be included in future studies to determine the validity of using

these models trained solely with peptides/peptidomimetics. Furthermore, in this study, only the one-hot encoding vectorization method was used for molecule representation. However, this one-hot encoding method is sparse and high-dimensional, and similar entities are not placed close to one another in vector space by this method.[54] As is shown in Figure 3.1.5, small molecules and peptides/peptidomimetics were plotted in different spaces from t-SNE analyses. Such dissimilarity rendered by the one-hot encoding method also led to the poor performance of ML models built solely with peptides/peptidomimetics as with strategy 2. This is the major limitation in this study. As ML models' ability to learn is affected by the conversion from a molecular structure to a vector representation, better representations of molecular structure (i.e., graph embedding representation)[55] need to be developed in the future studies in order to effectively 'transform' peptide-based molecules into small molecules. Nevertheless, it has been shown that some of the models in this study have great potential to guide the drug design of small molecules, with key attributes of both peptides and small molecules, for the ghrelin receptor. The discovery and development of compounds displaying advantages of both peptides and small molecules remains a difficult task. Therefore, the trained RF and SVM models from strategy 4 are made available upon request.

5. References

1. Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; Zhao, S., Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov* **2019,** *18* (6), 463-477.

2. Yang, X.; Wang, Y.; Byrne, R.; Schneider, G.; Yang, S., Concepts of Artificial Intelligence for Computer-Assisted Drug Discovery. *Chem Rev* **2019,** *119* (18), 10520–10594.

3. Raschka, S.; Kaufman, B., Machine learning and AI-based approaches for bioactive ligand discovery and GPCR-ligand recognition. *Methods* **2020,** *180*, 89-110.

4. Carracedo-Reboredo, P.; Linares-Blanco, J.; Rodriguez-Fernandez, N.; Cedron, F.; Novoa, F. J.; Carballal, A.; Maojo, V.; Pazos, A.; Fernandez-Lozano, C., A review on machine learning approaches and trends in drug discovery. *Comput Struct Biotechnol J* **2021,** *19*, 4538-4558.

5. Kong, W.; Tu, X.; Huang, W.; Yang, Y.; Xie, Z.; Huang, Z., Prediction and Optimization of NaV1.7 Sodium Channel Inhibitors Based on Machine Learning and Simulated Annealing. *J Chem Inf Model* **2020,** *60* (6), 2739-2753.

6. Tan, X.; Li, C.; Yang, R.; Zhao, S.; Li, F.; Li, X.; Chen, L.; Wan, X.; Liu, X.; Yang, T.; Tong, X.; Xu, T.; Cui, R.; Jiang, H.; Zhang, S.; Liu, H.; Zheng, M., Discovery of Pyrazolo[3,4-d]pyridazinone Derivatives as Selective DDR1 Inhibitors via Deep Learning Based Design, Synthesis, and Biological Evaluation. *J Med Chem* **2022,** *65* (1), 103-119.

7. Miljkovic, F.; Rodriguez-Perez, R.; Bajorath, J., Machine Learning Models for Accurate Prediction of Kinase Inhibitors with Different Binding Modes. *J Med Chem* **2020,** *63* (16), 8738-8748.

8. Zhavoronkov, A.; Ivanenkov, Y. A.; Aliper, A.; Veselov, M. S.; Aladinskiy, V. A.; Aladinskaya, A. V.; Terentiev, V. A.; Polykovskiy, D. A.; Kuznetsov, M. D.; Asadulaev, A.; Volkov, Y.; Zholus, A.; Shayakhmetov, R. R.; Zhebrak, A.; Minaeva, L. I.; Zagribelnyy, B. A.; Lee, L. H.; Soll, R.; Madge, D.; Xing, L.; Guo, T.; Aspuru-Guzik, A., Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat Biotechnol* **2019,** *37* (9), 1038-1040.

9.   Hedegaard, M. A.; Holst, B., The Complex Signaling Pathways of the Ghrelin Receptor. *Endocrinology* **2020,** *161* (4), bqaa020.

10.   Müller, T. D.; Nogueiras, R.; Andermann, M. L.; Andrews, Z. B.; Anker, S. D.; Argente, J.; R.L. Batterham; Benoit, S. C.; Bowers, C. Y.; Broglio, F.; Casanueva, F. F.; D'Alessio, D.; Depoortere, I.; A. Geliebter; Ghigo, E.; Cole, P. A.; Cowley, M.; Cummings, D. E.; Dagher, A.; Diano, S.; Dickson, S. L.; Diéguez, C.; Granata, R.; Grill, H. J.; Grove, K.; Habegger, K. M.; Heppner, K.; Heiman, M. L.; Holsen, L.; Holst, B.; Inui, A.; Jansson, J. O.; Kirchner, H.; Korbonits, M.; Laferrère, B.; LeRoux, C. W.; Lopez, M.; Morin, S.; Nakazato, M.; Nass, R.; Perez-Tilve, D.; Pfluger, P. T.; Schwartz, T. W.; Seeley, R. J.; Sleeman, M.; Sun, Y.; Sussel, L.; Tong, J.; Thorner, M. O.; Lely, A. J. v. d.; Ploeg, L. H. T. v. d.; Zigman, J. M.; Kojima, M.; Kangawa, K.; Smith, R. G.; Horvath, T.; Tschöp, M. H., Ghrelin. *Mol Metab* **2015,** *4* (6), 437-460.

11.   Poher, A.-L.; Tschöp, M. H.; Müller, T. D., Ghrelin regulation of glucose metabolism. *Peptides* **2018,** *100*, 236–242.

12.   Lu, C.; McFarland, M. S.; Nesbitt, R.-L.; Williams, A. K.; Chan, S.; Gomez-Lemus, J.; Autran-Gomez, A. M.; Al-Zahrani, A.; Chin, J. L.; Izawa, J. I.; Luyt, L. G.; Lewis, J. D., Ghrelin receptor as a novel imaging target for prostatic neoplasms. *Prostate* **2012,** *72* (8), 825-833.

13.   Zhang, J.; Xie, T., Ghrelin inhibits cisplatin-induced MDA-MB-231 breast cancer cell apoptosis via PI3K/Akt/mTOR signaling. *Exp Ther Med* **2020,** *19* (3), 1633–1640.

14.   Gaytan, F.; Morales, C.; Barreiro, M. L.; Jeffery, P.; Chopin, L. K.; Herington, A. C.; Casanueva, F. F.; Aguilar, E.; Dieguez, C.; Tena-Sempere, M., Expression of growth hormone secretagogue receptor type 1a, the functional ghrelin receptor, in human ovarian surface epithelium, mullerian duct derivatives, and ovarian tumors. *J Clin Endocrinol Metab* **2005,** *90* (3), 1798-1804.

15.   Hanrahan, P.; Bell, J.; Bottomley, G.; Bradley, S.; Clarke, P.; Curtis, E.; Davis, S.; Dawson, G.; Horswill, J.; Keily, J.; Moore, G.; Rasamison, C.; Bloxham, J., Substituted azaquinazolinones as modulators of GHSr-1a for the treatment of type II diabetes and obesity. *Bioorg Med Chem Lett* **2012,** *22*, 2271-2278.

16.   Moulin, A.; Brunel, L.; Boeglin, D.; Demange, L.; Ryan, J.; M'Kadmi, C. l.; Denoyelle, S. v.; Martinez, J.; Fehrentz, J.-A., The 1,2,4-triazole as a scaffold for the design of

ghrelin receptor ligands: development of JMV 2959, a potent antagonist. *Amino Acids* **2013,** *44*, 301-314.

17. Hou, J.; Kovacs, M. S.; Dhanvantari, S.; Luyt, L. G., Development of Candidates for Positron Emission Tomography (PET) Imaging of Ghrelin Receptor in Disease: Design, Synthesis, and Evaluation of Fluorine-Bearing Quinazolinone Derivatives. *J Med Chem* **2018,** *61*, 1261-1275.

18. Luyt, L. G.; HOU, J. QUINAZOLINONE DERIVATIVES USEFUL FOR IMAGING. US 11186571. 2021.

19. Lau, J. L.; Dunn, M. K., Therapeutic peptides: Historical perspectives, current development trends, and future directions. *Bioorg Med Chem* **2018,** *26*, 2700–2707.

20. Jr, L. O.; Wade, J. D., Current challenges in peptide-based drug discovery. *Front Chem* **2014,** *2*, 1-4.

21. Lundquist, P.; Artursson, P., Oral absorption of peptides and nanoparticles across the human intestine: Opportunities, limitations and studies in human tissues. *Adv Drug Deliv Rev* **2016,** *106*, 256-276.

22. M'Kadmi, C.; Cabral, A.; Barrile, F.; Giribaldi, J.; Cantel, S.; Damian, M.; Mary, S.; Denoyelle, S.; Dutertre, S.; Péraldi-Roux, S.; Neasta, J.; Oiry, C.; Banères, J.-L.; Marie, J.; Perello, M.; Fehrentz, J.-A., N-Terminal Liver-Expressed Antimicrobial Peptide 2 (LEAP2) Region Exhibits Inverse Agonist Activity toward the Ghrelin Receptor. *J Med Chem* **2019,** *62* (2), 965-973.

23. Hou, J.; Charron, C. L.; Fowkes, M. M.; Luyt, L. G., Bridging computational modeling with amino acid replacements to investigate GHS-R1a-peptidomimetic recognition. *Eur J Med Chem* **2016,** *123*, 822-833.

24. Giorgioni, G.; Bello, F. D.; Quaglia, W.; Botticelli, L.; Cifani, C.; Bonaventura, E. M. D.; Bonaventura, M. V. M. D.; Piergentili, A., Advances in the Development of Nonpeptide Small Molecules Targeting Ghrelin Receptor. *J Med Chem* **2022,** *65*, 3098−3118.

25. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T., Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **2003,** *13* (22), 2498-2504.

26. Davies, M.; Nowotka, M.; Papadatos, G.; Dedman, N.; Gaulton, A.; Atkinson, F.; Bellis, L.; Overington, J. P., ChEMBL web services: streamlining access to drug discovery data and utilities. *Nucleic Acids Res* **2015,** *43*, W612–W620.

27. Siramshetty, V. B.; Chen, Q.; Devarakonda, P.; Preissner, R., The Catch-22 of Predicting hERG Blockade Using Publicly Accessible Bioactivity Data. **2018**.

28. Siramshetty, V. B.; Nguyen, D. T.; Martinez, N. J.; Southall, N. T.; Simeonov, A.; Zakharov, A. V., Critical Assessment of Artificial Intelligence Methods for Prediction of hERG Channel Inhibition in the "Big Data" Era. *J Chem Inf Model* **2020,** *60* (12), 6007-6019.

29. Fan, T.; Sun, G.; Zhao, L.; Cui, X.; Zhong, R., QSAR and Classification Study on Prediction of Acute Oral Toxicity of N-Nitroso Compounds. *Int J Mol Sci* **2018,** *19* (10), 3015.

30. Rogers, D.; Hahn, M., Extended-Connectivity Fingerprints. *J Chem Inf Model* **2010,** *50*, 742–754.

31. Morgan, H. L., The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *J Chem Doc* **1965,** *5* (2), 107-113.

32. MACCS structural keys. *Accelrys, San Diego, CA* **2011**.

33. The RDKit Book. *https://www.rdkit.org/docs/RDKit_Book.html*.

34. RDKit: Cheminformatics and Machine Learning Software, 2013. *http://www.rdkit.org*.

35. Yap, C. W., PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J Comput Chem* **2011,** *32* (7), 1466-1474.

36. Maaten, L. v. d.; Hinton, G., Visualizing Data using t-SNE. *J Mach Learn Res* **2008,** *9*, 2579-2605

37. Miljkovic, F.; Martinsson, A.; Obrezanova, O.; Williamson, B.; Johnson, M.; Sykes, A.; Bender, A.; Greene, N., Machine Learning Models for Human *In Vivo* Pharmacokinetic Parameters with In-House Validation. *Mol Pharm* **2021,** *18* (12), 4520-4530.

38. Hou, T.; Bian, Y.; McGuire, T.; Xie, X.-Q., Integrated Multi-Class Classification and Prediction of GPCR Allosteric Modulators by Machine Learning Intelligence. *Biomolecules* **2021,** *11* (6), 870.

39.   Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E., Scikit-learn: Machine Learning in Python. *J Mach Learn Res* **2011,** *12*, 2825-2830.

40.   JD, H., Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* **2007,** *9* (3), 90-95.

41.   Breiman, L., Random Forests. *Mach Learn* **2001,** *45*, 5–32.

42.   Vapnik, V. N., The Nature of Statistical Learning Theory. *Springer: New York* **2000**.

43.   Chen, T.; Guestrin, C., XGBoost: A Scalable Tree Boosting System. *arXiv:1603.02754v3* **2016**.

44.   Lundberg, S. M.; Lee, S.-I., A Unified Approach to Interpreting Model Predictions. NIPS, Long Beach, CA, USA. **2017**, 4768–4777.

45.   Rodríguez-Pérez, R.; Bajorath, J. r., Interpretation of Compound Activity Predictions from Complex Machine Learning Models Using Local Approximations and Shapley Values. *J Med Chem* **2020,** *63* (16), 8761−8777.

46.   Wang, K.; Tian, J.; Zheng, C.; Yang, H.; Ren, J.; Liu, Y.; Han, Q.; Zhang, Y., Interpretable prediction of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP. *Comput Biol Med* **2021,** *137*, 104813.

47.   Shiimura, Y.; Horita, S.; Hamamoto, A.; Asada, H.; Hirata, K.; Tanaka, M.; Mori, K.; Uemura, T.; Kobayashi, T.; Iwata, S.; Kojima, M., Structure of an antagonist-bound ghrelin receptor reveals possible ghrelin recognition mode. *Nat Commun* **2020,** *11* (1), 4160.

48.   Forli, S.; Huey, R.; Pique, M. E.; Sanner, M. F.; Goodsell, D. S.; Olson, A. J., Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nature Protocols* **2016,** *11*, 905-919.

49.   Cui, X.; Yang, R.; Li, S.; Liu, J.; Wu, Q.; Li, X., Modeling and insights into molecular basis of low molecular weight respiratory sensitizers. *Mol Divers* **2021,** *25* (2), 847-859.

50.   Kruskal, W. H.; Wallis, W. A., Use of Ranks in One-Criterion Variance Analysis. *J Am Stat Assoc* **2012,** *47* (260), 583-621.

51. Sanchez, J. E.; KC, G. B.; Franco, J.; Allen, W. J.; Garcia, J. D.; Sirimulla, S., BiasNet: A Model to Predict Ligand Bias Toward GPCR Signaling. *J Chem Inf Model* **2021,** *61* (9), 4190–4199.

52. Jasial, S.; Gilberg, E.; Blaschke, T.; Bajorath, J., Machine Learning Distinguishes with High Accuracy between Pan-Assay Interference Compounds That Are Promiscuous or Represent Dark Chemical Matter. *J Med Chem* **2018,** *61* (22), 10255–10264.

53. Galati, S.; Yonchev, D.; Rodríguez-Pérez, R.; Vogt, M.; Tuccinardi, T.; Bajorath, J., Predicting Isoform-Selective Carbonic Anhydrase Inhibitors via Machine Learning and Rationalizing Structural Features Important for Selectivity. *ACS Omega* **2021,** *6* (5), 4080–4089.

54. Yang, K. K.; Wu, Z.; Bedbrook, C. N.; Arnold, F. H., Learned protein embeddings for machine learning. *Bioinformatics* **2018,** *34* (15), 2642-2648.

55. Duvenaud, D.; Maclaurin, D.; Aguilera-Iparraguirre, J.; Gómez-Bombarelli, R.; Hirzel, T.; Aspuru-Guzik, A.; Adams, R. P., Convolutional Networks on Graphs for Learning Molecular Fingerprints. NIPS, Montreal, Canada. **2015**, 2215-2223.

Part two

Local Augmented Graph Neural Networks for the Classification of GPCR Modulators

Abstract:

G-protein coupled receptors (GPCRs), the largest family of cell transmembrane molecules, are distributed across various cell types and play crucial roles in signal transduction. GPCRs-mediated signaling pathways are involved in many pathological conditions. At least one third of marketed drugs target GPCRs, emphasizing their therapeutic importance. GPCR ligands can be classified as agonist, antagonist, and inverse agonist based on their functionality. The ability to accurately identify a ligand's functional profile is immensely useful for structural optimization of the ligand to selectively activate or block relevant pathways, which is important to accelerate drug discovery and reduce costs. In this study, a local augmented feature was incorporated into Graph Convolutional Network (LAGCN) and Graph Attention Network (LAGAT) in an attempt to predict the functional profile of GPCR ligands. A total of 5 representative GPCRs, including C-X-C chemokine receptor type 4 (CXCR4), ghrelin receptor (GR), histamine H3 receptor (H_3R), lysophosphatidic acid receptor 1 (LPA_1), and cannabinoid 1 receptor (CB_1) were included. The results showed that the LAGCN and LAGAT models achieved state-of-the-art accuracy for the classification of GPCR ligands. The inclusion of the local augmented feature consistently improved the accuracy, precision, recall, and F1 score in the dataset when compared to the base models without data augmentation. The average accuracies of LAGNN and LAGAN is 83% and 81%, respectively. This study showed that local augmented feature is a powerful tool to improve graph neural network model performance for GPCR functionality prediction. The models are promising tools to be used for screening novel therapeutic agents targeting GPCRs.

Graphical abstract:
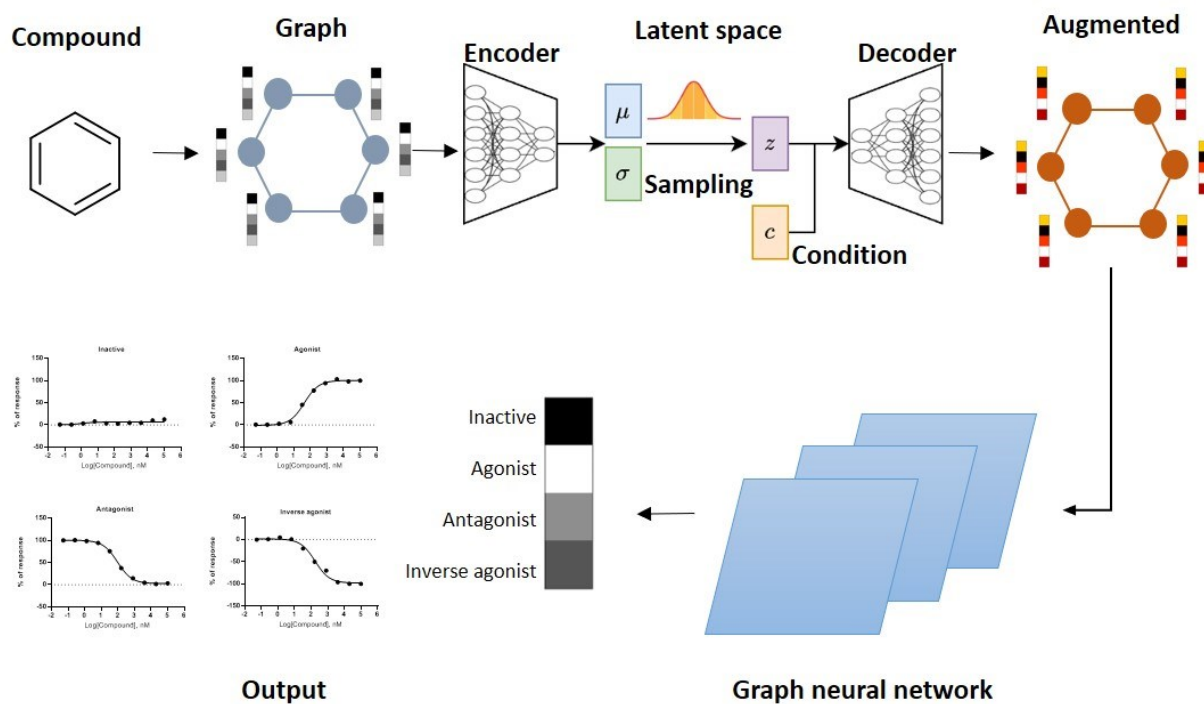
Key words: G-protein coupled receptors, local augmented graph convolutional network, local augmented graph attention network

1. Introduction

The emergence of large-scale and high-dimensional data has driven the successful application of artificial intelligence (AI) to medicinal chemistry in the past few years.[1-4] As a subset of AI, deep learning offers a versatile framework that enables the extraction of intricate patterns and relationships from complicated datasets, transcending the limitations of traditional methods in drug discovery. Deep learning methods, the driving force behind today's AI boom, possess immense potential to advance the drug discovery process and reduce failure rates, sparking a revolution in how the design and development of new pharmaceuticals are approached.[5-6] The commonly used deep learning methods in medicinal chemistry include feed-forward network, recurrent neural network (RNN), autoencoder (AE), generative adversarial network (GAN), and graph neural network (GNN). The feed-forward network represents the fundamental form of an artificial neural network, characterized by unidirectional connections that flow from the input layer to the hidden layers, and then onward to the output layer. It operates without loops or cycles within its architecture.[5] RNN is designed as an approach for modelling sequential data. In contrast to feed-forward network, RNN introduces feedback connections, allowing information to flow in cycles within its network architecture.[7] RNN and its variants, including long-short-term memory (LSTM) and gated recurrent unit (GRU), have gained prominence as powerful generative models used in *de novo* drug design.[8] An AE consists of two main components: an encoder and a decoder. The encoder transforms high-dimensional data into a low-dimensional representation, while the decoder reconstructs the low-dimensional representation back to the original input space. AE and its variants, including variational autoencoder (VAE) and adversarial autoencoder (AAE), are usually used to achieve unsupervised deep learning tasks.[9] A GAN is composed of two key parts: a generator and a discriminator, which undergo competitive training. The discriminator learns from its errors and enhances its ability to distinguish real from generated inputs, while the generator produces new samples that closely mimic real samples, trying to deceive the discriminator.[10] This adversarial interplay drives gradual enhancements in the abilities of both the generator and discriminator. Graph Neural Networks (GNNs) are specifically designed to tackle tasks that involve graph-structured data. They operate directly on the graph structure, making them ideal for analyzing relational information. The key idea for GNNs is to capture and leverage the local neighborhood information of nodes in a graph to learn informative representations.

157

In the field of medicinal chemistry, GNNs have captured significant attention due to their ability to represent chemical molecules using graphs in a highly intuitive and concise way.[11] In these graphs, atoms are represented as nodes, while bonds are represented as edges. This representation allows GNNs to capture the intricate relationships between atoms and bonds, enabling them to make accurate predictions. Over the past few years, there has been a surge in the utilization of various GNN variants, such as Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs), for a wide range of tasks in medicinal chemistry. These GNN variants have found successful applications in predicting drug-target interactions,[12-13] estimating ligand binding affinities,[14] and assessing the synthetic accessibility of organic molecules.[15] However, when dealing with nodes that have few neighbors, there may be challenges in adequately aggregating neighborhood information. The limited amount of available local information can result in incomplete or insufficient representations for such nodes. This crucial limitation of GNN hinders its broader utilization in drug discovery, necessitating attention and resolution to improve the capabilities and robustness of GNNs.

GPCRs, the largest family of cell transmembrane molecules, are extensively found in diverse cell types and play crucial roles in transducing a wide array of cell signals. Due to their significant involvement in cell signaling pathways, GPCRs have captured considerable attention in the field of drug discovery. These receptors have emerged as attractive targets for therapeutic/diagnostic intervention due to their ability to modulate important physiological processes.[16] Currently, approximately 30-35% of the FDA-approved drugs are designed to target GPCRs.[17] Based on the mechanism of action and interaction with the receptor, GPCR ligands can be categorized into several types: agonist, antagonist, inverse agonist and inactive molecules. Accurate classification of GPCR ligands into agonists, antagonists, inverse agonists, or inactive molecules is essential for comprehending their distinct effects on GPCRs and their implications in therapeutic applications. This classification provides insights into their biological effects, therapeutic potential, and possible side effects. Traditionally, the classification of GPCR ligands has been accomplished through costly and time-consuming biological experiments. While deep learning has found extensive application in the field of drug discovery, there have been no specific deep learning models developed to address this particular issue.

By harnessing the power of deep learning modeling, the objective is to explore the potential of deep learning in accurately classifying potential GPCR ligands, thereby addressing the aforementioned problem. In order to tackle the challenges associated with inadequate aggregation of neighborhood information in GNNs, the local augmented graph convolutional network (LAGCN) and local augmented graph attention network (LAGAT) were introduced as novel approaches to boost the performance of multi-classification tasks on graph-structured data. In LAGCN and LAGAT, a local augmentation mechanism was incorporated into the graphs in an attempt to enhance the graph topological representations and feature extraction processes.
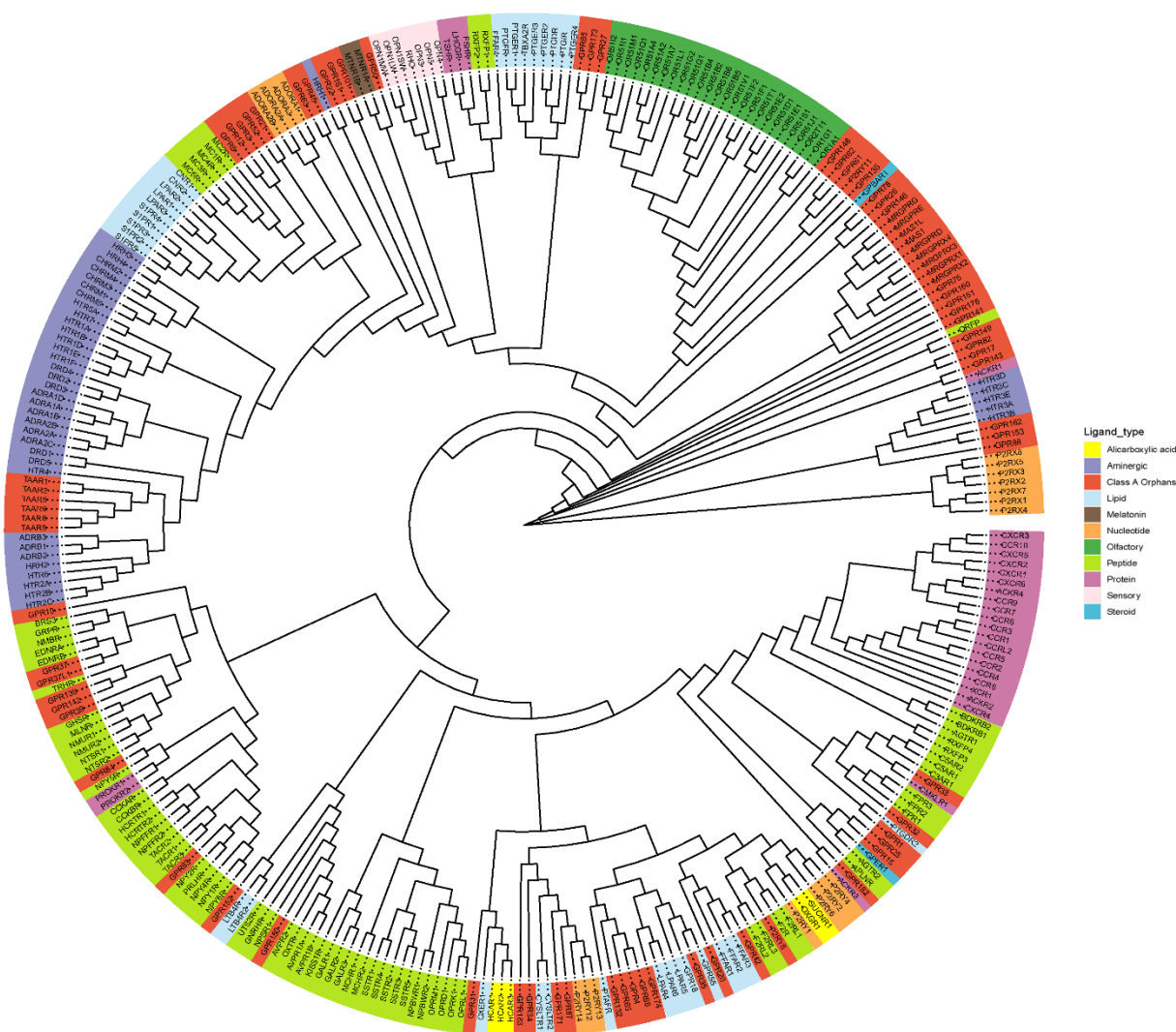
Figure 3.2.1. Phylogenetic tree of class A GPCRs as drug targets. Node represents GPCR named according to its gene name.

2. Methods

2.1 Data collection and preparation

Five GPCRs, namely C-X-C chemokine receptor type 4 (CXCR4), ghrelin receptor (GR), histamine H3 receptor (H_3R), lysophosphatidic acid receptor 1 (LPA_1), and cannabinoid 1 receptor (CB_1) with their endogenous ligand representing protein (CXCR4), peptide (GR), aminergic (H_3R), and lipid were selected (LPA_1 and CB_1) (Figure 3.2.1). The ChEMBL[18] database was employed to create ligand type datasets for the GPCRs. Duplicates and compounds with missing values were removed from the dataset. In cases where compounds had multiple values from different biological tests, the mean value was used. Compounds with an $IC_{50}/EC_{50} <$ 1000 nM were classified as active ligands and further categorized as agonists, antagonists, or inverse agonists. Compounds with an $IC_{50}/EC_{50} \geq$ 1000 nM were considered as inactive molecules. The dataset of all the receptors was summarized in Table 3.2.1. For each dataset, the compounds were randomly split into training and test set with a respective ratio of 7:3. All of the data were collected and processed by using the "chembl_webresource_client" package (https:// github.com/chembl/chembl_webresource_client) in Python. The chemophysical properties of all compounds were generated using the RDKit[19] package in Python.

Table 3.2.1. Summary of the dataset.

Target	No. of agonist	No. of antagonist	No. of inverse agonist	No. of inactive
LPA_1	48	234	0	73
CB_1	541	753	471	2213
H_3R	194	1024	239	291
GR	842	368	106	193
CXCR4	8	519	0	253

2.2 Graph construction

In this study, the Deep Graph Library (DGL)-LifeSci[20] was used to convert molecule objects from SMILES strings into a bidirectional graph format. The conversion process involved

representing each compound's atoms as nodes and covalent bonds as edges. To encode atom and bond information, the CanonicalAtomFeaturizer module and the CanonicalBondFeaturizer module from the DGL-LifeSci package in Python were utilized, respectively. The conversion tool provided eight atom features, resulting in a 1x74 dimensional atom feature matrix (Table 3.2.2). For a compound with N atoms, the resulting atom feature matrix would be N x 74. Additionally, four bond features were provided, resulting in a 1x12 dimensional bond feature matrix (Table 3.2.2). For a compound with N bonds, the conversion produced a bond feature matrix of size N x 12.

Table 3.2.2. Atom and bond features in the graph conversion tool from the DGL-LifeSci package.[20]

	No.	Description of features	No. of bits	Form
Atom	1	One hot encoding of the atom type	1-43	Binary
	2	One hot encoding of the atom degree	44-54	Binary
	3	One hot encoding of implicit H count on atoms	55-61	Binary
	4	Formal charge of the atom	62	Numerical
	5	Atom's radical electron count	63	Numerical
	6	One hot encoding of the atom hybridization.	64-68	Binary
	7	Whether the atom is aromatic	69	binary
	8	One-hot encoding of total H count on atoms	70-74	binary
Bond	1	One hot encoding of the bond type	1-4	binary
	2	Bond conjugation	5	binary
	3	Bond connectivity	6	binary
	4	One hot encoding of bond stereo configuration	7-12	binary

2.3 Local augmentation

To address issues associated with inadequate aggregation of neighborhood information in GNNs, a data augmentation framework, Local Augmentation for Graph Neural Networks (LAGNN) proposed by Liu *et al*[21] was used, in order to generate more features in the local neighborhood, enabling the acquisition of effective representations. Consequently, this approach enhances the expressive capacity of GNNs and leads to improved model performance. The term "local augmentation" denotes the process of generating neighborhood features using a conditional variational autoencoder (CVAE) model that is conditioned on local structures and node features.[21] The architecture of CVAE is shown in Figure 3.2.2.

A graph is represented as $G = (V, E)$, where V is a collection of N nodes $\{v_1, v_2, \ldots, v_n\}$, and E is the set of edges. The adjacency matrix, denoted as $\mathbf{A} \in \{0,1\}^{N \times N}$, indicates a value of 1 for $\mathbf{A}_{i,j}$ if and only if there exists an edge between nodes v_i and v_j in E. The neighborhood of a node is defined as $N_i = \{v_j | \mathbf{A}_{i,j} = 1\}$, representing the set of neighboring nodes connected to node v_i. The diagonal degree matrix is defined as $\mathbf{D}_{i,i} = \sum_{j=0}^{n} \mathbf{A}i, j$. It calculates the sum of all the edges connected to node v_i from $j = 0$ to n. The feature matrix is represented as $\mathbf{X} \in \mathbb{R}^{N \times F}$, where each node v is related to an F-dimensional feature vector $\mathbf{X}v$.

For a node v and its neighbors Nv, neighboring pairs $(\mathbf{X}v, \mathbf{X}\mu)$ was extracted as input for CVAE, where μ belongs to Nv. In the inference stage, the latent space vector z was sampled from a standard normal distribution N (0, I). Subsequently, the condition vector c and z were concatenated and fed into the decoder. This process allows us to obtain the generation of augmented feature vector $\bar{\mathbf{X}}v$.
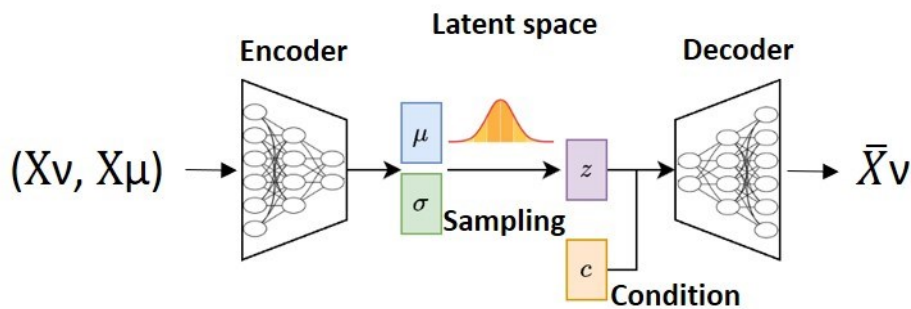
Figure 3.2.2. A schematic representation of CVAE.

2.4 LAGCN and LAGAT

In this work, LAGCN and LAGAT were implemented as proposed by Liu *et al.*[21] The first layer of LAGCN is defined as:

$$H^{(1)} = \sigma(\widetilde{P}XW_0^{(1)})\|\sigma(\widetilde{P}\overline{X}W_1^{(1)}),$$

where $\widetilde{P} = \widetilde{D}^{-1/2}\widetilde{A}\widetilde{D}^{-1/2}$. The symbol $\|$ denotes the concatenation applied to matrices along the second dimension. In the weigh matrix $\mathbf{W}$, the subscripts and superscripts indicate the ordinal numbers of the layers and parameters, respectively.

The first layer of LAGAT is defined as:

$$H^{(1)} = (\|_{d=1}^{D/2}\sigma\sum_{\mu \in N\nu}\left(\alpha_{\nu\mu}^{d}\,W_d^{(1)}\,\mathbf{X}\mu\right)) \,\| \, (\|_{d=D/2+1}^{D}\sigma\sum_{\mu \in N\nu}\left(\alpha_{\nu\mu}^{d}\,W_d^{(1)}\,\overline{\mathbf{x}}\mu\right)),$$

where $\alpha_{\nu\mu}^{d}$ is computed on $\mathbf{X}$ ($1 \leq d \leq D/2$) or $\overline{\mathbf{X}}$ ($D/2+1 \leq d \leq D$).

2.5 Model architecture and performance evaluation

The architectures of the LAGCN and LAGAT are shown in Figure 3.2.3. For LAGCN, local augmentation graphs were employed as input, which were then passed through three graph convolution layers interconnected by a ReLU function. The resulting information was subsequently utilized as input for a fully connected layer composed of two linear layers. Finally, a classifier was employed to calculate the ultimate prediction. For LAGAT, it utilized local augmentation graphs as the initial input, which were then propagated through three graph attention layers. The resulting information was further utilized as input for a fully connected layer consisting of two linear layers. Lastly, a classifier was employed to calculate the final prediction.

The model performances were evaluated using multiple metrics, including accuracy (ACC), precision (Prec), recall (Rec), and F1 scores. The equations that define these metrics are provided below:

$$ACC = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Prec = \frac{TP}{TP+FP}$$

163

$$\text{Rec} = \frac{TP}{TP+FN}$$

$$F1 = 2 \times \frac{TP}{2TP+FP+FN}$$

where TP represents true positives; TN, true negatives; FP, false positives; FN, false negatives.

In addition, the confusion matrix is used to show the performance of multiclassification models by offering comprehensive details about the model's performance for each individual class.

2.6 Implementation detail

In this study, all implementations and experiments were conducted in an environment with the following software/package: Python 3.10.12, Scikit-learn 1.2.2, RDKit v2022.09.5. PyTorch 2.0.1 with CUDA 11.8 was used as the basic framework. The deep learning models were implemented using DGL 1.1.0[22] and its supplementary package DGL-LifeSci 0.3.2.[20] The Adam optimizer with a learning rate of 0.001 was used for model training. The data was shuffled for each epoch, and a batch size of 32 was used. Figure 3.2.S1 demonstrate the training of the LAGCN and LAGAT models.

3. Results and discussion

3.1 Dataset overview

The overall chemical profiles of the whole dataset were reviewed prior to data splitting. A flowchart for the entire study is shown in Figure 3.2.3. To investigate the chemophysical properties of the compounds in the dataset, various parameters were calculated across agonist, antagonist, inverse agonist, and inactive molecules within a receptor (Figure 3.2.4). These parameters include molecular weight (MW), topological polar surface area (TPSA), lipid–water partition coefficient (ALogP), number of hydrogen bond acceptors (NHA) and donors (NHD), number of rotatable bonds (NRB), number of NHs or OHs (NHOH), and fractions of sp^3 hybridized carbons (sp^3). These properties are important for the functional activity of a molecule and have been extensively used in previous studies to compare compounds belonging to different categories.[23]
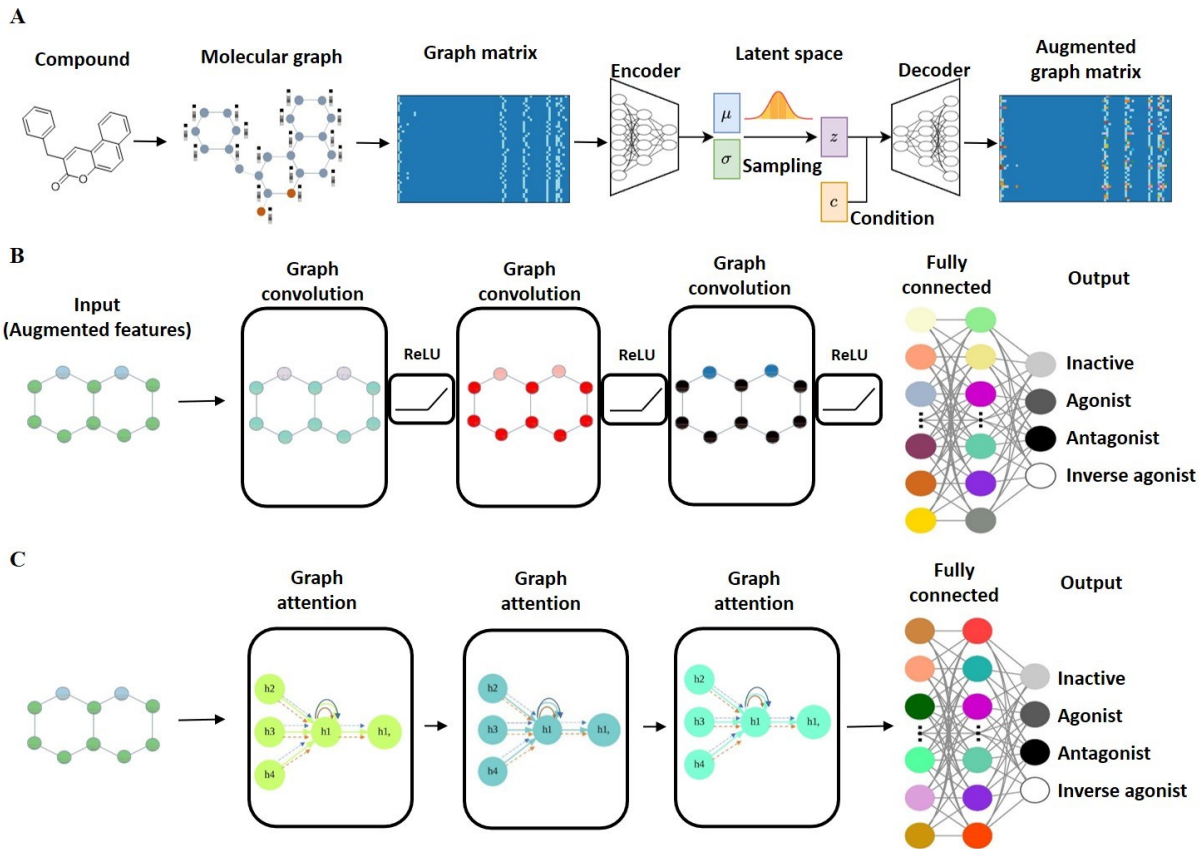
Figure 3.2.3. Overview of the LAGCN and LAGAT framework. A. Features on the graph were augmented by a conditional variational autoencoder (CVAE). B. Graphs with augmented features were fed into LAGCN. C. Graphs with augmented features were fed into LAGAT.
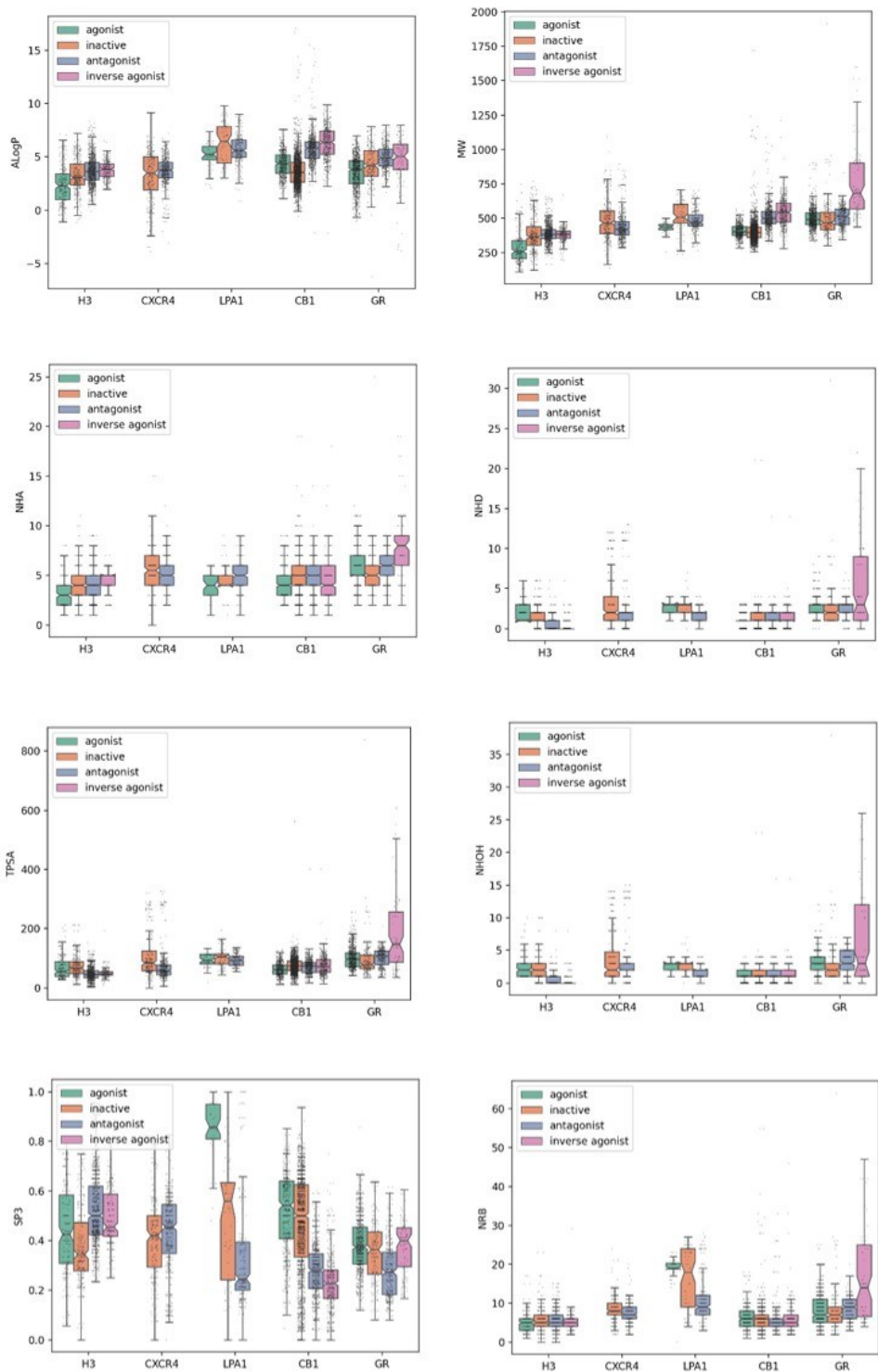
Figure 3.2.4. Chemical space analysis. Box plots illustrating the distribution of molecular properties across agonist (green), antagonist (blue), inverse agonist (purple), and inactive molecules (orange) of the five GPCRs.

3.2 Model performance

For each receptor examined in this study, both LAGCN and LAGAT models were constructed. The metric scores for each model are shown in Table 3.2.3. Overall, the LAGCN and LAGAT models consistently demonstrated good performance in terms of accuracy, with values ranging from 0.76 to 0.82 and 0.70 to 0.80, respectively. Across all receptors, the LAGCN model exhibited slightly higher accuracy scores compared to the LAGAT model. The largest difference was observed in the case of GR, where the LAGCN model achieved an accuracy score 8% higher than the LAGAT model. Additionally, the LAGAT models generally displayed higher precision, recall, and F1 scores compared to the LAGCN models.

Table 3.2.3. Summary of model performance.

		ACC	Precision	Recall	F1
CB$_1$	GAT	0.77	0.72	0.73	0.72
	GCN	0.80	0.77	0.76	0.76
	LAGAT	0.78	0.75	0.71	0.72
	LAGCN	0.81	0.77	0.80	0.78
LPA$_1$	GAT	0.76	0.63	0.75	0.67
	GCN	0.75	0.62	0.73	0.65
	LAGAT	0.77	0.64	0.74	0.68
	LAGCN	0.77	0.73	0.70	0.70
H$_3$R	GAT	0.70	0.67	0.66	0.64
	GCN	0.79	0.72	0.78	0.74
	LAGAT	0.80	0.72	0.74	0.73

	LAGCN	0.82	0.76	0.77	0.77
	GAT	0.72	0.62	0.65	0.61
GR	GCN	0.70	0.56	0.61	0.57
	LAGAT	0.70	0.59	0.61	0.59
	LAGCN	0.78	0.66	0.64	0.63
	GAT	0.74	0.45	0.45	0.45
CXCR4	GCN	0.68	0.42	0.43	0.42
	LAGAT	0.75	0.62	0.57	0.59
	LAGCN	0.76	0.69	0.70	0.70

In comparison to their counterpart models without data augmentation, LAGCN demonstrated an average improvement of 4% in accuracy, 10% in precision, 7% in recall, and 8% in F1 score. On the other hand, LAGAT exhibited an average improvement of 2% in accuracy, 5% in precision, 2% in recall, and 4% in F1 score. The results suggest that the augmented graph features generated by the CVAE effectively captured the underlying patterns in the dataset. By incorporating these augmented representations into the graph neural network models, namely GCN and GAT, the models acquired the abilities to recognize and leverage latent features critical for accurate predictions. This enhancement enabled the models to make informed decisions and improve their overall performance.

The model performances were further analyzed using the confusion matrix in an attempt to assess the class-specific performance. The confusion matrix allowed for easy observation of well-predicted classes and those that exhibited confusion and misclassification. Given the imbalanced nature of the dataset for each receptor, a normalized confusion matrix was employed.

In this matrix, each cell represents the percentage of instances classified into the true class. The corresponding results are depicted in Figure 3.2.5 and Figure 3.2.S2.
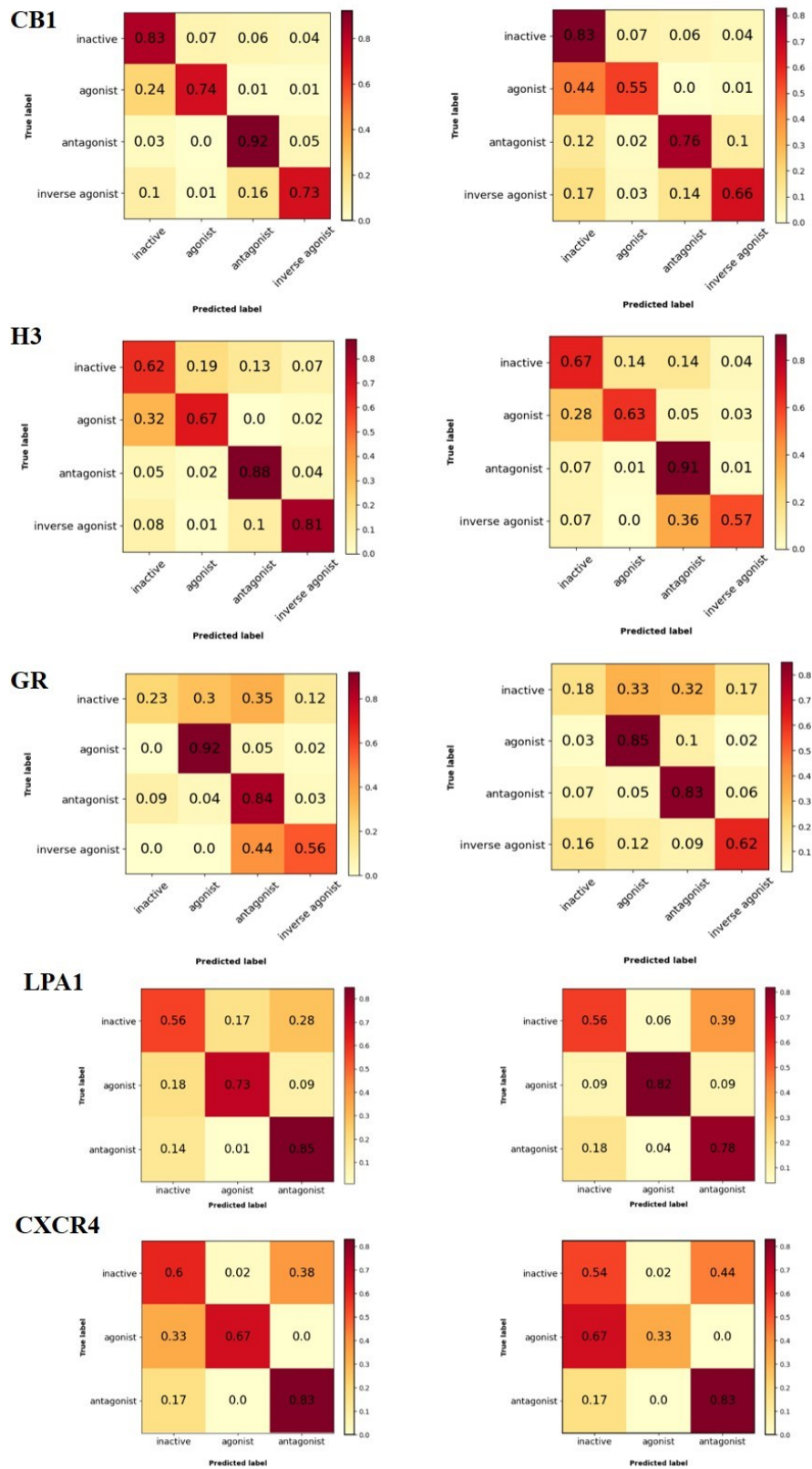
Figure 3.2.5. Confusion matrix for LAGCN (left) and LAGAT (right) models.

For the CB_1 and H_3R datasets, both LAGCN and LAGAT demonstrated satisfactory performance in classifying all four ligand types. LAGCN slightly outperformed LAGAT. However, when using LAGAT on the CB_1 dataset, a small number of agonists were misclassified as inactive molecules, and in the H_3R dataset, some inverse agonists were misclassified as antagonists. In the GR dataset, both LAGCN and LAGAT exhibited excellent ability in classifying agonists and antagonists, achieving an overall accuracy of more than 0.83. However, both models struggled to accurately classify inactive molecules, with over 60% of them being misclassified as either agonists or antagonists. For the LPA_1 dataset, both LAGCN and LAGAT exhibited correct classification of agonists and antagonists. Despite the small number of agonists in the overall dataset, both models were able to correctly identify the majority of true agonists, achieving an accuracy of 0.73 for LAGCN and 0.82 for LAGAT. In the CXCR4 dataset, LAGCN outperformed LAGAT in classifying agonists and inactive molecules. Although the number of agonists in this dataset was very limited, LAGCN achieved an accuracy of 0.67 in correctly classifying agonists. However, LAGAT misclassified more than 60% of the true agonists as inactive molecules. The limited number of trainable compounds in the CXCR4 dataset contributes to the model's inability to capture underlying patterns for precise predictions. In summary, both LAGCN and LAGAT demonstrated state-of-the-art performance based on the confusion matrix.

The fundamental concept behind GNNs is to iteratively update node representations in a graph by aggregating information from neighboring nodes, enabling the model to capture complex interactions and propagate information throughout the graph structure. In a GCN, predetermined weights are used for aggregating information from neighboring nodes,[24] while GAT employs attention mechanisms to assign different weights to neighboring nodes during aggregation.[25] This enables the model to selectively emphasize more informative nodes while de-emphasizing less important nodes, improving its ability to capture important patterns and relationships in the graph.

In this study, the base models of GCN and GAT exhibited comparable performances when the data was not locally augmented. However, upon local augmentation of the input data, both models demonstrated significantly enhanced performance. In particular, LAGCN displayed a more substantial improvement in overall metrics compared to LAGAT. This is partially due to the fact that GAT models typically have a higher number of parameters compared to GCN

models, as they involve attention mechanisms that introduce additional complexity. Data augmentation introduced additional diversity and variability into the input data, which may lead to a relatively higher risk of overfitting for GAT models. In this study, LAGCN demonstrated a superior ability to leverage the augmented features, effectively utilizing the additional information to improve performance.

4. Conclusion

In this research, local data augmentation was applied to enhance the performance of two graph neural network variations, GCN and GAT, in predicting GPCR ligand types. The utilization of locally augmented graphs generated by the CVAE resulted in a significant enhancement in model performance. On average, the accuracy of LAGCN improved by 4%, while LAGAT saw a 2% improvement. Accurate classification of GPCR ligands based on their ligand type is crucial for drug design and development. This study represents a pioneering effort in this field, as two advanced deep learning models have been developed, LAGCN and LAGAT, specifically designed for this precise classification task. By integrating a local augmented mechanism into LAGCN and LAGAT, the aim is to enhance their performance in multi-classification tasks involving graph-structured data. This incorporation empowers the models to effectively utilize the inherent structure and connectivity within the graphs, resulting in more precise and effective predictions. Ultimately, this advancement significantly improves the overall efficiency of the drug discovery pipeline.

5. References

1.   Liu, W.; Hopkins, A. M.; Yan, P.; Du, S.; Luyt, L. G.; Li, Y.; Hou, J., Can machine learning 'transform' peptides/peptidomimetics into small molecules? A case study with ghrelin receptor ligands. *Mol Divers* **2022**.

2.   Wang, M.; Hsieh, C.-Y.; Wang, J.; Wang, D.; Weng, G.; Shen, C.; Yao, X.; Bing, Z.; Li, H.; Cao, D.; Hou, T., RELATION: A Deep Generative Model for Structure-Based De Novo Drug Design. *J Med Chem* **2022,** *65* (13), 9478-9492.

3.   Burés, J.; Larrosa, I., Organic reaction mechanism classification using machine learning. *Nature* **2023,** *613* (7945), 689-695.

4.   Soleimany, A. P.; Amini, A.; Goldman, S.; Rus, D.; Bhatia, S. N.; Coley, C. W., Evidential Deep Learning for Guided Molecular Property Prediction and Discovery. *ACS Cent Sci* **2021,** *7* (8), 1356-1367.

5.   Yang, X.; Wang, Y.; Byrne, R.; Schneider, G.; Yang, S., Concepts of Artificial Intelligence for Computer-Assisted Drug Discovery. *Chem Rev* **2019,** *119* (18), 10520-10594.

6.   Chen, H.; Engkvist, O.; Wang, Y.; Olivecrona, M.; Blaschke, T., The rise of deep learning in drug discovery. *Drug Discov Today* **2018,** *23* (6), 1241-1250.

7.   Medsker, L. R.; Jain, L. C., Recurrent neural networks. Design and Applications. *Design and Applications* **2001 Dec,** *5*, 64-67.

8.   Tong, X.; Liu, X.; Tan, X.; Li, X.; Jiang, J.; Xiong, Z.; Xu, T.; Jiang, H.; Qiao, N.; Zheng, M., Generative Models for De Novo Drug Design. *J Med Chem* **2021,** *64* (19), 14011-14027.

9.   Martinelli, D. D., Generative machine learning for de novo drug discovery: A systematic review. *Comput Biol Med* **2022,** *145*, 105403.

10.   Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y., Generative Adversarial Networks. In *Advances in neural information processing systems*, 2014; p 2672−2680.

11.   Xiong, J.; Xiong, Z.; Chen, K.; Jiang, H.; Zheng, M., Graph neural networks for automated de novo drug design. *Drug Discov Today* **2021,** *26* (6), 1382-1393.

12.   Torng, W.; Altman, R. B., Graph Convolutional Neural Networks for Predicting Drug-Target Interactions. *J Chem Inf Model* **2019,** *59* (10), 4131-4149.

13. Li, Y.; Qiao, G.; Wang, K.; Wang, G., Drug-target interaction predication via multi-channel graph neural networks. *Brief Bioinform* **2022,** *23* (1), bbab346.

14. Li, X.-S.; Liu, X.; Lu, L.; Hua, X.-S.; Chi, Y.; Xia, K., Multiphysical graph neural network (MP-GNN) for COVID-19 drug design. *Brief Bioinform* **2022,** *23* (4), bbac231.

15. Yu, J.; Wang, J.; Zhao, H.; Gao, J.; Kang, Y.; Cao, D.; Wang, Z.; Hou, T., Organic Compound Synthetic Accessibility Prediction Based on the Graph Attention Mechanism. *J Chem Inf Model* **2022,** *62* (12), 2973-2986.

16. Liu, W.; Hopkins, A. M.; Hou, J., The development of modulators for lysophosphatidic acid receptors: A comprehensive review. *Bioorg Chem* **2021,** *117*, 105386.

17. Hauser, A. S.; Attwood, M. M.; Rask-Andersen, M.; Schiöth, H. B.; Gloriam, D. E., Trends in GPCR drug discovery: new agents, targets and indications. *Nat Rev Drug Discov* **2017,** *16* (12), 829–842.

18. David Mendez, A. G., A Patrícia Bento, Jon Chambers, Marleen De Veij, Eloy Félix, María Paula Magariños, Juan F Mosquera, Prudence Mutowo, Michał Nowotka, María Gordillo-Marañón, Fiona Hunter, Laura Junco, Grace Mugumbate, Milagros Rodriguez-Lopez, Francis Atkinson, Nicolas Bosc, Chris J Radoux, Aldo Segura-Cabrera, Anne Hersey, Andrew R Leach, ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res* **2019,** *47* (D1), D930-D940.

19. RDKit: Cheminformatics and Machine Learning Software, 2013. *http://www.rdkit.org*.

20. Li, M.; Zhou, J.; Hu, J.; Fan, W.; Zhang, Y.; Gu, Y.; Karypis, G., DGL-LifeSci: An Open-Source Toolkit for Deep Learning on Graphs in Life Science. *ACS Omega* **2021,** *6* (41), 27233–27238.

21. Songtao Liu, R. Y., Hanze Dong, Lanqing Li, Tingyang Xu, Yu Rong, Peilin Zhao, Junzhou Huang, Dinghao Wu, Local Augmentation for Graph Neural Networks. In *39th International Conference on Machine Learning*, 2022; Vol. 162, pp 14054-14072.

22. Wang, M.; Yu, L.; Zheng, D.; Gan, Q.; Gai, Y.; Ye, Z.; Li, M.; Zhou, J.; Huang, Q.; Ma, C.; Huang, Z.; Guo, Q.; Zhang, H.; Lin, H.; Zhao, J.; Li, J.; Smola, A.; Zhang, Z., Deep graph library: Towards efficient and scalable deep learning on graphs. In *ICLR workshop on representation learning on graphs and manifolds*, 2019.

23. Kong, W.; Tu, X.; Huang, W.; Yang, Y.; Xie, Z.; Huang, Z., Prediction and Optimization of NaV1.7 Sodium Channel Inhibitors Based on Machine Learning and Simulated Annealing. *J Chem Inf Model* **2020,** *60* (6), 2739-2753.

24. Thomas N. Kipf, M. W., Semi-Supervised Classification with Graph Convolutional Networks. *arXiv:1609.02907* **2016**.

25. Petar Veličković, G. C., Arantxa Casanova, Adriana Romero, Pietro Liò, Yoshua Bengio, graph attention networks. *arXiv:1710.10903* **2017**.

Chapter 4

Conclusions and Outlook

1. **Conclusions and future work**

The connection between the activation of LPA$_1$ and the progression of breast cancer is becoming increasingly apparent. The mRNA expression of LPA$_1$ is significantly increased across multiple breast cancer cell lines,[1] with a particularly pronounced increase observed in the triple-negative breast cancer cell line MDA-MB-231, where LPA$_1$ expression is approximately 86 times higher than in non-tumorigenic cells.[1] LPA$_1$ is recognized for its role in mediating LPA-induced cell migration and invasion, demonstrated in both *in vitro* breast cancer cell lines and *in vivo* mouse models of breast cancer.[2-4] Our bioinformatic analysis further supports the correlation between LPA$_1$ and the development of breast cancer. The convergence of these pieces of evidence establishes LPA$_1$ as a promising target for the advancement of drug development for breast cancer treatment.

Chapter one offers a comprehensive review of drug development efforts targeting LPA$_1$ and its associated family members, LPA$_{2-6}$. This review serves as a valuable reference for fellow researchers engaged in LPA$_1$-related studies.

In chapter two, a comprehensive description of the synthesis and assessment of novel carbamate-based and urea-based LPA$_1$ antagonists is presented. The study delves into the elucidation of the structure-activity relationship (SAR) concerning the previously reported LPA$_1$ antagonist, **RO6842262**. Notably, successful incorporation of fluorine was achieved, enhancing our understanding of the compound's SAR. Among the array of carbamate-derived analogues of **RO6842262**, compound **12f** emerged as the most potent. This compound underwent further evaluation on the MDA-MB-231 breast cancer cell line, where it exhibited a dose-dependent inhibition of LPA-induced cell survival, migration, and invasion.

Given the promising outcomes observed with compound **12f** on MDA-MB-231, the subsequent phase involves *in vivo* experiment employing breast cancer mouse models. To begin with this study, a breast cancer mouse model will be established through the injection of 4T1 cells into the mammary fat pads of female mice. Once the successful creation of the breast cancer mouse model is accomplished, the administration of compound **12f** will ensue. This aims to comprehensively assess the properties of **12f**, including its drug-like properties, *in vivo* potency, efficacy, and therapeutic impact.

Subsequently, the benzyl alcohol building block was changed into a benzylamine building block during the Curtius rearrangement reaction, leading to the production of a second series of LPA$_1$ antagonists featuring a urea moiety. A total of eight urea-based LPA$_1$ antagonists were synthesized and subjected to evaluation. Notably, this series of urea-derived compounds generally exhibited lower activity in comparison to the initial series of carbamate-derived compounds. Nevertheless, these urea-derived compounds hold potential as improved drug candidates due to the inherent stability of urea over carbamate. Within this series, compound **17f** emerged as the most potent candidate. Therefore, it was selected for biological evaluation using the breast cancer cell line MDA-MB-231.

The results suggested that **17f** effectively inhibits LPA-induced cell migration and invasion in a dose-dependent manner, thus showing promising therapeutic implications for addressing breast cancer metastasis. Subsequent investigations will involve subjecting compound **17f** to *in vivo* evaluations using breast cancer mouse models. Comprehensive assessments will be carried out to determine the drug-like properties and *in vivo* potency of this compound, providing valuable insights into its potential as a therapeutic agent.

In the third section of this chapter, modifications were made to the ester form of compound **12h**, initially discovered in the first section, in an endeavor to create PET tracers. The successful synthesis of the fluorinated counterpart, designated as **12k**, was achieved. Notably, this fluorinated compound, **12k**, maintains a high degree of selectivity for LPA$_1$, although its activity has shown a slight reduction. Moving ahead, efforts will focus on the synthesis and purification of the corresponding ^{18}F-labeled hot compound, **^{18}F-12k**.

To assess the binding affinity of **^{18}F-12k** toward various breast cancer tumor cells, a cell uptake assay will be performed. Following this, an *in vivo* evaluation using breast cancer mouse models is planned. In this phase, the mice will receive intravenous administration of the radiotracer **^{18}F-12k** via the tail vein, and subsequent imaging will be performed using a microPET scanner. This process is expected to yield valuable insights into the biodistribution characteristics of the radiotracer.

Upon the successful completion of animal studies, these ligands could then move forward into clinical trials. Notably, compounds **12f** and **17f** have the potential to emerge as promising treatments for breast cancer patients, particularly those with metastasis and recurrence. Given

that metastasis is the primary contributor to breast cancer-related fatalities,[5] compounds **12f** and **17f** could play a crucial role in enhancing the prognosis for individuals affected by breast cancer.

Furthermore, the radiolabeled compound ^{18}**F-12k** holds the promise of finding application in breast cancer diagnosis through PET scans. Early diagnosis and accurate staging would offer significant benefits to breast cancer patients, and the innovative tracer ^{18}**F-12k** possesses the potential to facilitate this critical aspect of patient care.

During the course of this study, conventional drug discovery methodologies were exclusively used for the development of LPA$_1$ antagonists. The design of these antagonists was based on the structure of a potent LPA$_1$ antagonist previously reported, and molecular docking studies were used to guide the process of drug design. It is our hope that more cutting-edge methods can be applied to LPA$_1$ drug design in order to improve success rate and reduce failure.

With the advent of extensive, high-dimensional data, our society has entered a 'big data' era powered by artificial intelligence (AI). AI-related technologies have become ubiquitous in daily life, including applications like facial recognition, chatbots, and language translation. In the field of medicinal chemistry, AI-assisted drug discovery is rapidly emerging as a new methodology.

To harness the potential of this emerging technology and explore its use in drug discovery, the first part of chapter three focused on developing various machine learning models to predict ligand binding for a GPCR. In the second part of this chapter, we delved deeper into this field and created state-of-the-art deep learning models to predict the ligand type for various GPCRs. This study represents the first attempt to classify GPCR ligand using deep learning techniques. Overall, the machine learning and deep learning models developed in these studies achieved high accuracy in predicting ligand binding and ligand type. In upcoming research, these models will be helpful in shaping LPA$_1$ drug design. Using the best AI model developed in these studies, virtual screening will be performed with a large compound library to discover novel LPA$_1$ ligands with unique scaffolds. Additionally, the next step will involve building deep generative models, a prospective approach for generating potent molecules targeting LPA$_1$.
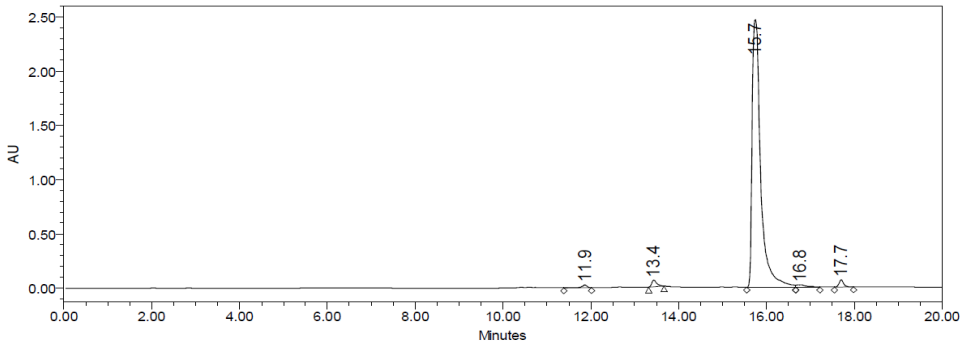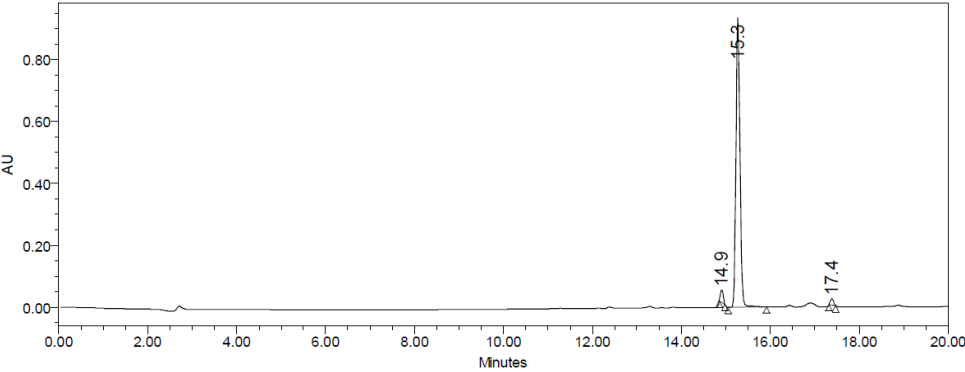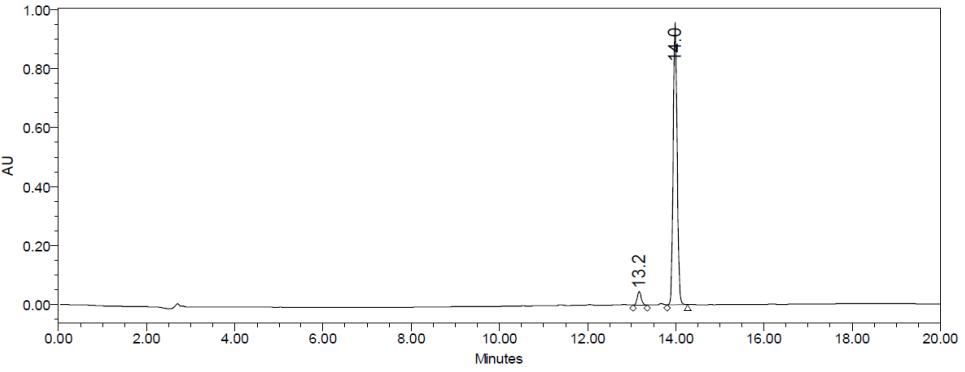
The discovery and development of LPA$_1$ ligands has proven to be a difficult task. However, the integration of cutting-edge AI technology holds the potential to streamline the drug discovery

pipeline, enhancing success rates. Considering the potential benefits of exploiting these novel technologies, such research emerges as pivotal in advancing the medical toolkit.

2. References

1.   Li, T. T.; Alemayehu, M.; Aziziyeh, A. I.; Pape, C.; Pampillo, M.; Postovit, L.-M.; Mills, G. B.; Babwah, A. V.; Bhattacharya, M., Beta-arrestin/Ral signaling regulates lysophosphatidic acid-mediated migration and invasion of human breast tumor cells. *Mol Cancer Res* **2009,** *7* (7), 1064-77.

2.   Hama, K.; Aoki, J.; Fukaya, M.; Kishi, Y.; Sakai, T.; Suzuki, R.; Ohta, H.; Yamori, T.; Watanabe, M.; Chun, J.; Arai, H., Lysophosphatidic acid and autotaxin stimulate cell motility of neoplastic and non-neoplastic cells through LPA1. *J Biol Chem* **2004,** *279* (17), 17634-17639.

3.   David, M.; Ribeiro, J.; Descotes, F.; Serre, C.-M.; Barbier, M.; Murone, M.; Clézardin, P.; Peyruchaud, O., Targeting lysophosphatidic acid receptor type 1 with Debio 0719 inhibits spontaneous metastasis dissemination of breast cancer cells independently of cell proliferation and angiogenesis. *Int J Oncol* **2012,** *40* (4), 1133-41.

4.   Boucharaba, A.; Serre, C.-M.; Guglielmi, J.; Bordet, J.-C.; Clézardin, P.; Peyruchaud, O., The type 1 lysophosphatidic acid receptor is a target for therapy in bone metastases. *Proc Natl Acad Sci U S A* **2006,** *103* (25), 9643-8.

5.   Chaffer, C. L.; Weinberg, R. A., A perspective on cancer cell metastasis. *Science* **2011,** *331* (6024), 1559-64.

Appendix A: U-HPLC Chromatograms and Purities

Compds.	HPLC Chromatograms	Purity (%)
12b		96
12c		95
12d		96

12e   100

12f   99

12g   100
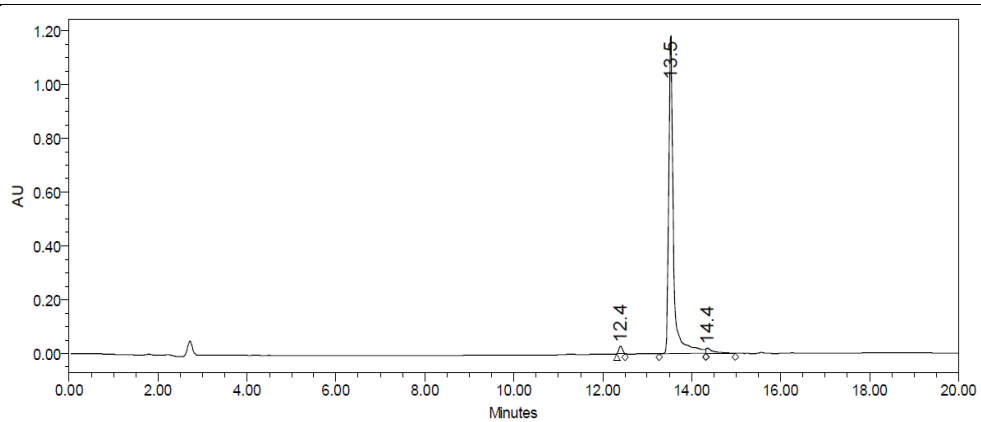
12h   96
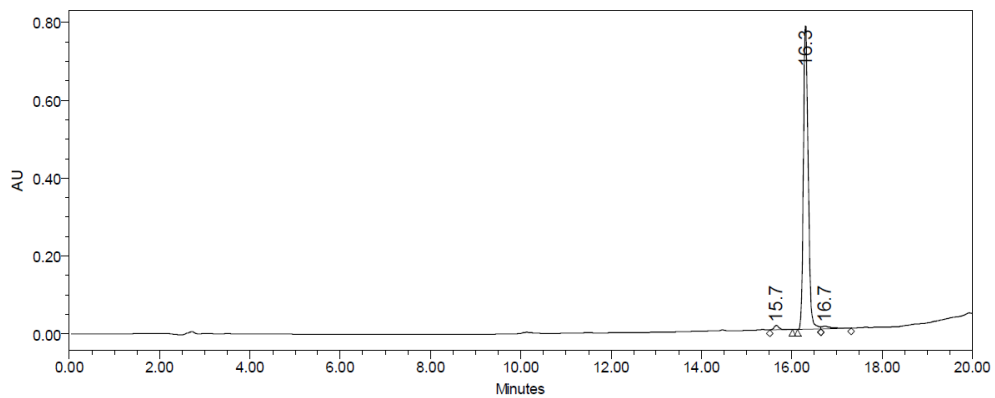
12i   **97**
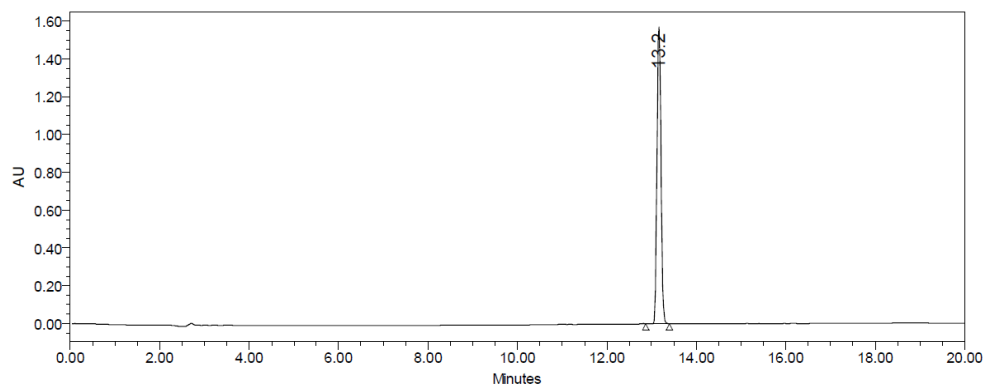
14a   **100**
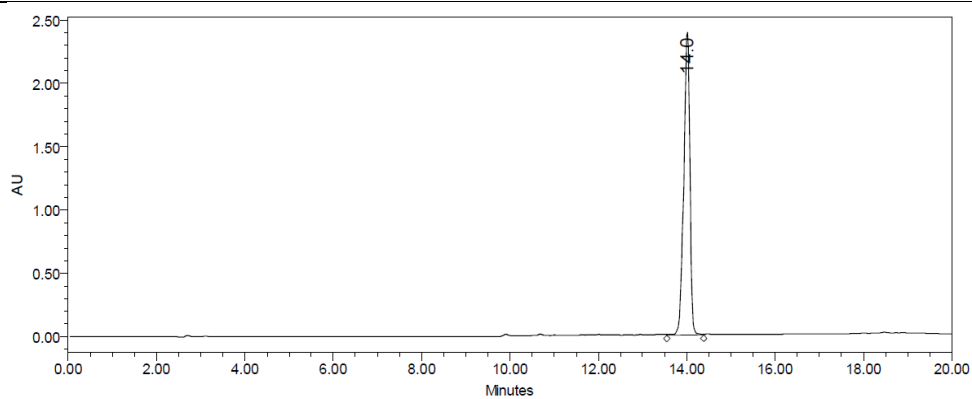
12k   **100**
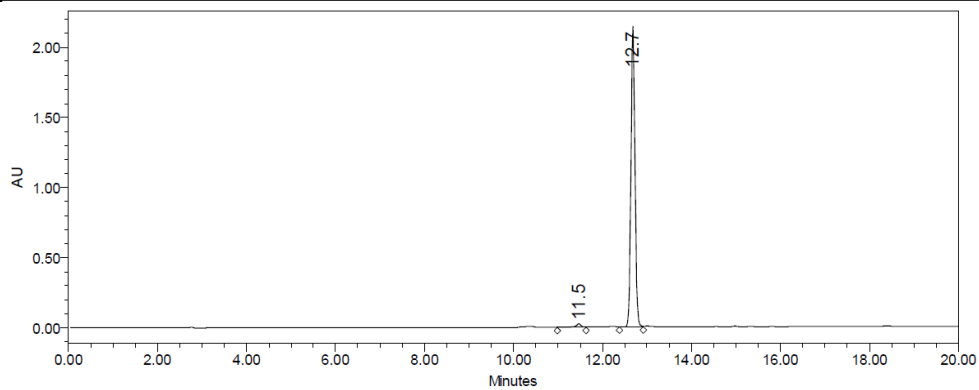
17a   **98**
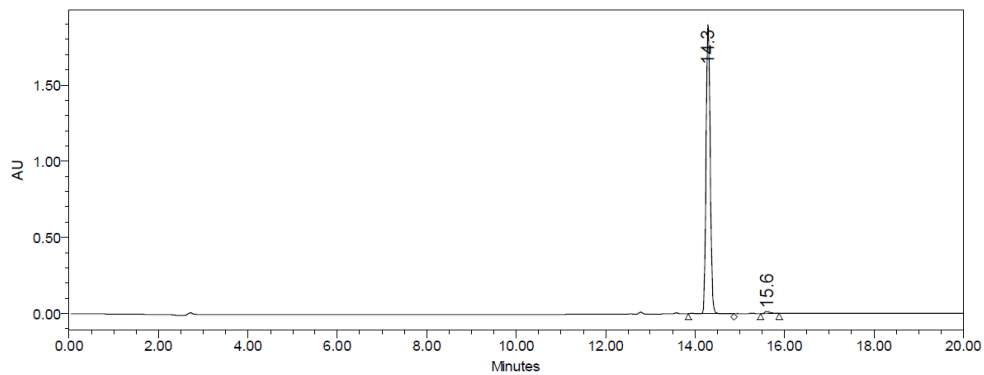
17b   **99**
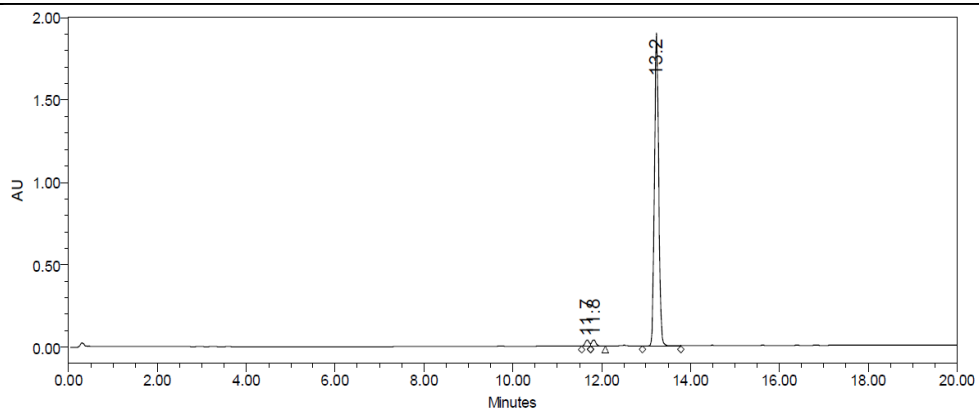
17c   **97**
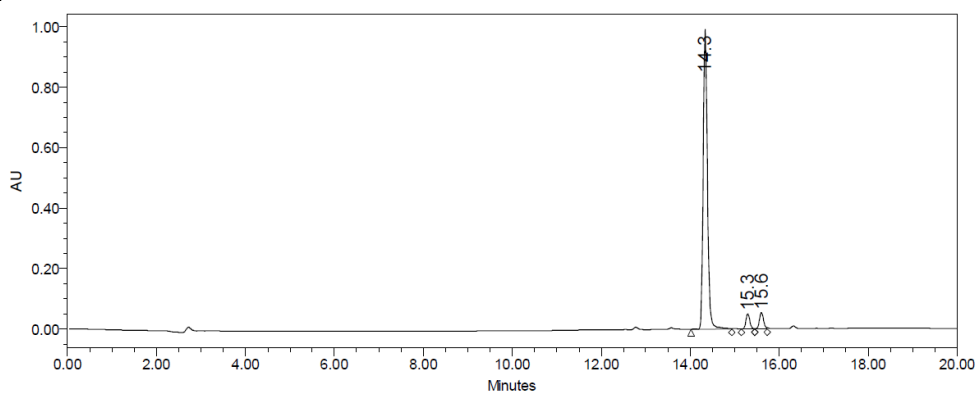
17d   **92**
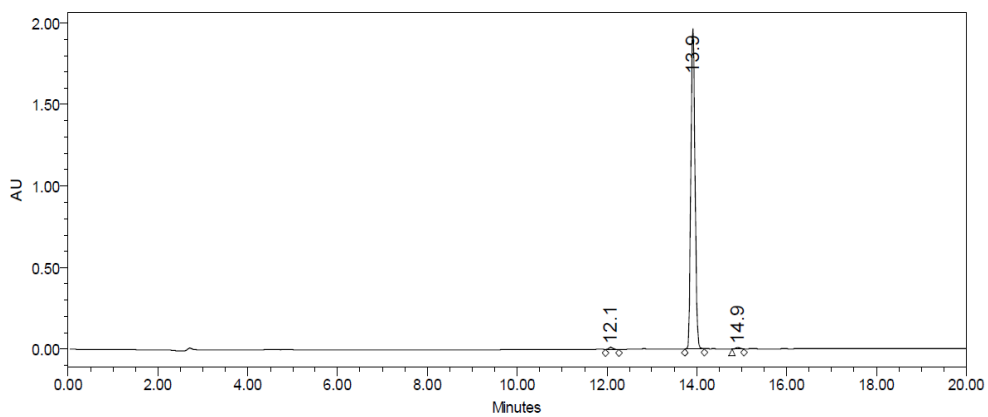
17e   **99**
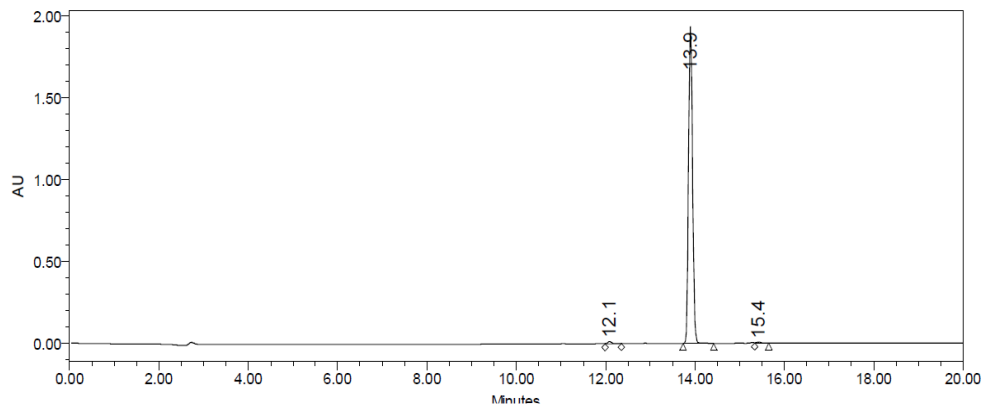
17f <!-- --> **99**
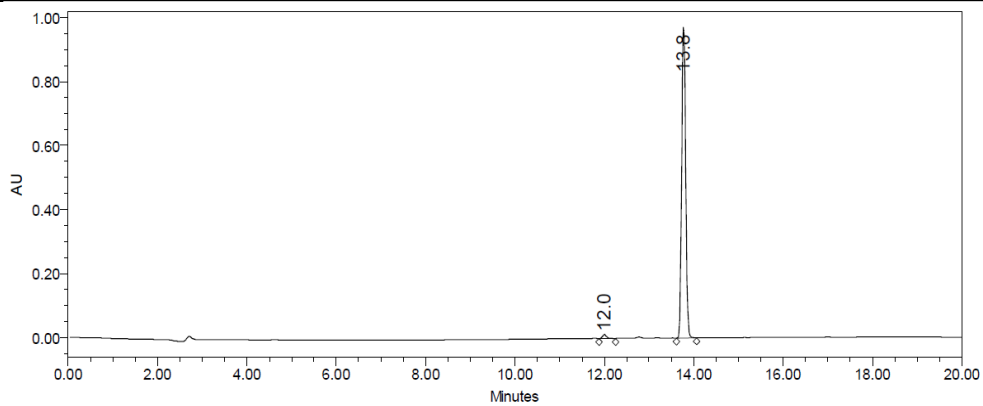
17g <!-- --> **99**
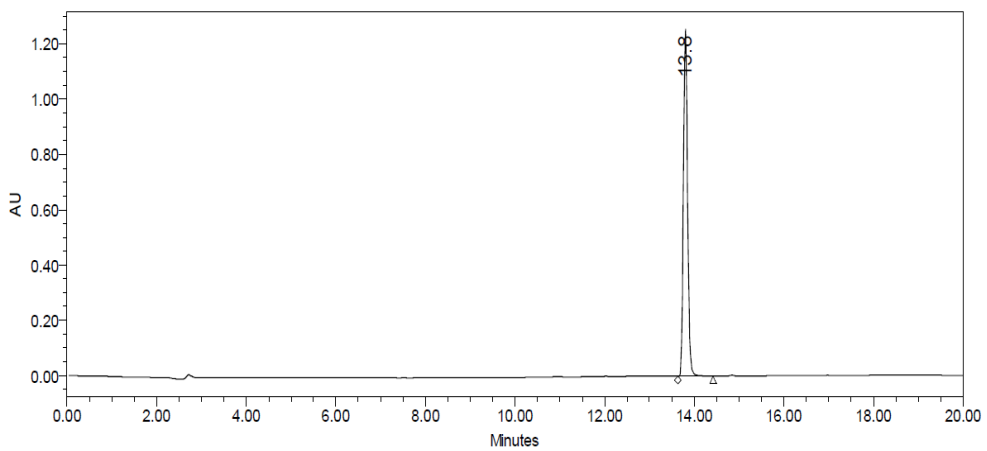
17h **100**

Appendix B: ^{1}H and ^{13}C NMR spectra

12b
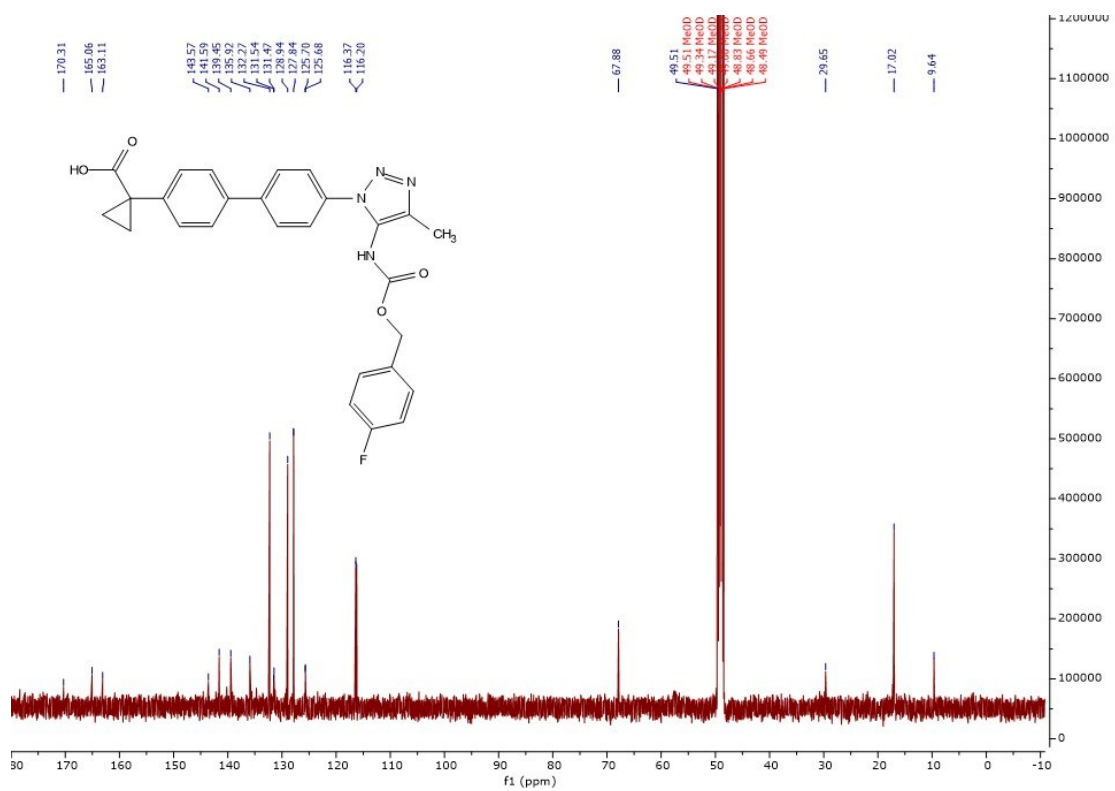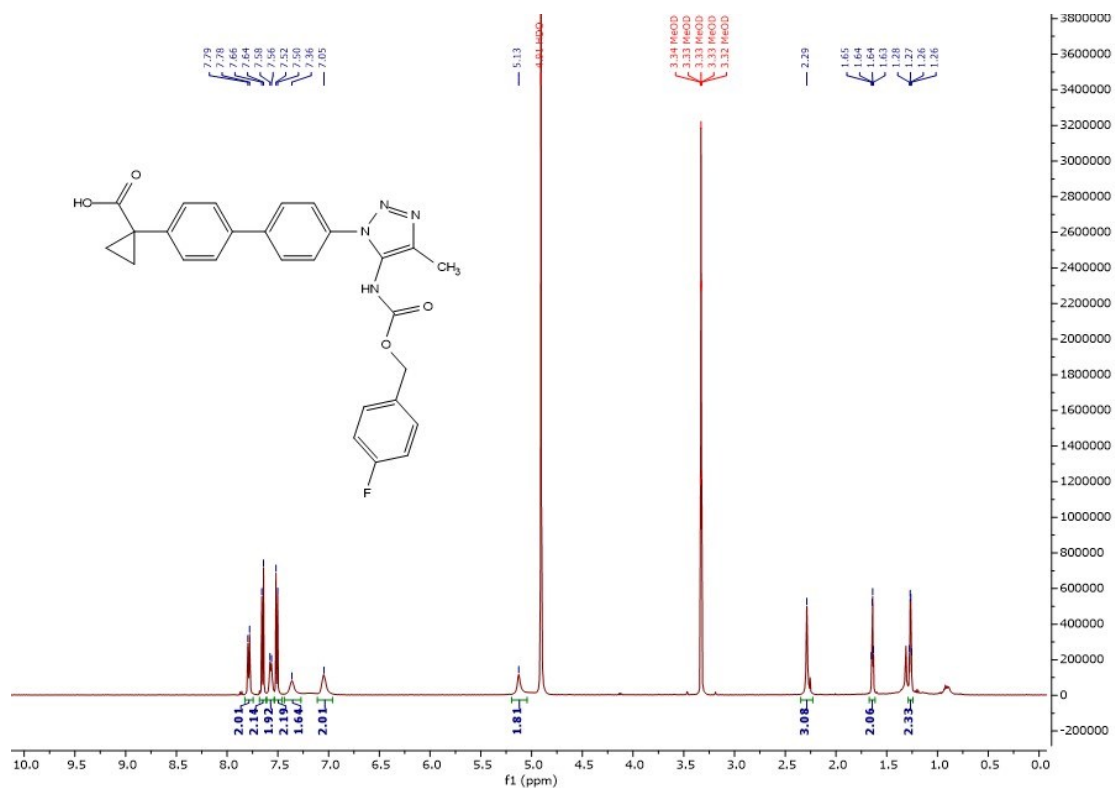
12c
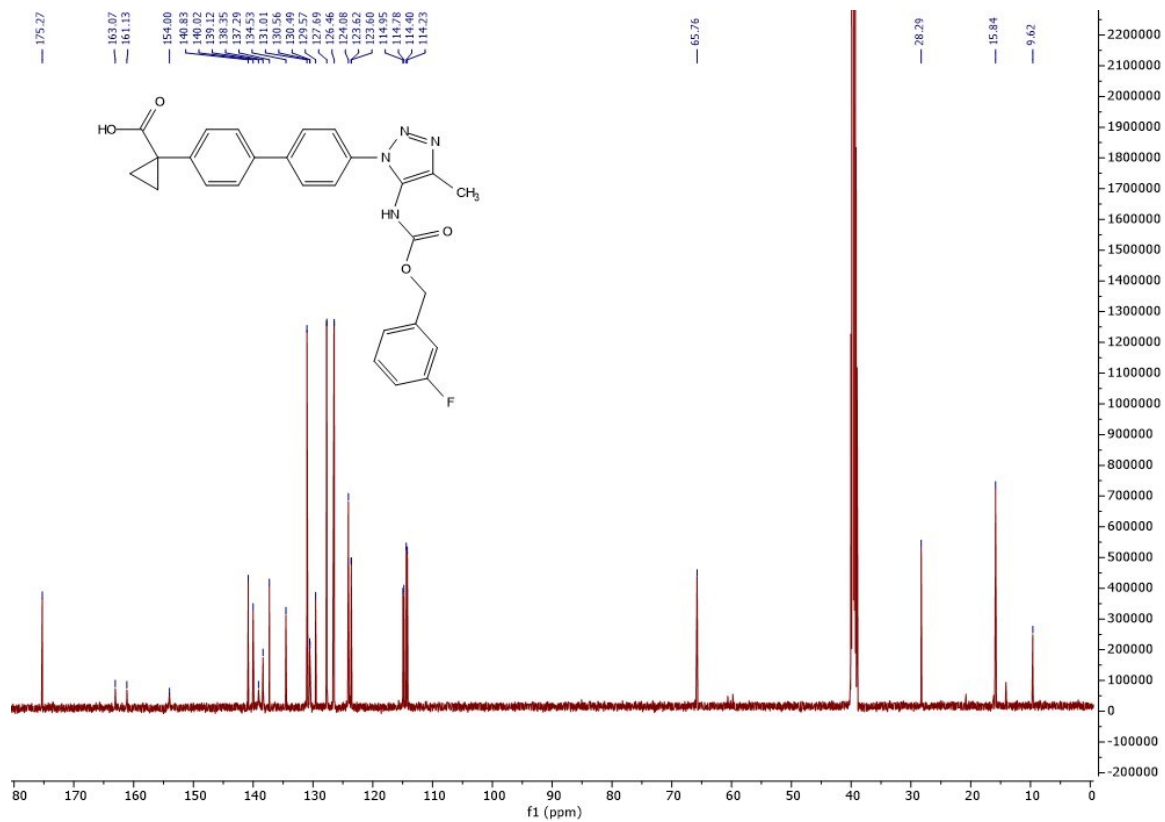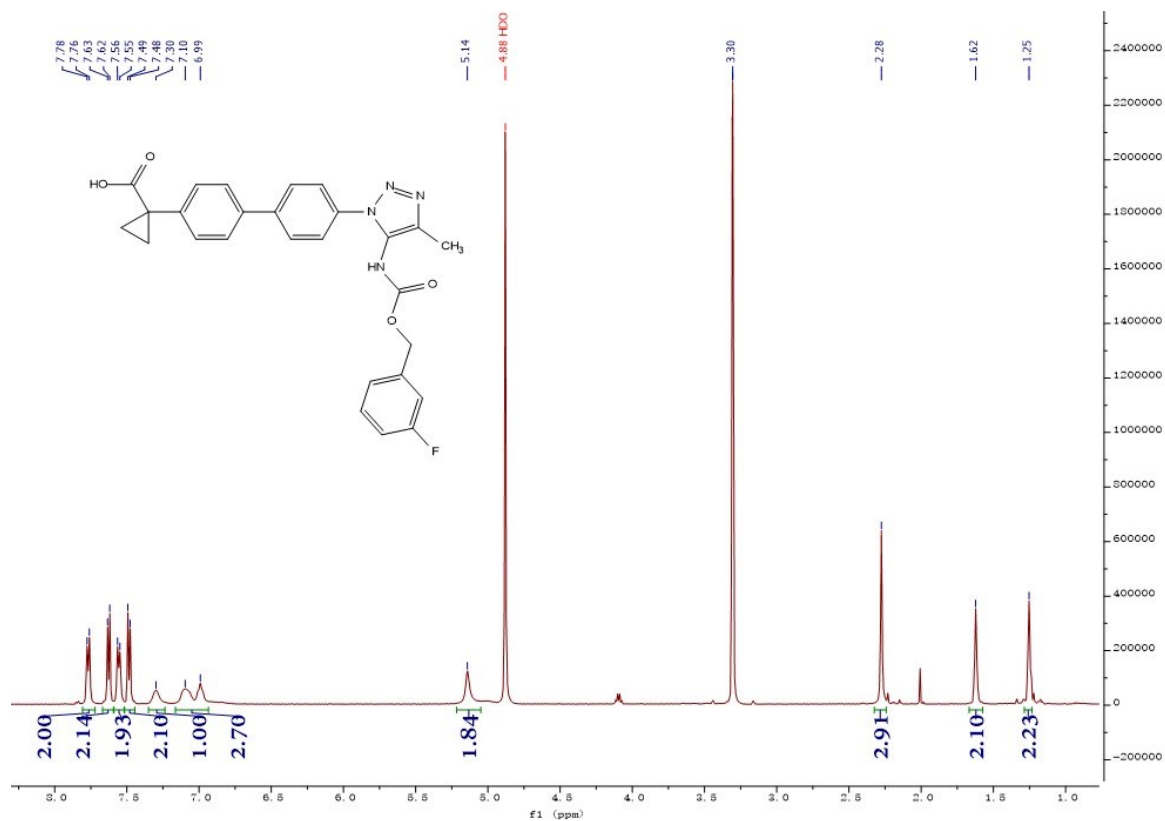
12d
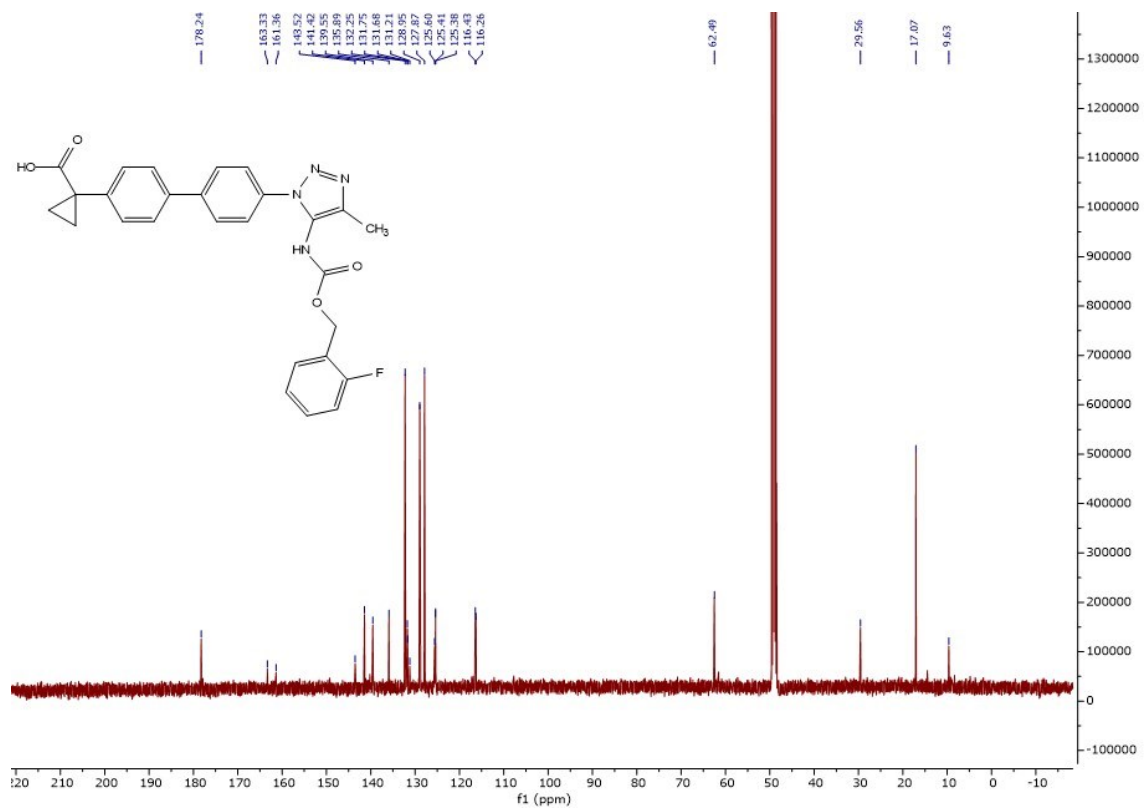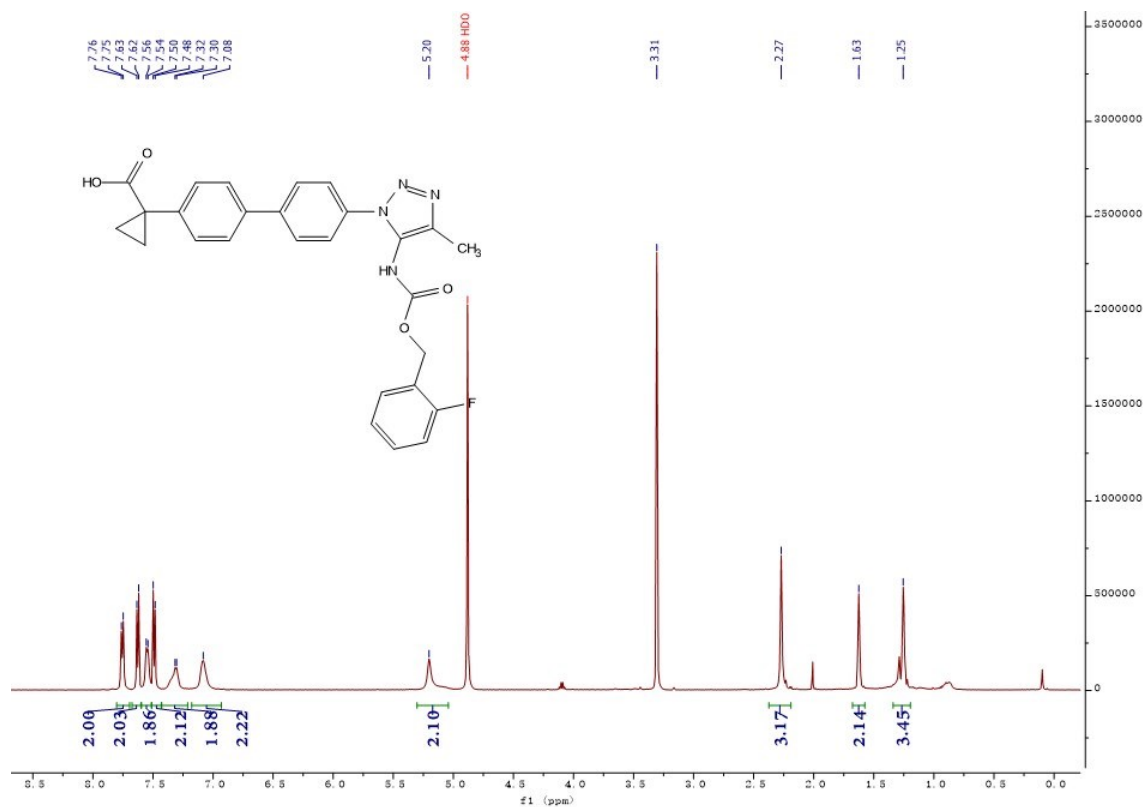
12e
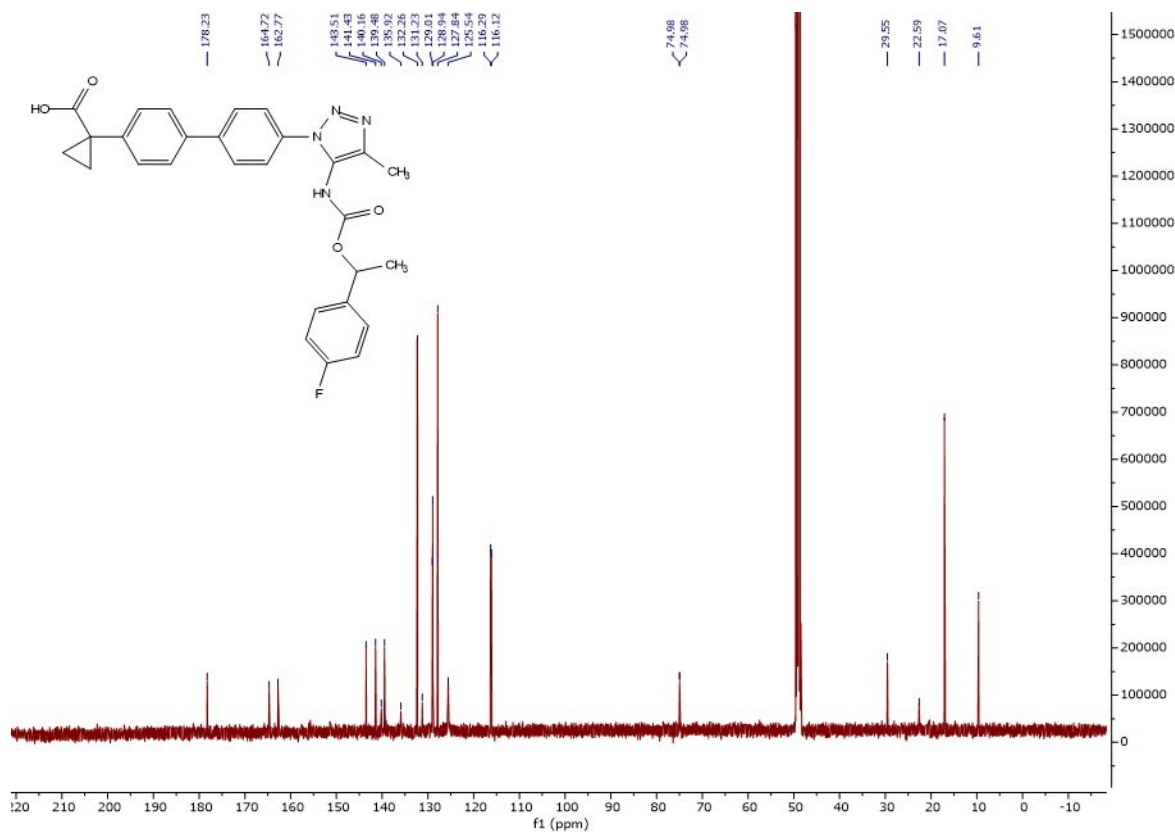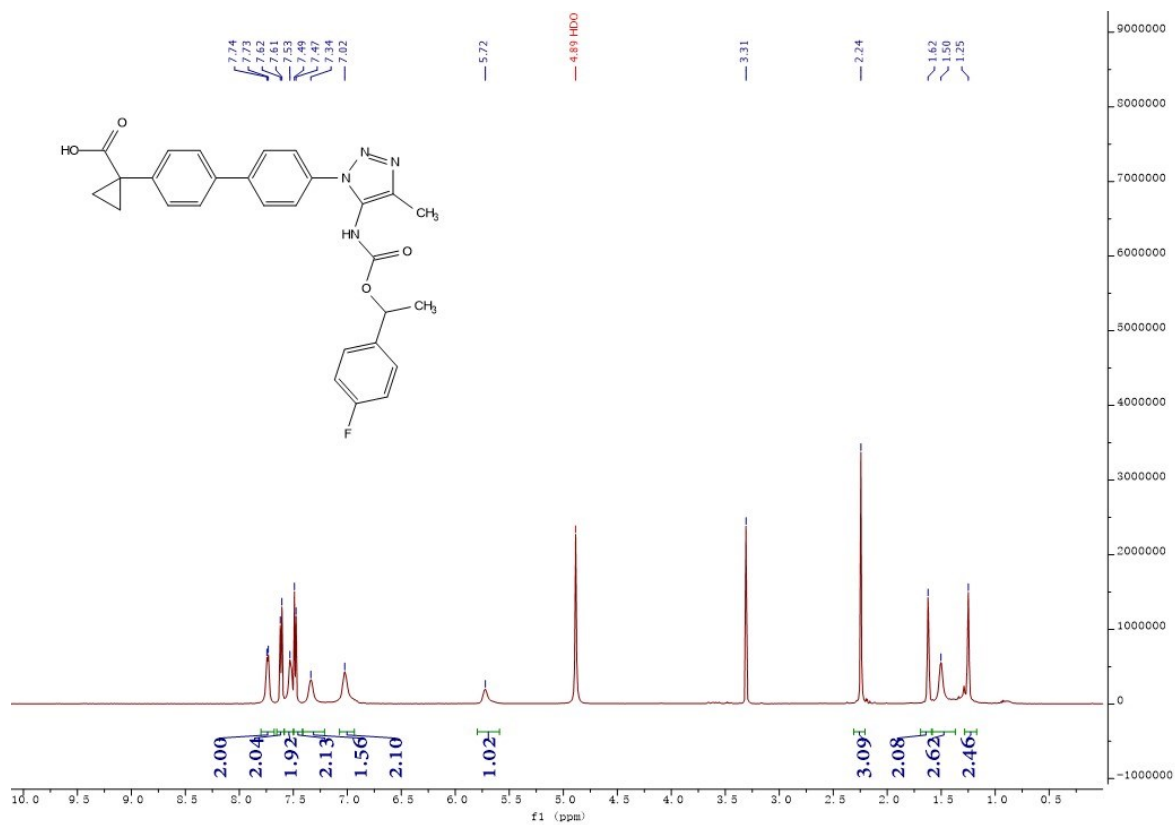
12f
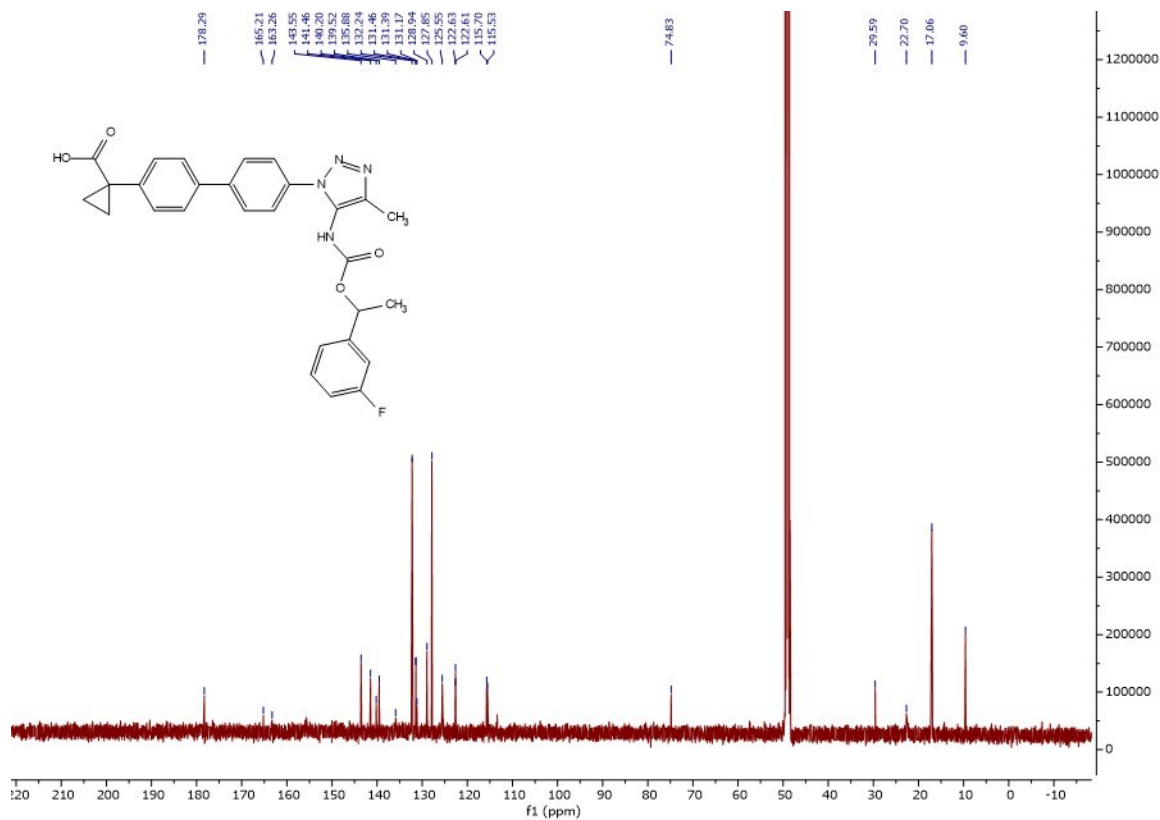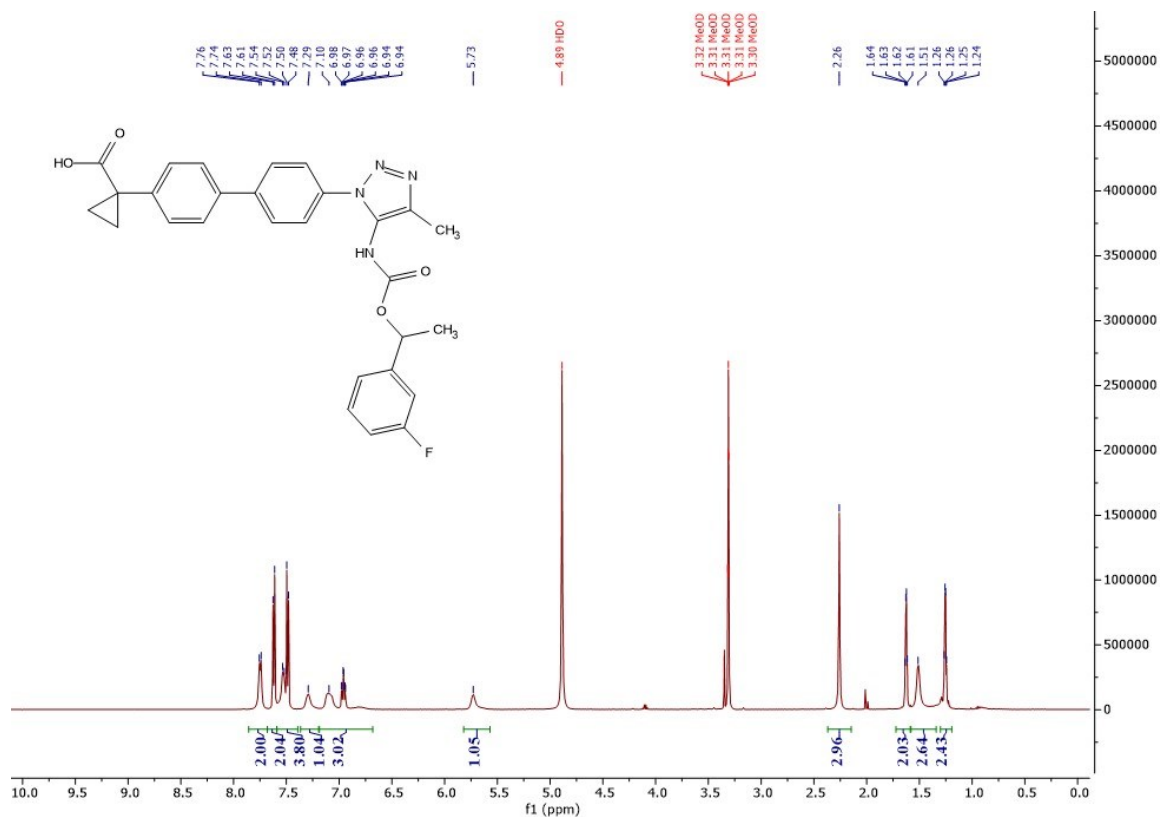
12g
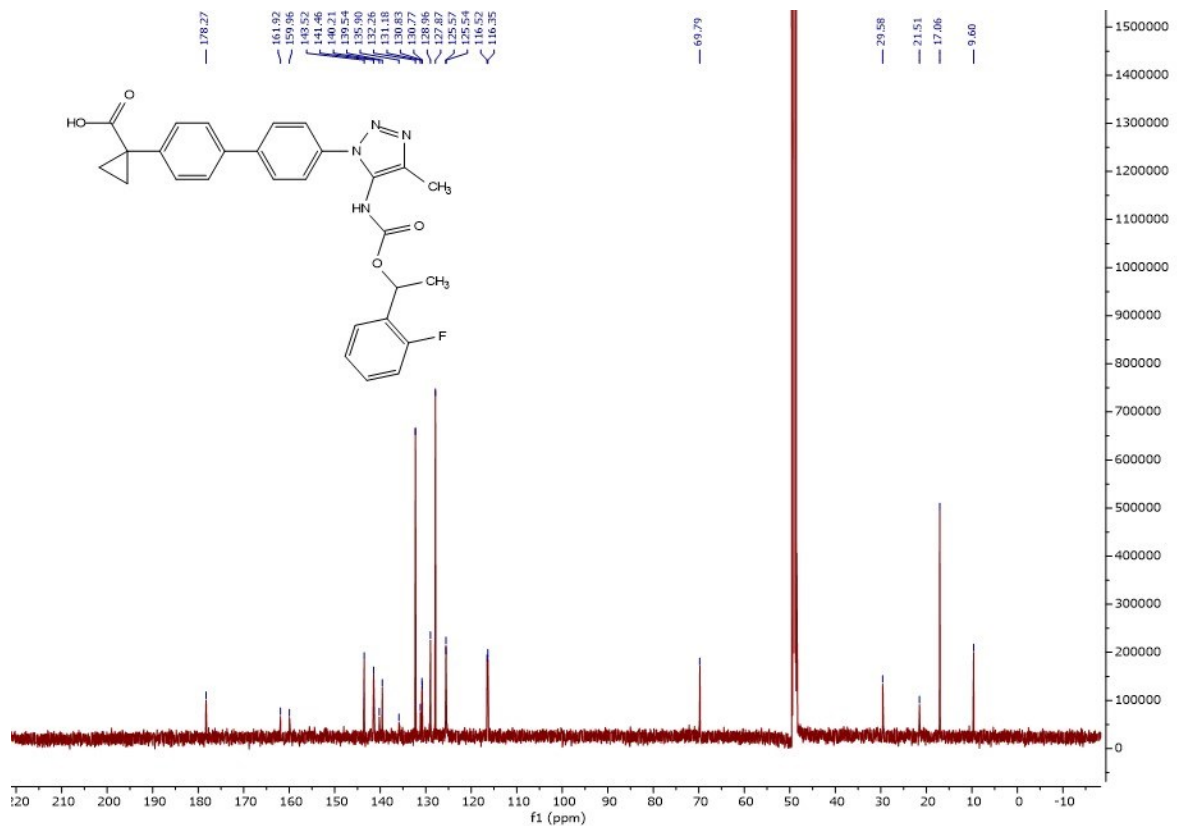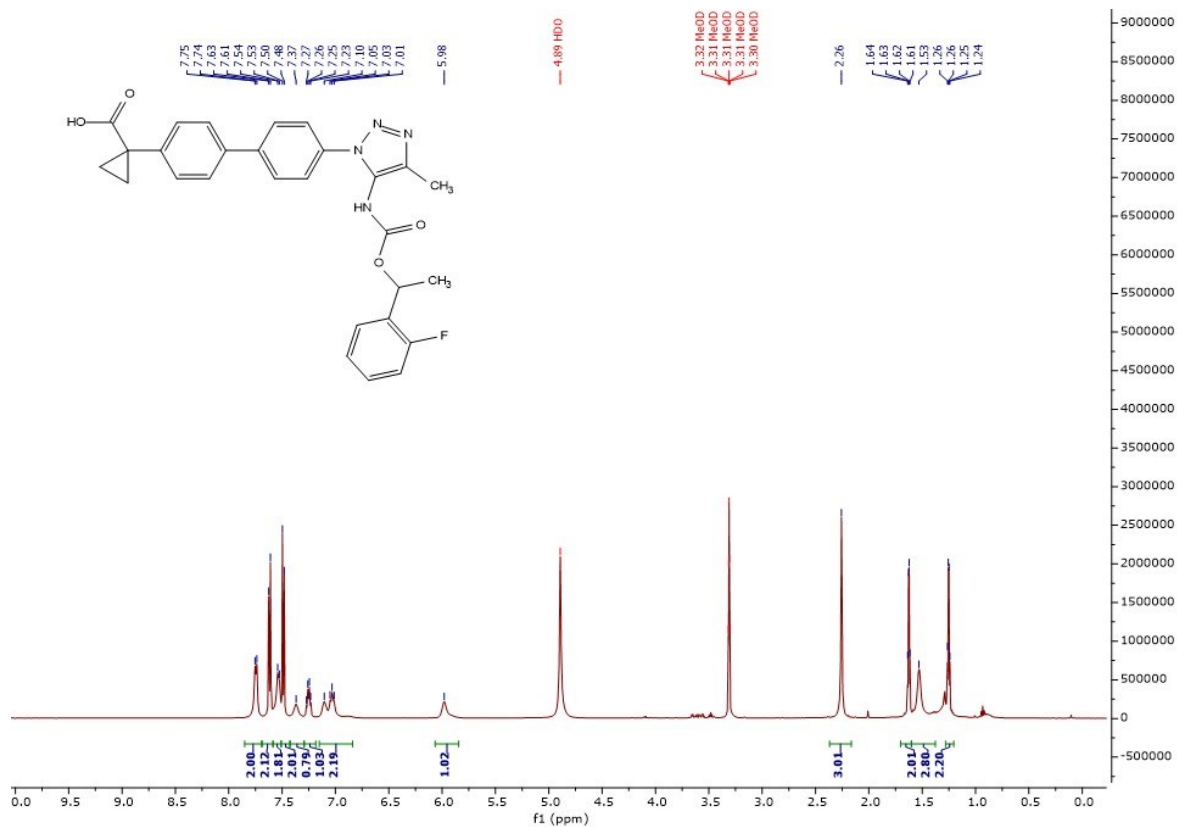
12h
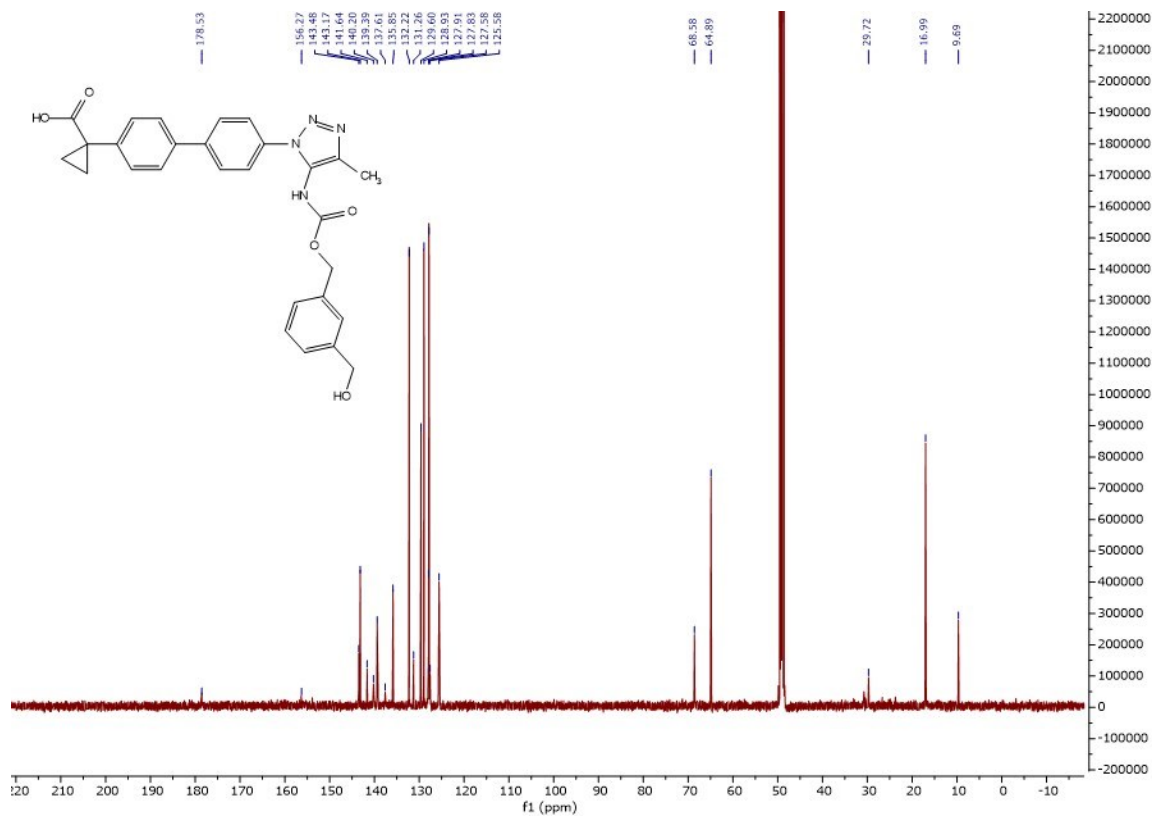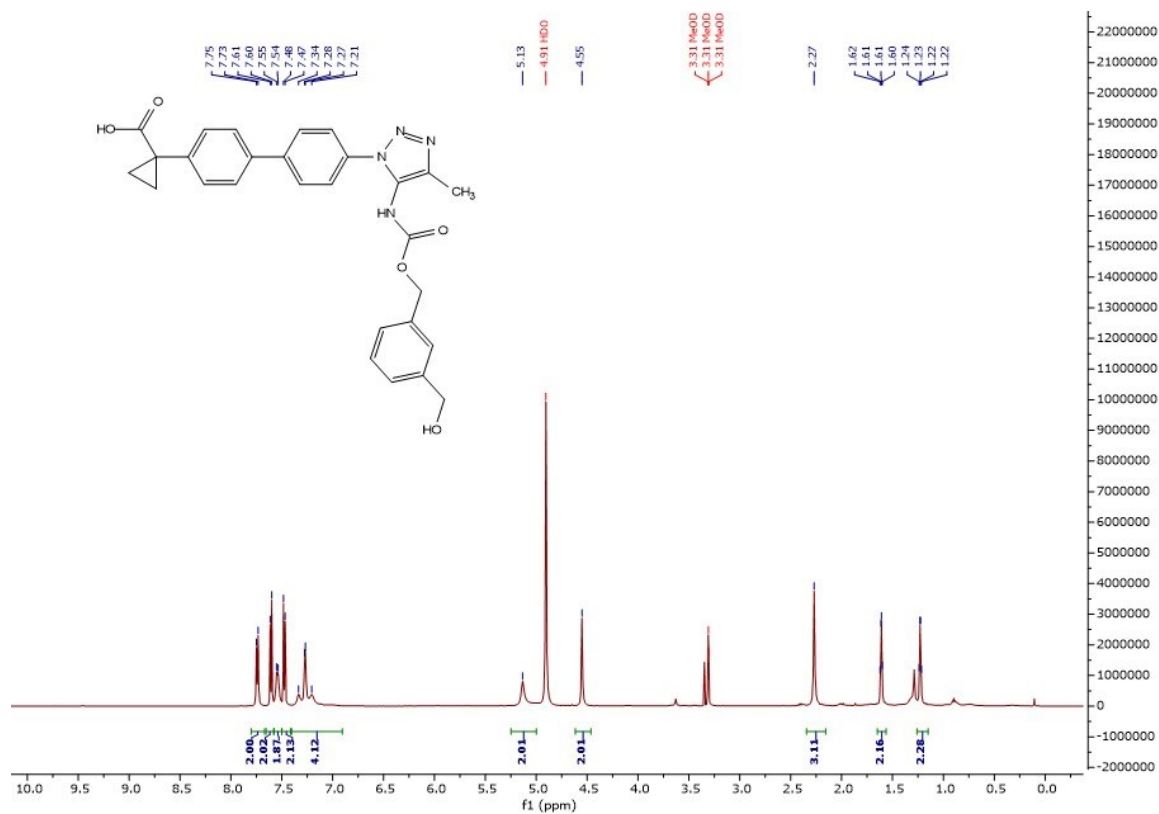
12i
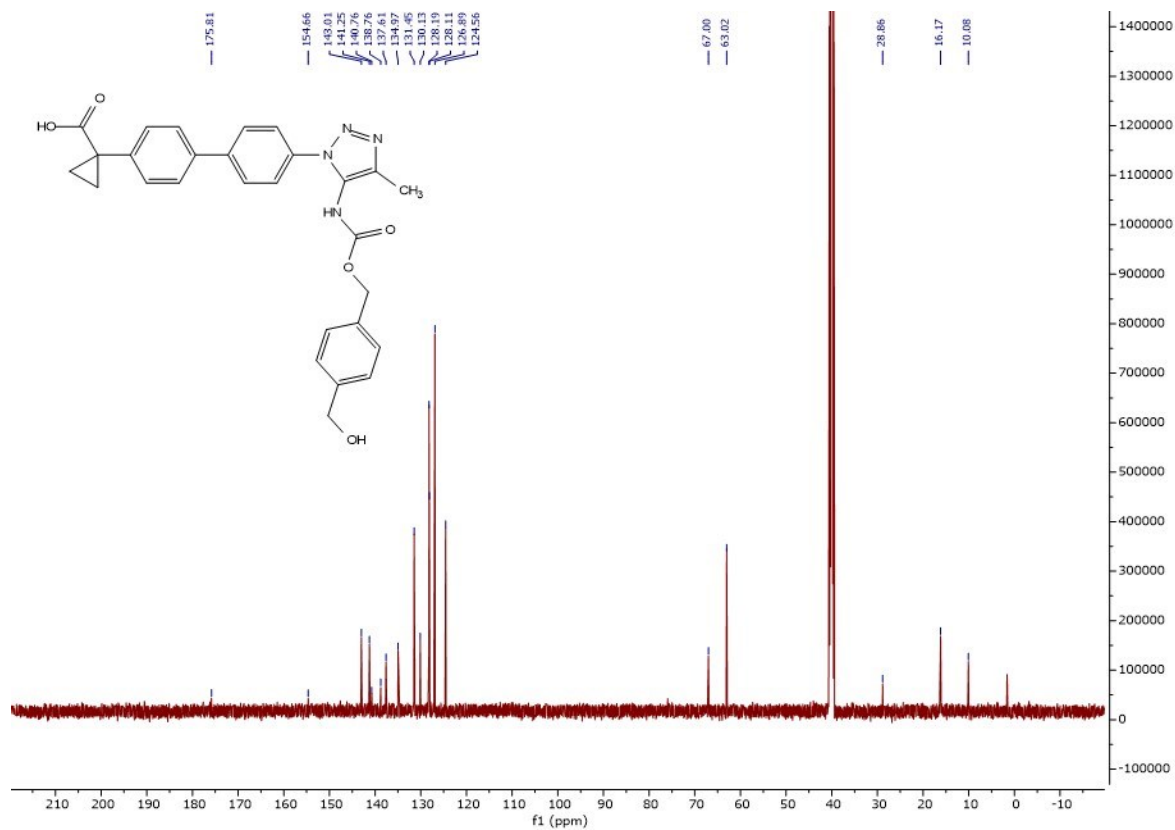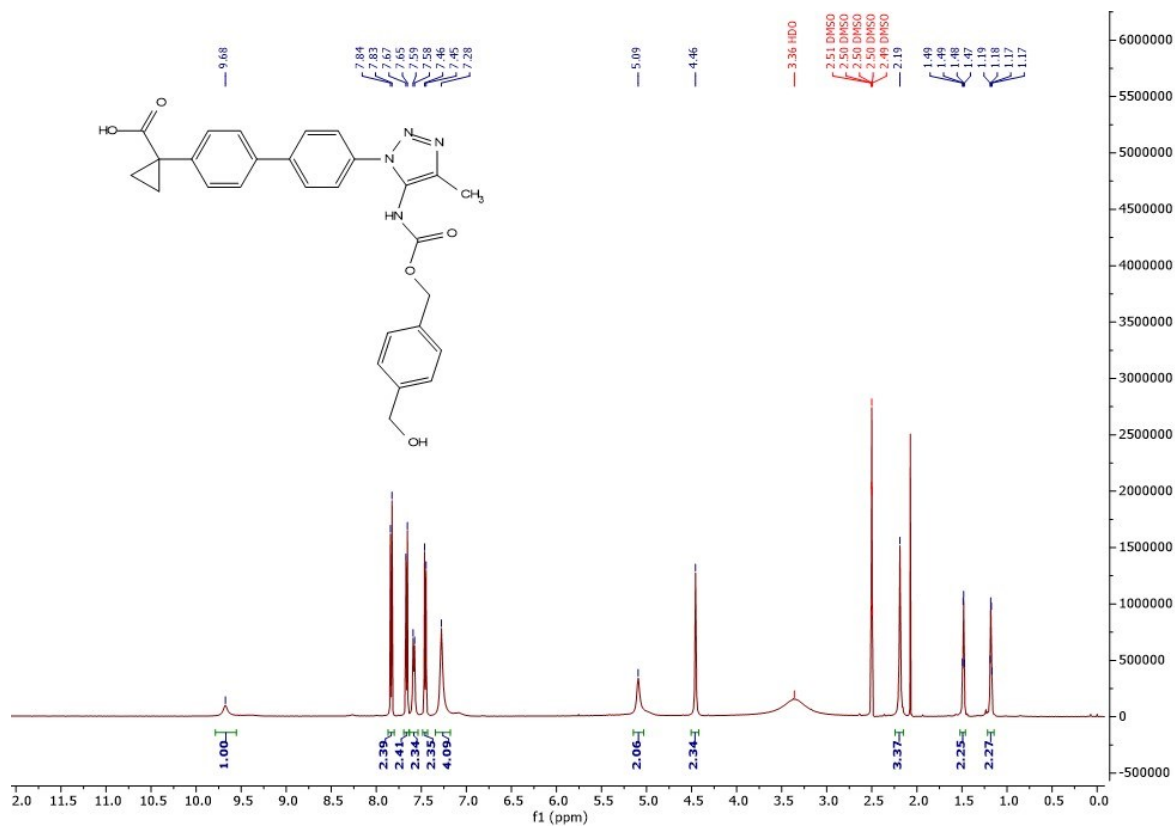
14a
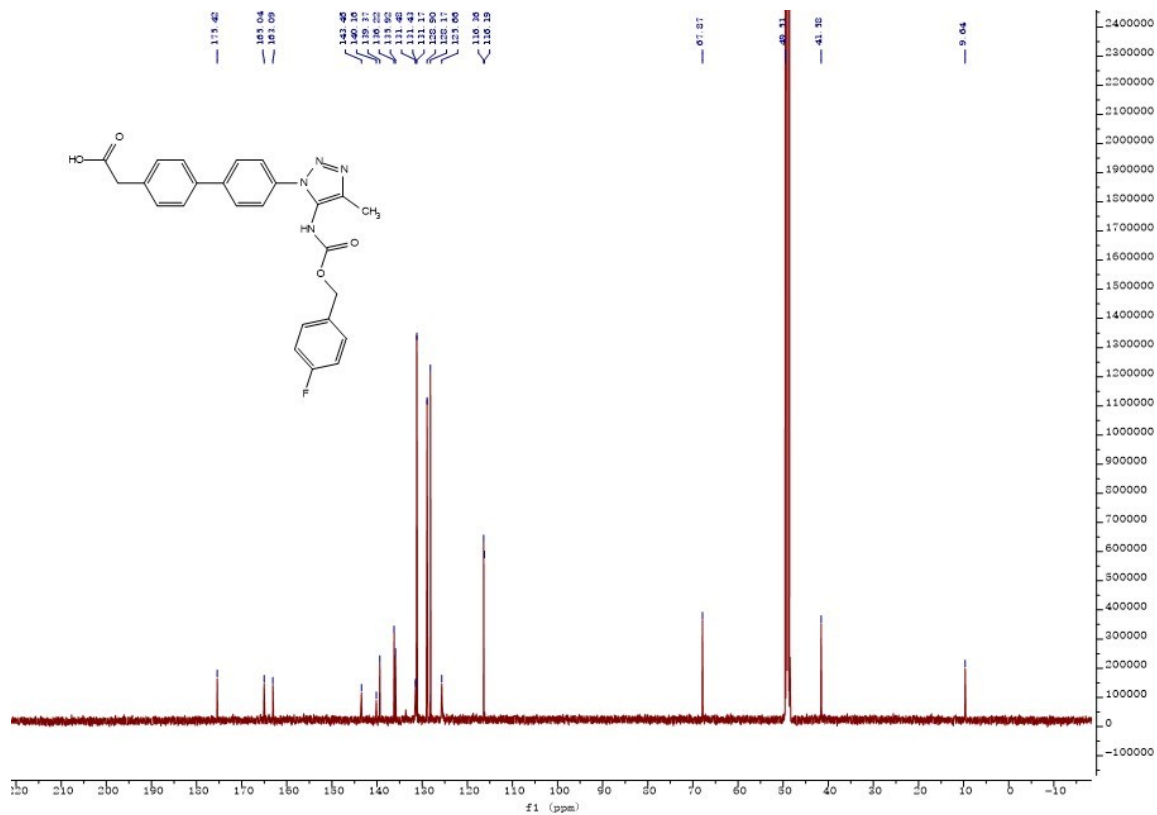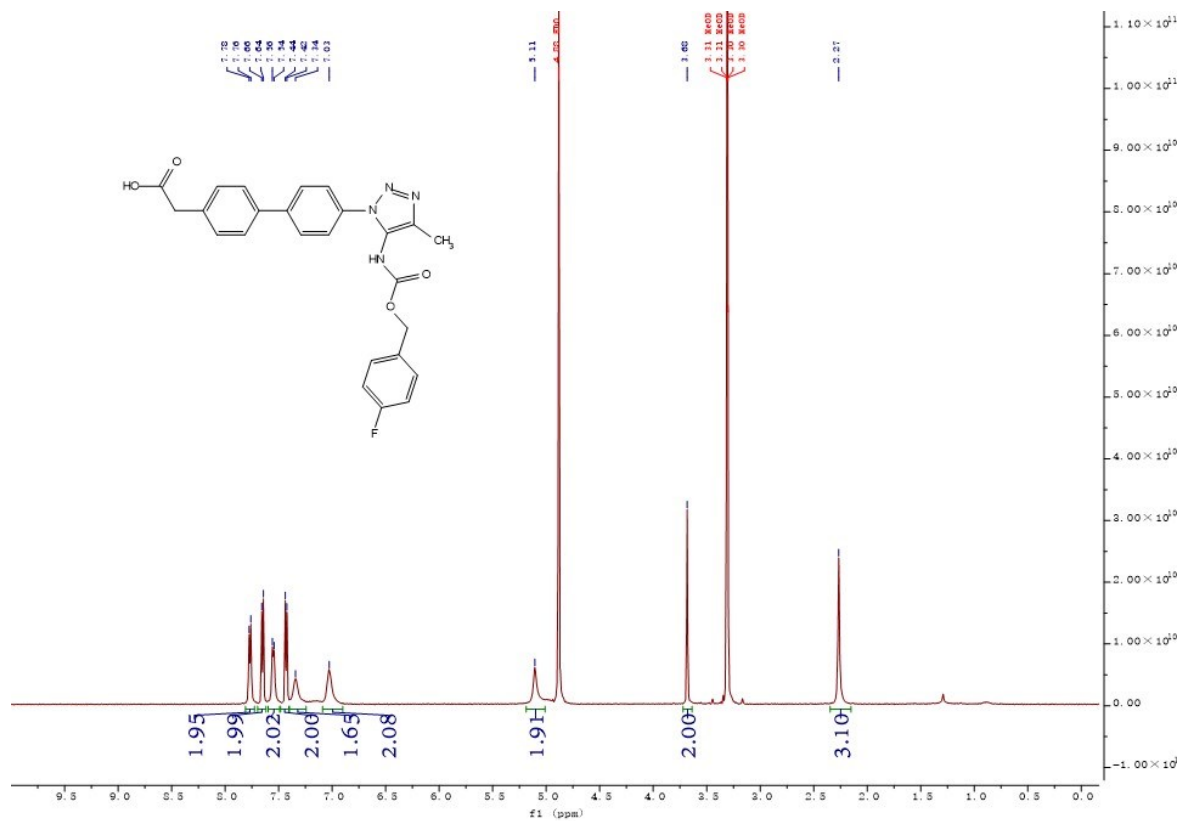
12k
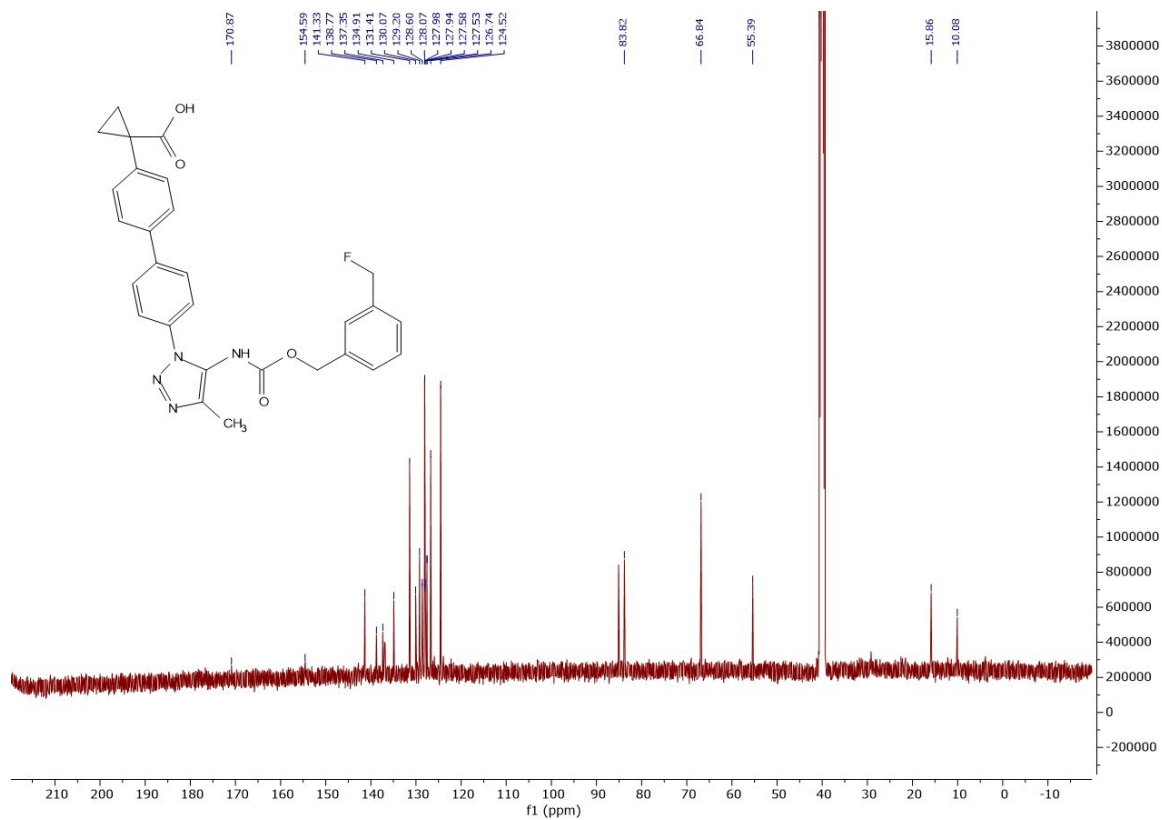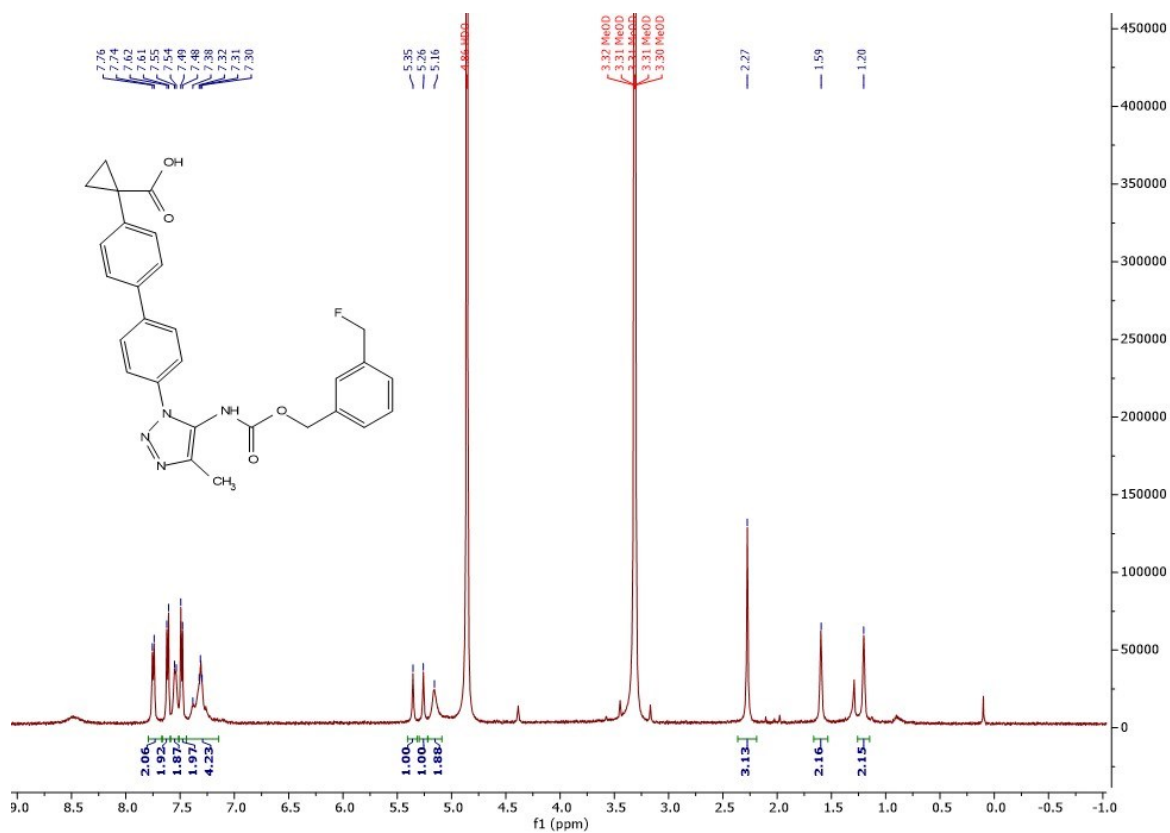
17a

17b

17c

17d

17e

17f
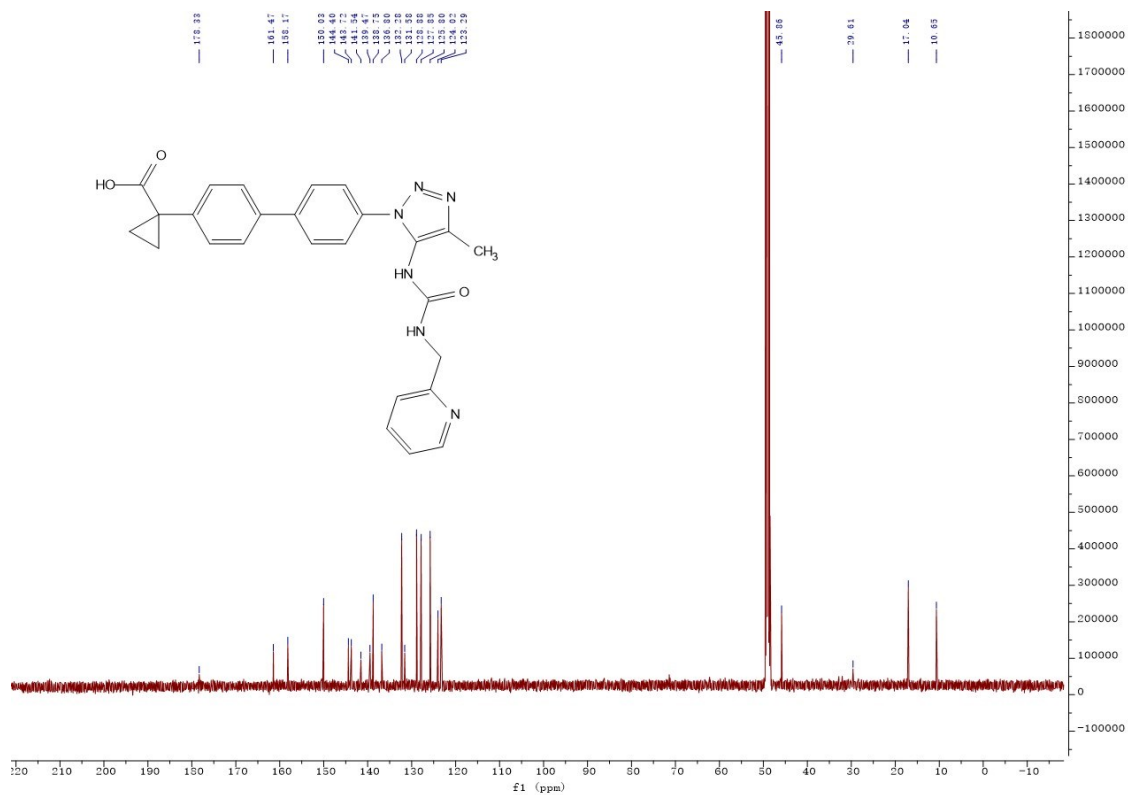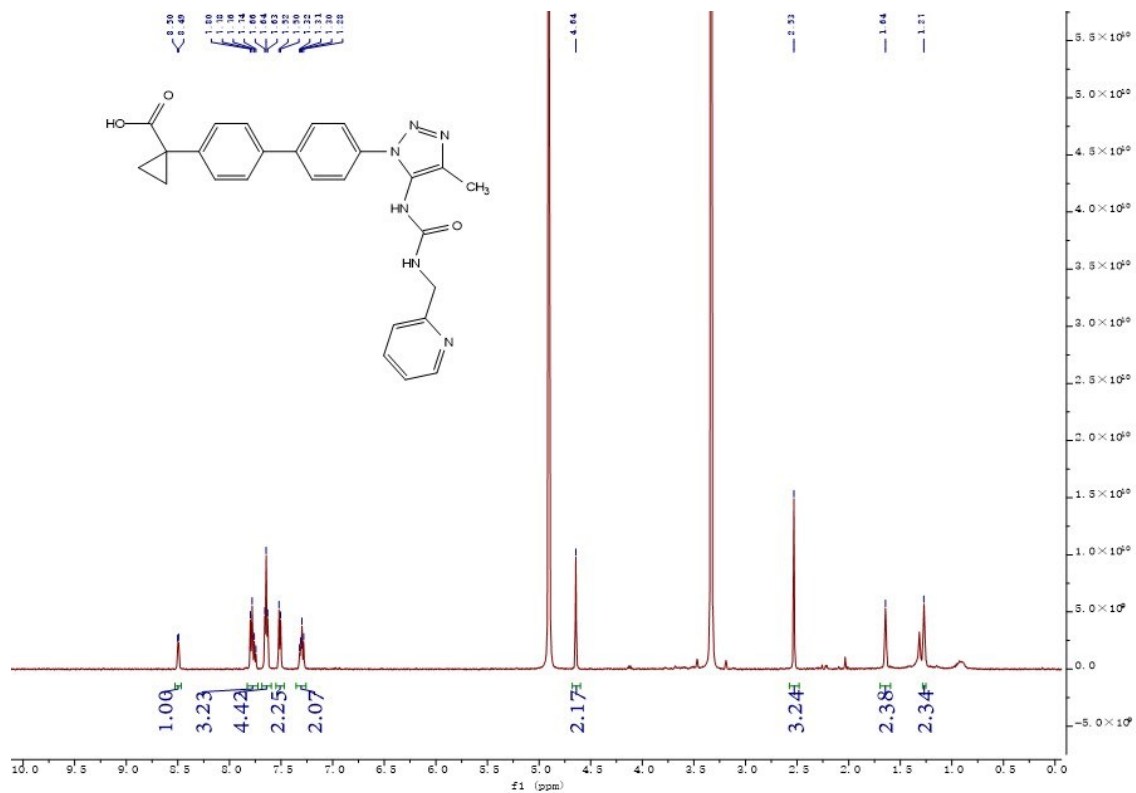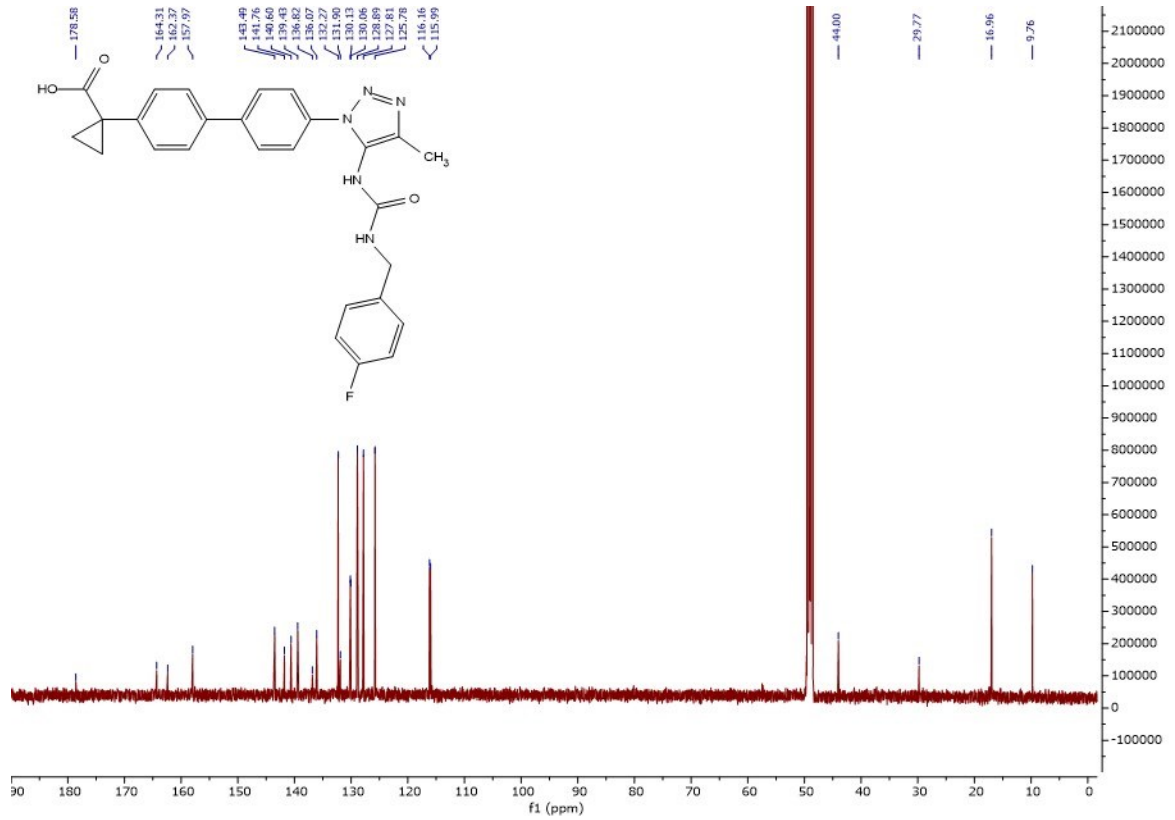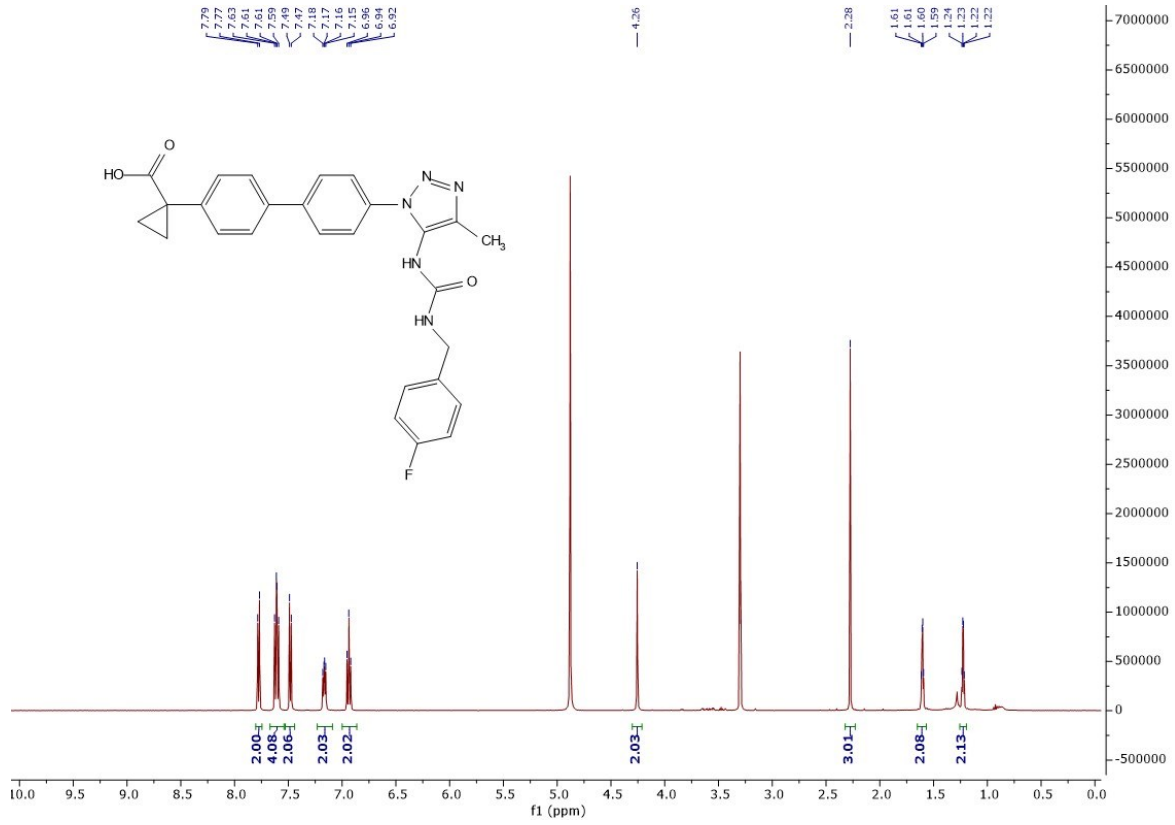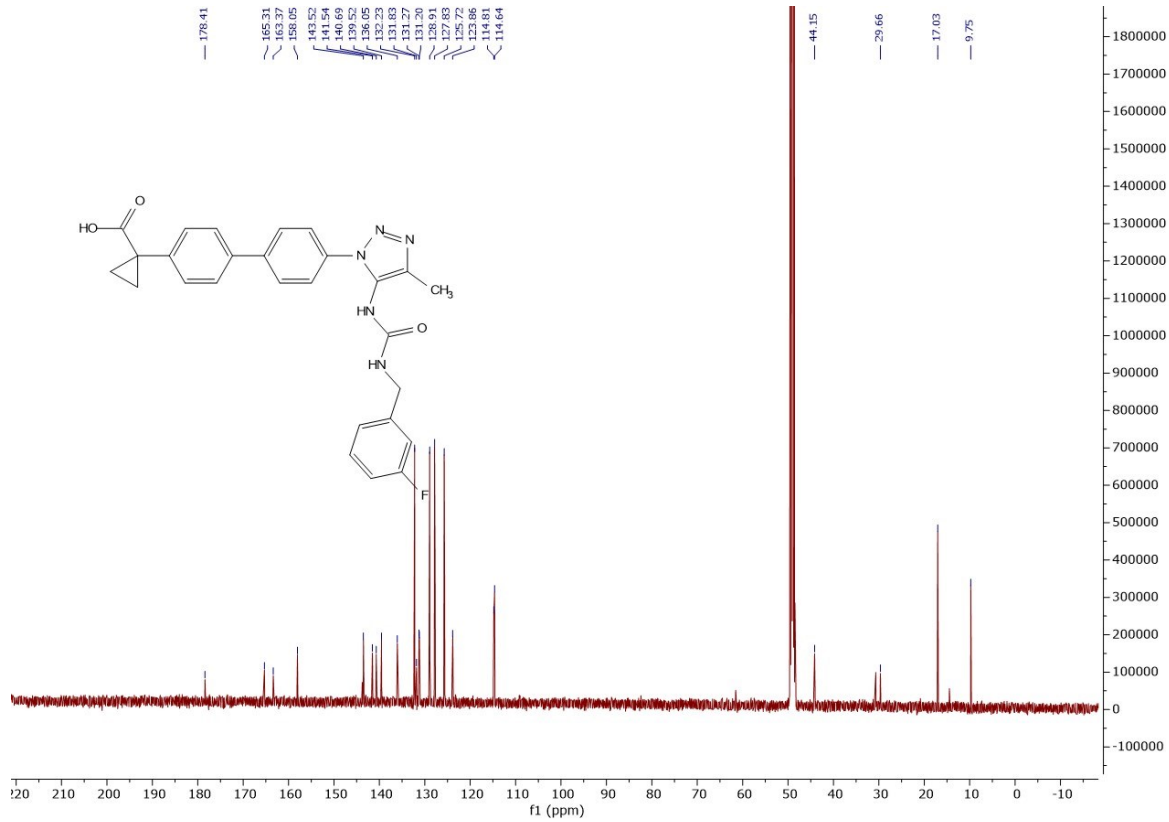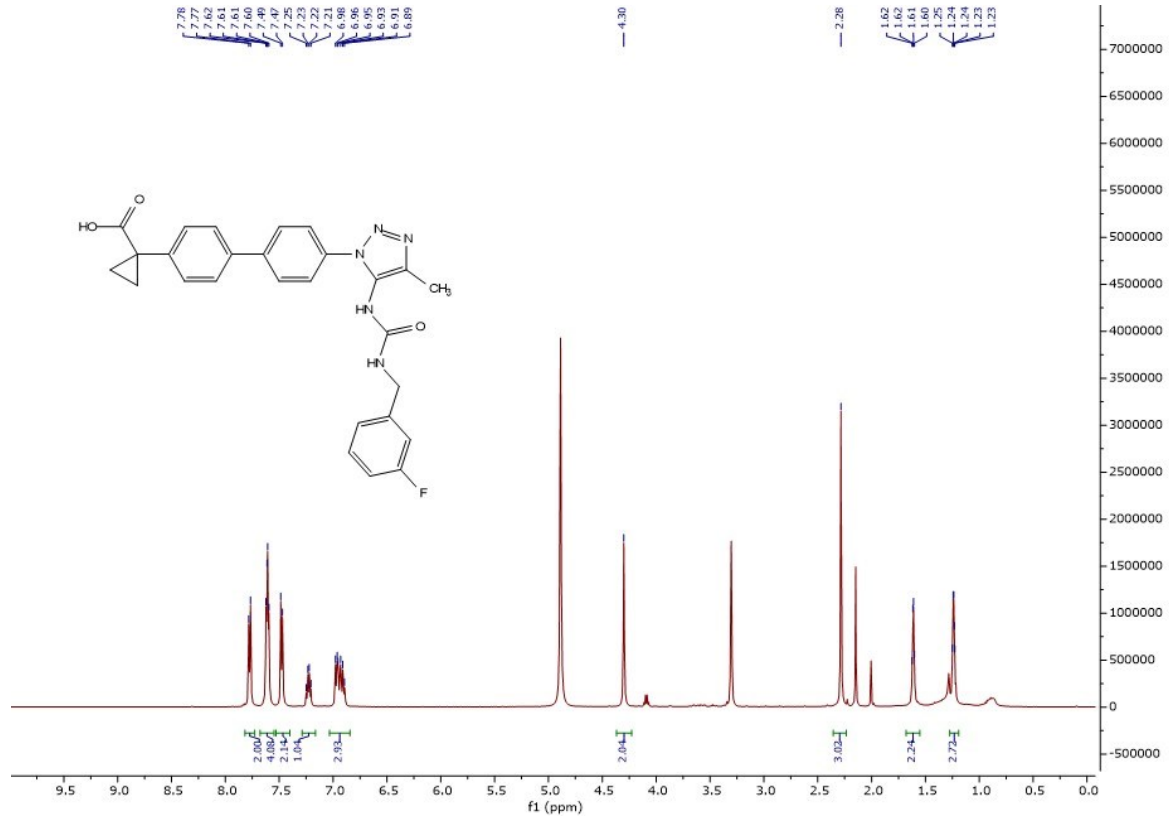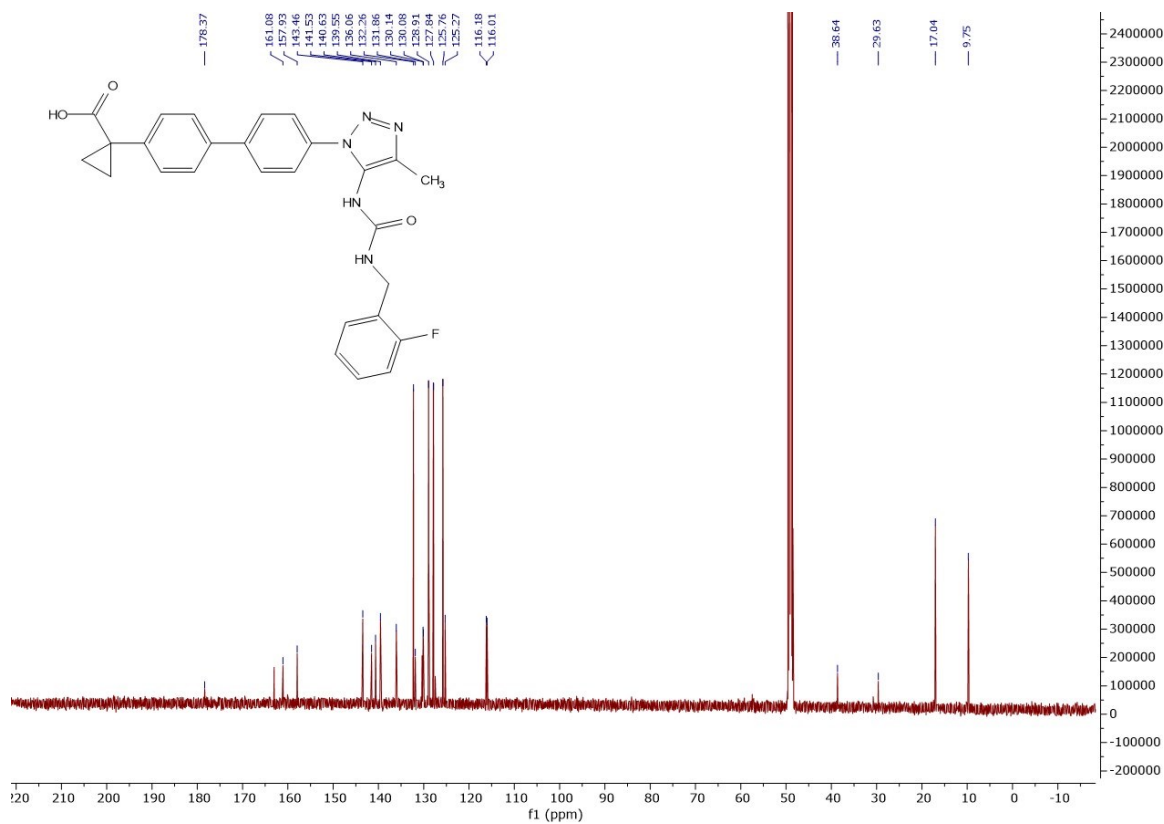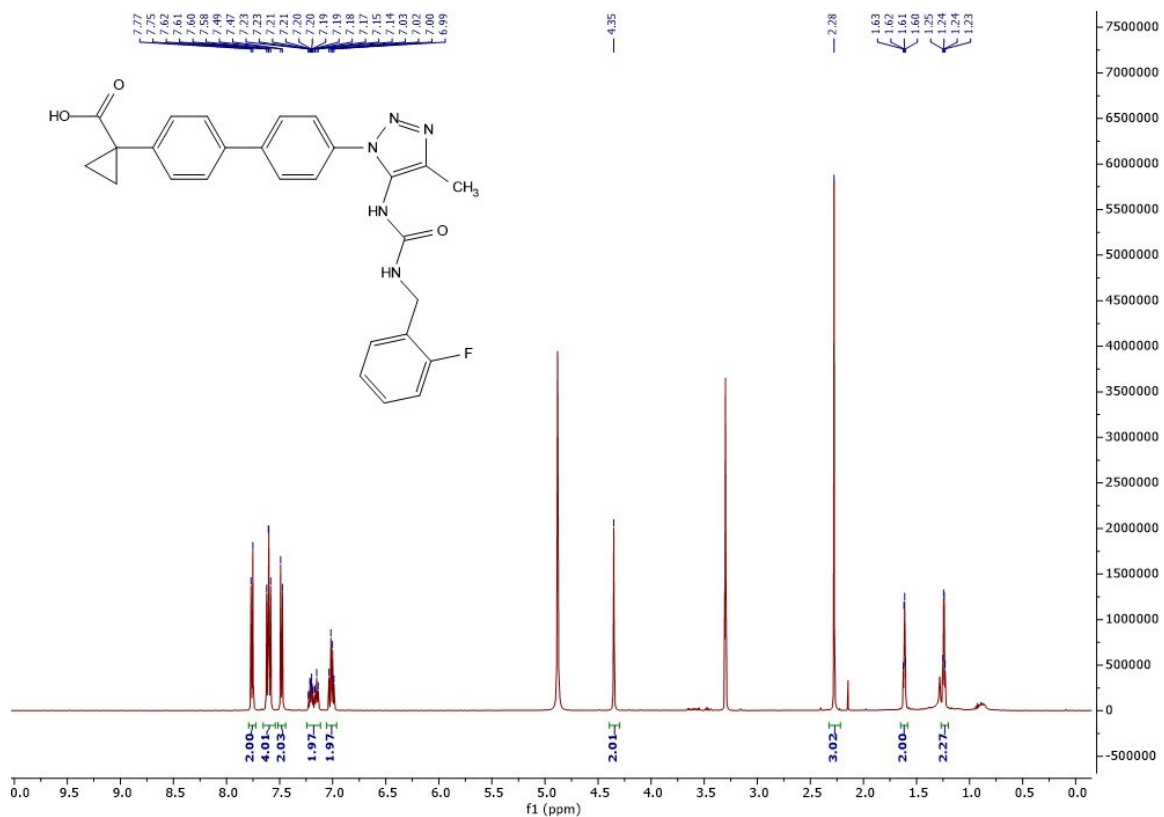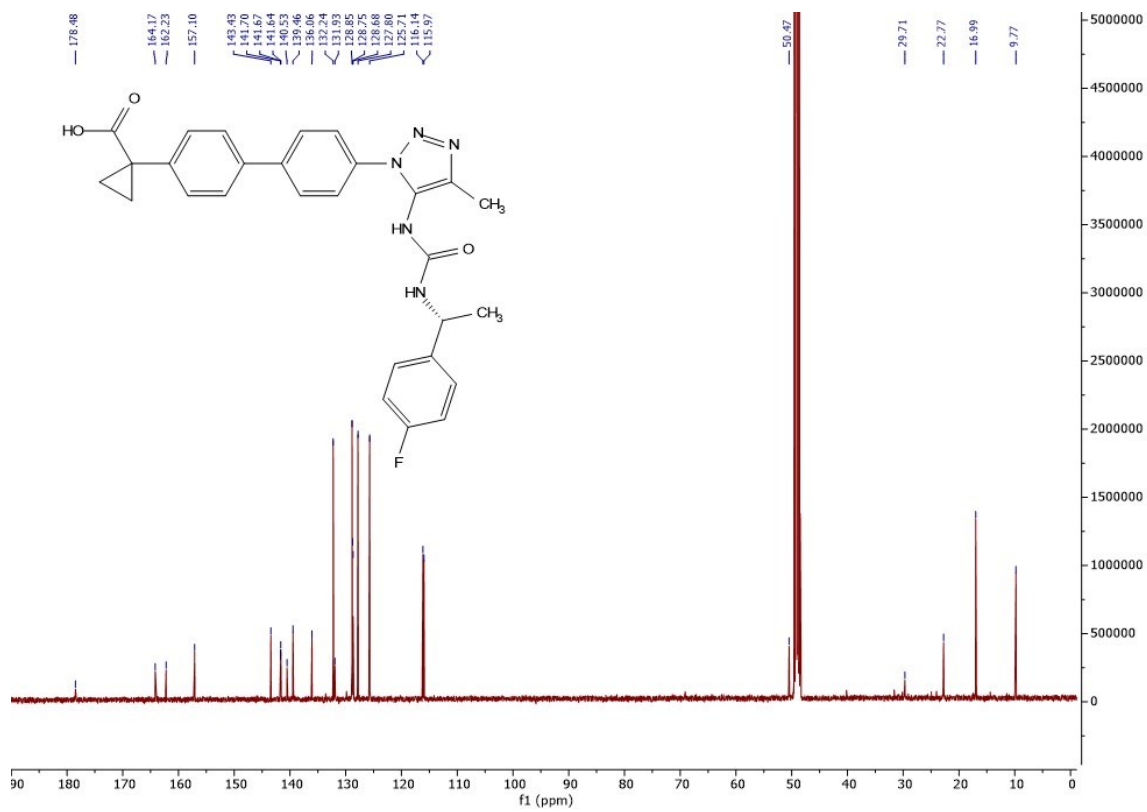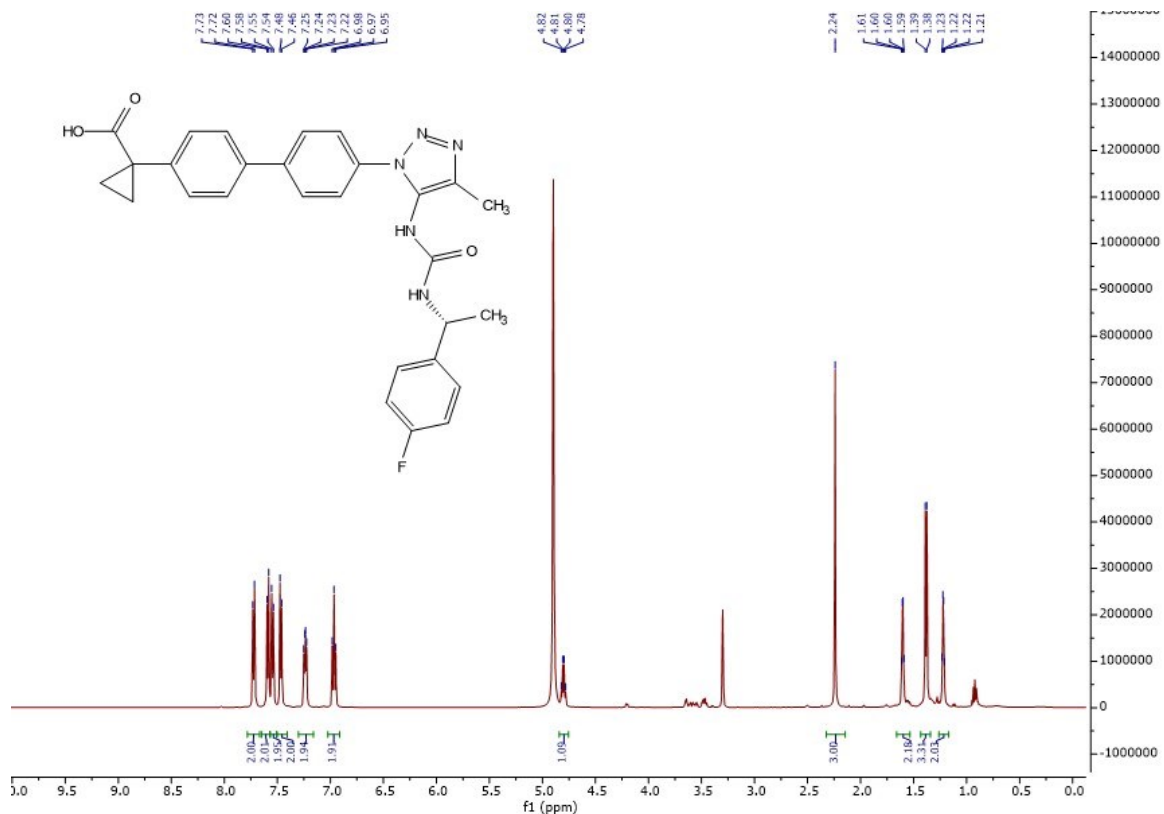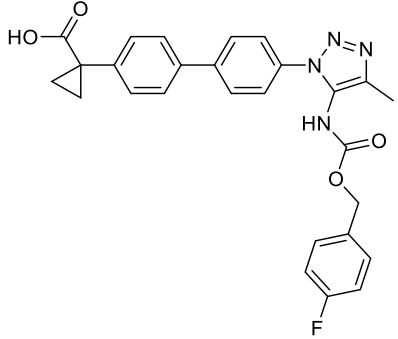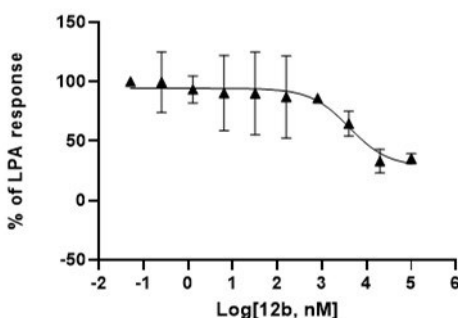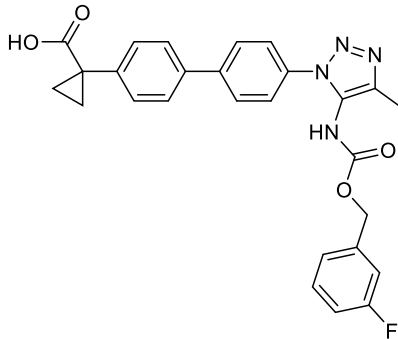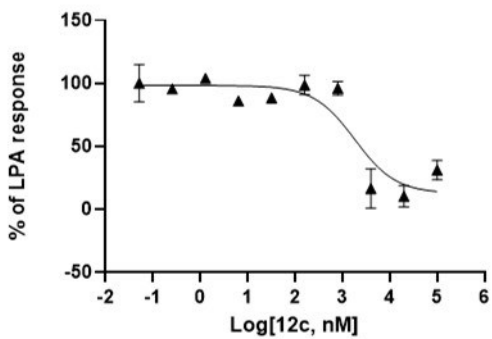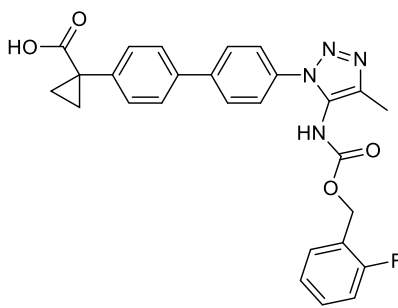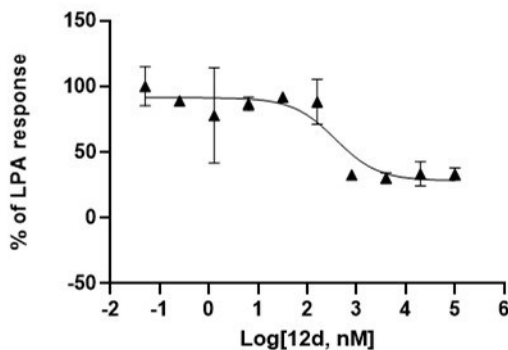
17g
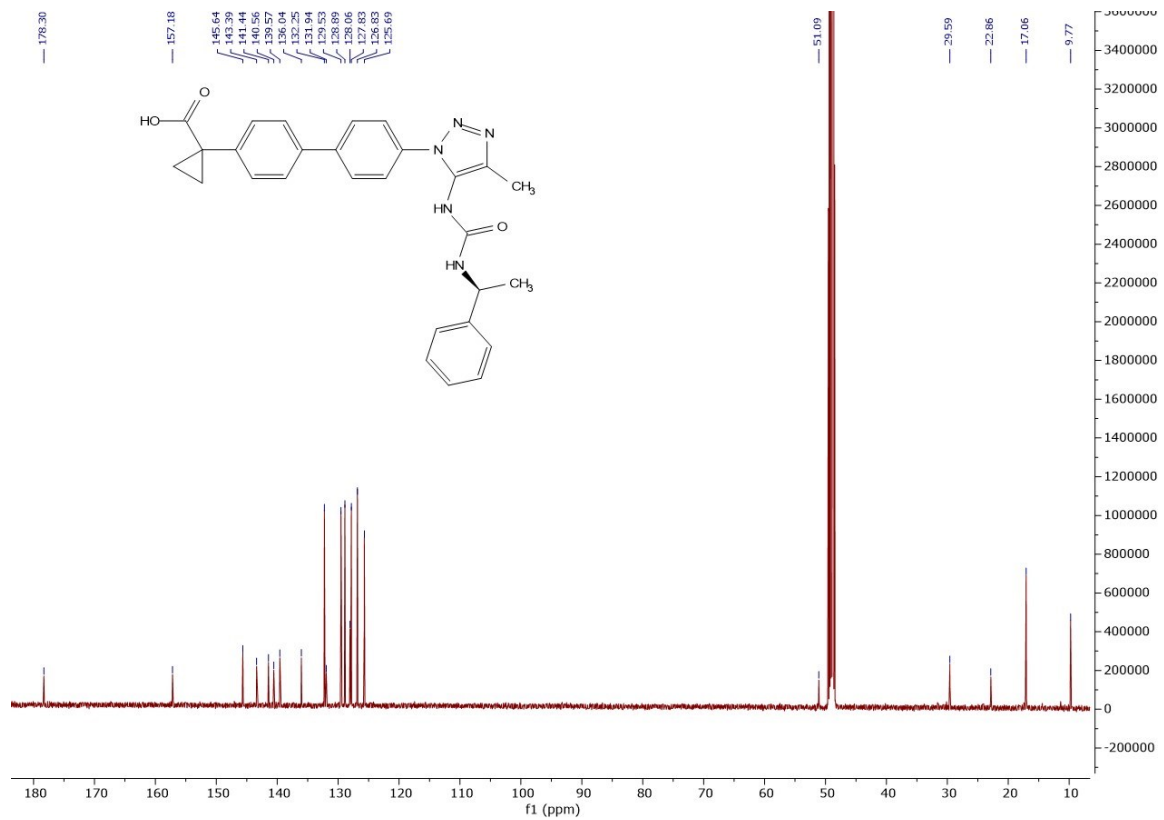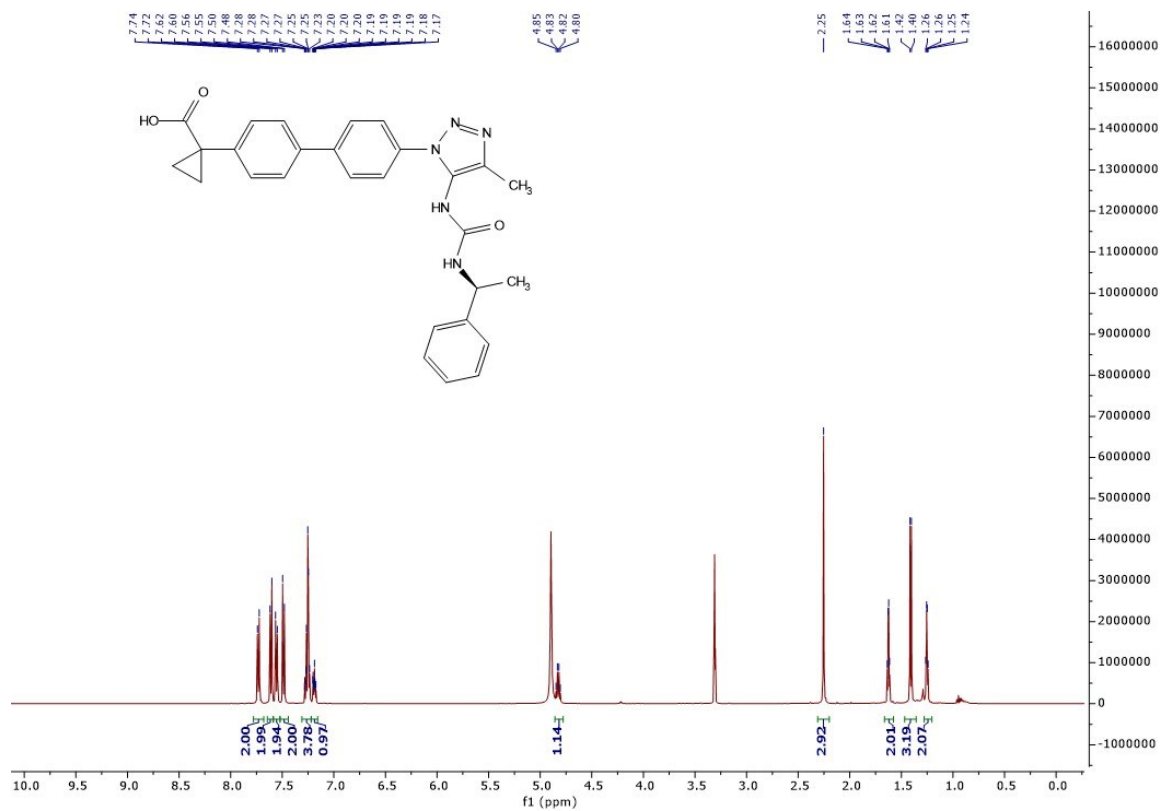
17h
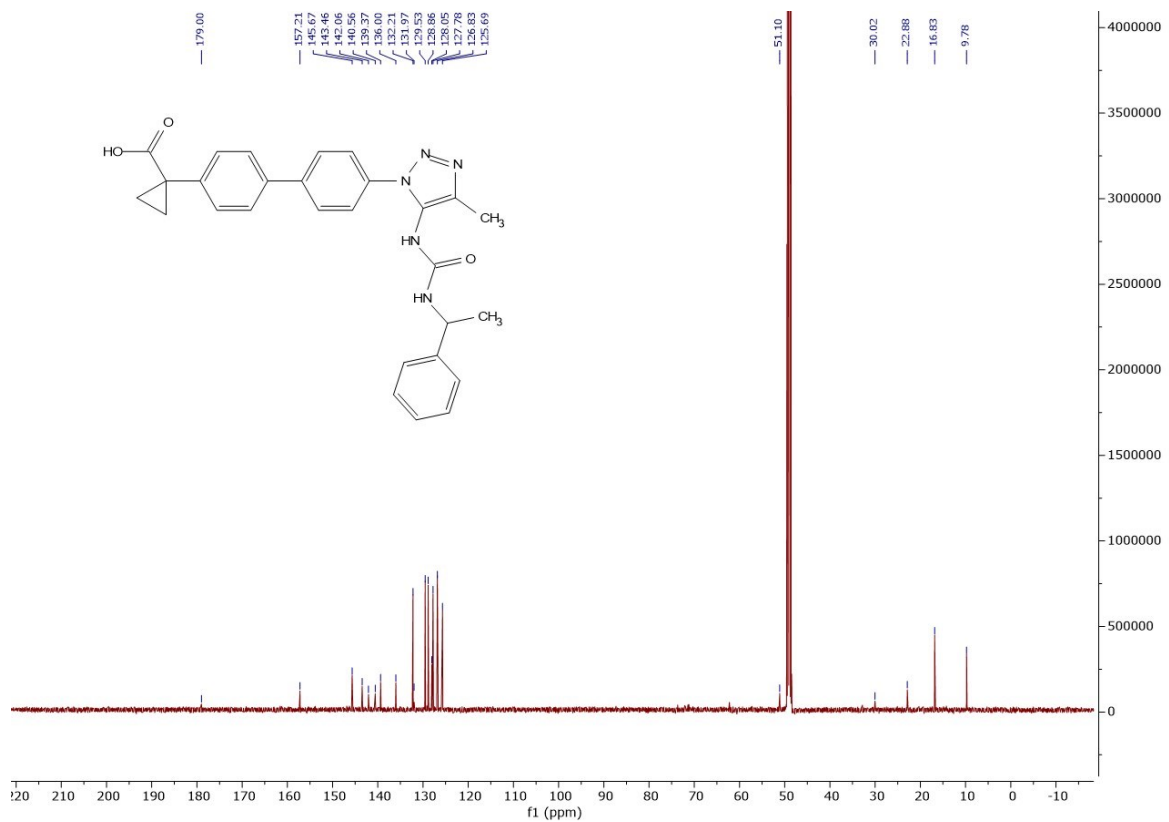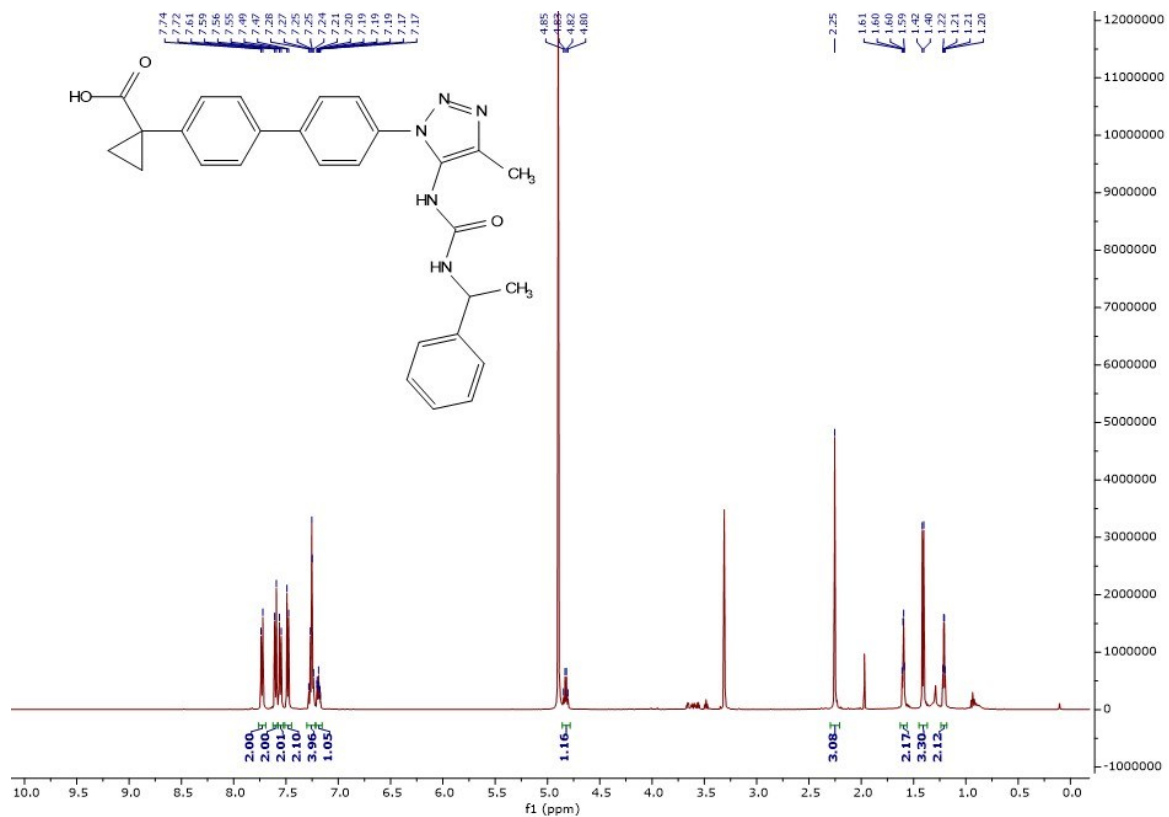
Appendix C: Dose-response curves

ID	Compds.	Dose-response curves
12b		
12c		
12d		

17b

CH$_3$

HN

O

HN

F

% of LPA response

150

100

50

0

-50

-2 -1 0 1 2 3 4 5 6

Log[17b, nM]

17c

HO

O

CH$_3$

HN

O

HN

F

% of LPA response

150

100

50

0

-50

-2 -1 0 1 2 3 4 5 6

Log[17c, nM]

17d

HO

O

CH$_3$

HN

O

HN

F

% of LPA response

150

100

50

0

-50

-2 -1 0 1 2 3 4 5 6

Log[17d, nM]

17e

HO

O

CH$_3$

HN

O

HN

CH$_3$

F

% of LPA response

150

100

50

0

-50

-2 -1 0 1 2 3 4 5 6

Log[17e, nM]
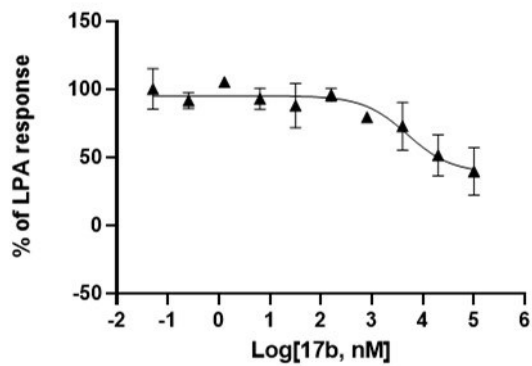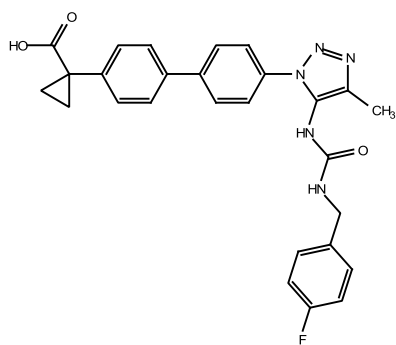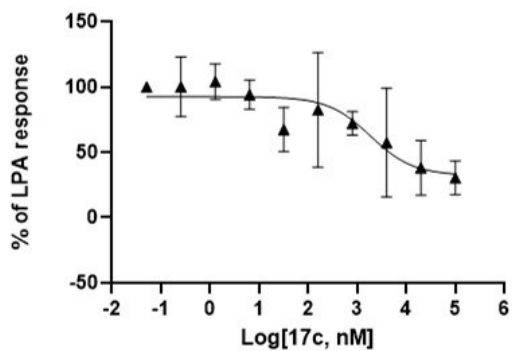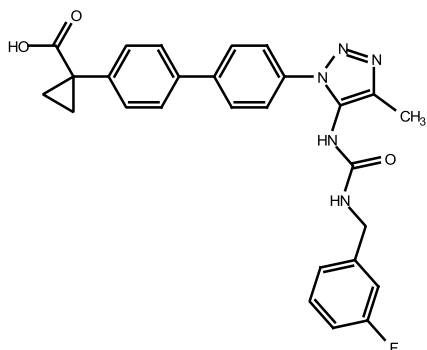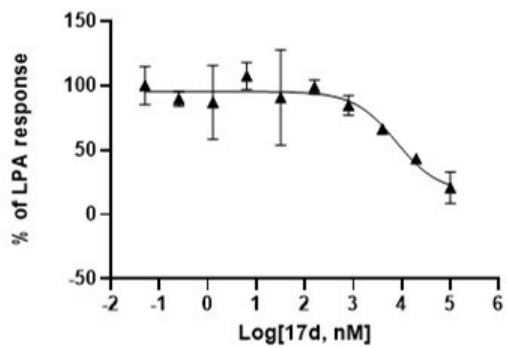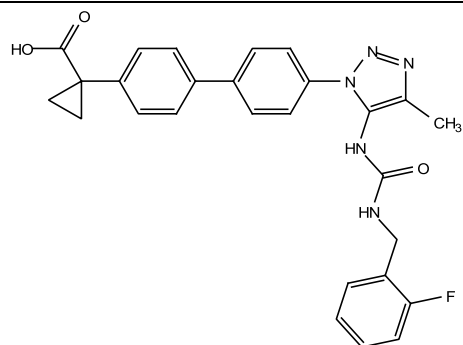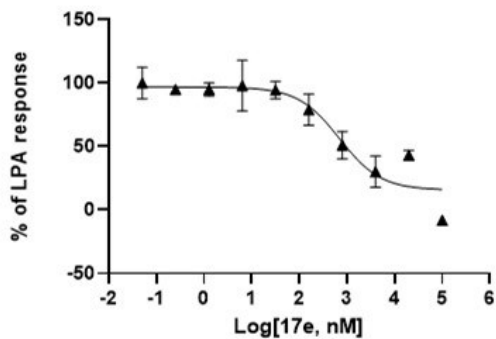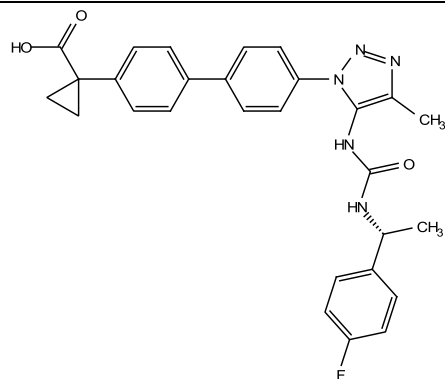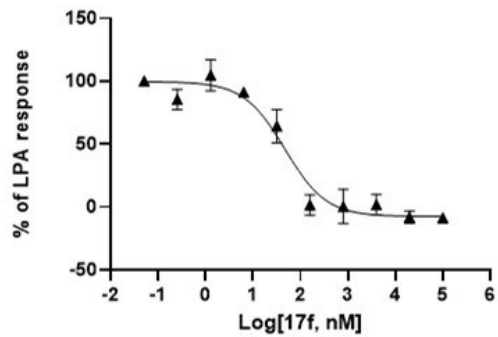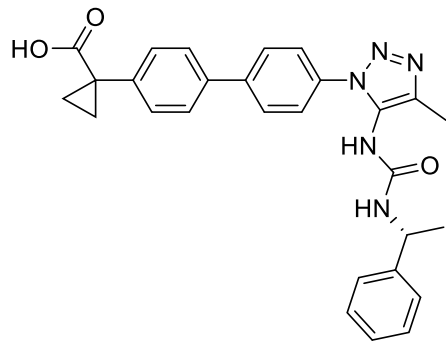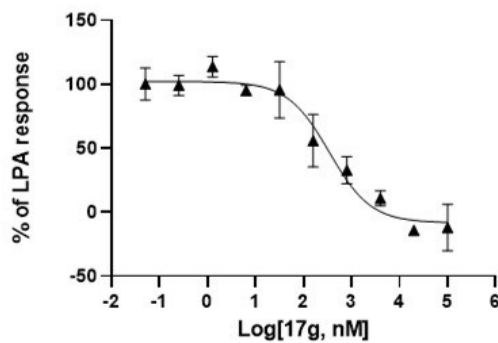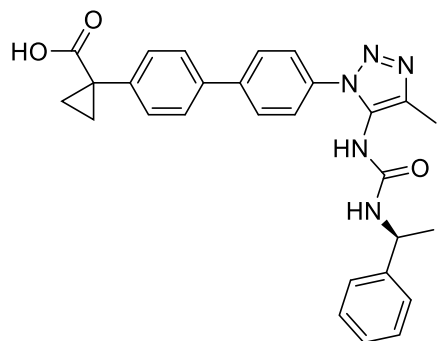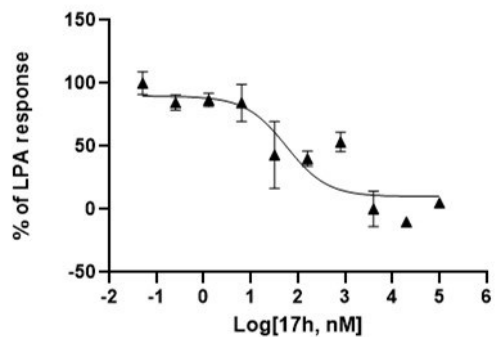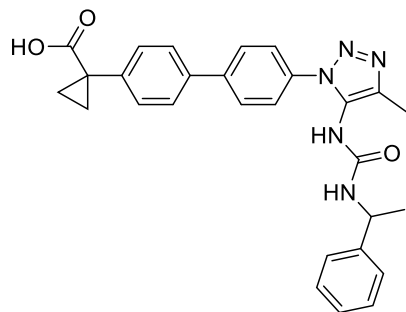
17f		
17g		
17h		
RO68 42262		
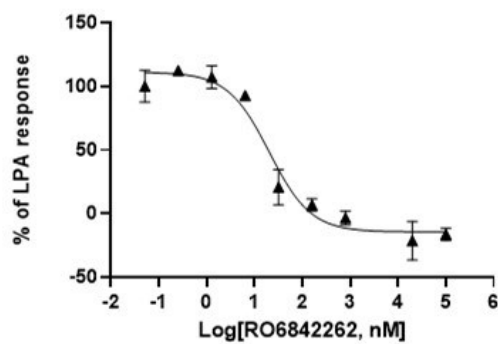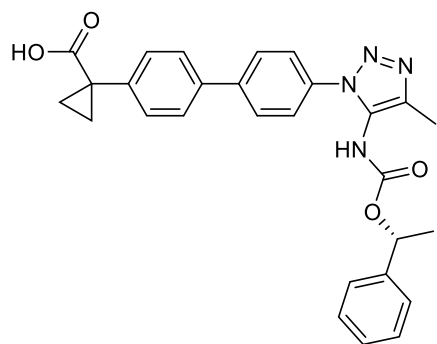
Appendix D: High-resolution mass spectrometry (HRMS)

ID	Formula	Calcd	Found
12b	$C_{27}H_{23}FN_4NaO_4$	509.1596	509.1600

ID	Formula	Calcd	Found
12c	$C_{27}H_{23}FN_4NaO_4$	509.1596	509.1598

ID	Formula	Calcd	Found
12d	$C_{27}H_{23}FN_4NaO_4$	509.1596	509.1588

12e $C_{28}H_{25}FN_4NaO_4$ 523.1752   523.1768

12f $C_{28}H_{25}FN_4NaO_4$ 523.1752   523.1748
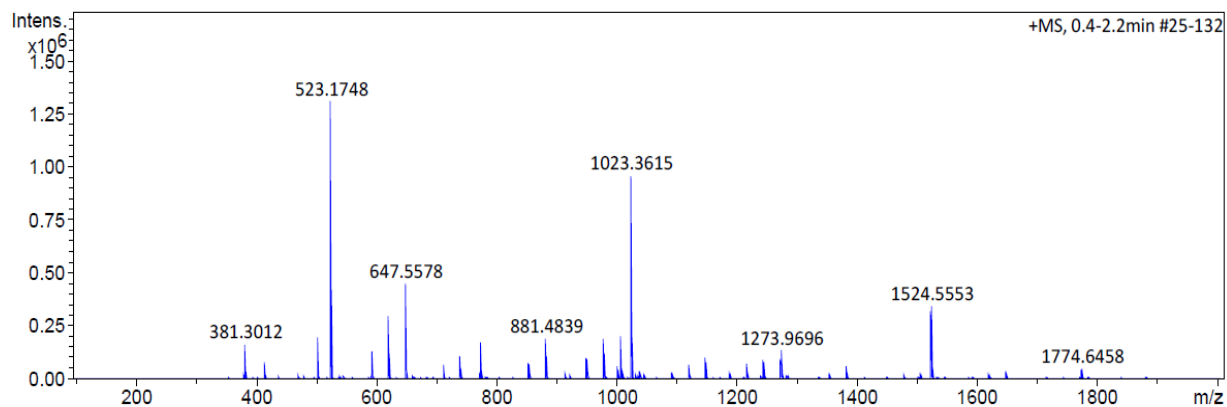
12g $C_{28}H_{25}FN_4NaO_4$ 523.1752   523.1759
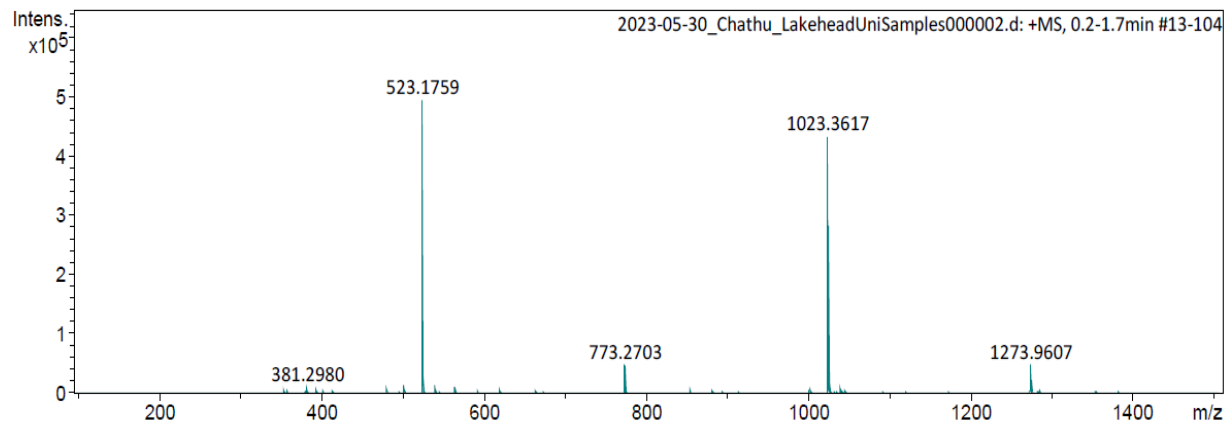
12h $C_{28}H_{26}N_4NaO_5$ 521.1795   521.1826
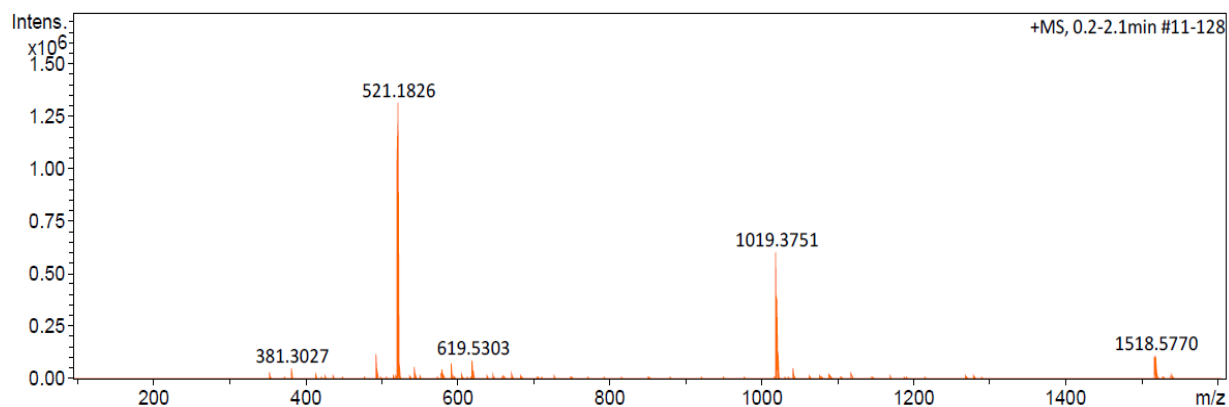
12i $C_{28}H_{26}N_4NaO_5$ 521.1795   521.1790
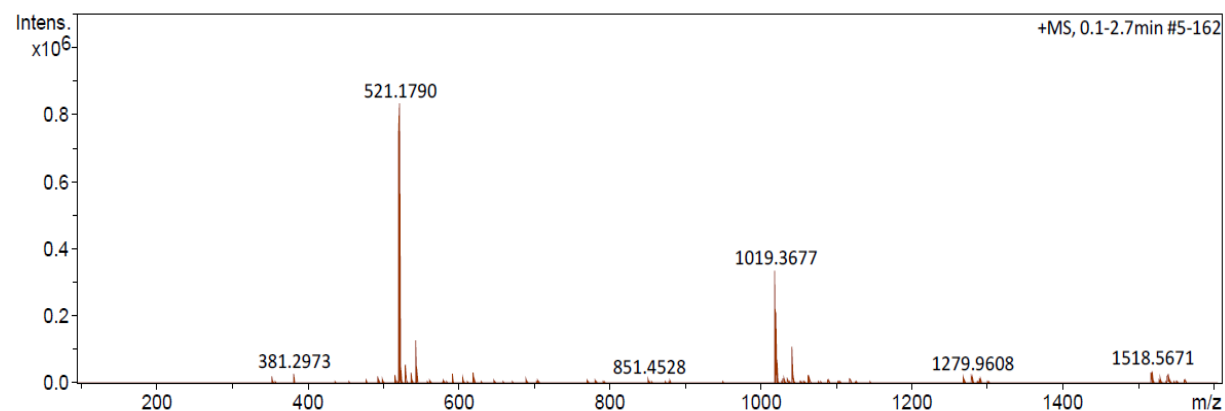
14a $C_{25}H_{21}FN_4NaO_4$ 483.1439   483.1453
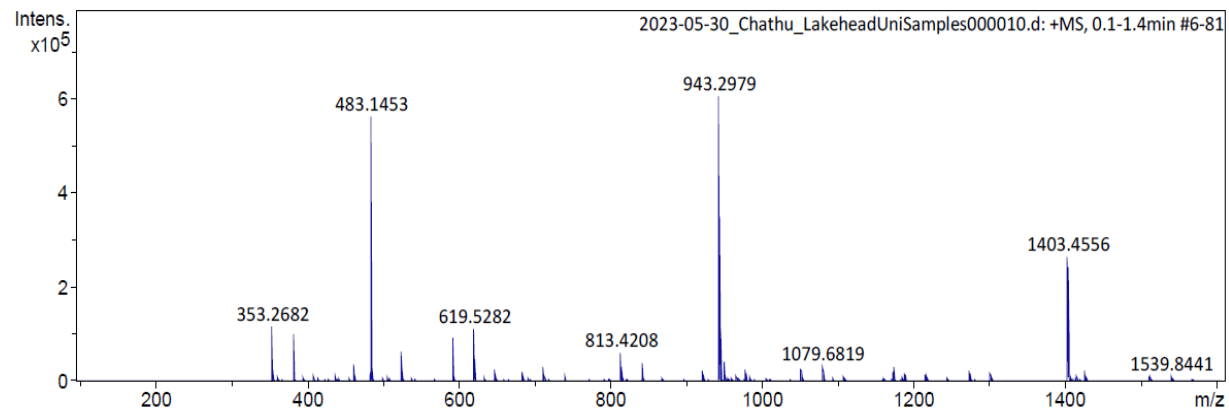
17b $C_{27}H_{24}FN_5NaO_3$ 508.1755   508.1770
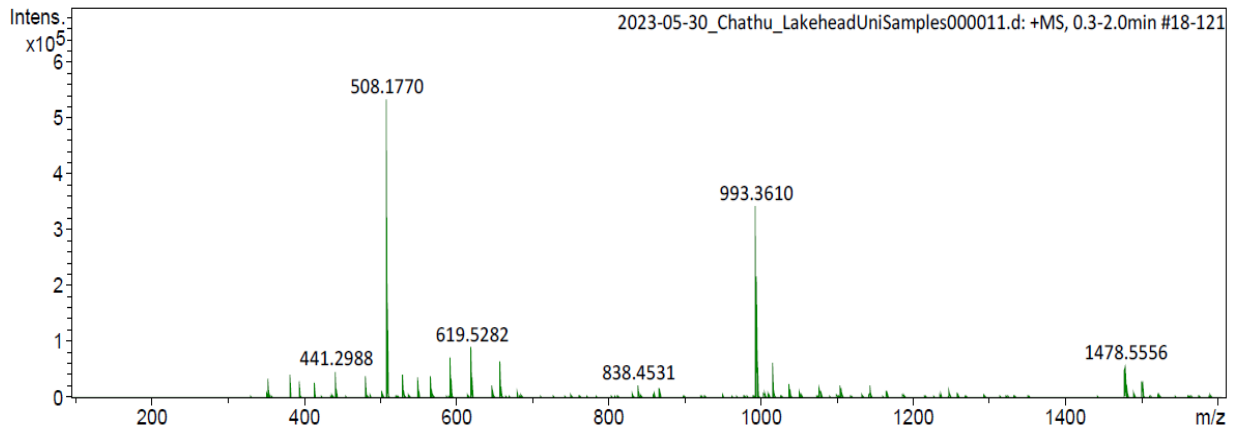
17c $C_{27}H_{24}FN_5NaO_3$ 508.1755   508.1776
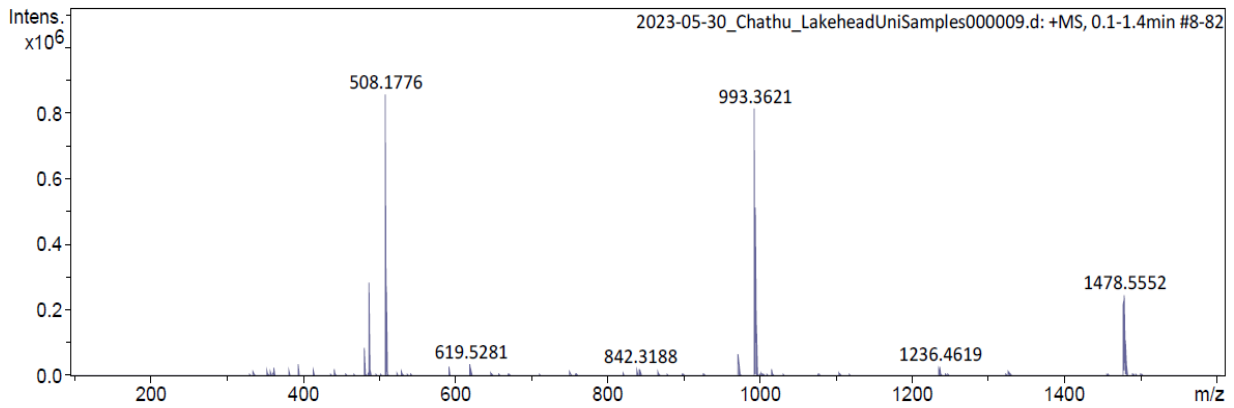
17d $C_{27}H_{24}FN_5NaO_3$ 508.1755   508.1770
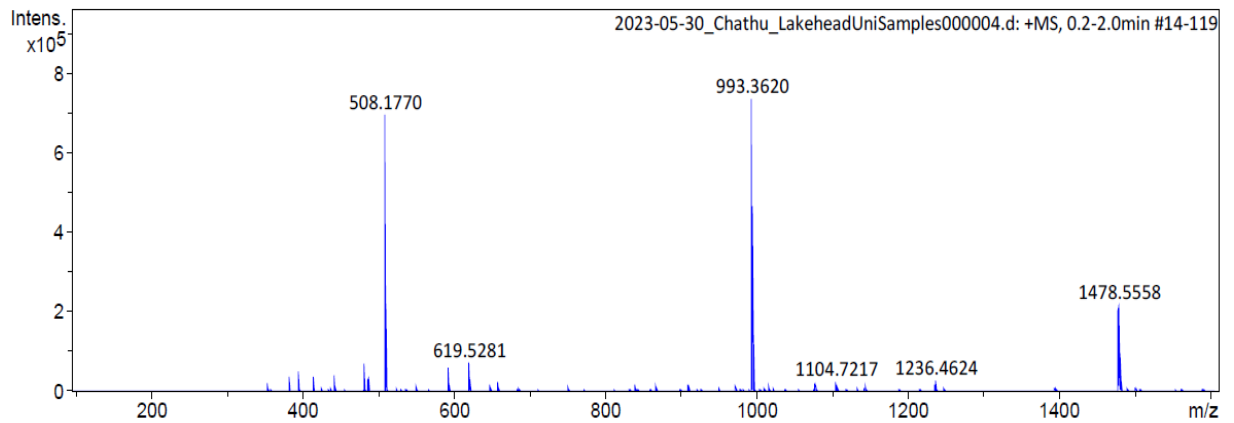
17e C$_{28}$H$_{26}$FN$_5$NaO$_3$ 522.1912 522.1909
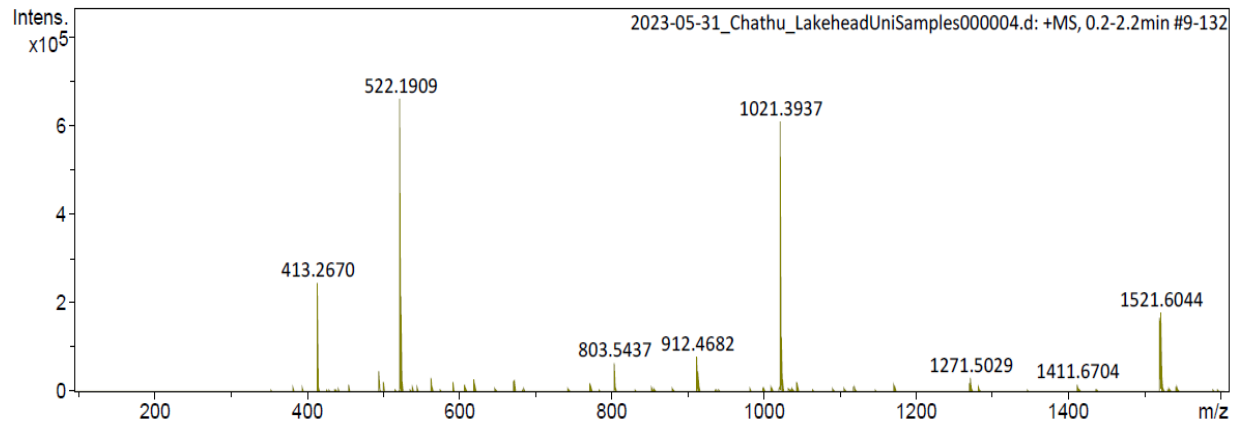
17f C$_{28}$H$_{27}$N$_5$NaO$_3$ 504.2006 504.2002
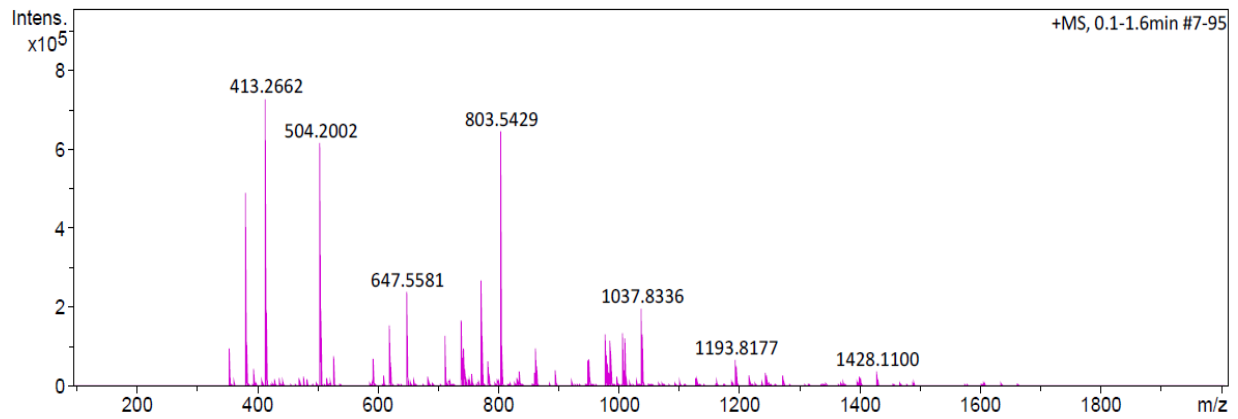
17g C$_{28}$H$_{27}$N$_5$NaO$_3$ 504.2006 504.2034
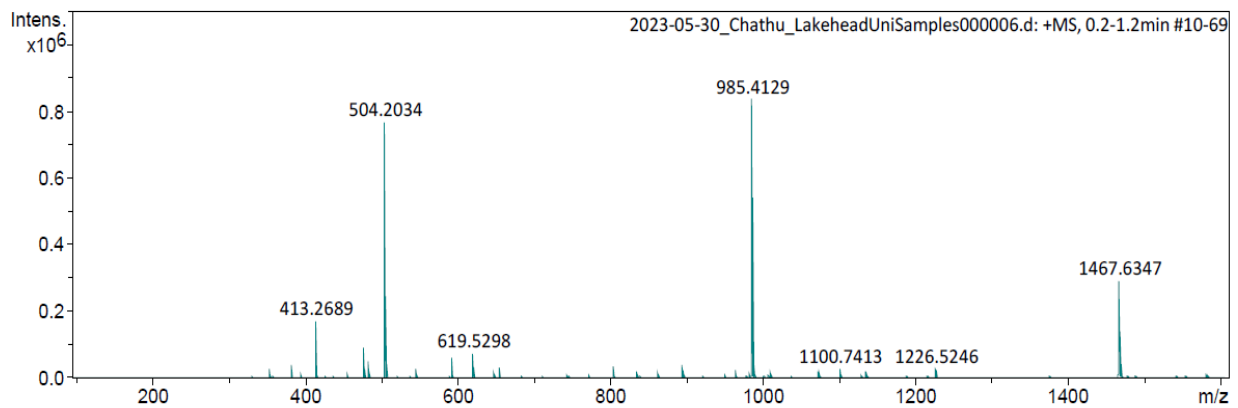
17h $C_{28}H_{27}N_5NaO_3$ 504.2006   504.2001
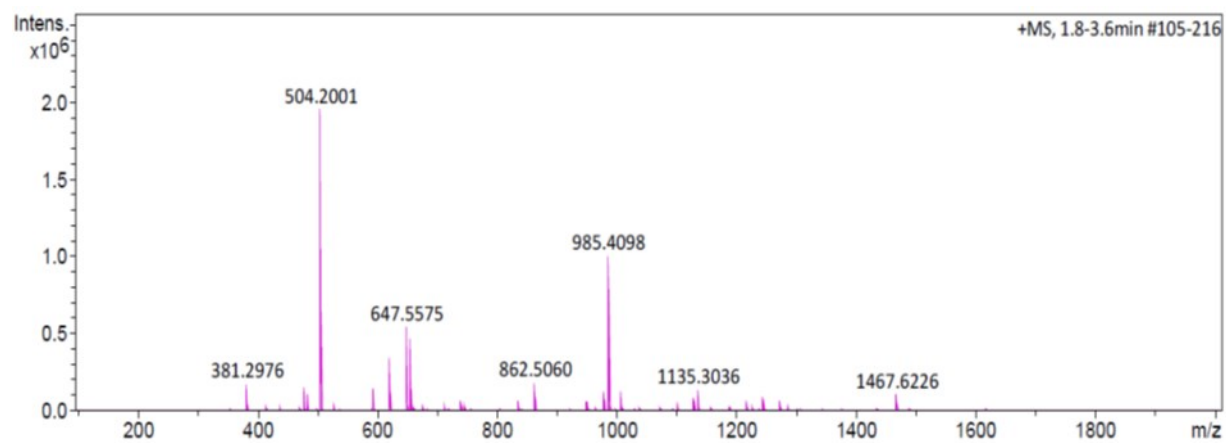
Appendix E

Sanger DNA sequencing/Whole plasmid sequencing Data:

pCMV6-AC-GFP-LPAR$_1$

ACCGCCGTTGCGCAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTG
AACCGTCAGAATTTTGTAATACGACTCACTATAGGGCGGCCGGGAATTCGTCGACTGGATCCGGTACC
GAGGAGATCTGCCGCCGCGATCGCCATGGCTGCCATCTCTACTTCCATCCCTGTAATTTCACAGCCCCA
GTTCACAGCCATGAATGAACCACAGTGCTTCTACAACGAGTCCATTGCCTTCTTTTATAACCGAAGTGG
AAAGCATCTTGCCACAGAATGGAACACAGTCAGCAAGCTGGTGATGGGACTTGGAATCACTGTTTGTA
TCTTCATCATGTTGGCCAACCTATTGGTCATGGTGGCAATCTATGTCAACCGCCGCTTCCATTTTCCTAT
TTATTACCTAATGGCTAATCTGGCTGCTGCAGACTTCTTTGCTGGGTTGGCCTACTTCTATCTCATGTTC
AACACAGGACCCAATACTCGGAGACTGACTGTTAGCACATGGCTCCTTCGTCAGGGCCTCATTGACAC
CAGCCTGACGGCATCTGTGGCCAACTTACTGGCTATTGCAATCGAGAGGCACATTACGGTTTTCCGCAT
GCAGCTCCACACACGGATGAGCAACCGGCGGGTAGTGGTGGTCATTGTGGTCATCTGGACTATGGCCA
TCGTTATGGGTGCTATACCCAGTGTGGGCTGGAACTGTATCTGTGATATTGAAAATTGTTCCAACATGG
CACCCCTCTACAGTGACTCTTACTTAGTCTTCTGGGCCATTTTCAACTTGGTGACCTTTGTGGTAATGGT
GGTTCTCTATGCTCACATCTTTGGCTATGTTCGCCAGAGGACTATGAGAATGTCTCGGCATAGTTCTGG
ACCCCGGCGGAATCGGGATACCATGATGAGTCTTCTGAAGACTGTGGTCATTGTGCTTGGGGCCTTTAT
CATCTGCTGGACTCCTGGATTGGTTTTGTTACTTCTAGACGTGTGCTGTCCACAGTGCGACGTGCTGGC
CTATT

CCCGCGGNNTTCAGGGCCAGGAGAGGCACTGGGGAGGGGTCACAGGGATGCCACCCGGGATCTGTTC
AGGAAACAGCTATGACCGCGGCCGGCCGTTTAAACTCTTTCTTCACCGGCATCTGCATCCGGGGTCTTG
AAGGCGTGCTGGTACTCCACGATGCCCAGCTCGGTGTTGCTGTGATCCTCCTCCACGCGGCGGAAGGC
GAACATGGGGCCCCCGTTCTGCAGGATGCTGGGGTGGATGGCGCTCTTGAAGTGCATGTGGCTGTCCA
CCACGGAGCTGTAGTAGCCGCCGTCGCGCAGGCTGAAGGTGCGGGTGAAGCTGCCATCCAGATCGTTA
TCGCCCATGGGGTGCAGGTGCTCCACGGTGGCGTTGCTGCGGATGATCTTGTCGGTGAAGATCACGCT
GTCCTCGGGGAAGCCGGTGCCCATCACCTTGAAGTCGCCGATCACGCGGCCGGCCTCGTAGCGGTAGC
TGAAGCTCACGTGCAGCACGCCGCCGTCCTCGTACTTCTCGATGCGGGTGTTGGTGTAGCCGCCGTTGT
TGATGGCGTGCAGGAAGGGGTTCTCGTAGCCGCTGGGGTAGGTGCCGAAGTGGTAGAAGCCGTAGCC
CATCACGTGGCTCAGCAGGTAGGGGCTGAAGGTCAGGGCGCCTTTGGTGCTCTTCATCTTGTTGGTCAT
GCGGCCCTGCTCGGGGGGTGCCCTCTCCGCCGCCCACCAGCTCGAACTCCACGCCGTTCAGGGTGCCGG
TGATGCGGCACTCGATCTCCATGGCGGGCAGGCCGCTCTCGTCGCTCTCCATCTCGAGCGGCCGCGTA
CGCGTAACCACAGAGTGGTCATTGCTGTGAACTCCAGCCAAGATGGTGTGGTTGAGGGAGGAAGCCG
AGCGGTCTGAGCCTTCTGTGGGGCCGGTGGGGTTCTCACTGCGCTGGCAGCAGAGGATCTGCCTAAAG
GTGGCGCTCATTTCTTTGTCGCGGTAGGAGTAAATGATGGGGTTCATGGCAGAGTTGAATTCAGCAAG
GAGAAGGAAGAATTTCTCATAGGCCAGCACGTCGCACTGTGGACAGCACACGTCTAGAAGA

pCMV6-AC-GFP-LPAR$_2$

TCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTG
CGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGC
TCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTT
CGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCA
TAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATT
CTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCAC
ATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTA
CCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTC
ACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACA
CGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCA

TGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGA
AAAGTGCCACCTGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGT
GACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTC
GCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCAC
CTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTT
TCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAA
CCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAG
CTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGCTTACAATTTCCATTCGCCATTCA
GGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGG
GGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGAC
GGCCAGTGCCAAGCTGATCTATACATTGAATCAATATTGGCAATTAGCCATATTAGTCATTGGTTATAT
AGCATAAATCAATATTGGCTATTGGCCATTGCATACGTTGTATCTATATCATAATATGTACATTTATAT
TGGCTCATGTCCAATATGACCGCCATGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTAC
GGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTG
GCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATA
GGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGT
GTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCC
AGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGG
TGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTC
CACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAA
TAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTC
GTTTAGTGAACCGTCAGAATTTTGTAATACGACTCACTATAGGGCGGCCGGGAATTCGTCGACTGGAT
CCGGTACCGAGGAGATCTGCCGCCGCGATCGCCATGGTCATCATGGGCCAGTGCTACTACAACGAGAC
CATCGGCTTCTTCTATAACAACAGTGGCAAAGAGCTCAGCTCCCACTGGCGGCCCAAGGATGTGGTCG
TGGTGGCACTGGGGCTGACCGTCAGCGTGCTGGTGCTGCTGACCAATCTGCTGGTCATAGCAGCCATC
GCCTCCAACCGCCGCTTCCACCAGCCCATCTACTACCTGCTCGGCAATCTGGCCGCGGCTGACCTCTTC
GCGGGCGTGGCCTACCTCTTCCTCATGTTCCACACTGGTCCCCGCACAGCCCGACTTTCACTTGAGGGC
TGGTTCCTGCGGCAGGGCTTGCTGGACACAAGCCTCACTGCGTCGGTGGCCACACTGCTGGCCATCGC
CGTGGAGCGGCACCGCAGTGTGATGGCCGTGCAGCTGCACAGCCGCCTGCCCCGTGGCCGCGTGGTCA
TGCTCATTGTGGGCGTGTGGGTGGCTGCCCTGGGCCTGGGGCTGCTGCCTGCCCACTCCTGGCACTGCC
TCTGTGCCCTGGACCGCTGCTCACGCATGGCACCCCTGCTCAGCCGCTCCTATTTGGCCGTCTGGGCTC
TGTCGAGCCTGCTTGTCTTCCTGCTCATGGTGGCTGTGTACACCCGCATTTTCTTCTACGTGCGGCGGC
GAGTGCAGCGCATGGCAGAGCATGTCAGCTGCCACCCCCGCTACCGAGAGACCACGCTCAGCCTGGTC
AAGACTGTTGTCATCATCCTGGGGGCGTTCGTGGTCTGCTGGACACCAGGCCAGGTGGTACTGCTCCT
GGATGGTTTAGGCTGTGAGTCCTGCAATGTCCTGGCTGTAGAAAAGTACTTCCTACTGTTGGCCGAGG
CCAACTCACTGGTCAATGCTGCTGTGTACTCTTGCCGAGATGCTGAGATGCGCCGCACCTTCCGCCGCC
TTCTCTGCTGCGCGTGCCTCCGCCAGTCCACCCGCGAGTCTGTCCACTATACATCCTCTGCCCAGGGAG
GTGCCAGCACTCGCATCATGCTTCCCGAGAACGGCCACCCACTGATGGACTCCACCCTTACGCGTACG
CGGCCGCTCGAGATGGAGAGCGACGAGAGCGGCCTGCCCGCCATGGAGATCGAGTGCCGCATCACCG
GCACCCTGAACGGCGTGGAGTTCGAGCTGGTGGGCGGCGGAGAGGGCACCCCCGAGCAGGGCCGCAT
GACCAACAAGATGAAGAGCACCAAAGGCGCCCTGACCTTCAGCCCCTACCTGCTGAGCCACGTGATG
GGCTACGGCTTCTACCACTTCGGCACCTACCCCAGCGGCTACGAGAACCCCTTCCTGCACGCCATCAA
CAACGGCGGCTACACCAACACCCGCATCGAGAAGTACGAGGACGGCGGCGTGCTGCACGTGAGCTTC
AGCTACCGCTACGAGGCCGGCCGCGTGATCGGCGACTTCAAGGTGATGGGCACCGGCTTCCCCGAGGA
CAGCGTGATCTTCACCGACAAGATCATCCGCAGCAACGCCACCGTGGAGCACCTGCACCCCATGGGCG
ATAACGATCTGGATGGCAGCTTCACCCGCACCTTCAGCCTGCGCGACGGCGGCTACTACAGCTCCGTG
GTGGACAGCCACATGCACTTCAAGAGCGCCATCCACCCCAGCATCCTGCAGAACGGGGGCCCCATGTT
CGCCTTCCGCCGCGTGGAGGAGGATCACAGCAACACCGAGCTGGGCATCGTGGAGTACCAGCACGCC
TTCAAGACCCCGGATGCAGATGCCGGTGAAGAAAGAGTTTAAACGGCCGGCCGCGGTCATAGCTGTTT
CCTGAACAGATCCCGGGTGGCATCCCTGTGACCCCTCCCCAGTGCCTCTCCTGGCCCTGGAAGTTGCCA
CTCCAGTGCCCACCAGCCTTGTCCTAATAAAATTAAGTTGCATCATTTTGTCTGACTAGGTGTCCTTCT
ATAATATTATGGGGTGGAGGGGGGTGGTATGGAGCAAGGGGCAAGTTGGGAAGACAACCTGTAGGGC

217

CTGCGGGGTCTATTGGGAACCAAGCTGGAGTGCAGTGGCACAATCTTGGCTCACTGCAATCTCCGCCT
CCTGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTTGTTGGGATTCCAGGCATGCATGACCAGGC
TCAGCTAATTTTTGTTTTTTTGGTAGAGACGGGGTTTCACCATATTGGCCAGGCTGGTCTCCAACTCCT
AATCTCAGGTGATCTACCCACCTTGGCCTCCCAAATTGCTGGGATTACAGGCGTGAACCACTGCTCCCT
TCCCTGTCCTTCTGATTTTAAAATAACTATACCAGCAGGAGGACGTCCAGACACAGCATAGGCTACCT
GGCCATGCCCAACCGGTGGGACATTTGAGTTGCTTGCTTGGCACTGTCCTCTCATGCGTTGGGTCCACT
CAGTAGATGCCTGTTGAATTGGGTACGCGGCCAGCTTGGCTGTGGAATGTGTGTCAGTTAGGGTGTGG
AAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGT
GTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACC
ATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCAT
GGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAG
TGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGA
TCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCC
GGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCG
CCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGA
ATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTG
CTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCT
GTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCT
TGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGG
AAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTC
GCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCC
GAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCCGACC
GCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGC
TTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGT
TCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATT
TCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGA
TCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAATG
GTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTG
GTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGT
AATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCG
GAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCA
CTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAG
AGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCT
GCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCA
GGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCG
TTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAA
CCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGAC
CCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACG
CTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCA
GCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCC
ACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTG
AAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGT
TACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTT
TTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACG
GGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGAT
CTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTG
GTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCAT
AGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTG
CAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAG
GGCCGAGCGCAGAAGTGGTCCTGCAACTTTA

pCMV6-AC-GFP-LPAR₃

CACCTTTTTCACATGCTGCACGCCACACTGCCTGCAGTTCAGGCCGTCGAGGAGCAGAACCACCAGGCC
CGGGGTCCAGCATACCACAAACGCCCCTAAGACAGTCATCACCGTCTTCATTAGCTTCATGGGTGTCCT
CCGGCGGCTGATGGACCCACTTGTATGCGGAGACAAGACGTTGGTTTTCCTCTTGACGTACACGTAGA
TCCGCAGGTACACCACAACCATGATGAGGAAGGCCATGAGGTTGGACACTGTCCAGAAAACAAGGTA
ACTCCTGCTGTAAATGGGGGCCAGGGAAGAGCAGGCAGAGATGTTGCAGAGGCAATTCCAGCCCAGT
GTGGGGACCGCCCCCATAAAAATGGCGATGGCCCAGACAAGCAAAATGAGCAGTGTCACCCTCTTTTT
GGTCAGGTTGCTATGGACCCGCATCCTCATGATTGACATGTGCCTCTCCACGGCGATAACCAGCAAGT
TGGTGAGGGAAGCAGTCAAGCTACTGTCCAGAAGCCCCTGACGGAGAAACCAGCGGTTGACAGTCAA
AGTTTTTGAAACTGGGCCTGTGTTAAACATCAGGAATACATAGGCAATTCCAGCGAAGAAATCGGCAG
CAGCTAAATTAGCCAACAGGTAGTAGAAGGGGAAATGAAATTTTCTGTTTTTGATCACTGCCGCGATG
ACCAGAGAATTAGAAAAAAAAAATAAACAGGCAGAAAAACGTCCCAACACACAAAACAATCACAAG
CTTTGTTCCTGTCCAGTCATCGACAGTATCAGTGTTGCTCCTATTATAAAAAAAGTCCATGTGCTTGTC
ATAGTGACACTCATTCATGGCGATCGCGGCGGCAGATCTCCTCGGTACCGGATCCAGTCGACGAATTC
CCGGCCGCCCTATAGTGAGTCGTATTACAAAATTCTGACGGTTCACTAAACGAGCTCTGCTTATATAGA
CCTCCCACCGTACACGCCTACCGCCCATTTGCGTCAACGGGGCGGGGTTATTACGACATTTTGGAAAG
TCCCGTTGATTTTGGTGCCAAAACAAACTCCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCCG
TGAGTCAAACCGCTATCCACGCCCATTGGTGTACTGCCAAAACCGCATCACCATGGTAATAGCGATGA
CTAATACGTAGATGTACTGCCAAGTAGGAAAGTCCCGTAAGGTCATGTACTGGGCATAATGCCAGGCG
GGCCATTTACCGTCATTGACGTCAATAGGGGGCGGACTTGGCATATGATACACTTGATGTACTGCCAA
GTGGGCAGTTTACCGTAAATACTCCACCCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGG
GAACATACGTCATTATTGACGTCAATGGGCGGGGGTCGTTGGGCGGTCAGCCAGGCGGGCCATTTACC
GTAAGTTATGTAACGCGGAACTCCATATATGGGCTATGAACTAATGACCCCGTAATTGATTACTATTA
ATAACTAGTCAATAATCAATGTCAACATGGCGGTCATATTGGACATGAGCCAATATAAATGTACATAT
TATGATATAGATACAACGTATGCAATGGCCAATAGCCAATATTGATTTATGCTATATAACCAATGACT
AATATGGCTAATTGCCAATATTGATTCAATGTATAGATCAGCTTGGCACTGGCCGTCGTTTTACAACGT
CGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGG
CGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGA
AATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAA
TAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCC
AGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTAT
CAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGC
ACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCG
AGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTG
CGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGG
AAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACA
ATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCG
CCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAA
AGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCC
TTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGG
TATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTG
GTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGC
TGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGC
TAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAAT
GAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAAC
TATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAA
GTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGT
GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTAT
CTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCA
CTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCAT
TTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAG
TTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTTCT
GCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAG

219

AGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAG
TGTAGCCGTAGTTAGGCCACCCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCC
TGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTA
CCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGA
CCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAA
GGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGG
AAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATG
CTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTT
GCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTT
TGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCG
GAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGA
CAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGG
CACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTC
ACACAGGAAACAGCTATGACCATGATTACGCCAAGCTCTAGCTAGAGGTCGACGGTATACAGACATG
ATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGA
AATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTGGGGTGGGCGAAG
AACTCCAGCATGAGATCCCCGCGCTGGAGGATCATCCAGCCGGCGTCCCGGAAAACGATTCCGAAGCC
CAACCTTTCATAGAAGGCGGCGGTGGAATCGAAATCTCGTGATGGCAGGTTGGGCGTCGCTTGGTCGG
TCATTTCGAACCCCAGAGTCCCGCTCAGAAGAACTCGTCAAGAAGGCGATAGAAGGCGATGCGCTGC
GAATCGGGAGCGGCGATACCGTAAAGCACGAGGAAGCGGTCAGCCCATTCGCCGCCAAGCTCTTCAG
CAATATCACGGGTAGCCAACGCTATGTCCTGATAGCGGTCGGCCACACCCAGCCGGCCACAGTCGATG
AATCCAGAAAAGCGGCCATTTTCCACCATGATATTCGGCAAGCAGGCATCGCCATGGGTCACGACGAG
ATCCTCGCCGTCGGGCATGCGCGCCTTGAGCCTGGCGAACAGTTCGGCTGGCGCGAGCCCCTGATGCT
CTTCGTCCAGATCATCCTGATCGACAAGACCGGCTTCCATCCGAGTACGTGCTCGCTCGATGCGATGTT
TCGCTTGGTGGTCGAATGGGCAGGTAGCCGGATCAAGCGTATGCAGCCGCCGCATTGCATCAGCCATG
ATGGATACTTTCTCGGCAGGAGCAAGGTGAGATGACAGGAGATCCTGCCCCGGCACTTCGCCCAATAG
CAGCCAGTCCCTTCCCGCTTCAGTGACAACGTCGAGCACAGCTGCGCAAGGAACGCCCGTCGTGGCCA
GCCACGATAGCCGCGCTGCCTCGTCCTGCAGTTCATTCAGGGCACCGGACAGGTCGGTCTTGACAAAA
AGAACCGGGCGCCCCTGCGCTGACAGCCGGAACACGGCGGCATCAGAGCAGCCGATTGTCTGTTGTGC
CCAGTCATAGCCGAATAGCCTCTCCACCCAAGCGGCCGGAGAACCTGCGTGCAATCCATCTTGTTCAA
TCATGCGAAACGATCCTCATCCTGTCTCTTGATCAGATCCGAAAATGGATATACAAGCTCCCGGGAGC
TTTTTGCAAAAGCCTAGGCCTCCAAAAAAGCCTCCTCACTACTTCTGGAATAGCTCAGAGGCCGAGGC
GGCCTCGGCCTCTGCATAAATAAAAAAAATTAGTCAGCCATGGGGCGGAGAATGGGCGGAACTGGGC
GGAGTTAGGGGCGGGATGGGCGGAGTTAGGGGCGGGACTATGGTTGCTGACTAATTGAGATGCATGC
TTTGCATACTTCTGCCTGCTGGGGAGCCTGGGGACTTTCCACACCTGGTTGCTGACTAATTGAGATGCA
TGCTTTGCATACTTCTGCCTGCTGGGGAGCCTGGGGACTTTCCACACCCTAACTGACACACATTCCACA
GCCAAGCTGGCCGCGTACCCAATTCAACAGGCATCTACTGAGTGGACCCAACGCATGAGAGGACAGT
GCCAAGCAAGCAACTCAAATGTCCCACCGGTTGGGCATGGCCAGGTAGCCTATGCTGTGTCTGGACGT
CCTCCTGCTGGTATAGTTATTTTAAAATCAGAAGGACAGGGAAGGGAGCAGTGGTTCACGCCTGTAAT
CCCAGCAATTTGGGAGGCCAAGGTGGGTAGATCACCTGAGATTAGGAGTTGGAGACCAGCCTGGCCA
ATATGGTGAAACCCCGTCTCTACCAAAAAAACAAAAATTAGCTGAGCCTGGTCATGCATGCCTGGAAT
CCCAACAACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCGGAGATTGCAGTGAGCC
AAGATTGTGCCACTGCACTCCAGCTTGGTTCCCAATAGACCCCGCAGGCCCTACAGGTTGTCTTCCCAA
CTTGCCCCTTGCTCCATACCACCCCCCTCCACCCCATAATATTATAGAAGGACACCTAGTCAGACAAAA
TGATGCAACTTAATTTTATTAGGACAAGGCTGGTGGGCACTGGAGTGGCAACTTCCAGGGCCAGGAGA
GGCACTGGGGAGGGGTCACAGGGATGCCACCCGGGATCTGTTCAGGAAACAGCTATGACCGCGGCCG
GCCGTTTAAACTCTTTCTTCACCGGCATCTGCATCCGGGGTCTTGAAGGCGTGCTGGTACTCCACGATG
CCCAGCTCGGTGTTGCTGTGATCCTCCTCCACGCGGCGGAAGGCGAACATGGGGCCCCGTTCTGCAG
GATGCTGGGGTGGATGGCGCTCTTGAAGTGCATGTGGCTGTCCACCACGGAGCTGTAGTAGCCGCCGT
CGCGCAGGCTGAAGGTGCGGGTGAAGCTGCCATCCAGATCGTTATCGCCCATGGGGTGCAGGTGCTCC
ACGGTGGCGTTGCTGCGGATGATCTTGTCGGTGAAGATCACGCTGTCCTCGGGGAAGCCGGTGCCCAT
CACCTTGAAGTCGCCGATCACGCGGCCGGCCTCGTAGCGGTAGCTGAAGCTCACGTGCAGCACGCCGC

CGTCCTCGTACTTCTCGATGCGGGTGTTGGTGTAGCCGCCGTTGTTGATGGCGTGCAGGAAGGGGTTCT
CGTAGCCGCTGGGGTAGGTGCCGAAGTGGTAGAAGCCGTAGCCCATCACGTGGCTCAGCAGGTAGGG
GCTGAAGGTCAGGGCGCCTTTGGTGCTCTTCATCTTGTTGGTCATGCGGCCCTGCTCGGGGGTGCCCTC
TCCGCCGCCCACCAGCTCGAACTCCACGCCGTTCAGGGTGCCGGTGATGCGGCACTCGATCTCCATGG
CGGGCAGGCCGCTCTCGTCGCTCTCCATCTCGAGCGGCCGCGTACGCGTGGAAGTGCTTTTATTGCAG
ACTGCACCTTGGCTAATACTATCCTCTATGTACTGGCTGCCTGTGTCACTCCTGCTGAGGACTGTGGAG
GGGATGCGAGAGGGACGCCTCTCTGGGTTCTCCTGAGAGAAGCAGCAGATCATCTTCTTCATGGTGCC
ATACATGTCCTCGTCCTTGTAGGAGTAGATGATGGGGTTCACGACGGAGTTGAGCAGCGCCAGCAGCA
GGAAC

Table 3.1.S1. List of the fingerprints used in this study for building ML classification models.

Name	Description	Bit	Form
Morgan	Representing fragments enumerating layered atom environments	1024	binary
Pubchem	Representing substructures defined by PubChem	881	binary
MACCS	Representing chemical patterns defined by MACCS keys	166	binary
RDkit	Representing subgraphs defined by RDKit	1024	binary
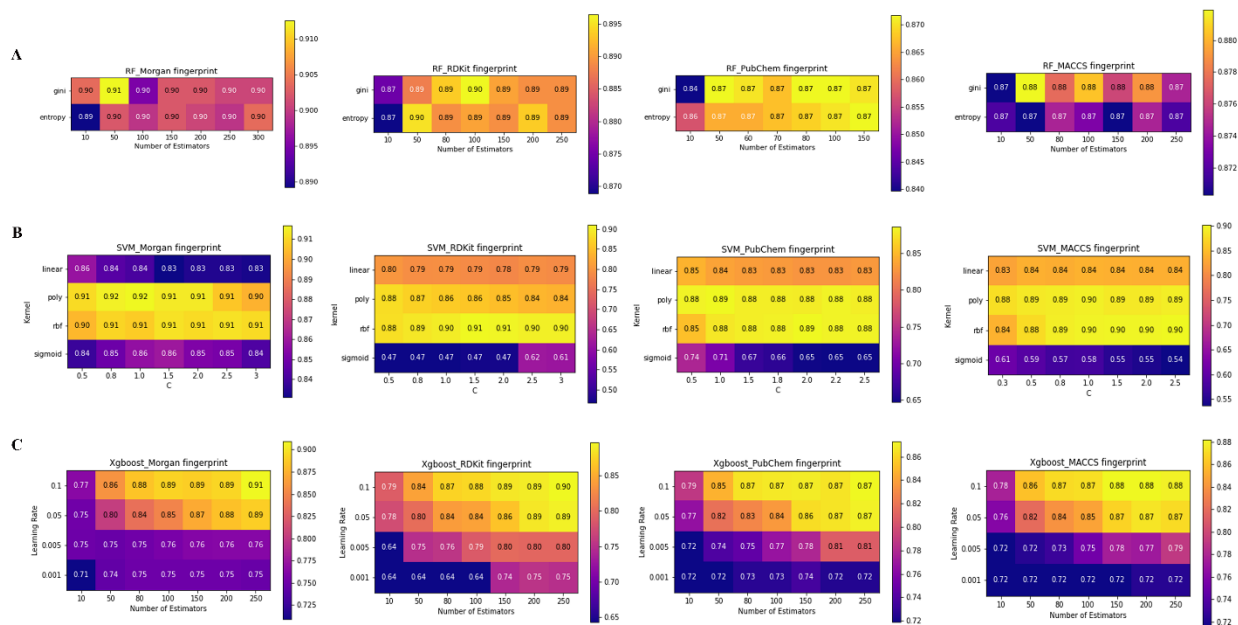
Figure 3.1.S1. Heatmaps demonstrate accuracy scores on the validation set with different hyperparameter combinations from strategy 1. A. RF models built with the four fingerprints. The heatmap illustrates the results from Grid search regarding number of estimators and criterion. B. SVM models built with the four fingerprints. The heatmap illustrates the results from Grid search regarding kernel and C. C. XGBoost models built with the four fingerprints. The heatmap illustrates the results from Grid search regarding number of estimators and learning rate.
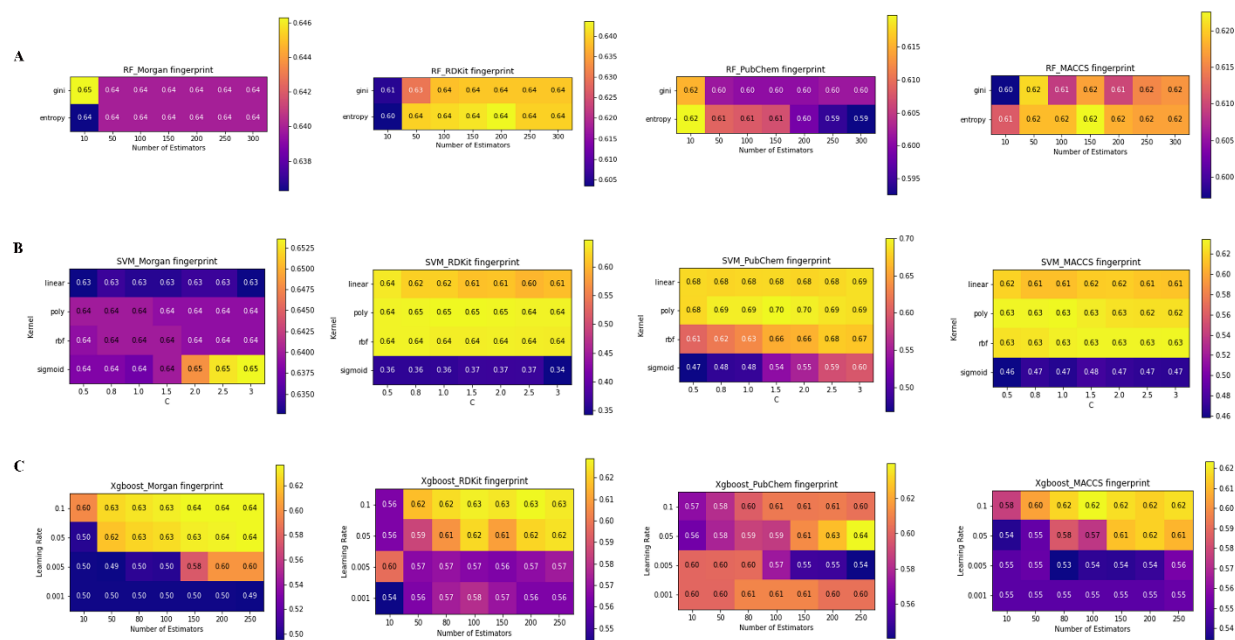
Figure 3.1.S2. Heatmaps demonstrate accuracy scores on the validation set with different hyperparameter combinations from strategy 2. A. RF models built with the four fingerprints. The heatmap illustrates the results from Grid search regarding number of estimators and criterion. B. SVM models built with the four fingerprints. The heatmap illustrates the results from Grid search regarding kernel and C. C. XGBoost models built with the four fingerprints. The heatmap illustrates the results from Grid search regarding number of estimators and learning rate.
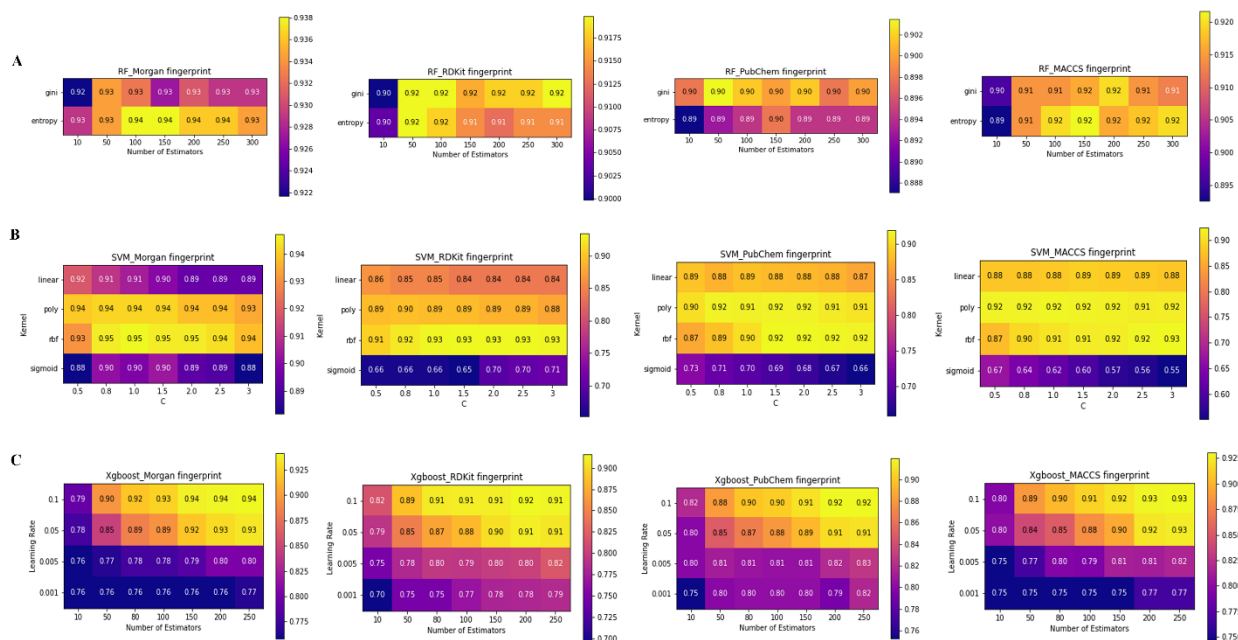
Figure 3.1.S3. Heatmaps demonstrate accuracy scores on the validation set with different hyperparameter combinations from strategy 3. A. RF models built with the four fingerprints. The heatmap illustrates the results from Grid search regarding number of estimators and criterion. B. SVM models built with the four fingerprints. The heatmap illustrates the results from Grid search regarding kernel and C. C. XGBoost models built with the four fingerprints. The heatmap illustrates the results from Grid search regarding number of estimators and learning rate.
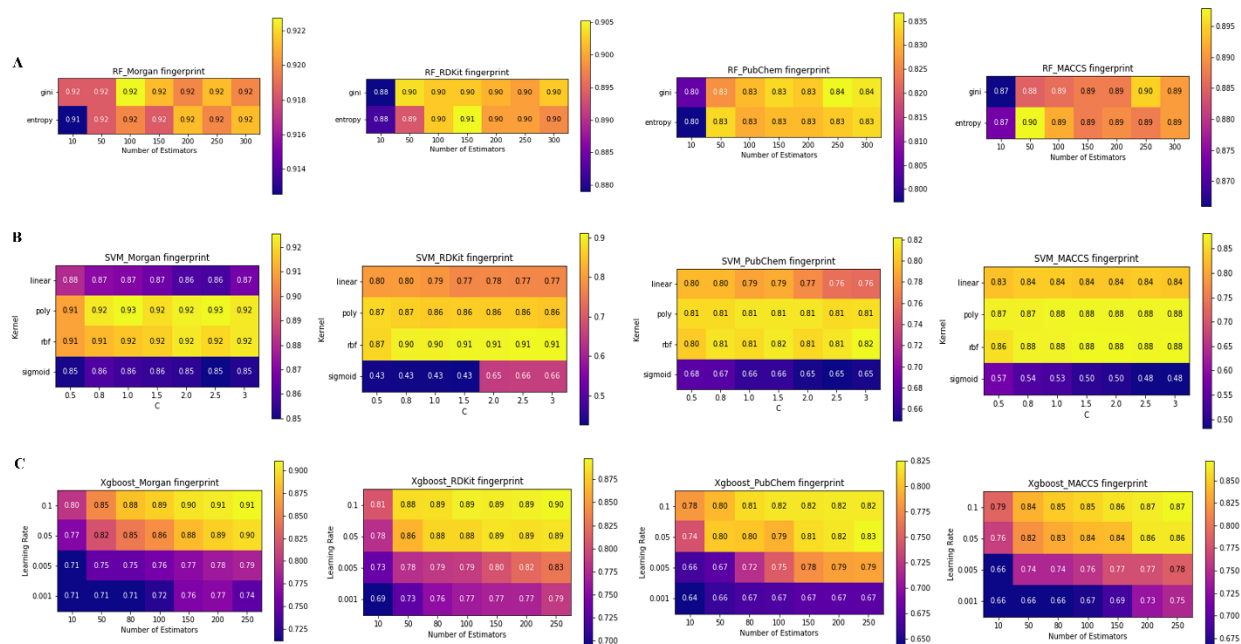
Figure 3.1.S4. Heatmaps demonstrate accuracy scores on the validation set with different hyperparameter combinations from strategy 4. A. RF models built with the four fingerprints. The heatmap illustrates the results from Grid search regarding number of estimators and criterion. B. SVM models built with the four fingerprints. The heatmap illustrates the results from Grid search regarding kernel and C. C. XGBoost models built with the four fingerprints. The heatmap illustrates the results from Grid search regarding number of estimators and learning rate.
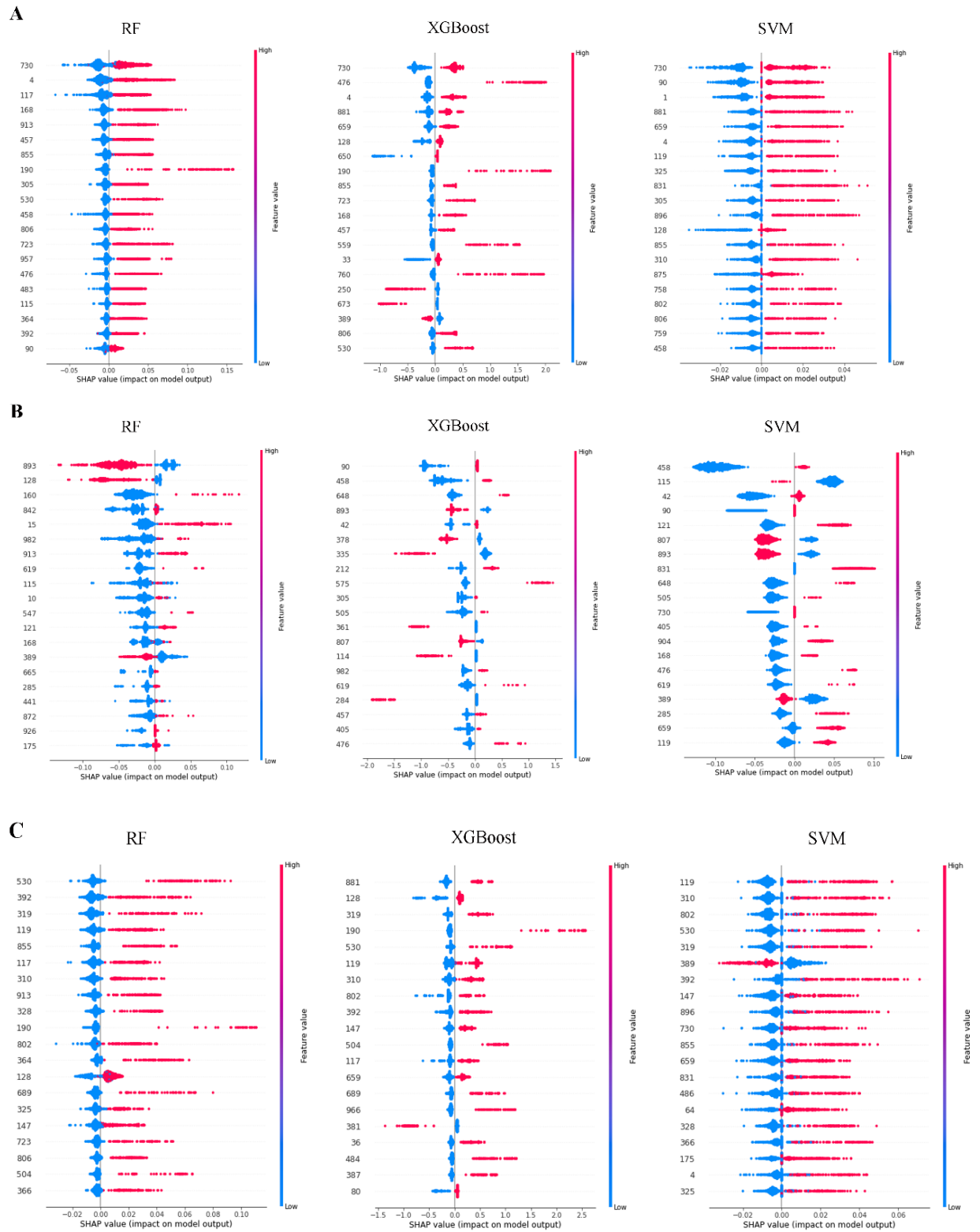
Figure 3.1.S5. The top 20 important features ranked by SHAP for the RF, XGBoost, and SVM models using strategy 1 (A), strategy 2 (B), and strategy 3 (C).
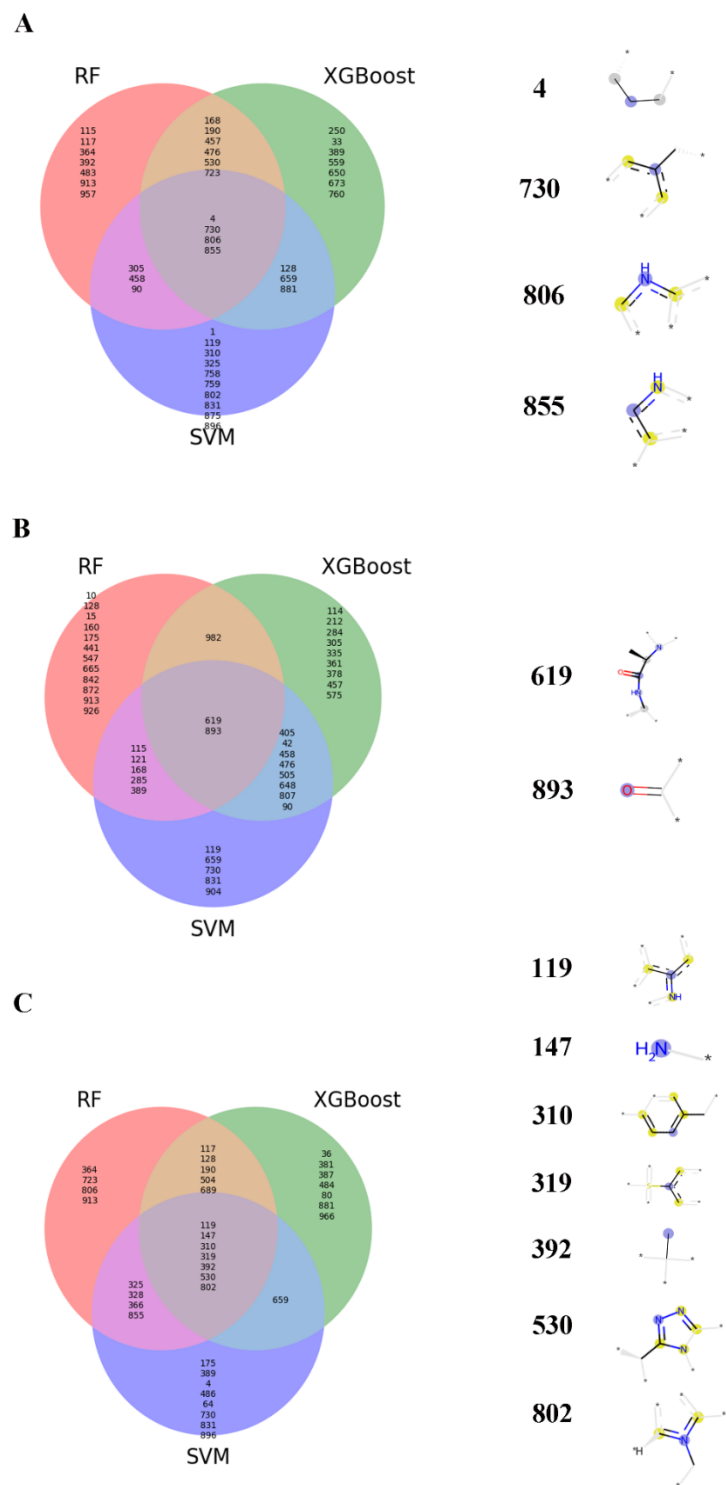
Figure 3.1.S6. Venn diagrams demonstrating the features of bit information that are commonly present among the RF, XGBoost, and SVM models using strategy 1 (A), strategy 2 (B), and strategy 3 (C) and their corresponding substructures.

Table 3.1.S2. The corresponding substructures of the top 20 important features ranked by SHAP for the RF, XGBoost, and SVM models using strategy 4.

| Bit	Structure	Bit	Structure	Bit	Structure
457		802		831	
730		730		90	
802		831		730	
4		212		305	
90		650		119	
913		457		457	
328		648		659	
659		119		758	
806		476		121	
310		695		325	

119		366		881	
212		341		168	
831		319		285	
476		659		310	
364		33		117	
320		389		913	
777		189		128	
104		881		855	
319		4		896	
291		751		878	

1. The central atom is highlighted in blue, and aromatic atoms are highlighted in yellow. Atoms/bonds in light gray indicate pieces of the structure that affect the connectivity invariants of the atoms, but they are not directly part of the fingerprint.

Table 3.1.S3. Detailed information about the external data and the prediction made by the best model from this study.

Ligand	Canonical smiles	IC50 (nM)	Strategy 2 prediction	Strategy 3 prediction	Strategy 4 prediction
19	Cc1ccccc1-c1nc2ccc(C(=O)c3ccc(F)cc3)cc2c(=O)n1C[C@H]1CCCNC1	4	non-binding	non-binding	non-binding
20	Cc1ccccc1-c1nc2ccc(C(=O)c3ccc(F)cc3)cc2c(=O)n1C[C@H]1CCCN(C(C)C)C1	3.8	non-binding	binding	binding
17	Cc1ccccc1-c1nc2ccc(Cc3ccc(F)cc3)cc2c(=O)n1C[C@H]1CCCN(C(C)C)C1	0.3	non-binding	binding	binding
10a	CC(C)N1CCC[C@H](Cn2c(-c3ccccc3F)nc3ccc(Oc4ccc(F)cn4)cc3c2=O)C1	70.5	non-binding	binding	binding
10b	CC(C)N1CCC[C@H](Cn2c(-c3ccccc3F)nc3ccc(Oc4ccc5ncsc5c4)cc3c2=O)C1	0.02	non-binding	binding	binding
10c	O=c1c2cc(Oc3ccc4ncsc4c3)ccc2nc(-c2ccccc2F)n1C[C@H]1CCCN(CCF)C1	8.1	non-binding	binding	binding
14a	O=c1c2cc(Oc3ccc(F)cn3)ccc2nc(-c2ccc(OCCO)cc2F)n1C[C@H]1CCCNC1	353.1	non-binding	non-binding	non-binding
14b	O=c1c2cc(Oc3ccc(F)cc3)ccc2nc(-c2ccc(OCCO)cc2F)n1C[C@H]1CCCNC1	19.8	non-binding	non-binding	non-binding
15a	CC(C)N1CCC[C@H](Cn2c(-c3ccc(OCCO)cc3F)nc3ccc(Oc4ccc(F)cn4)cc3c2=O)C1	40.1	non-binding	binding	binding
15b	CC(C)N1CCC[C@H](Cn2c(-c3ccc(OCCO)cc3F)nc3ccc(Oc4ccc(F)cc4)cc3c2=O)C1	2.3	non-binding	binding	binding
17a	CC(C)N1CCC[C@H](Cn2c(-c3ccc(OCCF)cc3F)nc3ccc(Oc4ccc(F)cn4)cc3c2=O)C1	17.6	non-binding	binding	binding
17b	CC(C)N1CCC[C@H](Cn2c(-c3ccc(OCCF)cc3F)nc3ccc(Oc4ccc(F)cc4)cc3c2=O)C1	0.33	non-binding	binding	binding
4a	Cc1ccccc1-c1nc2ccc(Oc3ccc(F)cc3)cc2c(=O)n1C[C@H]1CCCNC1	5.6	non-binding	non-binding	binding

4b	Cc1ccccc1-c1nc2ccc(Oc3ccc(F)cn3)cc2c(=O)n1C[C@H]1CCCNC1	52	non-binding	non-binding	binding
4c	Cc1ccccc1-c1nc2ccc(Oc3ccc(F)nc3)cc2c(=O)n1C[C@H]1CCCNC1	68.5	non-binding	binding	binding
4d	Cc1ccccc1-c1nc2ccc(Oc3ccc(Oc4ccc(F)nc4)nc3)cc2c(=O)n1C[C@H]1CCCNC1	82.9	non-binding	binding	binding
4e	Cc1ccccc1-c1nc2ccc(Oc3ccc4ncsc4c3)cc2c(=O)n1C[C@H]1CCCNC1	5.3	non-binding	binding	binding
4f	CC(=O)N1CCN(c2ccc(Oc3ccc4nc(-c5ccccc5C)n(C[C@H]5CCCNC5)c(=O)c4c3)cc2)CC1	587	non-binding	binding	binding
5a	Cc1ccccc1-c1nc2ccc(Oc3ccc(F)cc3)cc2c(=O)n1C[C@H]1CCCN(CCF)C1	29.5	non-binding	binding	binding
5b	Cc1ccccc1-c1nc2ccc(Oc3ccc(F)cn3)cc2c(=O)n1C[C@H]1CCCN(CCF)C1	20.6	non-binding	binding	binding
5c	Cc1ccccc1-c1nc2ccc(Oc3ccc(F)nc3)cc2c(=O)n1C[C@H]1CCCN(CCF)C1	64.2	non-binding	binding	binding
5d	Cc1ccccc1-c1nc2ccc(Oc3ccc(Oc4ccc(F)nc4)nc3)cc2c(=O)n1C[C@H]1CCCN(CCO)C1	65.2	non-binding	binding	binding
5e	Cc1ccccc1-c1nc2ccc(Oc3ccc4ncsc4c3)cc2c(=O)n1C[C@H]1CCCN(CCF)C1	9.3	non-binding	binding	binding
5f	Cc1ccccc1-c1nc2ccc(Oc3ccc(F)cc3)cc2c(=O)n1C[C@H]1CCCN(C(C)C)C1	3.7	non-binding	binding	binding
5g	Cc1ccccc1-c1nc2ccc(Oc3ccc4ncsc4c3)cc2c(=O)n1C[C@H]1CCCN(C(C)C)C1	1.8	non-binding	binding	binding
5h	Cc1ccccc1-c1nc2ccc(Oc3ccc(F)cn3)cc2c(=O)n1C[C@H]1CCCN(CCO)C1	18.2	non-binding	binding	binding
5i	Cc1ccccc1-c1nc2ccc(Oc3ccc4ncsc4c3)cc2c(=O)n1C[C@H]1CCCN(CCO)C1	0.4	non-binding	binding	binding

5j	Cc1ccccc1-c1nc2ccc(Oc3ccc(F)cn3)cc2c(=O)n1C[C@H]1CCCN(C(=O)C(C)F)C1	1000	non-binding	binding	binding
9a	O=c1c2cc(Oc3ccc(F)cn3)ccc2nc(-c2ccccc2F)n1C[C@H]1CCCNC1	935.2	non-binding	non-binding	binding
9b	O=c1c2cc(Oc3ccc4ncsc4c3)ccc2nc(-c2ccccc2F)n1C[C@H]1CCCNC1	16	non-binding	non-binding	non-binding
1r	CCN1C(C(=O)N[C@@H](Cc2ccccc2)C(=O)C(=O)NCCCN2CCOCC2)Cc2cc3c(cc2S1(=O)=O)OCCO3	non-binding	non-binding	non-binding	non-binding
2r	C[C@@H]1CC(C)(C)O[C@@]2(CCC3=Cc4c(cnn4-c4ccc(F)cc4)C[C@@]32C)O1	non-binding	non-binding	non-binding	non-binding
3r	O=C(Cc1ccccc1)Nc1ccc(-c2ccc3ncc4ccc(=O)n(-c5cccc(C(F)(F)F)c5)c4c3c2)cn1	non-binding	non-binding	non-binding	non-binding
4r	CC(C)[C@H](NC(=O)OCc1cncs1)C(=O)N[C@@H](Cc1ccccc1)[C@H](O)CN1CCN(Cc2ccccc2)C[C@H]1C(=O)NC(C)(C)C	non-binding	non-binding	non-binding	non-binding
5r	Cc1ncc([N+](=O)[O-])n1CCOC(c1ccccc1)c1ccc(Cl)s1	non-binding	non-binding	non-binding	non-binding
6r	COc1ccc(C2CNC(=O)N2C)cc1OC1CC2C3CCC(C3)C2C1	non-binding	non-binding	non-binding	non-binding
7r	COc1ccc(Cl)cc1N=C(S)NNC(=O)CC(C)C	non-binding	non-binding	non-binding	non-binding
8r	Cc1cccc(Nc2cc(C(F)(F)F)nc(SCC(=O)O)n2)c1C	non-binding	non-binding	non-binding	non-binding
9r	COc1ccc2c(c1)CN(C(=O)N(C)C)N=C2Cc1c(Cl)cncc1Cl	non-binding	non-binding	non-binding	non-binding
10r	CCO/C(O)=C1\C(COCCN2CCOCC2)=NC(C)=C(C(=O)OC)C1c1ccccc1Cl	non-binding	non-binding	non-binding	non-binding
11r	Cc1cc(-c2nn(-c3ccccc3F)cc2CN2CCCC(CO)C2)c(C)s1	non-binding	non-binding	non-binding	non-binding
12r	Nc1nc(O)nc2ccc(Sc3ccc4ccccc4c3)cc12	non-binding	non-binding	non-binding	non-binding
13r	O=C(O)Cn1c(=O)c(=O)[nH]c2cc([N+](=O)[O-])c(-n3ccc(CC(=O)N4CCCCC4)c3)cc21	non-binding	non-binding	non-binding	non-binding
14r	NC(=O)c1cccc(N2C(=O)N(c3cccc(C(N)=O)c3)[C@@H](Cc3ccccc3)[C@@H](O)[C@@H]2CCc2ccccc2)c1	non-binding	non-binding	non-binding	non-binding
15r	O=C1OCC2=C1C(c1ccc3c(c1)OC(F)(F)O3)c1cc3c(cc1N2)OCO3	non-binding	non-binding	non-binding	non-binding

|					
16r	Clc1ccc(Nc2nnc(Cc3ccncc3)c3ccncc23)cc1	non-binding	non-binding	non-binding	non-binding
17r	N=C(N)c1ccc2[nH]c(-c3ccc(OCc4cn(-c5ccc(I)cc5)nn4)cc3)nc2c1	non-binding	non-binding	non-binding	non-binding
18r	O=C(Nc1ccc(F)cc1-c1ccc(F)c(Cl)c1)OCC1CCNCC1	non-binding	non-binding	non-binding	non-binding
19r	N/C(=N/O)c1cncc(N2CC[C@@H]3CN[C@@H]3C2)c1	non-binding	non-binding	non-binding	non-binding
20r	c1ccc(B2OCc3ccccc32)cc1	non-binding	non-binding	non-binding	non-binding
21r	Cc1nnc2n1-c1ccc(NC(=O)C3CCN(Cc4ccccc4F)CC3)cc1CC2	non-binding	non-binding	non-binding	non-binding
22r	Cc1cc(C)c2sc(NC(=O)c3ccc(C#N)cc3)nc2c1	non-binding	non-binding	non-binding	non-binding
23r	Cc1noc(C)c1COc1ccc(C(=O)OCC(=O)Nc2ccc3c(c2)OCO3)cc1	non-binding	non-binding	non-binding	non-binding
24r	O=S1(=O)NCc2ccccc2N1C1CCN(C2CCc3ccccc3C2)CC1	non-binding	non-binding	non-binding	non-binding
25r	Cc1nc2cnc(Oc3ccccc3)nc2n(CCc2ccccc2)c1=O	non-binding	non-binding	non-binding	non-binding
26r	O=C(/C=C/c1ccccc1)NC(NC(=S)Nc1cccc2cccnc12)C(Cl)(Cl)Cl	non-binding	non-binding	non-binding	non-binding
27r	CC(C)CN(Cc1ccccn1)C(=O)Cc1c(-c2ccccc2)[nH]c2ccccc12	non-binding	non-binding	non-binding	non-binding
28r	CCCCc1ccc2[nH]c(-c3ccccc3)c(C3=C(Br)C(=O)NC3=O)c2c1	non-binding	non-binding	non-binding	non-binding
29r	C[C@]1(c2ccccc2Cl)N[C@H](C(=O)O)Cc2c1[nH]c1ccccc21	non-binding	non-binding	non-binding	non-binding
30r	CC1(C)C[C@H](O)C[C@](C)(C/N=C(\S)c2cccc(Cl)c2)C1	non-binding	non-binding	non-binding	non-binding
31r	O=C(NC1CCN(C2CCCCC2)CC1)c1cc(-c2cccnc2)on1	non-binding	non-binding	non-binding	non-binding
32r	CC(C)CCOc1ccccc1-c1ccccc1OCc1cn([C@H](CCCNC(=N)N)C(=O)NCC2CCCCC2)nn1	non-binding	non-binding	non-binding	non-binding

CB1
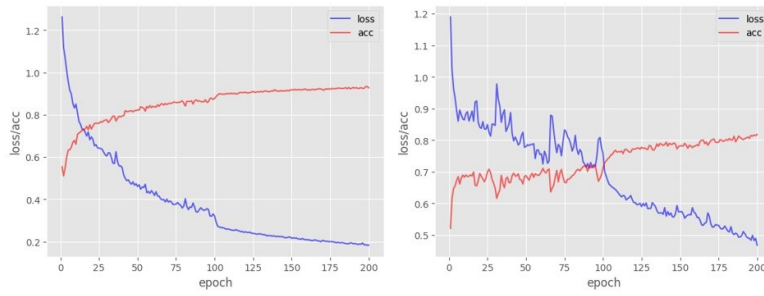
CXCR4
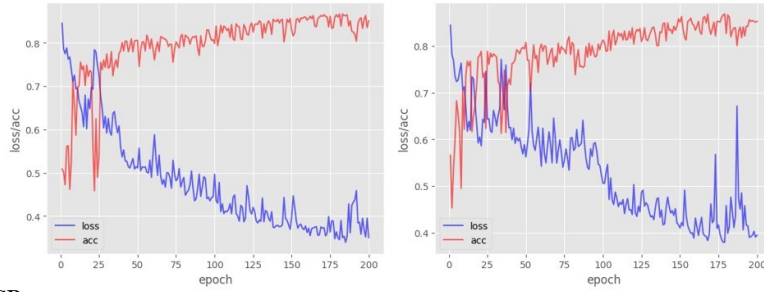
GR
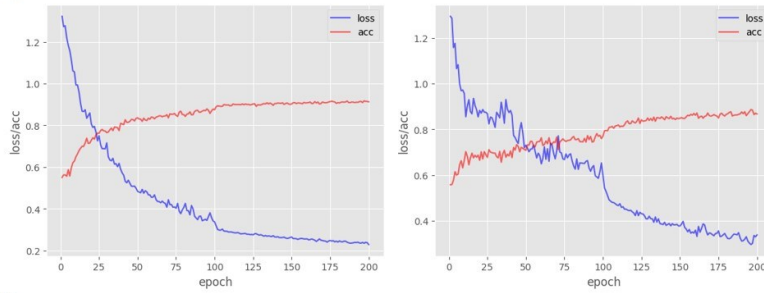
H3
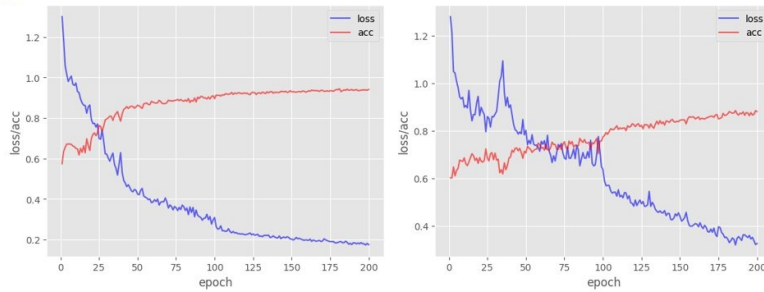
LPA1
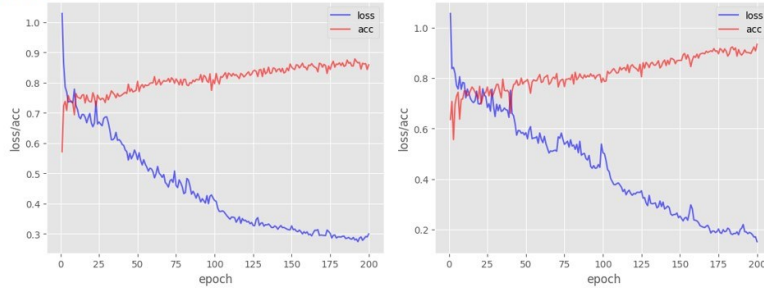
Figure 3.2.S1. Training of LAGCN (left) and LAGAT (right) models. The curves show the training accuracy and loss across each epoch of training on the five datasets.
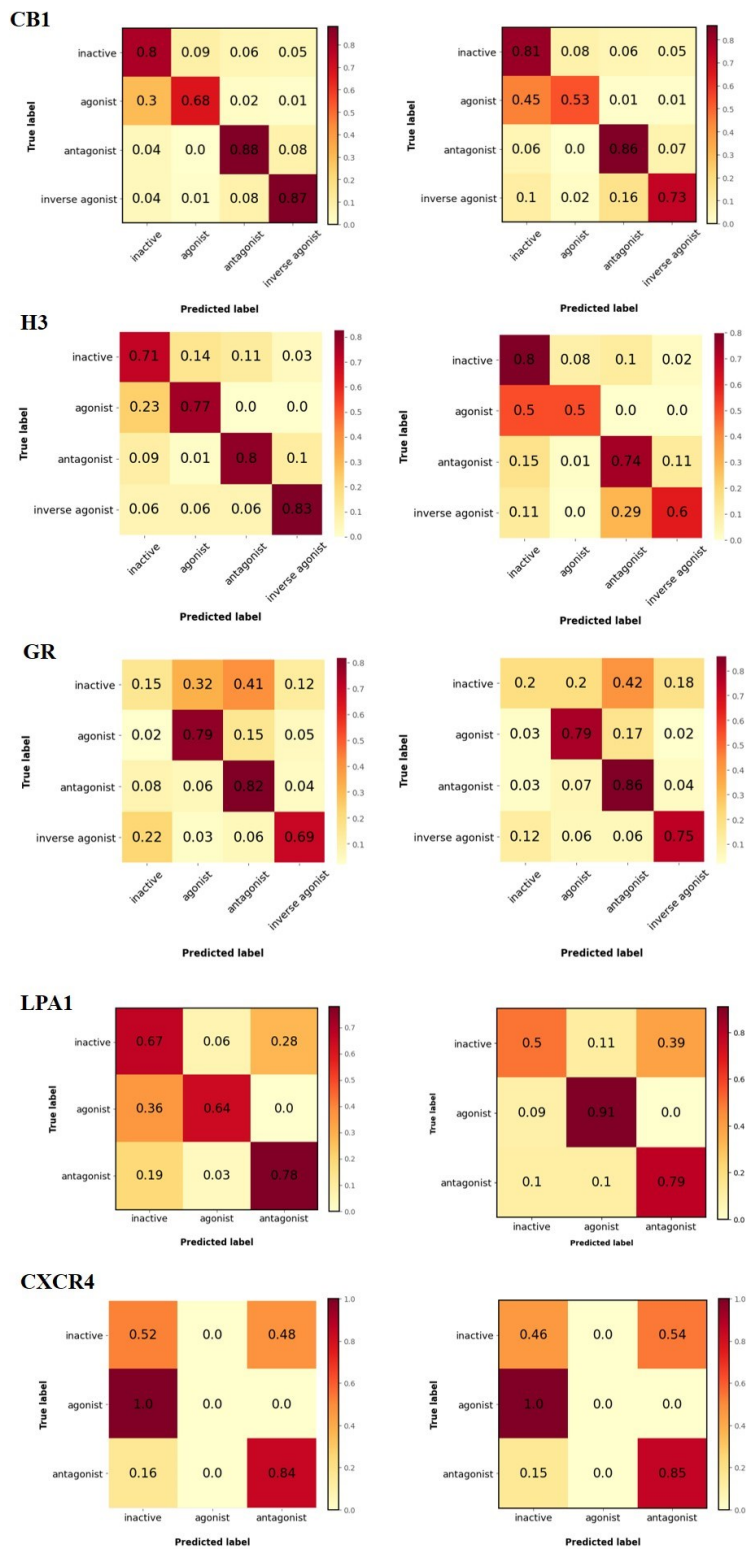
Figure 3.2.S2. Confusion matrix for the base GCN (left) and GAT (right) models.

List of publications during PhD

First-author publications:

1.   **Liu, W.**; Hopkins, A. M.; Yan, P.; Du, S.; Luyt, L. G.; Li, Y.; Hou, J., Can machine learning 'transform' peptides/peptidomimetics into small molecules? A case study with ghrelin receptor ligands. *Mol Divers* 2022.

2.   **Liu, W.**; Hopkins, A. M.; Hou, J., The development of modulators for lysophosphatidic acid receptors: A comprehensive review. *Bioorg Chem* 2021, *117*, 105386.

3.   **Liu, W.**; Yan, P.; Du, S.; Hou, J., Local Augmented Graph Neural Networks for the Classification of GPCR Modulators (Pending)

4.   **Liu, W.**; Lees, S.; Ramachandran, R.; Campbell, M.; Thu, Kelsie.; Hopkins, A. M.; Hou, J., Design, synthesis and biological evaluation of fluorine-containing triazole derivatives targeting lysophosphatidic acid receptor 1 for the treatment of breast cancer (Pending)

5.   **Liu, W.**; Lees, S.; Ramachandran, R.; Campbell, M.; Hopkins, A. M.; Hou, J., Design, synthesis and biological evaluation of urea triazole derivatives targeting lysophosphatidic acid receptor 1 for the treatment of breast cancer (Pending)

Publications with collaborators:

1.   Zhang, Y.-H.; Huang, X.-H.; Wong, W.-L.; Luo, J.-R.; Guo, X.-C.; **Liu, W**.; Hou, J.; She, M.-T.; Jiang, W.-H.; Sun, N.; Lu, Y.-J., A red fluorescent small-molecule for visualization of higher-order cyclic dimeric guanosine monophosphate (c-di-GMP) structure in live bacterial cells and real-time monitoring of biofilm formation on biotic and abiotic surfaces. *Sensors and Actuators B: Chemical* 2023, *376*, 132992.

2.   Zheng, B.-X.; She, M.-T.; Long, W.; Xu, Y.-Y.; Zhang, Y.-H.; Huang, X.-H.; **Liu, W.**; Hou, J.-Q.; Wong, W.-L.; Lu, Y.-J., A small-sized benzothiazole–indolium fluorescent probe: the study of interaction specificity targeting c-MYC promoter G-quadruplex structures and live cell imaging. *Chem Commun* 2020, *56* (95), 15016-15019.

3.   Xu, Y.; Huang, Z.; Wu, G.; Jin, F.; Lin, S.; Zhang, C.; Zheng, J.; **Liu, W.**; Hou, J.; Lu, Y.-J., Bioinformatic assay reveal the potential mechanism of Guizhi-Shaoyao-Zhimu decoction

against rheumatoid arthritis and mild-to-moderate COVID-19. *Comput Methods Programs Biomed* 2023, *238*, 107584.

4.   Zhu, Y.; Hou, J.; Huang, X.-H.; Zhong, D.-X.; Long, W.; **Liu, W.**; Lu, Y.-J.; Zhang, K.; Wong, W.-L., Structural modification of nonspecific thiazole orange for ligand-DNA interaction study: Understanding the ligand recognition selectivity towards G4-DNA over duplex-DNA. *Journal of Luminescence* 2020, *226*, 117488.

Patent:

Hou J, **Liu WJ**. Lysophosphatidic Acid Receptor Antagonists for Breast Cancer Treatment. United States Provisional Application No: 63/488,599. 2023

Conferences:

1. **Wenjie Liu**, Jinqiang Hou, (2021, August 13-20). Development of PET Molecular Probes for Lysophosphatidic Acid Receptor Type 1. Poster presented at Canadian Chemistry Conference and Exhibition (CSC 2021), Virtual

2. **Wenjie Liu**, Peizhi Yan, Shan Du, Jinqiang Hou, (2023, June 4-8). Local Augmented Graph Neural Networks for the Classification of GPCR Modulators. Poster presented at Canadian Chemistry Conference and Exhibition (CSC 2023), Vancouver, Canada

3. **Wenjie Liu**, Simon Lees, Rithwik Ramachandran, Jinqiang Hou, (2023, June 4-8). Design, synthesis and biological evaluation of fluorine-containing triazole derivatives targeting lysophosphatidic acid receptor 1 for the treatment of breast cancer. Poster presented at Canadian Chemistry Conference and Exhibition (CSC 2023), Vancouver, Canada