Computationally aided rational design, synthesis and evaluation of PFKFB3 ligands for atherosclerotic plaque stabilisation

Thesis presented for the degree of

Doctor of Philosophy at

THE UNIVERSITY OF ABERDEEN

and

UNIVERSITA' DEGLI STUDI DI MILANO
FACOLTA' DI SCIENZE DEL FARMACO
Department of Pharmaceutical Sciences
PhD Course in Pharmaceutical Sciences (XXXI Cycle)

Virginia Cristofori (MSc, University of Pisa)

Academic years 2016/2019

Supervisors: Prof. Matteo Zanda
Dr Sergio Dall'Angelo
Prof. Alessandro Contini

This PhD programme was founded in the frame of

HORIZON 2020 - Marie Skłodowska-Curie

ITN-European Joint Doctorate

MOGLYNET-programme

“MOGLYNet: “*Mo*dulation of *gly*colytic flux as a *ne*w approach for *t*reatment of atherosclerosis and plaque stabilization: a multidisciplinary study”

The project leading to this application has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 67552

Declaration

I hereby declare that this thesis has been composed by myself and it has not been accepted in any previous application for a degree at these, or any other University. The work described herein has been performed by myself, except where expressly indicated otherwise. All sources of information have been specifically acknowledged by means of reference, and verbatim extracts have been distinguished by quotation marks.

Virginia Cristofori, February 2019

Virginia Cristofori

Acknowledgements

Firstly, I am thankful to Professor Matteo Zanda for giving me this unique and personal development opportunity, and for his supervision and guidance throughout my Ph.D study and research. I thank Dr Sergio Dall'Angelo, who became my acting supervisor at my Home Institution, the University of Aberdeen, for his help since the beginning of the project.

I also would like to express my sincere gratitude to my Host Institution supervisor, Professor Alessandro Contini for his continuous support, guidance and patience, not only during my Academic Secondment at the University of Milan, but for all the rest of the project.

I thank my fellow labmates, flatmates and friends Ilaria Patruno, Carlo de Dominicis, Beatrice Longo, Federica Rizzi and Alessandro Colombano for sharing with me the ups and downs of this incredible journey. My advisors and mentors Dr Monica Piras, Dr Chih-Chung 'Jimy' Tseng and Dr Chiara Zanato for sharing their knowledge, giving me experimental tips every day in the lab and helping me over my trying and disappointing times with setbacks and doubts. I thank Dr Francesca Stanzione for her contribution regarding the modelling part that was performed in Aberdeen.

Moreover, this work would have not been possible without the financial support of Horizon 2020 MSCA European Joint Doctorate programme, the Moglynet project. Hence, a distinct acknowledgment goes to the project coordinator Prof. Maria Luisa Gelmi as well as all the Professors of the Moglynet Consortium and my colleagues, Early Stage Researchers for their encouragement, insightful comments during our meetings. Very important to me was the collaboration with my colleagues and friends, Helena Macut and Xiao Hu whose work was indispensable for my research.

My sincere thanks also go to Dr Chiara Piacenza, of Bayer HealthCare Manufacturing for offering me the industrial internship opportunity and for providing me with trainings and courses.

Finally, my deepest gratitude goes to all my family for always being there for me and for their immense love, support, care and affection.

CONTENTS

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 8

List of Figures

List of Tables

List of Schemes

Abstract

Several types of cells actively involved in atherosclerosis undergo a metabolic reprogramming that comprises an accelerated glycolytic flux to the detriment of mitochondrial oxidative phosphorylation. This phenomenon, known as the Warburg effect, allows cells to produce ample energy and biomass with negligible oxygen consumption, resulting in uncontrolled high-speed proliferation. This phenotypic dysfunction underlies both inflammation and angiogenesis, two processes that are foundation of the pathological behaviour of the plaque and promote its destabilisation. The preponderant kinase domain of inducible PFKFB3 enzyme catalyses the synthesis of F2,6P, which simultaneously promotes glycolysis and inhibits gluconeogenesis. PFKFB3 has been found to be one of the most abundant and overly expressed isoenzymes in the Warburg effect, suggesting its key role in the pathology. From a clinical perspective, PFKFB3 represents an emerging biological target, and its inhibition would be an innovative therapeutic strategy for the treatment of atherosclerotic lesions.

Our study revolves around the identification of new PFKFB3 kinase inhibitors. The rational design was carried out with the aid of computational methods, based on the 3D-structures of both PFKFB3 protein and known active ligands. In particular, two classes of inhibitors were used as reference ligands for the development of two distinct virtual screening workflows that led to the identification of two generations of candidate inhibitors. The first hierarchical workflow consisted of a pharmacophore-based library filtration, followed by molecular docking and molecular dynamic simulations, whereas the second consisted of a ligand-optimisation process based on the design of a focussed synthesisable library of analogues that were submitted to molecular docking along with a concomitant similarity search for scaffold hopping. The first- and the second-generation candidate compounds, which were predicted to have good *in silico* affinity towards PFKFB3, were synthesised and biologically evaluated through kinase assay, providing further insights about these selected drug-like molecules and new structure-activity relationship information.

Abbreviations

1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional
3PO	3-(3-pyridinyl)-1-(4-pyridinyl)-2-propen-1-one
Å	Angstrom
ACE	Acetyl group
ACO	Ant colony optimisation
ACN	Acetonitrile
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
Ar	Aryl
Asn	Asparagine
AUC	Area under the curve
AZI	AstraZeneca phenoxy-indole/indazole inhibitors
AZPP	AstraZeneca dihydropyrrolopyrimidone inhibitors
Boc	*Tert*-butoxycarbonyl group
Bu	Butyl
bs	Broad signal
cat.	Catalytic amount
*c*logP	Calculated partition coefficient
conc.	Concentrated
COSY	Correlation spectroscopy
crys-	Crystallised binding mode
Da	Daltons
DAD	Diode-array detector

DB	Database
DCM	Dichloromethane
d	doublet
DAD	Diode-array detector
dd	doublet of doublets
ddd	doublet of doublet of doublet
DIPEA	*N,N'*-Diisopropylethylamine
DMA	Dimethylacetamide
DMAP	4-Dimethylaminopyridine
DMF	*N,N'*-Dimethylformamide
DMFDMA	*N,N'*-Dimethylformamide dimethyl acetal
DMSO	Dimethyl sulfoxide
DMSO-d_6	Deuterated dimethyl sulfoxide
dock-	Docked binding mode
DUD•E	Data of Useful Decoys: Enhanced
EC	Endothelial cell
EDC	1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride
EF	Enrichment Factor
ESI	Electrospray ionisation
ESI-MS	Electrospray ionisation mass spectrometer
et al.	And others
Et	Ethyl
Et_2O	Diethyl ether
EtOAc	Ethyl acetate
EtOH	Ethanol
F6P	Fructose-6-phosphate
F1,6P	Fructose-1,6-bisphosphate
F2,6P	Fructose-2,6-bisphosphate

FBPase	Fructose-1,6-bisphosphatase
g	gram(s)
Gly	Glycine
GPU	Graphical processing unit
h	Hour(s)
HATU	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium-3-oxid hexafluorophosphate
H-bond	Hydrogen bond
HBA	Hydrogen bond acceptor
HBD	Hydrogen bond donor
Hex	Hexane
HMBC	Heteronuclear multiple-bond correlation spectroscopy
HOBt	1-Hydroxybenzotriazole
HPC	High Performance Computing
HPLC	High performance liquid chromatography
HPLC-MS	High performance liquid chromatography-mass spectrometry
HSQC	Heteronuclear single-quantum correlation spectroscopy
HTS	High-throughput screening
Hz	Hertz(s)
IC_{50}	Half maximal inhibitory concentration
Ile	Isoleucine
J	Coupling constant
IUPAC	International Union of Pure and Applied Chemistry
Kase	Kinase
L × ID	Length × internal diameter
Leu	Leucine
LDL	Low-density lipoprotein
logP	Partition coefficient

m	multiplet
m/z	mass-to-charge ratio
Me	Methyl
MeOD	Deuterated methanol
MeOH	Methanol
mg	Milligram(s)
MHz	Megahertz(s)
min	Minute(s)
mL	Millilitre(s)
MM	Molecular Mechanics
MM-GBSA	Molecular Mechanics – Generalised Born and surface area continuum-solvation
MM-PBSA	Molecular Mechanics – Poisson-Boltzmann and surface area continuum-solvation
MMP	Matrix metalloproteinase
mM	millimolar
mmol	Millimole(s)
Mφs	Macrophages
MOE	Molecular Operating Environment
MS	Mass spectrometry
MW	Molecular weight
mw	microwave irradiation
n-Hex	Normal hexane
nM	nanomolar
NME	*N*-methyl amide group
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser effect spectroscopy
NPT	Isothermal-isobaric ensemble: constant number of particles, pressure and temperature

NVT	Canonical ensemble: constant number of particles, volume and temperature
Nwat	Number of water molecules
Pase	Phosphatase
PBS	Portable Batch System
PDB	Protein Data Bank
PDB ID	Protein Data Bank identifier
PFK15	1-(4-Pyridinyl)-3-(2-quinolinyl)-2-propen-1-one
PFK158	1-(4-pyridinyl)-3-[7-(trifluoromethyl)-2E-quinolinyl]-2-propen-1-one
Ph	Phenyl
Pi	Phosphate ion
PKA	Protein Kinase A
PLANTS	Protein-Ligand ANT System
PFKFB	6-Phosphofructo-2-kinase/Fructose-2,6-bisphosphatase
PFK-1	6-Phosphofructo-1-kinase
ppm	parts per million
Pr	Propyl
PSA	Polar Surface Area
Py	Pyridine
q	quartet
RA%	Residual activity
r^2	correlation coefficient
R_f	Retardation factor
r.l.s.	Rate limiting step
RLU	Relative Light Units
RMSD	Root mean square deviation
RPBS	Ressource Parisienne en Bioinformatique Structurale
RP-HPLC	Reverse phase high performance liquid chromatography

Ro5	Rule of five (of Lipinski)
ROC	Receiver Operating Characteristic
ROS	Reactive oxygen species
r.t.	Room temperature
s	singlet
SA	Surface area accessible to the solvent
SAR	Structure-activity relationship
Ser	Serine
SMILES	Simplified molecular-input line entry-system
SMC	Smooth muscle cell
S_NAr	Nucleophilic Aromatic Substitution
SPORES	Structure Protonation and Recognition System
t	triplet
TCA	Tricarboxylic acid
TEA	Triethylamine
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
Thr	Threonine
TLC	Thin layer chromatography
TM	Target molecule
TPSA	Topological polar surface areas
t_R	Retention time
UV	Ultraviolet
v/v	Volume of solute to volume of solution ratio
Val	Valine
vs.	*versus*
VS	Virtual screening
VSMC	Vascular smooth muscle cell

w/v	Weight of solute to volume of solution ratio
ΔG	Gibbs free energy variation
ΔS	Entropy variation
δ	Chemical shift
ε	Dielectric constant
μM	Micromolar

1

Introduction

1 Introduction

1.1 Biological background

1.1.1 Atherosclerosis

Atherosclerosis is a chronic life-threatening inflammatory condition characterised by the build-up of a waxy substance, called plaque, inside the arterial walls. Atherosclerotic plaque is predominantly made of lipids and, over time, can grow, leading to peripheral artery diseases and serious cardiovascular complications. The swelling of plaque causes the arteries to harden and narrow, which disrupts and limits the flow of oxygen-rich blood to the heart and to the rest of the body. Eventually, an area of the plaque may break open after intra-plaque hemorrhages, leading to a clotting cascade and occlusive thrombosis with, at times, fatal complications. In fact, thrombosis-prone areas of the plaque often arise in unpredictable myocardial infarction and sudden ischemic stroke. Experimental work has elucidated some key molecular and cellular pathways that promote atherogenesis, as well as critical factors, such as inflammation and angiogenesis that contribute to the progression of the plaque. Nonetheless, atherosclerosis still represents a widespread health problem and an unmet medical need, as it causes cardiovascular diseases that continue to be the major cause of death in the Western world.[1–5]

Atherosclerosis starts when the endothelium, namely the thin layer of cells (i.e. endothelial cells, ECs) that keeps the artery smooth and allows blood to flow, becomes damaged. The main risk factors causing such dysfunction of the endothelium include hypertension, diabetes, high levels of cholesterol and sugar in the blood, toxins from cigarette smoking and accelerated vascular aging. Those areas of the endothelium that are injured are likely to undergo a build-up of plaque, initiating a process called atheroma formation or atherogenesis. Specifically, the activation of inflammation among ECs, along with several other compensatory responses, augment the vascular permeability of the endothelium. As a consequence, lipids and low-density lipoproteins (LDL) (i.e. the so-called "bad" cholesterol) are able to enter and then accumulate under the injured endothelium, in the tunica intima layer.[4] This instigates mononuclear phagocyte cells and other leukocytes to stream in and differentiate into tissue macrophages to internalise the lipid and lipoprotein particles through phagocytosis. Once trapped in the artery, LDL and fatty acids particles can undergo progressive oxidation, triggering an uncontrolled and relentless inflammatory process. The hallmark of a nascent atheroma, indeed, arises from the oxidation of cholesterol that causes the death of macrophage cells and the formation of voluminous foam cells, initiating the vicious inflammatory cycle.[5]

Expanding lesions are also promoted by signalling proteins such as chemokines, pro-inflammatory cytokines, reactive oxygen species (ROS) and other mediators secreted by macrophages and apoptotic foam cells. These substances not only contribute to the recruitment of further inflammatory cells which augment the pre-existing inflammatory environment, but also stimulate the migration and proliferation of vascular smooth muscle cells (VSMCs) into the tunica intima.[6] VSMCs secrete extracellular macromolecules, such as collagen, elastin and fibrin, that harden and enclose the plaque. Alongside this process of fibrosis, an increment of intracellular concentration of calcium is also observed, causing the calcification of the necrotic core.[7,8]

A mature atherosclerotic plaque is made up of a mass of intracellular and extracellular lipids and lipoproteins forming the "fatty steak" (comprising cholesterol esters, free cholesterol crystals, phospholipids, triglycerides), a high number of inflammatory phagocytic cells such as monocytes, macrophages and other immune cells (T-lymphocytes), foam cells and debris, VSMCs and their waste productions, as well as apoptotic cells which constitute the necrotic core. This cluttered heap is stuck under the endothelium, in the tunica intima layer, covered by a fibrous cap, predominately composed of VSMCs and collagen proteoglycan matrix (**Figure 1**).[3,7,9,10]

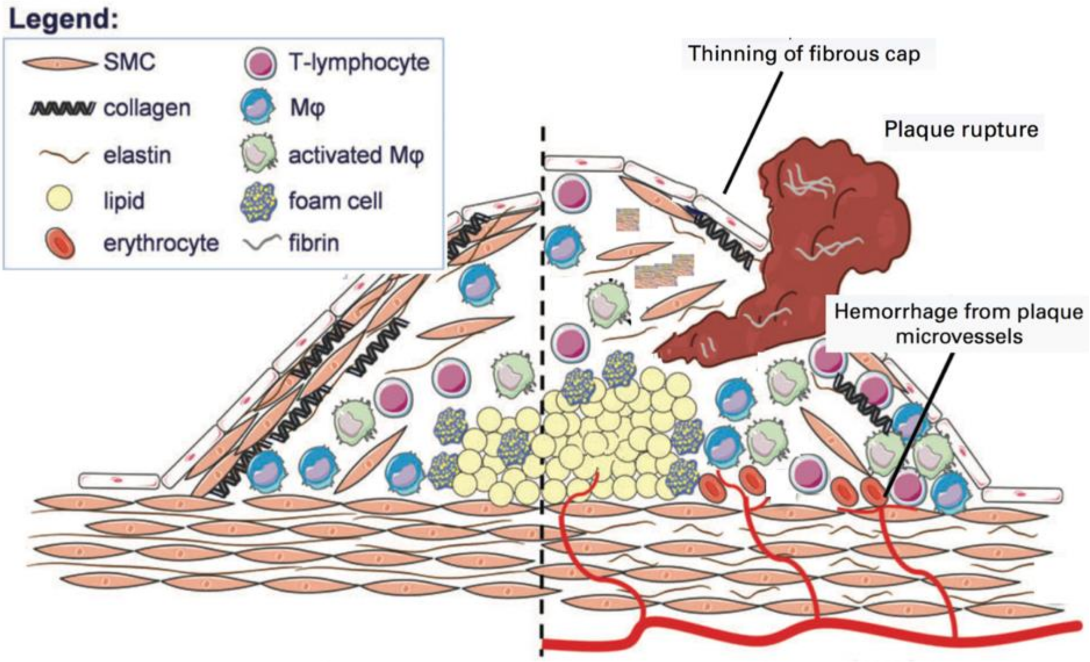


Figure 1: (*left)* Overview of mature atherosclerotic plaque constituents including a lipid-rich necrotic core*,* foam cells, immune cells such as macrophages (Mφs) and activated Mφs, as well as T-lymphocytes. These are encapsulated by a fibrous cap which is mainly made of vascular smooth muscle cells (SMCs) and fibrous tissues (collagen, elastin and fibrin). (*right*) Prototype of vulnerable and active thrombogenic plaque characterised by a thin-cap fibroatheroma. Intra-plaque microhemorrhages due to neovascularisation (angiogenesis) and consequent intrusion of erythrocytes inside the plaque (intra-plaque bleeding). Plaque rupture with extravasation of blood and thrombus formation. (Image was adapted with permission from reference [**11**]).

Plaque rupture mainly occurs among lesions of unstable areas defined as “thin-cap fibroatheromas” or “vulnerable plaque”. These areas are characterised by particularly thin fibrous caps (< 65 µm) consisting of few collagen bundles and rare VSMCs (**Figure 1**).[5,10–13] The mechanisms by which the fibrous caps thin, predisposing the plaque to rupture, include the death by apoptosis of VSMCs and the consequent breakdown of the collagen and extracellular matrix synthesis.[14] Macrophages are also pivotally involved in weakening and degrading the already thin fibrous cap, through secretion of proteolytic enzymes such as matrix metalloproteinases (MMP).[12]

In addition to the thickness of the fibrous cap and the degree of inflammation, the pathological angiogenesis promoted by ECs contributes to render the atherosclerotic plaque unstable and thus prone to rupture.[9,15–18] In fact, neovascularisation occurs to nurture the atherosclerotic plaque and the newly formed blood vessels and microvessels are inherently immature and leaky, causing intra-plaque microhemorrhages. This permits the influx of blood constituents, such as plasma proteins and erythrocytes into the sub-endothelial space of the artery, as well as the infiltration of macrophages which further aggravate the inflammation.

When plaque rupture occurs, the local extravasation of blood, and the consequent interaction of overlying blood with plaque constituents, induces thrombotic changes resulting in the activation of a clotting cascade. In particular, when plasma enters into contact with the necrotic core, tissue factor procoagulants instigate the aggregation of platelets and the coagulation at the affected area, forming blood clots and resulting in luminal narrowing. A large blood clot can partially or completely obstruct the blood flow through the artery. Complicated thrombogenic plaques frequently show signs of multiple ruptures and repairs, and severe luminal thrombosis often occurs in patients dying of acute myocardial infarction or stroke.[7,17,19,20]

1.1.2 Metabolic dysfunction in atherosclerosis

As outlined above (**paragraph 1.1.1**), several cellular mechanisms are involved in plaque progression, and specifically inflammation and angiogenesis are the primary and critical factors in the pathogenesis of atherosclerosis and its complications. Currently, therefore, the condition of atherosclerosis is incontrovertibly recognised as a dynamic inflammatory process, aggravated by pathological angiogenesis.[4]

Macrophages play a fundamental role throughout all stages of the disease, driving the inflammation process. These activated immune cells are subjected to significant alterations in their metabolism and their glucose uptake.[21] Specifically, it has been reported that proinflammatory macrophages rely greatly on glycolysis to fulfil the high energy demand needed for the production of cytokines and ROS. In particular, the activation of macrophages is characterised by the switching of their metabolism from oxidative phosphorylation to anaerobic glycolysis, even in the presence of oxygen.[22–26] Studies have shown that the excessive glycolytic rate is responsible for the consequent pathologic behaviour of macrophages in the plaque.

Together with pro-inflammatory macrophages, ECs also actively partake in atherosclerosis by concomitantly directing pathological angiogenesis, thus contributing to intra-plaque bleeding and plaque destabilisation. Remarkably, also within this class of cells a similar metabolic change in favour of the glycolysis pathway is observed.[27] In fact, it has been reported that, stimulated by angiogenic factors, ECs change from quiescence to proliferation,[28] serving thereby to vascularise otherwise avascular tissues through sprouting and ensure presence of adenosine triphosphate (ATP) and oxygen also in intra-plaque *vasa vasorum*.[29,30] ECs have therefore a high bioenergetic demand in order to generate nimble filopodia and lamellipodia, and they rely on glycolysis to meet this demand.[31] Specifically, the glycolytic flux in ECs increases from 1% to 85% activity, with lactate production, while oxygen consumption reduces despite its immediate availability in blood vessels. Likewise, it has been described that also the phenotypic modulation, leading VSMCs to undergo proliferation and migration in atherosclerosis, is associated with the induction of similar compensatory glycolytic disorder, supported by increased glycolytic protein expression and glucose uptake with lactate production.[6,32–34]

1.1.2.1 The Warburg effect

Thereafter, inflammatory and angiogenic stimuli make macrophages, ECs and VSMCs opt for anaerobic glycolysis as their main source of energy, regardless that they are in the presence of a normal oxygen blood concentration (normoxic conditions). The propensity towards reliance on glycolysis for energy and biomass production, on the other hand, normally occurs as a defence mechanism against hypoxic conditions, namely when a tissue must function anaerobically. Eukaryotic cells produce energy in the form of ATP by cellular respiration, which is a sum of interrelated metabolic pathways, including cytoplasmic glycolysis followed by the oxygen-dependent Krebs cycle (also known as tricarboxylic acid (TCA)-cycle) and oxidative phosphorylation. During hypoxia, however, due to the scarcity of oxygen, the pyruvate generated from the metabolism of glucose through the glycolysis process, cannot be converted into acetyl coenzyme A and then enter the mitochondrial (TCA)-cycle. In particular, pyruvate gets reduced to lactate, instead of being metabolised oxidatively in the mitochondria. This metabolic reprogramming allows cells to produce ATP even in a low oxygen environment.[22] It is also well-known, however, that the efficiency of this alternative metabolic pathway, in terms of the number of synthesised ATP molecules per molecule of glucose, is far inferior to the normal aerobic glycolysis followed by oxidative phosphorylation. Indeed, the former process only provides two molecules of ATP per molecule of glucose, against the 36 ATP molecules produced by the latter.

The metabolic switching of macrophages, ECs and VSMCs toward anaerobic glycolysis, despite the ready availability of oxygen in the artery, seems therefore contradictory, considering the low ATP yield of the glycolytic process. In principle, there are two ascribable reasons that make glycolysis preferable over oxidative phosphorylation. Firstly, glycolysis is a prompt process and about 100 times faster than oxidative metabolism. Secondly, the glycolytic pathway additionally provides i) the biosynthetic precursors used by macrophages for the production of the pro-inflammatory proteins[35] and ii) the precursors required for cell proliferation.[26,30]

The analogous behaviour, characterised by an anomalous metabolism shift towards glycolysis in a normoxic conditions, was already observed in tumour cells as early as 1923[36] by Otto Warburg, who termed it the “Warburg effect”. Targeting the Warburg effect is in fact a known therapeutic strategy for fighting cancer and has been widely explored *in vitro*, *in vivo* and in clinical studies. Currently, multiple and growing lines of clinical evidence have identified this phenomenon as a common feature in several metabolic disorders and chronic morbidities, encompassing atherosclerosis and cardiovascular diseases.[22,37,38] In particular, this phenomenon is observed, more generically, within cells in a rapid growth phase, where ample energy production and essential biosynthetic precursors are urgently required for a high-speed proliferation with a negligible oxygen consumption. Hence, from a clinical perspective, harnessing the Warburg effect *via* glycolysis modulation, may now help target therapeutics not only for the oncology development but also for the treatment of inflammation or angiogenesis in atherosclerosis.

Several biological targets have been identified for regulating glycolysis and thus, ultimately, the metabolic dysfunctions due to the Warburg effect.[22] One emerging target is the tandem enzyme 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase (PFKFB3) which regulates the glycolytic rate (**Figure 2**). PFKFB3 has been thought to be responsible for the heightened glycolytic activity in cancer cells[39–41] and some small molecule PFKFB3 inhibitors have been developed to block the Warburg effect, mainly for cancer chemotherapy applications (**paragraph 1.1.3.1**). In recent years, however, PFKFB3 has also gained attention in the field of inflammatory dysfunctions and pathological angiogenesis, as it is overly expressed in macrophages[26] and is the most abundant isoenzyme in various EC subtypes.[29] In addition, a rise in PFKFB3 expression was also recently observed in proliferative VSMCs.[32]

After inflammatory stimuli, macrophages up-regulate the expression of PFKFB3 in order to increase the glycolytic flux and promote pro-inflammatory activation. Recent studies have reported that inhibition of PFKFB3 expression results in a decrease of glycolytic activity in macrophages leading to a reduction of pro-inflammatory activation.[26] Likewise, knockdown of PFKFB3 not only lowers the rate of proliferation of ECs, but also accelerates endothelial apoptosis, indicating that this protein is important in the physiological activity of ECs.[30,42]

Taken together, these arguments suggest that PFKFB3 should be considered as a promising target point for the treatment of atherosclerosis.

1.1.3 PFKFB3 Function

One of the three rate-limiting steps (r.l.s.) of the glycolytic flux is the conversion of fructose-6-phosphate (F6P) to fructose-1,6-bisphosphate (F1,6P) catalysed by 6-phosphofructo-1-kinase (PFK-1). The opposite reaction is the r.l.s. of the gluconeogenesis pathway, catalysed by fructose-1,6-bisphosphatase (FBPase). The isomeric metabolite of

F1,6P, fructose-2,6-bisphosphate (F2,6P), is a potent allosteric modulator of these controlling steps. Specifically, F2,6P guides both glycolytic and gluconeogenic pathways by respectively activating the gluconeogenic enzyme FBPase and inhibiting the glycolytic enzyme PFK-1 (**Figure 2**). Thus, an increased concentration of F2,6P would allow transformed cells to maintain a high glycolytic flux.[43]

The 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase enzymes (PFKFBs) constitute a family of four homodimeric bifunctional enzymes displaying two identical subunits. Each subunit comprises two functional domains which have mutually opposing catalytic activities: a kinase activity (Kase) in the *N*-terminal domain and a phosphatase activity (Pase) in the *C*-terminal domain. Hence, PFKFB enzymes modulate both i) the biosynthesis of F2,6P from the phosphoryl donor ATP where F6P is the phosphoryl acceptor (Kase reaction), and ii) the reverse reaction which is the hydrolysis of F2,6P with ADP to F6P and ATP (Pase reaction).[43,44] The two catalytic activities of PFKFBs are subtly regulated.

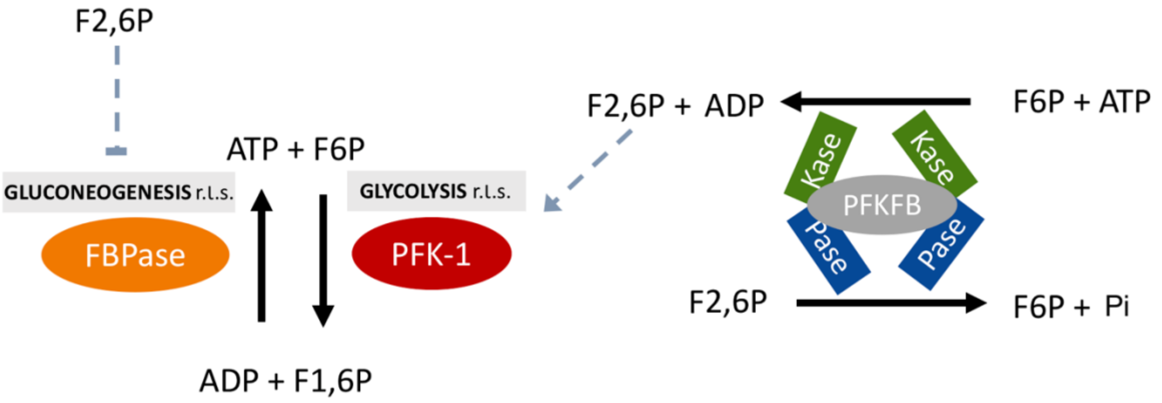


Figure 2: Allosteric modulation of the rate limiting steps (r.l.s.) of both the gluconeogenesis and glycolysis pathways by fructose-2,6-bisphosphate (F2,6P metabolite): inhibition of gluconeogenic fructose-1,6-bisphosphatase, (FBPase enzyme); activation of the glycolytic 6-phosphofructo-1-kinase (PFK-1 enzyme). Mutually opposing catalysis function of the tandem dimeric enzyme PFKFB: biosynthesis of F2,6P, promoted by kinase domains (Kase); hydrolysis of F2,6P promoted by phosphatase domains (Pase).

As reported by Kim *et al.*,[45] the Kase reaction mechanism proceeds through a direct substrate-substrate interaction between ATP and F6P that allows the phosphoryl transfer without formation of a phosphoenzyme intermediate. As a consequence, no significant conformational changes of the enzyme occur during the reaction. Also, the Pase reaction mechanism was outlined by the same authors through a crystallographic study.[46]

Every mammalian cell presents one of the four isoforms of PFKFB according to a structure and tissue-specific regulation of distinct encoding genes. PFKFB1 is present in the liver, PFKFB2 in the heart, PFKFB4 in the testis, while PFKFB3 is defined as the inducible or the ubiquitous form, since it is expressed in all tissues after particular conditions (mainly hypoxic stress).[30,47,48] Each isoform responds differently to various metabolic products, allosteric activators, and hormonal and growth factor signals that govern the final concentration of F2,6P and, ultimately, the rate of glycolysis.[49,50]

Although the amino acid sequences in the core catalytic domains within the proteins of the PFKFB family have an elevated homology (85 %),[51] peculiar structural features in each isoform differently tune their kinase vs. bisphosphatase activity ratio and the kinetic of their catalysis. Notably, the inducible isoform, PFKFB3, has unique sequence characteristics that make the Kase activity considerably predominant over the Pase activity. In particular, in contrast to the rather balanced activity ratio exhibited by the liver, heart and testis isoforms, PFKFB3 displays a kinase vs. bisphosphatase activity ratio larger than 700.[52,53] Crystallographic study revealed that only isoform PFKFB3 displays a β-hairpin structure in the *N*-terminal domain that appears to induce a self-inhibition of the Pase activity.[46] Additionally, the phosphorylation of the amino acid residue, Ser460, by other enzymes, such as protein kinase A (PKA) or AMP-dependent protein kinase, can increase the ratio in the inducible form further to > 3000.[54–56]

Taken together, these characteristics explain the reason why cells presenting an overexpression of the inducible isoform PFKFB3, have such a greatly enhanced glycolytic rate.

1.1.3.1 PFKFB3 Kinase inhibitors

Thus far only a few examples of small molecule inhibitors of the PFKFB3 kinase moiety have been reported in the literature. Both *in silico* and "traditional" high-throughput screenings have been described in order to find new drug candidates.[57–62] Representative examples of small molecule inhibitors are depicted in **Figure 3**.

The most studied inhibitors are undoubtedly 3-(3-pyridinyl)-1-(4-pyridinyl)-2-propen-1-one (3PO; **1**)[63] and its analogues PFK15 and PFK158, (compounds **2** and **3**) as well as the closely related benzo[e]indole **4** (**Figure 3**),[64–66] which are able to decrease the level of F2,6P and the glucose uptake, both *in vitro* and *in vivo*. The virtual screening (VS) that led to the discovery of 3PO was based on a computational model of the enzyme that was obtained by homology modelling using the X-ray structure of the rat testes, isoform PFKFB4 as a structural template. These compounds display rather poor drug-like properties such as low solubility and lack of selectivity. Nonetheless, the derivative **3**, PFK158, a more potent inhibitor having better pharmacokinetic properties than 3PO, recently underwent phase I clinical trials as an antitumor agent.[67] There is growing evidence, however, (including our internal data in the Moglynet Project Consortium) that the activity data previously reported for 3PO cannot be independently reproduced. Indeed, in repeated kinase assays with the isolated enzyme, 3PO resulted, contrarily, to be inactive.[60–62,68] These results, along with the absence of literature indications confirming the actual receptor binding, suggest that all the reported biological activity of 3PO, such as reduction of glycolysis, cell cycle regulation and tumour growth inhibition, are very unlikely to be linked to the inhibition of PFKFB3, but are more probably due to some alternative nonspecific mechanisms.

Seo *et al.* performed a structure-based virtual screen aimed at finding other inhibitors with high specificity. The study approach consisted of targeting the F6P binding site of PFKFB3, rather than the ATP binding site, in an attempt to avoid side effects due to cross activity with other kinases. The observed antiproliferative effect of the identified hit compound **5** N4A (**Figure 3**), however, was ascertained only for isoform PFKFB3 and the selectivity towards the other PFKFB isoforms was never assessed.[58]

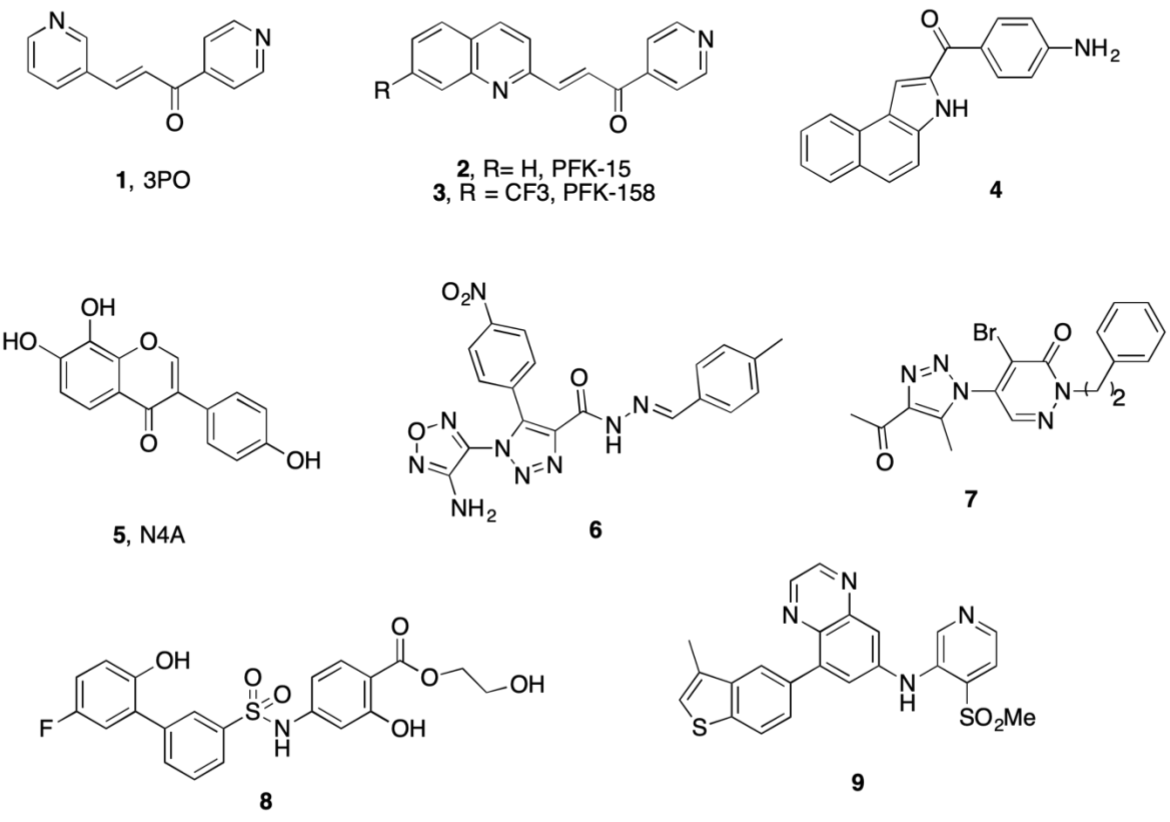


Figure 3: Chemical structure of small molecules reported in the literature as PFKFB3 inhibitors.

Another virtual screen on a rat PFKFB3 model led to a class of new aminofurazan-triazole series designed as PFKFB3 regulators, theoretically binding either the kinase (glycolysis inhibitors) or the phosphatase active site (glycolysis activators), according to a docking study.[69] The small molecule **6** represents a kinase inhibitor, able to reduce the lactate accumulation to 20% compared to the control.[69] The high topological polar surface areas (TPSA) and molecular weights (MW), however, render these small molecule modulators implausible lead compounds for a further optimisation. Similarly, the triazolo-arylpyridazinone **7** was selected following an investigation of the structure-activity relationships (SAR) around a class of inhibitors discovered using high-throughput screening.[60] Nonetheless, given the nature of the cellular *in vitro* assays used for testing **6**, **7** (**Figure 3**) and their derivatives, the authors of both studies also disclosed possible different modes of action, perhaps involving the interaction with other similar glycolytic enzymes.

Other related compounds have been patented as PFKFB3 inhibitors but their binding sites are unknown and they are also characterised by a rather mediocre half maximal inhibitory concentration (IC_{50}).[65,70,71] Moreover, due to the strong homology of the four isoforms of the PFKFBs family, the majority of these compounds show a general lack of selectivity.

An important piece of work published by the pharmaceutical company AstraZeneca (AZ) describes the SAR study that led to the development of a new class of potent and selective derivatives, bearing a phenoxy-indole or phenoxy-indazole scaffold.[68] These novel series of inhibitors (hereinafter referred to as the AZI compounds) are ATP competitors and display an IC_{50} in the low nanomolar (nM) concentration range. Furthermore, six chosen compounds from the AZI series (**Table 1**) were co-crystallised inside the human PFKFB3 enzyme, confirming their binding at the ATP pocket (PDB IDs: 5AJX, 5AJY, 5AJZ, 5AJW, 5AJV, 5AK0).

Two years later, AZ published another paper concerning the identification via high-throughput screening of a novel family of compounds having activity in the sub-µM concentration range and displaying promising lead-like properties.[72] The active aromatic core of these ligands (hereinafter referred to as the AZPP compounds) is a dihydropyrrolopyrimidinone heterocycle. Four compounds of the AZPP series (**Table 3**) were also co-crystallised and provided as three-dimensional (3D) X-ray structures in the Protein Data Bank (PDB IDs: 4D4K, 4D4J, 4D4M, 4D4L).

Both of these families of AZ compounds (i.e. the AZI and AZPP series), were used as primary reference ligands in the present research study for the rational design of new candidate PFKFB3 inhibitors following two distinct workflows (**Chapters 3 - 4**).

Very recently, another two new entries in the group of PFKFB3 inhibitors were identified and biologically evaluated *in vitro*: the phenylsulfonamido salicylic acid **8**[61] and the aminoquinoxaline **9**[62] (**Figure 3**). Compound **8** represents the most active hit among 50,000 compounds screened through high-throughput enzymatic and cellular assays. Confirmed by Saturation Transfer Difference NMR, this compound is an F6P competitor since it reversibly binds PFKFB3 at F6P binding site.[61] Compound **9**, meanwhile, is an ATP competitor that was developed after high-throughput screening, through an *in silico* hit-optimisation process with blind molecular docking as well as the aid of crystallographic studies. [62]

1.2 Virtual Screening in drug discovery

1.2.1 Introduction

The early stage of drug discovery comprises the identification of novel hit molecules that bind to a certain target, usually a protein, and exert the desired biological effect. The traditional approach consists of the high-throughput screening (HTS) of corporate libraries of compounds against the biological target of interest. Specifically, with the HTS approach, an extremely high number of experimental assays have to be performed until the identification of a hit. Although it remains a very effective process, it is extremely expensive and time-consuming. In this context, modern computational methods, such as virtual screening (VS), represent a powerful means for hastening this process, also considerably reducing the costs of the entire drug discovery campaign. The term “virtual screening” first appeared in the 1990s and, since then, VS has been increasingly employed, both by academic laboratories and pharmaceutical companies, as an effective tool in medicinal chemistry research and drug discovery and development. In fact, compared to the costly and lengthy HTS, the VS methods allow mining and evaluation of large *in silico* databases of compounds in a reasonable amount of time, using molecular modelling and without consuming pricey substance materials. The result of the search is a filtered or ranked database where the affinity of each ligand towards a biological target is calculated, then compared with the other ligands of the library and sorted accordingly.

These computationally-aided predictions can be performed to complement the HTS, usually during a stage prior to the experimental screening. Indeed, one of the principal goals of VS is to enrich the starting search library and provide valuable clues to prioritise potential active molecules, in order to narrow down the actual number of compounds that are tested in experimental validation.

To date, a tremendous number of chem- and bioinformatics tools have been developed and proven to be quite successful in drug discovery and drug design. The existing broad screening strategies are conventionally divided into two different subcategories: i) ligand-based virtual screening (LBVS) and ii) structure-based virtual screening (SBVS). The first approach takes information from the structure and the properties of known active molecules, while the second approach takes information from the structure and the properties of a target receptor and its active site. LBVS methods are mainly based on similarity and diversity measures between molecules and reference active compound(s) that are assessed by comparing topological descriptors or putative pharmacophores to find new hypothetically active scaffolds. Whereas SBVS methods are based on the systematic measurements associated with the likelihood of a ligand to interact with a critical target in the disease of interest. Depending on the information available, the two methods can be used independently or in combination. Clearly, when both 3D structures of known active compounds and target protein are available, the combination of LBVS and SBVS can potentially be favourable to any individual screening method. However, *a priori* neither approach can be peremptorily considered better than the other, in terms of efficacy and efficiency.[73–75] These methods are data hungry procedures and the workflow of the VS can vary depending on the availability of structures and the knowledge of experimental biological data. Today, however, a number of structure databases of small molecules and protein targets have been created and provided to the entire scientific community to respond to the need for chemical and biological information.

1.2.1.1 Molecular Database

In any virtual screen, the choice of the initial chemical library (otherwise referred to as a chemical or molecular database, DB) to be scanned is of critical importance and can greatly affect the outcome of the drug discovery campaign. Both library size and extent of chemical diversity, which are key factors in the evaluation of the quality of a library, have to be set against the cost, time and the purposes of the VS method. For instance, databases embracing the chemical space as fully as possible will be more likely to contain true-active molecules but may be ‘heavy’ to manage and their screenings may necessitate massive computational cost and storage space. The estimated size of the chemical space is indeed greater than 10^{60} molecules, making it impossible to tackle with exhaustive search tools.[76]

On the other hand, small libraries displaying poor chemical diversity can only result in few redundant hits. Thus, clustering similar scaffolds within a screening library is usually necessary to reduce structural redundancy. Likewise, upfront filtering is also indispensable for prioritising molecules with appropriate physicochemical properties and for getting rid of problematic and reactive functionalities or improbable and non-synthesisable molecules. Traditional properties filters are, for example, based on the well-known Lipinski's "Rule of Five" (Ro5)[77] which delineates the portion of molecules in the chemical space that most likely can be flagged as "drug-like" molecules, namely having more chance to be orally bioavailable. There are also related rules adapted from Ro5, such as "Rule of Three",[78] or the guidelines suggested by Hann and Oprea for "lead-likeness", regarding the suitability of a compound to undertake follow-up and optimisation processes.[79,80]

Furthermore, databases formed by one (1D) or two-dimensional (2D) representations of molecules must typically undergo a 3D sampling and often require the generation of the possible multiconformers of each molecule, before performing the screen. In fact, in many VS methods, sampling represents a critical issue because, during the DB mining experiment, no new conformation is generated on-the-fly. Nonetheless, molecules are flexible entities and therefore pre-computed conformations must be already associated with the molecules in the input search DB.

Along with a search library to be screened, often a suitable test DB is also needed to validate the screening technique. In fact, as described in the following **paragraph 1.2.1.2**, when looking for new hit compounds *via* VS, it is desirable to test the method in question with respect to its ability to enrich the search DB. Specifically, the VS method is applied to a dedicated set library comprising known active ligands and a benchmarking of inactive ligands. From the outcome it is then appreciated how the former would rank against the latter. Obviously, the validation of the screening method requires the knowledge of both active and non-active ligands.

1.2.1.2 Validation of Virtual Screening Method

The evaluation of the outcome of a virtual screening experiment corresponds to its ability to discriminate biologically active molecules (true positive, TP) over non-active molecules (true negative, TN), in relation to the number of overestimated non-actives (false positive, FP) and discarded actives (false negative, FN). In other words, the success of a screening method is quantified by the Enrichment Factor (EF, **Equation 1**). The EF of a VS is its capacity to *enrich* the maximum number of TP in a defined percentage of top-ranked ligands from the screened library with respect to a random approach to selecting molecules. It is determined through the following equation:

$$\mathrm{EF} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}}\,\frac{\boldsymbol{N}}{\boldsymbol{n}}$$ **Equation 1**

where N is the total number of molecules in the search database and n is the portion of molecules retrieved by the screening.

Another forthright measure of that is expressed by the area under the curve value (AUC) derived from the Receiver Operating Characteristic curve (ROC). The ROC curve is the plot which correlates the 'true positive rate', also indicated as 'Sensitivity', with the 'false positive rate', also indicated as '1 – Specificity'. The Sensitivity is the ratio between the number of TP and the total number of active molecules (**Equation 2**), while Specificity is the ratio between the number of TN and the total number of non-active molecules (**Equation 3**).

$$\text{Sensitivity} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}}$$ **Equation 2**

$$\text{Specificity} = \frac{\boldsymbol{TN}}{\boldsymbol{FP+TN}}$$ **Equation 3**

The maximum value of AUC is 1, which corresponds to the best possible performance in which the VS search returned every active and correctly excluded every non-active compound (Sensitivity = Specificity = 1). While AUC values of 0.5, with a ROC curve corresponding to the bisecting diagonal line (true positive rate = false positive rate), indicate that the selection of hit molecules by the VS method is no better than a random guess.

The following paragraphs describe the most common computer-aided drug design methods and techniques for both LBVS and SBVS, with a special focus on the methodologies used in the present work.

1.2.2 Similarity Searching

One of the simplest and most widely used cheminformatics technique is similarity searching, in which a reference molecule is compared against a database of compounds. This field is extremely wide and encompasses numerous methods and applications.[81] For example, similarity-based measures are frequently used to cluster redundant molecules of a large database into similarity groups for reducing chemotype bias, as mentioned above (**paragraph 1.2.1.1**).

In the context of VS and drug discovery, similarity is a key concept and it is based on the assumption that similar molecules are likely to exhibit similar biological activity on a given target. This type of LBVS can be applied in a lead-optimisation campaign for the search for related compounds that are chemically similar to a given drug or lead molecule but may display improved potency and/or pharmacological properties or they may be more suitable for large-scale synthesis. Alternatively, one objective of a similarity search is to identify the nearest-neighbour molecules displaying similar physiochemical properties of a drug but having structurally novel molecular frameworks. This type of search has been termed ‘scaffold hopping’,[82,83] since it facilitates the identification of novel hits with different chemical scaffolds. This strategy can be also a valid means to break out of a protected patent space and create new intellectual property, or to overcome a “flat” structure-activity relationship (SAR).

To date, a number of computational methods have been developed to capture and encode the properties of ligands and to compare them in a systematic and consistent manner, through similarity search algorithms.[84] The search can be computed based upon topological indices, steric, electronic or physical properties, and can be grouped into two main categories: 2D and 3D methods.[85] In 2D similarity methods, bit strings called 'molecular fingerprints' are used to encode the compound chemical patterns. The Tanimoto coefficient (T_c) is probably the most widely used metric to compare the fingerprints of two molecules and quantify the degree of their similarity. It is defined by the fraction of shared chemical patterns, in the range between 0 (dissimilar molecules) and 1 (identical or very similar molecules). Clearly, although they have the advantage of a short processing time, 2D approaches might be too simplistic for the purpose of scaffold hopping. In this context, the more exhaustive 3D molecular representations are generally preferentially used as the queries of reference, due to their additional conformational and shape information content.[86–88] Some 3D methods define common substructures of molecules based on the spatial alignment of their sets of coordinates. These methods use iterative optimisation algorithms to stochastically search the maximal common substructure between molecules, taking into account the atomic nature.[83,89,90] Similarly, other 3D approaches rely upon shape information or pharmacophoric descriptors.[82,88]

1.2.3 Pharmacophore-Based Virtual Screening

The pharmacophore modelling technique is a powerful tool that has been used in medicinal chemistry for several years. The term *pharmacophore* is sometimes used incorrectly in this field, however. The consistent definition provided by the International Union of Pure and Applied Chemistry (IUPAC) is as follows:[91]

> *"A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and trigger (or to block) its biological response."*

And an additional note also elucidates that:

> *"A pharmacophore does not represent a real molecule or a real association of functional groups, but a purely abstract concept that accounts for the common molecular interaction capacities of a group of compounds towards their target structure. The pharmacophore can be considered as the largest common denominator shared by a set of active molecules. This definition discards a misuse often found in the medicinal chemistry literature which consists of naming as pharmacophore simple functionalities such as guanidines, sulphonamides or dihydroimidazoles (formerly imidazolines), or typical structural skeletons such as flavones, phenothiazines, prostaglandins or steroids."*

As explained, a pharmacophore model is an abstract 3D map characterised by steric and electronic molecular features that depicts the physiochemical nature and the coordinate position of chemical functional groups that are involved in the ligand-protein interactions. Conversely, it is not a real molecule or chemical portion of an active ligand, as erroneously stated at times in the literature. Hydrogen bond donors (HBD), hydrogen bond acceptors (HBA), positively or negatively charged groups, hydrophobic regions, etc. are features typically annotated and used, along with their spatial location and orientation, as pharmacophoric representations of small molecules. Also, the shape of a molecule can be embodied in a pharmacophore query, in order to specify the suitable geometric constraint required to fit the active site of the protein target.

Despite its apparent intuitive interpretation, a pharmacophore query model is never unequivocally designed, and its elucidation remains a conceptual challenge that requires knowledge and expertise of medicinal chemistry. Usually, to derive a pharmacophore model, a training set of active ligands is used as a template to determine which common features might be indispensable for the desired interaction with the protein. Preferably, the template ligands should be potent, chemically diverse and bind with similar modes in the same target pocket; however, theoretically only one active ligand could be sufficient to design a pharmacophore. Furthermore, if known, structural information about the conformation and the active site of the macromolecular target can also be of benefit for constructing or refining a good model. Hence, both ligand-based and structure-based strategies can be used for the expression of pharmacophore models.[92,93]

Currently there are several platforms being developed that facilitate the elucidation of a pharmacophore and that allow users to perform pharmacophore-based library screening.[94–98] Each platform application has its distinct criteria for encoding features into 3D queries, as well as different representations and customisation schemes. The software platform MOE (Molecular Operating Environment) of the Chemical Computing Group has different annotation schemes containing specific procedures for the abstraction of defined electronic and steric features.[98] MOE also allows multiple query hypotheses to be constructed after the flexible alignment of a given training set of known ligands. Queries elucidated in this way are made up of shared features within the ligands that can theoretically explain their structure-activity data.

Pharmacophore-based searching is a procedure for the screening of 3D libraries of small molecules that can be classified either as LBVS or SBVS, depending on the information used for the elucidation of the query map. In general, however, the search consists of aligning, with a certain tolerance, the ligand atoms onto the pharmacophoric site points, in order to retrieve only those molecules fitting the map. In this way, a pharmacophore can provide a powerful tool for pre-filtering large ligand libraries, potentially at any stage of the drug discovery process.

Although the pharmacophore search *per se* can be very fast, the DB subjected to screening must provide a broad coverage of conformers, the generation of which can be very slow. In fact, a critical aspect for the quality of matching that must be ascribed in this type of VS, is the 3D conformational flexibility of the molecules. Only a few software applications are able to tweak the conformation on-the-fly during the fitting process,[99] the majority treat the small molecules as rigid entities and hence pre-computed sampling of conformers is required. For instance, the MOE software application[98] adopts a rigid-body alignment between conformer and pharmacophore model and only translational and strain-free rotational movements are allowed. Alternatively, albeit more rarely, some software packages, comprising MOE, can be supplemented with dedicated multi-conformational databases.

1.2.4 Molecular Docking

Whenever the structure of a particular biochemical target is known, the molecular docking calculation is an SBVS method that has been very frequently adopted in drug discovery for hit identification and/or lead-optimisation campaigns.

The 3D structures of the macromolecular targets are generally determined with X-ray crystallography and/or nuclear magnetic resonance (NMR) spectroscopy and are deposited in the Protein Data Bank (PDB).[100] For crystal structures, a preliminary preparation procedure is required before they can be used in subsequent calculations. In fact, although representing an excellent source of information, macromolecular X-ray structures often suffer from missing or poorly resolved atoms, alternative geometries, or other crystallographic artefacts that need to be addressed and corrected. For instance, because of their low electron density, hydrogen atoms only weakly scatter X-ray radiation, making their detection difficult. For these reasons, an upfront computational assignment of protonation states is necessary. In other cases, a 3D model can be constructed *in silico* (i.e. by homology modelling), combining the accessible structural information of related homologous proteins with the amino acid sequence of the target of interest.

A 3D representation of the target can be used against a collection of compounds using molecular docking in order to predict their putative optimal binding mode and to calculate their tightness-to-fit. Docking algorithms involve the prediction of multiple conformations and orientations of a single ligand inside the binding site in order to select the most energetically preferred mode. This process is called ‘posing’, and its complexity is associated with the conformational degrees of freedom of the ligands. A computational scoring function is applied to achieve an approximate evaluation of how well the selected pose fits the binding site. The scoring function is a predictive mathematical model that produces a score which represents an estimation of the stability of the resulting docked ligand-protein complexes. In other words, the scoring function gives an approximation of the binding affinity on the basis of the tightness of the interaction, making it possible to prioritise the configuration of each ligand that would more likely bind to the receptor. Finally, in a post-processing step, the best geometries of the ligands are ranked according to their assigned docking score in order to discriminate promising compounds to take further as candidate actives.

Usually docking is site-specific, meaning that the centre coordinates and the radius of the binding pocket into which the ligands are docked are defined *a priori*. Alternatively, when the position of catalytic or allosteric sites of a target are not available, ‘blind docking’ is used also to find out which part of the receptor the ligands would more likely bind.

Although the simulation of the true pose of the bound ligand is usually obtainable with relatively good reliability, the more advanced steps of scoring and ranking are still very challenging with the available docking methods. The scoring functions are designed to reproduce experimental data, such as the binding free energy, as a sum of several parameterised functions that represent the thermodynamics of interaction of the protein-ligand system. Some of the difficulties with scoring functions come from the fact that several terms and events, for example some molecular interactions, are difficult to parameterise. However, relatively simple scoring functions continue to be heavily used for VS, especially because of their acceptable time-scale of computation.

Other factors weakening accuracy may be due to a limited resolution of the X-ray target structure or to the disregard of possible pivotal participation of water molecules in the binding. ‘Cross-docking’, in which multiple models of the same protein are used to reproduce ligand pose in non-native structures, can provide more information than the docking calculations that rely on a single structure.

The flexibility of the receptor can also be a crucial issue, especially while modelling induced fit systems. However, the inclusion of flexibility in the docking calculation (referred to as ‘soft docking’ or ‘flexible docking) is controversial because it is theoretically more accurate, yet computationally more expensive.[101–104] For example, soft docking may be desirable when the bound protein (*holo form*) presents important conformational rearrangements with respect to the protein with an empty binding site (*apo form*). Nonetheless, to solve docking problems, ‘rigid docking’, using an unmoving protein, is most of the time preferred in order to reduce the computational cost.

Methods for docking ligands have a long history, and in the literature there is a plethora of works and reviews describing the existing docking programs and their applications.[101,105] Today there are a number of docking programs providing different conformational sampling algorithms and a variety of scoring functions.[102,106–109]

The docking methodology used in the PLANTS (Protein-Ligand ANT System) program, developed by Korb *et al.*,[110,111] is based on a stochastic optimisation to find a minimum energy conformation of the ligands in a defined binding site. The empirical ChemPLP scoring function employed in PLANTS mainly takes into account steric complementarity and hydrogen bonding interactions between the protein and the ligands. It is characterised by variables that describe the degree of freedom (translational, rotational and torsional) of the ligands. To find the best ligand orientation and conformation, the PLANTS search algorithm minimises the scoring function by tuning the degree of freedom values through an iterative optimisation process. The optimisation technique uses the ant colony optimisation (ACO) metaheuristic algorithm. This mathematical method is inspired by the behaviour of real ants using pheromones to find the shortest path from their nest to a food source according to an iterative process. Similarly, PLANTS makes random pose changes and uses an "artificial pheromone" to mark low-energy ligand conformations that, in subsequent iterations, are modified, until an energy minimum is found.[110,111]

PLANTS also provides 'flexible receptor docking' (or 'soft docking') that addresses flexibility to the part of the target protein that adapts its conformation to an incoming ligand (e.g. the amino acid side chains of the binding site).[103] The program allows degrees of freedom in defined flexible side chains of amino acids by adopting the same ACO methods used to explore the conformational space of ligands. This approach can theoretically improve both pose prediction success and enrichment rates, especially if the *holo form* differs substantially from the *apo form*. As anticipated above in this paragraph, however, modelling the molecular flexibility of the target protein drastically increases the required computing time and cost.

1.2.5 Molecular Dynamic (MD) Simulations

The docked conformers pre-selected through scoring functions, can be further evaluated (re-scored and re-ranked) using more complex ranking schemes. Specifically, methods based on Molecular Dynamic (MD) simulations, involving a more detailed treatment of electrostatic and van der Waals interactions, and also capturing the solvation effect, might serve to improve the quality of affinity estimation. Moreover, the MD simulation of ligand-protein complex can also be used to validate and/or refine the posing outcome obtained with molecular docking, by analysing the stability of the trajectories of the atoms throughout the time of the simulation.

MD physical methods were first developed in the late 1970s[112] and classically have been used for studying the motion and the conformational rearrangements of molecules in evolving systems. In recent decades, however, they have been the subject of increasing interest and applications in drug discovery and VS. This is also due to the significant progress made in computational technology and in particular to the advent of the powerful hardware based on graphical processing unit (GPU). In fact, GPUs have substantially increased the affordability of very demanding computational calculations, such as the complex task of high-throughput simulation of macromolecular systems in SBVS.

The MD simulation technique uses approximation based on Molecular Mechanic energies (MM) to estimate the forces acting between atoms and molecules, and applies the Newtonian laws to calculate the successive positions and velocities of atoms. MM potential energies are expressed as the sum of interactions arising between bonded atoms (stretching, bending and torsions) and non-bonded atoms (van der Waals and electrostatic interactions). To calculate the contribution of these forces, a series of parameters, referred to as the ‘force field’, describing chemical bonding, atomic angles, atomic radii, partial atomic charges, etc., is assigned. The forces are assumed to be constant during discrete intervals of time, defined time-steps. The integration of the Newtonian law of motion for each time-step gives successive time-dependent configurations that altogether produce the trajectories of the evolving system. Several force fields have been developed and are commonly used in MD simulations, such as AMBER,[113,114] CHARMM,[115] GROMOS[116] and others, differing principally in their approach to parametrisation and their suitability to distinct type of systems.

Usually, the simulation is produced for one protein-ligand complex in periodic boundary conditions (PBC) so as to better mimic a homogeneous bulk system in which an infinite number of complexes take part. The unit box containing the simulating complex immersed in the solvent, is evenly repeated in every direction of three-dimensional space.

The result is an infinite network of identical boxes (ensemble), so that the number of particles in every virtual copy is constant and any change is perpetuated an infinite number of times over each box.

The conditions in which the system is simulated are controlled by the thermostat and the barostat algorithms that mimic a thermostatic bath and the isobaric conditions respectively. In the NVT canonical ensemble, the number of particles (N), volume (V) and temperature (T) are kept constant, allowing energy fluctuation. While in the NPT isothermal-isobaric ensemble volume and density can vary and pressure (P) is conserved, as well as the number of particles and the temperature.

The actual production simulation starts after an equilibration/stabilisation step in which the system is minimised and relaxed to reach a stationary state and so as to remove initial unfavourable contacts between solvent and solute. The production time can last up to a 100 ns or more, depending on the complexity of the system and the purpose of the study. To test the stability of biomolecular aggregates, simple MDs ranging up to few ns of timespan length are reasonably adequate.[117] The stability can be assessed by measuring the coordinate root mean squared deviation (RMSD) of the atoms trajectory, which measures the degree of fluctuation within the coordinates of a set of atoms across the whole production time. Clearly, a stable system will show little variations in the RMSD profile, whereas a particularly unstable system, in which the ligand may even leave the binding pocket, will generate a high variation in RMSD values.

1.2.6 MM-GB(PB)SA calculation

Another relatively accurate way for estimating the free energy of the binding of small molecules to a biological target ($\Delta G_{\mathrm{bind}}^{\mathrm{solv}}$) is the calculation of Molecular Mechanics energies combined with the Generalised Born or Poisson-Boltzmann, and surface area continuum-solvation (MM-GBSA or MM-PBSA) methods. These methods can be used for rescoring molecular DBs in a SBVS. They are based on the average energy contribution of ensemble MD trajectories of the solvated ligand-protein complex.[118,119]

ΔG_{bind}^{solv} is the difference in free energy of the solvated bound state (ligand-receptor complex in solvent) and the solvated unbound states (free ligand and free receptor in solvent) (**Equation 4**). However, this calculation would be very computationally demanding and inaccurate to perform, due to the huge energy contribution coming from the large number of solvent-solvent interactions and fluctuations. In fact, these interactions are an order of magnitude larger than binding energy, causing additional noise to the calculation.

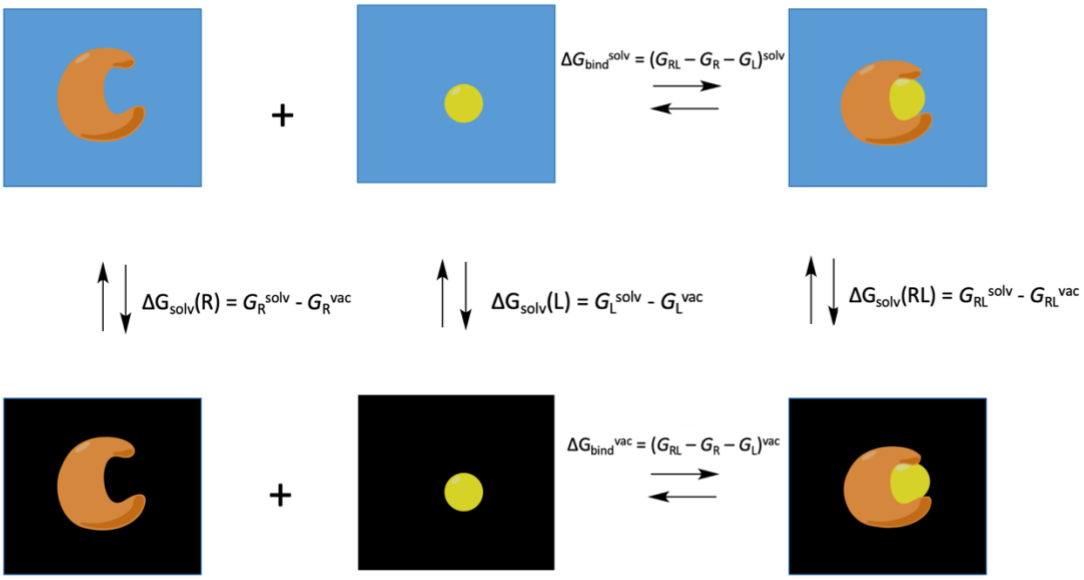


Figure 4: Schematic illustration of the thermodynamic cycle used to calculate the binding free energy (ΔG_{bind}^{solv}) associated with the chemical equilibrium between unbound receptor (R) and ligand (L) and their complex (RL), in their solvated states. ΔG_{bind}^{solv} is decomposed as a sum of binding free energy in the vacuum (ΔG_{bind}^{vac}) and solvation free energy (ΔG_{solv}).

To limit the complexity of the calculation, therefore, the interaction between ligand and receptor is considered as in the vacuum and corrected by adding the contribution of the solvation energy. As a result, ΔG_{bind}^{solv} is decomposed according to the thermodynamic equilibrium depicted in **Figure 4**, and approximated as a sum of individual uncorrelated terms: the changes upon binding free energy in the vacuum $\Delta G_{\mathrm{bind}}^{\mathrm{vac}}$, and the free energies of solvation (ΔG_{solv}) of each state (i.e. the difference in free energy of the solvated and unsolvated states) (**Equation 5**).

$$\Delta G_{\mathrm{bind}}^{\mathrm{solv}} = (G_{\mathrm{RL}} - G_{\mathrm{R}} - G_{\mathrm{L}})^{\mathrm{solv}} \qquad \textbf{Equation 4}$$

$$\Delta G_{\mathrm{bind}}^{\mathrm{solv}} = (G_{\mathrm{RL}} - G_{\mathrm{R}} - G_{\mathrm{L}})^{\mathrm{vac}} + [(G_{\mathrm{RL}}^{\mathrm{solv}} - G_{\mathrm{RL}}^{\mathrm{vac}}) - (G_{\mathrm{R}}^{\mathrm{solv}} - G_{\mathrm{R}}^{\mathrm{vac}}) - (G_{\mathrm{L}}^{\mathrm{solv}} - G_{\mathrm{L}}^{\mathrm{vac}})]$$

$$\Delta G_{\mathrm{bind}}^{\mathrm{solv}} = \Delta G_{\mathrm{bind}}^{\mathrm{vac}} + [\Delta G_{\mathrm{solv}}(\mathrm{RL}) - \Delta G_{\mathrm{solv}}\mathrm{R} - \Delta G_{\mathrm{solv}}\mathrm{L}]$$

$$\Delta G_{\mathrm{bind}}^{\mathrm{solv}} = \Delta G_{\mathrm{bind}}^{\mathrm{vac}} + \Delta G_{\mathrm{solv}} \qquad \textbf{Equation 5}$$

This approach was originally developed in the AMBER software.[120] Specifically, an ensemble average of vacuum binding free energy and solvation free energy is calculated from a given number of snapshots collected from the MD trajectories. In fact, the AMBER modules *tleap* and *cpptraj* create the required topology and trajectory data of the free ligand and free receptor, as well as the unsolvated states, by simply removing the appropriate atoms from the selected snapshots. With this approach, only a single production simulating the solvated receptor-ligand complex is required for the calculation (single trajectory approach). In less common cases, when the modelled system displays significant conformational changes and important protein rearrangements, like in an induced fit system, then three separated simulations of solvated unbound ligand, receptor and complex might be used (a three trajectories approach). The single trajectory approaches, however, often give more accurate results than those with three trajectories.[121]

The binding free energy in the vacuum $\Delta G_{\mathrm{bind}}^{\mathrm{vac}}$ is approximated by the average of all internal molecular mechanic (MM) interaction energies determined from the MD force field (stretching, bending, torsions, van der Waals and electrostatic) between receptor and ligand. A conformational entropy variation term, multiplied by absolute temperature ($T\Delta S$), should be added to the MM interaction energy for a more accurate estimation of the binding free energy. Because of the high computational cost required for its determination, however, and the additional uncertainty added by this calculation, most of the time, $T\Delta S$ is neglected from the calculation (**Equation 6**).[118,122,123]

The solvation term ΔG_{solv} is the sum of electrostatic and non-electrostatic term (**Equation 7**). The electrostatic solvation energy, which models the polar effects, is obtained by the Generalised Born (GB) or Poisson Boltzmann (PB) equation. Non-electrostatic contributions, which model the hydrophobic effects, are proportional to the surface area accessible to the solvent (SA). From here derives the MM-GBSA or MM-PBSA equation (**Equation 8**).

$$\Delta G_{\mathrm{bind}}{}^{\mathrm{vac}} = \Delta E_{\mathrm{MM}} - T\Delta \mathrm{S} \approx \Delta E_{\mathrm{MM}} \qquad \textbf{Equation 6}$$

$$\Delta G_{\mathrm{solv}} = \Delta G_{\mathrm{el}} + \Delta G_{\mathrm{non\text{-}el}} = \Delta G_{\mathrm{GB/PB}} + \Delta G_{\mathrm{SA}} \qquad \textbf{Equation 7}$$

$$\Delta G_{\mathrm{bind}}{}^{\mathrm{solv}} = \Delta E_{\mathrm{MM}} + \Delta G_{\mathrm{GB/PB}} + \Delta G_{\mathrm{SA}} \qquad \textbf{Equation 8}$$

In both the Generalised Born model and the linearized Poisson-Boltzamann equation, the solvent is regarded as an implicit continuous medium with a specific dielectric constant value (i.e. $\varepsilon_{\mathrm{water}} = 80$). This homogeneous medium solvation model, however, does not contain any information about water molecules in the binding site or their role in the ligand binding. In some systems, this lack of information can influence the result quite substantially, especially when water is pivotal for the ligand binding (e.g. water residue bridging through hydrogen bond) or when water is involved in transient interactions.

Some efforts have been made to account for the contribution of structural water molecules in the MM-GBSA or MM-PBSA equation by explicitly including a hydration shell around the ligand. The Nwat-MMGBSA method, which was adopted in the present work, allows important water residues to be selected as part of the receptor, providing improvements in correlation with the biological data and thus increasing the prediction power.[124–127]

2

Aim of the Research

2 Aim of the Research

As outlined in **Chapter 1**, the metabolic dysfunction characterised by an upturn of the glycolytic rate and a downregulation of oxidative phosphorylation regardless of the levels of oxygen is a well-known phenomenon called the Warburg effect. Recent studies have highlighted the importance of this cellular metabolism reprogramming as the origin of the inflammation and pathological neovascularisation in atherosclerosis. In fact, various types of cells pivotally involved in the progression of atherosclerotic plaque display much the same glycolysis-driven metabolic profile. The result of this change is the fast production of ATP and bioprecursors, with a relentless high proliferation that leads to an improper immune response and plaque destabilisation. Targeting the Warburg effect *via* the inhibition of glycolysis represents an attractive therapeutic approach to delay the progression of atherosclerotic plaque or even, possibly, induce its regression.

PFKFB dimeric bifunctional enzymes control the glycolytic flux through their Kase/Pase activities that regulate the level of F6P and F2,6P molecules. In fact, the latter allosterically activates the glycolytic enzyme PFK-1 and inhibits the gluconeogenic enzyme FBPase. Among the existing four PFKFB isoforms, the inducible ubiquitous form PFKFB3 presents an unbalanced Kase vs. Pase activity ratio in favour of the Kase activity, resulting in a high level of F2,6P and a consequent acceleration of the glycolytic rate.

From a clinical perspective, based on the substantial reported evidence suggesting the implication of PFKFB3 in the Warburg effect, this enzyme may serve as a very effective biological target. Specifically, inhibition of its prevailing kinase activity could reinstate the regular aerobic respiration, thereby restoring metabolism homeostasis and a normal glucose utilisation. This approach represents an innovative therapeutic strategy for the stabilisation of atherosclerotic plaque.

To date, only a handful of small molecule inhibitors of the PFKFB3 kinase domain have been reported in the literature. Furthermore, the majority of these require further biological validation or optimisation, since they display relatively poor potency, poor isoform selectivity or “non-druggable” properties. Thus, there is a huge demand for novel ligands or a class of ligands with elevated affinity towards PFKFB3, capable of selectively inhibiting its kinase activity. Moreover, a novel isoform specific inhibitor could serve as valid probe to pursue further understanding of the emerging biology of this interesting metabolic target, potentially removing doubts about presumed off-target modes of action.

Two important classes of PFKFB3 inhibitors, developed by AstraZeneca and displaying an interesting profile and promising properties, have been co-crystallised with the protein. The availability of the 3D crystal structures of the PFKFB3-inhibitor complexes represents an excellent source of information, paving the way for the identification of the key interactions that are responsible for the molecular recognition and bioactivity. Indeed, the identification of drug-enzyme interactions is pivotal for finding new candidate inhibitors with greater target affinity and improved properties.

In this scenario, computational tools, such as molecular modelling and VS, applied to the early stage of the drug discovery process, have demonstrated increasing reliability and prediction capacity for the identification of innovative biologically active molecules. Both ligand-based and structure-based virtual screening of chemical libraries considerably decrease the cost and the length of the hit identification process compared to the traditional expensive method of high-throughput screening.

The main objectives of the present multidisciplinary project are therefore to:

i. Use computational technologies to identify new compounds, or classes of compounds, capable of selectively inhibiting the kinase domain of PFKFB3;

ii. Design and carry out the organic synthesis of the identified compounds;

iii. Evaluate the biological activity of the selected compounds *in vitro*.

3

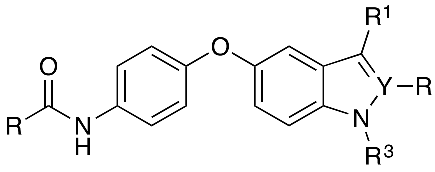
AZI-Based Rational Design

3 AZI-Based Rational Design

3.1 Overview

The phenoxy-indole and phenoxy-indazole series (AZI compounds) developed by the pharmaceutical company AstraZeneca, are a class of potent and selective kinase inhibitors of the PFKFB3 isoenzyme.[68] At the beginning of this project, only a small number of reports on PFKFB3 kinase inhibitors had been presented in the literature, with the newly published AZI series being the ones showing the highest potency in both isolated enzymatic assays and cellular assays. As described above (**paragraph** 1.1.3.1), these compounds are ATP competitive ligands, since they non-covalently bind at the ATP binding site. The structure of these compounds has already been optimised through a structure-activity relationship (SAR) study that led to the identification of the phenoxy-indazole AZ33 (compound **16**, **Table 1**), as the most potent compound, showing a single-digit nM IC_{50} (half maximal inhibitory concentration).[68] In addition, the isoform selectivity of these compounds has been assessed *in vitro*, with kinase assays performed on the isoforms PFKFB1 and PFKFB2. In addition, six crystallographic structures of PFKFB3 in complex with representative inhibitors of the AZI series (**Table 1**) have been elucidated and made available in the Protein Data Bank, as experimental evidence of their binding to the enzyme. From this early scenario, came the idea for this project to develop a virtual screen based on these compounds.

The computational search was carried out taking advantage of the knowledge of the 3D structures of both the ligands and the receptor target, derived from the available X-ray data of the complexes. A library containing over 1 million drug-like compounds, whose characteristics are described in **paragraph 3.2** was explored. The workflow of this study involved a tiered screening where a ligand-based approach preceded a structure-based approach, therefore representing overall a hierarchical strategy procedure that can be divided into three levels of filtration, as depicted in **Figure 5**.



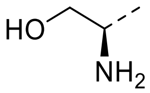
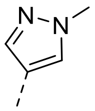
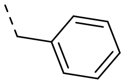
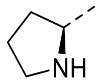
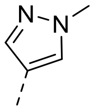
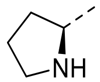
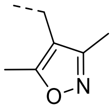
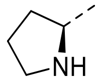
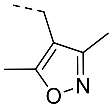
Entry	Compound	Y	R	R^1	R^2	R^3	Reported IC_{50} (µM)[68]
5ajz	**10**	C	CH_2NH_2	CN	H	H	2.01
5ajv	**11**	C		CN	NH_2	H	0.575
5ajy	**12**	C	CH_2NH_2		H	H	0.496
5ajw	**13**	C	CH_2NH_2	CN	NH_2		0.104
5ak0	**14**	C			H	Me	0.021
5ajx	**15**	C		CN	H		0.011
AZ33	**16**	N		CN	H		0.003

Table 1: Chemical structures and corresponding reported IC_{50} values of representative examples from the AZI class of ligands. Compounds **10-15** have been co-crystallised with the PFKFB3 isoenzyme (PDB IDs are shown), while compound **16** (reported as AZ33, according to reference [68]), represents the most potent inhibitor of this series.

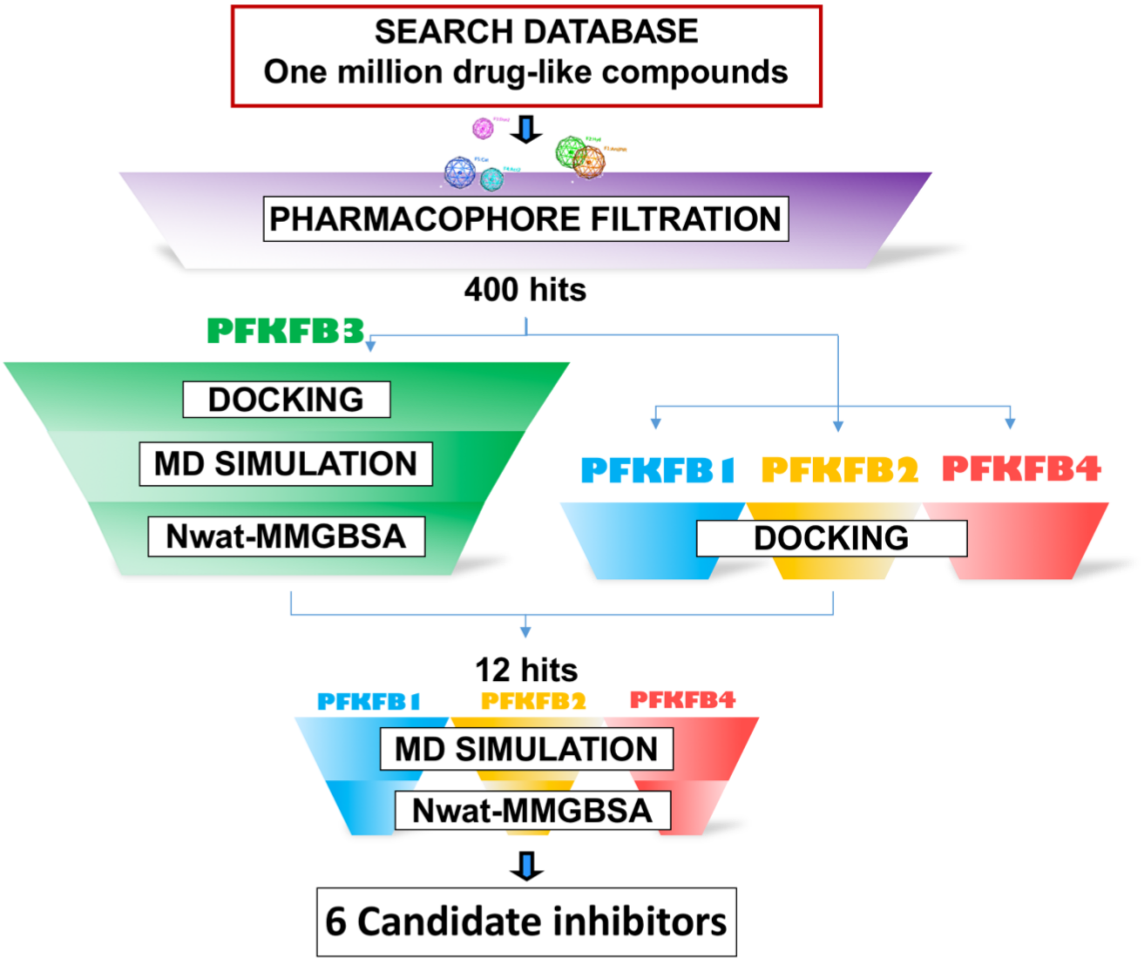


Figure 5: The overall computational strategy workflow of the tiered virtual screening that was based on the AZI compounds.

In the first level, a 3D pharmacophore model featuring the ligand structural elements deemed to be essential for the kinase inhibition, was used for a preliminarily filtration of the search database (DB) (**paragraph 3.3**). In the second level, the resulting sieved subset of 400 molecules that satisfied the pharmacophore map was submitted to structure-based molecular docking using a monomeric model of PFKFB3 (**paragraph 3.4**). Both the pharmacophore filtration and docking method were validated using a collection of theoretically inactive ligands (decoys). The docking poses were refined with short Molecular Dynamic (MD) simulations, and a modified version of the MM-GBSA method (termed Nwat-MMGBSA) was used to calculate the binding free energy and to re-rank the molecules accordingly (**paragraph 3.5**). Another aspect that was estimated *in silico* was the ligand selectivity over the other isoforms of the PFKFB family. For this purpose, the filtered library of compounds that satisfied the pharmacophore query were also re-submitted to molecular docking using monomeric models of PFKFB1, 2 and 4 isoforms. In the third level, twelve selected potential hits were re-evaluated for their binding affinity towards these isoforms, *via* Nwat-MMGBSA analysis of the respective MD trajectories.

In conclusion, the final triage of the compounds was made basing on a combination of predicted receptor affinity and selectivity, synthetic feasibility and commercial availability (**paragraph 3.6**). The selected six chemically diverse small molecules, shown in **Figure 14**, formed the pool of first-generation candidates that were pursued in further experimental tests.

3.2 Molecular Databases

3.2.1 Search database

The combination of two chemical libraries was explored. The first is an unpacked molecular database (DB), previously prepared by Contini's research group;[128] the second DB is a packed molecular DB provided within the software MOE.[98]

The first library consists of a 3D conformational database of about 1 million compounds derived from the public library of commercially available compounds, the ZINC database.[129] It contains chemically diverse small molecules covering of a wide part of the chemical space (redundancies were removed by filtration) and being suitable for future optimisation (i.e. the included molecules comply with the Oprea's test for lead-likeness). The compounds in the library are clustered on a similarity basis (Tanimoto coefficient, T_c = 0.8) and each cluster is represented by the molecule with the lowest molecular weight (MW). The DB also includes multi-conformational information for each small molecule previously generated using high-throughput fragment-based conformational search.

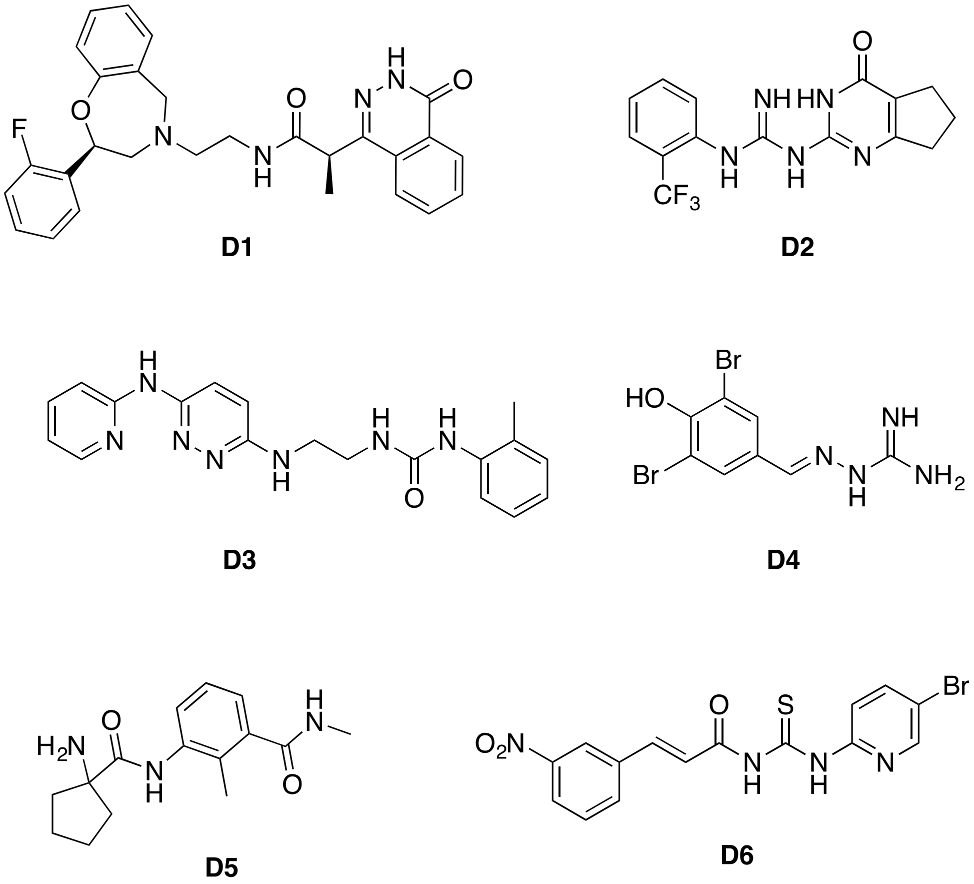
In addition, the DB supplied with MOE contains over 2000 biologically active molecules for a variety of targets. This library is specifically dedicated to pharmacophore searching applications and provides full conformational sampling, so that each molecule is packed with its associated multiple conformers.

The molecular database obtained by merging the two libraries, constituted the search database for the present virtual screening.

3.2.2 Test databases preparation

As explained in **paragraph 1.2.1.2**, in order to validate the VS method and assess its ability to discriminate between biologically active and inactive compounds, a suitable test DB, containing a set of biologically evaluated ligands, was required. Obviously, for the evaluation of VS methods, biological data must be available. Nonetheless, due to the lack of published negative reports about PFKFB3 non-binders, a framework of theoretically inactive ligands (decoys) had to be generated computationally so as to be able to test both the pharmacophore and docking method used herein. For this purpose, the online platform DUD•E (A Data of Useful Decoys: Enhanced) was employed.[85,130] Decoys generally are hypothetical structures that have not yet been tested experimentally but that are unlikely to bind to a given target; thus they are highly likely to be inactive. In particular, the decoys are generated such that they resemble the active ligands in terms of their physicochemical properties but display dissimilar 2D topologies.

Given a list of active ligands, the DUD•E server automatically generates a collection of decoys. The six co-crystallised AZI ligands, were used as the input ligand dataset. Each active ligand gave rise to 50 matched decoys, giving a total of 300 theoretically non-actives. The created list of compounds was processed to convert the 1D decoys into their 3D low energy conformers, with acceptable bond length and angles. In addition, hydrogen atoms were added, rebalancing strong acids and strong bases in their more probable protonation states in the aqueous environment, followed by a minimisation to find a low energy conformation close to the starting geometry. Representative examples of decoys, compounds **D1-6**, are depicted in **Figure 6**.



D1 **D2**

D3 **D4**

D5 **D6**

Figure 6: Chemical structures of six examples of decoys that were automatically generated through dissimilarity filter from the AZI chemical structures.

The library obtained by merging the six crystallised AZI inhibitor 3D structures and the 300 decoys constituted the test-DB that was used for theoretical validation of the docking method (**paragraph 3.4.1**).

In order to validate the pharmacophore model, whereas, an additional 3D conformational sampling of the decoys in the test database was needed. A tool provided by the software MOE was employed to divide the molecules into their fragments and compute the stochastic conformational search independently on each fragment before re-assembling them in the database in their molecule conformations.

The library obtained by merging the AZI inhibitor 3D structures, preserved in their crystallographic bound conformations, and the conformational sampled decoys, constituted the 3D-conformational-test-DB that was used for theoretical validation of the pharmacophore model (**paragraph** 3.3.1).

3.3 Pharmacophore Model

The pharmacophore model used in this work was designed basing on the AZI crystallised ligands, with the aid of the computational application 'Pharmacophore Elucidator' provided by MOE. This tool generates a collection of pharmacophore hypotheses from a given set of input ligand conformers, associated (if known) with their activity information. The methodology of the application involves the flexible alignment of the ligands (scored on the basis of their superimposability, i.e. their 'Overlap') and the determination of their shared chemical features.

As outlined in **paragraph 1.2.3**, in general, to ensure quality and good discriminatory power within a pharmacophore model, it is recommended that the template ligands employed for its elucidation are highly potent, and possibly chemically diverse, inhibitors. As stated previously, when we started the research, the AZI ligands have been proven to be the most potent and selective PFKFB3 inhibitors, hence it was mandatory to use this series of compounds regardless of their chemical similarity. In fact, it should be mentioned that when the training set of active ligands displays a little diversity, such as within the AZI ligands, an intrinsic risk arising from using the 'Pharmacophore Elucidator' tool is that some identified pharmacophore features may be redundant. In particular, the initial ligand alignment of the application will clearly identify many common features and geometries, nonetheless not necessarily all of them might be connected to the ligand activity. On the other hand, however, trying to overlap inhibitors displaying high chemical diversity would be meaningless if they bind the receptor in completely different modes or even at alternative binding sites. As demonstrated from the crystal structures, AZI ligands bind PFKFB3 at its ATP binding site, whereas other reported compounds were either known to bind at the F6P binding pocket[58] or their binding site was never experimentally located and/or confirmed.[63,65–67] Moreover, it should be also noted that other families of inhibitors have been disclosed[61,62] and/or evaluated[72] after the beginning of this project. Bearing these reasons in mind, it was reasonable to use only AZI ligands as a template for the pharmacophore design rather than including every PFKFB3 inhibitor present in the literature, since their inclusion would have misled the result of the search, potentially giving an inaccurate pharmacophore model.

Nevertheless, the risk of redundancy due to the evident similarity within the crystal AZI ligands, could have been to some extent compensated by the availability of the atomic details of the ligand-receptor complex structures. In fact, the additional knowledge of the location and directionality of the interactions in the binding pocket constituted valuable information used for the pharmacophore design and especially its subsequent validation. In particular, as it will be described in the following **paragraph 3.3.1**, once the collection of pharmacophore hypotheses was obtained from the MOE elucidator tool, two consecutive approaches were adopted to choose the most representative query:

i) a systematic assessment of the enrichment gained using each query model as a molecular filter of the 3D-conformational-test-DB; and;

ii) a careful visual comparison of the most restrictive queries with the chemical/structural details of the X-ray protein-ligand complexes.

As a result, the elucidation of the elucidation of the pharmacophore used for this study (**Figure 7**) can be considered to be not only ligand-based but also partially structure-based.

3.3.1 Pharmacophore Validation

To select the most representative pharmacophore, the automatically generated set of query hypotheses with best superimposability were assessed by running the ‘Pharmacophore Search’ filtering tool of the MOE software, followed by a visual inspection inside the binding site of PFKFB3.

The search algorithm of the 'Pharmacophore Search' aligns all the molecules of a given DB to a query, and outputs a filtered database of those ligands that matched the query as 'Hits'. The dedicated 3D-conformational-test-DB of active/non-active ligands (**paragraph 3.2.2**) was used as the input database, and the search was thoroughly repeated for each one of the computed queries. A fulfilment of all features of each query was required for a compound conformation to pass the filter. Only those pharmacophore queries that reported no more than five decoys (false positive, FP ≤ 5) in the output database (Specificity ≥ 0.98, **Equation 3**), showing thereby a good enrichment, were subsequently reviewed. One restrictive pharmacophore model (**Figure 7**) retrieved only two decoys in addition to the six true ligands (Sensitivity = 1, **Equation 2**; Specificity = 0.99, **Equation 3**).

A visual analysis reflecting on the structural details of the protein active site followed the enrichment assessment, so as to select only the query bearing features that explicitly mimicked real ligand-protein interactions that would most likely trigger the desired pharmacological effect. Query in **Figure 7** showed consistency and complementarity with the ligand-protein interactions experimentally observed in the X-ray structures. Indeed, from the visual inspection, each pharmacophoric feature was situated in correspondence of interacting functional groups inside the pocket.

On the other hand, we discarded other pharmacophoric hypotheses even though they gave good enrichment and high specificity in the decoys filtration, because their annotated features mimicked functional groups actually not involved in any interactions with the receptor, thus doubtfully related to activity or receptor recognition. For example, we dropped queries containing features corresponding to the cyano functional group at position-2 of the indole ring of some AZI compounds. Similarly, we discarded those queries with features mimicking the phenoxy portion of the AZI scaffold. In fact, as previously reported in the AZ paper[68] and then confirmed by our inspection of the X-ray complexes, those moieties of the molecules do not appear to interact with any residue of the kinase pocket.

Hence, eventually, the 5-features pharmacophore shown in **Figure 7** was selected and used for the pharmacophore-based virtual screening (filtration) of the search database (**paragraph** 3.3.3).

3.3.2 Pharmacophore Description

Figure 7 illustrates the selected pharmacophore query, superimposed onto the aligned AZI ligands with the relative distances between features. Each feature of the pharmacophore map annotates the electronic and steric nature and location of specific functional groups that were expected to be mandatory for the binding. The features are represented by sphere specified as a location in space, a tolerance radius and the annotation type. The pharmacophore model selected for the present study comprises the following feature types:

(a),(b) A hydrophobic centroid (*Hyd*) and an aromatic ring or a general π-system (*Aro|PiR*) required to mimic both the adenosine moiety of ADP/ATP and the indole/indazole ring of the AZI compounds. In the crystal structures, these fragments occupy a lipophilic cleft within the PFKFB3 kinase pocket and make an arene-H interaction with residues Val217 and Gly46 of the active site;

(c) A projected location of the H-bond acceptor (*Don2*) to annotate the implicit direction of the H-bond (hydrogen bond) of the ligand with the backbone carbonyl oxygen of residue Leu238 of the active site;

(d) A positively ionisable feature (*Cat*) for representing the basic group of the amine portion of the AZI that makes a solvent contact with a conserved water molecule of the crystal structures and a H-Bond with backbone carbonyl oxygen of residue Ile241;

(e) A projected location of H-bond donor (*Acc2*) corresponding to the implicit direction of another H-bond with conserved water molecules bridging between the backbone amidic hydrogen of residue Val243.

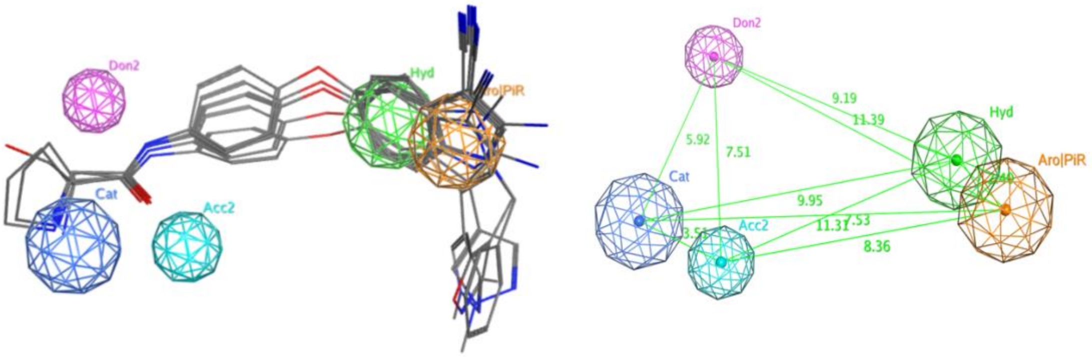
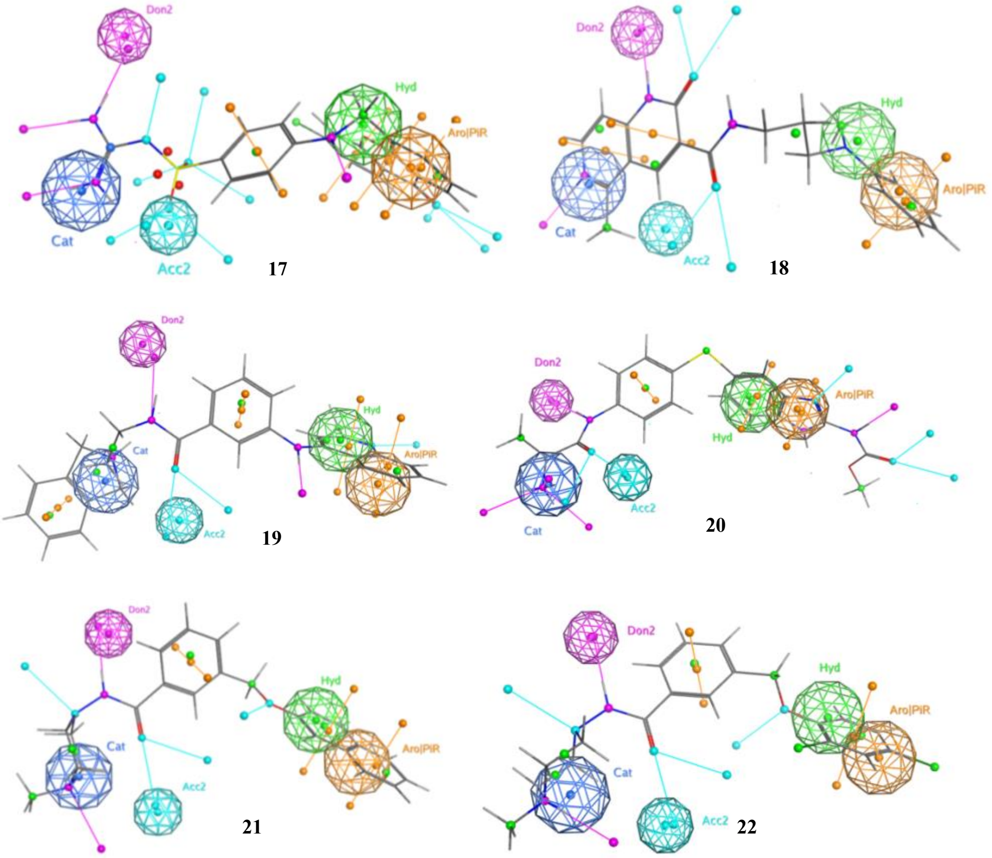


Figure 7: Graphical representation of the validated pharmacophore model. (*left*) 3D overlay of the pharmacophore map (shown in coloured spheres) onto the aligned AZI inhibitors (shown in green sticks); (*right*) relative distances between features. The model consists of: a hydrophobic centroid (*Hyd*, average radius 1.4 Å); a general aromatic ring (*Aro|PiR,* average radius 1.4 Å); a projected location of H-bond acceptor (*Don2*, average radius 1.0 Å); a cationic ionisable heavy atom (*Cat*, average radius 1.4 Å); a projected location of H-bond donor (*Acc2*, average radius 1.0 Å).

3.3.3 Pharmacophore Filtration

The validated pharmacophore model (**Figure 7**) was used for a preliminary filtration of the search DB (described in **paragraph 3.2.1**). For this purpose, the MOE ‘Pharmacophore Search’ filtering tool was applied. The search algorithm aligns all the molecules of the DB to the given pharmacophore query using a rigid-body superimposition and then reports only those ligands that satisfy the query constraint. To guarantee specificity, as was done in the pharmacophore validation phase (**paragraph 3.3.1**), all the five features of the query were required to be matched at the same time for a compound conformation to pass the filter (‘*systematic matching*’ in MOE terminology). This operation allowed to narrow down the original search DB to 380 hit compounds fitting the pharmacophore map.

Figure 8 shows 3D overlays of one representative conformation of compounds that matched the pharmacophore map and that were later selected for the experimental validation, **17**-**22** (2D chemical structures are shown in **Figure 14**).



17 18 19 20 21 22

Figure 8: Overlays of the pharmacophore with one representative matching conformation of compounds (from top-left to bottom-right corners) **17**, **18**, **19**, **20**, **21** and **22**, examples of molecules that were retrieved in the filtered search database.

All unique lowest energy conformations of every hit molecule were merged together, along with the six crystal ligands, to give a total of 386 ligands that were successively submitted to molecular docking for a SBVS.

3.4 Molecular Docking

The filtrated library of compounds fitting the pharmacophore map were submitted to molecular docking calculation. For comparison purposes, AZI crystallised ligands were also submitted with this set of compounds to the same protocol.

In order to avoid misleading or limited information that might come by relying on a single PDB structure, cross-docking was also performed. Thus, all six structures extracted from crystals of PFKFB3 in complex with the AZI ligands, along with another available PFKFB3 crystal structure (PDB ID: 2I1V) were used to model a 3D monomeric structure of PFKFB3, for the purposes of the molecular docking calculation.

Preliminary cleaning and correction of all PDB issues of every receptor structure was performed through common operations using MOE software. In particular, hydrogen atoms were added and adjusted to their optimised position, then uncertain ionisation states of polar atoms were fixed basing on physiological conditions. Furthermore, missing atomic data or ambiguous sequence identities were modelled or corrected. The system was then minimised according to the AMBER10:EHT[131,132] force field and the Generalised Born implicit solvation model.

In addition, the 3D models of PFKFB1, 2 and 4 were used for molecular docking in order to theoretically assess the potential isoforms selectivity of the ligands.

> *The monomeric models of the isoforms PFKFB1, 2 and 4 have been prepared by Francesca Stanzione (Cambridge Crystallographic Data Centre), in respect of a collaborative work with our research group.*

Since no human crystallographic structures were available in the Protein Data Bank, the 3D structural models of the isoform PFKFB1, 2 and 4 have been built *in silico* by homology modelling, basing on their known amino acid sequences. The monomeric form of PFKFB3 extracted from the crystal structure (PDB ID: 5AJX) was used as the template for the construction of the isoform models, which was undertaken using the software Modeller.[133]

The docking program employed for the structure-based screening was PLANTS (Protein-Ligand ANT System).[110,111] For this work the receptor was kept rigid (rigid docking), since we postulated that a potential new drug would bind to a reasonably similar protein conformation. This assumption was supported by evidence from previous studies suggesting that the bound state of the active pocket (which is itself rather rigid) causes no significant conformational changes to the enzyme.[45] Moreover, when a comparison was performed by superimposing every available 3D *holo form* structure of PFKFB3 binding different ligands, only slight residue movements were evident, mostly situated in peripheral regions of the protein (i.e. the β-hairpin autoregulatory domain), far away from the binding site. Furthermore, as already mentioned in **paragraph 1.2.4**, because of its complexity, soft docking is quite computationally demanding and seldom performs better than rigid docking.[104]

All tautomers, stereoisomers and ring conformations were generated using a fully automated tool and, after the docking process, only geometries with the lowest energy conformation were ranked and considered for a visual inspection.

The docking method was validated using the test-DB described above (**paragraph 3.2.2**) and the receptor model obtained from the 5AJX crystal structure. The ROC curve was generated and the corresponding area under the curve (AUC) and Enrichment Factor (EF) values were calculated (**Figure 9**).

3.4.1 Results and discussion

The Enrichment Factor of test-DB molecular docking was calculated (EF, **Equation 1**) for the ligands ranked in the top 1%, 10% and 20% score, obtaining, respectively: 37.5, 8.1 and 4.9. This result means that by sampling the top 1% (or 10% or 20%) of the ranked database, the probability of finding an active compound is 37.5 (or 8.1 or 4.9) times higher than the one associated with a random selection of 3 (or 30 or 60) molecules from the database.

Figure 9 shows the Receiver Operating Characteristic curve (ROC), which plots the true positive rate against the false positive rate. The value of the area under the curve (AUC) is 0.96, indicating that the docking method has good capacity to discriminate between true actives against the decoys.

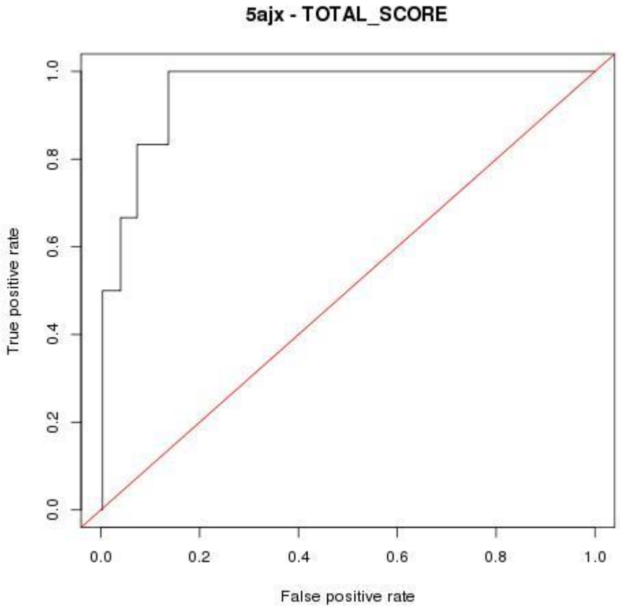


Figure 9: Enrichment assessment of the molecular docking protocol using the test-DB of active/non-active ligands. The AUC relative to ROC curve plot (black line) is 0.96. The reference red line indicates no enrichment, corresponding to a random selection; The Enrichment Factor of 1% top-ranked ligands is 37.5; Enrichment Factor of 10% top-ranked ligands is 8.1; Enrichment Factor of 20 % top-ranked ligands is 4.9.

Compound	**10**	**11**	**12**	**13**	**14**	**15**
RMSD value (Å)	1.0	1.0	1.1	2.1	1.2	0.7

Table 2: RMSD values associated with the deviation of the poses of AZI compounds **10**, **11**, **12**, **13**, **14** and **15**, predicted by molecular docking and the corresponding crystallographic poses.

A visual inspection after every docking and cross-docking experiments showed that the AZI ligands were docked correctly. In fact, a good overlay was observed from the superimposition of docked and crystallised ligands (**Figure 10**). The root mean square deviations (RMSD) associated with the deviation between the docked and the crystal binding modes are reported in

Table 2**.** Moreover, the overall trend of the docking scores of AZI ligands has good correspondence to the reported experimental activity data.

Figure 11 depicts the docking poses of two representative examples of the compounds of the search DB that were later selected for the experimental validation, molecules **17** and **21** (2D chemical structures are shown in **Figure 14**). The xanthenone moiety of **17** is predicted to be tightly bound by an H-bond between the backbone amidic proton of the Asn163 and the carbonyl oxygen of the ligand, as well as through a network of π-interaction with the sidechain protons of the residue Val217. This arene-H interaction with Val217 is a conserved interaction that is also observed in the ADP-receptor crystal structure with the adenine ring of ADP. The sulphaguanidine tail of **17** is presumed to be stretched out towards the opening of the binding cavity, providing additional contacts with both backbone and sidechain acceptor residues of the enzyme through several H-bonds.

Other ligands showed a binding mode strikingly close to the one observed in AZI crystal structures, almost perfectly fitting the AZI pharmacophore and making similar interactions. Docking binding mode of compound **21**, is shown in **Figure 11**. The naphthyl ring of **21** overlays the lipophilic cleft, sitting coplanar to the superimposed indole portion of AZI **15**. The central aromatic ring of **21**, meanwhile, occupies the phenoxy moiety pocket of AZI compounds and displaces the water molecules which are seen when ADP binds. The amide group interacts *via* an H-bond with the backbone of Leu238, similar to what has been observed in the crystal structures of AZI. These key interactions are common binding features observed also in other docked ligand-receptor complexes of this study.

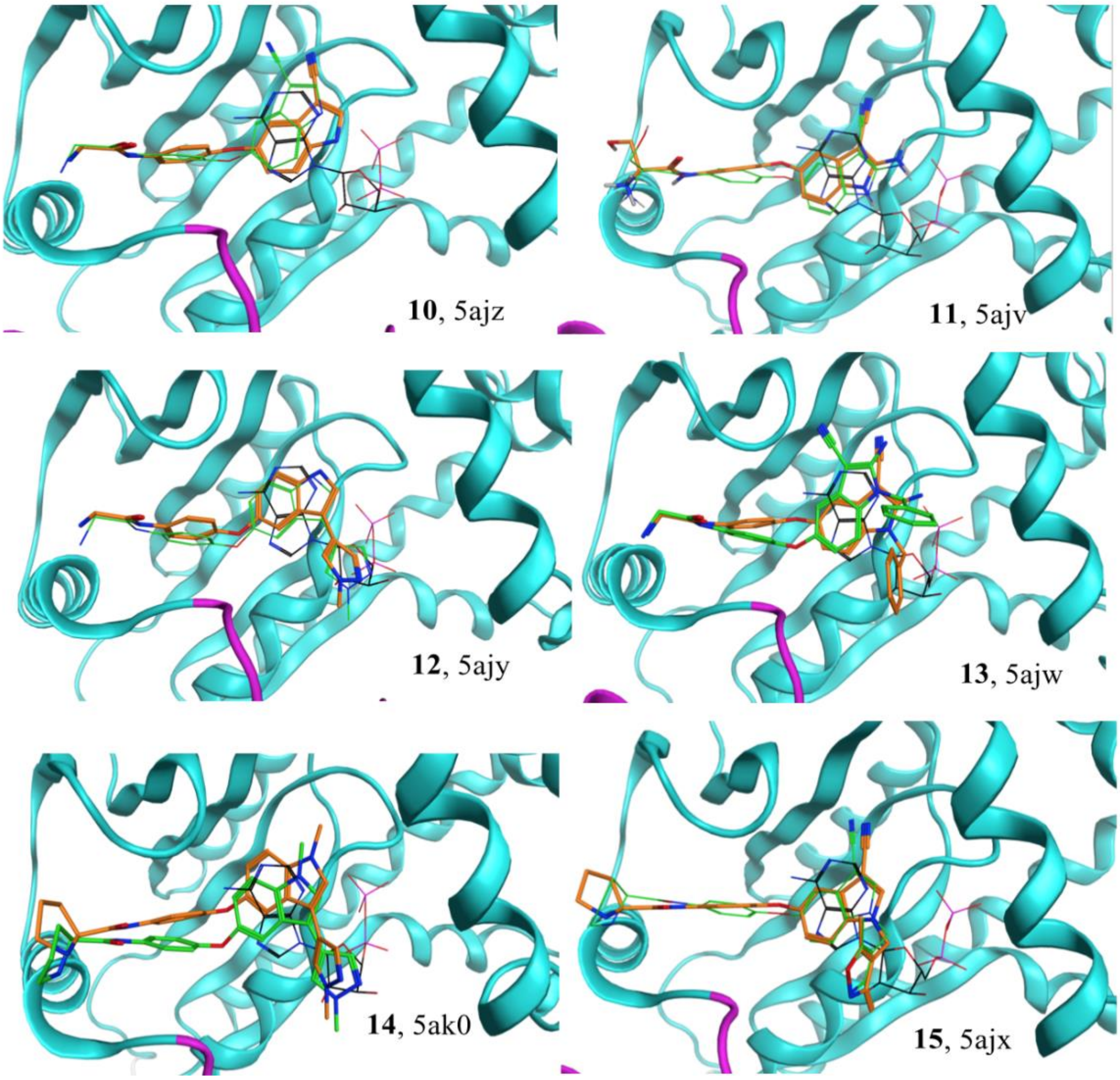


Figure 10: Comparison of the docking AZI ligands pose prediction (shown in green sticks) and the corresponding crystallographic poses (shown in orange sticks), by superimposition into the kinase catalytic domain of PFKFB3. From top-left to bottom-right corners: ligands **10**, **11**, **12**, **13**, **14** and **15**. Crystal structures are labelled with their PDB IDs. The natural ligand, ADP, is also in site in its crystallographic pose (shown in black sticks).

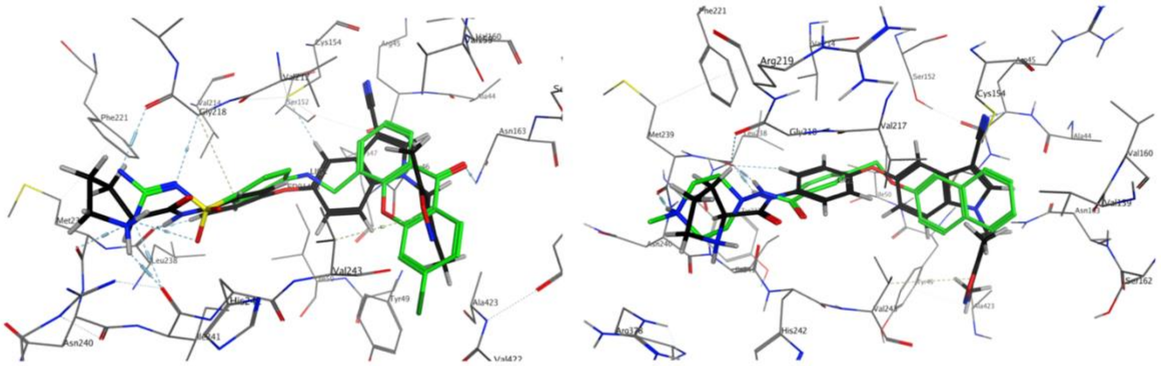


Figure 11: (*left*) Comparison by superimposition of the docking pose of ligand **17** (shown in green sticks) with the crystallographic pose of AZI **15** (shown in black sticks), into the kinase catalytic domain of PFKFB3 (PDB ID: 5AJX); (*right*) Comparison by superimposition of the docking pose of ligand **21** (shown in green sticks) with the crystallographic pose of AZI **15** (shown in black sticks), into the kinase catalytic domain of PFKFB3 (PDB ID: 5AJX).

3.5 Molecular Dynamic (MD) simulations and Nwat-MMGBSA calculation

After the docking process, compounds in complex with the receptor were subjected to molecular dynamic (MD) simulation. Among all the complexes generated through docking, those obtained using the 5AJX as the receptor model were arbitrarily chosen to be submitted to the MD simulations. Additionally, for comparison purposes, both docking-generated and crystallographic complexes of the AZI ligands were simulated through an identical protocol.

The aims of the MD simulation were to: i) carry out a refinement of the ligand poses generated by molecular docking and thereby assess the stability of the complexes ii) use the production trajectories to calculate binding affinities and re-score the ligands accordingly.

Considering the high computational expenses demanded by MD screening, and also bearing in mind the purposes of the simulations, only a narrow time-span of 1 ns production was simulated. Also, as it has been statistically assessed in previous studies, MDs production of 1 ns are considered sufficiently long for virtual screenings and rescoring purposes.[117,125] The MD simulations were run according to the protocol reported by Maffucci *et al*.[125,128]

The MD trajectories were analysed by evaluation of the root mean square deviation (RMSD) of i) the atoms of the ligands, ii) the backbone C-α of the receptor and iii) the backbone C-α of residues defining the kinase pocket, so as to measure the fluctuations during the simulation and to assess the system stability.

The binding energies were estimated, from a specific number of collected frames of MD trajectories, through a proprietary version of the MM-GBSA method, namely the Nwat-MMGBSA method.[124–126] As anticipated in **paragraph 1.2.6**, since the MM-GBSA (or MM-PBSA, if the Poisson Boltzmann equation is applied) calculation approximates the solvent to a continuum environment around the solute, it somehow underestimate the presence of eventually important water-mediated H-bond bridges between ligand and the protein. The Nwat-MMGBSA method developed by Maffucci *et al.*,[124–126] however, ameliorates the solvation term of the MM-GBSA equation (**Equation 7**) by adding the effects of binding-site water molecules. In particular, Nwat-MMGBSA combines an established number of water molecules (Nwat) taken from the MD snapshots, with the GB (or PB) continuous solvation model. Consequently, explicit water residues are considered in the ligand affinity calculation as part of the receptor. The water molecules automatically selected for the calculation are those closest to the ligand throughout each frame of the MD simulation. Since they are located around the ligand, the explicit water molecules considered in this calculation are the ones that are most likely to be involved in the receptor-ligand interaction.

It has been shown that this method provides better predictability than the classic MM-GBSA method with a negligible addition of computational cost.[124–126] As previously stated by Maffucci *et al.*, the extent of improvements achieved by adding explicit water can vary depending on the system, and in particular on the role of water in the ligand-receptor interaction.[124–126] As a consequence, the optimal value of Nwat also varies with the system, but usually Nwat > 60 gave worse results, besides the additional computational cost. Nevertheless, it has been reported that an average of Nwat ≈ 30 can provide a good trade-off.[125] In the present work, to explore a broad range of possibilities, the calculation was performed for each ligand in three different conditions, i.e. using 0, 30 and 60 number of waters in the hydration shell as part of the receptor (Nwat = 0, 30, 60). In order to assess the reproducibility of the protocol, molecular docking, MD simulation and Nwat-MMGBSA were performed in duplicate runs.

In order to evaluate the correlation between the predicted values of the binding free energies (ΔG_{bind}) and the reported experimental data for the inhibition activity of AZI ligands (-log10(IC_{50})), the Pearson's coefficients (r^2) were calculated for each of the three Nwat-MMGBSA calculations. In addition, since AZI ligands were simulated starting from both their crystallographic and their docked geometries, the two outcomes were compared: the ΔG_{bind} corresponding to the MD simulation of the docked geometries of AZI complexes are hereinafter referred to as *dock*-Nwat-MMGBSA; while the ΔG_{bind} corresponding to the MD simulation of the crystallographic geometries of AZI complexes are hereinafter referred to as *crys*-Nwat-MMGBSA (**Figure 12** and **Figure 13**).

In addition, in order to theoretically assess the isoform selectivity of the most promising ligands (12 compounds), 1 ns MD simulation, followed by Nwat-MMGBSA analysis, were also performed using 3D models of PFKFB1, 2 and 4 enzymes.

3.5.1 Results and discussion

Correlations between calculated and experimental bioactivity data were evaluated by the Pearson's coefficient r^2, as reported in the plots in **Figure 12** and **Figure 13**. Higher r^2 were obtained when the crystallographic poses of the AZI ligands were used as the starting frame for the MD simulation (*crys*-Nwat-MMGBSA), in comparison to the corresponding *dock*-Nwat-MMGBSA correlation values (r^2 of *crys*-Nwat-MMGBSA > r^2 of *dock*-Nwat-MMGBSA, for each Nwat values). Overall, however, both results provided satisfactory reproduction of the experimental trend.

The binding energies calculated with *crys*-Nwat-MMGBSA gave the best correlation (r^2 = 0.89) picking zero (Nwat = 0) as the number of molecules in the explicit hydration shell (**Figure 13**). On the other hand, when performing the simulation of the poses of docking, the best result ($r^2 = 0.61$) was obtained with Nwat = 60 (**Figure 12**). The overall accordance between the predicted affinity and the experimental activity data supported the validity of the computational method, since satisfactory correlation values r^2 were obtained for each calculation.

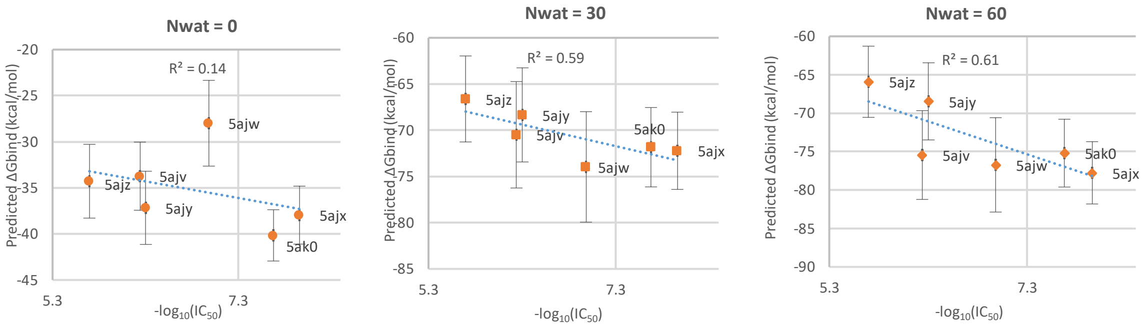


Figure 12: Correlations between the experimental $-\log_{10}(IC_{50})$ values**[68]** and the calculated binding energies (ΔG_{bind}) of AZI ligands (labelled according to their PDB IDs). ΔG_{bind} were obtained from MD trajectories of the AZI ligands simulated for 1 ns in their docked poses (*dock*-Nwat-MMGBSA). The number of water residues (Nwat) explicitly accounted for in the calculation were: (from left to right) 0, 30 and 60.

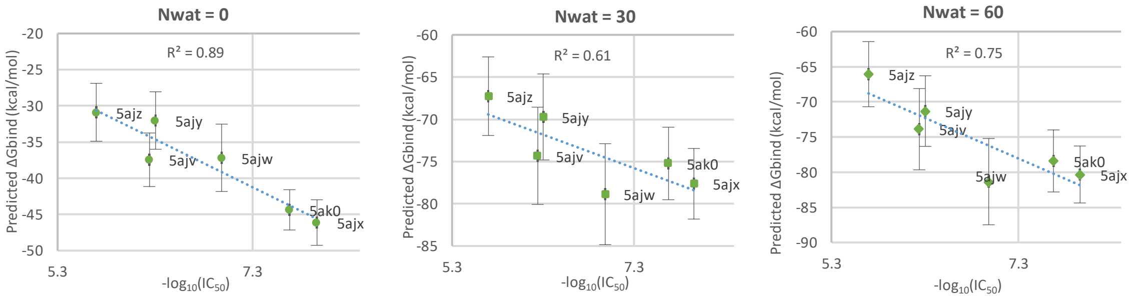


Figure 13: Correlations between the experimental $-\log_{10}(IC_{50})$ values**[68]** and the calculated binding energies (ΔG_{bind}) of AZI ligands (labelled according to their PDB IDs). ΔG_{bind} were obtained from the MD trajectories of the AZI ligands simulated for 1 ns in their crystallised poses (*crys*-Nwat-MMGBSA). The number of water residues (Nwat) explicitly accounted for in the calculation were: (from left to right) 0, 30 and 60.

3.6 Selection of ligands

In order to prioritise good candidate molecules to be experimentally evaluated *in vitro* for their bioactivity, all the docking results (seven calculations and one repeated run, with a total of eight docking scores for each compound) were combined with the calculated ΔG_{bind} data (0, 30 and 60-MMGBSA per repeat run, with a total of six ΔG_{bind} values for each compounds) and a consensus criterion of selection was adopted. This procedure made it possible to determine the extent of agreement between the results obtained from different calculation techniques and runs. This was also intended to balance any errors arising within each single calculation and thus to improve the probability of identifying 'true' ligands.

The 30th percentile of the predicted affinity for the Nwat-MMGBSA calculations were imposed as thresholds to mark compounds as potential good candidates (where a potential good candidate has a predicted ΔG_{bind} < threshold). The consensus approach was thus aimed at evaluating the frequency that a compound appears as potential good candidate. In particular, only ligands with Nwat-MMGBSA "full consensus" (i.e. ligands with ΔG_{bind} < threshold in both the first and second run for each of the 0, 30 and 60-MMGBSA) were kept in a reduced DB, leaving a total of ten ligands. Also, five out of six AZI true active ligands showed "Nwa-MMGBSA full consensus", which supported the validity of the method.

Likewise, the 35th percentile was defined as the threshold for the docking scores calculated with PFKFB3 receptor. The choice of a slightly less restrictive threshold than the one assigned for Nwat-MMGBSA, was due to the intrinsic lower accuracy of docking calculation. Indeed, a docking score "full consensus" retained four out of six AZI true actives.

From the set of ten "Nwat-MMGBSA-full consensus" ligands, four were subsequently excluded since they were suspected to be false positives, judging from the low consensus in the docking scores (i.e. the corresponding docking scores for the PFKFB3 isoform did not reach the 35th-percentile threshold in at least three out of eight docking reiterations). The remaining six molecules were predicted to have high inhibitory activity, as characterised by the predicted binding free energy and docking score values.

It is worth noting that the initial strategy of selection was instead on the basis of a “partial consensus” with respect to the Nwat-MMGBSA calculations. Hence, at first, the rejection was applied in a relative rather than categorical way. Specifically, a predicted ΔG_{bind} < threshold for at least four out of six ΔG_{bind} values was required for a compound to be picked for further *in vitro* tests. This led to roughly 60 molecules being marked as potential active and selected for experimental assays. It was discovered, however, that unfortunately, only a very small number of molecules were still commercial, and the majority were either depleted or had been removed from the market. Therefore, a theoretical refinement of the accuracy of calculation, and a more restrictive selection scheme were needed, in order to limit the number of compounds considered worth synthesising and testing. In particular, a “full consensus” and further *in silico* evaluation of the ligands selectivity were required to reduce the number of candidates and so to alleviate the otherwise enormous synthetic work. The selectivity towards the other isoforms of the PFKFBs family was taken account of as follows.

A penalty was assigned to non-selective ligands based on the scores obtained from molecular docking into either PFKFB1, PFKFB2 or PFKFB4 isoform models. In particular, ligands whose isoform docking score surpassed the 30th-percentile threshold were penalised as non-selective ligands. Another six ligands were subsequently reconsidered as potential candidates because resulted quite selective versus related isoforms (only one penalty), although still showing relatively good predicted potency (i.e. showing a “partial consensus” with at least five out of eight docking scores higher than the threshold and at least four out of six Nwat-MMGBSA ΔG_{bind} values higher than the threshold).

The twelve compounds retained in this way were submitted to 1 ns MD simulations using PFKFB1, PFKFB2 and PFKFB4 models as the receptor, and trajectories were analysed with the Nwat-MMGBSA method according to the same protocol (Nwat = 0, 30, 60). As a result, three ligands were excluded because they also showed high affinity towards PFKFB2 and PFKFB4 (poor selectivity). Compounds that showed promising metrics were selected for the prospective follow-up tests. However, because of the high homology within the isoforms, it was not universally defined whether a compound would preferentially bind either one of them.

Eventually, six chemically diverse compounds showing a convincing predicted affinity and selectivity profile, were selected. It should also be noted that the selection was also partially influenced by the commercial availability and synthetic feasibility of the molecules. Hence, the final choice of candidates was driven by a trade-off between affinity and selectivity prediction, synthetic feasibility and commercial availability. The six picked compounds, shown in **Figure 14** were either synthesised (compounds **17**-**20**) or purchased (compounds **21** and **22**) in order to test their affinity and biological activity towards the PFKFB3 enzyme in the wet lab.

Paragraphs 5.1 - **5.4** describe the organic syntheses of these AZI-derived target molecules and, **paragraph** 6.1 sets out their biological evaluation.

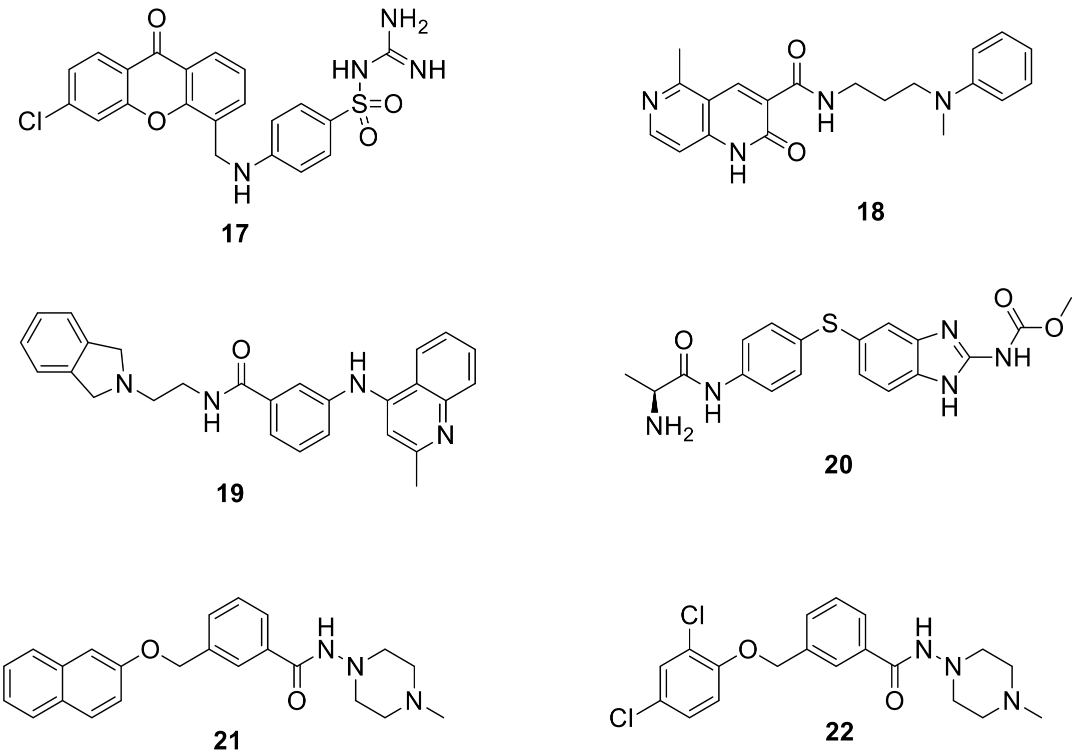


Figure 14: Chemical structures of the chemically diverse small molecules selected from the AZI-based rational design as candidate PFKFB3 inhibitors.

.

4

AZPP-Based Rational Design

4 AZPP-Based Rational Design

4.1 Overview

An alternative strategy for the identification of new PFKFB3 inhibitors was focussed on another class of ligands, the dihydropyrrolopyrimidinone series (AZPP), which were evaluated, after the beginning of this project, by St-Gallay *et al.*.[72] As described above (**Paragraph 1.1.3.1**), although exhibiting quite a low inhibitory activity (low µM concentration range of IC_{50}, being about 10 to 1000-fold less potent than AZI ligands), these compounds display interesting lead-like properties and improvements to them might lead to an increase in potency and selectivity. Because of their low molecular weight (MW), average *c*logP < 4 and modifiable structure, these molecules bear the potential for further optimisation and their scaffold represents a promising starting point to build accurate structure-activity relationship studies (SAR). Moreover, the crystal structures of PFKFB3 bound to four AZPP compounds (**Table 3**) are stored and freely available in the Protein Data Bank (PDB IDs: 4D4K, 4D4J, 4D4M, 4D4L). Despite the low resolution, it was observed from the X-ray structures that AZPPs bind at the PFKFB3 *N*-terminal domain into the ATP pocket of the kinase domain, thereby mimicking the adenosine moiety of ATP. In fact, the active core of this compounds is a fused heterocycle bearing a pyrrole and a pyrimidinone portion. Due to its structural resemblance to pyrimidines and purines, this isosteric scaffold has been the focus of significant interest in medicinal chemistry and a wide range of pharmacological applications have already reviewed,[134] highlighting the remarkable potentiality of these compounds as candidate drugs. Moreover, their synthesis is relatively tractable, allowing the parallel preparation of several analogues using different starting materials that can be bought or prepared cheaply.

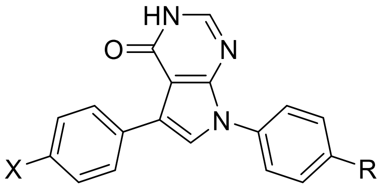
Their optimisable properties and interesting profile prompted us to investigate this series of compounds further, embracing the possibility of exploring the chemical space around them. The new search plan was therefore founded on the synthetic opportunity associated with an exploitation/lead-optimisation process using the crystallised AZPP ligands as new probe compounds.

For this purpose, a new structure-based virtual screen was focussed on a dedicated chemical library, covering the more restricted chemical space around these compounds. **Paragraph 4.2.1** describes the ligand optimisation for the rational design of AZPP derivatives that constituted the new library.

The availability of X-ray structures again facilitated the ligand optimisation process, which was primarily aimed at improving the interaction and filling the pocket. In fact, from the crystal structures the ligand binding appears rather weak, mostly directed by van der Waals interactions. Also, a weak σ-hole interaction is showed when one of the two phenyl rings bears a halogen atom. As a consequence, two flipped binding modes are observed depending on the position of the halogen. Based on that, modifications to the AZPP structure were made by adding functional groups onto the phenyl rings in order to lock the ligands in a single binding mode. The library of analogues bearing the same heterocyclic active scaffold obtained in this way was then submitted to docking virtual screening with the dimeric model of PFKFB3 in order to prioritise the most promising compounds (**paragraph 4.2.2**).

The AZPP analogues **29**-**32**, along with the original AZPP compounds **23**, **27** and **28** shown in **Figure 19** were selected for follow-up synthesis and biological testing with the recombinant PFKFB3 enzyme.

In parallel to this structure-guided ligand optimisation strategy, a similarity search for scaffold hopping was carried out in order to identify new potential binders having physicochemical properties similar to the AZPP series but displaying chemically diverse structures. The molecule **24** (**Table 3**) was used as the query. A web server of the RPBS (Ressource Parisienne en Bioinformatique Structurale) web portal providing different small chemical banks to screen, was employed for the 3D similarity search (**paragraph 4.3**).[135] The hit molecules were docked inside the kinase domain of the dimeric model of PFKFB3, and the most promising compound **33** (**Figure 20**) was synthesised and tested.



Entry	Compound	X	R	Reported IC_{50} (µM)[72]
4d4k	**23**	H	OMe	5.012
4d4j	**24**	Br	H	0.398
4d4m	**25**	H	Br	0.398
4d4l	**26**	Cl	H	0.631

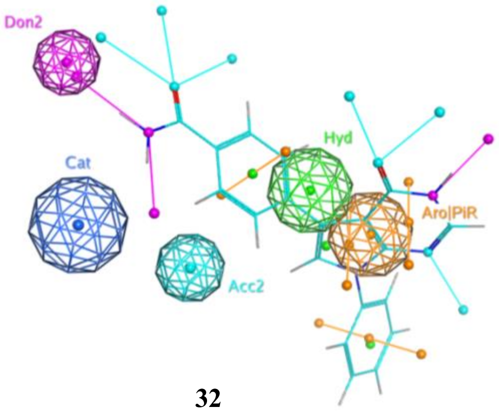
Table 3: Chemical structures and corresponding reported IC_{50} values[72] of the four ligands of the AZPP series that have been co-crystallised with the PFKFB3 isoenzyme (PDB IDs are shown).

4.2 AZZPP Ligand Optimisation

4.2.1 Molecular Database Design

A manual construction based on rational design/synthetic feasibility, combined with automatic generation of analogues, was aimed at building a focused chemical database. This allowed us to collate in a database a broad spectrum of synthesisable derivatives to be screened in order to investigate the chemical space surrounding the AZPP series. Molecules from the new library were drawn in a way that conserved much of the original active core, with small-scale structural changes. In particular, modifications were introduced mainly by replacing and/or functionalising, with a variety of substituents, the phenyl portions attached to the pyrrole moiety. Bioisosteric replacements were also made at the structure whenever possible by switching groups with similar electrochemical properties and preserving the geometry. In addition, in order to satisfy synthetic accessibility needs, compounds amenable to be produced preferably through a divergent synthetic strategy (namely from a common key precursor), were also added in the library.

The MOE 'Database Viewer' was used to build a 3D database of the set of molecules drawn from 2D Sketcher. In addition, the 'MOE builder' tool was employed to further modify the structures of the AZPP in situ, i.e. inside the kinase pocket, so as to check the goodness of fit. As mentioned, the goal of the optimisation was to increase the strength of the ligand-protein interaction, which for the AZPP series are mostly van der Waals rather than electrostatic or polar interaction. In particular, it was hypothesised that the functionalisation of the aromatic rings with judiciously positioned hydrogen bond donor (HBD) substituents, would offer more opportunities for hydrogen bonding to the backbone amide residue Leu238 and would increase binding site occupancy. Moreover, this modification would meet three out of five pharmacophore features of the previously used pharmacophore model. In particular, in addition to the *Hyd*, *Aro|PiR* features matching the central heterocyclic scaffold, the *Don2* features would be also satisfied (**Figure 15**).



32

Figure 15: Overlay of the pharmacophore with one representative conformation matching three out of five features of compound **32**, an example of molecule that was included in the database of AZPP analogues.

In addition, the 'MedChem Transformations' application of MOE software was employed to expedite the process of generating several analogues. To a given input structure, known as an active ligand, the tool applies a set of transformation rules usually made in medicinal chemistry, through a match-and-replace algorithm. Then, the generated molecules were filtered upon their chemical attributes before being output in a 3D database. Each AZPP ligand extracted from the crystal structures (compounds **23** – **26**) gave rise to 37 newly created analogues. The four output databases obtained in this way were merged together with the other manually constructed sub-set of analogues, to give a total of 654 compounds that were then submitted to docking calculation.

4.2.2 Molecular Docking

Molecular docking of the constructed database of AZPP analogues was performed using the PLANTS program, without additional pre-processing, so that no new conformations were searched. Also in this case the receptor was kept rigid for the same reasons discussed in **paragraph 3.4**. For a more realistic receptor model, however, the head-to-head dimeric PFKFB3 structure (modelled using the crystal structure PDB ID: 2AXN) was used as the protein for the virtual screen.

The dimeric model the PFKFB3 enzyme has been prepared by Xiao Hu, Early Stage Researcher of the Moglynet Project (University of Milan, Italy).

The PFKFB3 endogenous substrates were kept in place during docking, so that the bisphosphatase catalytic domain contained F2,6P molecules (in of both chains of the dimer), while ATP and F6P occupied only one kinase pocket of the dimer. Docking was focussed on the whole empty kinase site, defined by both ATP and F6P pockets. With an approach similar to a blind docking, molecules from the database were evaluated exploring both pockets and, since only the most favourable mode was chosen, it was ultimately defined, in this way, which pocket each ligand would preferably bind.

For comparison purposes, with our library of analogues, the 21 published AZPP ligands (comprising the compounds **23** – **26**) were also submitted to the same docking protocol.

4.2.3 Results and Discussion

From a visual inspection of the docking poses, three AZPP ligands displayed a 180° flipped binding mode with respect to the crystallised poses (**Figure 16**). This observation was not totally unexpected, however, because, as anticipated in **paragraph 4.1**, the ambiguity in respect to the orientation of the AZPP compounds was already stated in the original AZ paper[72] and also detailed in the validation report of the PDB files. Specifically, alternative ligand atom coordinates displaying lower occupancy from the electron density map have already been found in the X-ray structures, confusing the overall binding mode picture.

On the other hand, the general position of the ligands was, however, reasonably well estimated by docking, since ligands nicely overlaid to the hydrophobic cleft of ATP, coplanar to the adenine portion. The RMSD associated with the deviation between the docked and the crystal binding modes of AZPP are reported in **Table 4**.

Curiously, a few compounds were predicted to have propensity to bind at the F6P pocket rather than the ATP one, including compound **27** which experimentally proved to be the most potent among the AZPP analogues (**Figure 17**).

The obtained docking results were compared with the reported experimental activity values of the AZPP; the plot at **Figure 18** shows the correlation between reported pIC_{50} and docking scores. The attained correlation was quite low, meaning that even though docking posing was acceptably correct, the discrimination between fairly similar structures, such us this congeneric series, remained quite challenging during the scoring phase. Nonetheless, as already discussed in **paragraph 1.2.4**, several previous comparative studies have revealed rather poor accuracy in respect to docking scoring functions, as it is still a common limitation in this kind of calculation.[136]

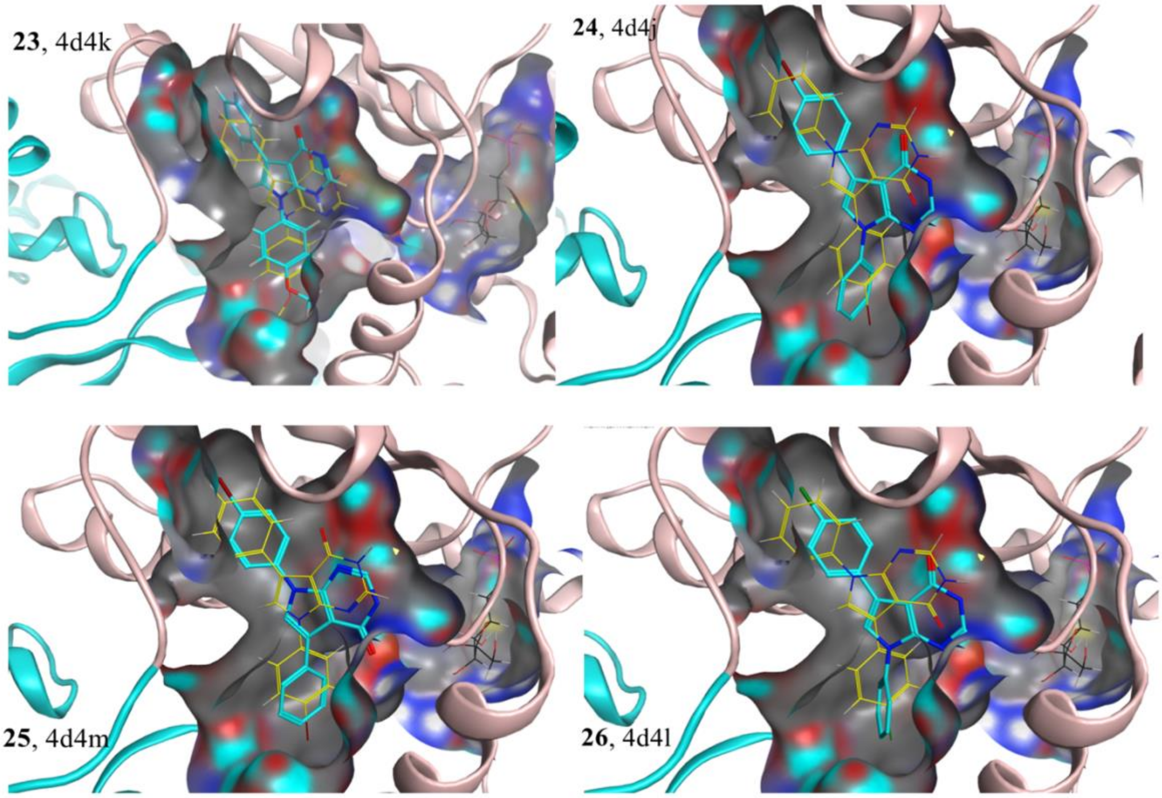


Figure 16: Comparison of the docking AZPP ligands pose prediction (shown in cyan sticks) and the corresponding crystallographic poses (shown in yellow sticks), by superimposition into the kinase catalytic domain of PFKFB3. From top-left to bottom-right corners: ligands **23**, **24**, **25** and **26**. Crystal structures are labelled according with their PDB IDs. The natural ligand, F6P, is also shown in its crystallographic pose (in black sticks). An approximation to the molecular solvent-excluded surface is shown coloured by atom colours (grey for carbon, red for oxygen, blue for nitrogen and cyan for H-bond hypotheses).

Compound	**23**	**24**	**25**	**26**
RMSD value (Å)	1.9	6.5	6.5	6.6

Table 4: RMSD values associated with the deviation of the poses of AZPP compounds **23**, **24**, **25** and **26**, predicted by molecular docking and the corresponding crystallographic poses.

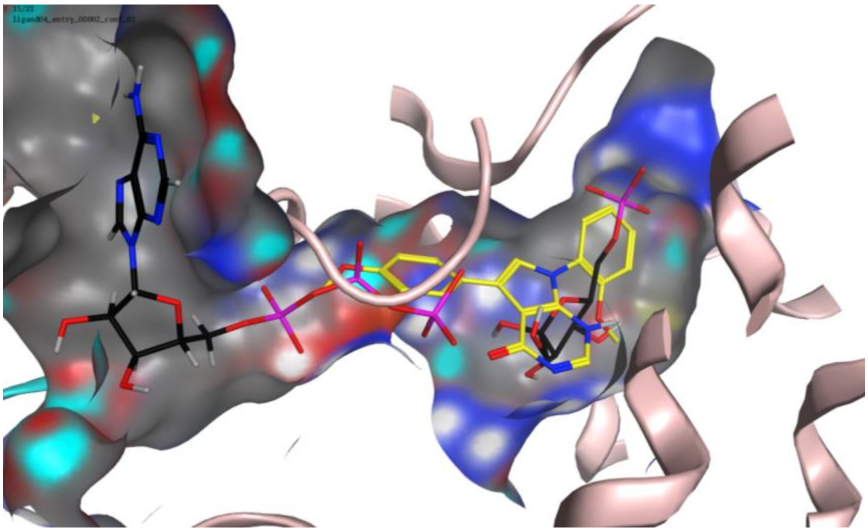


Figure 17: Site view of ligand **27** (shown in yellow sticks) in its predicted binding mode and superimposed with both endogenous ligands, ATP and F6P (shown in black sticks), into the kinase catalytic domain of PFKFB3. An approximation to the molecular solvent-excluded surface is shown coloured by atom colours (grey for carbon, red for oxygen, blue for nitrogen and cyan for H-bond hypotheses).

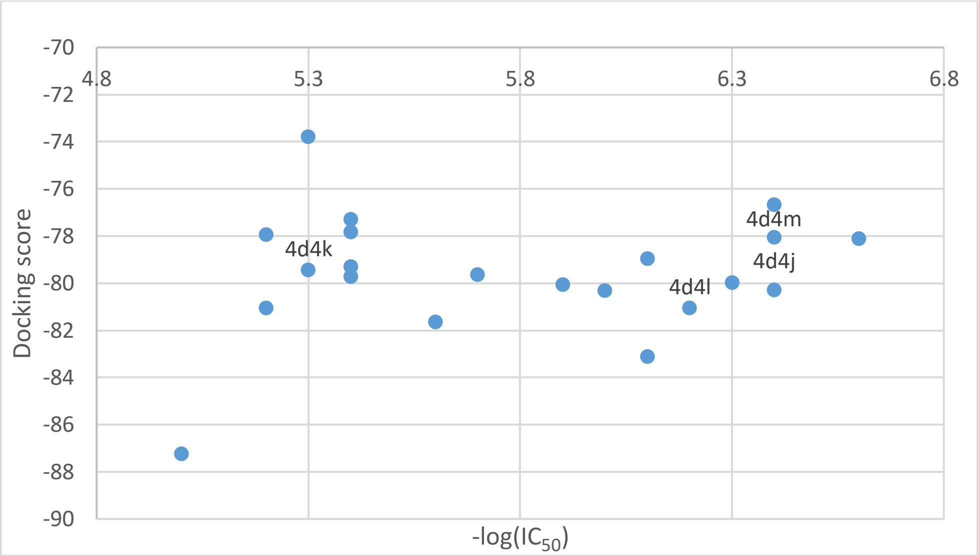
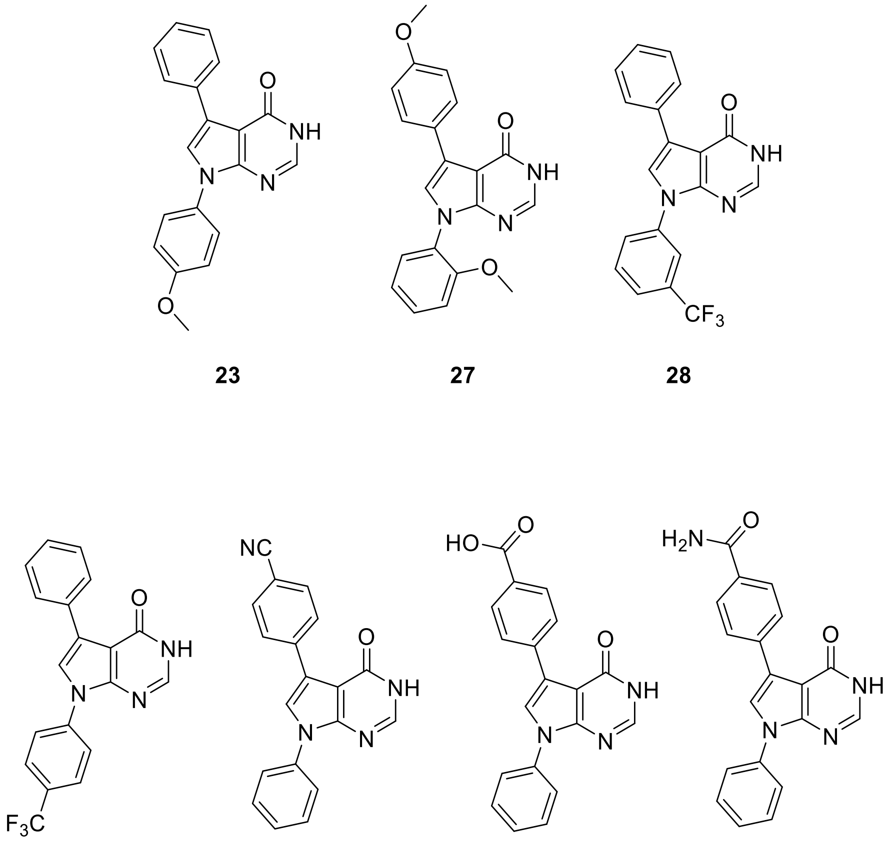


Figure 18: Correlation between reported $-\log_{10}(IC_{50})$ of AZPP series[72] and the docking score values. The crystallised AZPP ligands are labelled according to their PDB IDs.

Eventually, supplemented with qualitative docking calculations and principally guided by observation of interactions with the receptor, compounds **29**-**32** (**Figure 19**) were proposed to be empirically evaluated and thus synthesised. In order to compare activities, along with the selected compound, three compounds from the original collection, molecules **23**, **27** and **28**, were also synthesised and used as basis of comparison for the biological tests (**Figure 19**).

Paragraph 5.5 describes the organic synthesis of these AZPP-derived target molecules and **paragraph** 6.1 their biological evaluation.


23 **27** **28**

29 **30** **31** **32**

Figure 19: Chemical structures of the synthesised and tested AZPP analogues. (*top*) Reported compounds **23**, **27** and **28**.[72] (*bottom*) Compounds **29**-**32** are the molecules selected from the AZPP-based ligands optimisation as candidate PFKFB3 inhibitors.

4.3 Similarity Search and Molecular Docking

The 3D similarity search around the AZPP series was carried out using the wwLigCSRre online facility.[89] The server allows a specified conformational library of compounds to be mined and returns a subset of ranked hit structures matching a given query. In this case the query represents the 3D atom coordinates of an active ligand. The system molecular search is based on a 3D maximal common chemical patterns that takes into account both atomic nature and bond connectivity information. The alignment algorithm consists of an iterative and stochastic procedure for maximising the number of paired bonds with atomic compatibilities. The screened database of the present search was supplied with the server and consisted of a suitable chemically diverse library derived from the DrugBank source collection,[137] containing synthesisable drug-like small molecules already stored with their conformational information. The 3D geometrical features of molecule **24** (**Table 3**), were used as the query ligand. Compounds from the library, were compared to the 3D structure of **24**, and a maximum of 50 top-ranked matching ligands were retrieved. The obtained reduced database of hit compounds were then submitted to the molecular docking calculation into the kinase pocket using the PFKFB3 dimeric model described in **paragraph 4.2.2**.

4.3.1 Results and Discussion

After the similarity search mining was processed, the hit molecules were visually examined, comparing the selected 3D conformations by superimposition onto the query molecule **24**.

Figure 20 illustrates that the alignment of one hit molecule, compound **33**, was reasonably correct. The dimethylthiazole moiety of **33** shows complementarity with the N-aromatic ring attached to the pyrrole scaffold of **24**, which, on the X-ray structure sits on the cleft that hosts the ribose fragment of ATP. The pyrimidine and the *p*-trifluoromethylphenyl substituent of **33** overlap the heterocyclic core and the *p*-halophenyl ring of **24**, respectively.

A visual analysis of the results of posing followed molecular docking, to assess the quality of the predicted binding modes. The most probable binding site for compound **33** was predicted to be the binding channel occupied by phosphate groups when ADP/ATP is bound. In particular, the docked pose of **33** fits the cleft establishing contact interactions with hydrogen-bond-rich side chains. Moreover, the π-electron clouds of the planar aromatic groups of the molecule was predicted to engage with the hydrophobic side chain of Thr48 (**Figure 21**).

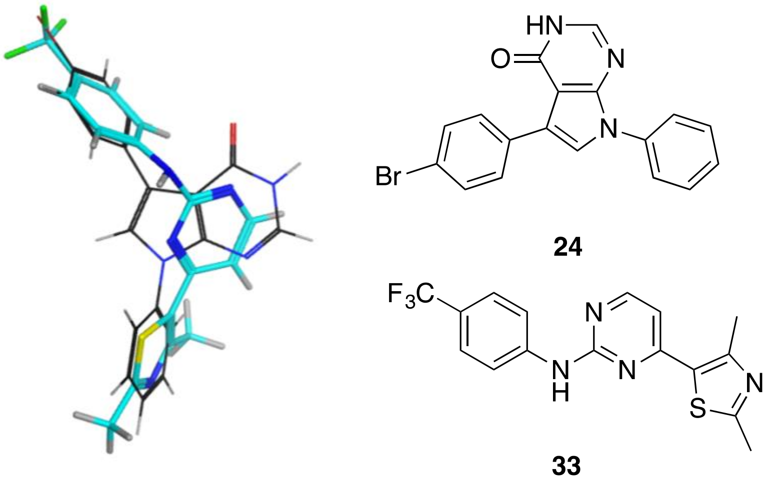


Figure 20: (*left)* Superimposition of the similarity search hit, compound **33** (shown in cyan sticks), and the AZPP query, molecule **24** (shown in black sticks). (*right)* 2D chemical structures of the compounds **33** and **24**.

Also in this case, analysing the result of virtual screening and, from that, identifying the “right” compounds to take forward for further validation, represented a key challenge. As was done previously, we deemed it important to not underestimate synthetic feasibility and chemical reliability, in view of later stages of ligand validation. Among the other molecules, compound **33** showed the most convincing alignment with the AZPP **24** as well as a reasonable docking pose. Other molecules were instead discarded, even though showing good alignment and docking score, because either i) they were too similar to ATP, ADP, purines and pyrimidines thus very likely unselective towards PFKFB3, or ii) because of their poor synthetic tractability. Therefore molecule **33** was chosen and synthesised for further biological studies.

Paragraph 5.6 describes the organic synthesis of this target molecule and **paragraph** 6.1 its biological evaluation.

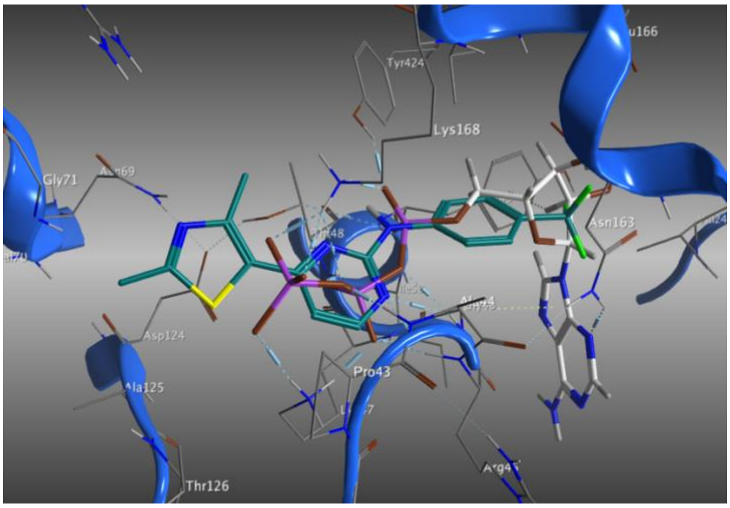


Figure 21: Predicted binding mode by docking of compound **33** (shown in turquoise sticks), into the kinase catalytic domain. ATP is also shown in its crystallographic position (shown in white sticks). Hit molecule **33** was selected through the AZPP-based similarity search, and then synthesised for further biological evaluation.

5

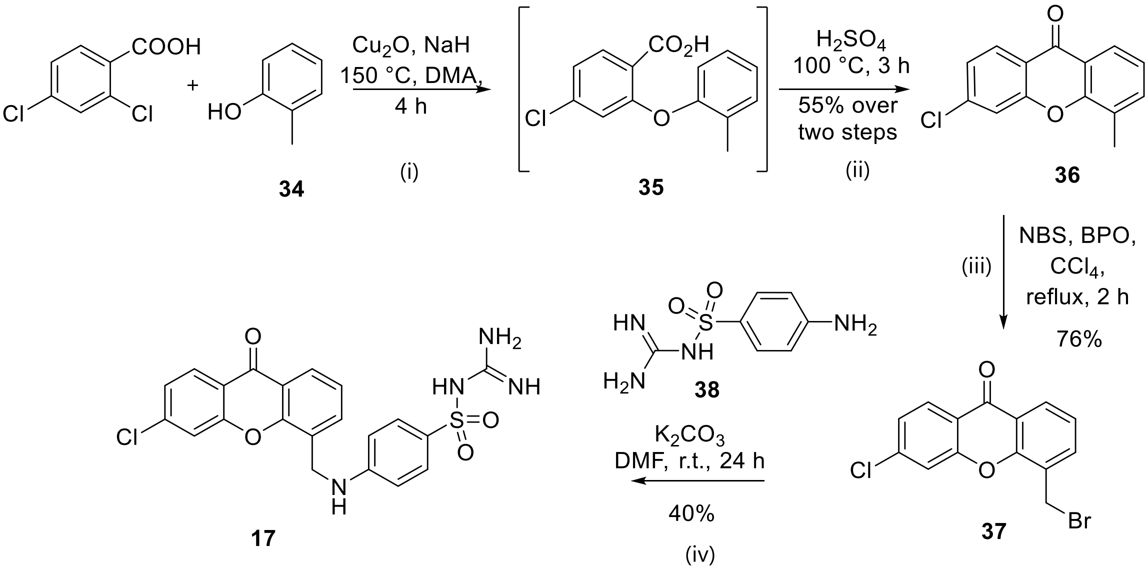
Organic Syntheses

5 Organic Syntheses

5.1 Synthesis of TM1, compound 17

The first target molecule (TM1), compound **17**, was prepared through a multistep synthesis as illustrated below in **Scheme 1**, which is a refinement of previously published procedures.[138–141]

The first step is a copper-mediated Ullmann's condensation of the substituted chlorobenzoic acid and *o*-cresol **34**, followed by cyclisation, resulting in the corresponding xanthone **36**, through the carboxylic acid **35**.



Scheme 1: Optimised synthesis of TM1, compound **17**. *Reagents and conditions*: (i) Cu_2O, NaH (60 % mineral oil), DMA, 150 °C, 4 h; (ii) conc. H_2SO_4, 100 °C, 3 h; (iii) NBS, benzoyl peroxide, CCl_4, reflux, 2 h; (iv) K_2CO_3(s), sulphaguanidine **38**, DMF, r.t., 24 h.

The Ullmann-type coupling usually occurs in rather harsh conditions such as high temperature and addition of strong base, especially with sterically hindered substrates, such as the compound **34**; thus high-boiling polar solvents were necessary. In addition, the rather strong air sensitivity of the phenol **34** required a nitrogen atmosphere. Very low yields with (mostly homocoupling) by-products were obtained when performing the reaction under microwave irradiation (mw), with caesium carbonate and dimethylglycine as the coordinating agent. Whereas the catalytic system Cu/CuI in the presence of pyridine and an inorganic base, in water, gave no-reaction. The best copper source used for catalysing the condensation step was found to be copper(I) oxide which could be used without any ligand, in the presence of NaH. Yet, because of its low solubility, the copper oxide was used in stoichiometric amounts (1 equivalent) in order to improve the yield. Also, utilisation of strong base could be avoided by using as starting materials the corresponding sodium salts of both chlorobenzoic acid and *o*-cresol. As has previously been well documented,[142] the *ortho*-carboxylic group onto the aryl chloride makes the C2 more electrophilic, allowing good regioselectivity at this position. The crude material of the Ullman arylation, containing intermediate **35**, could be used without purification for the following intramolecular Friedel Crafts acylation in concentrated sulphuric acid.

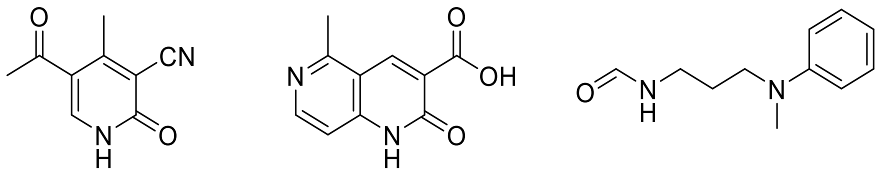
Next, the free-radical bromination of the methyl group in carbon tetrachloride, using benzoyl peroxide (BPO) as a radical initiator and equimolar *N*-bromosuccinimide (NBS), gave the bromomethyl derivative **37** in relatively good yield, with only little formation of the dibrominated species. Last, the nucleophilic substitution of the benzylic bromide with the commercial sulphaguanidine **38** in basic conditions afforded the desired TM1, compound **17**.

5.2 Synthesis of TM2, compound 18

The synthesis of the second target molecule (TM2), compound **18**, is shown in **Scheme 2**.

The dimethylformamide dimethyl acetal (DMFDMA) was used as the key building block for the preparation of the naphthyridine heterocycle.[143–147] As reported in the literature, DMFDMA was used twice in the first and third step as a one-carbon synthon, and the enamine intermediates **40** and **42** did not need to be isolated for the following steps. Enaminedione **40** was directly reacted with freshly prepared sodium cyanocetamide in dry DMF at room temperature (23 °C) in a one-pot process. After acidification and work-up, the substituted pyridinone **41** was obtained as a precipitate.

The structure of the desired pyridinone **41** was confirmed by a combination of both HMBC and NOESY NMR analysis. A four-bond coupling between the aromatic proton and the methyl carbon in position-6 of the ring (W-coupling) was observed, moreover a diagnostic NOESY signal confirmed the spatial proximity between the methyl protons and the downfield N-H of the lactam. On the other hand, there was no evidence of coupling between N-H and the aromatic proton, excluding the hypothesis of having the isomer **45** (**Figure 22**). Moreover, further confirmation was supported by the correspondence between the reported and obtained chemical shift data of the desired compound.[146]



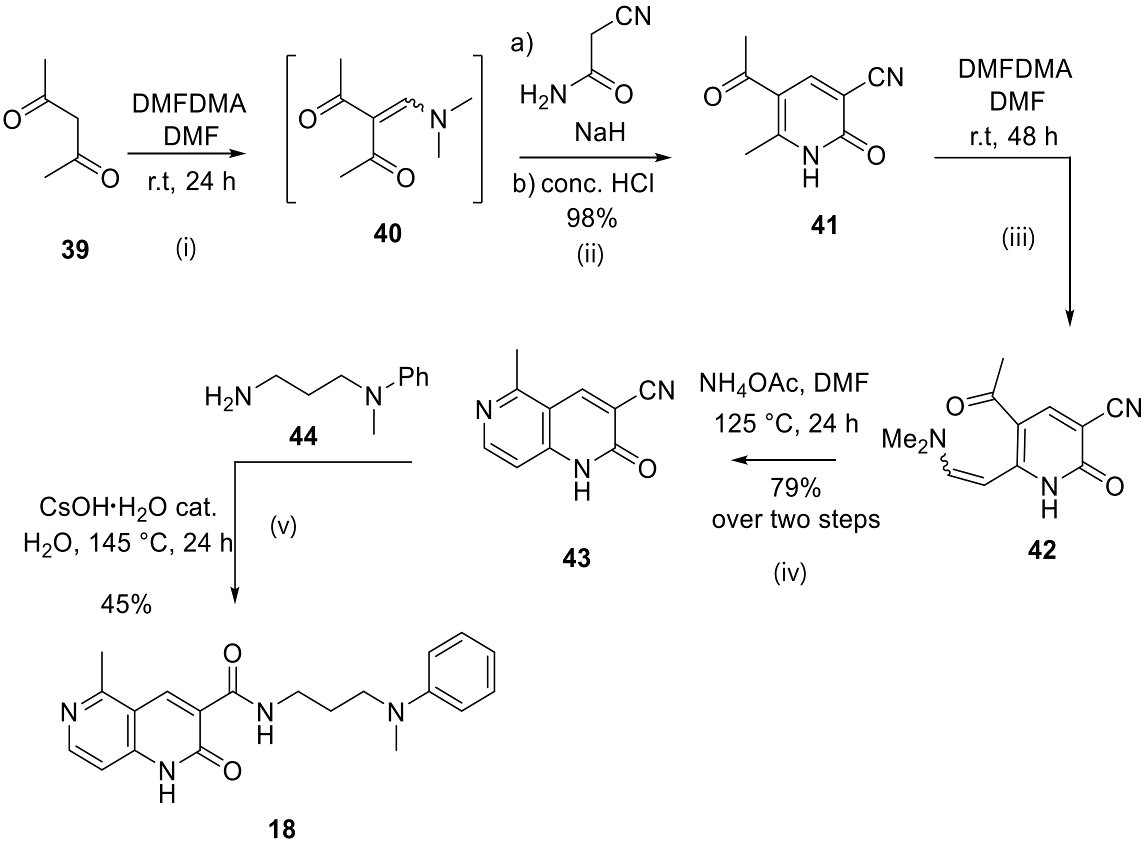
45
Not observed

46

47

Figure 22: Chemical species related to the synthesis of compound **18**.

Next, adduct **42**, obtained from the reaction of **41** with DMFDMA in DMF, was treated with an excess of ammonium acetate resulting in the desired napthyridine derivative **43**.



Scheme 2: Optimised synthesis of TM2, compound **18**. *Reagent and conditions*: (i) DMF, DMFDMA, r.t., 24 h; (ii) a) cyanocetamide, NaH (60% mineral oil), r.t., 20 h; b) H_2O/EtOH (1:1), conc. HCl, 24 h; (iii) DMF, DMFDMA, r.t., 48 h; (iv) NH_4OAc, DMF, 125 °C, 24 h; (v) *N-(*3-aminopropyl)-*N*-methylaniline **44**, $CsOH{\cdot}H_2O$ cat, H_2O, 145 °C, 24 h.

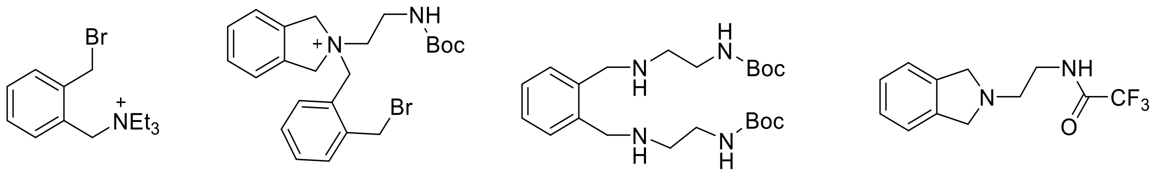
Finally, target molecule TM2, **18** was obtained in good yield in a sole step, directly from the nitrile, by a CsOH-catalysed aminolysis in water with amine **44**, in a sealed tube at high temperature.[148] This reaction is a green alternative approach that conveniently allowed to avoid two troublesome steps: i) the hydrolysis of nitrile **43** to the corresponding carboxylic acid **46** (**Figure 22**), and ii) the following amidic coupling of napthytidine acid and the amine **44**. Indeed, this latter two-step route in which the amide coupling was catalysed by HOBt and EDC·HCl was unsuccessful since the purification of the substrate **46** was rather problematic.

Interestingly, along with the secondary carboxyamide **18**, trace of the formylated adduct of amine **47** was also isolated from the crude mixture of the final aminolysis step (**Figure 22**). Further investigation by performing a control reaction in the same conditions, without nitrile **43** and introducing a catalytic amount of DMF, suggested that the formylation of the primary amine group of **44** occurred when trace of DMF were present in the reaction mixture.

5.3 Synthesis of TM3, compound 19

Scheme 3 shows the optimised multistep synthesis of the third target molecule (TM3), compound **19**.

Attempts to obtain the intermediate **53** by reductive amination of the *o*-pthalaldehyde with $NaCNBH_3$ and monoprotected *N*-Boc-ethylenediamine **52** were unsuccessful, leading mostly to the corresponding diol and a series of undesired compounds. Thus, the isoindoline fragment of compound **53** was prepared from the commercially available dibromo-xylene and ethylenediamine **52** in presence of an organic base. The quaternary ammonium salt of TEA with bromide (adduct **48**, **Figure 23**) was detected by mass spectrometry in the raw material. Even though **48** was easily removed by washing several times with water, the amount of the ammonium salt could be minimised by using a more hindered organic base, such as DIPEA. When two equivalents of bromide were used, after completion of the reaction, during concentration phase to remove solvent, the excess of bromide acted as a scavenger, trapping the product within the species **49** (**Figure 23**). Hence, the optimised conditions consist of using an excess of amine **52** and DIPEA with respect to the bromide, which enabled to obtain **53** in good yield and no formation of the dialkylated xylylenediamine (adduct **50**, **Figure 23**).



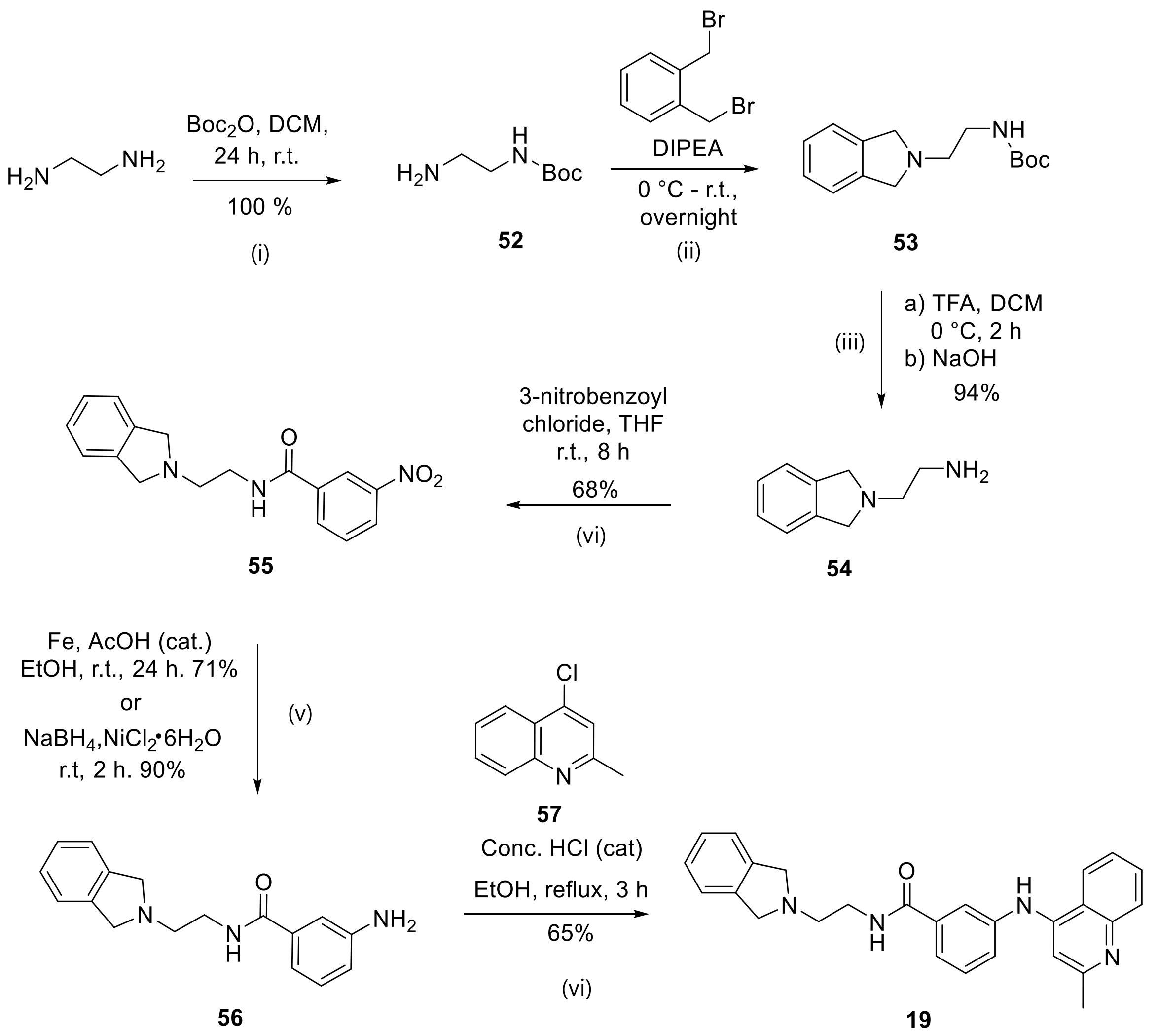
48 **49** **50** **51**

Figure 23: By-product obtained throughout the synthesis of compound **19**.

Deprotection of the amine was performed in standard acidic conditions with TFA, although ~30% of the corresponding trifluoroacetamide of the amine **54** (adduct **51**, **Figure 23**), presumably formed during the basic work-up, was also detected and isolated from the crude mixture as a by-product.

In the following step, the free amine **54** was reacted with 3-nitrobenzoyl chloride to obtain the carboxamide **55**. The intrinsic basicity of the molecule **54** allowed the reaction to occur without the necessity of any base. On the contrary, unexpectedly, the addition of either TEA or DIPEA alongside DMAP (even in catalytic amount of the latter) was actually ineffective, meanwhile favouring the formation of unidentified by-products of competing side reactions. The nitro group of **55** was then reduced rather smoothly to aniline **56**, either with Fe and acetic acid as the catalyst, or in the presence of nickel(II) chloride and sodium borohydride, obtaining good yields. Whereas other reduction conditions, such as Zn dust and ammonium chloride in methanol, did not afford the desired amino group.

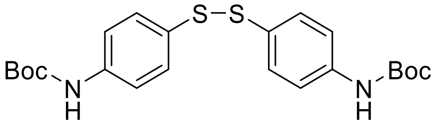
Finally, in the last step the 4-chloroquinaldine **57** underwent a S_NAr reaction (Nucleophilic Aromatic Substitution) with **65** and a catalytic amount of concentrated HCl to afford TM3, **19**.



Scheme 3: Optimised synthesis of TM3, compound **19**. *Reagent and conditions*: (i) Boc$_2$O, DCM, r.t. 24 h; (ii) α,α'-Dibromo-*o*-xylene, DIPEA, CHCl$_3$, r.t. overnight; (iii) a) TFA/DCM (40% v/v), 0 °C, 2 h, b) NaOH aq. (vi) 3-nitrobezoyl chloride, THF, r.t., 8 h; (v) Fe, AcOH cat, EtOH, r.t., 24 h or NaBH$_4$, NiCl$_2$·6H$_2$O, MeOH/EtOAc (20:1), r.t., 2 h; (vi) 4-chloroquinaldine **57**, conc. HCl (cat), EtOH.

5.4 Synthesis of TM4, compound 20

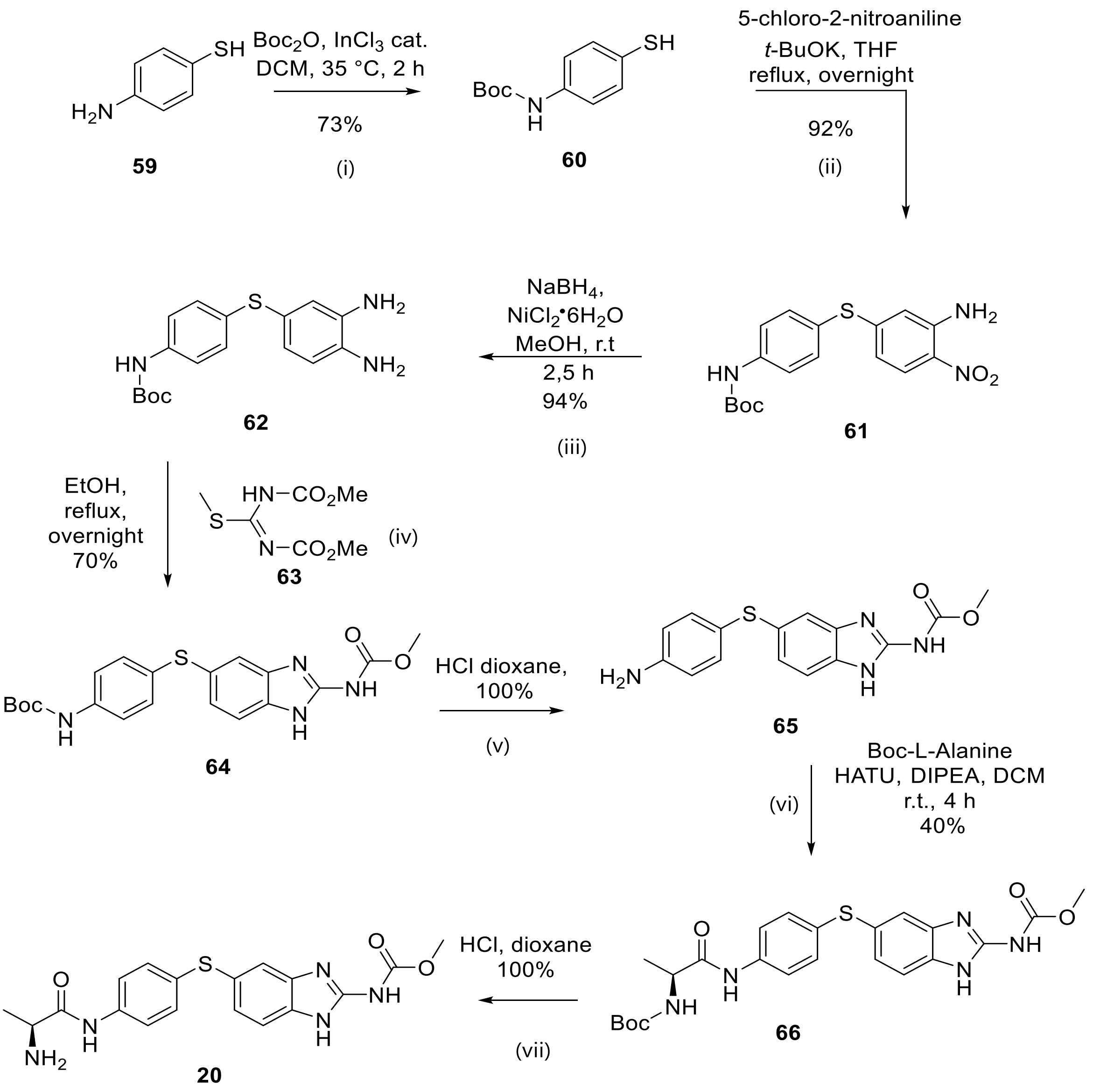
The synthesis of target molecule TM4, compound **20** is shown in **Scheme 4**. Hasegawa *et al.* have already published an expeditious synthetic route for **20** where the benzimidazole ring is prepared by treating the dianiline **62** with the thiopseudourea **63**.[149] We implemented the described pathway which starts from the commercially available 5-chloro-2-nitroaniline and the *N*-protected 4-thioaniline **60**. Instead of an acetyl group we opted to use *tert*-butyloxycarbonyl (Boc) as protecting group because of its more convenient removal conditions. The preparation of *N*-Boc-aminothiophenol was carried out in the presence of catalytic indium(III) chloride, in a good yield. Only a small amount of disulphide dimer as a by-product (compound **58**, **Figure 24**), derived from the homo-coupling of two thiol groups, was found and isolated from the reaction mixture.



58

Figure 24: Disulphide by-product.

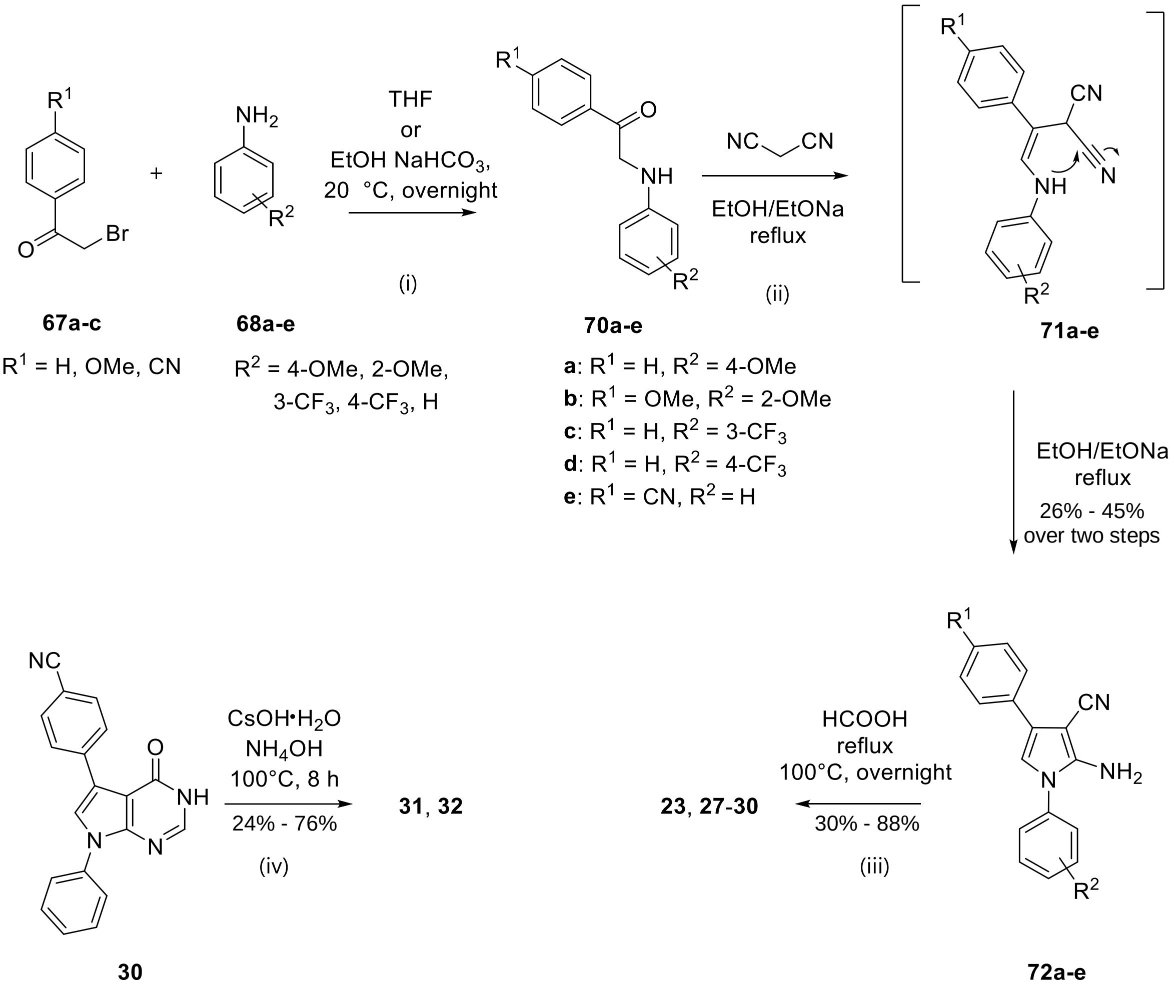
The compound **60** was reacted with 5-chloro-2-nitroaniline in the presence of a base to provide the intermediate **61**. Next, contrary to what has previously reported,[149] in our hands, treating the nitro compound with sodium dithionite ($Na_2S_2O_4$) did not give the expected aniline **62**, with only starting material being recovered after work-up. Nevertheless, the reduction with the $NaBH_4$/$NiCl_2$ catalytic system afforded **62** in excellent yield. After cyclisation to benzo[d]imidazolyl methylcarbamate **64**, *via* condensation with **63**, the Boc protecting group was removed in acidic conditions. The aniline **24** was then acylated with Boc-L-alanine through standard amidic coupling conditions, namely with HATU and DIPEA. Finally, deprotection of alanine gave the TM4 (**Scheme 4**).



Scheme 4: Optimised synthesis of TM4, compound **20**. *Reagent and conditions*: (i) Boc$_2$O, InCl$_3$ cat., DCM, 35 °C. 2 h; (ii) 5-chloro-2-nitroaniline, Bu^{t}OK, THF, reflux, overnight (iii) NiCl$_2$·6H$_2$O, NaBH$_4$, MeOH/EtOAc (20:1), r.t., 2.5 h; (vi) 1,3-bis(methoxycarbonyl)-2-methyl-thiopseudourea **63**, EtOH, reflux, overnight; (v) 4M HCl in dioxane, r.t. 24 h; (vi) Boc-L-Alanine, HATU, DIPEA, DCM, r.t. 4 h; (vii) 4M HCl in dioxane, r.t. 24 h.

5.5 Synthesis of AZPP series

The synthesis of the dihydropyrrolopyrimidinone analogues was carried out as reported previously, through the formation of 2-amino-3-cyano-1,4-disubstituted pyrroles **72a-e**, followed by cyclocondensation with formic acid (**Scheme 5**).[150–155]

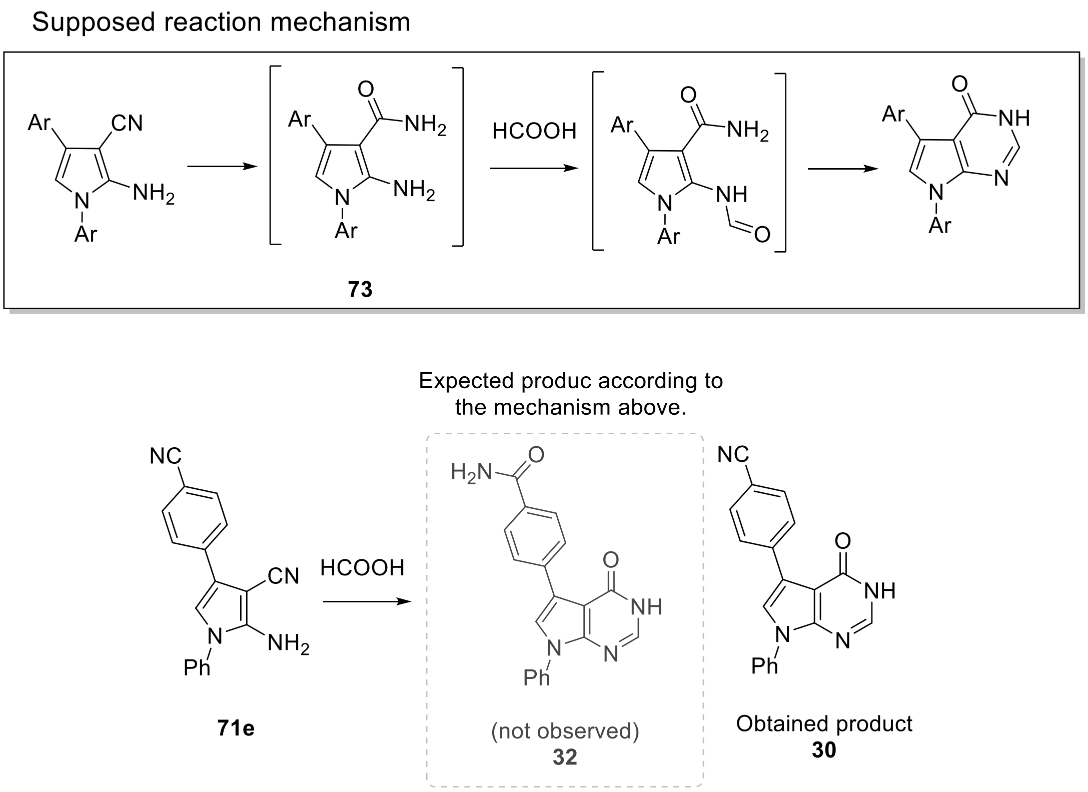


Scheme 5: Synthesis of AZPP analogues. *Reagent and conditions*: (i) THF, 2 equivalents of aniline, 20 °C, overnight or EtOH, $NaHCO_3$(s), r.t. overnight; (ii) Malononitrile, EtONa (21 wt.% in EtOH), 85 °C, overnight; (iii) HCOOH 85%, reflux, overnight; (iv) $CsOH·H_2O$ cat, NH_4OH, H_2O, 100°C, 8 h.

Several procedures for the preparation of the key α-aminoketone precursors **70** are described in the literature, and the vast majority starts from the commercially available substituted phenacyl bromides **67** and anilines **68** using different solvent and stoichiometric ratio.[156–158] However, contrary to what is reported in some procedures, the reaction must be carried out thoroughly at room temperature, since the heat was found to facilitate the formation of the di-alkylated amine as a by-product. The optimal conditions consist of setting-up the reaction in EtOH with a 1:1:1 molar ratio of phenacyl bromide, aniline and $NaHCO_3$. Furthermore, the α-aminoketones **70** were found quite unstable making it difficult its handling and isolation. Both acidity from silica gel and exposure to light degrade these intermediates. They thereby must be kept under an inert atmosphere and be immediately submitted to the following step. Particular care was required for the extremely unstable substrate **70a**, and the highest yields were obtained by using THF as reaction solvent and adding two equivalents (one acting as a base) of the electron-rich *p*-anisidine **67a**. The product was then recovered after precipitation as a hydrobromide salt from THF and was reacted immediately through the following step.

When compounds **70** were heated in the presence of equimolar amount of malononitrile and fresh EtONa, they underwent cyclisation to afford the substituted pyrroles **72**, following a Knoevenagel condensation-type reaction, through the formation of adducts **71** (**Scheme 5**). The isolation of the intermediate **72a** by flash chromatography was more troublesome due to its instability, thus purification by crystallisation was required instead, which was not beneficial for the yields. Compounds **72d** and **72e**, similarly, could not be isolated and were used immediately for the next step. For most of the substrates, the yields were actually affected by the reaction scale of starting materials.

Finally, prolonged heating of **72** in formic acid gave the desired products. The reaction was believed to proceed through the corresponding 2-amino-3-carboxamidopyrroles **73**, resulting from the hydrolysis of the nitrile group followed by condensation with formic acid, which causes the cyclisation to occur (**Scheme 6**). This mechanism was hypothesised after the detection by mass spectrometry of the α-amido amides **73** in the reaction mixtures of each substrate and it was also previously proposed in the literature.[152,156]



Scheme 6: (*top*) Supposed reaction mechanism for the formation of pyrimidinone heterocycle: hydrolysis of the nitrile group, followed by the condensation of formic acid and cyclisation;[152,156] (*bottom*) cyclocondensation of dinitrile compound **71e** in boiling formic acid: only compound **30** was afforded, whereas no trace of **32** was observed.

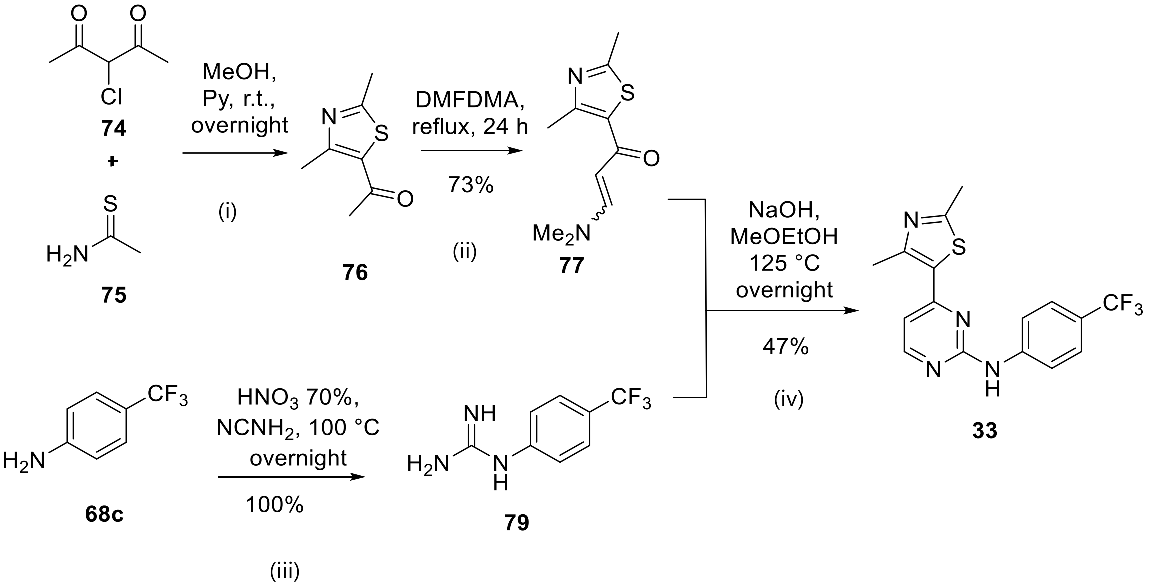
In view of this supposed reaction mechanism of the pyrimidinone cycle, the synthesis of the primary amide **32** was, therefore, retro-synthetically designed as a one-pot process involving the hydrolysis in boiling formic acid of both pyrrolic and phenylic nitrile groups of **71e**, prior to the cyclisation. Unexpectedly, the nitrile group attached to the phenyl moiety, however, did not react, not even with a large excess of HCOOH(aq) and running the reaction at high temperature for several hours (**Scheme 6**). Hence, a different reaction mechanism may have been involved in this case.

The obtained compound **30** was isolated, characterised and tested. Nitrile **30** was further reacted with NH_4OH and catalytic CsOH to give the corresponding amide **32** by aminolysis. The overhydrolysis of the nitrile led to carboxylic acid **31** which was also isolated, characterised and biologically evaluated (**Scheme 5**).

5.6 Synthesis of Similarity Search-Hit, compound 33

For the preparation of similarity search hit, compound **33**, the pyrimidine condensation method of Bredereck was applied, according to a convergent synthetic strategy (**Scheme 7**).[159,160]

The chloroacetyl acetone **74** and the thioacetamide **75** were used as starting materials to build the 5-acetyl-thiazole intermediate **76**, which was subsequently treated with DMFDMA. The obtained enaminone **77**, was condensed with anilino guanidine **79** to give the desired pyrimidine **33**. The anilino guanidine was derived from the corresponding *para*-trifluoromethyl aniline **68c** by treatment of its nitrate salt with cyanamide (**Scheme 7**).



Scheme 7: Synthesis of the target molecule designed from AZPP through a similarity search, compound **33**. *Reagent and conditions*: (i) Pyridine (Py), MeOH, r.t., overnight; (ii) DMFDMA, reflux, 24 h; (iii) HNO_3 (70%), cyanamide 50 wt.% in H_2O, 100 ˚C, overnight; (iv) NaOH, 2-methoxyethanol, 125 ˚C, overnight.

6

Biological Evaluation

6 Biological Evaluation

All biological experiments reported in this manuscript were carried out by Helena Macut, Early Stage Researcher of the Moglynet Project (University of Milan, Italy), in respect of a collaborative programme.

6.1 Kinase Assay

The kinase assays of the compounds **17**-**20** (**Figure 14**), **23**, **27-32** (**Figure 19**) and **33** (**Figure 20**), were performed using the ADP-Glo™ format[161] according to literature protocol.[68] The test method consists of a two-step process in which the kinase activity is quantified by the amount of ADP produced during the enzymatic kinase reaction, in presence of the ligand. In the first step, the enzyme and the substrates (ATP and F6P) are incubated with the ligand, and after 1 h (determined by PFKFB3 kinetics) a reagent, provided with the kit, is added to stop the reaction and deplete any unreacted ATP. In the second step, ADP is converted back to ATP and the newly synthesised ATP reacts with the luciferase/luciferin system producing a luminescence signal that is proportional to the ADP concentration. The generated light is measured by luminometer and it is directly correlated by a linear relationship with kinase activity.

The assay was performed using human recombinant PFKFB3 that was produced and purified in-house. Synthesised AZ33, compound **16** (**Table 1**) was used as positive control since it proved to almost completely inhibit the kinase reaction. Compounds were screened at two different concentrations (100 µM and 10 µM). The corresponding residual activity (RA%), which compares the kinase activity of PFKFB3 in presence of the ligand with respect to the blank (i.e. in absence of ligand), was calculated.

6.2 Results and Discussion

Table 5 reports the residual activities of the compounds that were designed based on the AZI ligands (**Chapter** 3). Analysis of the RA% data obtained from this generation of candidate inhibitors revealed that unfortunately rather poor or no inhibition was observed for most of the ligands.

Compound **20**, amongst the others, however, showed a more significant signal change with respect to the blank signal, and gave a residual activity of 31.3% and 60.5%, at 100 µM and 10 µM respectively. Looking at its chemical structure (**Figure 14**), it is the closest compound resembling the AZI ligands (for instance compared to AZI compound **11**, **Table 1**). Indeed, it displays shape similarity and a comparable central hydrophobic scaffold formed by two aromatic portions connected with a heteroatom. The sulphur linker atom in **20**, instead of the oxygen of the AZIs, might have increased the lipophilicity, however, the solubility was improved by the benzimidazole and the methyl carbamate moiety replacing the AZI indole. Moreover, although showed less potency compared to the low nM activity of some AZI ligands, inhibition activity of compound **20** was still within the active concentration range observed for the AZPP compounds. Compound **19** also revealed a relatively comparable inhibition, but with a concentration-independent behaviour, leading us to suppose that some kind of artefactual influences might have affected the result.

Initially, we ascribed the surprisingly slight activating behaviour of compounds **21** and **22** to a possible allosteric modulation. Indeed, as already mentioned, the PFKFB3 enzyme has a distinctive dual nature, due to its two mutually contrasting activities. Besides, since they bind the same endogenous ligands (ATP/ADP and F6P/F2,6P), the bisphosphatase and the kinase domains of this enzyme intrinsically display similarities in their structural and physicochemical properties. As a consequence, since these evaluated molecules were designed to bind at the kinase domain, an indirect stimulatory effect on the enzymatic activity, through the binding at the bisphosphatase domain or other allosteric sites, cannot be excluded. Also, it has been observed previously that small molecules belonging to the same family of compounds can induce opposite effects towards this enzyme, and only slight differences in their structures have proved to be discriminative characteristics for their biological effect.[69] Taking our data from a relative perspective, however, it is also true that an allosteric activator would have showed a much higher increment of the activity, while compounds **21** and **22** only showed a two-fold activation which is probably too little effect, and thus negligible.

Compound	**RA% (at 100 µM)**	**RA% (at 10 µM)**	**Reported IC_{50} (µM)[68]**
17	110.2 ± 1.6	96.4 ± 0.7	-
18	100.4 ± 3.3	59.4 ± 3.9	-
19	35.96 ± 1.03	34.6 ± 0.2	-
20	31.3 ± 1.1	60.5 ± 2.8	-
21	197.3 ± 2.9	200.4 ± 0.8	-
22	286.4 ± 5.4	107.8 ± 6.3	-
16	-	7.2 ± 0.1	0.003

Table 5: Kinase assay result of the compounds selected from the AZI-based virtual screening. Residual activity (RA%) is reported at 100 µM and 10 µM of ligand concentration).

In some cases, the residual activity obtained at 100 µM was slightly higher than the corresponding value measured at 10 µM of inhibitor concentration. This could be an artefact due to the noise and a too low inhibition potency, or it may be due to possible interference by the compounds with the measurements.

Table 6 reports the residual activity related to compounds designed basing on AZPP ligands (**Chapter 4**). The original AZPP ligands, compounds **23**, **27** and **28**, have a reported IC_{50} of 5.012 µM, 0.215 µM and 10.000 µM respectively,[72] which is overall in agreement with the trend of our obtained residual activity results. After testing the close analogue of molecule 28, molecule **29**, we found that modification of the trifluoromethyl group from *meta* position to *para* position of the phenyl ring, was not beneficial for potency, since **29**

showed lower inhibition. The compound **30** was also tested as an organic salt, with trifluoroacetate (TFA) as counter-ion, in order to improve its solubility. This resulted, however, in a loss of activity of the salt form of **30** at 100 µM.

The carboxylic acid and the primary amide derivatives **31** and **32** were designed specially to increase the poor hydrophilicity of the AZPP ligands and ensue additional polar interaction with the PFKFB3 enzyme, filling the kinase pocket. Unfortunately, however, the inhibition was shown to be comparable to the inhibition observed with the original AZPP compounds. Within this generation of candidates, compound **33**, showing promising profile *in silico*, was synthesised and tested. Unexpectedly, however, this compound did not exhibit improved activity at the tested concentrations, since it reduced PFKFB3 kinase activity only partially (**Table 6**).

Compound	**RA% (at 100 µM)**	**RA% (at 10 µM)**	**Reported IC$_{50}$ (µM)**[72]
23	9.31 ± 0.01	23.1 ± 0.75	5.012
27	4.24 ± 0.13	8.18 ± 0.50	0.215
28	30.2 ± 0.06	30.3 ± 0.73	10.000
29	65.1 ± 1.21	37.5 ± 1.54	-
30-TFA	40.2 ± 1.29	40.4 ± 0.26	-
30	25.8 ± 0.80	39.3 ± 1.32	-
31	55.7 ± 1.0	61.7 ± 1.37	-
32-TFA	27.3 ± 0.06	49.0 ± 0.97	-
33	73.4 ± 0.16	67.7 ± 1.89	-

Table 6: Kinase assay result of the compounds selected from the APP-based virtual screenings. Residual activity (RA%) is reported at 100 µM and 10 µM of ligand concentration.

7

Conclusion

7 Conclusion

In summary, with this study we have presented our contribution to the challenging aim of identifying new potent and selective kinase inhibitors of the kinase domain of the PFKFB3 enzyme. Different tiered computational methods, such as both ligand-based and structure-based virtual screenings were applied to facilitate the process of the rational design of the prioritisation of promising compounds. Two classes of known active ligands, herein referred to as AZI[68] and AZPP[72] inhibitors, in combination with the 3D structure of the receptor, were used as a source of information for the development of molecular model used for virtual screenings, leading to two generations of new candidate inhibitors. Those selected compounds showing good predicted affinity towards the enzyme were synthesised, fully characterised and tested. Unfortunately, despite our consistent efforts and attempts, only low or no activity was detected within the tested compound concentrations.

From the information presented in the literature at the onset of this project, we attempted to extract the best possible understandings to build a new model and develop virtual screenings (VS), also according with our resources available. It should be noted though, that despite the numerous positive studies, the discovery of new drugs or hit active molecules via VS is undeniably not a trivial task and unfortunately it is not always a successful process. It is known that these types of complex calculations inherently have a margin of error due to the approximations of parameters, besides the fact that other important aspects, such as the solubility in buffers, non-specific bindings, etc. cannot be simulated and predicted in VS. Thus, care must be taken with computational methods and VS results because the power of these tools primarily lies in the elimination of molecules that are unlikely to succeed, rather than the selection the “perfect” compounds. In fact, the effectiveness and impact of LBVS and SBVS are also very reliant on the actual number of compounds that are picked and tested. Pharmaceutical companies test thousands of hit molecules or continue to rely heavily on traditional high-throughput screening (HTS), since its efficacy still remains dominant in this field.

Given these considerations, it is worth remarking that unfortunately the vast majority of the hit molecules identified in our study, from the AZI-based rational design, were no more commercially available. This represented a big limitation since we necessarily had to take the risk of starting the inherently time-demanding organic syntheses before having the ready support of a biological validation of our computational model. In fact, the massive synthetic work, required for the individual preparation of the selected compounds, that besides are chemically different from one another, with no common precursors (**Figure 14**), delayed the important experimental feedback. As a consequence, the number of hit molecules that we were able to test, in a PhD project timespan, could not quite reach the sufficient percentage of compounds that needs to be tested to succeed in this type of campaign.

However, with our piece of work we offer further insights in a molecular level into the facets of PFKFB3 ligand investigation. We provided a framework of our study on the structural and electronic characteristics of the ligand-protein interaction and we presented our set of observations used to build our model. Overall our results may still serve as an endorsement for future progress around the state-of-the-art of PFKFB3 ligand discovery. For instance, the presented pharmacophore model, which represents the crucial part of the screening cascade, can be refined in future studies. Modification of the tolerance radii of and the number of the matched features or the addition of volume and shape constraints for taking into account the steric allowance of the kinase pocket may represents a valuable step forward. We also speculate that optimisation of the pharmacophore model could be addressed to the inclusion of ligand occupancy in correspondence of the ribose cleft of ATP pocket. This part, matching with the N-substituent of the AZI ligand indole, in fact is reported to have an impact to the potency. For example, additional aromatic ring features could be imbedded in the model in order to offer more opportunities of π-stacking interactions with this portion of the protein which internal surface is defined by electron-rich aromatic side chains.

Moreover, even though no high-affinity ligands were identified, this work also provides additional information about the chemically diverse tested molecules that display interesting scaffolds and drug-like properties. In particular, with compound **20** at 100 µM, the kinase activity of PFKFB3 was inhibited by ∼70%, and ∼40% at 10 µM, with respect to the blank, representing an encouraging result that might lead further investigations around this interesting core template.

We then moved to the other class of more recently evaluated molecules, the AZPP derivatives, as a fresh starting point to carry out a ligand optimisation guided by a SAR study. In particular, we carried out a further investigation of their interesting properties, taking advantage of the breadth of feasible synthetic modifications. We built and explored a dedicated focused library of synthesisable and drug-like analogues. However, also in this case, the prediction of compound modifications that substantially enable improvements in inhibition potency is a complex task that is difficult to appreciate and, even more so to simulate on a computer. In fact, by molecular docking although we obtained reasonably good posing prediction, the discriminatory power between close analogues of a congeneric series of the same family of molecules was unavoidably rather weak. For this reason, the ligand selection was principally guided by the qualitative observation of the interactions with the receptor in the complex models, in order to fit the model. As such attempts to improve their affinity towards PFKFB3 and their solubility, by adding hydrogen bond donors and polar functional groups to best fit (even though only partially) the initial pharmacophore model, or by using compound salt forms were unfortunately ineffective (**Figure 19**), probably due to difficult desolvation before access the lipophilic pocket. Thus, the poor solubility of this generation of AZPP analogues in addition to the need for a new scaffold, induced us to undertake another concomitant strategy. As described, we performed a similarity search for scaffold hopping followed by molecular docking, taking advantage of the relatively limited computational time-scale required for these screening methods. The search led to a compound showing a more promising affinity towards the enzyme *in silico*, albeit it exhibited unfortunately only comparable low potency (**Figure 20**).

However, in conclusion with this study we present relevant SAR outcomes and property indications associated with new derivatives of the known AZPP compounds. This little headway may guide follow-up studies as well as integrate the knowledge about this class of purine isosteres into the field of kinases inhibition.

8

Experimental

8 Experimental

Computational Study

8.1 AZI-Based computational part

8.1.1 Test Databases Preparation

A benchmarking of theoretically non-actives was built through the Automated Decoy Generation Method provided by the online platform DUD•E (A Data of Useful Decoys: Enhanced).[85,130] The six co-crystallised AZI ligands expressed as 1D SMILES strings (simplified molecular-input line entry-system)[162] were used as input ligand dataset. Each active ligand gave rise to 50 matched decoys, giving a total of 300 theoretically non-actives. The obtained list of compounds was downloaded as 1D SMILES strings and then converted into 3D unique low energy conformers using the MOE 'Database Viewer' application. The 3D structures of six crystallised AZI inhibitor were merged with the generated decoys in the DB. The 'wash' tool of MOE application was applied followed by an energy minimisation with 'energy minimise', with conjugate gradient method to RMS gradient 0.1 kcal/mol·Å. Both 'wash' and 'energy minimise' were run with MMff94x force field[163] and the Generalised Born implicit solvation model, to form the test-DB.

8.1.2 3D-Conformational Test Database Preparation

A conformational sampling of the test-DB was performed using the 'Conformation Import' of MOE. Only decoys were selected into the Input dataset, in order to preserve the AZI ligands in their crystallographic bound conformations. The application was run with default settings, MMff94x force field and related parameters, Generalised Born implicit solvation model. A maximum of 15 conformations with the lowest energy were retained. Finally, the molecules were refined with MOE software by applying the upfront 'wash' and 'minimisation' preparation protocol in MMff94x force field and Generalised Born implicit solvation model, as described above (**paragraph 8.1.1**), giving the 3D-conformational-test-DB.

8.1.3 Pharmacophore Elucidation

A collection of pharmacophore queries was automatically generated through the 'Pharmacophore Elucidator' tool provided by MOE software. The input set of template ligands contained the six co-crystallised AZI ligands, all given as active, since they were proven to be potent inhibitors. In order to maintain their crystallographic information, the 3D conformations of the AZI ligands were kept untouched (*'As-is'*, in MOE terminology) and no further conformational search was run. The annotation scheme *'Unified'* was set for the assignment of the pharmacophore features, and 5 was set as maximum number of features per query. The program computed a collection of 121 query hypotheses. Only those queries displaying a good ligand alignment (Overlap score > 4.2) and matching each one of the six active compounds (Coverage = 6), giving a total of 80 queries, were taken into account for following validation.

8.1.4 Pharmacophore Filtration

The 'Pharmacophore Search' filtering tool of MOE was used for the pharmacophore filtration. The *'systematic matching'* of MOE was applied, by masking as 'essential' every features of each query. All conformations of the Hit molecules were retrieved in the output database by selecting 'results: molecules' in the Pharmacophore Search panel.

All unique lowest energy conformations of every hit were merged together, along with the six crystal ligands, to give a total of 386 ligands. Before subjecting the ligands to molecular docking calculation, the molecules were refined with MOE software by applying the upfront 'wash' and 'minimisation' preparation protocol like described above (**paragraph 8.1.1**).

For the validation of the pharmacophore query the 3D-conformational-test-DB was used as input dataset and 'Pharmacophore Search' was applied using same settings.

8.1.5 Structures Preparation

The computational models of monomeric PFKFB3 receptors were derived from crystal structures downloaded from the Protein Data Bank (PDB ID: 5AJX, 5AJY, 5AJZ, 5AJW, 5AJV, 5AK0, 2I1V). All eventual co-crystallised non-standard residues and cofactors (such as ADP/ATP, F6P/F2,6P, phosphoric acid, phosphate ion or other bound substrates) were manually removed from the structure before docking. The 'Structure Preparation' of MOE was used to correct issues, whereas hydrogen were added with 'Protonate 3D' at pH = 7, T = 300 K and salinity = 0.1 M. Chain breaks and termini were appropriately capped with NME (*N*-methyl amide group) for the *C*-terminus, and ACE (acetyl group) for the *N*-terminus; missing atomic data or ambiguous sequence identities were modelled or corrected (e.g. replace selenomethionine by methionine). Following, an energy refinement of the system with AMBER10:EHT[131,132] force field and the Generalised Born implicit solvation model was performed, keeping the backbone atoms and pocket residues (up to 8 Å from the ligand) constrained to original position.

The following homology modelling description has been provided by Dr Francesca Stanzione (Cambridge Crystallographic Data Centre).

The models of 3D monomeric structures of the isoform PFKFB1, 2 and 4 have been built *in silico* by homology modelling with Chimera graphical interface of MODELLER, basing on their known amino acid sequences (UniProt codes: P16118, O60825, Q16877).[133] The monomeric form of PFKFB3 extracted from the crystal structure (downloaded from Protein Data Bank, PDB ID: 5AJX) was used as template. Before processing, the missing loops in the template structure were built using the loop/refinement tool of MODELLER. For each target sequence we created a new alignment with the template sequence based on the Needleman-Wunsch global alignment algorithm.[164] The homology modelling, EM fitting were launched from Chimera and executed remotely via MODELLER. The system then calculated a set of plausible structures containing all non-hydrogen atoms. For each PFKFB isoform, 5 models were created and the model returning with the least restraints was selected for further analysis. Each model has been adapted to match the sequence length of the template. *C*- and *N*-termini were capped and limited energy minimization of the retained isoforms models was performed in order to ensure absence of atomic clashes. The stereochemical quality was assessed using Ramachandran φ-ψ dihedral plot. Subsequently 'Protonate 3D' of MOE was used to add hydrogen atoms.

8.1.6 Molecular Docking Protocol

The calculations were performed on a multiprocessor HPC cluster (High Performance Computing provided by University of Milan), subdividing the task on a chosen number of processors for a PBS queuing system. Docking was focused in the ATP binding site, which 3D coordinates were defined by a sphere of 13.5 Å radius, and centred on the centre of mass of the co-crystallised ligand. The protein structure was kept rigid. PLANTS program was used in a low search speed/high accuracy mode and ChemPLP was specified as scoring function.

All ligands were automatically pre-processed by the SPORES software (Structure Protonation and Recognition System)[165] to assign atom and bond type and to define stereoisomers for non-specified asymmetric centres. Alongside, all tautomers and ring conformations of the ligands were generated with the MOE software. After the posing process, the best ChemPLP score poses of each docked ligands of the database (i.e. the lowest energy geometry) was kept for further calculations. The outcome was re-scored according to scoring function values, resulting in a ranked database.

For the validation of docking method, the test-DB of the active/non-active ligands was used. In addition, the R-script integrated in the virtual screening protocol, was used to automatically plot the ROC curve and calculate the corresponding AUC.

8.1.7 MD Simulations Protocol

The simulations were carried out on a GPU workstation with the *pmemd.cuda* module of the Amber14 package.[120] The initial geometry of the complexes for the MD simulations were obtained using the docked poses of the ligands and the receptor structure modelled from 5AJX PDB file. For AZI ligands, both docked and crystallographic poses were simulated in complex with the receptor.

The topological files and a set of initial coordinates were prepared for each complex using the *tleap* module of the AMBER package according with the *ff14SB* force field[166] for the proteins, and *gaff* force field[113] for the ligands. Each complex was inserted in an octahedral box in periodic boundary conditions (PBC). The system was solvated with the explicit TIP3P water model[167] up to a distance of 10 Å from the solute and then was neutralized by replacing water molecules with an adequate number of Na^+ or Cl^- ions. In order to reach a stationary state and remove unwanted bad contacts or gaps between solvent and solute, the system was equilibrated and minimised with the steepest descent method. According to the reported protocol,[126] in the first step (1000 cycles with the steepest descent algorithm and 5000 cycles with the conjugated gradient algorithm) the geometry of hydrogen atoms was optimised, followed by optimisation of ions and water residues. The water box was then equilibrated at 300 K, first in NVT ensemble for 45 ps, then in NPT ensemble for other 45 ps of simulation. In the following steps i) sidechains, waters and ions were minimised and ii) the whole system was minimised (2500 cycles with the steepest descent algorithm and 5000 cycles o with the conjugated gradient algorithm). Restraints of 10.0 kcal/mol and 5.0 kcal/mol were applied respectively on the receptor backbone and the ligand atoms. Then the system was progressively heated, from 0 to 300 K, in six steps of 20 ps ($\Delta T = 50$ K). During heating, the restraints of both backbone and ligand were gradually reduced to 5.0 kcal/mol and 0.5 kcal/mol respectively. The restraints were then maintained for 100 ps cycle of system equilibration in NVT ensemble and for other 100 ps cycle of equilibration in NPT ensemble. Then, followed three steps of 100 ps each of system equilibration in NPT ensemble, in which restraints were gradually decreased, to 2.0 kcal/mol for backbone and 0.2 kcal/mol for ligand. Final equilibration was reached in 1 ns of NPT ensemble, where restraints were gradually removed (first 500 ps with 1.0 kcal/mol and 0.1 kcal/mol restraints for backbone and ligand respectively, and then no restraints during last 500 ps of equilibration). After these steps, the unrestrained complex was simulated over 1 ns production in NPT ensemble at 300 K. The time-step used was 2 fs, with a total of 500’000 MD time-steps. To reduce computational complexity, the hydrogen motion was restrained with the SHAKE algorithm[168] and an electrostatic cut-off of 8.0 Å was applied.

8.1.8 Nwat-MMGBSA Protocol

An ensemble of 100 snapshot equally spaced out were collected from the MD production runs for predicting the binding affinities through the Nwat-MMGBSA method. The Nwat-MMGBSA script was set up defining Nwat = 0, 30 and 60 explicit water molecules in the hydration shell. Molecular mechanics interaction energy and solvation free energy were calculated for all specified frames using the MMPBSA.py script of the AMBER package.[120] For the solvation free energy calculations, the unsolvated topology and trajectory files of complex, receptor and ligand, were created according to the Nwat-MMGBSA protocol.[126] Specifically, the command 'closest' was implemented in *cpptraj* module to retain within the topology and trajectory files of both unsolvated receptor and complex, the selected number of water molecules surrounding the ligand, defined by the "Nwat" variable. In such way, the calculation of $\Delta G_{\text{bind}}^{\text{solv}}$ explicitly takes into account, as part of the receptor, the hydration shell comprising the Nwat water residues (0, 30 and 60) that are the closest to the ligand in each frame of the MD simulation.

8.2 AZPP-Based computational part

8.2.1 AZPP Analogues Database Preparation

The MOE 'Database Viewer' was used to convert the set of molecules, drawn from 2D Sketcher and expressed as 1D SMILES strings, into 3D single low energy conformer database. MOE 'Builder' application was also used to implement and refine the structures. In addition, the 'MedChem Transformations' of MOE software was employed. The four AZPP ligands were extracted from the crystal structures, available in the Protein Data Bank (PDB IDs: 4D4K, 4D4J, 4D4M, 4D4L), and used as template input structures. The application was run with using default settings for technical parameters and filtering criteria, comprising: molecular weights lower than 450 Da, TPSA (topological polar surface area) in the range of 40 - 100 Å^2 and absence of reactive groups. Each ligand gave rise to 37 newly created analogues collected in an output database. The 4 output databases were merged together with the other unique sub-set of analogues manually constructed. Finally, the molecules were prepared for docking by applying the upfront 'wash' and 'minimisation' with MOE software, like described above (**paragraph 8.1.1**)

8.2.2 Similarity Search and Database Preparation

The 3D similarity search was performed through the LigCSRre method provided by the online engine wwLigCSRre of the RPBS web portal.[83,89,135] The three-dimensional mol2 file of compound **24**, extracted from crystal structure (PDB ID: 4D4J), was used as input ligand query. The search was performed over a dedicated 3D conformational library of 942 drug-like small molecules derived from DrugBank source collection.[137] A maximum of 50 best-ranked ligands were retrieved in their matching conformation. After the mining process, the resulting list of aligned hits, expressed in 3D format (mol2 file), was downloaded and merged in a database using the 'Database Viewer' of MOE. After a visual inspection by superimposition onto the query, the molecules were refined by applying the 'wash' and 'minimisation' preparation protocol with MOE software like described above (**paragraph 8.1.1**).

8.2.3 Molecular Docking Protocol

The calculations were performed on a multiprocessor HPC cluster (High Performance Computing service provided by University of Aberdeen), subdividing the task on a chosen number of processors for a PBS queuing system. Molecular docking of the constructed database of AZPP analogues was performed using PLANTS software and the ChemPLP scoring function in a low speed/high accuracy mode. No new conformations of the ligands were generated before docking.

The dimeric model the PFKFB3 enzyme has been prepared by Xiao Hu, Early Stage Researcher of the Moglynet Project (University of Milan, Italy).

The tandem model of PFKFB3 (extracted from crystal structure PDB ID: 2AXN) was used as the rigid receptor for the virtual screening. Docking was focussed on the whole empty kinase site which 3D coordinates were defined by the centre of mass of crystallographic ATP and F6P, and limited inside a sphere of 15.0 Å radius. The endogenous ligands were kept *in situ* inside the other binding sites of the dimer: F2,6P in the phosphatase domains; ATP and F6P in one of the two kinase domains.

After the docking posing process, the best ChemPLP score poses of each docked ligands of the database (i.e. the lowest energy geometry) was kept for further calculations. The outcome was re-scored according to scoring function values, resulting in a ranked database.

Biology

The following description has been provided by Helena Macut, Early Stage Researcher of the Moglynet Project (University of Milan, Italy).

8.3 Kinase Assay

The Kinase assay was performed with ADP-Glo^{TM} Kinase assay purchased from Promega, [161] on a 384-well white plate, purchased from Greiner. The kinase base buffer containing 100 mM HEPES (pH 7.0), 400 mM KCl, 10 mM potassium phosphate (pH 7.0), 10 mM $MgCl_2$, 1 mM dithiothreitol, 0.2 mM Triton X100 was prepared prior to the assay. The recombinant PFKFB3 enzyme, expressed in *Escherichia coli*, was added to the base buffer immediately prior to starting the assay. Each well received 8 µL of enzyme diluted in the base buffer (final concentration in the assay was 150 nM).

The tested compound was added to each well at 100 µM and at 10 µM (compounds were dissolved in kinase buffer, with max. 5% DMSO). After 30 min incubation, each well received 8 µL of a mixture containing 100 µM ATP and 200 µM F6P. The reactions were left to incubate for 1 hour before the addition of 16 µL of ADP-Glo^{TM} Reagent that terminates the kinase reaction and depletes the remained ATP. After the second incubation of 50 minutes at room temperature, 32 µL Kinase Detection Reagent was added. This reagent converts ADP to ATP and allows the newly synthesised ATP to be measured using a luciferase/luciferin reaction. After an additional hour, reaction mixtures were transferred to a Greiner white 384 well plate in triplicates (AZI-derivatives) or duplicates (AZPP-derivatives), and Tecan Magellan spectrometer® with enhanced luminescence module was used to measure the luminescence signal generated in each well. The signal, expressed in RLU (relative light units), is proportional to the ADP concentration produced and is correlated with kinase activity. Since compounds showed low activity in comparison the blank, only residual activity was calculated.

8.3.1.1 Sensitivity and linearity test

Prior to compounds evaluation, a sensitivity and linearity test of the kinase assay were performed. Three ATP to ADP standard conversion curves were prepared at 1 mM, 100 µM and 10 µM concentration of (ATP + ADP) standards (**Figure 25**), in 25 µL of reaction buffer. The reactions for standard curves were prepared in 96-well plate. The preparation of the 1 mM series of (ATP + ADP) standards in wells A1-12 is presented in **Table 7**. The wells B1-12 and C1-12 received 90 µL of the kinase reaction buffer. The 100 µM series was prepared by transferring 10 µL of the sample in well A1 to well B1, well A2 to B2, etc. The 10 µM series was prepared by transferring 10 µL of the sample in well B1 to well C1, well B2 to C2, etc.

Well number	**A1**	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**A10**	**A11**	**A12**
1 mM ADP (µL)	100	80	60	40	20	10	5	4	3	2	1	0
1 mM ATP (µL)	0	20	40	60	80	90	95	96	97	98	99	100

Table 7 : Preparation of the 1 mM series of (ATP + ADP) standards.

25 µL of standard from each well were transferred into separate wells of the new assay plate, and 25 µL of ADP-Glo^{TM} Reagent were added to each well. After 40 min incubation at room temperature, 50 µL of Kinase Detection Reagent were added to each well. Luminescence was measured after additional 1 h incubation at room temperature and standard curves were generated. A linear relationship was confirmed between the luminescent signal and the amount of ADP in the reaction buffer for all three (ATP + ADP) concentration series tested.

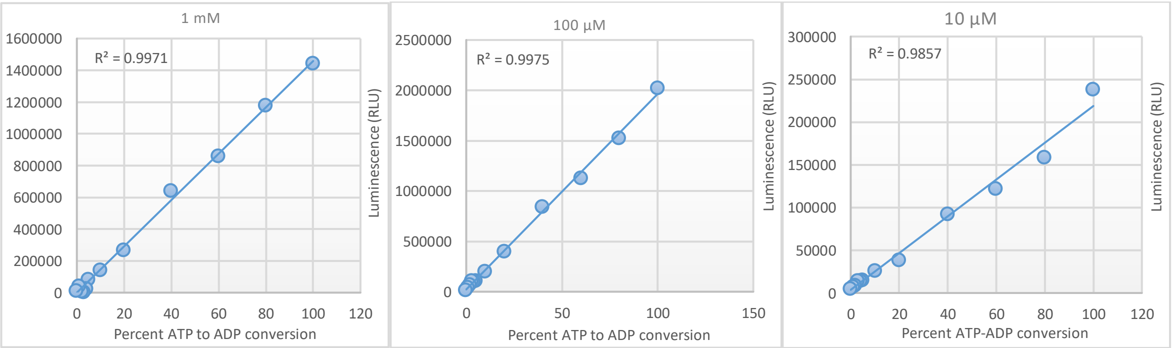


Figure 25: 1 mM, 100 µM and 10 µM standard curves.

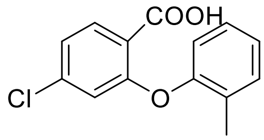
Chemistry

General methods

Dry solvents were obtained from commercial sources and used without further purification. Reactions performed in dry solvents were carried out under nitrogen atmosphere. NaH was used as a 60% dispersion in mineral oil. Reactions were monitored by thin-layer chromatography (TLC), unless otherwise noted. TLCs, included those used for purification, were performed on Merck silica gel glass plates (60 F_{254}). Visualisation was accomplished by UV light (254 nm) and/or staining with a ceric ammonium molybdate, $KMnO_4$ or ninidrine solution. Flash chromatography was performed using Silica gel (60 Å, particle size 40–63 µm) purchased from Merck. ^1H NMR and ^{13}C NMR spectra were recorded on a Bruker AVANCE III 400 NMR spectrometer and calibrated using residual undeuterated solvent as internal reference. ^{19}F NMR spectra were recorded on a Bruker AVANCE III 400 NMR spectrometer and were referenced to $CFCl_3$. ^{13}C NMR spectra were recorded with complete proton decoupling. Chemical shifts (δ) are reported in parts per million (ppm) and coupling constants (*J*) are given in Hertz (Hz). The following abbreviations are used for spin multiplicity: s = singlet, d = doublet, t = triplet, q = quartet, dd = doublet of doublet, ddd = doublet of doublet of doublet, m = multiplet, bs = broad signal. When necessary, resonances were assigned using two-dimensional experiments (COSY, HMBC and HSQC). Mass Analyses (MS) were performed using Agilent 1200 HPLC system coupled to Agilent G6120 single quadrupole detector equipped with an electrospray ionisation (ESI) source in direct infusion modality. ESI-MS spectra were recorded in positive mode unless otherwise noted. RP- (reverse phase)-HPLC-MS analyses were performed with an Agilent 1200 HPLC system equipped with a DAD and an ESI-MS detector. HPLC conditions for analytical analyses: Phenomenex Luna C18 column, 5 µm, 100 Å, 250 × 4.6 mm (L × ID) injected volume 10 µL, flow rate 1 mL/min, unless otherwise specified. HPLC preparative purification were performed using a Agilent 1260 HPLC system equipped with DAD detector with reverse phase, Phenomenex Luna C18 column, 5 µm, 100 Å, 250 × 21.2 mm (L × ID), flow rate 20 mL/min.

8.4 Target Molecule 1

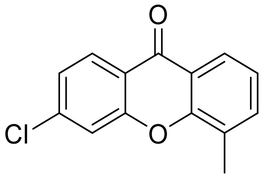
8.4.1 4-Chloro-2-phenoxybenzoic acid (35)



35

To a round bottom flask containing the carboxylic acid (300 mg, 1.6 mmol) and *o*-cresol **34** (340 mg, 3.1 mmol) dissolved in DMA (3 mL), 2.5 equivalents of NaH (60% in mineral oil), were added at 0 °C. The mixture was stirred for 20 minutes under nitrogen atmosphere at r.t., then Cu_2O (224.6 mg, 1.6 mmol) was added and the temperature was increased to 150 °C. After 4 h, the mixture was diluted with EtOAc, neutralized with a solution of HCl 1N and washed three times with water. In order to remove the excess of unreacted *o*-cresol, a saturated solution of $NaHCO_3$ (pH 8) was added to the organic phase. The basic aqueous layer containing the sodium salt of **35** was separated and then acidified to pH 3 with an aqueous solution of HCl 2N and extracted with DCM (3 x 10 mL). The organic phase was dried over Na_2SO_4, filtered and the DCM was removed under reduced pressure, to obtain 320 mg of reaction crude containing **35** that was immediately used for the following step, without further purification. MS (ESI, *m/z*): $C_{14}H_{12}ClO_3^+$ $[M+H]^+$ calc.: 263.05, found: 263.0 $[M+H]^+$, 285.0 $[M+Na]^+$, 547.1 $[2M+Na]^+$. ^1H NMR (400 MHz, $CDCl_3$) δ 8.16 (d, *J* = 8.5 Hz, 1H, aromat.), 7.38 – 7.22 (m, 3H, aromat.), 7.16 (dd, *J* = 8.5, 1.9 Hz, 1H, aromat.), 7.05 (dd, *J* = 7.8, 1.3 Hz, 1H, aromat.), 6.63 (d, *J* = 1.9 Hz, 1H, aromat.), 2.23 (s, 3H, CH_3). ^{13}C NMR (101 MHz, $CDCl_3$) δ 165.0, 158.0, 151.4, 141.0, 134.9, 132.5, 130.6, 128.2, 126.9, 123.7, 121.4, 116.9, 115.9, 16.2. The ^1H and ^{13}C NMR data matched with the data reported in the literature.[139,140]

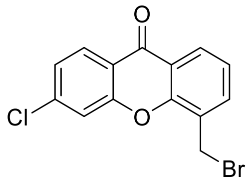
8.4.2 3-chloro-5-methyl-9H-xanthen-9-one (36)



36

An ice-cooled conc. H_2SO_4 (98% w/v, 8 mL) was added dropwise at 0 °C to the crude mixture containing acid **35**. The reaction mixture was gradually heated up to 100 °C and stirred for 3 h. The reaction was stopped, diluted with water and cooled down to r.t., then the pH was adjusted to 9 by carefully adding a 2N solution of NaOH at 0 °C. Finally, the product was extracted with DCM (3 x 10 mL) and dried giving **36** as a white solid (242 mg, 55% yield over two steps). TLC R_f = 0.5 (solvent system: *n*-Hex/EtOAc, 10:1). MS (ESI, *m/z*): $C_{14}H_{10}ClO_2^+$ $[M+H]^+$ calc.: 245.04, found: 245.0. ^1H NMR (400 MHz, $CDCl_3$) δ 8.25 (d, *J* = 8.5 Hz, 1H, aromat.), 8.15 (d, *J* = 8.0 Hz, 1H, aromat.), 7.61 – 7.49 (m, 2H, aromat.), 7.33 (dd, *J* = 8.5, 1.9 Hz, 1H, aromat.), 7.31 – 7.25 (m, 1H, aromat.), 2.53 (s, 3H, CH_3). ^{13}C NMR (101 MHz, $CDCl_3$) δ 176.8, 156.3, 154.5, 140.8, 136.1, 128.2, 127.4, 124.9, 124.4, 123.9, 121.7, 120.3, 118.2, 15.8. The ^1H and ^{13}C NMR data matched with the data reported in the literature.[139]

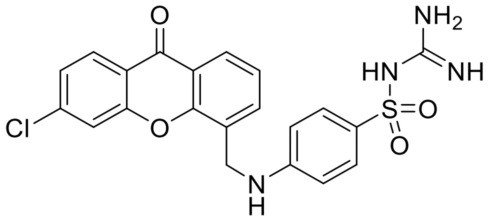
8.4.3 5-(bromomethyl)-3-chloro-9H-xanthen-9-one (37)



37

A mixture of **36** (50 mg, 0.2 mmol), NBS (41 mg, 0.3 mmol) and BPO (75% remainder water) in CCl_4 was refluxed for 2 h. After cooling down to r.t. the solvent was removed under reduced pressure and the crude mixture was purified by isocratic flash chromatography (solvent system: *n*-Hex/EtOAc, 4:1) to afford **37** as a white solid (50 mg, 76% yield). MS (ESI, *m/z*): $C_{14}H_9BrClO_2^+$ $[M+H]^+$ calc.: 322.95 found: $[M+H]^+$ 322.9. ^1H NMR (400 MHz, $CDCl_3$) δ 8.31 (dd, *J* = 8.0, 1.7 Hz, 1H, aromat.), 8.28 (d, *J* = 8.5 Hz, 1H, aromat.), 7.79 (dd, *J* = 7.4, 1.7 Hz, 1H, aromat.), 7.65 (d, *J* = 1.9 Hz, 1H, aromat.), 7.42 – 7.35 (m, 2H, aromat.), 4.80 (s, 2H, CH_2). ^{13}C NMR (101 MHz, $CDCl_3$) δ 176.1, 156.0, 153.8, 141.3, 136.2, 128.3, 127.7, 127.1, 125.4, 124.3, 122.4, 120.4, 118.3, 26.3. The ^1H and ^{13}C NMR data matched with the data reported in the literature.[139]

8.4.4 *N*-carbamimidoyl-4-(((6-chloro-9-oxo-9H-xanthen-4-yl) methyl)amino)benzenesulfonamide (17)

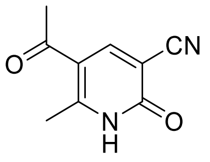


17

To a solution of **37** (55 mg, 0.17 mmol) in dry DMF (1.5 mL), the suphaguanidine **38** (38.4 mg, 0.18 mmol) and K_2CO_3 (24 mg, 0.17 mmol) were added at room temperature. After stirring for 24 h, the mixture was diluted with EtOAc and washed three times with water (3 x 10 mL). The organic phase was dried over Na_2SO_4, filtered and the solvent was removed under reduced pressure. The crude mixture was purified by gradient elution flash chromatography (solvent system: DCM/MeOH, gradient: from 0% to 3% of MeOH) to afford **17** as a white solid (31 mg, 40% yield). Analytical RP-HPLC: Phenomenex Gemini C18 column, 5 µm, 110 Å, 250 × 4.60 mm (L × ID) (solvent system: H_2O/ACN, gradient: from 40 % to 80% of ACN in 15 min, flow: 1 mL/min; t_R = 9.4 min). MS (ESI, *m/z*): $C_{21}H_{18}ClN_4O_4S^+$ $[M+H]^+$ calc.: 457.07, found: 457.1 $[M+H]^+$, 479.1 $[M+Na]^+$, 495.0 $[M+K]^+$, 913.2 $[2M+H]^+$. ^1H NMR (400 MHz, DMSO-d_6) δ 8.19 (d, *J* = 8.6 Hz, 1H, aromat.), 8.08 (dd, *J* = 7.9, 1.6 Hz, 1H, aromat.), 8.00 (d, *J* = 1.9 Hz, 1H, aromat.), 7.80 (dd, *J* = 7.3, 1.6 Hz, 1H, aromat.), 7.55 (dd, *J* = 8.6, 1.9 Hz, 1H, aromat.), 7.49 – 7.39 (m, 3H, aromat.), 6.99 (t, *J* = 6.2 Hz, 1H, N*H*), 6.66 (d, *J* = 8.8 Hz, 2H, aromat.), 6.62 – 6.44 (bs, 4H, guanid.), 4.69 (d, *J* = 6.2 Hz, 2H, C*H*$_2$). ^{13}C NMR (101 MHz, DMSO-d_6) δ 175.4, 157.8, 155.7, 153.6, 150.4, 139.7, 133.9, 131.4, 128.5, 127.8, 127.2, 124.9, 124.6, 124.2, 121.2, 119.9, 118.5, 111.0, 40.6. The ^1H and ^{13}C NMR data matched with the data reported in the literature.[138]

8.5 Target Molecule 2

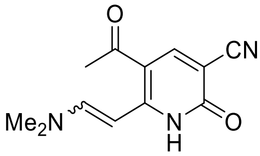
8.5.1 5-Acetyl-6-methyl-2-oxo-1,2-dihydropyridine-3-carbonitrile (41)



41

A mixture of acetylacetone **39** (100 mg, 1mmol) and *N*,*N*’-dimethylformamide dimethyl acetal (DMFDMA; 133 µL, 1 mmol) in dry DMF was stirred at r.t under nitrogen atmosphere for 24 h. The resulting yellow mixture, containing the adduct dimetylamino pentane dione **40** was used immediately for the following step. TLC R_f= 0.4 (solvent system: DCM/MeOH, 10:1) MS (ESI, *m/z*): $C_8H_{14}NO_2^+$ $[M+H]^+$ calc.: 156.10, found: 156.1 $[M+H]^+$, 178.1 $[M+Na]^+$. In another flask, a mixture of NaH (60% in mineral oil) (80 mg, 2 mmol) and cyanocetamide (84 mg, 1 mmol) in dry DMF was stirred at r.t. under nitrogen atmosphere for 10 minutes. The adduct **40** from the first flask was added to the mixture and it was allowed to react at r.t. for 20 h. 5 mL of H_2O/EtOH (1:1) were added followed by concentrated HCl (37%) to adjust the pH to 3. After 24 h the acidic solution was diluted with H_2O and was extracted with EtOAc (3 x 8 mL). The organic phase was washed several times with H_2O dried over Na_2SO_4, filtered and the solvent was removed under reduced pressure. The product was isolated by crystallisation from EtOH to obtain 175 mg of **41** (98 % yield) as a white solid. The ^{1}H NMR data matched with the data reported in the literature. MS (ESI, *m/z*): $C_9H_9N_2O_2^+$ $[M+H]^+$ calc.: 177.07, found: 177.1 $[M+H]^+$, 199.1 $[M+Na]^+$. ^{1}H NMR (400 MHz, $CDCl_3$) δ 12.36 (bs, 1H, N*H*), 8.33 (s, 1H, aromat.), 2.81 (s, 3H, C*H*$_3$), 2.52 (s, 3H, COC*H*$_3$). ^{1}H NMR (400 MHz, Acetone-d$_6$) δ 11.62 (bs, 1H, N*H*), 8.60 (s, 1H, aromat.), 2.72 (s, 3H, C*H*$_3$), 2.54 (s, 3H, COC*H*$_3$). ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 12.94 (s, 1H, N*H*), 8.65 (s, 1H, aromat.), 2.55 (s, 3H, C*H*$_3$), 2.45 (s, 3H, COC*H*$_3$). ^{13}C NMR (101 MHz, Acetone-d$_6$) δ 195.16, 160.23, 159.29, 150.43, 116.45, 116.08, 101.30, 28.97, 20.70. ^{13}C NMR (101 MHz, DMSO-d_6) δ 195.0, 159.7, 158.6, 149.8, 115.9, 115.3, 99.1, 28.9, 19.9. The ^{1}H and ^{13}C NMR data matched with the data reported in the literature.[146]

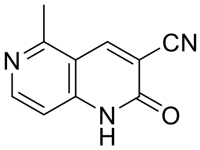
8.5.2 (E,Z)-5-Acetyl-6-(2-(dimethylamino)vinyl)-2-oxo-1,2-dihydropyridine-3-carbonitrile (42)



42

A mixture containing the pyridinone **41** (250 mg, 1.4 mmol) and DMFDMA (207 µL, 1.6 mmol) in DMF was stirred under nitrogen atmosphere at r.t. over-weekend. After dilution with MeOH, a bright yellow solid precipitated from the mixture. The solid was filtrate, washed with acetone, dried over Na_2SO_4, filtered and used immediately for the following step. MS (ESI, *m/*z): $C_{12}H_{14}N_3O_2^+$ [M+H]$^+$ calc.: 232.11, found : [M+H]$^+$ 232.1, [M+Na]$^+$ 254.1. ^{1}H NMR (400 MHz, DMSO-d_6) δ 11.37 (bs, 1H, N*H*), 8.33 (s, 1H, aromat.), 8.22 (d, *J* = 13.2 Hz, 1H, aromat.), 6.63 (d, *J* = 13.2 Hz, 1H, aromat.), 3.04 (bs, 6H, N(C*H*$_3$)$_2$), 2.38 (s, 3H, C*H*$_3$). The ^{1}H and ^{13}C NMR data matched with the data reported in the literature.[145]

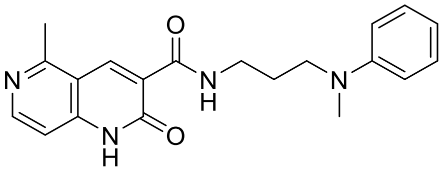
8.5.3 5-Methyl-2-oxo-1,6-naphthyridine-3-carbonitrile (43)



43

Four equivalents of NH_4OAc (438 mg, 5.6 mmol) were added to a solution of enamine **42** (320 mg, 1.4 mmol) in DMF. The mixture was gradually heated in an oil bath to 125 °C for 24 h. After cooling down to r.t., the brown cake mixture was washed several times with H_20 and extracted with Et_2O. The organic phase was dried over Na_2SO_4, filtered and the Et_2O was removed under reduced pressure to obtain 147 mg of **43**, a white solid (79% yield over two steps). The ^1H NMR data matched with the data reported in the literature. Analytical RP-HPLC (solvent system: H_2O/ACN, gradient: from 3% to 10% of ACN in 10 min, flow: 1 mL/min; $t_R = 8.3$). $C_{10}H_8N_3O^+$ $[M+H]^+$ calc.: 186.07, found: $[M+H]^+$ 186.1, $[M+Na]^+$ 208.2, $[M+K]^+$ 224.0. ^1H NMR (400 MHz, MeOD) δ 8.89 (s, 1H, aromat.), 8.40 (d, *J* = 6.0 Hz, 1H, aromat.), 7.16 (d, *J* = 6.0 Hz, 1H), 2.80 (s, 3H, C*H*$_3$). ^1H NMR (400 MHz, DMSO-d$_6$) δ 12.57 (bs, 1H, N*H*), 8.98 (s, 1H, aromat.), 8.42 (d, *J* = 5.8 Hz, 1H, aromat.), 7.10 (d, *J* = 5.8 Hz, 1H, aromat.), 2.72 (s, 3H, C*H*$_3$). ^{13}C NMR (101 MHz, DMSO-d$_6$) δ 159.2, 158.7, 150.8, 147.5, 145.1, 115.5, 112.5, 108.6, 106.6, 21.5. The ^1H and ^{13}C NMR data matched with the data reported in the literature.[145]

8.5.4 5-Methyl-*N*-(3-(methyl(phenyl)amino)propyl)-2-oxo-1,6-naphthyridine-3-carboxamide (18)

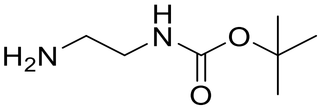


18

A mixture of nitrile **43** (100 mg, 0.5 mmol), *N*-(3-aminopropyl)-*N*-methylaniline **44** (105 mg, 0.6 mmol) and $CsOH{\cdot}H_2O$ (8.1 mg, 0.05 mmol) in H_2O (0.5 mL) was sealed in a Schlenk tube under nitrogen atmosphere and gradually heated up to 145 °C. After 24 h the reaction mixture was diluted with water and extracted three times with DCM and once with EtOAc. The organic phases were collected and dried over Na_2SO_4, filtered and, after removing the solvent, the resulting crude was purified by gradient elution flash chromatography (solvent system: DCM/EtOH, from 0% to 5% of EtOH) to give **18** (86 mg, 45% yield) as pale-yellow solid. Analytical RP-HPLC (solvent A: H_2O + 0.1% TFA, solvent B: ACN + 0.1% TFA, gradient: from 10% to 50% of B in 10 min flow: 1 mL/min; t_R = 6.5 min). MS (ESI, *m/z*): $C_{20}H_{23}N_4O_2^+$ $[M+H]^+$ calc.: 351.18, found: 176.2 $[M+2H]^{2+}$, 351.2 $[M+H]^+$, 373.1 $[M+Na]^+$, 389.1 $[M+K]^+$. ^1H NMR (400 MHz, $CDCl_3$) δ 10.62 (s, 1H, N*H*), 9.52 (t, *J* = 5.5 Hz, 1H, N*H*), 9.21 (s, 1H, aromat.), 8.47 (d, *J* = 5.8 Hz, 1H, aromat.), 7.24 – 7.17 (m, 2H, aromat.), 7.00 (d, *J* = 5.8 Hz, 1H, aromat.), 6.77 – 6.70 (m, 2H, aromat.), 6.70 – 6.64 (m, 1H, aromat.), 3.62 – 3.50 (m, 2H, C*H*$_2$), 3.51 – 3.41 (m, 2H, C*H*$_2$), 2.96 (s, 3H, C*H*$_3$), 2.90 (s, 3H, C*H*$_3$), 2.02 – 1.90 (m, 2H, C*H*$_2$). ^{13}C NMR (101 MHz, $CDCl_3$) δ 163.1, 162.8, 160.5, 150.3, 149.3, 143.9, 142.5, 129.4, 122.8, 116.6, 114.3, 112.6, 108.1, 50.6, 38.7, 37.9, 27.0, 22.1.

8.6 Target Molecule 3

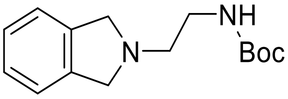
8.6.1 *tert*-Butyl (2-aminoethyl)carbamate (52)



52

A solution of di-*tert*-butyl dicarbonate, Boc_2O (1 g, 4.6 mmol) in DCM was added, through a dropping funnel, over 1 h, to a solution of ethylenediamine (6 equivalents 2 mL, 27.5 mmol) pre-dissolved in DCM while maintaining vigorous stirring. The mixture was stirred at r.t. for 24 h, then, after concentration, the resulting oil was taken into aqueous solution of $NaHCO_3$ (5% w/v) and extracted three times with DCM (3 x 20 mL). The organic phase was dried over Na_2SO_4 and the DCM was removed under reduced pressure to afford 725 g of **52** in quantitative yield. MS (ESI, *m/z*): $C_7H_{17}N_2O_2^+$ $[M+H]^+$ calc.: 161.13, found: 161.1. ^1H NMR (400 MHz, $CDCl_3$) δ 5.34 (bs, 1H, N*H*), 3.13 – 2.87 (m, *J* = 43.3 Hz, 2H, CH_2), 2.70 – 2.53 (m, 2H, CH_2), 1.28 (s, 9H, OC(CH_3)$_3$). The ^1H and ^{13}C NMR data matched with the data reported in the literature.[169]

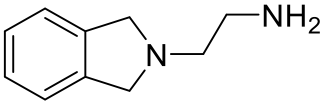
8.6.2 *tert*-Butyl (2-(isoindolin-2-yl)ethyl)carbamate (53)



53

To a solution of α,α'-Dibromo-*o*-xylene (104.3 mg, 0.39 mmol) in $CHCl_3$ (2 mL), a solution of DIPEA (172 µL) in $CHCl_3$ (1 mL) was added dropwise at 0 °C followed by another solution of *N*-Boc ethylenediamine **16** (95 mg, 0.59 mmol) in $CHCl_3$ (1 mL). The resulting mixture was stirred at r.t. overnight. When bromine was no longer detected by TLC, the reaction was stopped, and the solvent was removed under reduced pressure. The crude was washed three times with water and extracted with EtOAc. The organic phase was dried over Na_2SO_4 and the EtOAc was removed under reduced pressure to afford 100 mg of crude that was used without purification for the next step. TLC $R_f = 0.3$ (solvent system: *n*-Hex/EtOAc 1:1). MS (ESI, *m/z*): $C_{15}H_{23}N_2O_2^+$ $[M+H]^+$ calc.: 263.18, found: 263.2. ^1H NMR (MeOD, 400 MHz): $\delta = 7.23$ (m, 4H, aromat.), 3.97 (s, 4H, benzylic), 3.29 (m, 2H, CH_2), 2.85 (m, 2H, CH_2), 1.46 (s, 9H, $OC(CH_3)_3$); ^1H NMR ($CDCl_3$, 400 MHz): $\delta = 7.22$ (s, 1H, aromat.), 5.15 (bs, 1H, N*H*), 3.97 (s, 4H, benzylic), 3.34 (m, 2H, CH_2), 2.89 (m, 2H, CH_2), 1.46 (s, 9H, $OC(CH_3)_3$).^{13}C NMR (101 MHz, MeOD) $\delta = 158.4$, 140.6, 128.1, 123.3, 80.1, 60.0, 56.4, 40.2, 28.8.

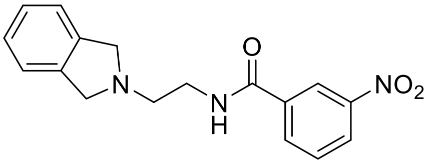
8.6.3 2-(Isoindolin-2-yl)ethan-1-amine (54)



54

Compound **53** (80 mg, 0.3 mmol) was treated with a solution of TFA in DCM (40% v/v), at 0 °C, for 2 h. The solvent was removed by bubbling nitrogen and stripping with DCM several times. 8 mL of aqueous solution of HCl 1N was added and the mixture was lyophilised. The resulting crude was treated with a 1N solution of NaOH then extracted with DCM to remove the counter ion and give the free amine of **54** (45 mg, 94% yield). MS (ESI, *m/z*): $C_{10}H_{15}N_2^+$ $[M+H]^+$ calc.: 163.12 found: 163.1. ^1H NMR (400 MHz, MeOD) δ 7.21 (qd, *J* = 6.7, 4.3 Hz, 4H, aromat.), 3.99 (s, 4H, benzylic), 3.12 – 3.04 (m, 2H, CH_2), 3.03 – 2.94 (m, 2H, CH_2). The ^1H and ^{13}C NMR data matched with the data reported in the literature.[170]

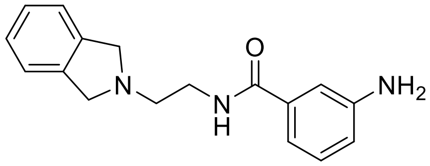
8.6.4 *N*-(2-(Isoindolin-2-yl)ethyl)-3-nitrobenzamide (55)



55

The amine **54** (40 mg, 0.14 mmol) was dissolved in THF at 0 °C then 3-nitrobenzoyl chloride (50 mg, 0.3 mmol) was added to the solution and the reaction was warmed up to r.t. and stirred for 8 h. The reaction was quenched with a saturated aqueous solution of $NaHCO_3$ and extracted with EtOAc. The crude material was purified by isocratic flash chromatography (solvent system: DCM/MeOH, 9.5:0.5) to give **55** (30 mg, 68% yield). TLC R_f = 0.3 (solvent system DCM/MeOH 9.5:0.5). MS (ESI, *m/z*): $C_{17}H_{18}N_3O_3^+$ $[M+H]^+$ calc.: 312.13, found: 312.2. ^1H NMR (400 MHz, MeOD) δ 8.74 (t, *J* = 2.0 Hz, 1H, aromat.), 8.37 (ddd, *J* = 8.0, 2.0, 1.0 Hz, 1H, aromat.), 8.28 – 8.21 (m, 1H, aromat.), 7.71 (t, *J* = 8.0 Hz, 1H, aromat), 7.27 – 7.16 (m, 4H, aromat.), 4.02 (s, 4H, benzylic), 3.67 (t, *J* = 6.6 Hz, 2H, C*H*$_2$), 3.02 (t, *J* = 6.6 Hz, 2H, C*H*$_2$). ^{13}C NMR (101 MHz, MeOD) δ 167.7, 149.7, 140.6, 137.3, 134.4, 130.9, 128.2, 126.9, 123.3, 123.3, 59.9, 55.8, 39.8.

8.6.5 3-Amino-*N*-(2-(isoindolin-2-yl)ethyl)benzamide (56)



56

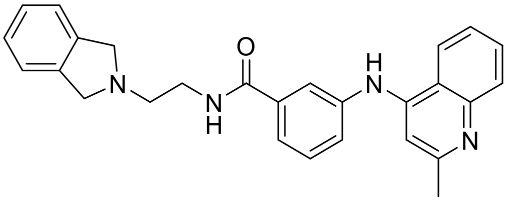
METHOD A

The nitro-compound **55** (30 mg, 0.1 mmol) was dissolved in EtOH (1.5 mL) and AcOH (0.5 mL) was added at 0 °C, followed by Fe (80 mg, 1.4 mmol). The reaction was warmed up to r.t. and vigorously stirred for 24 h. After dilution with EtOAc, the mixture was washed three times with water and three time with an aqueous solution of $NaHCO_3$ (5% w/v). The organic phase was dried over Na_2SO_4, filtered and the reaction crude was purified by gradient elution flash chromatography (solvent system: DCM/MeOH, from 0% to 10% of MeOH) to afford the aniline **56** (20 mg, 71% yield).

METHOD B

The nitro-compound **55** (100 mg, 0.32 mmol) was dissolved in MeOH/EtOAc (20:1), then $NiCl_2{\cdot}H_2O$ (125.8 mg, 0.64 mmol) was added to the solution. The mixture was stirred for 20 min then $NaBH_4$ (73 mg, 1.9 mmol) was added at 0 °C. The reaction was warmed up to r.t. and stirred for 1 h, then quenched with a saturated aqueous solution of NH_4Cl and extracted three times with EtOAc (3 x 20 mL). The organic layer was dried over Na_2SO_4, filtered and the solvent was removed under reduced pressure to give the aniline **56** (80 mg, 90% of yield). MS (ESI, *m/z*): $C_{17}H_{20}N_3O^+$ $[M+H]^+$ calc.: 282.16, found: 141.6 $[M+2H]^{2+}$, 282.1 $[M+H]^+$, 304.1 $[M+Na]^+$, 585.3 $[2M+Na]^+$. ^1H NMR (400 MHz, MeOD) δ 7.26 – 7.19 (m, 4H, aromatic.), 7.18 – 7.08 (m, 3H, aromatic.), 6.84 (ddd, *J* = 8.0, 2.2, 1.2 Hz, 1H, aromatic), 4.01 (s, 4H, benzylic), 3.60 (t, *J* = 6.8 Hz, 2H, C*H*$_2$), 2.98 (t, *J* = 6.8 Hz, 2H, C*H*$_2$). ^{13}C NMR (101 MHz, MeOD) δ 170.9, 149.4, 140.6, 136.5, 130.2, 128.1, 123.3, 119.3, 117.3, 114.9, 60.0, 55.9, 39.6.

8.6.6 *N*-(2-(Isoindolin-2-yl)ethyl)-3-((2-methylquinolin-4-yl)amino)benzamide (19)

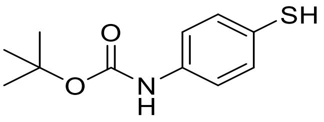


19

To a solution of **56** (30 mg, 0.1 mmol) and 4-chloroquinaldine **57** (36 mg, 0.2 mmol) in EtOH, 300 µL of concentrated HCl (37%) were added. The mixture was refluxed under nitrogen atmosphere for 3 h. The mixture was allowed to cool down to r.t. and, after removal of the solvent under reduced pressure, the residue was basified with a saturated solution of $NaHCO_3$ aq. and extracted three times with EtOAc (3 x 15 mL). The organic phase was dried over Na_2SO_4, filtered and, after removing the solvent under reduced pressure, the product was purified by preparative RP-HPLC (solvent A = H_2O + 0.1% TFA, solvent B = ACN + 0.1% TFA, gradient from 15% B to 40% B in 10 min; flow 20 mL/min; t_R: 8.8 min) to give **19** (30 mg, 65% yield). Analytical RP-HPLC (solvent A: H_2O + 0.1% of TFA, solvent B: ACN + 0.1% of TFA, gradient: from 15% to 40% of B in 10 min, flow: 1 mL/min, t_R = 9.2 min). MS (ESI, *m/z*): $C_{27}H_{27}N_4O^+$ $[M+H]^+$ calc.: 423.22, found: 212.2 $[M+2H]^{2+}$, 423.2 $[M+H]^+$, 445.2 $[M+Na]^+$. ^1H NMR (400 MHz, MeOD) δ 8.53 (d, *J* = 8.1 Hz, 1H, aromat.), 8.05 – 7.64 (m, 7H, aromat.), 7.41 (s, 4H, aromat.), 6.75 (s, 1H, aromat.), 4.86 (s, 4H, benzylic) 3.90 (t, *J* = 5.7 Hz, 2H, C*H*$_2$), 3.73 (t, *J* = 5.7 Hz, 2H, C*H*$_2$), 2.62 (s, 3H, C*H*$_3$). ^{13}C NMR (101 MHz, MeOD) δ 169.8, 156.8, 156.3, 140.0, 139.0, 136.9, 135.2, 134.9, 131.6, 130.2, 130.2, 128.3, 127.6, 126.0, 123.9, 123.9, 120.8, 117.8, 101.5, 60.2, 55.8, 37.4, 20.3.

8.7 Target Molecule 4

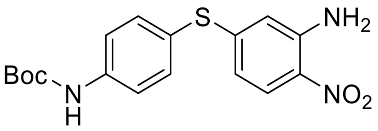
8.7.1 *tert*-Butil 4-mercaptophenylcabamate (60)



60

4-aminotiophenol **59** (200 mg, 1.6) was added to a solution of Boc$_2$O (349 mg) and catalytic amount (10 %) of InCl$_3$ (35.4 mg, 0.16 mmol) in DCM at r.t under nitrogen atmosphere. The yellow mixture was stirred at 35 °C for 2 h. The thick yellow oily mixture was then diluted with EtOAc and washed three times with H$_2$O. The organic layer was dried over Na$_2$SO$_4$, filtered and solvents were removed under reduced pressure. The crude mixture was purified by isocratic flash chromatography (solvent system: *n*-Hex/EtOAc, 8:1) to obtain 263 mg (1.16 mmol, 73% yield) of **60** as a yellow oil. ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.29 – 7.19 (m, 4H, aromat.), 6.62 (s, 1H, N*H*), 3.40 (s, 1H, S*H*), 1.51 (s, 9H, OC(C*H*$_3$)$_3$). ^{13}C NMR (101 MHz, CDCl$_3$) δ 152.5, 138.4, 131.1, 131.0, 130.8, 118.9, 80.9, 28.3. The ^{1}H and ^{13}C NMR data matched with the data reported in the literature.[171]

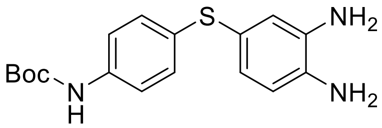
8.7.2 *tert*-Butyl (4-((3-amino-4-nitrophenyl)amino)-phenyl)carbamate (61)



61

To a mixture of Boc-protected aniline **60** (260 mg, 1.15 mmol) and Bu^{t}OK (142.5 mg, 1.27 mmol) in THF, 5-chloro-2-nitroaniline (368 mg, 1.27 mmol) was added at r.t. under nitrogen atmosphere. The red mixture was refluxed overnight. When the reaction was shown to be complete by TLC, it was stopped and cooled down to r.t.. After removal of the solvent, the reaction crude was purified by isocratic flash chromatography (solvent system *n*-Hex/EtOAc, 8:1) to obtain **61** as a yellow solid (382 mg, 92% yield). TLC R_f = 0.35 (solvent system *n*-Hex/EtOAc, 8:1). MS (ESI, *m/z*): $C_{17}H_{20}N_3O_4S^+$ [M+H]$^+$ calc.: 362.11, found: 362.1. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.96 (d, *J* = 9.1 Hz, 1H, aromat.), 7.47 (m, 4H, aromat.), 6.61 (bs, 1H, N*H*), 6.39 (dd, *J* = 9.1, 2.0 Hz, 1H, aromat.), 6.26 (d, *J* = 2.0 Hz, 1H, aromat.), 6.01 (bs, 2H, NH_2), 1.54 (s, 9H, OC(CH_3)$_3$). ^{13}C NMR (101 MHz, $CDCl_3$) δ 152.6, 150.5, 145.0, 140.2, 136.6, 129.9, 126.7, 122.9, 119.6, 115.1, 113.6, 81.4, 28.4.

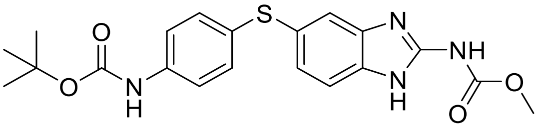
8.7.3 *tert*-Butyl (4-((3,4-diaminophenyl)amino) phenyl)carbamate (62)



62

Compound **61** (180 mg, 0.5 mmol) was dissolved in MeOH/EtOAc (20:1), then $NiCl_2 \cdot H_2O$ (237 mg, 1 mmol) was added to the solution. The mixture was stirred for 20 min then $NaBH_4$ (113.17 mg, 3 mmol) was added portion-wise at 0 °C. The reaction was warmed up to r.t. and stirred for 2.5 h, then quenched with aq. NH_4Cl and extracted with EtOAc. The organic layer was dried over Na_2SO_4, filtered and, after removing the solvent under reduced pressure, the crude was purified by gradient elution flash chromatography (solvent system: *n*-Hex/EtOAc, from 3.5:6.5 to 3:7) to give **62** (156 mg, 94% yield). TLC R_f = 0.5 (solvent system: *n*-Hex/EtOAc, 3:7). MS (ESI, *m/z*): $C_{17}H_{22}N_3O_2S^+$ $[M+H]^+$ calc.: 332.14, found: 332.0 $[M+H]^+$, 354.1 $[M+Na]^+$, 663.3 $[2M+H]^+$, 685.2 $[2M+Na]^+$. ^1H NMR (400 MHz, MeOD) δ 7.26 (d, *J* = 8.7 Hz, 2H, aromat.), 7.10 – 7.04 (m, 2H, aromat.), 6.78 (dd, *J* = 1.7, 0.5 Hz, 1H, aromat.), 6.67 (dd, *J* = 2.1, 1.1 Hz, 2H, aromat.), 1.50 (s, 9H, $OC(CH_3)_3$). ^1H NMR (400 MHz, $CDCl_3$) δ 7.24 – 7.14 (m, 3H, aromat.), 6.83 – 6.50 (m, 4H, aromat.), 3.42 (s, 4H, N*H*), 1.50 (s, 9H, $OC(CH_3)_3$). ^{13}C NMR (101 MHz, MeOD) δ 155.1, 138.4, 136.6, 136.5, 133.8, 130.4, 125.8, 123.9, 122.0, 120.3, 117.7, 80.9, 28.6.

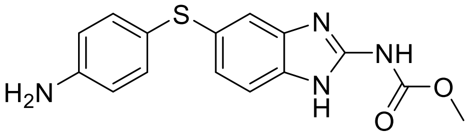
8.7.4 *tert*-Butyl (4-((2-((methoxycarbonyl)amino)-1H-enzo[d]imidazol-5-yl)amino)phenyl)carbamate (64)



64

The dianiline **62** (156 mg, 0.47 mmol) was dissolved in EtOH, then 1,3-bis(methoxycarbonyl)-2-methyl-thiopseudourea **63** (102 mg, 0.49 mmol) was added. The mixture was stirred at reflux temperature overnight. After cooling down to r.t., the product **64** was isolated as brown precipitate by filtration and it was used without further purification for the following step (136 mg, 70% yield). MS (ESI) $C_{20}H_{23}N_4O_4S^+$ $[M+H]^+$ calc. 415.14 found 415.1 $[M+H]^+$, 829.2 $[2M+H]^+$. ^{1}H NMR (400 MHz, DMSO-d_6) δ 11.62 (bs, 2H, N*H*C(N*H*)), 9.42 (bs, 1H, N*H*), 7.46 – 7.32 (m, 4H, aromat.), 7.24 – 7.15 (m, 2H, aromat.), 7.07 (dd, *J* = 8.2, 1.8 Hz, 1H, aromat.), 3.74 (s, 3H, OCH_3), 1.46 (s, 9H, OC(CH_3)$_3$). ^{13}C NMR (101 MHz, DMSO-d_6) δ 154.7, 152.7, 148.0, 138.7, 131.1, 128.8, 126.4, 124.6, 118.9, 79.2, 52.5, 28.1.

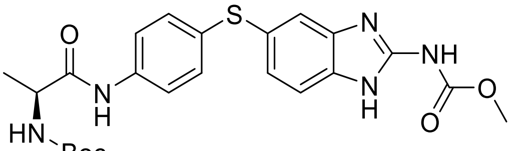
8.7.5 Methyl (5-((4-aminophenyl)amino)-1H-benzo[d]imidazol-2-yl)carbamate (65)



65

Compound **64** (100 mg, 0.24 mmol) was dissolved in a 4M solution of HCl in 1,4-dioxane (4 mL) at 0 °C. The mixture was warmed up to r.t. and stirred for 24 h. The solvent was removed under reduced pressure and the resulting **65** was immediately used for the next step (160 mg, quantitative yield). MS (ESI, *m/z*): $C_{15}H_{15}N_4O_2S^+$ $[M+H]^+$ calc.: 315.09, found: 315.1. ^1H NMR (400 MHz, MeOD) δ 7.73 (s, 1H, aromat.), 7.67 (d, *J* = 8.2 Hz, 1H, aromat.), 7.53 (d, *J* = 8.2 Hz, 1H, aromat.), 7.40 (dd, *J* = 17.7, 8.3 Hz, 4H, aromat.), 3.97 (s, 3H, OCH_3). ^{13}C NMR (101 MHz, MeOD) δ 154.0, 146.3, 139.6, 132.1, 131.9, 131.1, 131.1, 130.8, 130.3, 125.2, 118.2, 115.2, 54.8. The ^1H and ^{13}C NMR data matched with the data reported in the literature.[149]

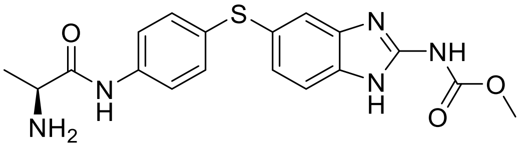
8.7.6 Methyl (S)-(5-((4-(2-((tert-butoxycarbonyl)amino)propanamido)phenyl)amino)-1H-benzo[d]imidazol-2-yl)carbamate (66)



66

Boc-L-Alanine (14 mg, 0.1 mmol) was dissolved in DCM then HATU (56 mg, 0.2 mmol) and DIPEA (40 µL, 0.3 mmol) were added to the solution at r.t.. The reaction mixture was stirred for 20 min, then the aniline **65** (24 mg, 0.1 mmol) was added. After 4 h an 1N aqueous solution of HCl was added and the mixture was extracted with EtOAc. The crude material was purified by preparative RP-HPLC (solvent A = H_2O + 0.1% TFA, solvent B = ACN + 0.1% TFA, gradient from 40% to 80% of B in 15 min; flow 20 mL/min; t_R: 7.2 min) to give **66•TFA** (24 mg, 40% yield). Analytical RP-HPLC (solvent A: H_2O + 0.1% of TFA, solvent B: ACN + 0.1% of TFA, gradient: from 40% to 80% of B in 15 min, flow: 1 mL/min, t_R = 7.6 min). MS (ESI, *m/z*): $C_{23}H_{28}N_5O_5S^+$ $[M+H]^+$ calc.: 486.18, found: 486.2 $[M+H]^+$, 971.4 $[2M+H]^+$, 993.3 $[2M+Na]^+$. ^1H NMR (400 MHz, DMSO-d_6) δ 11.70 (bs, 2H, N*H*C(N*H*)), 9.99 (s, 1H, N*H*), 7.56 (d, *J* = 8.7 Hz, 2H, aromat.), 7.44 – 7.34 (m, 1H, aromat.), 7.20 (d, *J* = 8.7 Hz, 2H, aromat.), 7.12 – 6.95 (m, 2H, aromat.), 4.09 (m, 1H, C*H*), 3.75 (s, 3H, OC*H*$_3$), 1.37 (s, 9H, OC(C*H*$_3$)$_3$), 1.25 – 1.22 (m, 3H, C*H*$_3$). ^1H NMR (400 MHz, MeOD) δ 7.53 – 7.37 (m, 4H, aromat.), 7.24 – 7.16 (m, 3H, aromat.), 4.23 – 4.10 (m, 1H, C*H*), 3.84 (s, 3H, OC*H*$_3$), 1.44 (s, 9H, OC(C*H*$_3$)$_3$), 1.36 (d, *J* = 7.1 Hz, 3H, C*H*$_3$).^{13}C NMR (101 MHz, DMSO-d_6) δ 171.9, 155.2, 148.1, 137.9, 130.4, 125.7, 125.1, 119.9, 78.0, 52.5, 50.4, 28.2, 17.9.

8.7.7 Methyl (5-((4-aminophenyl)amino)-1H-benzo[d]imidazol-2-yl)carbamate (20)

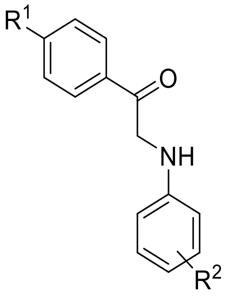


20

Compound **66** (100 mg, 0.24 mmol) was dissolved in a 4M solution of HCl in 1,4-dioxane (4 mL) at 0 °C. The mixture was warmed up to r.t. and stirred for 24 h. The solvent was removed under reduced pressure to give **20** (100 mg, quantitative yield). Preparative RP-HPLC (solvent A = H_2O + 0.1% TFA, solvent B = ACN + 0.1% TFA, gradient from 15% to 60% B in 10 min; flow 20 mL/min; t_R: 7.2 min). Analytical RP-HPLC (solvent A: H_2O + 0.1% of TFA, solvent B: ACN + 0.1% of TFA, gradient: from 15% to 60% of B in 10 min, flow: 1 mL/min, t_R = 7.9 min). MS (ESI, *m/z*): $C_{18}H_{20}N_5O_3S^+$ $[M+H]^+$ calc.: 386.13, found: 386.2 $[M+H]^+$, 408.0 $[M+Na]^+$, 771.3 $[2M+H]^+$, 793.3 $[2M+Na]^+$. ^{1}H NMR (400 MHz, MeOD) δ 7.61 – 7.53 (m, 2H, aromat.), 7.54 – 7.44 (m, 2H, aromat.), 7.35– 7.29 (m, 3H, aromat.), 4.04 (q, *J* = 7.1 Hz, 1H, C*H*), 3.91 (s, 3H, OC*H*3), 1.58 (d, *J* = 7.1 Hz, 3H, C*H*3). ^{13}C NMR (101 MHz, MeOD) δ 169.3, 154.5, 139.0, 134.2, 133.6, 132.0, 128.3, 122.0, 116.0, 114.8, 54.5, 50.9, 17.6. The ^{1}H and ^{13}C NMR data matched with the data reported in the literature.[149]

8.8 AZPP series

8.8.1 General procedure for preparation of α-aminoacetophenone



70a-e

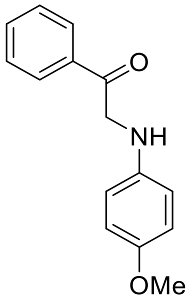
METHOD A

α-Bromoacetophenone **67** (1-3 mmol) in dry THF (1,5-3 mL) was added dropwise to a flame-dried and nitrogen-purged flask containing a solution of aniline **68** (2 equivalents) in dry THF. The mixture was stirred overnight at 20 °C, under N_2. The excess of aniline was removed as insoluble salts by filtration and the resulting supernatant was dried under reduced pressure and immediately used for the following step, without further purification.

METHOD B

To a solution of aniline **68** in dry EtOH, 1 equivalent of α-bromoacetophenone **67** was added portion wise followed by 1 equivalent of $NaHCO_3$(s). The mixture was stirred at room temperature overnight in nitrogen atmosphere. The precipitate formed from the reaction was collected by filtration and immediately used for the following step, without further purification.

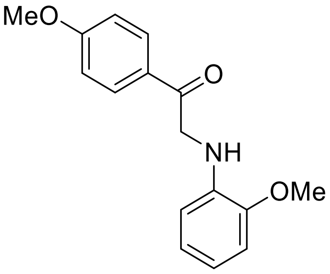
*8.8.1.1 2-((4-Methoxyphenyl)amino)-1-phenylethan-1-one (**70a**)*



70a

Prepared according to METHOD A. Quantitative yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.07 – 7.98 (m, 2H, aromat.), 7.67 – 7.58 (m, 1H, aromat.), 7.57 – 7.47 (m, 2H, aromat.), 6.83 (d, *J* = 9.0 Hz, 2H, aromat.), 6.71 (d, *J* = 9.0 Hz, 2H, aromat.), 4.61 (s, 2H, C*H*$_2$), 3.76 (s, 3H, OC*H*$_3$). The ^{1}H and ^{13}C NMR data matched with the data reported in the literature.[153]

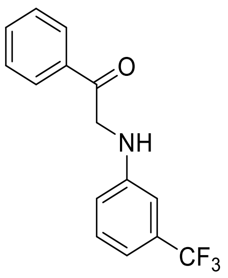
*8.8.1.2 1-(4-Methoxyphenyl)-2-((2-methoxyphenyl)amino)ethan-1-one (**70b**)*



70b

Prepared according to METHOD B. Quantitative yield. MS (ESI, *m/z*): $C_{16}H_{18}NO_3^+$ $[M+H]^+$ calc.: 272.13, found: $[M+H]^+$ 272.1. ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.01 (dd, *J* = 9.3, 2.4 Hz, 2H, aromat.), 6.98 (dd, *J* = 9.3, 2.4 Hz, 2H, aromat.), 6.90 (td, *J* = 7.7, 1.4 Hz, 1H, aromat.), 6.82 (dd, *J* = 7.7, 1.3 Hz, 1H, aromat.), 6.72 (td, *J* = 7.7, 1.3 Hz, 1H, aromat.), 6.61 (dd, *J* = 7.7, 1.4 Hz, 1H, aromat.), 4,57 (s, 2H, C*H*$_2$), 3.90 (s, 3H, OC*H*$_3$), 3.88 (s, 3H, OC*H*$_3$). ^{13}C NMR (101 MHz, CDCl$_3$) δ 193.6, 164.0, 147.3, 137.5, 130.1, 128.2, 121.2, 117.0, 114.1, 110.2, 109.7, 55.6, 55.6, 49.9.

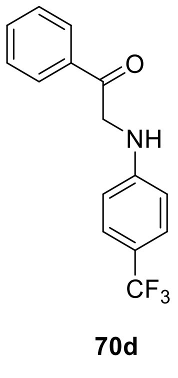
*8.8.1.3 1-Phenyl-2-((3-(trifluoromethyl)phenyl)amino)ethan-1-one (**70c**)*



70c

Prepared according to METHOD B. Quantitative yield. TLC R_f = 0.5 (solvent system: DCM/MeOH, 10:1).MS (ESI, *m/z*): $C_{15}H_{13}F_3NO^+$ $[M+H]^+$ calc.: 280.09, found: $[M+H]^+$ 280.1. ^1H NMR (400 MHz, $CDCl_3$) δ 8.09 – 8.01 (m, 2H, aromat.), 7.69 – 7.61 (m, 1H, aromat.), 7.59 – 7.50 (m, 2H, aromat.), 7.31 (t, *J* = 7.8 Hz, 1H, aromat.), 6.99 (d, *J* = 7.6 Hz, 1H, aromat.), 6.94 – 6.78 (m, 2H, aromat.), 5.18 (bs, 1H), 4.64 (d, *J* = 4.2 Hz, 2H, CH_2). ^{13}C NMR (101 MHz, $CDCl_3$) δ 194.4, 147.2, 134.7, 134.1, 131.7 (q, *J* = 31.9 Hz), 129.7, 128.99, 127.8, 124.3 (d, *J* = 272.3 Hz), 116.4, 114.2 (q, *J* = 3.8 Hz), 108.8 (q, *J* = 3.9 Hz), 50.0. The ^1H and ^{13}C NMR data matched with the data reported in the literature.[172]

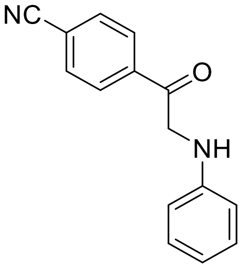
*8.8.1.4 1-Phenyl-2-((4-(trifluoromethyl)phenyl)amino)ethan-1-one (**70d**)*



70d

Prepared according to METHOD B, quantitative yield. MS (ESI, *m/z*): $C_{15}H_{13}F_3NO^+$ $[M+H]^+$ calc.: 280.09, found: $[M+H]^+$ 280.1. ^1H NMR (400 MHz, $CDCl_3$) δ 8.09 – 7.98 (m, 2H, aromat.), 7.69 – 7.61 (m, 1H, aromat.), 7.58 – 7.50 (m, 2H, aromat.), 7.46 (d, J = 8.4 Hz, 2H, aromat.), 6.72 (d, J = 8.4 Hz, 2H, aromat.), 4.64 (d, J = 4.3 Hz, 2H, C*H*$_2$). ^{13}C NMR (101 MHz, CDCl$_3$) δ, 194.3, 149.5, 134.8, 134.3, 129.1, 127.9, 126.9 (q, *J* = 3.3 Hz), 126.2 (q, *J* = 296.6 Hz), 119.5 (q, *J* = 32.6 Hz), 114.3, 112.3, 49.7. The ^{1}H and ^{13}C NMR data matched with the data reported in the literature.[173]

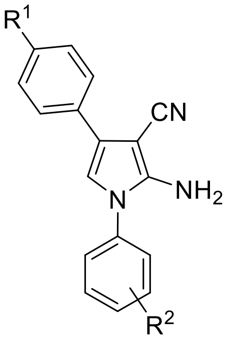
*8.8.1.5 4-(Phenylglycyl)benzonitrile (**70e**)*



70e

Prepared according to METHOD B, quantitative yield. TLC R_f = 0.5 (solvent system: *n*-Hex/EtOAc, 7:3).MS (ESI, *m/z*): $C_{15}H_{13}N_2O^+$ [M+H]$^+$ calc.: 237.10, found: [M+H]$^+$ 237.1. ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.12 (d, *J* = 8.2 Hz, 2H, aromat.), 7.83 (d, *J* = 8.2 Hz, 2H, aromat.), 7.26 – 7.20 (m, 2H, aromat.), 6.79 (t, *J* = 7.3 Hz, 1H, aromat.), 6.72 (d, *J* = 8.3 Hz, 2H, aromat.), 4.65 (s, 2H, C*H*$_2$). ^{13}C NMR (101 MHz, CDCl$_3$) δ 194.2, 146.8, 138.0, 132.9, 129.6, 128.4, 118.4, 117.8, 117.3, 113.2, 51.0.

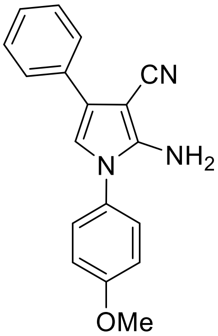
8.8.2 General procedure for preparation of 2-Amino-3-cyan-1,4-disubstituted pyrroles



72a-e

To a solution of (1 -3 mmol) α-aminoacetophenone **70** in dry EtOH and the 1 equivalent of malononitrile was added followed by a 21 wt.% solution of EtONa in EtOH (1 equivalent). The mixture was gradually heated up to 85 °C for 3 h. The solvent was removed under reduced pressure and the residue was purified by flash chromatography or crystallisation from EtOH.

*8.8.2.1 2-Amino-1-(4-methoxyphenyl)-4-phenyl-1H-pyrrole-3-carbonitrile (**72a**)*

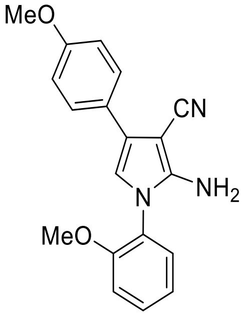


72a

Purified by crystallisation from EtOH. White crystal, 43% yield over two steps. MS (ESI, *m/z*): $C_{19}H_{18}N_3O_2^+$ [M+H]$^+$ calc.: 290.13, found: [M+H]$^+$ 290.1. ^{1}H NMR (400 MHz, CDCl$_3$)

δ 7.67 – 7.60 (m, 2H, aromat), 7.41 – 7.34 (m, 2H, aromat.), 7.35 – 7.30 (m, 2H, aromat.), 7.28 – 7.23 (m, 1H, aromat.), 7.06 – 6.99 (m, 2H, aromat.), 6.49 (s, 1H, aromat.), 4.16 (bs, 2H, N*H₂*), 3.86 (s, 3H, OC*H*₃). ^{13}C NMR (101 MHz, $CDCl_3$) δ 159.76, 146.68, 133.20, 129.45, 128.89, 126.99, 126.96, 125.99, 123.63, 117.79, 115.25, 113.19, 72.80, 55.75. The ^{1}H and ^{13}C NMR data matched with the data reported in the literature.[153]

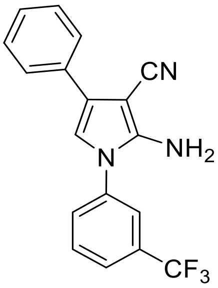
*8.8.2.2 2-Amino-1-(2-methoxyphenyl)-4-(4-methoxyphenyl)-1H-pyrrole-3-carbonitrile (**72b**)*



72b

Purified by gradient elution flash chromatography (solvent system: *n*-Hex/EtOAc, from 3:2 to 2:3). Pale yellow solid, 35% yield over two steps. MS (ESI, *m/z*): $C_{19}H_{18}N_3O_2^+$ $[M+H]^+$ calc.: 320.14, found: $[M+H]^+$ 320.1. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.58 (d, *J* = 8.7 Hz, 2H, aromat.), 7.46 – 7.39 (m, 1H, aromat.), 7.30 (d, *J* = 7.9 Hz, 1H, aromat.), 7.08 (t, *J* = 7.8 Hz, 2H, aromat.), 6.92 (d, *J* = 8.7 Hz, 2H, aromat.), 6.41 (s, 1H, aromat.), 4.22 (bs, 2H, N*H₂*), 3.87 (s, 3H, OC*H*₃), 3.82 (s, 3H, OC*H*₃). ^{13}C NMR (101 MHz, $CDCl_3$) δ 158.6, 154.0, 147.4, 130.3, 128.3, 127.1, 125.9, 125.7, 123.4, 121.5, 118.1, 114.2, 113.0, 112.64, 6.12, 55.3.

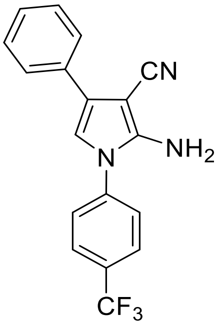
*8.8.2.3 2-Amino-4-phenyl-1-(3-(trifluoromethyl)phenyl)-1H-pyrrole-3-carbonitrile (**72c**)*



72c

Purified by gradient elution flash chromatography (solvent system: *n*-Hex/DCM, from 2:3 to 3:7). 34% yield over two steps. Analytical RP-HPLC (solvent A: H_2O, solvent B: ACN + 0.1% TFA, gradient: from 20% to 80% of B in 20 min flow: 1 mL/min; t_R= 22.25 min). MS (ESI, *m/z*): $C_{18}H_{13}F_3N_3^+$ $[M+H]^+$ calc.: 328.11, found: $[M+H]^+$. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.78 – 7.60 (m, 6H, aromat.), 7.40 (t, *J* = 7.6 Hz, 2H, aromat.), 7.32 – 7.27 (m, 1H, aromat.), 6.57 (s, 1H, aromat.), 4.15 (bs, 2H, N*H*$_2$). ^{13}C NMR (101 MHz, $CDCl_3$) δ 146.0, 137.6, 132.7, 131.0, 129.0, 128.6, 127.4, 126.2, 125.4 (q, *J* = 3.6 Hz), 124.9, 122.4 (q, *J* = 3.9 Hz,), 117.1, 112.7, 29.8.

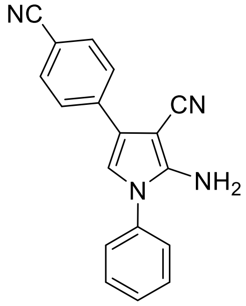
*8.8.2.4 2-Amino-4-phenyl-1-(4-(trifluoromethyl)phenyl)-1H-pyrrole-3-carbonitrile (**72d**)*



72d

The crude material was immediately used for the following step. MS (ESI, *m/z*): $C_{18}H_{13}F_3N_3^+$ $[M+H]^+$ calc.: 328.11, found: $[M+H]^+$ 328.1. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.85 – 7.77 (m, 2H, aromat.), 7.69 – 7.57 (m, 4H, aromat.), 7.35 – 7.45 (m, 2H, aromat.), 7.29 (t, *J* = 8.7 Hz, 1H, aromat.), 6.57 (s, 1H, aromat.), 4.20 (bs, 2H, N*H*$_2$).

*8.8.2.5 2-Amino-4-(4-cyanphenyl)-1-phenyl-1H-pyrrole-3-carbonitrile (**72e**)*

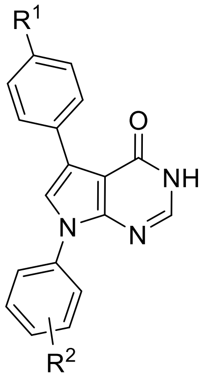


72e

Purified by gradient elution flash chromatography (solvent system: *n*-Hex/EtOAc, from 7:3 to 1:1) to obtain 74 mg of **72e** with 26% yield over two steps. MS (ESI, *m/z*): C18H13N4+ [M+H]+ calc.: 285.11, found: [M+H]+ 285.1. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.77 – 7.70 (m, 2H, aromat.), 7.68 – 7.62 (m, 2H, aromat.), 7.59 – 7.53 (m, 2H, aromat.), 7.52 – 7.46 (m, 1H, aromat.), 7.45 – 7.40 (m, 2H, aromat.), 6.64 (s, 1H, aromat.), 4.27 (bs, 2H, N*H*$_2$). ^{13}C NMR (101 MHz, $CDCl_3$) δ 147.1, 137.8, 136.4, 132.8, 130.4, 129.9, 129.2, 126.2, 125.5, 122.1, 120.1, 119.2, 119.0, 117.2, 114.4, 110.1.

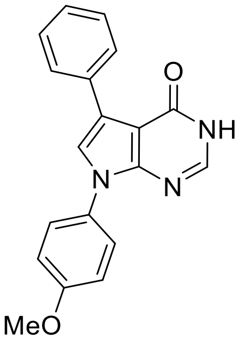
.

8.8.3 General procedure for preparation of 5,7-disubstituted-7H-pyrrolo[2,3-d]pyrimidin-4(3H)-ones



A mixture of (1-3 mmol) 2-Amino-3-cyan-1,4-disubstituted pyrrole and a large excess (20 equivalents) of formic acid 85 % (v/v) was refluxed in a sealed Schlenk tube overnight. The reaction mixture was then allowed to cool down to r.t., ice-cooled water was added, and the precipitate was filtered and crystallised.

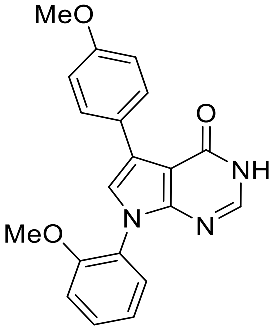
*8.8.3.1 7-(4-Methoxyphenyl)-5-phenyl-3,7-dihydro-4H-pyrrolo[2,3-d]pyrimidin-4-one (**23**)*



23

Amber solid. Purified by precipitation in water then crystallisation, 46%. MS (ESI, *m/z*): $C_{19}H_{16}N_3O_2^+$ [M+H]$^+$ calc.: 318.12, found: [M+H]$^+$ 318.1, [M+Na]$^+$ 340.1, [M+K]$^+$ 356.0, [2M+Na]$^+$ 657.2. Analytical RP-HPLC: Phenomenex Jupiter C18 column, 5 µm, 300 Å, 150 × 4.6 mm (L × ID) (solvent A: H_2O, solvent B: ACN + 0.1% TFA, gradient: from 30% to 80% of B in 15 min, flow: 1 mL/min; t_R = 9.5 min). ^{1}H NMR (400 MHz, $CDCl_3$) δ 12.42 (bs, 1H, N*H*), 793 – 7.94 (m, 3H, aromat.), 7.51 (d, *J* = 8.9 Hz, 2H, aromat.), 7.43 (t, *J* = 7.5 Hz, 2H, aromat.), 7.34 – 7.29 (m, 1H, aromat.), 7.06 (d, *J* = 8.9 Hz, 2H, aromat.), 3.87 (s, 3H, OC*H*$_3$). ^{13}C NMR (101 MHz, $CDCl_3$) δ 165.8, 161.2, 159.2, 148.8, 143.23 133.3, 130.3, 128.7, 128.5, 127.1, 126.4, 122.4, 121.8, 114.8, 106.2, 55.7. The ^{1}H and ^{13}C NMR data matched with the data reported in the literature.[153]

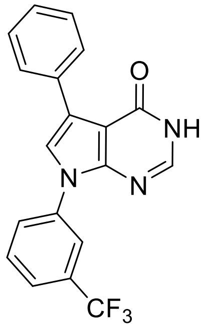
*8.8.3.2 7-(2-Methoxyphenyl)-5-(4-methoxyphenyl)-3,7-dihydro-4H-pyrrolo[2,3-d]pyrimidin-4-one (**27**)*



27

Pale yellow solid. Purified by gradient elution flash chromatography (solvent system: *n*-Hex/EtOAc, from 3.5:6.5 to 1:4), 30% yield. Preparative RP-HPLC (solvent A: H_2O + 0.1% TFA, solvent B: ACN + 0.1% TFA; gradient: from 40% to 88% of B in 8 min, flow: 20 mL/min; t_R = 7.4 min). Analytical RP-HPLC (solvent A: H_2O, solvent B: ACN + 0.1% TFA; gradient: from 40% to 100% of B in 10 min, flow: 1 mL/min; t_R = 7.6 min). MS (ESI, *m/z*): $C_{20}H_{18}N_3O_3^+$ [M+H]$^+$ calc.: 348.13, found: [M+H]$^+$ 348.1, [M+Na]$^+$ 370.1, [M+K]$^+$ 386.0, [2M+Na]$^+$ 717.3. ^{1}H NMR (400 MHz, $CDCl_3$) δ 11.05 (bs, 1H, N*H*), 7.85 (s, 1H, aromat.), 7.81 (d, *J* = 8.2 Hz, 2H, aromat.), 7.49 – 7.40 (m, 2H, aromat.), 7.15 (s, 1H, aromat.), 7.14 – 7.07 (m, 2H, aromat.), 6.96 (d, *J* = 8.0 Hz, 2H, aromat.), 3.84 (s, 3H, OC*H*$_3$), 3.82 (s, 3H, OC*H*$_3$). ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 11.98 (d, *J* = 3.0 Hz, 1H, N*H*), 7.93 – 7.87 (m, 2H, aromat.), 7.83 (d, *J* = 3.0 Hz, 1H, aromat.), 7.51 – 7.44 (m, 1H, aromat.), 7.45 – 7.40 (m, 2H, aromat.), 7.27 – 7.22 (m, 1H, aromat.), 7.09 (td, *J* = 7.6, 1.0 Hz, 1H, aromat.), 6.95 – 6.90 (m, 2H, aromat.), 3.78 (s, 3H, OC*H*$_3$), 3.76 (s, 3H, OC*H*$_3$). ^{1}H NMR (400 MHz, Acetone-d$_6$) δ 10.80 (s, 1H), 8.03 – 7.97 (m, 2H), 7.86 (s, 1H), 7.52 – 7.44 (m, 2H), 7.33 (s, 1H), 7.25 (dd, *J* = 8.8, 1.2 Hz, 1H), 7.11 (td, *J* = 7.6, 1.2 Hz, 1H), 6.95 – 6.89 (m, 2H), 3.82 (s, 3H, OC*H*$_3$), 3.82 (s, 3H, OC*H*$_3$). ^{13}C NMR (101 MHz, Acetone-d$_6$) δ 55.5, 56.3, 106.4, 113.3, 114.2, 121.2, 121.5, 122.9, 127.2, 127.4, 129.9, 130.4, 144.3, 150.1, 155.8, 159.4, 159.4. ^{13}C NMR (101 MHz, DMSO-d$_6$) δ 158.6, 157.9, 154.4, 148.9, 143.9, 129.8, 129.3, 128.9, 126.0, 125.6, 122.4, 120.4, 119.6, 113.5, 112.6, 104.8, 55.8, 55.0. ^{13}C NMR (101 MHz, CDCl$_3$) δ 55.5, 56.0, 112.5, 113.9, 121.0, 121.1, 123.3, 125.8, 126.0, 128.9, 130.2, 142.7, 149.3, 154.34, 158.9, 160.7.

*8.8.3.3 5-Phenyl-7-(3-(trifluoromethyl)phenyl)-3,7-dihydro-4H-pyrrolo[2,3-d]pyrimidin-4-one (**28**)*

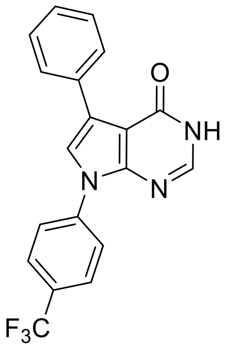


28

Pale yellow solid. Purified by precipitation in water then crystallisation, 48% of yield. MS (ESI, *m/z*): $C_{19}H_{13}F_3N_3O^+$ $[M+H]^+$ calc.: 356.10, found: $[M+H]^+$ 356.1, $[M+Na]^+$ 378.0, $[M+K]^+$ 394.0. Analytical RP-HPLC (solvent A: H_2O, solvent B: ACN + 0.1% TFA, gradient: from 50% to 90% of B in 20 min, flow: 1 mL/min; t_R = 11.8 min). ^{1}H NMR (400 MHz, CDCl$_3$) δ 11.33 (bs, 1H, N*H*), 8.00 – 7.84 (m, 5H, aromat.), 7.74 – 7.67 (m, 2H, aromat.), 7.45 (t, *J* = 7.6 Hz, 2H, aromat.), 7.40 – 7.31 (m, 2H, aromat.); ^{19}F NMR (376 MHz, CDCl$_3$) δ -62.67; ^{13}C NMR

(101 MHz, $CDCl_3$) δ 160.8, 148.8, 143.4, 137.9, 132.7, 132.2 (q, *J* = 32.7 Hz), 130.3, 128.9, 128.6, 128.1, 127.5, 124.5 (q, *J* = 3.6 Hz), 123.7 (q, *J* = 272.6 Hz), 123.1, 121.8 (q, *J* = 3.8 Hz), 121.6, 107.0.

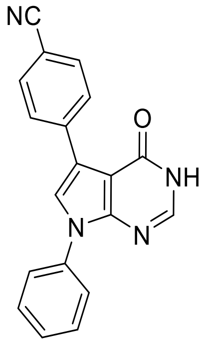
*8.8.3.4 5-Phenyl-7-(4-(trifluoromethyl)phenyl)-3,7-dihydro-4H-pyrrolo[2,3-d]pyrimidin-4-one (**29**)*



29

Purified by precipitation in water then preparative RP-HPLC, 23% of yield (over compound 70). Preparative RP-HPLC (solvent A = H_2O + 0.1% TFA, solvent B = ACN + 0.1% TFA, gradient from 55% to 100% B in 15 min; flow 20 mL/min; t_R: 8.8 min). MS (ESI, *m/z*): $C_{19}H_{13}F_3N_3O^+$ $[M+H]^+$ calc.: 356.10, found: $[M+H]^+$ 356.1, $[M+Na]^+$ 378.0, $[M+K]^+$ 394.0. Analytical RP-HPLC (solvent A: H_2O + 0.1% TFA, solvent B: ACN + 0.1% TFA, gradient: from 55% to 100% of B in 10 min, flow: 1 mL/min; t_R = 9.0 min). ^1H NMR (400 MHz, MeOD) δ 8.01 (d, *J* = 8.4 Hz, 2H, aromat.), 7.97 (s, 1H, aromat.), 7.88 (dd, *J* = 13.2, 5.0 Hz, 4H, aromat.), 7.65 (s, 1H, aromat.), 7.38 (dd, *J* = 10.4, 4.7 Hz, 2H, aromat.), 7.29 (dd, *J* = 10.5, 4.3 Hz, 1H, aromat.); ^{19}F NMR (376 MHz, MeOD) δ -63.92; ^{13}C NMR (101 MHz, MeOD) δ 150.1, 145.2, 142.0, 134.4, 130.2, 129.9, 129.8, 129.1, 127.9, δ 127.5 (q, *J* = 3.9 Hz), 126.1, 124.1, 122.6.

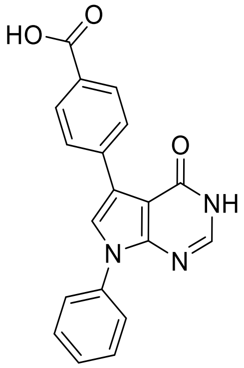
*8.8.3.5 4-(4-Oxo-7-phenyl-4,7-dihydro-3H-pyrrolo[2,3-d]pyrimidin-5-yl)benzonitrile (**30**)*



30

Pale yellow solid. Purified by precipitation in water and preparative RP-HPLC, 88% yield. MS (ESI, *m/z*): $C_{19}H_{13}N_4O^+$ [M+H]$^+$ calc.: 313.11, found: 313.1 [M+H]$^+$, 647.2 [2M+Na]$^+$, 663.3 [2M+K]$^+$.b Analytical RP-HPLC (solvent A: H_2O + 0.1% of TFA, solvent B: ACN + 0.1% of TFA, gradient: from 50 % to 95% of B in 15 min, flow: 1 mL/min, t_R = 7.2 min). Preparative RP-HPLC (solvent A: H_2O + 0.1% TFA solvent B: ACN + 0.1% TFA; gradient: from 30% to 90 % of B in 10 min, flow: 20 mL/min; t_R = 9.20 min). ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 12.29 (d, *J* = 3.2 Hz, 1H, N*H*), 8.36 – 8.24 (m, 2H, aromat.), 8.09 (s, 1H, aromat.), 8.01 (d, *J* = 3.2 Hz, 1H, aromat.), 7.90 – 7.80 (m, 2H, aromat.), 7.82 – 7.72 (m, 2H, aromat.), 7.62 – 7.52 (m, 2H, aromat.), 7.50 – 7.41 (m, 1H, aromat.). ^{13}C NMR (101 MHz, DMSO-d$_6$) δ 158.6, 148.6, 144.9, 138.3, 136.9, 132.0, 129.2, 128.7, 127.5, 124.7, 123.7, 119.3, 119.0, 108.4, 106.0.

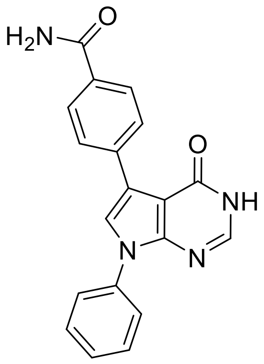
8.8.4 4-(4-Oxo-7-phenyl-4,7-dihydro-3H-pyrrolo[2,3-d]pyrimidin-5-yl)benzoic acid (31)



31

A mixture of nitrile (262 mg, 0.84 mmol), NH_4OH (3.0 mL) and $CsOH·H_2O$ (12.58 mg, 0.08 mmol) in H_2O (0.5 mL) was sealed in a Schlenk tube under nitrogen atmosphere and gradually heated up to 100 °C. After 8h, the reaction mixture was cooled down to r.t., diluted with water EtOAc. The two phases were separated. The basic aqueous layer was acidified with a solution of HCl 1N (pH 3) end extracted with DCM and EtOAc giving **31** as pale yellow solid. The organic layer was dried over Na_2SO_4, filtered and the solvent was removed under reduced pressure, giving a mixture of the compounds, **31** and **32** in ratio 24% - 76%. The two compounds were separated by preparative RP-HPLC (solvent A: H_2O + 0.1% TFA solvent B: ACN + 0.1% TFA; gradient: from 30% to 90% of B in 10 min, flow: 20 mL/min; t_R = 7.20 min). MS (ESI, *m/z*): $C_{19}H_{14}N_3O_3^+$ $[M+H]^+$ calc.: 332.10, found: 332.1 $[M+H]^+$, 663.3 $[2M+H]^+$, 685.2 $[2M+Na]^+$. Analytical RP-HPLC (solvent A: H_2O+ 0.1% TFA, solvent B: ACN + 0.1% TFA, gradient: from 30% to 90% of B in 10 min, flow: 1 mL/min; t_R = 7.6 min). ^1H NMR (400 MHz, DMSO-d_6) δ 12.82 (bs, 1H, O*H*), 12.23 (d, *J* = 3.7 Hz, 1H, N*H*), 8.25 – 8.09 (m, 2H, aromatic.), 8.00 (d, *J* = 3.8 Hz, 1H, aromatic.), 7.99 (s, 1H, aromatic.), 7.97 – 7.89 (m, 2H, aromatic.), 7.83 – 7.73 (m, 2H, aromatic.), 7.64 – 7.50 (m, 2H, aromatic.), 7.44 (tt, *J* = 7.4, 1.1 Hz, 1H, aromatic.). ^{13}C NMR (101 MHz, DMSO-d_6) δ 157.6, 148.8, 138.6, 135.0, 128.2, 127.3, 119.5, 119.4, 118.6, 118.4, 117.7, 115.0, 113.3, 110.0, 96.3.

8.8.5 4-(4-Oxo-7-phenyl-4,7-dihydro-3H-pyrrolo[2,3-d]pyrimidin-5-yl)benzamide (32)

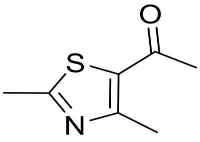


32

Pale yellow solid. Purified by Preparative RP-HPLC (solvent A: H_2O + 0.1% TFA solvent B: ACN + 0.1% TFA; gradient: from 30% to 90% of B in 10 min, flow: 20 mL/min; t_R = 5.90 min). MS (ESI, *m/z*): $C_{19}H_{15}N_4O_2^+$ $[M+H]^+$ calc.: 331.12, found: 331.1 $[M+H]^+$, 353. $[M+Na]^+$, 369.0 $[M+K]^+$, 661.3 $[2M+H]^+$, 683.3 $[2M+Na]^+$. Analytical RP-HPLC (solvent A: H_2O+ 0.1% TFA, solvent B: ACN + 0.1% TFA, gradient: from 30% to 90% of B in 10 min, flow: 1 mL/min; t_R = 6.3 min). ^1H NMR (400 MHz, DMSO-d_6) δ 12.20 (d, *J* = 3.3 Hz, 1H, N*H*), 8.12 (d, *J* = 8.5 Hz, 2H, aromatic.), 8.02 – 7.93 (m, 3H, aromatic.), 7.88 (d, *J* = 8.5 Hz, 2H, aromatic.), 7.81 – 7.75 (m, 2H, aromatic.), 7.62 – 7.52 (m, 2H, aromatic.), 7.47 – 7.40 (m, 1H, aromatic.), 7.31 (bs, 1H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 167.8, 158.56, 148.2, 144.6, 137.1, 136.3, 131.8, 129.2, 127.8, 127.3, 124.7, 122.6, 119.9, 106.1.

8.9 Similarity Search-Hit

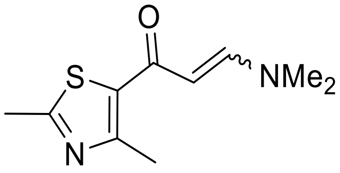
8.9.1 1-(2,4-Dimethylthiazol-5-yl)ethan-1-one (76)



76

150 µL of 3-chloro-2,4-pentadionewere **74** were added dropwise to a precooled solution (at 0 °C) of thioacetamide **75** (100 mg, 1.33 mmol) in MeOH (1 mL) and 40 µL of pyridine (Py) and the reaction was stirred overnight at r.t.. After dilution with EtOAc, the mixture was washed three times with water and with aqueous solution of HCl 1N. The organic phases were dried over Na_2SO_4, filtered and the solvent was removed under reduced pressure, to obtain **76** in quantitative yield. Analytical RP-HPLC (solvent A: H_2O, solvent B: ACN + 0.1% of TFA, gradient: from 10% to 70% of B in 15 min, t_R = 9.38 min). MS (ESI, *m/z*): $C_7H_{10}NOS^+$ $[M+H]^+$ calc.: 156.05, found: 156. ^1H NMR (400 MHz, $CDCl_3$) δ 2.69 (s, 3H, CH_3), 2.68 (s, 3H, CH_3), 2.50 (s, 3H, CH_3). The ^1H and ^{13}C NMR data matched with the data reported in the literature.[174]

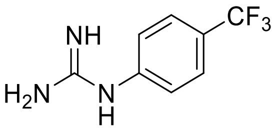
8.9.2 (E,Z)-3-(Dimethylamino)-1-(2,4-dimethylthiazol-5-yl)prop-2-en-1-one (77)



77

A solution of ketone **76** (318 µL, 2.6 mmol) in DMFDMA (660 µL, 5 mmol) was refluxed under nitrogen atmosphere for 24 h. After cooling down to r.t., the mixture was diluted with Et_2O (30 mL) and basified with an aqueous solution of NaOH 1N (10 mL). The ethereal layer was dried over Na_2SO_4, filtered and the solvent was removed under reduced pressure. The crude mixture was purified by gradient elution flash chromatography (solvent system: *n*-Hex/EtOAc, from 2:3 to 1:4) to afford 400 mg of a yellow solid (73% yield). Analytical RP-HPLC (solvent A: H_2O, solvent B: ACN + 0.1% of TFA, gradient: from 10% to 70% of B in 15 min, $t_R = 7.49$ min). MS (ESI, *m/z*): $C_{10}H_{15}N_2OS^+$ $[M+H]^+$ calc.: 211.09, found: 211. ^1H NMR (400 MHz, $CDCl_3$) δ 7.65 (d, *J* = 12.2 Hz, 1H, C*H*), 5.30 (d, *J* = 12.2 Hz, 1H, C*H*), 3.07 (s, 3H, NCH_3), 2.83 (s, 3H, NCH_3), 2.63 (s, 3H, CH_3), 2.59 (s, 3H, CH_3). ^{13}C NMR (101 MHz, $CDCl_3$) δ 181.3, 165.6, 154.7, 153.6, 133.1, 95.0, 45.07, 37.3, 19.3, 17.8. The ^1H and ^{13}C NMR data matched with the data reported in the literature.[159]

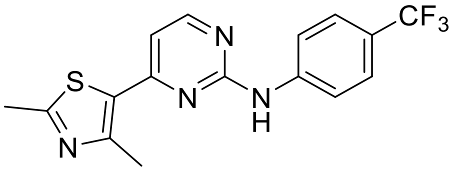
8.9.3 1-(4-Trifluoromethyl)phenyl)guanidine (79)



79

To an ice-cooled mixture solution of the 4-(trifluoromethyl)aniline **68c** (313 µL, 2.5 mmol) in EtOH (1 mL), a 70% aqueous solution of HNO_3 (180 µL) was added at 0 °C dropwise. The cyanamide 50 wt.% in H_2O (500 µL, 13 mmol) was added and the mixture was gradually heated up to 100 °C overnight under nitrogen atmosphere. The mixture was diluted with Et_2O and basified with a solution of NaOH 1N. The ethereal phase was separated, the aqueous layer was extracted three time with EtOAc. The combined organic phases were dried over Na_2SO_4, filtered and the solvent was removed under reduced pressure, to obtain a crude mixture containing **79** that was immediately used for the following step. Analytical RP-HPLC (solvent A: H_2O, solvent B: ACN + 0.1% of TFA, gradient: from 10% to 100% of B in 10 min, t_R = 6.83 min). TLC (solvent system: EtOAc/Hexane 3:7, R_f = 0.3) MS (ESI, *m/z*): $C_8H_9F_3N_3^+$ $[M+H]^+$ calc.: 204.07, found: 204.1. ^1H NMR (400 MHz, $CDCl_3$) δ 7.44 (d, *J* = 8.3 Hz, 2H, aromat.), 6.93 (d, *J* = 8.2 Hz, 2H, aromat.), 4.98 (bs, 4H, guanidine). The ^1H and ^{13}C NMR data matched with the data reported in the literature.[159]

8.9.4 [4-(2,4-Dimethylthiazol-5-yl)pyrimidin-2-yl]-(4-trifluoromethylphenyl)amine (33)



33

A mixture of enaminone **77** (450 mg, 2.13 mmol), the 4-trifluoromethyl guanidine **79** (896 mg, 4.41 mmol) and NaOH (80 mg, 2 mmol) in 2-methoxyethanol was heated up to 125 °C under nitrogen atmosphere. The mixture was stirred overnight and then, after cooling down to r.t., the solvent was evaporated under reduced pressure and the residue was purified by isocratic flash chromatography (solvent system: *n*-Hex/EtOAc 1:1) to obtain 355 mg of **33** (47% of yield). Analytical RP-HPLC (solvent A: H_2O, solvent B: ACN + 0.1% of TFA, gradient: from 10% to 100% of B in 10 min, t_R = 11.91 min). MS (ESI, *m/z*): $C_{16}H_{14}F_3N_4S^+$ $[M+H]^+$ calc.: 351.09, found: 351.1.^{1}H NMR (400 MHz, $CDCl_3$) δ 8.46 (d, *J* = 5.3 Hz, 1H, aromat.), 7.78 (d, *J* = 8.5 Hz, 2H, aromat.), 7.60 (d, *J* = 8.5 Hz, 2H, aromat.), 7.42 (bs, 1H, N*H*), 7.02 (d, *J* = 5.3 Hz, 1H, aromat.), 2.72 (s, 3H, C*H*$_3$), 2.71 (s, 3H, C*H*$_3$); ^{19}F NMR (376 MHz, $CDCl_3$) δ -61.77; ^{13}C NMR (101 MHz, $CDCl_3$) δ 167.4, 159.5, 159.3, 158.7, 152.8, 142.6, 131.2, 126.3 (q, *J* = 3.8 Hz), 124.5 (q, *J* = 271.3 Hz), 124.2 (q, *J* = 32.7 Hz), 118.5, 109.8, 19.6, 18.4.

REFERENCES

(1) Ruparelia, N.; Chai, J. T.; Fisher, E. A.; Choudhury, R. P. Inflammatory Processes in Cardiovascular Disease: A Route to Targeted Therapies. *Nat Rev Cardiol* **2016**, *advance online publication*. https://doi.org/10.1038/nrcardio.2016.185.

(2) Davies, M. J.; Thomas, A. C. Plaque Fissuring--the Cause of Acute Myocardial Infarction, Sudden Ischaemic Death, and Crescendo Angina. *Heart* **1985**, *53* (4), 363–373. https://doi.org/10.1136/hrt.53.4.363.

(3) Wang, J. C.; Bennett, M. Aging and Atherosclerosis. *Circulation Research* **2012**, *111* (2), 245–259. https://doi.org/10.1161/CIRCRESAHA.111.261388.

(4) Ross, R. Atherosclerosis — An Inflammatory Disease. *New England Journal of Medicine* **1999**, *340* (2), 115–126. https://doi.org/10.1056/NEJM199901143400207.

(5) Bentzon, J. F.; Otsuka, F.; Virmani, R.; Falk, E. Mechanisms of Plaque Formation and Rupture. *Circulation Research* **2014**, *114* (12), 1852–1866. https://doi.org/10.1161/CIRCRESAHA.114.302721.

(6) Chiong, M.; Morales, P. E.; Torres, G.; Gutiérrez, T.; García, L.; Ibacache, M.; Michea, L. Influence of Glucose Metabolism on Vascular Smooth Muscle Cell Proliferation. *Vasa* **2013**, *42* (1), 8–16. https://doi.org/10.1024/0301-1526/a000243.

(7) Ylä-Herttuala, S.; Bentzon, J. F.; Daemen, M.; Falk, E.; Garcia-Garcia, H. M.; Herrmann, J.; Hoefer, I.; Jukema, J. W.; Krams, R.; Kwak, B. R.; *et al*. Stabilisation of Atherosclerotic Plaques: Position Paper of the European Society of Cardiology (ESC) Working Group on Atherosclerosis and Vascular Biology. *Thrombosis and Haemostasis* **2011**, *106* (07), 1–19. https://doi.org/10.1160/TH10-12-0784.

(8) Otsuka, F.; Sakakura, K.; Yahagi, K.; Joner, M.; Virmani, R. Has Our Understanding of Calcification in Human Coronary Atherosclerosis Progressed? *Arteriosclerosis, Thrombosis, and Vascular Biology* **2014**, *34* (4), 724–736. https://doi.org/10.1161/ATVBAHA.113.302642.

(9) Finn, A. V.; Jain, R. K. Coronary Plaque Neovascularization and Hemorrhage: A Potential Target for Plaque Stabilization. *JACC: Cardiovascular Imaging* **2010**, *3* (1), 41–44. https://doi.org/10.1016/j.jcmg.2009.11.001.

(10) Virmani, R.; Kolodgie, F. D.; Burke, A. P.; Farb, A.; Schwartz, S. M. Lessons From Sudden Coronary Death. *Arteriosclerosis, Thrombosis, and Vascular Biology* **2000**, *20* (5), 1262–1275. https://doi.org/10.1161/01.ATV.20.5.1262.

(11) Van der Donckt, C.; Van Herck, J. L.; Schrijvers, D. M.; Vanhoutte, G.; Verhoye, M.; Blockx, I.; Van Der Linden, A.; Bauters, D.; Lijnen, H. R.; Sluimer, J. C.; *et al*. Elastin Fragmentation in Atherosclerotic Mice Leads to Intraplaque Neovascularization, Plaque Rupture, Myocardial Infarction, Stroke, and Sudden Death. *Eur. Heart J.* **2015**, *36* (17), 1049–1058. https://doi.org/10.1093/eurheartj/ehu041. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

(12) Moreno, P. R. Vulnerable Plaque: Definition, Diagnosis, and Treatment. *Cardiology Clinics* **2010**, *28* (1), 1–30. https://doi.org/10.1016/j.ccl.2009.09.008.

(13) Alsheikh-Ali, A. A. The Vulnerable Atherosclerotic Plaque: Scope of the Literature. *Annals of Internal Medicine* **2010**, *153* (6), 387. https://doi.org/10.7326/0003-4819-153-6-201009210-00272.

(14) Bennett, M. R.; Evan, G. I.; Schwartz, S. M. Apoptosis of Human Vascular Smooth Muscle Cells Derived from Normal Vessels and Coronary Atherosclerotic Plaques. *Journal of Clinical Investigation* **1995**, *95* (5), 2266–2274. https://doi.org/10.1172/JCI117917.

(15) Virmani, R.; Kolodgie, F. D.; Burke, A. P.; Finn, A. V.; Gold, H. K.; Tulenko, T. N.; Wrenn, S. P.; Narula, J. Atherosclerotic Plaque Progression and Vulnerability to Rupture. *Arteriosclerosis, Thrombosis, and Vascular Biology* **2005**, *25* (10), 2054–2061. https://doi.org/10.1161/01.ATV.0000178991.71605.18.

(16) Carmeliet, P. Angiogenesis in Health and Disease. *Nat. Med.* **2003**, *9* (6), 653–660. https://doi.org/10.1038/nm0603-653.

(17) Kaartinen, M.; Penttilä, A.; Kovanen, P. T. Mast Cells Accompany Microvessels in Human Coronary Atheromas: Implications for Intimal Neovascularization and Hemorrhage. *Atherosclerosis* **1996**, *123* (1), 123–131. https://doi.org/10.1016/0021-9150(95)05794-3.

(18) Herrmann, J.; Lerman, L. O.; Mukhopadhyay, D.; Napoli, C.; Lerman, A. Angiogenesis in Atherogenesis. *Arteriosclerosis, Thrombosis, and Vascular Biology* **2006**, *26* (9), 1948–1957. https://doi.org/10.1161/01.ATV.0000233387.90257.9b.

(19) Davies, M. J.; Thomas, A. Thrombosis and Acute Coronary-Artery Lesions in Sudden Cardiac Ischemic Death. *New England Journal of Medicine* **1984**, *310* (18), 1137–1140. https://doi.org/10.1056/NEJM198405033101801.

(20) Schaar, J. Terminology for High-Risk and Vulnerable Coronary Artery Plaques. *European Heart Journal* **2004**, *25* (12), 1077–1082. https://doi.org/10.1016/j.ehj.2004.01.002.

(21) O’Neill, L. A. J.; Hardie, D. G. Metabolism of Inflammation Limited by AMPK and Pseudo-Starvation. *Nature* **2013**, *493* (7432), 346–355. https://doi.org/10.1038/nature11862.

(22) Palsson-McDermott, E. M.; O’Neill, L. A. J. The Warburg Effect Then and Now: From Cancer to Inflammatory Diseases. *BioEssays* **2013**, *35* (11), 965–973. https://doi.org/10.1002/bies.201300084.

(23) Garedew, A.; Henderson, S. O.; Moncada, S. Activated Macrophages Utilize Glycolytic ATP to Maintain Mitochondrial Membrane Potential and Prevent Apoptotic Cell Death. *Cell Death & Differentiation* **2010**, *17* (10), 1540–1550. https://doi.org/10.1038/cdd.2010.27.

(24) Jiang, H.; Shi, H.; Sun, M.; Wang, Y.; Meng, Q.; Guo, P.; Cao, Y.; Chen, J.; Gao, X.; Li, E.; *et al*. PFKFB3-Driven Macrophage Glycolytic Metabolism Is a Crucial Component of Innate Antiviral Defense. *The Journal of Immunology* **2016**, *197* (7),

2880–2890. https://doi.org/10.4049/jimmunol.1600474.

(25) Nishizawa, T.; Kanter, J. E.; Kramer, F.; Barnhart, S.; Shen, X.; Vivekanandan-Giri, A.; Wall, V. Z.; Kowitz, J.; Devaraj, S.; O'Brien, K. D.; *et al*. Testing the Role of Myeloid Cell Glucose Flux in Inflammation and Atherosclerosis. *Cell Reports* **2014**, *7* (2), 356–365. https://doi.org/10.1016/j.celrep.2014.03.028.

(26) Tawakol, A.; Singh, P.; Mojena, M.; Pimentel-Santillana, M.; Emami, H.; MacNabb, M.; Rudd, J. H. F.; Narula, J.; Enriquez, J. A.; Través, P. G.; *et al*. HIF-1α and PFKFB3 Mediate a Tight Relationship Between Proinflammatory Activation and Anerobic Metabolism in Atherosclerotic MacrophagesSignificance. *Arteriosclerosis, Thrombosis, and Vascular Biology* **2015**, *35* (6), 1463–1471. https://doi.org/10.1161/ATVBAHA.115.305551.

(27) Pircher, A.; Treps, L.; Bodrug, N.; Carmeliet, P. Endothelial Cell Metabolism: A Novel Player in Atherosclerosis? Basic Principles and Therapeutic Opportunities. *Atherosclerosis* **2016**, *253*, 247–257. https://doi.org/10.1016/j.atherosclerosis.2016.08.011.

(28) Schoors, S.; De Bock, K.; Cantelmo, A. R.; Georgiadou, M.; Ghesquière, B.; Cauwenberghs, S.; Kuchnio, A.; Wong, B. W.; Quaegebeur, A.; Goveia, J.; *et al*. Partial and Transient Reduction of Glycolysis by PFKFB3 Blockade Reduces Pathological Angiogenesis. *Cell Metabolism* **2014**, *19* (1), 37–48. https://doi.org/10.1016/j.cmet.2013.11.008.

(29) De Bock, K.; Georgiadou, M.; Schoors, S.; Kuchnio, A.; Wong, B. W.; Cantelmo, A. R.; Quaegebeur, A.; Ghesquière, B.; Cauwenberghs, S.; Eelen, G.; *et al*. Role of PFKFB3-Driven Glycolysis in Vessel Sprouting. *Cell* **2013**, *154* (3), 651–663. https://doi.org/10.1016/j.cell.2013.06.037.

(30) Xu, Y.; An, X.; Guo, X.; Habtetsion, T. G.; Wang, Y.; Xu, X.; Kandala, S.; Li, Q.; Li, H.; Zhang, C.; *et al*. Endothelial 6-Phosphofructo-2-Kinase (PFKFB3) Plays a Critical Role in Angiogenesis. *Arterioscler Thromb Vasc Biol* **2014**, *34* (6), 1231–1239. https://doi.org/10.1161/ATVBAHA.113.303041.

(31) De Bock, K.; Georgiadou, M.; Carmeliet, P. Role of Endothelial Cell Metabolism in Vessel Sprouting. *Cell Metabolism* **2013**, *18* (5), 634–647. https://doi.org/10.1016/j.cmet.2013.08.001.

(32) Wang Yong; Xu Yiming; Yan Siyuan; Liu Zhiping; Zhou Yaqi; Zeng Xianqiu; Huo Yuqing. Abstract 719: PFKFB3 Mediated Glycolysis in Vascular Smooth Muscle Cells Is Essential for Arterial Neointima Formation. *Arteriosclerosis, Thrombosis, and Vascular Biology* **2015**, *35* (suppl_1), A719–A719. https://doi.org/10.1161/atvb.35.suppl_1.719.

(33) Werle, M.; Kreuzer, J.; Höfele, J.; Elsässer, A.; Ackermann, C.; Katus, H. A.; Vogt, A. M. Metabolic Control Analysis of the Warburg-Effect in Proliferating Vascular Smooth Muscle Cells. *J Biomed Sci* **2005**, *12* (5), 827–834. https://doi.org/10.1007/s11373-005-9010-5.

(34) Docherty, C. K.; Carswell, A.; Friel, E.; Mercer, J. R. Impaired Mitochondrial Respiration in Human Carotid Plaque Atherosclerosis: A Potential Role for Pink1 in

Vascular Smooth Muscle Cell Energetics. *Atherosclerosis* **2018**, *268*, 1–11. https://doi.org/10.1016/j.atherosclerosis.2017.11.009.

(35) Folco, E. J.; Sheikine, Y.; Rocha, V. Z.; Christen, T.; Shvartz, E.; Sukhova, G. K.; Di Carli, M. F.; Libby, P. Hypoxia But Not Inflammation Augments Glucose Uptake in Human Macrophages: Implications for Imaging Atherosclerosis With 18Fluorine-Labeled 2-Deoxy-D-Glucose Positron Emission Tomography. *Journal of the American College of Cardiology* **2011**, *58* (6), 603–614. https://doi.org/10.1016/j.jacc.2011.03.044.

(36) Warburg, O. On the Origin of Cancer Cells. *Science* **1956**, *123* (3191), 309–314. https://doi.org/10.1126/science.123.3191.309.

(37) Abdel-Haleem, A. M.; Lewis, N. E.; Jamshidi, N.; Mineta, K.; Gao, X.; Gojobori, T. The Emerging Facets of Non-Cancerous Warburg Effect. *Frontiers in Endocrinology* **2017**, *8*. https://doi.org/10.3389/fendo.2017.00279.

(38) Chen, Z.; Liu, M.; Li, L.; Chen, L. Involvement of the Warburg Effect in Non-Tumor Diseases Processes. *Journal of Cellular Physiology* **2018**, *233* (4), 2839–2849. https://doi.org/10.1002/jcp.25998.

(39) Chesney, J.; Mitchell, R.; Benigni, F.; Bacher, M.; Spiegel, L.; Al-Abed, Y.; Han, J. H.; Metz, C.; Bucala, R. An Inducible Gene Product for 6-Phosphofructo-2-Kinase with an AU-Rich Instability Element: Role in Tumor Cell Glycolysis and the Warburg Effect. *Proc Natl Acad Sci U S A* **1999**, *96* (6), 3047–3052.

(40) Ros, S.; Schulze, A. Balancing Glycolytic Flux: The Role of 6-Phosphofructo-2-Kinase/Fructose 2,6-Bisphosphatases in Cancer Metabolism. *Cancer Metab* **2013**, *1*, 8. https://doi.org/10.1186/2049-3002-1-8.

(41) Chesney, J. 6-Phosphofructo-2-Kinase/Fructose-2,6-Bisphosphatase and Tumor Cell Glycolysis: *Current Opinion in Clinical Nutrition and Metabolic Care* **2006**, *9* (5), 535–539. https://doi.org/10.1097/01.mco.0000241661.15514.fb.

(42) Schoors, S.; Cantelmo, A. R.; Georgiadou, M.; Stapor, P.; Wang, X.; Quaegebeur, A.; Cauwenberghs, S.; Wong, B. W.; Bifari, F.; Decimo, I.; *et al*. Incomplete and Transitory Decrease of Glycolysis. *Cell Cycle* **2014**, *13* (1), 16–22. https://doi.org/10.4161/cc.27519.

(43) Pilkis, S. J.; Claus, T. H.; Kurland, I. J.; Lange, A. J. 6-Phosphofructo-2-Kinase/Fructose-2,6-Bisphosphatase: A Metabolic Signaling Enzyme. *Annu. Rev. Biochem.* **1995**, *64*, 799–835. https://doi.org/10.1146/annurev.bi.64.070195.004055.

(44) Kim, S.-G.; Manes, N. P.; El-Maghrabi, M. R.; Lee, Y.-H. Crystal Structure of the Hypoxia-Inducible Form of 6-Phosphofructo-2-Kinase/Fructose-2,6-Bisphosphatase (PFKFB3) A POSSIBLE NEW TARGET FOR CANCER THERAPY. *J. Biol. Chem.* **2006**, *281* (5), 2939–2944. https://doi.org/10.1074/jbc.M511019200.

(45) Kim, S.-G.; Cavalier, M.; El-Maghrabi, M. R.; Lee, Y.-H. A Direct Substrate–Substrate Interaction Found in the Kinase Domain of the Bifunctional Enzyme, 6-Phosphofructo-2-Kinase/Fructose-2,6-Bisphosphatase. *Journal of Molecular Biology*

2007, *370* (1), 14–26. https://doi.org/10.1016/j.jmb.2007.03.038.

(46) Cavalier, M. C.; Kim, S.-G.; Neau, D.; Lee, Y.-H. Molecular Basis of the Fructose-2,6-Bisphosphatase Reaction of PFKFB3: Transition State and the C-Terminal Function. *Proteins* **2012**, *80* (4), 1143–1153. https://doi.org/10.1002/prot.24015.

(47) Minchenko, A.; Leshchinsky, I.; Opentanova, I.; Sang, N.; Srinivas, V.; Armstead, V.; Caro, J. Hypoxia-Inducible Factor-1-Mediated Expression of the 6-Phosphofructo-2-Kinase/Fructose-2,6-Bisphosphatase-3 (PFKFB3) Gene ITS POSSIBLE ROLE IN THE WARBURG EFFECT. *J. Biol. Chem.* **2002**, *277* (8), 6183–6187. https://doi.org/10.1074/jbc.M110978200.

(48) Obach, M.; Navarro-Sabaté, À.; Caro, J.; Kong, X.; Duran, J.; Gómez, M.; Perales, J. C.; Ventura, F.; Rosa, J. L.; Bartrons, R. 6-Phosphofructo-2-Kinase (Pfkfb3) Gene Promoter Contains Hypoxia-Inducible Factor-1 Binding Sites Necessary for Transactivation in Response to Hypoxia. *J. Biol. Chem.* **2004**, *279* (51), 53562–53570. https://doi.org/10.1074/jbc.M406096200.

(49) Okar, D. A.; Lange, A. J. Fructose-2,6-Bisphosphate and Control of Carbohydrate Metabolism in Eukaryotes. *Biofactors* **1999**, *10* (1), 1–14.

(50) Okar, D. A.; Lange, A. J.; Manzano, À.; Navarro-Sabatè, A.; Riera, L.; Bartrons, R. PFK-2/FBPase-2: Maker and Breaker of the Essential Biofactor Fructose-2,6-Bisphosphate. *Trends in Biochemical Sciences* **2001**, *26* (1), 30–35. https://doi.org/10.1016/S0968-0004(00)01699-6.

(51) El-Maghrabi, M. R.; Noto, F.; Wu, N.; Manes, N. 6-Phosphofructo-2-Kinase/Fructose-2,6-Bisphosphatase: Suiting Structure to Need, in a Family of Tissue-Specific Enzymes. *Curr Opin Clin Nutr Metab Care* **2001**, *4* (5), 411–418.

(52) Rider, M. H.; Bertrand, L.; Vertommen, D.; Michels, P. A.; Rousseau, G. G.; Hue, L. 6-Phosphofructo-2-Kinase/Fructose-2,6-Bisphosphatase: Head-to-Head with a Bifunctional Enzyme That Controls Glycolysis. *Biochemical Journal* **2004**, *381* (3), 561–579. https://doi.org/10.1042/BJ20040752.

(53) Sakakibara, R.; Kato, M.; Okamura, N.; Nakagawa, T.; Komada, Y.; Tominaga, N.; Shimojo, M.; Fukasawa, M. Characterization of a Human Placental Fructose-6-Phosphate, 2-Kinase/Fructose-2,6-Bisphosphatase. *J. Biochem.* **1997**, *122* (1), 122–128.

(54) Atsumi, T.; Chesney, J.; Metz, C.; Leng, L.; Donnelly, S.; Makita, Z.; Mitchell, R.; Bucala, R. High Expression of Inducible 6-Phosphofructo-2-Kinase/Fructose-2,6-Bisphosphatase (IPFK-2; PFKFB3) in Human Cancers. *Cancer Res.* **2002**, *62* (20), 5881–5887.

(55) Marsin, A.-S.; Bouzin, C.; Bertrand, L.; Hue, L. The Stimulation of Glycolysis by Hypoxia in Activated Monocytes Is Mediated by AMP-Activated Protein Kinase and Inducible 6-Phosphofructo-2-Kinase. *J. Biol. Chem.* **2002**, *277* (34), 30778–30783. https://doi.org/10.1074/jbc.M205213200.

(56) Manes, N. P.; El-Maghrabi, M. R. The Kinase Activity of Human Brain 6-Phosphofructo-2-Kinase/Fructose-2,6-Bisphosphatase Is Regulated via Inhibition by

Phosphoenolpyruvate. *Archives of Biochemistry and Biophysics* **2005**, *438* (2), 125–136. https://doi.org/10.1016/j.abb.2005.04.011.

(57) Hirata, T.; Watanabe, M.; Miura, S.; Ijichi, K.; Fukasawa, M.; Sakakibara, R. Inhibition of Tumor Cell Growth by A Specific 6-Phosphofructo-2-Kinase Inhibitor, N-Bromoacetylethanolamine Phosphate, and Its Analogues. *Bioscience, Biotechnology, and Biochemistry* **2000**, *64* (10), 2047–2052. https://doi.org/10.1271/bbb.64.2047.

(58) Seo, M.; Kim, J.-D.; Neau, D.; Sehgal, I.; Lee, Y.-H. Structure-Based Development of Small Molecule PFKFB3 Inhibitors: A Framework for Potential Cancer Therapeutic Agents Targeting the Warburg Effect. *PLoS ONE* **2011**, *6* (9), e24179. https://doi.org/10.1371/journal.pone.0024179.

(59) Crochet, R. B.; Cavalier, M. C.; Seo, M.; Kim, J.-D.; Yim, Y.-S.; Park, S.-J.; Lee, Y.-H. Investigating Combinatorial Approaches in Virtual Screening on PFKFB3: A Case Study for Small Molecule Kinases. *Anal Biochem* **2011**, *418* (1), 143–148. https://doi.org/10.1016/j.ab.2011.06.035.

(60) Brooke, D. G.; van Dam, E. M.; Watts, C. K. W.; Khoury, A.; Dziadek, M. A.; Brooks, H.; Graham, L.-J. K.; Flanagan, J. U.; Denny, W. A. Targeting the Warburg Effect in Cancer; Relationships for 2-Arylpyridazinones as Inhibitors of the Key Glycolytic Enzyme 6-Phosphofructo-2-Kinase/2,6-Bisphosphatase 3 (PFKFB3). *Bioorg. Med. Chem.* **2014**, *22* (3), 1029–1039. https://doi.org/10.1016/j.bmc.2013.12.041.

(61) Gustafsson, N. M. S.; Färnegårdh, K.; Bonagas, N.; Ninou, A. H.; Groth, P.; Wiita, E.; Jönsson, M.; Hallberg, K.; Lehto, J.; Pennisi, R.; *et al*. Targeting PFKFB3 Radiosensitizes Cancer Cells and Suppresses Homologous Recombination. *Nature Communications* **2018**, *9* (1), 3872. https://doi.org/10.1038/s41467-018-06287-x.

(62) Boutard, N.; Białas, A.; Sabiniarz, A.; Guzik, P.; Banaszak, K.; Biela, A.; Bień, M.; Buda, A.; Bugaj, B.; Cieluch, E.; *et al*. Discovery and Structure-Activity Relationships of N-Aryl 6-Aminoquinoxalines as Potent PFKFB3 Kinase Inhibitors. *ChemMedChem* **2018**. https://doi.org/10.1002/cmdc.201800569.

(63) Clem, B.; Telang, S.; Clem, A.; Yalcin, A.; Meier, J.; Simmons, A.; Rasku, M. A.; Arumugam, S.; Dean, W. L.; Eaton, J.; *et al*. Small-Molecule Inhibition of 6-Phosphofructo-2-Kinase Activity Suppresses Glycolytic Flux and Tumor Growth. *Mol. Cancer Ther.* **2008**, *7* (1), 110–120. https://doi.org/10.1158/1535-7163.MCT-07-0482.

(64) Clem, B. F.; O'Neal, J.; Tapolsky, G.; Clem, A. L.; Imbert-Fernandez, Y.; Kerr, D. A.; Klarer, A. C.; Redman, R.; Miller, D. M.; Trent, J. O.; *et al*. Targeting 6-Phosphofructo-2-Kinase (PFKFB3) as a Therapeutic Strategy against Cancer. *Mol. Cancer Ther.* **2013**, *12* (8), 1461–1470. https://doi.org/10.1158/1535-7163.MCT-13-0097.

(65) Chand, P.; Chesney, J. A.; Clem, B. F.; Tapolsky, G. H.; Telang, S.; Trent, J. O. Small Molecule Inhibitors of Pfkfb3 and Glycolytic Flux and Their Methods of Use as Anti-Cancer Therapeutics. WO2011103557A1, August 25, 2011.

(66) Zhu, W.; Ye, L.; Zhang, J.; Yu, P.; Wang, H.; Ye, Z.; Tian, J. PFK15, a Small Molecule Inhibitor of PFKFB3, Induces Cell Cycle Arrest, Apoptosis and Inhibits

Invasion in Gastric Cancer. *PLoS One* **2016**, *11* (9). https://doi.org/10.1371/journal.pone.0163768.

(67) Redman, R. A.; Pohlmann, P. R.; Kurman, M. R.; Tapolsky, G.; Chesney, J. A. A Phase I, Dose-Escalation, Multi-Center Study of PFK-158 in Patients with Advanced Solid Malignancies Explores a First-in-Man Inhbibitor of Glycolysis. *JCO* **2015**, *33* (15_suppl), TPS2606–TPS2606. https://doi.org/10.1200/jco.2015.33.15_suppl.tps2606.

(68) Boyd, S.; Brookfield, J. L.; Critchlow, S. E.; Cumming, I. A.; Curtis, N. J.; Debreczeni, J.; Degorce, S. L.; Donald, C.; Evans, N. J.; Groombridge, S.; *et al*. Structure-Based Design of Potent and Selective Inhibitors of the Metabolic Kinase PFKFB3. *J. Med. Chem.* **2015**, *58* (8), 3611–3625. https://doi.org/10.1021/acs.jmedchem.5b00352.

(69) Pyrkov, T. V.; Sevostyanova, I. A.; Schmalhausen, E. V.; Shkoporov, A. N.; Vinnik, A. A.; Muronetz, V. I.; Severin, F. F.; Fedichev, P. O. Structure-Based Design of Small-Molecule Ligands of Phosphofructokinase-2 Activating or Inhibiting Glycolysis. *ChemMedChem* **2013**, *8* (8), 1322–1329. https://doi.org/10.1002/cmdc.201300154.

(70) Angbrant, J.; Homan, E.; LUNDBÄCK, T.; Martinsson, J.; Sari, M.; JÖNSSON, M.; FÄRNEGÅRDH, K.; Hallberg, K. Bisarylsulfonamides Useful as Kinase Inhibitors in the Treatment of Inflammation and Cancer. WO2011161201A1, December 29, 2011.

(71) Aay, N.; Bajjalieh, W.; Huang, P.; Kabigting, L.; Kearney, P.; Marlowe, C. K.; Moon, K.; Nuss, J. M.; TAMBO-ONG, A.; Tesfai, Z.; *et al*. Inhibitors of Inducible Form of 6-Phosphofructose-2-Kinase. WO2012149528A1, November 1, 2012.

(72) St-Gallay, S. A.; Bennett, N.; Critchlow, S. E.; Curtis, N.; Davies, G.; Debreczeni, J.; Evans, N.; Hardern, I.; Holdgate, G.; Jones, N. P.; *et al*. A High-Throughput Screening Triage Workflow to Authenticate a Novel Series of PFKFB3 Inhibitors. *SLAS Discov* **2018**, *23* (1), 11–22. https://doi.org/10.1177/2472555217732289.

(73) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual Screening—an Overview. *Drug Discovery Today* **1998**, *3* (4), 160–178. https://doi.org/10.1016/S1359-6446(97)01163-X.

(74) Yu, W.; MacKerell, A. D. Computer-Aided Drug Design Methods. *Methods Mol Biol* **2017**, *1520*, 85–106. https://doi.org/10.1007/978-1-4939-6634-9_5.

(75) Lavecchia, A.; Di Giovanni, C. Virtual Screening Strategies in Drug Discovery: A Critical Review. *Curr. Med. Chem.* **2013**, *20* (23), 2839–2860.

(76) Kirkpatrick, P.; Ellis, C. Chemical Space. *Nature* **2004**, *432*, 823. https://doi.org/10.1038/432823a.

(77) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings1PII of Original Article: S0169-409X(96)00423-1. The Article Was Originally Published in Advanced Drug Delivery Reviews 23 (1997) 3–25.1. *Advanced Drug Delivery Reviews* **2001**, *46* (1), 3–26. https://doi.org/10.1016/S0169-409X(00)00129-0.

(78) Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A “rule of Three” for Fragment-Based Lead Discovery? *Drug Discov. Today* **2003**, *8* (19), 876–877.

(79) Hann, M. M.; Oprea, T. I. Pursuing the Leadlikeness Concept in Pharmaceutical Research. *Curr Opin Chem Biol* **2004**, *8* (3), 255–263. https://doi.org/10.1016/j.cbpa.2004.04.003.

(80) Oprea, T. I. Property Distribution of Drug-Related Chemical Databases. *J. Comput. Aided Mol. Des.* **2000**, *14* (3), 251–264.

(81) Maggiora, G.; Vogt, M.; Stumpfe, D.; Bajorath, J. Molecular Similarity in Medicinal Chemistry. *J. Med. Chem.* **2014**, *57* (8), 3186–3204. https://doi.org/10.1021/jm401411z.

(82) Böhm, H.-J.; Flohr, A.; Stahl, M. Scaffold Hopping. *Drug Discovery Today: Technologies* **2004**, *1* (3), 217–224. https://doi.org/10.1016/j.ddtec.2004.10.009.

(83) Quintus, F.; Sperandio, O.; Grynberg, J.; Petitjean, M.; Tuffery, P. Ligand Scaffold Hopping Combining 3D Maximal Substructure Search and Molecular Similarity. *BMC Bioinformatics* **2009**, *10*, 245. https://doi.org/10.1186/1471-2105-10-245.

(84) Andreas Bender, J.-L. F. Handbook of Chemoinformatics Algorithms https://www.crcpress.com/Handbook-of-Chemoinformatics-Algorithms/Faulon-Bender/p/book/9781420082920.

(85) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55* (14), 6582–6594. https://doi.org/10.1021/jm300687e.

(86) Thimm, M.; Goede, A.; Hougardy, S.; Preissner, R. Comparison of 2D Similarity and 3D Superposition. Application to Searching a Conformational Drug Database. *Journal of Chemical Information and Computer Sciences* **2004**, *44* (5), 1816–1822. https://doi.org/10.1021/ci049920h.

(87) Bauer, R. A.; Bourne, P. E.; Formella, A.; Frömmel, C.; Gille, C.; Goede, A.; Guerler, A.; Hoppe, A.; Knapp, E.-W.; Pöschel, T.; *et al*. Superimposé: A 3D Structural Superposition Server. *Nucleic Acids Res* **2008**, *36* (Web Server issue), W47–W54. https://doi.org/10.1093/nar/gkn285.

(88) Rush, T. S.; Grant, J. A.; Mosyak, L.; Nicholls, A. A Shape-Based 3-D Scaffold Hopping Method and Its Application to a Bacterial Protein-Protein Interaction. *J. Med. Chem.* **2005**, *48* (5), 1489–1495. https://doi.org/10.1021/jm040163o.

(89) Sperandio, O.; Petitjean, M.; Tuffery, P. WwLigCSRre: A 3D Ligand-Based Server for Hit Identification and Optimization. *Nucleic Acids Res.* **2009**, *37* (Web Server issue), W504-509. https://doi.org/10.1093/nar/gkp324.

(90) Petitjean, M. Interactive Maximal Common 3D Substructure Searching with the Combined SDM/RMS Algorithm. *Computers & Chemistry* **1998**, *22* (6), 463–465. https://doi.org/10.1016/S0097-8485(98)00017-5.

(91) Wermuth, C. G.; Ganellin, C. R.; Lindberg, P.; Mitscher, L. A. Glossary of Terms

Used in Medicinal Chemistry (IUPAC Recommendations 1998). *Pure and Applied Chemistry* **1998**, *70* (5), 1129–1143. https://doi.org/10.1351/pac199870051129.

(92) Langer, T.; Wolber, G. Pharmacophore Definition and 3D Searches. *Drug Discovery Today: Technologies* **2004**, *1* (3), 203–207. https://doi.org/10.1016/j.ddtec.2004.11.015.

(93) Strategies for 3D Pharmacophore-Based Virtual Screening. *Drug Discov Today Technol* **2010**, *7* (4), e203-270. https://doi.org/10.1016/j.ddtec.2010.11.004.

(94) Greene, J.; Kahn, S.; Savoj, H.; Sprague, P.; Teig, S. Chemical Function Queries for 3D Database Search. *J. Chem. Inf. Comput. Sci.* **1994**, *34* (6), 1297–1308. https://doi.org/10.1021/ci00022a012.

(95) Kurogi, Y.; Güner, O. F. Pharmacophore Modeling and Three-Dimensional Database Searching for Drug Design Using Catalyst. *Curr. Med. Chem.* **2001**, *8* (9), 1035–1055.

(96) Seidel, T.; Bryant, S. D.; Ibis, G.; Poli, G.; Langer, T. 3D Pharmacophore Modeling Techniques in Computer-Aided Molecular Design Using LigandScout. In *Tutorials in Chemoinformatics*; Varnek, A., Ed.; John Wiley & Sons, Ltd: Chichester, UK, 2017; pp 279–309. https://doi.org/10.1002/9781119161110.ch20.

(97) Dixon, S. L.; Smondyrev, A. M.; Knoll, E. H.; Rao, S. N.; Shaw, D. E.; Friesner, R. A. PHASE: A New Engine for Pharmacophore Perception, 3D QSAR Model Development, and 3D Database Screening: 1. Methodology and Preliminary Results. *J. Comput. Aided Mol. Des.* **2006**, *20* (10–11), 647–671. https://doi.org/10.1007/s10822-006-9087-6.

(98) Molecular Operating Environment (MOE), 2013.08; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2018 https://www.chemcomp.com/announcements/2015-12-21.htm.

(99) Hurst, T. Flexible 3D Searching: The Directed Tweak Technique. *J. Chem. Inf. Comput. Sci.* **1994**, *34* (1), 190–196. https://doi.org/10.1021/ci00017a025.

(100) RCSB PDB: Homepage https://www.rcsb.org/.

(101) Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Curr Comput Aided Drug Des* **2011**, *7* (2), 146–157.

(102) Miller, M. D.; Kearsley, S. K.; Underwood, D. J.; Sheridan, R. P. FLOG: A System to Select “quasi-Flexible” Ligands Complementary to a Receptor of Known Three-Dimensional Structure. *J. Comput. Aided Mol. Des.* **1994**, *8* (2), 153–174.

(103) Korb, O.; Stützle, T.; Exner, T. E. An Ant Colony Optimization Approach to Flexible Protein–Ligand Docking. *Swarm Intell* **2007**, *1* (2), 115–134. https://doi.org/10.1007/s11721-007-0006-9.

(104) Elokely, K. M.; Doerksen, R. J. Docking Challenge: Protein Sampling and Molecular Docking Performance. *J Chem Inf Model* **2013**, *53* (8), 1934–1945. https://doi.org/10.1021/ci400040d.

(105) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat Rev Drug Discov* **2004**, *3* (11), 935–949. https://doi.org/10.1038/nrd1549.

(106) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A Geometric Approach to Macromolecule-Ligand Interactions. *J. Mol. Biol.* **1982**, *161* (2), 269–288.

(107) Böhm, H. J. LUDI: Rule-Based Automatic Design of New Substituents for Enzyme Inhibitor Leads. *J. Comput. Aided Mol. Des.* **1992**, *6* (6), 593–606.

(108) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved Protein-Ligand Docking Using GOLD. *Proteins* **2003**, *52* (4), 609–623. https://doi.org/10.1002/prot.10465.

(109) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *Journal of Computational Chemistry* **1998**, *19* (14), 1639–1662. https://doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B.

(110) Korb, O.; Stützle, T.; Exner, T. E. Empirical Scoring Functions for Advanced Protein−Ligand Docking with PLANTS. *J. Chem. Inf. Model.* **2009**, *49* (1), 84–96. https://doi.org/10.1021/ci800298z.

(111) Korb, O.; Stützle, T.; Exner, T. E. PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design. In: Dorigo M., Gambardella L.M., Birattari M., Martinoli A., Poli R., Stützle T. (Eds) Ant Colony Optimization and Swarm Intelligence. ANTS 2006. Lecture Notes in Computer Science, Vol 4150. Springer, Berlin, Heidelberg. In *Ant Colony Optimization and Swarm Intelligence*; Dorigo, M., Gambardella, L. M., Birattari, M., Martinoli, A., Poli, R., Stützle, T., Eds.; Lecture Notes in Computer Science; Springer Berlin Heidelberg, 2006; pp 247–258.

(112) McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of Folded Proteins. *Nature* **1977**, *267* (5612), 585. https://doi.org/10.1038/267585a0.

(113) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J Comput Chem* **2004**, *25* (9), 1157–1174. https://doi.org/10.1002/jcc.20035.

(114) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179–5197. https://doi.org/10.1021/ja00124a002.

(115) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry* **1983**, *4* (2), 187–217. https://doi.org/10.1002/jcc.540040211.

(116) Christen, M.; Hünenberger, P. H.; Bakowies, D.; Baron, R.; Bürgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Kräutler, V.; Oostenbrink, C.; *et al*. The GROMOS Software for Biomolecular Simulation: GROMOS05. *Journal of Computational*

Chemistry **2005**, *26* (16), 1719–1751. https://doi.org/10.1002/jcc.20303.

(117) Xu, L.; Sun, H.; Li, Y.; Wang, J.; Hou, T. Assessing the Performance of MM/PBSA and MM/GBSA Methods. 3. The Impact of Force Fields and Ligand Charge Models. *The Journal of Physical Chemistry B* **2013**, *117* (28), 8408–8421. https://doi.org/10.1021/jp404160y.

(118) Homeyer, N.; Gohlke, H. Free Energy Calculations by the Molecular Mechanics Poisson−Boltzmann Surface Area Method. *Molecular Informatics* **2012**, *31* (2), 114–122. https://doi.org/10.1002/minf.201100135.

(119) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opinion on Drug Discovery* **2015**, *10* (5), 449–461. https://doi.org/10.1517/17460441.2015.1032936.

(120) Case, D. A.; Babin, V.; Berryman, J.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham III, T. E.; Darden, T. A.; Duke, R. E.; Gohlke, H.; *et al*. Amber 14. **2014**.

(121) Genheden, S.; Ryde, U. Comparison of End-Point Continuum-Solvation Methods for the Calculation of Protein–Ligand Binding Free Energies. *Proteins: Structure, Function, and Bioinformatics* **2012**, *80* (5), 1326–1342. https://doi.org/10.1002/prot.24029.

(122) Weis, A.; Katebzadeh, K.; Söderhjelm, P.; Nilsson, I.; Ryde, U. Ligand Affinities Predicted with the MM/PBSA Method: Dependence on the Simulation Method and the Force Field. *Journal of Medicinal Chemistry* **2006**, *49* (22), 6596–6606. https://doi.org/10.1021/jm0608210.

(123) Steinbrecher, T.; Case, D. A.; Labahn, A. A Multistep Approach to Structure-Based Drug Design: Studying Ligand Binding at the Human Neutrophil Elastase †. *Journal of Medicinal Chemistry* **2006**, *49* (6), 1837–1844. https://doi.org/10.1021/jm0505720.

(124) Maffucci, I.; Contini, A. Improved Computation of Protein-Protein Relative Binding Energies with the Nwat-MMGBSA Method. *J Chem Inf Model* **2016**, *56* (9), 1692–1704. https://doi.org/10.1021/acs.jcim.6b00196.

(125) Maffucci, I.; Contini, A. Explicit Ligand Hydration Shells Improve the Correlation between MM-PB/GBSA Binding Energies and Experimental Activities. *J Chem Theory Comput* **2013**, *9* (6), 2706–2717. https://doi.org/10.1021/ct400045d.

(126) Maffucci, I.; Hu, X.; Fumagalli, V.; Contini, A. An Efficient Implementation of the Nwat-MMGBSA Method to Rescore Docking Results in Medium-Throughput Virtual Screenings. *Front Chem* **2018**, *6*. https://doi.org/10.3389/fchem.2018.00043.

(127) Mikulskis, P.; Genheden, S.; Ryde, U. Effect of Explicit Water Molecules on Ligand-Binding Affinities Calculated with the MM/GBSA Approach. *Journal of Molecular Modeling* **2014**, *20* (6). https://doi.org/10.1007/s00894-014-2273-x.

(128) Ferri, N.; Corsini, A.; Bottino, P.; Clerici, F.; Contini, A. Virtual Screening Approach for the Identification of New Rac1 Inhibitors. *J. Med. Chem.* **2009**, *52* (14), 4087–4090. https://doi.org/10.1021/jm8015987.

(129) Irwin, J. J.; Shoichet, B. K. ZINC – A Free Database of Commercially Available Compounds for Virtual Screening. *J Chem Inf Model* **2005**, *45* (1), 177–182. https://doi.org/10.1021/ci049714.

(130) DUD-E: A Database of Useful (Docking) Decoys — Enhanced http://dude.docking.org/.

(131) Gerber, P. R.; Müller, K. MAB, a Generally Applicable Molecular Force Field for Structure Modelling in Medicinal Chemistry. *J. Comput. Aided Mol. Des.* **1995**, *9* (3), 251–268.

(132) AMBER 12, University of California, San Francisco. https://www.researchgate.net/publication/313442058_AMBER_12_University_of_California_San_Francisco.

(133) Fiser, A.; Šali, A. Modeller: Generation and Refinement of Homology-Based Protein Structure Models. In *Methods in Enzymology*; Macromolecular Crystallography, Part D; Academic Press, 2003; Vol. 374, pp 461–491. https://doi.org/10.1016/S0076-6879(03)74020-8.

(134) De Coen, L. M.; Heugebaert, T. S. A.; García, D.; Stevens, C. V. Synthetic Entries to and Biological Activity of Pyrrolopyrimidines. *Chemical Reviews* **2016**, *116* (1), 80–139. https://doi.org/10.1021/acs.chemrev.5b00483.

(135) wwLigCSRre: ligand similarity search. http://bioserv.rpbs.univ-paris-diderot.fr/services/wwLigCSRre/.

(136) Plewczynski, D.; Łaźniewski, M.; Augustyniak, R.; Ginalski, K. Can We Trust Docking Results? Evaluation of Seven Commonly Used Programs on PDBbind Database. *Journal of Computational Chemistry* **2011**, *32* (4), 742–755. https://doi.org/10.1002/jcc.21643.

(137) Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: A Comprehensive Resource for in Silico Drug Discovery and Exploration. *Nucleic Acids Res.* **2006**, *34* (Database issue), D668-672. https://doi.org/10.1093/nar/gkj067.

(138) Abadi, A. H.; el-Subbagh, H. I.; al-Khamees, H. A. Synthesis, Antitumor and Antitubercular Evaluation of Certain New Xanthenone and Acridinone Analogs. *Arzneimittelforschung* **1999**, *49* (3), 259–266. https://doi.org/10.1055/s-0031-1300411.

(139) Gobbi, S.; Rampa, A.; Bisi, A.; Belluti, F.; Valenti, P.; Caputo, A.; Zampiron, A.; Carrara, M. Synthesis and Antitumor Activity of New Derivatives of Xanthen-9-One-4-Acetic Acid. *J. Med. Chem.* **2002**, *45* (22), 4931–4939. https://doi.org/10.1021/jm020929p.

(140) Rewcastle, G. W.; Atwell, G. J.; Baguley, B. C.; Calveley, S. B.; Denny, W. A. Potential Antitumor Agents. 58. Synthesis and Structure-Activity Relationships of Substituted Xanthenone-4-Acetic Acids Active against the Colon 38 Tumor in Vivo. *J. Med. Chem.* **1989**, *32* (4), 793–799. https://doi.org/10.1021/jm00124a012.

(141) Filippatos, E.; Papadaki-Valiraki, A.; Todoulou, O.; Jacquemin-Sablon, A.

Synthesis of N-(9H-Xanthen-9-Yl)Aminoalkanamide and N-(9H-Thioxanthen-9-Yl)Aminoalkanamide Derivatives and Their in Vitro Evaluation as Potential Intercalators and Antitumor Drugs. *Arch. Pharm. (Weinheim)* **1994**, *327* (2), 61–66. https://doi.org/10.1002/ardp.19943270202.

(142) Sambiagio, C.; P. Marsden, S.; John Blacker, A.; C. McGowan, P. Copper Catalysed Ullmann Type Chemistry: From Mechanistic Aspects to Modern Development. *Chemical Society Reviews* **2014**, *43* (10), 3525–3550. https://doi.org/10.1039/C3CS60289C.

(143) Jones, W. D.; Schnettler, R. A.; Huber, E. W. A Convenient Synthesis of 5-acyl-6-substituted 3-cyano-2(1H)-pyridinones. *Journal of Heterocyclic Chemistry* **1990**, *27* (3), 511–518. https://doi.org/10.1002/jhet.5570270307.

(144) Li, G.; Watson, K.; Buckheit, R. W.; Zhang, Y. Total Synthesis of Anibamine, a Novel Natural Product as a Chemokine Receptor CCR5 Antagonist. *Org. Lett.* **2007**, *9* (10), 2043–2046. https://doi.org/10.1021/ol070748n.

(145) Singh, B.; Pennock, P. O.; Lesher, G. Y.; Bacon, E. R.; Page, D. F. An Efficient and Novel Synthesis of Fused Thiazol-2(3H)-Ones. *Heterocycles* **1993**, *36* (1), 133–144.

(146) Abu-Shanab, F. A.; Redhouse, A. D.; Thompson, J. R.; Wakefield, B. J. Synthesis of 2,3,5,6-Tetrasubstituted Pyridines from Enamines Derived from N,N-Dimethylformamide Dimethyl Acetal. *Synthesis* **1995**, *1995* (05), 557–560. https://doi.org/10.1055/s-1995-3954.

(147) Abu-Shanab, F. A.; Mousa, S. A. S.; Eshak, E. A.; Sayed, A. Z.; Al-Harrasi, A. Dimethylformamide Dimethyl Acetal (DMFDMA) in Heterocyclic Synthesis: Synthesis of Polysubstituted Pyridines, Pyrimidines, Pyridazine and Their Fused Derivatives. *International Journal of Organic Chemistry* **2011**, *01* (04), 207. https://doi.org/10.4236/ijoc.2011.14030.

(148) Li, Y.; Chen, H.; Liu, J.; Wan, X.; Xu, Q. Clean Synthesis of Primary to Tertiary Carboxamides by CsOH-Catalyzed Aminolysis of Nitriles in Water. *Green Chem.* **2016**, *18* (18), 4865–4870. https://doi.org/10.1039/C6GC01565D.

(149) Hasegawa, M.; Nishigaki, N.; Washio, Y.; Kano, K.; Harris, P. A.; Sato, H.; Mori, I.; West, R. I.; Shibahara, M.; Toyoda, H.; *et al*. Discovery of Novel Benzimidazoles as Potent Inhibitors of TIE-2 and VEGFR-2 Tyrosine Kinase Receptors. *J. Med. Chem.* **2007**, *50* (18), 4453–4470. https://doi.org/10.1021/jm0611051.

(150) Vaitla, J.; Bayer, A.; Hopmann, K. H. Synthesis of Indoles and Pyrroles Utilizing Iridium Carbenes Generated from Sulfoxonium Ylides. *Angewandte Chemie International Edition* **2017**, *56* (15), 4277–4281. https://doi.org/10.1002/anie.201610520.

(151) Mohamed, M. S.; Abd El-Hameed, R. H.; Sayed, A. I.; Soror, S. H. Novel Antiviral Compounds against Gastroenteric Viral Infections. *Arch. Pharm. (Weinheim)* **2015**, *348* (3), 194–205. https://doi.org/10.1002/ardp.201400387.

(152) Dave, C. G.; Shah, R. D. Synthesis of 7H-Tetrazolo[1,5-c]Pyrrolo[3,2-

e]Pyrimidines and Their Reductive Ring Cleavage to 4-Aminopyrrolo[2,3-d]Pyrimidines. *Journal of Heterocyclic Chemistry* **1998**, *35* (6), 1295–1300. https://doi.org/10.1002/jhet.5570350609.

(153) Mohamed, M. S.; Hussein, W. M.; McGeary, R. P.; Vella, P.; Schenk, G.; Abd El-Hameed, R. H. Synthesis and Kinetic Testing of New Inhibitors for a Metallo-β-Lactamase from Klebsiella Pneumonia and Pseudomonas Aeruginosa. *Eur J Med Chem* **2011**, *46* (12), 6075–6082. https://doi.org/10.1016/j.ejmech.2011.10.030.

(154) Barton, D. H. R.; Chern, C.-Y.; Tachdjian, C. Preparation of New Thiohydroxamic Acid Derivatives: Synthesis of Substituted 1-Hydroxy-1,2-Dihydroimidazole-2-Thiones. *Heterocycles* **1994**, *37* (2), 793. https://doi.org/10.1002/chin.199429171.

(155) Fatahala, S. S.; Mahgub, S.; Taha, H.; Hameed, R. H. A.-E. Synthesis and Evaluation of Novel Spiro Derivatives for Pyrrolopyrimidines as Anti-Hyperglycemia Promising Compounds. *Journal of Enzyme Inhibition and Medicinal Chemistry* **2018**, *33* (1), 809–817. https://doi.org/10.1080/14756366.2018.1461854.

(156) Dave, C. G.; Patel, K. J. Synthesis of 5,7-Disubstituted 7H-Pyrrolo[2,3-d]Pyrimidin-4(3H)-Ones and Their N-Alkylation's under Phase Transfer Conditions. *Journal of Heterocyclic Chemistry* **2014**, *51* (4), 943–947. https://doi.org/10.1002/jhet.1561.

(157) Kugelman, M.; Gala, D.; Jaret, R. S.; Nyce, P. L.; McPhail, A. T. Synthesis of Azetidinones from L-Threonine: Formation of Unusually Stable Bicyclic Hemiketals and Cis-Azetidinones. *Synlett* **1990**, *1990* (7), 431–432. https://doi.org/10.1055/s-1990-21119.

(158) Mohamed, M. S.; Rashad, A. E.; Zaki, M. E. A.; Fatahala, S. S. Synthesis and Antimicrobial Screening of Some Fused Heterocyclic Pyrroles. *Acta Pharm* **2005**, *55* (3), 237–249.

(159) Wang, S.; Meades, C.; Wood, G.; Osnowski, A.; Anderson, S.; Yuill, R.; Thomas, M.; Mezna, M.; Jackson, W.; Midgley, C.; *et al.* 2-Anilino-4-(Thiazol-5-Yl)Pyrimidine CDK Inhibitors: Synthesis, SAR Analysis, X-Ray Crystallography, and Biological Activity. *Journal of Medicinal Chemistry* **2004**, *47* (7), 1662–1675. https://doi.org/10.1021/jm0309957.

(160) Bredereck, H.; Effenberger, F.; Botsch, H. Säureamid-Reaktionen, XLV. Untersuchungen über die Reaktionsfähigkeit von Formamidinen, Dimethylformamid-diäthylacetal (Amidacetal) und Bis-dimethyl-amino-methoxy-methan (Aminalester). *Chemische Berichte* **1964**, *97* (12), 3397–3406. https://doi.org/10.1002/cber.19640971220.

(161) Zegzouti, H.; Zdanovskaia, M.; Hsiao, K.; Goueli, S. A. ADP-Glo: A Bioluminescent and Homogeneous ADP Monitoring Assay for Kinases. *Assay Drug Dev Technol* **2009**, *7* (6), 560–572. https://doi.org/10.1089/adt.2009.0222.

(162) D. Weininger Delany J., C. A. J. Daylight Theory Manual; Daylight Chemical Information System, Inc.: Mission Viejo. **2004**, 59.

(163) Halgren, T. A. MMFF VII. Characterization of MMFF94, MMFF94s, and Other Widely Available Force Fields for Conformational Energies and for Intermolecular-Interaction Energies and Geometries. *Journal of Computational Chemistry* **1999**, *20* (7), 730–748. https://doi.org/10.1002/(SICI)1096-987X(199905)20:7<730::AID-JCC8>3.0.CO;2-T.

(164) Needleman, S. B.; Wunsch, C. D. A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins. *Journal of Molecular Biology* **1970**, *48* (3), 443–453. https://doi.org/10.1016/0022-2836(70)90057-4.

(165) ten Brink, T.; Exner, T. E. Influence of Protonation, Tautomeric, and Stereoisomeric States on Protein-Ligand Docking Results. *J Chem Inf Model* **2009**, *49* (6), 1535–1546. https://doi.org/10.1021/ci800420z.

(166) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.

(167) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935. https://doi.org/10.1063/1.445869.

(168) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *Journal of Computational Physics* **1977**, *23* (3), 327–341. https://doi.org/10.1016/0021-9991(77)90098-5.

(169) Wang, J.; Uttamchandani, M.; Li, J.; Hu, M.; Yao, S. Q. “Click” Synthesis of Small Molecule Probes for Activity-Based Fingerprinting of Matrix Metalloproteases. *Chem. Commun.* **2006**, *0* (36), 3783–3785. https://doi.org/10.1039/B609446E.

(170) Ohwada, T.; Ishikawa, S.; Mine, Y.; Inami, K.; Yanagimoto, T.; Karaki, F.; Kabasawa, Y.; Otani, Y.; Mochizuki, M. 7-Azabicyclo[2.2.1]Heptane as a Structural Motif to Block Mutagenicity of Nitrosamines. *Bioorganic & medicinal chemistry, Bioorganic & medicinal chemistry., Bioorganic and medicinal chemistry* **2011**, *19* (8), 2726–2741. https://doi.org/10.1016/j.bmc.2011.02.049.

(171) Sarkar, A.; Roy, S. R.; Parikh, N.; Chakraborti, A. K. Nonsolvent Application of Ionic Liquids: Organo-Catalysis by 1-Alkyl-3-Methylimidazolium Cation Based Room-Temperature Ionic Liquids for Chemoselective N-Tert-Butyloxycarbonylation of Amines and the Influence of the C-2 Hydrogen on Catalytic Efficiency. *J. Org. Chem.* **2011**, *76* (17), 7132–7140. https://doi.org/10.1021/jo201102q.

(172) Chatterjee, N.; Arfeen, M.; Bharatam, P. V.; Goswami, A. A Metal and Base-Free Chemoselective Primary Amination of Boronic Acids Using Cyanamidyl/Arylcyanamidyl Radical as Aminating Species: Synthesis and Mechanistic Studies by Density Functional Theory. *J. Org. Chem.* **2016**, *81* (12), 5120–5127. https://doi.org/10.1021/acs.joc.6b00671.

(173) Martinez, G. E.; Nugent, J. W.; Fout, A. R. Simple Nickel Salts for the Amination of (Hetero)Aryl Bromides and Iodides with Lithium Bis(Trimethylsilyl)Amide.

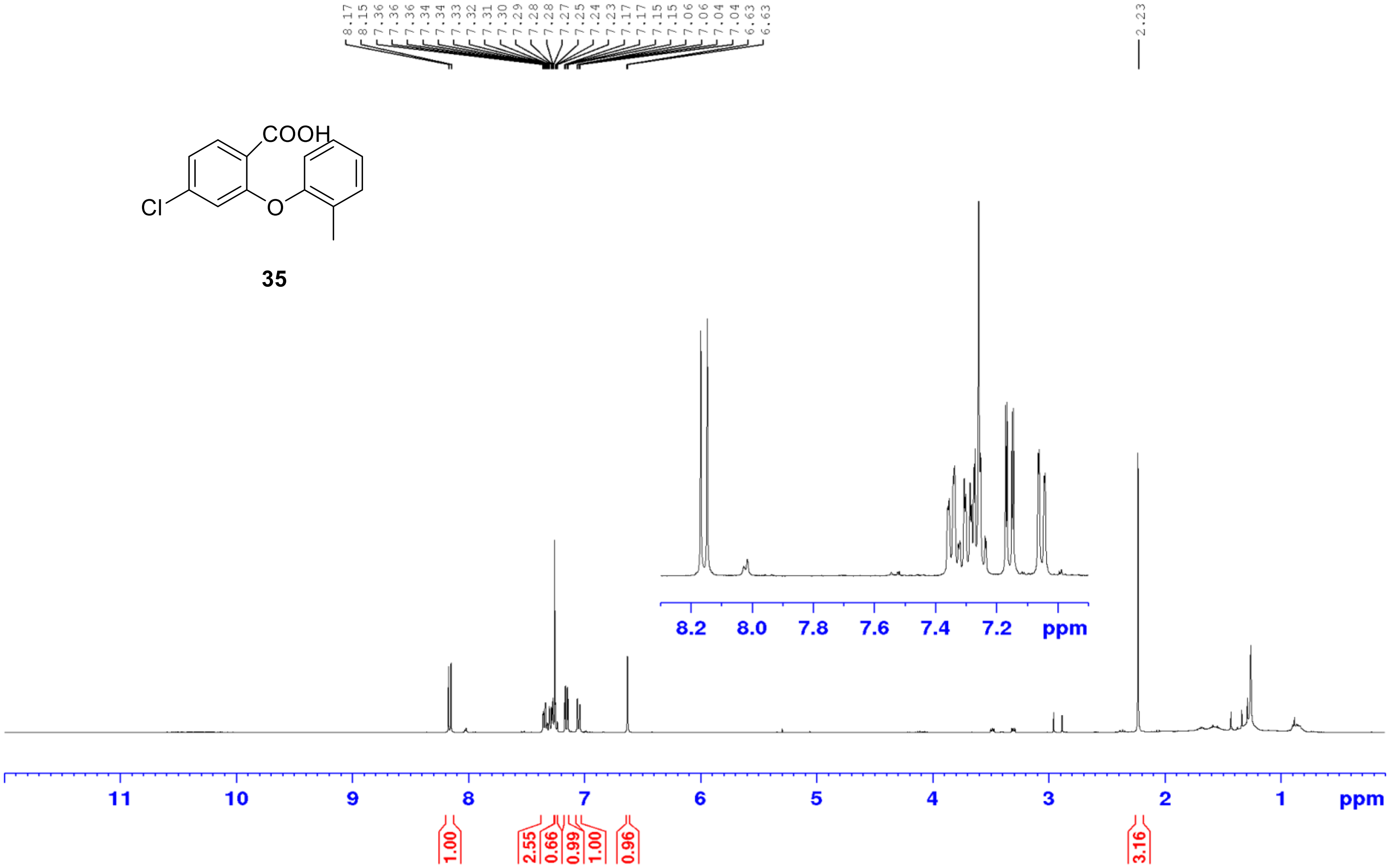
Organometallics **2018**, *37* (18), 2941–2944. https://doi.org/10.1021/acs.organomet.8b00567.

(174) Amir, E.; Rozen, S. Easy Access to the Family of Thiazole N-Oxides Using HOF·CH3CN. *Chem. Commun.* **2006**, *0* (21), 2262–2264. https://doi.org/10.1039/B602594C.

APPENDIX

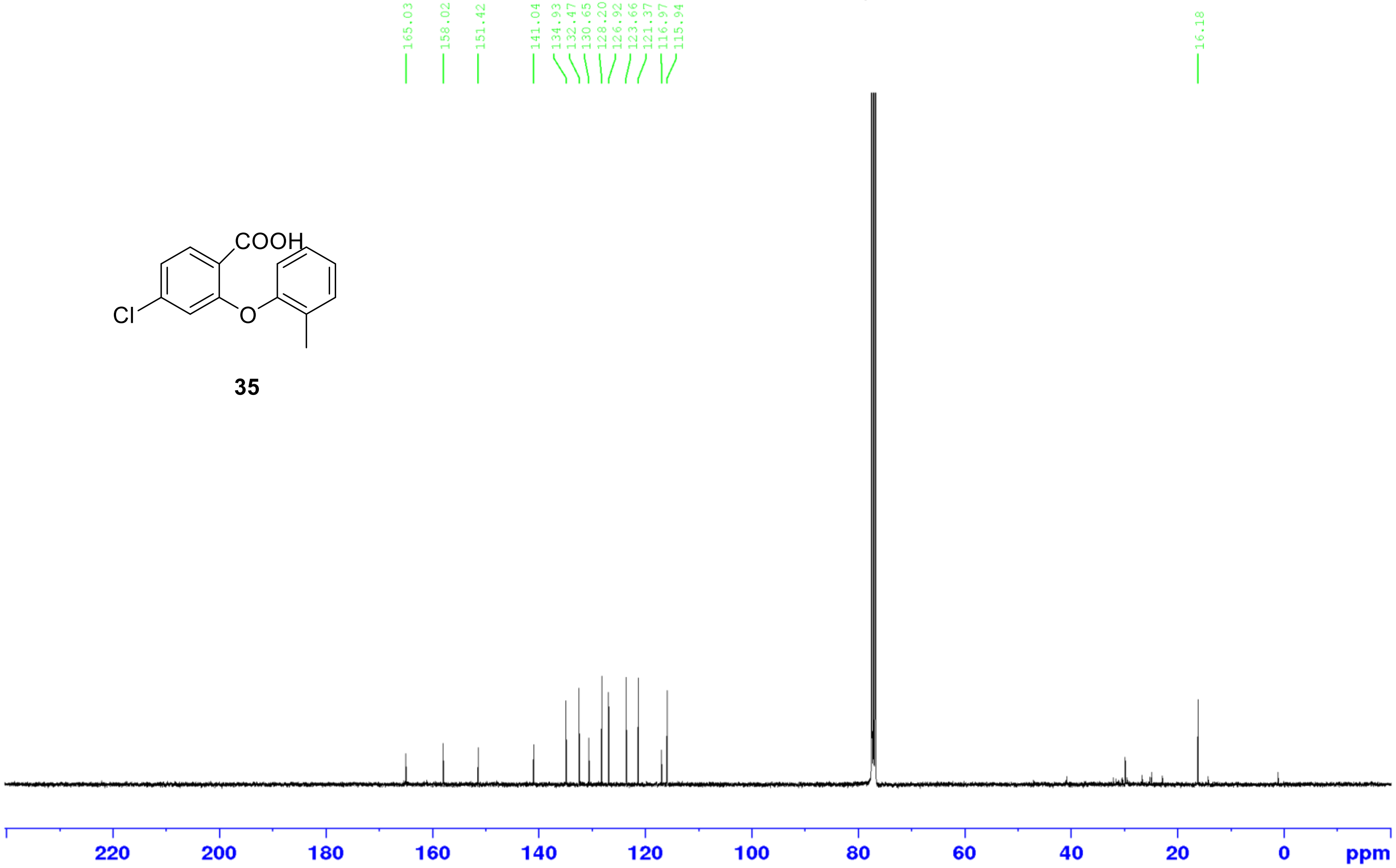
^{1}H, ^{13}C and ^{19}F NMR spectra

^{1}H NMR (400 MHz, CDCl$_3$)

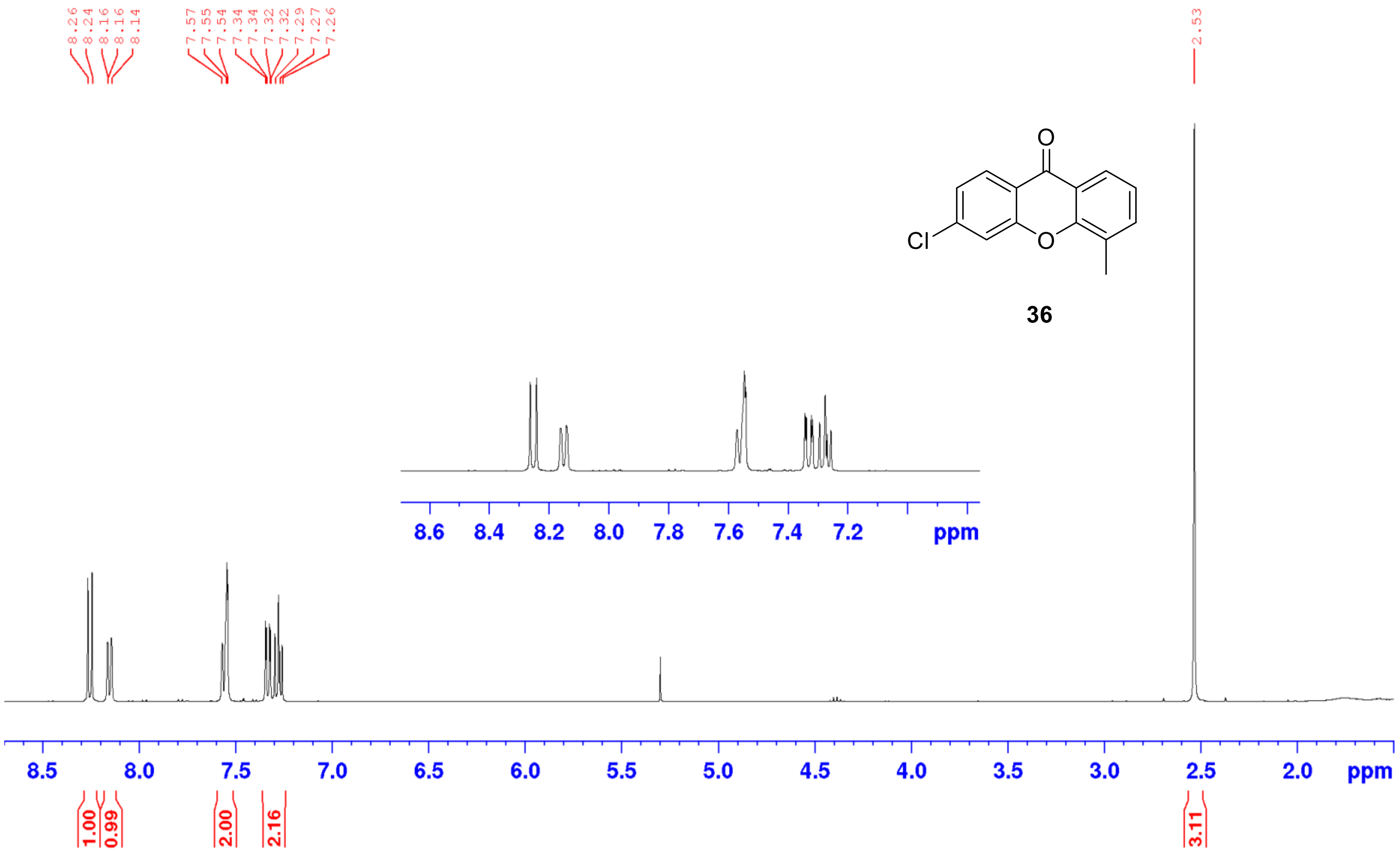


35

^{13}C NMR (101 MHz, $CDCl_3$)

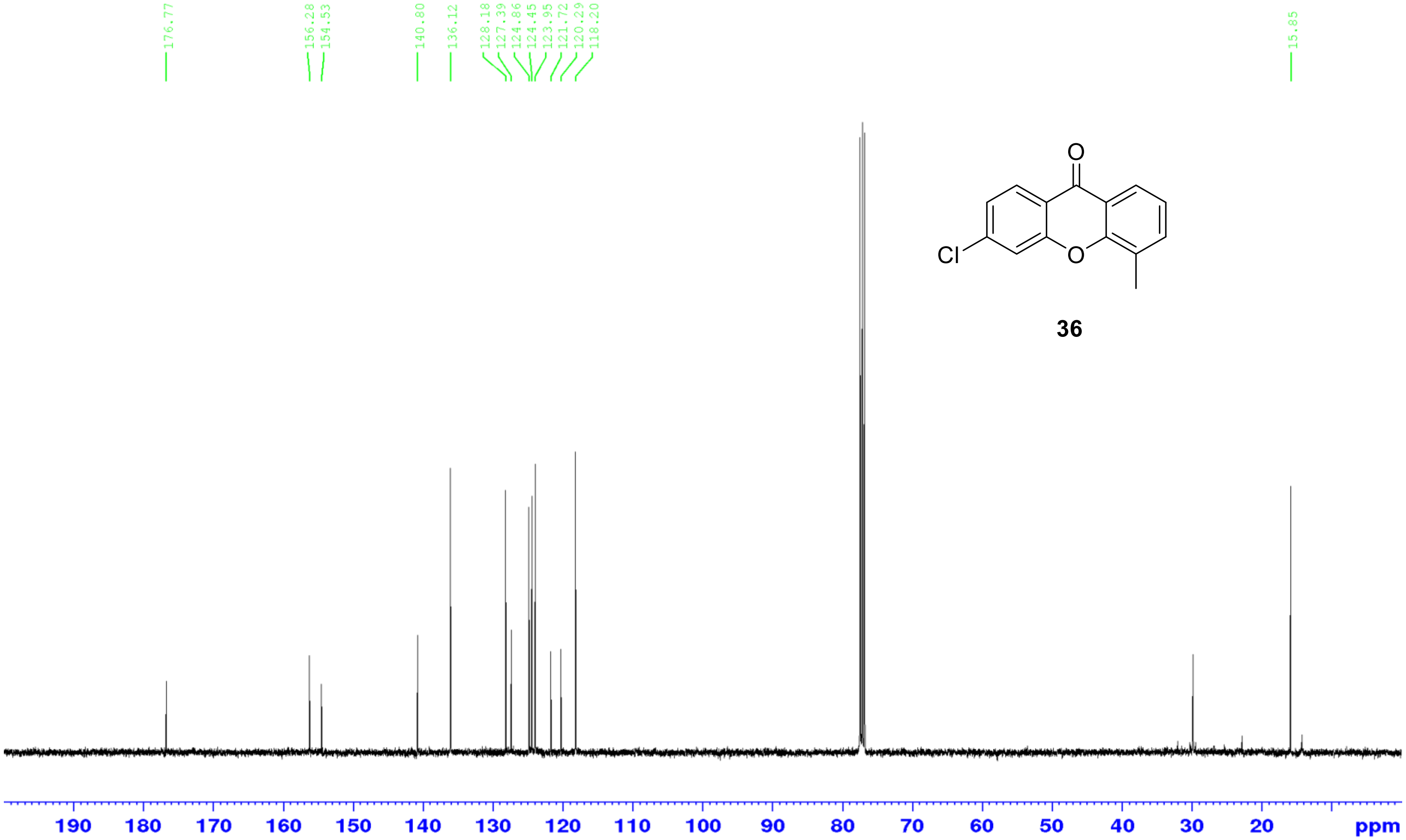


^{1}H NMR (400 MHz, $CDCl_3$)



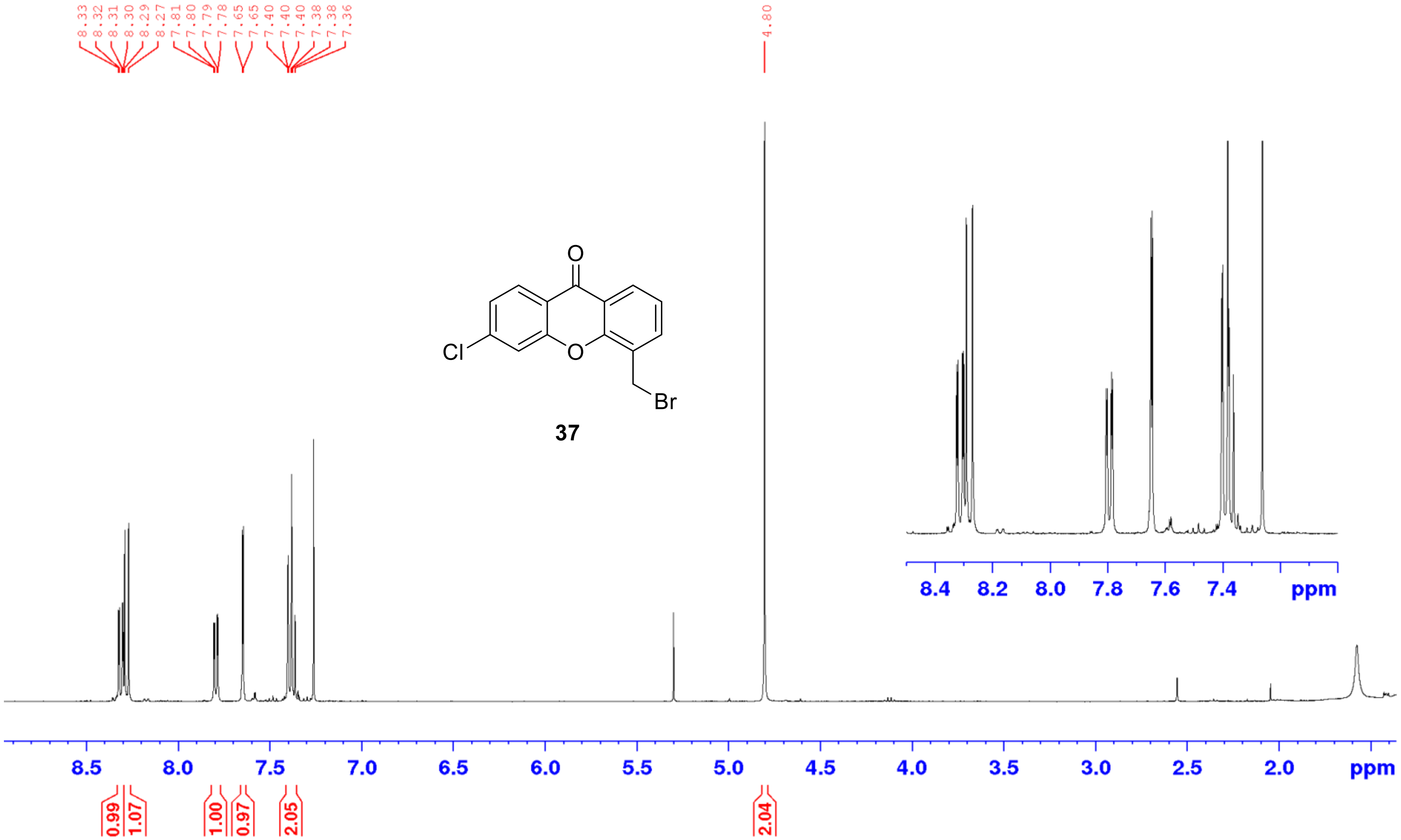
36

^{13}C NMR (101 MHz, $CDCl_3$)



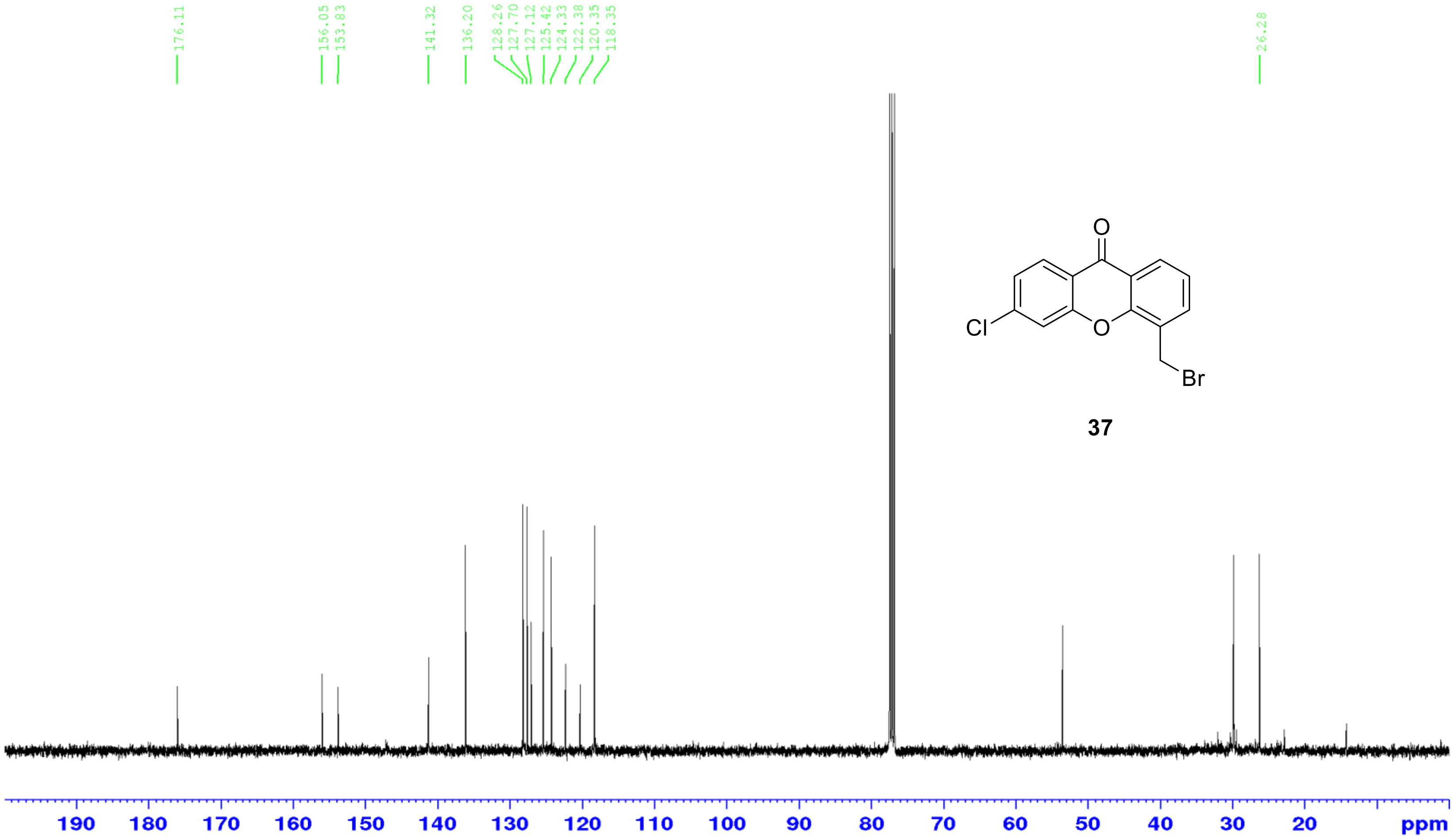
36

^{1}H NMR (400 MHz, $CDCl_3$)



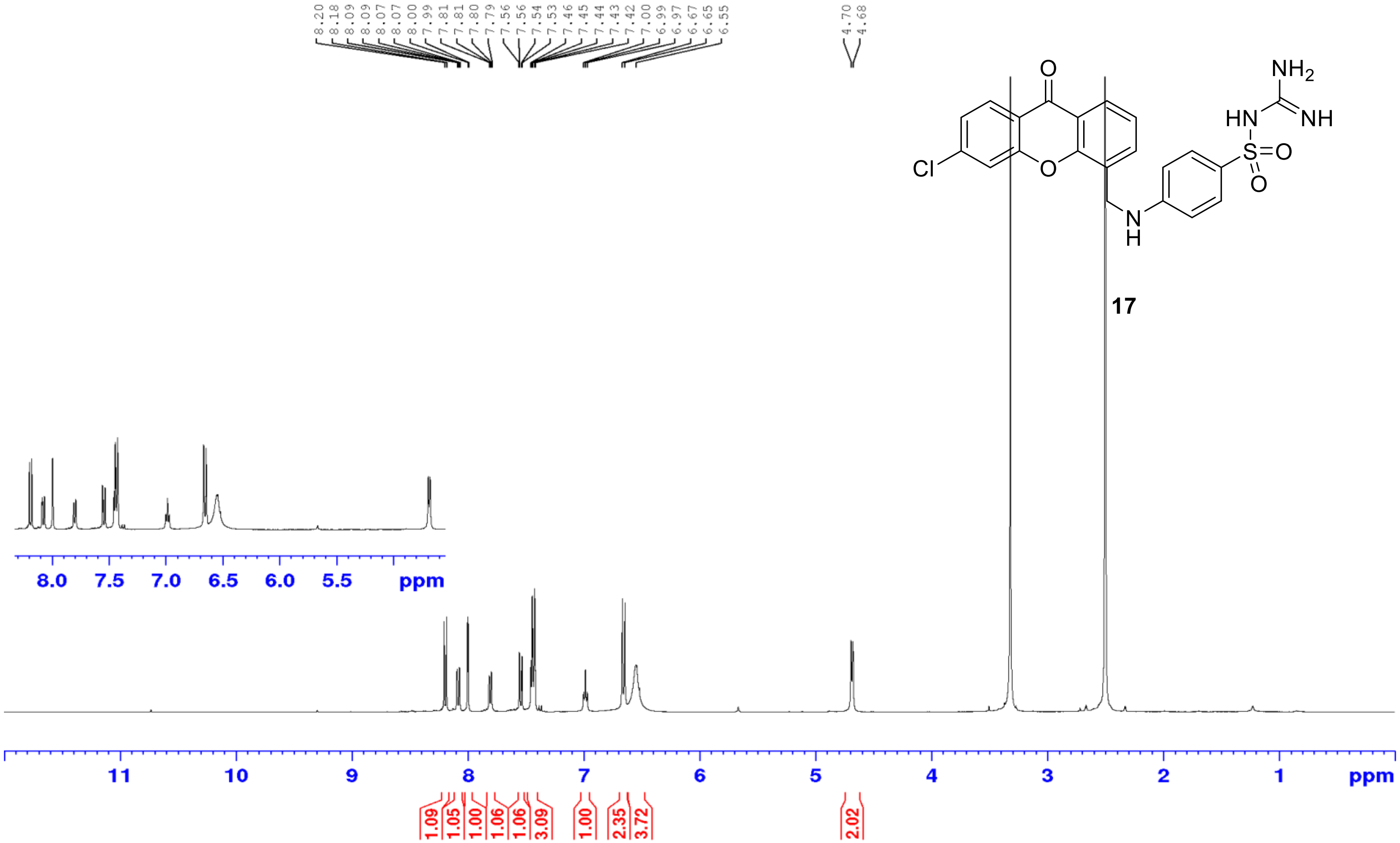
37

^{13}C NMR (101 MHz, $CDCl_3$)



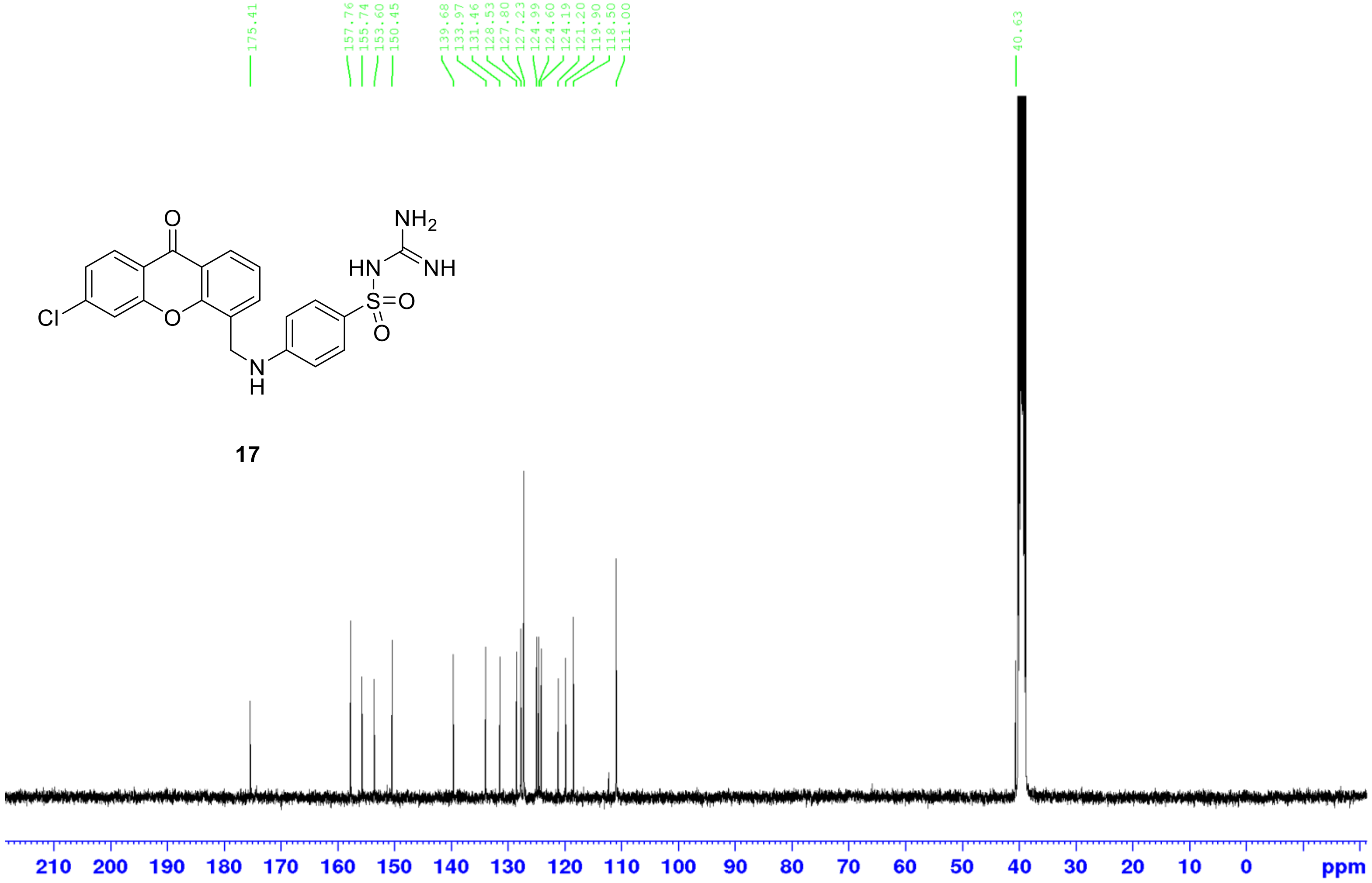
37

^{1}H NMR (400 MHz, DMSO-d$_6$)

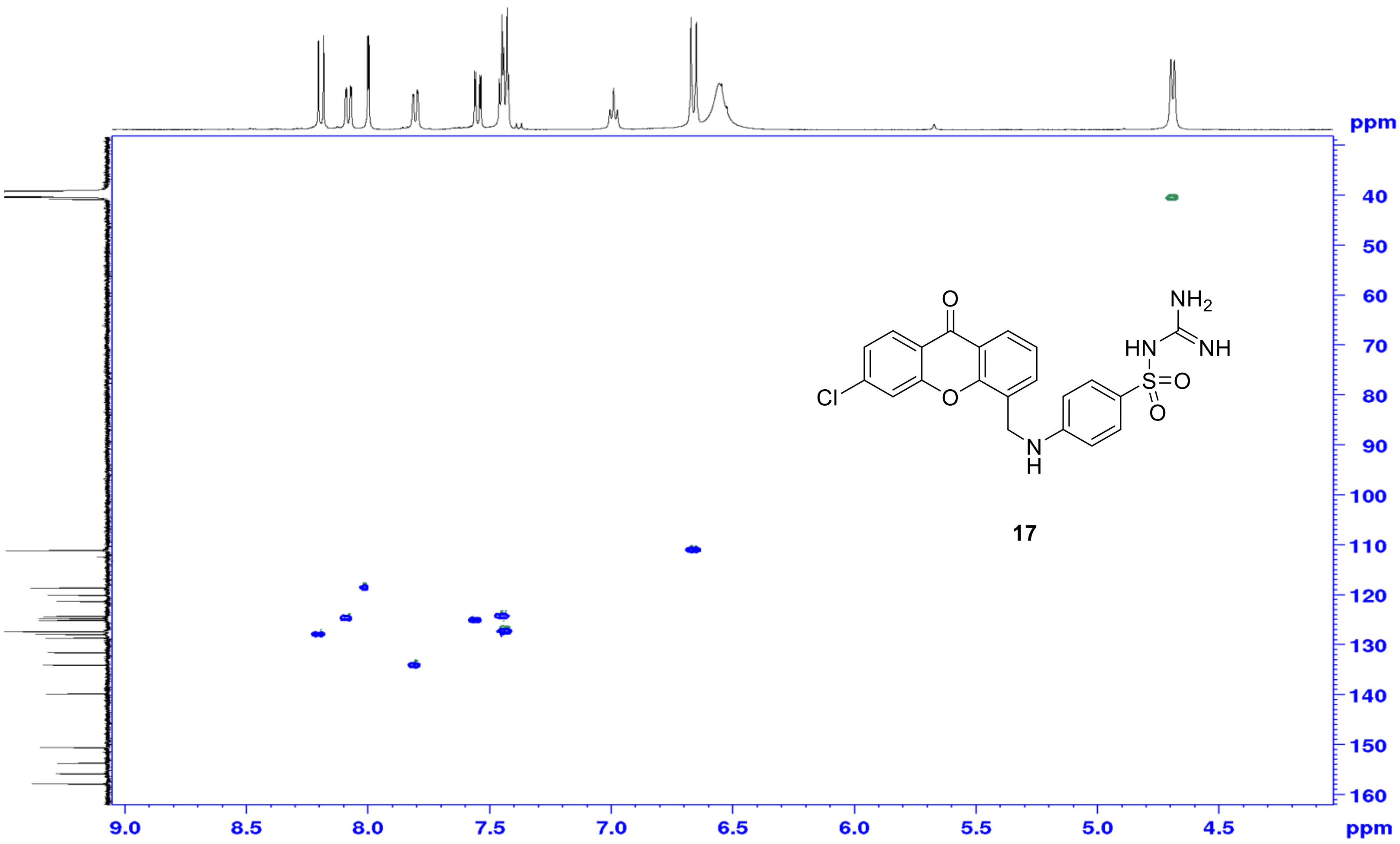


17

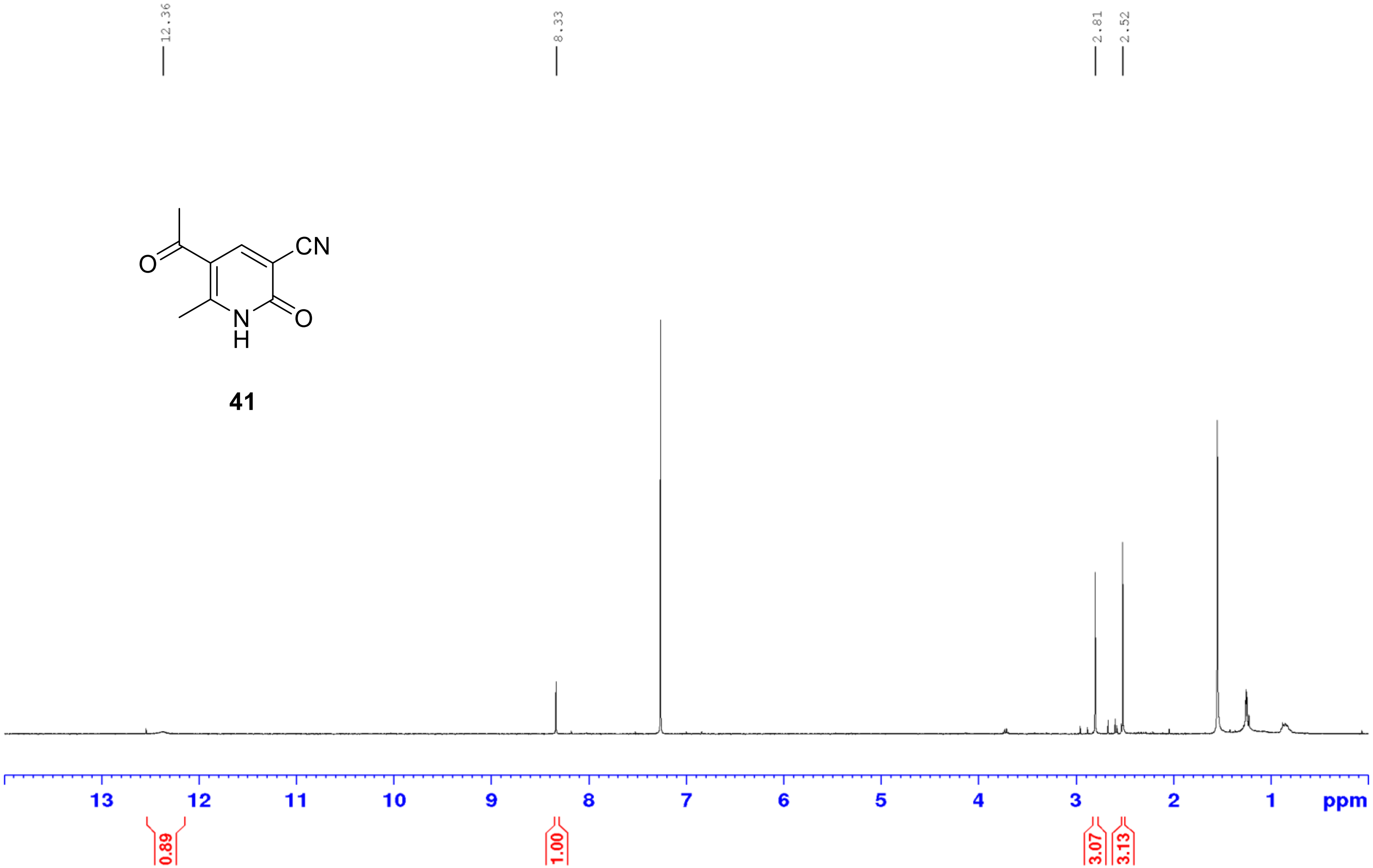
^{13}C NMR (101 MHz DMSO-d_6)



HSQC-NMR, DMSO

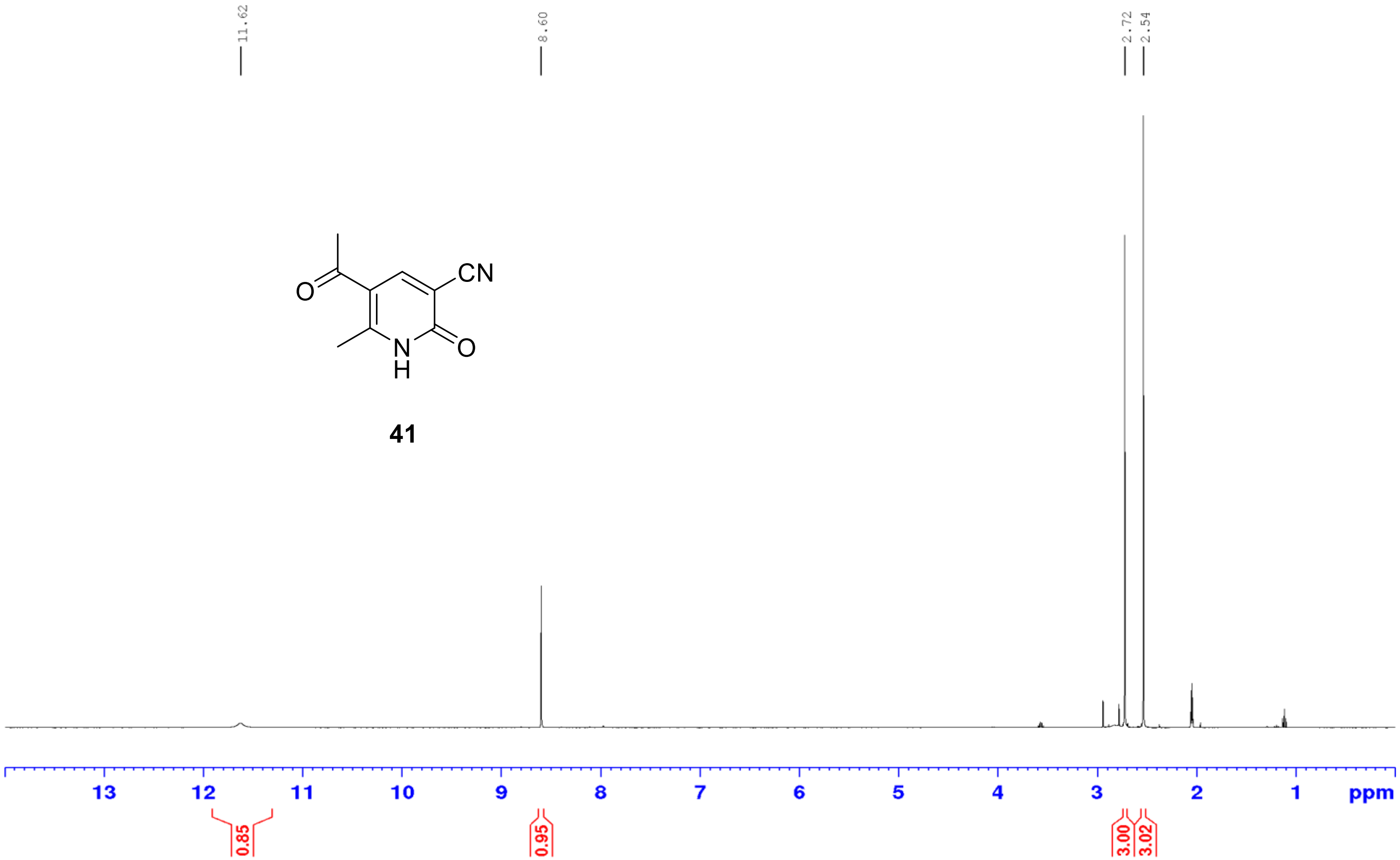


^{1}H NMR (400 MHz, $CDCl_3$)



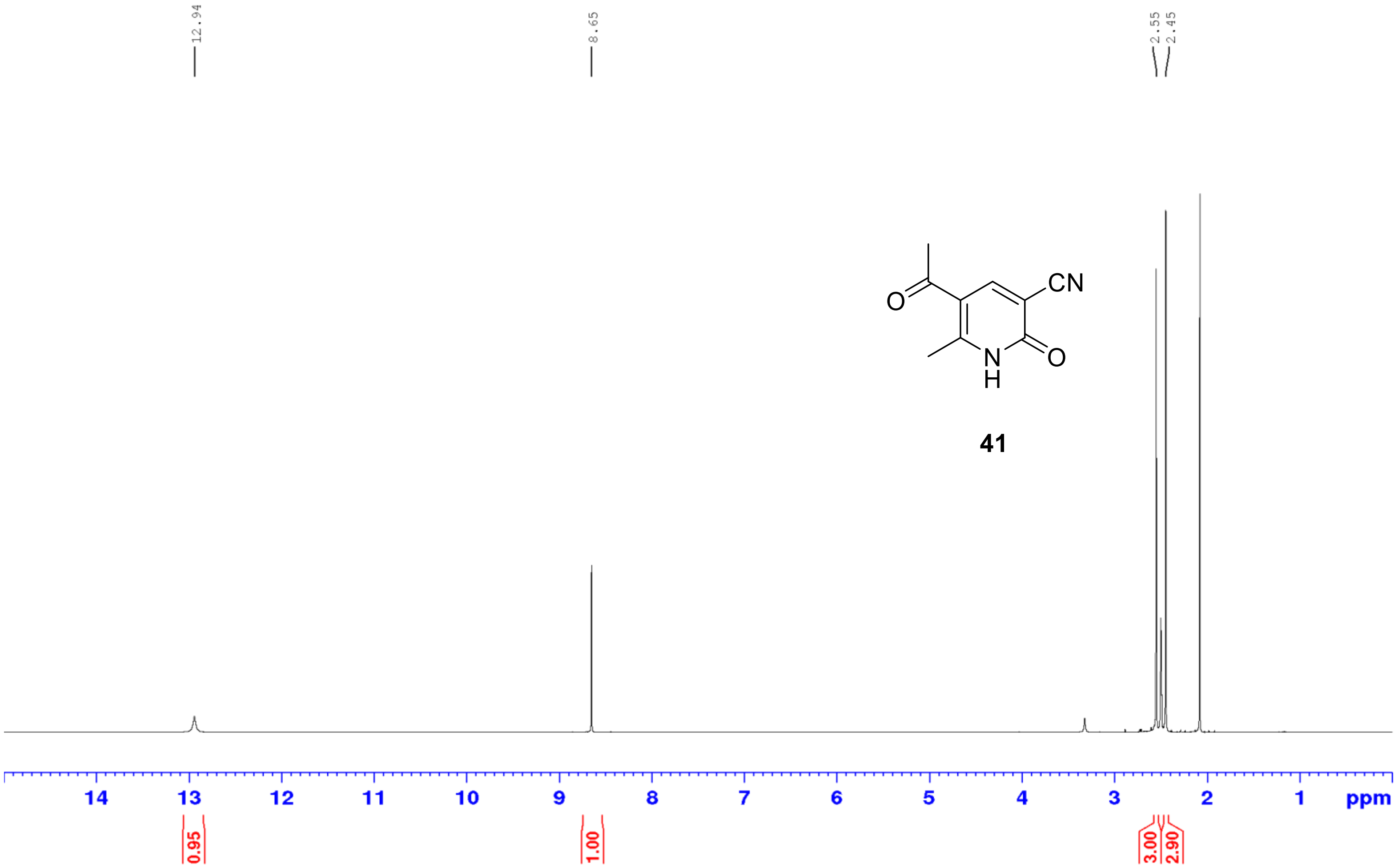
41

^{1}H NMR (400 MHz, Acetone-d$_6$)

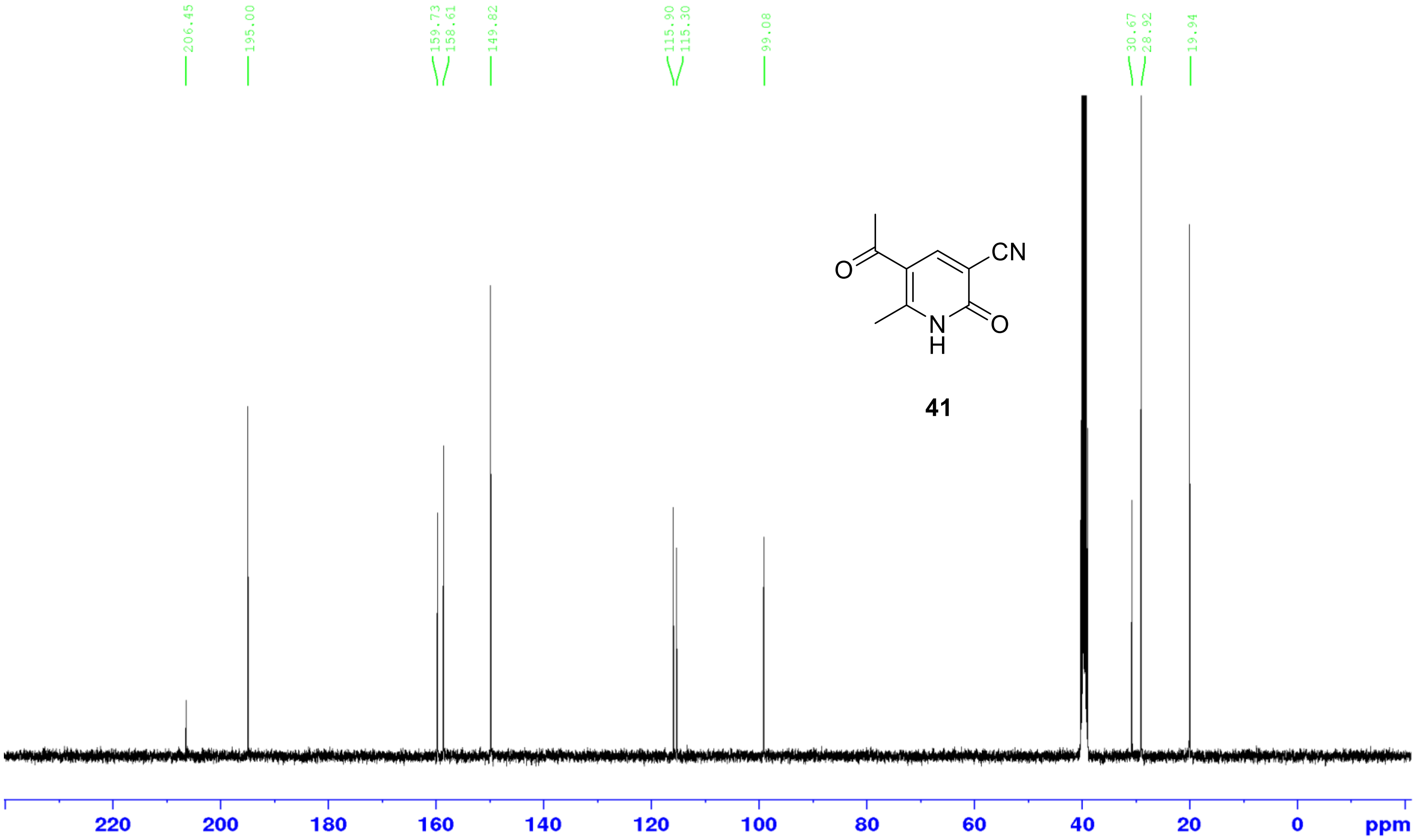


41

^1H NMR (400 MHz, DMSO-d_6)

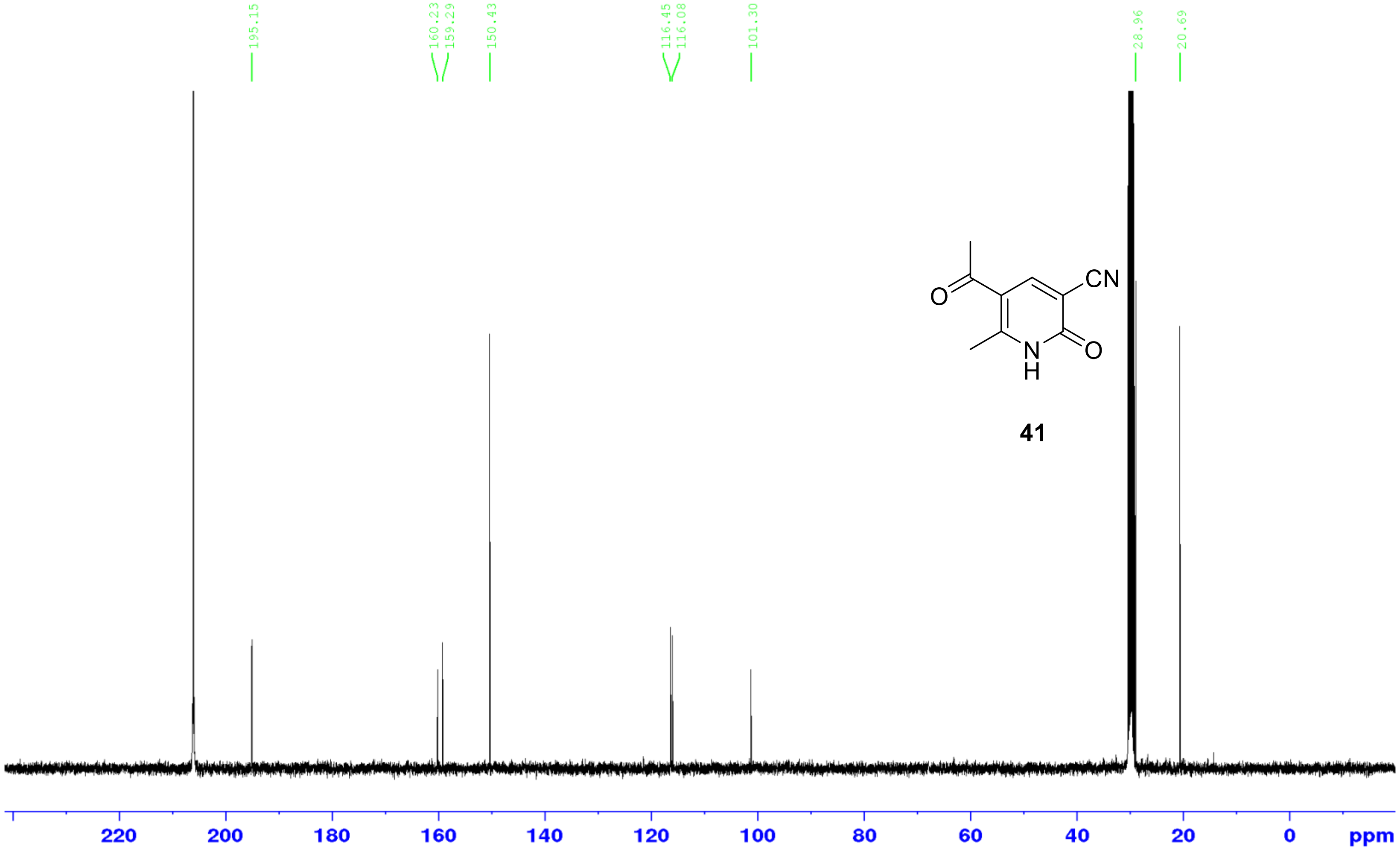


^{13}C NMR (101 MHz, DMSO-d_6)



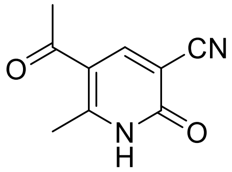
41

^{13}C NMR (101 MHz, Acetone-d_6)

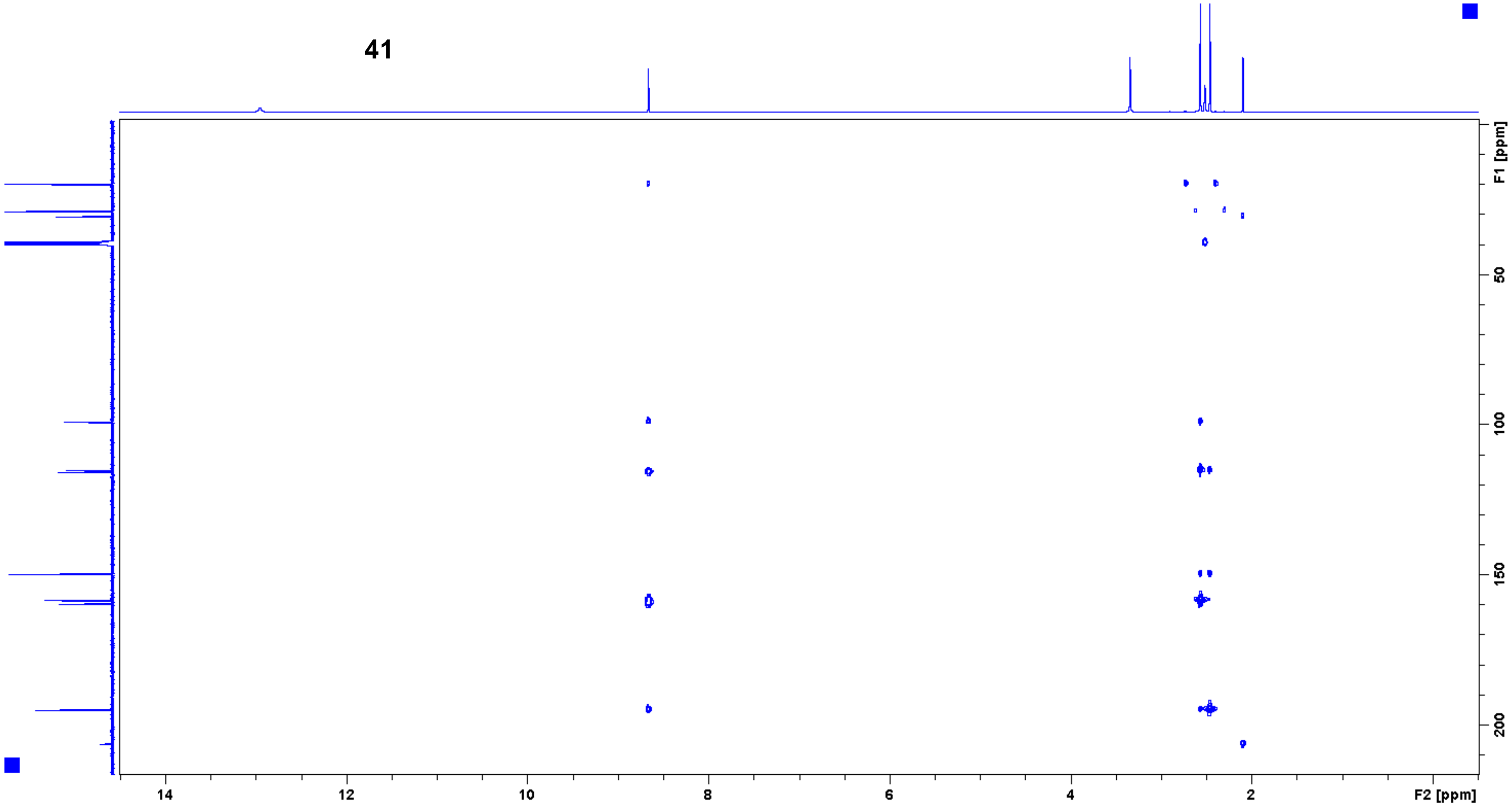


41

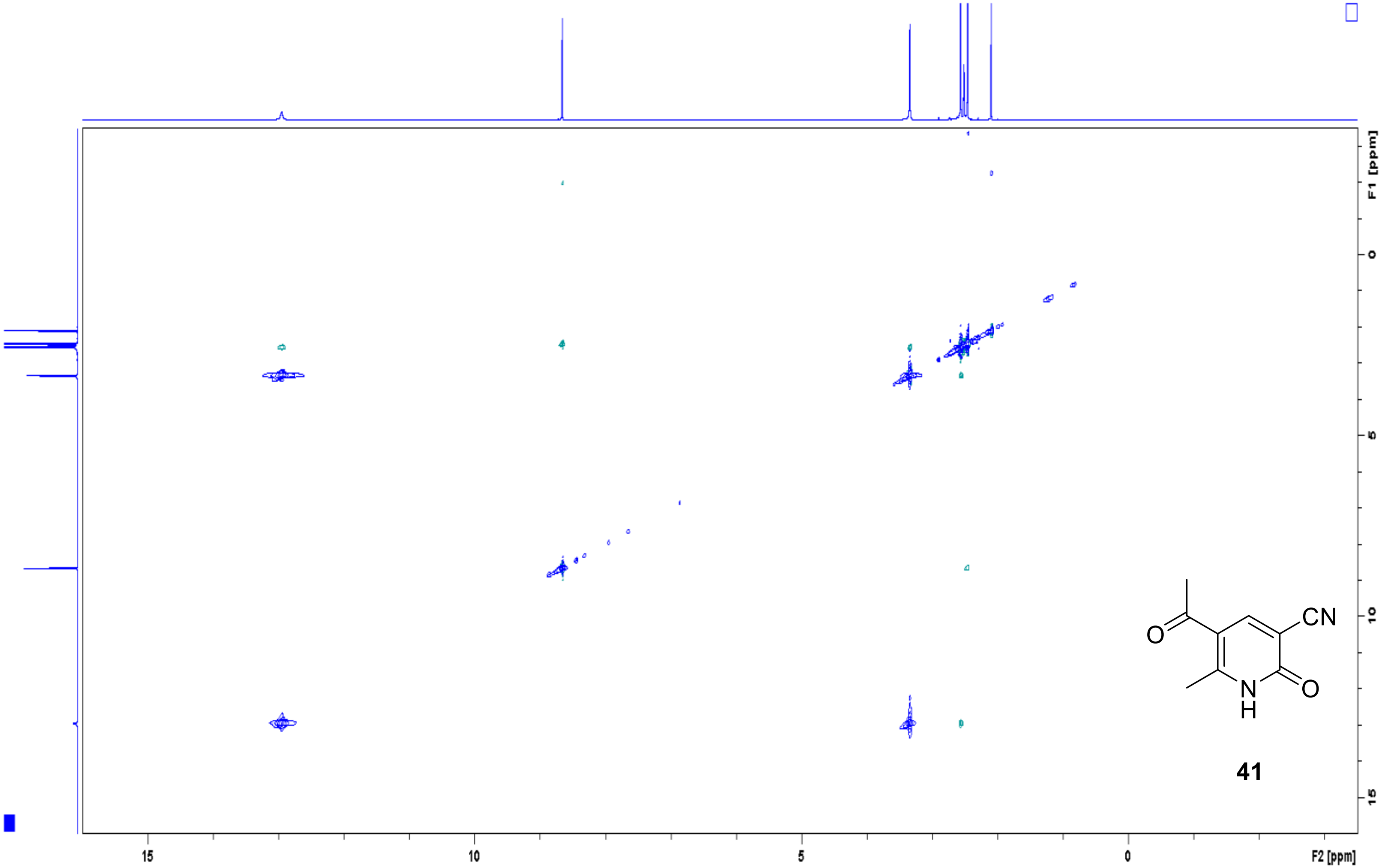
HMBC NMR (DMSO-d_6)



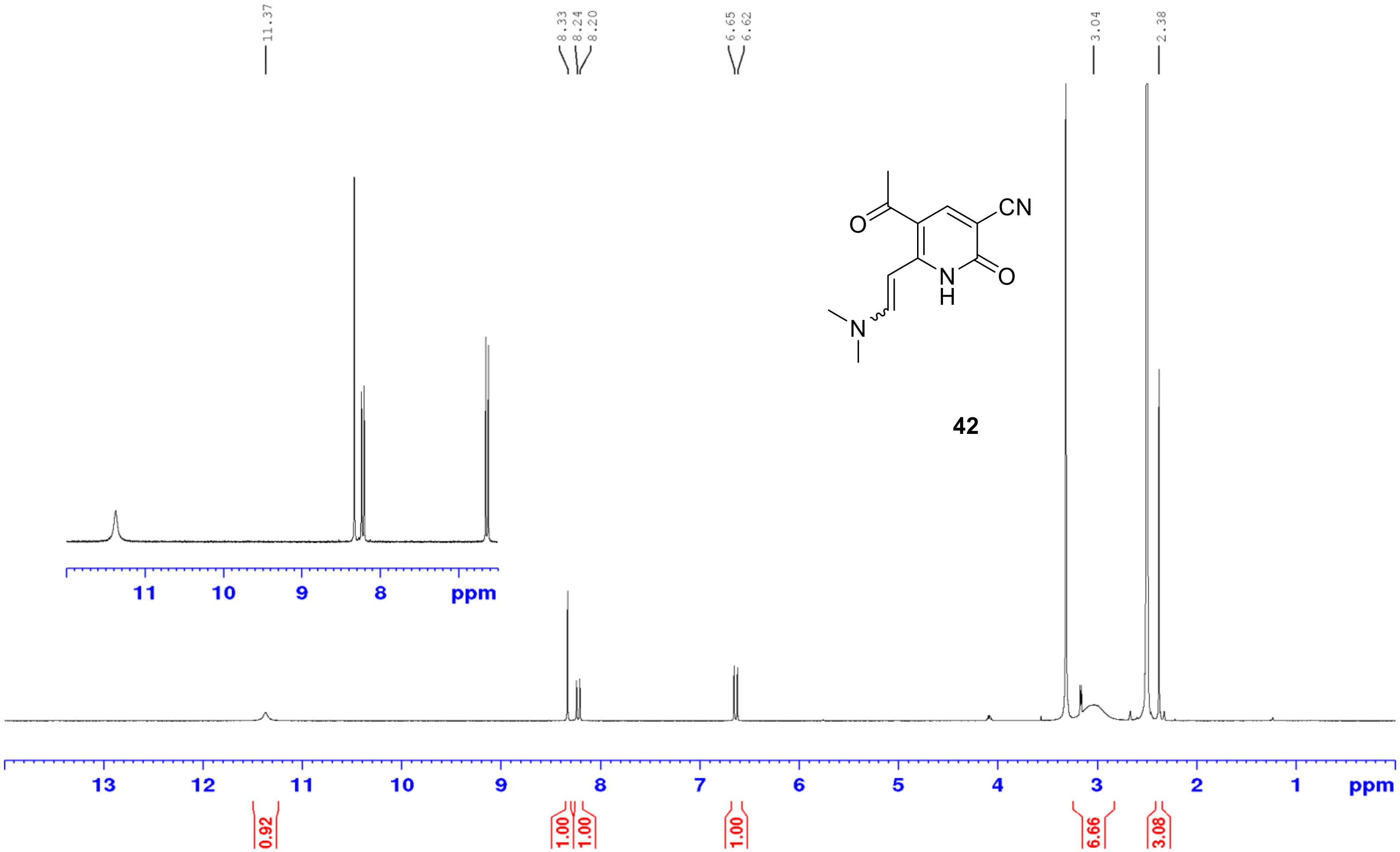
41



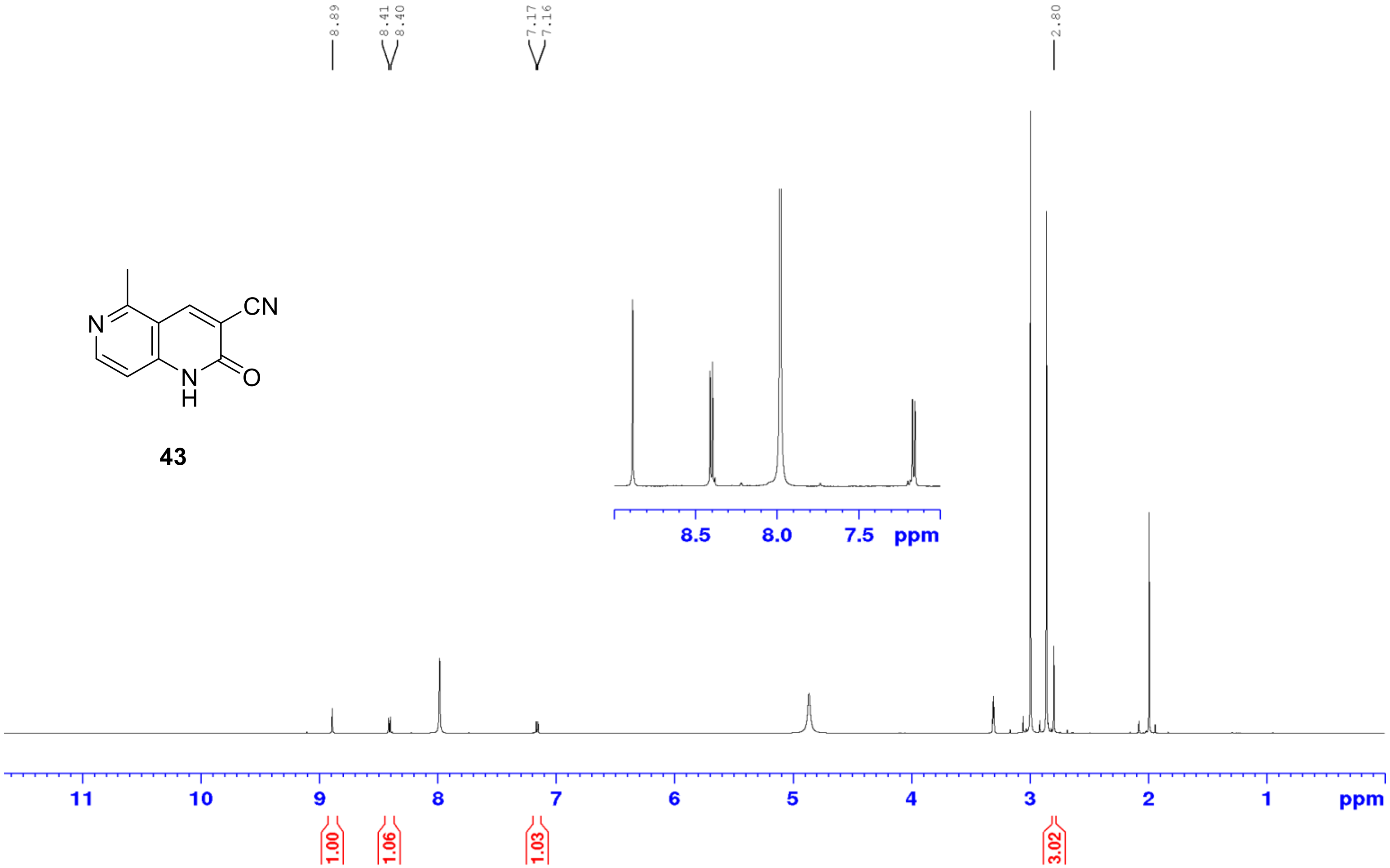
NOESY NMR (400 MHz, DMSO-d_6)



^{1}H NMR (400 MHz, DMSO-d_6)

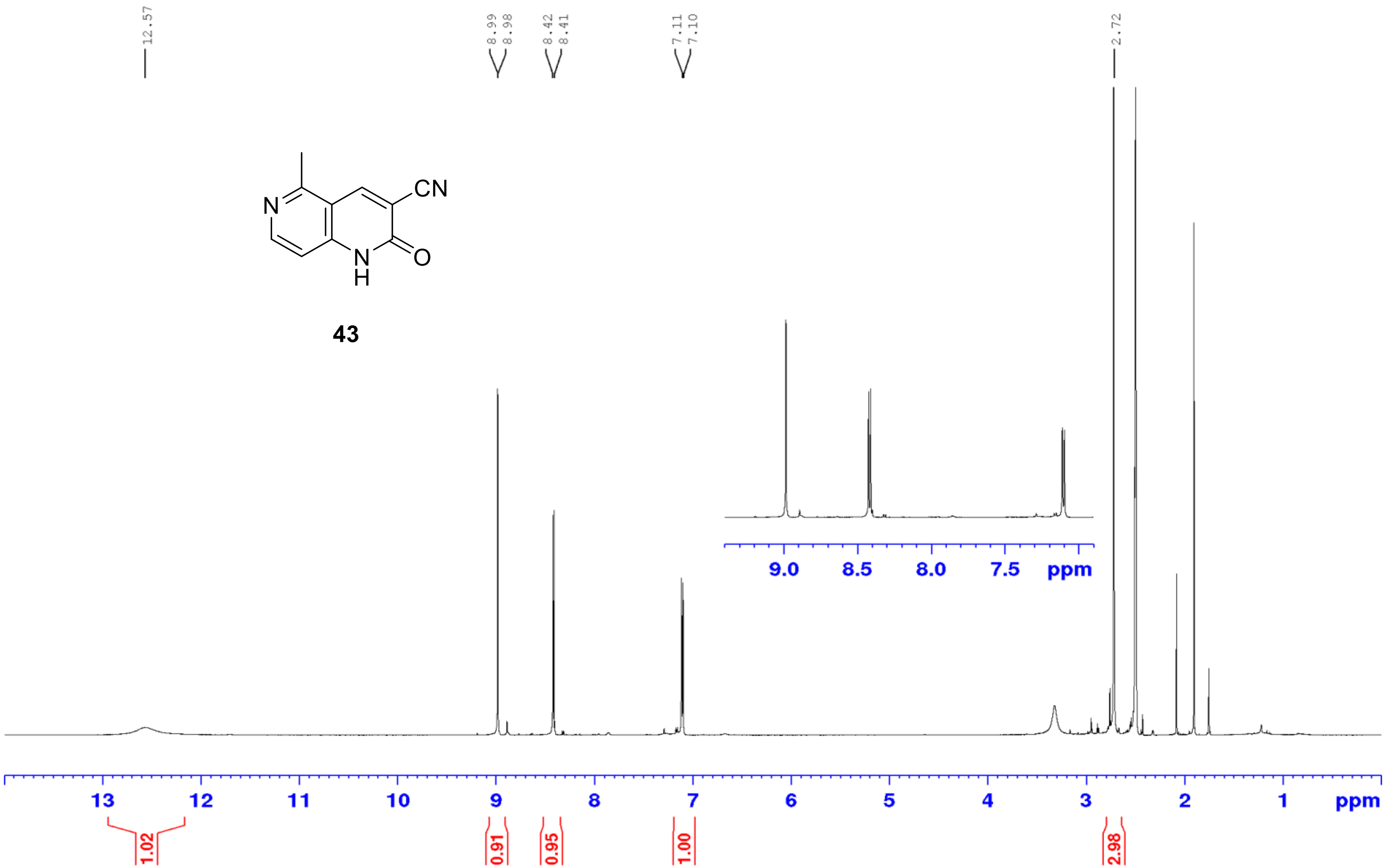


^{1}H NMR (400 MHz,MeOD)

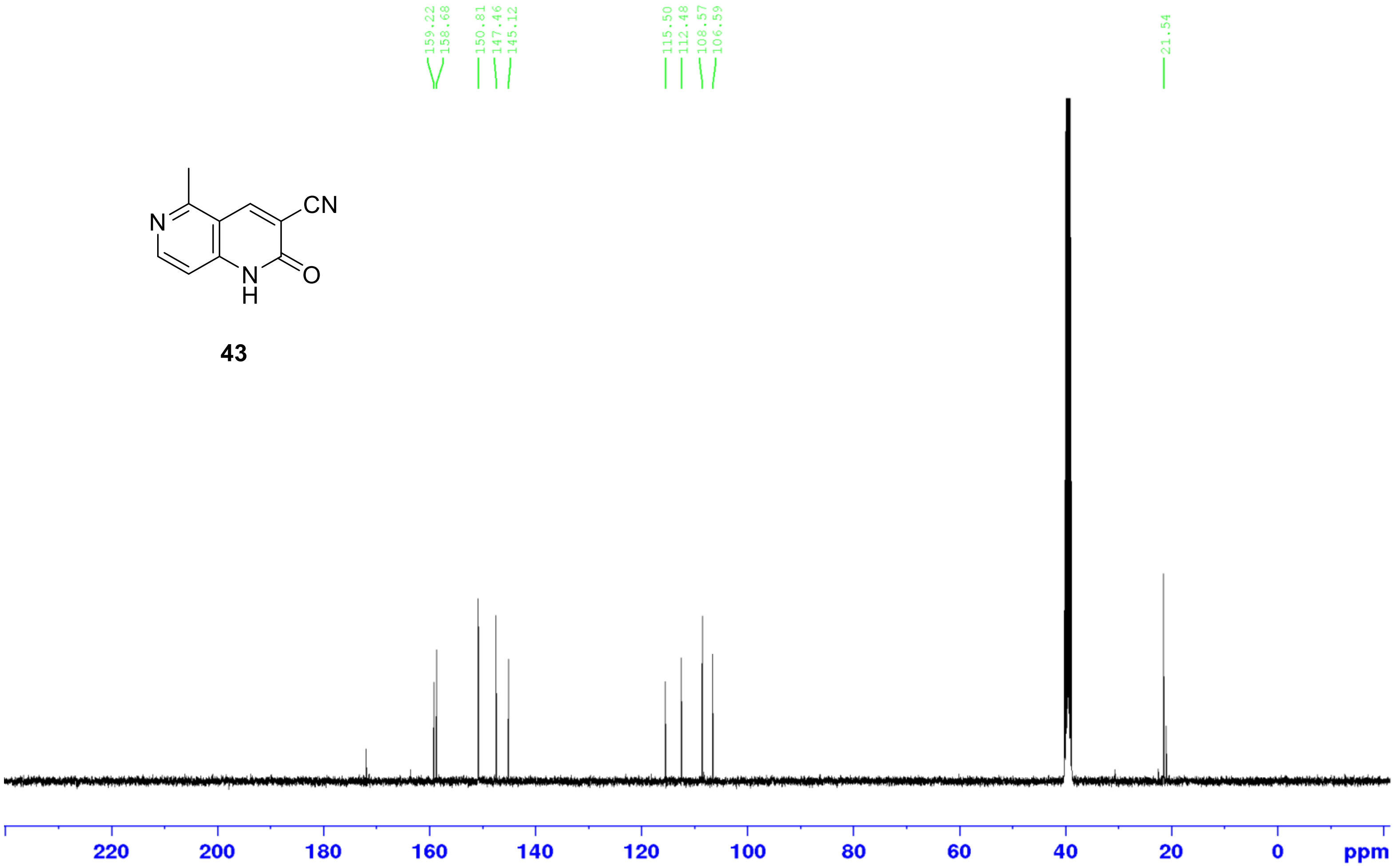


43

^{1}H NMR (400 MHz, DMSO-d_6)

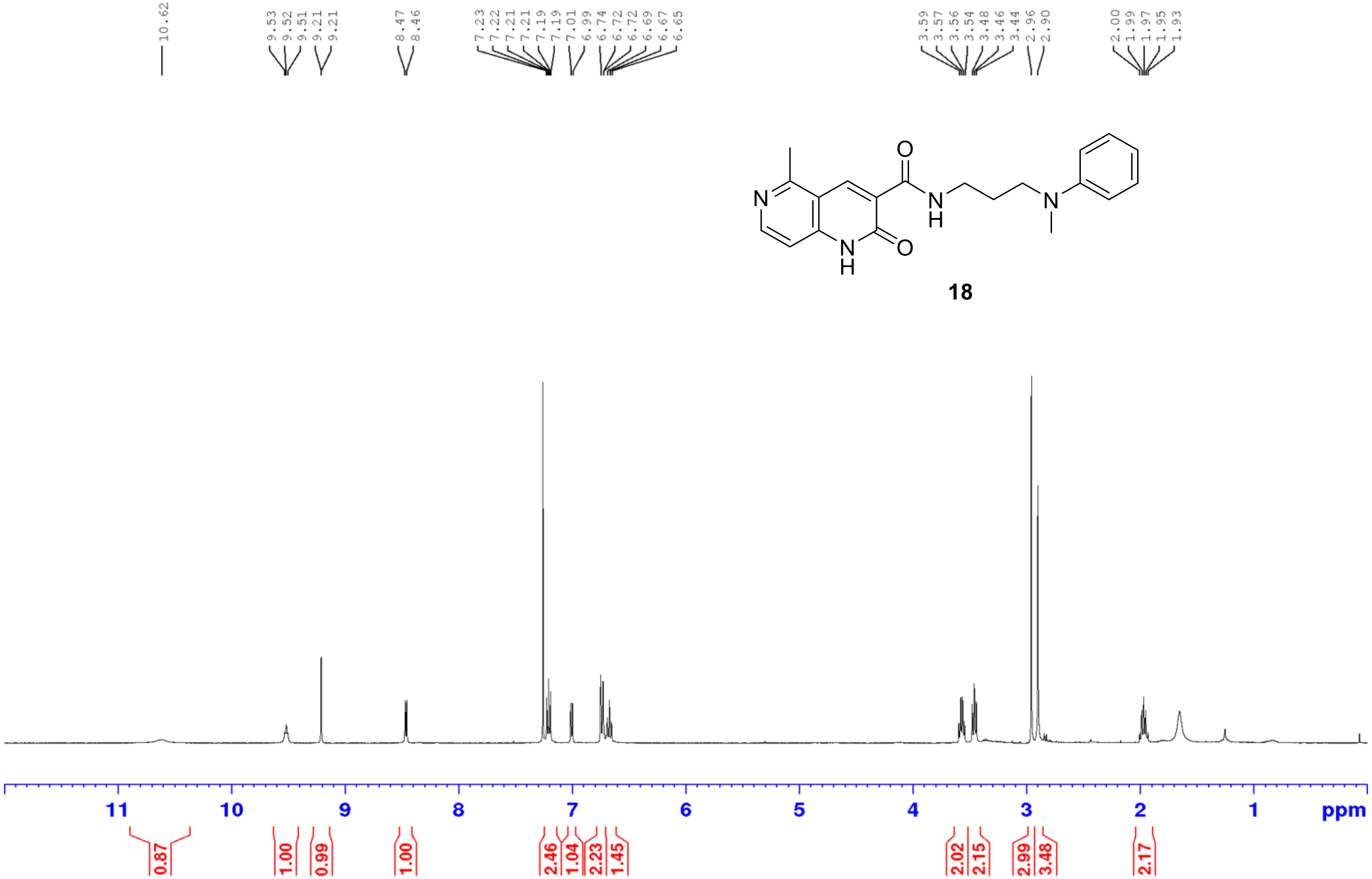


^{13}C NMR (101 MHz, DMSO-d_6)



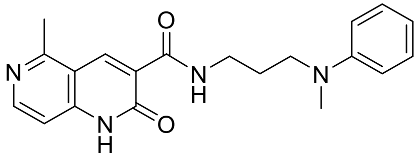
43

^{1}H NMR (400 MHz, $CDCl_3$)

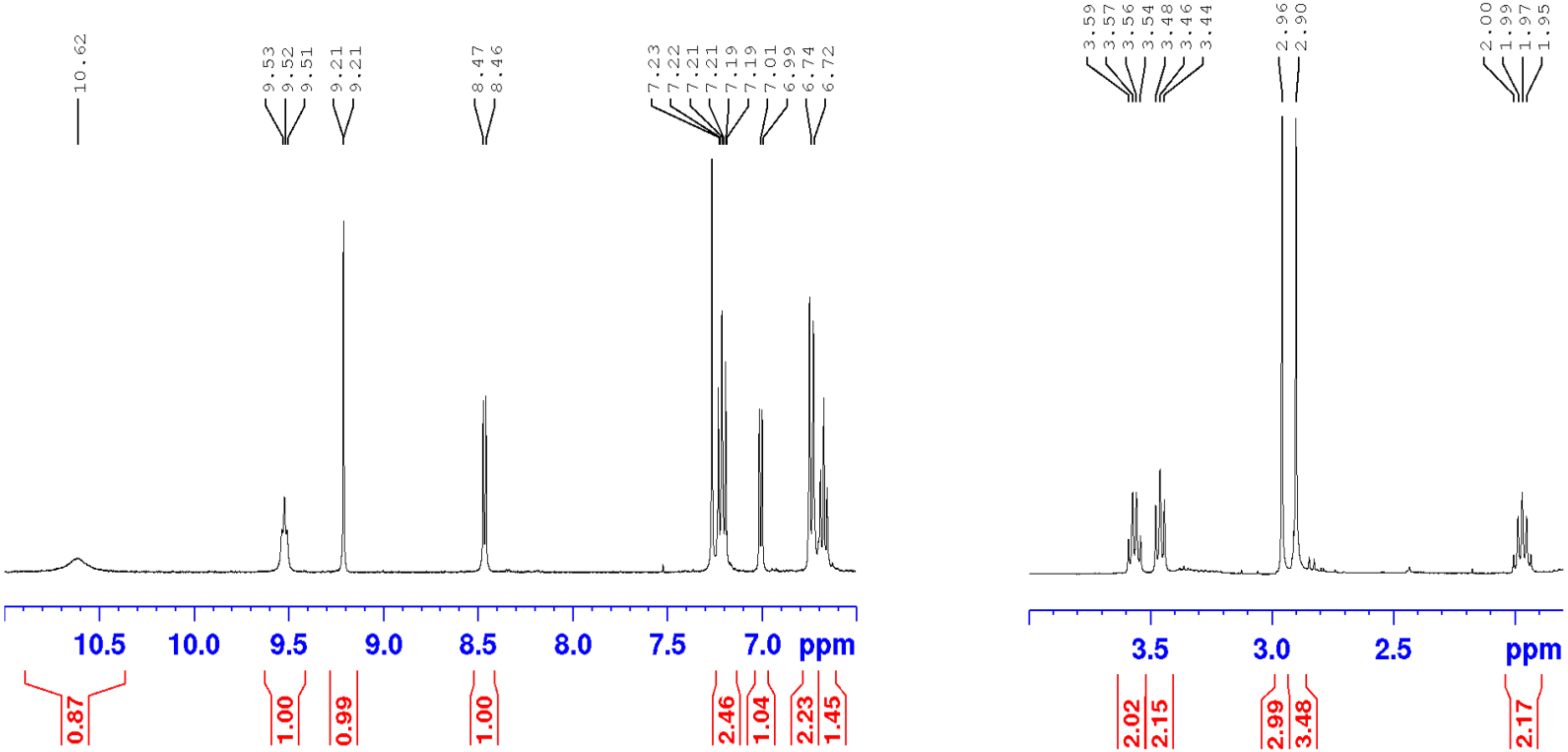


18

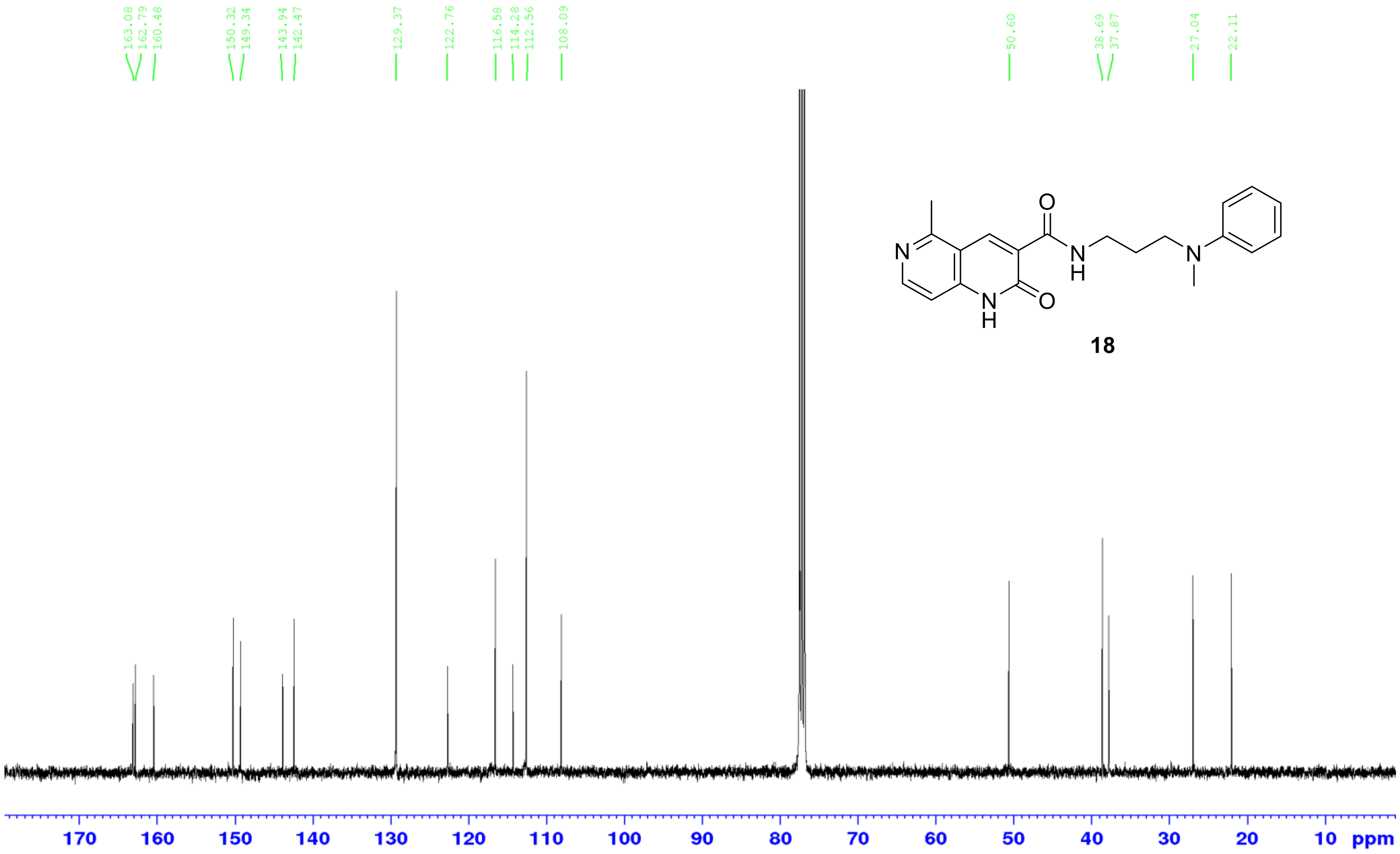
^{1}H NMR (400 MHz, CDCl$_3$)



18

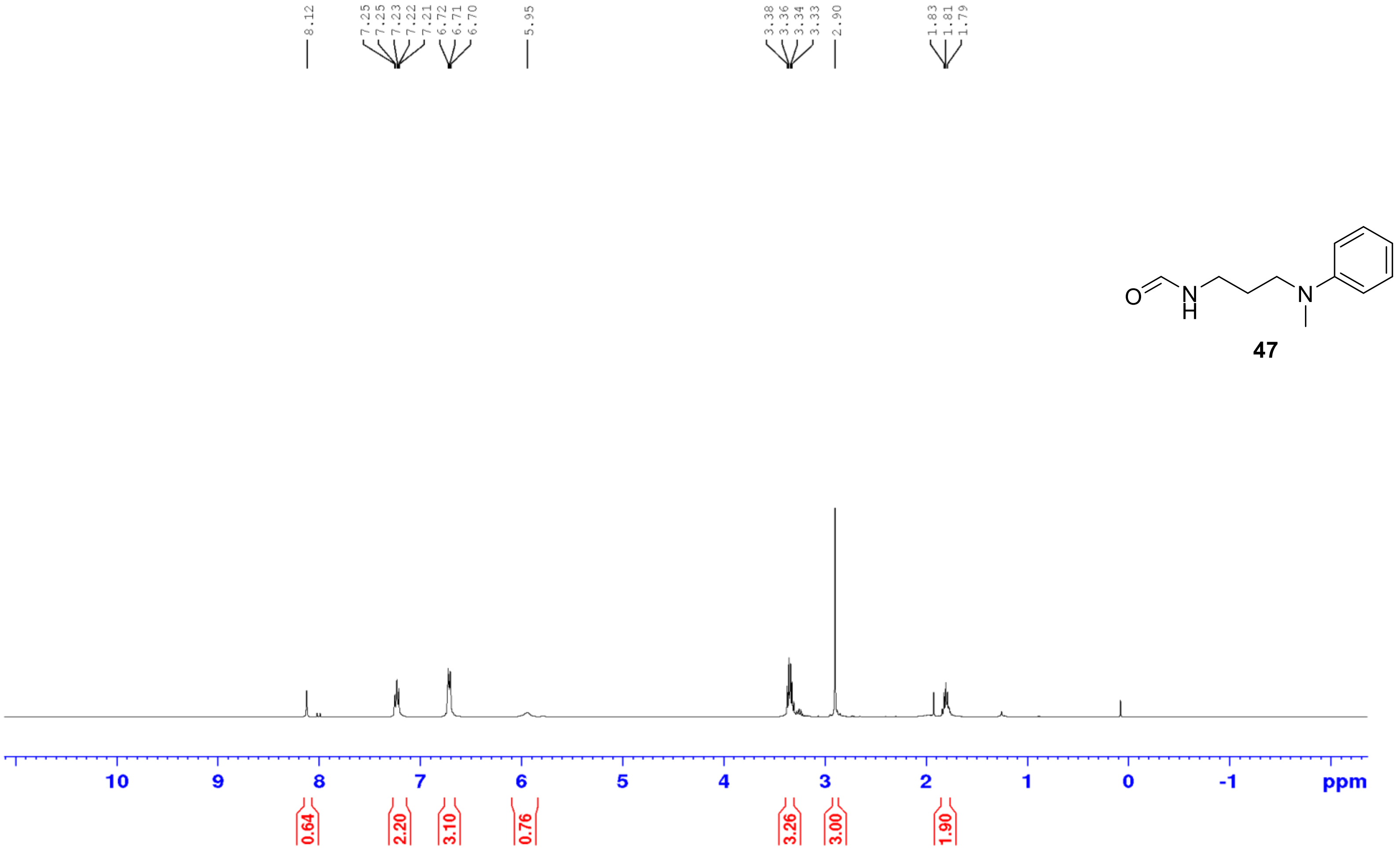


^{13}C NMR (101 MHz, $CDCl_3$)



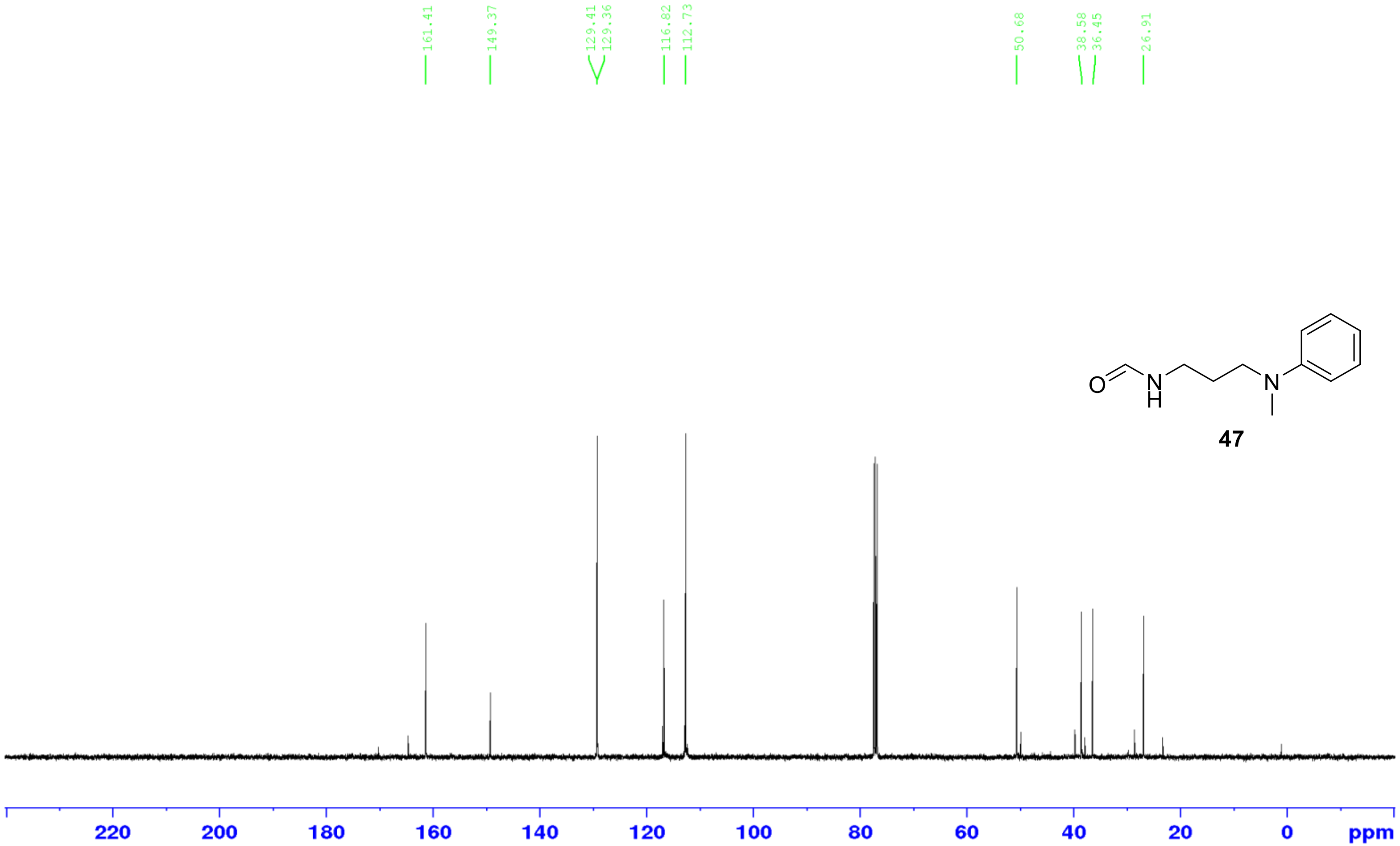
18

^{1}H NMR (400 MHz, CDCl$_3$)

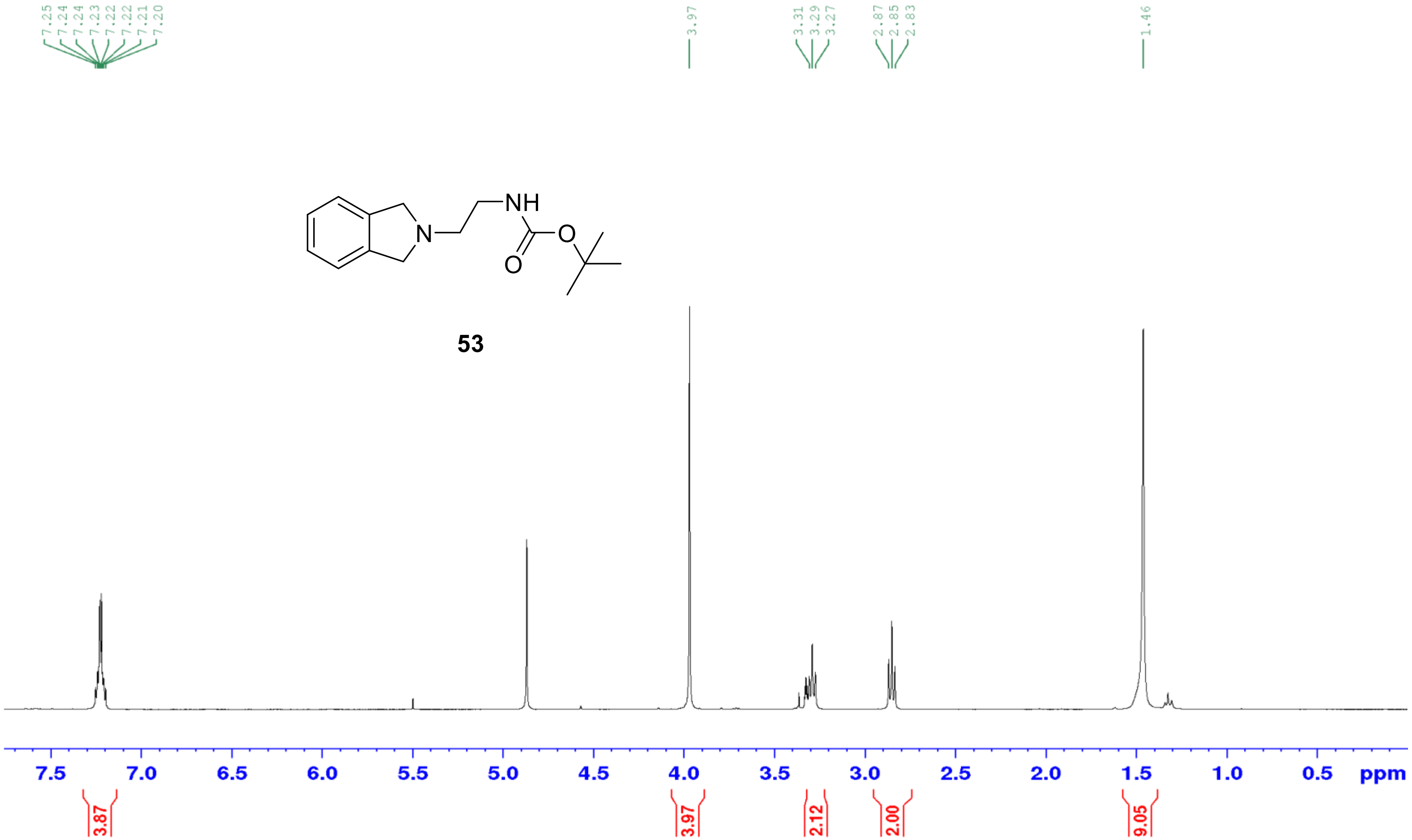


47

^{13}C NMR (101 MHz, $CDCl_3$)

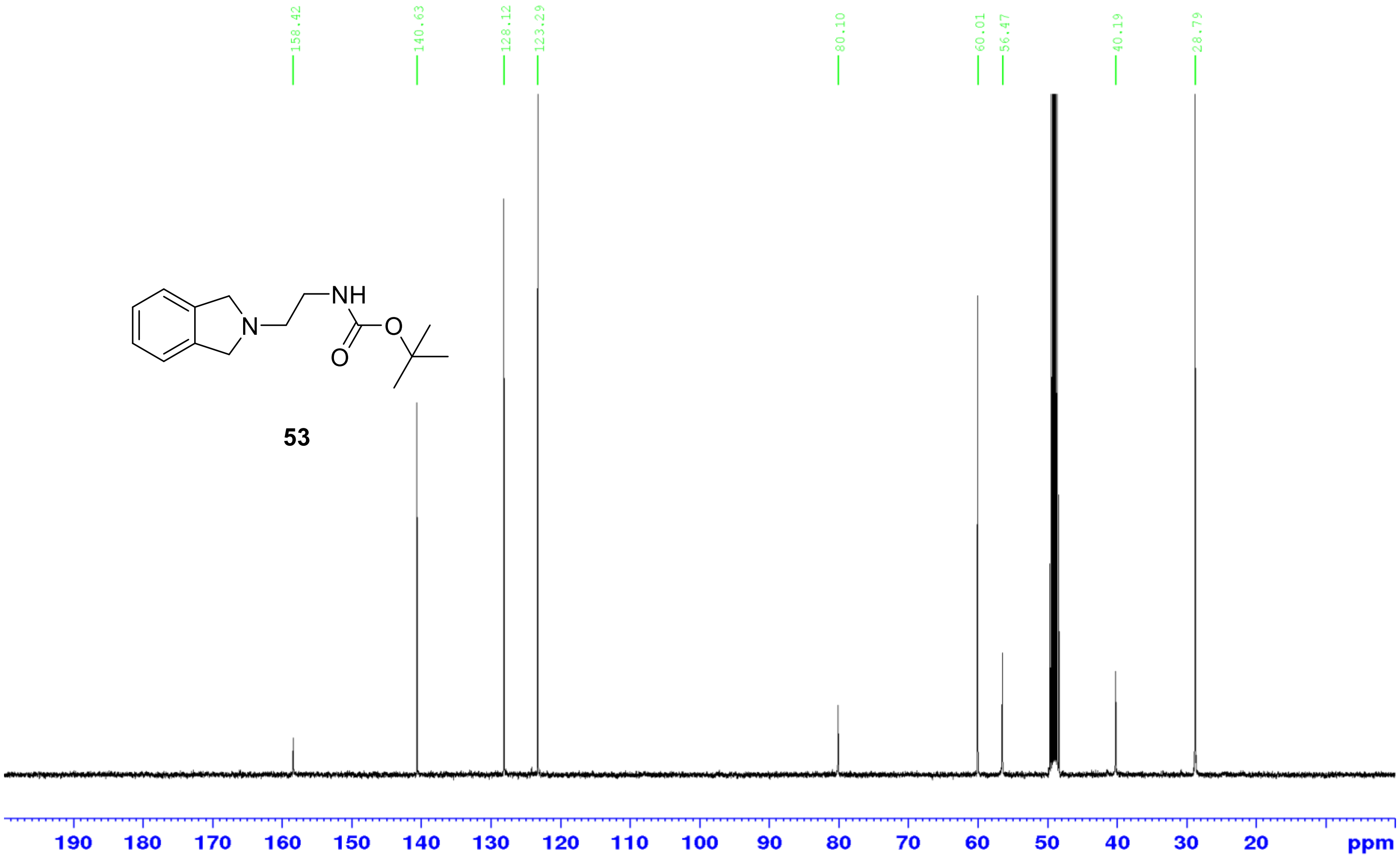


^{1}H NMR (400 MHz,MeOD)



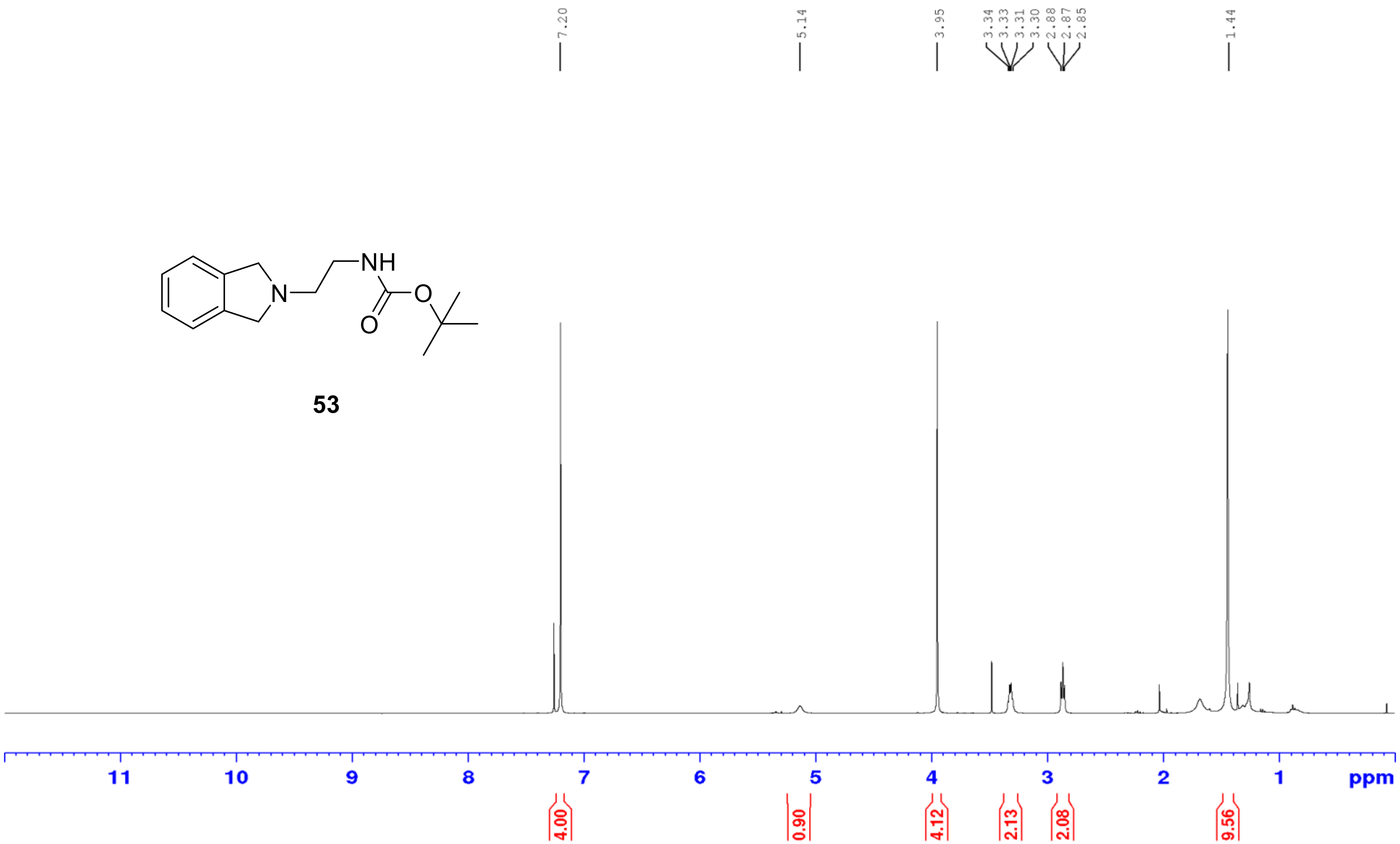
53

^{13}C NMR (101 MHz, MeOD)

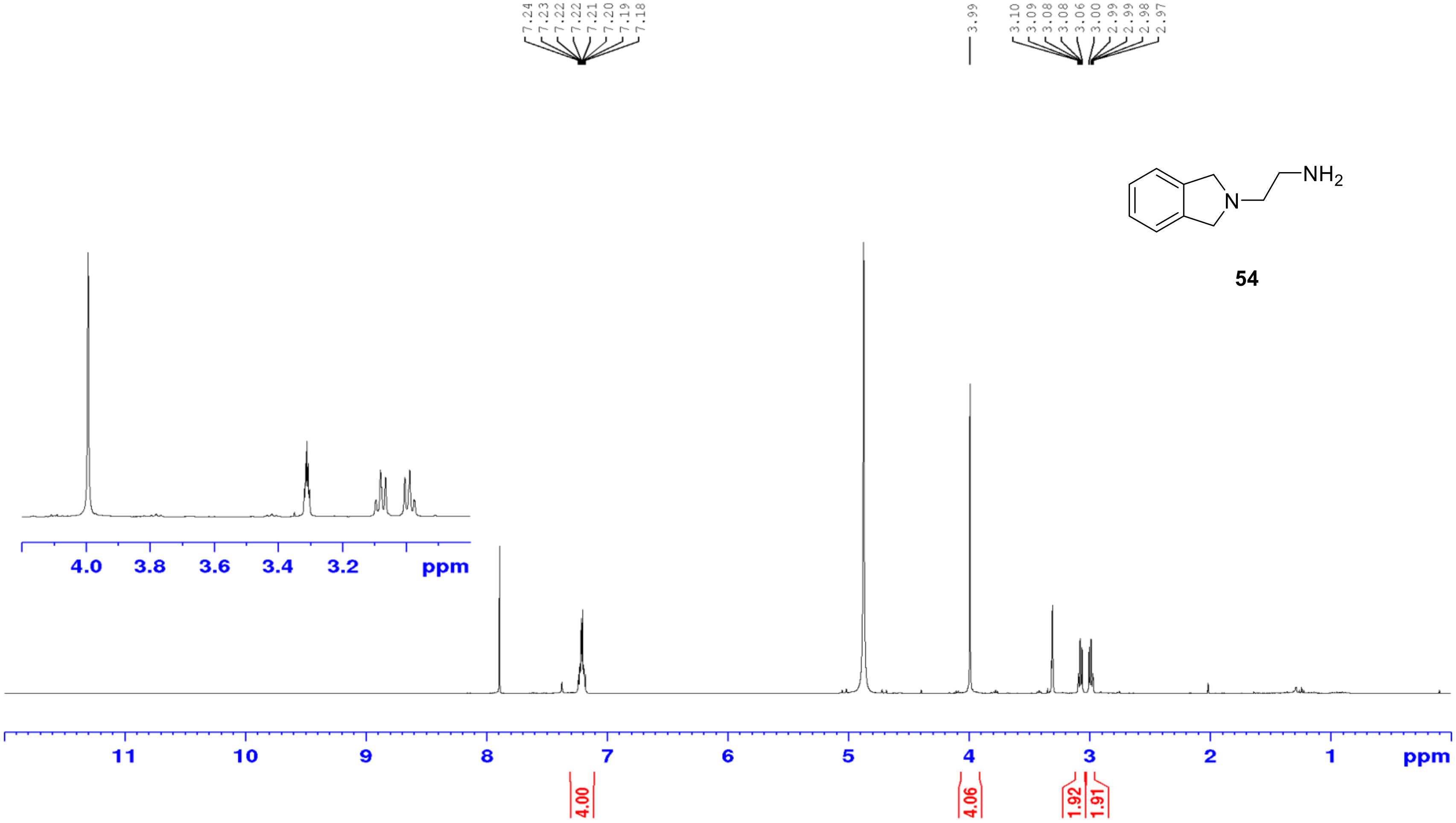


53

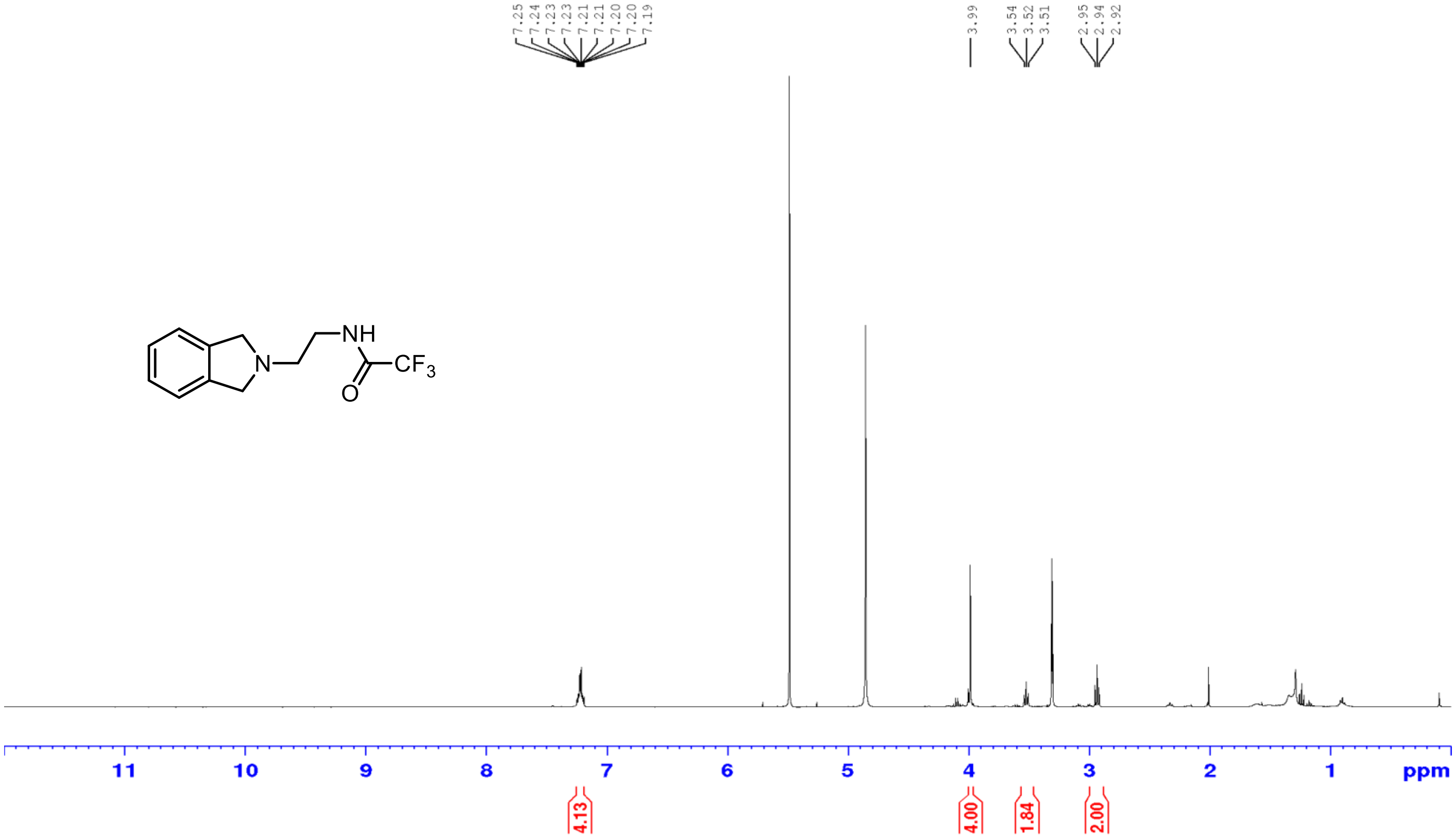
^{1}H NMR (400 MHz, $CDCl_3$)



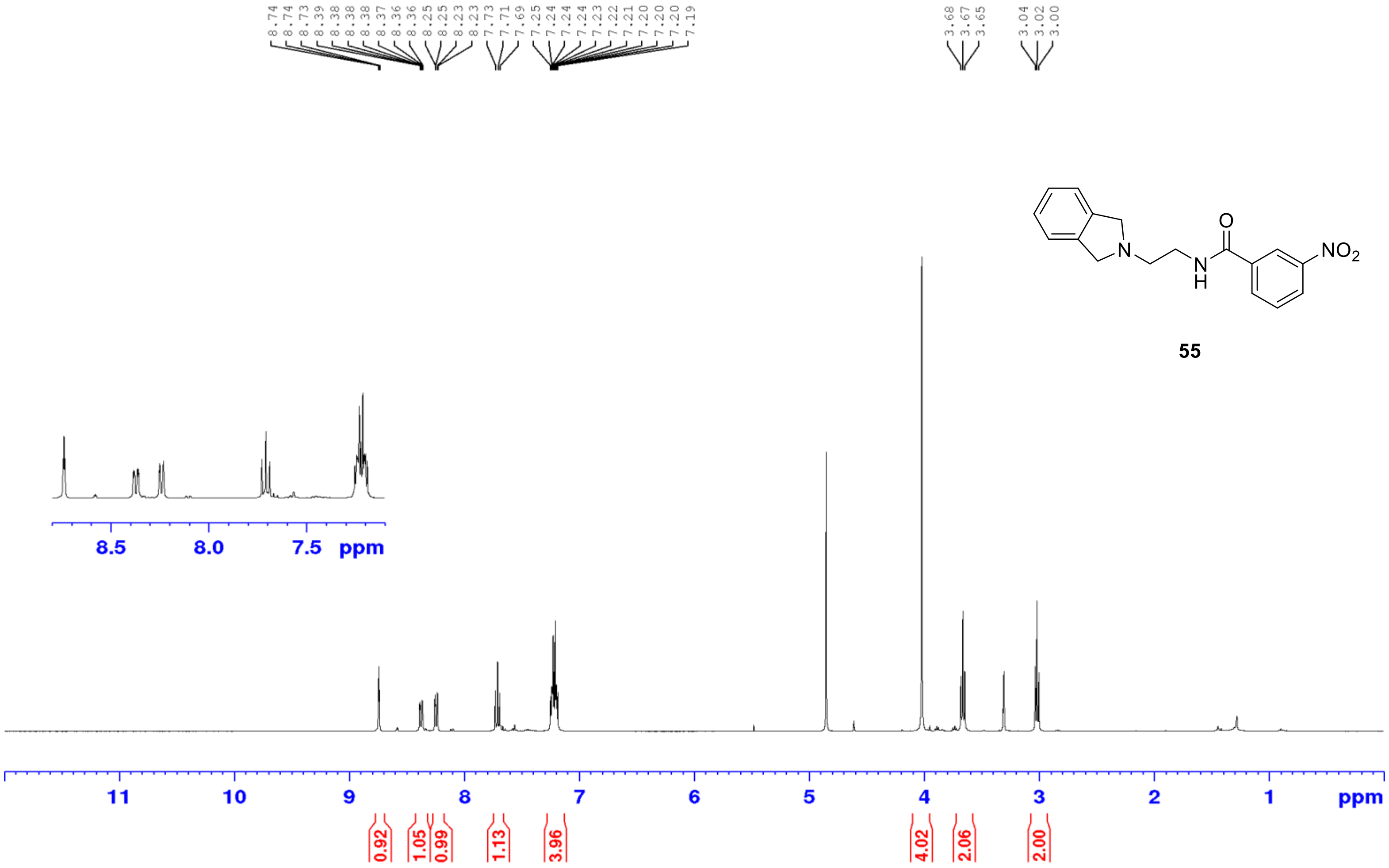
^1H NMR (400 MHz,MeOD)



^{1}H NMR (400 MHz,MeOD)

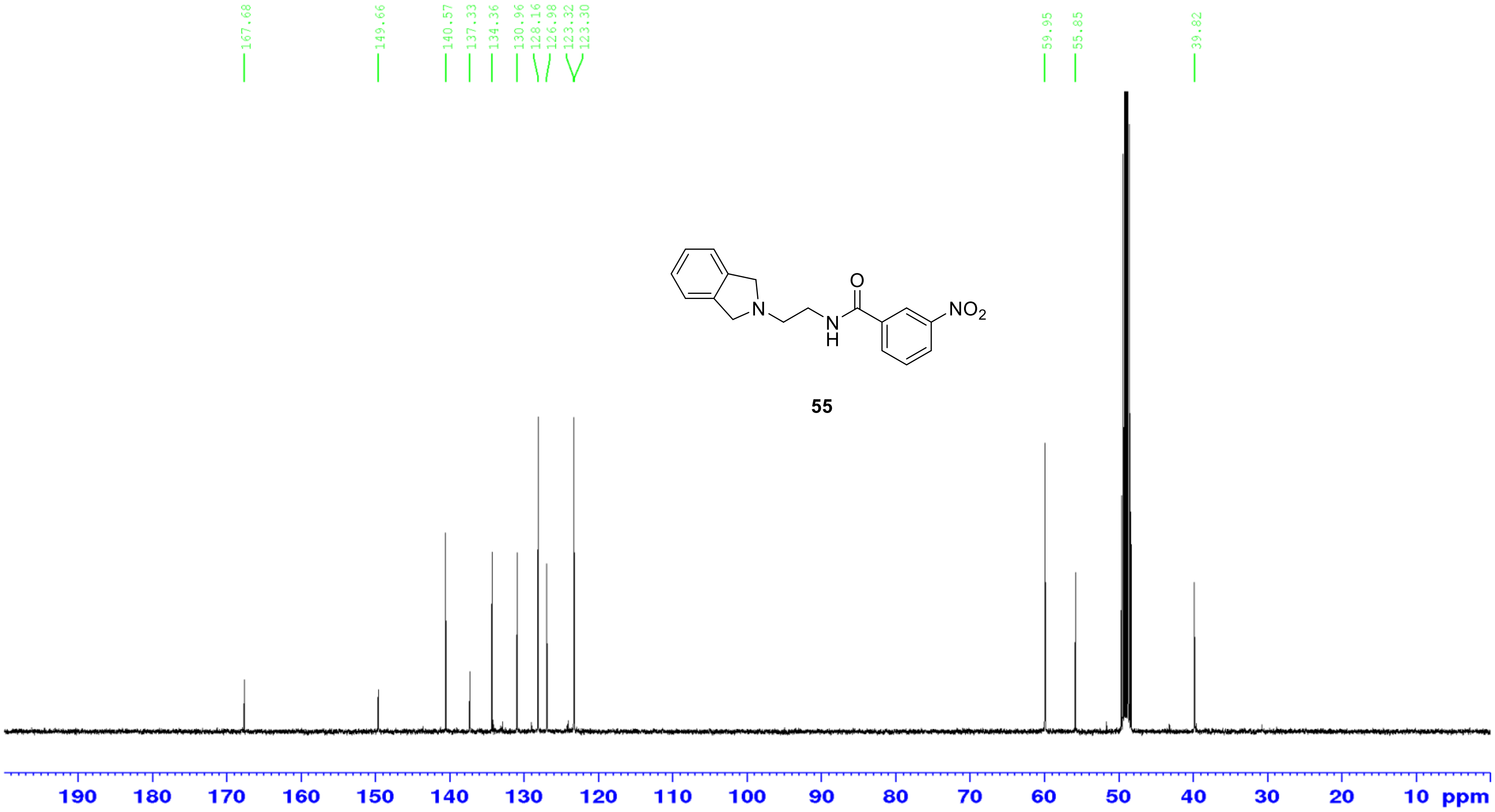


^{1}H NMR (400 MHz, MeOD)



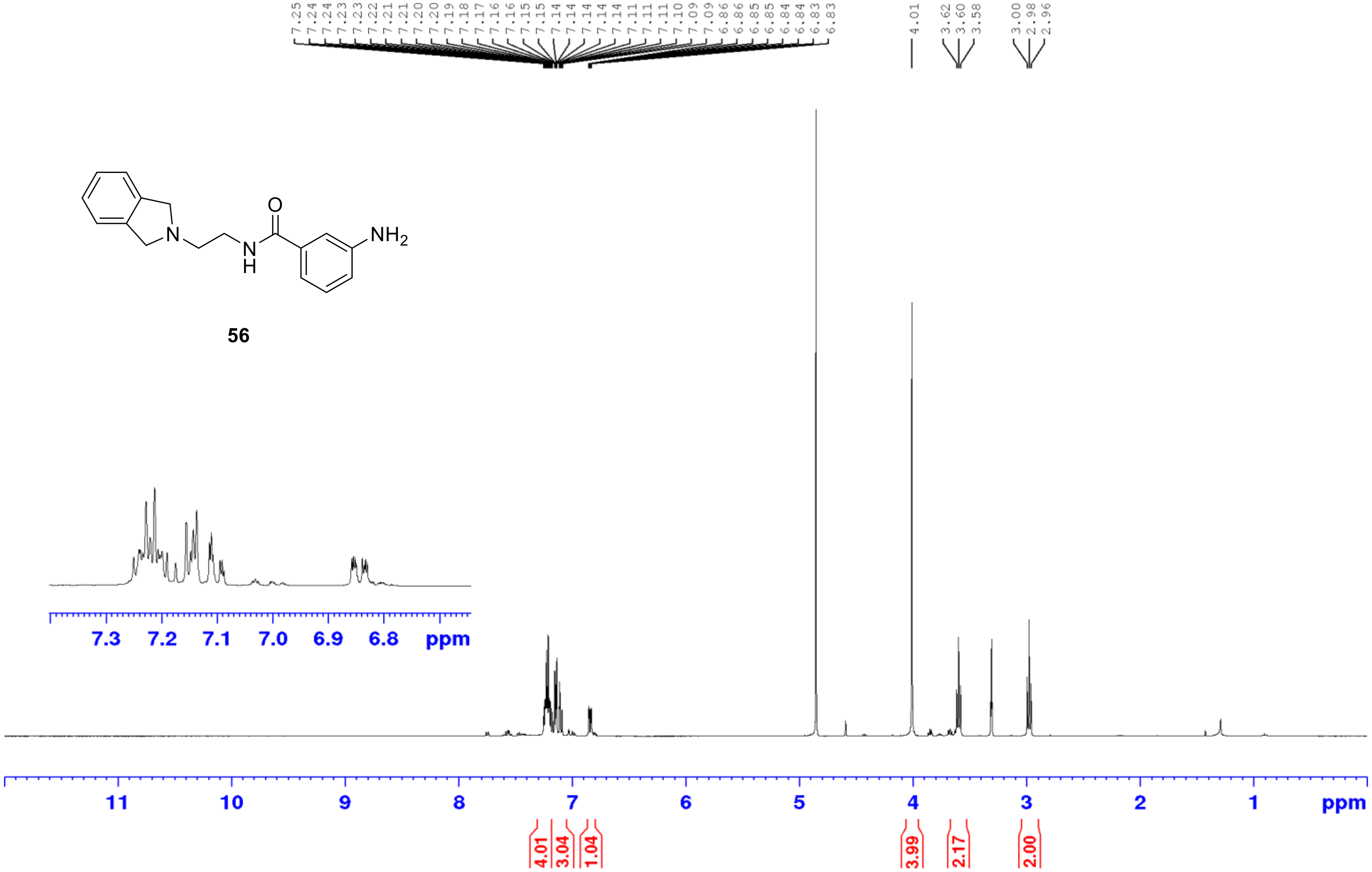
55

^{13}C NMR (101 MHz, MeOD)



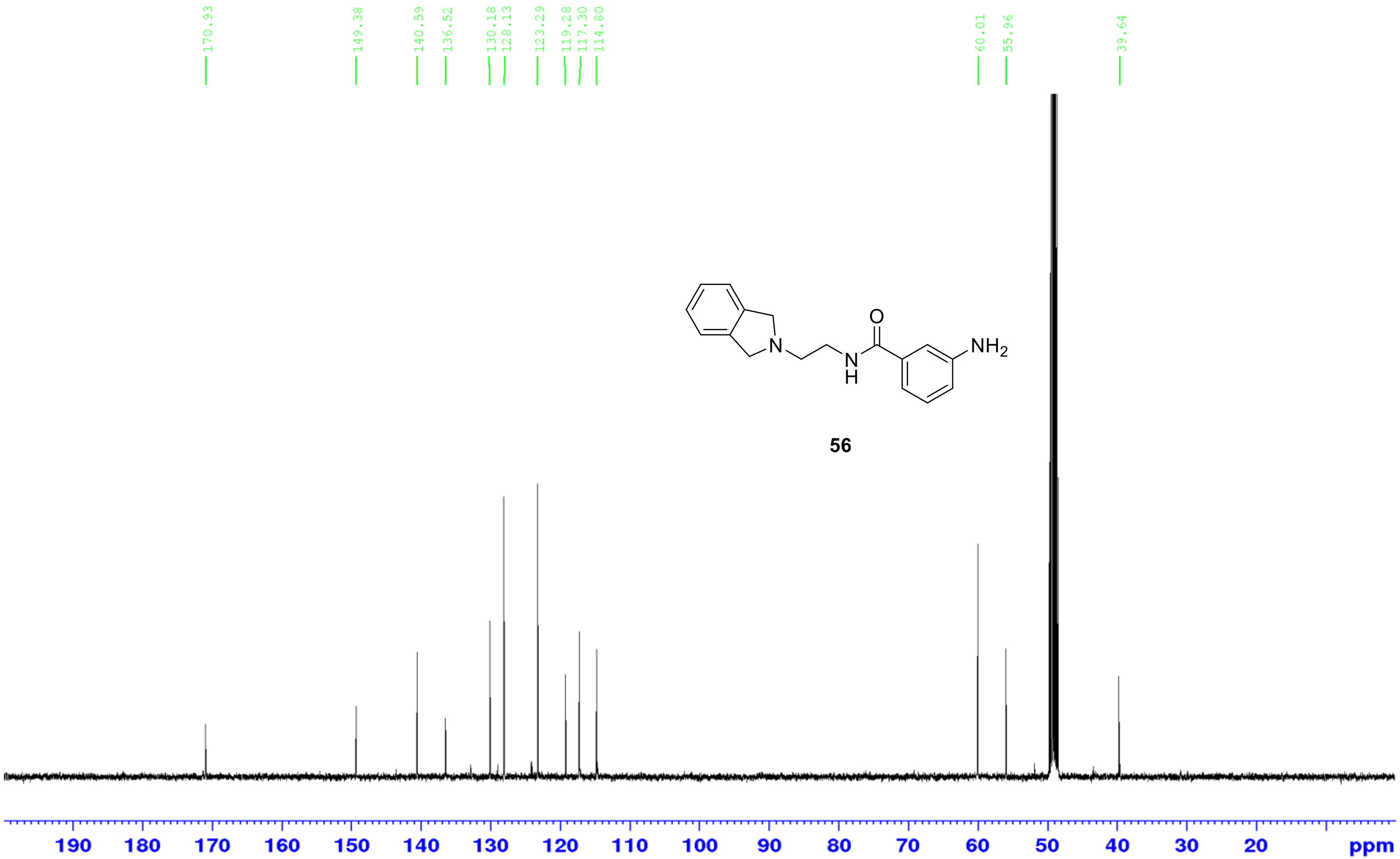
55

^{1}H NMR (400 MHz,MeOD)



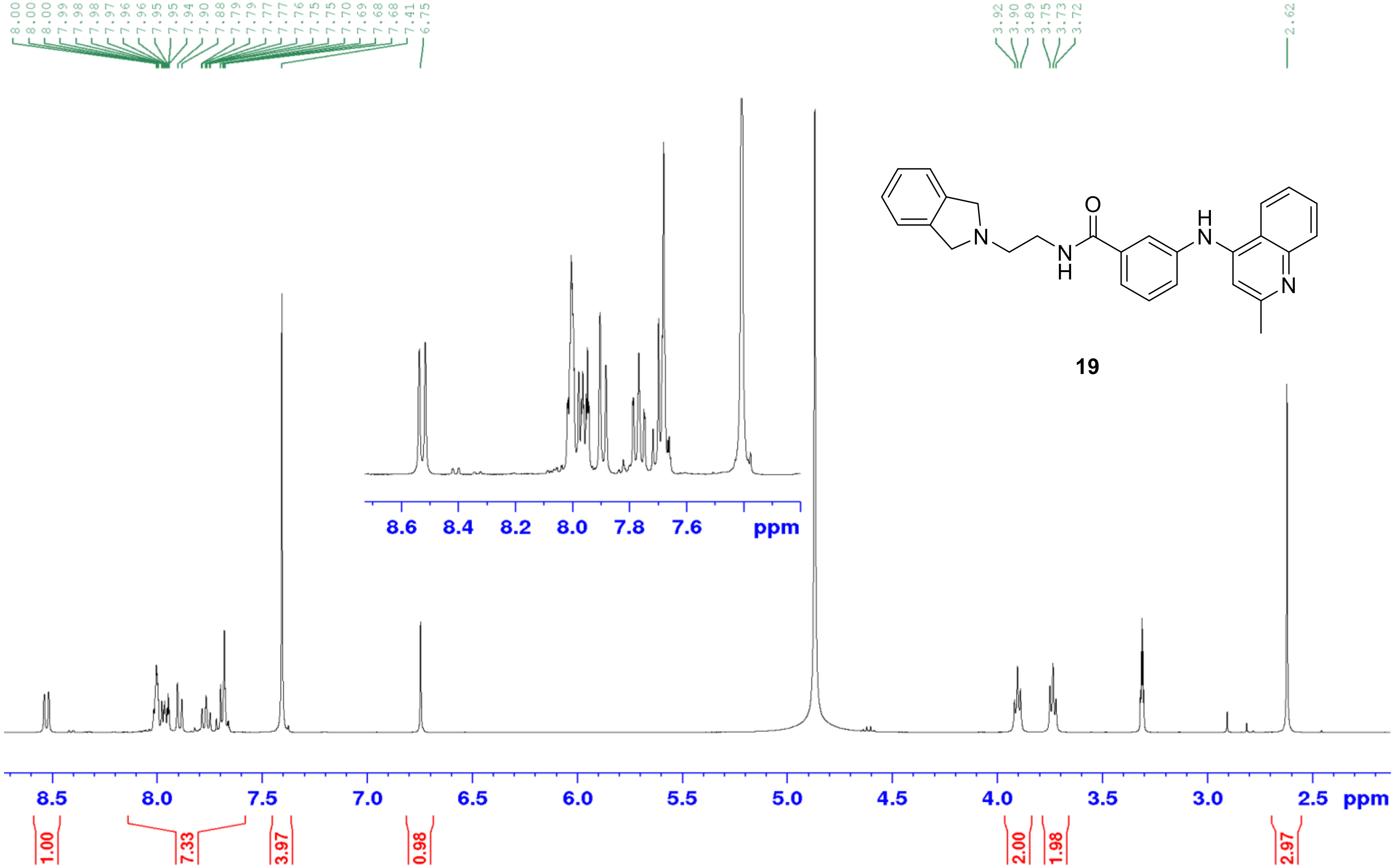
56

^{13}C NMR (101 MHz, MeOD)



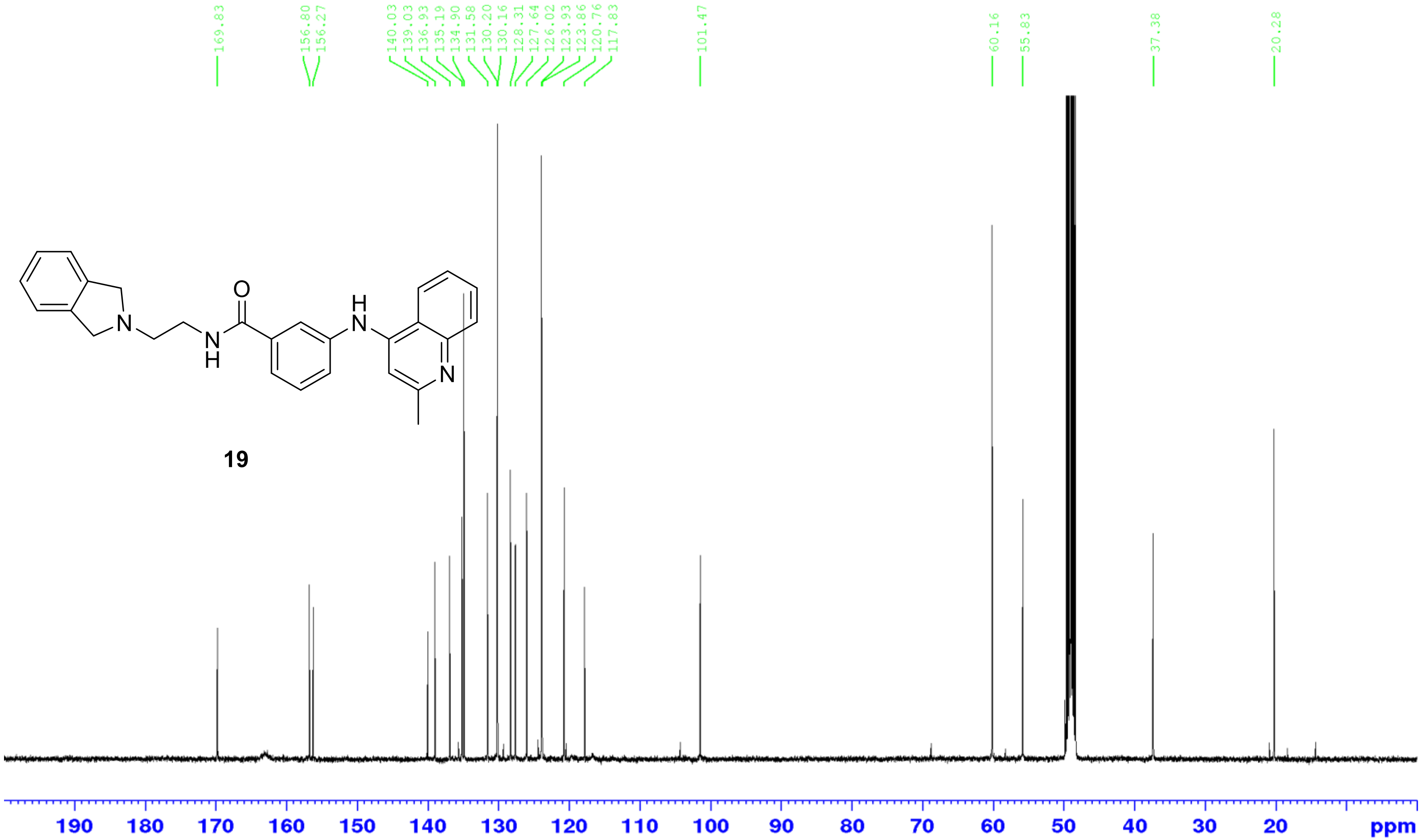
56

^{1}H NMR (400 MHz,MeOD)



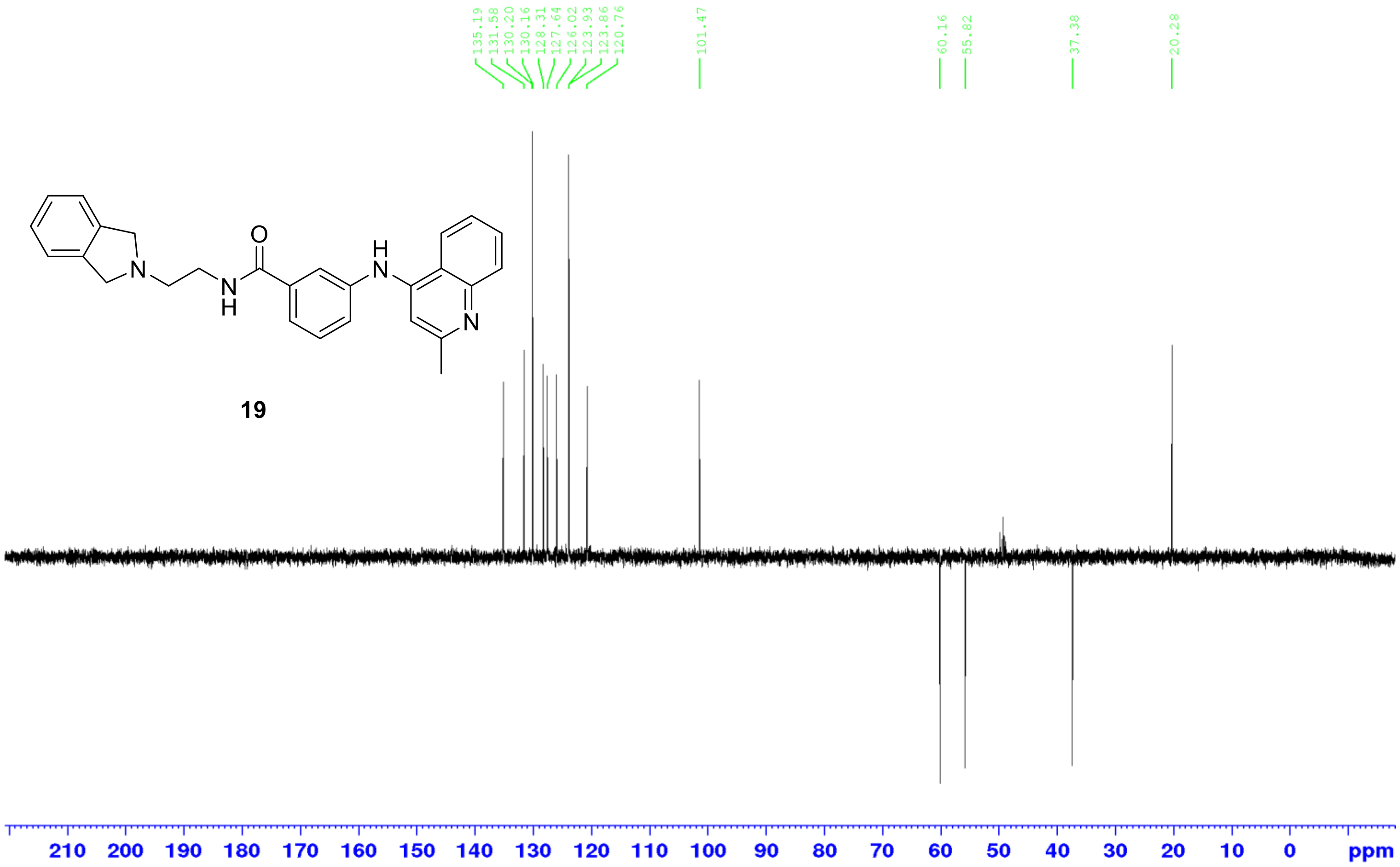
19

^{13}C NMR (101 MHz, MeOD)



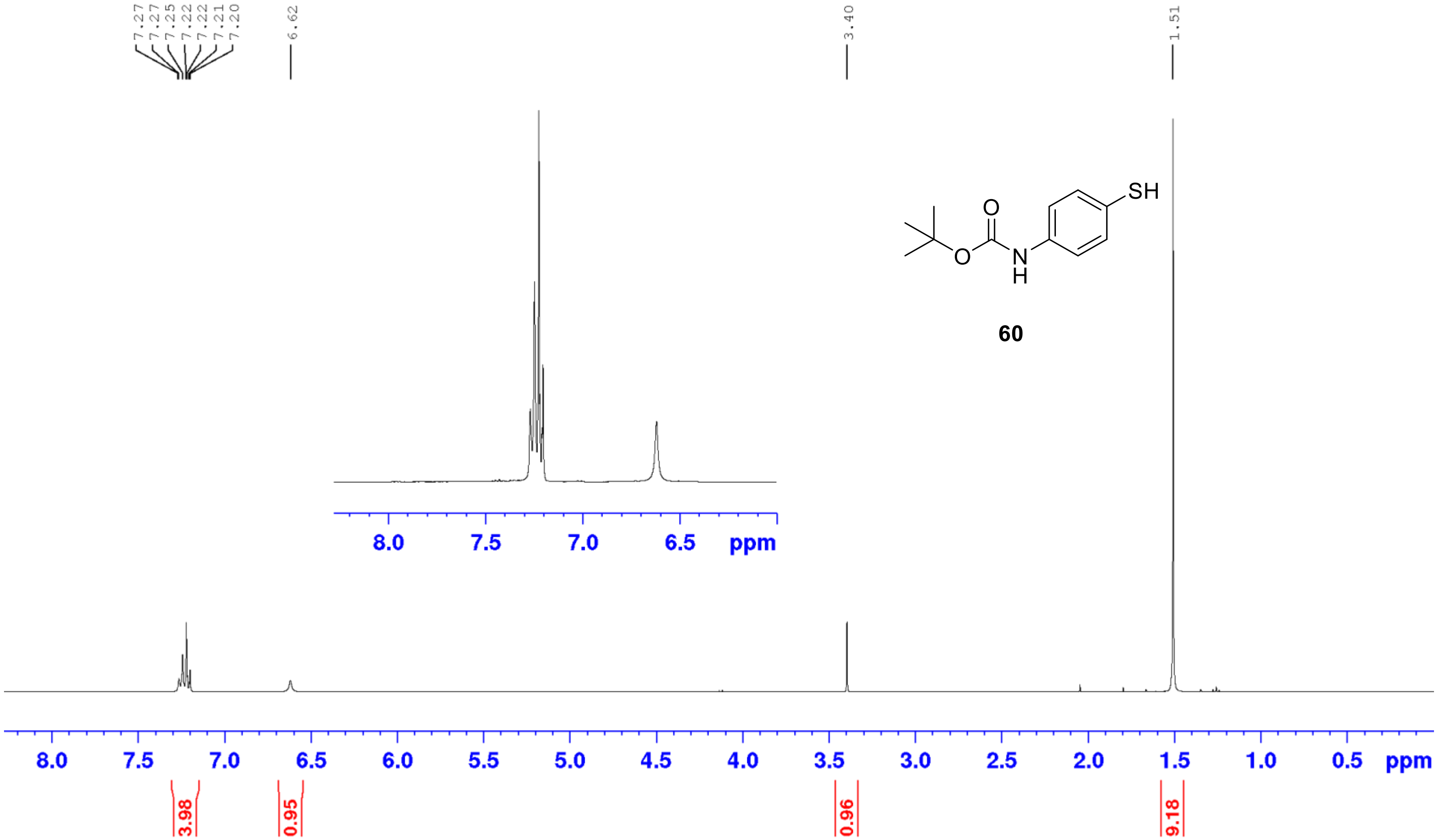
19

DEPT, $CDCl_3$



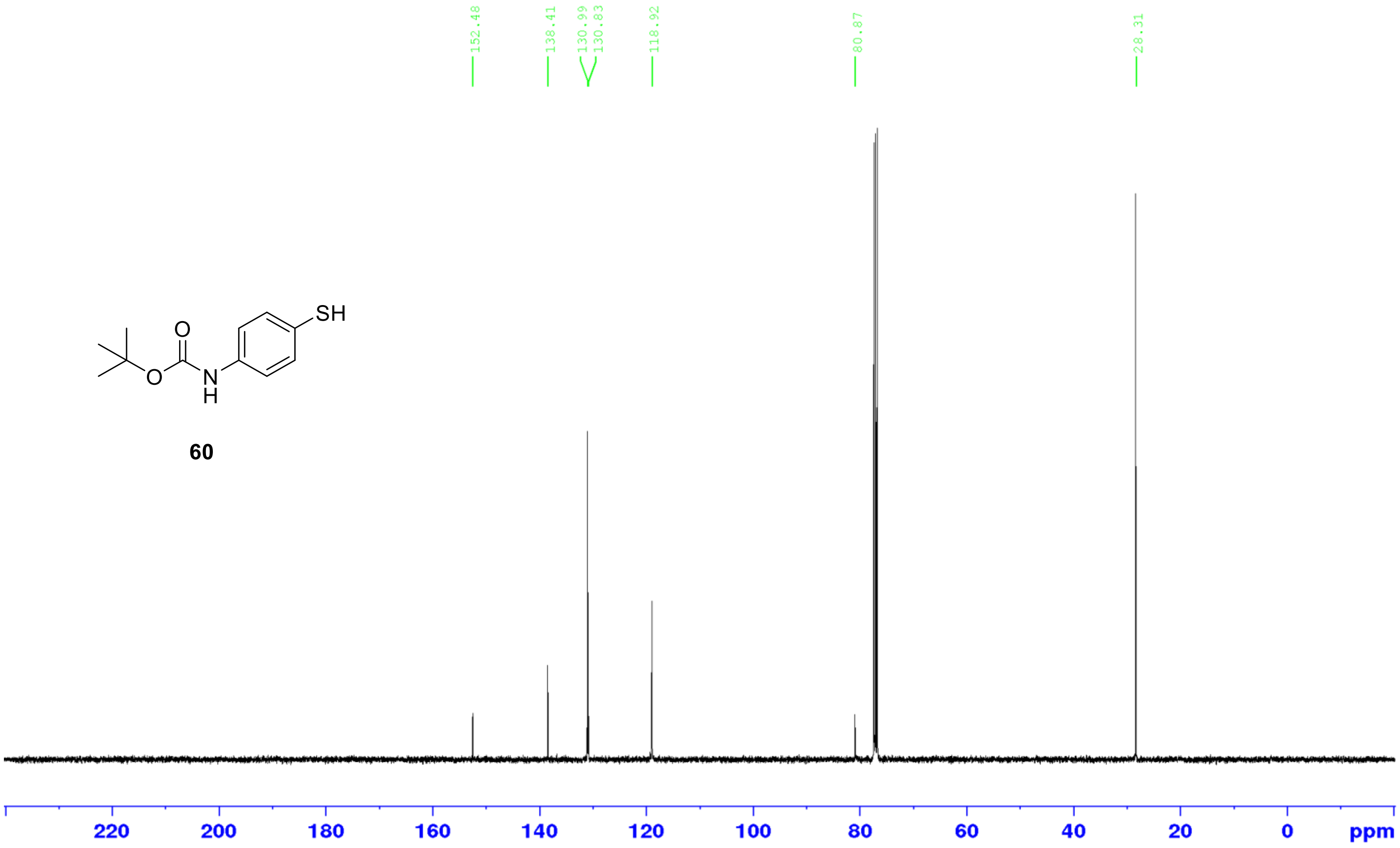
19

^1H NMR (400 MHz, $CDCl_3$)

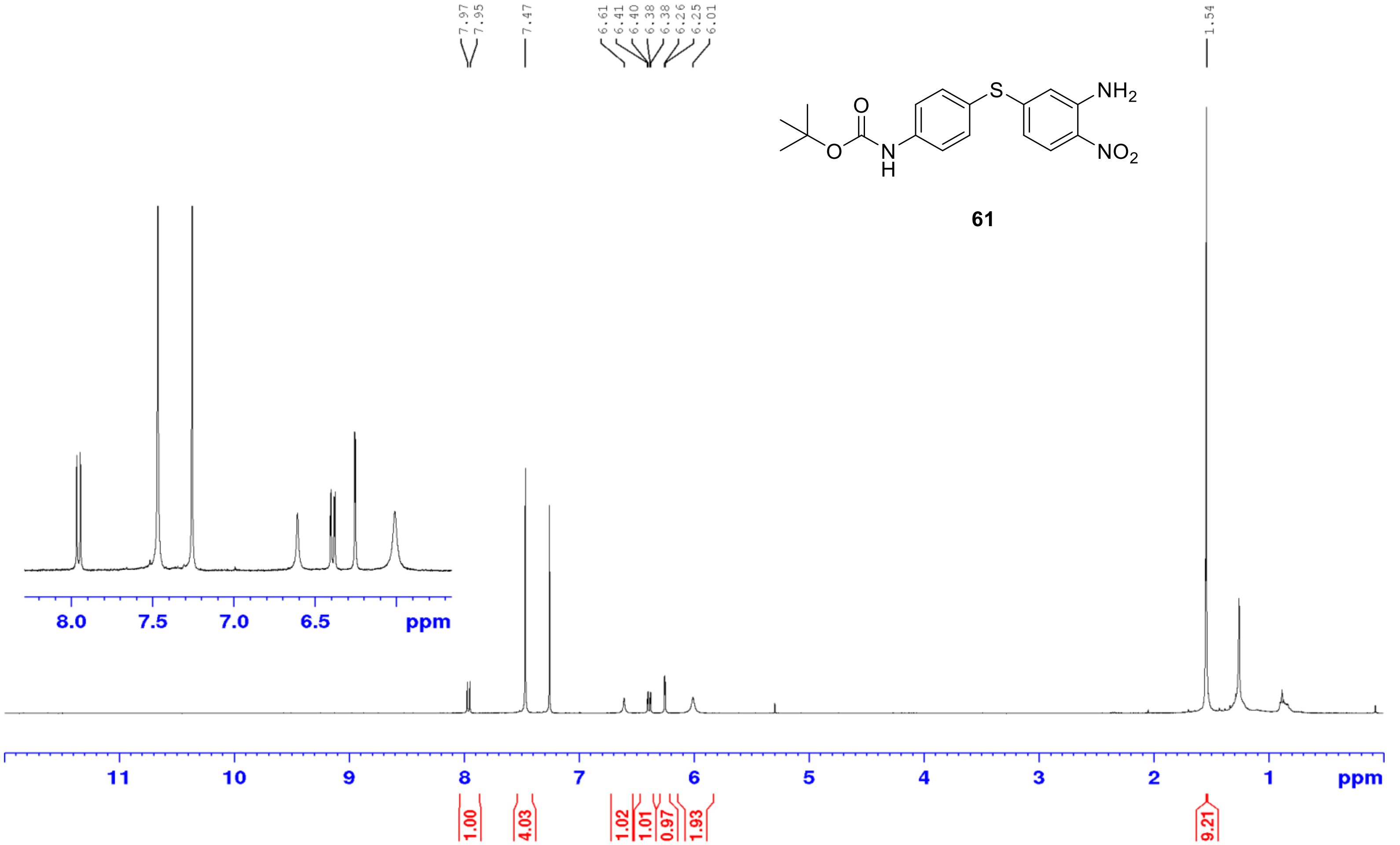


60

^{13}C NMR (101 MHz, CDCl$_3$)

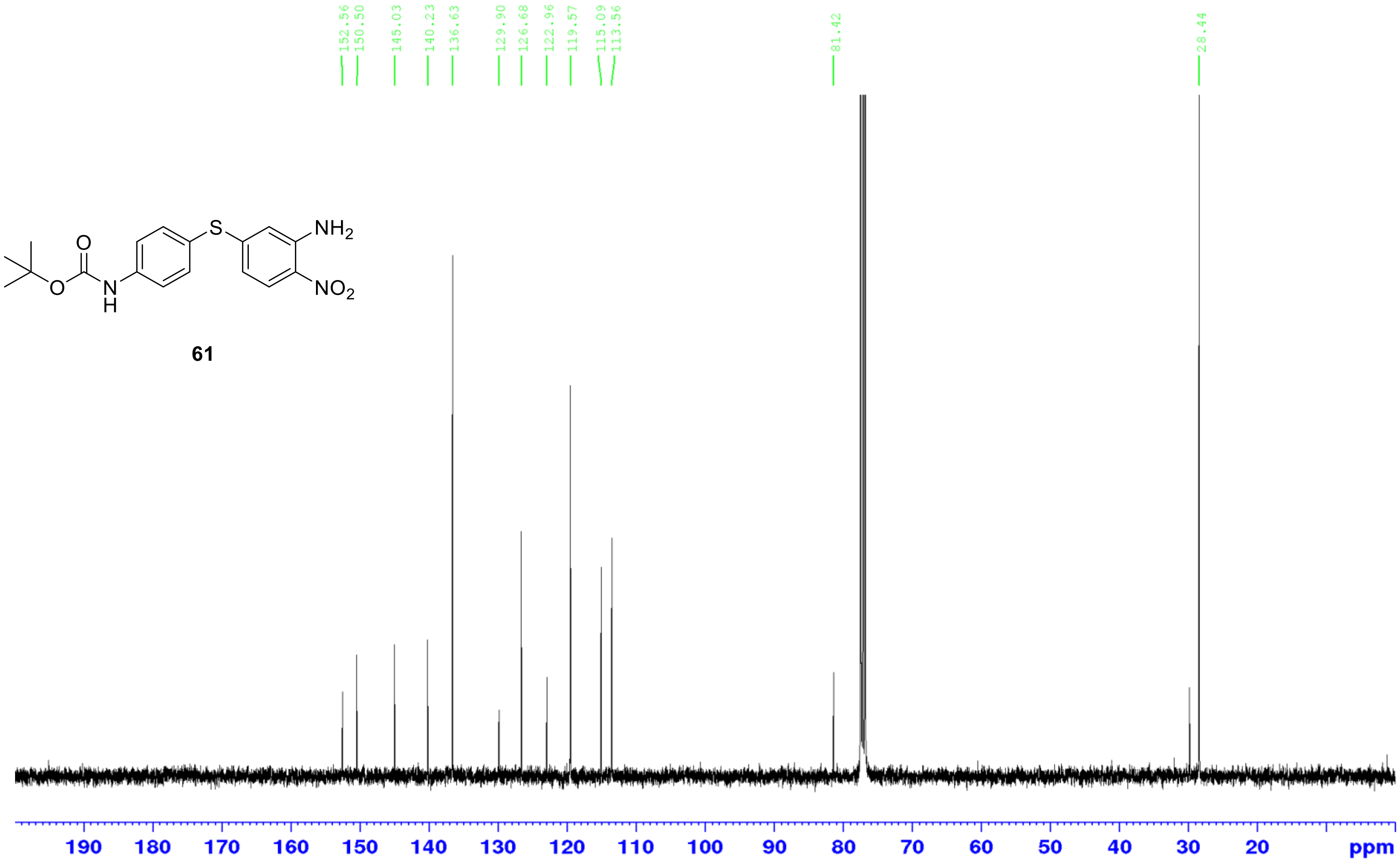


^{1}H NMR (400 MHz, $CDCl_3$)

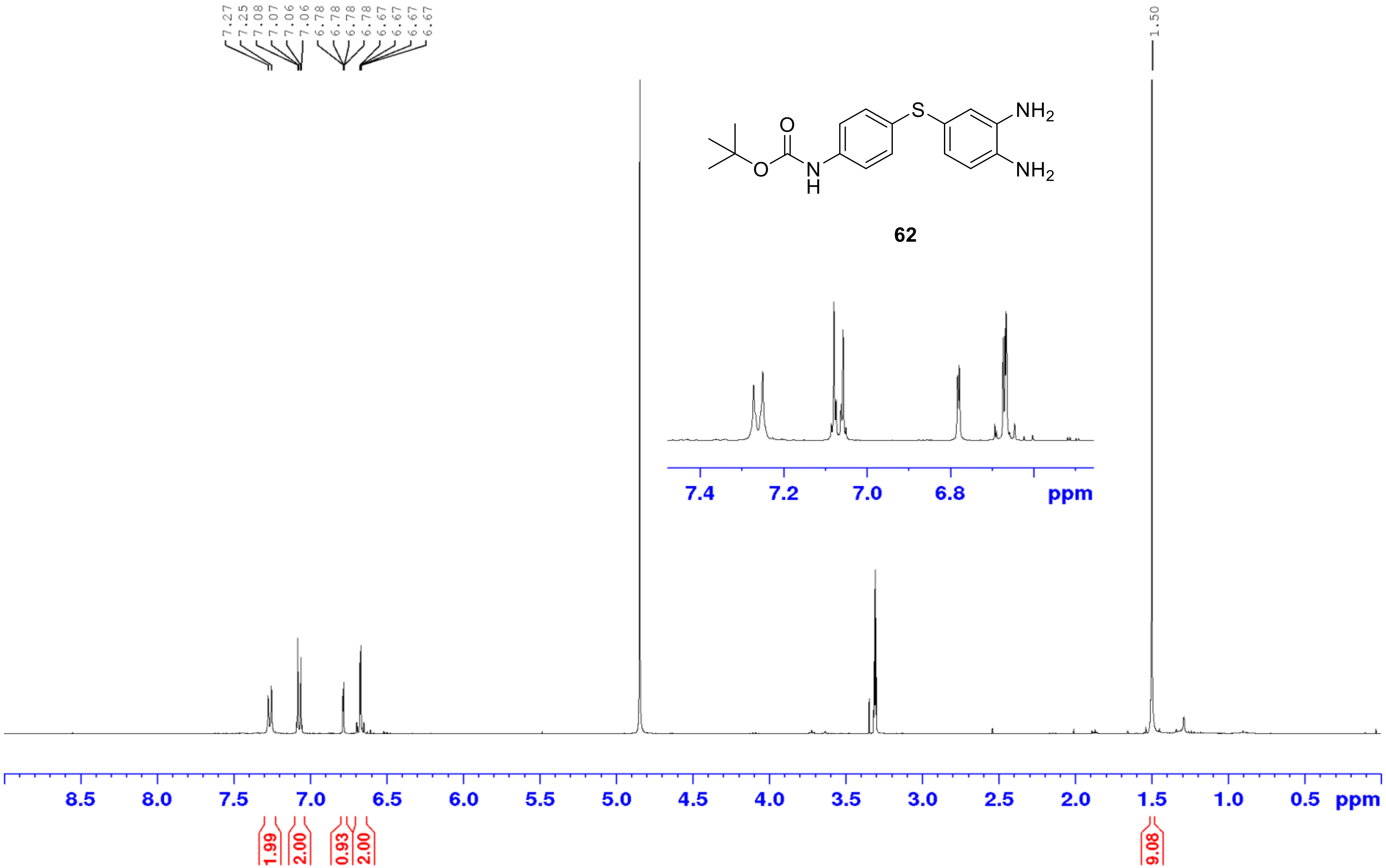


61

^{13}C NMR (101 MHz, $CDCl_3$)

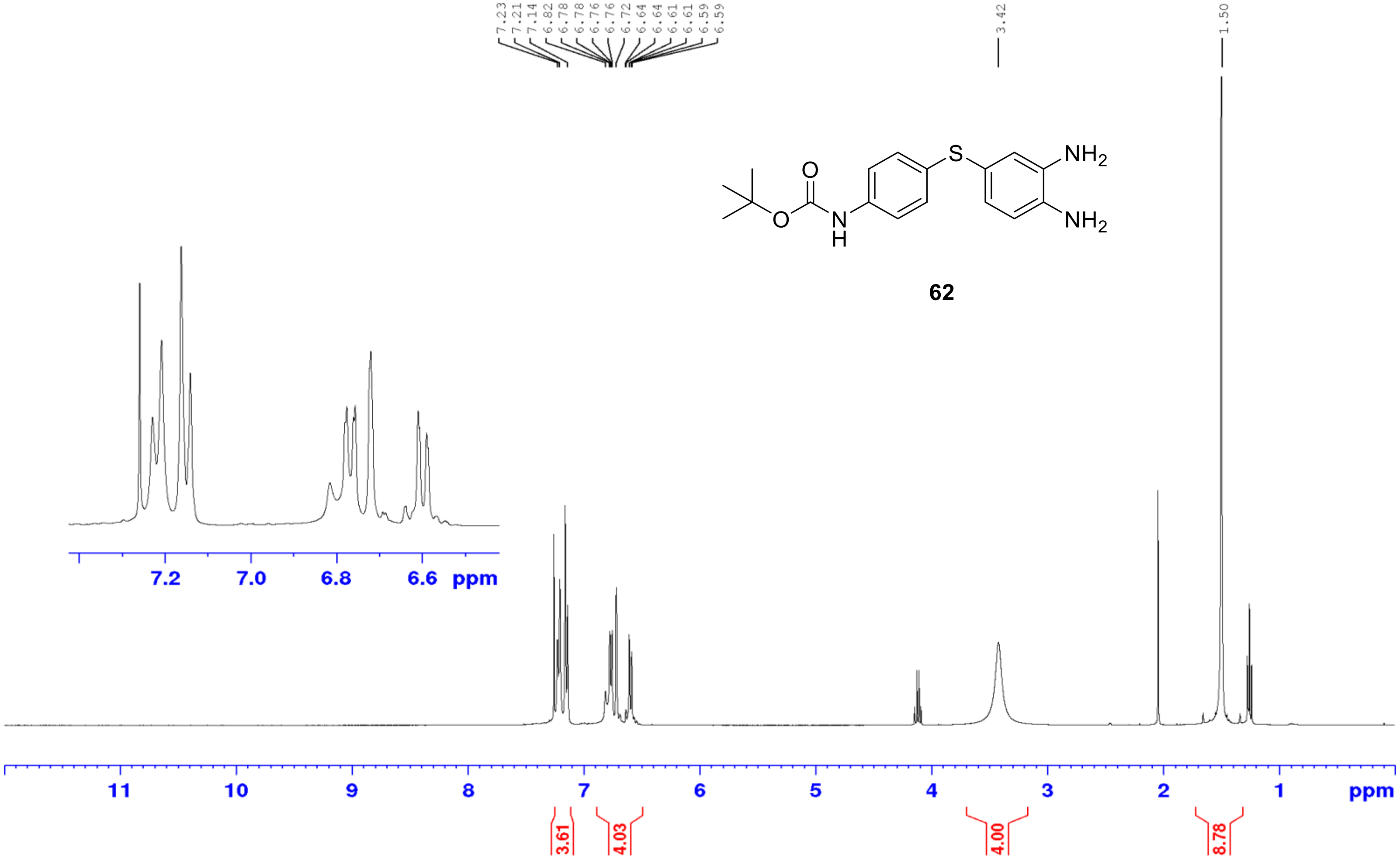


^{1}H NMR (400 MHz,MeOD)



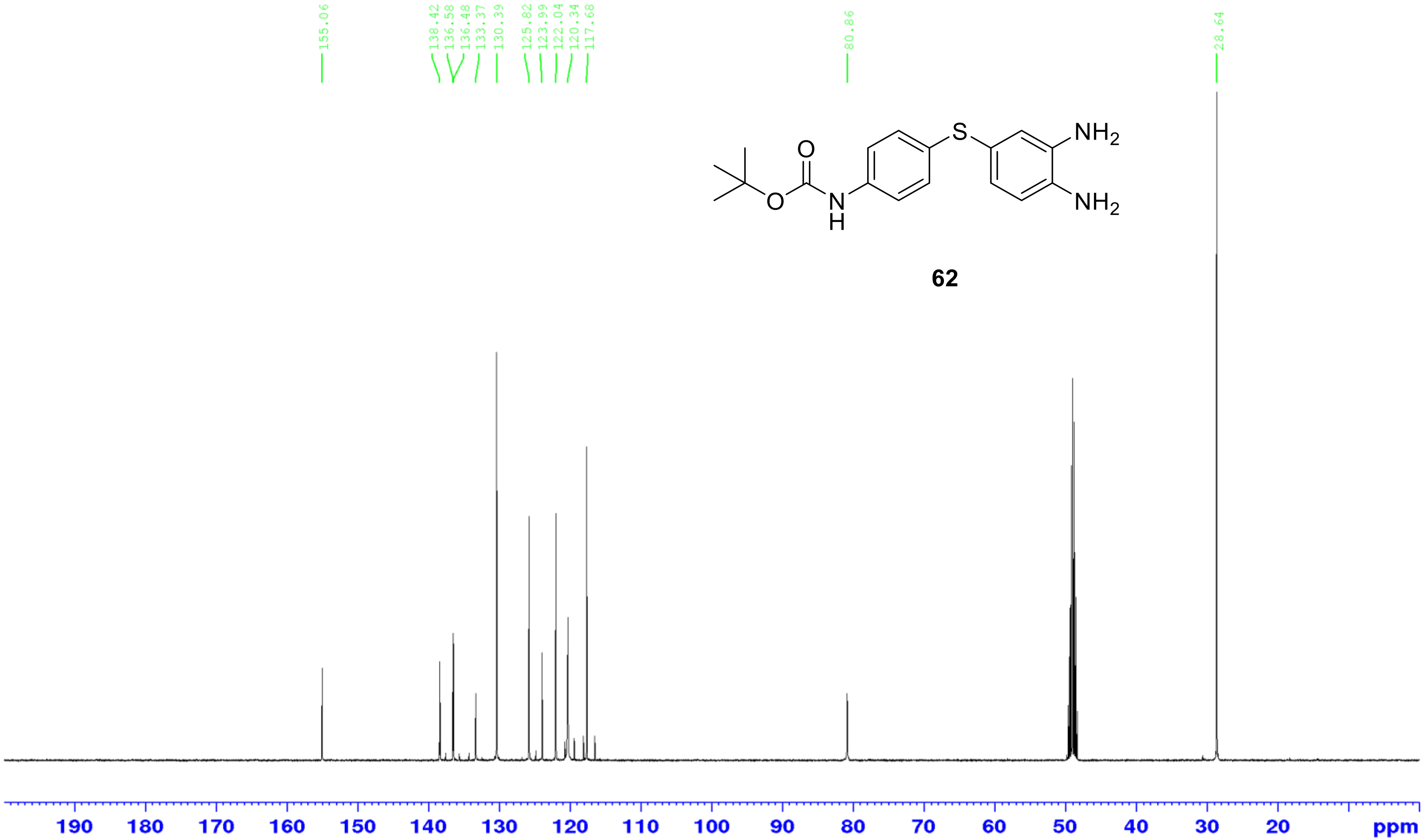
62

^{1}H NMR (400 MHz, $CDCl_3$)



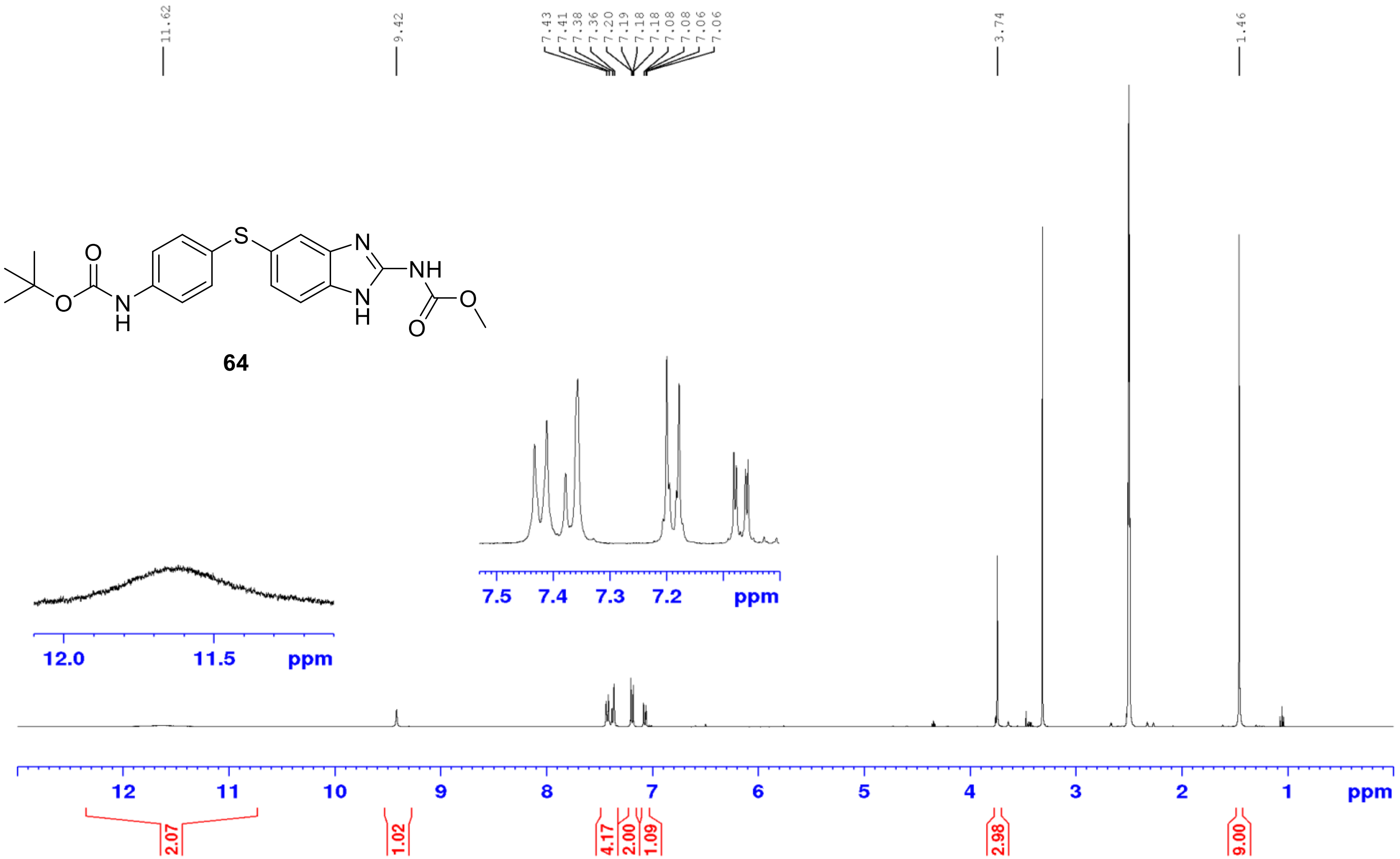
62

^{13}C NMR (101 MHz, MeOD)



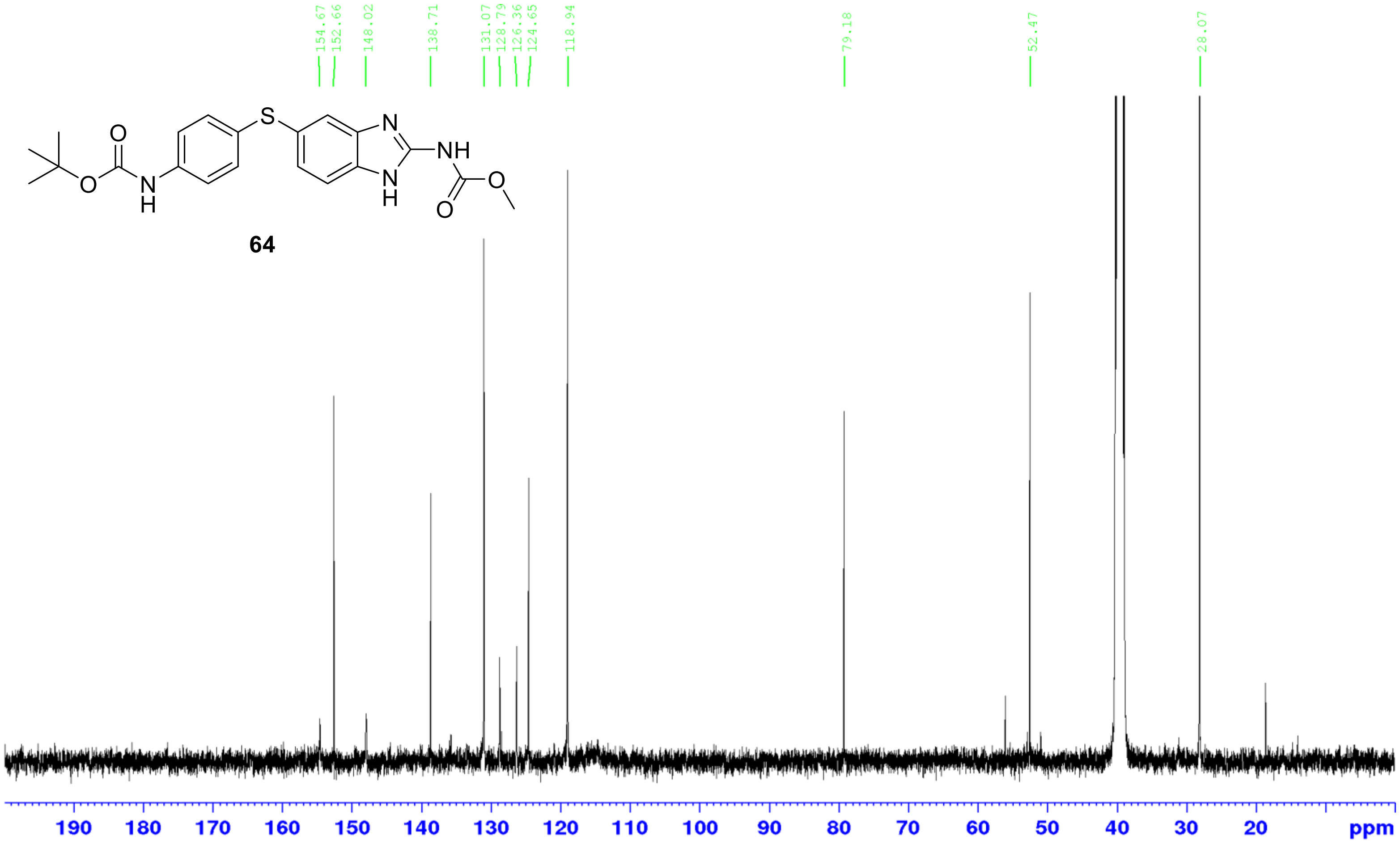
62

^{1}H NMR (400 MHz, DMSO-d_6)



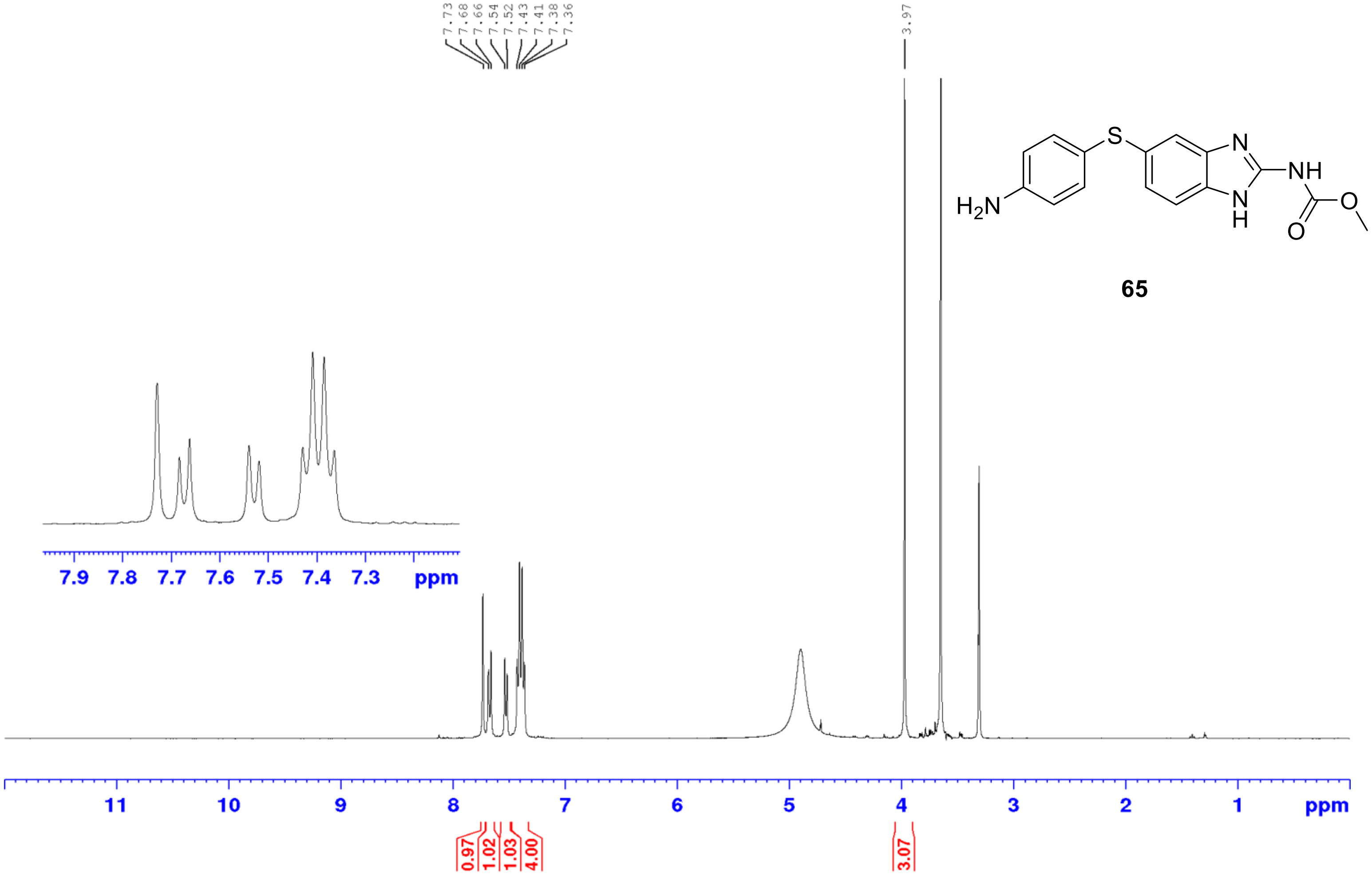
64

^{13}C NMR (101 MHz, DMSO-d_6)

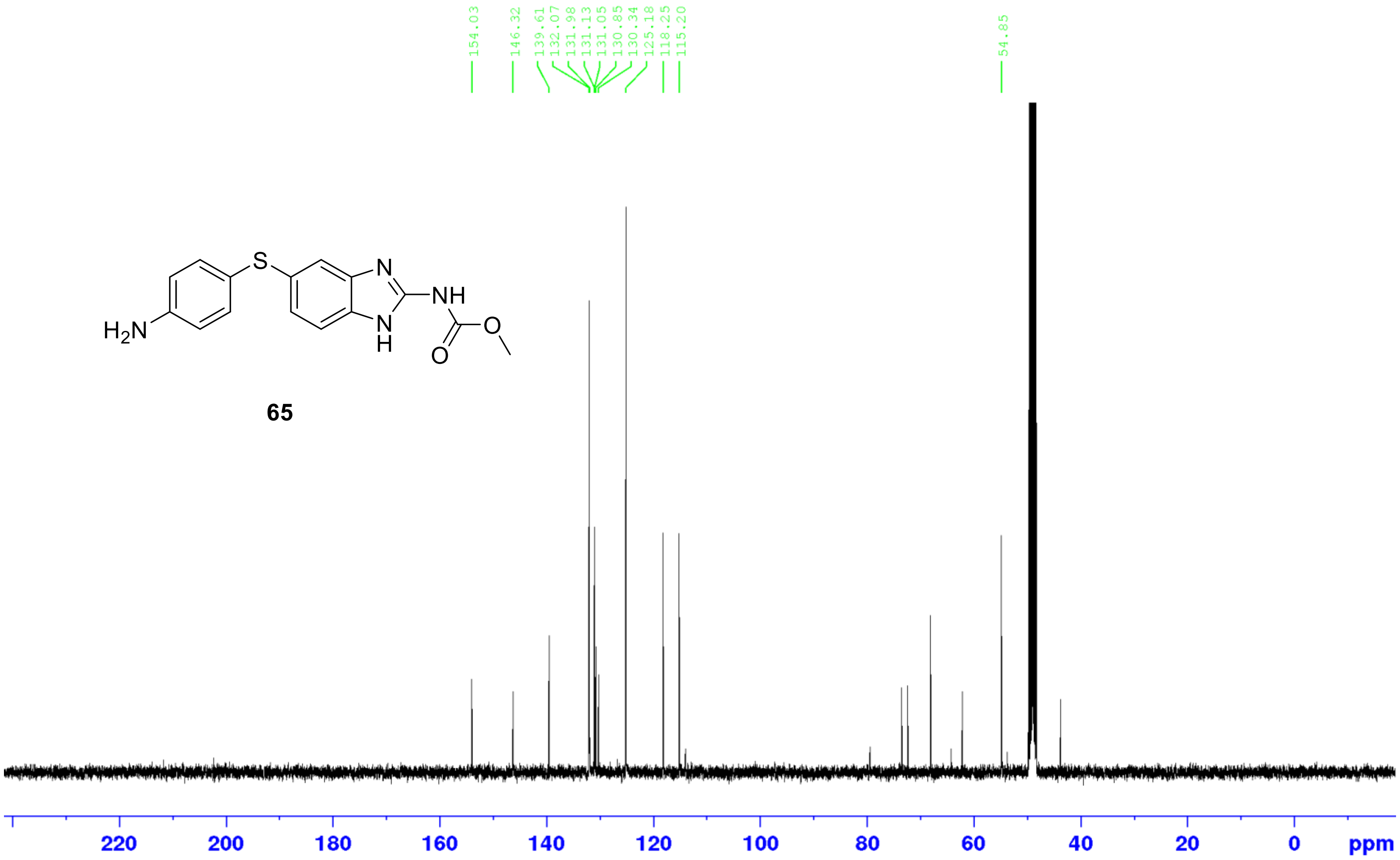


64

^{1}H NMR (400 MHz,MeOD)

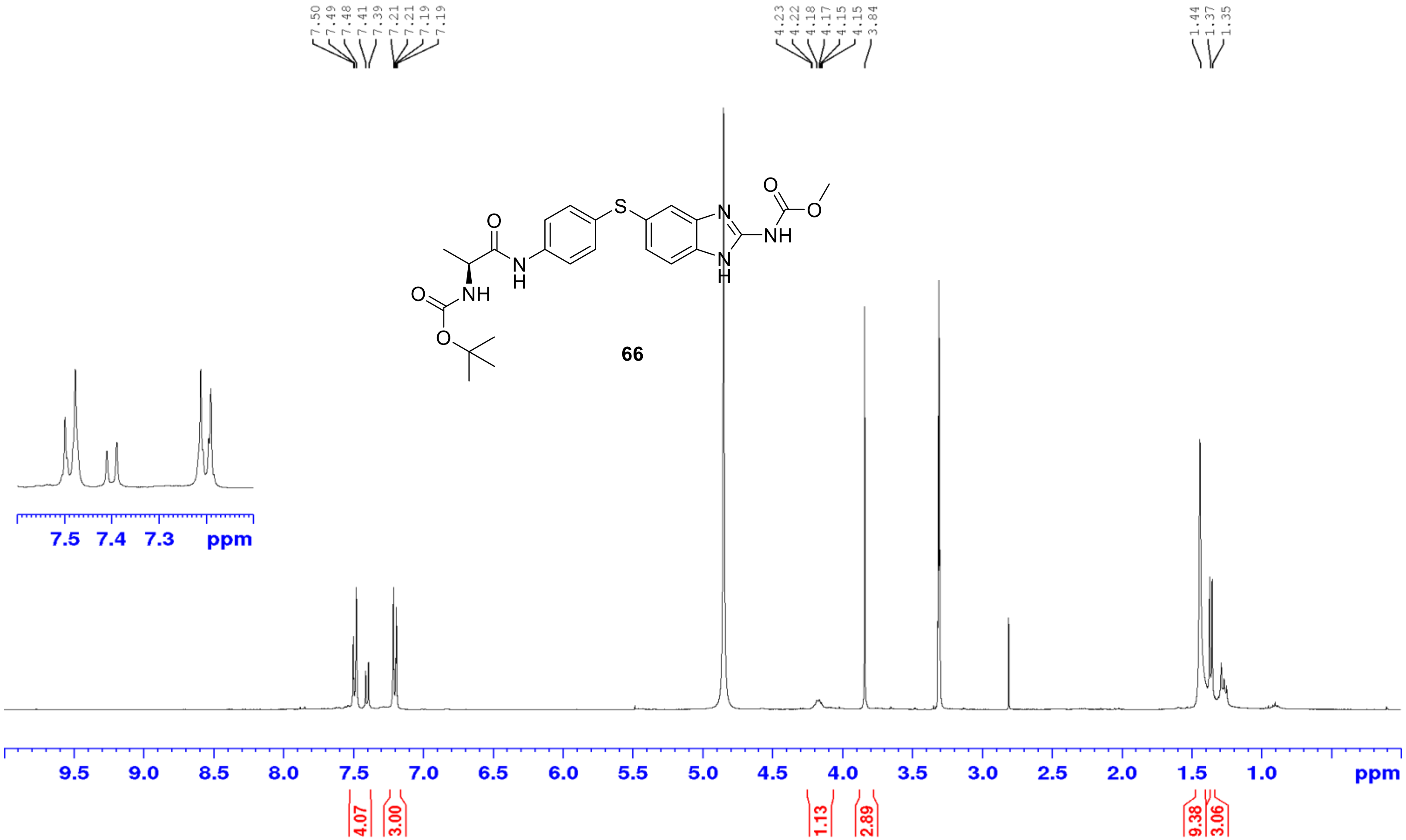


^{13}C NMR (101 MHz, MeOD)



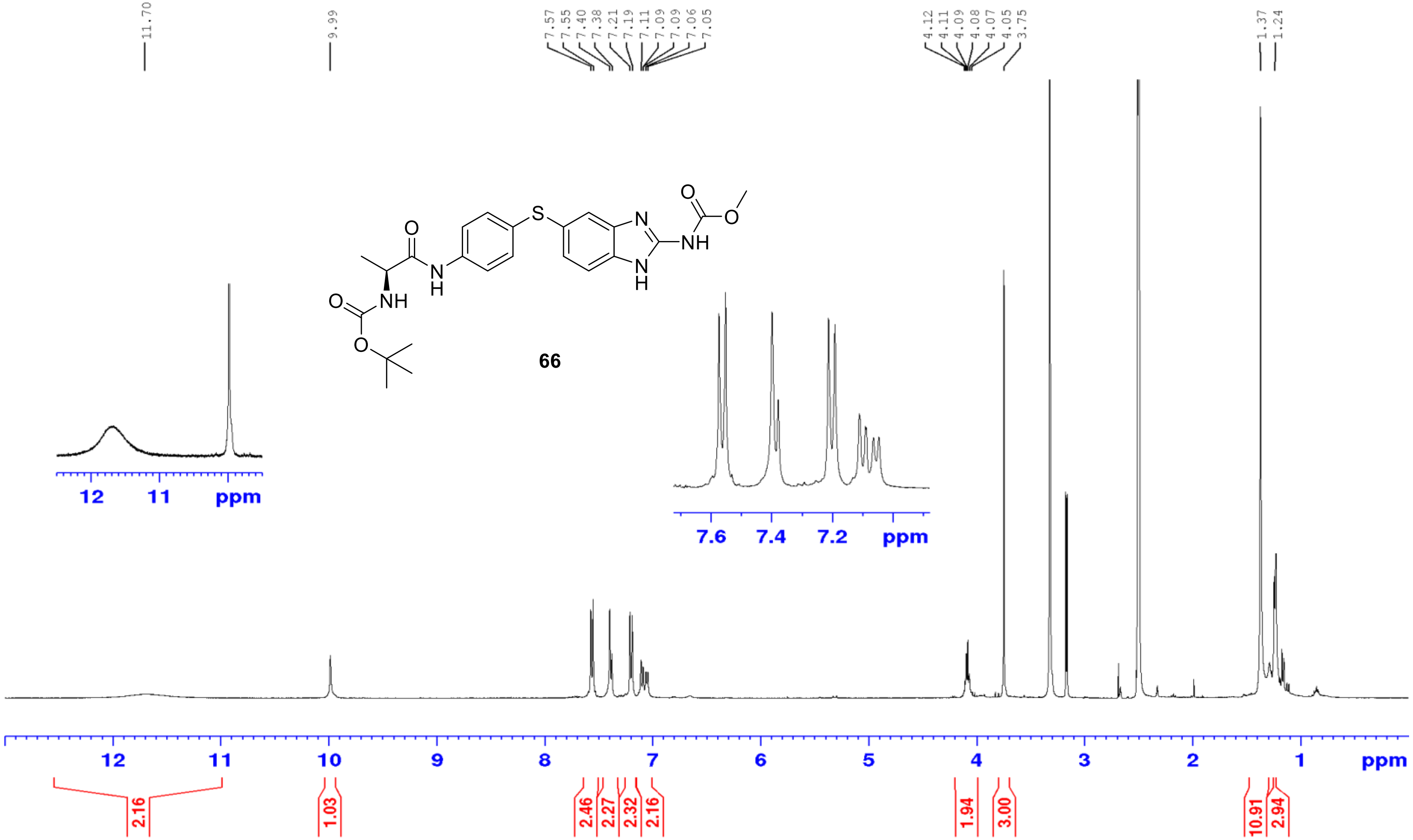
65

^{1}H NMR (400 MHz,MeOD)



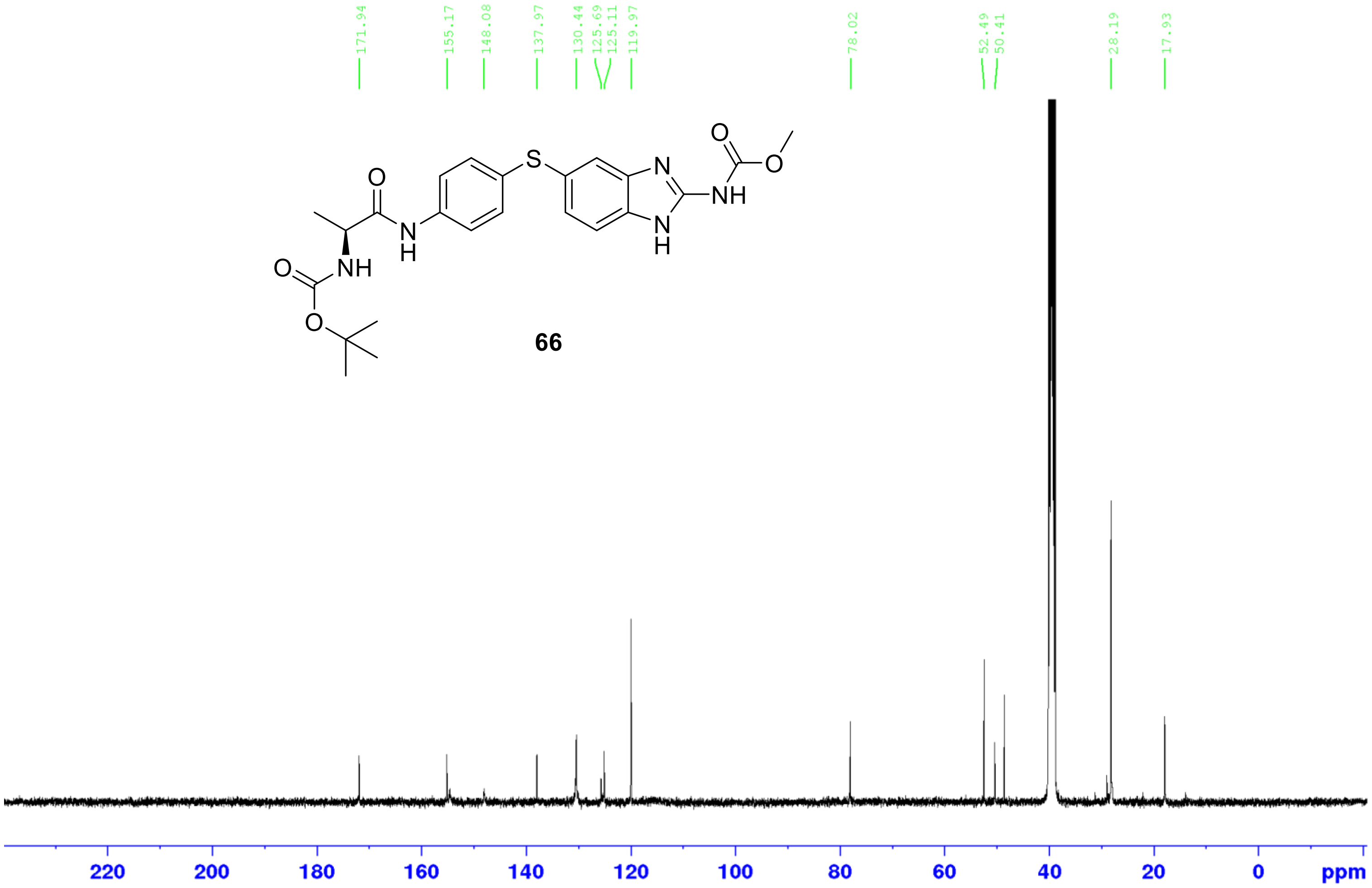
66

^{1}H NMR (400 MHz, DMSO-d$_6$)



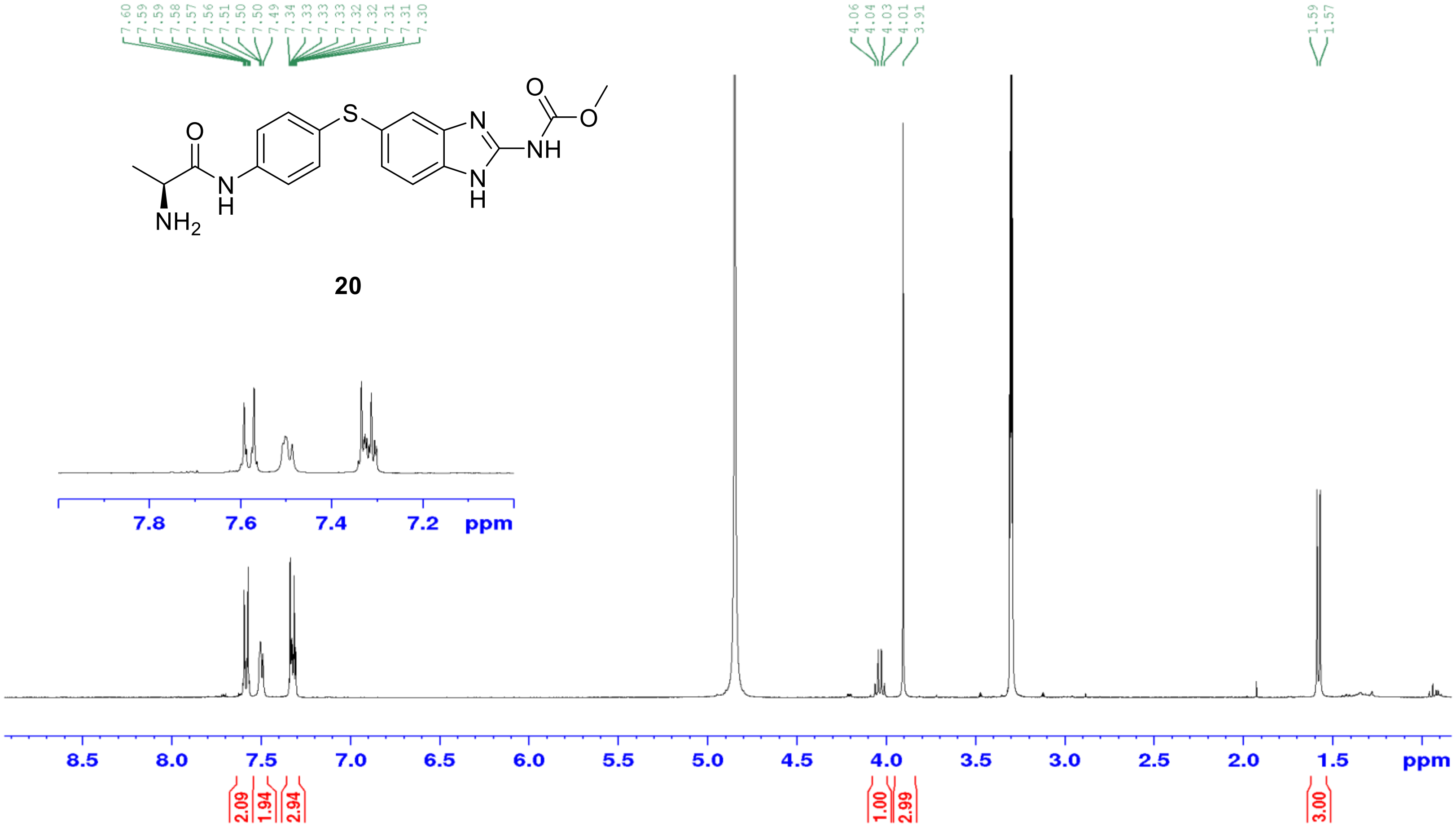
66

^{13}C NMR (101 MHz, DMSO-d_6)

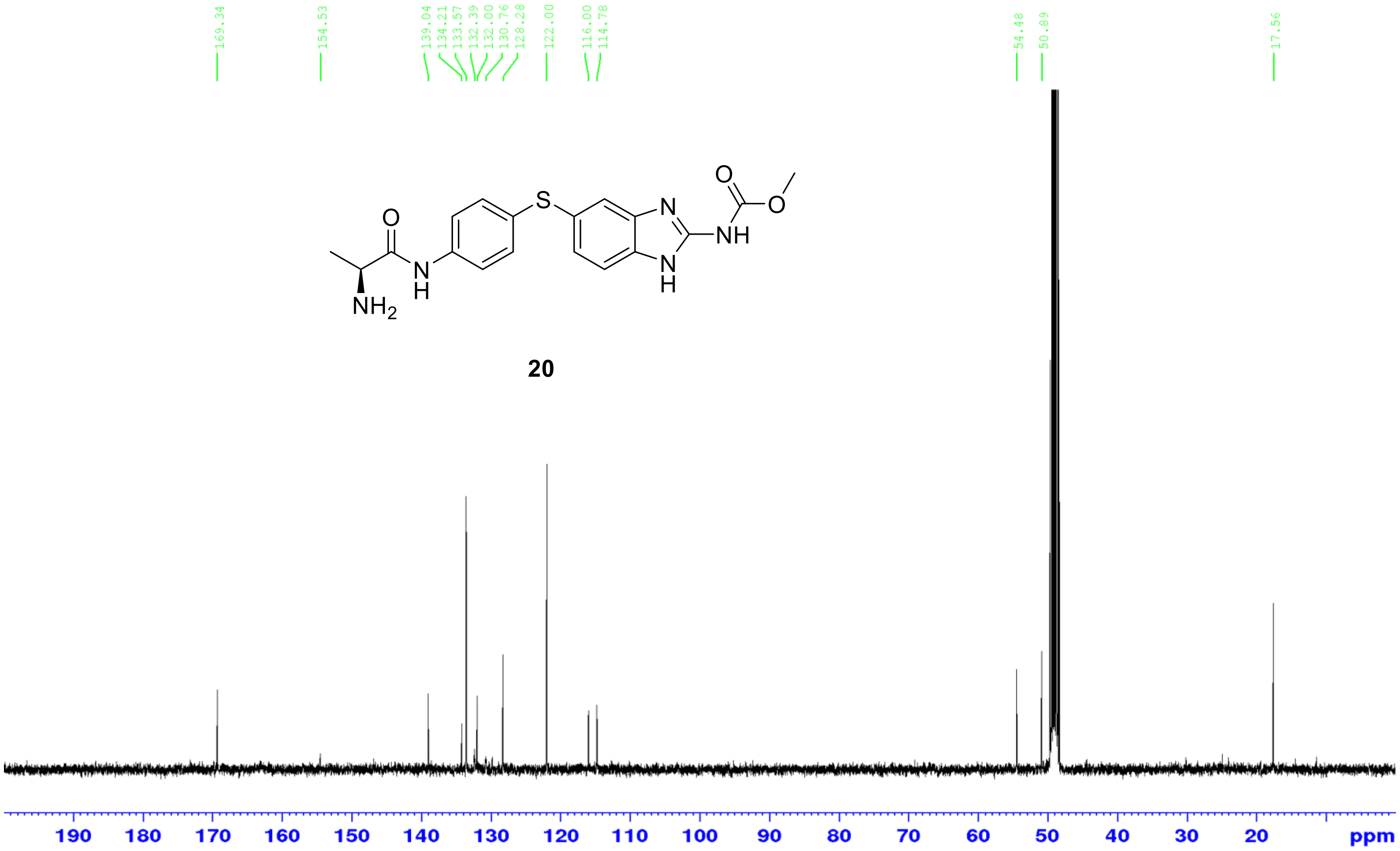


66

^{1}H NMR (400 MHz, MeOD)

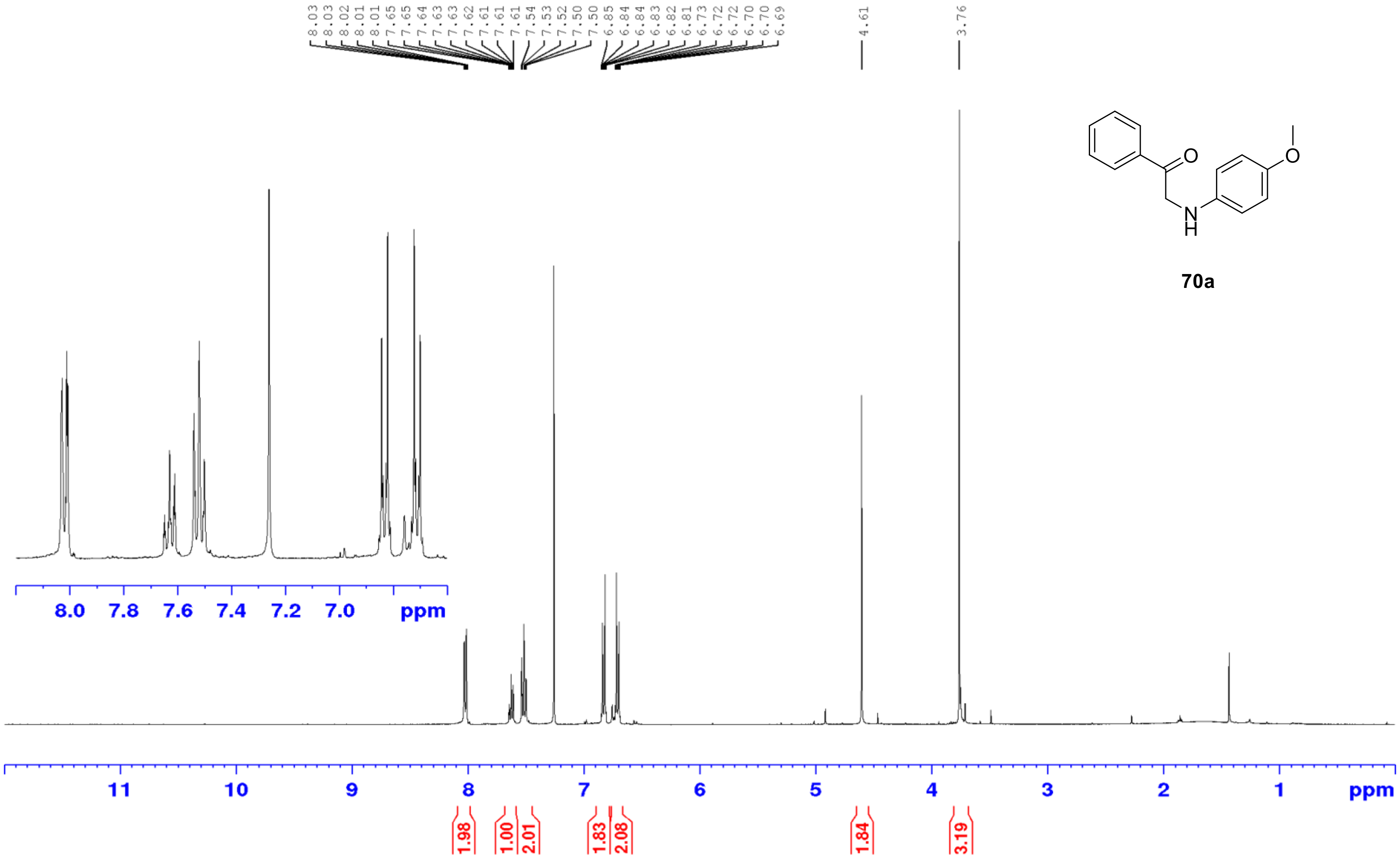


^{13}C NMR (101 MHz, MeOD)



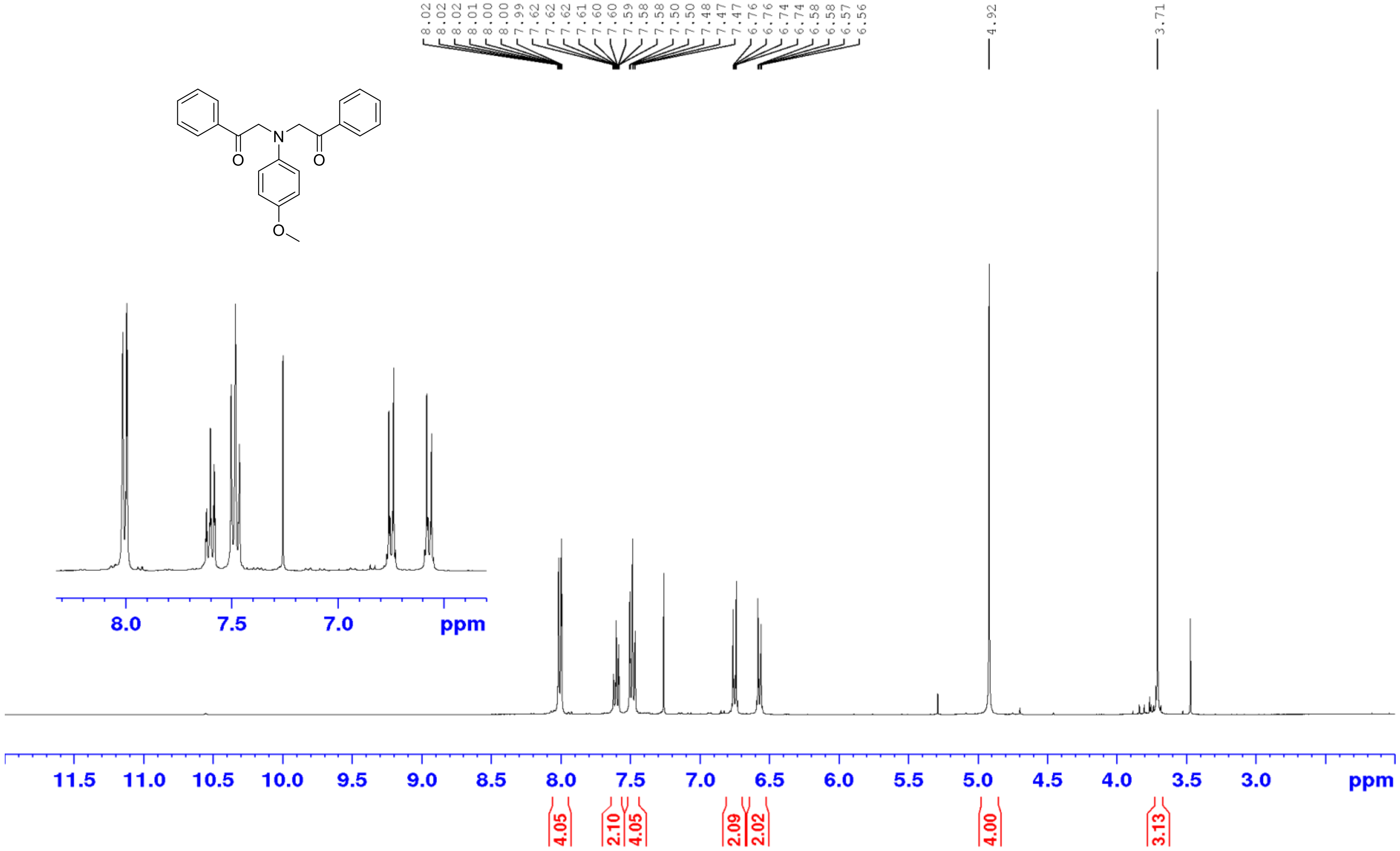
20

^1H NMR (400 MHz, $CDCl_3$)

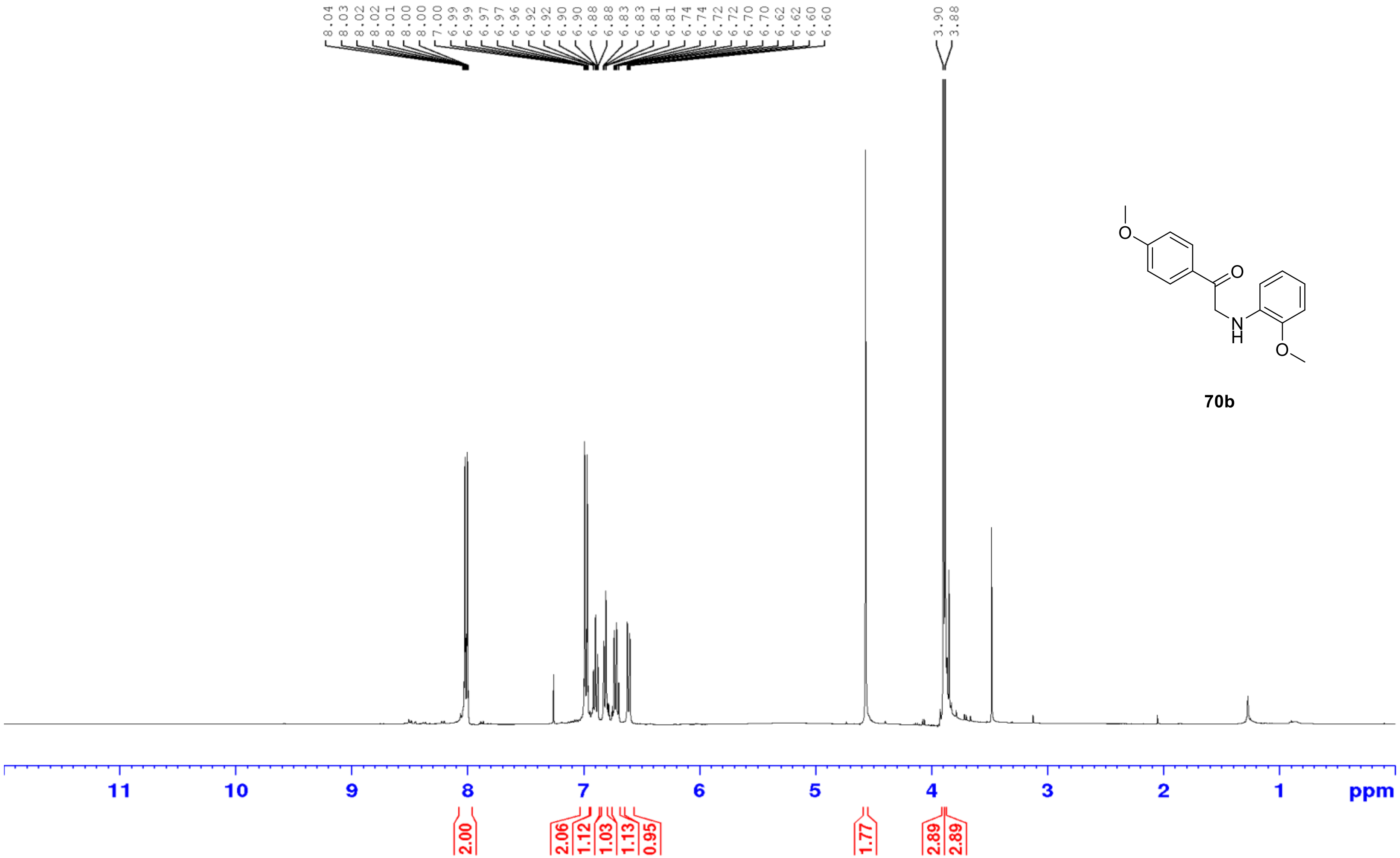


70a

^{1}H NMR (400 MHz, CDCl$_3$)

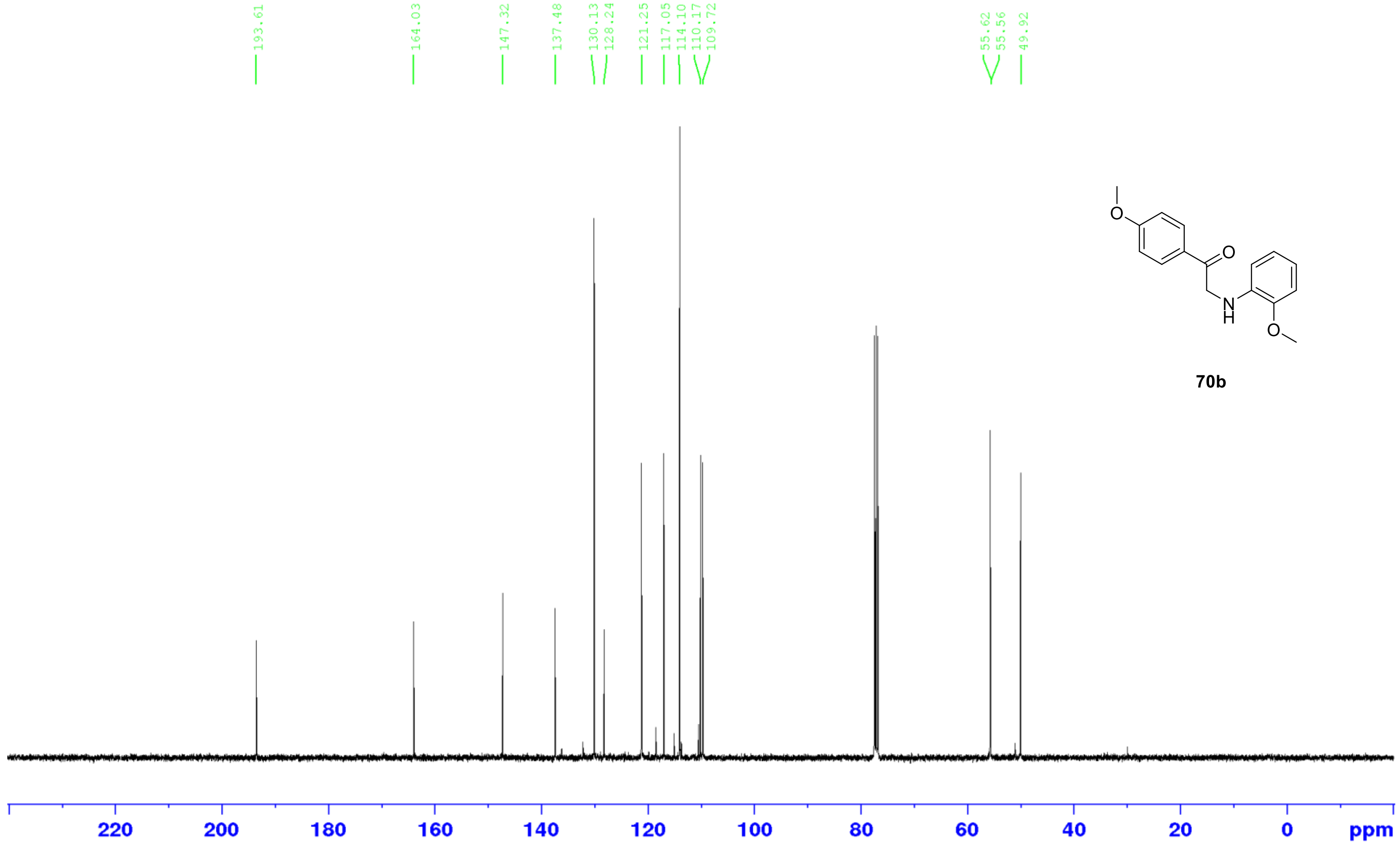


^{1}H NMR (400 MHz, $CDCl_3$)



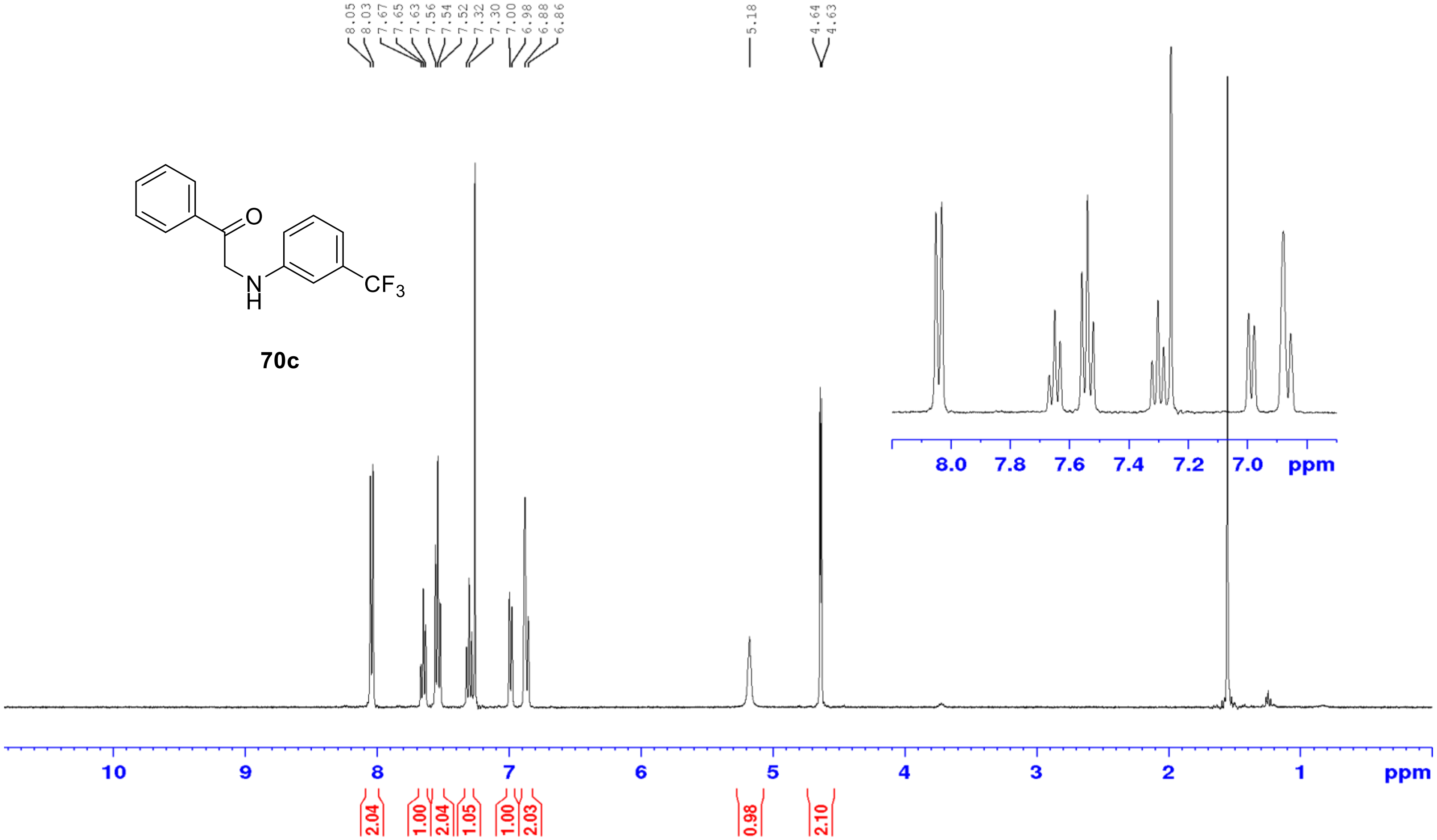
70b

^{13}C NMR (101 MHz, $CDCl_3$)



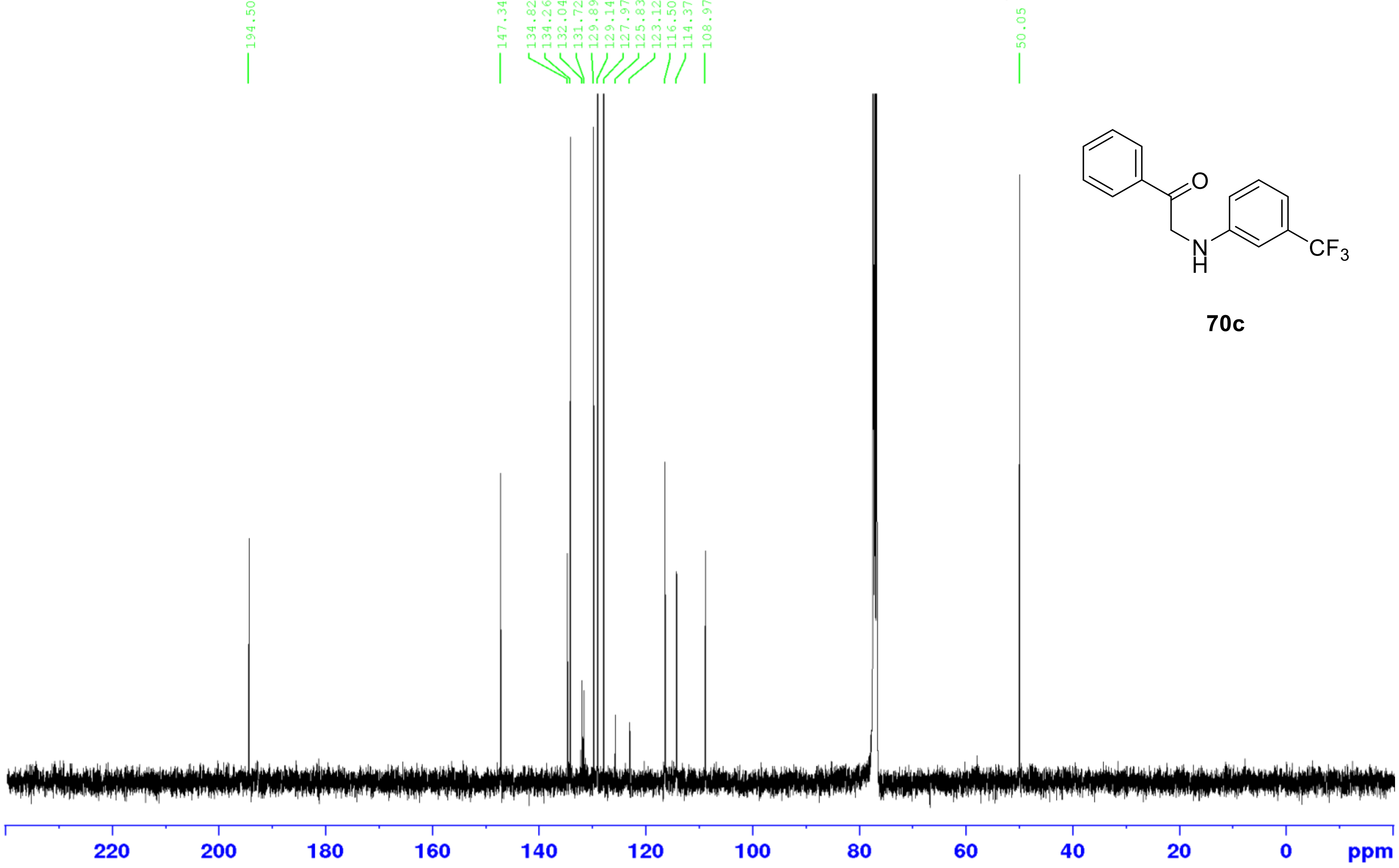
70b

^{1}H NMR (400 MHz, $CDCl_3$)



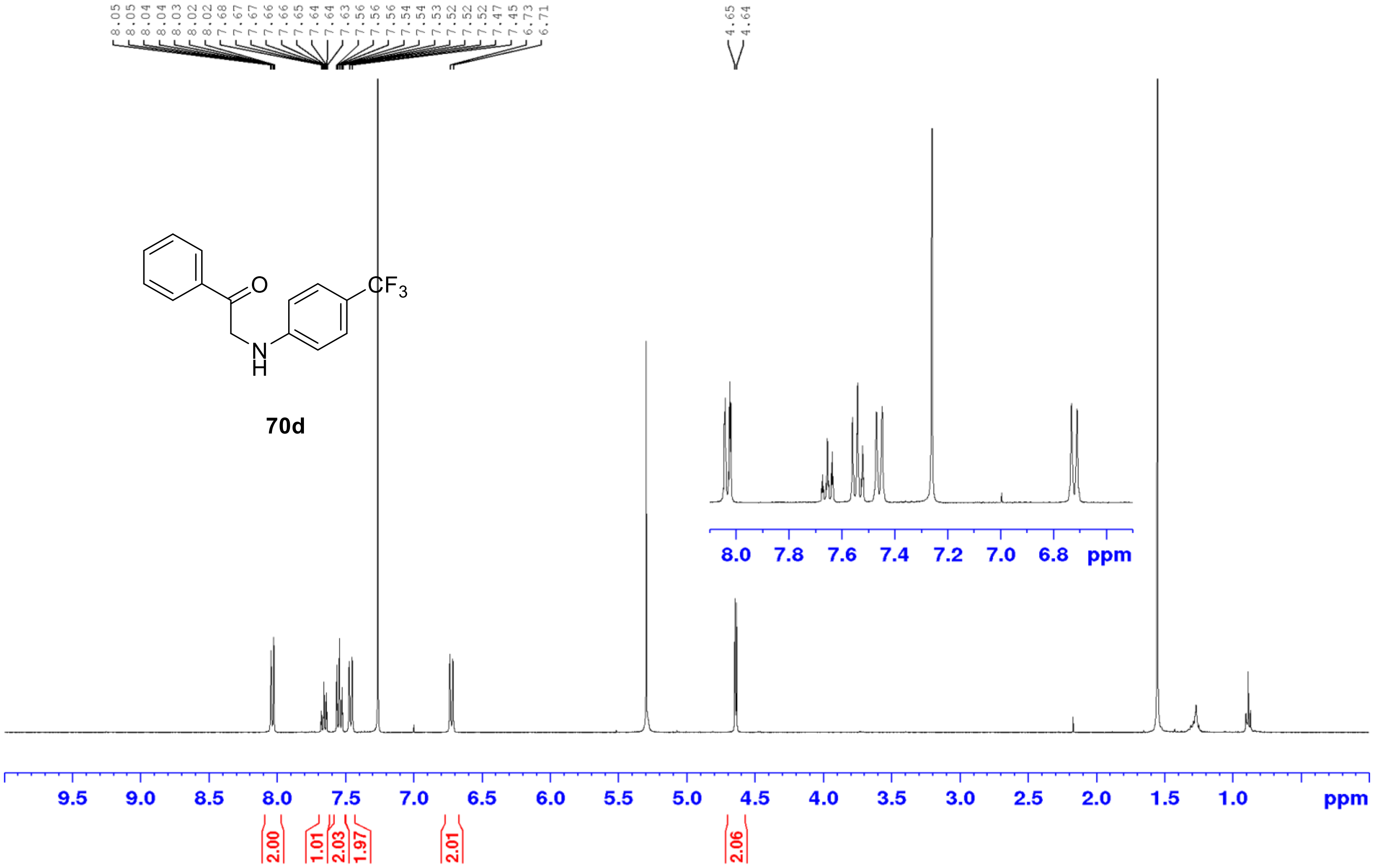
70c

^{13}C NMR (101 MHz, CDCl$_3$)



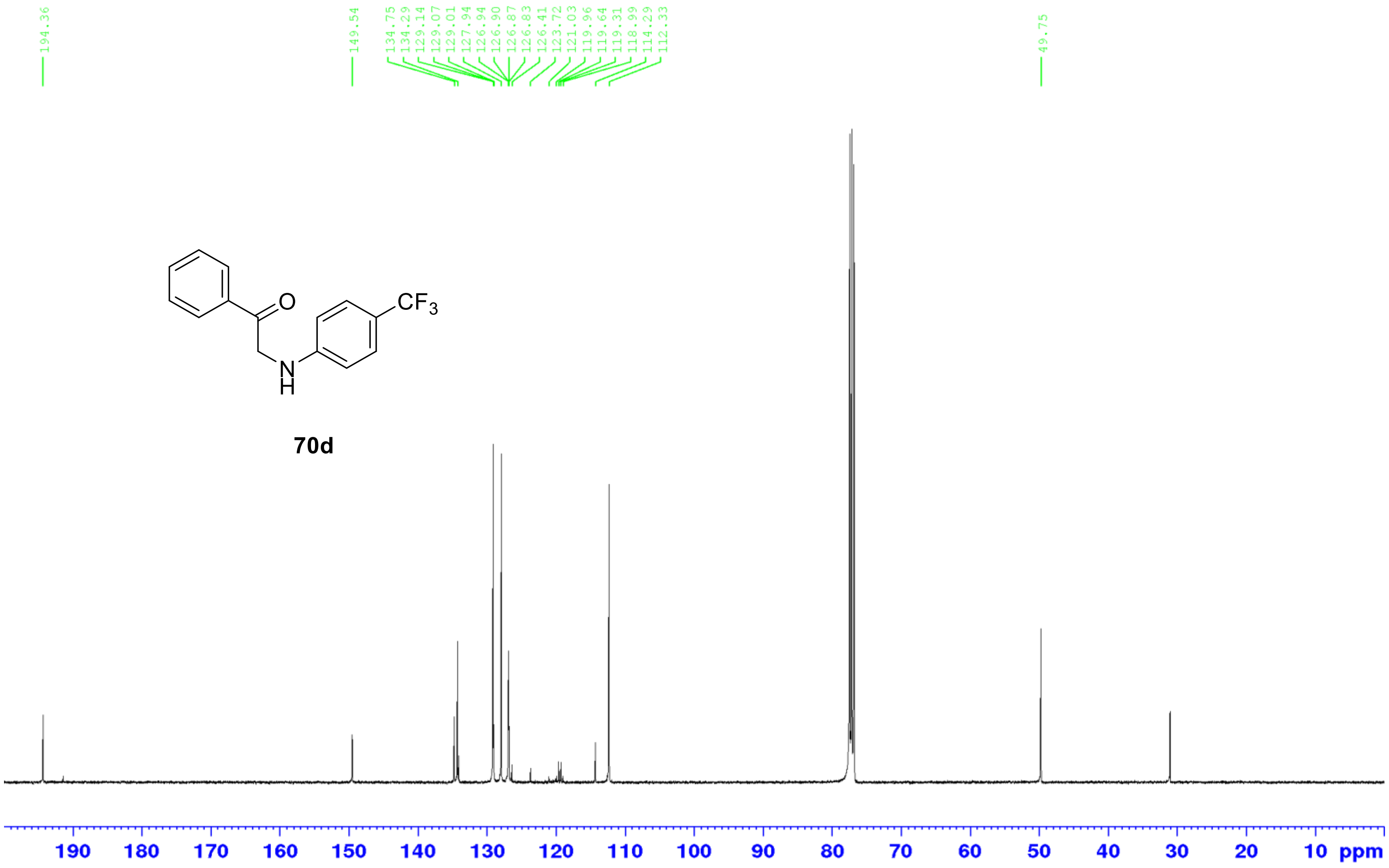
70c

^1H NMR (400 MHz, $CDCl_3$)



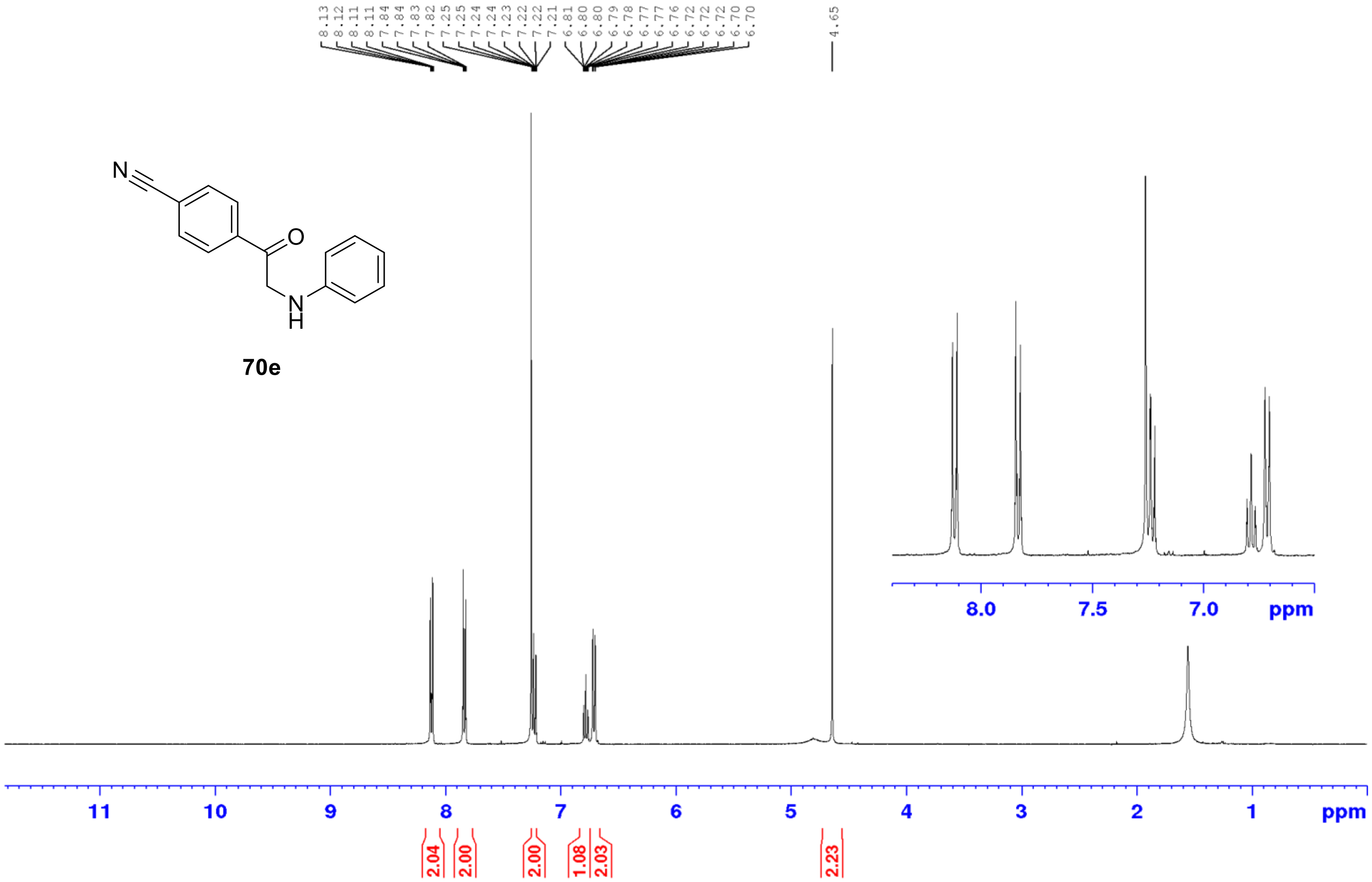
70d

^{13}C NMR (101 MHz, CDCl$_3$)

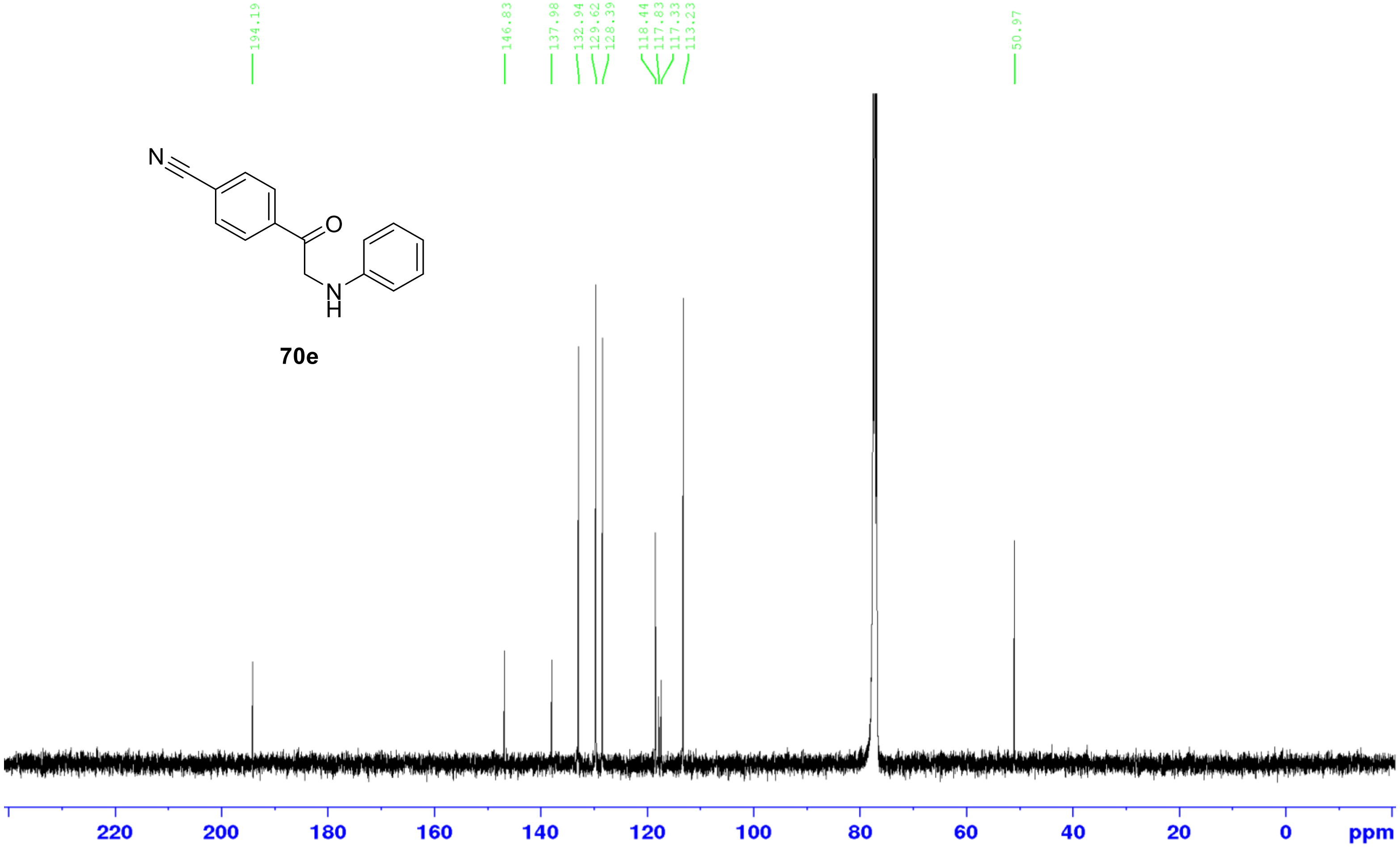


70d

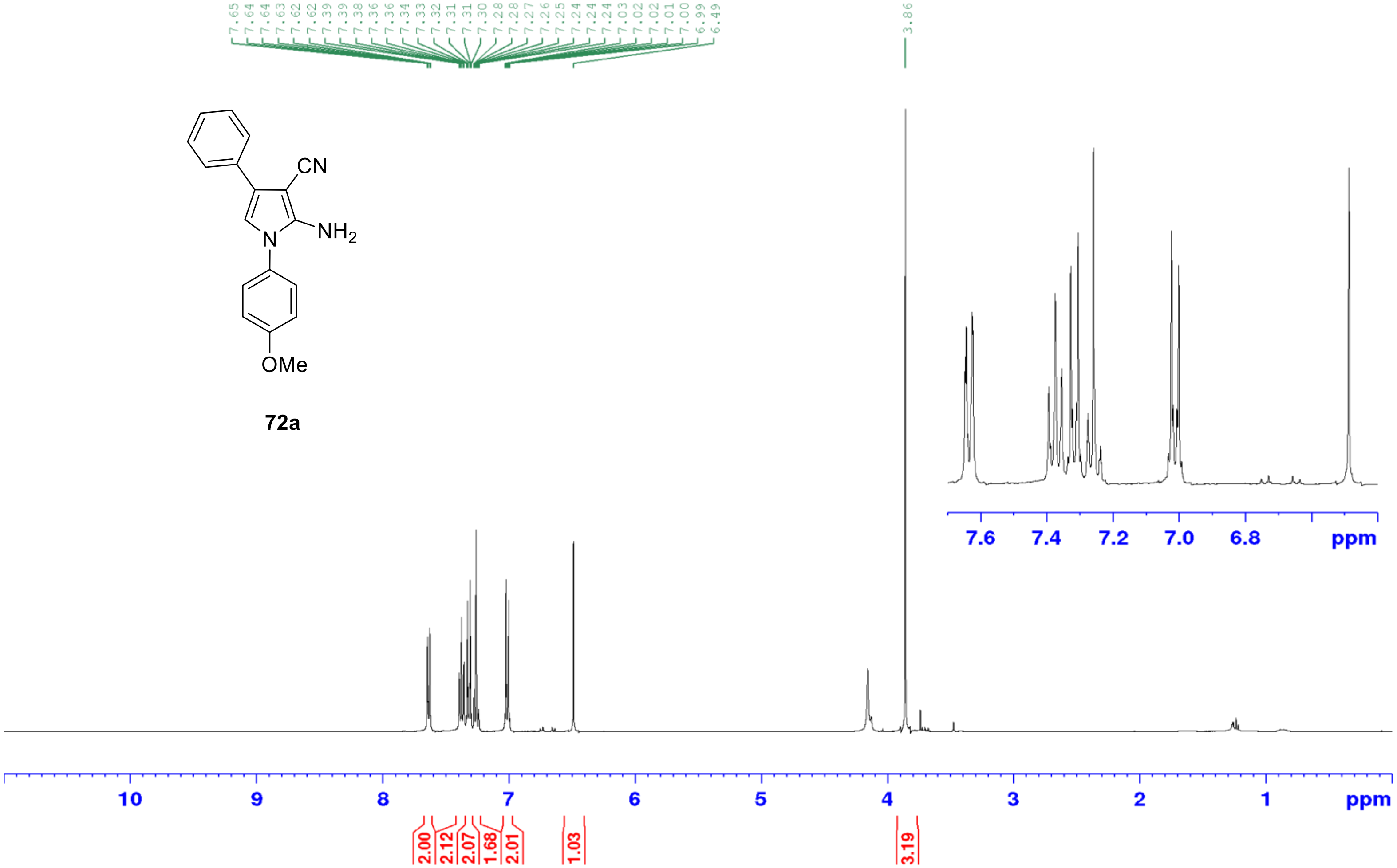
^{1}H NMR (400 MHz, CDCl$_3$)



^{13}C NMR (101 MHz, $CDCl_3$)

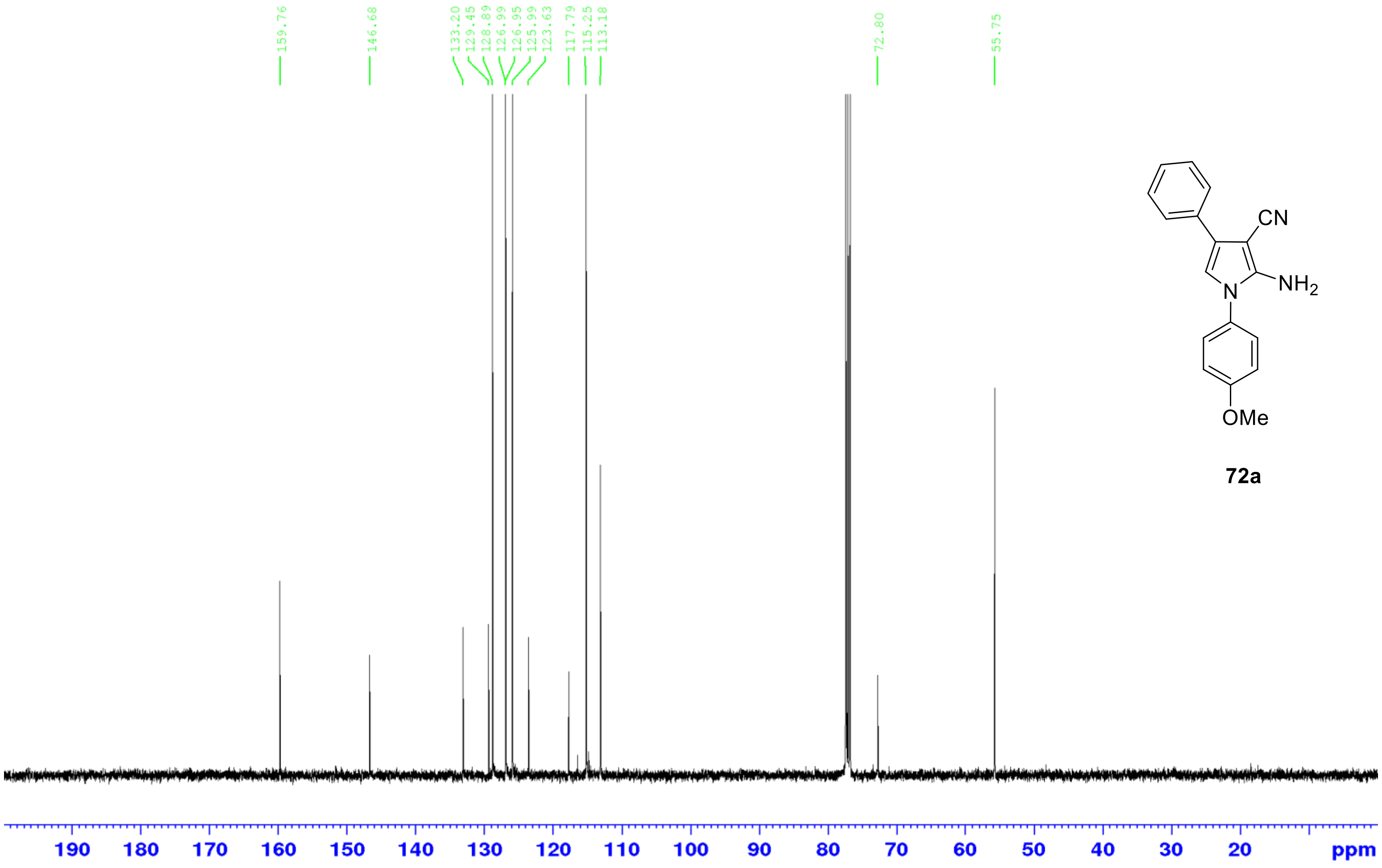


^{1}H NMR (400 MHz, $CDCl_3$)



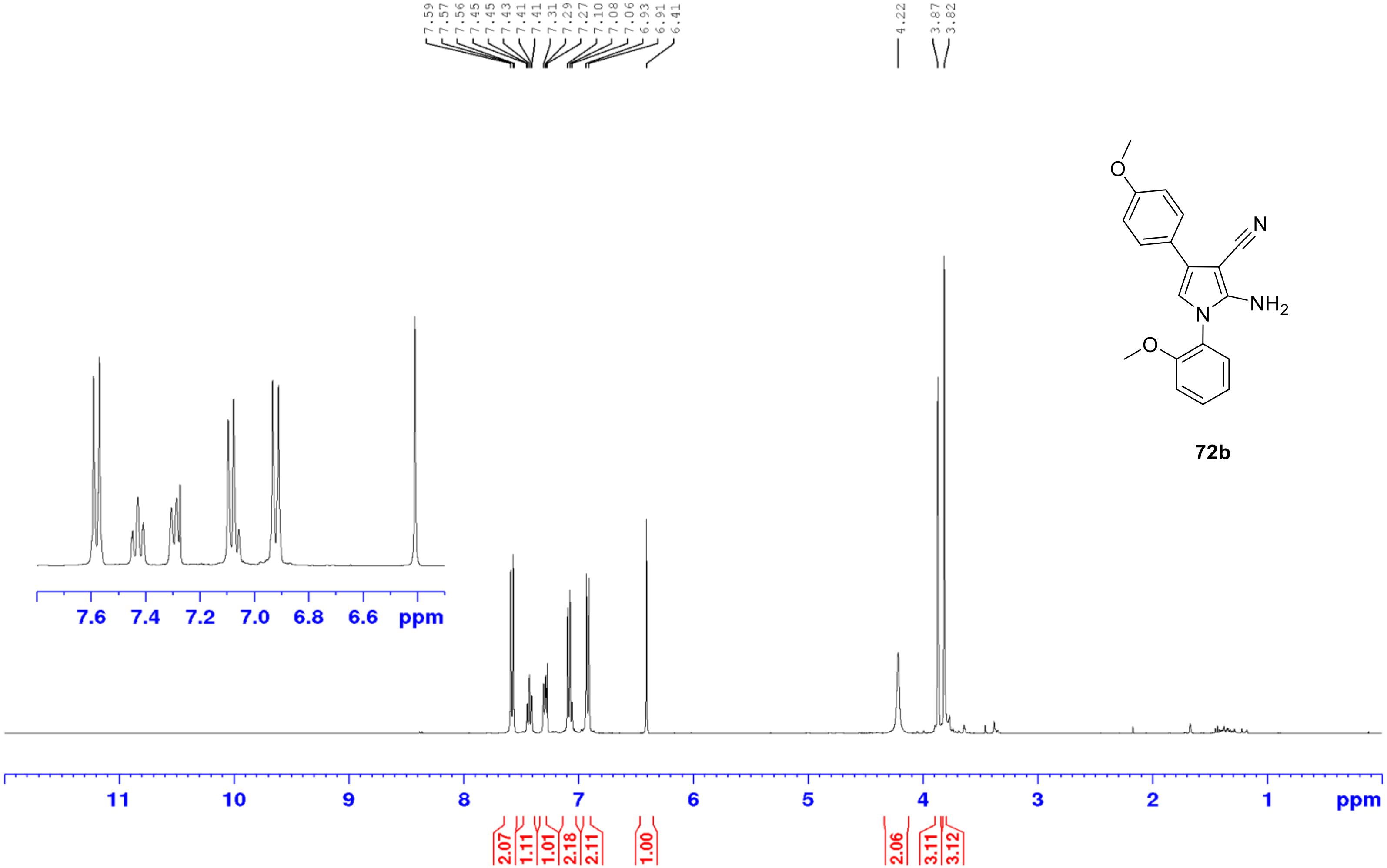
72a

^{13}C NMR (101 MHz, CDCl$_3$)



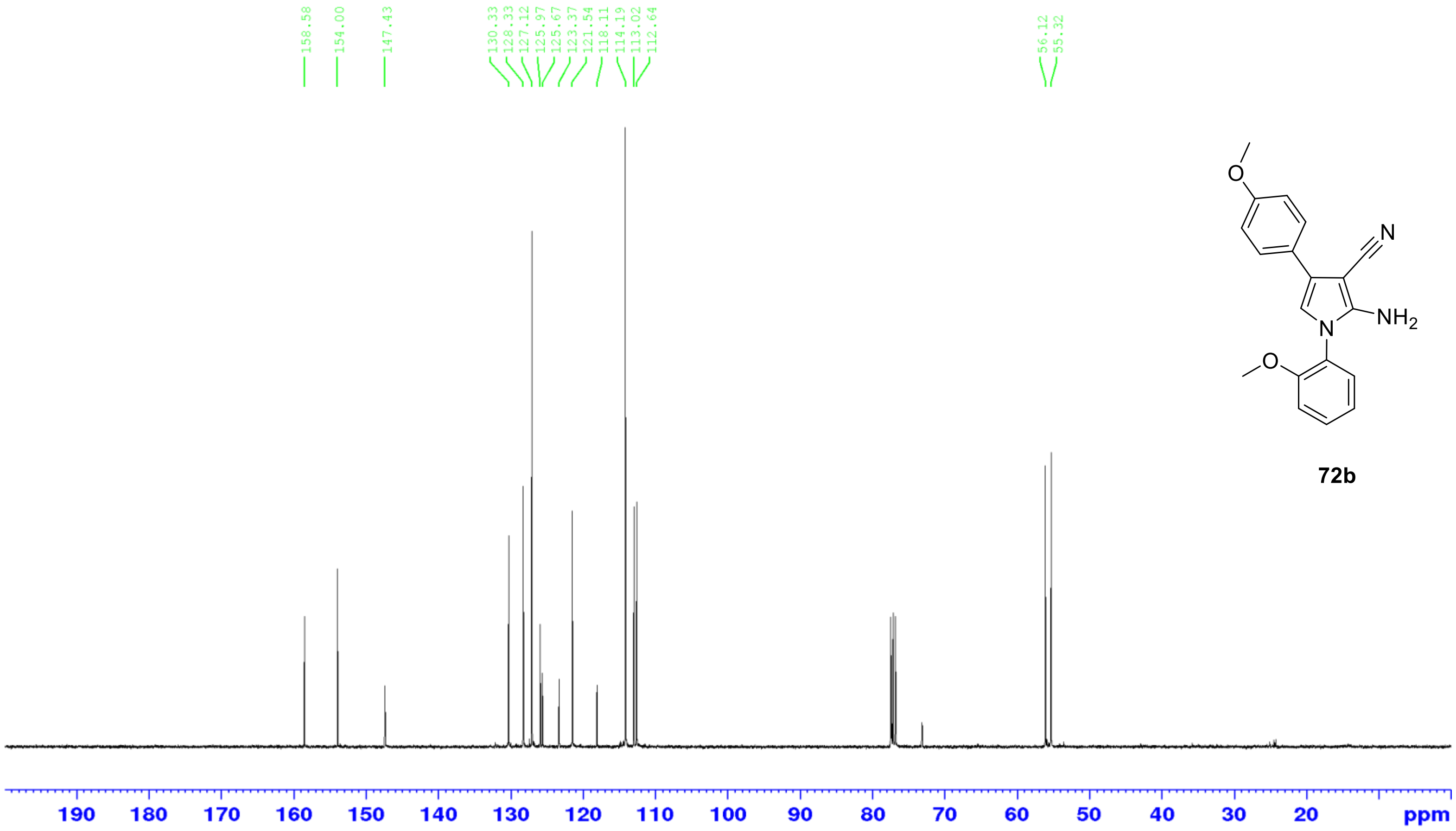
72a

^1H NMR (400 MHz, $CDCl_3$)



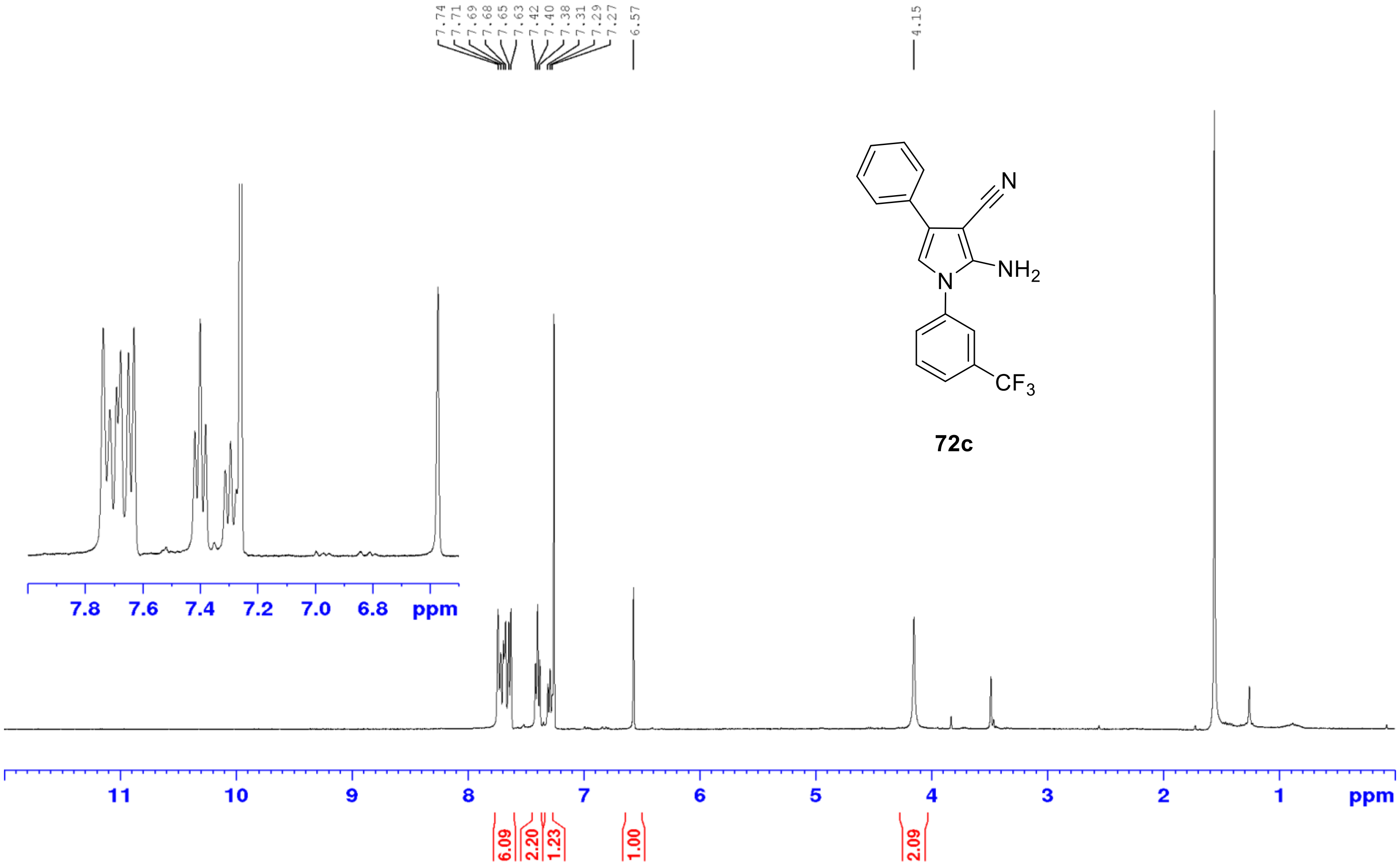
72b

^{13}C NMR (101 MHz, $CDCl_3$)



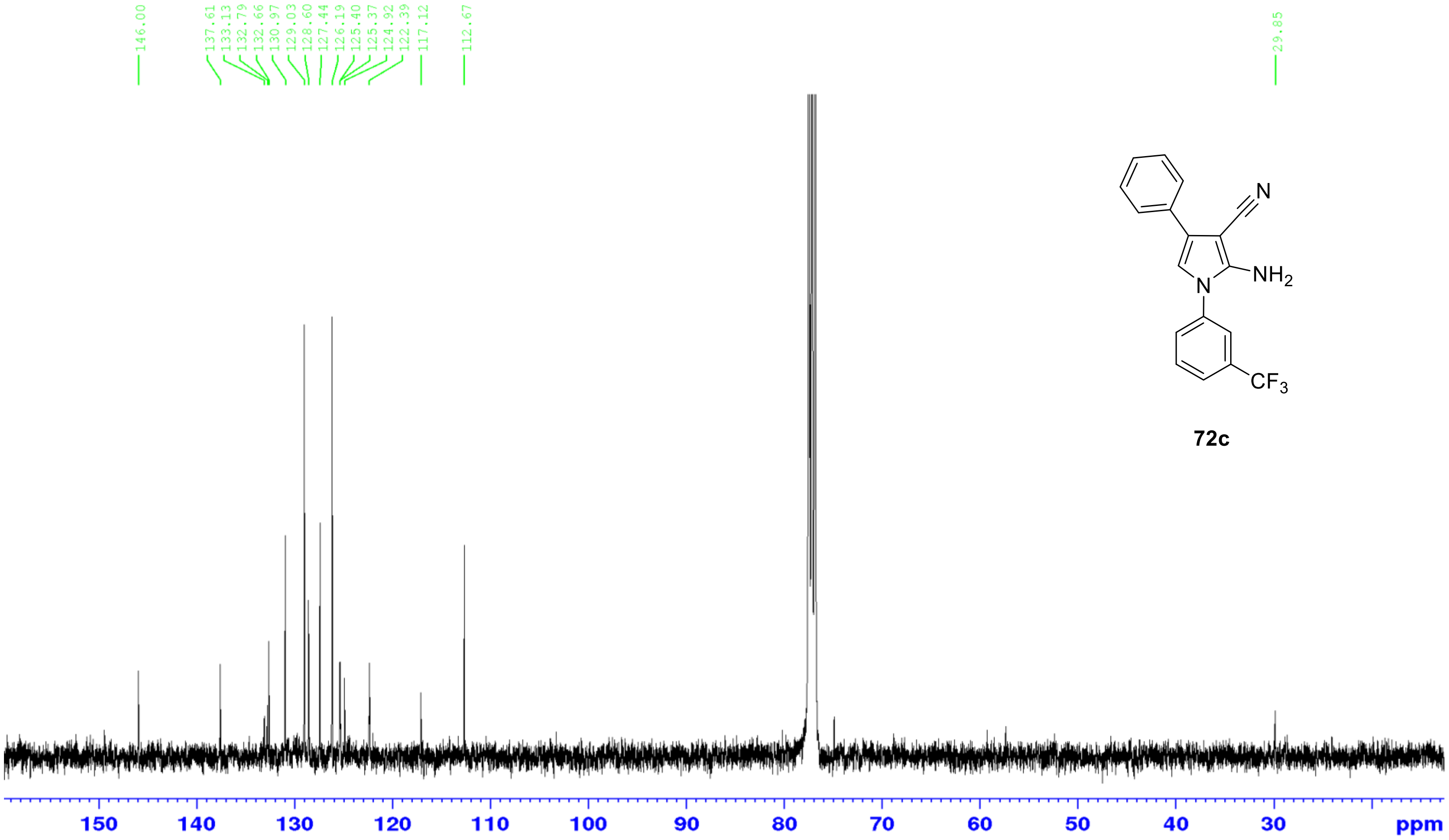
72b

^{1}H NMR (400 MHz, $CDCl_3$)



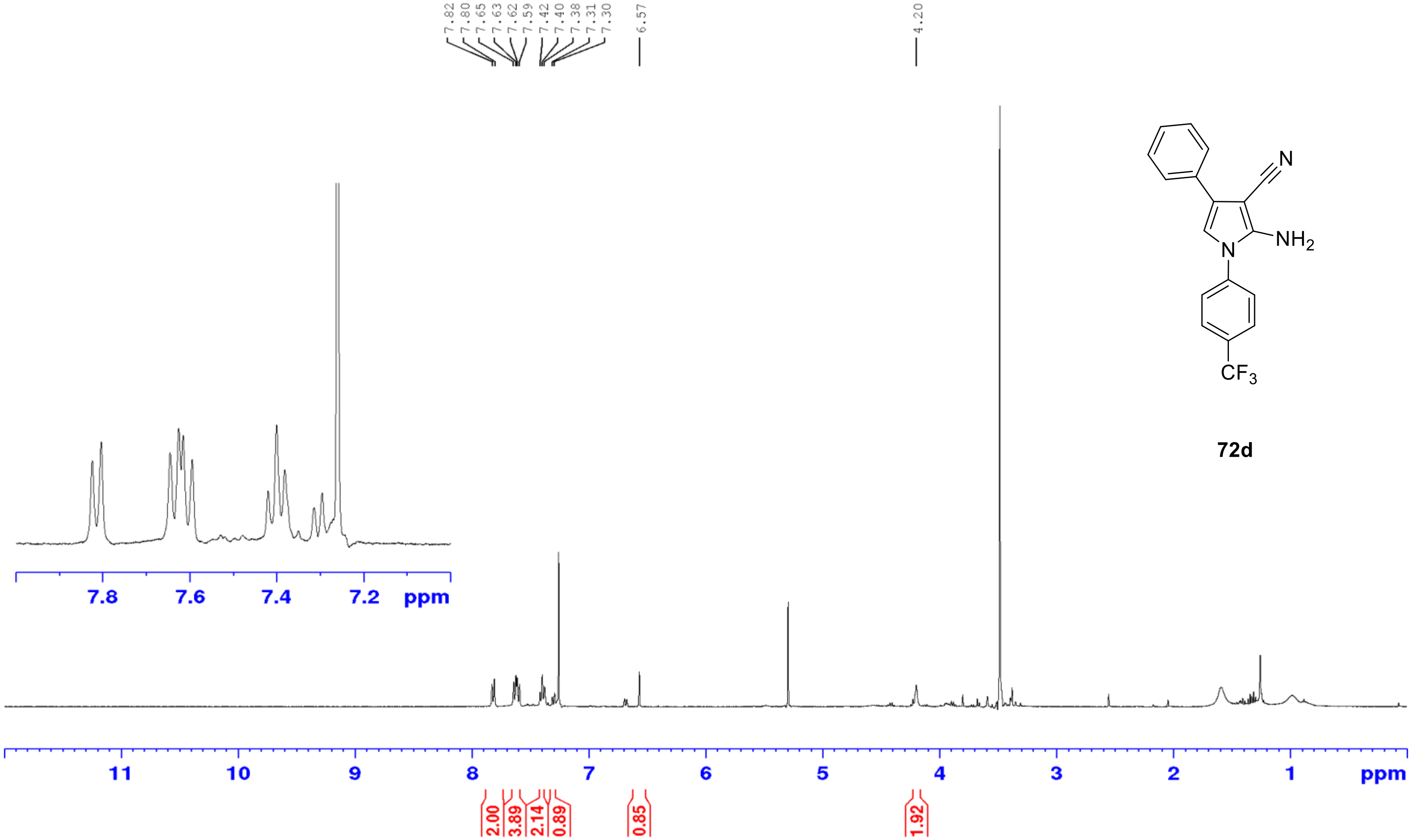
72c

^{13}C NMR (101 MHz, $CDCl_3$)



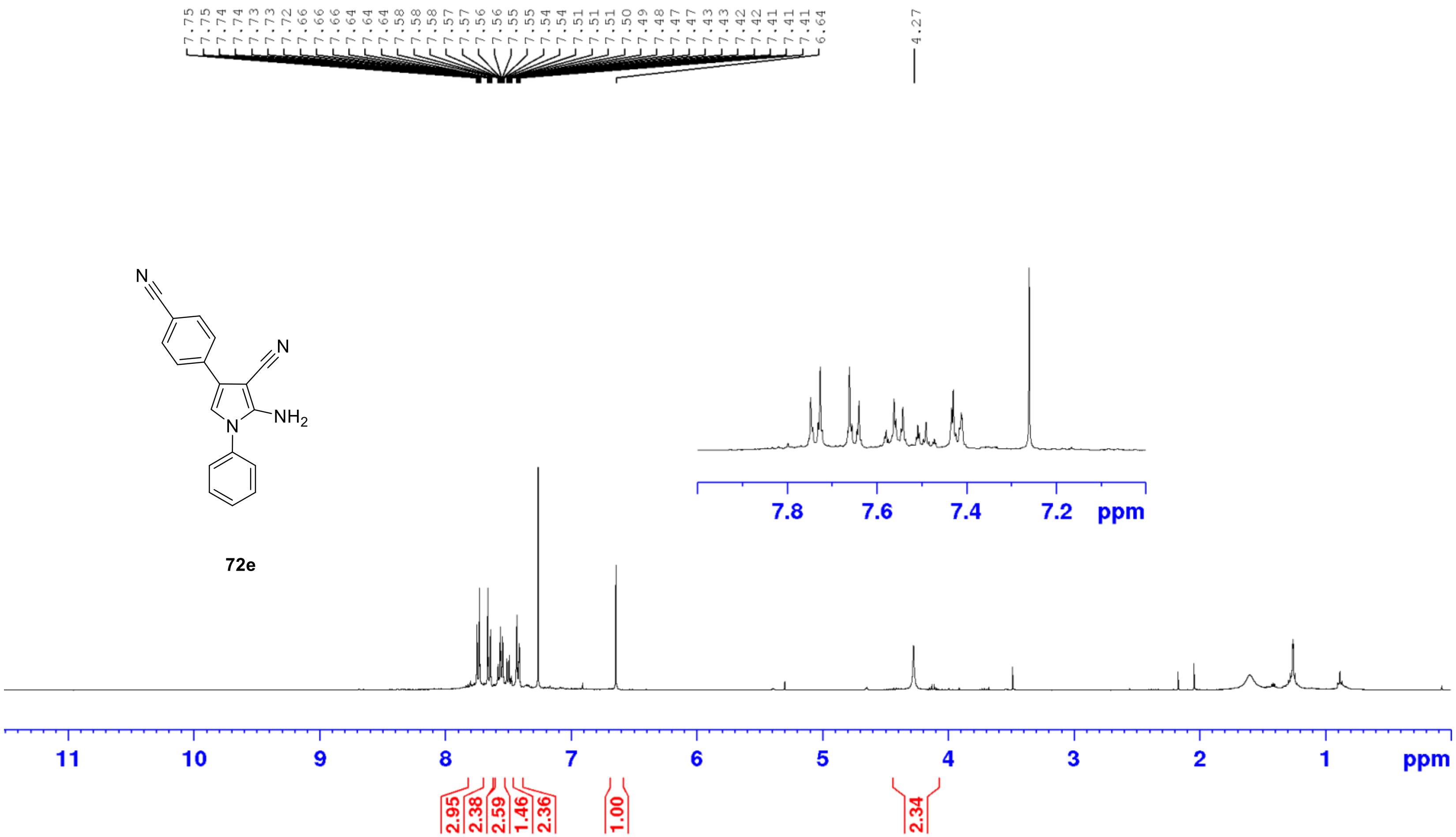
72c

^1H NMR (400 MHz, $CDCl_3$)



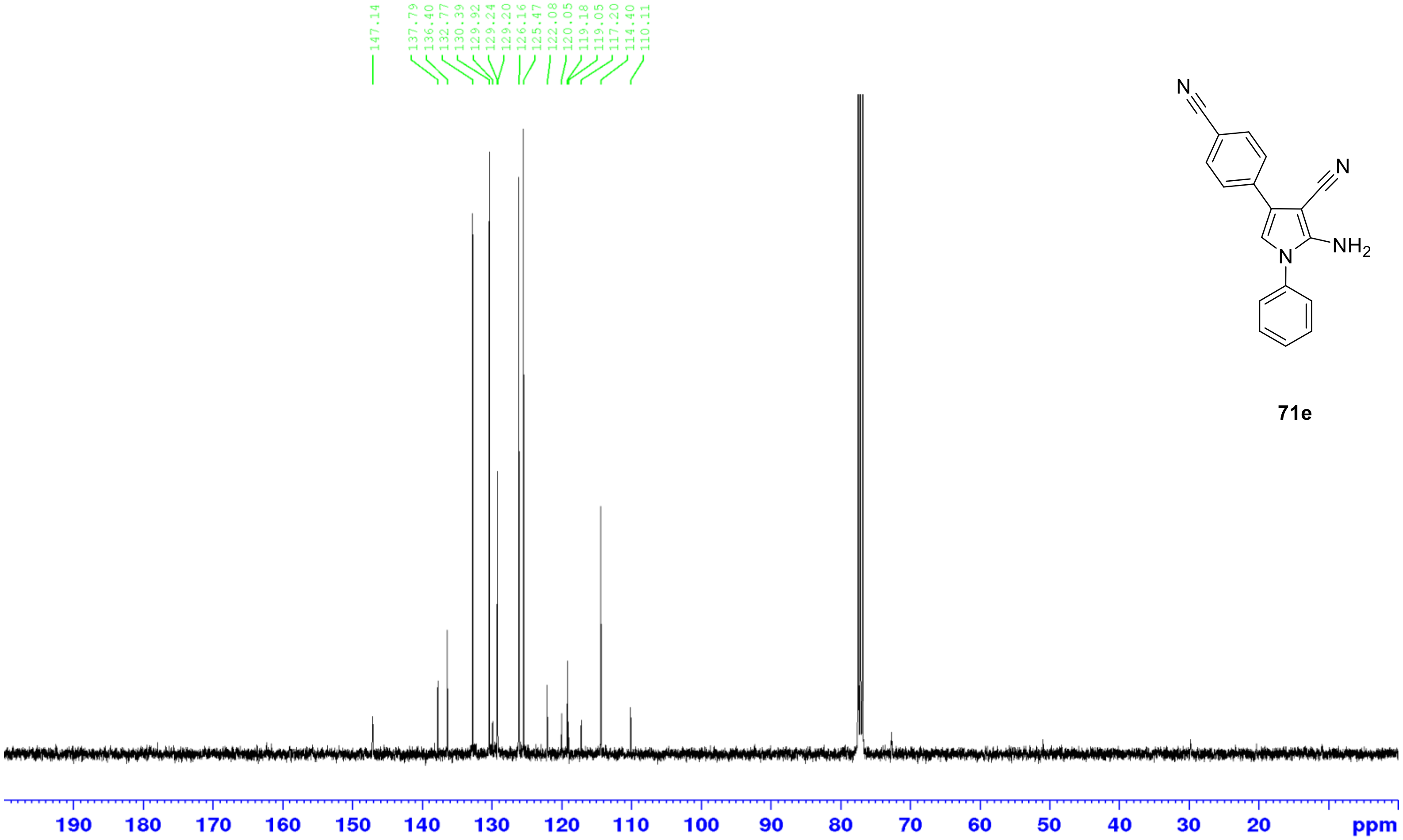
72d

^1H NMR (400 MHz, $CDCl_3$)



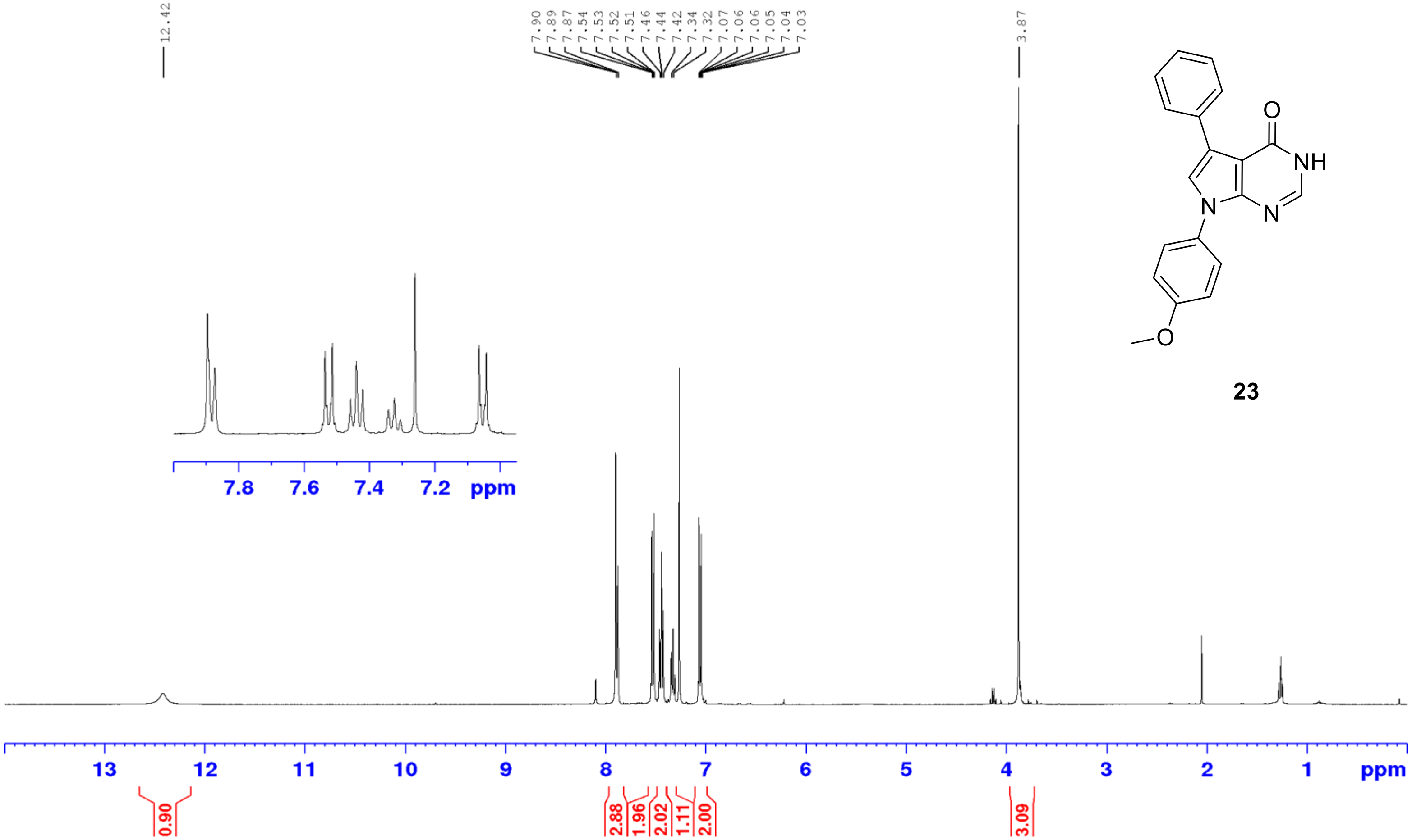
72e

^{13}C NMR (101 MHz, $CDCl_3$)



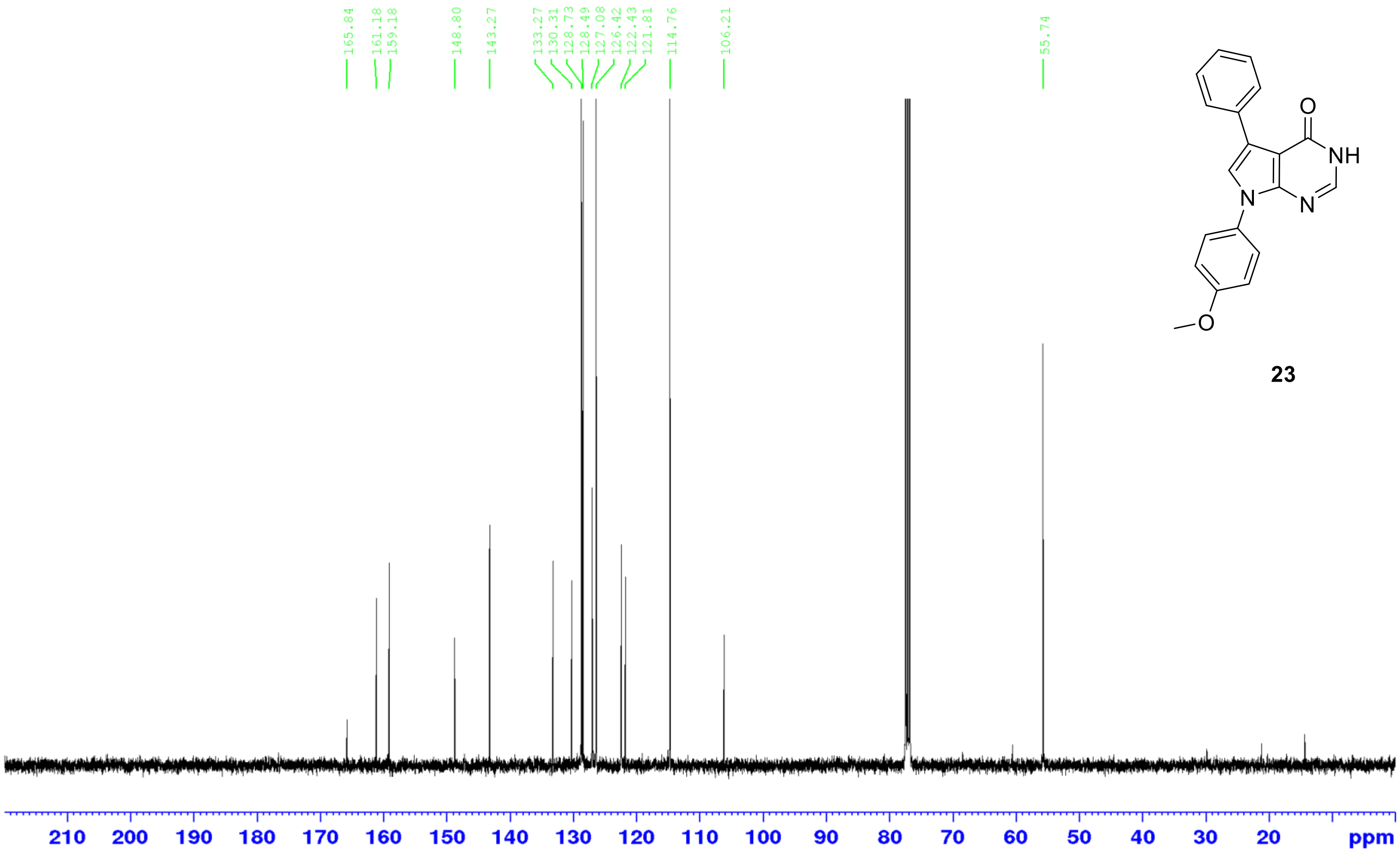
71e

^{1}H NMR (400 MHz, $CDCl_3$)



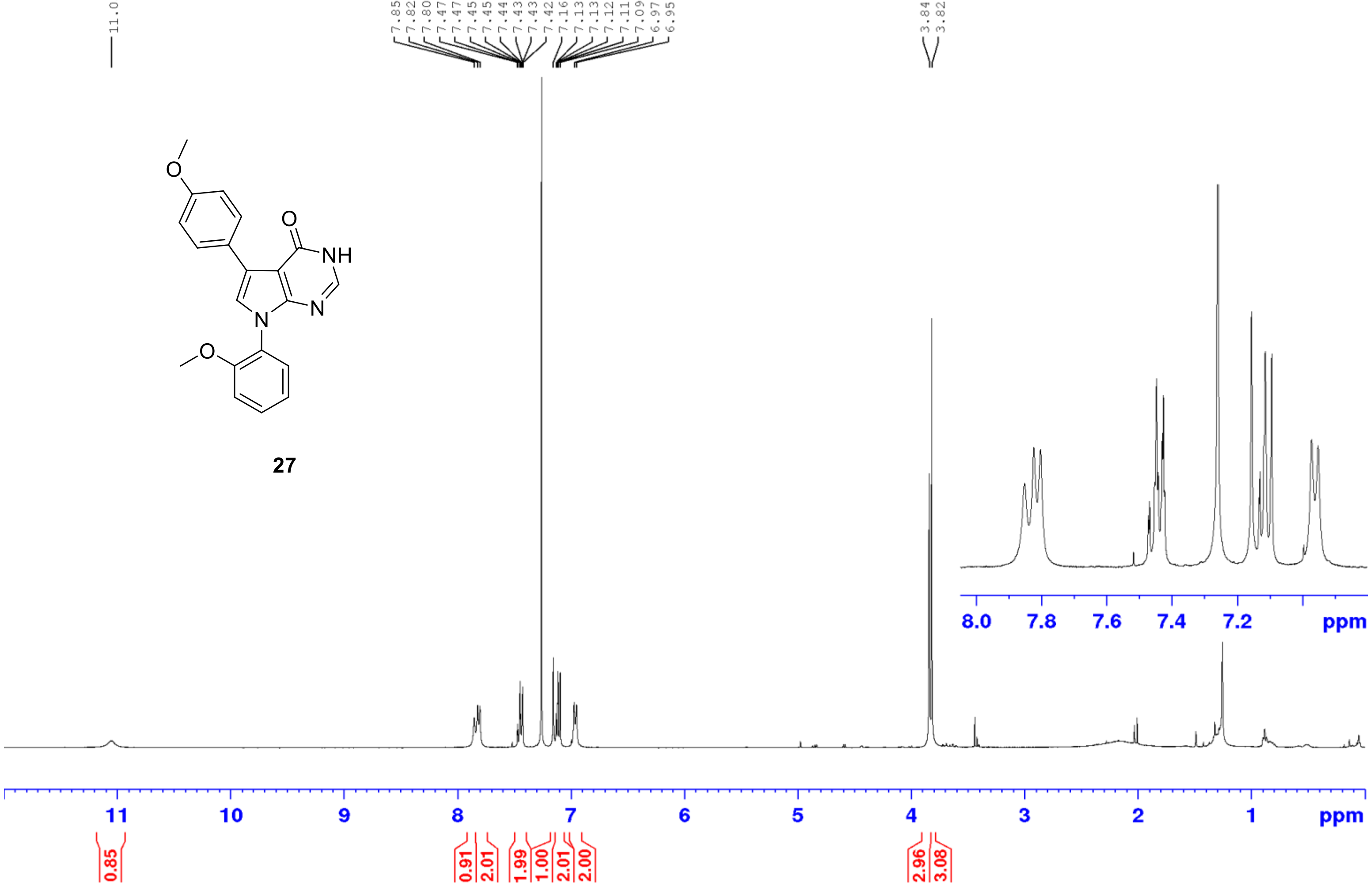
23

^{13}C NMR (101 MHz, $CDCl_3$)

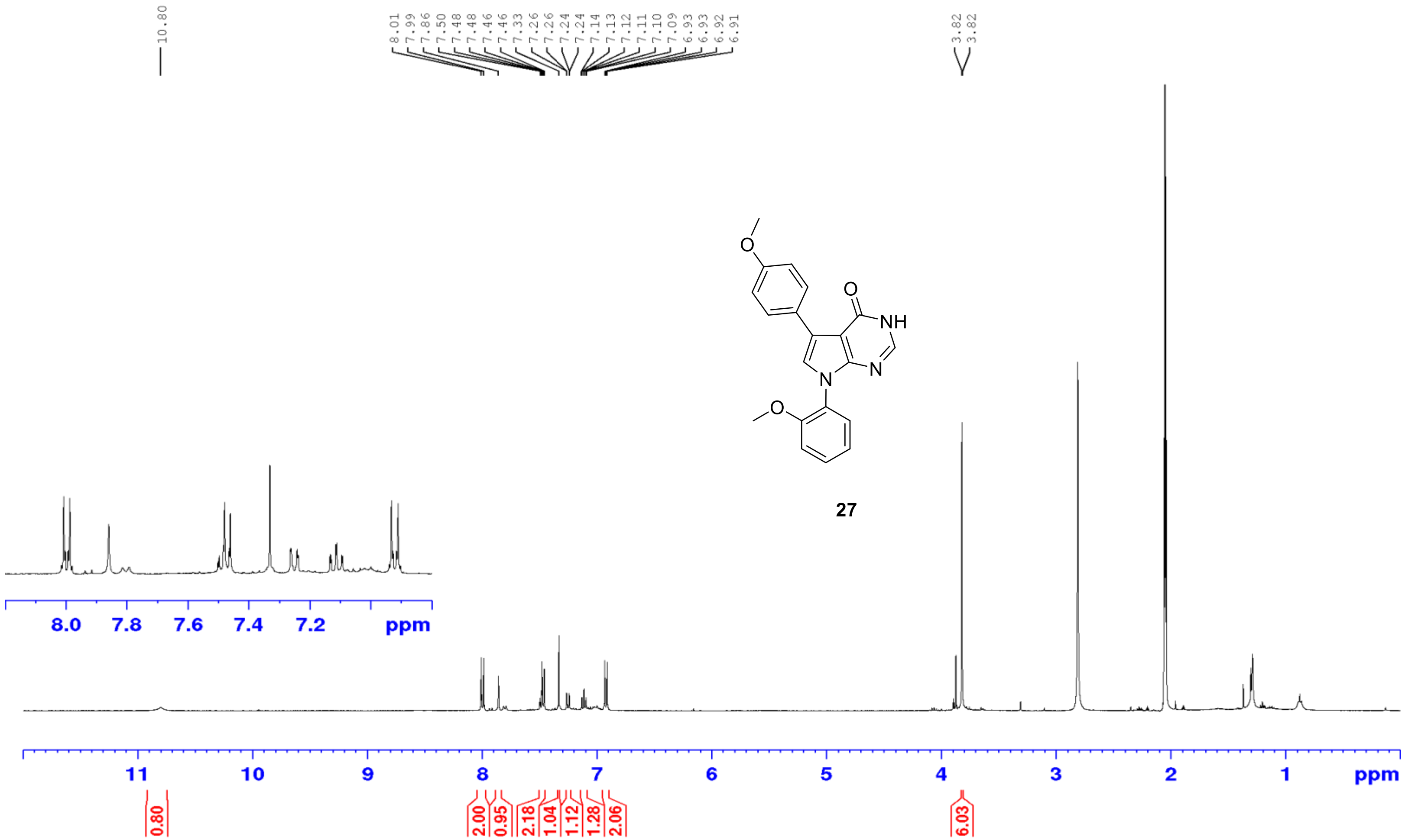


23

^{1}H NMR (400 MHz, $CDCl_3$)

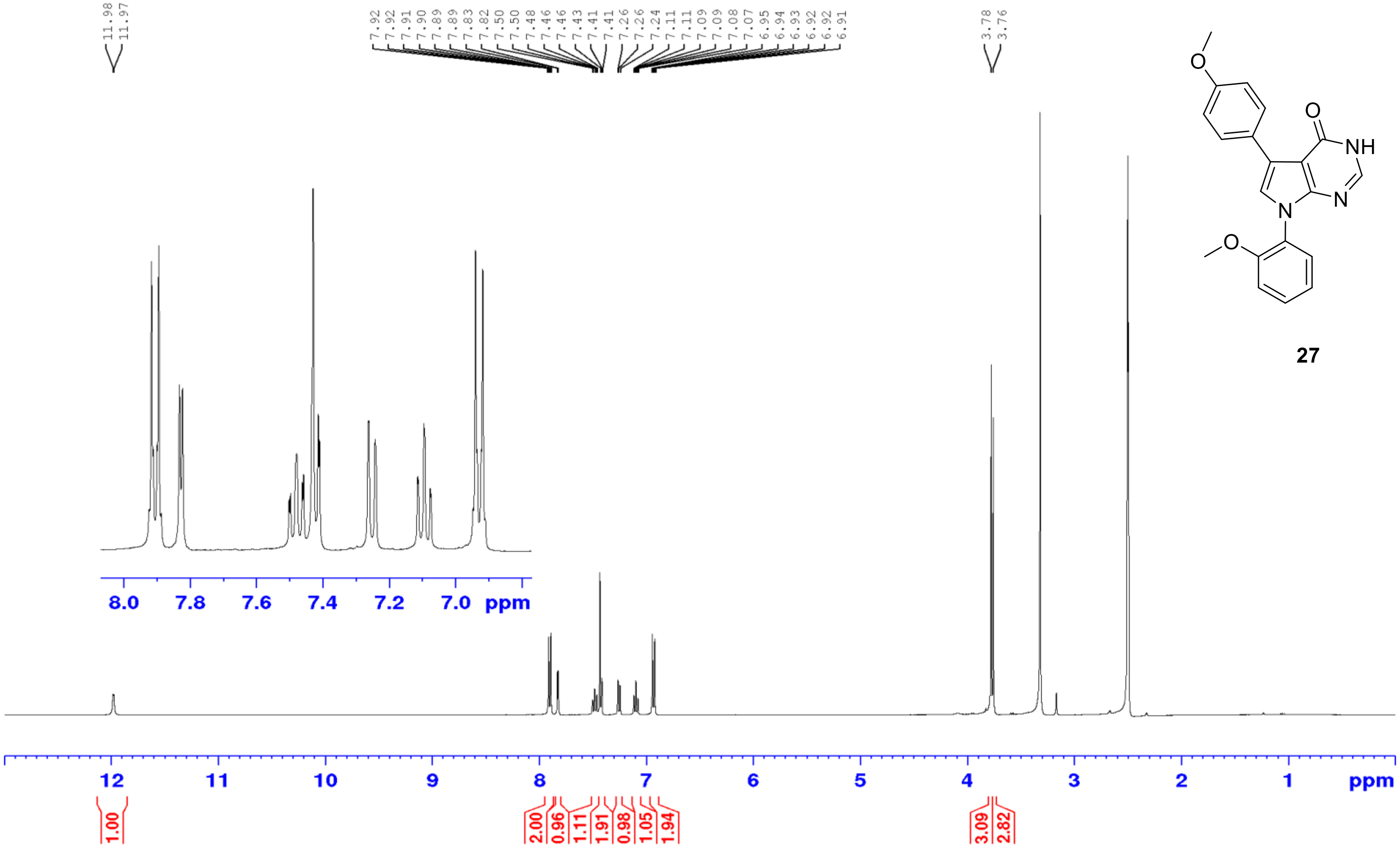


^{1}H NMR (400 MHz, Acetone-d_6)



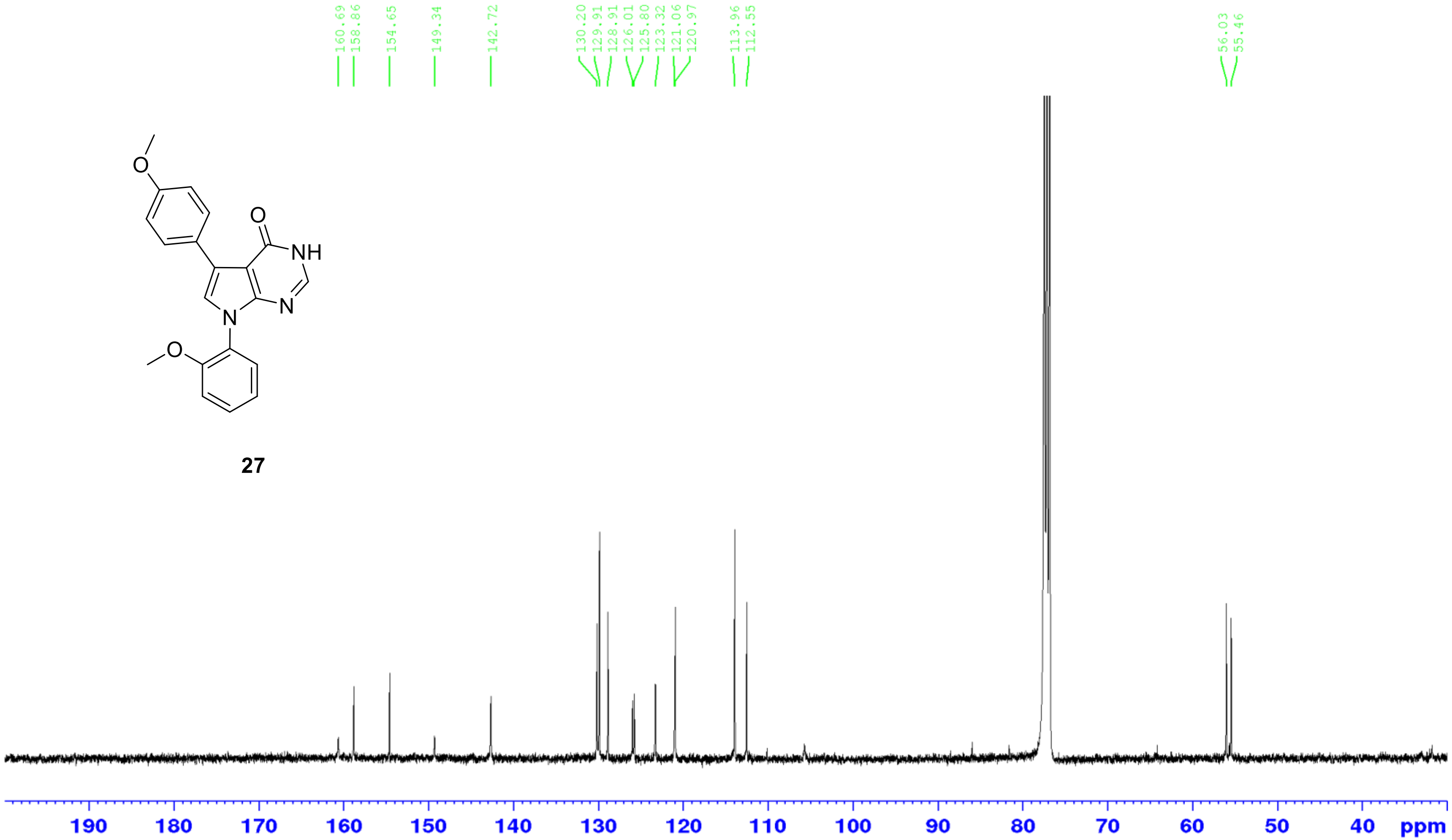
27

^{1}H NMR (400 MHz, DMSO-d$_6$)



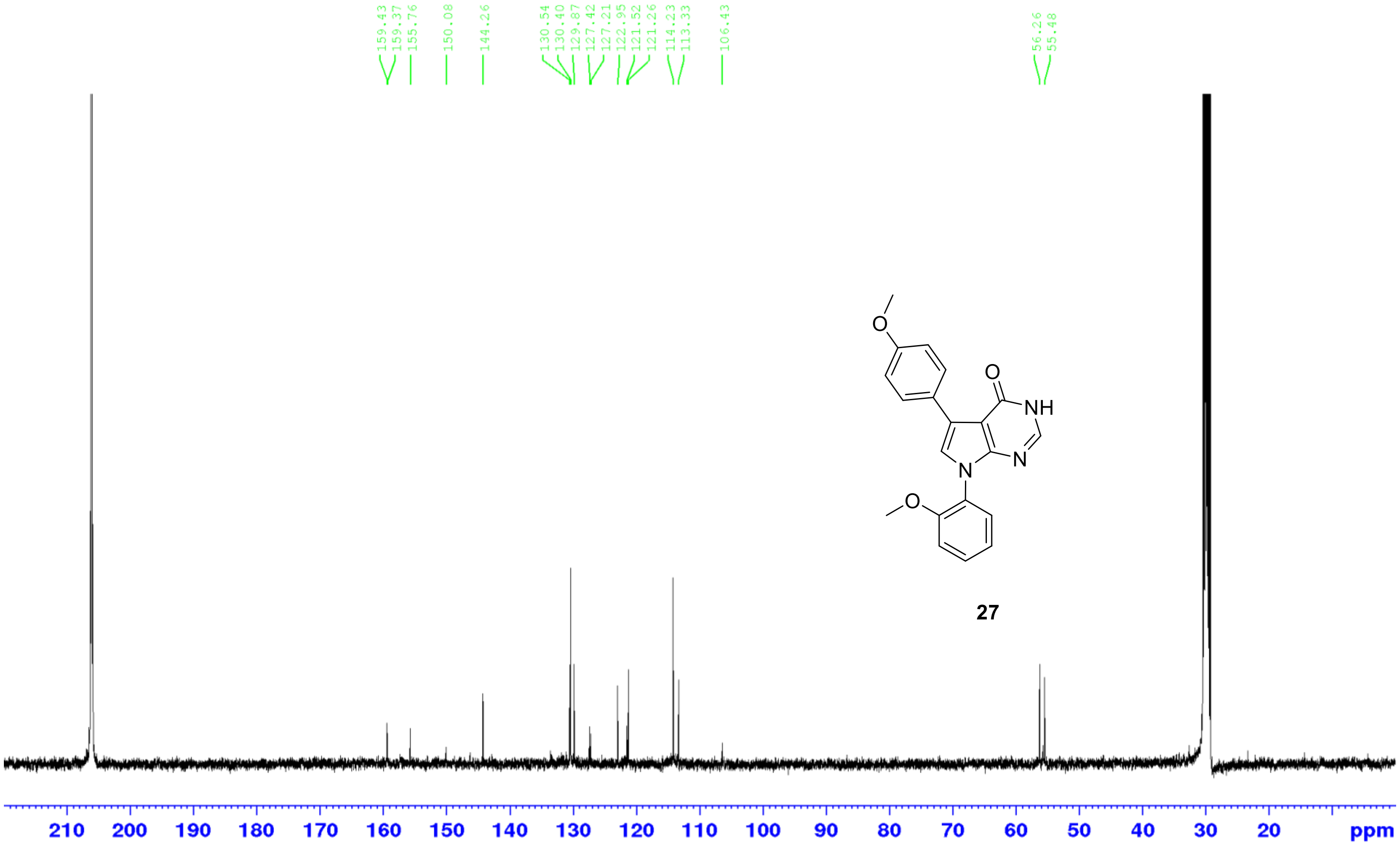
27

^{13}C NMR (101 MHz, $CDCl_3$)



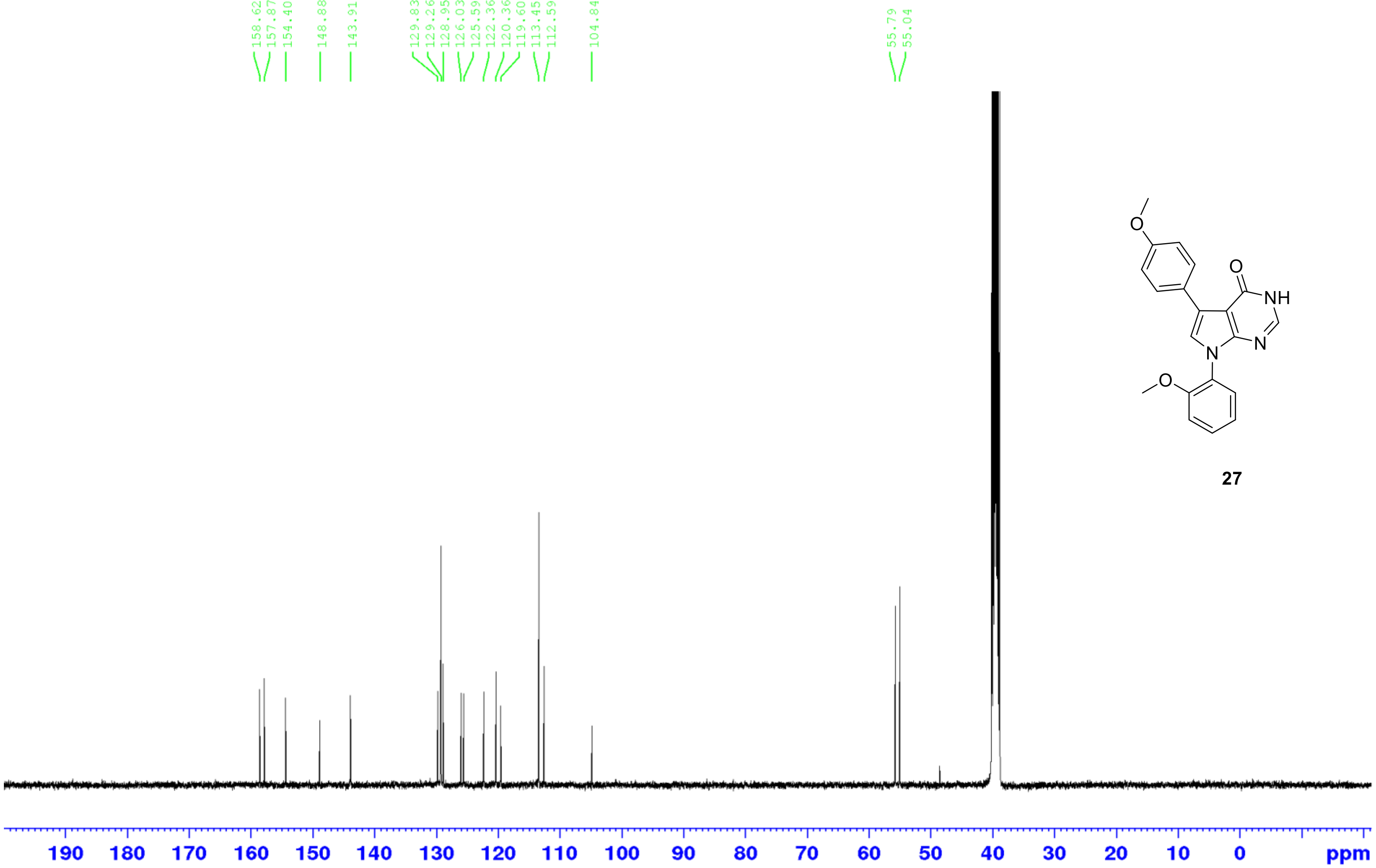
27

^{13}C NMR (101 MHz, Acetone-d$_6$)



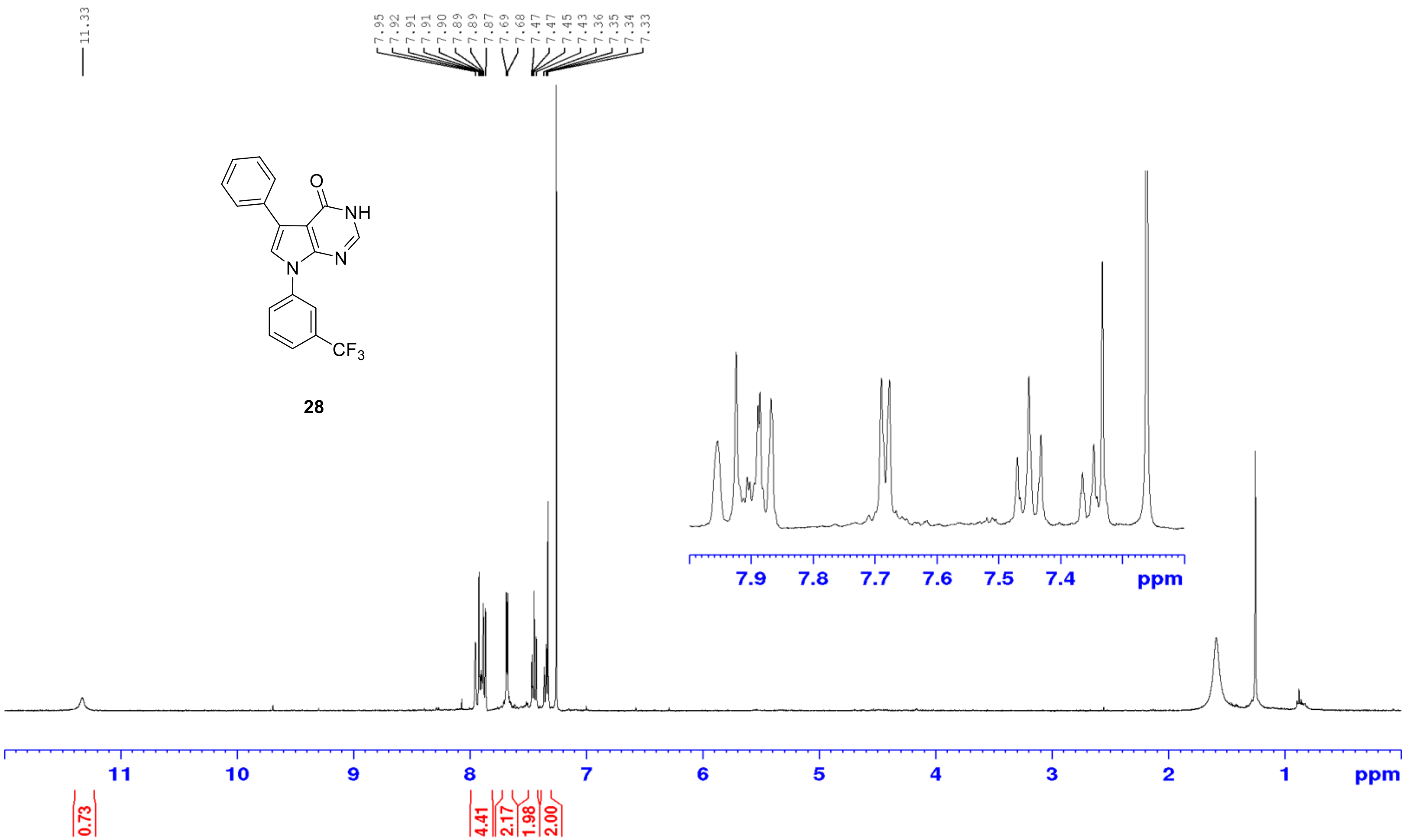
27

^{13}C NMR (101 MHz, DMSO-d_6)

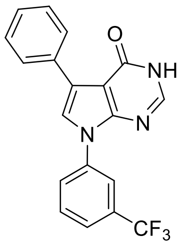


27

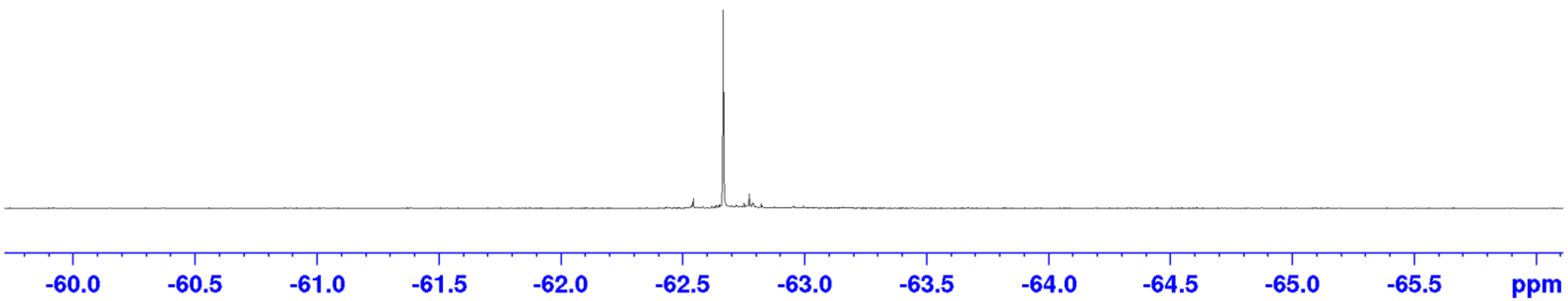
^{1}H NMR (400 MHz, $CDCl_3$)



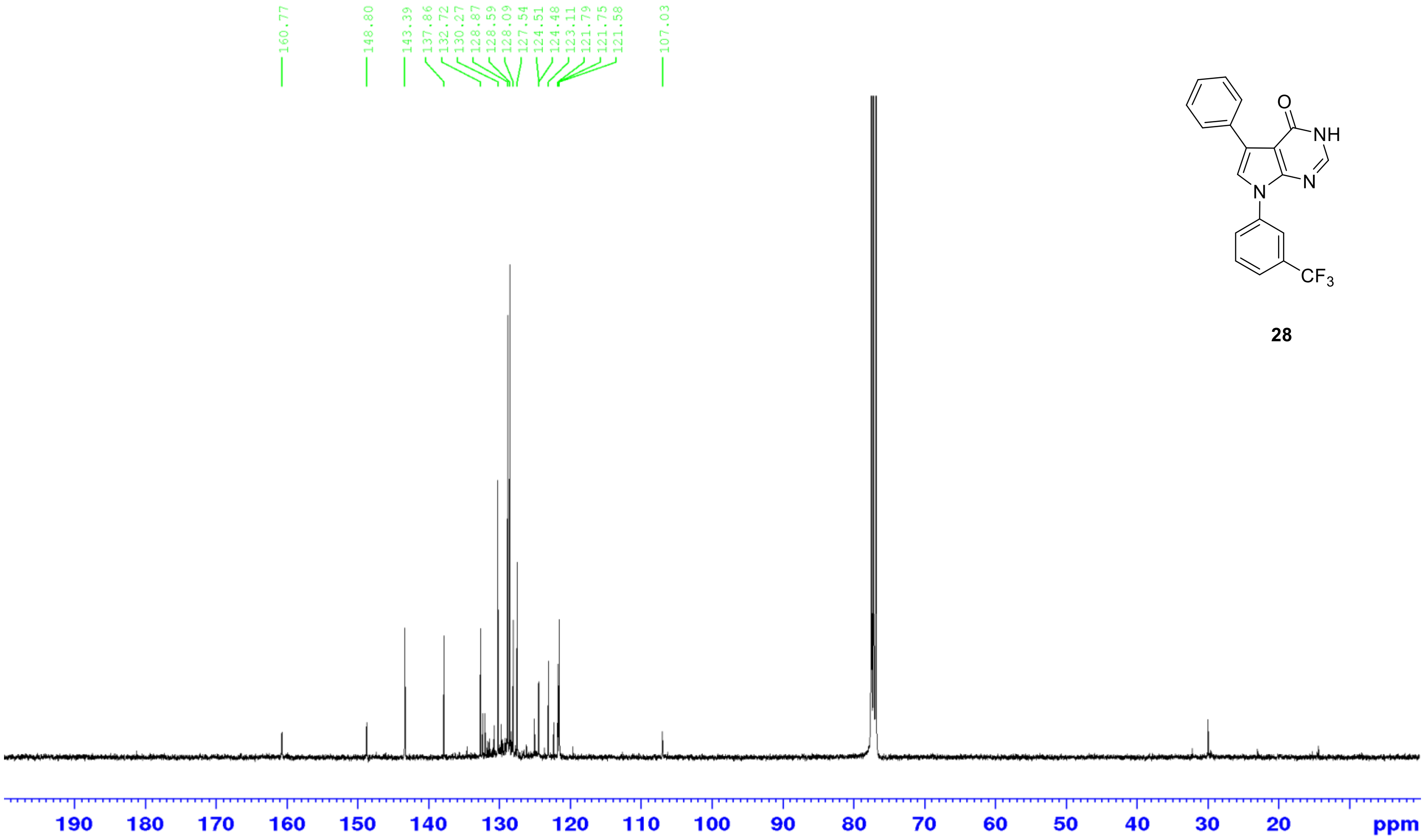
^{19}F NMR (376 MHz, $CDCl_3$)



28

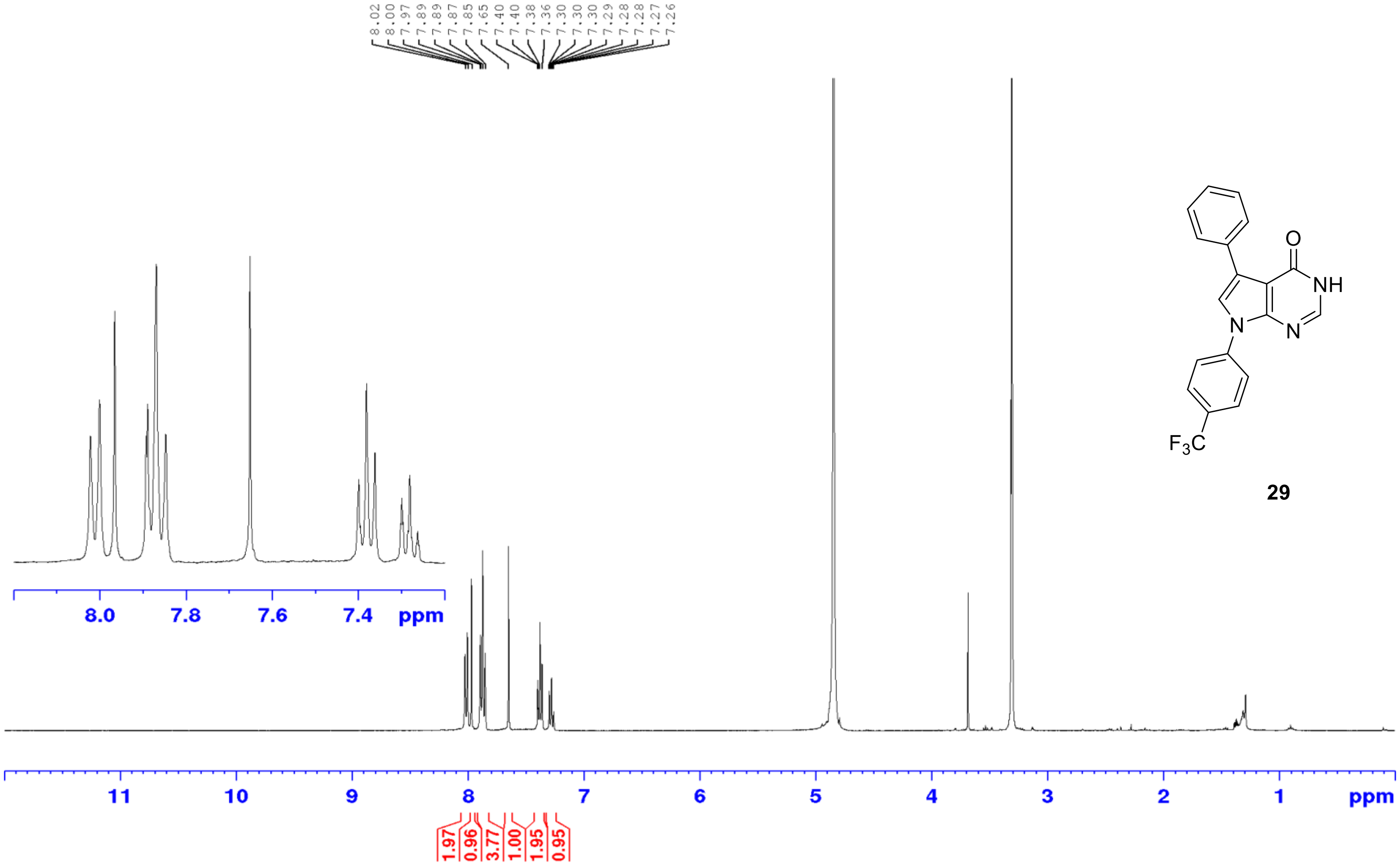


^{13}C NMR (101 MHz, CDCl$_3$)



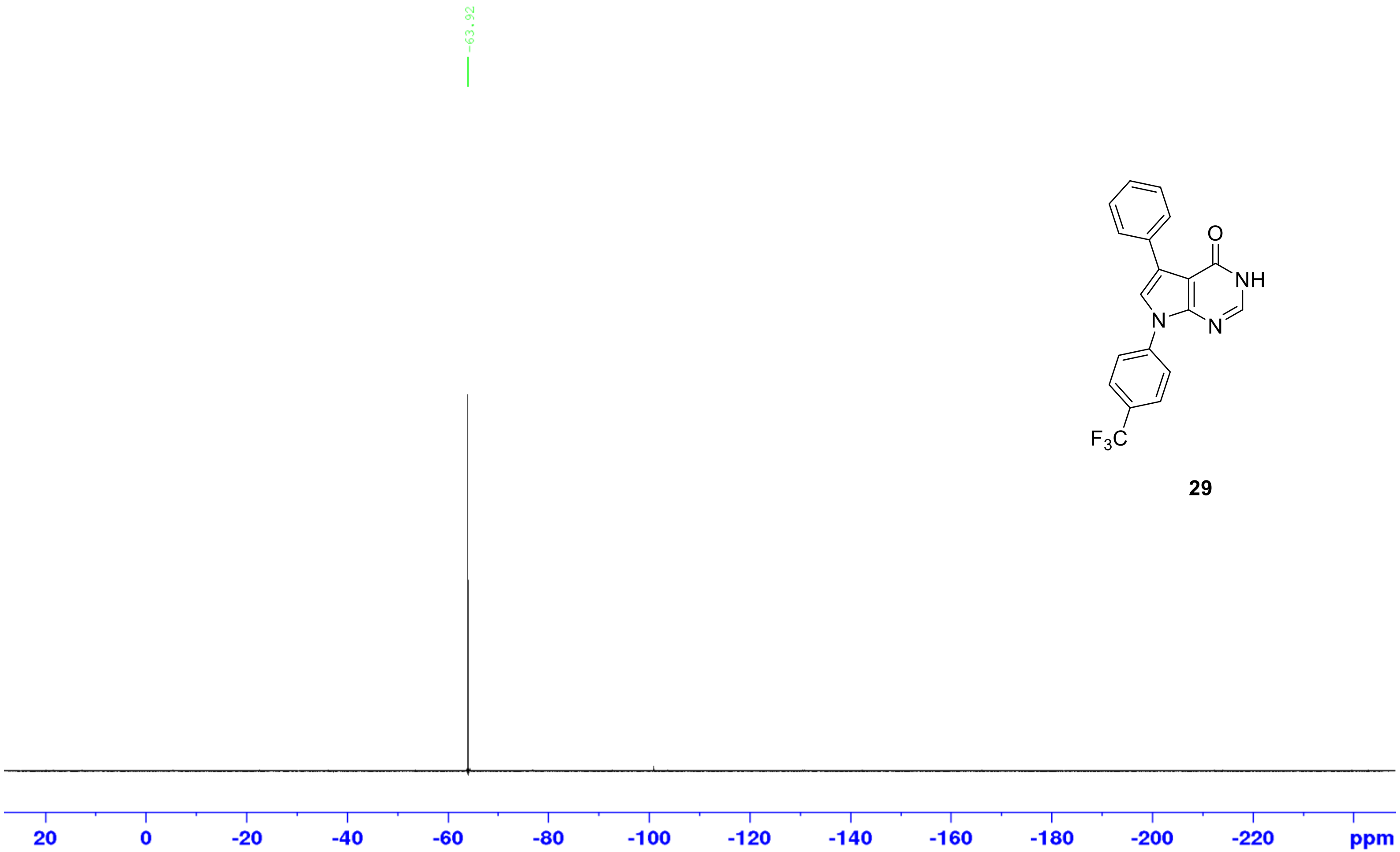
28

^{1}H NMR (400 MHz,MeOD)



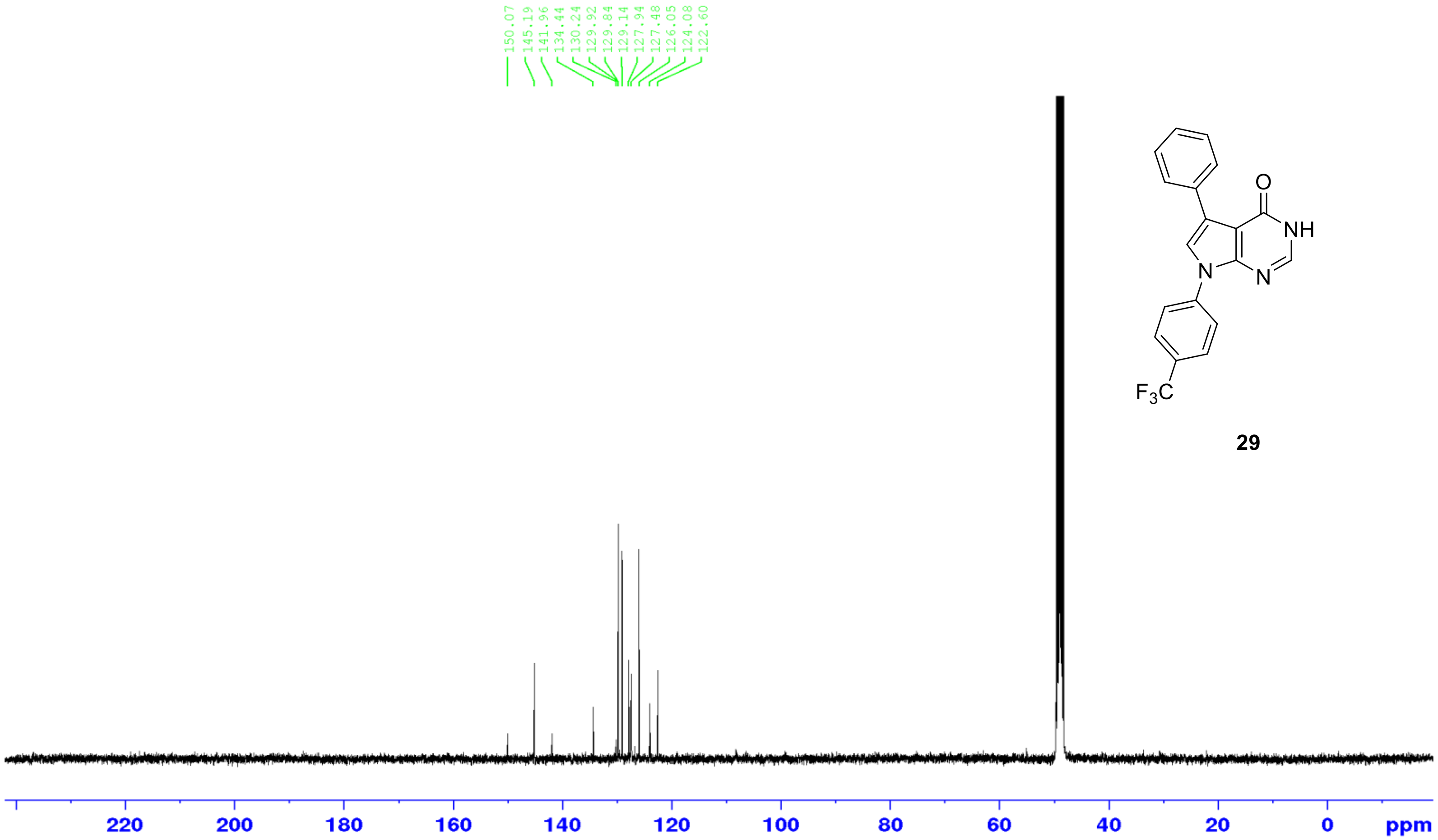
29

^{19}H NMR (376 MHz, CDC_3)



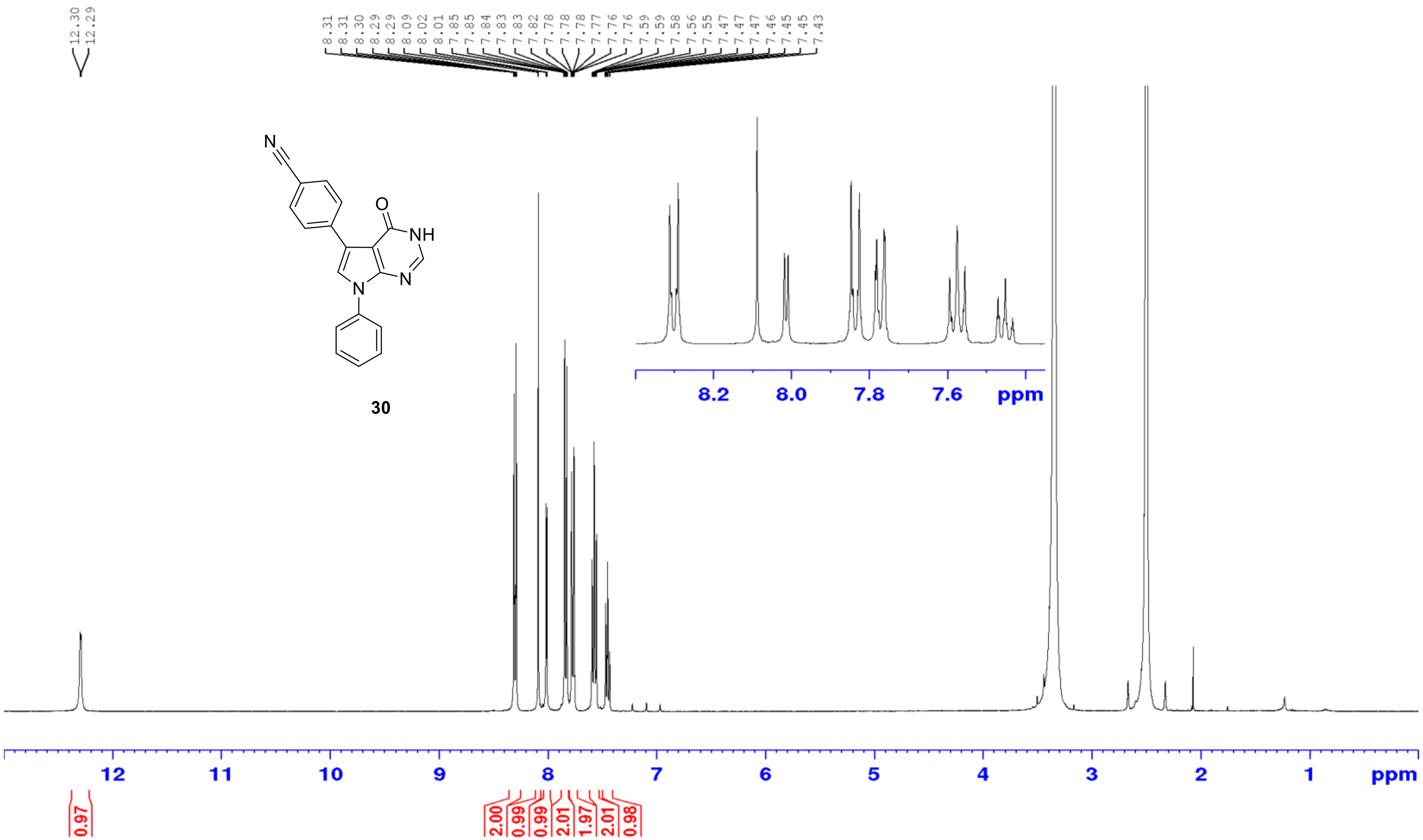
29

^{13}C NMR (101 MHz, MeOD)



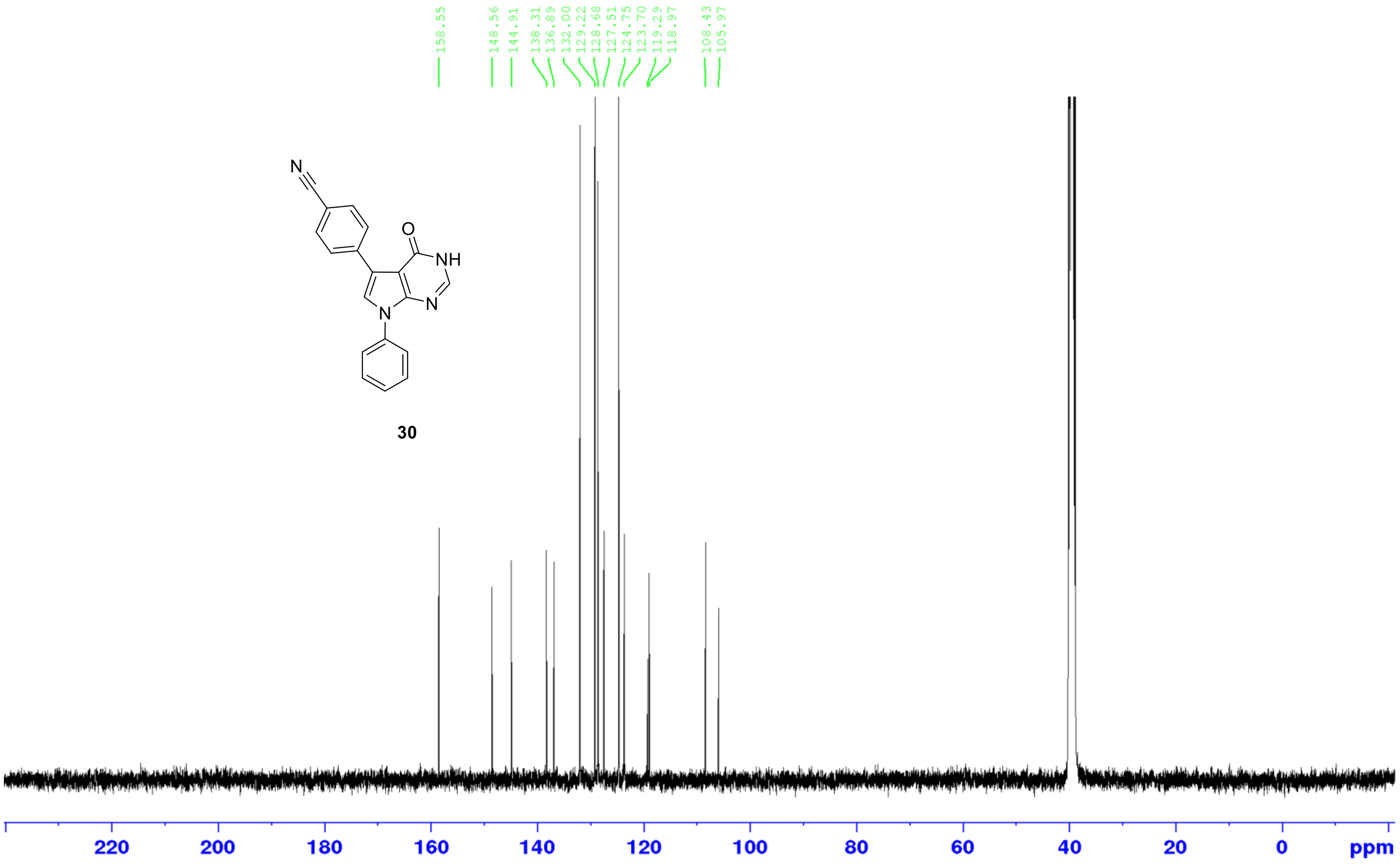
29

^{1}H NMR (400 MHz, DMSO-d_6)



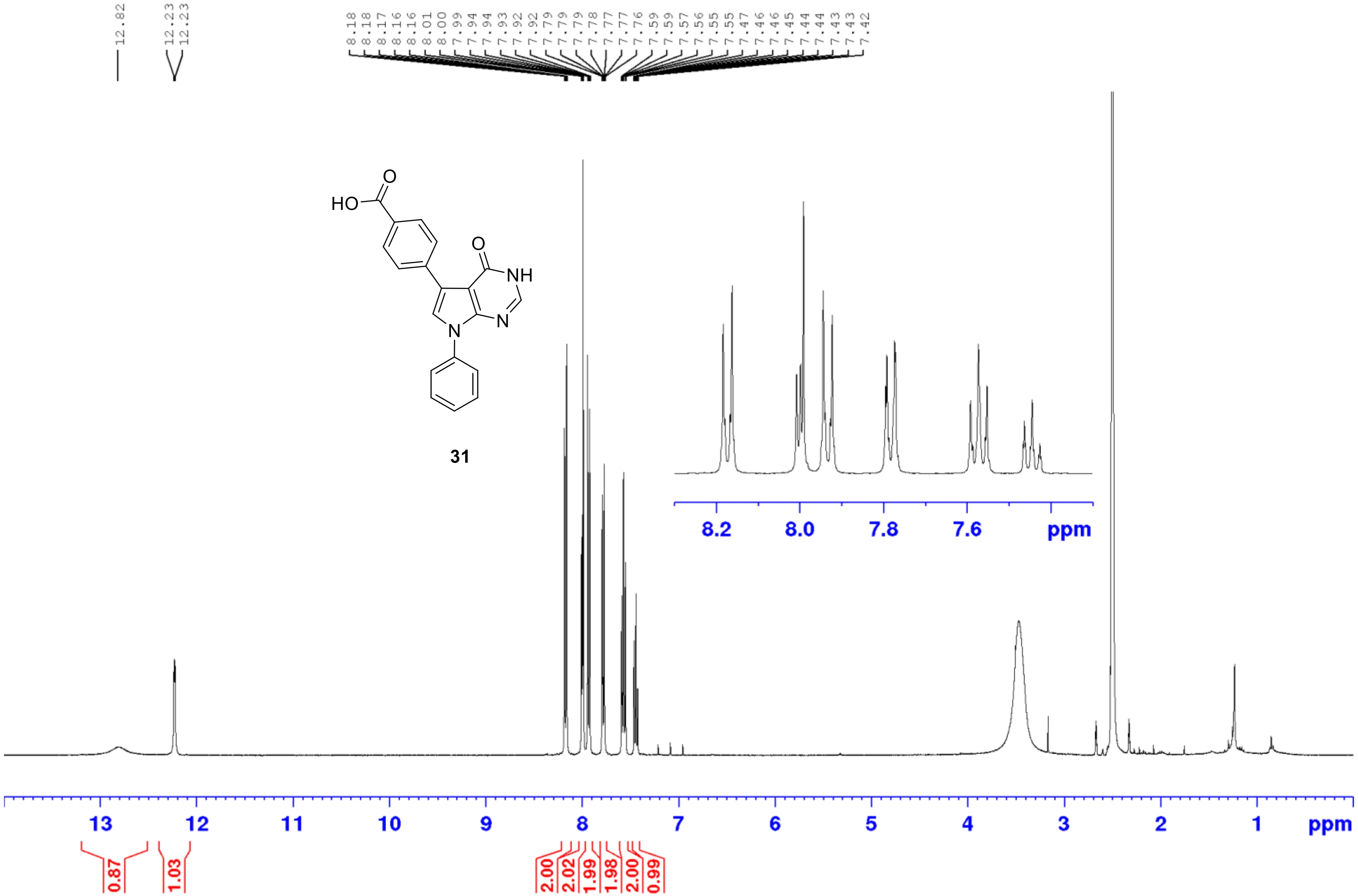
30

^{13}C NMR (101 MHz, DMSO-d_6)



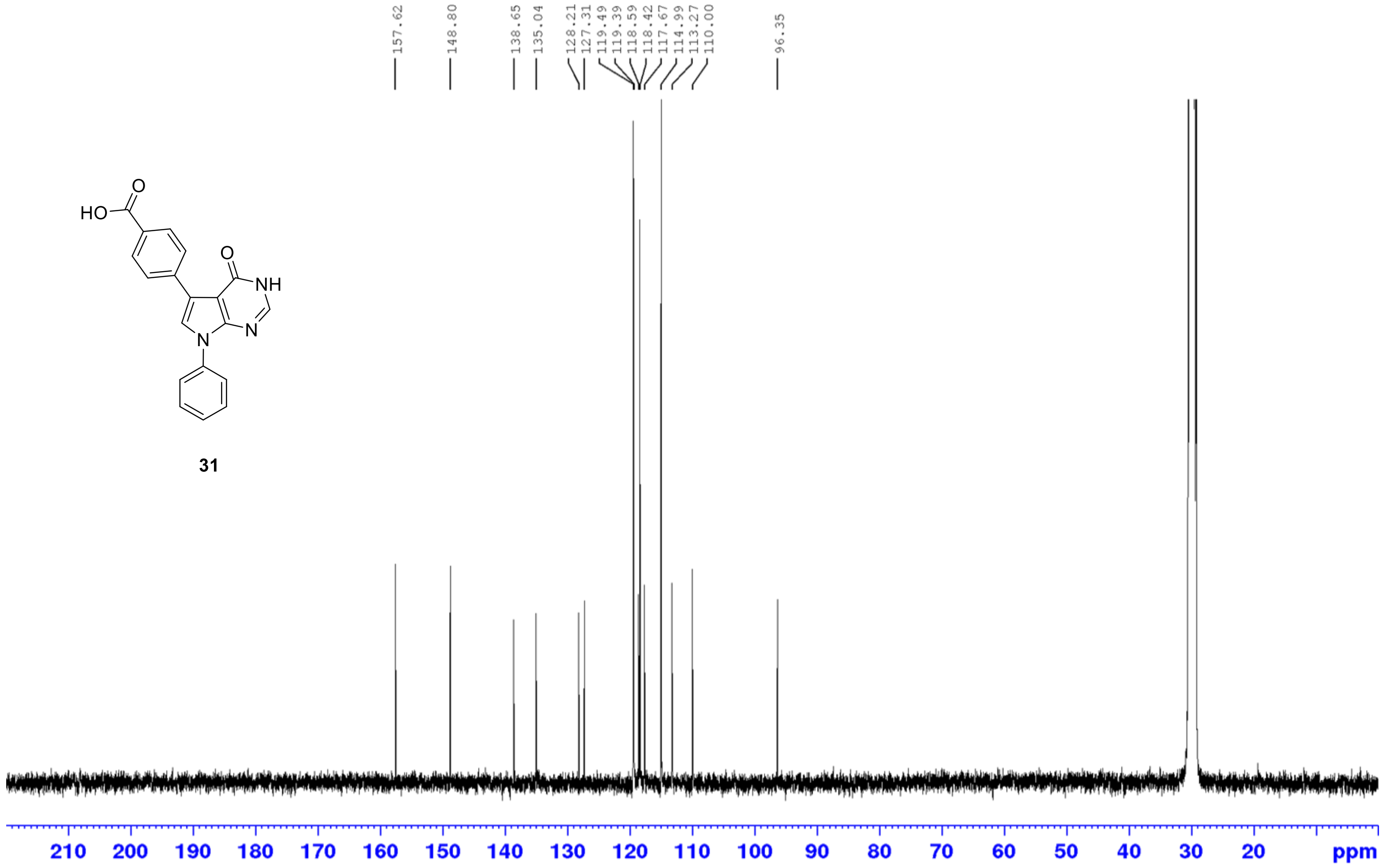
30

^{1}H NMR (400 MHz, DMSO-d_6)



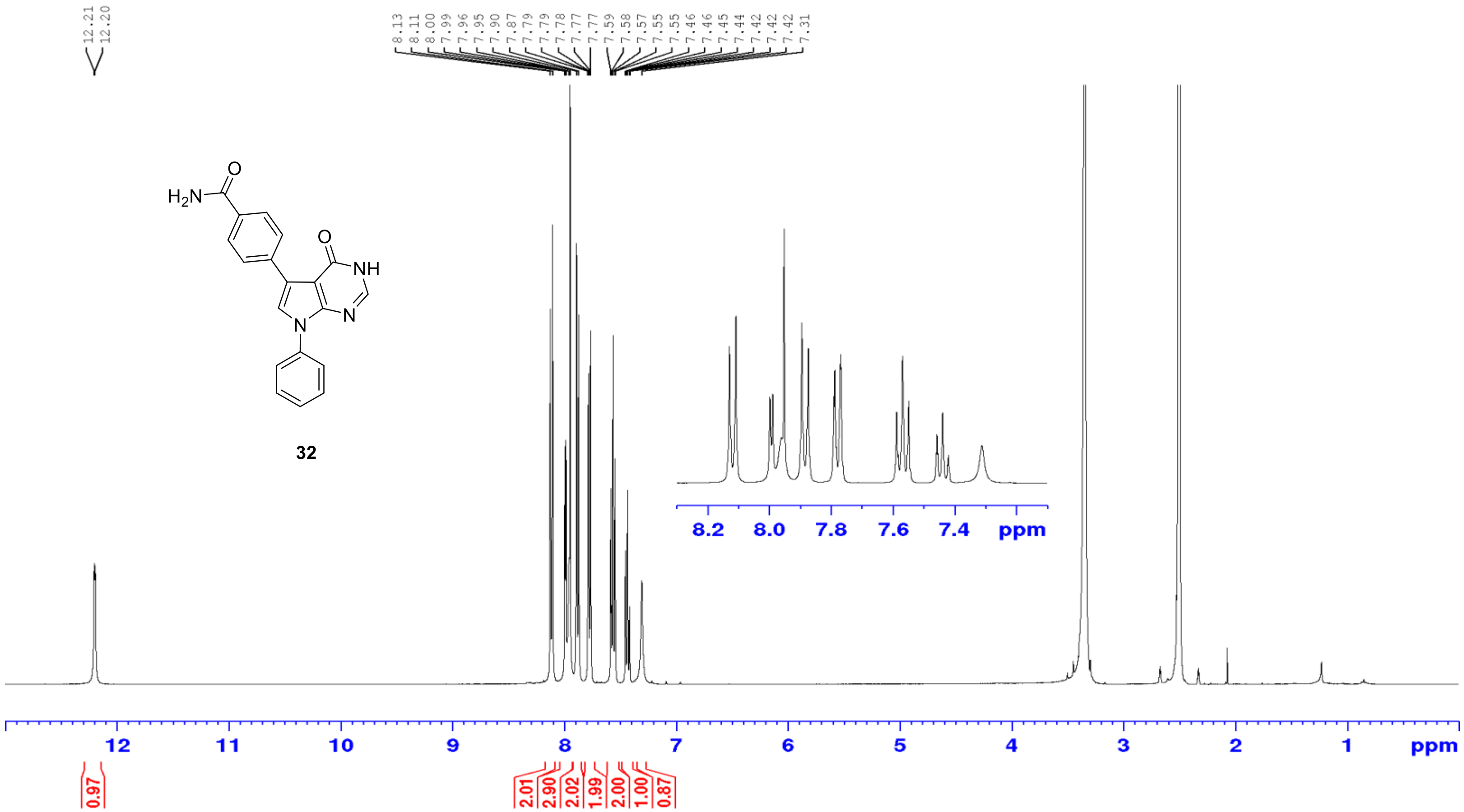
31

^{13}C NMR (101 MHz, DMSO-d_6)



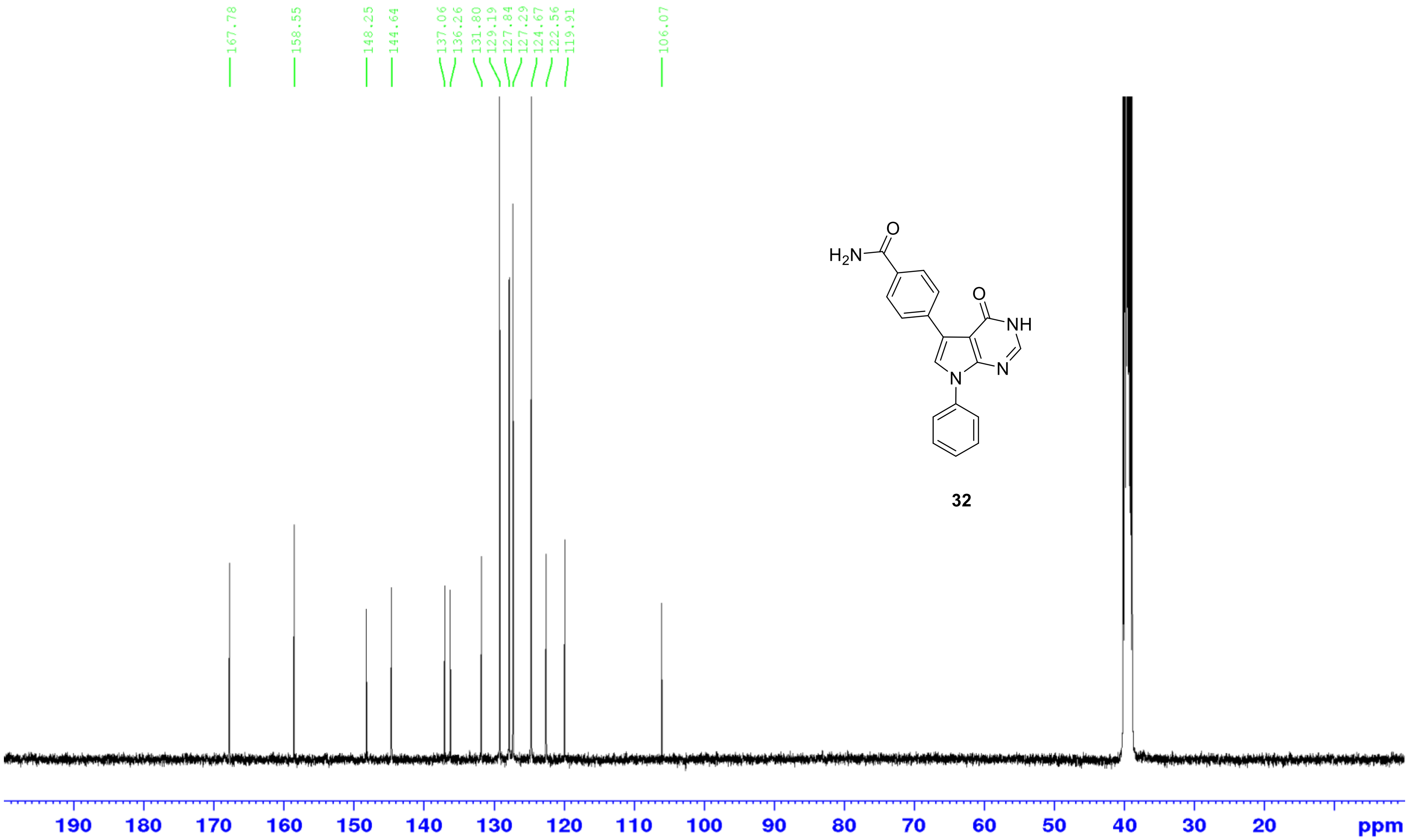
31

^1H NMR (400 MHz, DMSO-d_6)



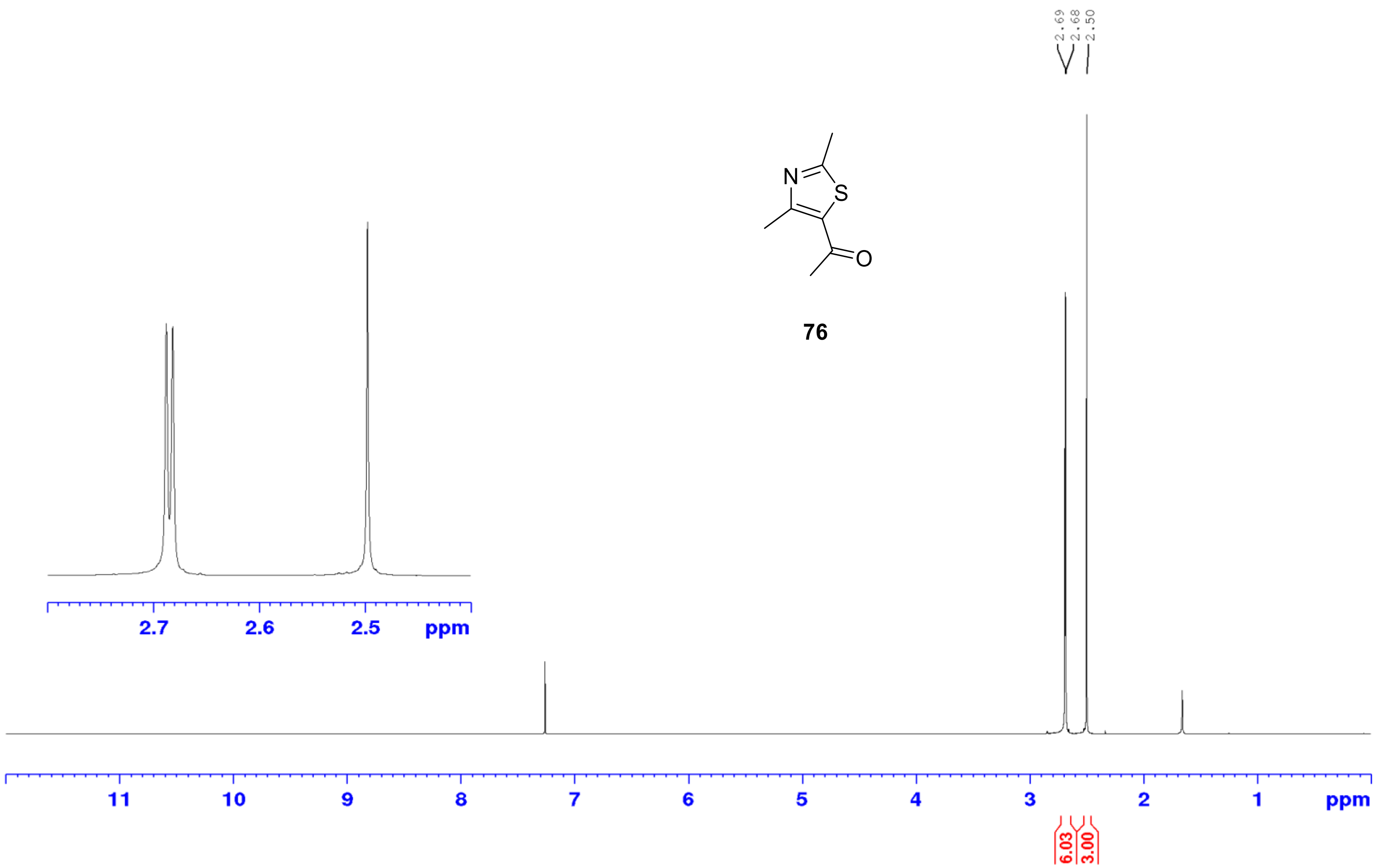
32

^{13}C NMR (101 MHz, DMSO-d_6)



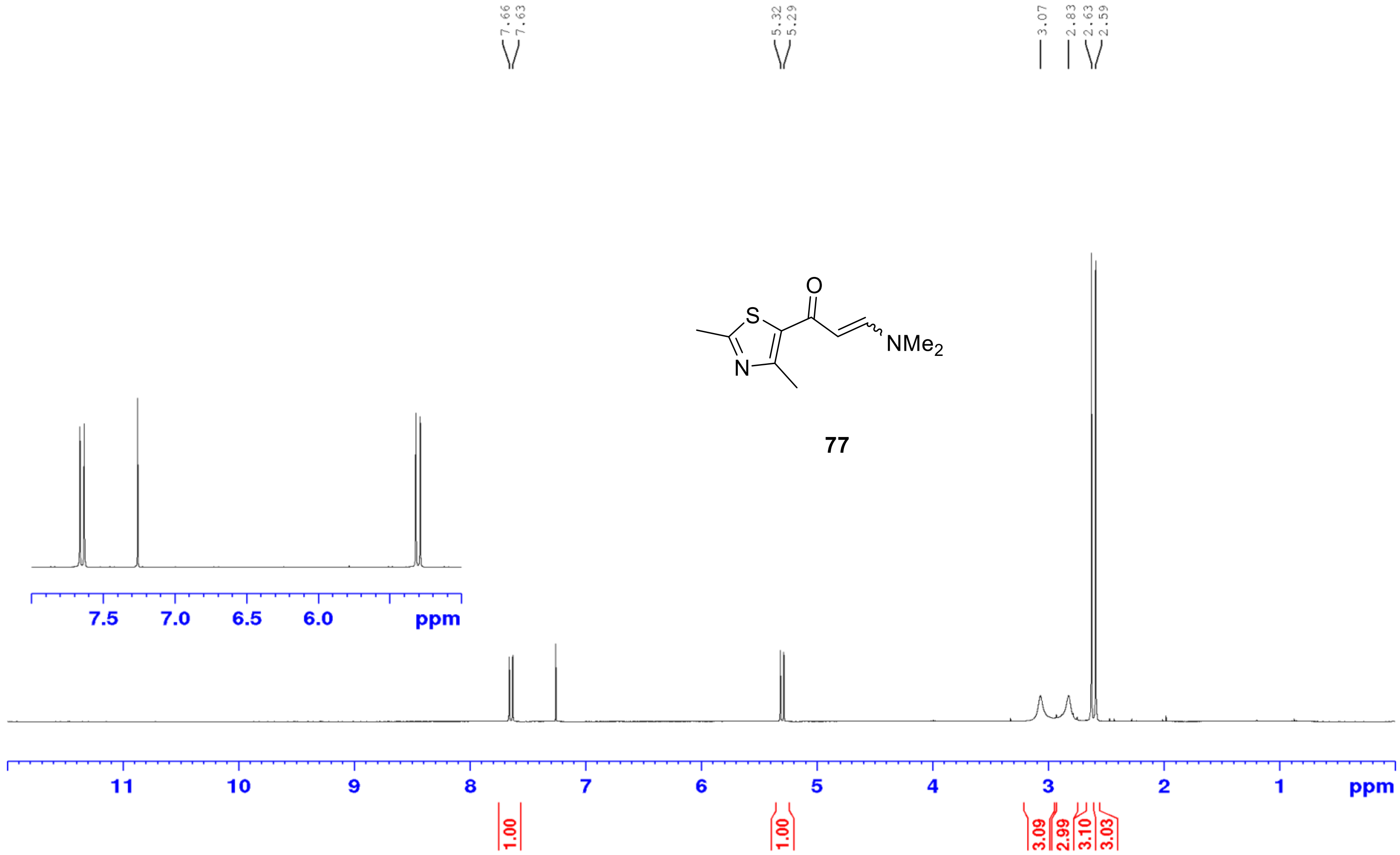
32

^{1}H NMR (400 MHz, $CDCl_3$)



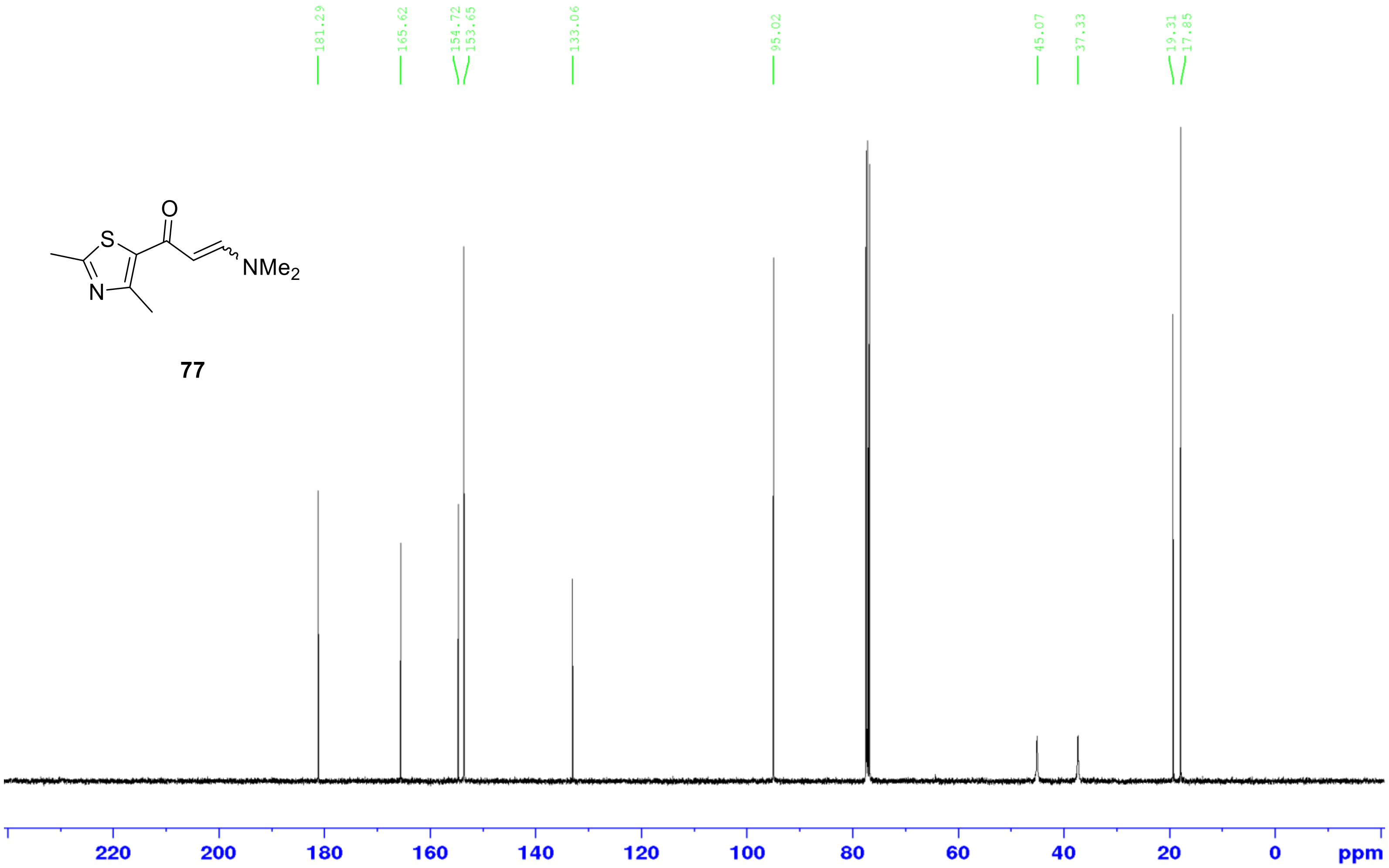
76

^1H NMR (400 MHz, $CDCl_3$)



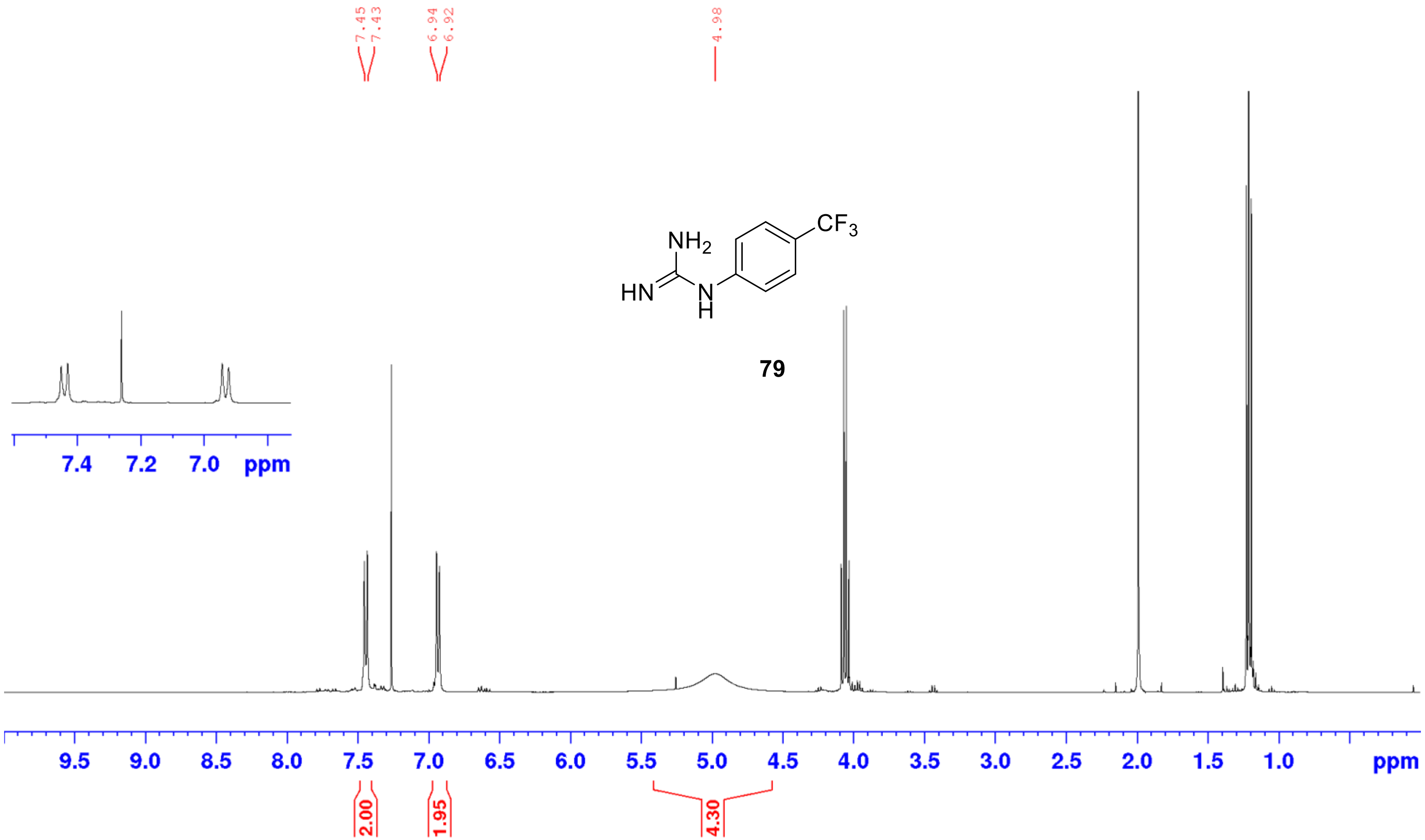
77

^{13}C NMR (101 MHz, $CDCl_3$)



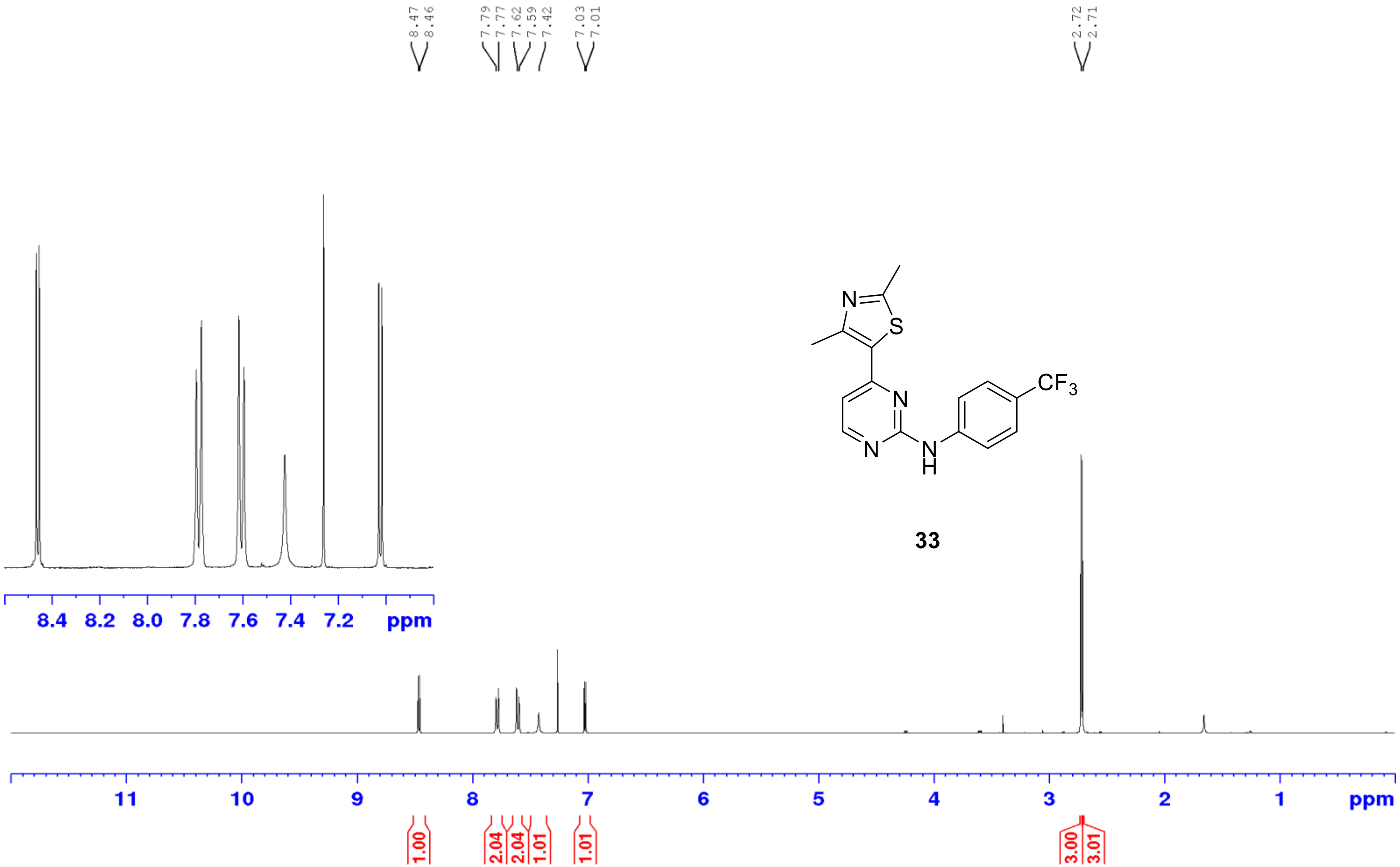
77

^{1}H NMR (400 MHz, CDCl$_3$)



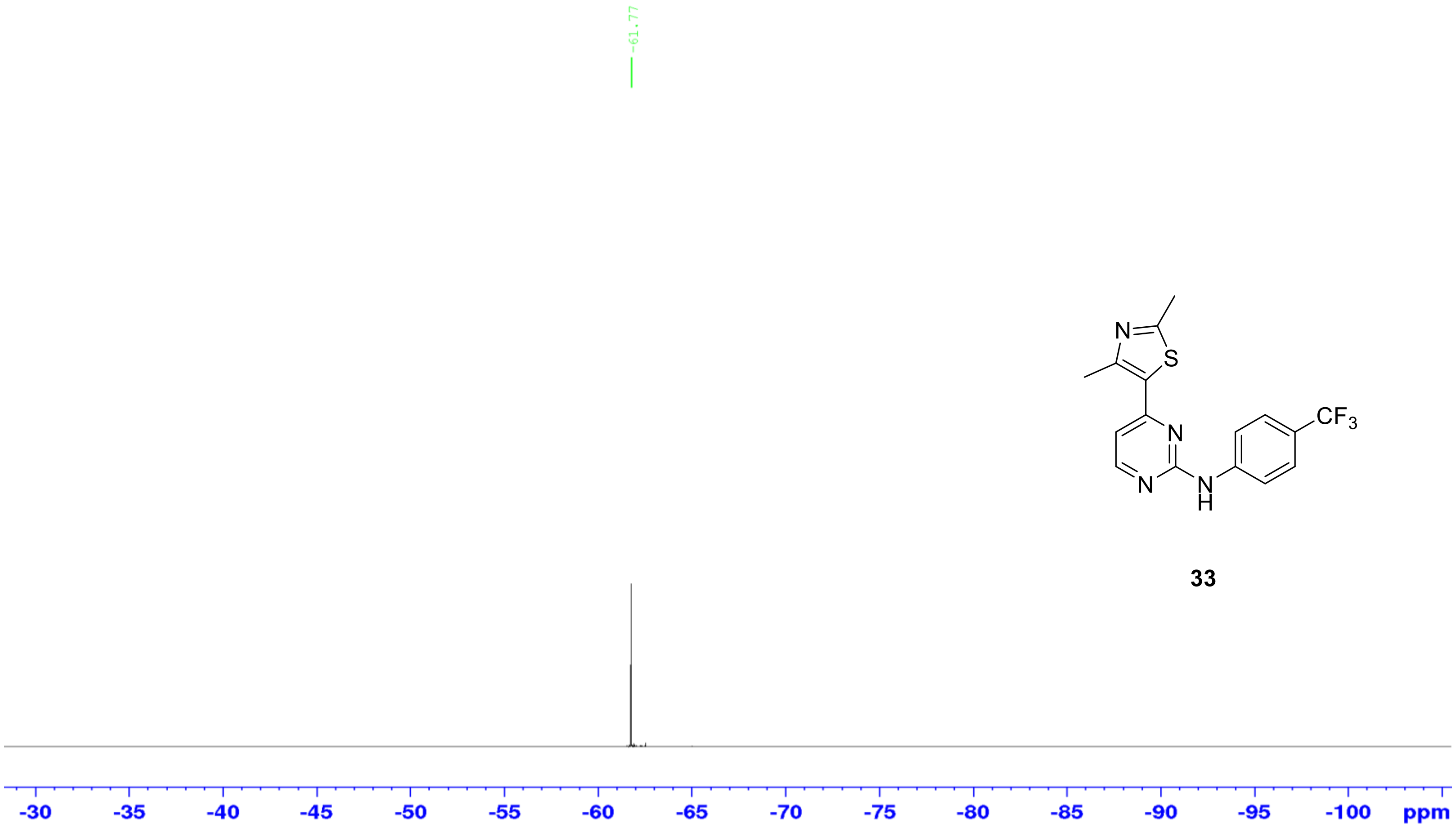
79

^1H NMR (400 MHz, $CDCl_3$)



33

^{19}F NMR (376 MHz, $CDCl_3$)



33

^{13}C NMR (101 MHz, $CDCl_3$)