

University of Montana

## ScholarWorks at University of Montana

---

Graduate Student Theses, Dissertations, &  
Professional Papers

Graduate School

---

2014

### VIRTUAL SCREENING AND DISCOVERY OF LEAD COMPOUNDS AS POTENTIAL DNA METHYLTRANSFERASE 1 INHIBITORS AND ANTICANCER AGENTS

Ogar Ofuka Ichire  
*The University of Montana*

Follow this and additional works at: <https://scholarworks.umt.edu/etd>

**Let us know how access to this document benefits you.**

---

#### Recommended Citation

Ichire, Ogar Ofuka, "VIRTUAL SCREENING AND DISCOVERY OF LEAD COMPOUNDS AS POTENTIAL DNA METHYLTRANSFERASE 1 INHIBITORS AND ANTICANCER AGENTS" (2014). *Graduate Student Theses, Dissertations, & Professional Papers*. 4390.  
<https://scholarworks.umt.edu/etd/4390>

This Dissertation is brought to you for free and open access by the Graduate School at ScholarWorks at University of Montana. It has been accepted for inclusion in Graduate Student Theses, Dissertations, & Professional Papers by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact [scholarworks@mso.umt.edu](mailto:scholarworks@mso.umt.edu).

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled:

VIRTUAL SCREENING AND DISCOVERY OF LEAD COMPOUNDS AS  
POTENTIAL DNA METHYLTRANSFERASE 1 INHIBITORS AND  
ANTICANCER AGENTS

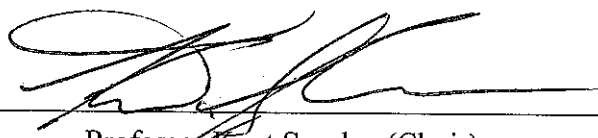
presented by Ogar Ofuka Leonard Ichire,

a candidate for the degree of doctor of philosophy and hereby certify that, in their opinion, it is worthy of acceptance.



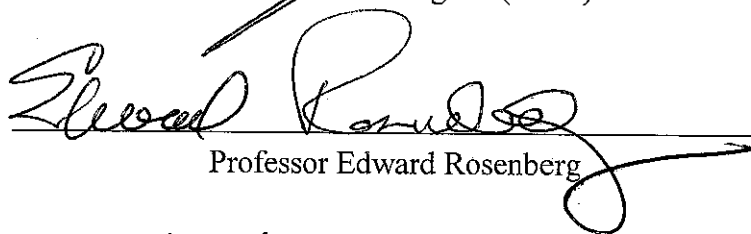
---

Professor Nigel Priestley (Advisor)



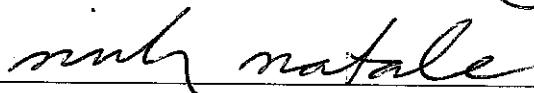
---

Professor Kent Sugden (Chair)



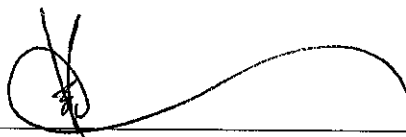
---

Professor Edward Rosenberg



---

Professor Nicholas Natale



---

Professor Valeriy Smirnov

VIRTUAL SCREENING AND DISCOVERY OF LEAD COMPOUNDS AS  
POTENTIAL DNA METHYLTRANSFERASE 1 INHIBITORS AND  
ANTICANCER AGENTS

By

OGAR OFUKA LEONARD ICHIRE

B.S Chemistry (Industrial) University of Ibadan - Nigeria, 2006

Dissertation

presented in partial fulfillment of the requirements  
for the degree of

Doctor of Philosophy  
Organic Chemistry

The University of Montana  
Missoula, MT

July 2014

Approved by:

Dr. Alexander J.B. Ross  
Dean of Graduate School

Dr. Nigel D. Priestley, Advisor  
Department of Chemistry and Biochemistry

Dr. Kent Sugden, Committee Chair  
Department of Chemistry and Biochemistry

Dr. Edward Rosenberg  
Department of Chemistry and Biochemistry

Dr. Nicholas Natale  
Department of Biomedical & Pharmaceutical Sciences

Dr. Valeriy Smirnov  
Department of Chemistry and Biochemistry

VIRTUAL SCREENING AND DISCOVERY OF LEAD COMPOUNDS AS POTENTIAL  
DNA METHYLTRANSFERASE 1 INHIBITORS AND ANTICANCER AGENTS

Advisor: Dr. Nigel D. Priestley

Chairperson: Dr. Kent Sugden

**ABSTRACT**

Epigenetic changes consist of DNA methylation, histone modification, micro RNA and genome imprinting. DNA methylation of the CpG islands is one of the main methods of epigenetic inactivation of genes and aberrant methylations at promoter regions of tumor suppressor genes can alter gene expression and play an important role in cancer development. DNA methyltransferase I (Dnmt1) is the enzyme responsible for maintaining methylation patterns during cell division and it is overexpressed in many cancers. Thus, Dnmt1 is a promising therapeutic target for development of novel anticancer agents and epigenetic modulators. We have developed two promising class of lead candidates, compounds 5-hydroxy-2-(4-hydroxyphenethyl)-3-oxo-N-pentyl-4-(4-(trifluoromethyl)phenyl)isoindoline-1-carboxamide **47**, 2-(2-(1H-indol-3-yl)ethyl)-5-hydroxy-3-oxo-N-pentyl-4-(4-(trifluoromethyl)phenyl)isoindoline-1-carboxamide **51** and 1-(4-isopropylphenyl)-2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indole **96**, as potential leads compounds that can be optimized for pharmaceutical applications..

The first class of lead compounds are isoindolinones; though originally discovered as hits from virtual screening we have designed and synthesized compound 5-hydroxy-2-(4-hydroxyphenethyl)-3-oxo-N-pentyl-4-(4-(trifluoromethyl)phenyl)isoindoline-1-carboxamide **47** and **51** as potential Dnmt1 inhibitors with promising anticancer properties. Compound **47** and 2-



(2-(1H-indol-3-yl)ethyl)-5-hydroxy-3-oxo-N-pentyl-4-(4-(trifluoromethyl)phenyl)isoindoline-1-carboxamide **51** serves as important lead compounds because of their efficient synthesis. The four-component single pot reaction used to prepare these compounds allows a variety of combinations of amines and isocyanide groups important for fine-tuning this class of compounds for structural activity relationship. Also, compound **47** and **51** displayed minimal toxicity to MCF10, a normal mammary gland cell line, but it was lethal to breast cancer MCF7 cells, which is an indication of selectivity. As a result, isoindolinones are being studied further as novel small molecules with potential therapeutic application.

Furthermore, using a combined study of pharmacophore modeling and 3-dimensional quantitative structure activity relationship (CoMFA), we discovered  $\beta$ -carbolines as another class of compounds with promising Dnmt1 inhibition and potent anticancer properties. From a small library of seven compounds synthesized, compound 1-(4-isopropylphenyl)-2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indole **96** was found to show potent anticancer activity against MCF7 with an  $IC_{50}$  close to Tamoxifen a commercial breast cancer drug. Compound **96** is in its early stage of development and can be studied and developed further to increase potency.

## ACKNOWLEDGEMENT

Coming to the University of Montana, I have received an enormous support and boost from a great number of individuals in the chemistry department. Dr. Nigel Priestley has been a mentor and inspiration all through my study. His guidance over the years has made this an insightful and rewarding experience. I am also very thankful to my dissertation committee: Edward Rosenberg, Kent Sudgen, Nick Natale and Valeriy Smirnov for their support and direction throughout my study. Special thanks to Dr. Edward Rosenberg for allowing me to kick start my career in his laboratory and for mentoring me the first two semesters of this journey. Also worth adding to this list of mentors is my undergraduate advisor, Dr. Idowu Iweibo, who believed and encouraged me to pursue a graduate degree in chemistry.

During my research, Dr. Brooke Martin, Adrienne Sochia and Jeremy Alverson spent countless hours testing my compounds and collecting data, I will like to say thank you. My computation and synthesis depended heavily on your test results and it was instrumental to getting me to this stage in graduate school. In the same light, I will like to appreciate Dr. Mike Braden for guiding me with the computational aspect of my project and allowing me to use the biomedical science computational facilities.

My list will not be complete without expressing my gratitude to Dr. Mike DeGrandpre for that special phone call that got me in graduate school in the first place; Mark Cracolice for TA assignments and guidance; Holly Thompson for her support; and Orion Berryman for the opportunity to work with him as a TA. Lastly, I will like to appreciate the chemistry department for the opportunity to work and learn with you all. It has been a great experience and I am more than thankful.

Of course, I have not forgotten my family and friends. I am grateful to my parents for their support and encouragement. I wish to thank Asen Oyongha and Nsor Obaji – none of these would have been possible without your help. I am also Thankful to all my friends: Yelebe Birhanu, Mainul Hossain, John Hoody, Ayesha Sharmin, Amanda Ross and all others I forgot to mention here – Thank You.

*“The most fundamental and lasting objective of synthesis is not production of new compounds but production of properties”*

**George S. Hammond**  
*Norris Award Lecture, 1968*

## Table of Contents

LIST OF FIGURES .....	VIII
LIST OF TABLES AND SCHEMES.....	X
LIST OF ABBREVIATIONS .....	XI
CHAPTER 1.....	1
<b>INTRODUCTION AND BACKGROUND OF EPIGENETICS.....</b>	<b>1</b>
1.0 EVOLUTION OF EPIGENETICS AND MECHANISM .....	1
1.1 EVOLUTION OF EPIGENETIC AND MECHANISM.....	2
1.2 GENETIC AND EPIGENETIC PARADIGM.....	2
1.3 CHROMATIN REMODELING AND DNA METHYLATION .....	5
1.4 EPIGENETICS AND HUMAN DISEASE .....	9
1.5 EPIGENETICS AND CANCER .....	10
1.6 DNA METHYLATION AND REGULATION IN MAMMALIAN GENETICS .....	15
1.6 DNA METHYL TRANSFERASE I – STRUCTURE AND FUNCTION .....	19
1.7 REGULATION OF DNMT1 .....	33
1.9 SMALL MOLECULES .....	38
1.10 BREAST AND COLON CANCER AND DNA METHYLATION.....	40
CHAPTER 2.....	46
<b>DRUG DISCOVERY AND DESIGN OF DNMT1 INHIBITORS.....</b>	<b>46</b>
2.0 INTRODUCTION.....	46
2.1 STRUCTURE AIDED DRUG DESIGN (SADD) .....	47
2.2 STRUCTURE-BASED VIRTUAL SCREENING .....	55
2.3 COMPUTATION BASED DRUG DISCOVERY OF DNMT1 INHIBITORS .....	70
CHAPTER 3.....	86
<b>DESIGN AND SYNTHESIS OF SMALL INHIBITOR MOLECULES .....</b>	<b>86</b>
3.0 INTRODUCTION.....	86
3.1 ZINC DATABASE SCREENING .....	86
3.2 PROMILIAD DATABASE SCREENING .....	90
3.3 PHARMACOPHORE MODELING.....	95
3.4 COMPARATIVE MOLECULAR FIELD ANALYSIS (CoMFA).....	98

3.5 TRANSITION STATE BASED INHIBITOR DESIGN .....	102
3.6 PROCAINAMIDE ANALOGUES .....	106
CHAPTER 4.....	109
<b>RESULT AND DISCUSSIONS .....</b>	<b>109</b>
4.1 TRANSITION STATE INHIBITORS .....	109
4.2 VIRTUAL SCREENING .....	113
4.3 ANALOGS OF ZINC04710208 (88SC).....	115
4.4 ISOINDOLINONES COMPOUNDS .....	118
4.5 BETA – CARBOLINES .....	125
CHAPTER 5 .....	129
<b>EXPERIMENTAL: CHEMISTRY AND BIOLOGY .....</b>	<b>129</b>
5.1 CHEMISTRY EXPERIMENTAL.....	129
5.1.1 General procedure for the naphthalenyl acrylamide .....	130
5.1.2 General Procedure for $\alpha$ -(N-Acylamino)- $\alpha,\beta$ -unsaturated acids .....	138
5.1.3 Isoindolinone Synthesis .....	144
5.1.4 Nucleoside Derivatives .....	175
5.1.5 Procainamide Derivatives .....	184
5.1.6 Beta-Carbolines .....	188
BIBLIOGRAPHY.....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
APPENIX: SELECTED NMR SPECTRA FOR SYNTHESIZED COMPOUNDS.....	207

## List of Figures

FIGURE 1.1: WADDINGTON'S EPIGENETIC LANDSCAPE.....	3
FIGURE 1.2: LINKS BETWEEN DNA METHYLATION, HISTONE MODIFICATION AND CHROMATIN REMODELLING.....	7
FIGURE 1.3: DEPICTION OF THE PROFILE OF GENE PROMOTER HYPERMETHYLATION ACROSS HUMAN TUMOR.....	13
FIGURE 1.4: DOMAIN ARCHITECTURE OF MAMMALIAN DNA MTASES.....	20
FIGURE 1.5: STRUCTURAL OVERVIEW OF MDNMT1.....	21
FIGURE 1.6: SEQUENCE ALIGNMENT OF THE C-TERMINAL SEGMENTS OF DNMT1 AND MDNMT1.....	22
FIGURE 1.7: THE CATALYTIC CORE AND TRD OF THE METHYLTRANSFERASE DOMAIN OF THE MDNMT1.....	23
FIGURE 1.8: INTRAMOLECULAR CONTACTS BETWEEN BOTH BAH DOMAINS AND THE Mtase DOMAIN.....	25
FIGURE 1.9: THE ZINC FINGER CLUSTERS OBSERVED IN THE CRYSTAL STRUCTURE OF THE MDNMT1.....	26
FIGURE 1.10: INTERMOLECULAR CONTACTS BETWEEN THE CATALYTIC DOMAIN OF MDNMT1.....	27
FIGURE 1.11: COMPARISON OF MDNMT1 WITH M.HHAI IN THEIR DNA-BOUND COMPLEXES.....	28
FIGURE 1.12: RIBBON MODEL OF MOUSE DNMT1(291–1620).....	34
FIGURE 1.13: EXPRESSION OF DNMT1 AND BRCA1 IN HUMAN PRIMARY BREAST CANCERS.....	35
FIGURE 1.14: COMPOUNDS TESTED FOR INHIBITORY ACTIVITY AGAINST DNMT1.....	39
FIGURE 2.1: MODELLING MOLECULAR RECOGNITION.....	49
FIGURE 2.2: CLASSIFICATION OF VARIOUS METHODS OF LIGAND BASED VIRTUAL.....	53
FIGURE 2.3: DRUG DISCOVERY AND RATIONAL DRUG DESIGN.....	54
FIGURE 2.4: CLASSIFICATION OF VARIOUS METHODS OF STRUCTURE-BASED VIRTUAL SCREENING.....	63
FIGURE 2.5: IDENTIFICATION OF TWO NOVEL CANDIDATE DNA MTase INHIBITORS BY COMPUTATIONAL SCREENING.....	72
FIGURE 2.6: OPTIMIZED DOCKING MODELS OF 2'-DEOXYCYTIDINE.....	74
FIGURE 2.7: OPTIMIZED DOCKING MODELS OF PROCAINAMIDE.....	74
FIGURE 2.8: COMPARISON OF THE OPTIMIZED BINDING MODELS OF HYDRALAZINE.....	75
FIGURE 2.9: MOLECULAR MODELING OF THE INTERACTION BETWEEN EGCG AND DNMT.....	76
FIGURE 2.10: OPTIMIZED DOCKING MODEL OF NSC 14778 WITH HUMAN DNMT1.....	77
FIGURE 2.11: OPTIMIZED BINDING MODEL OF VIRTUAL SCREENING HIT ZINC 00082754 WITH HUMAN DNMT1.....	77
FIGURE 2.12: COMPARISON OF THE BINDING MODES WITH PHARMACOPHORE HYPOTHESIS.....	79
FIGURE 2.13: EXAMPLE OF ENERGY-OPTIMIZED HYPOTHESIS FOR ARICEPT, 14, IN PDB 1EVE.....	81
FIGURE 2.14: STRUCTURE-BASED PHARMACOPHORE MODEL FOR DNMT INHIBITORS.....	82
FIGURE 2.15: COMPARISON BETWEEN THE BINDING MODE AND PHARMACOPHORE HYPOTHESIS.....	84
FIGURE 2.16: CHEMICAL STRUCTURE OF OTHER NOVEL DNMT1 INHIBITOR.....	85
FIGURE 3.1: TOP RANKED STRUCTURES FROM SCREENING ZINC DATABASE LIBRARY USING AUTODOCK 4.2.....	87
FIGURE 3.2: AMIDE ANALOGUES FROM (Z)-2-ACETAMIDO-3-(NAPHTHALENE-2-YL) ACRYLIC ACID (88SC).....	90

FIGURE 3.3: NAPHTHYL GROUP ANALOGUES FROM (Z)-2-ACETAMIDO-3-(NAPHTHALENE-2-YL) ACRYLIC ACID.....	90
FIGURE 3.4: DNMT1 ACTIVITIES OF ISOINDOLINONE COMPOUNDS FROM THE PROMILIAD LIBRARY.....	91
FIGURE 3.5: ISOINDOLINONE ANALOGUES FROM VARIATION OF ISOCYANIDES AND PRIMARY AMINE.....	93
FIGURE 3.6: PHARMACOPHORE MODEL USING LIGANDSCOUT.....	96
FIGURE 3.7: TOP SCORING NATURAL PRODUCT STRUCTURES PREDICTED BY THE PHARMACOPHORE MODEL....	97,98
FIGURE 3.8: STRUCTURES OF NATURAL PRODUCT PREDICTED BY COMFA QSAR.....	100
FIGURE 3.9: PROPOSED SCHEMATIC REPRESENTATION OF THE METHYLATION REACTION CATALYZED BY DMT....	103
FIGURE 3.10: PROCAINAMIDE ANALOGUES USING DIVERSE AMINE GROUPS.....	107
FIGURE 4.1: ELISA ASSAY I.....	109
FIGURE 4.2: RADIOACTIVE ASSAY I.....	111
FIGURE 4.3: RADIOACTIVE ASSAY II.....	115
FIGURE 4.4: MCF7 ASSAY II.....	116
FIGURE 4.5: RADIOACTIVE ASSAY III.....	118
FIGURE 4.6: IC50 VALUES OF ACTIVE ISOINDOLINONE ANALOGS.....	119
FIGURE 4.7: POTENTIAL B-CARBOLINE NATURAL PRODUCT FOR DEVELOPMENT AS DNMT1 INHIBITORS.....	124
FIGURE 4.8: FIRST LINE OF SIMPLE B-CARBOLINE TO TEST PHARMACOPHORE MODEL.....	125
FIGURE4.9: RADIOACTIVE ASSAY IV.....	126

## List of Tables and Schemes

TABLE 1.1 SUMMARY OF MOLECULES CURRENTLY KNOWN TO INTERACT WITH DNMT1.....	30
TABLE 1.2 SUMMARY OF AVAILABLE IC50 AND KI VALUES OF KNOWN DNMT1 INHIBITORS.....	40
TABLE 1.3 FREQUENCY OF METHYLATION OF REPRESENTATIVE TUMOR SUPPRESSOR.....	42
TABLE 2.1: A LIST OF WIDELY USED PROTEIN-LIGAND DOCKING SOFTWARE.....	62
TABLE 3.1 DOCKING RESULTS USING AUTODOCK4.2.....	88
TABLE 3.2: DNMT1 INHIBITORS PIC50 AND COMFA VALUES AND PREDICTED.....	99
TABLE 4.1: RADIOACTIVE ASSAY OF HIT COMPOUNDS FROM ZINC DATABASE.....	113
TABLE 4.2: IC50 VALUES OF ISOINDOLINONES OBTAINED FROM MCF7 AND MCF10 .....	120
TABLE 4.4: IC50 VALUES OF B – CARBOLINES AND TAMOXIFEN OBTAINED FOR MCF7 AND MCF10 CELL ASSAY.....	125
SCHEME 3.1: SYNTHESIS OF (Z)-2-ACETAMIDO-3-(NAPHTHALEN-2-YL)ACRYLIC ACID AND ANALOGS.....	88
SCHEME 3.2: SYNTHESIS FO NAPHTHALENYL ACRYLAMIDEDERIVATIVES.....	89
SCHEME 3.3 SYNTHESIS OF 3-(4-(TRIFLUOROMETHYL)PHENYL)PROPIOLIC ACID.....	92
SCHEME 3.4: SYNTHESIS OF SOINDOLINONES.....	92
SCHEME 3.5: SYNTHESIS OF S-((1-BENZYL-1H-1,2,3-TRIAZOL-4-YL)METHYL) ETHANETHIOATE.....	104
SCHEME 3.6: SYNTHESIS OF 4-AMINO-5-(((1-BENZYL-1H-1,2,3-TRIAZOL-4-YL)METHYL)THIO)METHYL)-1-(4-HYDROXY-5-(HYDROXYMETHYL)TETRAHYDROFURAN-2-YL)PYRIMIDIN-2(1H)-ONE.....	105
SCHEME 3.7: SYNTHESIS OF NUCLEOSIDE BASED INHIBITORS.....	106
SCHEME 3.8: SYNTHESIS OF PROCAINAMIDE DERIVATIVES.....	107



## List of Abbreviations

$^{13}\text{C}$	NMR 13-carbon nuclear magnetic resonance spectrum
$^1\text{H}$	NMR proton nuclear magnetic resonance spectrum
$\delta$	chemical shift
$\mu\text{M}$	micromolar
$\text{Ac}_2\text{O}$	acetic anhydride
AIBN	azobisisobutyronitrile
Bn	benzyl
$\text{CH}_2\text{Cl}_2$	methylene chloride
$\text{CuSO}_4$	copper (II) sulphate
DMAP	4-dimethylaminopyridine
DMSO	dimethylsulfoxide
d	doublet
DMF	dimethylformamide
DIPEA	diisopropylethyl amine
dd	doublet of doublets
EtOAc	ethyl acetate
$\text{Et}_3\text{N}$	triethylamine
g	gram
h	hours
HTS	high throughput screening
$\text{IC}_{50}$	50% of the inhibitory concentration
KOH	potassium hydroxide
m	molar
mL	milliliter
mg	milligram
$\text{MgSO}_4$	magnesium sulfate
MeCN	acetonitrile
MCF	Michigan Cancer Foundation
Pd	palladium
PDB	protein data bank
QSAR	quantitative structural activity relationship
s	singlet
THF	tetrahydrofuran
TLC	thin layer chromatography
TFA	trifluoroacetic acid
TBDMSCl	tert-butyl dimethylsilyl chloride
UV	ultraviolet/light
VS	virtual screening

# CHAPTER 1

## INTRODUCTION AND BACKGROUND OF EPIGENETICS

### 1.0 Historical Evolution of Epigenetics

DNA methylation at the C5 position of cytosine plays an important role in epigenetic regulation and it has been observed that the alteration of DNA methylation patterns is closely associated with gene silencing and epigenetic mutations. Also because the enzyme DNA methyltransferase1 (Dnmt1) is essential for maintaining methylation patterns during mammalian development the enzyme has become an important target for cancer therapies<sup>[1]</sup>. Therefore, the theme of this dissertation concerns the development of small molecules that reversibly bind to and inhibit the activity of Dnmt1 and their application as anticancer agents. Herein, we describe the molecular modeling, design and synthesis of DNA methyltransferase1 (Dnmt1) inhibitors and the use of these compounds as potential cancer therapeutics in MCF-7 breast cancer cell lines. The inhibitors are based on nucleoside and non-nucleoside compounds. The nucleoside inhibitors discussed are analogs of cytosine and thymidine, while the non-nucleoside structures are derivatives of isoindolinones,  $\beta$ -carboline and acrylic acid. This introductory chapter provides a thorough review on epigenetics, gene regulation and cancer. However, we will initially acquaint ourselves with the steady progression of knowledge from Darwinian genetics to the prevailing epigenetic model, particularly how the two approaches complement each other and help explain the relationship between DNA methylation and gene silencing.

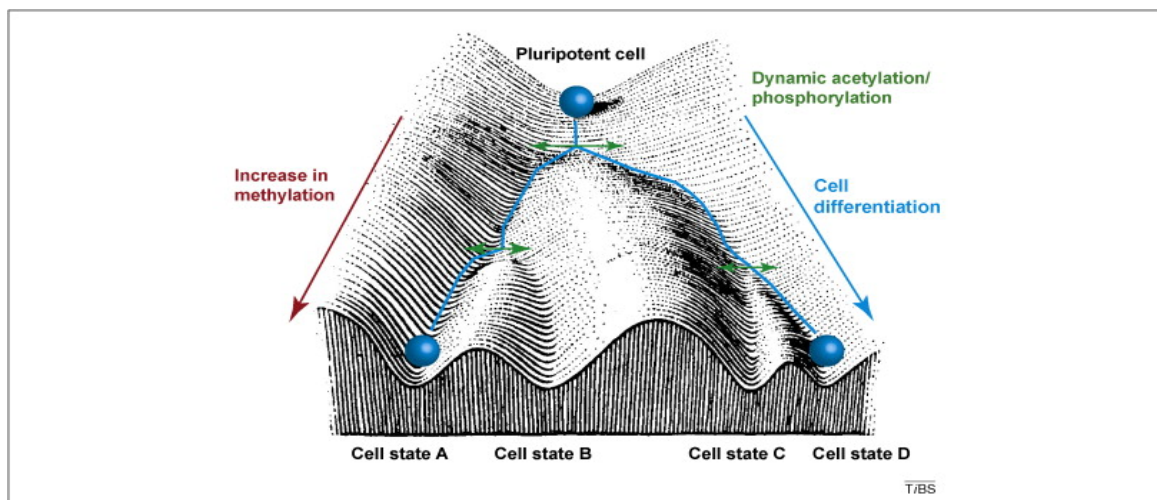
## 1.1 Evolution of Epigenetic and Mechanism

Darwinian genetics established the principle of change through natural selection; this change is seen in traits that show variation from parents to offspring. However, modern synthesis (Neo-Darwinism) has advanced Darwinianism to include other paths to variation and heredity apart from natural selection such as, random genetic drift, DNA recombination and mutation<sup>[2]</sup>. But the wind of change did not stop there; it has evolved, and probably still evolving, to a new concept termed Epigenetic Synthesis<sup>[3]</sup>. This new approach describes phenotypic variation due to changes in DNA expression without altering DNA sequence, which is determined largely by various epigenetic mechanisms<sup>[4]</sup>. These mechanisms use chromatin structure, DNA methylation and microRNA expressions to determine whether certain genes are expressed or specific sequences are partially or fully accessible by transcription factors during development without altering DNA sequence<sup>[5]</sup>; and these mechanisms in turn control phenotypic variations.

## 1.2 Genetic and Epigenetic Paradigm

In the early to mid-1900s, the views of neo-Darwinian genetics<sup>[6]</sup>, and by extension Darwinism, regarding the genetic paradigm of heredity and variation were summarized under some assumptions including: **(a)** genes determine characters in a straightforward and additive way, **(b)** genes are stable and passed on to the next generation, except for rare random mutations, and **(c)** genes are not affected by the environment – that is, no feedback from the environment on gene is ever observed<sup>[7]</sup>. These assumptions, however, have been met with various challenges over the years. For example, assumption **(a)** was revised by Sewell Wright, who argued that selection relates to a whole organism, but not to a single gene<sup>[8]</sup>. Also, great deal of knowledge has been added to assumption **(b)** by the fact that genes undergo mutations, insertions, deletion,

amplifications, rearrangements, recombinations, gene-jumping, and gene-conversion in response to stress, starvation, or adaptation to new situations. Lastly, if assumption (c) is considered then a gene would be said to be completely separated from the other network of genes and regulatory machinery, which is not the case as we now know today<sup>[1]</sup>. In addition, evidence that genes are unstable and are directly influenced by the environment is growing. Pollard and Rennie's studies of gene expression revealed some of the complexity and dynamism involve in cellular and genetic processes, which may serve to destabilize and alter genomes within the lifetime of an organism.<sup>[8-9]</sup>



**Figure 1.1: Waddington's Epigenetic Landscape.** The model presented by Conrad Waddington shows how a cell becomes more and more determined during development and that the possibility for differentiation decreases later in development. He compared the development of a cell with a ball (depicted in blue) rolling down the illustrated landscape and making its way through different valleys and elevations to different end points. Several types of histone modifications in this model can influence how cells develop. Various histone modifications such as acetylation and phosphorylation can change the fate of the cell to a small degree at a certain time point (green arrows), whereas methylations accumulation during development can have more influence on final destiny of a cell (red arrow). Although modifications with a fast turnover might not be involved in shaping a cell's identity directly, they can still influence long-term memory via interactions with methylation readers and writers.<sup>[10]</sup> *Permission requested*

Meanwhile, as the genetic paradigm undergoes some modifications, there is a revival of the epigenetic approach which was first proposed by Jean-Baptiste Lamarck. And according to

the Lamarckian theory, the organism is capable of interacting with the environment, and can acquire transformations during this process which can be internalized.<sup>[11]</sup> But the term ‘epigenetics’, which describes Lamarck’s idea, was not coined until Conrad H. Waddington (1942) used ‘epigenetic landscape’ as a metaphor to describe the flexibility and plasticity of gene regulation during development in his theory of genetic assimilation. In brief, Waddington visualized a cell in an embryo as a ball rolling down the ‘landscape’ (**Figure 1.1**). As it rolls, the ball has several ‘options’ or ‘choices’ as to which way to go – just as a developing embryo is influenced by genetic and environmental factors to take certain ‘paths’ so also so is the cell. By the time it reaches the bottom, the cell would have made several such ‘choices’. The importance of the epigenetic landscape, as described by Ho and Saunder, is that its topography is determined by all of the genes that are interlinked, and not on specific alleles of particular genes.<sup>[8]</sup> Ho and Saunder’s concept implies that when populations of organisms experience a new environmental influence during development a large proportion of the organism give a novel response to these new environmental stimuli; it will mean a normal developmental pathway being ‘pushed’ over a threshold or in some cases a new pathway emerges in the epigenetic landscape. Now if this response is adaptive, then a natural selection for it will be established and it becomes regulated so that more or less uniform response results from a range of intensity of the environmental stimulus. Ultimately, after some generations, the response become genetically assimilated, that is, it arises even without the stimulus.

Put together, Waddington defined epigenetics as ‘the causal interactions between genes and their products, which bring the phenotype into being.’<sup>[12]</sup> His definition was initially centered on embryonic development as guided by epigenetic landscape. Today, however, epigenetics has advanced to accommodate a wide variety of biological processes and describes heritable changes

in gene expression that occur independent of changes in DNA sequence. Also, while Waddington's genetic assimilation was demonstrated, at the time, by changes in cytoplasmic organization that could be stably inherited independent of nuclear or cytoplasmic DNA,<sup>[13]</sup> his argument on his epigenetic theory did not specify the mechanism by which the assimilation occurs, or as the measure of the ability of a population to produce the same phenotype regardless of the environment. But, today from the work of Holliday and colleagues,<sup>[14]</sup> the mechanism by which epigenetics modifications occur have been studied and they include: chromatin remodeling, DNA methylation, RNA transcription, MicroRNAs synthesis, and prions formation. We shall look at three of these epigenetic regulation mechanisms in more detail.

Nevertheless, from the ongoing discussion, it is only logical to view epigenetics not as a replacement of neo-Darwinism but as an advancement of our current understanding of genetics and its many ramifications. Or simply put, epigenetics introduces a chemical switchboard for a network of genes to be turned on and off without altering the DNA sequence. Epigenetics describes a new concept the 'epigenome' that is dynamic and highly sensitive to the environment, which complements the already known sequence based DNA molecule.

### **1.3 Chromatin Remodeling and DNA Methylation**

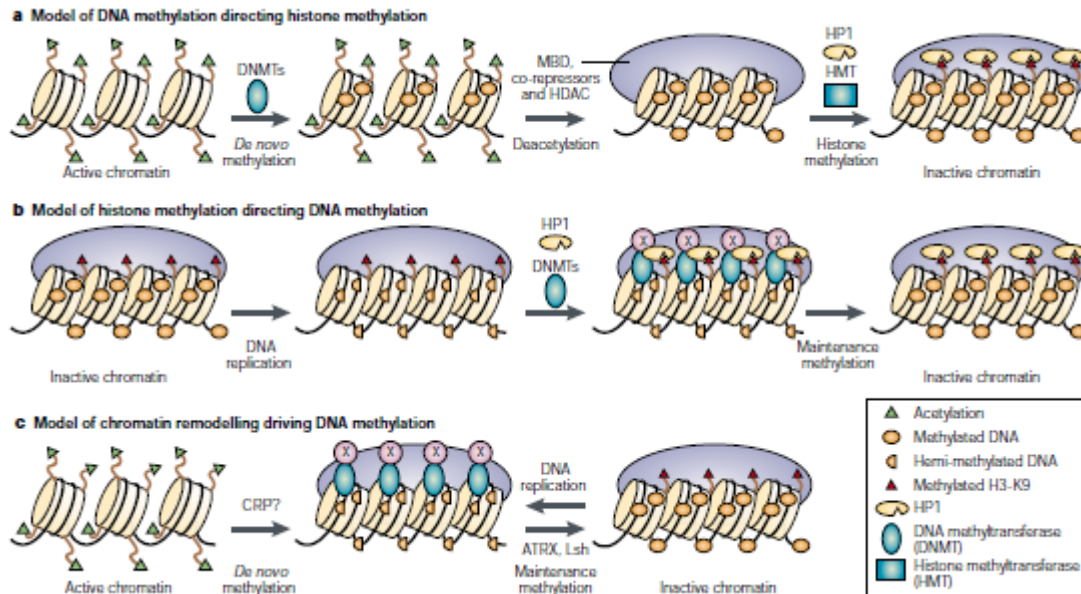
Eukaryotic DNA is packaged into highly organized chromatin structure, which is important to preserve a whole genome inside the nucleus of a cell. On the other hand, the compactness represents an obstacle for DNA transcription, replication, repair and recombination<sup>[15]</sup> because these processes require unwound DNA for them to occur. Consequently, there is a dynamic balance between genome packaging and genome accessibility. The dynamic balance would allow, for example, specific transcription factors to bind to a target sequence on DNA for a

particular gene transcription process, or would prevent access to that DNA sequence. Chromatin remodeling proteins, histone modification proteins and DNA methylation are multiple mechanisms that tightly regulate gene transcription processes.<sup>[15]</sup> In this section we shall summarize the first two processes and discuss DNA methylation extensively.

The basic structural unit of a chromatin is the nucleosome, which consist of approximately 146 base pairs of DNA wrapped 1.65 turns around a histone octamer core containing two molecules in each of core, the histones H2A, H2B, H3 and H4.<sup>[16]</sup> Neighboring nucleosomes are separated by, on average 50 bp of free DNA. Positive charged residues in the histone contact the phosphate backbone of the DNA every 10.4 bp, providing fourteen relatively weak histone-DNA interactions.<sup>[16]</sup> In particular, the histone-DNA interactions are influenced by nucleosome-modifying and nucleosome-remodeling protein complexes. The former makes covalent modifications on the histone protein, mostly at the tail of histone H3 and H4, and creates markers that will be identified later by transcription regulation proteins.<sup>[17]</sup> While the later, chromatin remodeling complexes restructures, mobilizes and ejects the nucleosomes in order to regulate access to the DNA, without covalently modifying the histone protein.<sup>[18]</sup> Remodeling involves the breaking and reforming of the histone-DNA contacts, which results in the mobilization of nucleosome on the nucleosome template. The regulators, though believed to be independent, are interconnected with DNA methylation, since they are also involved in chromatin silencing. There are some speculations that all three epigenetic mechanisms might influence each other to some extent.<sup>[19]</sup>

## DNA Methylation

If we shift our focus from covalent modification of histone core proteins to modification of DNA base pairs, then we will visit an important epigenetic mechanism by which tissue-specific gene-expression patterns and global gene silencing are established and maintained.



**Figure 1.2: Links between DNA methylation, histone modification and chromatin remodeling. a) A model of DNA methylation directing histone methylation.** DNA methylation patterns are established through *de novo* methylation by the DNA methyltransferases Dnmt3A and Dnmt3B, and are maintained by Dnmt1. Methyl-CpG-binding proteins (MBD) and histone deacetylase (HDAC) complexes, such as the MECP2–Sin3a–HDAC complex, are believed to then be recruited to the methylated region to induce histone deacetylation and silencing. The chromatin then attracts histone methyltransferases (HMTs), such as Suv39h or G9a, which methylate the lysine 9 residue on histone H3 (H3-K9) and stabilize the inactive state of the chromatin. **b) A model of histone methylation directing DNA methylation.** Methyl H3-K9 acts as a signal for inactive chromatin by recruiting HP1 to methylated histones, which might in turn recruit DNA methyltransferases directly or indirectly (through an unknown factor, factor X) to the silent chromatin to maintain DNA methylation and stabilize the inactive chromatin<sup>21,22</sup>. **c) A model of chromatin remodelling driving DNA methylation.** The ATP-dependent chromatin-remodelling and DNA-helicase activities of proteins, such as ATRX and Lsh, might facilitate DNA methylation and histone modification by unwinding nucleosomal DNA to increase its accessibility to Dnmts, HDACs and HMTs. The disrupted function of these proteins impairs both DNA methylation and histone methylation, as has been shown in plants. The chromatin-remodelling protein (CRP) that is involved in *de novo* methylation has yet to be identified.<sup>119</sup> *Permission requested*



DNA methylation occurs on both cytosine and adenine bases in prokaryotes, but in humans, methylation is restricted to cytosine bases<sup>[20]</sup>; it is mostly absent from the genome especially at short genomic regions called CpG islands, that is, regions with more than 500 base pairs and with a GC content greater than 55%<sup>[21]</sup>. Such stretches of DNA are located within the promoter regions of 40% of mammalian genes and are usually free of methylation. Aberrant *de novo* methylation can cause heritable transcriptional silencing, and is usually the hallmark of most human cancers<sup>[1]</sup>. Broadly put, DNA methylation is associated with chromatin compactness and gene repression, however, it plays a key role in cell development which we will discuss later<sup>[20b]</sup>. So far, two general mechanisms have been proposed for inhibition of gene expression. First, methylated cytosine bases can inhibit the association of DNA binding factors with their specific DNA recognition sequence.<sup>[22]</sup> Second, methyl-CpG-binding proteins (MBPs) can recognize methyl-CpG and recruit transcriptional co-repressive molecules<sup>[23]</sup> to the site of transcription. Epigenetic mechanisms are suspected to be interconnected (**Figure 1.2**), as in the above case, MBPs can use co-repressors molecules to silence transcription and possibly modify the surrounding chromatin<sup>[24]</sup> as noticed with chromatin remodelers.

### **DNA Methyltransferases**

The methylation of CpG sites within the human genome is maintained by a number of DNA methyl transferase (enzymes). Dnmts fit into two general classes based on their preferred DNA substrate and function. Dnmt3a and Dnmt3b are known as *de novo* methyl transferases; they introduce cytosine methylation at previously unmethylated CpG sites and establish new DNA-methylation patterns.<sup>[25]</sup> While maintenance methyl transferase, Dnmt1, copies pre-existing methylation patterns onto the new DNA strand during replication, it restores these DNA-

methylated patterns by methylating hemi-methylated CpG sites.<sup>[25]</sup> Dnmt3L is a Dnmt-related protein that does not possess known methyl transferase catalytic action but it is known to physically associate with Dnmt3a and Dnmt3b, and modulate their activity.<sup>[26]</sup> Lastly, there is a fourth type DNA methyltransferase, Dnmt2, which only shows a weak DNA methylase activity *in vitro* but is known to methylate small RNA (e.g. tRNA<sup>Asp</sup>).<sup>[27]</sup> The major activity and function of Dnmt2 is not yet fully understood, however the targeted deletion of the *Dnmt2* gene in embryonic stem cells had no noticeable effect on global methylation implying that Dnmt2 might not be involved in *de novo* methylation.<sup>[28]</sup>

So far, DNA methylation is one of the best understood epigenetic modifications of the chromatin. It will be important to consider the structure and mechanism of action of Dnmts; particularly we shall revisit Dnmt1 in more detail to discuss its mode of action and implication in cancer.

#### **1.4 Epigenetics and Human Disease**

There are several indications that dysfunctions and mutations in epigenetic mechanism can result in a number of human diseases, especially cancer.<sup>[29]</sup> Epigenetic disease could be as a result of changes in global or localized methylation patterns, or incorrect histone modification. Feinberg has reviewed the subject in terms monogenic epigenetic disease, and further classified them into two. The first type includes those diseases involving mutation in genes that are epigenetically regulated, (e.g. imprinted gene), and the second type are those that affect the epigenome as a whole, such as the modification of epigenetic mechanism (e.g. DNA methylation).<sup>[30]</sup>

Human diseases related to monogenic epigenetic machinery can either be single-gene disorder such as Rubinstein Taybi syndrome<sup>[31]</sup>, ATRX syndrome<sup>[32]</sup> and Retts syndrome, or a network of genes as in the case of disrupted imprinted genes. In the former case, Retts syndrome is an X-linked neurodevelopmental disorder that involves mutation of the CpG binding protein 2 (*MeCP2*), a protein that binds to methylated DNA sequence.<sup>[33]</sup> On the other hand, gene imprinting disorders can lead well known syndromes like Beckwith Weidemann syndrome (BWS), Prader-Willie syndrome and Angelman syndrome<sup>[34]</sup>; but we will not discuss them further, instead we will turn our attention to cancer related epigenetic changes.

## 1.5 Epigenetics and Cancer

### DNA Methylation and Cancer

The development of cancer involves a complex succession of events, from incipient cells acquiring alterations in oncogenes, tumor-suppressor genes and instability of genes to disruption of cellular regulatory pathways.<sup>[5]</sup> Epigenetic alterations in tissues-specific genes or housekeeping genes of either oncogenes or tumor-suppressor genes, can lead to a number of different malignancies by either hypermethylation or hypomethylation of these genes. We will examine some of these gene alterations and the type of cancers involved.

### Hypermethylation

Since the initial discovery of CpG island hypermethylation of the *Rb* promoter (tumor suppressor gene associated with retinoblastoma)<sup>[35]</sup>, hypermethylation of the CpG islands in the promoter regions of (tumor suppressor) genes is now a widely accepted feature in human cancers. Epigenetic silencing of tumor suppressor genes can also lead to tumor initiation by serving as the second hit in the Knudson's two hit model for tumor development.<sup>[36]</sup> Numerous genes are

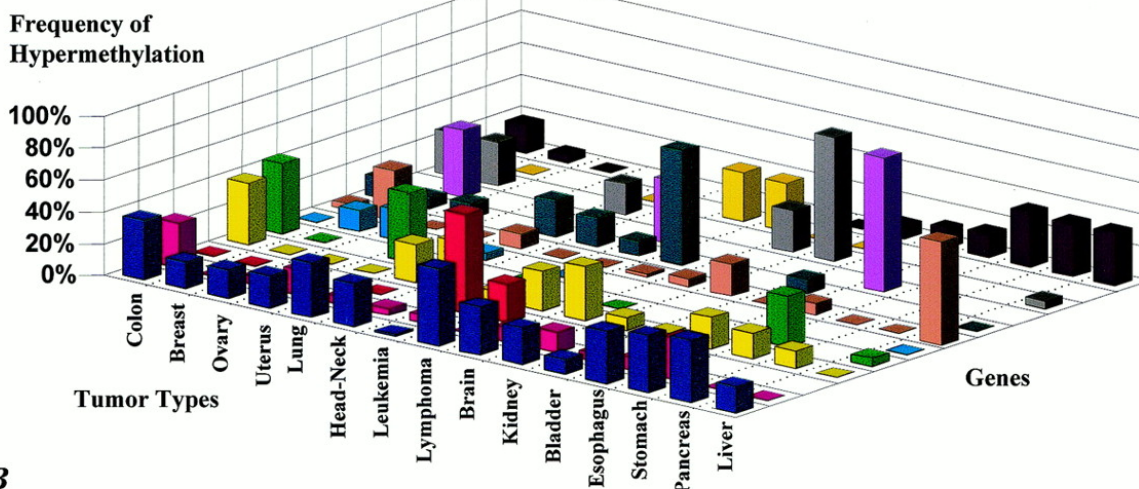
susceptible to hypermethylation, and they include, cell cycle regulation ( $p^{16INK4a}$ ,  $p^{15INK4a}$ ,  $Rb$ ,  $p^{14ARF}$ ), DNA repair ( $hMLH1$ ,  $MGMT$ ,  $WRN$ ,  $BRCA1$ ), apoptosis ( $DAPK$ ,  $TMS1$   $SFRP1$ ,  $WIF-1$ ), p53 network ( $p14ARF$ ,  $TP73$ ,  $HIC-1$ ), and angiogenesis.<sup>[37]</sup> For example, in many cases of leukemia and other hematologic related diseases,  $p^{15INK4B}$ ,  $p21^{Cip1/Waf1}$ , the  $ER$  gene,  $SDC4$ ,  $MDR$  group of genes were seen to be highly methylated in various hematologic cancers.<sup>[38]</sup> Also, Yang noted that in breast cancer development gene silencing includes steroid receptor genes, cell adhesion gene and inhibitors of metalloproteinases.<sup>[39]</sup> Some of the genes highly methylated are estrogen receptor ( $ER$ ) alpha, progesterone receptor ( $PR$ ),  $p16INKA$ ,  $BRCA1$ ,  $GESTP1$  and E-cadherin. The  $BRCA1$  gene is one of the more commonly associated genes in breast cancer, and aberrant DNA methylation results in reduced or complete absence of the BCRA1 protein.<sup>[40]</sup>

Esteller et al. conducted a gene hypermethylation profile of human cancer with a total of 12 genes, including well-characterized tumor suppressor genes ( $p16INK4a$ ,  $p15INK4b$ ,  $p14ARF$ ,  $p73$ ,  $APC$ , and  $BRCA1$ ), DNA repair genes ( $hMLH1$ ,  $GSTP1$ , and  $MGMT$ ), and genes related to metastasis and invasion ( $CDH1$ ,  $TIMP3$ , and  $DAPK$ ).<sup>[41]</sup> Each gene possesses a CpG island in the 5' region, which is normally nonmethylated in healthy cells.<sup>[20b]</sup> The study sampled over 600 specimens covering over 15 major tumor types (colon, stomach, pancreas, liver, kidney, lung, head and neck, breast, ovary, endometrium, kidney, bladder, brain, and leukemia and lymphomas). Their results showed that some genes, such as the cell cycle inhibitor  $p16^{INK4a}$ , are hypermethylated across many tumor types including colorectal, lung, and breast carcinomas.<sup>[42]</sup> Other studies expanded on  $p16^{INK4a}$  epigenetic silencing to include neoplasm (**Figure 1.3**), and other types of tumor such as bladder<sup>[43]</sup> and cervical tumors<sup>[44]</sup> or melanomas<sup>[45]</sup> and gliomas.<sup>[46]</sup> Also, DNA repair gene  $MGMT$  and  $SAPK$  share the same wide distribution in cancer types.<sup>[47]</sup> Conversely, hypermethylation of  $p14ARF$  and  $APC$  are most prevalent in gastrointestinal tumors

(i.e., colon and stomach).<sup>[48]</sup> Likewise, *GSTP1* methylation aberration is characteristic of steroid-related neoplasm such as breast, liver and prostate.<sup>[49]</sup> Equally noteworthy is the aberrant methylations that are very specific in selected tumor types, which may be inheritable. For example, *BRCA1* hypermethylation is found only in breast and ovarian carcinomas.<sup>[50]</sup> Female individuals carrying a mutated *BRCA1* allele have an estimated risk of 87% for breast cancer and 44% for ovarian cancer by age 70.<sup>[51]</sup> Another very specific gene is the mismatch repair gene *hMLH1*, which is restricted to three sporadic tumor types characteristic is the hereditary nonpolyposis colorectal cancer syndrome: colorectal, endometrial, and gastric tumors with microsatellite instability.<sup>[52]</sup> Likewise, hypermethylation of *p73* and *p15<sup>INK4b</sup>* is only observed in hematological malignancies.<sup>[53]</sup>

Interestingly, some epigenetic inactivation may affect all of the molecular pathways involved in cell immortalization and transformation. As a result, in any given tumor it is possible to find simultaneous inactivation of several pathways by aberrant methylation silencing in cell cycle (*p16<sup>INK4a</sup>* and *p15<sup>INK4b</sup>*), DNA repair (*hMLH1*, *MGMT*, and *BRCA1*), cell adherence and metastasis process (*CDH1*, *TIMP3*, *DAPK*), p53 network (*p14ARF* and *p73*), metabolic enzymes (*GSTP1*), and the *APC/b-catenin* route (*APC*).<sup>[41]</sup> For example, a colorectal tumor may have disruption of cell cycle, DNA repair, and metastasis-related process by hypermethylation of *p16<sup>INK4a</sup>*, *hMLH1*, and *TIMP-3*, respectively.<sup>[41]</sup> Similarly, mammary tumors may affect the same pathways by silencing *p16<sup>INK4a</sup>*, *BRCA1*, and *CDH*, or lung tumor affecting *p16<sup>INK4a</sup>*, *MGMT*, and *DAPK*.<sup>[41]</sup>

A



B

	p16 <sup>INK4a</sup>	p14 <sup>ARF</sup>	p15 <sup>INK4b</sup>	MGMT	hMLH1	BRCA1	GSTP1	DAPK	CDH1	TIMP-3	p73	APC
Colon	37%, 41/110	28%, 37/132	0%, 0/19	39%, 127/323	44%, 15/34*	0%, 0/18	4%, 1/23	13%, 2/23	N.D.	27%, 6/22	0%, 0/10	18%, 20/108
Breast	17%, 11/66	0%, 0/20	0%, 0/16	0%, 0/36	0%, 0/10	13%, 11/84	31%, 24/77	7%, 1/15	42%, 37/88	27%, 8/29	0%, 0/15	5%, 1/19
Ovary	18%, 4/22	5%, 1/20	N.D.	0%, 0/23	N.D.	19%, 11/58	0%, 0/10	9%, 2/23	N.D.	N.D.	N.D.	0%, 0/20
Uterus	20%, 6/29	16%, 4/25	N.D.	0%, 0/17	43%, 24/56*	N.D.	0%, 0/20	N.D.	N.D.	N.D.	N.D.	N.D.
Lung	31%, 28/89	6%, 4/62	0%, 0/21	21%, 18/83	0%, 0/20	4%, 1/22	9%, 2/21	16%, 10/64	N.D.	19%, 4/21	0%, 0/22	0%, 0/17
Head-Neck	27%, 26/95	4%, 1/25	N.D.	32%, 37/116	N.D.	N.D.	0%, 0/106	18%, 17/92	N.D.	N.D.	N.D.	0%, 0/10
Leukemia	1%, 1/150	5%, 1/20	62%, 93/150	6%, 2/31	6%, 3/51	0%, 0/19	0%, 0/10	9%, 8/86	40%, 30/75	N.D.	31%, 11/35	N.D.
Lymphoma	48%, 12/25	0%, 0/22	24%, 6/25	25%, 15/61	N.D.	N.D.	2%, 1/47	72%, 21/29	N.D.	N.D.	30%, 3/10	N.D.
Brain	30%, 3/10	9%, 2/22	N.D.	34%, 74/213	0%, 0/15	N.D.	5%, 1/20	N.D.	N.D.	26%, 20/77	0%, 0/22	0/10
Kidney	23%, 6/25	13%, 5/38	N.D.	8%, 1/12	N.D.	N.D.	20%, 8/35	N.D.	N.D.	78%, 28/36	0%, 0/10	8%, 1/12
Bladder	9%, 1/11	5%, 1/20	N.D.	4%, 2/44	N.D.	N.D.	0%, 0/24	9%, 1/11	N.D.	N.D.	N.D.	10%, 2/19
Esophagus	33%, 5/15	8%, 3/37	N.D.	20%, 3/14	N.D.	N.D.	7%, 1/14	N.D.	84%, 26/31	N.D.	N.D.	15%, 4/27
Stomach	36%, 8/22	26%, 31/118	N.D.	16%, 10/60	32%, 21/65*	N.D.	0%, 0/22	N.D.	N.D.	N.D.	N.D.	34%, 13/38
Pancreas	39%, 7/18	0%, 0/20	N.D.	11%, 2/18	N.D.	N.D.	0%, 0/18	N.D.	N.D.	N.D.	N.D.	33%, 6/18
Liver	15%, 3/20	0%, 0/20	N.D.	0%, 0/59	5%, 2/20	0%, 0/18	65%, 13/20	0%, 0/20	N.D.	5%, 1/20	N.D.	33%, 6/18

Figure 1.3: A, depiction of the profile of gene promoter hypermethylation across human tumor types. All cases represent random and unselected populations of each particular tumor type, except p, where hMLH1 methylation was determined in colorectal, endometrial, and gastric tumors enriched in microsatellite-unstable samples. Analysis of the methylation status was studied in most cases by sodium bisulfite modification of DNA and subsequent PCR using primers designed for either methylated or unmethylated DNA (PCR conditions and sequences are available upon request). Additional samples were analyzed by Southern blot with methyl-sensitive enzymes, restriction cut analysis, and bisulfite genomic sequencing. B, numerical distribution of promoter hypermethylation according to gene and tumor type<sup>[41]</sup>. Permission requested

Also obvious from the profile (Figure 1.3) is that certain tumor type share similar gene hypermethylation types. Gastrointestinal tumor (colon and gastric) share a set of genes undergoing hypermethylation characterized by  $p16^{INK4a}$ ,  $p14^{ARF}$ ,  $MGMT$ ,  $APC$ , and  $hMLH1$ , while other aerodigestive tumor types, such as lung and head and neck, have a different pattern of hypermethylated genes including  $DAPK$ ,  $MGMT$ , and  $p16^{INK4a}$ , but not  $hMLH1$  or  $p14^{ARF}$ . A pattern is also observed for ovarian and breast cancers, where  $BRCA1$ ,  $GSTP1$ , and  $p16^{INK4a}$

genes are mostly hypermethylated.<sup>[41]</sup> Such observations allow a small subset of genes to be used as biomarkers for certain cancer types. This represents only a partial picture of the complex methylation changes in cancer.

Whereas, our current focus is on hypermethylation of the CpG islands in the promoter region, certain reports suggest that most of the aberrant DNA methylation occurs in CpG island shores – that is, a short distance within 2 kb from the CpG island as observed in some colon cancers.<sup>[54]</sup> Also noteworthy is the fact that most changes in the CpG island shore (45-65%) are associated with regions highly methylated during tissue differentiation and lastly, differential methylation pattern correlates with gene expression at CpG island shores just as it does with CpG islands.<sup>[55]</sup> Whether it is the CpG island or the CpG island shores, global hypomethylation and local hypermethylation patterns observed in most cancers are caused by aberrant methylation resulting from dysfunctional DNA methyltransferase. High expressions of Dnmt1 and Dnmt3b have been observed in many tumor types, however these methylases are regulated by miRNAs as well.<sup>[56]</sup> So, though there are indications of a strong correlation amongst epigenetic regulatory machineries, this finding also suggests the involvement of high methylase enzyme in most cancers.

### **Hypomethylation**

In contrast to hypermethylation, global hypomethylation is also observed in a wide variety of cancers (e.g. prostate, cervical and hepatocellular carcinogenesis), and it occurs mainly at repetitive sequences.<sup>[57]</sup> Hypermethylation promotes chromosomal instability, translocation, gene disruption and activation of oncogene. Simple repeat sequences such as DNA satellites contain repeated DNA sequences arranged in tandem. Such satellites are commonly found in pericentromeric or subtelomeric heterochromatin and are normally methylated in the genome of



a healthy individual. It has also been observed that hypomethylation of DNA satellites, such as the centromeric *sat $\alpha$* , juxtacentromeric *sat2* and *sat3*, have been linked to ICF syndrome and cancers including Wilms tumor, ovarian and breast cancer.<sup>[58]</sup> In addition, DNA repeats *NBL2* was observed to be hypomethylated in neuroblastomas and hepatocellular cancer.<sup>[59]</sup>

Hypomethylation has been linked to oncogenesis by activation of oncogenes such as *cMYC*, *H-RAS*, *MAPSIN* in gastric cancer, *S-100* in colon cancer and *MAGE* (melanoma-associated antigen) in melanoma.<sup>[60]</sup> It is important to state here that oncogene activation is highly complex and the mechanism of action remains a subject of intense investigation. Current findings, however, have shown that hypomethylation plays a prominent role in genomic instability and increase mutation rates.<sup>[61]</sup>

## 1.6 DNA Methylation and Regulation in Mammalian Genetics

DNA methylation fundamentally modifies the functional organization of the human genome. The first evidence of DNA methylation was found in 1948<sup>[62]</sup> and decades later, we now understand, though not fully, the important role it plays in development, gene regulation and disease. It involves an enzymatic covalent transfer of a methyl group from *S*-Adenosyl methionine (SAM) to the N or C atoms of two DNA bases to form methylated bases such as N<sup>6</sup>-methyladenine, 5-methylcytosine, and N<sup>4</sup>-methylcytosine, which become natural components of the DNA. Modifications add extra information to the DNA that is not coded in the original sequence. In this chapter we will focus on C5 methylation of cytosine. Particularly, we will discuss the basic functions of DNA methylation, and then we will look at structure and function of *Dnmt1*.



## **DNA Methylation and its Basic Functions**

In mammals, during embryogenesis and development, complex and dynamic DNA methylation transformations are observed. For instance, active demethylation occurs in the male pronucleus shortly after fertilization, through a process that seems to be independent of DNA replication.<sup>[63]</sup> Also, after formation of the zygote, both maternal and paternal chromosomes undergo progressive passive demethylation from both gametes, with the exception of the methylation marks at the imprinted loci.<sup>[64]</sup> Furthermore, embryonic DNA methylation patterns are established after implantation through lineage-specific *de novo* methylation that begins in the inner cell mass of blastocyst.<sup>[64a, 64b, 65]</sup> DNA methylation levels increase rapidly in the primitive ectoderm, which gives rise to the entire embryo, whereas methylation is either inhibited or not maintained in the trophoblast and the primitive endoderm lineage, which give rise to the placenta and yolk sac membrane, respectively.<sup>[66]</sup> Also demethylation and *de novo* methylation occur during gametogenesis and they are important for parental-specific marks in imprinted loci (genomic imprint)<sup>[67]</sup>, which we shall discuss in a later section. It is important, from the ongoing, to acknowledge the critical role DNA methylation plays in cellular differentiation and the normal functioning of differentiated cells as well as development in general.

## **DNA Methylation and Genomic Imprinting**

In mammals, a small number of genes (50-80) carry an imprint that allows asymmetric expression of genes inherited from paternal or maternal copies of the chromosome.<sup>[68]</sup> Imprinting refers to genes that are either silenced when paternally inherited but expressed when maternally inherited or vice versa.<sup>[68]</sup> For example, monoallelic expression of the maternal copy of the *H19* gene and monoallelic expression of the paternal copy of *IGF2* gene is seen to be active in

offspring.<sup>[69]</sup> The involvement of DNA methylation in imprinting has been observed in several studies: genetic studies of the Dnmt3 family members have provided persuasive evidence that DNA methylation is essential for genomic imprinting. For example, the disruption of *Dnmt3a*, but not *Dnmt3b*, in primordial germ cells results in loss of paternal and maternal imprinting.<sup>[70]</sup> Also, Kaneda's results show that offspring of mutant female mice die *in utero* if *Dnmt3a* is disrupted and they lack methylation and allele-specific expression at all maternally imprinted loci. Furthermore, *Dnmt3a* conditional mutant male mice show impaired spermatogenesis and they lack correct methylation at two paternally imprinted loci. Therefore, appropriate methylation and expression of imprinted genes is important for normal development because of the developmental and genetic diseases that are associated with imprinting defects. Falls et al.<sup>[71]</sup> reviewed genetic imprinting and human diseases such as, Beckwith–Wiedemann syndrome, Silver–Russell syndrome, Angelman syndrome and Prader–Willi syndrome.<sup>[71]</sup>

### **DNA Methylation and X-chromosome Inactivation**

DNA methylation also plays an important role in X-chromosome inactivation. It is well known that in mammals, females carry two X chromosomes while males have only one. X-chromosome inactivation is described as a dosage compensation mechanism that results in the transcriptional silencing of one of the two X chromosome in the female during early embryogenesis. The process of inactivation involves specific expression of the Xist (inactive X-specific transcript) RNA from the inactivated chromosome, as well as increased methylation and histone deacetylation of the inactive chromosome.<sup>[72]</sup> Beard et al. and Panny et al. have revealed that DNA methylation is not essential for the initiation and development of X-inactivation but very much required for stable maintenance of X-chromosome inactivation.<sup>[73]</sup> Also, the choice of which X-chromosome to inactivate is made early in embryogenesis. In embryonic tissues of one

chromosome is selected at random<sup>[72]</sup>, whereas in extraembryonic tissues the paternal chromosome is always the choice for inactivation.<sup>[74]</sup> Lastly, X-chromosome inactivation depends on CTCF, a candidate trans-acting factor that detects the methylation site of the protein.<sup>[75]</sup>

### **DNA Methylation and Selfish Genes**

Genomes are vulnerable to selfish genetic elements (SGEs) such as transposons, retrotransposons, and viruses, which enhance their own transmission genome wide but are neutral or harmful to the individual. Startlingly, about 40% of the human genome is composed of transposable elements (TEs).<sup>[76]</sup> Since such transposable element are not stable, and random integration into the genome is a key source of mutation, preventing the insertion of transposable element by transcription silencing is important for life and avoidance of disease. DNA methylation is known to suppress TEs in both animals and plants<sup>[77]</sup>; transposons and other repetitive DNA sequence are usually relatively rich in GC sequences and are heavily methylated.<sup>[76a]</sup> Also, an increase in the transcription of transposon is known for *Dnmt1* knock-out embryonic stem cells and cell lines.<sup>[77-78]</sup> It is important to add here that DNA methylation is only one of the many machineries involved in SGE suppression, there are other mechanisms such as repeat-induce point mutation in fungi, RNAi, and small RNA suppression pathways that are also involved in regulating transposable elements.<sup>[76b, 79]</sup>

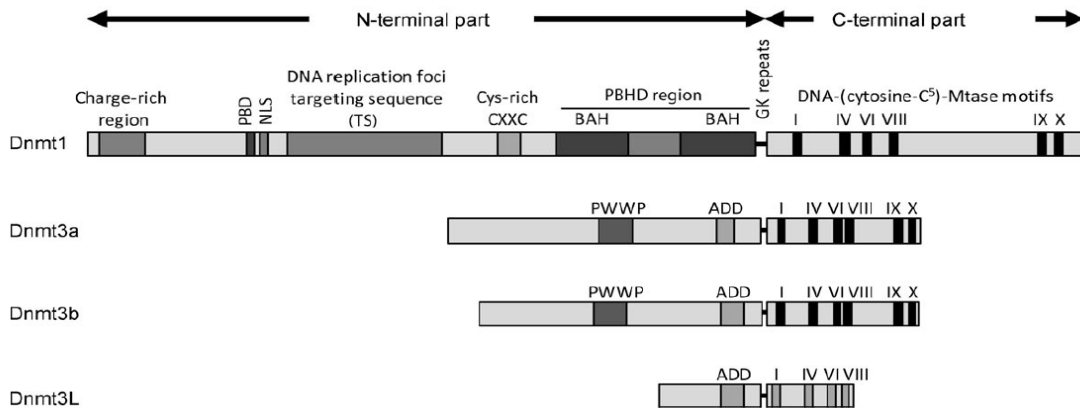
Interestingly, there are other important functions of DNA methylation such as reduction in transcriptional noise associated with the spurious transcription of genes.<sup>[80]</sup> The noise is based on the long-standing hypothesis, which posits that ‘gene body’ (transcriptional unit) methylation suppresses false transcription within coding regions. Adding that, by so doing, gene body

methylation can potentially reduce transcriptional noise. We will not go in depth with this subject matter but it is significant to highlight that in general, the levels of gene expression vary between cells even with the same genetic materials and under the same biological conditions. Understanding the nature and mechanism of this variability, which is commonly referred to as ‘transcriptional noise’ is crucial and it has a strong link with DNA methylation since most of the gene body region is extensively methylated.<sup>[79, 81]</sup>

## 1.7 DNA Methyl transferase I – Structure and Function

Dnmt1 is the principal DNA methyltransferase (Dnmt) in mammalian cells. It is a very large multimodular protein which comprises 1620 amino acids; the actual length depends on the species and the expression of tissue-specific exons.<sup>[82]</sup> The protein is highly dynamic with multiple regulatory features that control DNA methylation. In general, Dnmt enzymes are divided into a smaller catalytic domain and a larger N-terminal regulatory domain (**Figure 1.4**). Prokaryotic orthologs of Dnmt1 lack the regulatory N-terminal domain; however they share the same specificity for CpG sites.<sup>[83]</sup> Dnmt1 follow the same structural organization; composed of the N-terminal, which includes the replication foci-targeting domain (RFD), a DNA-binding CXXC domain, a pair of bromo-adjacent homology (BAH) domains, and a conserved C-terminal catalytic domain. It is important to differentiate hDnmt1 (human Dnmt1) from mDnmt1 (mouse Dnmt1) and M.HhaI (bacterial Dnmt1 form *Haemophilus haemolyticus*) because although the catalytic pockets are highly conserved within the enzymes, there some subtle difference important for drug design and we are concerned with hDnmt1. Song et al. solved the crystal structure (**Figure 1.5**) of mouse Dnmt1 (650-1602 amino acids) at 3.0 Å and HDnmt1 (646-1600 amino acids) at 3.6Å respectively, and both structures show 85% sequence similarity **Figure 1.6.**<sup>[84]</sup> We shall use different sections to discuss the structure and function of both the

catalytic and the regulatory N-terminal mainly of hDnmt1 while contrasting it with mDnmt1 and M.HhaI, since the project concerns design and synthesis of hDnmt1 inhibitors.



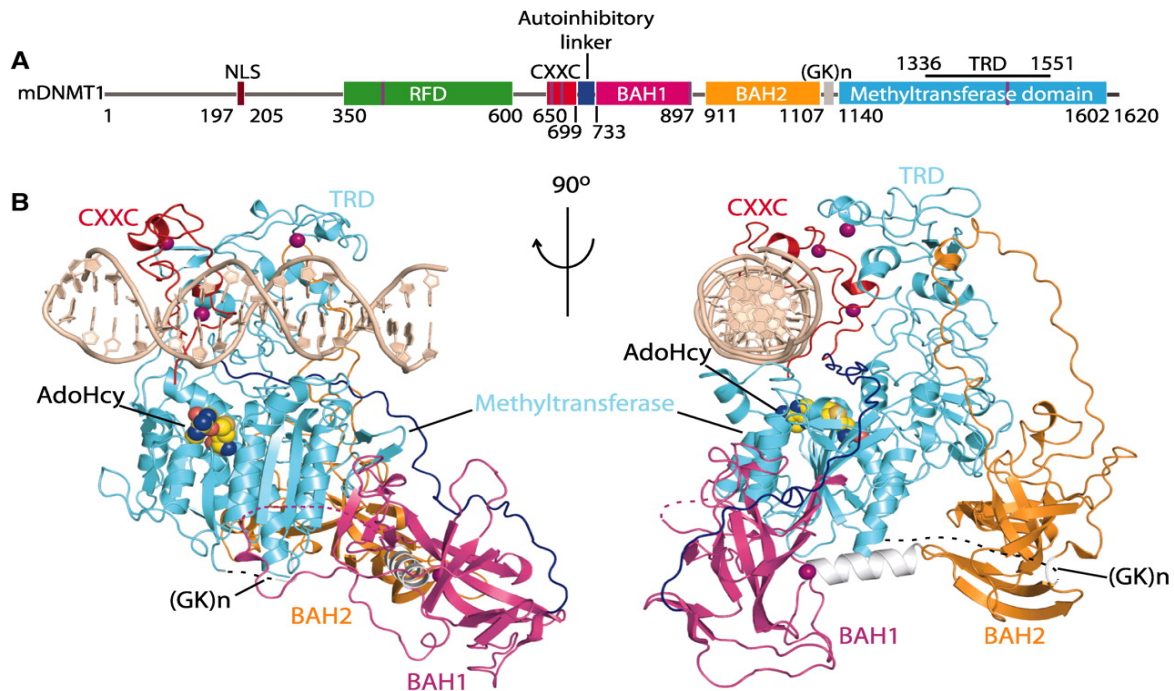
**Figure 1.4: Domain architecture of mammalian DNA MTases. Functional domains in the N-terminal part of the proteins are shown and the conserved C5 DNA MTase motifs in the C-terminal part are labeled. *Permission requested***

### The catalytic C-terminal Domain

The crystal structure of HhaI<sup>[85]</sup> shows that the catalytic domain needs three structural elements to support catalytic activity: an AdoMet-binding domain, a target based-binding cavity, and a target sequence recognition domain (TRD). The first two domains are highly conserved sequence domains and are easily recognizable in most Dnmt enzymes<sup>[86]</sup>, however, the target recognition/element domain is not precisely defined by conserved sequence motifs. Similarly, the catalytic domain of Dnmt1 folds into two subdomains, designated as the catalytic core and the TRD.<sup>[84]</sup> The catalytic core is dominated by seven-stranded  $\beta$ -sheet that is flanked by three  $\alpha$ -helices on either side (**Figure 1.5&1.7**). The central  $\beta$ -sheet is further joined by a two-stranded anti-parallel  $\beta$ -sheet from the BAH1 domain (**Figure 1.5 & 1.8C**). In the TRD, there is a hairpin-like fold at the start of the domain that forms hydrophobic contacts with the catalytic core and the BAH1 domain (**Figure 1.7B**). While, the majority of the TRD folds into an independent

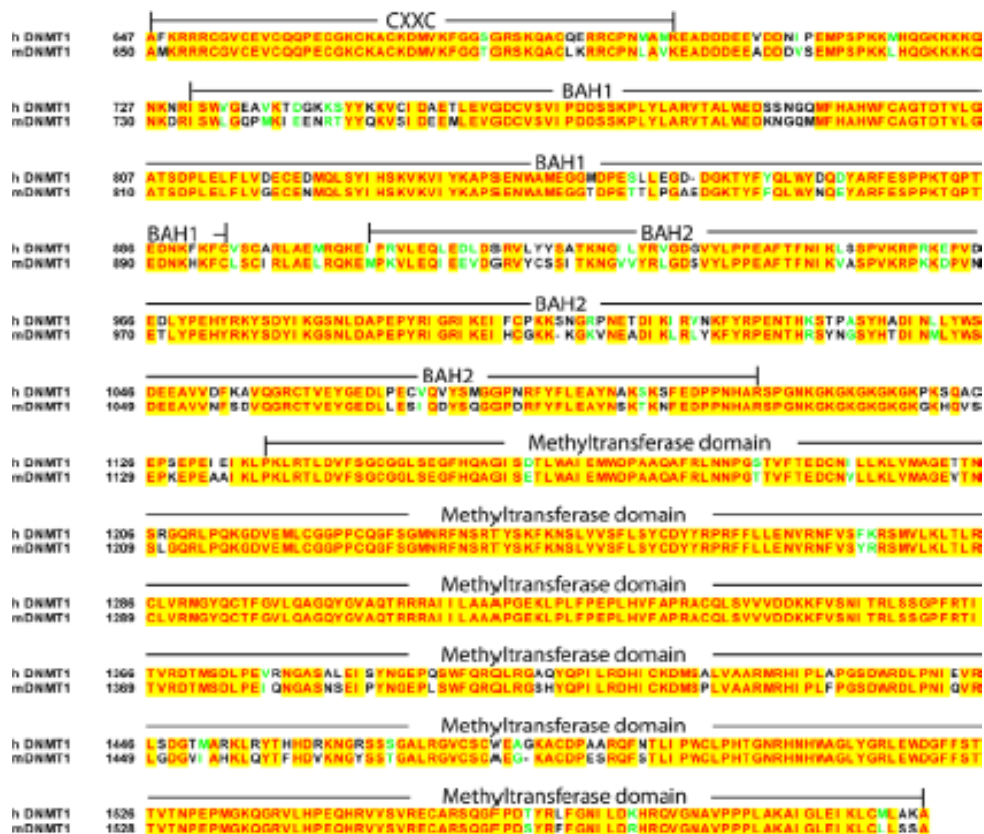
structural unit and is stabilized, in part, by a zinc finger residue, Cys<sub>3</sub>-His coordinated to Zn<sup>2+</sup> (Figure 1.8D & 1.9). There are intermolecular contacts between amino acid side chains described by Song et al. as primary contacts between arginine residue of the catalytic core, and phosphate groups flanking the (C4pG5)•(C5'pG4') segment of the unmethylated DNA duplex in the complex (Figure 1.10).

For catalytic activity, Dnmt1 needs at least 1000 amino acid residues from N-terminal to the catalytic region. So far, studies have shown that enzymes missing the first 508, or 621 amino acids are still active, but those missing 672 amino acids or more are no longer active.<sup>[87]</sup>



**Figure 1.5: Structural overview of mDnmt1(650–1602)–DNA 19-nucleotide oligomer complex with bound AdoHcy. (A) Color-coded domain architecture and numbering of mDnmt1 sequence. The thin vertical light blue bars indicate binding positions of zinc ions. (B) Ribbon representation of the complex in two orthogonal views. The CXXC, BAH1, BAH2, and methyltransferase domain are colored in red, light purple, orange and light blue, and DNA and zinc ions are colored in light brown and dark purple, respectively; CXXC-BAH1 linker in dark blue, BAH1-BAH2 linker in silver, (GK)<sub>n</sub>-containing BAH2-methyltransferase linker in black, and bound AdoHcy as in space-filling representation.<sup>[84]</sup> Permission requested**

Valuable insights of the catalytic activity and mechanism of reaction at this site have come from enzymatic kinetics, computational studies and the crystal structure of the bacterial enzyme M.HhaI bound to substrates and cofactor. In brief, the mechanism involves DNA binding, AdoMet (cofactor) binding, base flipping and the transfer of methyl group from AdoMet to the C5 position of Cytosine in the DNA.<sup>[85, 88]</sup> It is assumed that M.HhaI and Dnmt1 share a similar mechanism of reaction at the active site based on the conserved nature of the sequence motifs<sup>[86, 89]</sup>, the results of <sup>3</sup>H exchange reaction<sup>[90]</sup> and their inhibition by 5-fluorocytosine.<sup>[91]</sup>



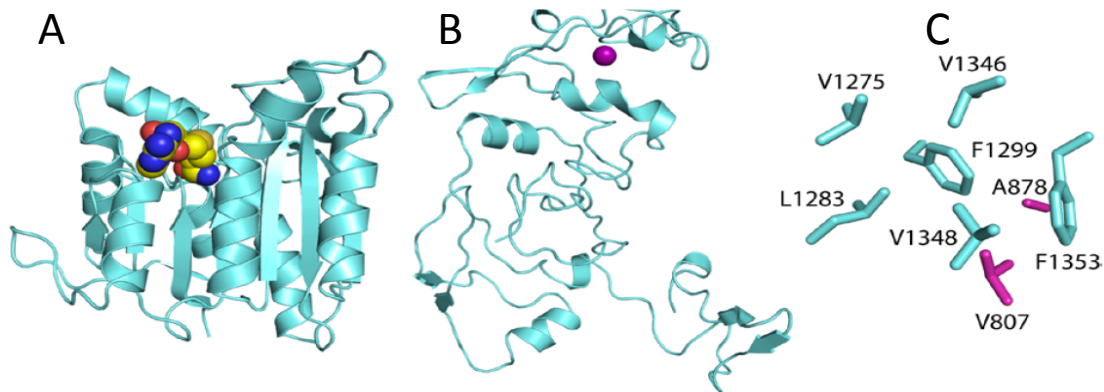
**Figure 1.6: Sequence alignment of the C-terminal segments of hDnmt1 and mDnmt1<sup>[84]</sup>**  
*Permission requested*

Furthermore, similar studies also suggest that the M.HaI and Dnmt1 share very similar catalytic processes at the C5 of the activated target base and including the methyl transfer step.<sup>[90b, 91]</sup> The methyl transfer step has been labeled the rate-limiting step for both enzymes<sup>[90b, 91]</sup>, however,



M.HhaI has ~100-fold faster catalytic rate than Dnmt1. It might be safe to speculate that the difference in catalytic rate might be due to the N-terminal in Dnmt1, which is absent in M.HhaI.<sup>[90b]</sup> Notwithstanding, some studies have attributed the difference in catalytic rate to be due to difference in rapid equilibrium between the initial steps of target base attack, particularly, the target recognition step, base-flipping step and formation of the unstable covalent adduct intermediate.<sup>[90a]</sup>

In addition, the allosteric regulation of Dnmt1 might lead to changes in the methylation rates as a result of changes in the rapid equilibrium during the early catalytic steps.<sup>[90b, 91]</sup> Lastly, the fact that HhaI flips the target cytosine base into the catalytic pocket has been considered as possible reason for difference in methylation rates between HhaI and hDnmt1, however, no base flipping experiments have been reported for Dnmt1, which if present might have a huge impact on hDnmt1 kinetics.<sup>[85]</sup>



**Figure 1.7: The catalytic core and TRD of the methyltransferase domain of the mDnmt1 (650-1602)-DNA complex 19-mer with bound AdoHcy. (A) Ribbon representation of mDnmt1 catalytic core, with bound AdoHcy shown in space filling representation (B) Ribbon representation of mDnmt1 TRD. (C), Hydrophobic contacts between mDnmt1 catalytic core (residues Val1275, Leu1283 and Phe1299 in cyan), TRD (residues Val1346, Val1348 and Phe1353 in cyan) and BAH1 (residues Val807 and Ala878 in magenta).<sup>[84]</sup> Permission requested**



## The Regulatory N-Terminal Domain

The first three quarters of the hDnmt1 contain the regulatory domains: the DNA methyltransferase associated protein 1 (DMAP1)-binding domain, the Replication Foci Targeting Sequence (RFTS) domain, the CXXC zinc binding domain and two Bromo-Adjacent Homology (BAH) domains<sup>[92]</sup> (**Figure 1.5**).

All sequence specific contacts with DNA are made via the **CXXC** domain. The CXXC domain, also called zinc domain, targets both the major and minor groove of the DNA over a CpG-containing 4-bp segment.<sup>[84]</sup> The CXXC residues (Arg684-Ser685-Lys686-Gln687) penetrate into the major groove to form base-specific and phosphodiester intermolecular interactions. Both guanine and cytosine bases in the CpG dinucleotide are recognized by side-chain interactions or the four CXXC residues mentioned above. Also, the DNA recognition is further strengthened by the salt bridges between arginine side chains of the CXXC domain and the phosphodiester backbone of the DNA. In summary, CXXC contributes to the catalytic action of Dnmt1 by recognizing and binding nonmethylated CpGs in the DNA.<sup>[84, 93]</sup>

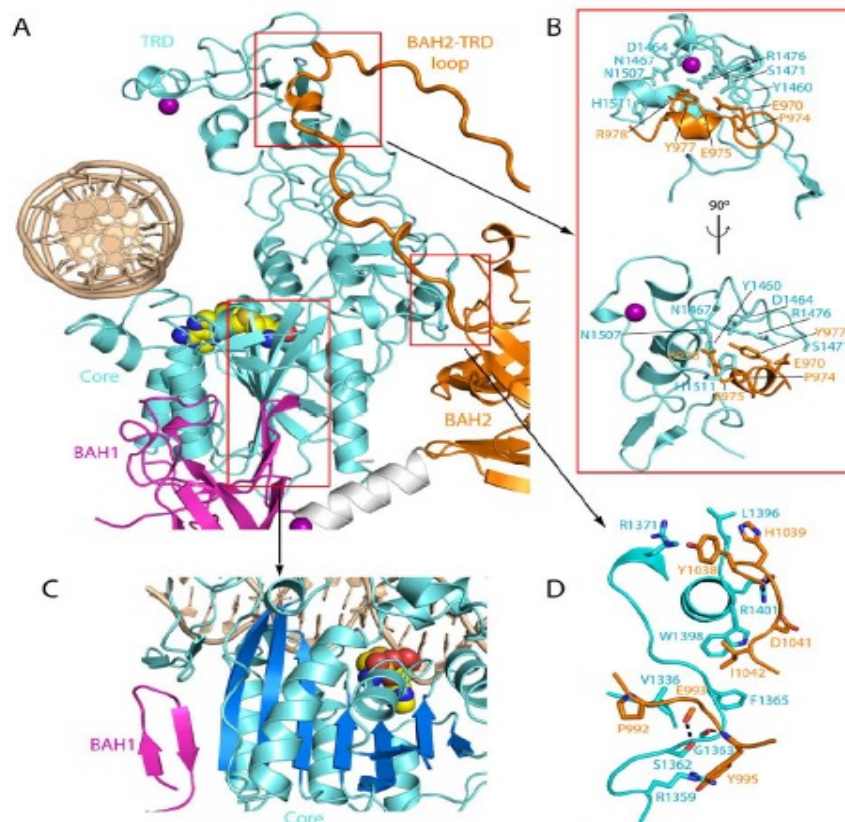
The BAH1 and BAH2 domains are separated by an  $\alpha$ -helical linker; both domains are on the surface of the protein remote from the bound DNA. Moreover, they are connected to the CXXC domain and the catalytic site via flexible loops and BAH1 is connected to a Zn ion in Cys<sub>3</sub>-His-coordination. This domain might contribute to the dynamic equilibrium required for Dnmt1 activity, but its functional role in Dnmt1 is unknown.

The function of the RFTS domain is also unclear but it is suspected to be involved in the proper targeting of Dnmt1 to the replicative foci during S phase of the cell cycle<sup>[94]</sup> and in the homodimerization of Dnmt1 which aids hemimethylated DNA binding.<sup>[95]</sup> Lastly, the DMAP1

binding domain of Dnmt1 binds DMAP which act as co-repressor of transcription by interacting with histone deacetylase 2 (HDAC2).<sup>[92]</sup>

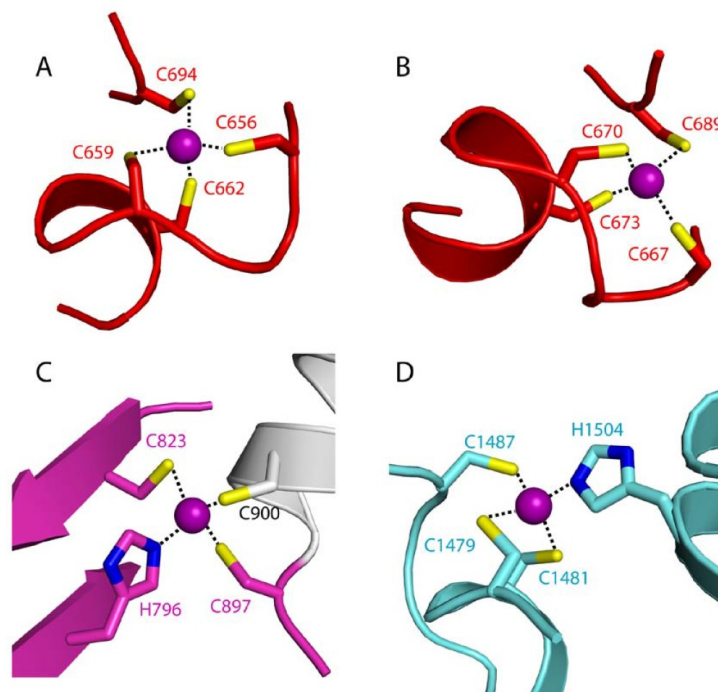
### N-Terminal Domain and Function

The N-terminal domain has multiple regulatory mechanisms that control the activity and specificity of DNA methylation. Its many roles can be summarized into three key functions: (i) allosteric regulation of the catalytic activity (ii) multiple phosphorylation and methylation sites

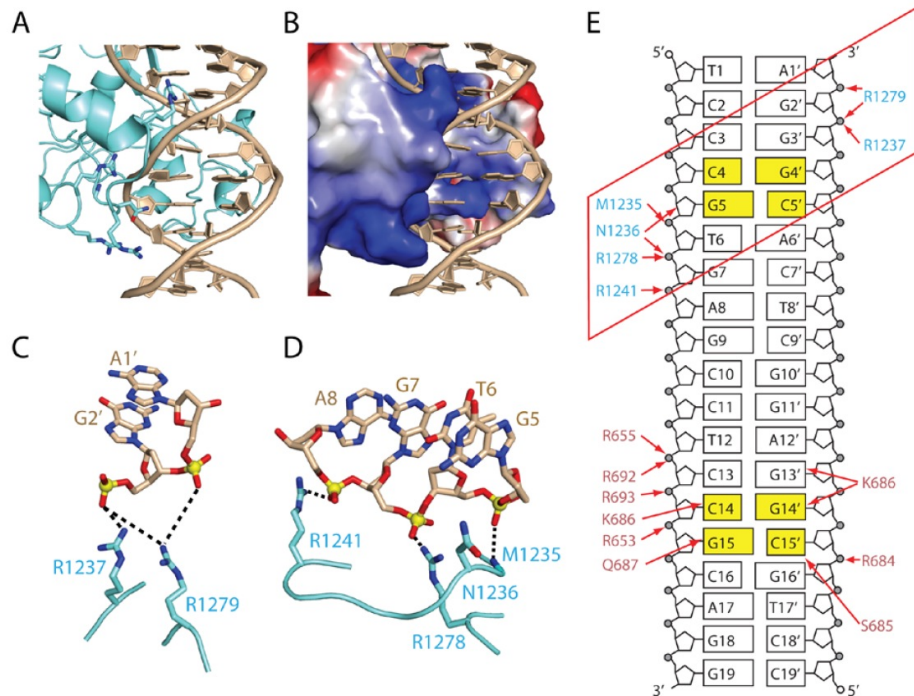


**Figure 1.8: Intramolecular contacts between both BAH domains and the methyltransferase domain in the mDnmt1(650-1602)-DNA 19-mer structure. A, Close-up view of the methyltransferase domain (in cyan), highlighting its interactions with DNA and both BAH1 (in magenta) and BAH2 (in orange) domains. B, Two views of anchoring of BAH2-TRD loop to the TRD of the methyltransferase domain. C, Two  $\beta$ -strands from the BAH1 domain (magenta) are paired with the central  $\beta$ -sheet (dark blue) of the catalytic core of the methyltransferase. The H-bonds are indicated by dashed lines. D, Residues around the C-terminus of the BAH2-TRD loop interact with the TRD of methyltransferase domain to further anchor the BAH2 domain to the TRD. The interactions of BAH1 and BAH2 with the methyltransferase domain lead to a buried surface area of 1148.7  $\text{\AA}^2$  and 887.3  $\text{\AA}^2$ , respectively.<sup>[84]</sup> Permission requested**

that regulate the catalytic activity of Dnmt1, and (iii) the many interactions between Dnmt1 and at least 31 other molecules that affect methylation.<sup>[96]</sup> The allosteric regulation of the catalytic pocket of Dnmt1 can be divided into two parts. One area of understanding was proposed by Song et al., who described an autoinhibition regulation of Dnmt1 at nonmethylated CpG sites on the DNA. Unmethylated DNA is excluded from the active site of mDnmt1 by binding with by CXXC domain; in addition, the presence of CXXC-BAH1 acidic linker prevents entrance of DNA into the catalytic pocket. The acidic linker is positioned between the DNA and the active site (Figure 1.11). Thus, the linker acts as a ‘gate keeper’ to the catalytic site.



**Figure 1.9:** The zinc finger clusters observed in the crystal structure of the mDnmt1(650-1602)-DNA 19-mer complex. A, The  $Zn^{2+}$  ion is coordinated by cysteine residues from both the N-terminal and Cterminal segments of the CXXC domain. B, The  $Zn^{2+}$  ion is coordinated by cysteine residues from the middle segment of the CXXC domain. C, The Cys3His-coordinated  $Zn^{2+}$  ion connects the BAH1 domain to the linker helix between BAH domains. D, The Cys3His coordinated  $Zn^{2+}$  ion in the TRD. The  $Zn^{2+}$  ions are shown in purple. The nitrogen and sulfur atoms are dark blue and yellow, respectively.<sup>[84]</sup> *Permission requested*

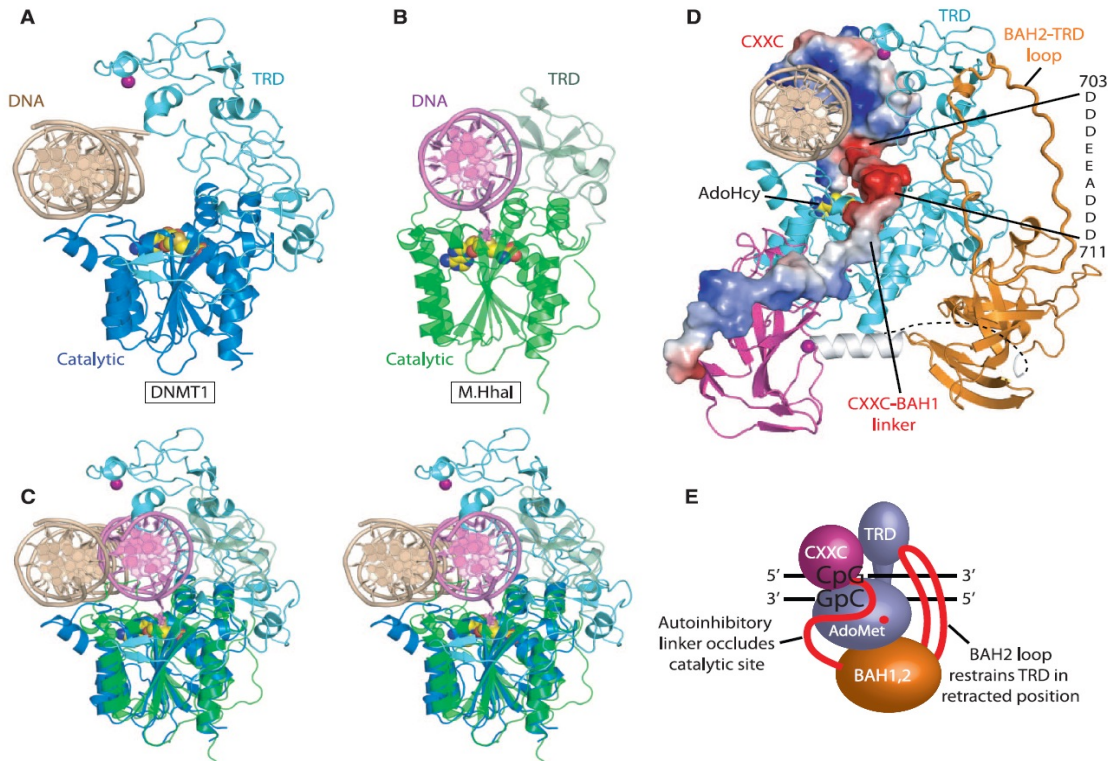


**Figure 1.10: Intermolecular contacts between the catalytic domain of mDnmt1(650-1602) and bound DNA 19-mer.** A, Ribbon representation of mDnmt1 catalytic domain bound to DNA 19-mer, with the side chains of DNA-interacting residues shown in stick representation. B, Surface electrostatic view of mDnmt1 catalytic domain bound to DNA. C, D, Intermolecular hydrogen bonding interactions between amino acid side chains and the DNA phosphodiester backbone. The nitrogen, oxygen and phosphate atoms are shown in dark blue, red and yellow, respectively. E, Schematic representation of intermolecular contacts in the mDnmt1(650-1602)-DNA 19-mer complex. The DNA region in contact with the catalytic domain is boxed by a red rectangle. The residue labels are colored according to their respective domains in Figure 1.11 <sup>184</sup> Permission requested

Furthermore, Song et al. proposed that nonmethylated CpG sites are protected from *de novo* methylation through binding with the CXXC domain<sup>[84]</sup> and this binding increases the efficiency of maintenance methylation through inhibition of *de novo* methylation. Song's proposal might support other studies that show Dnmt1's catalytic preference for hemimethylated DNA compared to unmethylated DNA<sup>[97]</sup>, however, Dnmt1 is also known to have a high kinetic preference for a poly(dI-dC) substrate<sup>[97a]</sup>. The other area of allosteric regulation of Dnmt1 is the activation of Dnmt1 with fully methylated DNA substrates. Fatemi et al. demonstrated that catalytic activity of murine Dnmt1 can be controlled by interaction of the N-terminal with fully methylated DNA.<sup>[98]</sup> Also, there are a number of other subtle pieces of evidence that have been



reported.<sup>[99]</sup> Some of these results vary due to enzyme purification procedures, but there is a general agreement of increase catalytic activity of Dnmt1 in the presence of fully methylated DNA



**Figure 1.11: Comparison of mDnmt1 with M.HhaI in their DNA-bound complexes. (A)** The crystal structure of the mDnmt1(650–1602)–DNA 19-nucleotide oligomer complex. The CXXC, BAH1, and BAH2 domains and CXXC-BAH1 linker of mDnmt1 have been removed for clarity. The bound DNA is in light brown, with the TRD and catalytic core in light and dark blue, respectively. **(B)** The crystal structure of the M.HhaI-DNA complex [PDB: 1MHT (8)]. The bound DNA is in light purple, with the TRD and catalytic core in pale and dark green, respectively. **(C)** Structural comparison of mDnmt1(650–1602)–DNA 19-nucleotide oligomer complex and M.HhaI-DNA complex in a stereo view looking down the DNA helix axis, after superposition of their methyltransferase domains. AdoHcy is shown in space-filling view. The everted cytosine in the M.HhaI complex is shown in ball-and-stick view in dark purple. **(D)** Electrostatic surface representation of mDnmt1 CXXC domain and the CXXC-BAH1 linker in the context of the structure of the mDnmt1(650–1602)–DNA 19-nucleotide oligomer complex. The BAH2-TRD loop is highlighted with thicker lines. **(E)** The proposed model for autoinhibitory mechanism in maintenance DNA methylation. In the autoinhibitory state, the CXXC domain and the autoinhibitory linker (in red) occlude the active site. In addition, the BAH2-TRD loop (in red) restrains the TRD in a retracted position so that it does not interact with CpG sites on the DNA.<sup>[84]</sup>

*Permission requested*

Another important process of the N-terminal regulatory action of Dnmt1 is phosphorylation. Interestingly, the N-terminal has multiple phosphorylation sites that can control the catalytic activity of Dnmt1.<sup>[100]</sup> It is speculated that phosphorylation sites can cause conformational changes that affect the flexibility of the protein structure allowing different interaction between the DNA and Dnmt1. For example, phosphorylation of Ser515 can inhibit the activity of mouse Dnmt1 by altering the interactions between different domains.<sup>[100]</sup> Another report, suggest that phosphorylation appears to inhibit Dnmt1 as a result of destabilization of the Dnmt1-DNA complex.<sup>[101]</sup> Also interesting is the fact that Dnmt1 missing the 501 amino acids have been shown to have higher activity than the full-length enzyme<sup>[102]</sup> and the missing phosphorylation regions might be a partial reason for this. In summary, the importance of phosphorylation of certain serine residues of the first 580 amino acid for catalytic activity of Dnmt1 is not yet conclusive but current evidence points to its involvement in the catalytic action of Dnmt1.

Lastly, we shall discuss the overwhelming number of interactions between the N-terminal of Dnmt1 and other molecules (**Table 1.1**). There have been some reports of interaction with the C-terminal of Dnmt1; one of such interaction is with co-chaperone p23, which binds to the C-terminal domain.<sup>[103]</sup> So far, Svedruzic has grouped the interactions of N-terminal Dnmt1 with other molecules into four groups based on their physiological functions: (i) Dnmt1 dimerization; (ii) core chromatin replication complex; (iii) interaction with molecules involved in DNA repair, cell cycle control, and apoptosis; and (iv) interaction with RNA Polymerase II, RNA-binding proteins, and specific RNA molecules.<sup>[96]</sup> We will discuss some examples from the four group mentioned above, in the coming paragraphs.

## Core chromatin replication complex

As part of the core chromatin replication complex, which preserves the existing epigenetic organization and chromatin structure, Dnmt1 interacts directly with proteins involved in

**Table 1.1 Summary of Molecules Currently Known to Interact Physically and/or Functional with Dnmt1<sup>[96]</sup>**

Core chromatin replication complex	DNA repair, cell cycle control, and regulation of apoptosis	RNA-directed DNA methylation
<ul style="list-style-type: none"> <li>_ Dnmt3a and Dnmt3b</li> <li>_ SNF2h-containing chromatin-remodeling complex NoRC</li> <li>_ LSH protein (lymphoid-specific helicases) protein related to the SNF2 family of chromatin-remodeling ATPases</li> <li>_ PCNA, DNA clamp processivity-promoting factor</li> <li>_ UHRF1</li> <li>_ HP1b, heterochromatin protein 1b isoform (chromobox protein)</li> <li>_ SUV39H1, histone-lysine N-methyltransferase</li> <li>_ G9a histone methyltransferase</li> <li>_ HDAC1 and HDAC2, histone deacetylase 1/2</li> <li>_ PML-RAR promyelocytic leukemia-retinoic acid receptor, oncogene transcription factor</li> <li>_ RIP140, metabolic repressor, also known as NRIP1 (nuclear receptor interacting protein 1)</li> <li>_ CFP1 CXXC finger protein 1 (PHD domain)</li> <li>_ MBD2/MBD3, methyl-CpG-binding domain protein</li> <li>_ PcG-EZH2 Polycomb-group proteins enhancer of Zeste homolog 2</li> </ul>	<ul style="list-style-type: none"> <li>_ PARP-1 (poly(ADP-ribose) polymerase 1) and poly(ADP-ribose)</li> <li>_ pRb/E2F1, Retinoblastoma tumor suppressor protein, control of G1/S transition and S-phase</li> <li>_ p53, tumor suppressor, regulation of cell cycle and apoptosis</li> <li>_ DMAP1, DNA methyltransferase 1-associated protein 1</li> <li>_ RGS6 member of mammalian RGS (regulator of G-protein signaling) proteins</li> <li>_ CK1d/E kinase that phosphorylates Dnmt1</li> <li>_ Annexin V, scaffolding proteins that anchors other proteins to the cell membrane and participates in apoptosis</li> <li>_ Hsp90, chaperon</li> <li>_ p23, cochaperone</li> <li>_ SET 7, protein lysine methyltransferase</li> <li>_ ATK1, serine/threonine protein kinase</li> </ul>	<ul style="list-style-type: none"> <li>_ MeCP2, methyl-CpG binding domain protein 2 (Rett syndrome)</li> <li>_ RNA Pol II</li> <li>_ Specific tRNA and mRNA</li> </ul>

*Permission requested*

chromatin remodeling. Dnmt1 functionally associates with (sucrose nonfermenting) SNF2h-containing chromatin-remodeling complex NoRC – chromatin remodelers machineries.<sup>[104]</sup> Also, the enzyme interacts with lymphoid-specific helicase (LSH) another chromatin remodeling enzyme of the SNF2 family. Interestingly, LSH function is so dependent on Dnmt1 that even inactive Dnmt1 is essential for it to function.<sup>[105]</sup> Such interactions with chromatin remodelers have been associated with possible interaction of Dnmt1 with two histone deacetylase proteins, HDAC1 and HDAC2 since histone modification and DNA methylation share a common function – gene silencing. Studies have confirmed direct interaction of Dnmt1 with the two histone

deacetylase.<sup>[106]</sup> Equally interesting, is the interaction of Dnmt1 with Dnmt3a and Dnmt3b. the latter interaction is possible because many proteins involved in histone modification that interact with Dnmt1 also interact with Dnmt3a and Dnmt3b.<sup>[82, 107]</sup> This finding questions our current understanding of classifying Dnmt1 as a maintenance methyl transferase and the Dnmt3 family of enzymes as *de novo* methyl transferases. Cell-based and enzyme-based assay suggest that both Dnmt1 and Dnmt3 enzymes are involved in *de novo* methylation reactions through an unknown mechanism<sup>[82, 102b]</sup>; maybe there is some cooperativity between the methyl transferase enzymes. Furthermore, Dnmt1 can have more than ten times *de novo* activity than Dnmt3 enzymes<sup>[108]</sup>, such finding compels researchers to suggest that the catalytic product of Dnmt3 can stimulate catalytic activity of Dnmt1.<sup>[109]</sup> Also, it has been observed that Dnmt3 bind the regulatory domain of Dnmt1<sup>[110]</sup>, which support the argument for possible stimulation of Dnmt1 by Dnmt3 enzymes.

The chromatin replication complex also houses an important enzyme ubiquitin-like PHD and ring finger domain1 (UHRF1), which recognizes a hemimethylated site on the DNA as Dnmt1 does. UHRF1 is a large multidomain protein that is central in the epigenetic code complex.<sup>[111]</sup> It can bind Dnmt1 and histone methyl transferase G9a through the SRA domain.<sup>[111-112]</sup> Two studies mapped the interaction parts of Dnmt1 and UHRF1 to either the SRA<sup>[113]</sup> or the PHD<sup>[111]</sup> domain of UHRF1 and the N-terminal part of Dnmt1. Separate research have shown that UHRF1 is highly important and it plays a central role in maintaining DNA methylation. UHRF1 knock-out embryos die shortly after gastrulation, which implies a significant reduced level of DNA methylation.<sup>[112]</sup> Similarly, UHRF1 knockout ES cell are viable despite global demethylation but fail to differentiate properly.<sup>[112]</sup> The current model based on available data, describes a mechanism in which UHRF1 recruits Dnmt1 to hemimethylated DNA to facilitate its efficient



remethylation. However, it is also known that Dnmt1 is recruited to active DNA replication sites through its interaction with Proliferating Cell Nuclear Antigen (PCNA), a processivity factor for DNA polymerases  $\delta$  and  $\beta$ <sup>[114]</sup>, which then facilitates the methylation of hemimethylated DNA in S phase.<sup>[115]</sup>

### **DNA repair, cell cycle control, and apoptosis**

The N-terminal of Dnmt1 interacts with protein complex associated with DNA repair, cell cycle control and apoptosis. Experiments have shown that the molecules that are involved in DNA damage response can stop DNA methylation at the level of Dnmt1 transcription<sup>[116]</sup> or by direct interaction with Dnmt1<sup>[101a]</sup> or regulate apoptosis in complex with Dnmt1.<sup>[106]</sup> For example, retinoblastoma protein (Rb) prevents excessive cell growth by inhibiting cell cycle progression until a cell is ready to divide. Interaction of Dnmt1 and Rb results in Dnmt1 inhibition and dissociation of Dnmt1-DNA complex.<sup>[117]</sup> The N-terminal of Dnmt1 binds to a specific pocket in Rb protein, and this interaction affects the allosteric regulation of Dnmt1.<sup>[87b]</sup> Another important interaction with the N-terminal of Dnmt1 is with p53, a tumor suppressor protein, which is important for DNA repair and apoptosis. Dnmt1 interaction with p53 affects the expression of the *surviving* gene, an inhibitor of apoptosis.<sup>[118]</sup> Also, Dnmt1 interaction with p53 might be involved in downregulation of protein phosphatases that regulate cell cycle.<sup>[119]</sup>

### **Interaction with RNA polymerase**

Lastly, Dnmt1 interactions with RNA and there is evidence that RNA molecules are involved in control of DNA methylation in mammalian cells.<sup>[120]</sup> Several studies demonstrate the interaction of Dnmt1 and RNA Pol II; and RNA Pol II is involved in synthesis of non-coding miRNA.<sup>[121]</sup> There is support for the interaction of Dnmt3 with non-coding micro RNAs.<sup>[122]</sup> Some other

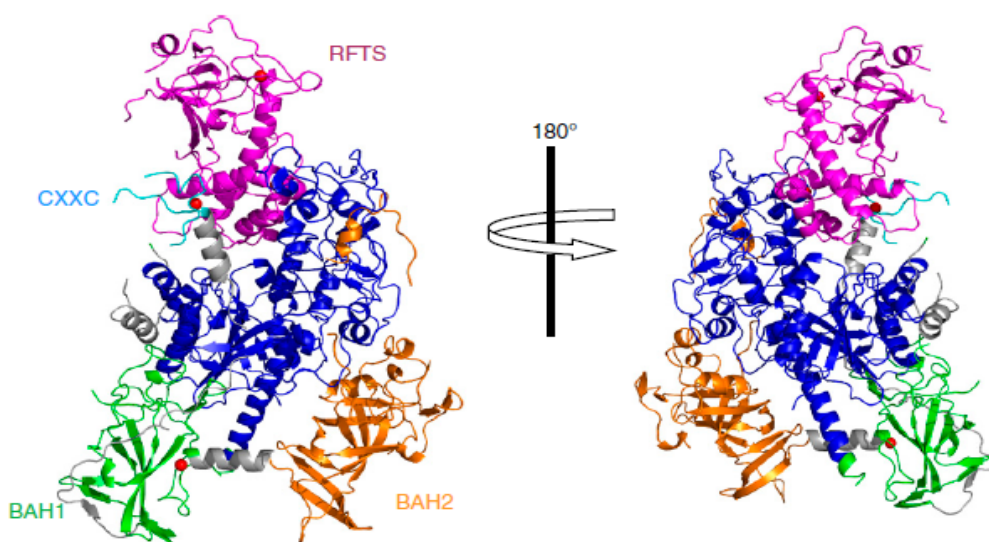
secondary interactions with Dnmt1 exist such as interaction with methyl-CpG-binding domain (MBD) family of proteins, which then bind several types of RNA molecules.<sup>[123]</sup> One notable example is Dnmt1 interaction with MeCP2, which is functionally important in neuronal development<sup>[33a]</sup>, as we discussed under Rett's syndrome. Despite the number studies done, the interaction of Dnmt1 and RNAs remains the least understood, but there is a strong regulatory connection between RNA molecules and the allosteric site on Dnmt1.

## 1.8 Regulation of Dnmt1

In the previous section we discussed various significant regulatory interactions of the N-terminal of Dnmt1 with a host of key proteins and complexes. We will discuss other specific examples of the N-terminal regulatory mechanism of mDnmt1, particularly the self-regulatory N-terminal of Dnmt1 and other interactions with BRCA, SET7 and then pharmacological inhibitors of Dnmt1.

Takeshita and colleagues solved the crystal structure of large fragment of Dnmt1 lacking only the first 290 residues of the N-terminal<sup>[124]</sup>, this large structure gave the first insight into the molecular mechanism of Dnmt1 N-terminal regulation by the TS domain. In this structure, the TS domain (Figure 1.12) is found deeply buried inside the catalytic domain, where it forms several hydrogen bonds, which is thought to prevent Dnmt1 from binding the hemimethylated DNA substrate.<sup>[124]</sup> This arrangement of domain would imply that activation of Dnmt1 requires several conformational changes or actual displacement of TS from the catalytic pocket to allow substrate binding. As a consequence, O'Gara, Syeda and colleagues have shown that deletion of TS lowers the activation energy, thereby increasing methylation<sup>[125]</sup>, and addition of purified TS domains inhibits Dnmt1 activity *in vitro*.<sup>[126]</sup> These findings clearly support the model of a self-regulatory N-terminal of Dnmt1, and corroborate the autoinhibitory function of TS domain

described by Takeshita et al.<sup>[124]</sup> In addition, the crystal structure obtained by Song et al. show that the CXXC motif is in a position where the DNA binds (**Figure 1.5**)<sup>[84]</sup>, which is also an auto-regulatory device that prevents binding of nonmethylated DNA to the catalytic pocket. In summary, from the mDnmt1 crystal structures obtained by Song and Takeshita, there is a strong indication that multiple structural changes occur, especially, in the N-terminal and probably modest structural changes in the C- terminal before faithful DNA methylation is achieved. Thus, Dnmt1 has a self-regulatory mechanism located in the N-terminal that controls catalytic activity of the C-terminal.

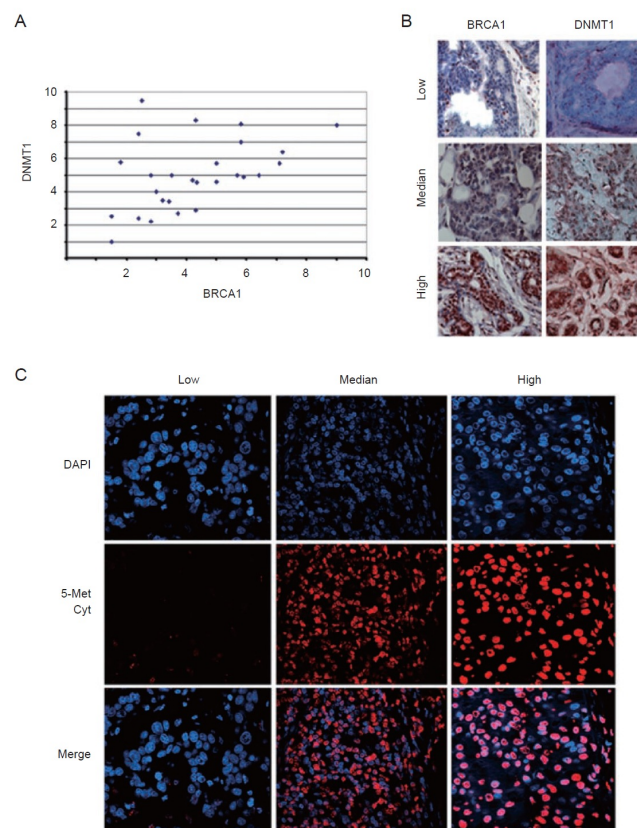


**Figure 1.12: Ribbon model of mouse Dnmt1(291–1620).** Around the C-terminal catalytic domain (blue), the other domains including the RFTS (magenta), CXXC motif (cyan), and two BAH domains BAH1 (green) and BAH2 (orange) are shown. Four zinc ions are shown in red spheres. All of the zinc ions are in a motif similar to Zn-finger motif (Fig. S10). The KG-repeat (1112–1124) linker connecting the N-terminal region and the C-terminal catalytic domain is in a flexible structure as the density map showed disorder.<sup>[124]</sup> *Permission requested*

In addition to the auto-regulatory action of Dnmt1, there is evidence that reveals an important role *BRCA1* plays in DNA methylation<sup>[116]</sup>, as well as many other biological processes including cell cycle regulation, DNA damage repair, transcription regulation, and centrosome duplication.

Shukla and colleagues discovered that the gene *Dnmt1* is a transcriptional target of BRCA1.<sup>[116]</sup>

BRCA1 and OCT1 (another transcriptional factor) bind to the promoter region of *Dnmt1* through a potential OCT1-binding site, AACGTTAA, which is found to be essential for maintaining a transcriptional active configuration of the promoter region in both mouse and human cells. Also, there is a positive correlation in the levels of global DNA methylation and expression levels of BRCA1 in human clinical samples (**Figure 1.13**)<sup>[116]</sup>



**Figure 1.13: Expression of Dnmt1 and BRCA1 in human primary breast cancers. (A) BRCA1 and Dnmt1 protein levels in 31 human sporadic breast cancers. Immunohistochemical staining of BRCA1 and Dnmt1 protein levels were detected by using antibodies against BRCA1 and Dnmt1. Intensity of the staining was scored by division into 10 arbitrary units based on the graded intensities. The x axis and y axis are for Dnmt1 and BRCA1 levels, respectively. (B) Examples of immunohistochemical staining showing positive correlation between expression levels of BRCA1 and Dnmt1 in human breast cancers. (C) Immuno-fluorescent images of varying methylation levels after 5-methyl cytosine staining, which was also proportional to levels of BRCA1 staining in all analyzed samples.**<sup>[116]</sup> *Permission requested*

Shukla further demonstrated that reduced function of BRCA1 leads to global DNA hypomethylation, loss of genomic imprinting, and an open chromatin configuration as observed

in many mutant mouse models. Their results indicate that BRCA1 is a positive regulator of Dnmt1 transcription through prevention of global hypomethylation – a hallmark of many cancers. Likewise, it supports the observation noticed in BRCA1 deficient cells, which show dire effects on cell viability and genomic stability, similar to effects noticed in Dnmt1 deficient cells<sup>[127]</sup>. Therefore, the outcome of the results of Shulka et al. further straightens the argument for a strong correlation between BRCA1 and Dnmt1 expression.

Equally considered, apart from basic interactions of Dnmt1 with proteins and other complexes are certain regulators that covalently modify Dnmt1. Two good examples of such modifications are phosphorylation and methylation.<sup>[100-101]</sup> So far, we briefly looked at phosphorylation of Ser515<sup>[100]</sup> at the beginning of this chapter, however, the protein kinases responsible for the phosphorylation of Dnmt1 were not discussed. We will look at the kinases elucidated recently by Sugiyama and colleagues.<sup>[101a]</sup> In brief, using monoclonal antibodies, multi-PK antibodies, Sugiyama identified CDKL5 (cyclin-dependent kinase-like 5) as the kinase protein responsible for phosphorylation of Dnmt1. But, the phosphorylation action of CDKL5 is weak and the importance of this phosphorylation remained uncertain, which then prompted the search for another kinase, CK1 $\delta/\epsilon$  (casein kinase 1  $\delta/\epsilon$ ) with stronger phosphorylation activity. Their results show that the casein kinase binds to and phosphorylates Dnmt1 at the N-terminal much more significantly than CDK15. Also, they identified the major phosphorylation site in the N-terminal to be Ser 146. In addition, DNA-binding affinity of the N-terminal of Dnmt1 was greatly reduced by phosphorylation with CK1 $\delta$ . In summary, the results from Sugiyama indicate that CK1  $\delta/\epsilon$  bind to and phosphorylates the N-terminal of Dnmt1, thereby regulating the functions of Dnmt1 through a reduction of DNA-binding activity.

Interestingly, Dnmt1 is also suffers methylation from other methyl transferase enzymes. A histone methyltransferase enzyme SET7 is known to colocalize and directly interact with Dnmt1, and it is responsible for the monomethylation of Lys 142<sup>[128]</sup> on the N-terminal of Dnmt1. In support of this modification is the fact that over expression of SET7 led to reduced levels of Dnmt1 and siRNA-mediated knockdown of SET7 stabilized Dnmt1.<sup>[128]</sup> The consequence of methylation of Dnmt1 at Lys142 during S and G2 phases of the cell cycle is its susceptibility to proteasome-mediated degradation, as notice by Esteve et al.<sup>[128]</sup> Implying, there is a robust interaction between SET7 and Dnmt1, and that Dnmt1 stability is regulated by protein methylation couple to proteasome-mediated protein degradation. Although, more work still has to be done *in vivo* to confirm this association between Dnmt1 and SET7, it is evident that covalent modification such as methylation of Dnmt1 can affect its reactivity towards DNA – binding.

Dnmt1 is also inhibited by poly(ADP-ribose) and Poly(ADP-ribose) polymerase 1 (PARP1).<sup>[129]</sup> Though the mechanism of inhibition is not known, Dnmt1 is said to belongs to a class of proteins able to bind, in a noncovalent manner, long and branched ADP-ribose polymers. ADP-ribose polymer, either free or PARP1-bound are able to inhibit Dnmt1. The current study by Caiafa et al.<sup>[129a]</sup> show that poly(ADP-ribosyl)ated PARP1 and Dnmt1 form complexes *in vivo* which are catalytically inept in DNA methylation. Likewise, Dnmt1 can also be inhibited by poly(dA).poly(dT) and poly(dA-dT), but not by poly(A) or poly(dA)<sup>[130]</sup>. It seems that most of the inhibitory action of these polymers is competitive, since they bind to the enzyme which precludes the binding of DNA.<sup>[129a]</sup>

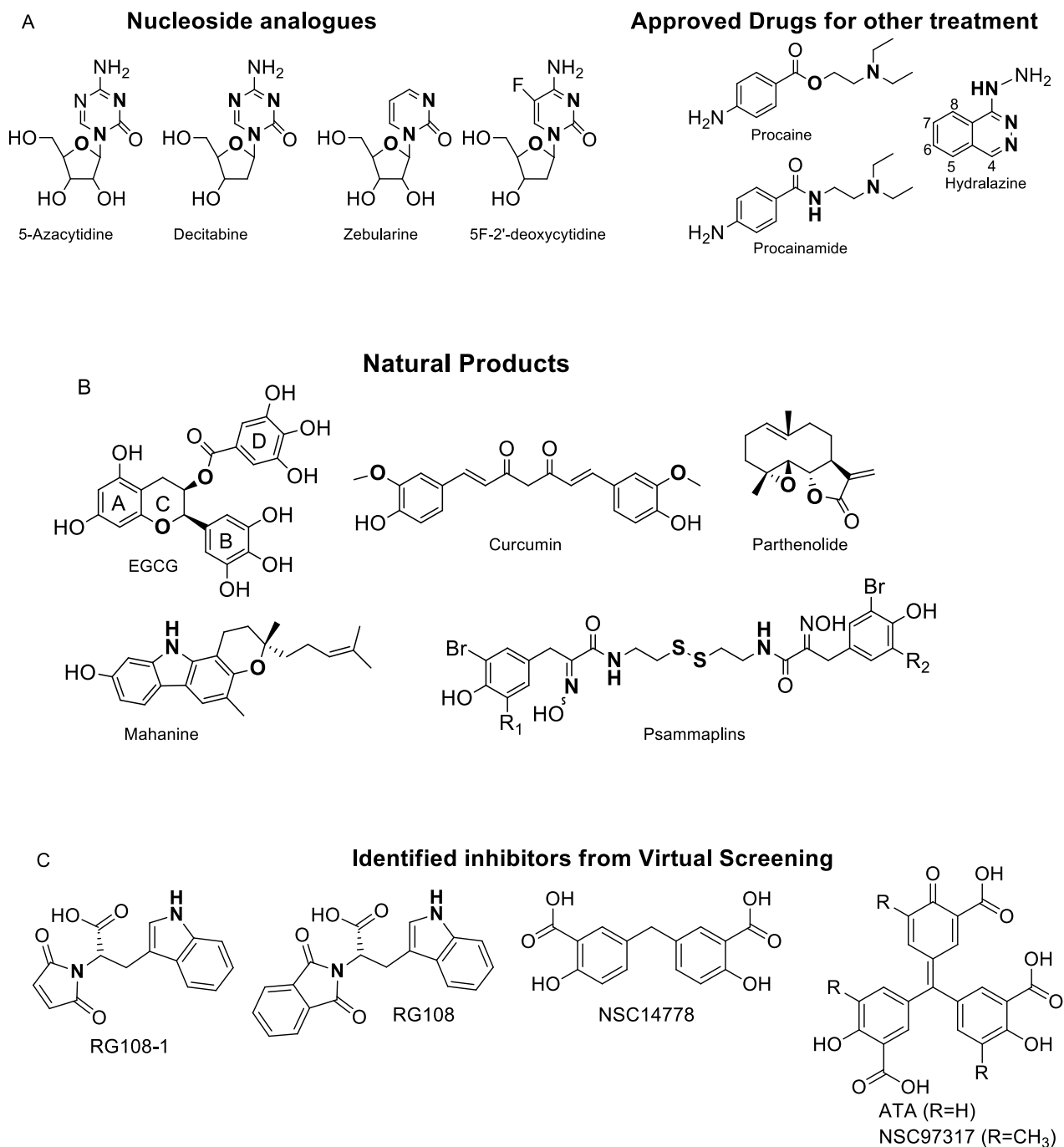
## 1.9 Small Molecules

Given the importance of Dnmt1 in various diseases including cancer, a number of other inhibitors have been designed and currently being optimized to take advantage of the mechanism of methylation and the conserved residues in the catalytic pocket of Dnmt1. Cytosine analogues are the best known active inhibitors based on the mechanism of Dnmt1 methylation.<sup>[131]</sup> They include classical nucleoside inhibitors such as 5-azacytidine (AzaC)<sup>[132]</sup>, 5-fluorocytosine, (FdC)<sup>[133]</sup>, and zebularine<sup>[134]</sup> that show antiproliferative activity against cancer cells. These inhibitors cannot inhibit Dnmt1 directly, but have to be incorporated into the DNA, where they covalently trap Dnmt1 in the process of DNA methylation.<sup>[134]</sup> They are referred to as “suicide inhibitors” and their use is associated with excessive DNA damage and a high level of toxicity, which in turn limits their dosage to low concentrations during cancer therapy.<sup>[88e]</sup> However, there are continuing efforts to improve pharmacology of cytosine analogues because of their features as lead compounds; particularly there are efforts to prevent the toxic incorporation into genomic DNA, but to still maintain the valuable suicide inhibitory property.

The other main class of inhibitors are non-nucleoside covalent or non-covalent inhibitors, which include natural products such as the dietary polyphenols like (–)-epigallocatechin-3-gallate (EGCG)<sup>[135]</sup>, curcumin<sup>[136]</sup>, parthenolide<sup>[137]</sup>, and mahanine<sup>[138]</sup>. Also, included in this class are small molecules and amino acid derivatives, for example the bisulfide bromotyrosine derivative psammaplins<sup>[139]</sup>, L-tryptophan derivative RG108<sup>[140]</sup>, hydralazine<sup>[141]</sup>, and the 4-aminobenzoic acid derivatives procaine and procainamide.<sup>[142]</sup> The last members of this class are novel inhibitors identified from virtual screening (**Figure 1.14**); they include compounds from various databases that are virtually screened against the crystal structure of Dnmt1. We will revisit this



class of compounds in chapter two and discussed in detail the principle behind computational identification of lead compounds in drug discovery.



**Figure 1.14: Compounds tested for inhibitory activity against Dnmt1 A. Promising nucleoside and approved drugs in clinical trials as Dnmt1 inhibitors B. Known natural product with great activity against various cancers C. New class of molecules from computation studies**



**Table 1.2 Summary of Available IC<sub>50</sub> and Ki values of Known Dnmt1 Inhibitors and how these compounds affect M.SssI and MCF-7 Cells.**

Dnmt Inhibitors	M.SssI		Dnmt1 from M.SssI		MCF-7	
	Ki/ μM	IC <sub>50</sub> / μM	Ki/ μM	IC <sub>50</sub> / μM	Ki/ μM	IC <sub>50</sub> / μM
Kazinol Q <sup>[143]</sup>	-	-	-	7.0	-	2.0
EGCG <sup>[143]</sup>	-	-	-	3.0	6.89	2.0
EGCG <sup>[135]</sup>	-	-	-	20.0	-	-
Clofarabin <sup>[144]</sup>	-	-	-	-	-	640 nM
5-aza-2'-deoxycytidine <sup>[145]</sup>	-	-	-	-	-	0.6
Fludarabine <sup>[145]</sup>	-	-	-	-	-	15
2-Chloro-2'-deoxyadenosine <sup>[145]</sup>	-	-	-	-	-	0.2
Psammaplins A <sup>[139]</sup>	-	-	-	18.6nM	-	-
Aurintricarboxylic acid <sup>[146]</sup>	-	-	-	0.68	-	-
RG108 <sup>[140a]</sup>	-	-	-	115nM	-	-
Trimethylaurintricarboxylic acid (NSC97317) <sup>[146]</sup>	-	-	-	4.79	-	-
5-aza-2'-deoxycytidine <sup>[145]</sup>	-	-	-	-	-	4.0
9-beta-D-arabinofuanosyl-2-fluoroadenine (Fludarabine) <sup>[145]</sup>	-	-	-	-	-	4.0
Procainamide <sup>[147]</sup>	-	5.0	-	3.5	-	-
Curcumin <sup>[136, 148]</sup>	-	30nM	-	-	-	<10.0
SGI-1027 <sup>[149]</sup>	-	-	-	6-12.5	-	-
methylenedisalicylic acid <sup>[146]</sup>	-	-	-	92	-	-
Parthenolide <sup>[150]</sup>	-	5.0	-	3.5	-	-
Δ <sup>2</sup> -isoxazoline <sup>[151]</sup>	-	-	-	150	-	-
Antroquinonol D <sup>[152]</sup>	-	-	-	5.0	-	-
Procainamide <sup>[147]</sup>	-	-	7.2±0.6	-	-	-

## 1.10 Breast and Colon Cancer and DNA Methylation

For this project we will narrow our focus to design and testing of Dnmt1 inhibitor in two cancer types: colorectal and breast cancer. Our discussion will be centered on key genes that are hypermethylated in both cancers and then center more on preliminary results involving breast cancer.

## Breast Cancer

Breast cancer is a common type of cancer occurring in both sexes, but it is a more common malignancy in women and it is greatly influenced by hormonal factors.<sup>[153]</sup> So far, DNA methylation has been implicated in breast cancer progression as well as in a number other cancers including colorectal cancer. Some important genes inactivated in breast cancer by aberrant DNA methylation are summarized in **Table 1.3**. For example, estrogen receptor (ER)-negative breast cancer cells show widespread methylation and elevated DNA methyltransferase expression compared with (ER)-positive cells, indicating the importance of the hormonal factor but implicating Dnmt enzymes in breast cancers.<sup>[154]</sup> Ottavia and colleagues found that the *ER* gene CpG island undergoes aberrant methylation, as noticed in established ER-negative breast cancer cell lines and primary tumors.<sup>[153]</sup> Also, an elaborate study by Parl et al. showed that the *ER* gene was unmethylated at the *NotI* site CpG island in all normal tissues such as thyroid, whole lung, bronchial epithelium, and cervix but are highly methylated and showed a unique pattern in ER-negative cell lines and MCF-7/AdrR. In addition, treatment of the ER-negative cell line MDA-MD231 with DNA methylation inhibitors led to demethylation of the ER CpG island and partial restoration of gene expression of the estrogen receptor<sup>[155]</sup>. Furthermore, detailed studies showed that expression of Dnmt1 at both RNA and protein level was high in ER-negative breast cancer cell lines (BCC) compared with ER-positive. However, while Dnmt1 expression was highly correlated with S-phase in ER-positive cells, ER-negative cells expressed Dnmt1 throughout the cell cycle; suggesting a dysregulation of Dnmt1 in ER-negative BCC.<sup>[154]</sup>

**Table 1.3 Frequency of methylation of representative tumor suppressor and growth regulatory gene in breast cancer**

Gene	Function	Incidence (%)	Reference
<i>P16<sup>INK4a</sup></i>	Cyclin-dependent Kinase	15	Herman et al. (1995)
<i>14-3-3<math>\sigma</math></i>	G2 checkpoint	91	Ferguson et al. (2000)
<i>ER<math>\alpha</math></i>	Steroid receptor	50	Ottavaino et al (1994)
<i>PR</i>	Steroid receptor	40	Lapidus et al. (1996)
<i>RAR<math>\beta</math>2</i>	Steroid receptor	25	Sirchia et al. (2000)
<i>BRCA1</i>	DNA damage repair	15	Dobrovic & Simfendorfer (1997)
<i>GSTP1</i>	Carcinogen detoxification	30	Esteller et al. (1998)
<i>E-cadherin</i>	Epithelial cell-cell adhesion	50	Graff et al. (1995)
<i>TIMP-3</i>	Inhibition of MMPs	25	Bachman et al. (199)

*Permission requested*

Moving down the list in **Table 1.3**, we come across *GSTP1*, another important gene in BCC, which, however, shows an inverse relationship with *ER*. This is because glutathione (GSH) and other cytosolic GSTs are involved in detoxification of xenobiotic and chemotherapeutic agents<sup>[156]</sup>, basically, they catalyze intracellular detoxification reactions by conjugating chemically reactive electrophiles to GSH, and thus inactivating the electrophilic carcinogen.<sup>[157]</sup> Experiments showed that *GSTP1* was expressed in ER-negative but not in ER-positive lines.<sup>[158]</sup> *GSTP1*-negative cell line MCF-7 treated with 5-aza-dC could induce mRNA expression and *de novo* synthesis of  $\pi$  - class GSP. Also, Esteller et al., demonstrated that *GSTP1* promoter methylation is associated with gene silencing in ~ 30% of primary breast carcinomas.<sup>[49a]</sup> Lastly, Cavalieri and colleagues proposed that methylation-associated inactivation of *GSTP1* can result in A or G base mutation by estrogen metabolites-DNA adduct formation that results in genetic instability.<sup>[159]</sup>

From **Table 1.3**, the *BRCA1* gene is a well-known breast cancer susceptibility gene and inherited mutations in this gene account for one-half of inherited carcinomas.<sup>[160]</sup> Methylation of *BRCA1* has been proposed to silence the gene and cause reduced expression of BRCA1 protein in sporadic breast cancer.<sup>[161]</sup> A study with 194 primary breast carcinomas showed that the *BRCA1* promoter is methylated in 13% of unselected primary breast tumors<sup>[50]</sup> and as expected the BRCA1 of normal cells remain unmethylated. So far, BRCA1 hypermethylation has been detected exclusively in sporadic breast and ovarian cancers but not in tumors or colon or liver or leukemia; however there are isolated sporadic lung cancer cases.<sup>[41]</sup>

An equally notable gene like *BRCA1* is *TIMP-3*, which codes for a family of molecules that inhibit the proteolytic activity of matrix metalloproteins (MMPs).<sup>[162]</sup> These proteins can suppress primary tumor growth by effects on tumor development, angiogenesis, invasion and metastasis.<sup>[163]</sup> Bechman et al., showed methylation of the promoter region of *TIMP-3* in ~ 30% of human BCC lines and also in ~ 30% of primary breast tumors.<sup>[39, 163]</sup> Lastly, the loss of expression of *TIMP-3* was restored by treating affected cells with 5-aza-dC, demonstrating the tight correlation between epigenetic gene silencing and cancer.

### **Colorectal Cancer**

Normally, colorectal cancer (CRC) develops through a multistage process as a result of the continuous accumulation of mutations and frequently in the Wnt pathway – a signal transduction pathway that pass signals from outside a cell to inside the cell. Recent studies are mounting evidence of the involvement of epigenetic alterations of the chromatin structure in CRC development; particularly, modifications that lead to tumor suppressor gene silencing and activation of oncogenes.<sup>[164]</sup> CRCs can be grouped according to genomic instability; for example,

one group is chromosomal instability (CIN) which is characterized by aneuploidy and chromosomal gains and loss; while microsatellite instable (MSI) is another and its identified by frequent insertion and deletion mutations in the repetitive DNA sequence.<sup>[165]</sup> Polyp progression starts initially with a mutation or inactivation of the adenomatous polyposis coli, *APC*, gene that results in activation of Wnt pathway, which leads to the progression of polyps to cancer by further mutations in other genes like *KRAS* and *TP53*<sup>[166]</sup> which unfortunately grapples apoptosis in the presence of DNA damage. However, apart from mutations, epigenetic modifications like DNA methylation and histone modification play key role in CRC. For example, methylation in the promoter region of mismatch repair gene like *MLH1* has been shown to be associated with MSI – high sporadic CRC<sup>[167]</sup> and DNA methylation-related deficiencies in the mismatch repair system can lead to mutation rates that are 100-fold greater than normal cells. Also, defects in mismatch repair system can lead to germ line mutations in human's genes such as *MHL1* and *MSH2* and sometimes, *MSH6*, *PMS1* and *PMS2* as noticed in hereditary nonpolyposis colorectal cancer (HNPCC), which ensures an increase rate of progression in adenomatous polyps.<sup>[168]</sup>

As a corollary, a distinct group of CRC called CpG island “methylator” phenotype (CIMP) has been introduced to differentiate it from “mutator” phenotype that was known before.<sup>[169]</sup> The most commonly used gene markers for this subgroup include *MLH1*, *p16*, *MINT1*, *MINT2*, *MINT31*, *CACNA1G*, *CRABP1*, *IGF2*, *NEUROG1*, *RUNX3*, *SOCS1*, *HIC1*, *IGFBP3*, or *WRN*, with no consensus on how many genes are required to be defined as CIMP positive.<sup>[170]</sup> However, CRC group can be classified as CIMP-high, CIMP-low and non-CIMP depending on the cancer-specific hypermethylation involved and the mutation rate of certain genes.<sup>[171]</sup>

In summary, from various studies involving CRCs, it is obvious that CRC-specific hypermethylation affects a number of genes and pathways including *TP53*, *P13K*, *WNT*, *IGF*

signaling DNA repair/stability, apoptosis, angiogenesis, invasion and metastasis.<sup>[170a, 172]</sup> Similarly, DNA methylation-mediated gene silencing, also targets and affects many of the signaling pathways involved in CRC just as genetic mutation. Thus, the biomarkers can be used to develop test for CRC diagnosis and epigenetic drugs for CRC therapy.

## CHAPTER 2

### DRUG DISCOVERY AND DESIGN OF Dnmt1 INHIBITORS

#### 2.0 Introduction

Modern drug discovery, especially at the early stage, involves finding new compounds for identified biological targets such as enzymes, receptors, ion channels, transporters and DNA targets.<sup>[173]</sup> Usually, the target is druggable – that is, the target can potentially bind with high affinity to the drug and elicit therapeutic response both *in vitro* and *in vivo* – and can pass a target validation screen, which implies the target is involved in the disease and can be acted upon by the drug. Target validation can be done *in vitro* such as finding drugs that modulate the target via enzyme-based assay or by using model interactions – that is using computational methods. Also, *in vivo* validation methods such as animal model testing and gene knockout methods.<sup>[174]</sup> Usually, a multi-validation approach increases the confidence in the observed outcome for validating a target. Following the target validation, the search for ‘hit’ molecules and the process of drug design begins. A hit molecule has been defined as a molecule that shows desired activity during screening and the activity is reproducible upon retesting.<sup>[175]</sup> A number of screening methods are used to identify hit molecules during drug discovery and we shall discuss a few of them. High throughput screening (HTS) is one method of hit identification which involves screening an entire diverse library of compounds directly, typically by cell-based assay, against a drug target. Usually, the conventional 96-well plates are replaced with 384-well microplates<sup>[176]</sup>, novel fluorescence-based detection system are used<sup>[176]</sup> and screening robots used are fully adapted to desktop environment.<sup>[177]</sup> The HTS method is made possible by the use of laboratory automation and robot technology to screen hundreds of thousands of compounds to discover a

small group of hits that will be later confirmed with a secondary assay for further development into leads.<sup>[178]</sup> On the other hand, focused or knowledge based screen involves testing a small class of compounds that are known to hit specific class of targets (e.g. kinases) and compounds with similar structure. The focus based approach is based on knowledge of protein target and literature studies that point to a chemical class of compounds likely to have chemical activity for the target protein.<sup>[179]</sup> There are other methods such as fragment screen<sup>[180]</sup>, physiological screens and NMR screens<sup>[181]</sup> which are equally important in drug discovery. As a complement to functional screening we have virtual screening, which are current rational drug design methods used in drug discovery and for the identification of Dnmt1 inhibitors. In general, rational drug design has been greatly accelerated by advances in genomics and proteomics<sup>[182]</sup> and bioinformatics<sup>[183]</sup>, which in turn has made the isolation of targets and determination of X-ray crystal structures possible. Also of great importance in rational drug design is the advancement in cheminformatics, computing power and molecular modeling<sup>[184]</sup>, which couple nicely to the structural data of the crystal structure and are immensely useful in drug discovery.

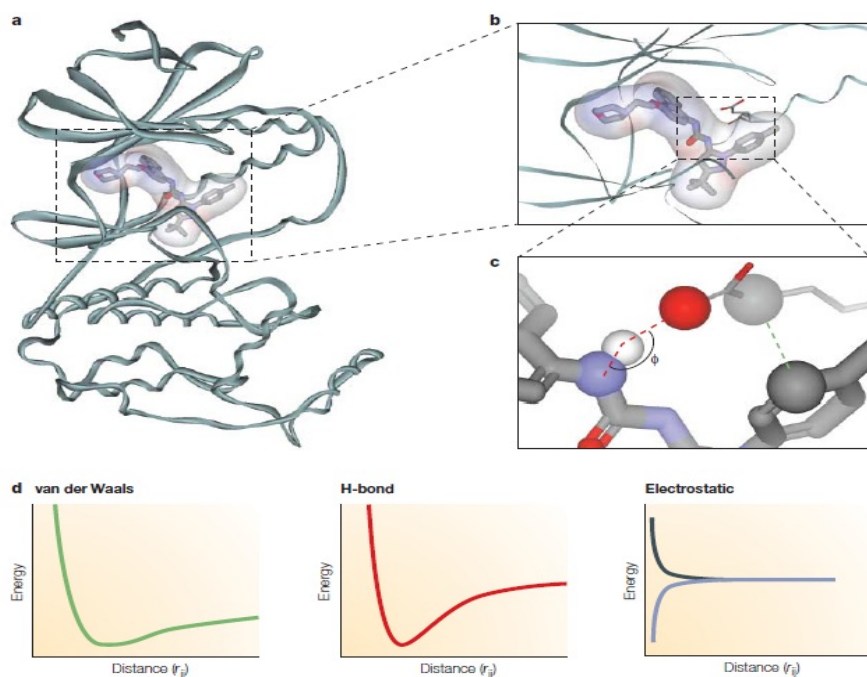
## 2.1 Structure Aided Drug design (SADD)

Structure aided drug design or structure-based drug design is an alternative to high throughput screening (HTS), and by contrast it is a knowledge-driven approach compared to HTS, which is a trial and error (experimental) based approach. The quality and amount of information regarding the target receptor and ligand is critical to SADD as well as in other computer-assisted drug design methods.<sup>[185]</sup> Structural information of the drug target can be resolved by X-ray crystallography or obtained from NMR spectroscopy. Also, homology model generated structures can be deduced from relevant structural data of target receptor family member(s), however, it is important that homology models are validated before use to reduce false positives



during screening for hits and lead optimization. SADD works best when the crystal structural data of the target receptor, and preferably to which the ligand is bound, is available. Such data would provide a starting point for structure activity relationship (SAR) since it gives insight into the geometric fit of the active ligand inside receptor and maybe information of a bound cofactor if present. Other important information from structure of the receptor include low-energy ligand conformation, molecular electrostatic potential<sup>[186]</sup>, types of hydrogen donor or acceptors, functional groups and hydrophobic interactions between lipophilic groups.<sup>[187]</sup> During drug design, the structural information gleaned from the ligand bound conformation can help guide rational subtle functional group alterations of new ligands either by visualization or the use of molecular modeling software. A good example is the rational design based substantially on molecular visualization and modeling is the identification of potent, bio-available, nonpeptide cyclic ureas as HIV protease inhibitor.<sup>[188]</sup> The success of the design was a single act of replacing and mimicking a structural water molecule by adding to the ligand carbonyl oxygen that mimics the hydrogen-bonding feature of the structural water, which was first hypothesized by careful molecular visualization. In generally, computer-aided drug design methods are used to model ligands *in silico* inside the active site of a receptor in order to achieve best theoretical fit between the defined binding sites on the ligand and the complementary molecular space in the target. More often than not, the original structural data of the ligand is removed from the binding site of the receptor and new molecular structures are computationally docked into the binding site. A typical *in silico* drug design process involves iterative cycles of docking, scoring and ranking based on how well the ligand docks with the active site in terms of alignment, hydrogen bonding, van der Waals forces, and electrostatics and hydrophobic interactions<sup>[189]</sup> (**Figure 2.1**). Docking entails conformational and orientational sampling of the ligand within the constraints of

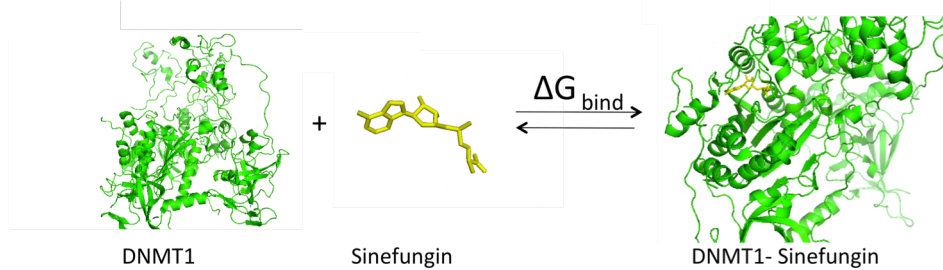
the receptor site. The scoring function for virtual screening selects the best pose such as ligand, orientation and translation for a given molecule and then ranks each ligand from top to bottom.<sup>[190]</sup> Notably, ranking involves a complex calculation that attempts to estimate the energy of the binding affinities (non-covalent interaction) of ligands to receptors. The estimated energies thus give an approximate representation of the underlying biochemical and biophysical phenomena, which is by no means perfect. So far, there are three main classes of scoring function which include force field-based, empirical or knowledge-based scoring function.<sup>[191]</sup>



**Figure 2.1: Modelling molecular recognition.** a | Structure of p38 mitogen-activated protein kinase with bound inhibitor BIRB796 (PDB code: 1KV2). The inhibitor is shown with its electrostatic potential surface. b | Enlarged view of the active site. c | Closeup view of the interaction between residue Glu71 and BIRB796. Hydrogen bonding (H-bond) and van der Waals interactions are colour-coded red and green, respectively. d | Schematic representation of functions used to model pair-wise interactions that contribute to binding. Interactions are calculated as a function of the distance ( $r_{ij}$ ) between two atoms  $i$  and  $j$ . Left of part d: van der Waals interaction given by a 12–6 Lennard–Jones potential (note the smoother attractive part of the potential compared to hydrogen bond term). Middle of part d: hydrogen-bond potential given by a ‘harder’ 12–10 Lennard–Jones potential (see also BOX 2). This term is angle-dependent (as indicated in c). Right of part d: electrostatic potential for two like (blue) or opposite (black) charges of same magnitude calculated using a distance-dependent dielectric constant of  $4r$ . *Permission requested*

## Theory of docking

The aim of docking is to correctly predict the structure of the protein-ligand complex  $[E + I] = [EI]$  under equilibrium conditions. For example, if sinefungin is docked in Dnmt1 (PDB: 3WSR) then:



$$\Delta G = -RT \ln K_A \quad K_A = K_i^{-1} = \frac{[EI]}{[E][I]} \quad (2,3)$$

The free energy of bind ( $\Delta G$ ) is related to binding affinity by equation 2 and 3:

*Where:*

G is the Gibbs free energy

R is the gas constant

$K_A$  is the equilibrium constant of association

[E] is the concentration free enzyme

[I] is the concentration of free ligand

[EI] is the concentration of the complex

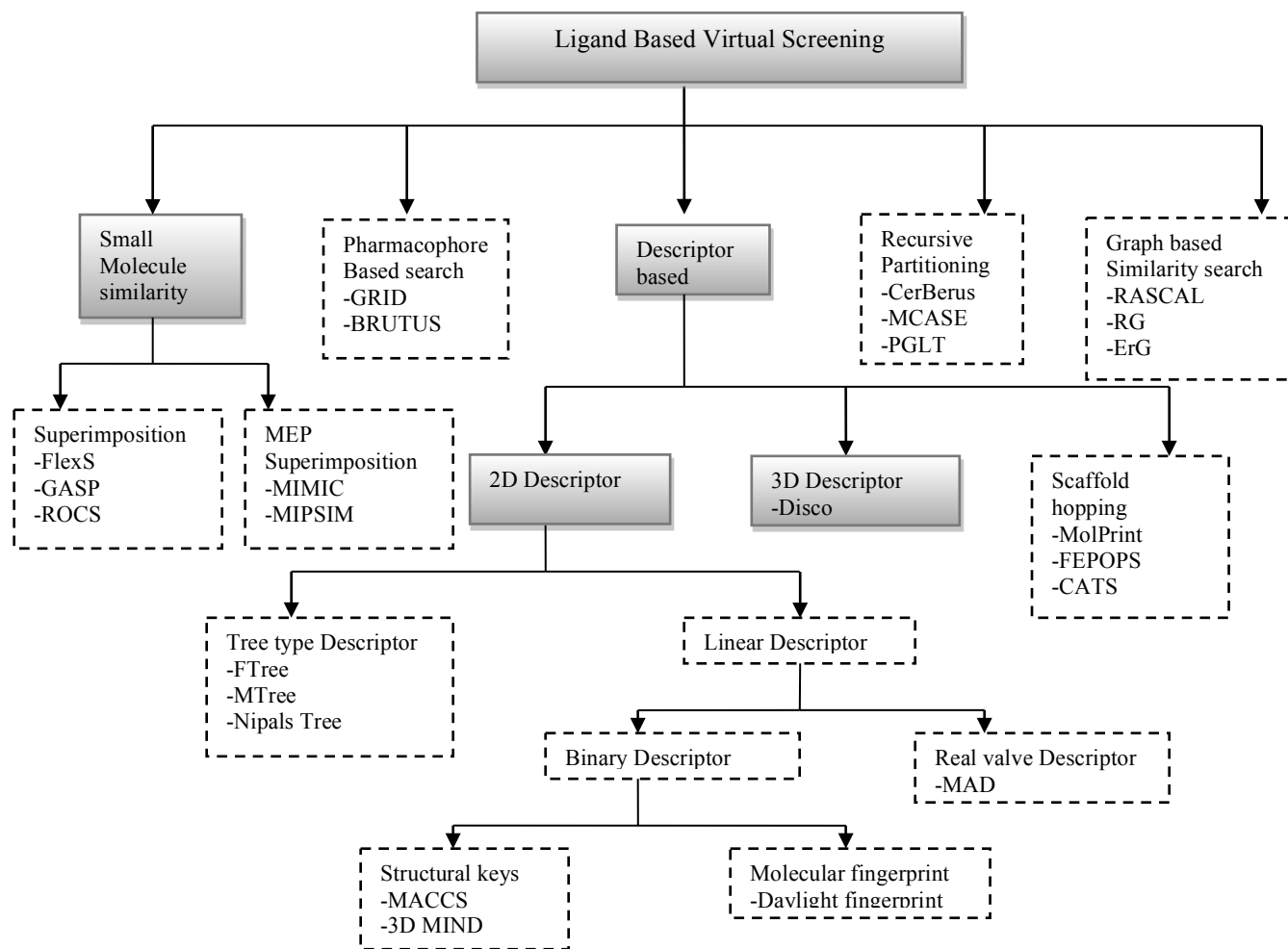
Docking must predict accurately two things relative to experimentally available data: the pose of the ligand structure and the binding affinity. Prediction of the correct structure (posing) of the [E+I] complex does not require information about  $K_A$ . However, prediction of biological activity (ranking) requires this information. Thus the scoring terms can also be divided in two: one that considers the term [EI], which takes certain factors into consideration such as steric, electrostatic hydrogen bonding, inhibitor strain (if flexible) and enzyme strain. The factors are considered as part of the contributing factors to a change in free energy that occurs during docking; this is because the scoring function predicts the ligand affinities by a free energy perturbation (FEB) method.<sup>[192]</sup> In the FEB method, the free energy change is calculated by changing one system into another (from the unbound state transition to a bound state) via a set of unphysical mixed

states using molecular dynamics (MD) or Monte Carlo simulations. Unfortunately, the results do not cover all the changes that occur during docking and virtual screening and thus it gives relative binding affinities for similar drugs docked in the same protein.<sup>[192-193]</sup> Undoubtedly, scoring functions for ligand affinities still remains a challenge for docking and drug design. On the other hand, from the **equation 1** the scoring function of the ligand poses inside the target at equilibrium do not require knowledge of  $K_A$ , however, a different set of factors become important such as desolvation (posing is in water), rotational entropy and translational entropy (degrees of freedom of the ligand) and the calculation is not as complex.<sup>[192]</sup> It is obvious that the final equation for the scoring functions would be a lot more complex than equation 2, however, we will not dwell much on it but a number of reviews are available<sup>[192, 194]</sup> for more in depth discussions.

One alternative approach that complements structure-based drug design is ligand-based drug design (LBDD) (**Figure 2.2**); the method uses the information present in known active ligand for lead identification and optimization. Normally, LBDD is computation based and requires structural information of both the ligand and the target receptor, however, it can be used even when the three-dimensional (3-D) protein structure of the target receptor is not available, typically as in the case of most G-protein-coupled receptors (GPCRs) targets<sup>[195]</sup> or protein structure resolve in an apo form<sup>[196]</sup>, that is a protein that is not ligand bound. In the presence of structural data of the ligand and maybe the target receptor, similar computational techniques such as virtual screening and molecular modeling can be applied in LBDD. By using the molecular fingerprint of known ligands, databases can be screened to find molecules with similar fingerprints.<sup>[197]</sup> The docking processes for LBDD can generally be divided into three classes (**Figure 2.2**): (i) quantitative structure activity relationship (QSAR) which is a method that

relates chemical structure to biological or chemical activity using mathematical models.<sup>[198]</sup> The quantitative part of QSAR is described by a set of properties, usually known as descriptors, which is computed from the structure and in turn used to describe the structure. In addition, by using structural descriptors as independent variables and activity as dependable variable, a model can be built to describe the relationship between the two.<sup>[199]</sup> Once the QSAR model is built and validated, then it can be used to screen the biological activity of novel molecules from their structural properties. Development of anti-allergic pyranen-amines by Cramer at Smith Kline and French Laboratories is one of the finest applications of QSAR methodology.<sup>[200]</sup> Also, in the area of cancer research QSAR has been used to develop a number of antitumor agents and inhibitors of cancer related enzymes<sup>[201]</sup> (ii) pharmacophore models try to describe the interaction between the active site of the target receptor and a ligand. It involves taking the essential features responsible for a drug's biological activity<sup>[202]</sup> and modeling them. Usually, types, position and direction of the active ligand are encoded into the pharmacophore model, along with possible steric constraints of the active compound.<sup>[203]</sup> The pharmacophore model has been applied to a number of drug design process, even as filters to reduce false positives during preprocessing of compounds before virtual screening; and for output processing after virtual screening protocols.<sup>[204]</sup> Lastly, pharmacophore model in addition with QSAR has been used to design a series of novel benzimidazole and imidazole inhibitors of histone deacetylase 2.<sup>[205]</sup> (iii) Similarity searching method operates on the principle that structurally related molecules are likely to have similar properties.<sup>[206]</sup> In addition, similarity search can be based on other features such as physiochemical properties as well as two and three-dimensional <sup>[207]</sup> features selected from QSAR or pharmacophore model. The chemical fingerprints is appropriate if only compounds structurally similar to the active compounds are desired, but if new scaffold/chemical

structure <sup>[208]</sup> are desired then physiochemical properties may be useful, One other usefulness of similarity search is when the number of ligands for which biological activity is known are too few to build a QSAR mode, then similarity search is applied.

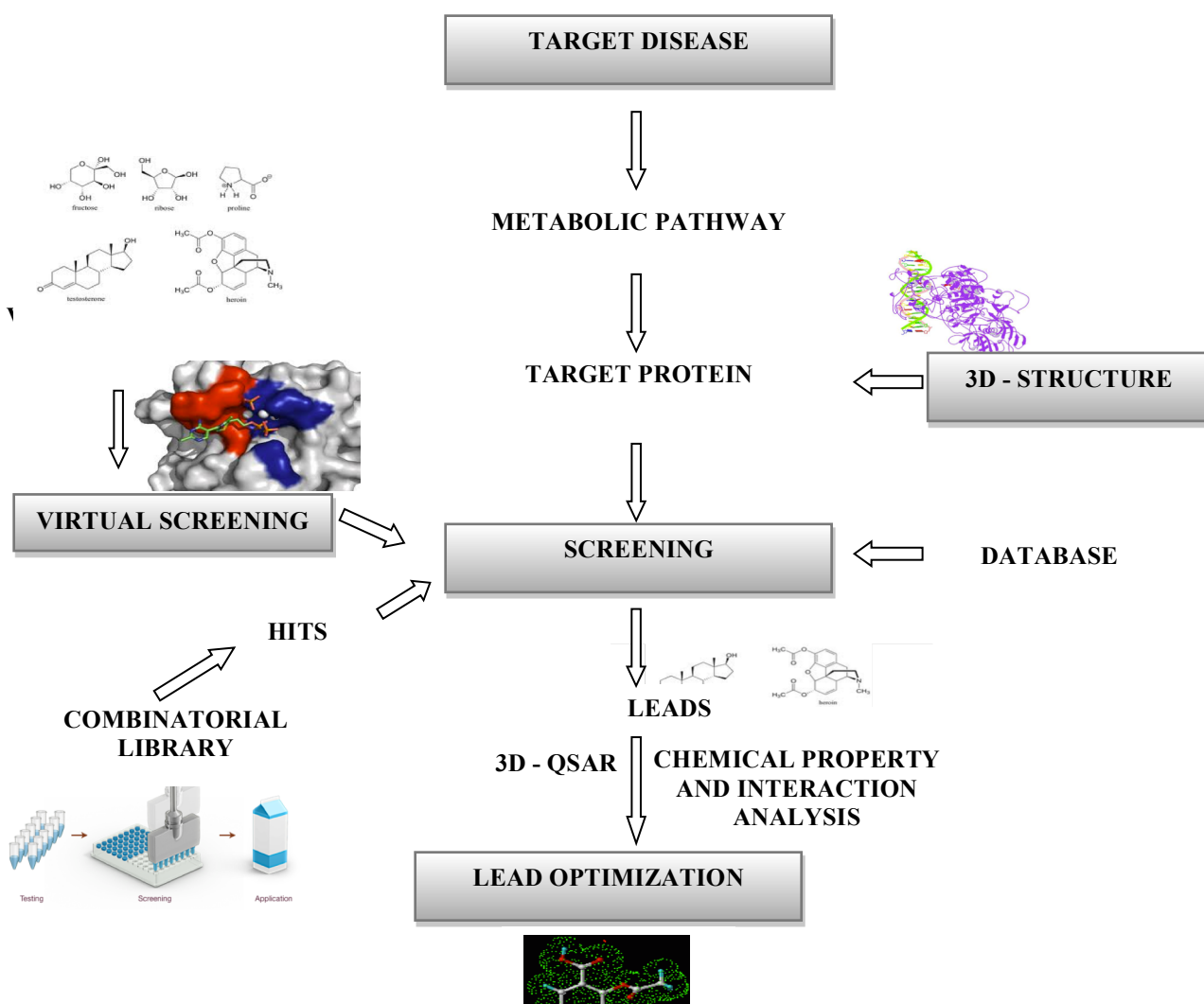


**Figure 2.2: Classification of various methods of Ligand based virtual**

## Virtual Screening

Drug discovery employs LBDD and SBDD approaches and both methods use virtual screening (VS) as a tool to search and optimize hits to potential leads; this makes VS an important

component in modern drug design and warrants further discussion. In general, the use of computer-assisted technique(s) to search through a large database of compounds for new biological active molecules is described as virtual screening.<sup>[206]</sup> The screening process is aimed at enriching a small subset of active compounds of pharmaceutical importance that can be developed into drug candidates.



**Figure 2.3: Drug discovery and rational drug design**

VS enrichment process in a way to complements HTS (**Figure 2.3**), and in some cases, serves as a good alternative to HTS, especially, when adequate knowledge of the three-dimensional

structure of the target receptor and small molecule binding modes are known. Given the structural data of receptor and ligand either ligand-based or structure-based virtual screening methods can be used to obtain potential hits using rapid docking algorithms to place available compounds into the active site of the receptor (as in the case of SBVS) or model the available compounds onto the active ligand template (as in the case of LBVS). For our discussion we are interested in VS, challenges, methodologies and the docking programs used in structure-based virtual screening. So far, only SBVS has been used for identification of Dnmt1 inhibitors, however, we believe that LBVS can also be utilized but we will save the proposal for future discussion.

## 2.2 Structure-based virtual screening

### Challenges

Structure-based virtual screening plays an important role in hit identification and lead optimization both in industry and academia. Over the years so many docking algorithms and scoring functions have been developed and refined to mimic true solution binding affinities of ligands to proteins, however, there are still some challenges. We shall highlight a couple of them in this section. Since docking and scoring functions are mathematic approximations of physical ligand-protein interactions, it is crucial that certain factors and conditions be considered before and during docking in order to obtain accurate results. One of such condition is the quality of the crystal structure used for virtual screening. Ideally, the receptor target for SBVS is an enzyme for which the crystal structure has been solve to a resolution of at most 2.5Å (preferably: placement of residues in the electron density map should be unambiguous) and the active site has been



characterized with a variety of probable hydrogen bond donors and acceptors, hydrophobic interactions, defined interstitial and structure water molecules (ordered water), and with molecular adherence surfaces.<sup>[208b]</sup> Although, this is not always the case with most receptors, in any case, a high quality receptor structure is fundamental to the success of SBVS in lead discovery. Also, apart from using a structure with high resolution, the algorithm of the docking programs should allow for receptor flexibility, which is one factor that was lacking previously in many docking programs until recently.<sup>[209]</sup> Most docking programs rely on the Lock-and-Key hypothesis for protein ligand binding instead of the induced fit model, however, it is now widely accepted that ligands and proteins are quite flexible in solution and may experience different assemblage of conformations compared to what is used during rigid *in silico* screening.<sup>[201c, 210]</sup> In addition to the flexibility of receptor structure, it is important to consider the effect of solvent during docking. Solvent plays a crucial role in docking, for instance ordered water molecule can be incorporated into the design of the ligand or treated as a bound ligand to which interactions with ligand of interest are probable.<sup>[188, 211]</sup> The solvent effect can also be incorporated into the scoring function approximation for binding affinities, however, most docking algorithm make the assumption that the molecules are in a vacuum instead of in a solvent. Furthermore, water molecules on the interface of target protein may facilitate increase in specificity or binding affinity of ligands<sup>[212]</sup> and so it is important that docking programs account for both bulk and specific water molecules. There is also the question of specific water molecules found in cavity of proteins. Thilagavethi and Mancera showed that inclusion of conserved water molecules during ligand-protein cross-docking led to a statistically significant improvement in the accuracy of predicted docking poses. The success rate of predicting correct pose went from 39% in the absence of water to 69% in presence of conserved water

molecules.<sup>[213]</sup> Also, Sotriffer and colleagues tested the structural prediction ability of two docking programs as a function of crystal water (ordered water) in the cavity of cytochrome *c* peroxidase W191G mutant and found very good results.<sup>[214]</sup> On the other hand, there is also the problem of sampling algorithm been, at least potentially, biased towards input ligand conformation.<sup>[190]</sup> Cross and colleagues compared the input conformation of ligands from CORINA (structure generation software) with known crystal ligands and, surprisingly, noticed that crystal structures, on average, had overall better docking results.<sup>[215]</sup> Kellenberger also noted that the input coordinates of ligands have an impact on docking poses.<sup>[216]</sup> Also, there is evidence that ligand speciation such as protonation, tautomerism and isomerism are important in docking. Ten et al. investigated the influence of protonated ligand state on docking and their result showed that different protomers and stereoisomers affected the scoring function; more so, docking programs are just learning how to deal with speciation of ligand species.<sup>[215]</sup> The above discussed challenges are by no means all the problems experienced by current virtual screening algorithms and programs but it succinctly describes the present performance state of VS in lead discovery and drug design. It is also important to remember that structure-based drug design methods directs the discovery of drug leads, which are not drug products, but are primarily active compounds enough with at least micromolar affinity to a target.<sup>[209a]</sup>

### **Methodology and docking programs**

Despite the challenges facing virtual screening, it is still a widely used tool to screen large libraries of compounds for identification of hits in drug design. The primary objective of virtual screening is to dock and score large libraries of compounds and rank them into active and inactive compounds, and the degree to which active compounds can be ranked ahead of inactive compounds is referred to as “enrichment”.<sup>[217]</sup> The most promising structures retrieved from

virtual screening depend on the method of docking and scoring function used. In this section, we will focus on the methodologies – preparation steps – docking and docking programs used to position ligands inside single receptor structures. In general, the preparation process in docking involves ligand and protein preparation, while the docking process involves posing a flexible ligand inside a rigid, or a partially flexible, receptor.<sup>[191]</sup>

## **Ligand Preparation**

Ligand preparation starts with the question of which variety of chemical structures should be included in the chemical library used to search the chemical space. Normally, a wide range of filters are applied to remove compounds with unfavorable pharmacodynamics or pharmacokinetics<sup>[218]</sup> and to reduce computational demands during the docking step. The initial selection criteria for organic molecules are drug-likeness, which implies that compounds are considered based on Lipinski's Rule of Five and the absorption, distribution, metabolism and excretion (ADME) restriction.<sup>[219]</sup> However, the rule does not predict if a compound would be active or not. Drug-likeness also includes filtering compounds based structure flexibility, most drugs show a balance between flexibility and rigidity. It is important to filter out toxic structural features but include important bio-recognizable groups in drug molecules.<sup>[220]</sup> An alternative or complementary filter method to drug-likeness is active-ligand similarity method.<sup>[221]</sup> Based on structural similarity search among small molecules, it is possible to retrieve privileged structures<sup>[221]</sup> containing identical substructures that share close binding affinity for the receptor. In addition, a three-dimensional descriptor of the active ligand can be included during similarity search.<sup>[222]</sup> The similarity search approach involves using appropriate molecular “fingerprints” to compute searches; an example of such fingerprint is Tanimoto coefficient<sup>[223]</sup> which has been used to improve the efficiency of similarity searches.

Furthermore, Bender and Glen<sup>[222a]</sup> have classified fingerprints according to their dimensionality ranging from one-dimension to three-dimension; Zhang and Muegge<sup>[224]</sup> have pointed out that two-dimensional fingerprint (most popular fingerprint) are more successful in retrieving active compounds. Another equally used similarity search method is the pharmacophore model, that is, geometric and topological constraints can be applied to the active compound, which can then be used as a search or filter criteria to screen for compounds.<sup>[225]</sup> Given the variability of target receptors and ligands in drug discovery filtering criteria varies widely and sometimes it depends on the quality and amount of knowledge available of ligand and receptor, but in general, it is advised to start with a vast and structurally diverse library of compounds and then follow the funnel strategy<sup>[226]</sup> to extract or achieve an enrichment of compounds for lead development.

Once a reliable and dedicated filter process is utilized and a diverse library has been selected, three-dimensional coordinate structures of all the compounds are generated and prepared in the same way as the ligand template to be used for the screening. The ligand preparation process can be as simple as adding hydrogen(s), formal charges and number of rotatable bonds or as thorough as including ionization/tautomerism<sup>[217]</sup> protonation, and stereoisomerism to the ligands, which can be done using ligand preparative software such as LigPrep.<sup>[217, 227]</sup>

### **Protein Preparation**

X-ray crystallography and NMR structures are obtained from the protein data bank (PDB)<sup>[228]</sup>, while homology models are generated using computer programs or devoted web servers such as SWISS-MODEL<sup>[229]</sup> and MODELLER.<sup>[230]</sup> Wallner and Ellofson have reviewed homology modeling programs such as SegMod/ENCAD, 3D-JIGSAW, NEST and builder giving the

strength and weakness of each program.<sup>[231]</sup> Normally, the starting point for protein preparation involves downloading and evaluating the receptor structures, usually, with programs such as PROCHECK.<sup>[232]</sup> Equally, homology models can be evaluated by similar software such as PROCHECK, but in addition there are a number of servers such as PSVS, Evals123D and JCSG Structure Validation Central available for three-dimensional model evaluation.<sup>[233]</sup> Once the receptor structure has been validated, further preparatory steps such as adding of hydrogens, checking for correct formal charges and bond orders are carried out. Additional protein preparation would depend on the docking software available and methodology; for instance Maestro<sup>[234]</sup> goes ahead and does hydrogen-bond network optimization. After which a protein minimization of the optimized structure is performed followed by the removal of water molecules, which is a similar step in the preparation protocol in AUTODOCK.<sup>[235]</sup> Most programs do not yet know how to treat water molecules in the target receptor, however the water molecule is removed because it is difficult to computationally determine which water molecule to retain; and primarily because the free energy of water molecule is not directly connect to crystallographic occupancy.<sup>[236]</sup> On the other hand, we had discussed earlier of the importance of structured water in drug design so; maybe future computation algorithms would be able to tackle the structure water problem.

## **Docking**

After filtering, selecting and preparing the target receptor and ligand for screening, a docking software and method is chosen. Currently, there are several free and commercially available docking programs; and the program of choice depend on the receptor target and which programs gives dependable results. From various reviews in the literature, the general consensus is that they is no universal docking program, the success of each program varies widely with different

ligands and receptors types. The strength and weakness of docking programs based on the work of Kellenberger et al. <sup>[214]</sup> is summarized in **Table 2.1**. In addition, Yuriev and Ramsland review on latest developments in molecular docking gives a more detailed and current review of docking programs and the novel features of each.<sup>[212]</sup>

The screening process involves docking of the prepared ligands into the catalytic site of the receptor crystallographic structure. Docking entails using computational programs to predict ligand conformation and orientation (or pose) that best matches the receptor's active site. Most docking programs generate several conformations for each ligand based on the number of rotatable bonds on the ligand (flexibility) by sampling the degrees of their freedom.<sup>[191]</sup> Initially both the target and the ligand were treated as rigid bodies during generation of poses; this was because the imposed rigidity reduces the computation requirement and simplifies the docking process. But the rigid treatment approach did not produce accurate results that reflect experimental observation. Consequently, most docking programs now treat ligand as flexible entities and protein as rigid structures but with some measure of flexibility, especially certain residues in the catalytic pocket. The treatment of ligand flexibility and poses can be divided into three categories<sup>[237]</sup>: systematic (incremental construction, conformational search, databases); random or stochastic methods (Monte Carlos, genetic algorithms, tabu search); and simulation methods (molecular dynamics, energy minimization)

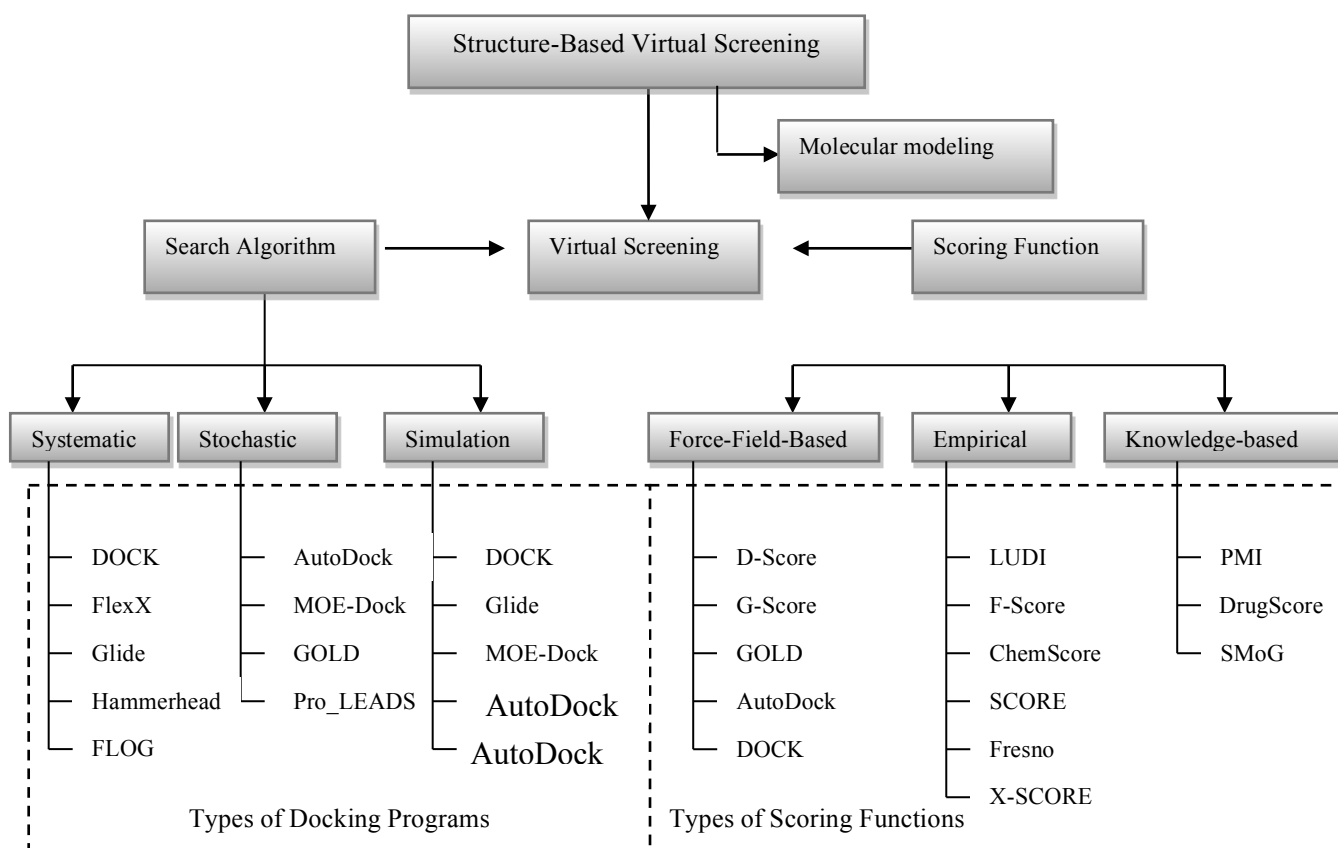
**Table 2.1: A list of widely used protein-ligand docking software. The main strengths and weaknesses are based on the work by Kellebeger et al., 2004<sup>[216]</sup>[371], and the selection of target studied by high-throughput virtual screening is followed favorable experimental testing. GA stands for Genetic Algorithm, HF for Hierachial Filtering, IC for Incremental Construction, MA for Matching Algorithm and MC for Monte Carlo sampling (Moitessier et al., 2009)**

Program and algorithm	Strengths	Weakness	Examples of recent successful virtual screening experiments
Autodock and AutoDock Vina (Morris et al., 1998; Osterberg, 2002; Trott and Olson, 2010) -GA	Small ligands Large binding sites Freely available	accuracy in highly flexible ligands low speed	Glutamate Transporter 1 (GLT1) inhibitors <sup>[238]</sup> Cdc25 phosphatase inhibitors <sup>[239]</sup> D-Ala:D-Ala ligase inhibitors <sup>[240]</sup> Cyclodextrin-based receptors <sup>[241]</sup>
DOCK (Ewing et al., 2001; Kuntz et al., 1982; Lang et al., 2009; Moustakas et al., 2006; Oshiro et al., 1995) -IC	Small binding sites opened cavities small hydrophobic ligands freely available	accuracy in highly flexible ligands highly polar ligands	Hepatitis C virus helicase inhibitors <sup>[242]</sup> S-CoV 3C-like proteinase inhibitors <sup>[243]</sup> Cyclooxygenase (COX-2) inhibitors <sup>[244]</sup>  Bacterial NAD synthetase inhibitors <sup>[245]</sup> Lymphoid phosphatase inhibitors <sup>[246]</sup> RNA polymerase inhibitors <sup>[247]</sup> ATP phosphoribosyltransferase (HisG) inhibitors <sup>[248]</sup> Human histamine H4 receptor ligands <sup>[249]</sup>
-FlexX (Rarey et al., 1996) -IC	Small binding sites small hydrophobic ligands	very flexible ligands	Liver X receptor modulators <sup>[250]</sup> HIV-1 reverse transcripts inhibitors <sup>[251]</sup>
Glide (Friesher et al., 2004) -HF+MC	flexible ligands small hydrophobic ligands	ranking for very polar ligands low speed	HIV-1: CD4-gp 120 binding inhibitors <sup>[252]</sup> Serotonin 5-HT(7) antagonists <sup>[253]</sup> Non-peptide $\beta$ -secretase inhibitors <sup>[254]</sup> Sarco/endoplasmic reticulum calcium ATPase inhibitors <sup>[255]</sup> <i>Trypanosoma cruzi</i> trans-sialidase inhibitors <sup>[256]</sup>
GOLD (Verdonk et al., 2003; Verdonk et al., 2005) -GA	small binding sites small hydrophobic ligands buried binding pockets	ranking for very polar ligands or large cavities	
Surflex (Jain, 2003; Jain, 2007) - IC+MA	large and opened cavities small binding sites very flexible ligands	low speed for large ligands	Triple helical DNA intercalators <sup>[257]</sup> ErmC methyltransferase inhibitors <sup>[258]</sup> Hepatitis C virus NS5B polymerase inhibitors <sup>[259]</sup>

### Systematic method (incremental)

The incremental construction method involves building the ligand “on-the-fly” within the constraints of the binding site, while at the same time exploring the ligand flexibility. The

approach is accomplished in a number of ways: (i) *E-Novo* protocol (or de Novo Design)<sup>[260]</sup> which involves docking various molecular fragments into the active-site region and then linking favorable matches covalently. An alternative approach is generating a scaffold core within the binding site, using a ligand-bound protein structure.



**Figure 2.4: Classification of various methods of Structure-based virtual screening**  
Adapted from Kitchen et al.

The scaffold core can then be used to construct ligand of interest by an R-group fragmentation/enumeration process. (ii) In the TrixX BMI method, the ligands are first split into fragments, which are then counted using triangular descriptors and stored in a database. During screening, the target-based query descriptors are used to extract matches and then reconstructed incrementally<sup>[261]</sup> to build the ligands. By and large, most systematic searches methods explore



all degrees of freedom in a molecule and it is the only search method that thoroughly samples the conformational space of the ligand. However, the method ultimately faces the problem of combinatorial explosion<sup>[262]</sup> – exhaustive search calculation– that is computationally prohibitive. Certain algorithms such as MOLSDOCK algorithm have been designed to deal with this problem<sup>[263]</sup> and some examples of docking programs (**Figure 2.4**) that utilize the search algorithm, though with a little modification, include DOCK4.0, FLOG<sup>[264]</sup> and FlexX.<sup>[265]</sup>

### **Random search**

Random search is a local search method that involves making random changes to either a single ligand or a population of ligands and then evaluates new ligands based on a pre-defined probability function with the hope of finding an improved solution or a better conformation. There are three approaches used in this method to generate new ligand conformations: (i) Monte Carlo methods use a random modification to generate alternative ligand poses; (ii) Genetic algorithm (GA) uses an evolutionary selection method to evolve poses that are favorable, which are passed to the next generation, but unfavorable poses are eliminated. The inherited poses are then passed through an iterative process where random or biased mutations can be made to increase genetic diversity and prevent premature convergence, and then the selection and crossover stages are carried out until favorable poses are obtained. (iii) Tabu search algorithm is a search method that takes into consideration already explored areas of the conformation space. In effect, the algorithm is able to mark solutions and those poses that have been rejected before with a “tabu” (forbidden) tag so that the algorithm does not repeat that conformation again. The algorithm uses root mean square deviation (RMSD) to differentiate between current molecular coordinates and previous recorded conformations, and to determine if a new molecular formation should be accepted or not. PRO\_LEADS<sup>[266]</sup> is good example of a docking program

that uses tabu search algorithms. While programs such as AUTODOCK<sup>[235, 267]</sup> uses a more highly effective search algorithm which combines Monte Carlo like perturbation with other search algorithms such as simulated annealing, but, DOCK and GOLD<sup>[268]</sup> are known to implement GA searches.

### **Simulated Search Method**

Simulation search method use molecular dynamics (MD) and energy minimization approaches to pose ligands in the active site of the enzyme. The major drawback of MD simulation is that it cannot cross high-energy barriers and so it can get trapped in a local minima when the calculations hits a rough potential energy landscape.<sup>[237]</sup> One way the problem has been solved is to use search algorithm that couples different degrees of freedom of the ligand to different temperatures; and during simulation the two temperatures that determine flexibility of the ligand are varied.<sup>[269]</sup> Other possible solutions tried include an attempt to simulate different parts of a protein-ligand system at different temperature<sup>[269b]</sup> or to start simulation from different point on the ligand. Unlike MD, energy minimization approach is a complementary method used by other search algorithms but not a standalone method. Docking programs such as DOCK and Monte Carlo algorithm have energy minimization incorporate into the search algorithm<sup>[270]</sup>, while other programs that use simulation based search either alone or in combination include Glide<sup>[271]</sup>, MOE-Dock<sup>[272]</sup>, AUTODUCK and Hammerhead.<sup>[273]</sup>

### **Scoring Function**

#### *Scoring Function*

Docking programs contain not only search algorithms but also complementary scoring and ranking functions. Even when the correct binding modes of the ligands are predicted, scoring

functions are still crucial in differentiating correct poses from incorrect ones. Scoring functions make it possible to determine the correct pose of the ligand in the receptor, predict the binding affinity and suggest other possible ligands that can bind to the active site. Usually, scoring functions score the poses by a rough estimate of how the ligand matches the active site, and then the poses are ranked and rescored for how tightly the ligands bind to the receptor using more rigorous parameters. Essentially, scoring functions play a key role in the success of many docking programs, particularly in the area of virtual screening and lead optimization. In this section we will review a number of scoring functions used by various docking programs and current challenges and possible improvements made to these functions. Basically, three types of scoring functions are been implemented most docking programs: Force fields, empirical and knowledge based scoring functions.

### **Force Fields Scoring Function**

The Force field score is based on physical atomic interactions<sup>[274]</sup>, it essentially sums atomic interactions such as van der Waals, electrostatic (using Lennard-Jones potential) and bond/stretching/binding/torsional forces. Usually in docking, the force field scoring functions attempts to predict binding affinities of ligands by adding individual contributions from different types of interactions. The major protein-ligand interaction is contributed by van der Waals and the electrostatic term<sup>[191]</sup>, which turns out to be first shortcoming of the scoring function. The Lennard-Jones potential is usually used to model van der Waals potential and as such the steep form at short interatomic distance on the potential well (**Figure 2.1**) can lead to strong repulsion, which might result in significant steric clashes that are unaccounted for during computation. Sometimes these clashes are noticeable in docking results, it is important to visually inspect each docking result to weed out such steric clashes during post processing of docked results.

Furthermore, the electrostatic term grossly over emphasizes the polar interaction which is due to the function not being able to deal with effect of solvent. Shoichet and Wei have established possible solution to ligand solvation problem but there is also the desolvation effect to compute, which deals with explicit water surrounding of the solute atoms, which influences ligand binding.<sup>[275]</sup> A rigorous approach to deal with the problem is reviewed by Wang and Adcock<sup>[276]</sup> but the solutions are too computationally expensive to be used in large scale virtual screening. Further solutions to the problem have been successfully implemented in virtual screening by using Poisson-Boltzman/Surface are (PB/SA)<sup>[277]</sup> model and the general-Born/surface area (GB/SA) model.<sup>[278]</sup> However, it is important to note that some of the force field scoring functions are based on different force field parameter sets and they differ from one docking program to another. For example, AUTODOCK uses the AMBER force field<sup>[279]</sup>, while G-score uses the Tripos force field<sup>[280]</sup> for scoring binding affinities.

### **Empirical Scoring Function**

An empirical scoring function binding affinities are estimated as a weighted sum of several parameterized functions such as explicit hydrogen bonding, ionic interaction and hydrophobic contact terms.<sup>[194d]</sup> The coefficients of the weighted terms are obtained from regression analysis determined experimentally by fitting the binding affinity data of a training set of protein-ligand complexes with known three-dimensional structures.<sup>[281]</sup> The terms used in empirical scoring functions, though similar to force field, are simpler and easy to evaluate. In fact, Tarasov and Tovbin developed an extremely simple empirical scoring function *NScore* in order to test how sophisticated a scoring function has to be for it to be successful in docking. Surprisingly, the results from using *NScore* were comparable to GOLD, DOCK and Glide with sophisticated scoring functions, if the scoring was based on fundamental physic principles only.<sup>[282]</sup> In general,

one shortcoming of these functions is their dependence on the molecular data set used to perform regression analysis and fitting. As a result, there is a different weighting coefficient for various terms, which makes it difficult to transfer or merge terms into new scoring functions. More so, each scoring function has to be calibrated with different dataset of protein-ligand complexes, but with the increase in the number of protein-ligand complexes with three-dimensional structures available in the database, it would be possible to develop a reasonable general empirical scoring function by training with thousands of known protein-ligand complexes. So far, well-known docking programs such as Surflex<sup>[283]</sup> FlexX<sup>[265]</sup>, and Glide<sup>[271]</sup> use empirical scoring function by combining various energy terms and the list of empirical scoring function is constantly growing with different training sets: LigScore, LUDI, ChemScore, SCORE, X-Score, ICM, MedusaScore, AIScore PLP, and SFCscore.<sup>[194d]</sup>

### **Knowledge-based Scoring Function**

Knowledge-based scoring functions are developed from a statistical analysis of information gleaned from protein-ligand experimentally determined structures.<sup>[191]</sup> They are also known as statistical-potential based scoring function. The principle behind knowledge-based scoring function is that pair-wise potentials or interatomic distances occurring more often than some average value should represent a favorable contact and the same is true for unfavorable interaction.<sup>[284]</sup> In addition, the atom-type interactions are usually selected based on their molecular environment in different experimental structures. The main appeal of these functions is the simplicity of the estimation, limited computational requirement and in addition to its ability to deal with low-resolution protein model with compactable long-scale simulation.<sup>[285]</sup> Although, the function is simple, the actual derivation using inverse Boltzmann relation<sup>[285]</sup> contains a reference state term that still remains a longstanding hurdle in deriving pair-wise

potentials.<sup>[286]</sup> In practical terms, derivation of these functions is challenging since it is based on implicit information from a limited data set, which affects its accuracy and it is tough to improve on the function in a stepwise manner. Despite the limitation inherent in knowledge-based scoring function, these functions have been implemented in the design and use of successful functions such as potential of mean force (PMF)<sup>[287]</sup>, DrugScore<sup>[288]</sup>, SMOG<sup>[289]</sup> and the well-known ITScore/SE.<sup>[290]</sup>

### **Consensus Scoring**

The ideal scoring function would score absolute binding affinities accurate binding poses and predict nanomolar hit compounds, but no one scoring function is perfect in terms of accuracy and general applicability. Consensus scoring function combines the information of multiple scoring functions to even out the errors and improve the probability of enriching the screening process with true hits.<sup>[191]</sup> Paramount to the success of these functions is the ability to design appropriate consensus strategies of individual scores so that true binding modes can be differentiated from non-binding modes.<sup>[291]</sup> Some of the strategies used include vote-by-number, rank-by-number, average rank etc<sup>[292]</sup> and reliable scores such as MultiScore<sup>[293]</sup>, X-Cscore<sup>[294]</sup>, SeleX-CS<sup>[295]</sup>, and SCS<sup>[296]</sup> have been developed and used in docking.

In summary, scoring function are typically part of a larger program such as DOCK, FlexX, Glide etc. and each performs specific implementation in these programs. The review by Zou et al.<sup>[194d]</sup> outlines certain general criteria used to evaluate the performance of a scoring function. One of such criteria is the ability of a scoring function to distinguish native binding modes from decoys. Root mean standard deviation (rmsd) values are commonly used by most scoring function to distinguish binders from non-binders; if the rmsd is  $\leq 2.0\text{\AA}$  for true ligands then prediction is

usually considered a success. A second criterion for a scoring function is its ability to predict the binding affinity of a complex, that is, how tightly does the ligand binds to the receptor. This is where most scoring functions find it challenging to achieve a score scale similar to experimental binding data. Lastly, the scoring function should be capable of selecting potential binders (hits) from a given library of compounds, which is in effect virtual screening in computer-based drug design. Equally notable in the implementation of scoring functions is that small changes during the preparatory process of the ligand or protein can make a big difference in binding affinity.<sup>[296]</sup> For example, changes in the number of rotatable bonds, protonation states, minimization etc. are very critical to obtaining reliable results during docking. Furthermore, scoring functions are not independent of docking algorithms that sample the chemical space, as a result scores are also greatly influenced by the binding site, method used to explore the conformational space, or methods used to sample and optimize the ligands.

### **2.3 Computation Based Drug Discovery of Dnmt1 Inhibitors**

The use of molecular modeling and virtual screening in drug discovery is a powerful tool in rational drug design and most techniques are geared towards understanding binding modes of potential drugs at the active site as well as identifying new leads with novel scaffold.<sup>[297]</sup> Understanding the binding modes will help fine-tune drug molecules to be more effective, selective, specific, and perhaps avoid preventable side effects, which currently plagues the known approved nucleoside Dnmt1 inhibitors. The cytosine based nucleosides approved by the FDA for myelodysplastic syndromes have low specificity and exert cellular and clinical toxicity.<sup>[298]</sup> Researchers are turning to computational methods to discover other potential Dnmt1 inhibitors with minimal side effects or to computationally reengineer the current suicidal nucleoside inhibitors in an effort to reduce the well-known side effects to nominal acceptable

limits. So far, a number of different virtual screening studies with Dnmt1 have provided novel potential drug-like inhibitors with unique binding modes to the enzyme. Herein, we shall discuss the discovery of non-nucleoside inhibitors of Dnmt1 by virtual screening and binding mode of both known nucleoside and non-nucleoside inhibitors.

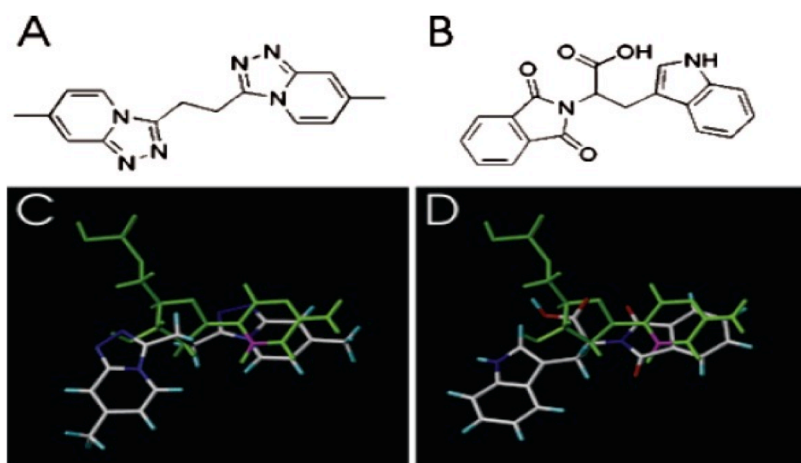
### **Homology Model**

Initially, before the Song et al. solved the crystal structure for hDnmt1, Siedlecki and colleagues had used a homology model<sup>[299]</sup> to discover two novel inhibitors, NSC3033530 and NSC404077<sup>[140b]</sup> (**Figure 2.5**). The model was accomplished by using MODELER module in INSIGHT2000 to convert the Dnmt1 sequence into a three dimensional structure based on the X-ray diffraction data of M HhaI, M. HaeIII and Dnmt2. Validation of the structure was done with the Profile3D module of INSIGHT2000 and PROCHECK.<sup>[232]</sup> After a short energy minimization process and removal of water molecules, the homology model of the Dnmt1 catalytic domain was ready for docking. Virtual screening of a set (the Diversity Set of 1990 compounds) of small molecules available from the National Cancer Institute (NCI) database was performed on the hDnmt1 homology model using the novel docking and scoring methodology in DOCK5. Interestingly, the two top-ranked compounds from this docking study, especially NSC404077 (RG108) still remain potent inhibitors of Dnmt1 despite the availability of the two crystallographic structures of hDnmt1 in the PDB database.

Though the Food and Drug Administration (FDA) has approved demethylating agents such as Vidaza and Dacogen, there is still an unwavering drive for new inhibitors for Dnmt1 because of the inherent toxicity of these nucleoside inhibitors<sup>[298]</sup>: 5-azacytidine (Vidaza) is cytotoxic, non-



specific and gets incorporated into RNA<sup>[300]</sup>, while 5-aza-2'-deoxycytidine (Dacogen) is also cytotoxic and gets incorporated into DNA.<sup>[301]</sup>



**Figure 2.5: Identification of two novel candidate DNA methyltransferase inhibitors by computational screening. (A, B) Chemical structures of NSC303530 and NSC401077. (C, D) Superimposed conformations of cytidine (green) and NSC303530 (C) or NSC401077 (D) docked into the active site of Dnmt1. The covalent bond between cytosine and the catalytic cysteine of Dnmt1 is formed at the cytosine-6 position (magenta) <sup>406</sup> Permission requested**

However, zebularine, a nucleoside analog, shows comparatively less toxicity and tumor specific but it follows similar inhibitory mechanism as other nucleosides<sup>[302]</sup> and as such gets incorporated into the DNA and covalently binds to Dnmt1. The above reason for cytotoxicity has fueled the search for new Dnmt1 inhibitors that are less toxic and do not get incorporated into the DNA and can reversibly bind to Dnmt1. Cornacchia and colleagues had asked the question whether drugs like procainamide and hydralazine are capable of inducing a lupus-like syndrome in humans as observed when cloned T cells were treated with 5-azacytidine, since these drugs

like 5-azacytidine bind to DNA too.<sup>[302]</sup> Their results led to the discovery of two new inhibitors of Dnmt1, procainamide and hydralazine a common anti-arrhythmias and a local antihypertensive respectively. Later, Villar-Garea and colleagues tested the potential of procaine (a local anesthetic) as an inhibitor of Dnmt1 based on its structural similarity to procainamide.<sup>[142b]</sup> Using the MCF-7 breast cancer cell line, they showed that procaine is an inhibitor of Dnmt1 and it produces a 40% reduction in 5-methylcytosine DNA. Furthermore, it was proposed that the main polyphenol component of green tea, (-)-epigallocatechin-3-gallate (EGCG) could also inhibit Dnmt1 since previous studies showed that EGCG inhibits catechol-*O*-methyltransferase, which belongs a superfamily of *S*-adenosylmethionine-dependent methyltransferases the same as Dnmt1.<sup>[135]</sup> EGCG inhibited Dnmt1 in a dose-dependent manner, showing competitive inhibition with a  $K_i$  of 6.89  $\mu\text{M}$ .<sup>[135]</sup> The newly found inhibitors are interesting but the mechanisms of action are still debatable, also the binding modes of these molecules to Dnmt1 are unknown. In order to understand the binding modes of these new inhibitors, Singh et al<sup>[303]</sup> conducted molecular modeling with the homology model provided by Lyko et al.<sup>[299]</sup> Their study compared the relative binding modes of two tautomeric forms of hydralazine, procainamide and procaine to known nucleoside inhibitors, 2'-deoxycytidine, 5-azacytidine and decitabine inside the homology model. Importantly, their docking results suggested that the new inhibitors shared similar binding modes as the nucleoside inhibitors, particularly, the interaction of all inhibitors with Glu128 and Arg174 (**Figure 2.6**). Also, superimposing the new inhibitors with known nucleoside inhibitors in the catalytic pocket showed they all shared similar position (**Figure 2.7**)

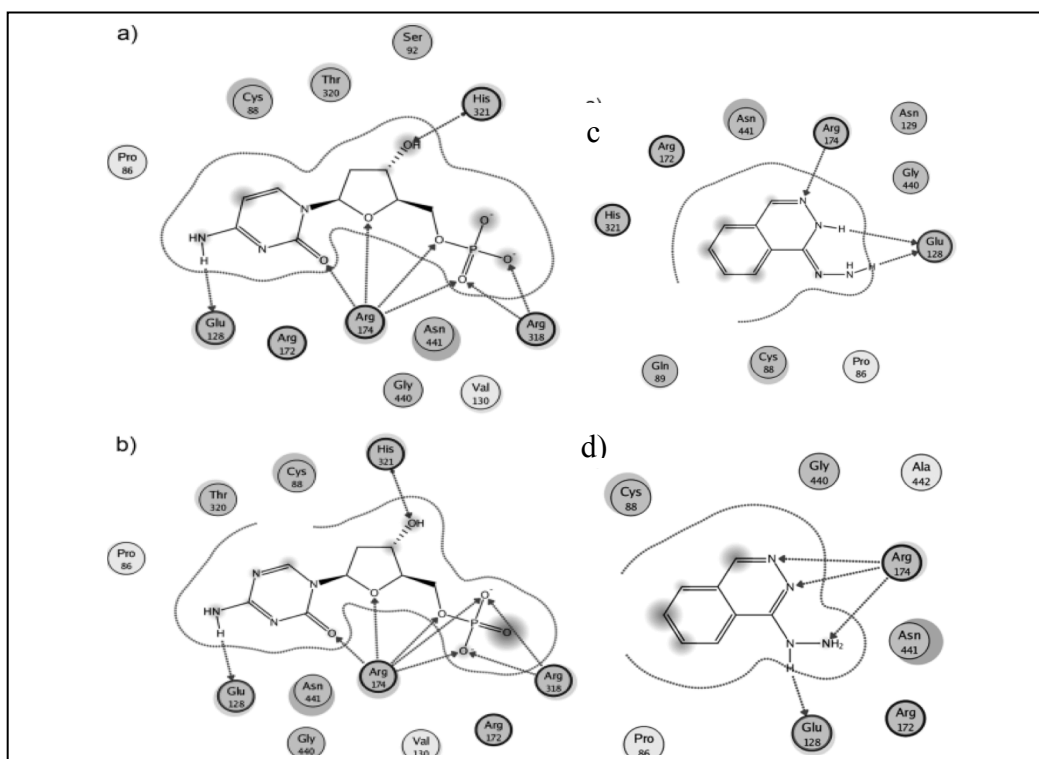


Figure 2.6: Optimized docking models of a) 2'-deoxycytidine and b) 5-aza-2'-deoxycytidine c) the hydrazone tautomer (imino form) and d) the hydrazine tautomer (amino form) with the homology model of human Dnmt1. Arrows indicate hydrogen bonding. "Clouds" on ligand atoms indicate the solvent-exposed surface area of ligand atoms (darker and larger clouds mean more solvent exposure). "Halos" around residues indicate the degree of interaction with ligand atoms (larger, darker halos mean greater interaction). The dotted contour reflects steric room for methyl substitution. The contour line is broken if it is closest to an atom which is fully exposed.<sup>410</sup> *Permission requested*

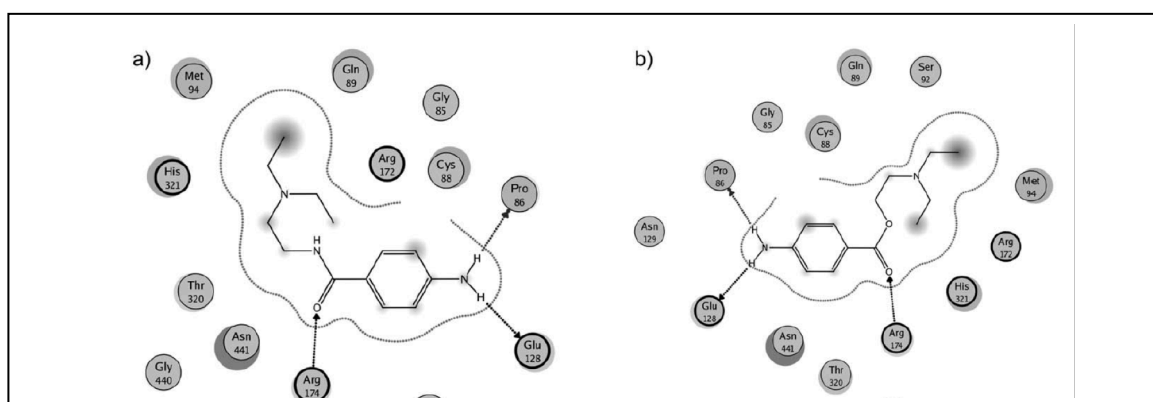
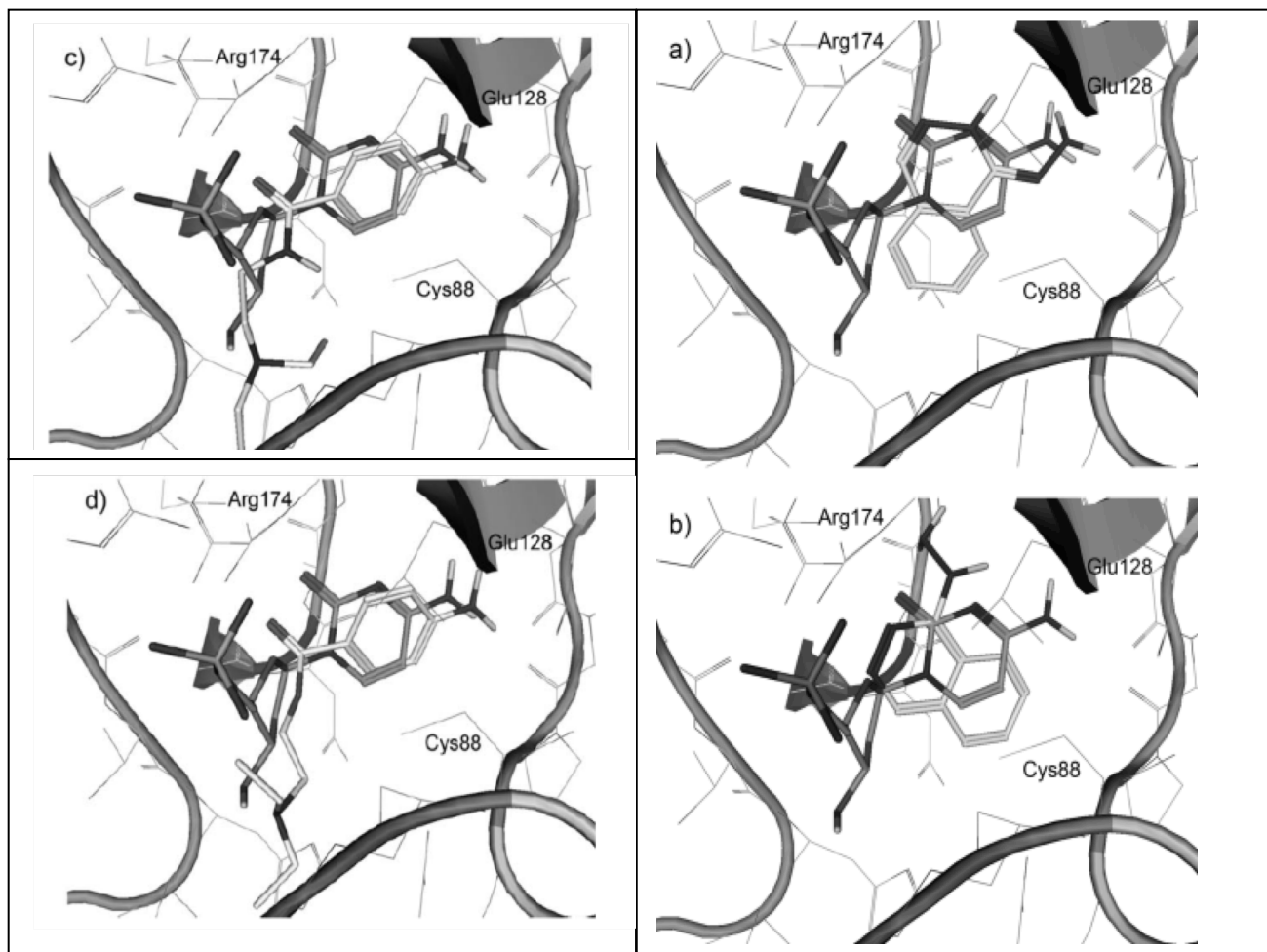


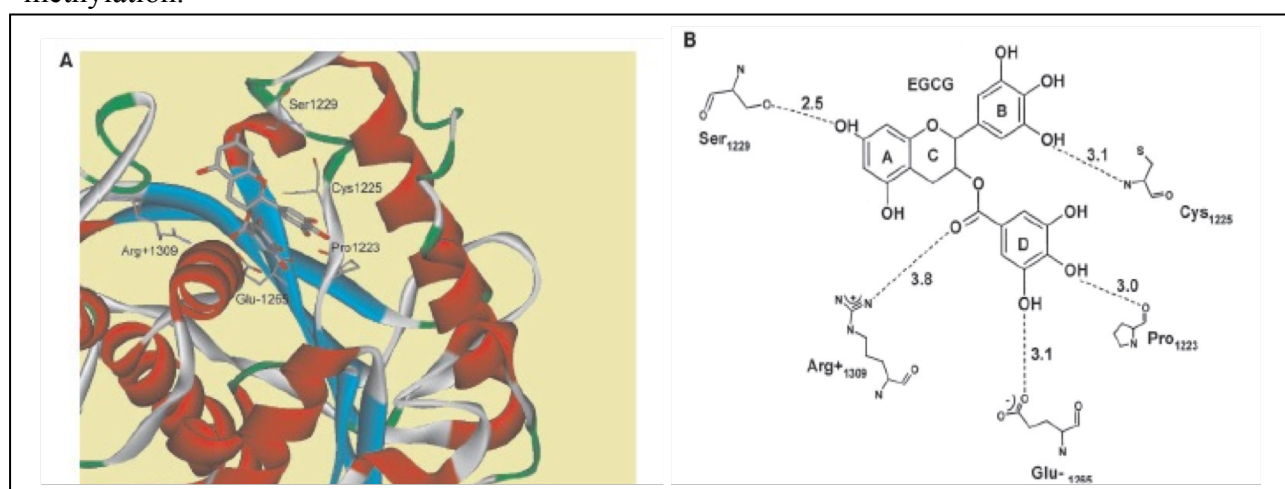
Figure 2.7: Optimized docking models of e) procainamide and f) procaine with the homology model of human Dnmt1. Arrows indicate hydrogen bonding. "Clouds" on ligand atoms indicate the solvent-exposed surface area of ligand atoms (darker and larger clouds mean more solvent exposure). "Halos" around residues indicate the degree of interaction with ligand atoms (larger, darker halos mean greater interaction). The dotted contour reflects steric room for methyl substitution. The contour line is broken if it is closest to an atom which is fully exposed.<sup>410</sup> *Permission requested*



**Figure 2.8** Comparison of the optimized binding models of hydralazine with the binding model of 2'-deoxycytidine (carbon atoms in dark gray): a) hydrazone tautomer, b) hydrazine tautomer c) procainamide and d) procaine (carbon atoms in light gray). Nonpolar hydrogen atoms are omitted for clarity. Note the similar positions of the amino groups and the very similar position of the carbonyl oxygen of procainamide and procaine with the ribose oxygen. These atoms form hydrogen bonds with Glu128 and Arg174, respectively. <sup>410</sup> *Permission requested*

In **Figure 2.8**, three of the nitrogen atoms of hydralazine in the hydrazone tautomer show a rough overlap with three of the heteroatoms of 2'-deoxycytidine. Also, the amino group occupies the same binding position as the amino group of cytosine, which is capable of binding with Glu128. Furthermore, the unprotonated nitrogen of the phthalazine ring in the hydrazone forms a rough overlap with the binding position of the carbonyl oxygen of cytosine and any of these atoms (N or O) can hydrogen bond with Arg174. Equally notable is the binding modes of

procaine and procainamide compared with the nucleoside inhibitors. The amino group of procaine and procainamide and that of 2'-deoxycytidine share similar positions; the carbonyl oxygen of procaine and procainamide occupy similar positions as the oxygen of the ribose sugar of 2'-deoxycytidine and are capable of making hydrogen bond with Arg174. In summary, their results highlight the importance of the interaction of inhibitors with key residues in the enzyme such as arginine and glutamic acid since they might play a major role in the mechanism of DNA methylation.



**Figure 2.9: Molecular modeling of the interaction between EGCG and Dnmt.** A, binding orientation of EGCG in Dnmt1. The close-up view of the consensus orientation for EGCG. The protein is depicted in ribbon representation and colored by secondary structures (i.e., helix, strand, and loop). Both EGCG and ligand contact residues are represented in stick form and colored by atom type with carbon in gray, oxygen in red, and sulfur in yellow. B, hydrogen-binding network of EGCG in Dnmt1. -- represents hydrogen bond. Figure depicts all potential H-bonding X-H . . . Y interactions for which the X . . . Y distance is  $4.0 \text{ \AA}$  <sup>199</sup> *Permission requested*

Furthermore, still using homology model, Fang and colleagues suggested the binding mode of EGCG and proposed a possible mechanism of action.<sup>[135]</sup> Using an in-house homology model created using HhaI Mtase as molecular template and docking with GOLD; they noticed certain interactions at the catalytic site (**Figure 2.9**). The gallate moiety of EGCG could form hydrogen bonds with Glu1265 and Pro1223. Also, the position of the D ring is oriented roughly at the same position as the pyrimidyl ring of cytosine. In addition, the hydroxyl groups in the A and B rings are positioned to form hydrogen bonds with Ser1229 and Cys1225 respectively. From these



findings they hypothesized that EGCG exerts its inhibitory effect on Dnmt1 function by blocking the entry of key nucleotide cytosine into the active site.<sup>[135]</sup>

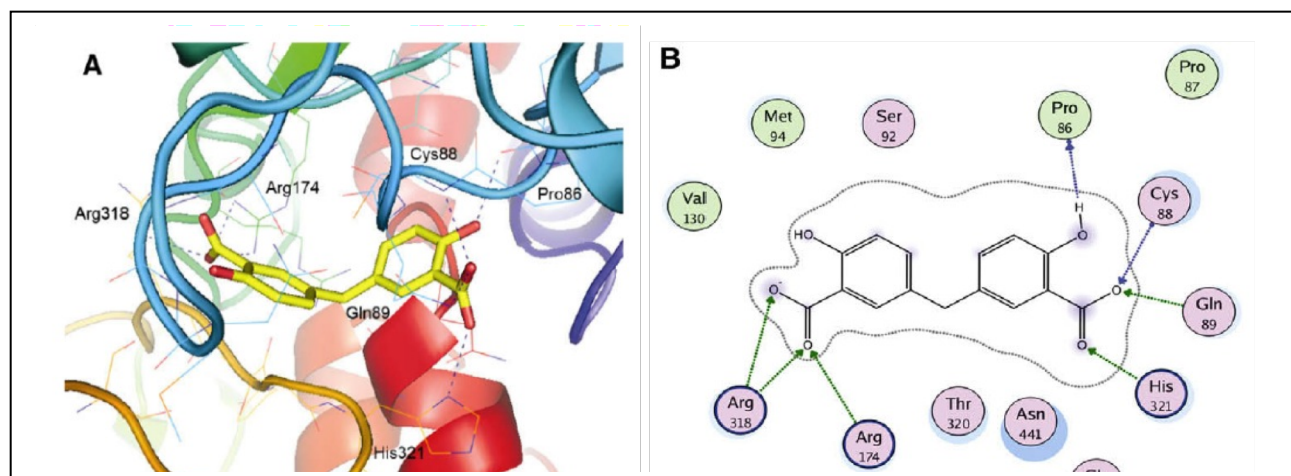


Figure 2.10: Optimized docking model of NSC 14778 with human Dnmt1. (A) 3D representation showing selected amino acid residues. Hydrogen bonds are indicated with blue dashes. Hydrogen atoms are omitted for clarity; (B) 2D interaction map displaying amino acid residues within 4.5 Å of the ligand. Green and blue arrows indicate hydrogen bonding to side chain and backbone atoms, respectively. Blue ‘clouds’ on ligand atoms indicate the solvent exposed surface area of ligand atoms. Light-blue ‘halos’ around residues indicate the degree of interaction with ligand atoms. The dotted contour reflects steric room for methyl substitution.<sup>411</sup> *Permission requested*

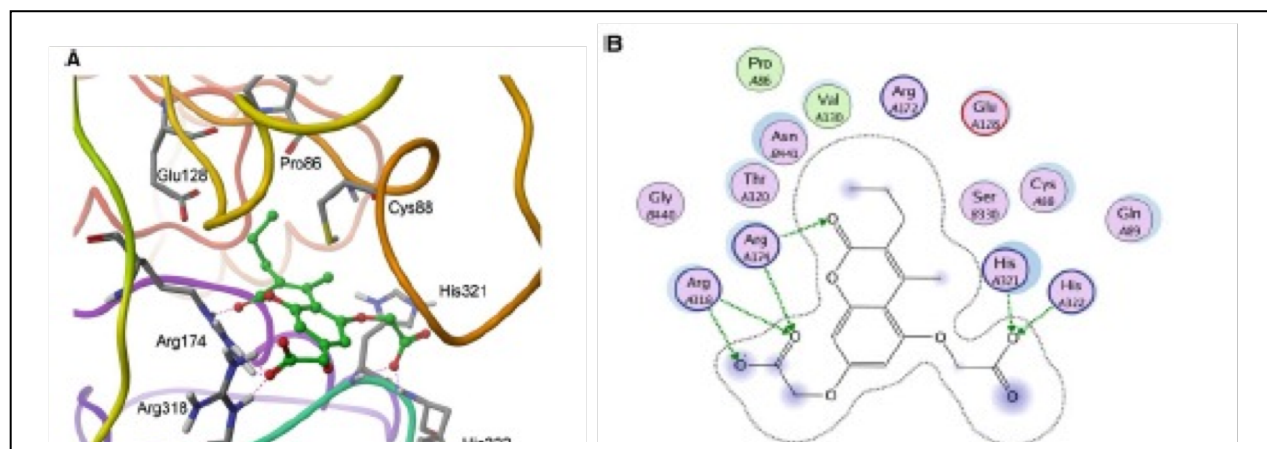


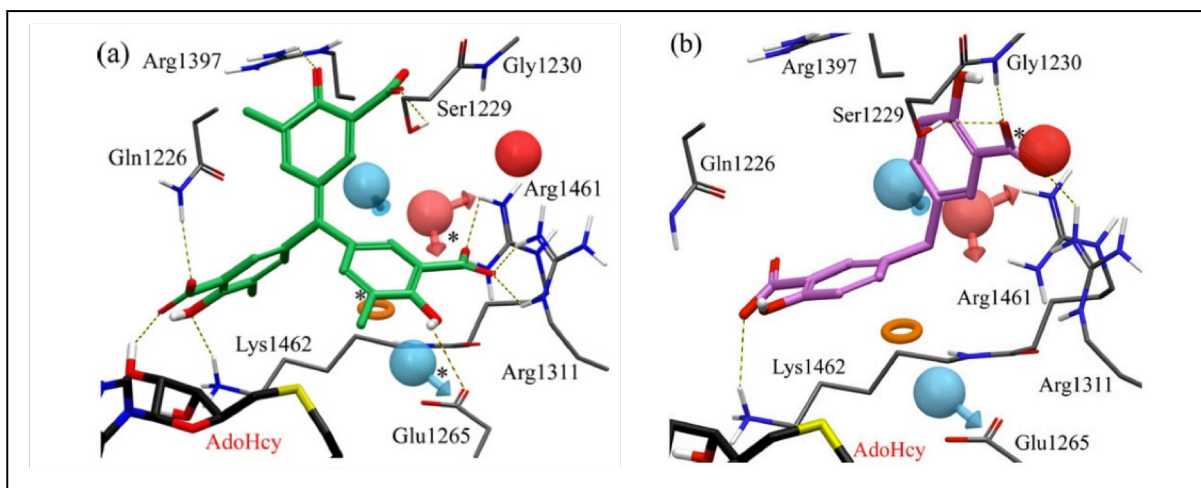
Figure 2.11: Optimized binding model of virtual screening hit ZINC 00082754 with human Dnmt1. a Three-dimensional representation displaying selected amino acid residues. Hydrogen bonds are indicated with magenta dashes. b Two-dimensional interaction map displaying amino acid residues within 4.5 Å of the ligand. The ligand proximity contour is depicted with a dotted line. The ligand solvent exposure is represented with blue circles; larger and darker circles on ligand atoms indicate more solvent exposure. The receptor solvent exposure differences—in the presence and absence of the ligand—are represented by the size and intensity of the turquoise discs surrounding the residues; larger and darker discs indicate residues highly exposed to solvent in the active site when the ligand is absent<sup>414</sup>

Notably, Kuck and colleagues used the homology model of Lyko and a multistep virtual screening methodology to uncover a diverse collection of novel and selective inhibitors of Dnmt1.<sup>[304]</sup> One of such hits was NSC14778 with an IC<sub>50</sub> of 92 μM. In this study, more than 65,000 lead-like compounds were obtained from the NCI Open Database<sup>[305]</sup> and prepared for docking using LigPrep and docked sequentially through GLIDE 5.0, GOLD 4.0 and AutoDock 3.0 to arrive at 24 consensus hits of which NSC14778 was most active after experimental testing. Optimized docking model of NSC14778 (**Figure 2.10**) showed a network of predictable hydrogen bonds with key residues in the catalytic pocket, most importantly Arg174 and Arg318, which corresponds with the binding modes of procaine, procainamide and hydralazine. Equally noteworthy is the discovery of a natural product inhibitor by Medina-Franco and colleagues.<sup>[306]</sup> After obtaining 89,425 natural product compounds from the ZINC database<sup>[307]</sup> [440] and passing them through a filter program and preparing them with LigPrep, Medina-Franco used a similar multistep docking protocol to uncover a coumarin based inhibitor, ZINC00082754. Also, optimized binding mode of this compound with Dnmt1 showed possible hydrogen bonds with the side chains of Arg174 and Arg318 in addition to interactions with His321 and His322 (**Figure 2.11**)

### **Crystal Structure Based Inhibitors**

Two crystallographic structures of hDnmt1 are available in the Protein Data Bank (PDB).<sup>[228]</sup> The first crystal structure was solved by Song et al.<sup>[84]</sup> depicting a hDnmt1 crystallographic structure with SAH and DNA containing unmethylated CpG sites solved at a resolution of 3.6Å (PDB:3PTA), which is not a very good resolution. A second structure was recently added to the PDB database by Hashimoto and Cheng (PDB: 3SWR); they published a structure hDnmt1

bound to the inhibitor, sinefungin (SFG) and with a resolution of 2.49Å. The two structures are similar with a root mean square deviation (RMSD) of 1.4Å.<sup>[308]</sup> Despite the available structures of hDnmt1 no virtual screening has been done entirely with either structure, however, modified structure<sup>[309]</sup> of PDB: 3PTA has been used for most docking experiments, while in some case 3SWR has been rebuilt for virtual screening studies.<sup>[308]</sup>



**Figure 2.12: Comparison of the binding modes with pharmacophore hypothesis for (a) NSC97317 and (b) 5,5-methylenedisalicylic acid (deprotonated forms). Hydrogen bonds are depicted with dashes. Nonpolar hydrogen atoms are omitted for clarity. Selected residues are displayed for reference. Red sphere: negative ionizable; pink sphere with vectors: hydrogen bond acceptor; blue sphere with vector: hydrogen bond donors; and orange ring: aromatic ring. Matching features, considering a distance matching tolerance of 2.0 Å, are marked with asterisks. *Permission requested***

Yoo and colleagues argue that the conformation of the crystal structure of human Dnmt1 by Song is in an inactive state (that is the target base from DNA is not flipped out) and so the geometry presented by Song does not necessarily reflect the catalytic mechanism of Dnmt1, he described the structure as being in an open configuration and thus, not optimal for docking computations. Yoo came to this conclusion by comparing model structures of Dnmts and other available crystal structure of Dnmts<sup>[310]</sup> and concluded that 3PTA was in an open configuration. Yoo and Medina-Franco have used the modified crystal structure to discover a potent inhibitor

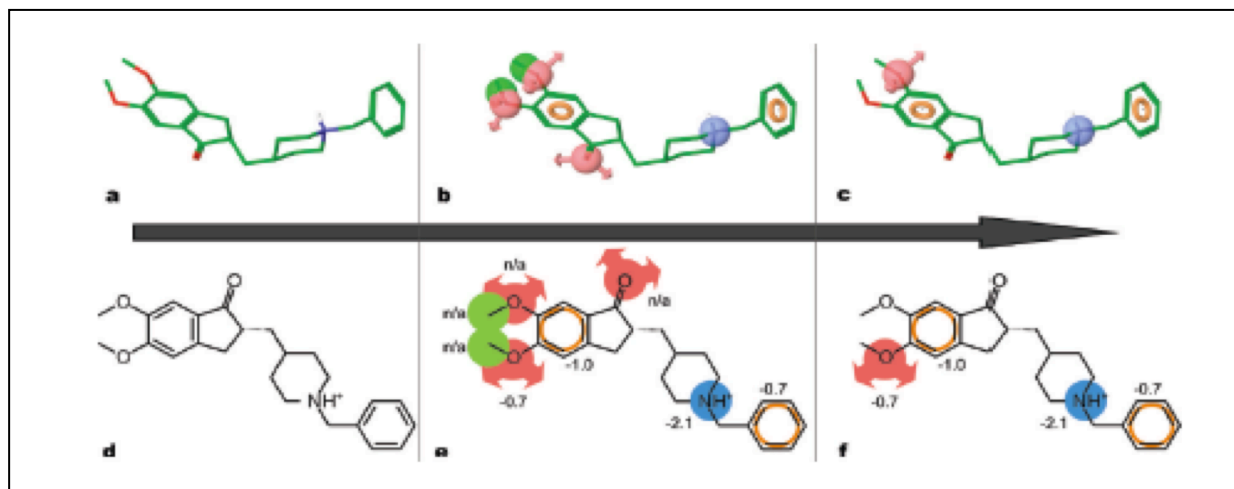


trimethylaurintricarboxylic acid (NSC97317) of Dnmt1. NSC97317 showed an  $IC_{50}$  value of 4.79  $\mu$ M, which is better than previously reported NSC14778<sup>[304]</sup> even though both compounds share similar chemical structure. However, the optimized docked model of NSC97317 show similar hydrogen bond interaction with Ser1229 and Arg1379 in addition to key mechanistic interaction with Glu1265 and Arg1311 (**Figure 2.12**). Also, hydrogen bonding with Glu1265, Arg1311 and Arg1461 have been predicted for other inhibitors and these interactions might be pharmacophoric.

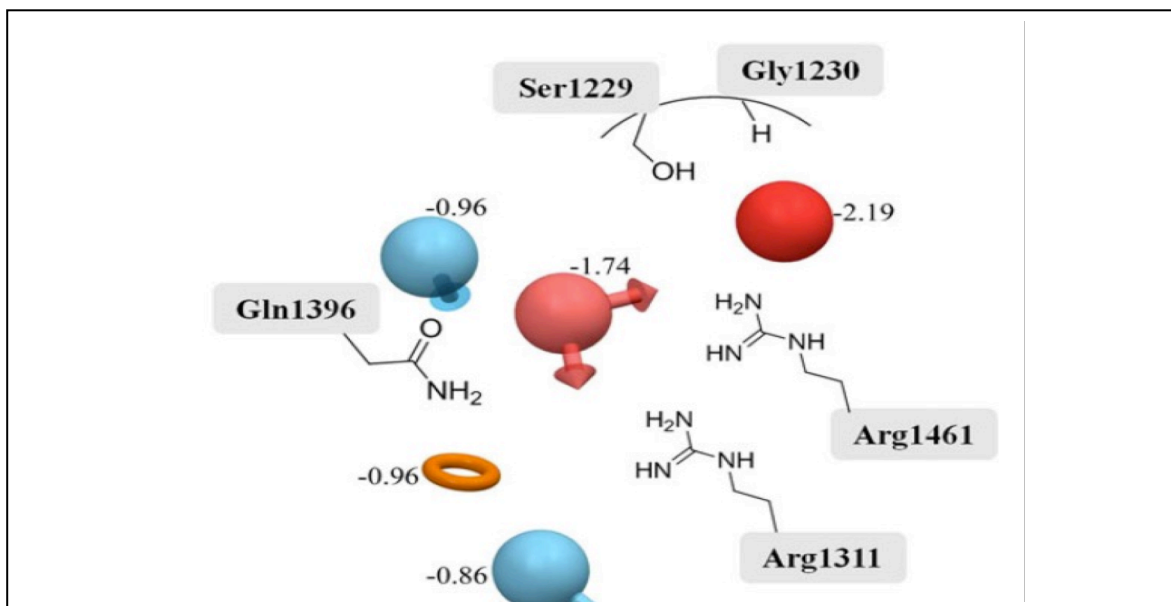
### Pharmacophore Model

In an effort to understand the basic structural requirements of Dnmt1 inhibitors, Yoo and Madina-Franco<sup>[311]</sup> docked a comprehensive list of 14 known Dnmt1 inhibitors (Figure 14) with the homology model of the catalytic domain of hDnmt1 and studied the key protein-ligand interactions. The docking results revealed that all known inhibitors have hydrogen bond interactions with specific glutamic and arginine residues, and suggested that these interactions could be pharmacophoric. As a result a pharmacophore hypothesis was developed using energy optimized pharmacophore (e-pharmacophore) method based on the work by Salam et al.<sup>[312]</sup> Briefly, the method entails the use of Glide to refine ligand-protein crystal structure, then starting with the refined X-ray ligand (Figure 24), pharmacophore sites are generated automatically with Phase (Phase, v3.0, Schrodinger, LLC, New York) based on a default set of six chemical features: hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobe (H), negative ionizable (N), positive ionizable (P), and aromatic ring (R). Each pharmacophore feature is assigned an energetic value equal to the sum of the Glide XP contributions the atom

comprising the site. The sites are then quantified and ranked on the basis of these energetic terms and the most favorable sites are selected for the pharmacophore hypothesis<sup>[312]</sup>; and the pharmacophore model is then used as queries for virtual screening. Applying the same e-pharmacophore model to the 14 known inhibitor of Dnmt1 Yoo and colleague developed a five-feature pharmacophore model (**Figure 2.13 & 2.14**) for the known inhibitors. The top-ranked feature is the negative charge next to residues Ser1229, Gly1230, and Arg1311, and then followed by an acceptor site that is in proximity to Arg 1311 and Arg 1461 (**Figure 2.14**). The third ranked feature is the aromatic ring, which in this case stabilizes the binding formation between AdoHcy and Cys1225, and a donor site that is proximal to Gln 1396 pharmacophore model captures key interactions of the 14 known inhibitors especially the highly conserved glutamic and arginine residues present in the catalytic domain of most DNA methyltransferases.<sup>[313]</sup>



**Figure 2.13: Example of energy-optimized hypothesis for Aricept, 14, in PDB 1eve. Arrow represents the evolution from the initial ligand (a,d) to the ligand with all sites (b,e) to the ligand with the most energetically favorable sites (c,f). Pink sphere/circle ) hydrogen bond acceptor, green sphere/circle ) hydrophobic group, orange ring ) aromatic ring, and blue sphere/circle ) positive ionizable.<sup>421</sup>**  
*Permission requested*

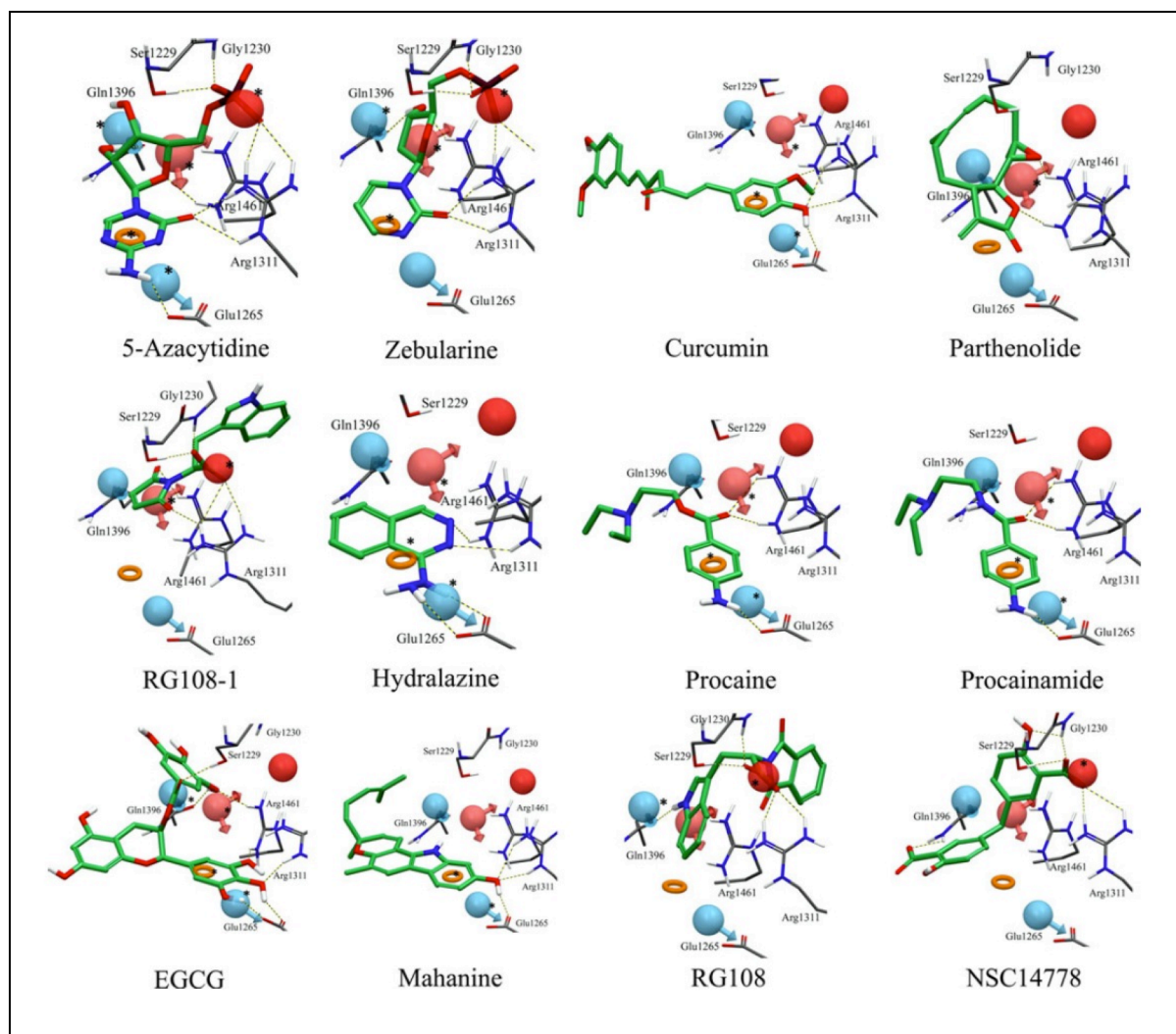


**Figure 2.14: Structure-based pharmacophore model for Dnmt inhibitors. Red sphere negative ionizable, pink sphere hydrogen bond acceptor, blue sphere hydrogen bond donors, and orange ring aromatic ring. Selected amino acid residues in the catalytic site are schematically depicted for reference. See the online version of the manuscript for colors.** <sup>420</sup> *Permission requested*

Most of the inhibitors matched several of the pharmacophore features in the proposed model, particularly the nucleoside analogues 5-azacytidine, decitabine, and 5-fluoro-2'-deoxycytidine matched all five pharmacophoric features (**Figure 2.15**) except zebularine with only four matches because of the absence of an amino group on the heterocyclic ring. Procaine, procainamide, curcumin and hydralazine match three features: the aromatic ring, donor and acceptor feature. While for EGCG, the B ring satisfied both the aromatic ring requirement and the donor feature near Glu1265. The D ring matched the donor close to Gln1396 and the acceptor. All other inhibitors matched at least two of the chemical features, except NSC14778 and parthenolide that matched only one the pharmacophore feature. The carboxylate feature of NSC 14778 matches the negative feature, which is the highest-ranked feature of the model. Whereas, parthenolide matched the acceptor feature interacting with key arginine residues

Arg1311 and Arg 1461 and it is possible that parthenolide might also interact with Glu1265 if it undergoes keto-enol tautomerisation.<sup>[313]</sup> In summary, the e-pharmacophore model shows good agreement with docking studies of most Dnmt1 inhibitors and it a vital step towards understanding the ligand-protein interactions and optimization of leads.

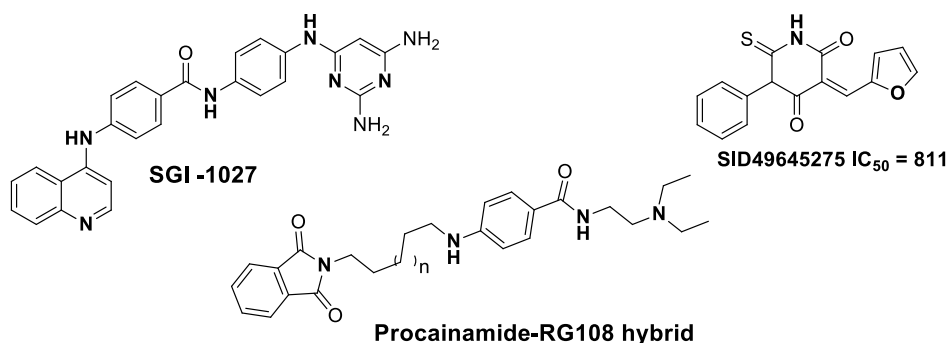
In addition to virtual screening and pharmacophore modeling, Jharma et al.<sup>[149]</sup> introduced a new class of quinoline-based Dnmt1 inhibitors. They reported a quinoline-based compound designated SGI-1027, which inhibits the activity of Dnmt1, Dnmt3A and Dnmt3B as well *M. SssI* with an IC<sub>50</sub> (6-13 μM). Further studies revealed that the mechanism of inhibition of Dnmt1 is competitive not with the nucleoside base target but with *S*-adenosylmethionine. Also, SGI-1207 is effective at reducing Dnmt1 concentration in different cancer cell types: hepatocellular carcinoma, (HepG2), human cervical cancer (HeLa), prostate cancer (LNCap), and breast cancer (MCF7) cell lines. The results show that the effectiveness of SGI-1027 is comparable with decitabine (95-96%) at inhibiting the proliferation of these cancer cells; similar growth-inhibitory effects were obtained when this compound was exposed to colon cancer cells (HCT116 and RKO). However, unlike decitabine, SGI-107 exerts growth-inhibitory effect with minimal toxicity, even its exposure to rat liver cells showed minimal toxicity.<sup>[149]</sup> Also notable is the fact that SGI-1027 is able to induce rapid proteasomal degradation of Dnmt1 in a variety of cancer cell types as observed for the nucleoside inhibitors<sup>[314]</sup>; in effect it can be classed as a suicidal inhibitor.



**Figure 2.15 Comparison between the binding mode and pharmacophore hypothesis for representative Dnmt inhibitors.** Red sphere negative ionizable, pink sphere hydrogen bond acceptor, blue sphere hydrogen bond donors, and orange ring aromatic ring. The asterisk denotes the matching site with the inhibitors. Selected amino acid residues in the catalytic site are shown for reference.<sup>420</sup> *Permission requested*

One more addition to the class of new structure-based non-nucleoside inhibitor is thioxodihydropyrimidine compound designated SID49645275 (**Figure 2.16**) obtained from a nonradioactive high-throughput activity assay for hDnmt1.<sup>[315]</sup> The compound showed an  $IC_{50}$  of 811 nM, and its structure has been used by Medina-Franco and Yoo in similarity search for other compounds within its class; also elucidating the binding modes for SID49645275 and SGI-107.<sup>[308, 315b]</sup>

Finally, a new synthetic inhibitor of Dnmt1 has been introduced by Champion et al.<sup>[316]</sup>; it is a hybrid compound based on the conjugation of procainamide to L-RG108 (**Figure 2.16**) via an alkyl linker. Six of these conjugates were obtained and tested for Dnmt1 activity and then validated with in-house tumor cell lines (DU145 and HCT116). The compounds (**Figure 2.16**: especially, n=12) showed substantial inhibitory activity over parent compounds; at least 50 times more active than parent compounds. In addition only mild toxicity was observed and they showed selectivity for Dnmt1 over Dnmt3A/3L.



**Figure 2.16: Chemical structure of other novel Dnmt1 inhibitor, compound SGI-107, Procainamide-RG108 and SID49645275**

In summary, a number of new scaffolds and compounds have been introduced as potential inhibitors of Dnmt1 many from various studies using virtual screening, binding mode and pharmacophore modeling. Also, some of the new compounds do show remarkable inhibitory effect comparable to known nucleoside inhibitors but with lesser toxicity. It follows that computational approach to drug discovery does offer some strong potential and will be used in this project to uncover new probable inhibitors of Dnmt1.

## CHAPTER 3

### DESIGN AND SYNTHESIS OF SMALL INHIBITOR MOLECULES

#### 3.0 Introduction

In order to uncover and develop novel inhibitors for Dnmt1, we used virtual screening in two independent studies; one study screened a diverse library of compounds from the ZINC<sup>[307]</sup> database using AutoDock 4.2<sup>[317]</sup>, while the second study docked ligands from the Promiliad (in-house) library of compounds using AUTODOCK Vina<sup>[267]</sup>. Both docking studies used the same crystallographic structure of Dnmt1 (PDB: 3PTA) solved by song et al.<sup>[84]</sup>. The choice of the docking program was based not only on the fact that they are open source software but also because the scoring functions used by AutoDock 4 (semi-empirical free energy force field) and Autodock Vina (modified X-score) are different but give closely related results. For any given problem, either program may provide better results; however, Autodock Vina is approximately two orders of magnitude faster than AutoDock and is deemed to have an improved accurate scoring function<sup>[267]</sup>. In addition to virtual screening, quantitative structure activity relationship and pharmacophore modeling were investigated to evaluate possible Dnmt1 inhibitors. Furthermore, transition state inhibitors were also designed based on cytosine and thymidine the inhibitors were made by modifying the C5 position of cytosine and thymidine. Lastly, analogues of known inhibitors were also investigated for possible inhibition of MCF-7 breast cancer cell.

#### 3.1 Zinc Database Screening

For this study we divided the approach into two, first we docked a diverse collection of 5,400 small molecules into the catalytic site of Dnmt1 (PDB: 3PTA) from which we obtained five hits

(Table 1.1) with distinct structures (Figure 3.1). Three of these compounds ZINC09425787, ZINC02835213 and ZINC04710208 were tested for Dnmt1 inhibition using an in-house high-throughput screening assay, which measures the direct incorporation of tritium ( $^3\text{H}$ ) from *S*-adenosyl-L-[methyl- $^3\text{H}$ ]methionine (AdoMet) to a synthetic oligonucleotide, poly (deoxyinosinic-deoxycytidylic) acid (poly(dI-dC)). The results showed that ZINC04710208 (88SC) inhibits Dnmt1 with an  $\text{IC}_{50}$  of 233  $\mu\text{M}$ . After testing of 88SC, amides derivatives were synthesis to check the importance of the carboxylic acid group and also to explore important amines used in drug design and development. Similarly, the naphthyl group was substituted with other flexible and rigid carbon scaffold.

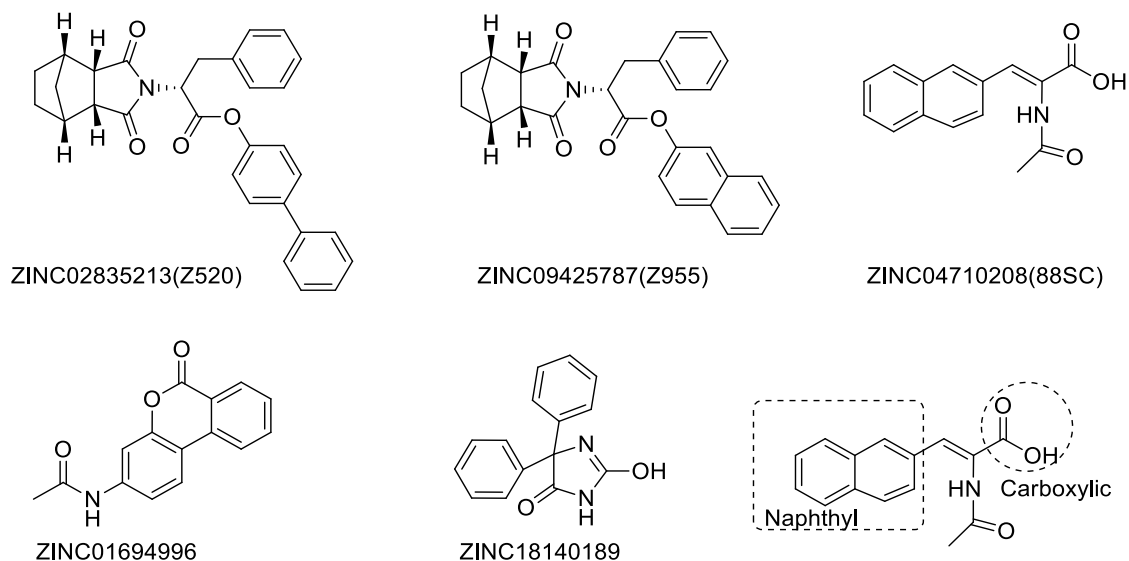


Figure 3.1: Top ranked structures from screening ZINC database library using AutoDock 4.2

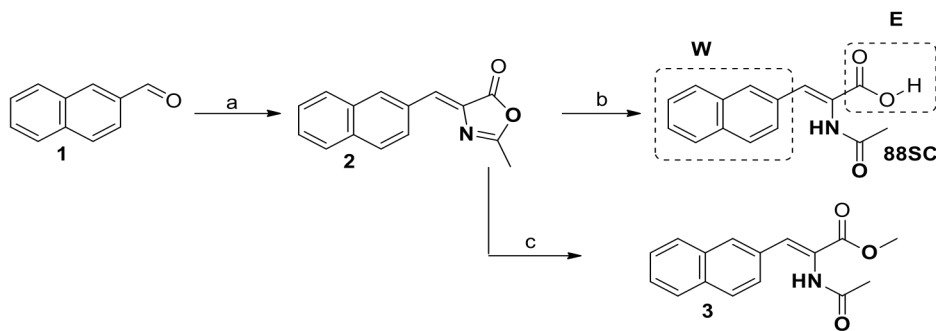


**Table 3.1 Docking results using AutoDock4.2**

Ligand	Binding Energy/ kcal/mol	Intermol Energy	Internal Energy	Torsional Energy	Unbound Energy
ZINC02835213	-8.70	-10.79	-1.23	2.09	-1.23
ZINC04710208	-7.51	-8.40	-0.67	0.89	-0.67
ZINC01694996	-6.98	-7.28	-0.08	0.30	-0.08
ZINC09425787	-7.34	-9.13	-1.30	1.79	-1.30
ZINC18140189	-7.00	-7.89	-0.32	0.89	-0.32

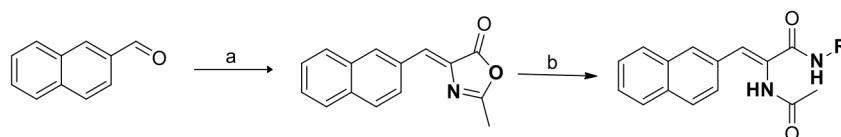
## Synthesis

A Plochl & Erlenmeyer type reaction <sup>[318]</sup> was used to synthesize analogues of (Z)-2-acetamido-3-(naphthalene-2-yl) acrylic acid (88SC). The first step of the reaction involves the condensation of a protect glycine and an aldehyde to give an oxazolone ring **2**, which then undergoes ring opening with acetone:water under reflux to give the acid, or the oxazolone can be recrystallized in methanol to give a ring opened methyl ester product **3** (Scheme 3.1).

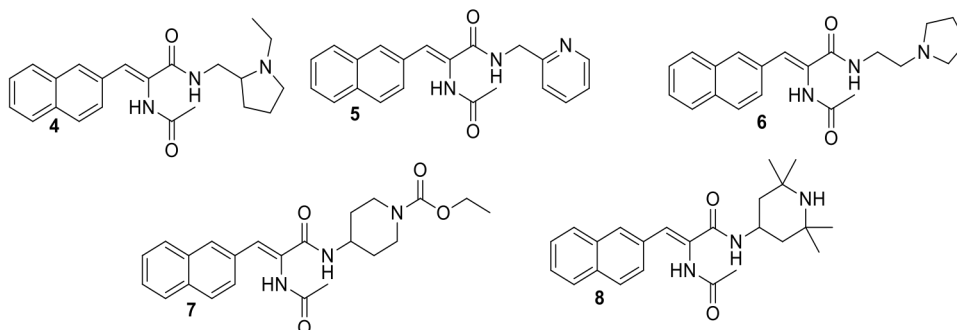


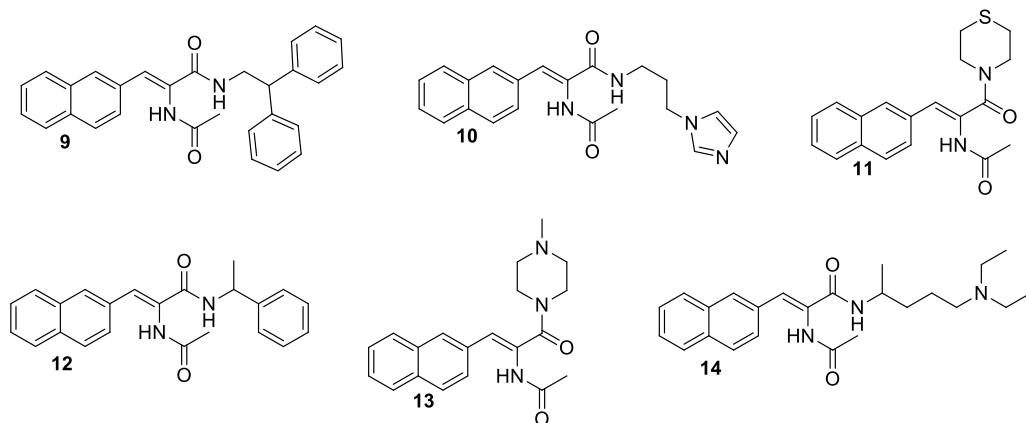
**Scheme 3.1:** (a) 2-acetamidoacetic acid, acetic anhydride, NaOAc 29% (b) acetone:water, reflux 83% (c) recrystallize in methanol

The oxazolone ring can also be opened with primary or secondary amines to give the corresponding amides (**Scheme 3.2**), this is done by heating the oxazolone at reflux with the corresponding amine. Based on the fact that piperazines, piperidine and non-cyclic amines were shown to have inhibitory activity against Dnmt1 (**Figure 3.4**) on their own, and because of their known bioactivity and frequent use in drug design amide analogues of ethyl 4-amino-1-piperidinecarboxylate (2,2,6,6-tetramethylpiperidin-1-yl)oxy (TEMPO) **7PL7C**, 4-Amino-2,2,6,6-tetramethylpiperidine **7PL8C** and 1-(2-Aminoethyl) piperazine **7PL11C** and other non-cyclic amine were used to build a library of amide derivatives of **88SC**. Apart from piperazines, piperadine, and non-cyclic amines, other sulfur based amine (thiomorpholine), straight chain and bulky amines were also used to expand the diversity of the amides library. On the other hand, the modification of the naphthyl group (West) (**Scheme 3.1**) was carried out by using a variety of single and flexible aldehydes for the Plochl & Erlenmeyer type reaction.<sup>[318]</sup> The modification on the West side of the molecule might give us a sense of the  $\pi$ - $\pi$  interaction of the naphthyl ring and aromatic residues in Dnmt1. The  $\pi$ - $\pi$  interaction could be perturbed by changing the naphthyl ring for biphenyl, diphenyl ether, and single phenyl groups while leaving the acidic group constant (**Figure 3.3**).

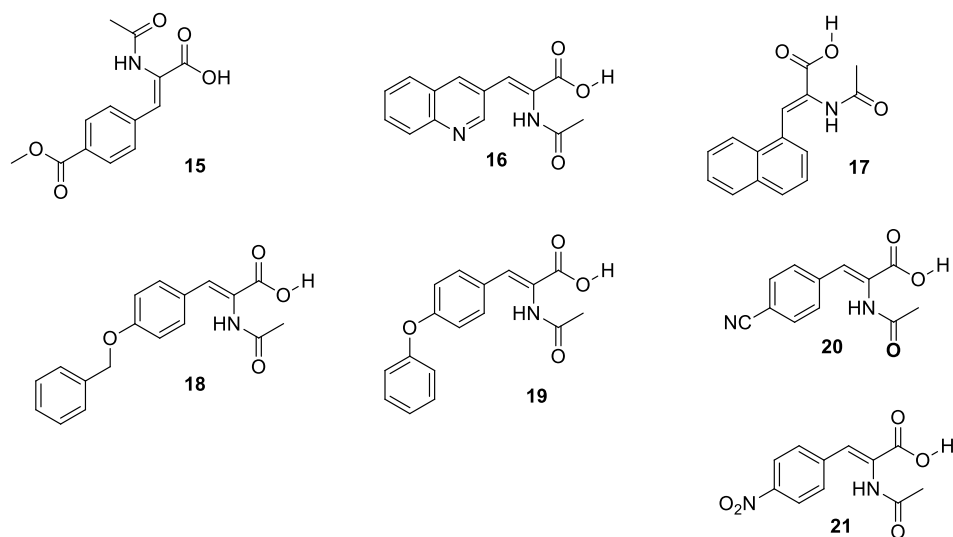


**Scheme 3.2:** (a) 2-acetamidoacetic acid, acetic anhydride, NaOAc 29% (b) CH<sub>2</sub>Cl<sub>2</sub> 80°C





**Figure 3.2: Amide analogues from (Z)-2-acetamido-3-(naphthalene-2-yl) acrylic acid (88SC)**

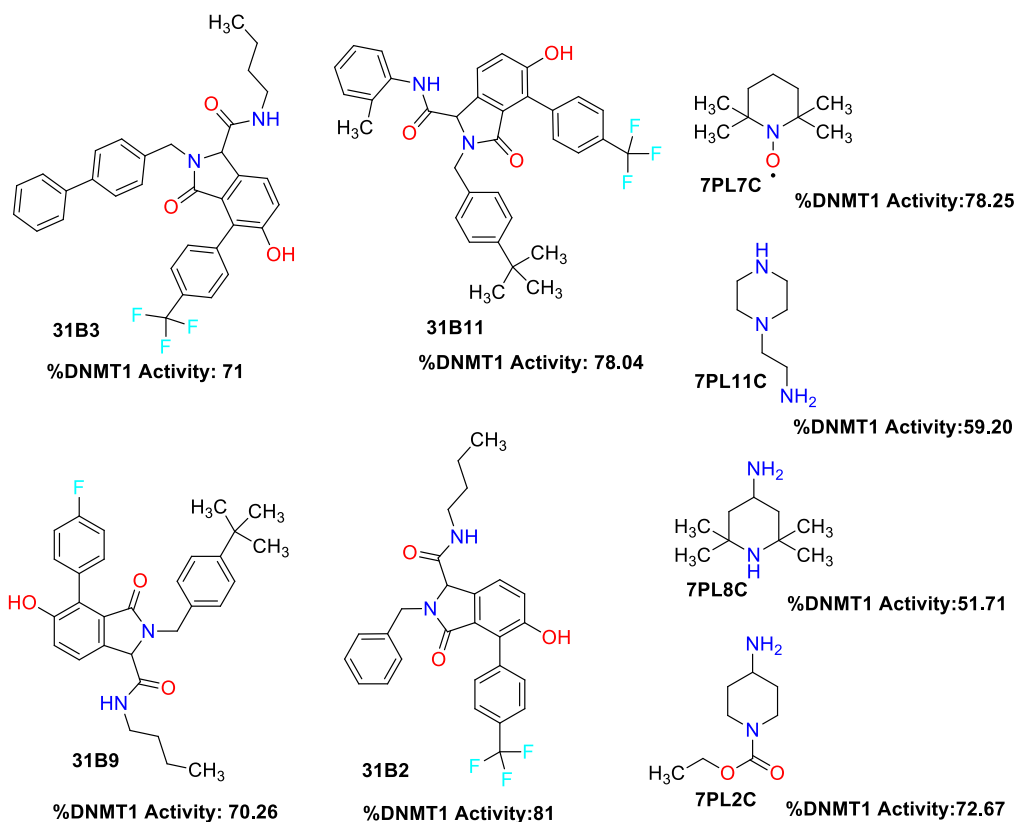


**Figure 3.3: Naphthyl group analogues from (Z)-2-acetamido-3-(naphthalene-2-yl) acrylic acid**

### 3.2 Promiliad Database Screening

Screening of Dnmt1 inhibitors using AutoDock Vina on the Promiliad database gave a group of isoindolinone compounds (**Figure 3.4**) with predicted Dnmt1 activity, from which analogues were further synthesized and docked to improve activity. A series of isoindolinone compounds (**Figure 3.5**) were synthesized from a sequential Ugi/intramolecular Diels Alder Furan (IMDAF)

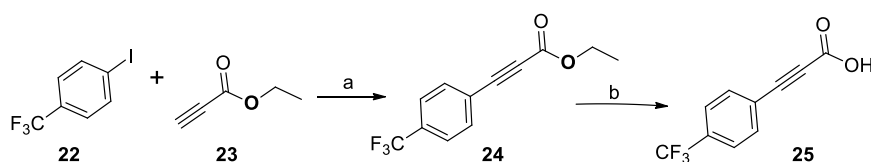
reaction using an acetylenic dienophile intermediate. The Ugi synthesis reaction involved the condensation of acetylenic acid, 2-furaldehyde, isocyanide, and primary amines (**Scheme 3.3**) and isoindolinone compounds were later obtained from an IMDAF reaction upon exposure of the Ugi product to catalytic amount of Yb(OTf)<sub>3</sub> in 1,4-dioxane at high temperature (**Scheme 3.4**).



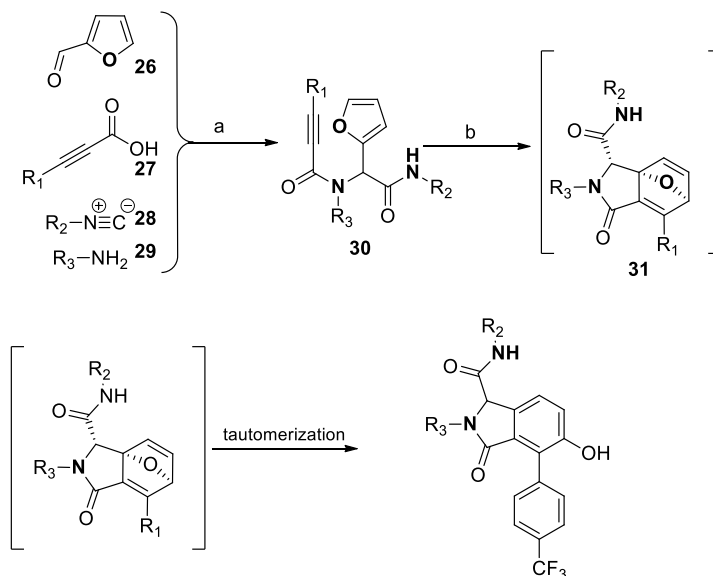
**Figure 3.4: Dnmt1 activities of Isoindolinone compounds from the Promiliad library, and the activity of selected piperazine and piperidine compounds.**

The acetylenic acid **27** and fufural **26** used for the isoindolinone analogues were kept constant while varying the isocyanide **28** and primary amine **29** components. Acetylic acid was obtained by direct Heck coupling of aryl iodide **22** and ethyl propiolate **23** in the presence of potassium carbonate (**Scheme 3.3**)<sup>[319]</sup> followed by hydrolysis of the ethyl arylpropiolates **24** with

potassium hydroxide in a water:tetrahydrofuran (1:1) mixture to give the acetylenic acid **25**. The importance of varying the amine and isocyanides component used in the Ugi reactions was to probe the different binding interactions of isoindolinone with Dnmt1 as a result of change of structure and possibly electrostatic charge around the compound. Thus, in the design of these isoindolinones certain isocyanides and amines were chosen to create structural difference in the library. Also some of the isocyanide derivatives used were components of isoindolinone with good Dnmt1 activity from previous assay (**Figure 3.4**), particularly cyclohexyl isocyanide, isopropyl isocyanide and tert-butyl isocyanide were used for the isoindolinone analogues.



**Scheme 3.3** (a)  $\text{Pd}(\text{PPh}_3)\text{Cl}_2$ ,  $\text{CuI}/\text{K}_2\text{CO}_3$ ,  $\text{THF}$ ,  $65^\circ\text{C}$  82% (b)  $\text{LiOH}\cdot\text{H}_2\text{O}$ ,  $\text{H}_2\text{O}:\text{MeOH}:\text{THF}$  (1:2:3), 78%



**Scheme 3.4** (a) 1 M in  $\text{MeOH}$ , RT overnight (b)  $\text{Yb}(\text{OTf})_3$ , dioxane  $100^\circ\text{C}$ , 12 h

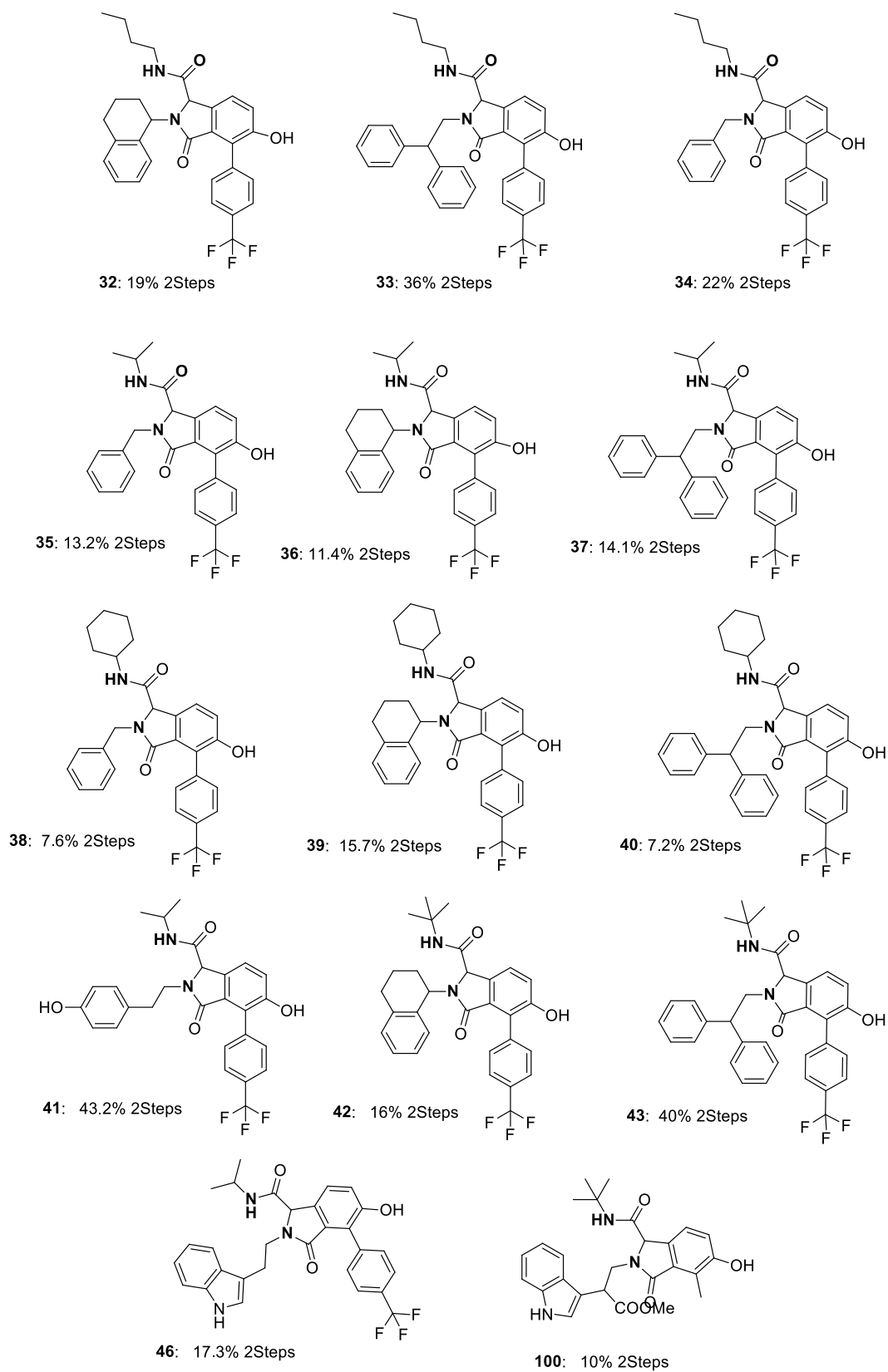


Figure 3.5: Isoindolinone analogues from variation of isocyanides and primary amine

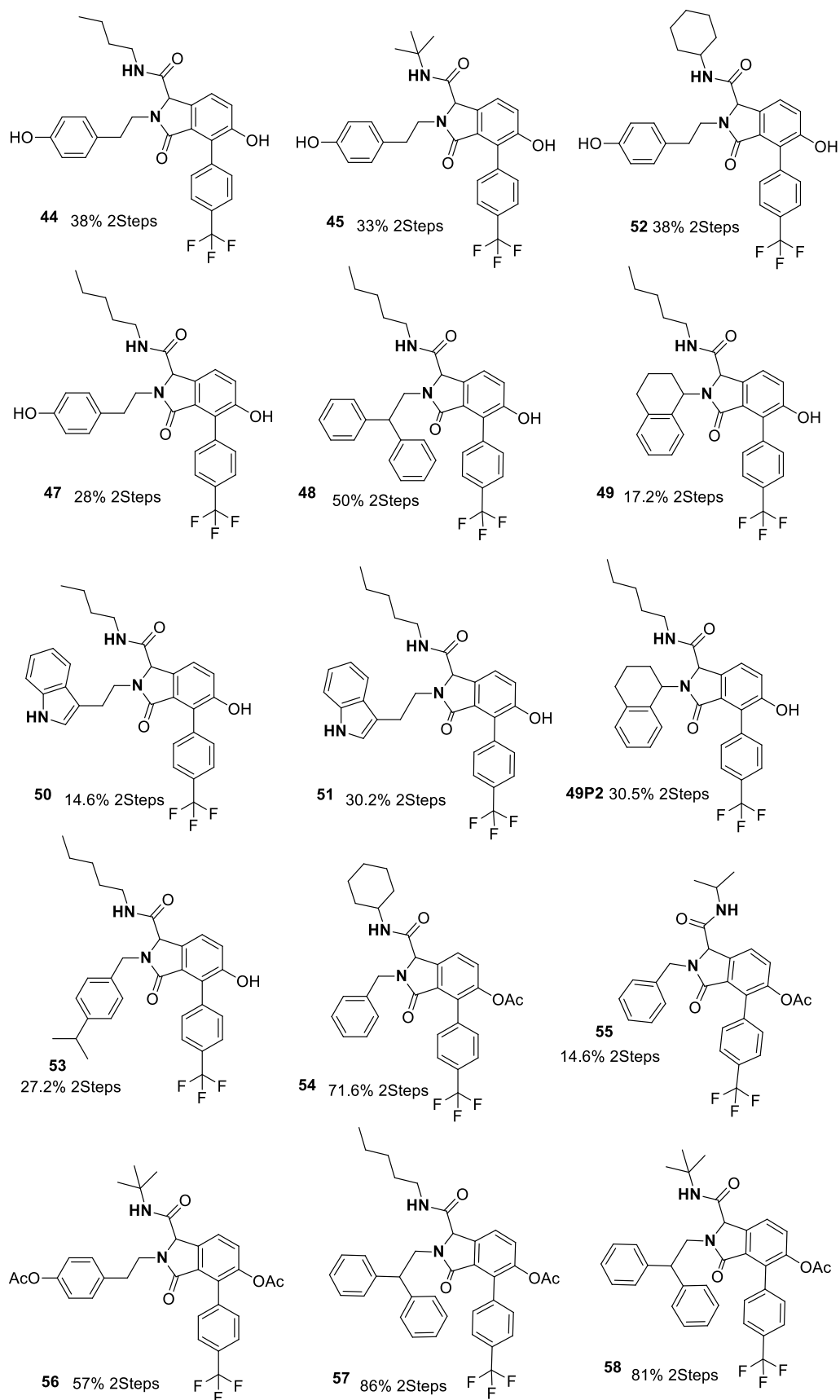


Figure 3.5: Isoindolinone analogues from variation of isocyanides and primary amine

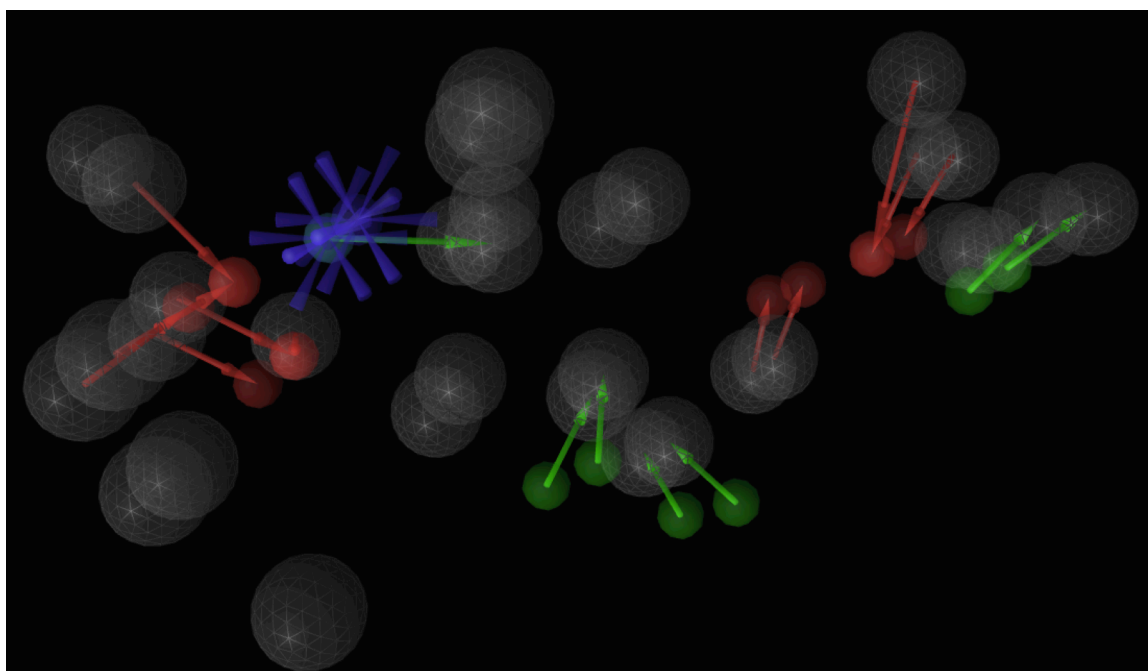
In summary, isoindolinone offers useful scaffold for drug design and it is present in many bioactive molecules<sup>[320]</sup> including both natural and synthetic compounds, and from virtual screening. In this project in this project it has been shown to have activity against Dnmt1. Consequently, we have used a three step reaction; synthesis of the acetylenic acid Ugi and IMDAF reactions produced a diverse set of compound library by varying two of the components. These compounds were later tested both on Dnmt1 and on breast cancer MCF7 cells.

### 3.3 Pharmacophore Modeling

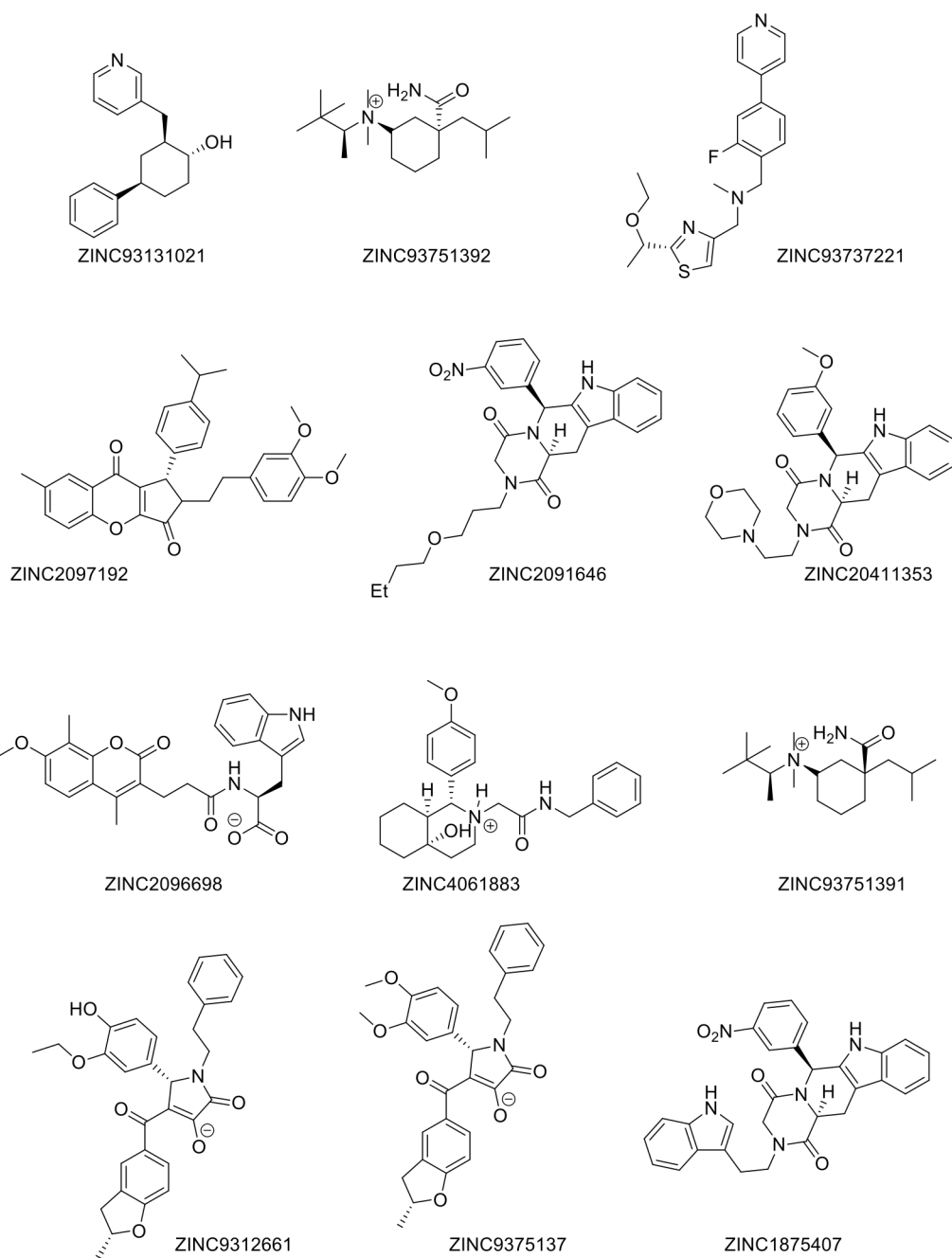
Pharmacophore modeling like virtual screening has been used in drug discovery for many years<sup>[321]</sup> and the models generated are either based on relevant chemical features derived from protein complex, as in the case of structure-based design or derived from common chemical features of known ligands (ligand-based design). Ligandscout<sup>[322]</sup> is a known software with tools that allow automatic generation of a 3D pharmacophore model either through structure-based or ligand-based design. The chemical features used by Ligandscout include lipophilic points (LIP), positive ionizable points (PI), negative ionizable points (NI), hydrogen bond acceptors vector (HBD), and finally hydrogen bond acceptor (HBA). In addition, the models generated by Ligandscout include inaccessible areas (steric restriction zones) to ligands based on defined interactions; these sterically forbidden regions are shown as excluded volume spheres and can be automatically generated and added to the pharmacophore model. For the design of a pharmacophore model for Dnmt1 inhibitors, two PDB entries (3PTA and 3SWR) containing *S*-Adenosyl-L-homocysteine and sinefungin bound to the DNA methyltransferase were used to automatically create two pharmacophore model, from which a common pharmacophore (**Figure**



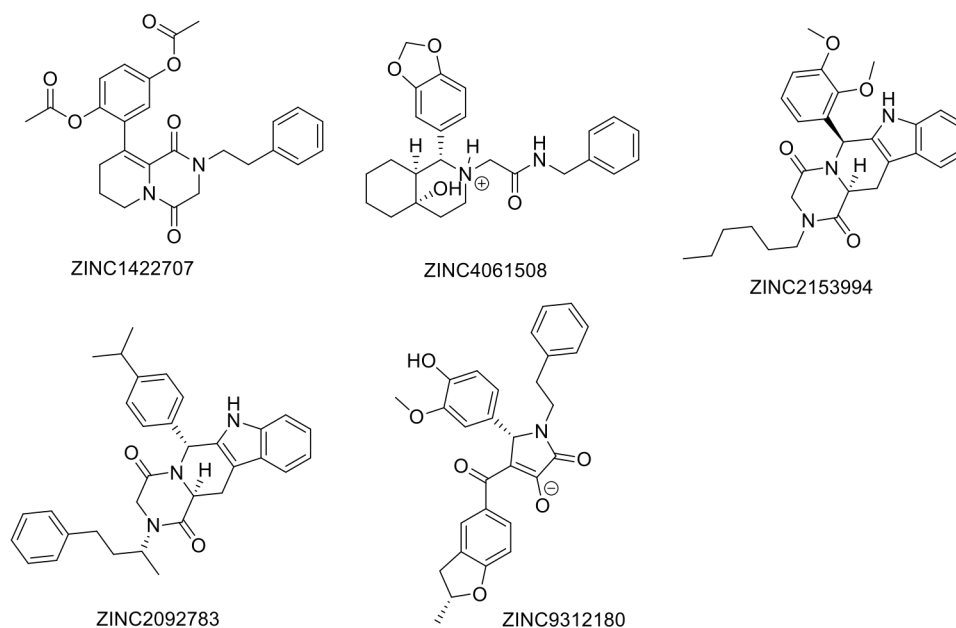
**3.6)** consisting of 1 positive ionization area, 7 H-bond donor, 1 H-bond acceptor and 22 excluded volume spheres is created. The model characterizes the protein environment surrounding the ligands in the structure. The performance metric used to assess the quality of a pharmacophore model is the Receiver Operating Characteristics (ROC) curves; this tells how good the model is able to discriminate between active and inactive ligand and the area under the curve (AUC) quantifies the ROC between 0 and 1, with values closer to 1 been favorable and vice versa. For this study, known inhibitors of Dnmt1 and inactive ligands were first docked into the model to determine the ROC and AUC value before screening of potential ligand was carried out. A 3D *in silico* screening of natural product compounds from the ZINC database was then carried out using the generated structure-based pharmacophore model with the aim of identifying structural templates that are similar to *S*-Adenosyl-L-homocysteine. The resulting hit molecules were ranked automatically based on maximum matching chemical features of the screened ligands with the pharmacophore (**Figure 3.7**).



**Figure 3.6:** Pharmacophore model using Ligandscout. Green arrows represent hydrogen bond donor, red are hydrogen bond acceptors, blue star is the positive ionization area.



**Figure 3.7: Top scoring natural product structures predicted by the pharmacophore model using LigandScout**



**Figure 3.7: Top scoring natural product structures predicted by the pharmacophore model using Ligandscout**

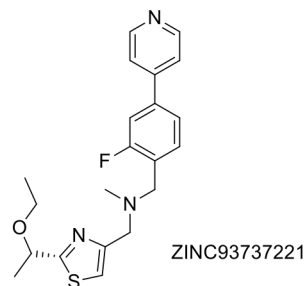
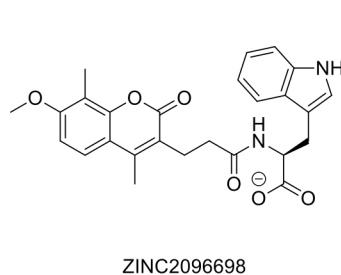
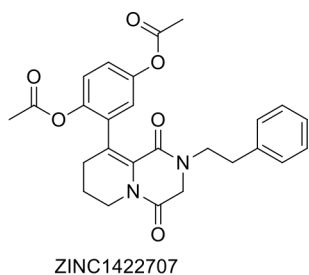
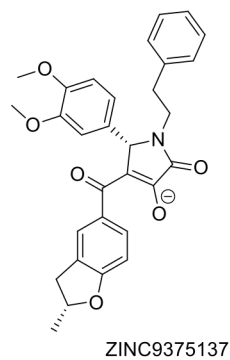
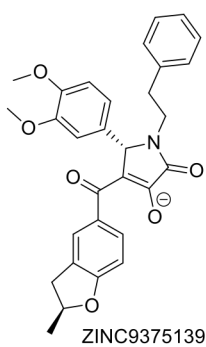
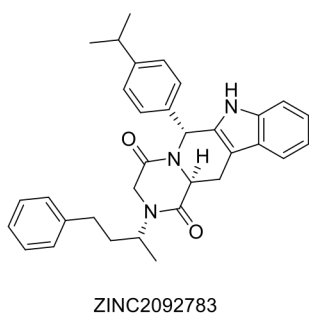
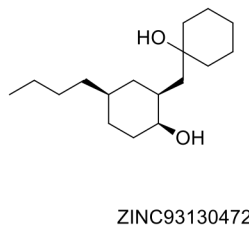
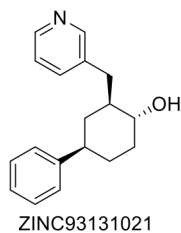
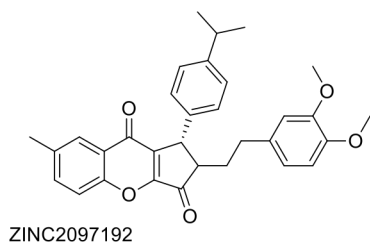
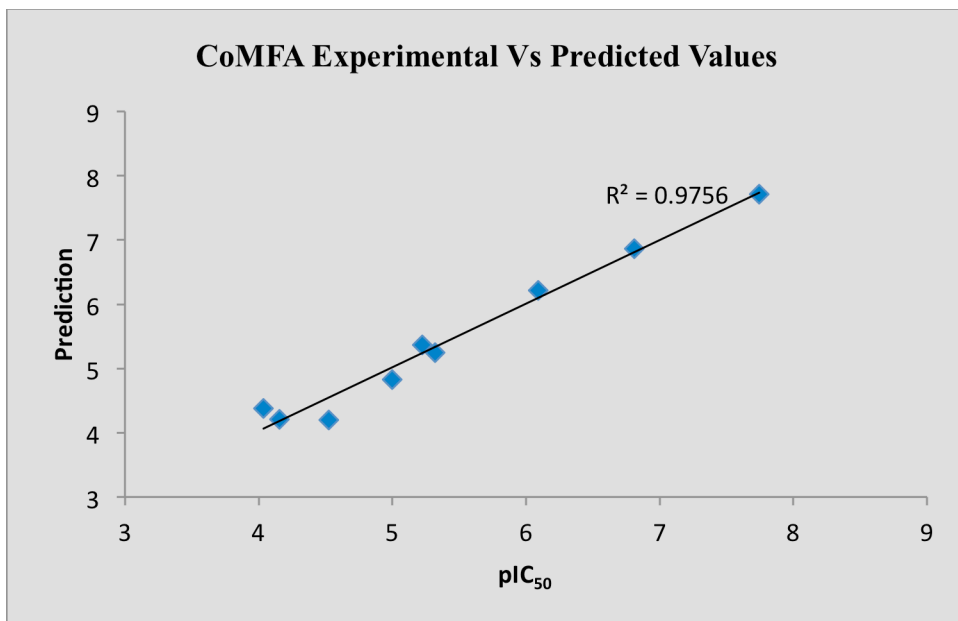
### 3.4 Comparative Molecular Field Analysis (CoMFA)

There is a correlation between structure and function and traditional quantitative structure-activity relationship (QSAR) correlates biological activities of active ligands with physiochemical or certain structural chemical features<sup>[323]</sup>. Also, ligand-receptor interactions depend on structural complementarity between the ligand and the binding site of the protein, the interactions are mostly described by 3-dimensional properties and the binding modes of ligands are determined by molecular structure<sup>[324]</sup>. Comparative Molecular Field Analysis (CoMFA) is a widely used 3D QSAR program that describes 3D structure-activity relationship in a quantitative manner, it involves calculating electrostatic and steric potentials energies of charged carbons atom located at each vertex of a rectangular grid and a series of other molecules. Certain atomic probes such as hydrogen bond donor or acceptor, or a lipophilic probe are then used to calculate fields values at each grid point, these values correspond to the energy values the probe would

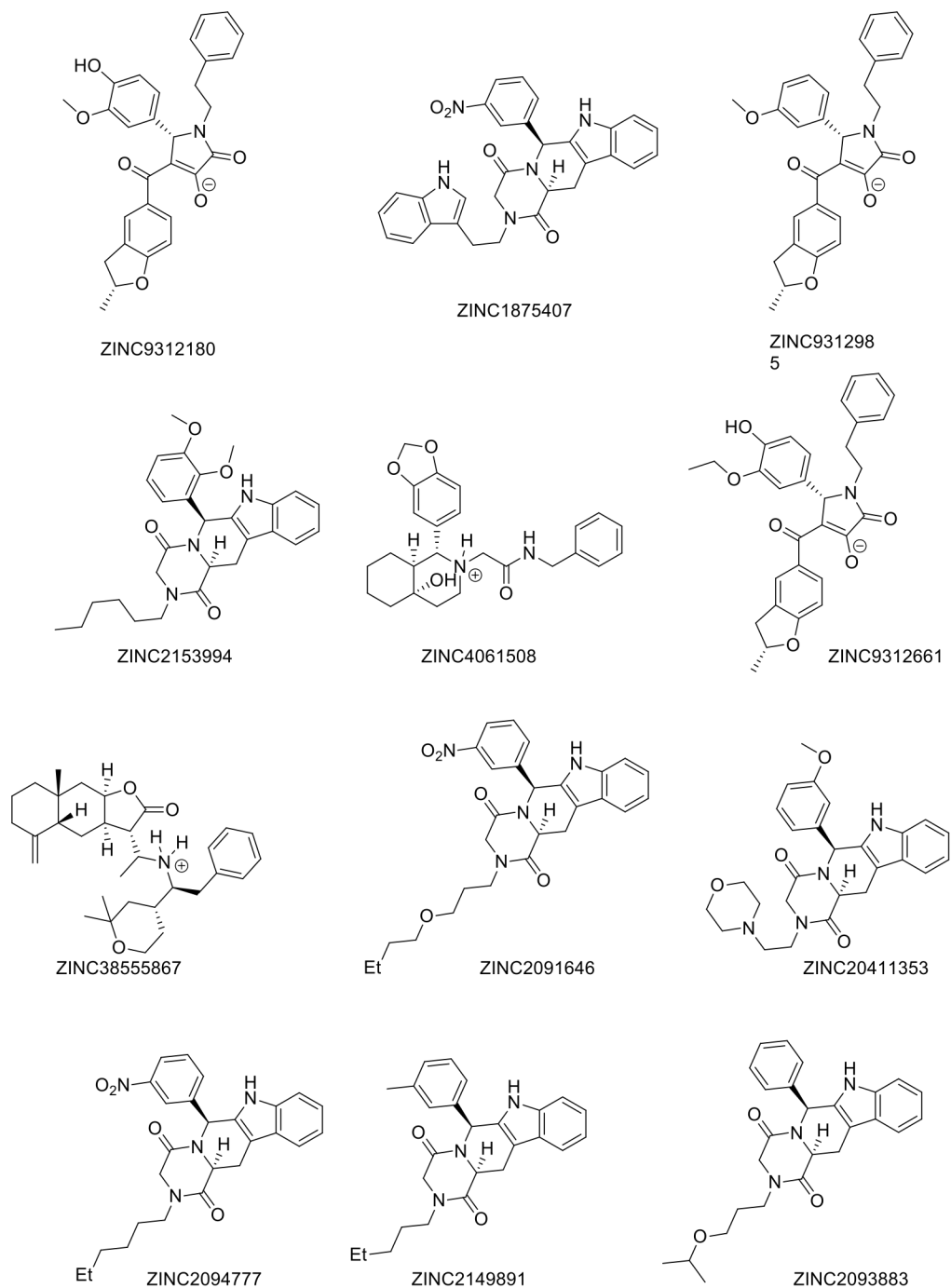
experience in a regular 3D lattice. The field values are correlated to binding affinity or biological activity through a partial least square (PLS) analysis. Usually, compounds with known biological activity are used to build a CoMFA model and test compounds can be screened using this model. For this work, in order to rationalize the structure activity relationship with natural product compounds from the ZINC database and compounds from ChEML database, CoMFA methodology was used to build a model from a training set of known Dnmt1 inhibitors with  $pIC_{50}$  values ranging from 3.3 to 7.7. The first step in building a model is usually structure preparation; structures of a set of nine Dnmt1 inhibitors was constructed using SYBYL sketcher and assigned Gasteiger-Huckel charges. The charged structures were later relaxed with Tripos force field. A CoMFA model (**Table 3.2**) was built after a substructure alignment of each compound and screening of a test set, (the same Dnmt1 inhibitors), revealed a squared correlation coefficient  $r^2$  of 0.9750, which is within the ( $> .60$ ) value for a good model. The potential predictive power of the constructed model was tested by screening of natural product compounds from IBScreen NP compounds obtained from the ZINC database and the top ranked structures are shown in **Figure 3.8**.

**Table 3.2: Dnmt1 inhibitors  $pIC_{50}$  and CoMFA values and predicted activities**

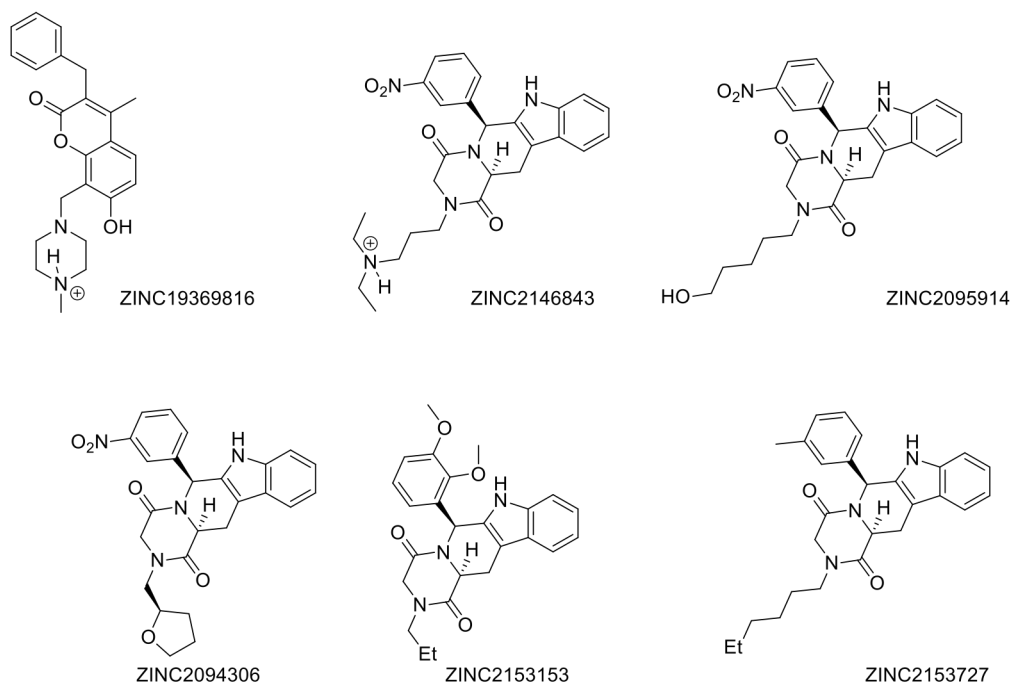
Dnmt1 Inhibitors	PKD	COMFA	PREDICTION	FUNCTIONAL_DATA
ATA	5.3196	872	5.035	0.2849
CURCUMIN	4.5228	1377	4.351	0.1723
EGCG	4.1549	1108	4.121	0.0335
NSC138419	3.3979	0996	4.432	1.0344
NSC14778	4.0362	0840	3.959	0.0774
NSC348926	3.3467	0760	3.202	0.1450
NSC408488	3.3010	1070	3.190	0.1114
NSC57278	3.3279	1130	3.799	0.4708
PROCAINAMIDE	5.0000	1372	4.443	0.5566
PSAMMAPLINS A	7.7447	1709	7.777	0.0322
RG108	6.9393	0850	7.033	0.0939
SGI-1027	5.2218	1861	5.073	0.1491
SID49645275	6.0969	0879	5.996	0.1012



**Figure 3.8: Structures of natural product predicted by CoMFA QSAR to have similar Dnmt1 inhibitory effect**



**Figure 3.8: Structures of natural product predicted by CoMFA QSAR to have similar Dnmt1 inhibitory effect**

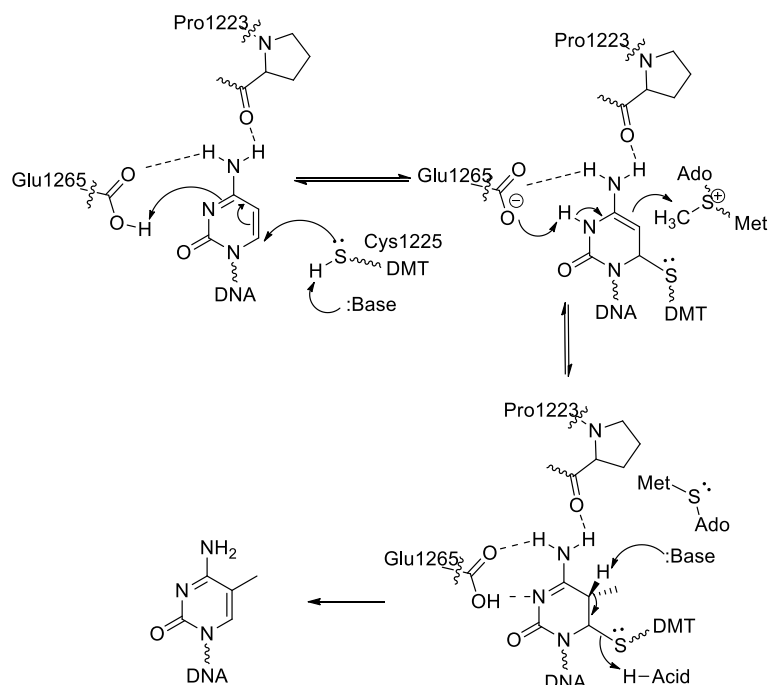


**Figure 3.8: Structures of natural product predicted by CoMFA QSAR to have similar Dnmt1 inhibitory effect**

### 3.5 Transition State Based Inhibitor Design

Initially, before applying computational methods to identify potential Dnmt1 inhibitors an investigation was carried out into the possibility of using transition state inhibitors. Since transition state theory suggest catalytic acceleration of enzymes is proportional to free energy released upon binding of transition state analogue, analogues that resemble the transition should give very powerful inhibitors. As a result, we designed and synthesized cytosine analogue with similar geometry and electronic features as the transition state during the methyl-transfer step from SAM to C5 position of cytosine (**Figure 3.9**). Cytosine with methylene-sulfur analogues at

the C5 position of cytosine might be one such structure that resembles the transition state during DNA methyltransferase especially one analogue that resemble the product.

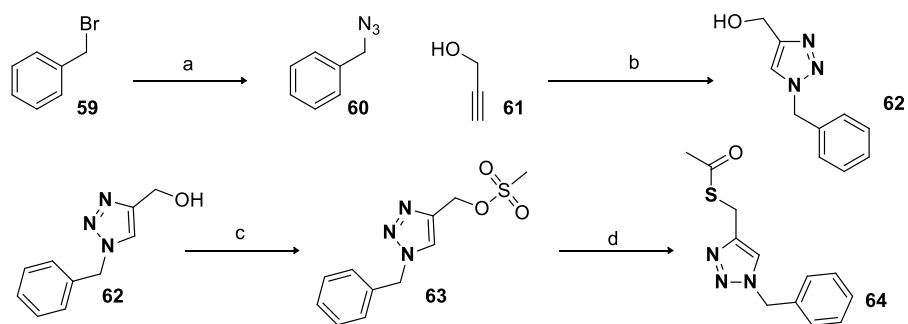


**Figure 3.9: Proposed schematic representation of the methylation reaction catalyzed by DMT.**

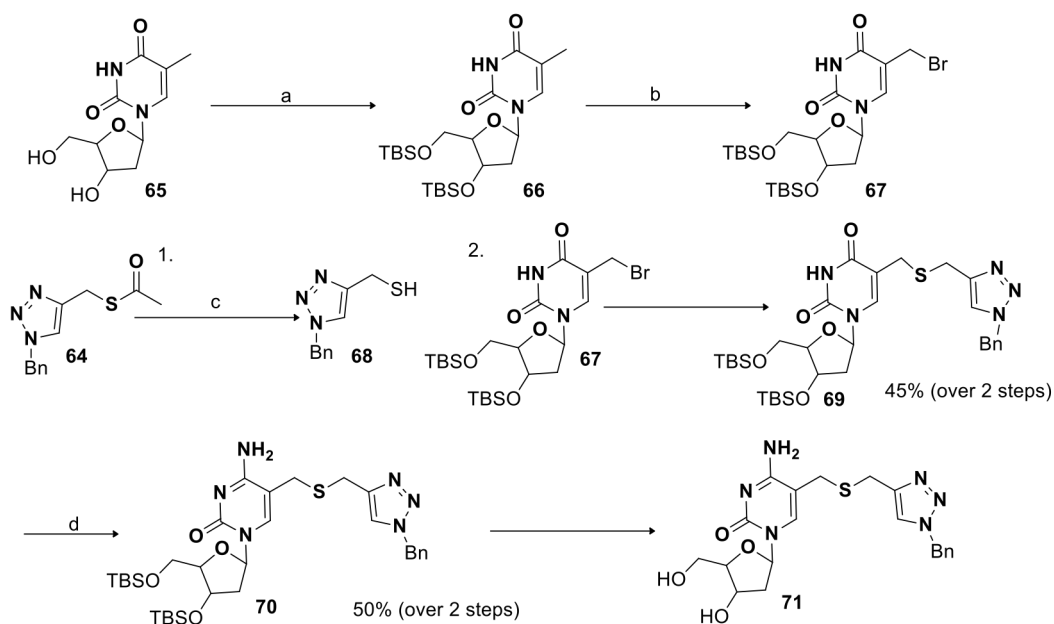
Kinetic and binding studies for Dnmt1 have not been well studied but steady-state studies of *M.HhaI*<sup>[325]</sup> and *M.MspI*<sup>[326]</sup> indicate an ordered binding of substrates, that is, DNA binds first before Adomet. In brief, the mechanism of the reaction shows that the key step is the formation of a Michael adduct between the sulfhydryl of a conserved cysteine (Cys-81 in *M.HhaI*) and C-6 of cytosine (**Figure 3.9**). The carbon-5 becomes nucleophilic for the S<sub>N</sub>2-type attack onto the methyl sulfonium center of AdoMet, which results in the capture of the methyl sulfonium center of AdoMet resulting in the transfer of the methyl group to the ring. The resulting intermediate undergoes β-elimination at the C-5 position to give the methylated cytosine (**Figure 3.9**). For the transition state inhibitor design in this study we considered cytosine structure resembling the first step in the mechanism (**Figure 3.9**) without the cysteine bound at the C-6 position. In



addition, we introduced a triazole ring with an attached benzyl group to mimic purine ring, that is, SAM at that position. The synthesis of the transition state inhibitor analogue was divided into two parts; one part involved the construction of the triazole moiety **65**, and the other a brominated thymidine analogue **68**, both parts were later coupled to give the final product. For the synthesis of the triazole, a simple  $S_N2$  reaction converts benzyl bromide **59** to benzyl azide<sup>[327]</sup> **60**, which can then undergo 1.3-dipolar cycloaddition<sup>[328]</sup> reaction with an propargyl alcohol **61** to give the corresponding triazole **62**. The reaction is catalyzed by copper (I) which is generated in-situ by the reduction of a less expensive copper (II) salt ( $CuSO_4 \cdot 5H_2O$ ) with ascorbic acid or sodium ascorbate. Subsequently, the resultant alcohol was easily converted to mesylate<sup>[329]</sup> **63** in high yield and then to the thioester<sup>[330]</sup> **64** using potassium thioethanoate. This thioester can be reduced to the thiol<sup>[331]</sup> **68** with pyrrolidine and then coupled to the brominated thymidine **67** via an  $S_N2$  type reaction. The brominated thymidine derivative was obtained by first protecting the hydroxyl groups in thymidine with tert-butyldimethylsilyl (TBDMS) to give 3',5'-bis-O-(tert-butyl dimethylsilyl) protected thymidine derivative<sup>[332]</sup> **66** and then a radical bromination<sup>[333]</sup> **67** in the benzylic position was achieved with N-bromo succinimide (NBS) and 2,2'-azo-bis-isobutyronitrile (AIBN). In the final step of this reaction, the thymidine sulfide derivative was converted to cytosine derivative via a two-step reaction. The first step involve the conversion thymidine analogue to a triazolyl derivative with 1,2,4- triazole and phosphoryl chloride, then after the triazole group at the C-4 position is replaced with amino group with aqueous ammonia followed by deprotection of the silyl to give the cytosine derivative **71** (Scheme 3.6).



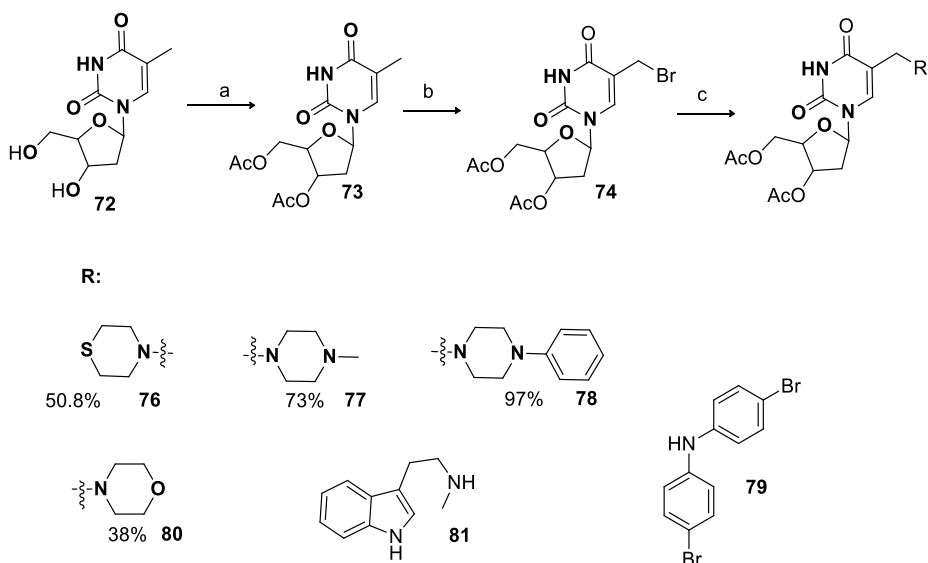
**Scheme 3.5:** (a)  $\text{NaN}_3$ , DMSO 91% (b)  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ , Na.Ascorbate  $\text{H}_2\text{O}:\text{tButyl alcohol}$ , 66% (c)  $\text{MeSO}_2\text{Cl}, \text{CH}_2\text{Cl}_2$  83% (d) potassium thioacetate, DMF, 54%



**Scheme 3.6:** (a)  $\text{TBDMSCl}$ , Imidazol, DMF, 91% (b) NBS, AIBN Benzene  $80^\circ\text{C}$  (c)  $\text{Et}_3\text{N}$ , Pyrrolidine, DMF (d) 1,2,4-triazole,  $\text{POCl}_3$ , DMAP, MeCN then  $\text{NH}_3$ , Dioxane

Secondary and tertiary amine functionalization at the C5 position of nucleoside gives rise to a special class of 5-substituted pyrimidine nucleosides with potent antiviral<sup>[334]</sup> and anticancer<sup>[335]</sup> properties and there is a continued interest for cyclic amines<sup>[336]</sup> at the C-5 position of pyrimidine. We evaluated the medicinal property of the amine derivatives of nucleoside using thymidine and several amines: cyclic amines, straight chain and aromatic amines were used and the synthetic steps involve an  $\text{S}_{\text{N}}2$  reaction of the amine with brominated thymidine. One

important amine used for this synthesis is bis(4-bromophenyl)amine **79** which showed good anticancer property and would be a good amine to use as a template for other thymidine derivatives.

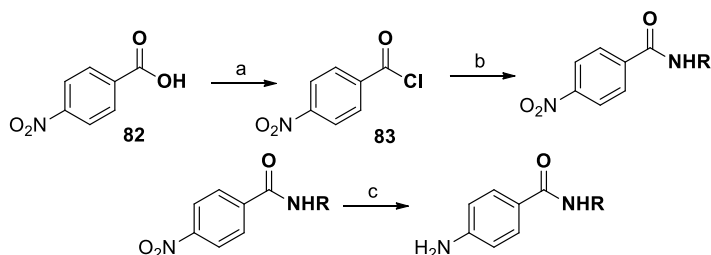


Scheme 3.7 (a)  $\text{Ac}_2\text{O}$ , DMAP (b) NBS, AIBN Benzene (c) DIPEA -THF

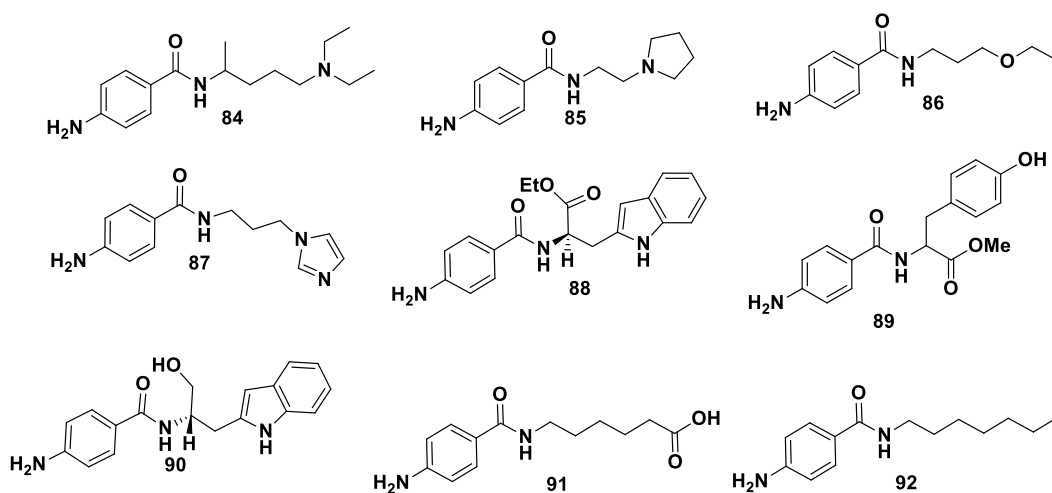
### 3.6 Procainamide Analogues

Unlike the transition state nucleoside analogues that bind to the enzyme, procainamide and its derivatives do not have a definitive mode of inhibition for Dnmt1. Procainamide is speculated to bind to CG-rich sequence, which then mediates demethylation, but they are widely accepted Dnmt1 inhibitors. In an effort to make other analogues of procainamide (**Figure 3.10**) we sort other amines that had the same 1,3-connection of nitrogen atoms as in procainamide, in addition, we used other cyclic and modified amines to test whether the inhibitory ability of procainamide could be improved. The synthesis started with conversion of 4-nitrobenzoic acid **82** to the acid chloride **83** using oxalyl chloride and dimethyl formamide, and then amide formation with respective amine groups in methylene chloride with catalytic amounts of triethylamine.

Subsequently, the nitro group was reduced with palladium on carbon at in methanol (**Scheme 3.8**).



**Scheme 3.8** (a) Oxalyl chloride, DMF, CH<sub>2</sub>Cl<sub>2</sub> (b) RNH<sub>2</sub>, Et<sub>3</sub>N, CH<sub>2</sub>Cl<sub>2</sub> (c) Pd/C, H<sub>2</sub> 1atm MeOH



**Figure 3.10:** Procainamide analogues using diverse amine groups.

Several important molecules have been proposed and tested as potential Dnmt1 inhibitors and in this study many more structures have been added. In particular, isoindolinones and transition state nucleosides have been uncovered as potent potential inhibitors of Dnmt1. Furthermore, a number of diverse natural products have been discovered that share structural and electronic similarity as S-adenyl-L-methionine and S-adenyl-homocystiene (SAH), which are known substrate and product of methylase reaction respective. Also, pharmacophore design agrees with

most of the structures proposed by CoMFA QSAR study and the structural scaffold is in line with tryptophan structures of **RG108**.

## CHAPTER 4

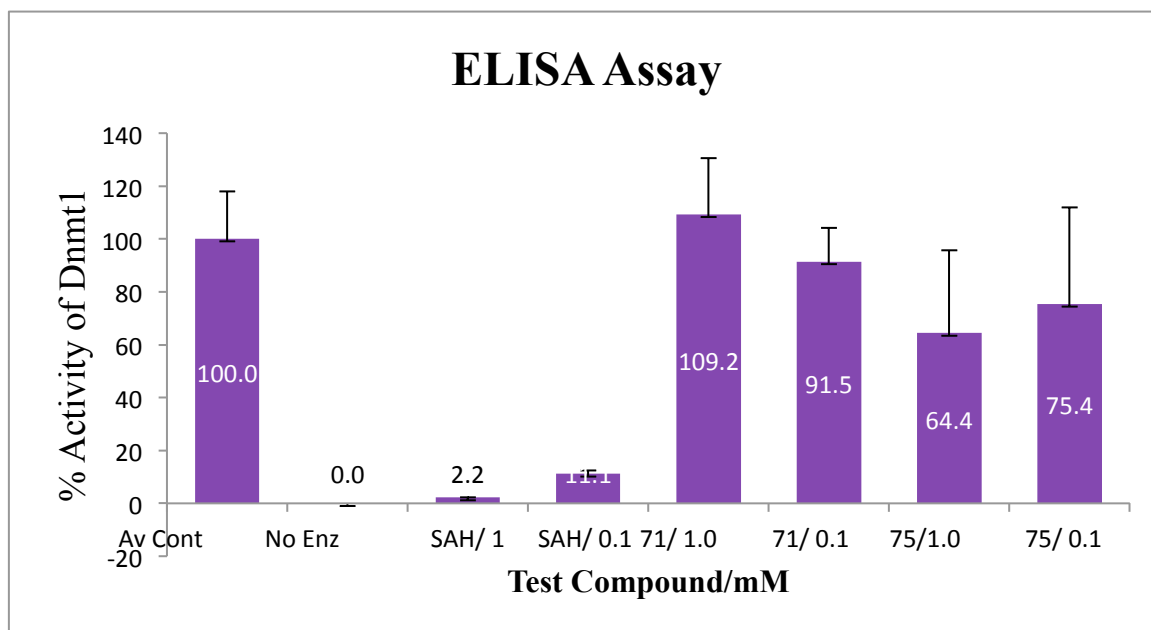
### RESULT AND DISCUSSIONS

#### 4.1 Transition State Inhibitors

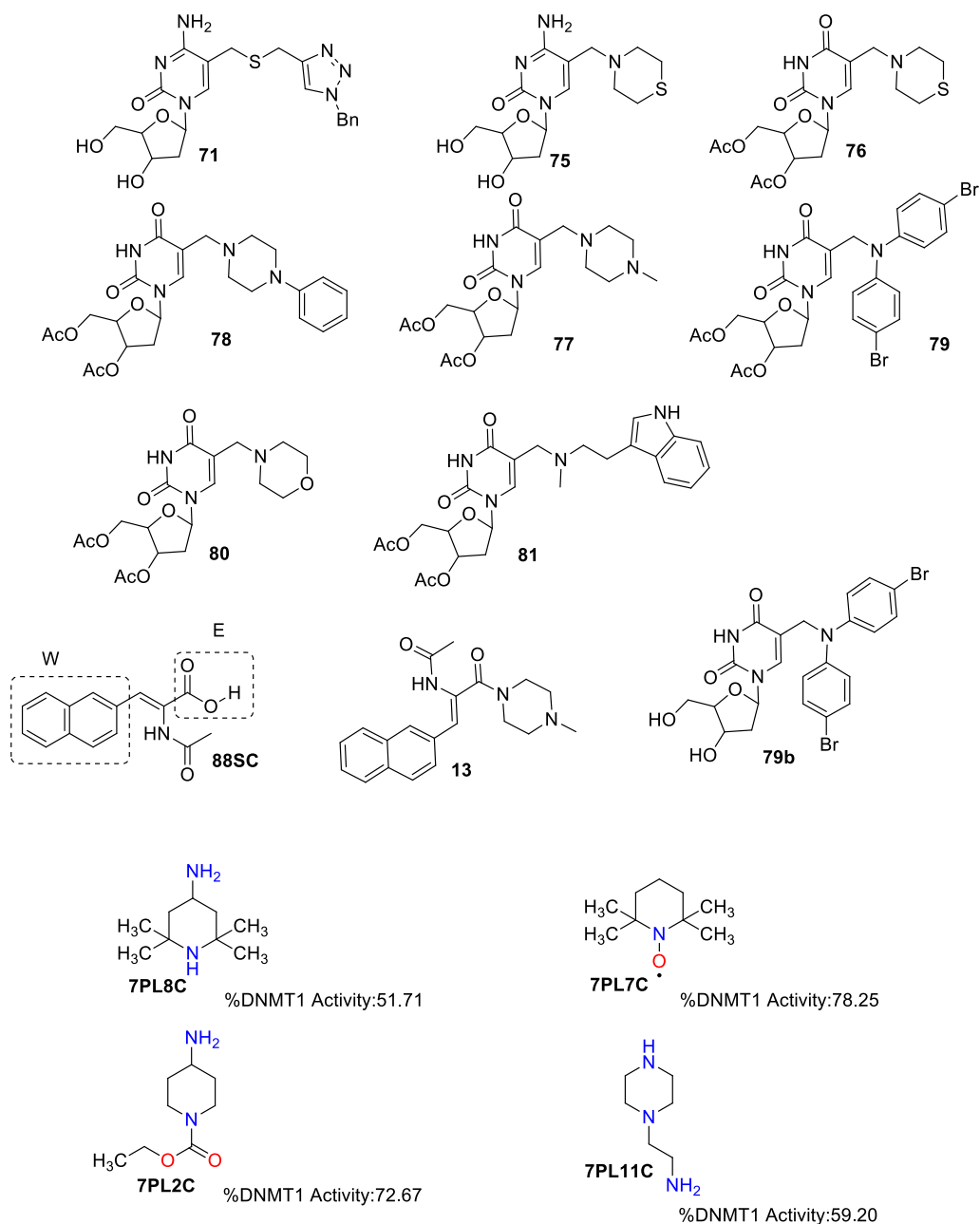
Design and development of transition state analogs as competitive inhibitors of Dnmt1 was our initial approach for discovery of novel epigenetic modulators. We hypothesized that deoxycytidine analogs with structures similar to the methyl-transfer step from the SAM cofactor to cytosine would have a greater binding affinity for the catalytic site of Dnmt1 compared to either the cytosine substrate or the SAM cofactor alone. Thus, we designed substrate analogs of deoxycytidine modified at the C5 position and compound **71** was the primary compound proposed. The compound contains three structural features resembling the transition state: first the cytosine substrate, the methylene-sulfur linker, and lastly a benzyl-1H-1,2,3-triazol heterocyclic moiety to represent the adenosyl portion of the cofactor SAM. At the time, we reasoned that these three features would represent a starting point for the molecular descriptors of the substrate and the cofactor. Also, Stefely *et al.*<sup>[337]</sup> had shown that aryl amides of benzyl-1H-1,2,3-triazole inhibit cancer cell growth of MCF-7 human breast tumor cells with IC<sub>50</sub> of 46 nM. In addition, methylation studies by Stein and Fatemi revealed that Dnmt1 binds preferable to hemi-methylated DNA<sup>[97b, 98]</sup>. In essence, our proposed inhibitor **71** captured the important triazole pharmacophore needed for anticancer activity and a cytosine substrate recognized by Dnmt1 for selectivity

To further expand our library and study of substrate analogs, we designed and synthesized C5-amino derivatives of cytosine and thymidine (compounds **75** and **76 – 81**) to complement compound **71**. Formerly, Ward<sup>[338]</sup> and Prusoff<sup>[339]</sup> highlighted pharmacological

interest of 5-substituted pyrimidine nucleoside for their antiviral and anticancer properties. However, for our study, we were interested in the possible binding and transfer of the methyl group from SAM to the C-5 amino pyrimidine derivatives. Since N-methylation can occur both at  $sp^2$  and  $sp^3$ -N-atom in the heterocyclic bases as well as the C(5) cytosine, we figured introduction of cyclic amines with  $sp^3$ -N-atom at C(5) position of cytosine will have interesting nucleophilic properties for methyl transfer. Also, we previously discussed how chromatin structure can be regulated through N-methylation of specific lysines in the N-terminal tails of histone H3 and H4. Mechanistically, we considered N-methylation of compounds **75** and **76 – 81** would be a non-covalent event with respect to Dnmt1 compared to the known aza-nucleotide suicidal inhibitors and we believed this class of compounds would be better inhibitors.



**Figure 4.1: ELISA Assay: screening of C5 pyrimidine analog 71 and 75 with an ELISA assay for activity. Data represent the average values of triplicate measurements. Error bars represent SD of the triplicate biological measurement**

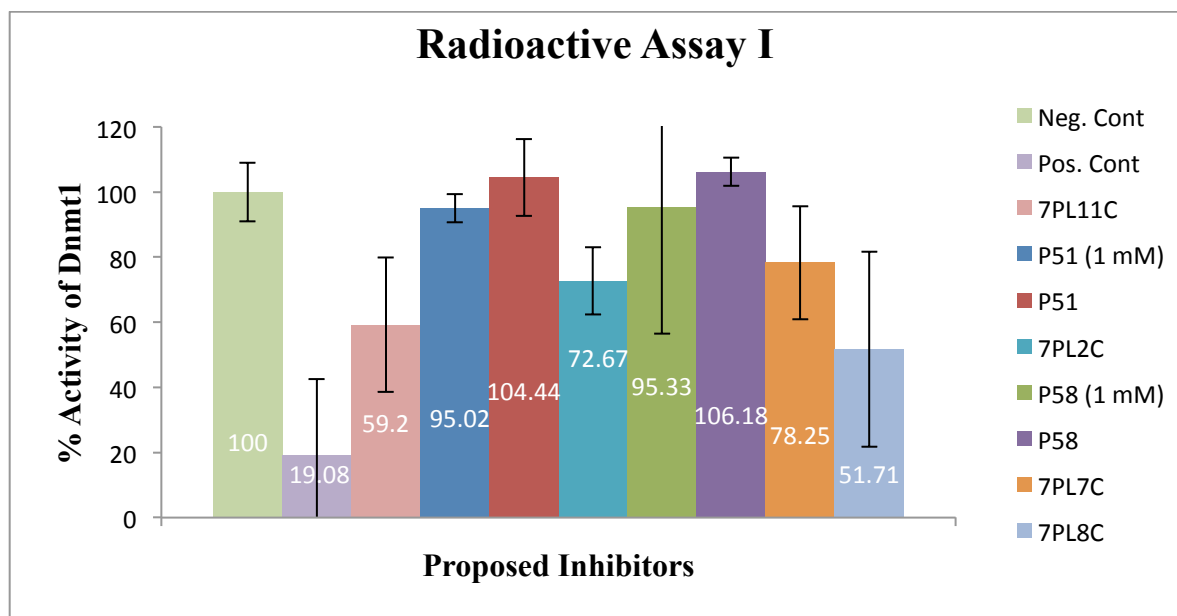


To evaluate the inhibitory activity of C(5)-pyrimidine, we used an enzyme-linked immunosorbent assay (ELISA) to screen the first set of compounds for Dnmt1 inhibition. The assay was performed by Dr. Brooke Martin using the EpiQuik Dnmt1 Assay Kit and showed that neither the thio-analogs nor the amino analog displayed any inhibitory activity at micro molar concentration (**Figure 4.1**). However, the C5-amino derivative **75** at 1 mM reduced the



activity of the enzyme by 35.6%. Though not a spectacular inhibitor, it shows that the amino functionalized nucleoside hold possible ability to inhibit Dnmt1.

A second evaluation method using radioactive methyl transfer assay developed by Jeremy Alverson of the Priestley group was used to test the ability of nucleoside inhibitors to influence human Dnmt1 activity. The method is a medium-throughput screening assay that measures the direct incorporation of tritium ( $^3\text{H}$ ) into the oligonucleotide poly-deoxyinosinic-deoxycytidylic acid (poly(dI-dC)). Once again the data reaffirms the inactivity of **71** against Dnmt1 even at 1 mM (**Figure 4.2**), however, **75** showed slight activity at the same high concentration (1 mM)



**Figure 4.2: Radioactive Assay I - In vitro radioactive assay to evaluate the effect of nucleoside and amines on Dnmt1. Unless state, all compounds are assayed at 100  $\mu\text{M}$ . Data represent the average values of triplicate measurements. Error bars represent SD of the triplicate biological measurement**

Though both **71** and **75** showed no measureable inhibitory activity at 100  $\mu\text{M}$ , hepatotoxicity studies of HepG2 cell revealed that both compounds have minimal to undetectable liver toxicity at concentrations greater than 500  $\mu\text{M}$ . Implying that further development of **75** with attention to electronic and steric might give rise to less toxic inhibitors.

## Discussion

In summary, nucleoside inhibitors such as 5-azacytidine, 5-aza-2'-deoxycytidine and zebularine are well established inhibitors of Dnmt1 and have been approved for the treatment of high-risk myelodysplastic syndrome (MDS) – a condition of ineffective production of myeloid in the blood. These drugs function as suicide inhibitors of Dnmt1 through irreversible covalent interaction with the enzyme and are incorporated into the DNA after phosphorylation. As such, we initiated an attempt to reengineer cytidine and thymidine at the C5 position to create compounds with similar activity but less toxic and more selective compared to the azanucleotides. Compound **75** showed minor activity against Dnmt1 however this compound displayed a remarkable lack of toxicity at high concentration (>500) better than the azanucleotides, which are cytotoxic at high concentration and thus limit their pharmaceutical use.

## 4.2 Virtual Screening

Our second approach towards discovering novel competitive inhibitors of Dnmt1 involved computational modeling. We obtained four unique organic structures (Figure 28) from docking over 5000 compounds in the catalytic pocket of Dnmt1 (PDB:3PTA) with the AutoDock virtual screening program and to validate the result, two of these compounds were tested for inhibition of enzymatic activity. The radioactive assay showed that ZINC04710208 (**88SC**) inhibits Dnmt1 with an  $IC_{50}$  of 233  $\mu$ M, while, **ZINC02835213** exhibited only slight inhibition even at 250  $\mu$ M (**Table 4.1**). Though **ZINC01694996** and **ZINC18140189** were not tested at the time, they are worth discussing a little further because of their peculiar steric and electronic features, which could interact with specific residues in an enzyme. Plus

**ZINC18140189** has remarkable pharmaceutical importance and it is a known drug, phenytoin – a major first line antiepileptic drug (AED).

**Table 4.1: Radioactive assay of hit compounds from ZINC database. %Dnmt1 inhibition is based on <sup>[340]</sup>radioactive SAM and CPM implies count per minute.**

Sample	Average CPM	Corrected CPM			% DNMT1 Activity	
Blank	2325.92	0.00	±	125.46	NA	
Neg. Control (No inhibition)	8632.55	6306.64	±	988.67	100.00	± 15.68
Pos. Control (10 µM SAH)	3446.08	1120.17	±	363.43	17.76	± 32.44
ZINC04710208- <b>88SC</b> (500 µM)	2071.88	-254.04	±	225.0226	-4.03	± -88.58
ZINC04710208 <b>88SC</b> (250 µM)	4890.06	2564.14	±	270.5037	40.66	± 10.55
ZINC04710208 (125 µM)	8309.89	5983.97	±	389.7078	94.88	± 6.51
ZINC02835213 (250 µM)	9396.28	7070.36	±	1016.382	112.11	± 14.38

Our interest in phenytoin is not for its antiepileptic properties but for one of its side reactions noticed with patients taking the drug. Phenytoin like procainamide, hydralazine, is known to cause drug induced-lupus<sup>[340]</sup>, a condition that is reversible. This observation aligns with our discussion in chapter 2 where Cornacchia and colleagues observed that procainamide and hydralazine are capable of inducing a lupus-like syndrome in humans similar to 5-azacytidine when treated with cloned T cells.<sup>[302]</sup> Also Rubin<sup>[340]</sup> has characterized 40 known drugs into three categories: high, low and very low risk of inducing lupus. His classification is based solely on the number of times these drugs appear in the literature in connection with drug-induced lupus, and this classification puts phenytoin in low risk category compared to procainamide and hydralazine. It is also worth adding that Eldredge<sup>[341]</sup> and colleagues had experimentally place phenytoin in the same low risk category based the interaction of phenytoin and other lupus-inducing drugs with DNA. Again procainamide and hydralazine remained top on the list because they altered the optical rotation of native DNA in addition to lowering the melting temperature of native DNA. A clear indication that lupus-inducing drug mode of action might involve DNA interaction.

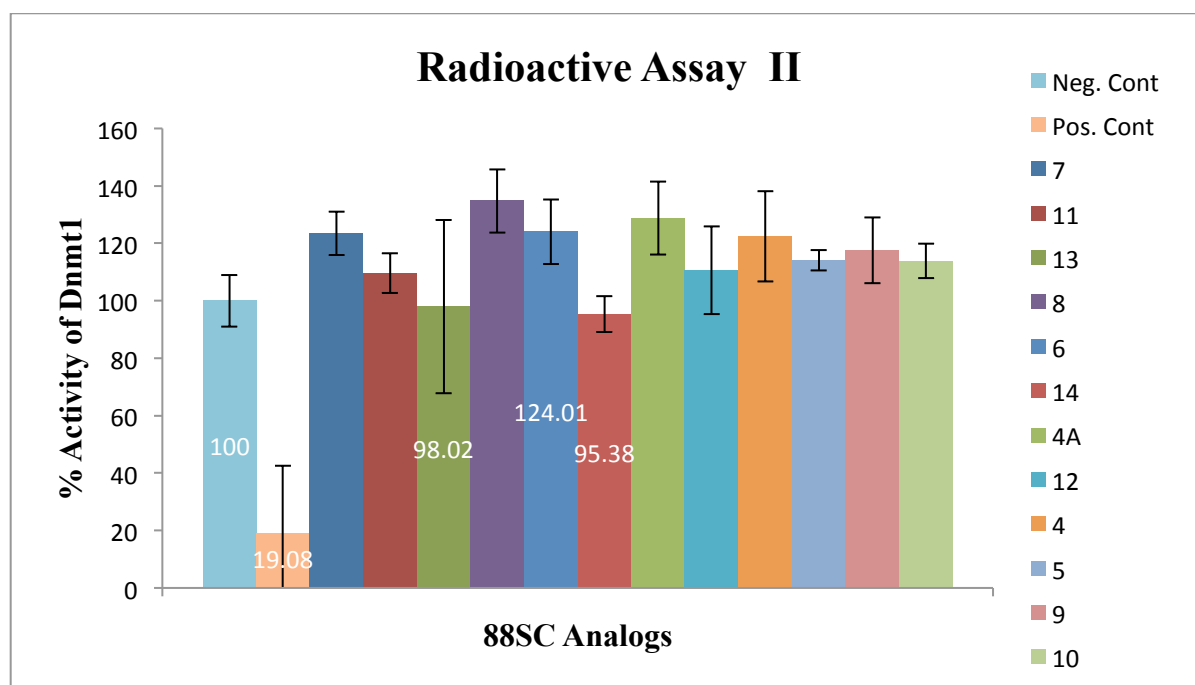
Finally, though Singh et al modeled procainamide and hydralazine interaction in the catalytic pocket of Dnmt1 as the main mode of inhibition, we cannot rule out DNA interaction with these molecules as another possible mode of action. However, for our current study, phenytoin will be a good starting molecule since it is a low drug-induced lupus compound, plus it has structural features and handles for lead development. Also, given that Singh et al computational studies implied that procaine and procainamide interact with similar residues as 5-azacytidine and it is possible that this group of drug-induced lupus might be competitive inhibitors of Dnmt1 as well.

### 4.3 Analogs of ZINC04710208 (88SC)

The next step of development for **88SC** was to make analogs using the inherent synthetic handles: the carboxylic acid group and the naphthalene ring. First we decided to increase the polarity of the carboxylic acid pharmacophore functionality by converting them to amides, particularly because two cyclic amines (piperidine and piperazine) showed reasonable Dnmt1 inhibition at a 100  $\mu$ M by 78% and 59% respectively. Thus using the Plochl and Erlenmeyer reaction we constructed the oxazolone ring which was later opened with different primary and secondary amines to give compounds **4 – 14**. Furthermore, to ascertain the importance of the interaction of the naphthalene ring to the activity of **88SC**, we replaced the naphthalene groups with a number of free rotating and smaller aromatic rings and groups, compounds **15 – 21**.

Compared to **SS8C**, the amide and naphthalene modification showed no activity (**Figure 4.3**). In fact, the amines or the carboxylic acid of **88SC** showed more inhibitory activity alone than the amide. This implies that the sterics and conformation inflexibility of the naphthalene group locks it inside a hydrophobic pocket that allows the carboxylic acid to make polar interactions with important enzyme residues necessary for methylation. If this is the case, it

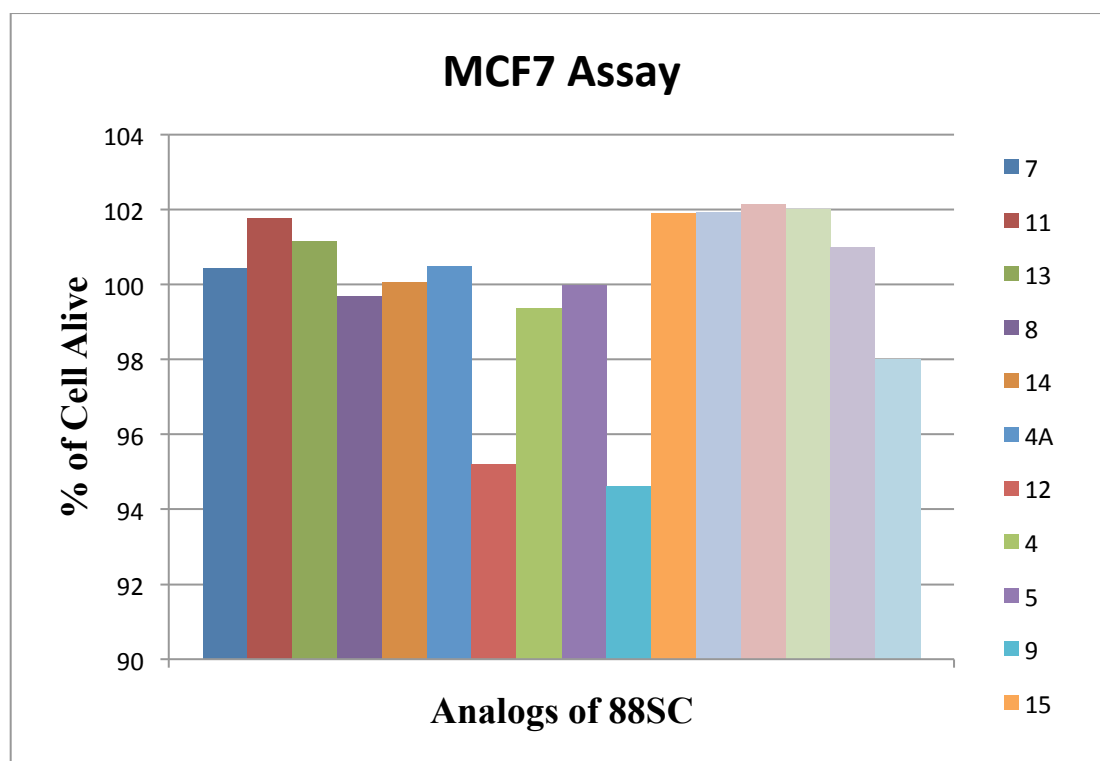
means the structure **88SC** can be optimized with functional groups that either contributes to the hydrophobicity or polarity on either ends of the molecule. However, the synthetic handles on the molecule are few and if the hydrogen bonding interactions of the carboxylic acid are already optimal changing this group with classical (sulphonic acid or N-hydroxyamide) or non-classical bioisoteres (sulfonamide or tetrazole) of  $-\text{COOH}$  might not significantly increase activity.



**Figure 4.3: Radioactive Assay II: In vitro radioactive inhibition assay of 88SC analogs on Dnmt1 activity. All compounds are assayed at 100  $\mu\text{M}$ . Data represent the average values of triplicate measurements. Error bars represent SD of the triplicate biological measurement**

As a step forward towards discovery of anticancer agents through inhibition of Dnmt1, we exposed our amide **88SC** analogs to human breast cancer cells MCF7 and normal breast cells MCF10. After treatment of compounds **4 – 14** for 72h, none of the analogs, even compound **6** (phenylpiperazine derivative), inhibited the growth of MCF-7 at 100  $\mu\text{M}$  (**Figure 4.4**) Unfortunately, **88SC** did not induce any inhibition of MCF7 and neither did **13**, a

methylpiperazine amide of **88SC**. However, piperazine derivatives have been reported to have antiproliferative activity against human cancer cell lines. Kumar et al.<sup>[342]</sup> reported antiproliferation activity of 1-benzhydrylpiperazine derivatives against MCF-7, HeLa, HT-29 carcinoma cell lines. Also, Cao and colleagues demonstrated that 4-substituted-piperazine-1-carbodithioate derivatives of 2,4-diaminoquinazoline exhibit a broad spectrum of inhibitory effects on MCF-7, HeLa HCT 116 and A549 with an IC<sub>50</sub> value in the range of 2.76 – 4.12 μM<sup>[343]</sup>. According to Cao, the 4-substituted – piperazine derivatives act by two possible mechanism, either as antifolates or DNA damage-inducing agents.



**Figure 41. MCF7 Assay: Treatment of 88SC analogs on on human breast cancer cell MCF -7. All compounds are assayed at 100 μM. Data represent the average values of triplicate measurements. Error bars represent SD of the triplicate biological measurement**

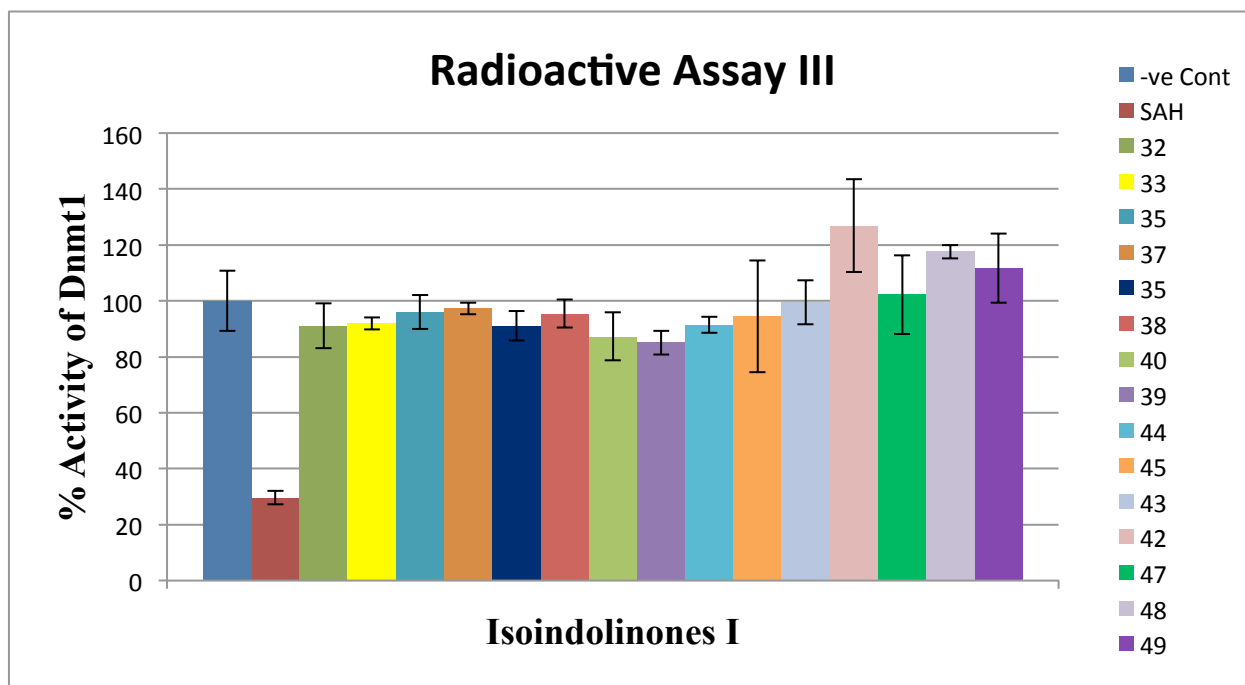
In summary, though compound **88SC** and piperazine amines showed moderate activity on Dnmt1 using radioactive assay, they failed to show corresponding effect on MCF7. Our result indicates that the targets for both **88SC** and its derivatives might differ in MCF7 cells in addition to the fact that they were only moderately active against Dnmt1. Also, the structure of our piperazine derivatives did not match similar targets as those discussed, which explains the absence of activity for our molecules since structure determines activity in this case.

#### 4.4 Isoindolinones Compounds

Another virtual screening study conducted by the Priestley group using AutoDock Vina docking software and the promiliad library of compounds uncovered isoindolinones, as a novel class of Dnmt1 inhibitors with promising biological activity. In an effort to study this class of compounds further, we synthesized analogs of isoindolinones using the Ugi type reaction followed by an IMDAF reaction. By varying the amine and the isocyanide component of the Ugi 4-component reaction, we synthesized a diverse library of isoindolinones analogs for testing (**Figure 3.5**). The synthesis was in two separate sets. The first set of indolinones had a free phenol group while the second set of compounds the phenolic OH was converted to an acetate group.

The first sets of compounds were tested for Dnmt1 inhibition using the radioactive based assay and in addition tested against MCF-7 and MCF10 cells. The results from the radioactive testing showed activity against Dnmt1 at 100  $\mu$ M (**Figure 4.5**) for any of the compounds, however compounds **32**, **38**, **44**, and **45** showed moderate inhibition of MCF7 cells with  $IC_{50}$  less than 100  $\mu$ M. (**Figure 4.6**). These results suggest that isoindolinones might be more active in the MCF-7 cells compared to enzymatic assay with Dnmt1 or this group of compounds might exhibit

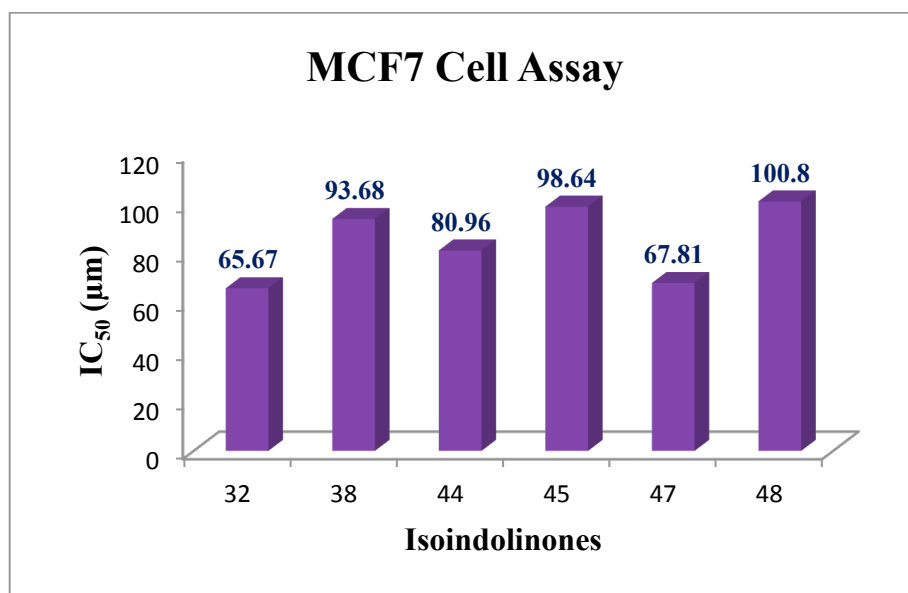
multiple targets in MCF7 cancer cells. The literature indicates that Watson et al.<sup>[344]</sup> had investigated the use of a different class of isoindolinones as anticancer agents; particularly they studied the versatility of isoindolinones as MDM2-p53 protein-protein interaction inhibitor. Briefly, in normal cells the balance between active p53 and inactive MDM2 is maintained by a negative feedback mechanism: the activity of p53 is tightly regulated by the MDM2 protein and vice versa<sup>[345]</sup>. Thus MDM2 can tightly bind to the transcriptional domain of p53 and block transcriptional activity. As such the use of isoindolinones as protein-protein modulator has the possibility to release p53 and promote apoptosis<sup>[346]</sup>, which justifies their use as anticancer agents. Thus for our study, in comparing the results from both the radioactive assay and the cell based there is a possibility that these class of isoindolinones might have two or more targets in the cell compared to our proposed selective inhibition of Dnmt1.



**Figure 4.5: Radioactive Assay III: In vitro radioactive inhibition assay of Isoindolinone analogs I on Dnmt1 activity. All compounds are assayed at 100  $\mu$ M. Data represent the average values of triplicate measurements. Error bars represent SD of the triplicate biological measurement**



Apart from its target, another interesting property of isoindolinones deduced from the cell based assay is their minimal toxicity to healthy breast cell MCF10 (**Table 4.2**). Generally, most of the isoindolinones with moderate activity against MCF7 only displayed minimum lethality against MCF10 even at concentration  $> 500 \mu\text{M}$ . This property or behavior of isoindolinones might indicate a selectivity for certain biological target(s) ( e.g MDM2 or Dnmt1) that is up regulated in cancer cells compared to non-cancerous cells, which is another indicator that isoindolinones can be considered as good leads for anticancer agents. Enhancing the binding interactions between isoindolinones with these biological target(s) would help increase the selectivity and activity of these compounds as anticancer therapeutics.



**Figure 4.6: IC<sub>50</sub> values of active isoindolinone analogs treated on breast cancer cell MCF -7**

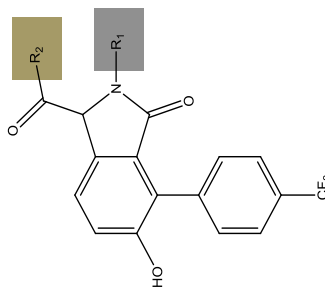
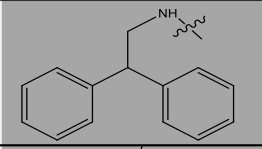
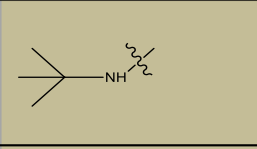
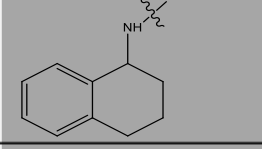
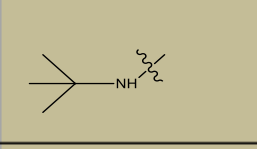
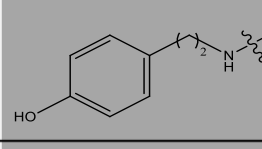
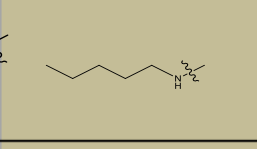
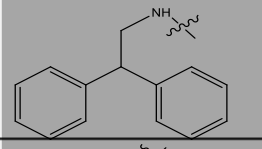
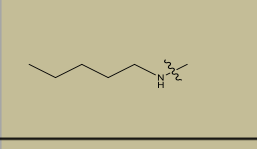
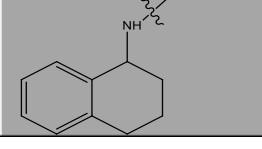
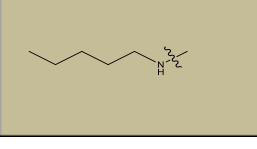
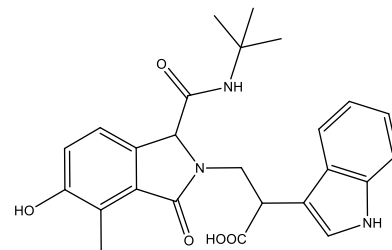
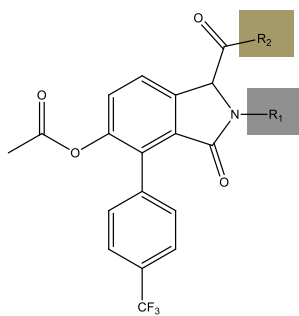
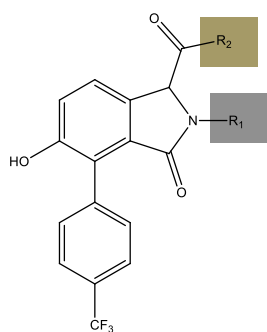


Table 4.2: IC<sub>50</sub> values of Isoindolinones obtained from MCF7 and MCF10 cell assay

Compounds	R1 Amine	R2 Isocyanide	MCF10 IC <sub>50</sub> (μM)	MCF7 IC <sub>50</sub> (μM)
32			>500	65.6.7±12.9
33			NA	No Activity
35			NA	No Activity
37			NA	No Activity
36			>500	No Activity
38			>500	93.6±0.8
40			NA	No Activity
39			NA	No Activity
44			>500	80.9.6±3.4
45			>500	98.6±0.4

Compounds	R1 Amine	R2 Isocyanide	MCF10 IC <sub>50</sub> (μM)	MCF7 IC <sub>50</sub> (μM)
43			NA	No Activity
42			NA	No Activity
47			>500	67.8±6.4
48			>500	100.8±0.3
49			NA	No Activity

The second set of isoindolinone involved the introduction of tryptamine in the R1 position (**Table 4.3**) and the modification of the phenolic OH to an acetate group. The major goal was to ascertain the importance of the OH group on isoindolinone for lethal activity against MCF7 and to investigate the possibility of using the OH group to introduce important moieties to increase binding affinity. Similarly, the choice of tryptamine for R1 was due in part to its pharmacological importance and its current use in RG108. Lastly, given the electronics of isoindolinones Tamoxifen (an antagonist of the estrogen receptor) was introduced as a positive control for the MCF7 cell assay. Our results indicate Tamoxifen is both toxic to MCF7 and MCF10 cells at our test concentration, which was 148 μM and we intend to redo the assay at much lower concentrations to obtain a selective toxicity concentration for MCF7 over MCF10.



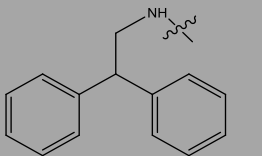
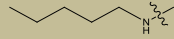
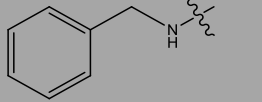
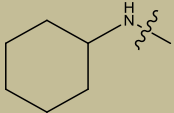
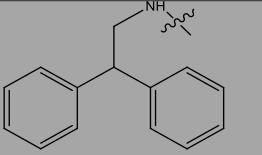
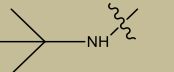
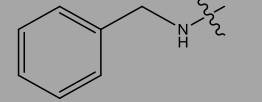
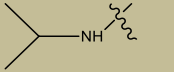
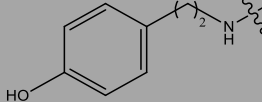
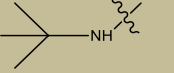
Compounds **34,41,49,50, 51,**

Compound **54 – 58**

**98**

**Table 4.3: R<sub>1</sub>, R<sub>2</sub> groups and IC<sub>50</sub> values of Isoindolinones obtained from MCF7 and MCF10 cell**

Compounds	R1 Amine	R2 Isocyanide	MCF10 IC <sub>50</sub> (μM)	MCF7 IC <sub>50</sub> (μM)
46			261.2±13.2	66
51			223.2±5.2	95
41			>500	No Activity
49			>500	No Activity
50			500	344.2±29.1
34			544.0±29	No Activity
53			>500	No Activity

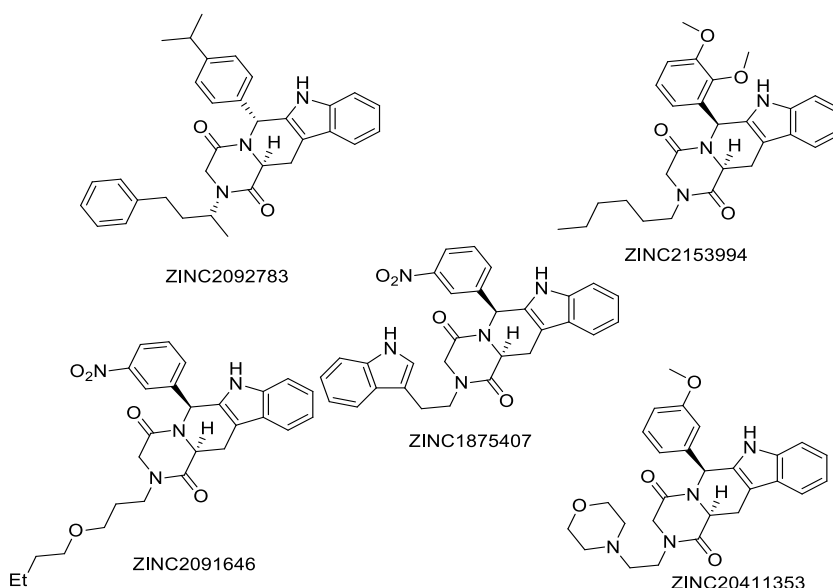
Compounds	R1 Amine	R2 Isocyanide	MCF10 IC <sub>50</sub> (μM)	MCF7 IC <sub>50</sub> (μM)
57			>500	No Activity
54			430.3±7.8	No Activity
58			>500	No Activity
55			411.8±4.9	250.07±47
56			323.2±5.0	80.0±28.2

The activity of the second set of isoindolinones against MCF7 gave similar results to the first set of compounds. Compound **46** had the lowest IC<sub>50</sub> value of 66 μM while the other active compounds were within 75 to 115 μM, an indication of similar biological targets but poor binding interactions which can be optimized if the target is known. However, one interesting observation from this class of compounds is that all isoindolinones with tryptamine at the R1 position (the amine) seem to exhibit some form of activity. It is possible that tryptamine might be a good pharmacophore to retain in the molecule or it can be used to modulate the activity of this class of isoindolinones. On the other hand, changing the phenolic OH to acetate on isoindolinone did not show a marked effect or – improvement in activity. Perhaps other

modification at this position might increase activity and selectivity but acetate group at this position did not give the desired boost expected from our investigation.

#### 4.5 Beta - Carbolines

In chapter 3 our study on pharmacophore modeling using LigandScout and CoMFA indicated that certain  $\beta$ -carboline natural product exhibit molecular descriptors similar to known Dnmt1 inhibitors (**Figure 4.7**). To test the results from this model, we designed and synthesized five simple  $\beta$ -carbolines intermediates (**Figure 4.8**) using the Pictet–Spengler condensation method and tested the products directly on MCF7 cancer cells. All  $\beta$ -carboline tested showed toxicity against MCF10 and MCF7 cells except for compound **95** (**Table 4.4**). One interesting compound in this group is **96**, which showed full lethality against MCF7 cells at 53 $\mu$ M but only had an  $IC_{50}$  value of 47.2 $\pm$ 1.3  $\mu$ M on MCF10. Further development of **96** might reveal selective toxicity concentration for MCF7 cells. In addition **96** showed a 25% inhibition of Dnmt1 at 250  $\mu$ M, this is 5% over compound 46, which was the most active compound against MCF7.



**Figure 4.7: Potential  $\beta$ -carboline natural product for development as Dnmt1 inhibitors obtain from pharmacophore modeling using LigandScout.**

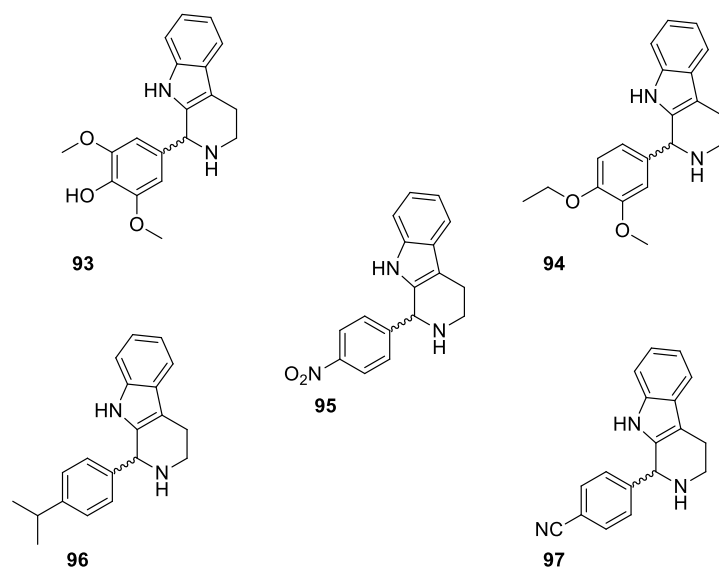
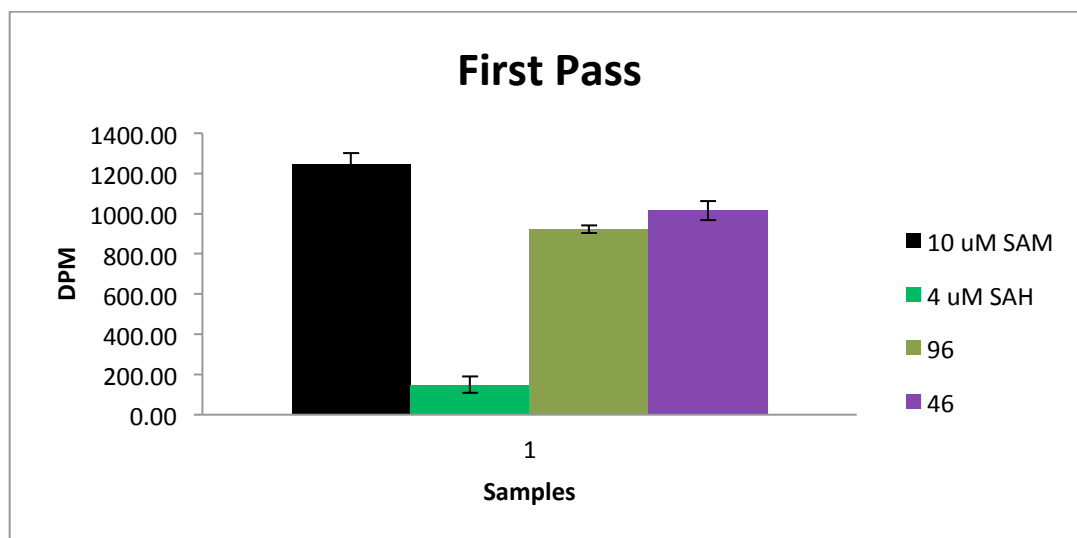


Figure 4.8: First line of simple  $\beta$ -carboline to test pharmacophore model results of LigandScout.

Table 4.4: IC<sub>50</sub> values of  $\beta$ -carbolines and Tamoxifen obtained for MCF7 and MCF10 cell assay

$\beta$ -Carboline	MCF10 IC <sub>50</sub> ( $\mu$ M)	MCF7 IC <sub>50</sub> ( $\mu$ M)
95	348.7 $\pm$ 23.3	288.3 $\pm$ 3.9
97	425 $\pm$ 11.7	417.2 $\pm$ 3.2
94	335.0 $\pm$ 54.5	306.4 $\pm$ 11.5
93	>500	>500
96	47.2 $\pm$ 1.3	53.8 $\pm$ 1.1
Tamoxifen	42.8 $\pm$ 2.1	40 $\pm$ 1.2



**Figure4.9: Radioactive Assay IV: Testing of  $\beta$ -carbolines for Dnmt1 activity at 250  $\mu$ M**

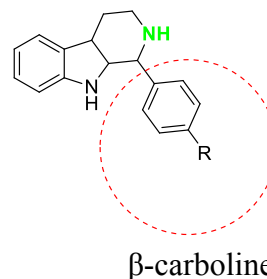
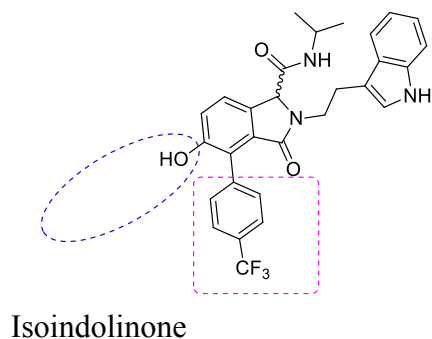
In summary, though we have only synthesized a few simple tetrahydro- $\beta$ -carbolines, the results from the cell assay suggest this class of compounds are promising anticancer agents. It is obvious that the activity of these compounds can be modulated by altering the substitution around the tetrahydro- $\beta$ -carboline moiety; in **96** cuminaldehyde gave the best results. The isopropyl group might be residing inside a hydrophobic pocket in the target and this position can be exploited or optimized with similar groups or using the Topliss scheme for aliphatic side chain substituents. Also, though we see toxicity with  $\beta$ -carbolines and the pharmacophore model suggest that these compounds share similar molecular descriptors with other Dnmt1 inhibitors; literature search indicates other targets exist. Skouta et al<sup>[347]</sup> researched  $\beta$ -carbolines as compounds with oncogene-*RAS* lethality while Xiao and colleagues<sup>[348]</sup> studied the intercalating property of  $\beta$ -carboline and Deveau et al<sup>[349]</sup> investigated the beta-carbolines as topoisomerase II inhibitors. This is a clear indication of the versatility of this class of compounds and why it will be too early to conclude our compounds only interact with Dnmt1 given the fact that beta-carbolines can be fine-tuned to induce desired effect for any one target.



### Future Work:

Given the minimal toxicity of isoindolinones it will be beneficially to harness this property further and synthesize analogues with greater potency. One approach would be to use the hydroxyl groups to added pharmacophore that can change binding of these molecules to their therapeutic targets. Also, given the structural rigidity of isoindolinones, the added pharmacophore would create more interactions with residues in the Dnmt1.

Since  $\beta$ -carboline are still in their early design and development stages, it will be important to use more steric or lipophilic aldehyde in for their synthesis in other to study the significance of these groups at that position. More so, the free amine can be optimized with different groups. It might be that the free amine is important for activity or that it can be utilize to improve activity.



## CHAPTER 5

### EXPERIMENTAL: CHEMISTRY AND BIOLOGY

#### 5.1 CHEMISTRY EXPERIMENTAL

##### General Methods

All reagents were obtained from commercial sources: Sigma, Acros, Fisher, Alfa Aesar and were used without further purification unless otherwise stated. Sensitive reaction solvents were purified by distillation over the drying agents indicated and kept under argon: anhydrous tetrahydrofuran (THF) and diethyl ether (Et<sub>2</sub>O) were distilled from Na/benzophenone. Anhydrous dimethylsulfoxide (DMSO) was distilled over calcium hydride (CaH<sub>2</sub>). Anhydrous methanol was distilled over magnesium wire and potassium permanganate. Anhydrous N,N-dimethylformamide (DMF), 1,1-dichloromethane, pyridine, benzene, triethylamine (Et<sub>3</sub>N TEA) and dioxane were purchased from commercial suppliers.

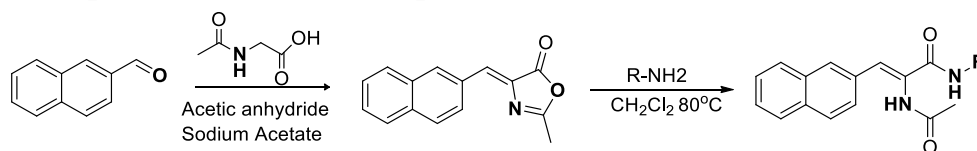
All air sensitive and non-aqueous reactions were carried out in oven-dried glassware or flamed dried and magnetic stir bar and under an atmosphere of argon unless otherwise stated. All aqueous solutions (sodium chloride, ammonium chloride, calcium carbonate) were prepared with deionized water; brine, refers to a saturated aqueous solution of non-iodized salt; water refers to deionized water. For all reactions, the drying of the organic phase was with anhydrous salts (K<sub>2</sub>CO<sub>3</sub>, MgSO<sub>4</sub>, Na<sub>2</sub>SO<sub>4</sub> or as indicated) and filtered through a filter paper (fluted or folded) or cotton plug. Also, all solvent evaporation or concentration of sample was done with a Buchi rotary evaporator under reduced pressure. Thin-layer chromatography (TLC) was performed with aluminum-backed Sigma-Aldrich Silica gel matrix with fluorescent indicator 254 nm added

for TLC plates visualization under Ultra-Violet light (254 and 365 nm). Column chromatography was performed with Siliclyce Silica gel 60 (230-400 mesh) and the solvent systems used are reported as percentages.

Compounds were characterized using Varian Unity Plus 500 MHz and Bruker Avance 400MHz. Nuclear magnetic resonance (NMR) and Mass Spectra (MS) – electrospray ionization spectroscopy (ESI) The radioactive assay was performed by Jeremy Alverson; ELISA by Brooke Martins and the MCF7 Assay was by Adrienne Sochia.

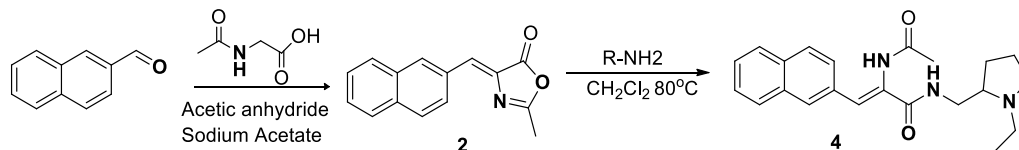
## 5.1 Chemistry Experimental

### 5.1.1 General procedure for the naphthalenyl acrylamide



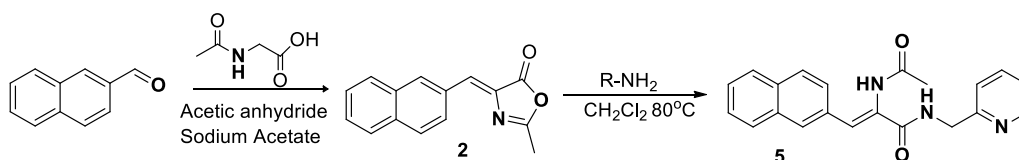
A suspension of N-acetylglycine (585 mg, 5 mmol), sodium acetate (410 mg, 5 mmol) and acetic anhydride (31mmol, 3mL) was added and stirred at room temperature for 30minutes. To this white suspension was added 2-naphthaldehyde (780 mg, 5 mmol). The resultant suspension was stirred at room temperature for 1 h and then at 60 °C for 5h. 75 mL of water was added after cooling to room temperature and stirred vigorously for 30min. The undissolved yellow solid was filtered and washed with water (3 x 3 mL) to give (Z)-2-methyl-4-(naphthalene-2-ylmethylene)oxazol-5(4H)-one (100 mg, 0.4 mmol) which was then refluxed with individual

amine in a methylene chloride for 2h. The solvent was removed under low pressure and the crude product purified by column chromatography



**4:** *(Z)-2-acetamido-N-((1-ethylpyrrolidin-2-yl)methyl)-3-(naphthalen-2-yl)acrylamide:*

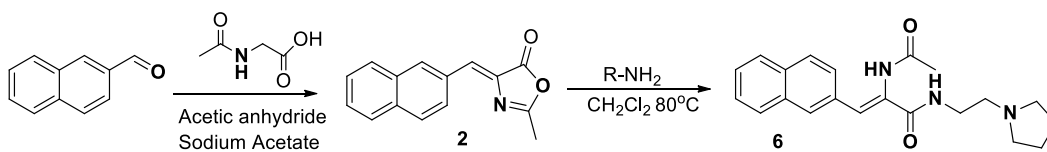
Compound **2** (*(Z)*-2-methyl-4-(naphthalene-2-ylmethylene)oxazol-5(4H)-one (100 mg, 0.4 mmol) and 2-(Aminomethyl)-1-ethylpyrrolidine (64 mg, 0.5 mmols)) were refluxed in a methylene chloride for 2h. The solvent was removed under low pressure and the crude purified by column chromatography was purified on silica gel in Methanol–EtOAc to give an off white solid (33 mg, 22.5%) <sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) δ 9.58 (br. s., 1H), 8.03 (s, 1H), 7.88 (d, *J* = 7.83 Hz, 2H), 7.84 (br. s., 1H), 7.67 (d, *J* = 8.31 Hz, 1H), 7.47 - 7.57 (m, 2H), 7.15 (s, 1H), 3.11 - 3.18 (m, 2H), 3.01 (m, *J* = 3.42 Hz, 1H), 2.96 (m, *J* = 5.38 Hz, 1H), 2.78 (dd, *J* = 7.34, 11.74 Hz, 1H), 2.21 (dd, *J* = 7.34, 11.74 Hz, 1H), 2.08 (q, *J* = 8.48 Hz, 1H), 2.01 (s, 3H), 1.69 - 1.79 (m, 1H), 1.61 (m., 2H), 1.50 - 1.57 (m, 1H), 1.03 (t, *J* = 7.34 Hz, 3H) <sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>) δ 169.8, 165.3, 133.2, 133.1, 132.3, 131.1, 129.7, 128.7, 128.3, 127.9, 127.7, 127.2, 126.9, 126.8, 62.8, 53.7, 48.5, 43.1, 28.9, 23.1, 14. MS ES+ (M+1) 366.1944



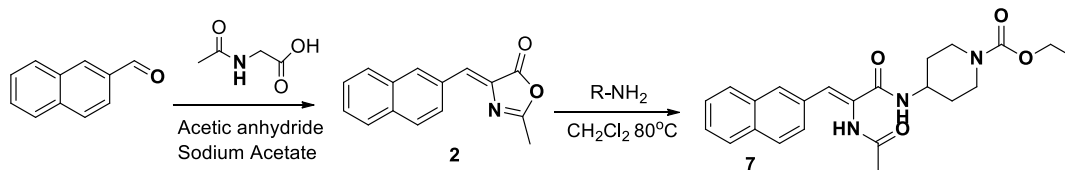
**5:** *(Z)-2-acetamido-3-(naphthalen-2-yl)-N-(pyridin-2-ylmethyl)acrylamide:* Compound **2**

*(Z)*-2-methyl-4-(naphthalene-2-ylmethylene)oxazol-5(4H)-one (100 mg, 0.4 mmol) and

2-pyridinylmethanamine (54 mg, 0.5 mmols)) were refluxed in a methylene chloride for 2h. The solvent was removed under low pressure and the crude purified by column chromatography was purified on silica gel in Methanol –EtOAc to give an off white solid (118mg, 81.1%) . <sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) d 9.68 (s, 1H), 8.71 - 8.80 (m, 1H), 8.49 (d, *J* = 4.40 Hz, 1H), 8.10 (s, 2H), 7.87 - 7.94 (m, 2H), 7.70 - 7.79 (m, 1H), 7.53 (dd, *J* = 3.18, 6.11 Hz, 2H), 7.43 (d, *J* = 7.83 Hz, 1H), 7.29 (s, 1H), 7.21 - 7.26 (m, 2H), 4.49 (d, *J* = 5.87 Hz, 2H), 2.08 (s, 3H) <sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>) d 170.3, 165.8, 159.4, 149.1, 137.0, 133.3, 133.1, 132.3, 130.9, 129.8, 128.7, 128.3, 127.9, 127.3, 126.9, 122.4, 121.2, 45.2, 45.9, 23.4 MS ES+(M+1) 346.1292

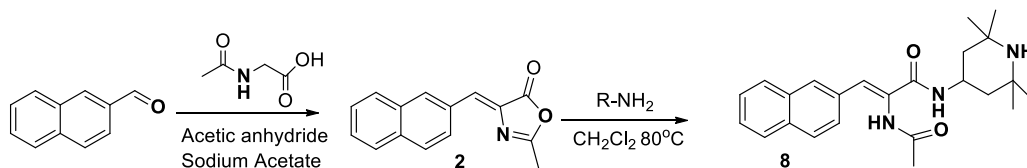


- 6:** *(Z)*-2-acetamido-3-(naphthalen-2-yl)-*N*-(2-(pyrrolidin-1-yl)ethyl)acrylamide: Compound **2** (*Z*)-2-methyl-4-(naphthalene-2-ylmethylene)oxazol-5(4H)-one (100 mg, 0.4 mmol) and 2-(1-pyrrolidinyl)ethanamine (57mg, 0.5mmols)) were refluxed in a methylene chloride for 2h. The solvent was removed under low pressure and the crude purified by column chromatography was purified on silica gel in Methanol–EtOAc to give an off white solid (61mg, 42%) <sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) d 9.50 (s, 1H), 8.03 (s, 1H), 7.96 (t, *J* = 5.87 Hz, 1H), 7.85 - 7.92 (m, 3H), 7.67 (d, *J* = 8.31 Hz, 1H), 7.45 - 7.55 (m, 2H), 7.16 (s, 1H), 3.25 (q, *J* = 6.68 Hz, 2H), 2.44 (t, 6H), 2.02 (s, 3H), 1.66 (t, 4H) <sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>) d 169.8, 165.3, 133.1, 132.4, 131.0, 129.6, 128.7, 128.2, 127.8, 127.2, 126.9, 126.8, 55.1, 54.1, 39.0, 23.6, 23.5 MS ES+(M+1) 352.1792



**7:** *(Z)-ethyl-4-(2-acetamido-3-(naphthalen-2-yl)acrylamido)piperidine-1-carboxylate:*

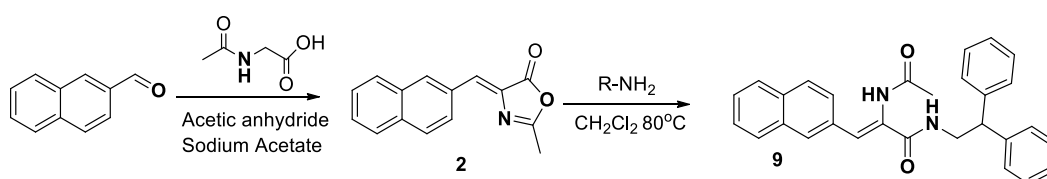
Compound **2** (*Z*)-2-methyl-4-(naphthalene-2-ylmethylene)oxazol-5(4H)-one (100 mg, 0.4 mmol) and ethyl 4-amino-1-piperidinecarboxylate (86mg, 0.5mmols)) were refluxed in a methylene chloride for 2h. The solvent was removed under low pressure and the crude purified by column chromatography was purified on silica gel in Methanol–EtOAc to give an off white solid (158mg, 97%) <sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) δ 9.49 (s, 2H), 8.03 (s, 2H), 7.94 (d, J = 7.83 Hz, 2H), 7.86 - 7.91 (m, 4H), 7.67 (d, J = 8.31 Hz, 2H), 7.51 (dd, J = 3.18, 6.11 Hz, 3H), 7.03 (s, 2H), 4.02 (q, J = 7.34 Hz, 4H), 1.94 - 2.06 (m, 5H), 1.59 - 1.75 (m, 4H), 1.32 - 1.49 (m, 4H), 1.06 - 1.23 (m, 6H) <sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>) δ 169.7, 165.0, 155.1, 133.3, 133.0, 132.5, 131.4, 129.5, 128.6, 128.2, 127.9, 126.9, 61.1, 46.8, 43.0, 31.5, 23.4, 15.1



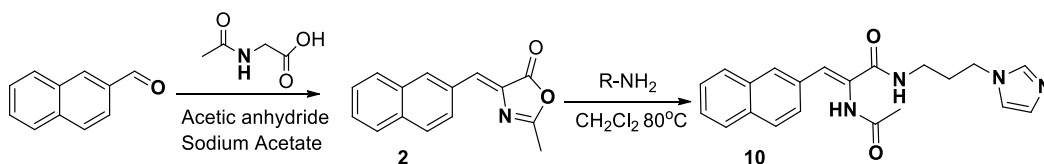
**8:** *(Z)-2-acetamido-3-(naphthalen-2-yl)-N-(2,2,6,6-tetramethylpiperidin-4-yl)acrylamide:*

Compound **2** (*Z*)-2-methyl-4-(naphthalene-2-ylmethylene)oxazol-5(4H)-one (100 mg, 0.4 mmol) and 2,2,6,6-tetramethyl-4-piperidinamine (94 mg, 0.5 mmols)) were refluxed in a methylene chloride for 2h. The solvent was removed under low pressure and the crude purified by column chromatography was purified on silica gel in Methanol–EtOAc to give an off white solid (128 mg, 81%) <sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) δ 9.47 (s, 1H),

8.02 (s, 1H), 7.88 (d, J = 8.31 Hz, 4H), 7.82 (d, J = 7.83 Hz, 1H), 7.67 (d, J = 8.31 Hz, 1H), 7.51 (dd, J = 3.18, 6.11 Hz, 3H), 7.01 (s, 1H), 4.13 (d, J = 8.31 Hz, 1H), 2.01 (s, 4H), 1.75 - 1.77 (m, 1H), 1.56 - 1.64 (m, 3H), 1.14 (s, 9H), 1.11 (br. s., 2H), 1.09 (br. s., 1H), 1.02 (s, 9H) <sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>) d 169.7, 164.8, 133.3, 133.0, 132.6, 131.7, 129.5, 128.6, 128.2, 127.9, 126.9, 126.5, 109.9, 50.9, 44.5, 35.1, 29.1, 23.4 MS ES+(M+1) 394.4022

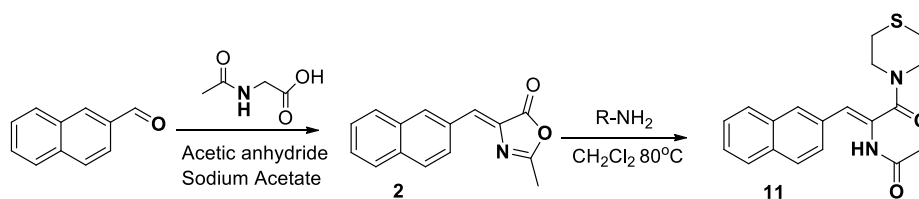


- 9:** *(Z)-2-acetamido-N-(2,2-diphenylethyl)-3-(naphthalen-2-yl)acrylamide*: Compound **2** (*Z*)-2-methyl-4-(naphthalene-2-ylmethylene)oxazol-5(4H)-one (100 mg, 0.4 mmol) and 2,2-diphenylethylamine (99 mg, 0.5 mmols) were refluxed in a methylene chloride for 2h. The solvent was removed under low pressure and the crude purified by column chromatography was purified on silica gel in Methanol–EtOAc to give an off white solid (105 mg, 61%). <sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) d 9.47 (br. s., 1H), 8.02 (br. s., 1H), 7.96 (br. s., 1H), 7.87 (d, J = 8.31 Hz, 2H), 7.61 (d, J = 8.80 Hz, 1H), 7.46 - 7.54 (m, 2H), 7.26 - 7.35 (m, 6H), 7.20 (d, J = 5.38 Hz, 4H), 6.90 (s, 1H), 4.37 (t, J = 7.34 Hz, 2H), 3.81 (t, J = 6.11 Hz, 2H), 1.98 (s, 3H) <sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>) d 169.8, 165.7, 143.4, 133.1, 132.3, 131.2, 129.5, 128.9, 128.7, 128.5, 128.3, 127.9, 127.1, 126.8, 50.3, 44.3, 23.3 MS ES+(M+1) 435.2386



**10:** *(Z)-N-(3-(1H-imidazol-1-yl)propyl)-2-acetamido-3-(naphthalen-2-yl)acrylamide:*

Compound **2** (*(Z)*-2-methyl-4-(naphthalene-2-ylmethylene)oxazol-5(4H)-one (100 mg, 0.4 mmol) and 3-(1H-imidazol-1-yl)-1-propanamine (63 mg, 0.5 mmols)) were refluxed in a methylene chloride for 2h. The solvent was removed under low pressure and the crude purified by column chromatography was purified on silica gel in Methanol–EtOAc to give an off yellow solid (140mg, 97%) <sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) δ 9.62 (s, 1H), 8.19 (br. s., 1H), 8.05 (s, 1H), 7.89 (d, J = 7.83 Hz, 3H), 7.69 (d, J = 8.80 Hz, 1H), 7.63 (s, 1H), 7.51 (dd, J = 3.18, 6.11 Hz, 2H), 7.19 (s, 1H), 7.07 (s, 1H), 6.87 (s, 1H), 3.99 (t, J = 6.85 Hz, 2H), 3.06 - 3.20 (m, 2H), 2.04 (s, 3H), 1.77 - 1.93 (m, 2H) <sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>) δ 170.1, 165.8, 137.8, 133.3, 133.0, 132.4, 131.5, 129.6, 128.7, 128.7, 128.3, 127.9, 127.1, 126.9, 126.7, 119.8, 43.9, 36.6, 31.3, 23.4 MS ES+(M+1) 363.1745

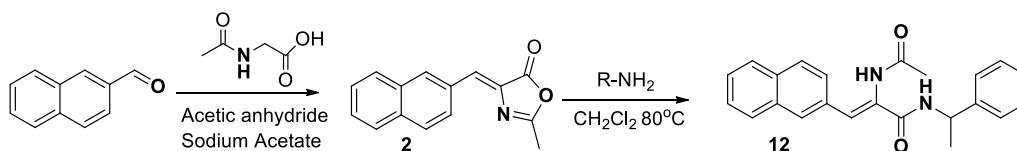


**11:** *(Z)-N-(1-(naphthalen-2-yl)-3-oxo-3-thiomorpholinoprop-1-en-2-yl)acetamide:*

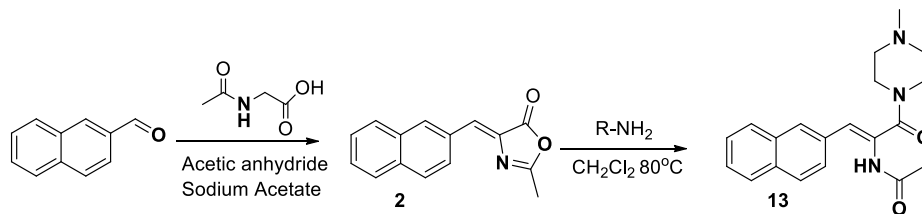
Compound **2** (*(Z)*-2-methyl-4-(naphthalene-2-ylmethylene)oxazol-5(4H)-one (100 mg, 0.4 mmol) and thiomorpholine (51 mg, 0.5 mmols)) were refluxed in a methylene chloride for 2h. The solvent was removed under low pressure and the crude purified by column chromatography was purified on silica gel in Methanol–EtOAc to give an off



white solid (93.8 mg, 97%) <sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) δ 9.95 (s, 1H), 8.04 (s, 1H), 7.83 - 7.93 (m, 3H), 7.59 - 7.70 (m, 1H), 7.45 - 7.56 (m, 2H), 6.24 (s, 1H), 3.65 - 3.91 (m, 4H), 2.53 - 2.76 (m, 4H), 1.93 - 2.08 (m, 3H) <sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>) δ 169.6, 167.9, 133.4, 132.7, 132.3, 130.7, 128.7, 128.5, 128.3, 127.9, 127.1, 126.8, 119.4, 49.8, 26.3, 22.8 MS ES+(M+1) 341.1450

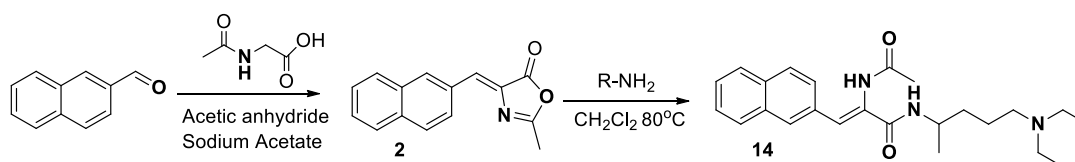


- 12:** *(Z)*-2-acetamido-3-(naphthalen-2-yl)-*N*-(1-phenylethyl)acrylamide: Compound **2** (*Z*)-2-methyl-4-(naphthalene-2-ylmethylene)oxazol-5(4H)-one (100 mg, 0.4 mmol) and 1-phenylethanamine (60 mg, 0.5 mmols) were refluxed in a methylene chloride for 2h. The solvent was removed under low pressure and the crude purified by column chromatography was purified on silica gel in Methanol–EtOAc to give an yellowish solid (128 mg, 97%) 70% yield <sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) δ 9.51 (d, J = 1.96 Hz, 1H), 8.46 (d, J = 5.38 Hz, 1H), 8.05 (br. s., 1H), 7.85 - 7.95 (m, 3H), 7.70 (d, J = 7.34 Hz, 1H), 7.51 (td, J = 3.36, 6.48 Hz, 2H), 7.37 - 7.44 (m, 2H), 7.18 - 7.35 (m, 3H), 7.10 (d, J = 2.45 Hz, 1H), 5.06 (dd, J = 4.89, 6.85 Hz, 1H), 1.88 - 2.09 (m, 3H), 1.37 - 1.51 (m, 3H) <sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>) δ 169.8, 165.0, 145.1, 133.3, 133.0, 132.5, 131.3, 129.6, 128.7, 128.5, 128.2, 127.9, 127.1, 127.0, 126.9, 126.6, 48.7, 23.4, 22.6 MS ES+(M+1) 359.1882



**13:** *(Z)-N-(3-(4-methylpiperazin-1-yl)-1-(naphthalen-2-yl)-3-oxoprop-1-en-2-yl)acetamide:*

Compound **2** (*(Z)*-2-methyl-4-(naphthalene-2-ylmethylene)oxazol-5(4H)-one (100 mg, 0.4 mmol) and 1-methylpiperazine (50 mg, 0.5 mmols)) were refluxed in a methylene chloride for 2h. The solvent was removed under low pressure and the crude purified by column chromatography was purified on silica gel in Methanol–EtOAc to give an off white solid (107.6mg, 80%). <sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) δ 9.90 (br. s., 1H), 8.03 (br. s., 1H), 7.83 - 7.95 (m, 3H), 7.66 (d, J = 8.80 Hz, 1H), 7.42 - 7.58 (m, 2H), 6.21 (d, J = 1.96 Hz, 1H), 3.37 - 3.71 (m, 4H), 2.36 (br. s., 3H), 2.20 (d, J = 2.93 Hz, 4H), 1.88 - 2.03 (m, 3H) <sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>) δ 169.4, 167.5, 133.4, 132.7, 132.3, 130.9, 128.7, 128.5, 128.3, 127.9, 127.1, 126.8, 119.7, 54.4, 51.7, 46.1, 43.2, 22.8 MS ES+(M+1) 338.1873

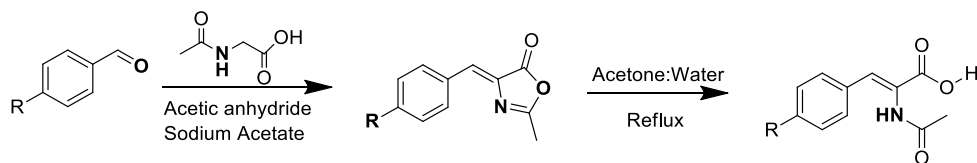


**14:** *(Z)-2-acetamido-N-(5-(diethylamino)pentan-2-yl)-3-(naphthalen-2-yl)acrylamide:*

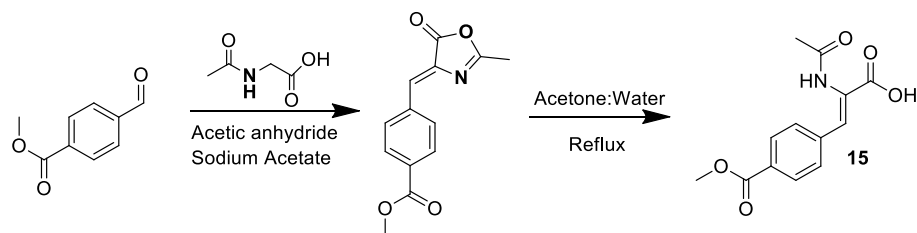
Compound **2** (*(Z)*-2-methyl-4-(naphthalene-2-ylmethylene)oxazol-5(4H)-one (100 mg, 0.4 mmol) and N<sup>1</sup>,N<sup>1</sup>-diethylpentane-1, 4-diamine (80mg, 0.51mmols)) were refluxed in a methylene chloride for 2h. The solvent was removed under low pressure and the crude

purified by column chromatography was purified on silica gel in Methanol–EtOAc to give an off white solid (152 mg, 92%)  $^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$  9.54 (br. s., 1H), 8.03 (br. s., 1H), 7.82 - 7.92 (m, 3H), 7.69 (d,  $J$  = 8.31 Hz, 1H), 7.51 (dd,  $J$  = 3.18, 6.11 Hz, 2H), 7.07 (s, 1H), 3.90 (m, 1H), 3.18 (s, 2H), 2.42 (q,  $J$  = 6.85 Hz, 4H), 2.04 (s, 3H), 1.31 - 1.53 (m, 4H), 1.10 (d,  $J$  = 6.36 Hz, 3H), 0.93 (t,  $J$  = 7.09 Hz, 6H)  $^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ )  $\delta$  169.7, 164.9, 133.3, 133.0, 132.6, 131.7, 129.5, 128.6, 127.9, 127.0, 126.9, 126.8, 126.6, 52.6, 46.7, 45.2, 34.4, 24.0, 23.4, 21.2, 12.2 MS ES $^+$ :( M+1) 396.2309

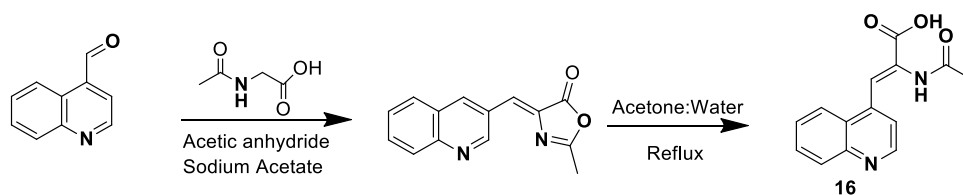
### 5.1.2 General Procedure for $\alpha$ -(N-Acylamino)- $\alpha,\beta$ -unsaturated acids



A suspension of N-acetylglycine (585 mg, 5 mmol), sodium acetate (410 mg, 5 mmol) and acetic anhydride (3 mL, 31 mmol) was added and stirred at room temperature for 30min to give a white suspension. phenyl-4-carboxaldehyde (911 mg, 5 mmol) was added and the resultant suspension was stirred at room temperature for 1hr and then at 60  $^{\circ}\text{C}$  for 5h. After 5h and disappearance of the aldehyde as noted by TLC, 75mL of water was added after cooling to room temperature and stirred vigorously for 30min. The undissolved yellow solid was filtered and washed with water (3 x 3mL) and then refluxed in a solution of Acetone:water (3:1) 12 mL for 2h. Solvent was removed under low pressure and purified using column chromatography 20% MeOH/EtOAc to give the corresponding  $\alpha$ -(N-Acylamino)- $\alpha,\beta$ -unsaturated acids.

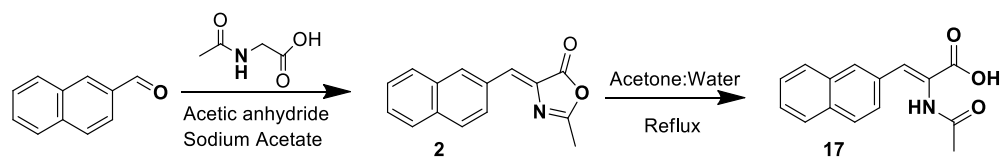


- 15:** *(Z)*-2-acetamido-3-(4-(methoxycarbonyl)phenyl)acrylic acid: A suspension of N-acetylglycine (468mg, 4mmol), sodium acetate (328 mg, 4 mmol) and acetic anhydride (2450 mg, 24 mmol,) was added and stirred at room temperature for 30min to give a white methyl 4-formylbenzoate (755 mg, 5 mmol) was added and the resultant suspension was stirred for 1hr at room temperature and then at 60° C for 12 h. After the disappearance of the aldehyde as noted by TLC, 75mL of water was added after cooling to room temperature and stirred vigorously for 30min. The undissolved yellow solid was filtered and washed with water (3 x 3mL) and then refluxed in a solution of Acetone:water (3:1) 12 mL for 2h. Solvent was removed under low pressure and purified using column chromatography 20% MeOH-EtOAc to give a yellow solid. (138mg, 13.2% - over two steps) <sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) δ 12.81 (br. s., 1H), 9.55 (br. s., 1H), 8.19 (br. s., 1H), 7.73 - 7.98 (m, 2H), 7.53 (br. s., 1H), 7.27 (br. s., 1H), 3.85 (br. s., 3H), 1.98 (br. s., 3H) <sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>) δ 169.5, 166.6, 166.4, 134.8, 130.4, 130.3, 130.1, 129.8, 129.5, 128.7, 52.7, 22.9



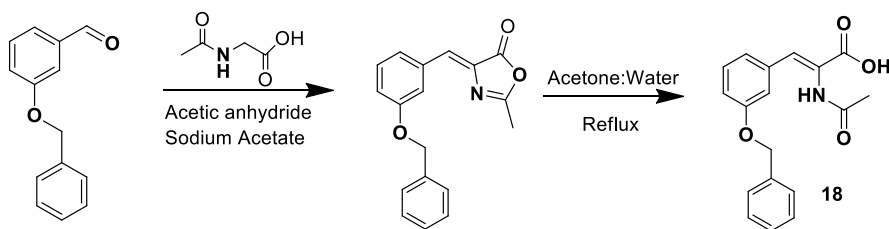
- 16:** *(Z)*-2-acetamido-3-(quinolin-4-yl)acrylic acid: A suspension of N-acetylglycine (585 mg, 5mmol), sodium acetate (410 mg, 5 mmol) and acetic anhydride (31 mmol, 3mL)

was added and stirred at room temperature for 30 minutes to give a white suspension. 4-quinolinecarbaldehyde (785 mg, 5 mmol) was added and the resultant suspension was stirred for 1 hr at room temperature and then at 60° C for 5h. After 5h and disappearance of the aldehyde as noted by TLC, 75mL of water was added after cooling to room temperature and stirred vigorously for 30min. The undissolved yellow solid was filtered and washed with water (3 x 3mL) and then refluxed in a solution of Acetone:water (3:1) 12 mL for 2h. Solvent was removed under low pressure and purified using column chromatography 20% MeOH-EtOAc to give yellow solid. (487 mg, 38% - over two steps). <sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) δ 13.07 (br. s., 1H), 9.54 (br. s., 1H), 8.89 (d, *J* = 4.40 Hz, 1H), 8.04 (d, *J* = 8.31 Hz, 1H), 7.97 (d, *J* = 8.31 Hz, 1H), 7.77 (t, *J* = 7.34 Hz, 1H), 7.61 (t, *J* = 7.34 Hz, 1H), 7.52 (br. s., 1H), 7.50 (d, *J* = 3.91 Hz, 1H), 1.83 (s, 3H). <sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>) δ 169.6, 166.2, 150.6, 148.4, 140.1, 132.7, 130.0, 127.3, 126.1, 124.9, 120.6, 109.2, 60.3, 22.8 MS ES+: (M+1)257.0090



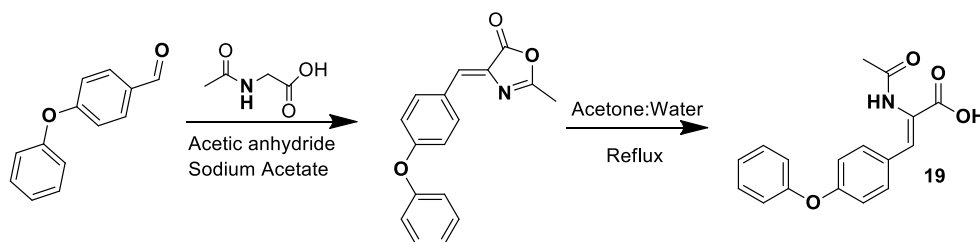
- 17:** *(Z)*-2-acetamido-3-(naphthalen-1-yl)acrylic acid: A suspension of N-acetylglycine (585 mg, 5 mmol), sodium acetate (410 mg, 5 mmol) and acetic anhydride (31 mmol, 3mL) was added and stirred at room temperature for 30 minutes to give a white suspension. 2-naphthaldehyde (785 mg, 5 mmol) was added and the resultant suspension was stirred for 1 hr at room temperature and then at 60° C for 5h. After 5h and disappearance of the aldehyde as noted by TLC, 75mL of water was added after cooling to room temperature and stirred vigorously for 30min. The undissolved yellow solid was filtered and washed

with water (3 x 3mL) and then refluxed in a solution of Acetone:water (3:1) 12 mL for 2h. Solvent was removed under low pressure and purified using column chromatography 20% MeOH-EtOAc to give yellow solid. (487 mg, 38% - over two steps). <sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) Shift 12.72 (br. s., 1H), 9.57 (s, 1H), 8.11 (s, 1H), 7.90 (d, J = 7.83 Hz, 3H), 7.74 (d, J = 8.80 Hz, 1H), 7.50 - 7.56 (m, 2H), 7.35 (s, 1H), 2.01 (s, 3H) <sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>) δ 169.7, 166.9, 133.3, 133.2, 131.9, 130.4, 128.8, 128.3, 128.0, 127.9, 127.5, 127.0, 126.9, 23.1

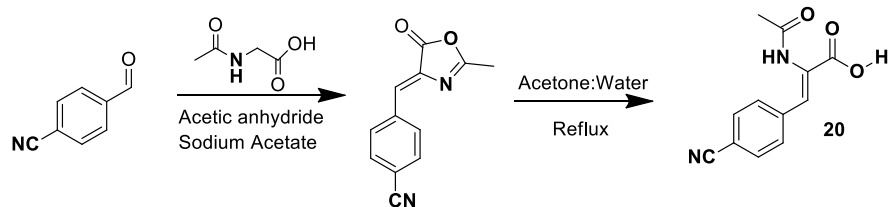


- 18:** *(Z)-2-acetamido-3-(3-(benzyloxy)phenyl)acrylic acid*: A suspension of N-acetylglycine (585 mg, 5 mmol), sodium acetate (410 mg, 5 mmol) and acetic anhydride (31 mmol, 3mL) was added and stirred at room temperature for 30minutes to give a white suspension 3-(benzyloxy)benzaldehyde (785 mg, 5 mmol) was added and the resultant suspension was stirred for 1hr at room temperature and then at 60° C for 12 h. After the disappearance of the aldehyde as noted by TLC, 75mL of water was added after cooling to room temperature and stirred vigorously for 30min. The undissolved yellow solid was filtered and washed with water (3 x 3mL) and then refluxed in a solution of Acetone:water (3:1) 12 mL for 2h. Solvent was removed under low pressure and purified using column chromatography 20% MeOH-EtOAc to give yellow solid. (338.4mg, 25% - over two steps). <sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) d 12.67 (br. s., 1H), 9.49 (s, 1H), 7.41 - 7.45 (m, 2H), 7.38 (t, J = 7.34 Hz, 2H), 7.24 - 7.34 (m, 3H), 7.19 (d, J = 7.83 Hz, 1H),

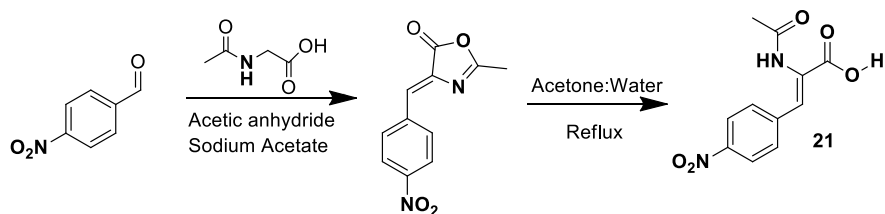
7.15 (s, 1H), 7.00 (d,  $J = 6.85$  Hz, 1H), 5.10 (s, 2H), 1.95 (s, 3H)  $^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ )  $\delta$  169.6, 166.8, 158.7, 137.4, 135.4, 131.1, 130.1, 128.9, 128.3, 128.1, 122.7, 116.2, 69.6, 22.9



- 19:** *(Z)*-2-acetamido-3-(4-phenoxyphenyl)acrylic acid: A suspension of N-acetylglycine (585 mg, 5 mmol), sodium acetate (410 mg, 5 mmol) and acetic anhydride (31 mmol, 3mL) was added and stirred at room temperature for 30minutes to give a white suspension 4-phenoxybenzaldehyde (991 mg, 5 mmol) was added and the resultant suspension was stirred for 1hr at room temperature and then at 60° C for 12 h. After the disappearance of the aldehyde as noted by TLC, 75mL of water was added after cooling to room temperature and stirred vigorously for 30min. The undissolved yellow solid was filtered and washed with water (3 x 3mL) and then refluxed in a solution of Acetone:water (3:1) 12 mL for 2h. Solvent was removed under low pressure and purified using column chromatography 20% MeOH-EtOAc to give a white solid. (460mg, 31% - over two steps).  $^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$  12.71 (br. s., 1H), 9.41 (s, 1H), 7.39 (td,  $J = 4.10, 7.95$  Hz, 3H), 7.31 (d,  $J = 7.34$  Hz, 1H), 7.23 (br. s., 1H), 7.12 - 7.19 (m, 2H), 6.96 - 7.06 (m, 3H), 1.80 (s, 3H)  $^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ )  $\delta$  169.4, 166.9, 157.3, 156.6, 136.0, 130.6, 128.4, 124.2, 119.7, 119.4, 118.9, 22. MS ES+: (M+1) 298.0468



**20:** *(Z)*-2-acetamido-3-(4-cyanophenyl)acrylic acid: A suspension of N-acetylglycine (585 mg, 5 mmol), sodium acetate (410 mg, 5 mmol) and acetic anhydride (31 mmol, 3mL) was added and stirred at room temperature for 30minutes to give a white suspension 4-cyanobenzaldehyde (755 mg, 5 mmol) was added and the resultant suspension was stirred for 1hr at room temperature and then at 60° C for 12 h. After the disappearance of the aldehyde as noted by TLC, 75mL of water was added after cooling to room temperature and stirred vigorously for 30min. The undissolved yellow solid was filtered and washed with water (3 x 3mL) and then refluxed in a solution of Acetone:water (3:1) 12 mL for 2h. Solvent was removed under low pressure and purified using column chromatography 20% MeOH-EtOAc to give a yellow solid. (174 mg, 15.2% - over two steps). <sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) δ 12.96 (br. s., 1H), 9.65 (br. s., 1H), 7.84 (d, *J* = 7.83 Hz, 2H), 7.75 (d, *J* = 7.83 Hz, 2H), 7.20 (br. s., 1H), 1.99 (br. s., 3H) <sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>) δ 169.7, 167.0, 133.4, 133.2, 132.0, 130.5, 128.9, 128.4, 128.1, 128.0, 127.6, 127.1, 126.9, 40.5, 40.4, 40.2, 40.0, 39.8, 39.7, 39.5, 38.8, 23.1, 0.6

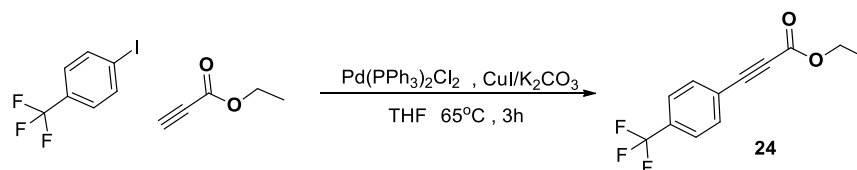


**21:** *(Z)*-2-acetamido-3-(4-nitrophenyl)acrylic acid: A suspension of N-acetylglycine (585 mg, 5 mmol), sodium acetate (410 mg, 5 mmol) and acetic anhydride (31 mmol, 3mL) was added and stirred at room temperature for 30minutes to give a white suspension 4-



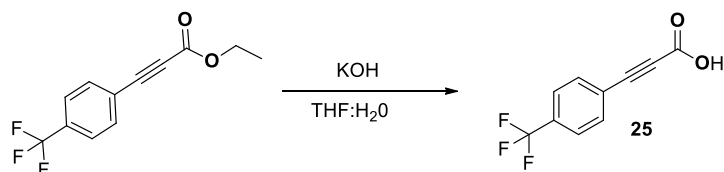
nitrobenzaldehyde (755 mg, 5 mmol) was added and the resultant suspension was stirred for 1hr at room temperature and then at 60° C for 12 h. After the disappearance of the aldehyde as noted by TLC, 75mL of water was added after cooling to room temperature and stirred vigorously for 30min. The undissolved yellow solid was filtered and washed with water (3 x 3mL) and then refluxed in a solution of Acetone:water (3:1) 12 mL for 2h. Solvent was removed under low pressure and purified using column chromatography 20% MeOH-EtOAc to give a yellow solid. (788mg, 63% - over two steps). <sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) δ 13.00 (br. s., 1H), 9.69 (br. s., 1H), 8.21 (d, *J* = 8.32 Hz, 2H), 7.80 (d, *J* = 8.31 Hz, 2H), 7.21 (br. s., 1H), 2.48 (br. s., 1H), 1.98 (br. s., 3H) <sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>) d 169.1, 166.0, 146.8, 140.9, 130.5, 130.4, 127.1, 123.5, 22.

### 5.1.3 Isoindolinone Synthesis



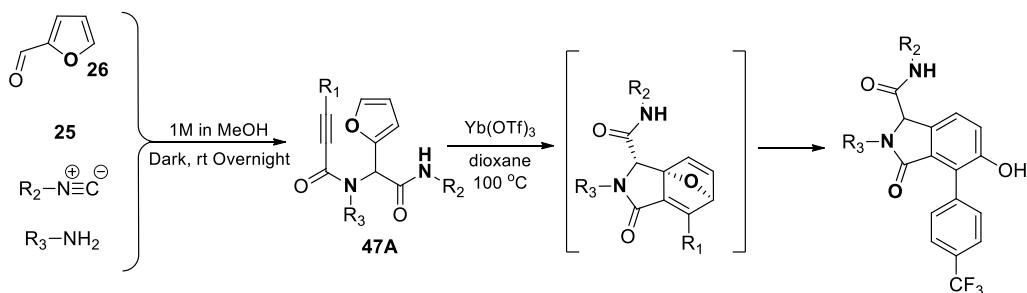
**24:** *ethyl 3-(4-(trifluoromethyl)phenyl)propiolate:* To a solution of 1-iodo-4-(trifluoromethyl)benzene (554 mg, 2 mmol) and ethyl propiolate (784.8 mg, 4 mmol) in THF (3 mL) were added bis(triphenylphosphine)palladium(II)chloride (28 mg, 0.02 mmol), copper iodide (15 mg, 0.04 mmol) and potassium carbonate (552 mg, 2 mmol). The mixture was stirred at 60° C for 3 h then cooled and filtered through celite (7 g) rinsing with diethyl ether. Subsequently, the solvent was removed by rotavap. Product was purified using column chromatography (10% Ethyl Acetate/Hexanes). Yield (82%). <sup>1</sup>H NMR (500 MHz,

CHLOROFORM-d  $\delta$  8.12 - 8.57 (m, 2H), 7.61 - 7.95 (m, 2H), 4.31 (br. s., 2H), 1.35 (br. s., 3H)  
 $^{13}\text{C}$  NMR (126 MHz, CHLOROFORM-d)  $\delta$  153.5, 133.1, 132.3, 132.0, 125.5, 124.6, 123.5,  
 83.8, 82.3, 62.3, 14.0



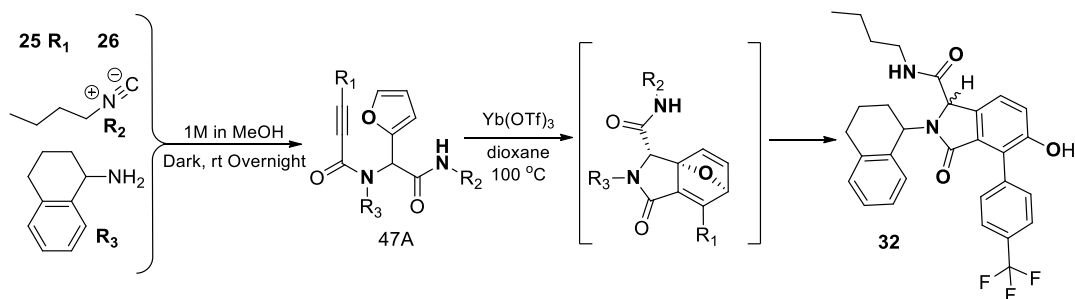
**25:** 3-(4-(trifluoromethyl)phenyl)prop-2-ynoic acid: Potassium Hydroxide (70 mg, 1.5 mmol) was added in small portions to ethyl 3-(4-(trifluoromethyl)phenyl) propionate (203 mg, 0.84 mmol) in THF:H<sub>2</sub>O (5 mL) mix solvent at 0°C and warmed up to room temperature. The solution was stirred at room temperature for 6h. The solvent was removed by rotavap and the crude diluted with water (2 mL) and pH adjusted to 3. It was then filtered and dried to give a white powder (142 mg, 74%).  $^1\text{H}$  NMR (500 MHz, DMSO-d<sub>6</sub>)  $\delta$  8.37 (br. s., 9H), 8.00 (br. s., 10H)  $^{13}\text{C}$  NMR (126 MHz, CHLOROFORM-d)  $\delta$  153.7, 148.5, 133.7, 126.2, 123.7, 83.8, 83.1

### General preparation of isoindolines



Step I: One molar solution of furfural (41  $\mu$ L 0.5 mmol), benzylamine (54  $\mu$ L, 0.5 mmol), 3-(4-(trifluoromethyl)phenyl) propionic acid (107 mg, 0.5 mmol) and n-butyliisocyanide (52  $\mu$ L, 0.5 mmol) was made up in MeOH (2 mL) and stirred in the dark under argon for 12 h. The Ugi product was purified by flash chromatography (40% EtoAc: Hexane) to give a mixture of three products that was used in the next step.

Step II: 0.02 M of the Ugi product (20 mg, 0.04 mmol) was prepared in Dioxane (2 mL), ytterbium(III) trifluoromethanesulfonate (5 mg, 0.008 mmol) was added and heated in a sealed tube at 100  $^{\circ}$ C for 12 h. The solvent was removed by rotavap and product purified by flash chromatography (70% EtoAc : Hexane) to the corresponding isoindolinone as a white solid.

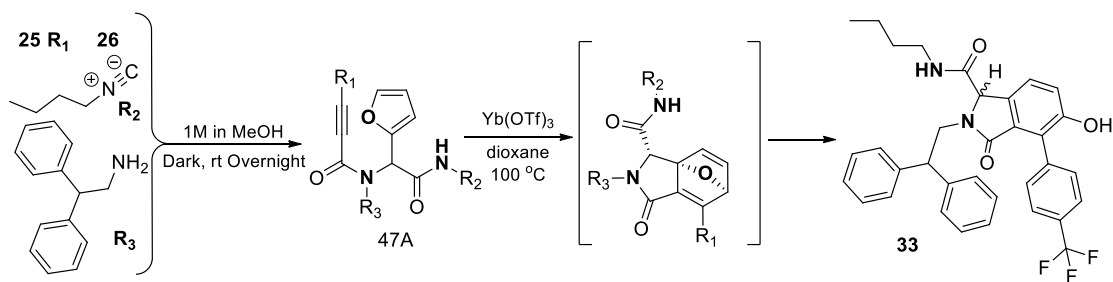


**32:** *N*-(*tert*-butyl)-5-hydroxy-3-oxo-2-(1,2,3,4-tetrahydronaphthalen-1-yl)-4-(4-

(trifluoromethyl)phenyl)isoindoline-1-carboxamide: **Step I:** One molar solution of

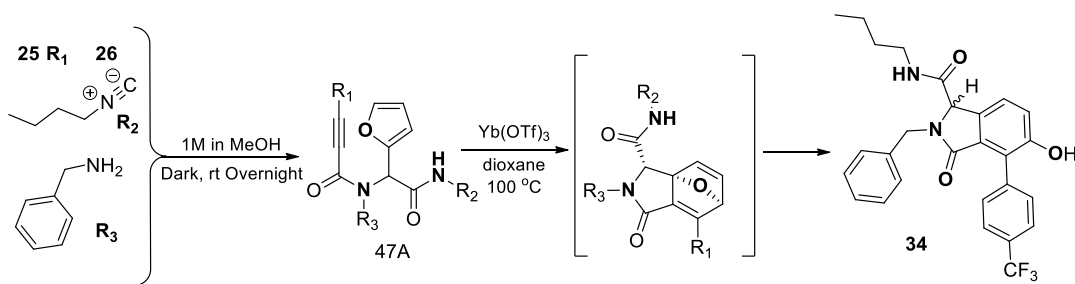
freshly distilled furfural (82  $\mu$ L 1 mmol), 1,2,3,4-tetrahydro-1-naphthalenamine (146  $\mu$ L, 1 mmol), 3-(4-(trifluoromethyl)phenyl) propionic acid (214 mg, 1 mmol) and n-butyliisocyanide (104  $\mu$ L, 1 mmol) was made up in MeOH (4 mL) and stirred in the dark under argon for 12 h. The Ugi product was purified by flash chromatography first with

CH<sub>2</sub>Cl<sub>2</sub> and then gradient elution with EtoAc-Hexane (10%, 20% & 40%)) to give a mixture of three Ugi products that was used in the next step. **Step II:** 0.02 M of the Ugi product (350 mg, 0.67 mmol) was prepared in Dioxane 14 mL), 20 mol% ytterbium(III) trifluoromethanesulfonate (103 mg, 0.167 mmol) was added and the reaction heated at 100° C in a sealed tube for 12 h. The solvent was removed by rotavap and product purified by flash chromatography twice (70% & 35% EtoAc-Hexane) to give a white solid (308 mg, 59% over two steps). <sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) Shift 9.95 (s, 1H), 8.49 (br. s., 1H), 7.73 (d, J = 7.83 Hz, 1H), 7.59 (d, J = 7.83 Hz, 1H), 7.16 - 7.26 (m, 1H), 7.12 (br. s., 1H), 7.04 - 7.09 (m, 1H), 6.70 (d, J = 7.34 Hz, 1H), 5.32 (d, J = 4.89 Hz, 1H), 4.71 (s, 1H), 3.34 (s, 1H), 3.07 (d, J = 3.91 Hz, 1H), 2.72 (br. s., 1H), 1.87 - 2.07 (m, 1H), 1.63 - 1.84 (m, 1H), 1.33 - 1.43 (m, 1H), 1.22 - 1.31 (m, 1H), 0.55 - 1.06 (m, 3H) <sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>) Shift 169.0, 168.7, 155.5, 138.3, 138.1, 135.7, 134.3, 132.1, 129.7, 129.4, 127.3, 126.6, 124.5, 124.1, 122.9, 119.9, 60.5, 51.3, 31.3, 29.4, 28.2, 22.3, 19.9, 14.0



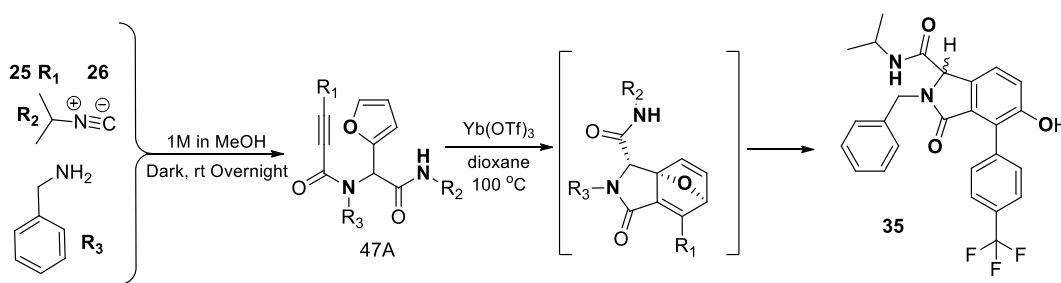
**33:** *N*-butyl-2-(2,2-diphenylethyl)-5-hydroxy-3-oxo-4-(4-(trifluoromethyl)phenyl)isoindoline-1-carboxamide: **Step I:** One molar solution of freshly distilled furfural (41  $\mu$ L 0.5 mmol), 2,2-diphenylethylamine (98 mg, 0.5 mmol), 3-(4-(trifluoromethyl)phenyl) propionic acid (107 mg, 0.5 mmol) and n-butylisocyanide (52  $\mu$ L, 0.5 mmol) was made up in MeOH (2

mL) and stirred in the dark under argon for 12 h. The Ugi product was purified by flash chromatography first with CH<sub>2</sub>Cl<sub>2</sub> and then gradient elution with EtoAc-Hexane (10%, 20% & 40%) to give a mixture of three Ugi products that was used in the next step. **Step II:** 0.02M of the Ugi product (160 mg, 0.27 mmol) was prepared in Dioxane (6 mL). 20mol% ytterbium(III) trifluoromethanesulfonate (43 mg, 0.067 mmol) was added and the reaction heated at 100° C in a sealed tube for 12 h. The solvent was removed by rotavap and product purified by flash chromatography twice (70% & 35% EtoAc-Hexane) to give a white solid (130 mg, 11.4% over two steps). <sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) Shift 9.87 (br. s., 1H), 8.54 (br. s., 1H), 7.67 (d, J = 7.34 Hz, 2H), 7.42 (d, J = 7.34 Hz, 2H), 7.14 - 7.31 (m, 8H), 7.10 (dd, J = 7.58, 19.81 Hz, 2H), 4.66 (br. s., 1H), 4.41 - 4.48 (m, 1H), 4.29 - 4.36 (m, 1H), 3.33 (dd, J = 6.85, 14.18 Hz, 1H), 3.07 - 3.15 (m, 2H), 2.46 (br. s., 2H), 1.36 - 1.45 (m, 2H), 1.21 - 1.33 (m, 2H), 0.43 - 0.99 (m, 3H) <sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>) Shift 167.7, 167.2, 155.4, 142.9, 141.9, 138.2, 133.2, 132.1, 131.9, 129.2, 129.1, 128.2, 128.1, 127.9, 127.8, 127.6, 127.1, 126.1, 124.2, 124.1, 123.9, 123.1, 119.5, 110.0, 62.0, 49.4, 45.4, 39.0, 31.5, 19.9, 14.0



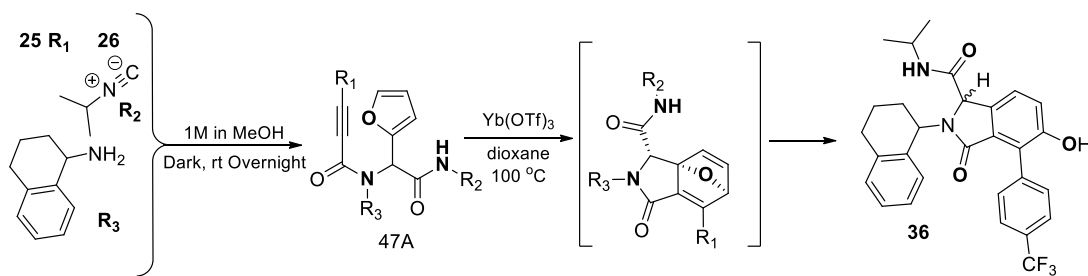
- 34:** *2-benzyl-N-butyl-4-(4-(trifluoromethyl)phenyl)-5-hydroxy-3-oxoisindoline-1-carboxamide*: **Step I:** One molar solution of freshly distilled furfural (62  $\mu$ L 0.75 mmol), benzylamine (81  $\mu$ L, 0.75 mmol), 3-(4-(trifluoromethyl)phenyl) propionic acid (160 mg,

0.75 mmol) and n-butyliisocyanide (78  $\mu$ L, 0.75 mmol) was made up in MeOH (2 mL) and stirred in the dark under argon for 12 h. The Ugi product was purified by flash chromatography first with  $\text{CH}_2\text{Cl}_2$  and then gradient elution with EtoAc-Hexane (10%, 20% & 40%) to give a mixture of three Ugi products that was used in the next step. **Step II:** 0.02M of the Ugi product (270 mg, 0.27 mmol) was prepared in Dioxane (6 mL). 20 mol% ytterbium(III) trifluoromethanesulfonate (86 mg, 0.13 mmol) was added and the reaction heated at 100 $^\circ$  C in a sealed tube for 12 h. The solvent was removed by rotavap and product purified by flash chromatography twice (70% & 35% EtoAc-Hexane) to give a white solid (130mg, 34.4% over two steps)  $^1\text{H}$  NMR (400 MHz,  $\text{DMSO-d}_6$ )  $\delta$  9.95 (s, 1H), 8.44 - 8.64 (m, 1H), 7.75 (d,  $J = 8.03$  Hz, 2H), 7.58 (d,  $J = 8.03$  Hz, 2H), 7.28 - 7.42 (m, 4H), 7.20 (d,  $J = 8.03$  Hz, 2H), 5.08 (d,  $J = 15.06$  Hz, 1H), 4.79 (s, 1H), 3.91 (d,  $J = 15.31$  Hz, 1H), 3.35 (s, 1H), 3.00 - 3.25 (m, 2H), 1.28 - 1.48 (m, 4H), 0.91 (t,  $J = 7.15$  Hz, 3H)  $^{13}\text{C}$  NMR (101 MHz,  $\text{DMSO-d}_6$ )  $\delta$  167.2, 166.8, 155.1, 137.9, 136.9, 132.9, 131.6, 128.8, 128.7, 127.9, 127.5, 127.4, 124.1, 123.7, 123.2, 122.7, 119.3, 61.0, 44.3, 38.5, 31.0, 19.5, 13.6



**35:** *2-benzyl-5-hydroxy-N-isopropyl-3-oxo-4-(4-(trifluoromethyl)phenyl)isoindoline-1-carboxamide*: **Step I:** One molar solution of freshly distilled furfural (41  $\mu$ L 0.5 mmol), benzylamine (54  $\mu$ L, 0.5 mmol), 3-(4-(trifluoromethyl)phenyl) propionic acid (107 mg,

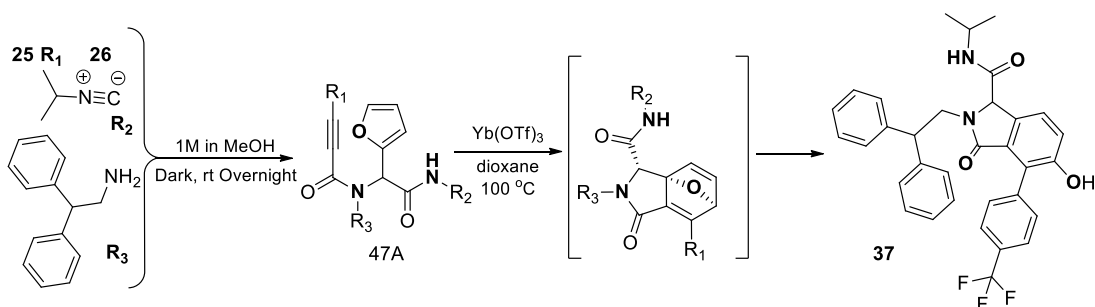
0.5 mmol) and isopropyl isocyanide (47  $\mu$ L, 0.5 mmol) was made up in MeOH (2 mL) and stirred in the dark under argon for 12 h. The Ugi product was purified by flash chromatography first with CH<sub>2</sub>Cl<sub>2</sub> and then gradient elution with EtoAc-Hexane (10%, 20% & 40%) to give a mixture of three Ugi products that was used in the next step. **Step II:** 0.02 M of the Ugi product (143 mg, 0.31 mmol) was prepared in Dioxane (6 mL). 20mol% ytterbium(III) trifluoromethanesulfonate (47 mg, 0.07 mmol) was added and the reaction heated at 100° C in a sealed tube for 12 h. The solvent was removed by rotavap and product purified by flash chromatography twice (70% & 35% EtoAc-Hexane) to give a white solid (62 mg, 7.0% over two steps). <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>)  $\delta$  9.95 (s, 1H), 9.80 - 10.10 (m, 1H), 8.47 (d,  $J$  = 7.53 Hz, 1H), 8.31 - 8.63 (m, 1H), 7.75 (d,  $J$  = 8.03 Hz, 1H), 7.58 (d,  $J$  = 8.03 Hz, 1H), 7.27 - 7.40 (m, 4H), 7.21 (d,  $J$  = 7.78 Hz, 3H), 5.07 (d,  $J$  = 15.06 Hz, 1H), 4.77 (s, 1H), 3.87 - 3.97 (m, 2H), 1.12 (dd,  $J$  = 6.78, 8.78 Hz, 6H) <sup>13</sup>C NMR (101 MHz, DMSO-d<sub>6</sub>)  $\delta$  167.2, 165.8, 155.0, 137.9, 136.9, 132.9, 131.6, 128.8, 128.7, 128.0, 127.5, 127.2, 125.9, 124.1, 123.7, 122.6, 119.3, 60.9, 44.3, 40.9, 22.2



**36:** *5-hydroxy-N-isopropyl-3-oxo-2-(1,2,3,4-tetrahydronaphthalen-1-yl)-4-(4-*

*(trifluoromethyl)phenyl)isoindoline-1-carboxamide:* **Step I:** One molar solution of freshly distilled furfural (62  $\mu$ L 0.75mmol), 1,2,3,4-tetrahydro-1-naphthalenamine (110

$\mu\text{L}$ , 0.75 mmol), 3-(4-(trifluoromethyl)phenyl) propionic acid (160 mg, 0.75 mmol) and isopropyl isocyanide (70  $\mu\text{L}$ , 0.75 mmol) was made up in MeOH (3 mL) and stirred in the dark under argon for 12fh. The Ugi product was purified by flash chromatography first with  $\text{CH}_2\text{Cl}_2$  and then gradient elution with EtoAc-Hexane (10%, 20% &40%) to give a mixture of three Ugi products that was used in the next step. **Step II:** 0.02M of the Ugi product (96 mg, 0.18 mmol) was prepared in Dioxane (4 mL). 20mol% ytterbium(III) trifluoromethanesulfonate (30 mg, 0.04 mmol) was added and the reaction heated at  $100^\circ\text{C}$  in a sealed tube for 12 h. The solvent was removed by rotavap and product purified by flash chromatography twice (70% & 35% EtoAc-Hexane) to give a white solid (63 mg, 9.3% over two steps)  $^1\text{H}$  NMR (500 MHz,  $\text{DMSO-d}_6$ )  $\delta$  9.93 (s, 1H), 8.41 (d,  $J = 7.34$  Hz, 1H), 7.71 (d,  $J = 7.83$  Hz, 3H), 7.57 (d,  $J = 7.83$  Hz, 1H), 7.14 - 7.27 (m, 1H), 7.02 - 7.12 (m, 2H), 6.71 (d,  $J = 7.34$  Hz, 1H), 5.28 (d,  $J = 5.38$  Hz, 1H), 4.68 (s, 1H), 3.80 (d,  $J = 6.36$  Hz, 1H), 3.34 - 3.34 (m, 1H), 3.33 (s, 1H), 2.71 (br. s., 3H), 2.48 (br. s., 1H), 1.61 - 2.03 (m, 4H), 1.01 - 1.11 (m, 6H)  $^{13}\text{C}$  NMR (126 MHz,  $\text{DMSO-d}_6$ )  $\delta$  168.9, 167.7, 155.5, 138.3, 138.1, 135.8, 134.3, 132.1, 129.7, 129.4, 127.3, 126.6, 126.5, 124.5, 124.1, 122.8, 119.9, 60.4, 51.2, 41.3, 29.4, 28.2, 22.6, 22.2

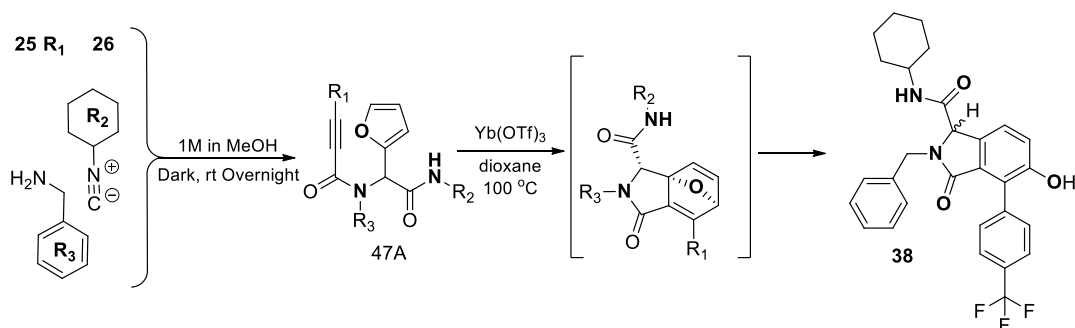


**37:** 2-(2,2-diphenylethyl)-5-hydroxy-N-isopropyl-3-oxo-4-(4-

(trifluoromethyl)phenyl)isoindoline-1-carboxamide: **Step I:** One molar solution of

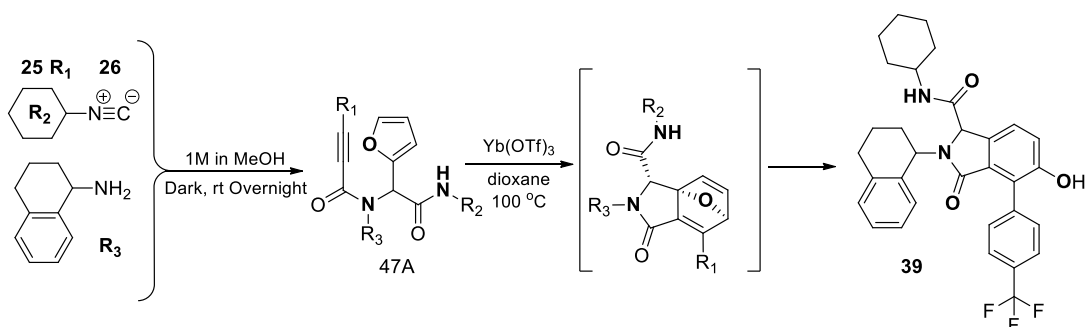


freshly distilled furfural (41  $\mu$ L 0.5 mmol), 2,2-diphenylethylamine (98 mg, 0.5 mmol), 3-(4-(trifluoromethyl)phenyl) propionic acid (107 mg, 0.5 mmol) and isopropyl isocyanide (47  $\mu$ L, 0.5 mmol) was made up in MeOH (2 mL) and stirred in the dark under argon for 12 h. The Ugi product was purified by flash chromatography first with  $\text{CH}_2\text{Cl}_2$  and then gradient elution with EtoAc-Hexane (10%, 20% & 40%) to give a mixture of three Ugi products that was used in the next step. **Step II:** 0.02M of the Ugi product (210 mg, 0.37 mmol) was prepared in Dioxane (8 mL). 20mol% ytterbium(III) trifluoromethanesulfonate (58 mg, 0.09 mmol) was added and the reaction heated at 100° C in a sealed tube for 12 h. The solvent was removed by rotavap and product purified by flash chromatography twice (70% & 35% EtoAc-Hexane) to give a white solid (190 mg, 25.5% over two steps).  $^1\text{H}$  NMR (500 MHz,  $\text{DMSO-d}_6$ ) Shift 9.90 (br. s., 1H), 8.55 (d,  $J = 7.34$  Hz, 1H), 7.70 (d,  $J = 7.34$  Hz, 1H), 7.44 (d,  $J = 6.85$  Hz, 2H), 7.09 - 7.34 (m, 10H), 4.68 (br. s., 1H), 4.45 (d,  $J = 13.21$  Hz, 1H), 4.35 (d,  $J = 6.36$  Hz, 1H), 3.90 (br. s., 1H), 3.34 (br. s., 2H), 1.08 - 1.18 (m, 6H)  $^{13}\text{C}$  NMR (126 MHz,  $\text{DMSO-d}_6$ )  $\delta$  167.7, 166.2, 155.4, 142.9, 142.0, 138.2, 133.3, 132.0, 129.3, 128.9, 127.8, 124.2, 61.9, 49.5, 45.4, 41.5, 22.7



**38:** 2-benzyl-N-cyclohexyl-5-hydroxy-3-oxo-4-(4-(trifluoromethyl)phenyl)isoindoline-1-carboxamide carboxamide: **Step I:** One molar solution of freshly distilled furfural (41

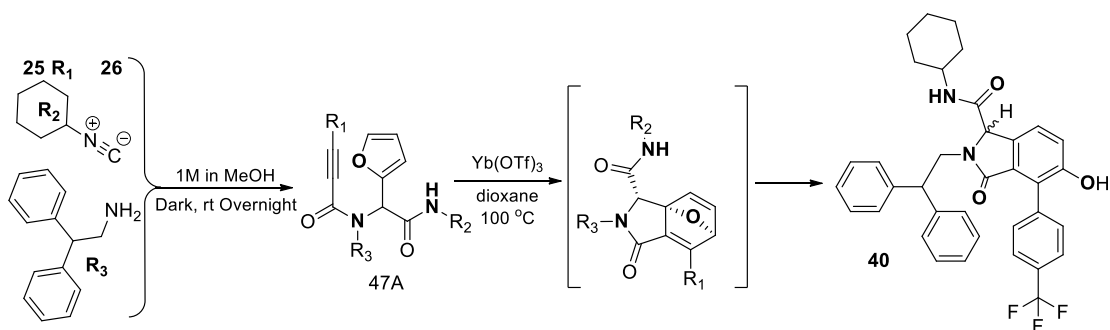
$\mu\text{L}$  0.5 mmol), benzylamine (54  $\mu\text{L}$ , 0.5 mmol), 3-(4-(trifluoromethyl)phenyl) propionic acid (107 mg, 0.5 mmol) isocyanocyclohexane (62 $\mu\text{L}$ , 0.5 mmol) was made up in MeOH (2 mL) and stirred in the dark under argon for 12 h. The Ugi product was purified by flash chromatography first with  $\text{CH}_2\text{Cl}_2$  and then gradient elution with EtoAc-Hexane (10%, 20% &40%) to give a mixture of three Ugi products that was used in the next step. **Step II:** 0.02 M of the Ugi product (121 mg, 0.23 mmol) was prepared in Dioxane (5 mL). 20mol% ytterbium(III) trifluoromethanesulfonate (36 mg, 0.07 mmol) was added and the reaction heated at  $100^\circ\text{C}$  in a sealed tube for 12 h. The solvent was removed by rotavap and product purified by flash chromatography twice (70% & 35% EtoAc-Hexane) to give a white solid (90 mg, 8.9% over two steps)  $^1\text{H}$  NMR (400 MHz, DMSO- $d_6$ )  $\delta$  9.95 (s, 1H), 8.47 (d,  $J = 8.03$  Hz, 1H), 7.75 (d,  $J = 8.28$  Hz, 2H), 7.58 (d,  $J = 8.28$  Hz, 2H), 7.31 - 7.39 (m, 4H), 7.20 (d,  $J = 8.03$  Hz, 2H), 5.07 (d,  $J = 15.31$  Hz, 1H), 4.79 (s, 1H), 3.91 (d,  $J = 15.06$  Hz, 1H), 3.60 (d,  $J = 7.78$  Hz, 1H), 3.36 (s, 1H), 1.53 - 1.87 (m, 5H), 1.13 - 1.38 (m, 5H)  $^{13}\text{C}$  NMR (101 MHz, DMSO- $d_6$ )  $\delta$  167.2, 165.8, 155.0, 138.3, 136.9, 133.0, 131.6, 128.8, 128.7, 128.0, 127.4, 125.9, 124.1, 123.7, 122.7, 119.3, 60.8, 47.9, 44.3, 32.3, 32.3, 25.1, 24.4



**39:** *N*-cyclohexyl-5-hydroxy-3-oxo-2-(1,2,3,4-tetrahydronaphthalen-1-yl)-4-(4-

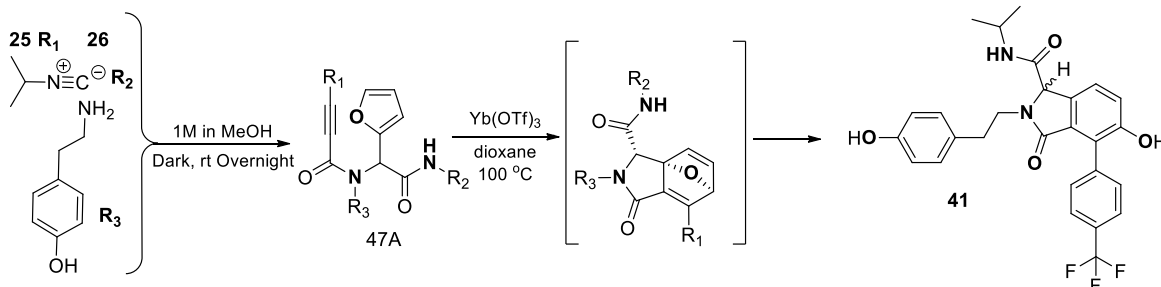
(trifluoromethyl)phenyl)isoindoline-1-carboxamide **P1:** **Step I:** One molar solution of

freshly distilled furfural (82  $\mu$ L 1 mmol), 1,2,3,4-tetrahydro-1-naphthalenamine (146  $\mu$ L, 1 mmol), 3-(4-(trifluoromethyl)phenyl) propionic acid (214 mg, 1 mmol) and isocyanocyclohexane (124  $\mu$ L, 1 mmol) was made up in MeOH (4 mL) and stirred in the dark under argon for 12 h. The Ugi product was purified by flash chromatography first with  $\text{CH}_2\text{Cl}_2$  and then gradient elution with EtoAc-Hexane (10%, 20% & 40%) to give a mixture of three Ugi products that was used in the next step. **Step II:** 0.02 M of the Ugi product (236 mg, 0.43 mmol) was prepared in Dioxane 9mL), 20 mol% ytterbium(III) trifluoromethanesulfonate (66 mg, 0.107 mmol) was added and the reaction heated at 100<sup>o</sup> C in a sealed tube for 12 h. The solvent was removed by rotavap and product purified by flash chromatography twice (70% & 35% EtoAc-Hexane) to give a white solid (114 mg, 20% over two steps). <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>)  $\delta$  9.95 (s, 1H), 8.45 (d,  $J$  = 7.78 Hz, 1H), 7.75 (d,  $J$  = 8.03 Hz, 2H), 7.61 (d,  $J$  = 8.03 Hz, 2H), 7.17 - 7.29 (m, 2H), 7.05 - 7.16 (m, 3H), 6.75 (d,  $J$  = 7.53 Hz, 1H), 5.33 (dd,  $J$  = 5.90, 9.66 Hz, 1H), 4.76 (s, 1H), 3.54 (d,  $J$  = 7.28 Hz, 1H), 3.36 (s, 1H), 2.74 (br. s., 2H), 1.50 - 2.09 (m, 8H), 1.08 - 1.35 (m, 5H) <sup>13</sup>C NMR (101 MHz, DMSO-d<sub>6</sub>)  $\delta$  168.5, 167.3, 155.0, 137.7, 135.4, 133.9, 131.7, 129.3, 129.0, 126.8, 126.1, 126.1, 124.0, 123.7, 122.3, 119.4, 59.9, 50.7, 47.9, 32.1, 29.0, 27.7, 25.1, 24.4, 21.8



**40:** *N*-cyclohexyl-2-(2,2-diphenylethyl)-5-hydroxy-3-oxo-4-(4-(trifluoromethyl)phenyl)isoindoline-1-carboxamide: **Step I:** One molar solution of freshly distilled furfural (41  $\mu$ L 0.5 mmol), 2,2-diphenylethylamine (98 mg, 0.5 mmol), 3-(4-(trifluoromethyl)phenyl) propionic acid (107 mg, 0.5 mmol) isocyanocyclohexane (62  $\mu$ L, 0.5 mmol) was made up in MeOH (2 mL) and stirred in the dark under argon for 12 h. The Ugi product was purified by flash chromatography first with  $\text{CH}_2\text{Cl}_2$  and then gradient elution with EtoAc-Hexane (10%, 20% & 40%) to give a mixture of three Ugi products that was used in the next step. **Step II:** 0.02M of the Ugi product (227 mg, 0.37 mmol) was prepared in Dioxane (8 mL). 20mol% ytterbium(III) trifluoromethanesulfonate (58 mg, 0.07 mmol) was added and the reaction heated at 100° C in a sealed tube for 12 h. The solvent was removed by rotavap and product purified by flash chromatography twice (70% & 35% EtoAc-Hexane) to give a white solid (156 mg, 13.0% over two steps)  $^1\text{H}$  NMR (500 MHz,  $\text{DMSO-d}_6$ )  $\delta$  9.90 (br. s., 1H), 8.55 (d,  $J = 6.85$  Hz, 1H), 7.70 (d,  $J = 6.85$  Hz, 2H), 7.45 (d,  $J = 6.85$  Hz, 2H), 7.18 - 7.32 (m, 7H), 7.09 - 7.17 (m, 2H), 4.73 (br. s., 1H), 4.46 (d,  $J = 12.72$  Hz, 2H), 4.34 (t,  $J = 7.09$  Hz, 2H), 3.59 (br. s., 2H), 3.32 - 3.39 (m, 2H), 1.47 - 1.87 (m, 4H), 1.11 - 1.32 (m, 4H)  $^{13}\text{C}$  NMR (126 MHz,  $\text{DMSO-d}_6$ )  $\delta$  167.7, 166.2, 155.4, 142.9, 142.0, 138.2,

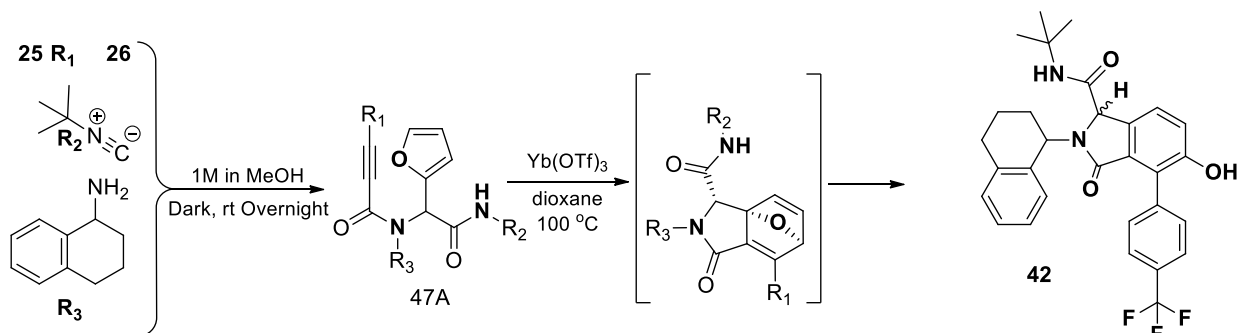
133.3, 132.1, 131.9, 129.3, 128.2, 128.1, 127.6, 124.2, 119.7, 119.6, 61.9, 61.9, 49.6, 48.5, 45.4, 32.9, 25.6, 24.9



**41:** *5-hydroxy-2-(4-hydroxyphenethyl)-N-isopropyl-3-oxo-4-(4-*

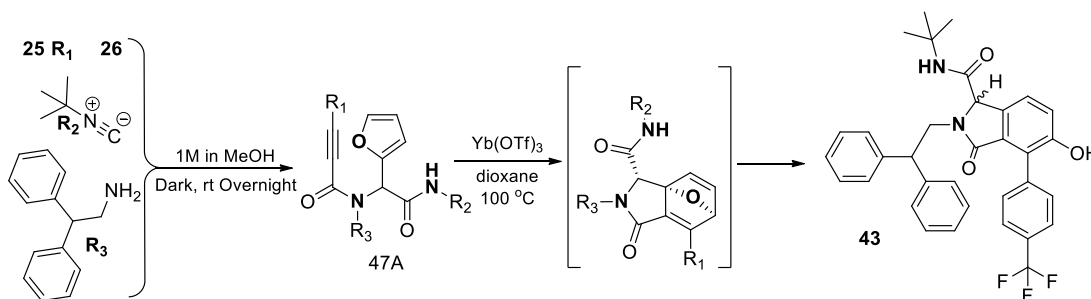
*(trifluoromethyl)phenyl)isoindoline-1-carboxamide: Step I:* One molar solution of freshly distilled furfural (41  $\mu$ L 0.5 mmol), tyramine (68 mg, 0.5 mmol), 3-(4-(trifluoromethyl)phenyl) propionic acid (107 mg, 0.5 mmol) and isopropyl isocyanide (47  $\mu$ L, 0.5 mmol) was made up in MeOH (2 mL) and stirred in the dark under argon for 12 h. The Ugi product was purified by flash chromatography first with  $\text{CH}_2\text{Cl}_2$  and then gradient elution with EtoAc-Hexane (10%, 20% & 40%) to give a mixture of three Ugi products that was used in the next step. **Step II:** 0.02M of the Ugi product (132 mg, 0.26 mmol) was prepared in Dioxane (6 mL). 20mol% ytterbium(III) trifluoromethanesulfonate (41 mg, 0.06 mmol) was added and the reaction heated at 100 $^\circ$  C in a sealed tube for 12 h. The solvent was removed by rotavap and product purified by flash chromatography twice (70% & 35% EtoAc-Hexane) to give a white solid (98 mg, 14.7% over two steps)  $^1\text{H}$  NMR (500 MHz,  $\text{DMSO-d}_6$ )  $\delta$  9.88 (br. s., 1H), 9.20 (br. s., 1H), 8.47 (d,  $J = 6.85$  Hz, 2H), 7.69 (d,  $J = 6.85$  Hz, 2H), 7.48 (d,  $J = 7.34$  Hz, 2H), 7.33 (d,  $J = 7.83$  Hz, 2H), 7.15 (d,  $J = 7.34$  Hz, 1H), 6.94 (d,  $J = 6.85$  Hz, 1H), 6.64 (d,  $J = 6.85$  Hz, 1H), 4.91 (br. s., 1H), 3.80 - 3.90 (m, 2H), 3.34 (s, 1H), 2.96 - 3.04 (m, 2H), 2.57 - 2.82 (m, 2H), 1.10 (t,  $J = 5.14$  Hz, 6H)  $^{13}\text{C}$  NMR (126 MHz,  $\text{DMSO-d}_6$ )  $\delta$  167.6,

166.6, 156.2, 155.3, 138.4, 133.4, 132.0, 129.7, 129.7, 129.3, 127.8, 127.6, 126.1, 124.2, 124.0, 123.1, 119.6, 115.7, 115.6, 62.2, 62.0, 41.4, 33.1, 22.8



**42:** *N*-(*tert*-butyl)-5-hydroxy-3-oxo-2-(1,2,3,4-tetrahydronaphthalen-1-yl)-4-(4-(trifluoromethyl)phenyl)isoindoline-1-carboxamide: **Step I:** One molar solution of freshly distilled furfural (82  $\mu$ L 1 mmol), 1,2,3,4-tetrahydro-1-naphthalenamine (146  $\mu$ L, 1 mmol), 3-(4-(trifluoromethyl)phenyl) propionic acid (214 mg, 1 mmol) and *tert*-Butyl isocyanide (112  $\mu$ L, 1 mmol) was made up in MeOH (4 mL) and stirred in the dark under argon for 12 h. The Ugi product was purified by flash chromatography first with  $\text{CH}_2\text{Cl}_2$  and then gradient elution with EtoAc-Hexane (10%, 20% & 40%) to give a mixture of three Ugi products that was used in the next step. **Step II:** 0.02 M of the Ugi product (154 mg, 0.43 mmol) was prepared in Dioxane 7 mL), 20 mol% ytterbium(III) trifluoromethanesulfonate (45 mg, 0.107 mmol) was added and the reaction heated at 100° C in a sealed tube for 12 h. The solvent was removed by rotavap and product purified by flash chromatography twice (70% & 35% EtoAc-Hexane) to give a white solid (76 mg, 14% over two steps)<sup>1</sup>H NMR (400 MHz, DMSO- $d_6$ )  $\delta$  9.92 (br. s., 1H), 8.23 (s, 1H), 7.75 (d,  $J$  = 8.28 Hz, 2H), 7.61 (d,  $J$  = 7.78 Hz, 2H), 7.28 - 7.32 (m, 1H), 7.19 - 7.24 (m, 1H), 7.07 - 7.17 (m, 2H), 6.76 (d,  $J$  = 7.78 Hz, 1H), 5.32 (br. s., 1H), 4.82 (s, 1H), 3.36 (br. s., 1H), 2.76 (br. s., 2H), 1.64 - 2.19 (m, 4H), 1.20 - 1.42 (m, 9H) <sup>13</sup>C

NMR (101 MHz, DMSO- $d_6$ )  $\delta$  168.5, 167.4, 154.9, 138.0, 137.7, 135.6, 134.2, 131.6, 129.2, 129.0, 126.8, 126.1, 124.0, 123.7, 123.7, 122.3, 119.4, 60.2, 50.7, 50.5, 29.0, 28.2, 27.6, 21.8

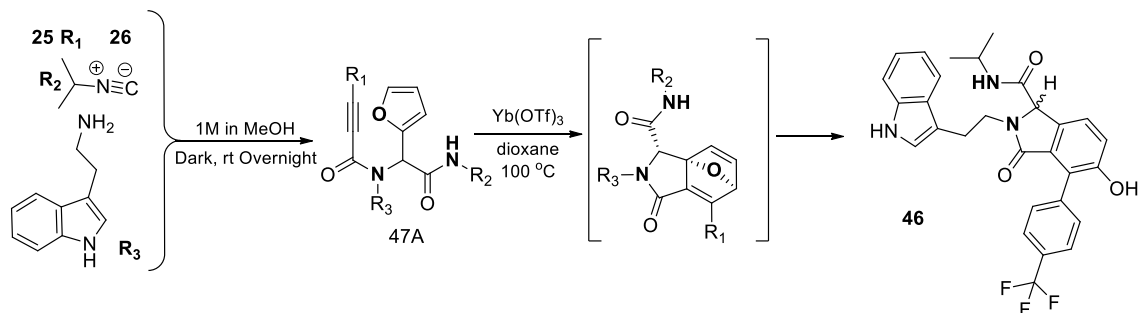


**43:** *N*-(*tert*-butyl)-2-(2,2-diphenylethyl)-5-hydroxy-3-oxo-4-(4-

(trifluoromethyl)phenyl)isoindoline-1-carboxamide: **Step I:** One molar solution of

freshly distilled furfural (41  $\mu$ L 0.5 mmol), 2,2-diphenylethylamine (98 mg, 0.5 mmol), 3-(4-(trifluoromethyl)phenyl) propionic acid (107 mg, 0.5 mmol) *tert*-Butyl isocyanide (56  $\mu$ L, 0.5 mmol) was made up in MeOH (2 mL) and stirred in the dark under argon for 12 h. The Ugi product was purified by flash chromatography first with  $CH_2Cl_2$  and then gradient elution with EtoAc-Hexane (10%, 20% & 40%) to give a mixture of three Ugi products that was used in the next step. **Step II:** 0.02M of the Ugi product (191 mg, 0.33 mmol) was prepared in Dioxane (8 mL). 20 mol% ytterbium(III) trifluoromethanesulfonate (51 mg, 0.07 mmol) was added and the reaction heated at 100 $^{\circ}$  C in a sealed tube for 12 h. The solvent was removed by rotavap and product purified by flash chromatography twice (70% & 35% EtoAc-Hexane) to give a white solid (125 mg, 10.0% over two steps)  $^1H$  NMR (400 MHz, DMSO- $d_6$ )  $\delta$  9.88 (s, 1H), 8.44 (s, 1H), 7.74 (d,  $J$  = 8.28 Hz, 2H), 7.50 (d,  $J$  = 8.28 Hz, 2H), 7.12 - 7.39 (m, 12 h), 4.89 (s, 1H), 4.52 (d,  $J$  = 13.80 Hz, 1H), 4.42 (d,  $J$  = 6.27 Hz, 1H), 3.30 - 3.39 (m, 1H), 1.37 (s, 9H)  $^{13}C$

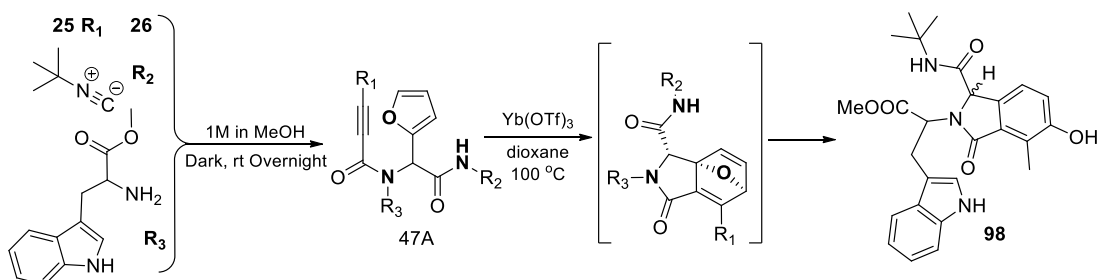
NMR (101 MHz, DMSO- $d_6$ )  $\delta$  167.4, 166.0, 154.9, 142.6, 141.4, 133.2, 131.6, 128.9, 128.8, 128.4, 127.8, 127.6, 126.7, 126.6, 123.7, 122.6, 119.2, 61.4, 50.9, 49.0, 44.7, 28.4



**46:** *2-(2-(1H-indol-3-yl)ethyl)-5-hydroxy-N-isopropyl-3-oxo-4-(4-(trifluoromethyl)phenyl)isoindoline-1-carboxamide*: **Step I:** One molar solution of freshly distilled furfural (41  $\mu$ L 0.5 mmol), tryptamine (80 mg, 0.5 mmol), 3-(4-(trifluoromethyl)phenyl) propionic acid (107 mg, 0.5 mmol) and isopropyl isocyanide (47  $\mu$ L, 0.5 mmol) was made up in MeOH (2 mL) and stirred in the dark under argon for 12 h. The Ugi product was purified by flash chromatography first with CH<sub>2</sub>Cl<sub>2</sub> and then gradient elution with EtoAc-Hexane (10%, 20% & 40%) to give a mixture of three Ugi products that was used in the next step. **Step II:** 0.02M of the Ugi product (135 mg, 0.25 mmol) was prepared in Dioxane (5 mL). 20mol% ytterbium(III) trifluoromethanesulfonate (40 mg, 0.06 mmol) was added and the reaction heated at 100° C in a sealed tube for 12 h. The solvent was removed by rotavap and product purified by flash chromatography twice (70% & 35% EtoAc-Hexane) to give a white solid (98 mg, 9.0% over two steps). <sup>1</sup>H NMR (400 MHz, DMSO- $d_6$ ) Shift 10.84 (br. s., 1H), 9.91 (s, 1H), 8.56 (d, J = 7.28 Hz, 1H), 7.74 (d, J = 7.53 Hz, 2H), 7.50 - 7.57 (m, 3H), 7.37 (dd, J = 8.53, 11.29 Hz, 2H), 7.20 (d, J = 8.28 Hz, 1H), 7.14 (s, 1H), 7.08 (t, J = 7.53 Hz, 1H), 6.94 - 7.02 (m, 1H), 5.10 (s, 1H), 3.89 - 4.03 (m, 2H), 2.85 - 3.25 (m, 3H), 1.16 (t, J =

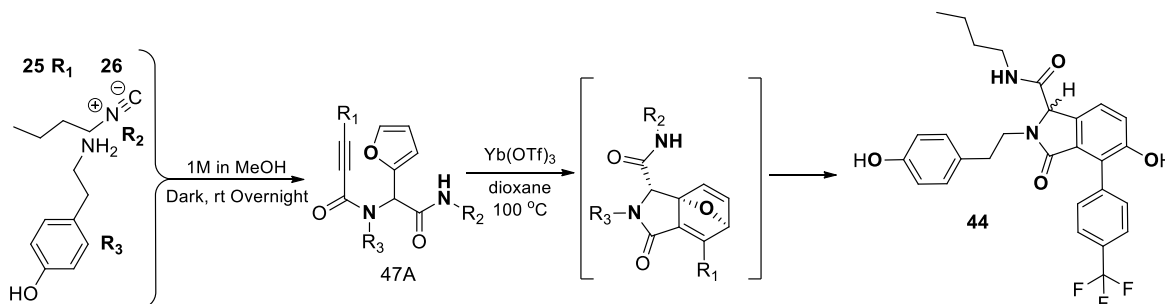


5.77 Hz, 6H)  $^{13}\text{C}$  NMR (101 MHz, DMSO- $d_6$ )  $\delta$  167.2, 166.3, 154.9, 138.0, 136.2, 133.1, 131.6, 129.3, 127.0, 123.8, 123.7, 123.2, 122.6, 122.5, 121.0, 119.1, 118.2, 118.0, 111.5, 111.1, 61.6, 41.6, 41.0, 23.5, 22.3



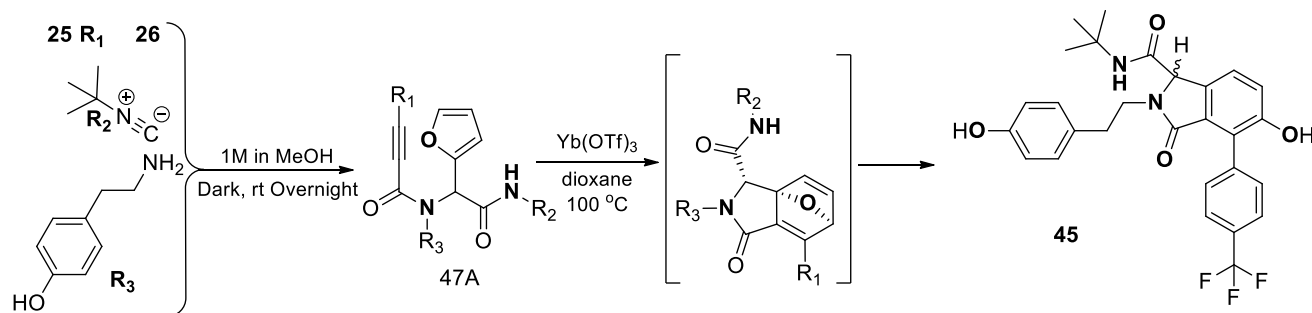
**98:** *methyl 3-(1-(tert-butylcarbamoyl)-5-hydroxy-4-methyl-3-oxoisindolin-2-yl)-2-(1H-indol-3-yl)propanoate*: **Step I:** One molar solution of freshly distilled furfural (331  $\mu\text{L}$  4 mmol), methyl 2-amino-4-(1H-indol-3-yl)butanoate (929 mg, 4 mmol), 3-(4-(trifluoromethyl)phenyl) propionic acid (107 mg, 0.5 mmol) and tert-Butyl isocyanide (452  $\mu\text{L}$ , 4 mmol) was made up in MeOH (16 mL) and stirred in the dark under argon for 12 h. The Ugi product was purified by flash chromatography first with  $\text{CH}_2\text{Cl}_2$  and then gradient elution with EtoAc-Hexane (10%, 20% & 40%) to give a mixture of three Ugi products that was used in the next step. **Step II:** 0.02 M of the Ugi product (185 mg, 0.39 mmol) was prepared in Dioxane (5 mL). 20 mol% ytterbium(III) trifluoromethanesulfonate (61 mg, 0.09 mmol) was added and the reaction heated at 100 $^\circ\text{C}$  in a sealed tube for 12 h. The solvent was removed by rotavap and product purified by flash chromatography twice (70% & 35% EtoAc-Hexane) to give a white solid (164 mg, 9.0% over two steps)  $^1\text{H}$  NMR (400 MHz, DMSO- $d_6$ )  $\delta$  10.82 (s, 1H), 9.67 (s, 1H), 8.19 (s, 1H), 7.63 (d,  $J = 7.78$  Hz, 1H), 7.34 (d,  $J = 8.03$  Hz, 1H), 7.17 (d,  $J = 8.28$  Hz, 1H), 6.99 - 7.13 (m, 5H), 5.19 (s, 1H), 5.05 (t,  $J = 7.65$  Hz, 1H), 3.54 (s, 3H), 3.36 (s, 1H),

2.46 (s, 3H), 1.31 (s, 9H)  $^{13}\text{C}$  NMR (101 MHz, DMSO- $d_6$ )  $\delta$  170.3, 169.3, 166.6, 155.8, 136.1, 133.3, 128.7, 127.0, 123.2, 122.4, 121.1, 119.6, 118.4, 118.2, 111.4, 109.4, 65.2, 60.6, 54.0, 51.7, 50.6, 28.4, 9.4



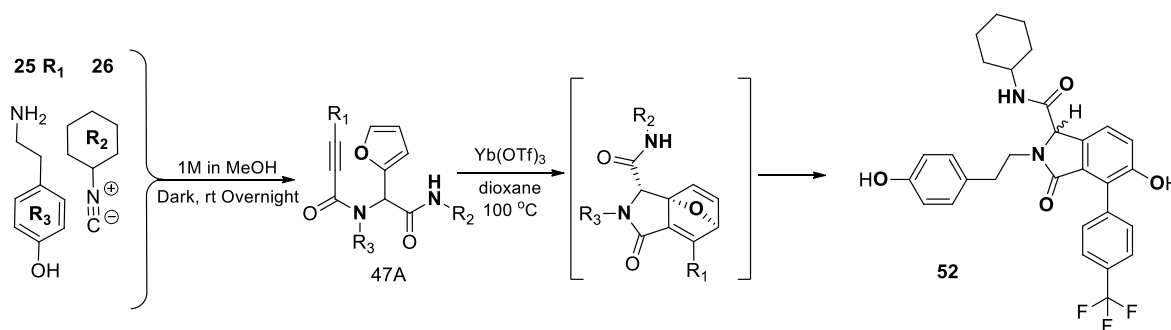
- 44:** *N*-butyl-5-hydroxy-2-(4-hydroxyphenethyl)-3-oxo-4-(4-(trifluoromethyl)phenyl)isoindoline-1-carboxamide **Step I:** One molar solution of freshly distilled furfural (41  $\mu\text{L}$  0.5 mmol), tyramine (68 mg, 0.5 mmol), 3-(4-(trifluoromethyl)phenyl) propionic acid (107 mg, 0.5 mmol) n-butylisocyanide (52  $\mu\text{L}$ , 0.5 mmol) was made up in MeOH (2 mL) and stirred in the dark under argon for 12 h. The Ugi product was purified by flash chromatography first with  $\text{CH}_2\text{Cl}_2$  and then gradient elution with EtoAc-Hexane (10%, 20% & 40%) to give a mixture of three Ugi products that was used in the next step. **Step II:** 0.02M of the Ugi product (139 mg, 0.27 mmol) was prepared in Dioxane (6 mL). 20mol% ytterbium(III) trifluoromethanesulfonate (42 mg, 0.07 mmol) was added and the reaction heated at 100 $^\circ$  C in a sealed tube for 12 h. The solvent was removed by rotavap and product purified by flash chromatography twice (70% & 35% EtoAc-Hexane) to give a white solid (84 mg, 12.3% over two steps)  $^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$  9.87 (br. s., 1H), 9.18 (br. s., 1H), 8.47 (br. s., 1H), 7.67 (d,  $J$  = 7.34 Hz, 2H), 7.46 (d,  $J$  = 7.34 Hz, 2H), 7.31 (d,  $J$  = 8.31 Hz, 1H), 7.12 (d,  $J$  = 8.31 Hz, 1H), 6.91 (d,  $J$  = 7.34 Hz, 2H), 6.62 (d,  $J$  = 7.34 Hz,

2H), 4.91 (br. s., 1H), 3.82 (d,  $J = 4.89$  Hz, 2H), 2.93 - 3.16 (m, 2H), 2.53 - 2.77 (m, 2H), 1.39 (d,  $J = 6.36$  Hz, 2H), 1.26 (d,  $J = 6.85$  Hz, 2H), 0.84 (t,  $J = 6.60$  Hz, 3H)  $^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ )  $\delta$  167.6, 167.5, 156.2, 155.3, 138.3, 133.4, 132.1, 131.9, 129.7, 129.6, 129.2, 124.1, 123.9, 119.5, 115.7, 115.6, 62.4, 43.1, 38.9, 33.0, 31.5, 19.9, 14.1



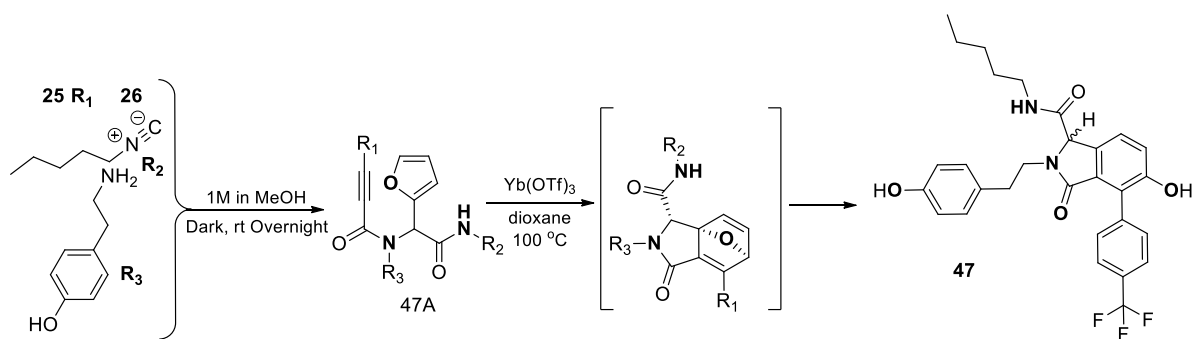
45: *N*-(*tert*-butyl)-5-hydroxy-2-(4-hydroxyphenethyl)-3-oxo-4-(4-(trifluoromethyl)phenyl)isoindoline-1-carboxamide: **Step I:** One molar solution of freshly distilled furfural (41  $\mu\text{L}$  0.5 mmol), tyramine (68 mg, 0.5 mmol), 3-(4-(trifluoromethyl)phenyl) propionic acid (107 mg, 0.5 mmol) *tert*-Butyl isocyanide (56  $\mu\text{L}$ , 0.5 mmol) was made up in MeOH (2 mL) and stirred in the dark under argon for 12 h. The Ugi product was purified by flash chromatography first with  $\text{CH}_2\text{Cl}_2$  and then gradient elution with EtoAc-Hexane (10%, 20% & 40%) to give a mixture of three Ugi products that was used in the next step. **Step II:** 0.02M of the Ugi product (264 mg, 0.51 mmol) was prepared in Dioxane (10 mL). 20 mol% ytterbium(III) trifluoromethanesulfonate (79 mg, 0.12 mmol) was added and the reaction heated at 100 $^\circ$  C in a sealed tube for 12 h. The solvent was removed by rotavap and product purified by flash chromatography twice (70% & 35% EtoAc-Hexane) to give a white solid (191 mg, 18.0% over two steps)  $^1\text{H}$  NMR (400 MHz, DMSO- $d_6$ )  $\delta$  9.90 (s, 1H), 9.22 (s, 1H), 8.52

(t,  $J = 5.65$  Hz, 1H), 7.73 (d,  $J = 8.28$  Hz, 2H), 7.53 (d,  $J = 8.03$  Hz, 2H), 7.38 (d,  $J = 8.28$  Hz, 1H), 7.19 (d,  $J = 8.28$  Hz, 1H), 6.98 (d,  $J = 8.53$  Hz, 2H), 6.69 (d,  $J = 8.28$  Hz, 2H), 4.98 (s, 1H), 3.83 - 3.94 (m, 1H), 3.00 - 3.24 (m, 2H), 2.62 - 2.84 (m, 2H), 1.25 - 1.51 (m, 5H), 0.91 (t,  $J = 7.28$  Hz, 3H)  $^{13}\text{C}$  NMR (101 MHz, DMSO- $d_6$ )  $\delta$  167.2, 167.1, 155.8, 154.9, 137.9, 133.0, 131.6, 129.3, 129.2, 128.8, 127.1, 123.8, 123.7, 123.6, 123.2, 122.6, 119.1, 115.2, 61.9, 42.8, 38.5, 32.6, 31.1



**52:** *N*-cyclohexyl-5-hydroxy-2-(4-hydroxyphenethyl)-3-oxo-4-(4-(trifluoromethyl)phenyl)isoindoline-1-carboxamide carboxamide: **Step I:** One molar solution of freshly distilled furfural (41  $\mu\text{L}$  0.5 mmol), tyramine (68 mg, 0.5 mmol), 3-(4-(trifluoromethyl)phenyl) propionic acid (107 mg, 0.5 mmol) isocyanocyclohexane (62  $\mu\text{L}$ , 0.5 mmol) was made up in MeOH (2 mL) and stirred in the dark under argon for 12 h. The Ugi product was purified by flash chromatography first with  $\text{CH}_2\text{Cl}_2$  and then gradient elution with EtoAc-Hexane (10%, 20% & 40%) to give a mixture of three Ugi products that was used in the next step. **Step II:** 0.02 M of the Ugi product (159 mg, 0.29 mmol) was prepared in Dioxane (6 mL). 20 mol% ytterbium(III) trifluoromethanesulfonate (46 mg, 0.07 mmol) was added and the reaction heated at 100 $^\circ$  C in a sealed tube for 12 h. The solvent was removed by rotavap and product purified by flash chromatography twice (70% & 35% EtoAc-Hexane) to give a white solid (119 mg,

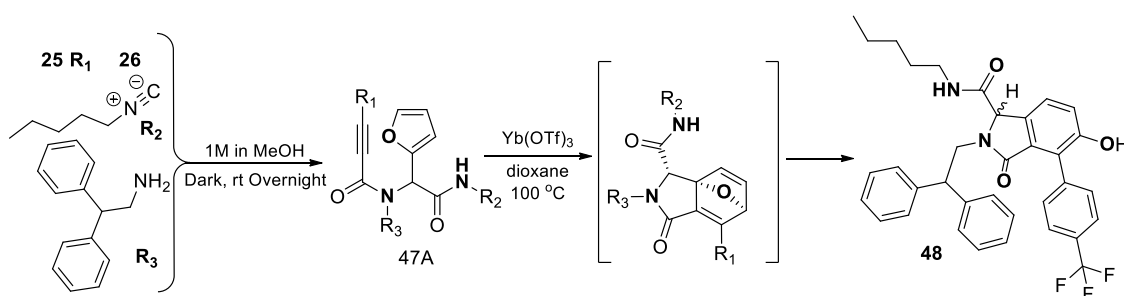
11.4% over two steps)  $^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$  9.90 (br. s., 1H), 9.22 (br. s., 1H), 8.50 (d,  $J = 7.34$  Hz, 1H), 7.67 - 7.72 (m, 2H), 7.49 (d,  $J = 7.34$  Hz, 2H), 7.31 - 7.36 (m, 1H), 7.16 (d,  $J = 7.83$  Hz, 1H), 6.91 - 6.97 (m, 2H), 6.66 (d,  $J = 8.31$  Hz, 2H), 4.97 (br. s., 1H), 3.83 (br. s., 1H), 3.57 (br. s., 2H), 3.36 (br. s., 2H), 3.01 (d,  $J = 5.87$  Hz, 2H), 2.73 (br. s., 2H), 2.64 (br. s., 2H), 1.52 - 1.81 (m, 4H), 1.14 - 1.31 (m, 4H)  $^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ )  $\delta$  167.5, 166.6, 156.2, 138.4, 133.5, 132.0, 130.3, 129.7, 129.2, 127.7, 126.1, 124.2, 124.0, 123.9, 119.4, 115.7, 115.6, 62.2, 48.5, 43.1, 33.1, 32.7, 25.6, 24.9



**47:** *5-hydroxy-2-(4-hydroxyphenethyl)-3-oxo-N-pentyl-4-(4-*

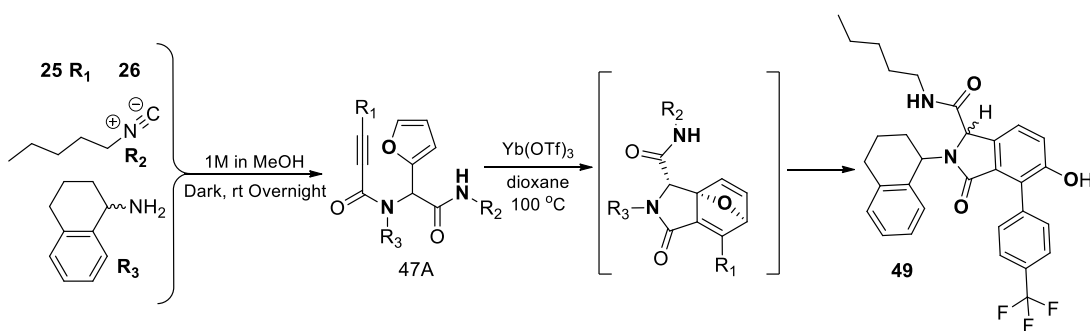
*(trifluoromethyl)phenyl)isoindoline-1-carboxamide:* **Step I:** One molar solution of freshly distilled furfural (41  $\mu\text{L}$  0.5 mmol), tyramine (68 mg, 0.5 mmol), 3-(4-(trifluoromethyl)phenyl) propionic acid (107 mg, 0.5 mmol) 1-isocyanopentane (62  $\mu\text{L}$ , 0.5 mmol) was made up in MeOH (2 mL) and stirred in the dark under argon for 12 h. The Ugi product was purified by flash chromatography first with  $\text{CH}_2\text{Cl}_2$  and then gradient elution with EtoAc-Hexane (10%, 20% & 40%) to give a mixture of three Ugi products that was used in the next step. **Step II:** 0.02 M of the Ugi product (203 mg, 0.38 mmol) was prepared in Dioxane (8 mL). 20mol% ytterbium(III) trifluoromethanesulfonate (59 mg, 0.09 mmol) was added and the reaction heated at 100°

C in a sealed tube for 12 h. The solvent was removed by rotavap and product purified by flash chromatography twice (70% & 35% EtoAc-Hexane) to give a white solid (125 mg, 12.0% over two steps)  $^1\text{H}$  NMR (400 MHz, DMSO- $d_6$ )  $\delta$  9.91 (br. s., 1H), 9.23 (s, 1H), 8.52 (t,  $J$  = 5.40 Hz, 1H), 7.73 (d,  $J$  = 7.78 Hz, 2H), 7.53 (d,  $J$  = 8.03 Hz, 2H), 7.37 (d,  $J$  = 8.28 Hz, 1H), 7.19 (d,  $J$  = 8.28 Hz, 1H), 6.98 (d,  $J$  = 8.03 Hz, 2H), 6.68 (d,  $J$  = 7.53 Hz, 2H), 4.98 (s, 1H), 3.85 - 3.94 (m, 1H), 3.38 (s, 1H), 3.00 - 3.23 (m, 3H), 2.62 - 2.84 (m, 2H), 1.45 - 1.50 (m, 2H), 1.29 (d,  $J$  = 3.01 Hz, 4H), 0.88 (t,  $J$  = 6.65 Hz, 3H)  $^{13}\text{C}$  NMR (101 MHz, DMSO- $d_6$ )  $\delta$  155.8, 154.9, 137.9, 133.0, 131.6, 129.3, 129.2, 128.8, 127.4, 127.1, 125.9, 123.8, 123.7, 123.2, 122.6, 119.1, 115.2, 61.9, 42.8, 38.8, 32.6, 28.6, 28.5, 21.8, 13.9



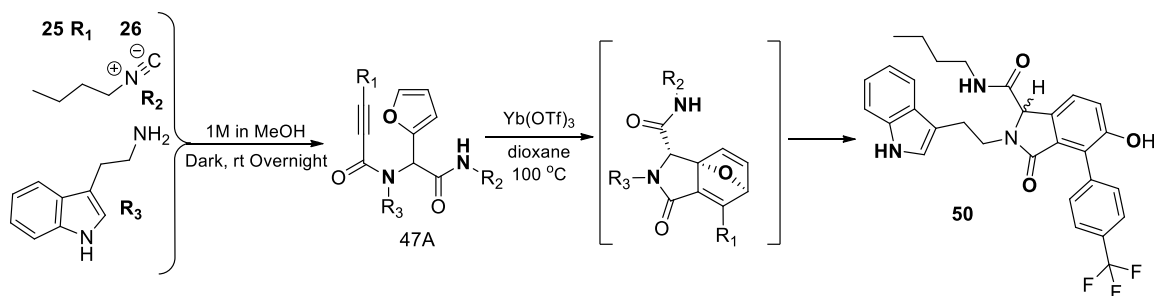
- 48:** *2-(2,2-diphenylethyl)-5-hydroxy-3-oxo-N-pentyl-4-(4-(trifluoromethyl)phenyl)isoindoline-1-carboxamide*: **Step I:** One molar solution of freshly distilled furfural (41  $\mu\text{L}$  0.5 mmol), 2,2-diphenylethylamine (98 mg, 0.5 mmol), 3-(4-(trifluoromethyl)phenyl) propionic acid (107 mg, 0.5 mmol) and 1-isocyanopentane (62  $\mu\text{L}$ , 0.5 mmol) was made up in MeOH (2 mL) and stirred in the dark under argon for 12 h. The Ugi product was purified by flash chromatography first with  $\text{CH}_2\text{Cl}_2$  and then gradient elution with EtoAc-Hexane (10%, 20% & 40%) to give a mixture of three Ugi products that was used in the next step. **Step II:** 0.02 M of the Ugi product (276 mg, 0.47 mmol) was prepared in Dioxane (6 mL). 20mol% ytterbium(III)

trifluoromethanesulfonate (73 mg, 0.11 mmol) was added and the reaction heated at 100° C in a sealed tube for 12 h. The solvent was removed by rotavap and product purified by flash chromatography twice (70% & 35% EtoAc-Hexane) to give a white solid (116 mg, 10.0% over two steps). <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>) δ 9.92 (s, 1H), 8.59 (t, *J* = 5.52 Hz, 1H), 7.75 (d, *J* = 8.28 Hz, 2H), 7.50 (d, *J* = 8.03 Hz, 2H), 7.14 - 7.37 (m, 12 h), 4.74 (s, 1H), 4.54 (dd, *J* = 9.29, 13.80 Hz, 1H), 4.38 - 4.47 (m, 1H), 3.37 - 3.46 (m, 1H), 3.13 - 3.21 (m, 2H), 1.44 - 1.54 (m, 2H), 1.30 (td, *J* = 3.58, 6.90 Hz, 4H), 0.83 - 0.93 (m, 3H) <sup>13</sup>C NMR (101 MHz, DMSO-d<sub>6</sub>) δ 167.3, 166.8, 155.0, 142.5, 141.5, 132.9, 131.6, 128.7, 128.7, 128.4, 127.7, 127.6, 126.8, 126.6, 123.8, 123.7, 122.7, 119.2, 64.9, 61.6, 49.0, 45.0, 38.8, 28.7, 28.5, 21.8, 13.9



- 49:** *5-hydroxy-3-oxo-N-pentyl-2-(1,2,3,4-tetrahydronaphthalen-1-yl)-4-(4-(trifluoromethyl)phenyl)isoindoline-1-carboxamide* P1: **Step I:** One molar solution of freshly distilled furfural (82 μL 1.0 mmol), 1,2,3,4-tetrahydro-1-naphthalenamine (146 μL, 1 mmol), 3-(4-(trifluoromethyl)phenyl) propionic acid (214 mg, 1.0 mmol) and 1-isocyanopentane (125 μL, 1.0 mmol) was made up in MeOH (4 mL) and stirred in the dark under argon for 12 h. The Ugi product was purified by flash chromatography first with CH<sub>2</sub>Cl<sub>2</sub> and then gradient elution with EtoAc-Hexane (10%, 20% & 40%) to give a mixture of three Ugi products that was used in the next step. **Step II:** 0.02M of the Ugi

product (235 mg, 0.43 mmol) was prepared in Dioxane (9 mL). 20 mol% ytterbium(III) trifluoromethanesulfonate (67 mg, 0.11 mmol) was added and the reaction heated at 100° C in a sealed tube for 12 h. The solvent was removed by rotavap and product purified by flash chromatography twice (70% & 35% EtoAc-Hexane) to give a white solid (194 mg, 36.2% over two steps) <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>) δ 9.83 - 10.14 (m, 1H), 8.49 - 8.61 (m, 1H), 7.72 - 7.80 (m, 2H), 7.66 (d, *J* = 8.28 Hz, 1H), 7.28 - 7.34 (m, 1H), 7.20 - 7.27 (m, 1H), 7.05 - 7.19 (m, 2H), 6.77 (d, *J* = 7.78 Hz, 1H), 5.39 (dd, *J* = 5.90, 9.91 Hz, 1H), 4.78 (s, 1H), 3.45 (s, 1H), 3.05 - 3.20 (m, 2H), 2.76 (br. s., 2H), 1.63 - 2.17 (m, 4H), 1.39 - 1.53 (m, 2H), 1.17 - 1.37 (m, 5H), 0.89 (t, *J* = 6.78 Hz, 3H) <sup>13</sup>C NMR (101 MHz, DMSO-d<sub>6</sub>) δ 168.6, 168.3, 155.1, 137.6, 135.3, 133.9, 131.7, 129.3, 129.0, 126.8, 126.1, 126.0, 124.1, 123.7, 123.2, 122.4, 119.5, 60.1, 50.9, 38.8, 28.9, 28.5, 28.5, 27.8, 21.8, 13.8

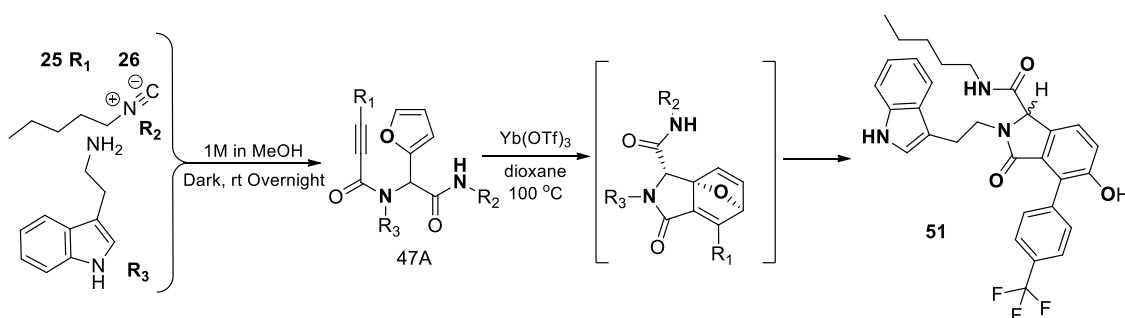


**50:** 2-(2-(1H-indol-3-yl)ethyl)-N-butyl-4-(4-(trifluoromethyl)phenyl)-5-hydroxy-3-

*oxisoindoline-1-carboxamide*: **Step I:** One molar solution of freshly distilled furfural (62 μL 0.75mmol), tryptamine (120 mg, 0.75 mmol), 3-(4-(trifluoromethyl)phenyl) propionic acid (160 mg, 0.75 mmol) and n-butylisocyanide (78 μL, 0.75 mmol) was made up in MeOH (2 mL) and stirred in the dark under argon for 12 h. The Ugi product was purified by flash chromatography first with CH<sub>2</sub>Cl<sub>2</sub> and then gradient elution with



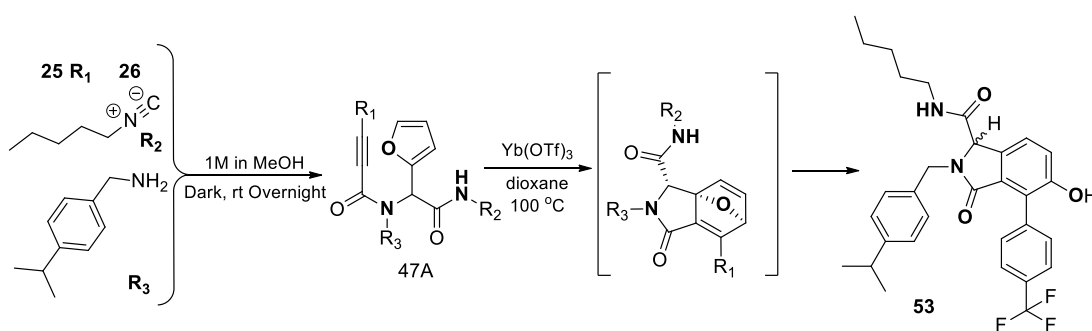
EtoAc-Hexane (10%, 20% &40%)) to give a mixture of three Ugi products that was used in the next step. **Step II:** 0.02M of the Ugi product (350 mg, 0.65 mmol) was prepared in Dioxane (13 mL). 20 mol% ytterbium(III) trifluoromethanesulfonate (101 mg, 0.13 mmol) was added and the reaction heated at 100° C in a sealed tube for 12 h. The solvent was removed by rotavap and product purified by flash chromatography twice (70% & 35% EtoAc-Hexane) to give a white solid (294 mg, 41.2% over two steps) <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>) δ 10.83 (s, 1H), 9.91 (s, 1H), 8.55 (t, *J* = 5.65 Hz, 1H), 7.73 (d, *J* = 8.03 Hz, 2H), 7.49 - 7.56 (m, 2H), 7.36 (dd, *J* = 8.16, 10.42 Hz, 2H), 7.19 (d, *J* = 8.28 Hz, 1H), 7.12 (d, *J* = 2.01 Hz, 1H), 7.08 (t, *J* = 7.15 Hz, 1H), 6.94 - 7.01 (m, 1H), 5.10 (s, 1H), 3.94 - 4.04 (m, 1H), 3.35 (s, 2H), 3.09 - 3.24 (m, 2H), 2.84 - 3.07 (m, 2H), 1.24 - 1.52 (m, 4H), 0.90 (t, *J* = 7.28 Hz, 3H) <sup>13</sup>C NMR (101 MHz, DMSO-d<sub>6</sub>) δ 167.3, 167.2, 154.9, 138.1, 136.2, 133.1, 131.6, 129.3, 127.0, 123.9, 123.7, 123.2, 122.6, 122.6, 121.0, 119.1, 118.3, 118.0, 111.5, 111.1, 61.8, 41.6, 38.6, 31.1, 23.5, 21.0, 19.5, 13.6



**51:** 2-(2-(1*H*-indol-3-yl)ethyl)-5-hydroxy-3-oxo-*N*-pentyl-4-(4-

(trifluoromethyl)phenyl)isoindoline-1-carboxamide: **Step I:** One molar solution of freshly distilled furfural (41 μL 0.5 mmol), tryptamine (80 mg, 0.5 mmol), 3-(4-(trifluoromethyl)phenyl) propionic acid (107 mg, 0.5 mmol) and 1-isocyanopentane (62

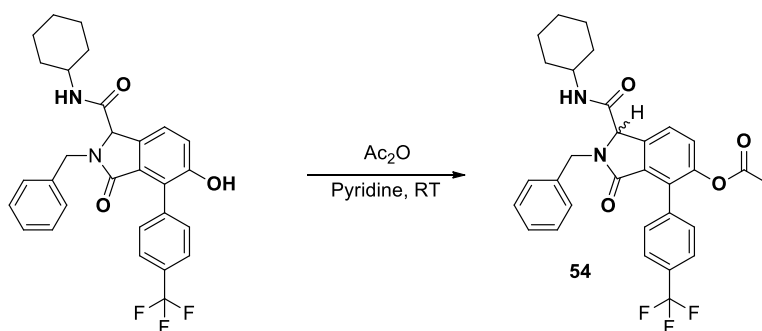
$\mu\text{L}$ , 0.5 mmol) was made up in MeOH (2 mL) and stirred in the dark under argon for 12 h. The Ugi product was purified by flash chromatography first with  $\text{CH}_2\text{Cl}_2$  and then gradient elution with EtoAc-Hexane (10%, 20% & 40%) to give a mixture of three Ugi products that was used in the next step. **Step II:** 0.02M of the Ugi product (168 mg, 0.31 mmol) was prepared in Dioxane (6 mL). 20mol% ytterbium(III) trifluoromethanesulfonate (47 mg, 0.07 mmol) was added and the reaction heated at  $100^\circ\text{C}$  in a sealed tube for 12 h. The solvent was removed by rotavap and product purified by flash chromatography twice (70% & 35% EtoAc-Hexane) to give a white solid (74 mg, 7.0% over two steps)  $^1\text{H}$  NMR (400 MHz,  $\text{DMSO-d}_6$ )  $\delta$  10.85 (br. s., 1H), 9.93 (s, 1H), 8.51 - 8.63 (m, 1H), 7.75 (d,  $J = 8.03$  Hz, 2H), 7.50 - 7.59 (m, 2H), 7.38 (dd,  $J = 8.16$ , 13.43 Hz, 2H), 7.18 - 7.24 (m, 1H), 7.06 - 7.15 (m, 2H), 6.95 - 7.03 (m, 1H), 5.09 - 5.15 (m, 1H), 3.96 - 4.20 (m, 1H), 3.36 - 3.43 (m, 2H), 2.87 - 3.28 (m, 4H), 1.48 (d,  $J = 6.27$  Hz, 2H), 1.30 (d,  $J = 3.26$  Hz, 4H), 0.83 - 0.94 (m, 3H)  $^{13}\text{C}$  NMR (101 MHz,  $\text{DMSO-d}_6$ )  $\delta$  167.3, 167.2, 155.0, 136.2, 133.1, 131.6, 129.3, 127.0, 123.9, 123.7, 123.7, 122.6, 122.6, 121.0, 119.1, 118.3, 118.0, 111.5, 111.1, 61.8, 41.7, 38.8, 28.7, 28.5, 23.5, 21.8, 13.9



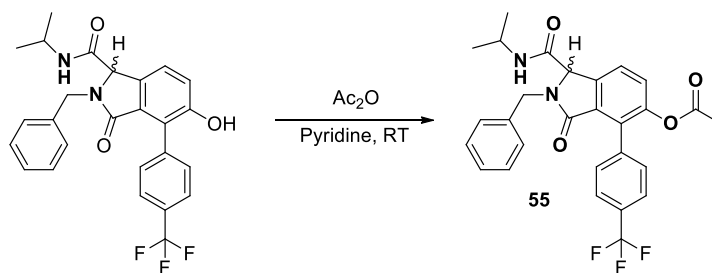
**53:** 5-hydroxy-2-(4-isopropylbenzyl)-3-oxo-N-pentyl-4-(4-

(trifluoromethyl)phenyl)isoindoline-1-carboxamide: **Step I:** One molar solution of

freshly distilled furfural (41  $\mu$ L 0.5 mmol), 4-isopropylbenzylamine (80  $\mu$ L, 0.5 mmol), 3-(4-(trifluoromethyl)phenyl) propionic acid (107 mg, 0.5 mmol) and 1-isocyanopentane (62  $\mu$ L, 0.5 mmol) was made up in MeOH (2 mL) and stirred in the dark under argon for 12 h. The Ugi product was purified by flash chromatography first with  $\text{CH}_2\text{Cl}_2$  and then gradient elution with EtoAc-Hexane (10%, 20% & 40%) to give a mixture of three Ugi products that was used in the next step. **Step II:** 0.02M of the Ugi product (159 mg, 0.29 mmol) was prepared in Dioxane (6 mL). 20mol% ytterbium(III) trifluoromethanesulfonate (45 mg, 0.07 mmol) was added and the reaction heated at 100 $^\circ$  C in a sealed tube for 12 h. The solvent was removed by rotavap and product purified by flash chromatography twice (70% & 35% EtoAc-Hexane) to give a white solid (113 mg, 10.5% over two steps)  $^1\text{H}$  NMR (400 MHz,  $\text{DMSO-d}_6$ )  $\delta$  9.94 (s, 1H), 8.52 (t,  $J = 5.40$  Hz, 1H), 7.75 (d,  $J = 8.03$  Hz, 2H), 7.58 (d,  $J = 8.03$  Hz, 2H), 7.34 (d,  $J = 8.28$  Hz, 1H), 7.17 - 7.25 (m, 3H), 7.12 (d,  $J = 8.03$  Hz, 2H), 4.79 (s, 1H), 3.36 (s, 2H), 3.07 - 3.19 (m, 2H), 2.88 (td,  $J = 6.90, 13.80$  Hz, 1H), 1.46 (quin,  $J = 6.78$  Hz, 2H), 1.25 - 1.36 (m, 4H), 1.19 (d,  $J = 7.03$  Hz, 6H), 0.86 - 0.94 (m, 3H)  $^{13}\text{C}$  NMR (101 MHz,  $\text{DMSO-d}_6$ )  $\delta$  167.1, 166.8, 155.1, 147.6, 137.9, 137.8, 134.3, 132.9, 131.6, 128.8, 128.0, 126.6, 125.9, 124.1, 123.7, 123.7, 122.7, 119.2, 60.9, 44.0, 38.8, 33.1, 28.5, 23.8, 21.8, 13.9

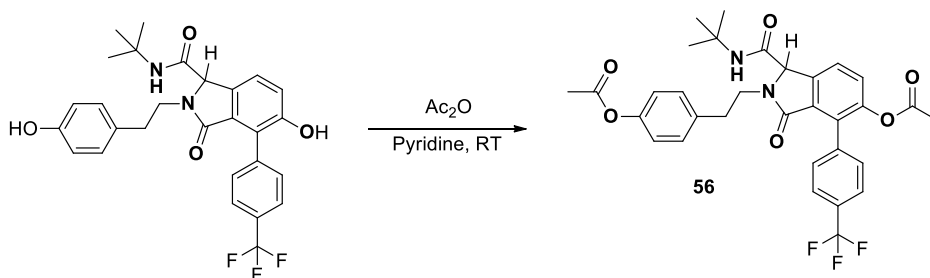


**54:** *2-benzyl-1-(cyclohexylcarbamoyl)-3-oxo-4-(4-(trifluoromethyl)phenyl)isoindolin-5-yl acetate*: To a stirring solution of acetic anhydride (100  $\mu$ L, 1.2mmol) in pyridine (2 mL) was added compound **38** (68 mg, 0.12 mmol) and stirred at room temperature overnight. After disappearance of starting material, the reaction mixture was diluted with ethyl acetate (3 mL) and anhydrous  $\text{CuSO}_4$  solution was added to remove the pyridine (3 x 3 mL). The organic portions were combined, washed with brine and dried over  $\text{Na}_2\text{SO}_4$ . The crude product was purified by column chromatography (35% EtOAc-hexanes) to give a white solid (66mg, 90%)  $^1\text{H}$  NMR (400 MHz,  $\text{DMSO-d}_6$ )  $\delta$  8.60 (d,  $J = 7.78$  Hz, 1H), 7.82 (d,  $J = 8.03$  Hz, 2H), 7.47 - 7.62 (m, 4H), 7.29 - 7.41 (m, 3H), 7.24 (d,  $J = 6.78$  Hz, 2H), 5.10 (d,  $J = 15.06$  Hz, 1H), 4.97 (s, 1H), 3.97 (d,  $J = 15.06$  Hz, 1H), 3.58 - 3.69 (m, 1H), 3.35 (s, 1H), 2.00 (s, 3H), 1.54 - 1.90 (m, 5H), 1.16 - 1.40 (m, 5H)  $^{13}\text{C}$  NMR (101 MHz,  $\text{DMSO-d}_6$ )  $\delta$  169.1, 166.2, 165.0, 148.1, 140.2, 136.6, 136.2, 130.8, 129.2, 128.7, 128.4, 128.1, 127.5, 126.6, 125.7, 124.2, 123.1, 61.2, 48.1, 44.4, 32.3, 32.2, 25.1, 24.3, 20.2



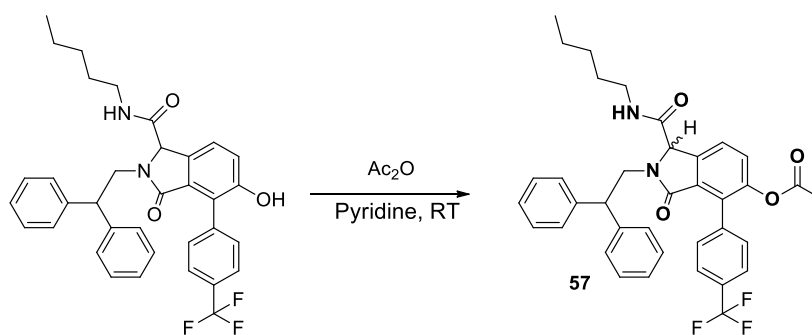
**55:** *2-benzyl-1-(isopropylcarbamoyl)-3-oxo-4-(4-(trifluoromethyl)phenyl)isoindolin-5-yl acetate*: To a stirring solution of acetic anhydride (100  $\mu$ L, 1.2 mmol) in pyridine (2 mL) was added compound **35** (86 mg, 0.18 mmol) and stirred at room temperature overnight. After disappearance of starting material, the reaction mixture was diluted with ethyl

acetate (3 mL) and anhydrous CuSO<sub>4</sub> solution was added to remove the pyridine (3 x 3 mL). The organic portions were combined, washed with brine and dried over Na<sub>2</sub>SO<sub>4</sub>. The crude product was purified by column chromatography (35% EtOAc-hexanes) to give a white solid (81mg, 86%) <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>) δ 8.60 (d, *J* = 7.53 Hz, 1H), 7.81 (d, *J* = 8.28 Hz, 2H), 7.57 - 7.62 (m, 1H), 7.52 (d, *J* = 8.03 Hz, 3H), 7.28 - 7.41 (m, 3H), 7.24 (d, *J* = 7.03 Hz, 2H), 5.08 (d, *J* = 15.31 Hz, 1H), 4.93 (s, 1H), 3.86 - 4.01 (m, 1H), 3.34 (s, 1H), 2.00 (s, 3H), 1.14 (dd, *J* = 6.65, 10.67 Hz, 6H) <sup>13</sup>C NMR (101 MHz, DMSO-d<sub>6</sub>) δ 169.1, 166.2, 165.0, 148.1, 140.1, 136.6, 130.8, 129.2, 128.7, 128.4, 128.0, 127.5, 126.6, 125.7, 124.2, 124.2, 123.1, 61.2, 44.4, 41.1, 22.3, 22.2, 20.2



- 56:** *4-(2-(5-acetoxy-1-(tert-butylcarbamoyl)-3-oxo-4-(4-(trifluoromethyl)phenyl)isoindolin-2-yl)ethyl)phenyl acetate*: To a stirring solution of acetic anhydride (100 μL, 1.2 mmol) in pyridine (2 mL) was added compound **45** (180 mg, 0.35 mmol) and stirred at room temperature overnight. After disappearance of starting material, the reaction mixture was diluted with ethyl acetate (3 mL) and anhydrous CuSO<sub>4</sub> solution was added to remove the pyridine (3 x 3 mL). The organic portions were combined, washed with brine and dried over Na<sub>2</sub>SO<sub>4</sub>. The crude product was purified by column chromatography (35% EtOAc-hexanes) to give a white solid (118 mg, 56%) <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>) δ 8.69 (t, *J* = 5.27 Hz, 1H), 7.81 (d, *J* = 8.03 Hz, 2H), 7.68 (d, *J* = 8.03 Hz, 1H), 7.47 - 7.55 (m, 3H),

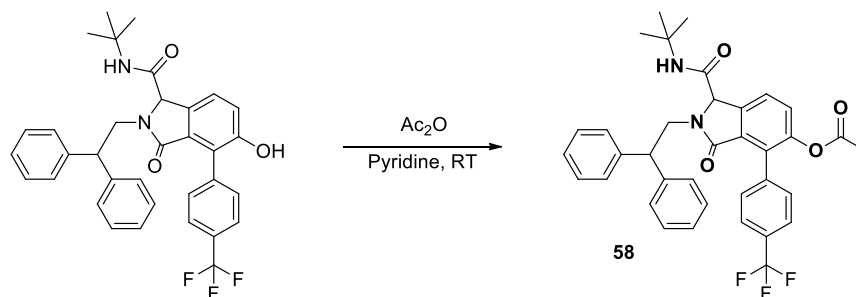
7.23 - 7.29 (m, 2H), 7.05 - 7.10 (m, 2H), 5.24 (s, 1H), 3.95 - 4.05 (m, 1H), 3.39 (s, 1H), 3.12 - 3.30 (m, 2H), 2.80 - 3.01 (m, 2H), 2.27 (s, 3H), 2.01 (s, 3H), 1.29 - 1.55 (m, 4H), 0.93 (t,  $J = 7.40$  Hz, 3H)  $^{13}\text{C}$  NMR (101 MHz, DMSO- $d_6$ )  $\delta$  169.1, 169.0, 166.4, 166.2, 149.0, 148.1, 140.2, 136.2, 130.7, 130.6, 129.4, 129.4, 126.5, 125.7, 124.1, 123.1, 121.8, 62.2, 42.5, 38.7, 32.7, 31.0, 20.8, 20.2



**57:** *2-(2,2-diphenylethyl)-3-oxo-1-(pentylcarbamoyl)-4-(4-*

*(trifluoromethyl)phenyl)isoindolin-5-yl acetate* To a stirring solution of acetic anhydride (100  $\mu\text{L}$ , 1.2mmol) in pyridine (2 mL) was added compound **48** (50 mg, 0.08 mmol) and stirred at room temperature overnight. After disappearance of starting material, the reaction mixture was diluted with ethyl acetate (3 mL) and anhydrous  $\text{CuSO}_4$  solution was added to wash the pyridine (3 x 3 mL). The organic portions were combined, washed with brine and dried over  $\text{Na}_2\text{SO}_4$ . The crude product was purified by column chromatography (35% EtOAc-hexanes) to give a white solid (44 mg, 82%)  $^1\text{H}$  NMR (400 MHz, DMSO- $d_6$ )  $\delta$  8.75 (t,  $J = 5.27$  Hz, 1H), 7.81 (d,  $J = 8.28$  Hz, 2H), 7.60 (d,  $J = 8.28$  Hz, 1H), 7.46 (dd,  $J = 8.03, 16.81$  Hz, 3H), 7.15 - 7.39 (m, 9H), 4.91 (s, 1H), 4.55 (dd,  $J = 9.41, 13.68$  Hz, 1H), 4.38 - 4.49 (m, 1H), 3.48 (dd,  $J = 7.03, 13.80$  Hz, 1H), 3.20 (q,  $J = 6.11$  Hz, 2H), 1.98 (s, 3H), 1.46 - 1.56 (m, 2H), 1.27 - 1.36 (m, 4H), 0.88 (t,  $J = 6.27$  Hz, 3H)  $^{13}\text{C}$  NMR (101 MHz, DMSO- $d_6$ )  $\delta$  169.0, 166.3, 166.0, 148.1, 142.4, 141.5, 140.0,

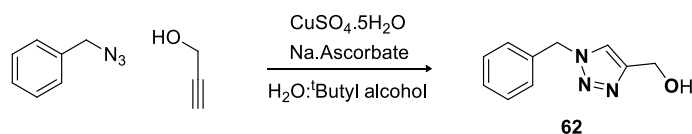
130.7, 129.0, 128.7, 128.5, 127.7, 127.6, 126.8, 126.6, 126.5, 124.2, 123.2, 61.9, 48.9, 45.1, 38.9, 28.6, 28.6, 21.8, 20.2, 13.9



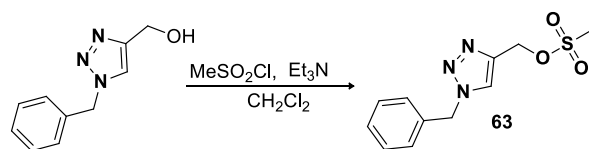
**58:** *1-(tert-butylcarbamoyl)-2-(2,2-diphenylethyl)-3-oxo-4-(4-*

*(trifluoromethyl)phenyl)isoindolin-5-yl acetate 81C* To a stirring solution of acetic anhydride (100  $\mu$ L, 1.2mmol) in pyridine (2 mL) was added compound **43** (77 mg, 0.13 mmol) and stirred at room temperature overnight. After disappearance of starting material, the reaction mixture was diluted with ethyl acetate (3 mL) and anhydrous CuSO<sub>4</sub> solution was added to wash the pyridine (3 x 3 mL). The organic portions were combined, washed with brine and dried over Na<sub>2</sub>SO<sub>4</sub>. The crude product was purified by column chromatography (35% EtOAc-hexanes) to give a white solid (78mg, 94%) <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>)  $\delta$  8.61 (s, 1H), 7.82 (d,  $J$  = 8.53 Hz, 2H), 7.67 (d,  $J$  = 8.28 Hz, 1H), 7.43 - 7.53 (m, 3H), 7.28 - 7.42 (m, 8H), 7.14 - 7.28 (m, 2H), 5.07 (s, 1H), 4.54 - 4.66 (m, 1H), 4.41 - 4.52 (m, 1H), 3.42 (dd,  $J$  = 6.65, 14.43 Hz, 1H), 1.98 (s, 3H), 1.40 (s, 9H) <sup>13</sup>C NMR (101 MHz, DMSO-d<sub>6</sub>)  $\delta$  169.0, 166.4, 165.2, 148.1, 142.5, 141.3, 140.3, 136.1, 130.7, 130.5, 129.2, 128.8, 128.5, 127.8, 127.6, 126.8, 126.6, 126.5, 124.2, 123.1, 61.8, 51.1, 49.0, 44.9, 28.3, 20.2

### 5.1.4 Nucleoside Derivatives

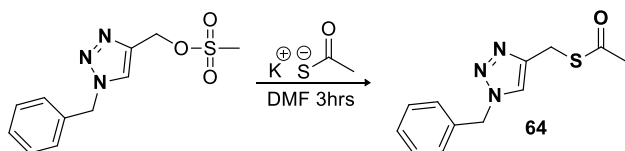


**62:** *(1-benzyl-1H-1,2,3-triazol-4-yl)methanol*: To a solution of benzyl azide (2.66 g, 20 mmol) and propargyl alcohol (1.28 mL, 22 mmol) in <sup>t</sup>BuOH:H<sub>2</sub>O (1:1, 10mL) at room temperature, sodium ascorbate (1.5844 g, 8 mmol) and CuSO<sub>4</sub>.5H<sub>2</sub>O (1.0 g, 2 mmol) were added and the resulting yellow mixture was stirred vigorously overnight. The reaction mixture was diluted and extracted with EtOAc (10 mL). The organic layer was dried over MgSO<sub>4</sub> and the solvents were evaporated under reduced pressure. The product was purified by column chromatography (2% MeOH-EtOAc) to give an amorphous solid (2.51 g, 66%). <sup>1</sup>H NMR (500 MHz, CHLOROFORM-d) δ 7.44 (bs, *J* = 3.18 Hz, 1H), 7.32 - 7.36 (m, 3H), 7.25 (dd, *J* = 1.10, 2.81, 4.89 Hz, 2H), 5.47 (bs, *J* = 3.18 Hz, 2H), 4.72 (bs, *J* = 2.69 Hz, 2H), 3.59 (bs, *J* = 2.20 Hz, 1H) <sup>13</sup>C NMR (126 MHz, CHLOROFORM-d) δ 148.2, 134.5, 129.1, 128.8, 128.1, 121.7, 56.2, 54.2

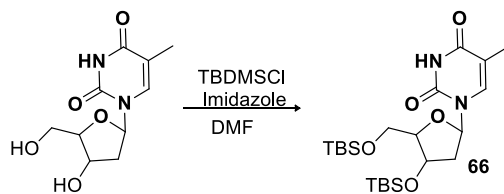




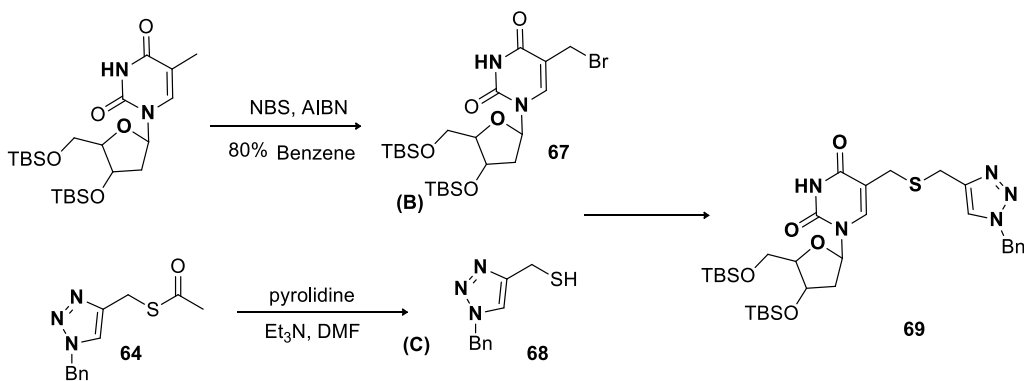
**63:** *(1-benzyl-1H-1,2,3-triazol-4-yl)methyl methanesulfonate*: Under argon, Et<sub>3</sub>N (0.55 mL, 3.9 mmol) and methanesulfonyl chloride (0.222 mL, 2.87 mmol) were added successively to a solution of 1-Benzyl-4-hydroxymethyl-1H-1,2,3-triazole **62** (500mg, 2.64mmol) in dry CH<sub>2</sub>Cl<sub>2</sub> (10 mL) at -30° C. After stirring for 30minutes TLC showed no more starting material. The mixture was poured into a mixture of water and CH<sub>2</sub>Cl<sub>2</sub> (and the separated organic layer was washed with NaHCO<sub>3</sub>, dilute HCl and brine, dried over MgSO<sub>4</sub> and rotavap to give a white solid (0.5873g, 83%) <sup>1</sup>H NMR (500 MHz, CHLOROFORM-d) δ 7.63 (br. s., 1H), 7.35 (br. s., 3H), 7.26 (br. s., 2H), 5.52 (br. s., 2H), 5.31 (br. s., 2H), 2.98 (br. s., 3H) <sup>13</sup>C NMR (126 MHz, CHLOROFORM-d) δ 141.3, 134.1, 129.1, 128.2, 124.2, 62.5, 54.3, 38.2, 31.1



**64:** *Ethanethioic acid, S-[(1-Benzyl-1H-1,2,3-triazol-4-yl)methyl] ester*: To a stirring solution of [1-(Benzyl)-4-<sup>[340]</sup>-methyl]-1H-1,2,3-triazole (400 mg, 1.49 mmol) in DMF (3 mL) under argon was added potassium thioacetate (195.69 mg, 1.713 mmol) and allowed to stir for 15minutes. The reaction mixture was dissolved in 20 mL CH<sub>2</sub>Cl<sub>2</sub> and washed with water (15mL x 3), the organic layer was dried with MgSO<sub>4</sub>, rotavap and purified by column (70% EtOAc/Hexanes) chromatography to give a white solid (133.3 mg, 54%); <sup>1</sup>H NMR (500 MHz, CHLOROFORM-d) δ 7.38 (br. s., 1H), 7.34 (br. s., 3H), 7.23 (d, *J* = 3.91 Hz, 2H), 5.45 (br. s., 2H), 4.13 (s, 2H), 2.30 (s, 3H) <sup>13</sup>C NMR (126 MHz, CHLOROFORM-d) δ 195.2, 144.7, 134.5, 129.1, 128.7, 128.0, 121.7, 54.1, 30.3, 24.0



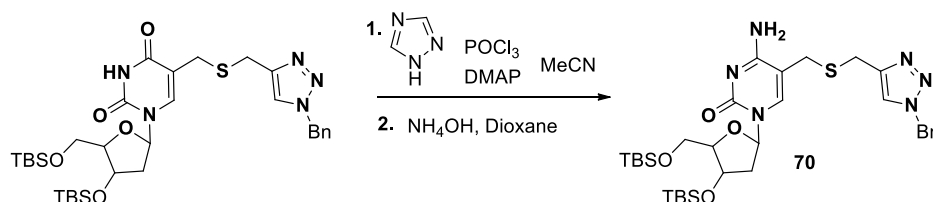
**66:** *1-(4-((tert-butyldimethylsilyl)oxy)-5-((tert-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)-5-methylpyrimidine-2,4(1H,3H)-dione*: A solution of thymidine (5 g, 20.7 mmol), imidazole (5.6 g, 82.8 mmol) and TBDMSCl (6.85 g, 45.5 mmol) in DMF (30 mL) was stirred under an argon atmosphere for 2 h at 50°C. The resultant mixture was then diluted with ethyl acetate (80 mL) and washed with saturated ammonium chloride solution (3 x 50 mL) and brine (3x 50 mL). The combined organic fractions were then dried over Na<sub>2</sub>SO<sub>4</sub> and solvent removed in vacuo to give 3', 5'-bis (O-tert-butyldimethylsilyl)-2'-deoxyuridine as a white solid (8.98 g, 92%). *R*<sub>f</sub> = 0.64 (MeOH/EtOAc, 2%); <sup>1</sup>H NMR (500 MHz, CHLOROFORM-*d*) δ 9.88 (br. s., 1H), 7.43 (d, *J* = 1.47 Hz, 1H), 6.27 - 6.35 (m, 1H), 4.36 (dd, *J* = 2.69, 5.62 Hz, 1H), 3.89 (d, *J* = 2.45 Hz, 1H), 3.67 - 3.85 (m, 2H), 1.92 - 2.28 (m, 2H), 1.87 (d, *J* = 0.98 Hz, 3H), 0.81 - 0.93 (m, 18H), -0.29 - 0.34 (m, 12 h) <sup>13</sup>C NMR (126 MHz, CHLOROFORM-*d*) δ 164.3, 150.6, 135.4, 110.8, 87.7, 84.7, 72.2, 62.9, 41.4, 25.8, 18.4, 12.5, -5.1



**67, 68 & 69: B)** 3', 5'-Bis (*O*-*tert*-butyldimethylsilyl)-2'-deoxyuridine (400 mg, 0.85mmol), *N*-bromosuccinimide(211.79 mg, 1.19 mmol) and azoisobutyronitrile (12.56 mg, 0.21 mmol) were dissolved in benzene (10 mL) under an inert atmosphere. This solution was then heated under reflux for 40 min at 80 °C. The particulates were removed by filtration (filter paper) while the reaction mixture was still hot. The filtrate was then evaporated *in vacuo* and the resulting crude oily product, 5-Bromo-3',5'-bis(*O*-*tert*-butyldimethylsilyl)-2'-deoxyuridine (B) was used without further purification

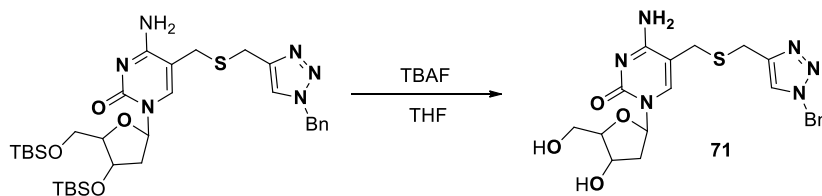
**C)**5-((((1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl)thio)methyl)-1-(4-((*tert*-butyldimethylsilyl)oxy)-5-(((*tert*-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)pyrimidine-2,4(1*H*,3*H*)-dione To a flame dried 10 mL three neck flask containing S-[(1-Benzyl-1*H*-1,2,3-triazol-4-yl)methyl] ester (A) (100 mg, 0.40 mmol) dissolved in anhydrous DMF (2 mL) was added pyrrolidine (36  $\mu$ L, 0.40 mmol) via syringe under argon. After stirring and monitoring with TLC for the disappearance of starting material, triethylamine (52  $\mu$ L, 0.38 mmol) was added and the resultant solution cannulated to a solution of 5-Bromo-3',5'-bis(*O*-*tert*-butyldimethylsilyl)-2'-deoxyuridine (B) in DMF (2 mL). The resulting solution stirred at 75° C for 2 h. The mixture was then diluted with CH<sub>2</sub>Cl<sub>2</sub>, washed with water (3 x 100 mL), NaHCO<sub>3</sub>, and brine and dried with MgSO<sub>4</sub>. The crude product was purified by column chromatography to give an off white solid (302 mg, 45% over two steps); <sup>1</sup>H NMR (500 MHz, CHLOROFORM-*d*)  $\delta$  8.76 (br. s., 1H), 7.80 (br. s., 1H), 7.53 (br. s., 1H), 7.35 (br. s., 3H), 7.23 - 7.30 (m, 2H), 6.32 (br. s., 1H), 5.49 (br. s., 2H), 4.43 (br. s., 1H), 3.76 - 3.97 (m, 3H), 3.73 (br. s.,

2H), 3.37 (d,  $J = 9.78$  Hz, 2H), 2.25 (br. s., 2H), 0.77 - 1.06 (m, 18H), -0.18 - 0.17 (m, 12 h)  $^{13}\text{C}$  NMR (126 MHz, CHLOROFORM-d)  $\delta$  162.7, 149.9, 145.7, 138.2, 134.6, 129.1, 128.7, 128.1, 122.0, 111.1, 88.0, 85.1, 72.4, 63.3, 54.2, 40.7, 26.8, 25.9, 25.7, 25.5, 18.2, -5.1

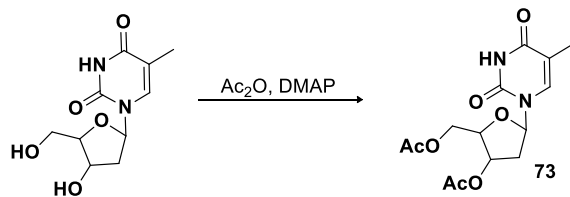


**70:** *4-amino-5-(((1-benzyl-1H-1,2,3-triazol-4-yl)methyl)thio)methyl)-1-(4-((tert-butyl)dimethylsilyl)oxy)-5-(((tert-butyl)dimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)pyrimidin-2(1H)-one* 1, 2, 4-triazole (2.229 g, 21.8mmol) was suspended in dry MeCN (25 mL), 5-(((1-benzyl-1H-1,2,3-triazol-4-yl)methyl)thio)methyl)-1-(4-((tert-butyl)dimethylsilyl)oxy)-5-(((tert-butyl)dimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)pyrimidine-2,4(1H,3H)-dione (1 g, 1.48 mmol) dissolved in MeCN (5 mL) and DMAP (3.97 g, 22.02 mmol) was added and stirred until dissolved. POCl<sub>3</sub> (0.688 mL, 7.62 mmol) was added drop wise and allowed to stir for 2 h. The reaction was diluted with EtOAc (50 mL) and quenched with NaHCO<sub>3</sub>, washed with brine and dried with Na<sub>2</sub>SO<sub>4</sub>. Solvent was removed under low pressure and used for the next step without further purification. The resultant mixture was dissolved in dioxane (20 mL) and NH<sub>4</sub>OH (25 mL) was added and stirred overnight. Excess solvent was removed and resultant purified by column chromatography to give a white solid (336 mg, 50% over two steps);  $^1\text{H}$  NMR (500 MHz, CHLOROFORM-d)  $\delta$  8.18 (br. s., 2H), 7.81 (br. s., 1H), 7.35 (d,  $J = 3.91$  Hz, 2H), 7.26 (br. s., 2H), 6.32 (br. s., 1H), 5.48 (br. s., 2H), 4.37 (br. s., 1H), 3.95 (br. s.,

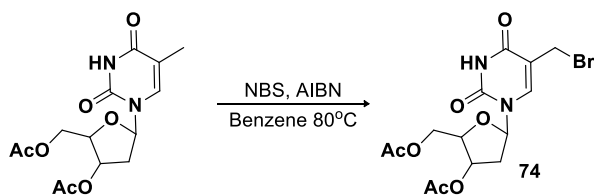
1H), 3.77 - 3.87 (m, 4H), 3.68 (br. s., 2H), 3.59 (br. s., 2H), 2.40 (d,  $J = 12.23$  Hz, 1H), 1.97 - 2.07 (m, 1H), 0.87 (d,  $J = 13.21$  Hz, 18H), 0.06 (d,  $J = 17.12$  Hz, 12 h)  $^{13}\text{C}$  NMR (126 MHz, CHLOROFORM-d)  $\delta$  164.6, 155.7, 145.7, 140.8, 134.3, 129.2, 128.9, 128.1, 121.4, 101.6, 87.9, 86.2, 72.1, 63.1, 54.3, 41.8, 30.9, 29.3, 26.0, 25.7, 17.9, -5.1



**71:** *4-amino-5-(((1-benzyl-1H-1,2,3-triazol-4-yl)methylthio)methyl)-1-((2R,4S,5R)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)pyrimidin-2(1H)-one* To a solution of 4-amino-5-(((1-benzyl-1H-1,2,3-triazol-4-yl)methylthio)methyl)-1-(4-((tert-butyldimethylsilyl)oxy)-5-(((tert-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)pyrimidin-2(1H)-one (235mg, 0.347mmol) in dry THF (3.5mL) was added TBAF (1M solution in THF, 0.875mL, 0.9022mmol). The mixture was stirred for 3 h at room temperature. The resulting mixture was evaporated and purified by silica gel (20% MeOH/EtOAc) to give a white solid (89mg, 58%)  $^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$  8.03 (s, 2H), 7.77 (s, 2H), 7.25 - 7.38 (m, 4H), 6.14 (t,  $J = 6.60$  Hz, 1H), 5.54 (s, 2H), 5.18 (d,  $J = 4.40$  Hz, 1H), 4.81 - 5.07 (m, 1H), 4.18 (br. s., 1H), 3.74 (d,  $J = 2.93$  Hz, 2H), 3.63 (s, 2H), 3.47 - 3.57 (m, 1H), 3.45 (d,  $J = 4.89$  Hz, 2H), 3.33 (br. s., 1H), 2.05 - 2.13 (m, 1H), 1.96 (td,  $J = 6.54, 12.84$  Hz, 1H)  $^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ )  $\delta$  164.4, 155.1, 145.1, 140.5, 136.7, 129.2, 128.6, 128.3, 123.5, 101.8, 87.6, 85.3, 70.9, 62.0, 53.3, 40.6, 28.5, 24.4

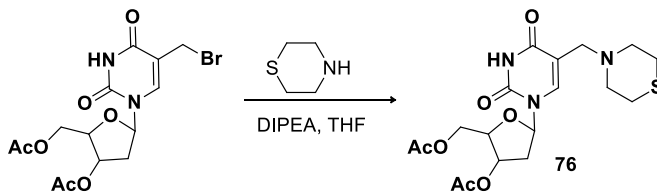


**73:** (3-acetoxy-5-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)tetrahydrofuran-2-yl)methyl acetate: Thymidine (1.0 g, 4.10 mmol), DMAP (spatula tip) and Ac<sub>2</sub>O (5 mL) were stirred under nitrogen until all the solid disappeared. Excess Ac<sub>2</sub>O was evaporated in vacuo and the residue was dissolved in 100 mL of DCM. The organic layer was washed with 2 x 100 mL of water, 100 mL of NaHCO<sub>3</sub>, and brine. The organic layer was dried over MgSO<sub>4</sub>, filtered and evaporated. The acetate **73** was obtained as a white solid yield (1.07 g, 81%) <sup>1</sup>H NMR (500 MHz, CHLOROFORM-d) Shift 9.55 (br. s., 1H), 6.32 (dd, J = 5.75, 8.44 Hz, 1H), 5.16 - 5.22 (m, 1H), 4.28 - 4.39 (m, 2H), 4.18 - 4.25 (m, 1H), 2.39 - 2.48 (m, 1H), 2.10 (dd, J = 0.61, 6.48 Hz, 7H), 1.92 (s, 3H) <sup>13</sup>C NMR (126 MHz, CHLOROFORM-d) δ 170.7, 163.2, 149.8, 134.9, 111.6, 84.8, 82.1, 74.1, 63.9, 36.6, 20.8, 11.0

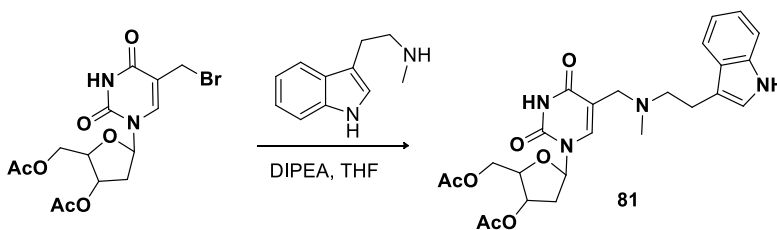


**74:** (3-acetoxy-5-(5-(bromomethyl)-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)tetrahydrofuran-2-yl)methyl acetate: After addition of NBS (166 mg, 0.93 mmol) and AIBN (8.3 mg, 0.05 mmol) to the refluxing solution of **73** (250 mg, 0.765 mmol) in 200 mL of dry benzene stirring with refluxing was kept for 5 h with monitoring the progress of bromination by tlc (EtOAc:Hexane 70%). Hot filtration of benzene solution to remove

solid particles through ordinary filter paper followed by evaporation of benzene gave crude monobromo- intermediate which was used directly for the next step.

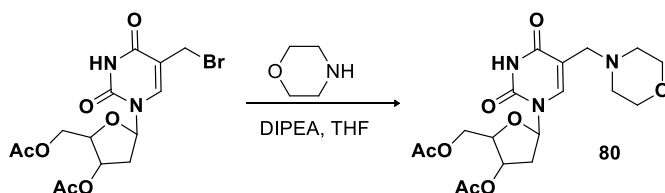


**76:** *(3-acetoxy-5-(2,4-dioxo-5-(thiomorpholinomethyl)-3,4-dihydropyrimidin-1(2H)-yl)tetrahydrofuran-2-yl)methyl acetate*: To a solution of crude monobromo-intermediate **74** and thiomorpholine (0.45 mL, 4.5 mmol) in dry THF was added diisopropylethylamine (174  $\mu$ L, 1 mmol) and refluxed for 3 h. The reaction was cool and solvent removed under reduced pressure. Purification was by column chromatography using methanol-Ethyl acetate (20%) to give a white foam (275 mg, 85% over two steps).  $^1\text{H}$  NMR (400 MHz, DMSO- $d_6$ )  $\delta$  11.44 (br. s., 1H), 7.53 (s, 1H), 6.19 (t,  $J = 7.03$  Hz, 1H), 5.22 (br. s., 1H), 4.18 - 4.33 (m, 3H), 3.12 - 3.29 (m, 2H), 2.62 (d,  $J = 9.54$  Hz, 8H), 2.28 - 2.36 (m, 2H), 2.09 (s, 6H)  $^{13}\text{C}$  NMR (101 MHz, DMSO- $d_6$ )  $\delta$  170.1, 163.2, 150.2, 138.1, 109.8, 84.5, 81.3, 74.1, 63.8, 54.1, 53.3, 35.9, 27.1, 20.8



**81:** *5-(5-(((2-(1H-indol-3-yl)ethyl)(methyl)amino)methyl)-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)-2-(acetoxymethyl)tetrahydrofuran-3-yl acetate*: To a solution of crude **74** monobromo-intermediate and N-methyltryptamine (786 mg, 4.5 mmol) in dry THF was

added diisopropylethylamine (174  $\mu$ L, 1 mmol) and refluxed for 3 h. The reaction was cool and solvent removed under reduced pressure. Purification was by column chromatography using methanol-Ethyl acetate (20%) to give a white foam (294 mg, 77% over two steps)  $^1\text{H}$  NMR (400 MHz, DMSO- $d_6$ )  $\delta$  11.46 (s, 1H), 10.78 (br. s., 1H), 7.56 (s, 1H), 7.49 (d,  $J = 7.78$  Hz, 1H), 7.34 (d,  $J = 8.03$  Hz, 1H), 7.14 (d,  $J = 1.51$  Hz, 1H), 7.07 (t,  $J = 7.53$  Hz, 1H), 6.94 - 7.01 (m, 1H), 6.21 (t,  $J = 7.28$  Hz, 1H), 5.22 (d,  $J = 4.52$  Hz, 1H), 4.22 (s, 3H), 3.25 - 3.33 (m, 2H), 2.79 - 2.90 (m, 2H), 2.59 - 2.69 (m, 2H), 2.31 - 2.39 (m, 2H), 2.23 - 2.29 (m, 3H), 2.08 (d,  $J = 10.04$  Hz, 6H)  $^{13}\text{C}$  NMR (101 MHz, DMSO- $d_6$ )  $\delta$  170.1, 163.1, 150.2, 137.7, 136.2, 127.2, 122.4, 120.8, 118.1, 112.5, 111.3, 111.1, 84.4, 81.3, 74.2, 63.8, 57.5, 52.1, 41.6, 35.9, 22.8, 20.8, 20.5



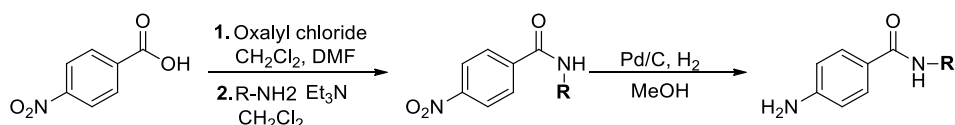
**80:** (3-acetoxy-5-(5-(morpholinomethyl)-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-

yl)tetrahydrofuran-2-yl)methyl acetate: To a solution of crude monobromo-intermediate **74** and morpholine (0.3 mL, 4.5 mmol) in dry THF was added diisopropylethylamine (174  $\mu$ L, 1 mmol) and refluxed for 3h. The reaction was cool and solvent removed under reduced pressure. Purification was by column chromatography using methanol-Ethyl acetate (20%) to give a white foam (203 mg, 64% over two steps)  $^1\text{H}$  NMR (400 MHz, DMSO- $d_6$ )  $\delta$  11.45 (s, 1H), 7.56 (s, 1H), 6.19 (t,  $J = 7.15$  Hz, 1H), 5.23 (d,  $J = 6.27$  Hz, 1H), 4.12 - 4.47 (m, 4H), 3.56 (br. s., 5H), 3.14 - 3.23 (m, 2H), 2.34 - 2.42 (m, 4H), 2.09



(d,  $J = 1.25$  Hz, 6H)  $^{13}\text{C}$  NMR (101 MHz, DMSO- $d_6$ )  $\delta$  172.0, 170.5, 163.1, 150.2, 138.3, 109.6, 84.6, 81.3, 74.1, 66.1, 63.8, 53.0, 52.8, 35.9, 20.7, 20.6

### 5.1.5 Procainamide Derivatives

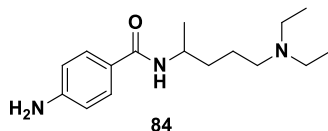


#### Synthesis of 3-nitrobenzamide derivatives.

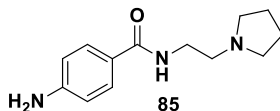
To an oven dried 25 mL round bottom flask was added p-nitrobenzoic acid (103 mg, 0.62mmol) and DCM (1.1 mL) under argon atmosphere. Oxalyl chloride (0.5 mL, 6.2mmol) was added drop wise followed by DMF (2 drops) resulting in vigorous bubbling. The mixture was stirred for 1 h after which no precipitate was remained and the starting material was not observed by TLC. The solvent was evaporated under reduced pressure to give a white solid, which was taken up in DCM (3 mL) without further purification and added immediately to the appropriate amine (1.2 equivalents) and triethylamine (2 equivalents) stirred overnight at room temperature.

**Synthesis of 3-aminobenzamide derivatives.** A solution of 3-nitro-benzamide (1 equivalent) in methanol was treated with 5% Pd/C (5 % w/w). The reaction was subjected to hydrogenation under 1atm hydrogen gas pressure at room temperature, using a balloon, and the reaction mixture stirred overnight. After completion of the reaction, the mixture was filtered through Celite pad (4

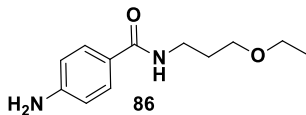
g) and concentrated by removal of solvent under reduced pressure. The resulting residue was purified by silica gel column chromatography.



**84:** *4-amino-N-(5-(diethylamino)pentan-2-yl)benzamide*: Yellow solid (91%)  $^1\text{H}$  NMR (500 MHz, CHLOROFORM- $d$ )  $\delta$  7.47 - 7.60 (m, 2H), 6.55 (d,  $J = 7.83$  Hz, 2H), 6.29 (d,  $J = 7.83$  Hz, 1H), 4.09 (br. s., 1H), 3.04 - 3.45 (m, 2H), 2.44 (q,  $J = 7.17$  Hz, 4H), 2.35 (br. s., 2H), 1.47 (br. s., 4H), 1.14 (d,  $J = 6.36$  Hz, 3H), 0.93 (t,  $J = 7.09$  Hz, 6H)  $^{13}\text{C}$  NMR (126 MHz, CHLOROFORM- $d$ )  $\delta$  166.8, 149.6, 128.4, 124.2, 113.9, 52.6, 46.7, 45.4, 34.9, 23.5, 20.9, 11.2

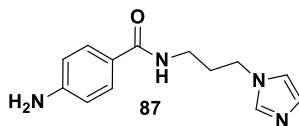


**85:** *4-amino-N-(2-(pyrrolidin-1-yl)ethyl)benzamide*: White solid (90.4%)  $^1\text{H}$  NMR (500 MHz, CHLOROFORM- $d$ )  $\delta$  7.59 (d,  $J = 7.83$  Hz, 2H), 6.70 (br. s., 1H), 6.62 (d,  $J = 8.31$  Hz, 2H), 4.04 (br. s., 2H), 3.50 (q,  $J = 5.54$  Hz, 2H), 2.66 (t,  $J = 6.11$  Hz, 2H), 2.52 (br. s., 4H), 1.76 (br. s., 4H)  $^{13}\text{C}$  NMR (126 MHz, CHLOROFORM- $d$ )  $\delta$  165.4, 149.4, 140.3, 128.2, 123.7, 54.3, 53.8, 38.7, 23.5

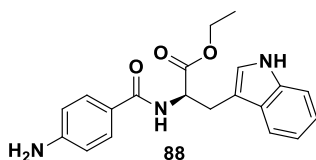


**86:** *4-amino-N-(3-ethoxypropyl)benzamide*: White solid (81%)  $^1\text{H}$  NMR (500 MHz, CHLOROFORM- $d$ )  $\delta$  7.60 (d,  $J = 8.31$  Hz, 1H), 6.91 (br. s., 2H), 6.65 (d,  $J = 8.31$  Hz, 2H), 3.94 (br. s., 2H), 3.47 - 3.63 (m, 6H), 1.82 - 1.90 (m, 2H), 1.24 (t,  $J = 6.85$  Hz, 3H)

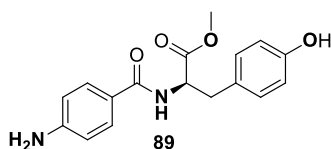
$^{13}\text{C}$  NMR (126 MHz, CHLOROFORM- $d$ )  $\delta$  166.9, 149.3, 128.5, 124.5, 114.1, 70.6, 66.6, 39.3, 28.9, 15.4



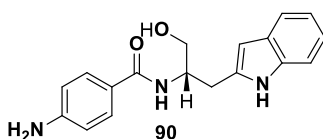
**87:** *N*-(3-(1*H*-imidazol-1-yl)propyl)-4-aminobenzamide: White solid (85% over 2 steps)  $^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$  8.03 (br. s., 1H), 7.60 - 7.66 (m, 1H), 7.54 (d,  $J$  = 8.31 Hz, 2H), 7.18 (s, 1H), 6.86 (s, 2H), 6.51 (d,  $J$  = 8.31 Hz, 1H), 5.59 (br. s., 2H), 3.96 (t,  $J$  = 6.60 Hz, 2H), 3.15 (d,  $J$  = 5.87 Hz, 2H), 1.83 - 1.94 (m, 2H)  $^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ )  $\delta$  166.8, 152.0, 137.8, 129.1, 128.8, 121.6, 119.8, 112.9, 44.3, 36.7, 31.5



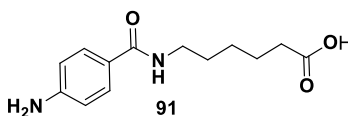
**88:** (*R*)-ethyl 2-(4-aminobenzamido)-3-(1*H*-indol-3-yl)propanoate White solid (81%)  $^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$  10.82 (br. s., 1H), 8.05 - 8.29 (m, 1H), 7.46 - 7.69 (m, 3H), 7.30 (d,  $J$  = 7.83 Hz, 1H), 7.17 (d,  $J$  = 1.47 Hz, 1H), 6.89 - 7.07 (m, 2H), 6.50 (d,  $J$  = 8.31 Hz, 2H), 5.64 (s, 2H), 4.50 - 4.68 (m, 1H), 4.03 (q,  $J$  = 7.17 Hz, 2H), 3.59 (s, 1H), 3.32 (s, 2H  $\text{H}_2\text{O}$ ), 3.06 - 3.22 (m, 1H), 1.08 (t,  $J$  = 7.09 Hz, 3H)  $^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ )  $\delta$  173.3, 166.8, 152.3, 136.7, 129.5, 127.6, 124.1, 121.4, 120.7, 118.8, 118.5, 112.8, 111.9, 110.5, 60.7, 54.1, 27.1, 14.4



- 89:** *(R)*-methyl 2-(4-aminobenzamido)-3-(4-hydroxyphenyl)propanoate White solids (73%)  
 $^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$  9.18 (s, 1H), 8.23 (d,  $J = 7.83$  Hz, 1H), 7.44 - 7.59 (m, 2H), 7.04 (d,  $J = 7.83$  Hz, 2H), 6.62 (d,  $J = 8.31$  Hz, 2H), 6.50 (d,  $J = 8.31$  Hz, 2H), 5.57 - 5.73 (m, 2H), 4.38 - 4.53 (m, 1H), 3.48 - 3.66 (m, 1H), 3.32 (s, 3H), 2.86 - 3.02 (m, 1H)  
 $^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ )  $\delta$  173.3, 166.8, 156.3, 152.3, 130.4, 129.5, 128.3, 120.6, 115.4, 112.8, 55.0, 52.2, 36.0

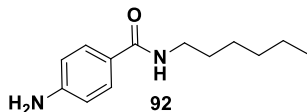


- 90:** *(R)*-4-amino-N-(1-hydroxy-3-(1H-indol-2-yl)propan-2-yl)benzamide White solid (74%)  
 $^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$  10.79 (br. s., 1H), 7.95 (d,  $J = 7.34$  Hz, 1H), 7.60 (d,  $J = 4.89$  Hz, 2H), 7.54 (d,  $J = 6.85$  Hz, 1H), 7.30 (d,  $J = 7.34$  Hz, 1H), 7.14 (br. s., 1H), 6.90 - 7.07 (m, 2H), 6.50 (br. s., 2H), 5.94 (br. s., 2H), 5.58 (br. s., 1H), 4.53 (br. s., 1H), 4.18 (br. s., 2H), 3.33 (br. s., 1H), 3.00 (br. s., 1H)  $^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ )  $\delta$  166.7, 153.9, 152.0, 136.6, 131.6, 129.2, 127.8, 123.7, 121.3, 118.7, 116.2, 113.0, 111.8, 65.4, 49.5, 27.6



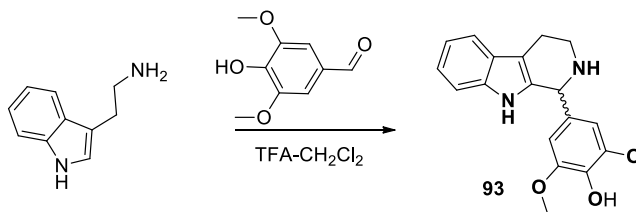
- 91:** 6-(4-aminobenzamido)hexanoic acid: White solid (80%)  $^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$  12.00 (br. s., 1H), 7.94 (br. s., 1H), 7.53 (d,  $J = 8.31$  Hz, 2H), 6.50 (d,  $J = 8.31$  Hz, 2H), 5.55 (br. s., 2H), 3.15 (d,  $J = 5.87$  Hz, 2H), 2.18 (t,  $J = 7.09$  Hz, 2H), 1.40 - 1.53 (m, 4H),

1.27 (d,  $J = 6.85$  Hz, 2H)  $^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ )  $\delta$  174.7, 166.5, 152.4, 129.0, 121.9, 112.9, 34.1, 29.6, 26.5, 24.8



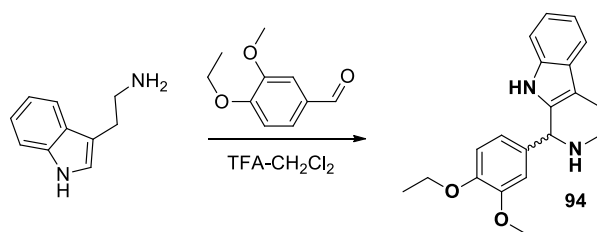
**92:** *4-amino-N-hexylbenzamide* White solid (56%)  $^1\text{H}$  NMR (500 MHz, CHLOROFORM- $d$ )  $\delta$  7.58 (d,  $J = 8.80$  Hz, 1H), 6.63 (d,  $J = 8.31$  Hz, 2H), 6.09 (br. s., 2H), 3.99 (br. s., 2H), 3.39 (q,  $J = 6.68$  Hz, 2H), 1.50 - 1.61 (m, 2H), 1.18 - 1.41 (m, 6H), 0.86 (t,  $J = 6.85$  Hz, 3H)  $^{13}\text{C}$  NMR (126 MHz, CHLOROFORM- $d$ )  $\delta$  167.2, 149.4, 128.5, 124.3, 114.1, 39.9, 31.8, 29.8, 29.0, 27.0, 22.6, 14.1

### 5.1.6 Beta-Carbolines



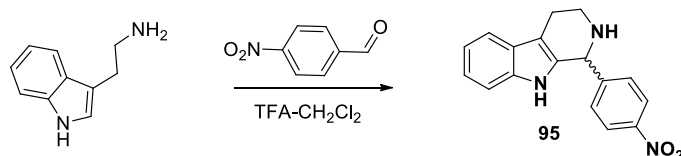
**93:** *2,6-dimethoxy-4-(2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indol-1-yl)phenol* Trifluoroacetic acid (0.229 mL, 3.0 mmol) was added to a solution of tryptamine (320 mg, 2mmol) and 4-hydroxy-3,5-dimethoxybenzaldehyde (437 mg, 2.4 mmol) in  $\text{CH}_2\text{Cl}_2$  at  $0^\circ\text{C}$  and stirred at room temperature for 24 h and then evaporated. The obtained residue was triturated with a 5%  $\text{K}_2\text{CO}_3$  aqueous solution (8 mL) and extracted with  $\text{CH}_2\text{Cl}_2$ . The organic layer was dried over magnesium sulfate and evaporated to dryness under reduced pressure. The

crude product was purified by column chromatography using methanol-EtOAc (10%) to give a yellow solid (302 mg, 46.2%)  $^1\text{H}$  NMR (400 MHz, DMSO- $d_6$ )  $\delta$  10.30 (s, 1H), 7.43 (d,  $J = 7.53$  Hz, 1H), 7.27 (d,  $J = 7.78$  Hz, 1H), 6.94 - 7.06 (m, 3H), 6.63 (s, 2H), 5.02 (s, 1H), 3.73 (s, 6H), 3.19 - 3.26 (m, 2H), 2.97 (td,  $J = 3.86, 8.09$  Hz, 1H), 2.79 (d,  $J = 7.28$  Hz, 1H), 2.64 - 2.71 (m, 1H)  $^{13}\text{C}$  NMR (101 MHz, DMSO- $d_6$ )  $\delta$  147.8, 136.0, 135.9, 135.0, 132.8, 126.9, 120.4, 118.1, 117.5, 111.1, 105.9, 57.4, 56.0, 42.2, 22.2

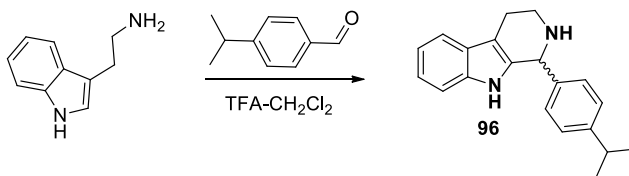


**94:** *1-(4-ethoxy-3-methoxyphenyl)-2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indole* Trifluoroacetic acid (0.229 mL, 3.0 mmol) was added to a solution of tryptamine (320 mg, 2 mmol) and 4-ethoxy-3-methoxybenzaldehyde (432 mg, 2.4 mmol) in  $\text{CH}_2\text{Cl}_2$  at  $0^\circ\text{C}$  and stirred at room temperature for 24h and then evaporated. The obtained residue was triturated with a 5%  $\text{K}_2\text{CO}_3$  aqueous solution (8 mL) and extracted with  $\text{CH}_2\text{Cl}_2$ . The organic layer was dried over magnesium sulfate and evaporated to dryness under reduced pressure. The crude product was purified by column chromatography using methanol-EtOAc (10%) to give a yellow solid (321 mg, 49.9%).  $^1\text{H}$  NMR (400 MHz, DMSO- $d_6$ )  $\delta$  10.43 (s, 1H), 7.45 (d,  $J = 7.53$  Hz, 1H), 7.29 (d,  $J = 7.78$  Hz, 1H), 6.97 - 7.08 (m, 3H), 6.91 (d,  $J = 8.03$  Hz, 1H), 6.78 (dd,  $J = 1.88, 8.16$  Hz, 1H), 5.10 (s, 1H), 4.01 (q,  $J = 6.94$  Hz, 2H), 3.75 (s, 3H), 3.11 - 3.22 (m, 1H), 2.99 (ddd,  $J = 4.89, 7.22, 12.23$  Hz, 1H), 2.67 - 2.85 (m, 2H), 1.35 (t,  $J = 6.90$  Hz, 3H)  $^{13}\text{C}$  NMR (101 MHz, DMSO- $d_6$ )  $\delta$  148.9, 147.4, 136.0, 135.5,

135.2, 126.9, 120.5, 118.2, 117.5, 112.6, 112.5, 111.1, 108.0, 63.8, 56.6, 55.4, 41.6, 22.1, 14.8

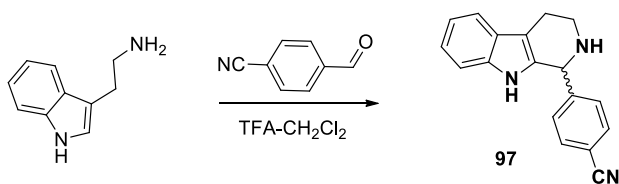


**95:** *1-(4-nitrophenyl)-2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indole* Trifluoroacetic acid (0.1 mL, 1.5 Mmol) was added to a solution of tryptamine (160 mg, 1mmol) and 4-Nitrobenzaldehyde (181 mg, 1.2 mmol) in CH<sub>2</sub>Cl<sub>2</sub> at 0° C and stirred at room temperature for 24 h and then evaporated. The obtained residue was triturated with a 5% K<sub>2</sub>CO<sub>3</sub> aqueous solution (3 mL) and extracted with CH<sub>2</sub>Cl<sub>2</sub>. The organic layer was dried over magnesium sulfate and evaporated to dryness under reduced pressure. The crude product was purified by column chromatography using methanol-EtOAc (10%) to give a yellow solid (181 mg, 61.7%) <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>) δ 10.53 (br. s., 1H), 8.12 - 8.37 (m, 2H), 7.60 (d, *J* = 7.53 Hz, 2H), 7.46 (d, *J* = 7.78 Hz, 1H), 7.26 (d, *J* = 7.53 Hz, 1H), 6.96 - 7.09 (m, 2H), 5.26 (s, 1H), 3.34 (br. s., 1H), 2.98 - 3.06 (m, 2H), 2.69 - 2.78 (m, 2H) <sup>13</sup>C NMR (101 MHz, DMSO-d<sub>6</sub>) δ 151.0, 146.7, 136.0, 134.1, 129.7, 126.8, 123.3, 120.8, 118.3, 117.7, 111.1, 108.6, 55.7, 41.0, 22.1



**96:** *1-(4-isopropylphenyl)-2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indole* Trifluoroacetic acid (0.229 mL, 3.0 mmol) was added to a solution of tryptamine (320 mg, 2mmol) and 4-

isopropylbenzaldehyde (355.68 mg, 2.4 mmol) in CH<sub>2</sub>Cl<sub>2</sub> at 0° C and stirred at room temperature for 24 h and then evaporated. The obtained residue was triturated with a 5% K<sub>2</sub>CO<sub>3</sub> aqueous solution (8 mL) and extracted with CH<sub>2</sub>Cl<sub>2</sub>. The organic layer was dried over magnesium sulfate and evaporated to dryness under reduced pressure. The crude product was purified by column chromatography using methanol-EtOAc (10%) to give a yellow solid ( 299.5 mg, 51.7%). <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>) δ 10.43 (s, 1H), 7.43 (d, *J* = 7.28 Hz, 1H), 7.18 - 7.27 (m, 5H), 6.93 - 7.05 (m, 2H), 5.07 (s, 1H), 3.05 - 3.13 (m, 2H), 2.85 - 2.99 (m, 2H), 2.64 - 2.80 (m, 2H), 1.19 - 1.25 (m, 6H) <sup>13</sup>C NMR (101 MHz, DMSO-d<sub>6</sub>) δ 147.3, 140.6, 135.9, 135.4, 128.4, 126.9, 126.0, 120.4, 118.1, 117.5, 111.0, 108.2, 56.3, 41.2, 33.2, 24.0, 22.3



**97:** 4-(2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indol-1-yl)benzonitrile Trifluoroacetic acid (0.1 mL, 1.5 mmol) was added to a solution of tryptamine(160 mg, 1 mmol) and 4-Cyanobenzaldehyde(157 mg, 1.2 mmol) in CH<sub>2</sub>Cl<sub>2</sub> at 0° C and stirred at room temperature for 24 h and then evaporated. The obtained residue was triturated with a 5% K<sub>2</sub>CO<sub>3</sub> aqueous solution (3 mL) and extracted with CH<sub>2</sub>Cl<sub>2</sub>. The organic layer was dried over magnesium sulfate and evaporated to dryness under reduced pressure. The crude product was purified by column chromatography using methanol-EtOAc (10%) to give a yellow solid (181 mg, 62.3%) <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>) δ 10.50 (s, 1H), 7.79 -



7.85 (m, 2H), 7.51 (d,  $J = 8.28$  Hz, 2H), 7.45 (d,  $J = 7.53$  Hz, 1H), 7.25 (d,  $J = 7.78$  Hz, 1H), 6.95 - 7.08 (m, 2H), 5.20 (s, 1H), 3.35 (br. s., 1H), 2.94 - 3.05 (m, 2H), 2.64 - 2.79 (m, 2H)  $^{13}\text{C}$  NMR (101 MHz, DMSO- $d_6$ )  $\delta$  148.9, 136.0, 134.2, 132.1, 129.4, 126.8, 120.8, 118.9, 118.3, 117.7, 111.1, 109.9, 108.6, 56.0, 41.0, 22.1

# Bibliography

## Primary Sources

## Secondary Sources

## Uncategorized References

- [1] P. A. Jones, S. B. Baylin, *Nat Rev Genet* **2002**, 3, 415-428.
- [2] aT. G. Dobzhansky, T. Dobzhansky, *Genetics and the Origin of Species*, Columbia University Press, **1937**; bJ. Huxley, M. Pigliucci, G. B. Müller, *Evolution: The Modern Synthesis: The Definitive Edition*, Mit Press, **2010**; cE. Mayr, *Systematics and the Origin of Species, from the Viewpoint of a Zoologist*, Harvard University Press, **1942**.
- [3] D. C. A. C. Gore, *F1000 Biology Reports* **2012**, 4, 1-6.
- [4] aP. Cubas, C. Vincent, E. Coen, *Nature* **1999**, 401, 157-161; bE. J. Richards, *Nat Rev Genet* **2006**, 7, 395-401; cO. Bossdorf, C. L. Richards, M. Pigliucci, *Ecology Letters* **2008**, 11, 106-115.
- [5] W. C. Hahn, R. A. Weinberg, *New England Journal of Medicine* **2002**, 347, 1593-1603.
- [6] aR. Goldschmidt, *The Material Basis of Evolution*, Yale University Press, **1940**; bC. Darwin, *On the Origin of Species, by Means of Natural Selection; Or, The Preservation of Favoured Races in the Struggle for Life*, G. Richards, **1902**.
- [7] G. Greenberg, M. M. Haraway, *Comparative Psychology: A Handbook*, Garland Pub., **1998**.
- [8] M. W. Ho, P. T. Saunders, *Beyond neo-Darwinism: an introduction to the new evolutionary paradigm*, Academic Press, **1984**.
- [9] R. J., *Scientific America* **1993**, 88-96.
- [10] T. K. Barth, A. Imhof, *Trends in Biochemical Sciences* **2010**, 35, 618-626.
- [11] aA. S. Packard, *Lamarck, the Founder of Evolution His Life and Work*, Wildhern Press, **2008**; bH. F. Osborn, *From the Greeks to Darwin: an outline of the development of the evolution idea*, Macmillan, **1902**; cR. W. Burkhardt, *The Spirit of System: Lamarck and Evolutionary Biology : Now with "Lamarck in 1995"*, Harvard University Press, **1995**.
- [12] W. C. H., *Endeavour* **1942**, 1, 18-20.
- [13] G. M. Malacinski, *Cytoplasmic organization systems*, McGraw-Hill Publishing Company, **1990**.
- [14] aR. Holliday, *Biological Reviews* **1990**, 65, 431-471; bE. Jablonka, M. J. Lamb, *Journal of Theoretical Biology* **1989**, 139, 69-83; cH. Weintraub, *Cell* **1985**, 42, 705-711.
- [15] B. Li, M. Carey, J. L. Workman, *Cell* **2007**, 128, 707-719.
- [16] K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, T. J. Richmond, *Nature* **1997**, 389, 251-260.
- [17] aT. Jenuwein, C. D. Allis, *Science* **2001**, 293, 1074-1080; bB. D. Strahl, C. D. Allis, *Nature* **2000**, 403, 41-45.
- [18] Owen-Hughes, *Biochem. Soc. Trans.* **2003**, 31, 893-905.
- [19] E. Li, *Nat Rev Genet* **2002**, 3, 662-673.
- [20] aG. G. Wilson, N. E. Murray, *Annual Review of Genetics* **1991**, 25, 585-627; bA. P. Bird, A. P. Wolffe, *Cell* **1999**, 99, 451-454.
- [21] aI. P. Ioshikhes, M. Q. Zhang, *Nat Genet* **2000**, 26, 61-63; bD. Takai, P. A. Jones, *Proceedings of the National Academy of Sciences* **2002**, 99, 3740-3745.
- [22] F. M. WATT, PL, *GENES & DEVELOPMENT* **1988** 2, 1136-1143.
- [23] aJ. Boyes, A. Bird, *Cell* **1991**, 64, 1123-1134; bB. Hendrich, A. Bird, *Molecular and Cellular Biology* **1998**, 18, 6538-6547.
- [24] S. A. Sarraf, I. Stancheva, *Molecular Cell* **2004**, 15, 595-605.
- [25] T. H. Bestor, *EMBO J.* **1992**, 11, 2611-2617.

- [26] I. S. Suetake, F ; Miyagawa, J ; et al., *JOURNAL OF BIOLOGICAL CHEMISTRY* **2004**, *279*, 27816-27823.
- [27] M. G. Goll, F. Kirpekar, K. A. Maggert, J. A. Yoder, C.-L. Hsieh, X. Zhang, K. G. Golic, S. E. Jacobsen, T. H. Bestor, *Science* **2006**, *311*, 395-398.
- [28] M. Okano, S. Xie, E. Li, *Nucleic Acids Research* **1998**, *26*, 2536-2540.
- [29] A. Portela, M. Esteller, *Nat Biotech* **2010**, *28*, 1057-1068.
- [30] A. P. Feinberg, *Nature* **2007**, *447*, 433-440.
- [31] F. Petrif, R. H. Giles, H. G. Dauwerse, J. J. Saris, R. C. M. Hennekam, M. Masuno, N. Tommerup, G.-J. B. van Ommen, R. H. Goodman, D. J. M. Peters, M. H. Breuning, *Nature* **1995**, *376*, 348-351.
- [32] H. D. Gibbons RJ, *Am J Med Genet* **2000**, *97*, 204-212.
- [33] aR. E. Amir, I. B. Van den Veyver, M. Wan, C. Q. Tran, U. Francke, H. Y. Zoghbi, *Nat Genet* **1999**, *23*, 185-188; bG. Egger, G. Liang, A. Aparicio, P. A. Jones, *Nature* **2004**, *429*, 457-463; cG.-L. Xu, T. H. Bestor, D. Bourc'his, C.-L. Hsieh, N. Tommerup, M. Bugge, M. Hulten, X. Qu, J. J. Russo, E. Viegas-Pequignot, *Nature* **1999**, *402*, 187-191; dC. M. Blanco-Betancourt, A ; Milili, M ; et al., *BLOOD* **2004**, *103*, 2683-2690.
- [34] E. L. Niemitz, M. R. DeBaun, J. Fallon, K. Murakami, H. Kugoh, M. Oshimura, A. P. Feinberg, *American journal of human genetics* **2004**, *75*, 844-849.
- [35] V. Greger, E. Passarge, W. Höpping, E. Messmer, B. Horsthemke, *Hum Genet* **1989**, *83*, 155-158.
- [36] P. A. Jones, P. W. Laird, *Nat Genet* **1999**, *21*, 163-167.
- [37] M. Esteller, *Human Molecular Genetics* **2007**, *16*, R50-R59.
- [38] G. T. Leone, L ; Voso, MT ; et al., *HAEMATOLOGICA* **2002**, *87*, 1324-1341.
- [39] X. Yang, L. Yan, N. E. Davidson, *Endocrine-Related Cancer* **2001**, *8*, 115-127.
- [40] A. Catteau, J. R. Morris, *Seminars in Cancer Biology* **2002**, *12*, 359-371.
- [41] M. Esteller, P. G. Corn, S. B. Baylin, J. G. Herman, *Cancer Research* **2001**, *61*, 3225-3229.
- [42] aH. J. Merlo A, Mao L, Lee DJ, Gabrielson E, Burger PC, Baylin SB, Sidransky D., *Nat Med* **1995**, *1*, 686-692; bJ. G. Herman, A. Merlo, L. Mao, R. G. Lapidus, J.-P. J. Issa, N. E. Davidson, D. Sidransky, S. B. Baylin, *Cancer Research* **1995**, *55*, 4525-4530.
- [43] C. Salem, G. Liang, Y. C. Tsai, J. Coulter, M. A. Knowles, A.-C. Feng, S. Groshen, P. W. Nichols, P. A. Jones, *Cancer Research* **2000**, *60*, 2473-2476.
- [44] G. J. Nuovo, T. W. Plaia, S. A. Belinsky, S. B. Baylin, J. G. Herman, *Proceedings of the National Academy of Sciences* **1999**, *96*, 12754-12759.
- [45] M. L. Gonzalgo, C. M. Bender, E. H. You, J. M. Glendening, J. F. Flores, G. J. Walker, N. K. Hayward, P. A. Jones, J. W. Fountain, *Cancer Research* **1997**, *57*, 5336-5347.
- [46] J. F. Costello, M. S. Berger, H.-J. S. Huang, W. K. Cavenee, *Cancer Research* **1996**, *56*, 2405-2410.
- [47] aM. Esteller, S. R. Hamilton, P. C. Burger, S. B. Baylin, J. G. Herman, *Cancer Research* **1999**, *59*, 793-797; bM. Sanchez-Cespedes, M. Esteller, L. Wu, H. Nawroz-Danish, G. H. Yoo, W. M. Koch, J. Jen, J. G. Herman, D. Sidransky, *Cancer Research* **2000**, *60*, 892-895.
- [48] aM. Esteller, S. Tortola, M. Toyota, G. Capella, M. A. Peinado, S. B. Baylin, J. G. Herman, *Cancer Research* **2000**, *60*, 129-133; bM. Esteller, A. Sparks, M. Toyota, M. Sanchez-Cespedes, G. Capella, M. A. Peinado, S. Gonzalez, G. Tarafa, D. Sidransky, S. J. Meltzer, S. B. Baylin, J. G. Herman, *Cancer Research* **2000**, *60*, 4366-4371.
- [49] aM. Esteller, P. G. Corn, J. M. Urena, E. Gabrielson, S. B. Baylin, J. G. Herman, *Cancer Research* **1998**, *58*, 4515-4518; bW. H. Lee, R. A. Morton, J. I. Epstein, J. D. Brooks, P. A. Campbell, G. S. Bova, W. S. Hsieh, W. B. Isaacs, W. G. Nelson, *Proceedings of the National Academy of Sciences* **1994**, *91*, 11733-11737.
- [50] M. Esteller, J. M. Silva, G. Dominguez, F. Bonilla, X. Matias-Guiu, E. Lerma, E. Bussaglia, J. Prat, I. C. Harkes, E. A. Repasky, E. Gabrielson, M. Schutte, S. B. Baylin, J. G. Herman, *Journal of the National Cancer Institute* **2000**, *92*, 564-569.

- [51] D. Ford, D. F. Easton, D. T. Bishop, S. A. Narod, D. E. Goldgar, *The Lancet* **1994**, *343*, 692-695.
- [52] aM. F. Kane, M. Loda, G. M. Gaida, J. Lipman, R. Mishra, H. Goldman, J. M. Jessup, R. Kolodner, *Cancer Research* **1997**, *57*, 808-811; bJ. G. Herman, A. Umar, K. Polyak, J. R. Graff, N. Ahuja, J.-P. J. Issa, S. Markowitz, J. K. V. Willson, S. R. Hamilton, K. W. Kinzler, M. F. Kane, R. D. Kolodner, B. Vogelstein, T. A. Kunkel, S. B. Baylin, *Proceedings of the National Academy of Sciences* **1998**, *95*, 6870-6875; cR. L. Manel Esteller, Stephen B Baylin, Lora Hedrick Ellenson, and James G Herman, *Oncogene* **1998**, *17*, 2413-2417; dA. S. Fleisher, M. Esteller, S. Wang, G. Tamura, H. Suzuki, J. Yin, T.-T. Zou, J. M. Abraham, D. Kong, K. N. Smolinski, Y.-Q. Shi, M.-G. Rhyu, S. M. Powell, S. P. James, K. T. Wilson, J. G. Herman, S. J. Meltzer, *Cancer Research* **1999**, *59*, 1090-1095.
- [53] aP. G. Corn, S. J. Kuerbitz, M. M. van Noesel, M. Esteller, N. Compitello, S. B. Baylin, J. G. Herman, *Cancer Research* **1999**, *59*, 3352-3356; bJ. G. Herman, J. Jen, A. Merlo, S. B. Baylin, *Cancer Research* **1996**, *56*, 722-727.
- [54] R. A. Irizarry, C. Ladd-Acosta, B. Wen, Z. Wu, C. Montano, P. Onyango, H. Cui, K. Gabo, M. Rongione, M. Webster, H. Ji, J. B. Potash, S. Sabunciyan, A. P. Feinberg, *Nat Genet* **2009**, *41*, 178-186.
- [55] aA. Doi, I.-H. Park, B. Wen, P. Murakami, M. J. Aryee, R. Irizarry, B. Herb, C. Ladd-Acosta, J. Rho, S. Loewer, J. Miller, T. Schlaeger, G. Q. Daley, A. P. Feinberg, *Nat Genet* **2009**, *41*, 1350-1353; bH. Ji, L. I. R. Ehrlich, J. Seita, P. Murakami, A. Doi, P. Lindau, H. Lee, M. J. Aryee, R. A. Irizarry, K. Kim, D. J. Rossi, M. A. Inlay, T. Serwold, H. Karsunky, L. Ho, G. Q. Daley, I. L. Weissman, A. P. Feinberg, *Nature* **2010**, *467*, 338-342.
- [56] R. Garzon, S. Liu, M. Fabbri, Z. Liu, C. E. A. Heaphy, E. Callegari, S. Schwind, J. Pang, J. Yu, N. Muthusamy, V. Havelange, S. Volinia, W. Blum, L. J. Rush, D. Perrotti, M. Andreeff, C. D. Bloomfield, J. C. Byrd, K. Chan, L.-C. Wu, C. M. Croce, G. Marcucci, *Blood* **2009**, *113*, 6411-6418.
- [57] aG. A. Kim YI, Hatch KD, Schneider A, Nour MA, Dallal GE, Selhub J, Mason JB., *Cancer* **1994**, *74*, 893-899; bC.-H. Lin, S.-Y. Hsieh, I.-S. Sheen, W.-C. Lee, T.-C. Chen, W.-C. Shyu, Y.-F. Liaw, *Cancer Research* **2001**, *61*, 4238-4243; cM. T. Bedford, P. D. van Helden, *Cancer Research* **1987**, *47*, 5274-5276.
- [58] aN. T. Hassan KM, Gimelli G, Gartler SM, Hansen RS., *Hum Genet* **2001**, *109*, 452-462; bM. Widschwendter, G. Jiang, C. Woods, H. M. Müller, H. Fiegl, G. Goebel, C. Marth, E. Müller-Holzner, A. G. Zeimet, P. W. Laird, M. Ehrlich, *Cancer Research* **2004**, *64*, 4472-4480.
- [59] R. Nishiyama, L. Qi, K. Tsumagari, K. Weissbecker, L. Dubeau, M. Champagne, S. Sikka, H. Nagai, M. Ehrlich, *Cancer biology & therapy* **2005**, *4*, 440-448.
- [60] aA. P. Feinberg, B. Vogelstein, *Biochemical and Biophysical Research Communications* **1983**, *111*, 47-54; bA. S. Wilson, B. E. Power, P. L. Molloy, *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **2007**, *1775*, 138-162.
- [61] aC. Lengauer, K. W. Kinzler, B. Vogelstein, *Proceedings of the National Academy of Sciences* **1997**, *94*, 2545-2550; bR. Z. Chen, U. Pettersson, C. Beard, L. Jackson-Grusby, R. Jaenisch, *Nature* **1998**, *395*, 89-93.
- [62] Weissbach, *EXS* **1993**, *64*, 1-10.
- [63] aW. Mayer, A. Niveleau, J. Walter, R. Fundele, T. Haaf, *Nature* **2000**, *403*, 501-502; bJ. Oswald, S. Engemann, N. Lane, W. Mayer, A. Olek, R. Fundele, W. Dean, W. Reik, J. Walter, *Current biology : CB* **2000**, *10*, 475-478.
- [64] aS. K. Howlett, W. Reik, *Development* **1991**, *113*, 119-127; bT. Kafri, M. Ariel, M. Brandeis, R. Shemer, L. Urven, J. McCarrey, H. Cedar, A. Razin, *Genes & Development* **1992**, *6*, 705-714; cN. Rougier, D. Bourc'his, D. M. Gomes, A. Niveleau, M. Plachot, A. Pàldi, E. Viegas-Péquignot, *Genes & Development* **1998**, *12*, 2108-2113.
- [65] F. Santos, B. Hendrich, W. Reik, W. Dean, *Developmental Biology* **2002**, *241*, 172-182.

- [66] aJ. Rossant, J. P. Sanford, V. M. Chapman, G. K. Andrews, *Developmental Biology* **1986**, *117*, 567-573; bV. Chapman, L. Forrester, J. Sanford, N. Hastie, J. Rossant, *Nature* **1984**, *307*, 284-286.
- [67] W. Reik, J. Walter, *Nat Genet* **2001**, *27*, 255-256.
- [68] W. Reik, J. Walter, *Nat Rev Genet* **2001**, *2*, 21-32.
- [69] aN. Giannoukakis, C. Deal, J. Paquette, C. G. Goodyer, C. Polychronakos, *Nat Genet* **1993**, *4*, 98-101; bK. D. Tremblay, K. L. Duran, M. S. Bartolomei, *Molecular and Cellular Biology* **1997**, *17*, 4322-4329.
- [70] M. Kaneda, M. Okano, K. Hata, T. Sado, N. Tsujimoto, E. Li, H. Sasaki, *Nature* **2004**, *429*, 900-903.
- [71] J. G. Falls, D. J. Pulford, A. A. Wylie, R. L. Jirtle, *The American journal of pathology* **1999**, *154*, 635-647.
- [72] P. Avner, E. Heard, *Nat Rev Genet* **2001**, *2*, 59-67.
- [73] aC. Beard, E. Li, R. Jaenisch, *Genes & Development* **1995**, *9*, 2325-2334; bB. Panning, R. Jaenisch, *Genes & Development* **1996**, *10*, 1991-2002.
- [74] N. Takagi, M. Sasaki, *Nature* **1975**, *256*, 640-642.
- [75] H. K. Chao W, Spencer RJ, Davidow LS, Lee JT., *Science* **2002**, *295*, 345-347.
- [76] aA. F. A. Smit, *Current Opinion in Genetics & Development* **1999**, *9*, 657-663; bJ. P. Blumenstiel, *Trends in genetics : TIG* **2011**, *27*, 23-31.
- [77] Y. Sekita, H. Wagatsuma, K. Nakamura, R. Ono, M. Kagami, N. Wakisaka, T. Hino, R. Suzuki-Migishima, T. Kohda, A. Ogura, T. Ogata, M. Yokoyama, T. Kaneko-Ishino, F. Ishino, *Nat Genet* **2008**, *40*, 243-248.
- [78] aC. P. Walsh, J. R. Chaillet, T. H. Bestor, *Nat Genet* **1998**, *20*, 116-117; bL. Jackson-Grusby, C. Beard, R. Possemato, M. Tudor, D. Fambrough, G. Csankovszki, J. Dausman, P. Lee, C. Wilson, E. Lander, R. Jaenisch, *Nat Genet* **2001**, *27*, 31-39.
- [79] L. Johnson, *Evol Biol* **2007**, *34*, 121-129.
- [80] J. Z. Iksoo Huh, Taesung Park and Soojin V Yi, *Epigenetics & Chromatin* **2013**, *6*, 1-10.
- [81] aR. Bahar, C. H. Hartmann, K. A. Rodriguez, A. D. Denny, R. A. Busuttil, M. E. T. Dolle, R. B. Calder, G. B. Chisholm, B. H. Pollock, C. A. Klein, J. Vijg, *Nature* **2006**, *441*, 1011-1014; bA. Novick, M. Weiner, *Proceedings of the National Academy of Sciences* **1957**, *43*, 553-566; cE. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, A. van Oudenaarden, *Nat Genet* **2002**, *31*, 69-73; dA. Raj, A. van Oudenaarden, *Cell* **2008**, *135*, 216-226.
- [82] J. Tost, *Epigenetics*, Caister Academic Press, **2008**.
- [83] S. Kumar, X. Cheng, S. Klimasauskas, M. Sha, J. Posfai, R. J. Roberts, G. G. Wilson, *Nucleic Acids Research* **1994**, *22*, 1-10.
- [84] J. Song, O. Rechkoblit, T. H. Bestor, D. J. Patel, *Science* **2011**, *331*, 1036-1040.
- [85] S. Klimasauskas, S. Kumar, R. J. Roberts, X. Cheng, *Cell* **1994**, *76*, 357-369.
- [86] J. Pósfai, A. S. Bhagwat, G. Pósfai, R. J. Roberts, *Nucleic Acids Research* **1989**, *17*, 2421-2435.
- [87] aM. Pradhan, P.-O. Estève, H. G. Chin, M. Samaranayke, G.-D. Kim, S. Pradhan, *Biochemistry* **2008**, *47*, 10000-10009; bS. Pradhan, P.-O. Estève, *Biochemistry* **2003**, *42*, 5321-5332; cS. Pradhan, P.-O. Esteve, *Clinical Immunology* **2003**, *109*, 6-16.
- [88] aB. Youngblood, F.-K. Shieh, F. Buller, T. Bullock, N. O. Reich, *Biochemistry* **2007**, *46*, 8766-8775; bM. O'Gara, S. Klimašauskas, R. J. Roberts, X. Cheng, *Journal of Molecular Biology* **1996**, *261*, 634-645; cH. Zhou, M. M. Purdy, F. W. Dahlquist, N. O. Reich, *Biochemistry* **2009**, *48*, 7807-7816; dF.-K. Shieh, B. Youngblood, N. O. Reich, *Journal of Molecular Biology* **2006**, *362*, 516-527; eJ. K. Christman, *Oncogene* **2002**, *21*, 5483-5495.
- [89] R.-W. C. Yen, P. M. Vertino, B. D. Nelkin, J. J. Yu, W. El-Deiry, A. Cumaraswamy, G. G. Lennon, B. J. Trask, P. Celano, S. B. Baylin, *Nucleic Acids Research* **1992**, *20*, 2287-2291.
- [90] aŽ. M. Svedružić, N. O. Reich, *Biochemistry* **2004**, *43*, 11460-11473; bŽ. M. Svedružić, N. O. Reich, *Biochemistry* **2005**, *44*, 9472-9485.

- [91] J. A. Yoder, N. S. Soman, G. L. Verdine, T. H. Bestor, *Journal of Molecular Biology* **1997**, *270*, 385-395.
- [92] M. R. Rountree, K. E. Bachman, S. B. Baylin, *Nat Genet* **2000**, *25*, 269-277.
- [93] M. D. Allen, C. G. Grummitt, C. Hilcenko, S. Y. Min, L. M. Tonkin, C. M. Johnson, S. M. Freund, M. Bycroft, A. J. Warren, *EMBO J* **2006**, *25*, 4503-4512.
- [94] H. Leonhardt, A. W. Page, H.-U. Weier, T. H. Bestor, *Cell* **1992**, *71*, 865-873.
- [95] K. Fellinger, U. Rothbauer, M. Felle, G. Längst, H. Leonhardt, *Journal of Cellular Biochemistry* **2009**, *106*, 521-528.
- [96] S. ŽM., *Prog Mol Biol Transl Sci* **2011**, *101*, 221-254.
- [97] aG. Pedrali-Noy, A. Weissbach, *Journal of Biological Chemistry* **1986**, *261*, 7600-7602; bR. Stein, Y. Gruenbaum, Y. Pollack, A. Razin, H. Cedar, *Proceedings of the National Academy of Sciences* **1982**, *79*, 61-65.
- [98] M. Fatemi, A. Hermann, S. Pradhan, A. Jeltsch, *Journal of Molecular Biology* **2001**, *309*, 1189-1199.
- [99] aR. Goyal, R. Reinhardt, A. Jeltsch, *Nucleic Acids Research* **2006**, *34*, 1182-1188; bS. ZM., *Curr Med Chem* **2008**, *15*, 92-106.
- [100] R. P. Goyal R, Laser H, Gowher H, Jeltsch A., *Epigenetics* **2007**, *2*, 155-160.
- [101] aY. Sugiyama, N. Hatano, N. Sueyoshi, I. Suetake, S. Tajima, E. Kinoshita, E. Kinoshita-Kikuta, T. Koike, I. Kameshita, *Biochemical Journal* **2010**, *427*, 489-497; bI. Kameshita, M. Sekiguchi, D. Hamasaki, Y. Sugiyama, N. Hatano, I. Suetake, S. Tajima, N. Sueyoshi, *Biochemical and Biophysical Research Communications* **2008**, *377*, 1162-1167; cJ. F. Glickman, J. G. Pavlovich, N. O. Reich, *Journal of Biological Chemistry* **1997**, *272*, 17851-17857.
- [102] aA. Bacolla, S. Pradhan, J. E. Larson, R. J. Roberts, R. D. Wells, *Journal of Biological Chemistry* **2001**, *276*, 18605-18613; bS. Pradhan, A. Bacolla, R. D. Wells, R. J. Roberts, *Journal of Biological Chemistry* **1999**, *274*, 33002-33010.
- [103] X. Zhang, G. L. Verdine, *FEBS Letters* **1996**, *392*, 179-183.
- [104] Y. Zhou, I. Grummt, *Current biology : CB* **2005**, *15*, 1434-1438.
- [105] K. Myant, I. Stancheva, *Molecular and Cellular Biology* **2008**, *28*, 215-226.
- [106] K. D. Robertson, S. Ait-Si-Ali, T. Yokochi, P. A. Wade, P. L. Jones, A. P. Wolffe, *Nat Genet* **2000**, *25*, 338-342.
- [107] X. Cheng, R. M. Blumenthal, *Biochemistry* **2010**, *49*, 2999-3008.
- [108] H. Gowher, C. J. Stockdale, R. Goyal, H. Ferreira, T. Owen-Hughes, A. Jeltsch, *Biochemistry* **2005**, *44*, 9899-9904.
- [109] M. Fatemi, A. Hermann, H. Gowher, A. Jeltsch, *European Journal of Biochemistry* **2002**, *269*, 4981-4984.
- [110] N. J. Kim GD, Kelesoglu N, Roberts RJ, Pradhan S., *EMBO J.* **2002**, *21*, 4183-4295.
- [111] M. Bostick, J. K. Kim, P.-O. Estève, A. Clark, S. Pradhan, S. E. Jacobsen, *Science* **2007**, *317*, 1760-1764.
- [112] J. Sharif, M. Muto, S.-i. Takebayashi, I. Suetake, A. Iwamatsu, T. A. Endo, J. Shinga, Y. Mizutani-Koseki, T. Toyoda, K. Okamura, S. Tajima, K. Mitsuya, M. Okano, H. Koseki, *Nature* **2007**, *450*, 908-912.
- [113] M. Achour, X. Jacq, P. Ronde, M. Alhosin, C. Charlot, T. Chataigneau, M. Jeanblanc, M. Macaluso, A. Giordano, A. D. Hughes, V. B. Schini-Kerth, C. Bronner, *Oncogene* **2007**, *27*, 2187-2197.
- [114] T. Miyata, T. Oyama, K. Mayanagi, S. Ishino, Y. Ishino, K. Morikawa, *Nat Struct Mol Biol* **2004**, *11*, 632-636.
- [115] T. Iida, I. Suetake, S. Tajima, H. Morioka, S. Ohta, C. Obuse, T. Tsurimoto, *Genes to Cells* **2002**, *7*, 997-1007.



- [116] V. Shukla, X. Coumoul, T. Lahusen, R.-H. Wang, X. Xu, A. Vassilopoulos, C. Xiao, M.-H. Lee, Y.-G. Man, M. Ouchi, T. Ouchi, C.-X. Deng, *Cell Res* **2010**, *20*, 1201-1215.
- [117] S. Pradhan, G.-D. Kim, *EMBO J* **2002**, *21*, 779-788.
- [118] P.-O. Estève, H. G. Chin, S. Pradhan, *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 1000-1005.
- [119] G. L. Gac, P.-O. Estève, C. Ferec, S. Pradhan, *Journal of Biological Chemistry* **2006**, *281*, 24161-24170.
- [120] aA. A. Aravin, R. Sachidanandam, D. Bourc'his, C. Schaefer, D. Pezic, K. F. Toth, T. Bestor, G. J. Hannon, *Molecular cell* **2008**, *31*, 785-799; bT. Kanno, E. Bucher, L. Daxinger, B. Huettel, G. Bohmdorfer, W. Gregor, D. P. Kreil, M. Matzke, A. J. M. Matzke, *Nat Genet* **2008**, *40*, 670-675; cT. Watanabe, Y. Totoki, A. Toyoda, M. Kaneda, S. Kuramochi-Miyagawa, Y. Obata, H. Chiba, Y. Kohara, T. Kono, T. Nakano, M. A. Surani, Y. Sakaki, H. Sasaki, *Nature* **2008**, *453*, 539-543.
- [121] aS. M. Carty, A. L. Greenleaf, *Molecular & Cellular Proteomics* **2002**, *1*, 598-610; bM. Zaratiegui, D. V. Irvine, R. A. Martienssen, *Cell* **2007**, *128*, 763-776.
- [122] aJ. P. Saito Y, *Cell Cycle* **2006**, *19*, 2220-2222; bM. Fabbri, R. Garzon, A. Cimmino, Z. Liu, N. Zanesi, E. Callegari, S. Liu, H. Alder, S. Costinean, C. Fernandez-Cymering, S. Volinia, G. Guler, C. D. Morrison, K. K. Chan, G. Marcucci, G. A. Calin, K. Huebner, C. M. Croce, *Proceedings of the National Academy of Sciences* **2007**, *104*, 15805-15810; cJ. M. Friedman, G. Liang, C.-C. Liu, E. M. Wolff, Y. C. Tsai, W. Ye, X. Zhou, P. A. Jones, *Cancer Research* **2009**, *69*, 2623-2629.
- [123] L. Jeffery, S. Nakielny, *Journal of Biological Chemistry* **2004**, *279*, 49479-49487.
- [124] K. Takeshita, I. Suetake, E. Yamashita, M. Suga, H. Narita, A. Nakagawa, S. Tajima, *Proceedings of the National Academy of Sciences* **2011**.
- [125] M. O'Gara, R. J. Roberts, X. Cheng, *Journal of Molecular Biology* **1996**, *263*, 597-606.
- [126] F. Syeda, R. L. Fagan, M. Wean, G. V. Avvakumov, J. R. Walker, S. Xue, S. Dhe-Paganon, C. Brenner, *Journal of Biological Chemistry* **2011**, *286*, 15344-15351.
- [127] aA. R. Venkitaraman, *Cell* **2002**, *108*, 171-182; bS. L. Lei Zheng, Thomas G Boyer and Wen-Hwa Lee, *Oncogene* **2000**, *19*, 6159-6175; cL. Cao, S. Kim, C. Xiao, R.-H. Wang, X. Coumoul, X. Wang, W. M. Li, X. L. Xu, J. A. De Soto, H. Takai, S. Mai, S. J. Elledge, N. Motoyama, C.-X. Deng, *EMBO J* **2006**, *25*, 2167-2177; dR. Bachelier, X. Xu, X. Wang, W. Li, M. Naramura, H. Gu, C.-X. Deng, *Oncogene* **2002**, *22*, 528-537; eC.-X. Deng, *Nucleic Acids Research* **2006**, *34*, 1416-1426.
- [128] P.-O. Estève, H. G. Chin, J. Benner, G. R. Feehery, M. Samaranyake, G. A. Horwitz, S. E. Jacobsen, S. Pradhan, *Proceedings of the National Academy of Sciences* **2009**, *106*, 5076-5081.
- [129] aA. Reale, G. D. Matteis, G. Galleazzi, M. Zampieri, P. Caiafa, *Oncogene* **2005**, *24*, 13-19; bG. Zardo, A. Reale, C. Passananti, S. Pradhan, S. Buontempo, G. De Matteis, R. L. P. Adams, P. Caiafa, *The FASEB Journal* **2002**.
- [130] aP. A. Jones, G. Liang, *Nat Rev Genet* **2009**, *10*, 805-811; bA. Bolden, C. Ward, J. A. Siedlecki, A. Weissbach, *Journal of Biological Chemistry* **1984**, *259*, 12437-12443.
- [131] W. M. Yu N, *Curr Med Chem* **2008**, *15*, 1350-1375.
- [132] P. A. Jones, S. M. Taylor, *Cell* **1980**, *20*, 85-93.
- [133] D. V. Santi, C. E. Garrett, P. J. Barr, *Cell* **1983**, *33*, 9-10.
- [134] L. Zhou, X. Cheng, B. A. Connolly, M. J. Dickman, P. J. Hurd, D. P. Hornby, *Journal of Molecular Biology* **2002**, *321*, 591-599.
- [135] M. Z. Fang, Y. Wang, N. Ai, Z. Hou, Y. Sun, H. Lu, W. Welsh, C. S. Yang, *Cancer Research* **2003**, *63*, 7563-7570.
- [136] Z. Liu, Z. Xie, W. Jones, R. E. Pavlovicz, S. Liu, J. Yu, P.-k. Li, J. Lin, J. R. Fuchs, G. Marcucci, C. Li, K. K. Chan, *Bioorganic & Medicinal Chemistry Letters* **2009**, *19*, 706-709.

- [137] Z. Liu, S. Liu, Z. Xie, R. E. Pavlovicz, J. Wu, P. Chen, J. Aimuwu, J. Pang, D. Bhasin, P. Neviani, J. R. Fuchs, C. Plass, P.-K. Li, C. Li, T. H.-M. Huang, L.-C. Wu, L. Rush, H. Wang, D. Perrotti, G. Marcucci, K. K. Chan, *Journal of Pharmacology and Experimental Therapeutics* **2009**, *329*, 505-514.
- [138] aS. Jagadeesh, S. Sinha, B. C. Pal, S. Bhattacharya, P. P. Banerjee, *Biochemical and Biophysical Research Communications* **2007**, *362*, 212-217; bS. Sinha, B. C. Pal, S. Jagadeesh, P. P. Banerjee, A. Bandyopadhyaya, S. Bhattacharya, *The Prostate* **2006**, *66*, 1257-1265.
- [139] I. C. Piña, J. T. Gautschi, G.-Y.-S. Wang, M. L. Sanders, F. J. Schmitz, D. France, S. Cornell-Kennon, L. C. Sambucetti, S. W. Remiszewski, L. B. Perez, K. W. Bair, P. Crews, *The Journal of Organic Chemistry* **2003**, *68*, 3866-3873.
- [140] aB. Brueckner, R. Garcia Boy, P. Siedlecki, T. Musch, H. C. Kliem, P. Zielenkiewicz, S. Suhai, M. Wiessler, F. Lyko, *Cancer Research* **2005**, *65*, 6305-6311; bP. Siedlecki, R. G. Boy, T. Musch, B. Brueckner, S. Suhai, F. Lyko, P. Zielenkiewicz, *Journal of Medicinal Chemistry* **2005**, *49*, 678-683.
- [141] aB. Segura-Pacheco, C. Trejo-Becerril, E. Perez-Cardenas, L. Taja-Chayeb, I. Mariscal, A. Chavez, C. Acuña, A. M. Salazar, M. Lizano, A. Dueñas-Gonzalez, *Clinical Cancer Research* **2003**, *9*, 1596-1603; bS.-P. B. Zambrano P, Perez-Cardenas E, Cetina L, Revilla-Vazquez A, Taja-Chayeb L, Chavez-Blanco A, Angeles E, Cabrera G, Sandoval K, Trejo-Becerril C, Chanona-Vilchis J, Duenas-González A., *BMC Cancer* **2005**, *29*, 44; cM. Candelaria, D. Gallardo-Rincón, C. Arce, L. Cetina, J. Aguilar-Ponce, Ó. Arrieta, A. González-Fierro, A. Chávez-Blanco, E. de la Cruz-Hernández, M. Camargo, C. Trejo-Becerril, E. Pérez-Cárdenas, C. Pérez-Plasencia, L. Taja-Chayeb, T. Wegman-Ostrosky, A. Revilla-Vazquez, A. Dueñas-González, *Annals of Oncology* **2007**, *18*, 1529-1538.
- [142] aX. Lin, K. Asgari, M. J. Putzi, W. R. Gage, X. Yu, B. S. Cornblatt, A. Kumar, S. Piantadosi, T. L. DeWeese, A. M. De Marzo, W. G. Nelson, *Cancer Research* **2001**, *61*, 8611-8616; bA. Villar-Garea, M. F. Fraga, J. Espada, M. Esteller, *Cancer Research* **2003**, *63*, 4984-4989; cJ. M. Scheinbart LS, Gross LA, Edelstein SR, Richardson BC., *J Rheumatol.* **1991**, *18*, 530 - 534.
- [143] J.-R. Weng, I. L. Lai, H.-C. Yang, C.-N. Lin, L.-Y. Bai, *Phytotherapy Research* **2014**, *28*, 49-54.
- [144] K. Lubecka-Pietruszewska, A. Kaufman-Szymczyk, B. Stefanska, B. Cebula-Obrzut, P. Smolewski, K. Fabianowska-Majewska, *European Journal of Pharmacology* **2014**, *723*, 276-287.
- [145] B. Krawczyk, K. Rudnicka, K. Fabianowska-Majewska, *Nucleosides, Nucleotides and Nucleic Acids* **2007**, *26*, 1043-1046.
- [146] J. Yoo, J. Medina-Franco, *J Mol Model* **2012**, *18*, 1583-1589.
- [147] B. H. Lee, S. Yegnasubramanian, X. Lin, W. G. Nelson, *Journal of Biological Chemistry* **2005**, *280*, 40749-40756.
- [148] L. Du, Z. Xie, L.-c. Wu, M. Chiu, J. Lin, K. K. Chan, S. Liu, Z. Liu, *Nutrition & Cancer* **2012**, *64*, 1228-1235.
- [149] J. Datta, K. Ghoshal, W. A. Denny, S. A. Gamage, D. G. Brooke, P. Phiasivongsa, S. Redkar, S. T. Jacob, *Cancer Research* **2009**, *69*, 4277-4285.
- [150] S. L. Zhongfa Liu, Zhiliang Xie, Ryan E. Pavlovicz, Jiejun Wu, Ping Chen, , J. P. Josephine Aimuwu, Deepak Bhasin, Paolo Neviani, , James R. Fuchs, P.-K. L. Christoph Plass, Chenglong Li, Tim H-M Huang, Lai-Chu Wu, Laura Rush, , D. P. Hongyan Wang, Guido Marcucci, and Kenneth K. Chan *Journal of Pharmacology and Experimental Therapeutics* **2009**.
- [151] S. Castellano, D. Kuck, M. Viviano, J. Yoo, F. López-Vallejo, P. Conti, L. Tamborini, A. Pinto, J. L. Medina-Franco, G. Sbardella, *Journal of Medicinal Chemistry* **2011**, *54*, 7663-7677.
- [152] S.-C. Wang, T.-H. Lee, C.-H. Hsu, Y.-J. Chang, M.-S. Chang, Y.-C. Wang, Y.-S. Ho, W.-C. Wen, R.-K. Lin, *Journal of Agricultural and Food Chemistry* **2014**.
- [153] Y. L. Ottaviano, J.-P. Issa, F. F. Parl, H. S. Smith, S. B. Baylin, N. E. Davidson, *Cancer Research* **1994**, *54*, 2552-2555.
- [154] A. T. F. Sharyl J Nass, Dorraya El-Ashry, William G Nelson and Nancy E Davidson, *Oncogene* **1999**, *18*, 7453-7461.



- [155] A. T. Ferguson, R. G. Lapidus, S. B. Baylin, N. E. Davidson, *Cancer Research* **1995**, *55*, 2279-2283.
- [156] D. V, *Crit Rev Biochem Mol Biol* **1993**, *28*, 173-207.
- [157] B. Mannervik, P. Alin, C. Guthenberg, H. Jensson, M. K. Tahir, M. Warholm, H. Jörnvall, *Proceedings of the National Academy of Sciences* **1985**, *82*, 7202-7206.
- [158] M. S. Jhaveri, C. S. Morrow, *Gene* **1998**, *210*, 1-7.
- [159] E. L. Cavalieri, D. E. Stack, P. D. Devanesan, R. Todorovic, I. Dwivedy, S. Higginbotham, S. L. Johansson, K. D. Patil, M. L. Gross, J. K. Gooden, R. Ramanathan, R. L. Cerny, E. G. Rogan, *Proceedings of the National Academy of Sciences* **1997**, *94*, 10937-10942.
- [160] aY. Miki, J. Swensen, D. Shattuck-Eidens, P. Futreal, K. Harshman, S. Tavtigian, Q. Liu, C. Cochran, L. Bennett, W. Ding, a. et, *Science* **1994**, *266*, 66-71; bL. S. Friedman, E. A. Ostermeyer, C. I. Szabo, P. Dowd, E. D. Lynch, S. E. Rowell, M.-C. King, *Nat Genet* **1994**, *8*, 399-404.
- [161] aA. Dobrovic, D. Simpfendorfer, *Cancer Research* **1997**, *57*, 3347-3350; bF. MAGDINIER, L.-M. BILLARD, G. WITTMANN, L. FRAPPART, M. BENCHAÏB, G. M. LENOIR, J. F. GUÉRIN, R. DANTE, *The FASEB Journal* **2000**, *14*, 1585-1594.
- [162] aD. E. Gomez, M. S. De Lorenzo, D. F. Alonso, Z. A. Andrade, *The American Journal of Tropical Medicine and Hygiene* **1999**, *61*, 9-13; bJ. R. Graff, J. G. Herman, R. G. Lapidus, H. Chopra, R. Xu, D. F. Jarrard, W. B. Isaacs, P. M. Pitha, N. E. Davidson, S. B. Baylin, *Cancer Research* **1995**, *55*, 5195-5199.
- [163] K. E. Bachman, J. G. Herman, P. G. Corn, A. Merlo, J. F. Costello, W. K. Cavenee, S. B. Baylin, J. R. Graff, *Cancer Research* **1999**, *59*, 798-802.
- [164] K. Bardhan, K. Liu, *Cancers* **2013**, *5*, 676-713.
- [165] J. R. Jass, *Surgical oncology* **2007**, *16*, 7-9.
- [166] aB. Vogelstein, E. R. Fearon, S. R. Hamilton, S. E. Kern, A. C. Preisinger, M. Leppert, A. M. M. Smits, J. L. Bos, *New England Journal of Medicine* **1988**, *319*, 525-532; bA. Vazquez, E. E. Bond, A. J. Levine, G. L. Bond, *Nat Rev Drug Discov* **2008**, *7*, 979-987.
- [167] T. L. Chan, S. T. Yuen, C. K. Kong, Y. W. Chan, A. S. Y. Chan, W. F. Ng, W. Y. Tsui, M. W. S. Lo, W. Y. Tam, V. S. W. Li, S. Y. Leung, *Nat Genet* **2006**, *38*, 1178-1183.
- [168] K. W. Kinzler, B. Vogelstein, *Cell* **1996**, *87*, 159-170.
- [169] aM. Toyota, N. Ahuja, M. Ohe-Toyota, J. G. Herman, S. B. Baylin, J.-P. J. Issa, *Proceedings of the National Academy of Sciences* **1999**, *96*, 8681-8686; bD. J. Weisenberger, K. D. Siegmund, M. Campan, J. Young, T. I. Long, M. A. Faasse, G. H. Kang, M. Widschwendter, D. Weener, D. Buchanan, H. Koh, L. Simms, M. Barker, B. Leggett, J. Levine, M. Kim, A. J. French, S. N. Thibodeau, J. Jass, R. Haile, P. W. Laird, *Nat Genet* **2006**, *38*, 787-793; cL. Shen, M. Toyota, Y. Kondo, E. Lin, L. Zhang, Y. Guo, N. S. Hernandez, X. Chen, S. Ahmed, K. Konishi, S. R. Hamilton, J.-P. J. Issa, *Proceedings of the National Academy of Sciences* **2007**, *104*, 18654-18659.
- [170] aC. Wu, T. Bekaii-Saab, *Chemotherapy Research and Practice* **2012**, *2012*, 10; bM. van Rijnsoever, H. Elsaleh, D. Joseph, K. McCaul, B. Iacopetta, *Clinical Cancer Research* **2003**, *9*, 2898-2903; cS. Ogino, J. Meyerhardt, T. Kawasaki, J. Clark, D. Ryan, M. Kulke, P. Enzinger, B. Wolpin, M. Loda, C. Fuchs, *Virchows Arch* **2007**, *450*, 529-537; dS. Lee, N.-Y. Cho, M. Choi, E. J. Yoo, J.-H. Kim, G. H. Kang, *Pathology International* **2008**, *58*, 104-113; eL. Barault, C. Charon-Barra, V. Jooste, M. F. de la Vega, L. Martin, P. Roinot, P. Rat, A.-M. Bouvier, P. Laurent-Puig, J. Faivre, C. Chapusot, F. Piard, *Cancer Research* **2008**, *68*, 8541-8546.
- [171] L. M. Ang PW, Liem N, Lim PL, Grieu F, Vaithilingam A, Platell C, Yong WP, Iacopetta B, Soong R., *BMC Cancer* **2010**, *10*, 227.
- [172] aA. Silver, N. Sengupta, D. Propper, P. Wilson, T. Hagemann, A. Patel, A. Parker, A. Ghosh, R. Feakins, S. Dorudi, N. Suraweera, *International Journal of Cancer* **2012**, *130*, 1082-1092; bT. K. Shuji Ogino, Gregory J Kirkner, Mutsuko Ohnishi and Charles S Fuchs *BMC Cancer* **2006**, *7*, 72; cS. Ogino, T. Kawasaki, K. Nosho, M. Ohnishi, Y. Suemoto, G. J. Kirkner, C. S. Fuchs, *International*

- Journal of Cancer* **2008**, *122*, 2767-2773; dJ.-P. J. Issa, P. M. Vertino, J. Wu, S. Sazawal, P. Celano, B. D. Nelkin, S. R. Hamilton, S. B. Baylin, *Journal of the National Cancer Institute* **1993**, *85*, 1235-1240.
- [173] T. Liljefors, P. Krogsgaard-Larsen, U. Madsen, *Textbook of Drug Design and Discovery, Third Edition*, Taylor & Francis, **2003**.
- [174] C. Smith, *Nature* **2003**, *422*, 341-347.
- [175] J. P. Hughes, S. Rees, S. B. Kalindjian, K. L. Philpott, *British Journal of Pharmacology* **2011**, *162*, 1239-1249.
- [176] A. Smith, *Nature* **2002**, *418*, 453-459.
- [177] J. S. Handen, *Drug Discov. World* **2002**, 47-50.
- [178] S. Fox, S. Farr-Jones, L. Sopchak, A. Boggs, H. W. Nicely, R. Khoury, M. Biros, *Journal of Biomolecular Screening* **2006**, *11*, 864-869.
- [179] K. Boppana, P. K. Dubey, S. A. R. P. Jagarlapudi, S. Vadivelan, G. Rambabu, *European Journal of Medicinal Chemistry* **2009**, *44*, 3584-3590.
- [180] R. Law, O. Barker, J. Barker, T. Hesterkamp, R. Godemann, O. Andersen, T. Fryatt, S. Courtney, D. Hallett, M. Whittaker, *J Comput Aided Mol Des* **2009**, *23*, 459-473.
- [181] aM. H. Zartler ER, *Curr Top Med Chem* **2007**, *7*, 1592-1599; bM. Pellecchia, D. S. Sem, K. Wuthrich, *Nat Rev Drug Discov* **2002**, *1*, 211-219.
- [182] E. C. Butcher, E. L. Berg, E. J. Kunkel, *Nat Biotech* **2004**, *22*, 1253-1259.
- [183] Y. P. C. Chen, F., *Expert Opinion on Therapeutic Targets* **2008**, *12*, 383-389.
- [184] T. Engel, *Journal of Chemical Information and Modeling* **2006**, *46*, 2267-2277.
- [185] G. Klebe, *Drug Discovery Today* **2006**, *11*, 580-594.
- [186] G. Naray-Szabo, G. G. Ferenczy, *Chemical Reviews* **1995**, *95*, 829-847.
- [187] K. KH, *J Comput Aided Mol Des* **2001**, *15*, 367-380.
- [188] J. P. Lam PY, Eyermann CJ, Hodge CN, Ru Y, Bacheler LT, Meek JL, Otto MJ, Rayner MM, Wong YN, et al., *Science* **1994**, *263*, 380-384.
- [189] R. Ng, *Drugs: From Discovery to Approval*, Wiley, **2011**.
- [190] E. Yuriev, M. Agostino, P. A. Ramsland, *Journal of Molecular Recognition* **2011**, *24*, 149-164.
- [191] D. B. Kitchen, H. Decornez, J. R. Furr, J. Bajorath, *Nat Rev Drug Discov* **2004**, *3*, 935-949.
- [192] D. L. Beveridge, F. M. DiCapua, *Annual Review of Biophysics and Biophysical Chemistry* **1989**, *18*, 431-492.
- [193] S. Miyamoto, P. A. Kollman, *Proteins: Structure, Function, and Bioinformatics* **1993**, *16*, 226-245.
- [194] aO. Takahashi, Y. Masuda, A. Muroya, T. Furuya, *SAR and QSAR in Environmental Research* **2010**, *21*, 547-558; bJ. A. Feng, G. R. Marshall, *Journal of Computational Chemistry* **2010**, *31*, 2540-2554; cR. Wang, Y. Lu, X. Fang, S. Wang, *Journal of Chemical Information and Computer Sciences* **2004**, *44*, 2114-2125; dS.-Y. Huang, S. Z. Grinter, X. Zou, *Physical Chemistry Chemical Physics* **2010**, *12*, 12899-12908.
- [195] aD. P. Marriott, I. G. Dougall, P. Meghani, Y.-J. Liu, D. R. Flower, *Journal of Medicinal Chemistry* **1999**, *42*, 3210-3216; bS. Flohr, M. Kurz, E. Kostenis, A. Brkovich, A. Fournier, T. Klabunde, *Journal of Medicinal Chemistry* **2002**, *45*, 1799-1805.
- [196] C. Jegerschöld, S.-C. Pawelzik, P. Purhonen, P. Bhakat, K. R. Gheorghe, N. Gyobu, K. Mitsuoka, R. Morgenstern, P.-J. Jakobsson, H. Hebert, *Proceedings of the National Academy of Sciences* **2008**, *105*, 11110-11115.
- [197] M. Vogt, J. Bajorath, in *Chemoinformatics and Computational Chemical Biology, Vol. 672* (Ed.: J. Bajorath), Humana Press, **2011**, pp. 159-173.
- [198] E. Esposito, A. Hopfinger, J. Madura, in *Chemoinformatics, Vol. 275* (Ed.: J. Bajorath), Humana Press, **2004**, pp. 131-213.
- [199] F. H. Perkins R, Tong W, Welsh WJ., *Environ Toxicol Chem* **2003**, *22*, 1666-1679.

[200] H. S. Jitender K Malik, Singhai A K & Harish Pandey, *International Journal of*

*Pharmaceutical Research &*

*Allied Sciences* **2013**, *2*, 1-13.

- [201] aP. A. N. Babu, M. Lakshmi; Srinivas, K., *Cancer Research* **2010**, *41*, 223-231; bK. C. ZHENG, J. C. CHEN, L. QIAN, W. J. WU, *Journal of Theoretical and Computational Chemistry* **2007**, *06*, 223-231; cC.-Y. Chen, C. Y.-C. Chen, *Journal of Molecular Graphics and Modelling* **2010**, *29*, 21-31.
- [202] P. Ehrlich, *Berichte der deutschen chemischen Gesellschaft* **1909**, *42*, 17-47.
- [203] v. D. JH., *Curr Pharm Des* **2003**, *9*, 1649-1664.
- [204] aG. Mustata, A. V. Follis, D. I. Hammoudeh, S. J. Metallo, H. Wang, E. V. Prochownik, J. S. Lazo, I. Bahar, *Journal of Medicinal Chemistry* **2009**, *52*, 1247-1250; bR. Petersen, K. Christensen, A. Assimopoulou, X. Fretté, V. Papageorgiou, K. Kristiansen, I. Kouskoumvekaki, *J Comput Aided Mol Des* **2011**, *25*, 107-116; cl. Barroso, M. Gurnell, V. E. F. Crowley, M. Agostini, J. W. Schwabe, M. A. Soos, G. L. Maslen, T. D. M. Williams, H. Lewis, A. J. Schafer, V. K. K. Chatterjee, S. O'Rahilly, *Nature* **1999**, *402*, 880-883; dP. D. Lyne, P. W. Kenny, D. A. Cosgrove, C. Deng, S. Zabudoff, J. J. Wendoloski, S. Ashwell, *Journal of Medicinal Chemistry* **2004**, *47*, 1962-1968; eM. L. P. a. M. C. Nicklaus, *J Cheminform.* **2009**, *6*.
- [205] Y. Xiang, Z. Hou, Z. Zhang, *Chemical Biology & Drug Design* **2012**, *79*, 760-770.
- [206] A. D. Andricopulo, R. V. C. Guido, G. Oliva, *Current Medicinal Chemistry* **2008**, *15*, 37-46.
- [207] aP. Willett, *Drug Discovery Today* **2006**, *11*, 1046-1053; bP. Willett, *Journal of Medicinal Chemistry* **2005**, *48*, 4183-4199.
- [208] aJ. Hert, P. Willett, D. J. Wilton, P. Acklin, K. Azzaoui, E. Jacoby, A. Schuffenhauer, *Journal of Chemical Information and Computer Sciences* **2004**, *44*, 1177-1185; bG. Schneider, P. Schneider, S. Renner, *QSAR & Combinatorial Science* **2006**, *25*, 1162-1171.
- [209] aA. C. Anderson, *Chemistry & Biology* **2003**, *10*, 787-797; bM. A. Lill, *Biochemistry* **2011**, *50*, 6157-6169; cC. A. Sotriffer, *Current Topics in Medicinal Chemistry* **2011**, *11*, 179.
- [210] A. C. Anderson, R. H. O'Neil, T. S. Surti, R. M. Stroud, *Chemistry & Biology* **2001**, *8*, 445-457.
- [211] R. A. Powers, F. Morandi, B. K. Shoichet, *Structure (London, England : 1993)* **2002**, *10*, 1013-1023.
- [212] E. Yuriev, P. A. Ramsland, *Journal of Molecular Recognition* **2013**, *26*, 215-239.
- [213] R. Thilagavathi, R. L. Mancera, *Journal of Chemical Information and Modeling* **2010**, *50*, 415-421.
- [214] D. Cappel, R. Wahlström, R. Brenk, C. A. Sotriffer, *Journal of Chemical Information and Modeling* **2011**, *51*, 2581-2594.
- [215] T. ten Brink, T. E. Exner, *Journal of Chemical Information and Modeling* **2009**, *49*, 1535-1546.
- [216] E. Kellenberger, J. Rodrigo, P. Muller, D. Rognan, *Proteins: Structure, Function, and Bioinformatics* **2004**, *57*, 225-242.
- [217] G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman, *J Comput Aided Mol Des* **2013**, *27*, 221-234.
- [218] A. Varnek, A. Tropsha, *Chemoinformatics Approaches to Virtual Screening*, Royal Society of Chemistry, **2008**.
- [219] aC. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Advanced Drug Delivery Reviews* **2001**, *46*, 3-26; bC. A. Lipinski, *Drug Discovery Today: Technologies* **2004**, *1*, 337-341.
- [220] aM. P. Gleeson, *Journal of Medicinal Chemistry* **2008**, *51*, 817-834; bR. S. Bohacek, C. McMartin, W. C. Guida, *Medicinal Research Reviews* **1996**, *16*, 3-50.
- [221] B. E. Evans, K. E. Rittle, M. G. Bock, R. M. DiPardo, R. M. Freidinger, W. L. Whitter, G. F. Lundell, D. F. Veber, P. S. Anderson, *Journal of Medicinal Chemistry* **1988**, *31*, 2235-2246.

- [222] aA. Bender, R. C. Glen, *Organic & Biomolecular Chemistry* **2004**, *2*, 3204-3218; bR. J. Zauhar, G. Moyna, L. Tian, Z. Li, W. J. Welsh, *Journal of Medicinal Chemistry* **2003**, *46*, 5674-5690.
- [223] L. Hamers, Y. Hemeryck, G. Herweyers, M. Janssen, H. Keters, R. Rousseau, A. Vanhoutte, *Information Processing & Management* **1989**, *25*, 315-318.
- [224] Q. Zhang, I. Muegge, *Journal of Medicinal Chemistry* **2006**, *49*, 1536-1548.
- [225] Z. Chen, H.-l. Li, Q.-j. Zhang, X.-g. Bao, K.-q. Yu, X.-m. Luo, W.-l. Zhu, H.-l. Jiang, *Acta Pharmacol Sin* **2009**, *30*, 1694-1708.
- [226] A. Beautrait, V. Leroux, M. Chavent, L. Ghemtio, M.-D. Devignes, M. Smail-Tabbone, W. Cai, X. Shao, G. Moreau, P. Bladon, J. Yao, B. Maignet, *J Mol Model* **2008**, *14*, 135-148.
- [227] aO. S. d. LigPrep v2.5. Portland, Inc.; , **2011**; bO. S. d. Epik v2.2. Portland, Inc.; , **2011**.
- [228] J. W. H.M. Berman, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, *Nucleic Acids Research* **2000**, *28*, 235-242.
- [229] K. Arnold, L. Bordoli, J. Kopp, T. Schwede, *Bioinformatics* **2006**, *22*, 195-201.
- [230] N. Eswar, B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M.-y. Shen, U. Pieper, A. Sali, in *Current Protocols in Protein Science*, John Wiley & Sons, Inc., **2001**.
- [231] B. Wallner, A. Elofsson, *Protein Science* **2005**, *14*, 1315-1327.
- [232] R. A. Laskowski, M. W. Macarthur, D. S. Moss, J. M. Thornton, *J. Appl. Cryst.* **1993**, *26*, 283-291.
- [233] aY. J. Huang, R. Powers, G. T. Montelione, *Journal of the American Chemical Society* **2005**, *127*, 1665-1674; bC. L. Gracy J, Sallantin J., *Protein Eng* **1993**, *6*, 821-829; cH. van den Bedem, G. Wolf, Q. Xu, A. M. Deacon, *Acta crystallographica. Section D, Biological crystallography* **2011**, *67*, 368-375.
- [234] O. S. d. Maestro v9.2. Portland, Inc.; , **2011**.
- [235] L. H. Goodsell DS, Stout CD, Olson AJ., *Proteins* **1993**, *17*, 1-10.
- [236] T. Beuming, Y. Che, R. Abel, B. Kim, V. Shanmugasundaram, W. Sherman, *Proteins: Structure, Function, and Bioinformatics* **2012**, *80*, 871-883.
- [237] N. Brooijmans, I. D. Kuntz, *Annual Review of Biophysics and Biomolecular Structure* **2003**, *32*, 335-373.
- [238] E. Luethi, K. T. Nguyen, M. Bürzle, L. C. Blum, Y. Suzuki, M. Hediger, J.-L. Reymond, *Journal of Medicinal Chemistry* **2010**, *53*, 7236-7250.
- [239] H. Park, Y. J. Bahn, S.-K. Jung, D. G. Jeong, S.-H. Lee, I. Seo, T.-S. Yoon, S. J. Kim, S. E. Ryu, *Journal of Medicinal Chemistry* **2008**, *51*, 5533-5541.
- [240] A. Kovač, J. Konc, B. Vehar, J. M. Bostock, I. Chopra, D. a. Janežič, S. Gobec, *Journal of Medicinal Chemistry* **2008**, *51*, 7442-7448.
- [241] A. Steffen, C. Thiele, S. Tietze, C. Strassnig, A. Kämper, T. Lengauer, G. Wenz, J. Apostolakis, *Chemistry – A European Journal* **2007**, *13*, 6801-6809.
- [242] C.-S. Chen, C.-T. Chiou, G. S. Chen, S.-C. Chen, C.-Y. Hu, W.-K. Chi, Y.-D. Chu, L.-H. Hwang, P.-J. Chen, D.-S. Chen, S.-H. Liaw, J.-W. Chern, *Journal of Medicinal Chemistry* **2009**, *52*, 2716-2723.
- [243] Z. Liu, C. Huang, K. Fan, P. Wei, H. Chen, S. Liu, J. Pei, L. Shi, B. Li, K. Yang, Y. Liu, L. Lai, *Journal of Chemical Information and Modeling* **2004**, *45*, 10-17.
- [244] J.-C. Mozziconacci, E. Arnoult, P. Bernard, Q. T. Do, C. Marot, L. Morin-Allory, *Journal of Medicinal Chemistry* **2005**, *48*, 1055-1068.
- [245] W. B. Moro, Z. Yang, T. A. Kane, C. G. Brouillette, W. J. Brouillette, *Bioorganic & Medicinal Chemistry Letters* **2009**, *19*, 2001-2005.
- [246] S. Wu, M. Bottini, R. C. Rickert, T. Mustelin, L. Tautz, *ChemMedChem* **2009**, *4*, 440-444.
- [247] J. Kim, K.-s. Kim, D.-E. Kim, Y. Chong, *Chemical Biology & Drug Design* **2008**, *72*, 585-591.
- [248] Y. Cho, T. R. Ioerger, J. C. Sacchettini, *Journal of Medicinal Chemistry* **2008**, *51*, 5984-5992.
- [249] R. Kiss, B. Kiss, Á. Könczöl, F. Szalai, I. Jelinek, V. László, B. Noszál, A. Falus, G. M. Keserű, *Journal of Medicinal Chemistry* **2008**, *51*, 3145-3153.

- [250] J.-F. Cheng, J. Zapf, K. Takedomi, C. Fukushima, T. Ogiku, S.-H. Zhang, G. Yang, N. Sakurai, M. Barbosa, R. Jack, K. Xu, *Journal of Medicinal Chemistry* **2008**, *51*, 2057-2061.
- [251] G. Barreiro, C. R. W. Guimarães, I. Tubert-Brohman, T. M. Lyons, J. Tirado-Rives, W. L. Jorgensen, *Journal of Chemical Information and Modeling* **2007**, *47*, 2416-2428.
- [252] J. M. LaLonde, M. A. Elban, J. R. Courter, A. Sugawara, T. Soeta, N. Madani, A. M. Princiotta, Y. D. Kwon, P. D. Kwong, A. Schön, E. Freire, J. Sodroski, A. B. Smith Iii, *Bioorganic & Medicinal Chemistry* **2011**, *19*, 91-101.
- [253] R. Kurczab, M. Nowak, Z. Chilmonczyk, I. Sylte, A. J. Bojarski, *Bioorganic & Medicinal Chemistry Letters* **2010**, *20*, 2465-2468.
- [254] W. Xu, G. Chen, W. Zhu, Z. Zuo, *Bioorganic & Medicinal Chemistry Letters* **2010**, *20*, 5763-5766.
- [255] J. Deye, C. Elam, M. Lape, R. Ratliff, K. Evans, S. Paula, *Bioorganic & Medicinal Chemistry* **2009**, *17*, 1353-1360.
- [256] J. Neres, M. L. Brewer, L. Ratier, H. Botti, A. Buschiazzi, P. N. Edwards, P. N. Mortenson, M. H. Charlton, P. M. Alzari, A. C. Frasch, R. A. Bryce, K. T. Douglas, *Bioorganic & Medicinal Chemistry Letters* **2009**, *19*, 589-596.
- [257] P. R. Patrick A. Holt, Lucjan Strekowski, Jonathan B. Chaires, and John O. Trent. , *Nucleic Acids Research* **2009**, 1280-1287.
- [258] M. Feder, E. Purta, L. Koscinski, S. Čubrilo, G. Maravic Vlahovicek, J. M. Bujnicki, *ChemMedChem* **2008**, *3*, 316-322.
- [259] I. Musmuca, A. Caroli, A. Mai, N. Kaushik-Basu, P. Arora, R. Ragno, *Journal of Chemical Information and Modeling* **2010**, *50*, 662-676.
- [260] B. C. Pearce, D. R. Langley, J. Kang, H. Huang, A. Kulkarni, *Journal of Chemical Information and Modeling* **2009**, *49*, 1797-1809.
- [261] aJ. Schlosser, M. Rarey, *Journal of Chemical Information and Modeling* **2009**, *49*, 800-809; bM. J. Hartshorn, M. L. Verdonk, G. Chessari, S. C. Brewerton, W. T. M. Mooij, P. N. Mortenson, C. W. Murray, *Journal of Medicinal Chemistry* **2007**, *50*, 726-741.
- [262] A. R. Leach, *Molecular Modelling: Principles And Applications*, 2/E, Pearson Education, **1996**.
- [263] S. N. Viji, P. A. Prasad, N. Gautham, *Journal of Chemical Information and Modeling* **2009**, *49*, 2687-2694.
- [264] aT. A. Ewing, S. Makino, A. G. Skillman, I. Kuntz, *J Comput Aided Mol Des* **2001**, *15*, 411-428; bM. D. M. a. S. K. K. a. D. J. U. a. R. P. Sheridan, *J Comput Aided Mol Des* **1994**, *8*, 153-174.
- [265] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, *Journal of Molecular Biology* **1996**, *261*, 470-489.
- [266] D. Westhead, D. Clark, C. Murray, *J Comput Aided Mol Des* **1997**, *11*, 209-228.
- [267] O. Trott, A. J. Olson, *Journal of Computational Chemistry* **2010**, *31*, 455-461.
- [268] C. Pérez, A. R. Ortiz, *Journal of Medicinal Chemistry* **2001**, *44*, 3768-3785.
- [269] aM. Mangoni, D. Roccatano, A. Di Nola, *Proteins: Structure, Function, and Bioinformatics* **1999**, *35*, 153-162; bA. Di Nola, D. Roccatano, H. J. C. Berendsen, *Proteins: Structure, Function, and Bioinformatics* **1994**, *19*, 174-182.
- [270] J.-Y. Trosset, H. A. Scheraga, *Proceedings of the National Academy of Sciences* **1998**, *95*, 8011-8015.
- [271] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, P. S. Shenkin, *Journal of Medicinal Chemistry* **2004**, *47*, 1739-1749.
- [272] Q. Chemical Computing Group. MOE. 2003. Montreal, Canada.
- [273] W. Welch, J. Ruppert, A. N. Jain, *Chemistry & biology* **1996**, *3*, 449-462.
- [274] N. Huang, C. Kalyanaraman, J. J. Irwin, M. P. Jacobson, *Journal of Chemical Information and Modeling* **2005**, *46*, 243-253.



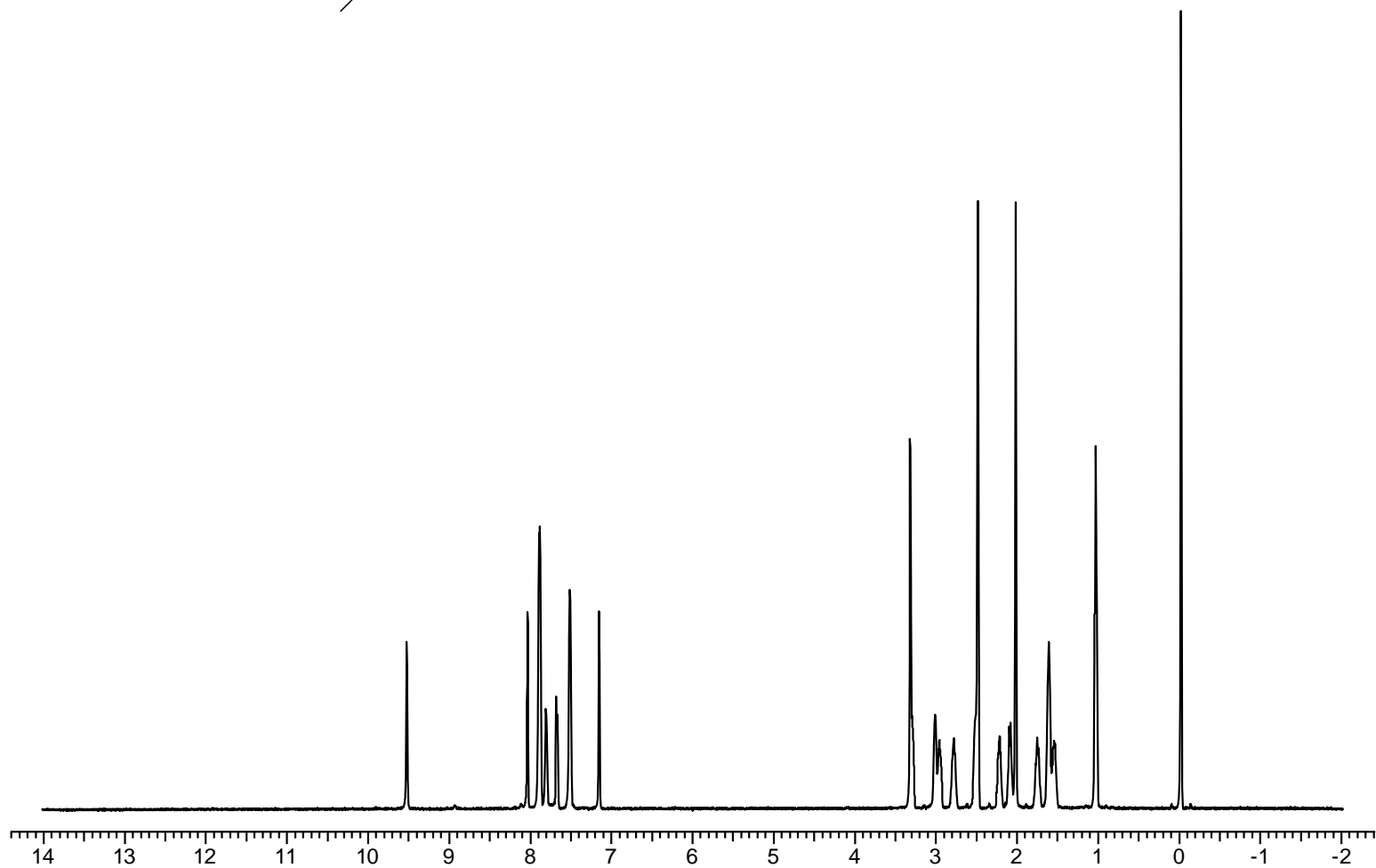
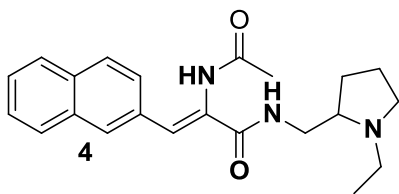
- [275] aB. Q. Wei, W. A. Baase, L. H. Weaver, B. W. Matthews, B. K. Shoichet, *Journal of Molecular Biology* **2002**, 322, 339-355; bB. K. Shoichet, A. R. Leach, I. D. Kuntz, *Proteins: Structure, Function, and Bioinformatics* **1999**, 34, 4-16.
- [276] aW. Wang, O. Donini, C. M. Reyes, P. A. Kollman, *Annual Review of Biophysics and Biomolecular Structure* **2001**, 30, 211-243; bS. A. Adcock, J. A. McCammon, *Chemical Reviews* **2006**, 106, 1589-1615.
- [277] aD. A. Pearlman, *Journal of Medicinal Chemistry* **2005**, 48, 7796-7807; bD. Huang, A. Caflisch, *Journal of Medicinal Chemistry* **2004**, 47, 5791-5797.
- [278] aH.-Y. Liu, I. D. Kuntz, X. Zou, *The Journal of Physical Chemistry B* **2004**, 108, 5453-5462; bH.-Y. Liu, S. Z. Grinter, X. Zou, *The Journal of Physical Chemistry B* **2009**, 113, 11793-11799.
- [279] S. J. Weiner, P. A. Kollman, D. T. Nguyen, D. A. Case, *Journal of Computational Chemistry* **1986**, 7, 230-252.
- [280] B. Kramer, M. Rarey, T. Lengauer, *Proteins: Structure, Function, and Bioinformatics* **1999**, 37, 228-241.
- [281] H.-J. Böhm, *J Comput Aided Mol Des* **1994**, 8, 243-256.
- [282] D. Tarasov, D. Tovbin, *J Mol Model* **2009**, 15, 329-341.
- [283] A. N. Jain, *Journal of Medicinal Chemistry* **2003**, 46, 499-511.
- [284] T. Schulz-Gasch, M. Stahl, *Drug Discovery Today: Technologies* **2004**, 1, 231-239.
- [285] P. D. Thomas, K. A. Dill, *Proceedings of the National Academy of Sciences* **1996**, 93, 11628-11633.
- [286] P. D. Thomas, K. A. Dill, *Journal of Molecular Biology* **1996**, 257, 457-469.
- [287] al. Muegge, in *Virtual Screening: An Alternative or Complement to High Throughput Screening?*, Vol. 20 (Ed.: G. Klebe), Springer Netherlands, **2002**, pp. 99-114; bl. Muegge, *Journal of Computational Chemistry* **2001**, 22, 418-425; cl. Muegge, Y. C. Martin, *Journal of Medicinal Chemistry* **1999**, 42, 791-804.
- [288] H. Gohlke, M. Hendlich, G. Klebe, *Journal of Molecular Biology* **2000**, 295, 337-356.
- [289] R. S. DeWitte, E. I. Shakhnovich, *Journal of the American Chemical Society* **1996**, 118, 11733-11744.
- [290] aS.-Y. Huang, X. Zou, *Journal of Computational Chemistry* **2006**, 27, 1876-1882; bS.-Y. Huang, X. Zou, *Journal of Chemical Information and Modeling* **2010**, 50, 262-273.
- [291] aR. Wang, S. Wang, *Journal of Chemical Information and Computer Sciences* **2001**, 41, 1422-1426; bR. D. Clark, A. Strizhev, J. M. Leonard, J. F. Blake, J. B. Matthew, *Journal of Molecular Graphics and Modelling* **2002**, 20, 281-295.
- [292] A. Oda, K. Tsuchida, T. Takakura, N. Yamaotsu, S. Hirono, *Journal of Chemical Information and Modeling* **2005**, 46, 380-391.
- [293] G. E. Terp, B. N. Johansen, I. T. Christensen, F. S. Jørgensen, *Journal of Medicinal Chemistry* **2001**, 44, 2333-2343.
- [294] L. L. Wang R, Wang S., *J Comput Aided Mol Des*. **2002**, 16, 11-26.
- [295] S. Bar-Haim, A. Aharon, T. Ben-Moshe, Y. Marantz, H. Senderowitz, *Journal of Chemical Information and Modeling* **2009**, 49, 623-633.
- [296] R. Teramoto, H. Fukunishi, *Journal of Chemical Information and Modeling* **2007**, 47, 526-534.
- [297] aF. Shah, J. Gut, J. Legac, D. Shivakumar, W. Sherman, P. J. Rosenthal, M. A. Avery, *Journal of Chemical Information and Modeling* **2012**, 52, 696-710; bH. Li, J. Huang, L. Chen, X. Liu, T. Chen, J. Zhu, W. Lu, X. Shen, J. Li, R. Hilgenfeld, H. Jiang, *Journal of Medicinal Chemistry* **2009**, 52, 4936-4940; cA. C. Pierce, M. Jacobs, C. Stuver-Moody, *Journal of Medicinal Chemistry* **2008**, 51, 1972-1975.
- [298] R. Jüttermann, E. Li, R. Jaenisch, *Proceedings of the National Academy of Sciences* **1994**, 91, 11797-11801.

- [299] P. Siedlecki, R. G. Boy, S. Comagic, R. Schirmacher, M. Wiessler, P. Zielenkiewicz, S. Suhai, F. Lyko, *Biochemical and Biophysical Research Communications* **2003**, *306*, 558-563.
- [300] C. A., *Oncology* **1974**, *30*, 405-422.
- [301] M. L. Momparler RL, Samson J., *Leuk Res.* **1984**, *8*, 1043-1049.
- [302] E. Cornacchia, J. Golbus, J. Maybaum, J. Strahler, S. Hanash, B. Richardson, *The Journal of Immunology* **1988**, *140*, 2197-2200.
- [303] N. Singh, A. Dueñas-González, F. Lyko, J. L. Medina-Franco, *ChemMedChem* **2009**, *4*, 792-799.
- [304] D. Kuck, N. Singh, F. Lyko, J. L. Medina-Franco, *Bioorganic & Medicinal Chemistry* **2010**, *18*, 822-829.
- [305] N. C. I. h. d. c. g. a. S. Developmental Therapeutics Program.
- [306] J. Medina-Franco, F. López-Vallejo, D. Kuck, F. Lyko, *Mol Divers* **2011**, *15*, 293-304.
- [307] J. J. Irwin, B. K. Shoichet, *Journal of Chemical Information and Modeling* **2004**, *45*, 177-182.
- [308] J. Yoo, S. Choi, J. L. Medina-Franco, *PLoS ONE* **2013**, *8*, e62152.
- [309] J. Yoo, J. H. Kim, K. D. Robertson, J. L. Medina-Franco, in *Advances in Protein Chemistry and Structural Biology, Vol. Volume 87* (Eds.: C. Christo, K.-C. Tatyana), Academic Press, **2012**, pp. 219-247.
- [310] aX. Cheng, R. M. Blumenthal, *Structure (London, England : 1993)* **2008**, *16*, 341-350; bH. S. Lan J, He X, Zhang Y., *Acta Biochim Biophys Sin (Shanghai)* **2010**, *42*, 243-252; cW. Sippl, M. Jung, *Epigenetic Targets in Drug Discovery*, Wiley, **2009**.
- [311] J. Yoo, J. Medina-Franco, *J Comput Aided Mol Des* **2011**, *25*, 555-567.
- [312] N. K. Salam, R. Nuti, W. Sherman, *Journal of Chemical Information and Modeling* **2009**, *49*, 2356-2368.
- [313] D. Evans, A. Bronowska, *Theor Chem Acc* **2010**, *125*, 407-418.
- [314] J. D. Kalpana Ghoshal, Sarmila Majumder, Shoumei Bai, Huban Kutay, Tasneem Motiwala, and Samson T. Jacob, *Mol Cell Biol* **2005**, *25*, 4727-4741.
- [315] aY. Ye, J. T. Stivers, *Analytical Biochemistry* **2010**, *401*, 168-172; bJ. Medina-Franco, J. Yoo, *Mol Divers* **2013**, *17*, 337-344.
- [316] L. Halby, C. Champion, C. Sénamaud-Beaufort, S. Ajjan, T. Drujon, A. Rajavelu, A. Ceccaldi, R. Jurkowska, O. Lequin, W. G. Nelson, A. Guy, A. Jeltsch, D. Guianvarc'h, C. Ferroud, P. B. Arimondo, *ChemBioChem* **2012**, *13*, 157-165.
- [317] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson, *Journal of Computational Chemistry* **2009**, *30*, 2785-2791.
- [318] B. S. Jursic, S. Sagiraju, D. K. Ancalade, T. Clark, E. D. Stevens, *Synthetic Communications* **2007**, *37*, 1709-1714.
- [319] T. Eckert, J. Ipaktschi, *Synthetic Communications* **1998**, *28*, 327-335.
- [320] C. Riedinger, J. A. Endicott, S. J. Kemp, L. A. Smyth, A. Watson, E. Valeur, B. T. Golding, R. J. Griffin, I. R. Hardcastle, M. E. Noble, J. M. McDonnell, *Journal of the American Chemical Society* **2008**, *130*, 16038-16044.
- [321] J. Drews, *Science* **2000**, *287*, 1960-1964.
- [322] G. Wolber, T. Langer, *Journal of Chemical Information and Modeling* **2004**, *45*, 160-169.
- [323] C. H. Hansch, A. Leo, D. H. Hoekman, *Exploring Qsar: Fundamentals and Applications in Chemistry and Biology*, American Chemical Society, **1995**.
- [324] L. A. O. B. Jeffrey J. Sutherland, and, and Donald F. Weaver, *Journal of Medicinal Chemistry* **2004**, *47*, 5541-5554.
- [325] J. C. W. a. D. V. Santi, *J. Biol. Chem.* **1987**, *262*, 4778-4786.
- [326] S. K. B. a. A. K. Dubey, *J. Biol. Chem.* **1999**, *274*, 14743-14749.
- [327] K. M. D. Edward J. O'Neil, and, and Bradley D. Smith, *Organic Letters* **2007**, *9*, 199-202.

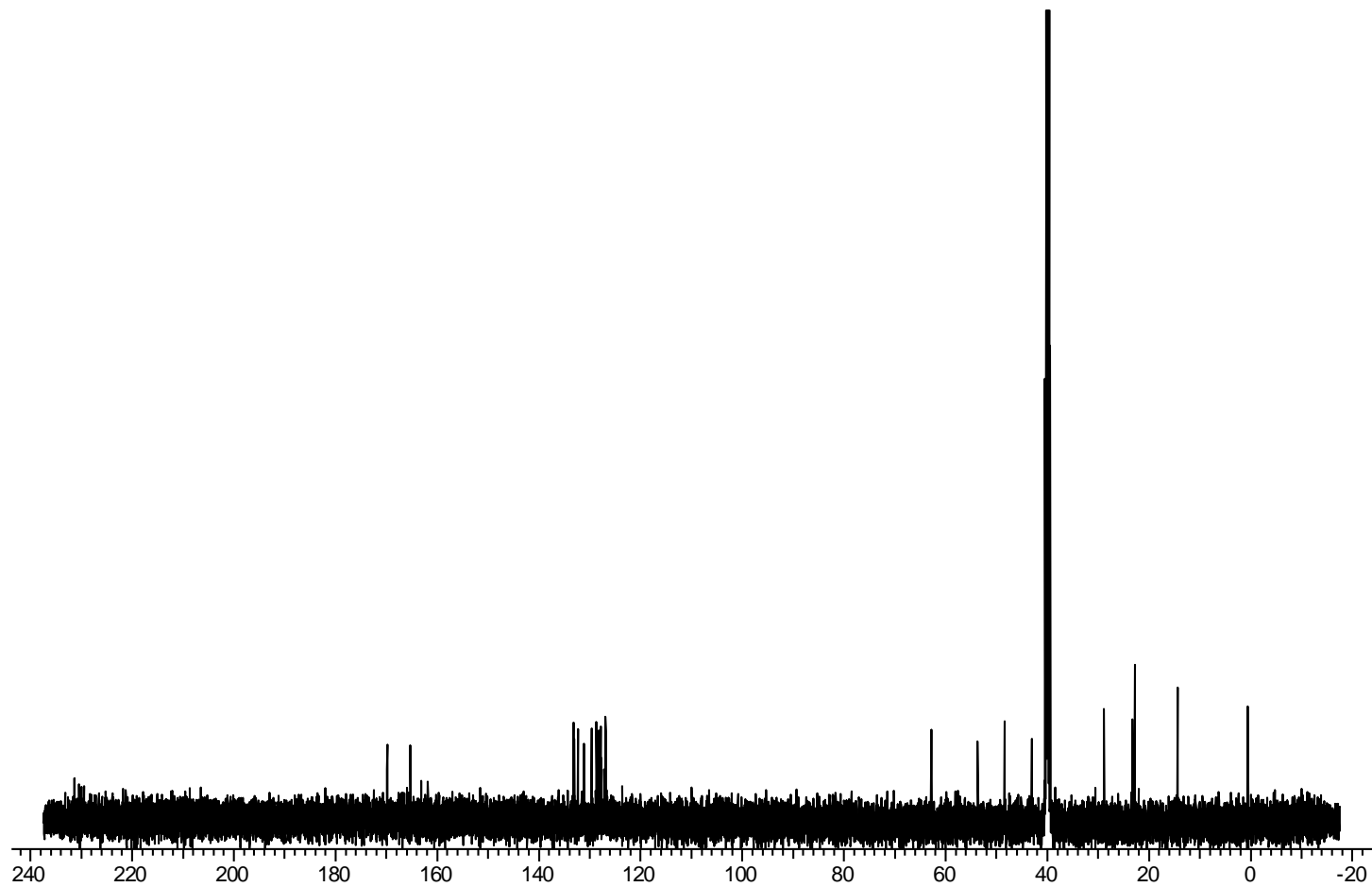
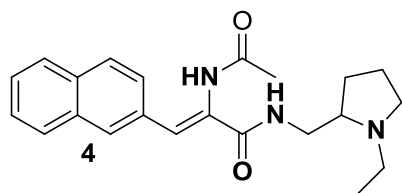
- [328] V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angewandte Chemie International Edition* **2002**, *41*, 2596-2599.
- [329] H. Yamamoto, T. Terasawa, A. Nakamura, K. Kawabata, H. Takasugi, H. Tanaka, S. Matsumoto, Y. Matsumoto, S. Tawara, *Bioorganic & Medicinal Chemistry* **2001**, *9*, 465-475.
- [330] S. Braverman, M. Cherkinsky, D. Meridor, M. Sprecher, *Tetrahedron* **2010**, *66*, 1925-1930.
- [331] K. E. Yelm, *Tetrahedron Letters* **1999**, *40*, 1101-1102.
- [332] H. Lusic, D. D. Young, M. O. Lively, A. Deiters, *Organic Letters* **2007**, *9*, 1903-1906.
- [333] B. Bornemann, A. Marx, *Bioorganic & medicinal chemistry* **2006**, *14*, 6235-6238.
- [334] W. H. Prusoff, D. C. Ward, *Biochemical Pharmacology* **1976**, *25*, 1233-1239.
- [335] J. A. Mekras, D. A. Boothman, L. M. Perez, S. Greer, *Cancer Research* **1984**, *44*, 2551-2560.
- [336] V. V. Filichev, E. B. Pedersen, *Tetrahedron* **2001**, *57*, 9163-9168.
- [337] J. A. Stefely, R. Palchaudhuri, P. A. Miller, R. J. Peterson, G. C. Moraski, P. J. Hergenrother, M. J. Miller, *Journal of Medicinal Chemistry* **2010**, *53*, 3389-3395.
- [338] A. D. Ward, B. R. Baker, *Journal of Medicinal Chemistry* **1977**, *20*, 88-92.
- [339] W. H. Prusoff, Ward, D.C., *Biochemical Pharmacology* **1976**, *25*, 1233-1239.
- [340] R. L. Rubin, *Toxicology* **2005**, *209*, 135-147.
- [341] N. T. Eldredge, W. van B. Robertson, J. J. Miller lii, *Clinical Immunology and Immunopathology* **1974**, *3*, 263-271.
- [342] C. S. A. Kumar, S. B. B. Prasad, K. Vinaya, S. Chandrappa, N. R. Thimmegowda, Y. C. S. Kumar, S. Swarup, K. S. Rangappa, *European Journal of Medicinal Chemistry* **2009**, *44*, 1223-1229.
- [343] S.-L. Cao, Y. Han, C.-Z. Yuan, Y. Wang, Z.-K. Xiahou, J. Liao, R.-T. Gao, B.-B. Mao, B.-L. Zhao, Z.-F. Li, X. Xu, *European Journal of Medicinal Chemistry* **2013**, *64*, 401-409.
- [344] A. F. Watson, J. Liu, K. Bennaceur, C. J. Drummond, J. A. Endicott, B. T. Golding, R. J. Griffin, K. Haggerty, X. Lu, J. M. McDonnell, D. R. Newell, M. E. M. Noble, C. H. Revill, C. Riedinger, Q. Xu, Y. Zhao, J. Lunec, I. R. Hardcastle, *Bioorganic & Medicinal Chemistry Letters* **2011**, *21*, 5916-5919.
- [345] Serge Y Fuchs, Victor Adler, Thomas Buschmann, X. Wu, Z. e. Ronai, *Oncogene* **1998**, *17*, 2543-2547.
- [346] in *Encyclopædia Britannica*, Encyclopædia Britannica Online, **2009**.
- [347] R. Skouta, M. Hayano, K. Shimada, B. R. Stockwell, *Bioorganic & Medicinal Chemistry Letters* **2012**, *22*, 5707-5713.
- [348] S. Xiao, W. Lin, C. Wang, M. Yang, *Bioorganic & Medicinal Chemistry Letters* **2001**, *11*, 437-441.
- [349] A. M. Deveau, M. A. Labroli, C. M. Dieckhaus, M. T. Barthen, K. S. Smith, T. L. Macdonald, *Bioorganic & Medicinal Chemistry Letters* **2001**, *11*, 1251-1255.



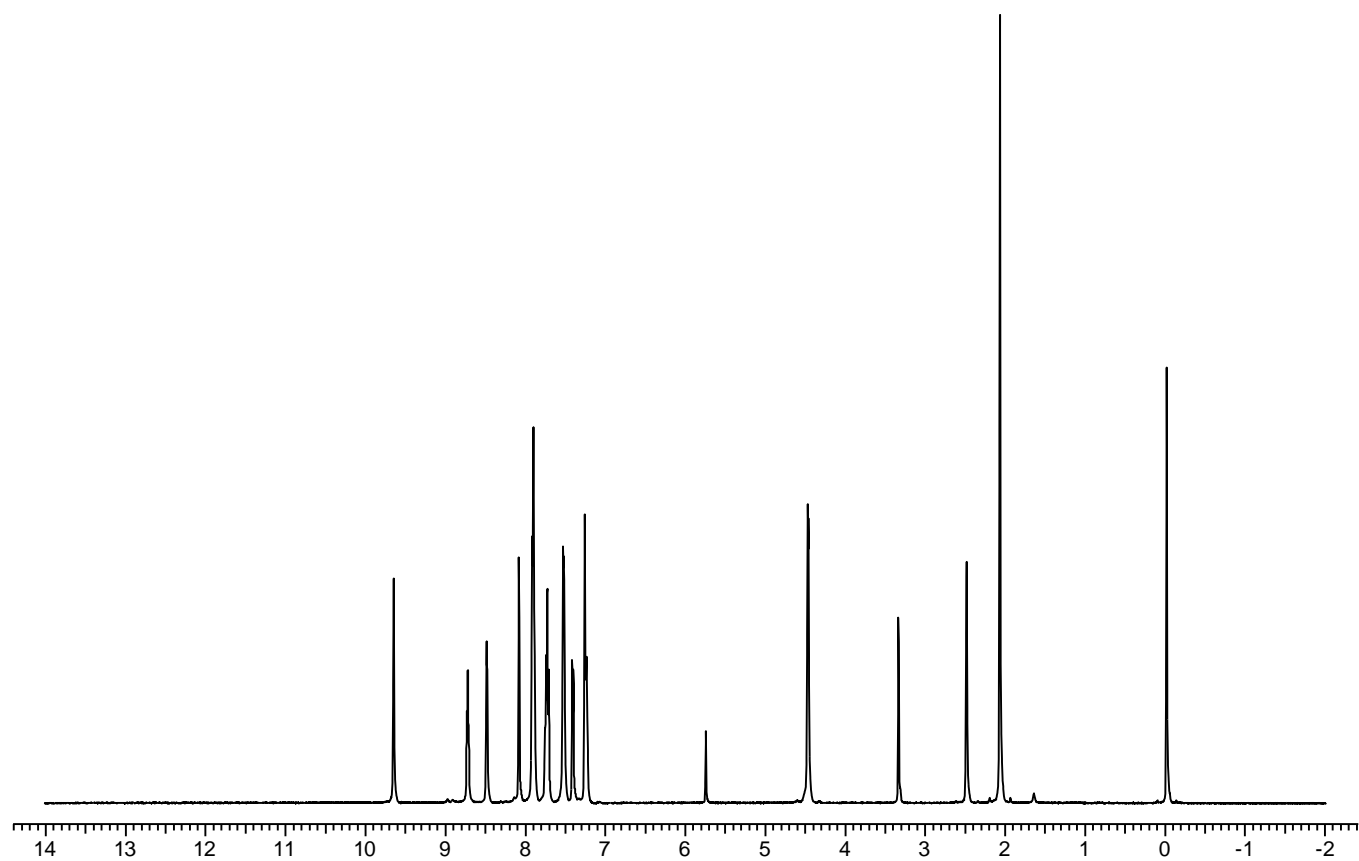
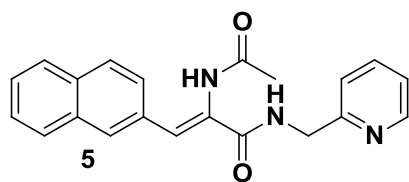
## APPENIX: SELECTED NMR SPECTRA FOR SYNTHESIZED COMPOUNDS



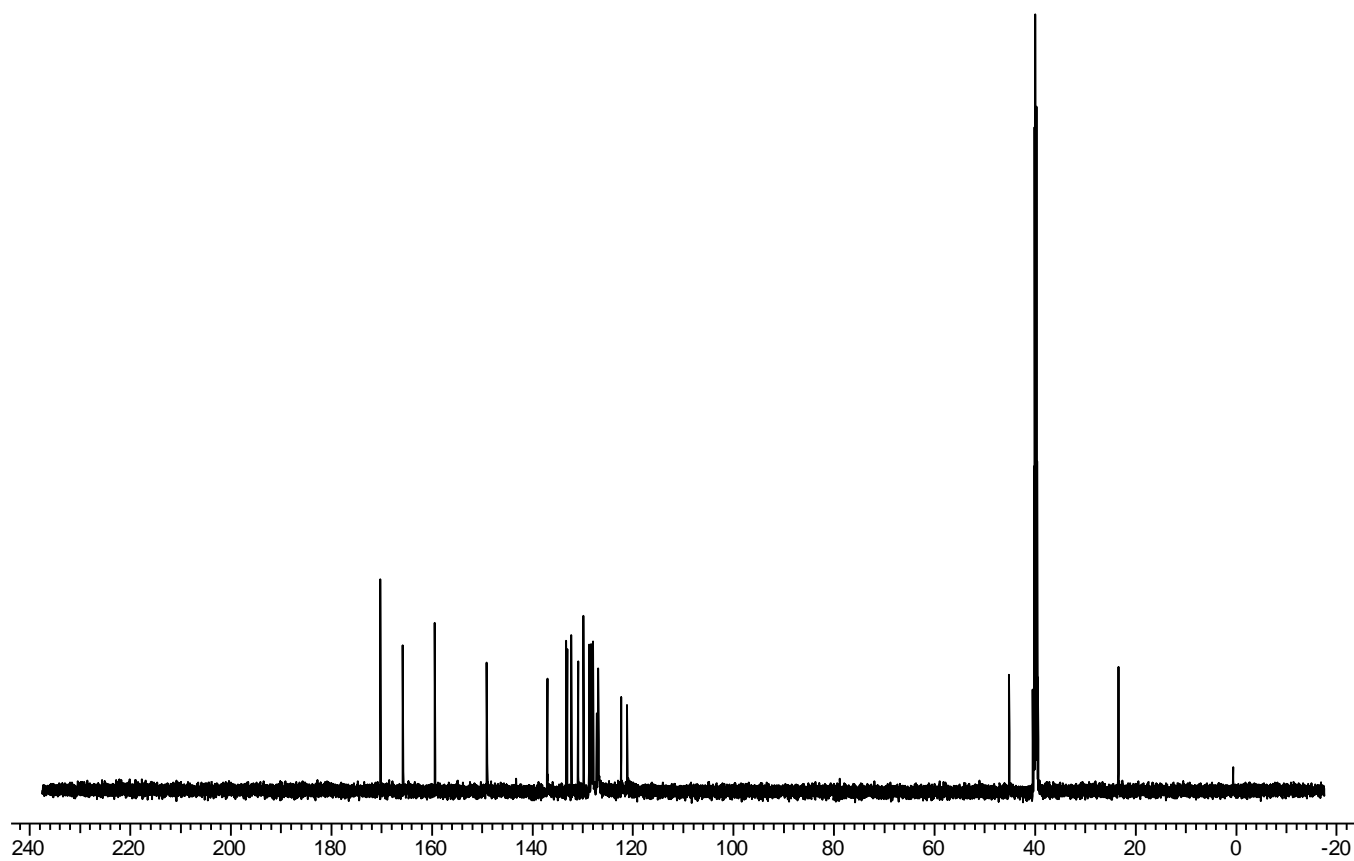
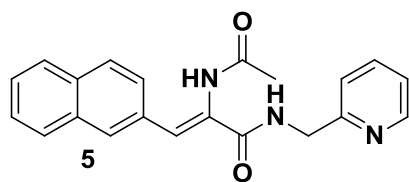
4: <sup>1</sup>H NMR



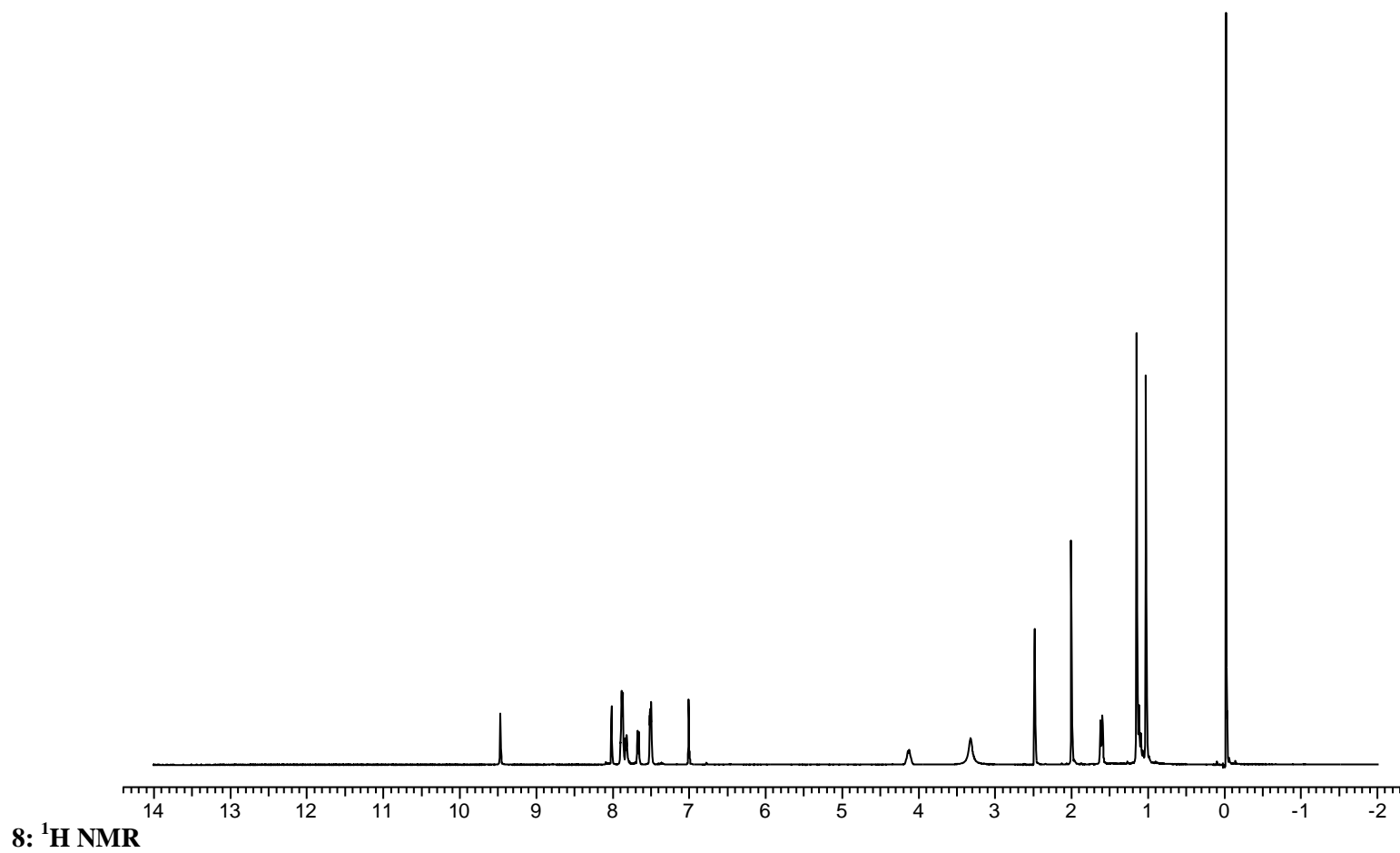
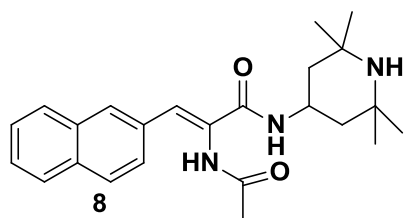
4: <sup>13</sup>C NMR

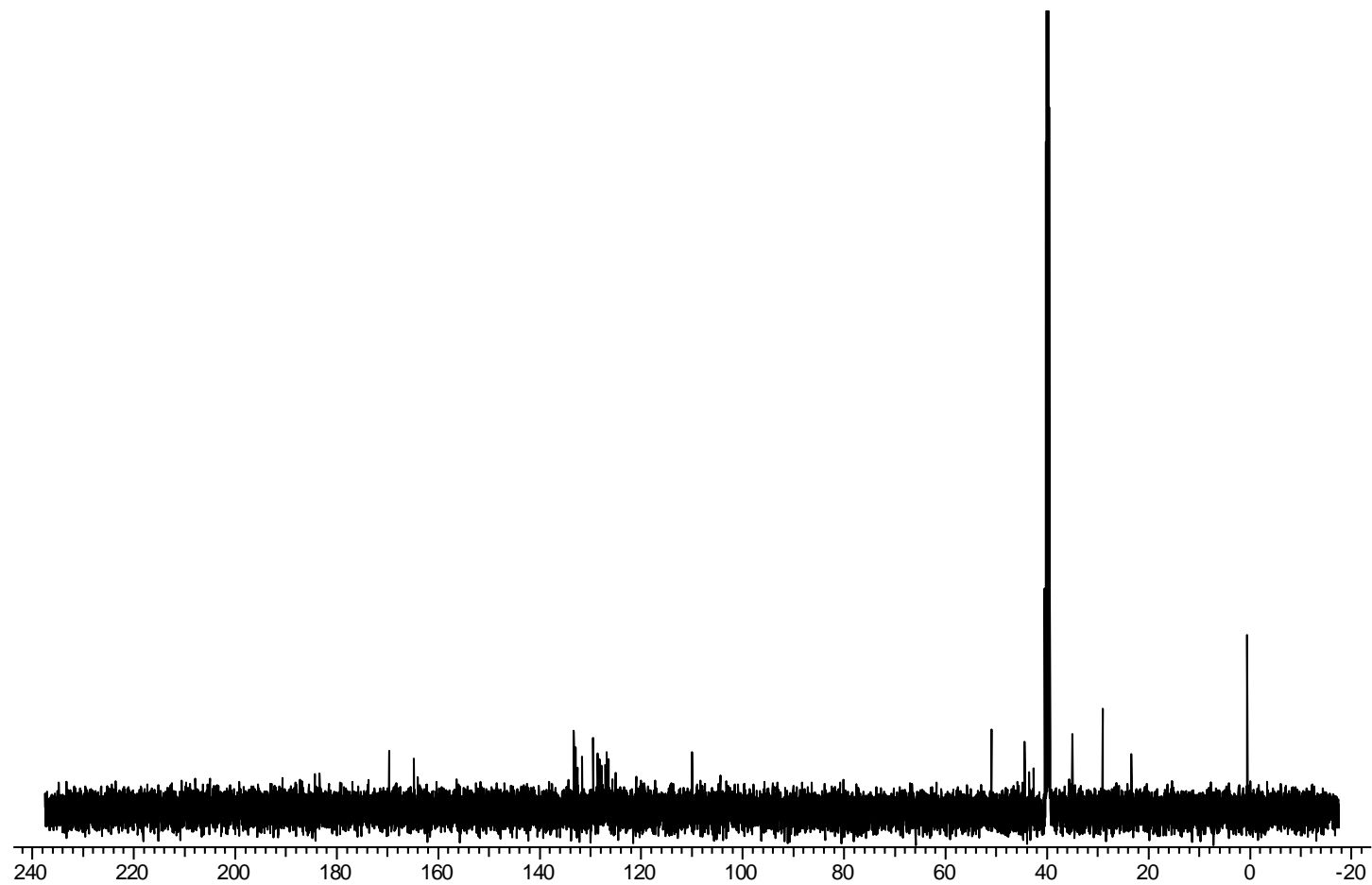
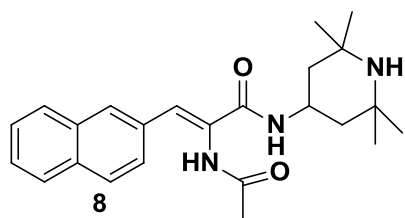


5: <sup>1</sup>H NMR

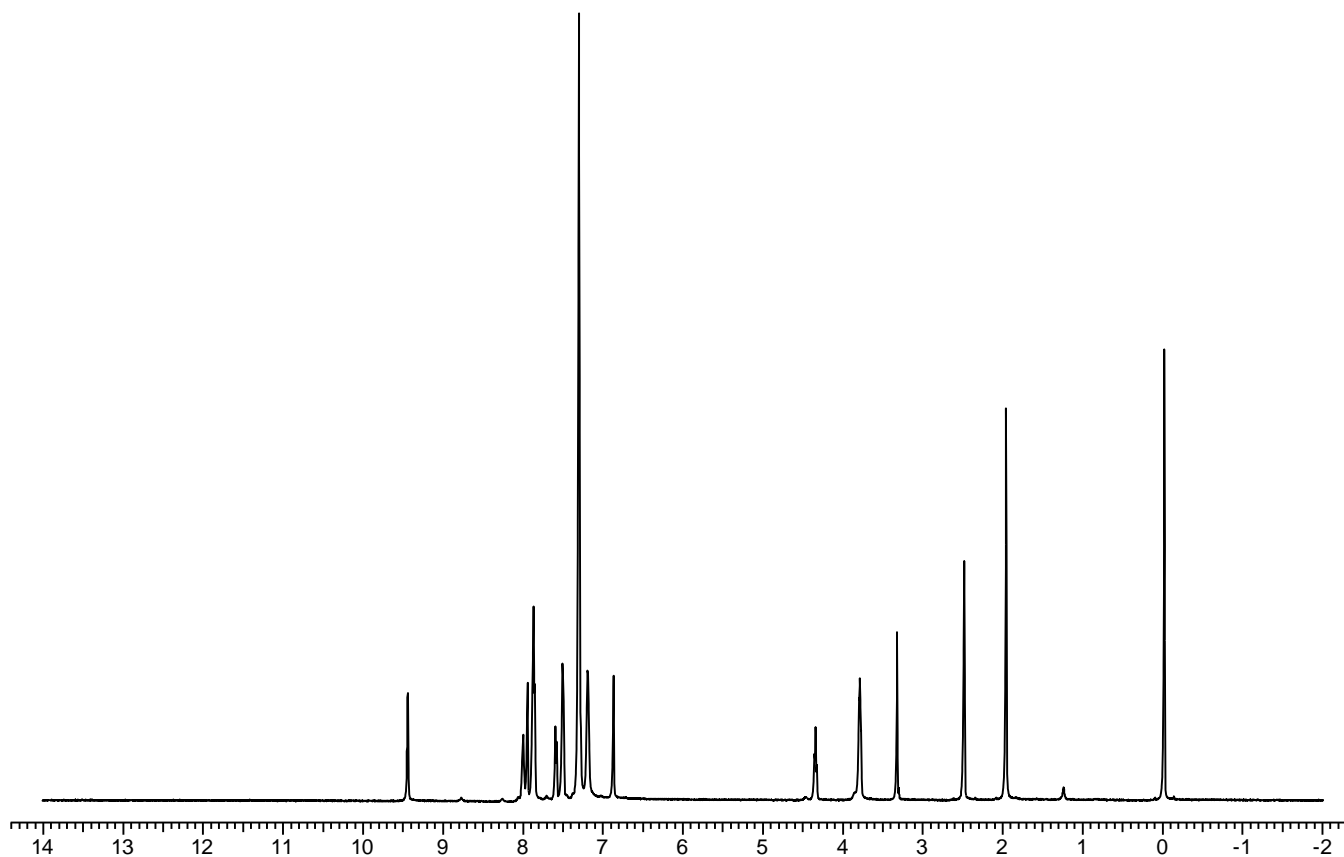
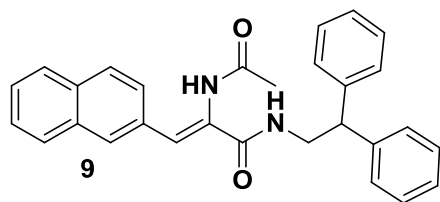


5:  $^{13}\text{C}$  NMR



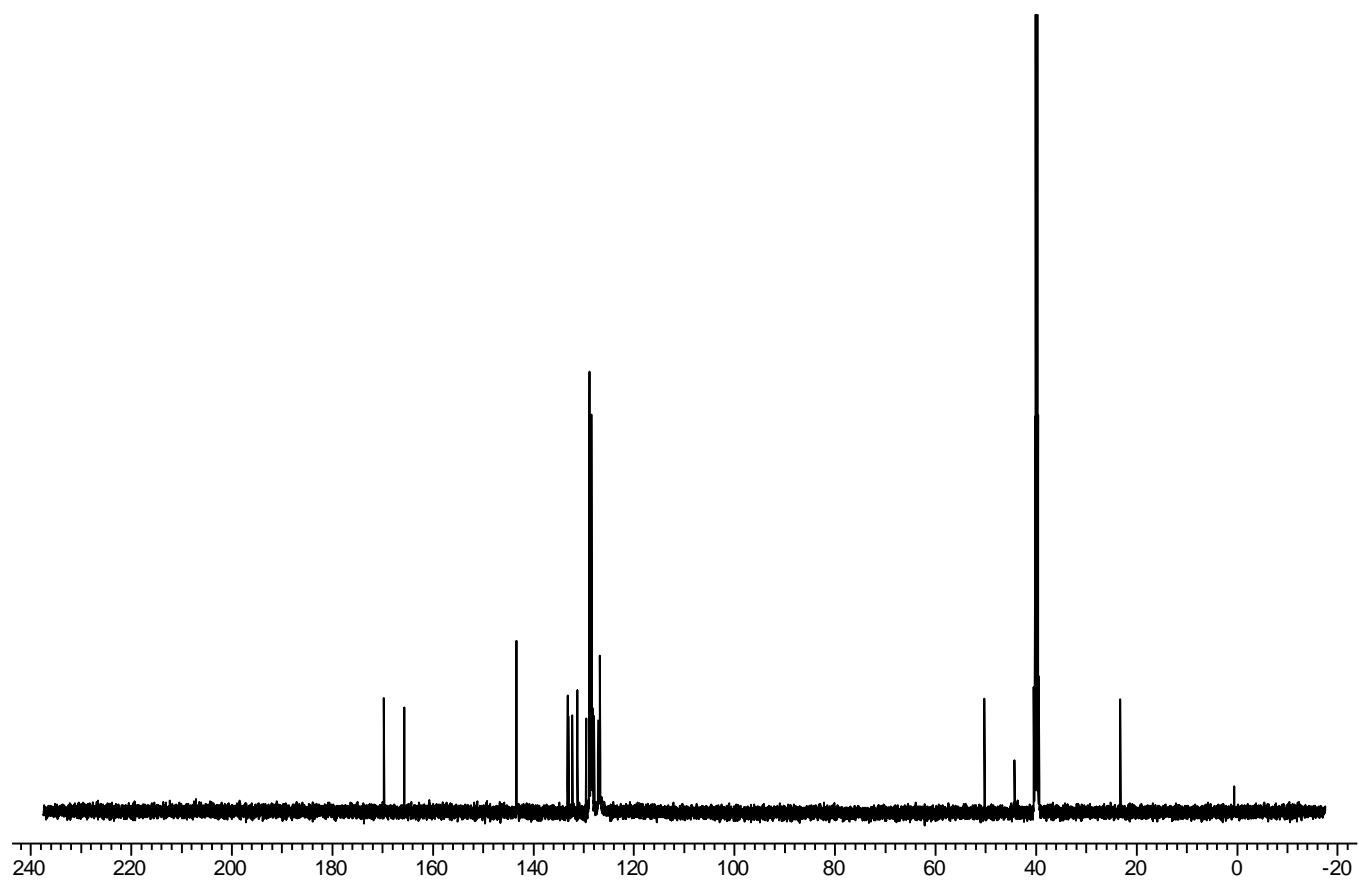
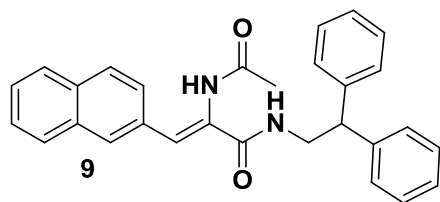


8:  $^{13}\text{C}$  NMR

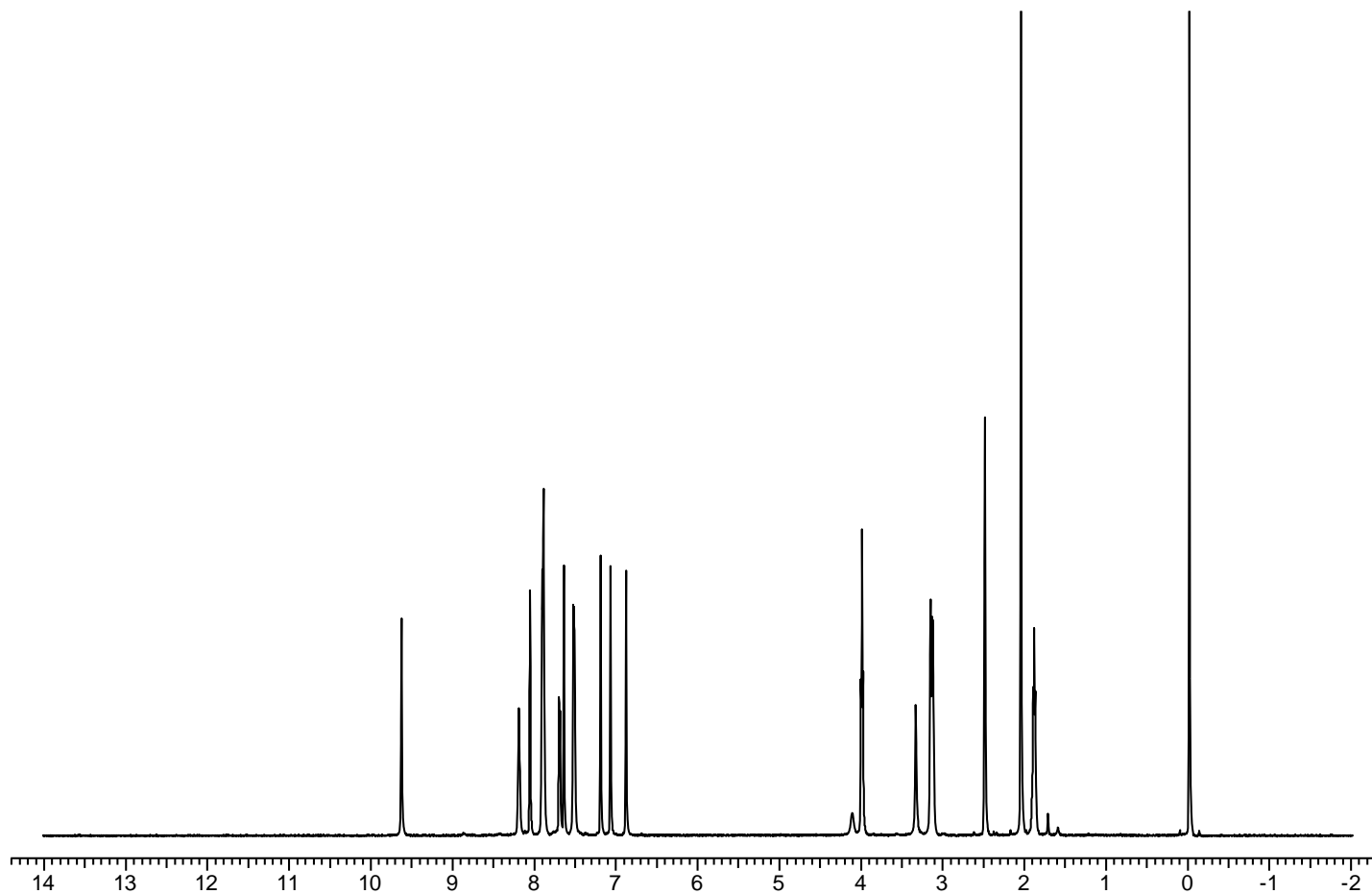
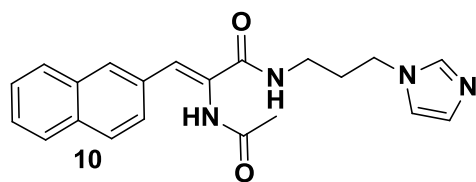


9:  $^1\text{H}$  NMR

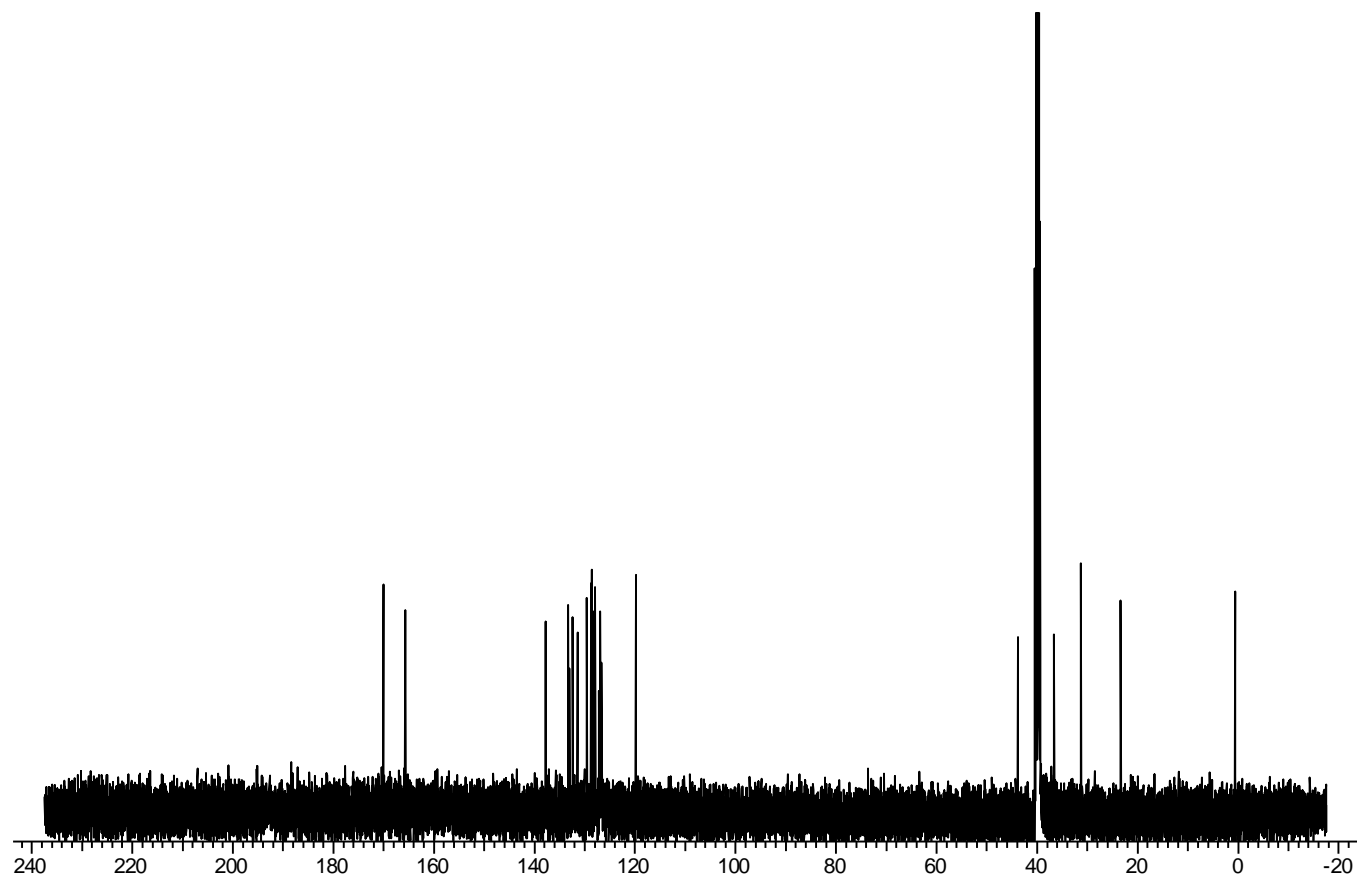
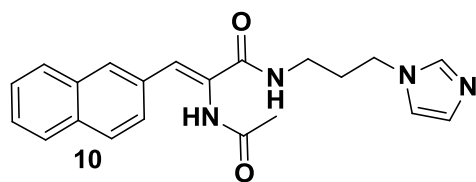




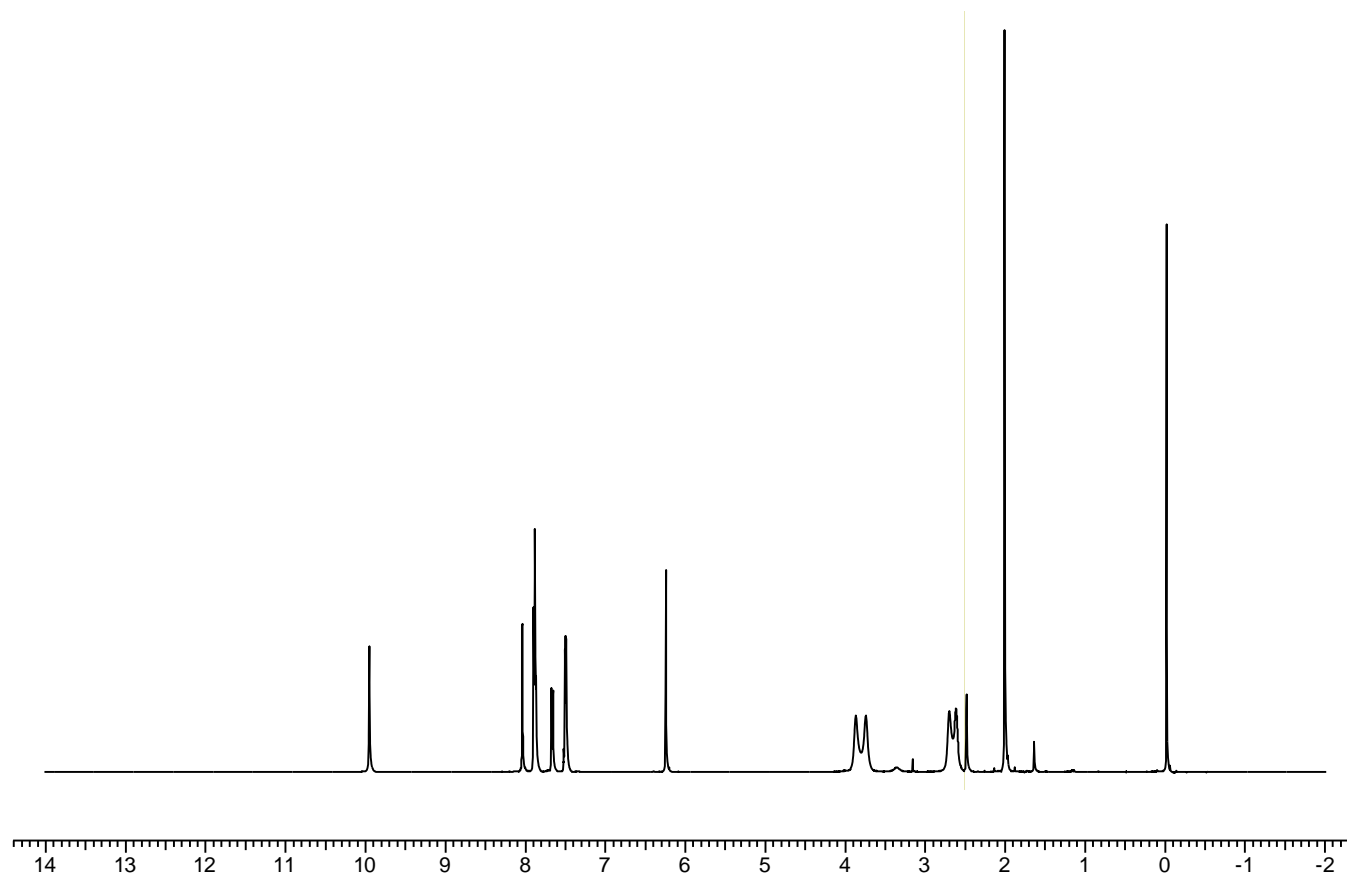
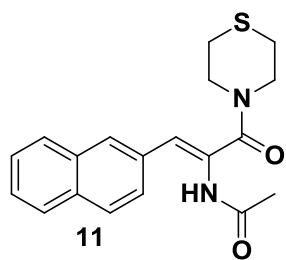
9:  $^{13}\text{C}$  NMR



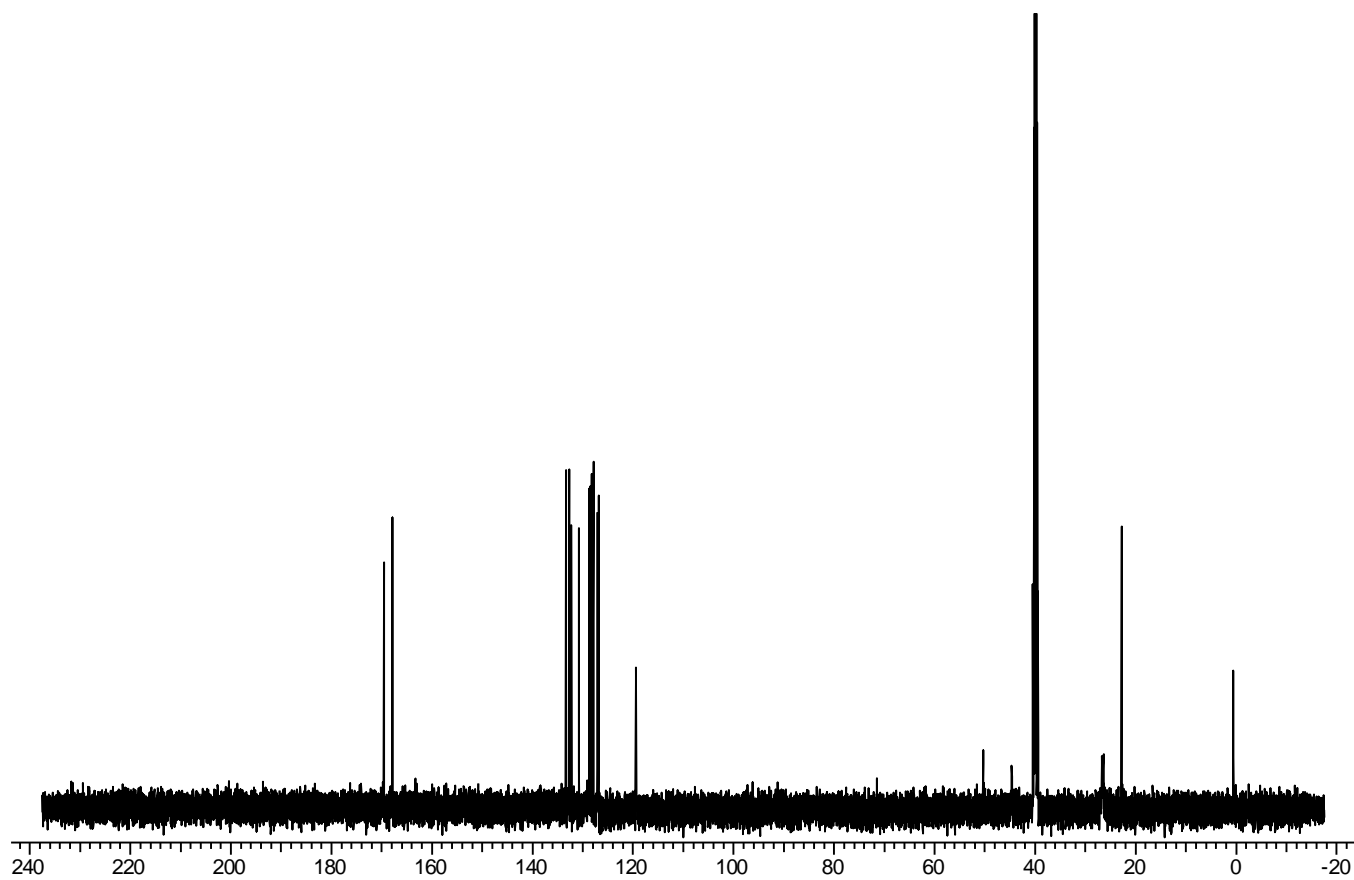
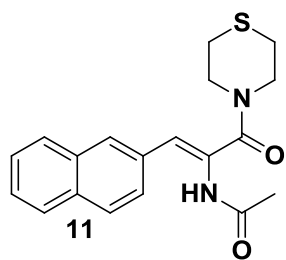
**10: <sup>1</sup>H NMR**



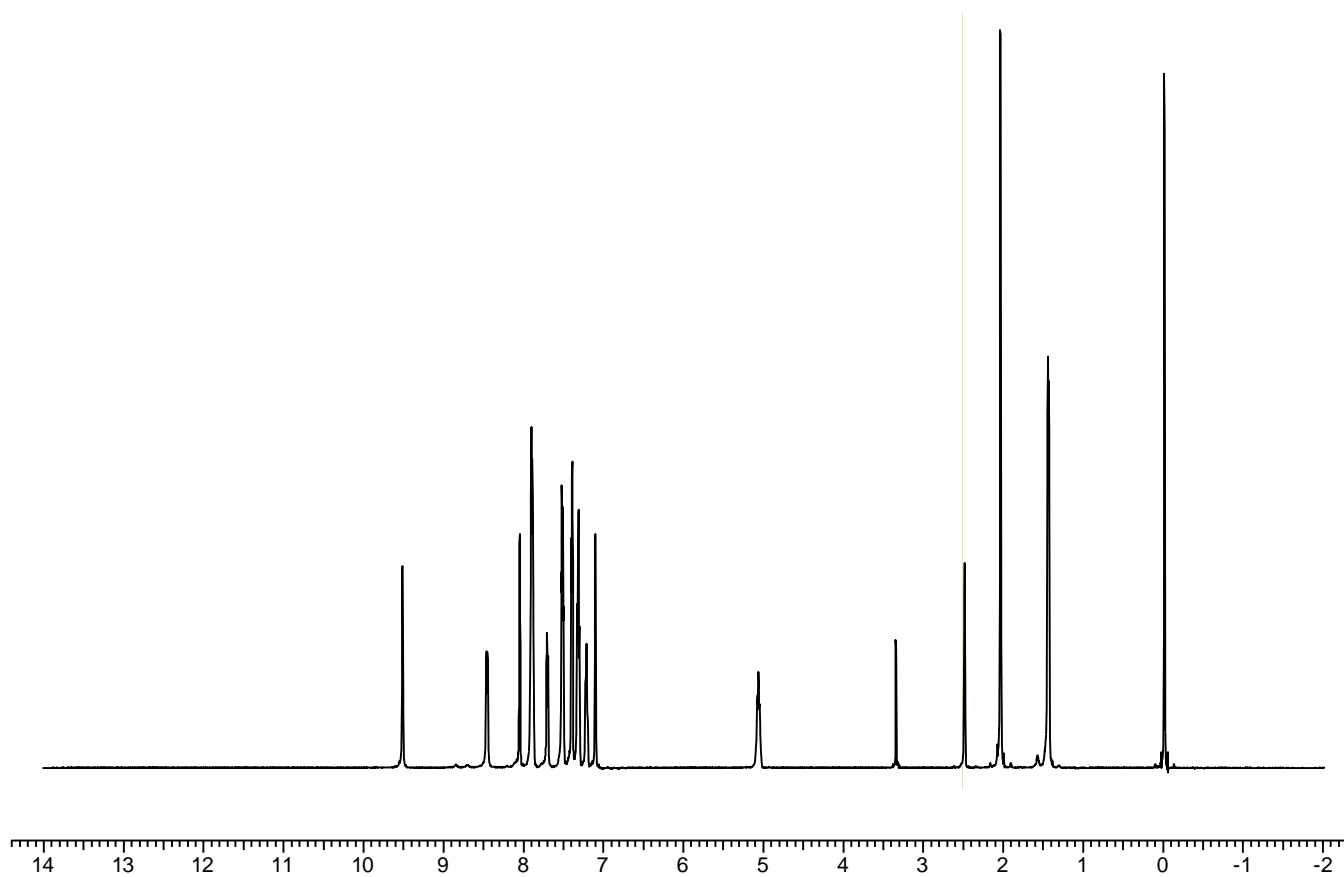
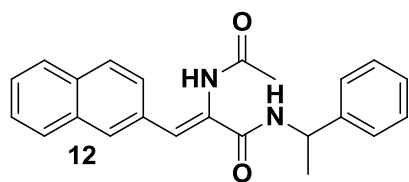
**10:**  $^{13}\text{C}$  NMR



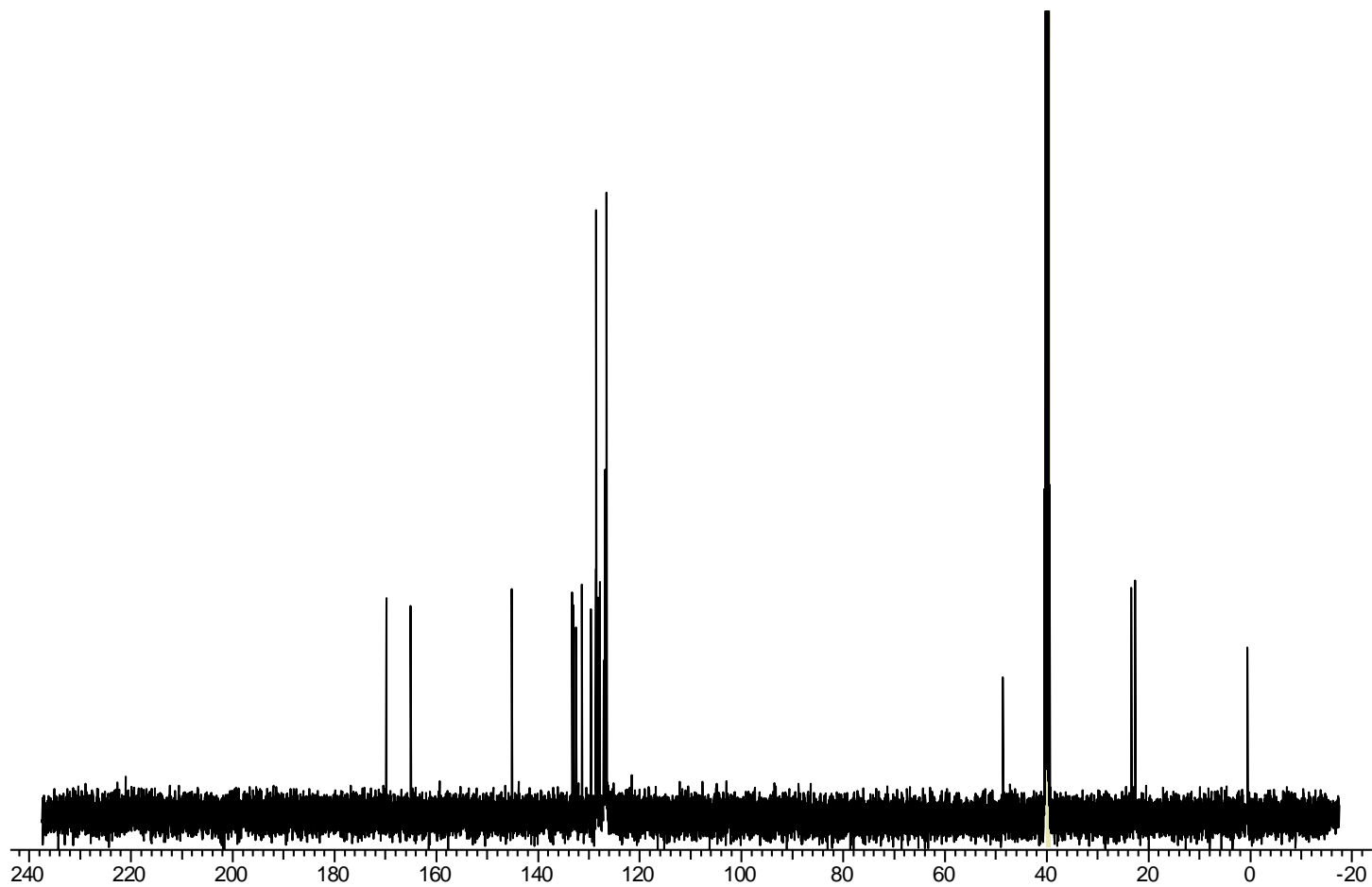
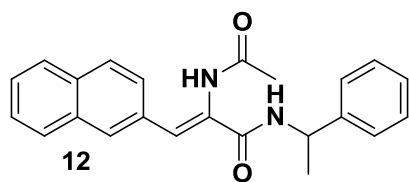
11:  $^1\text{H}$  NMR



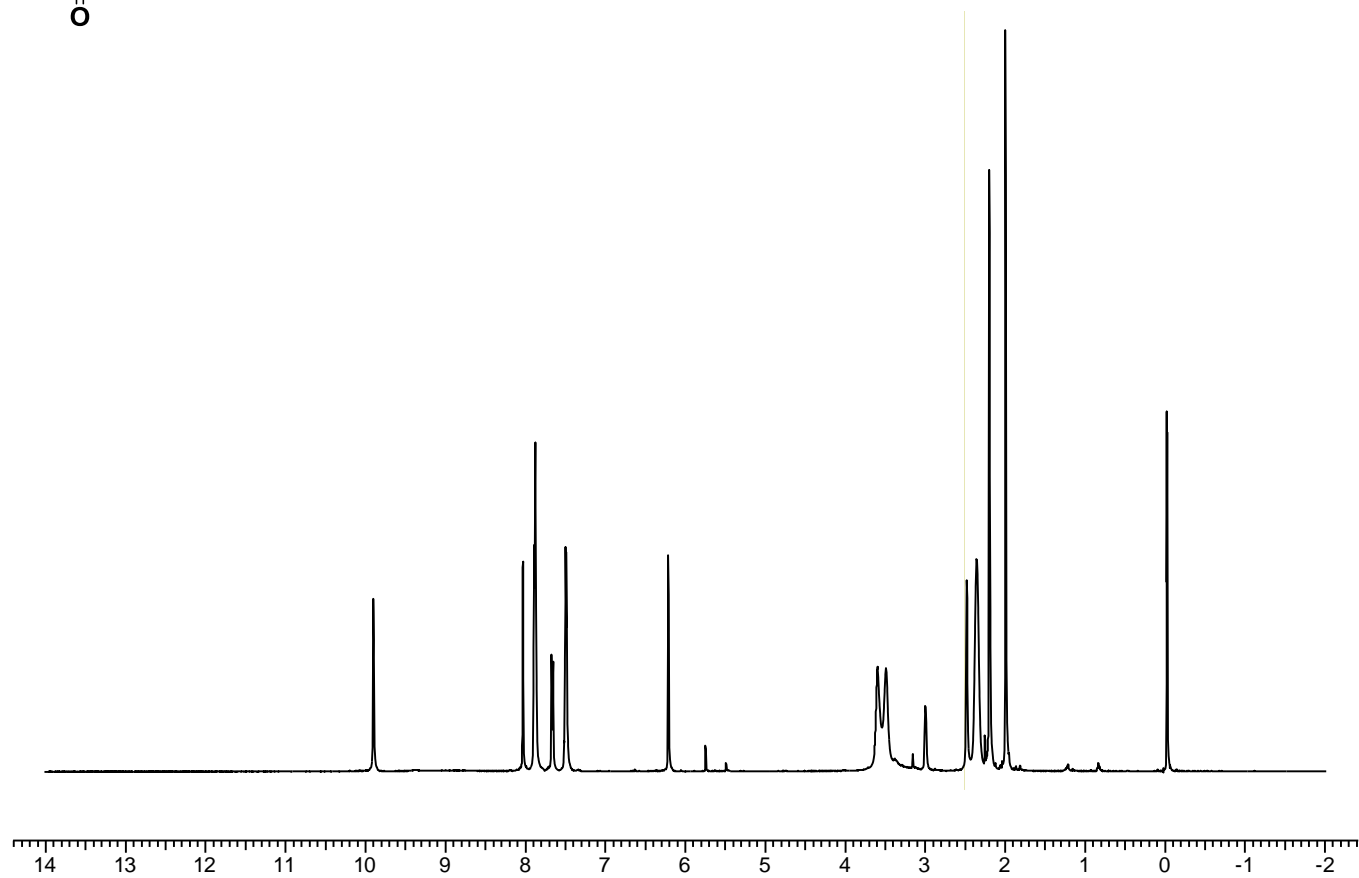
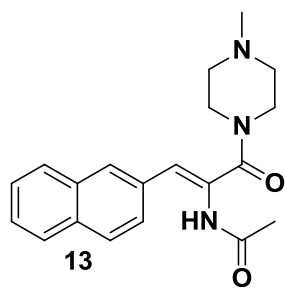
11:  $^{13}\text{C}$  NMR



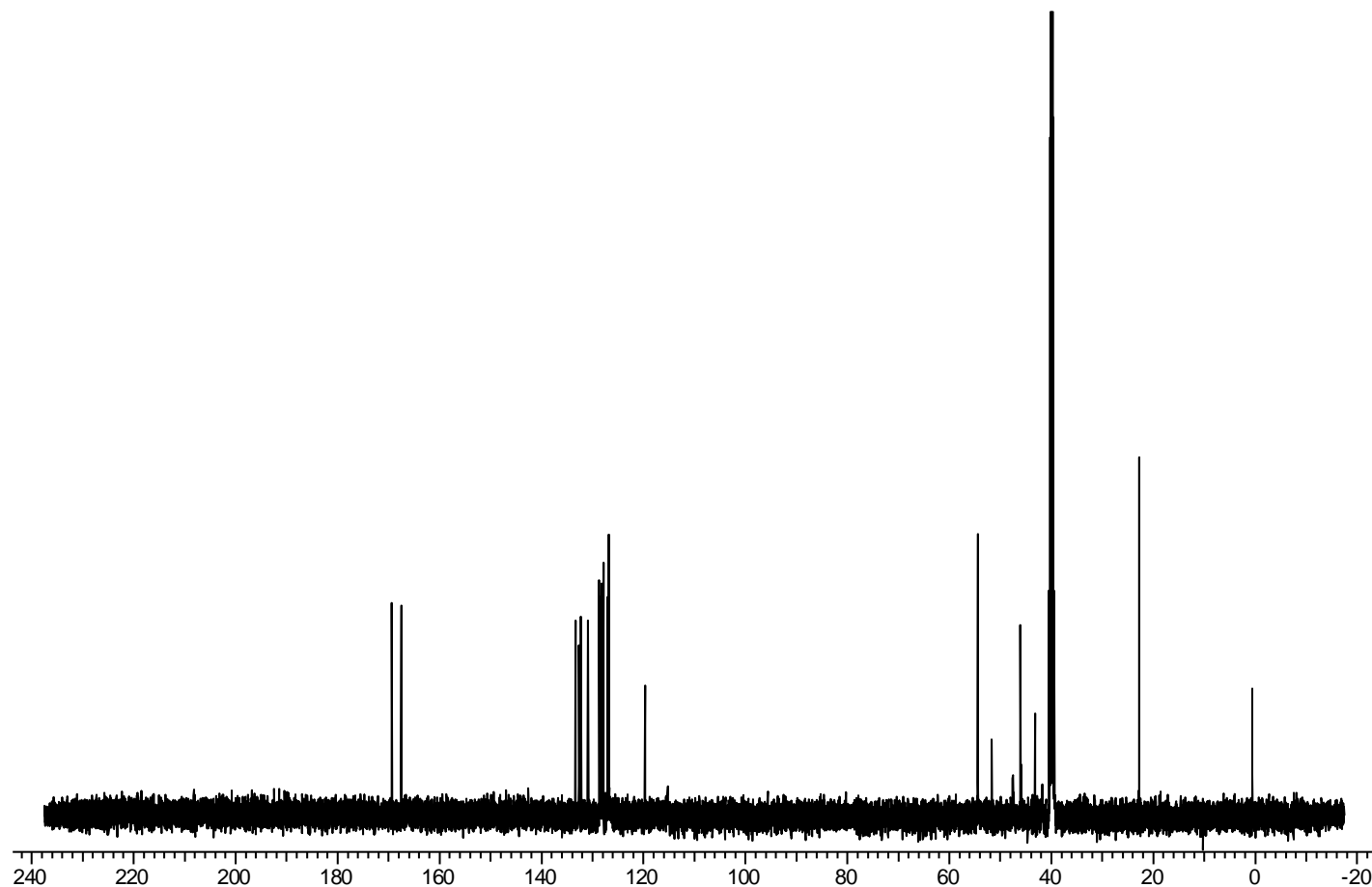
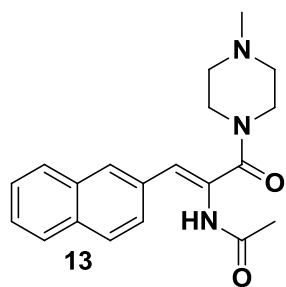
12: <sup>1</sup>H NMR



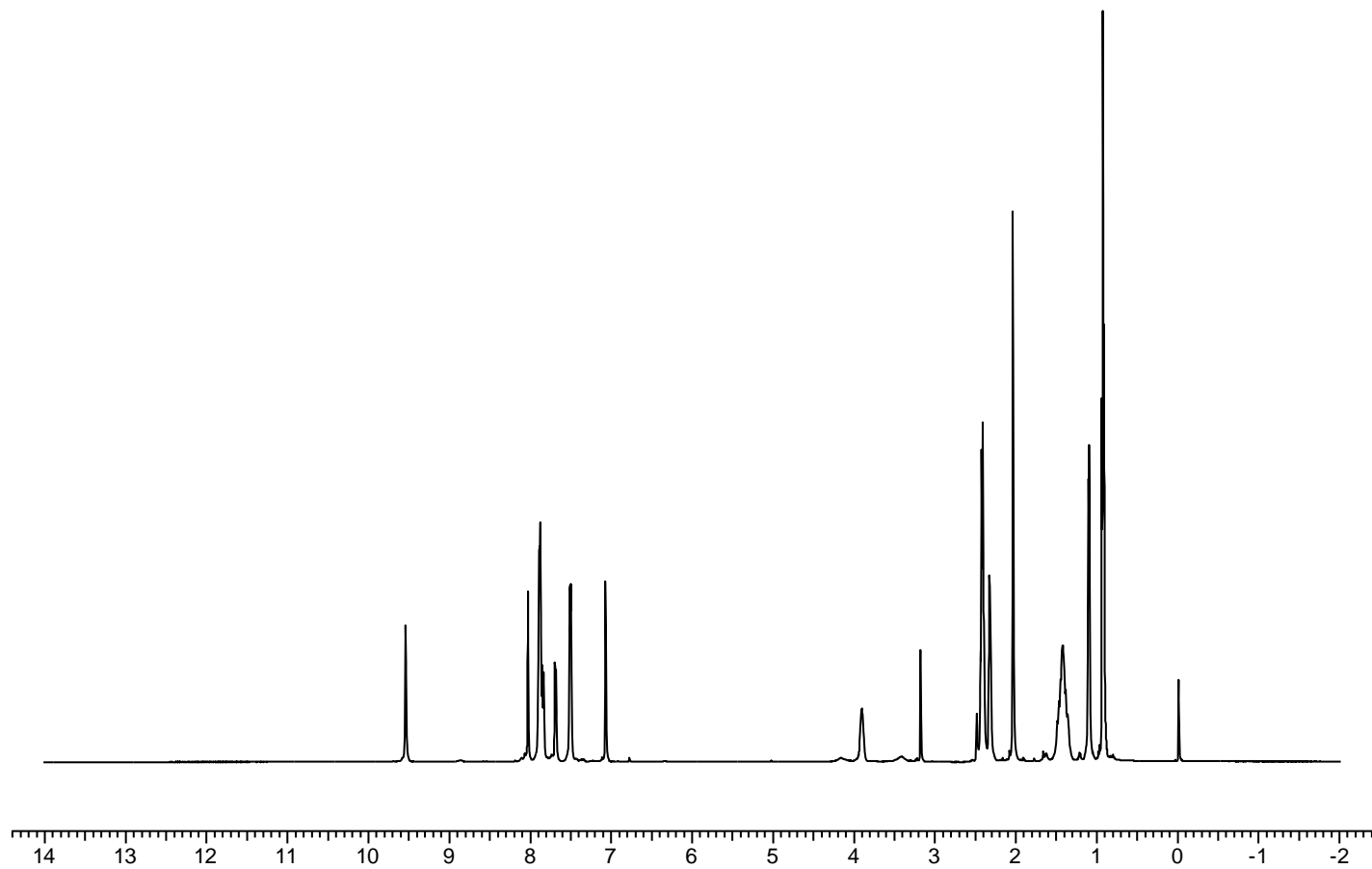
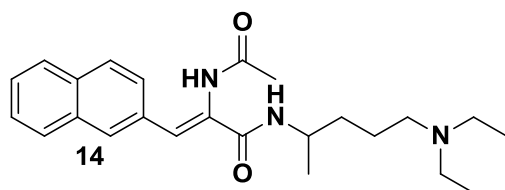
12:  $^{13}\text{C}$  NMR



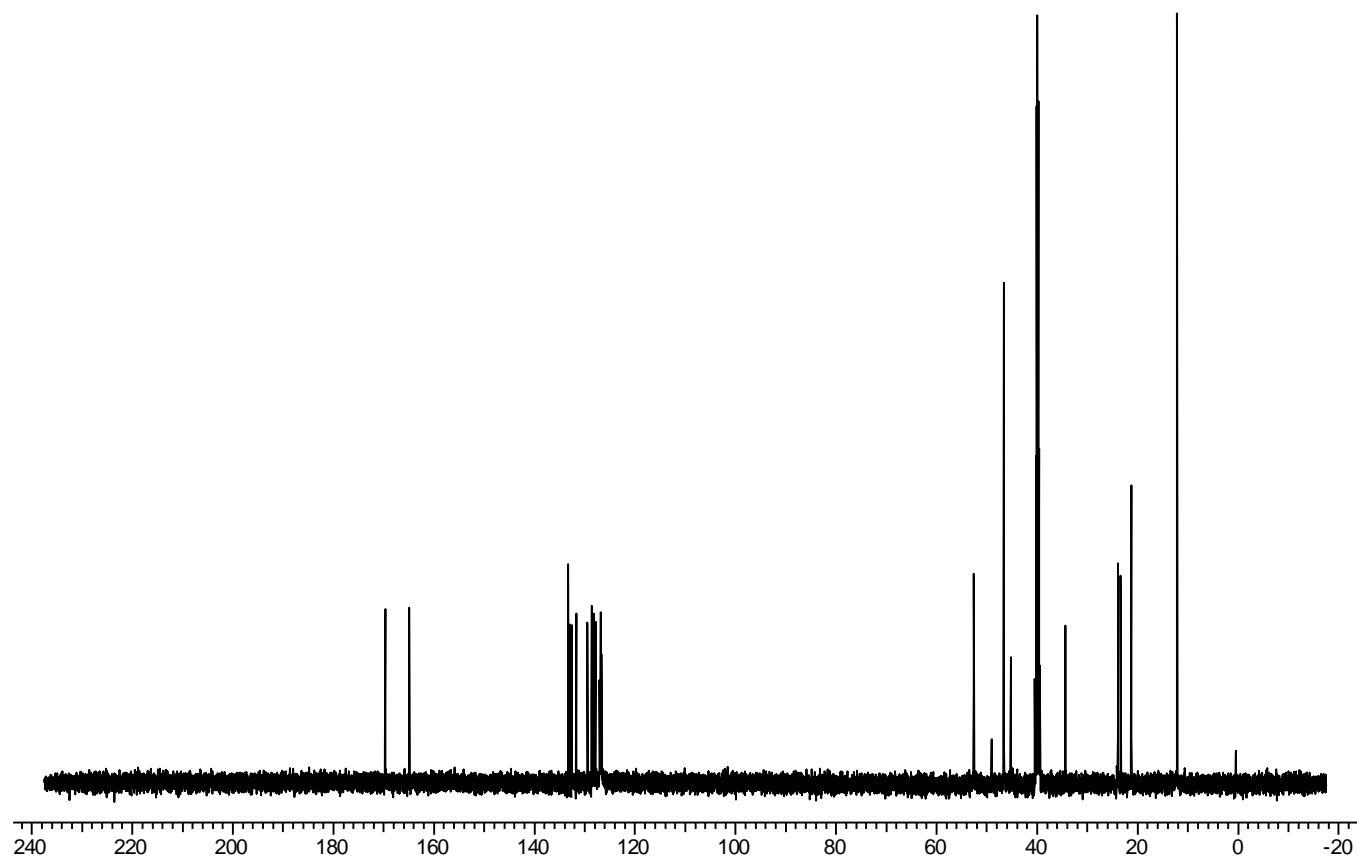
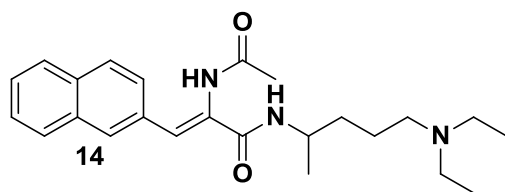




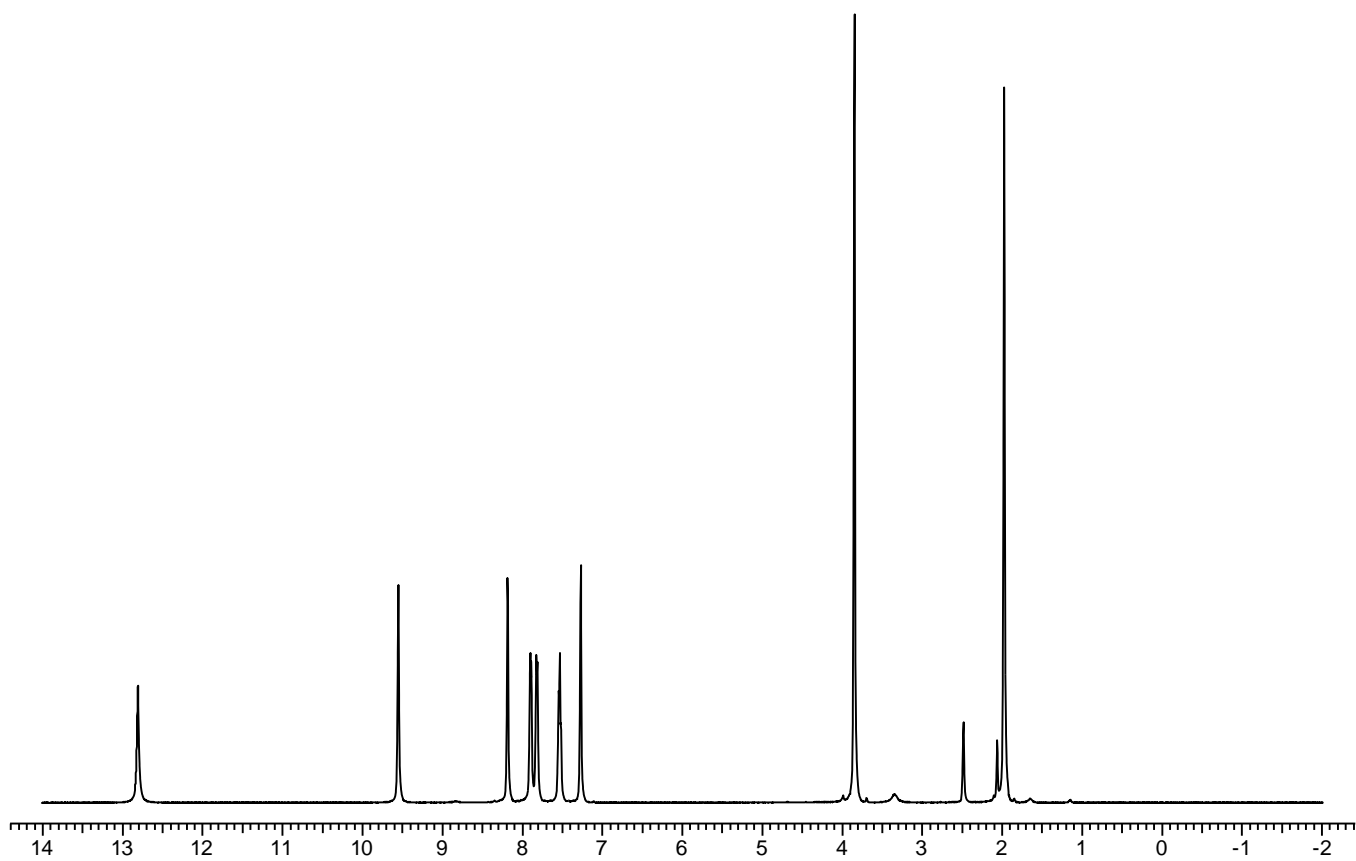
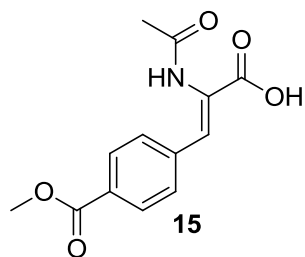
13:  $^{13}\text{C}$  NMR



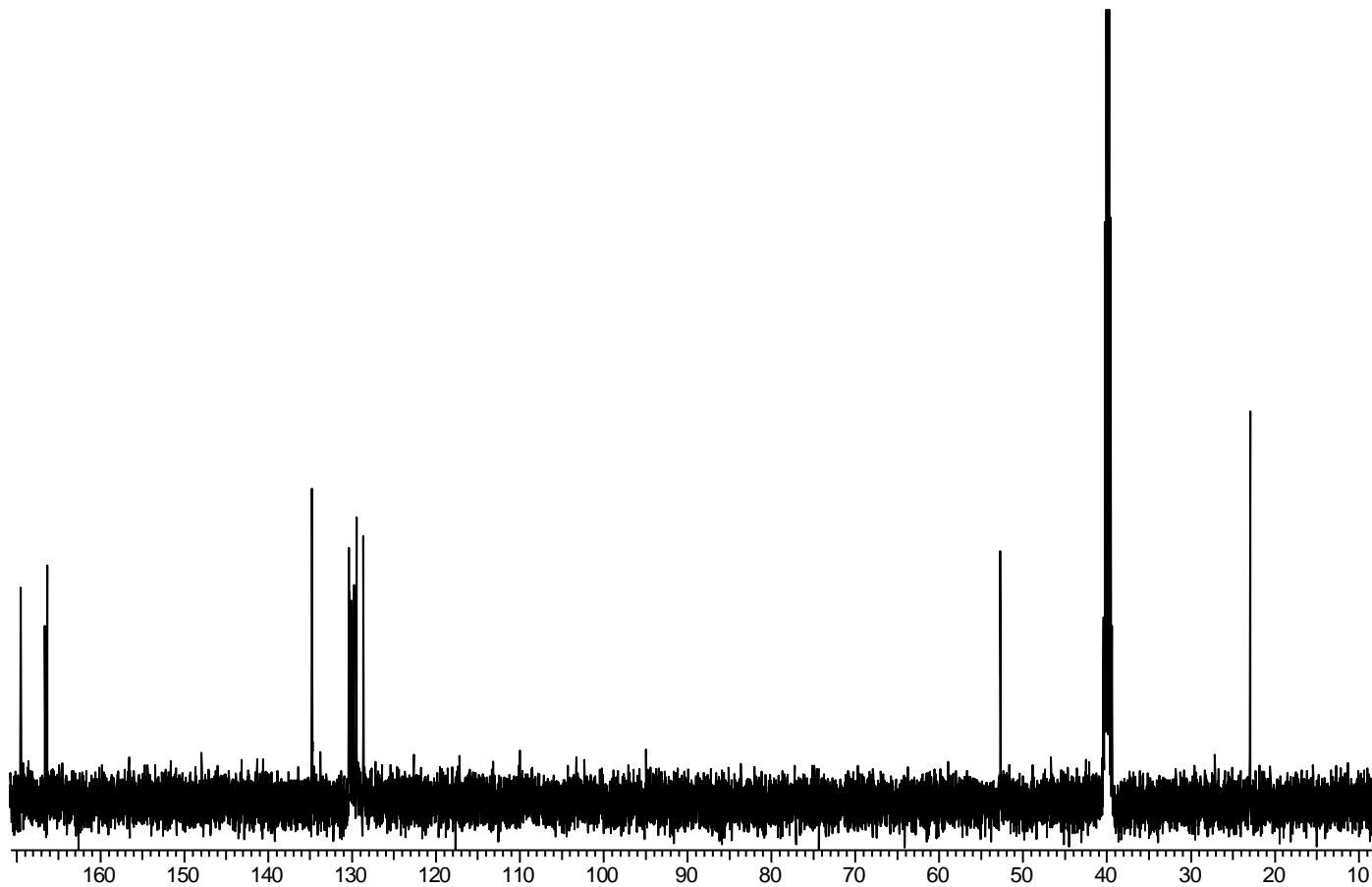
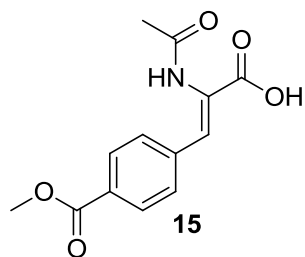
14:  $^1\text{H}$  NMR



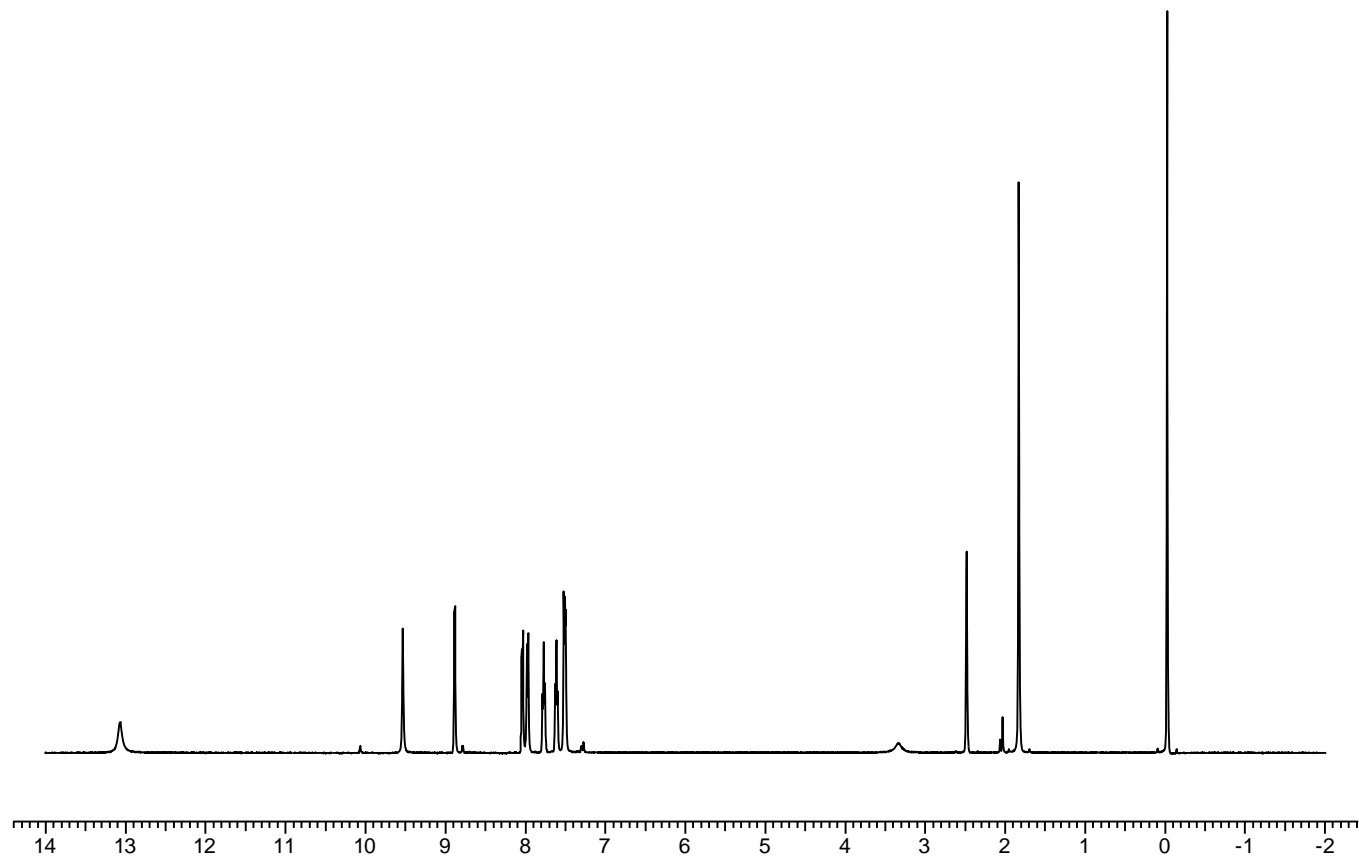
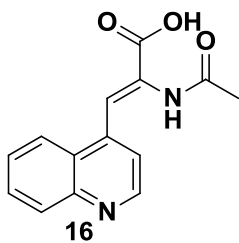
14:  $^{13}\text{C}$  NMR



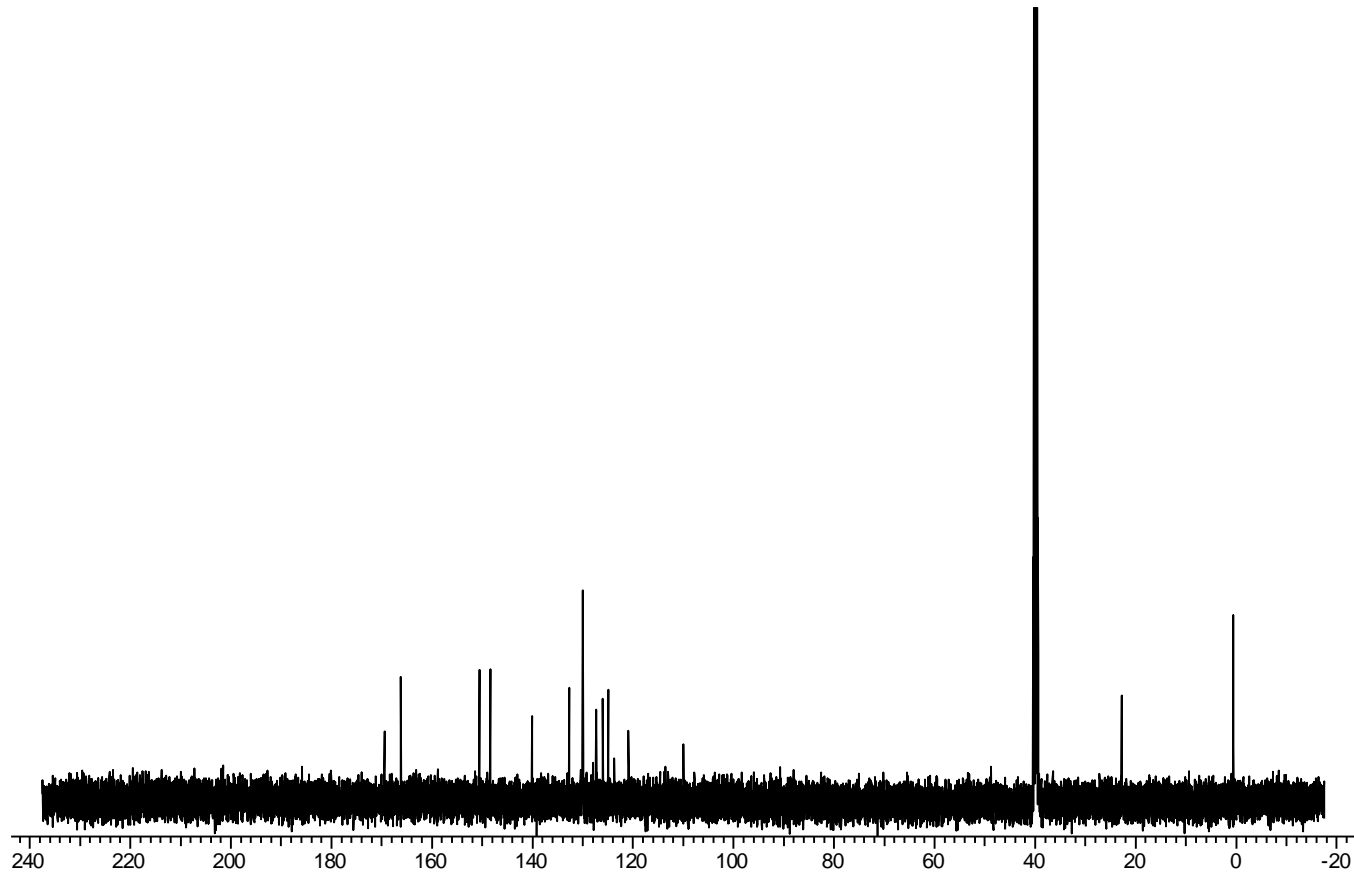
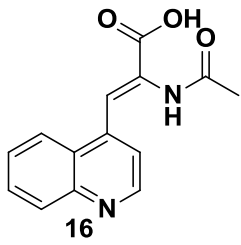
15:  $^1\text{H}$  NMR



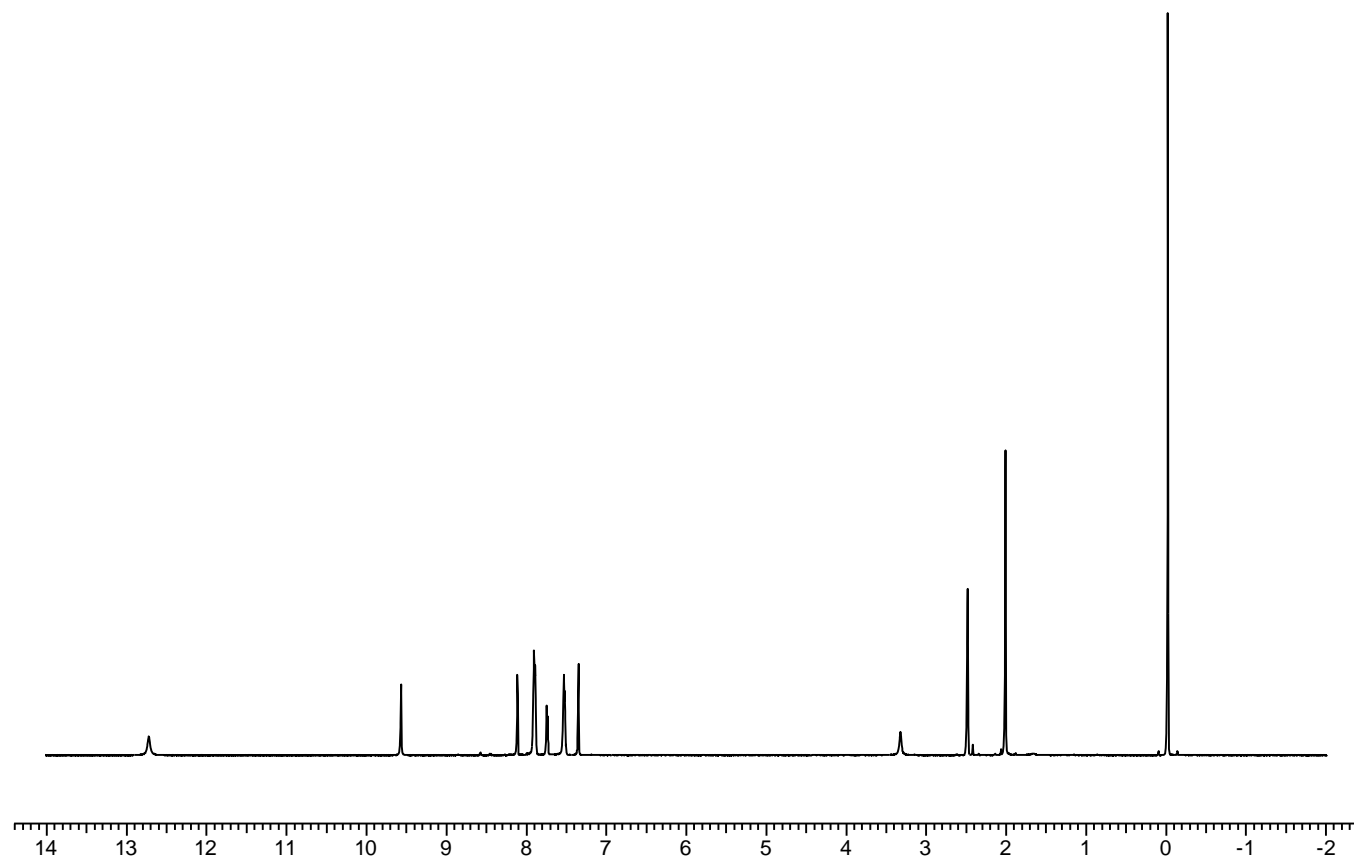
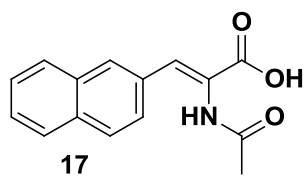
15:  $^{13}\text{C}$  NMR



16:  $^1\text{H}$  NMR

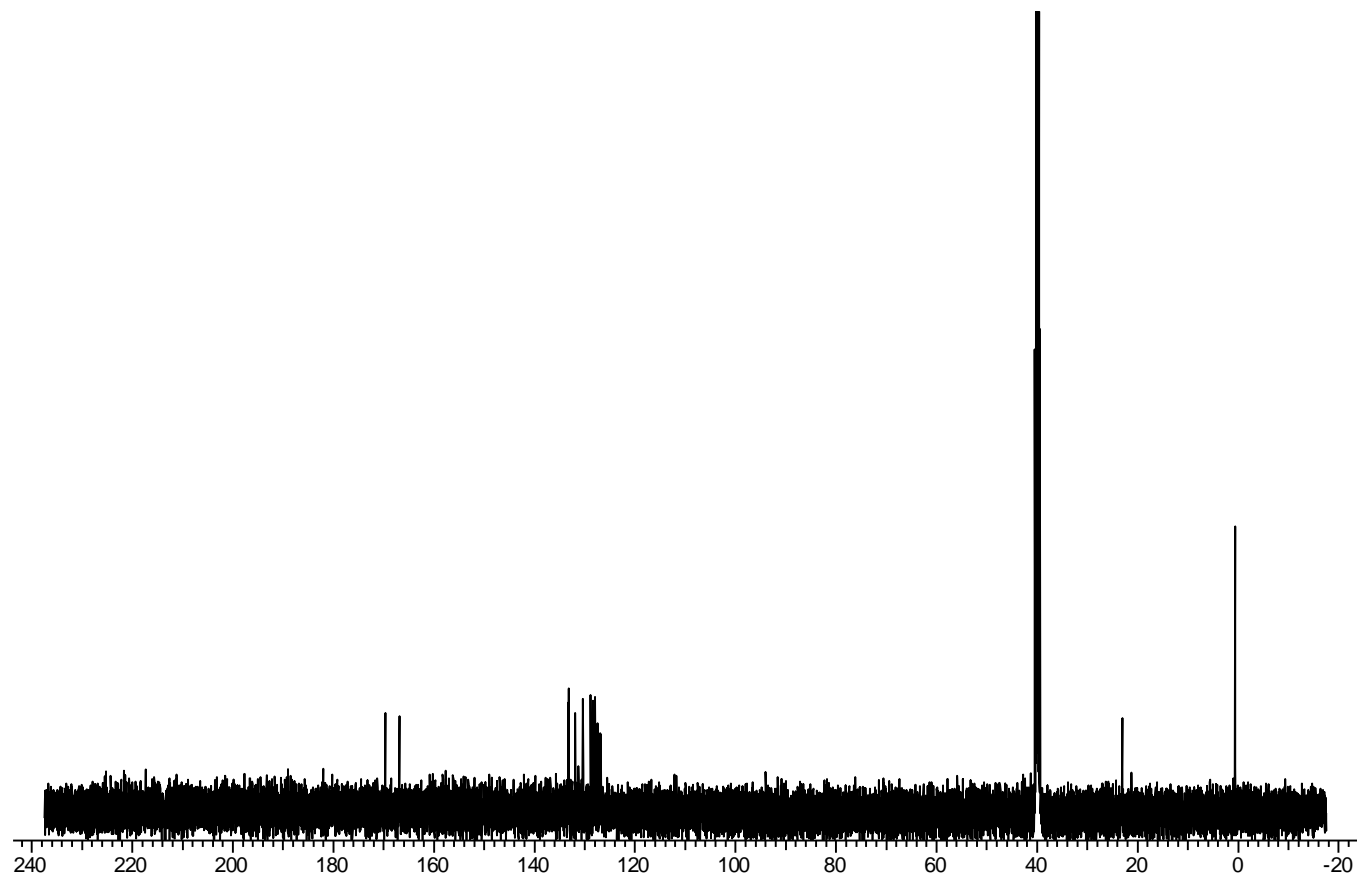
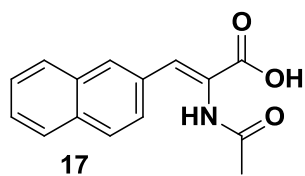


16:  $^{13}\text{C}$  NMR

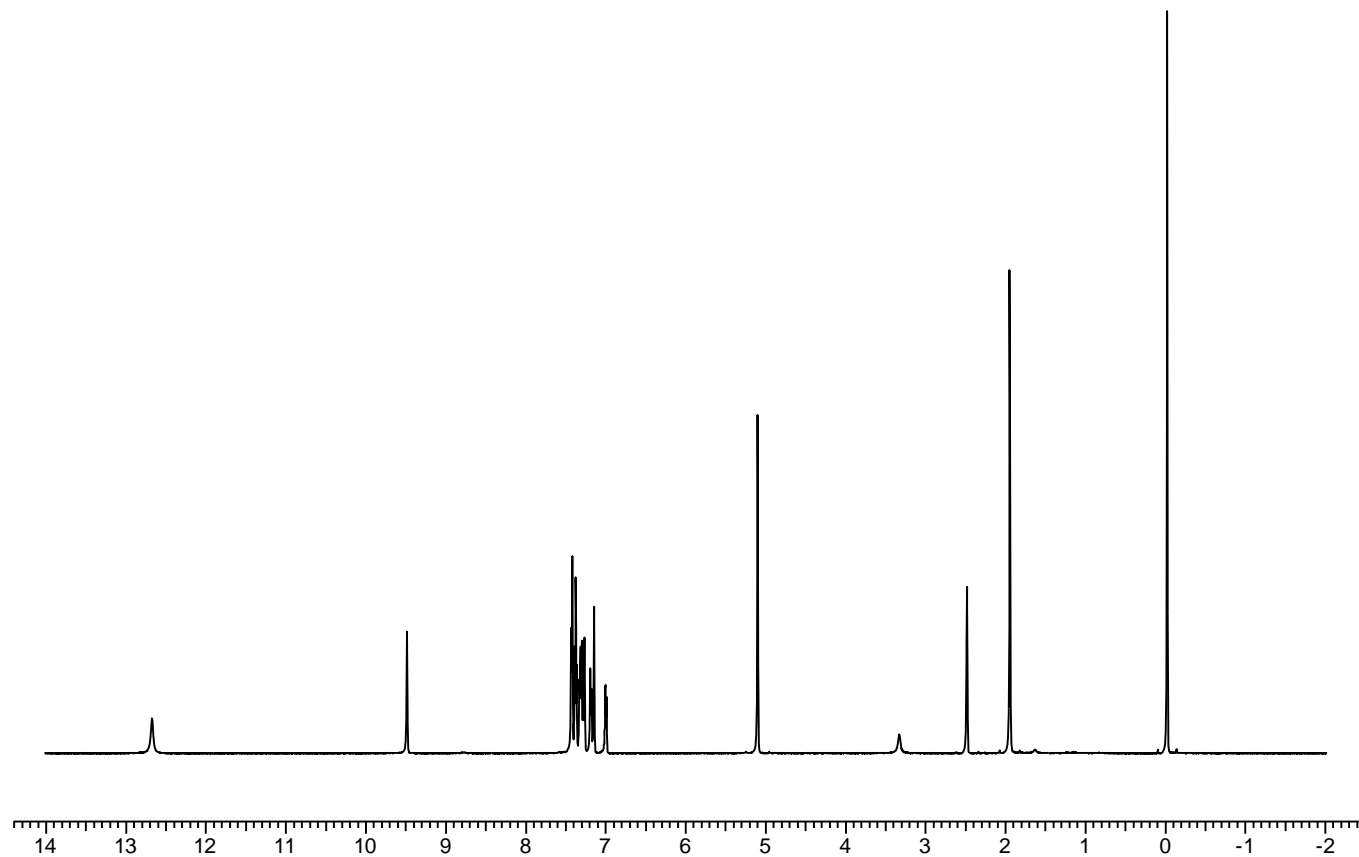
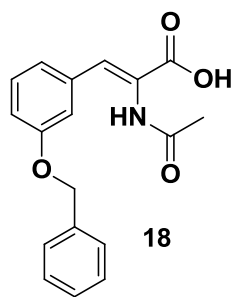


17: <sup>1</sup>H NMR

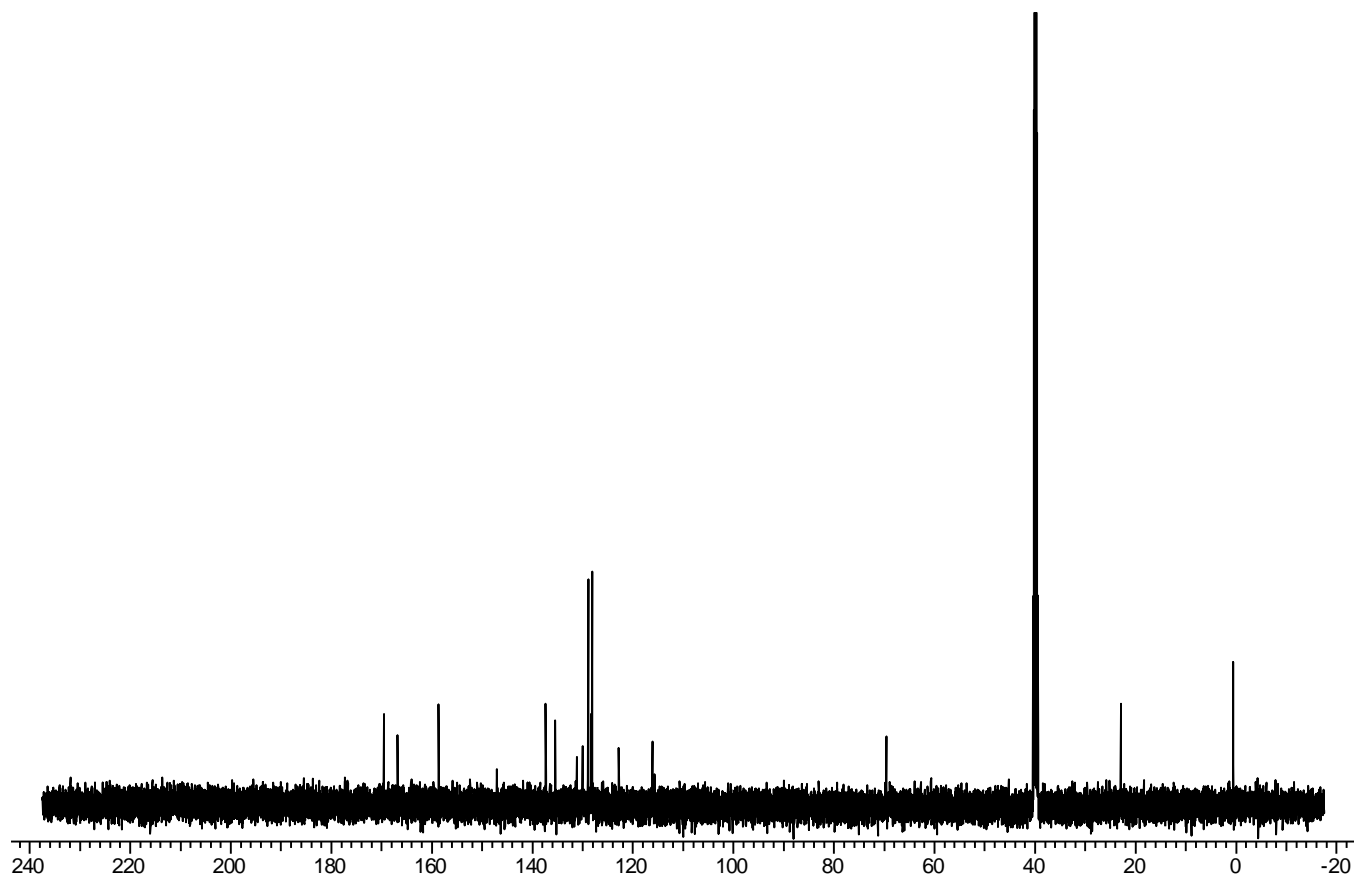
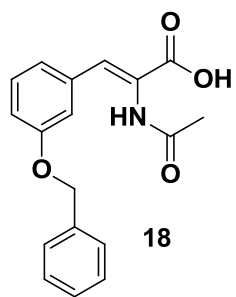




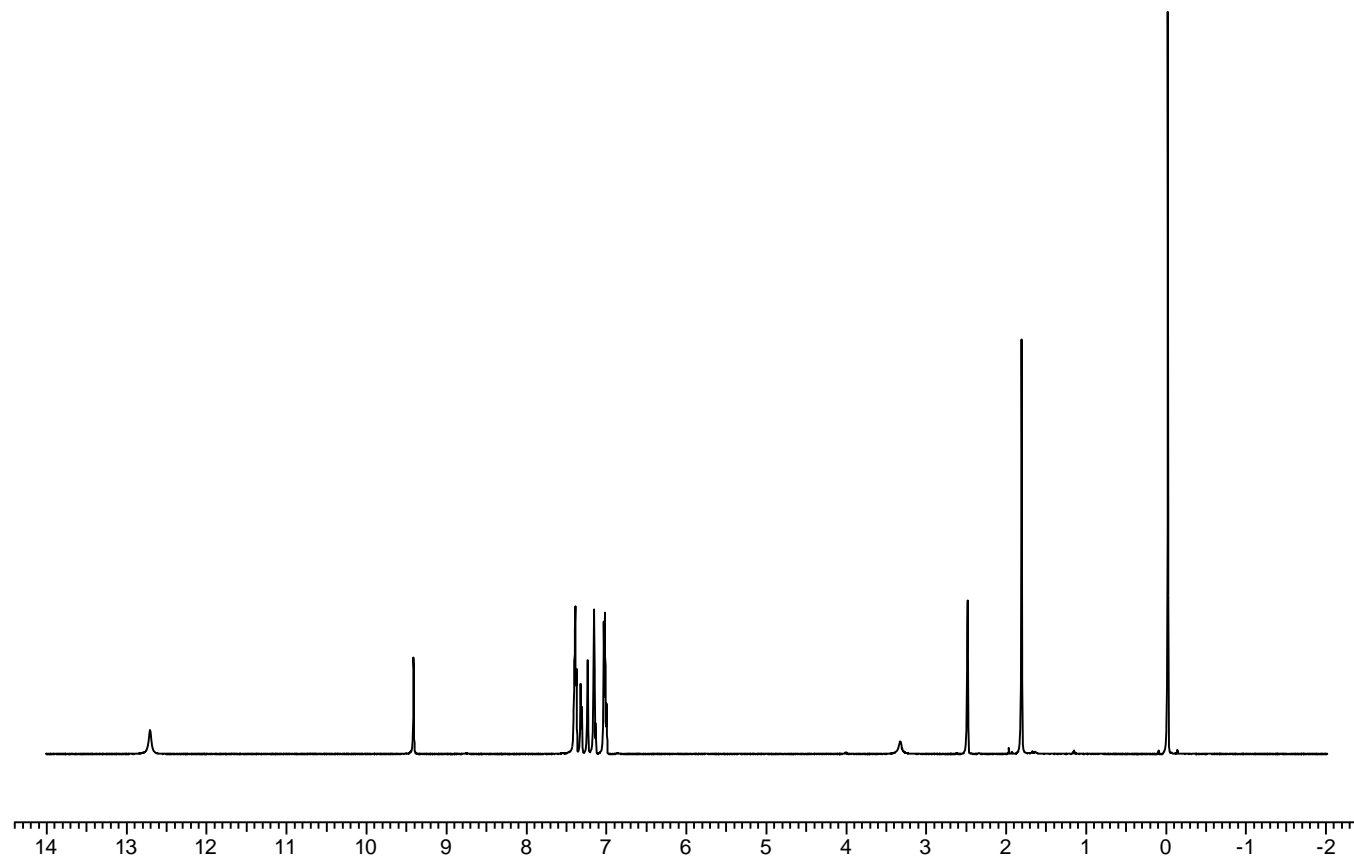
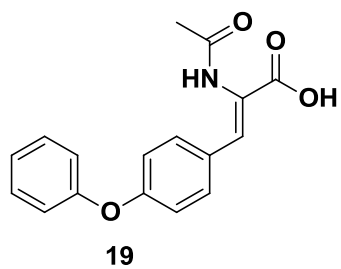
17:  $^{13}\text{C}$  NMR



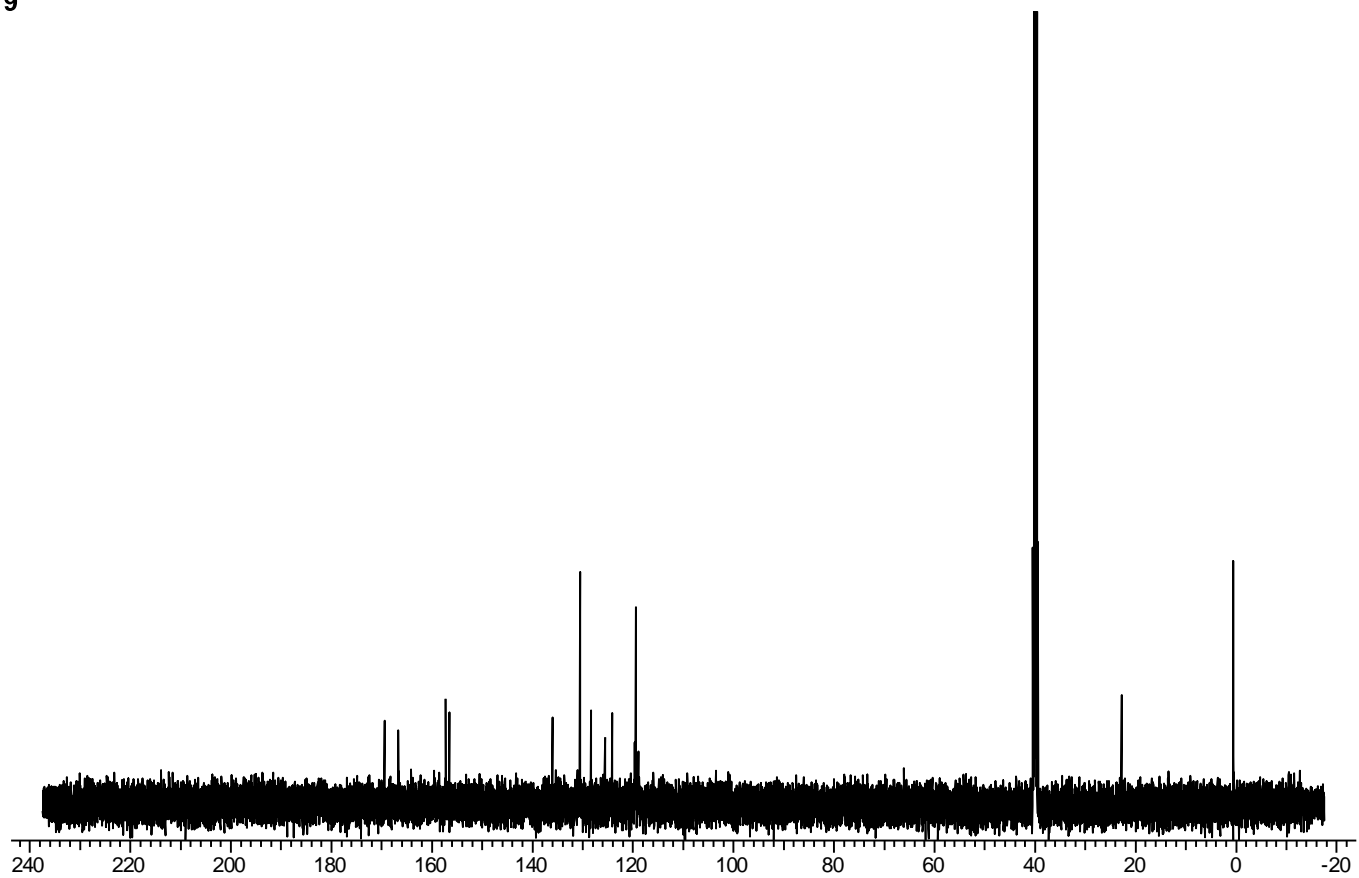
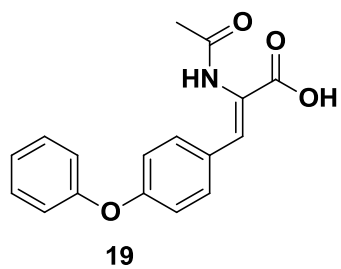
18:  $^1\text{H}$  NMR



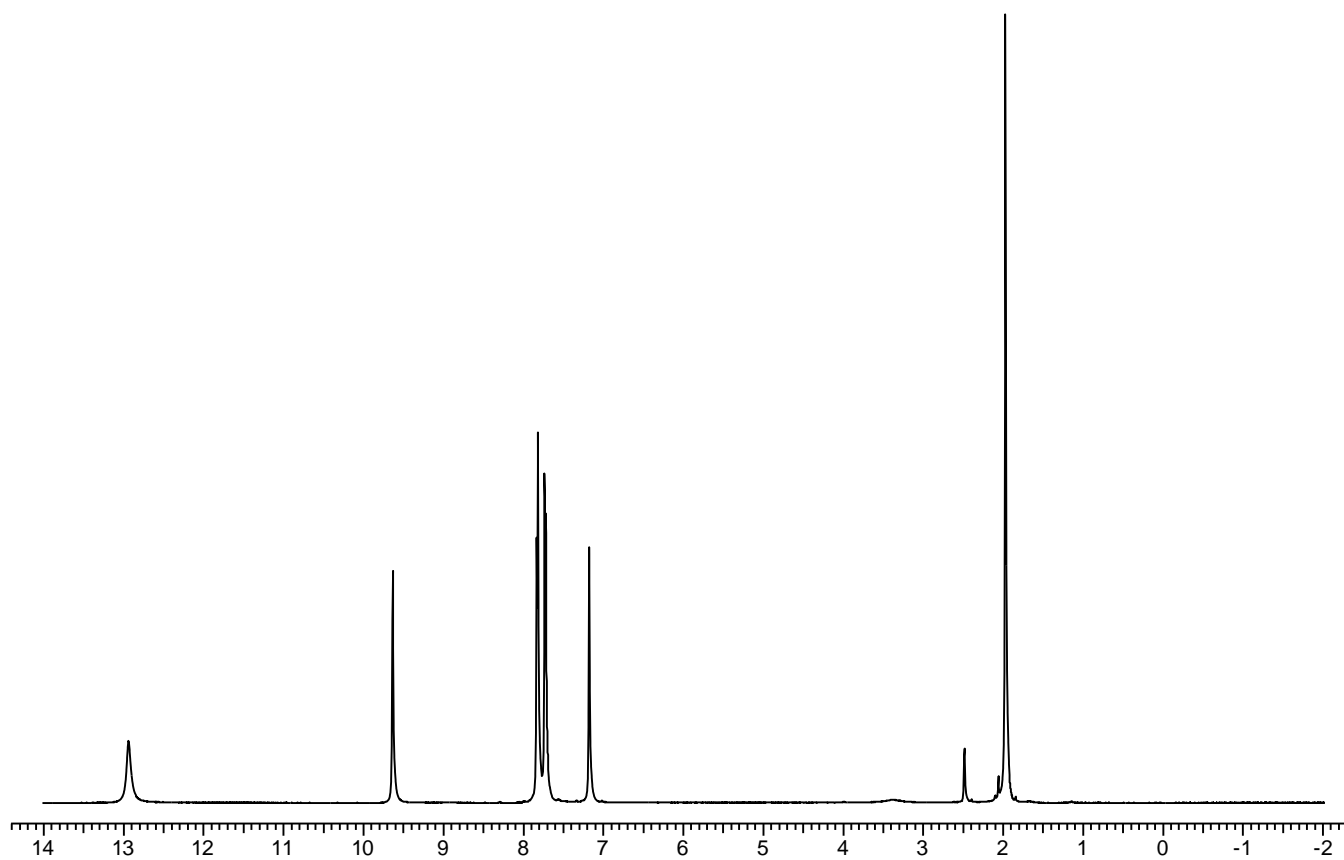
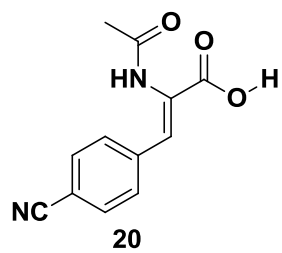
18:  $^{13}\text{C}$  NMR



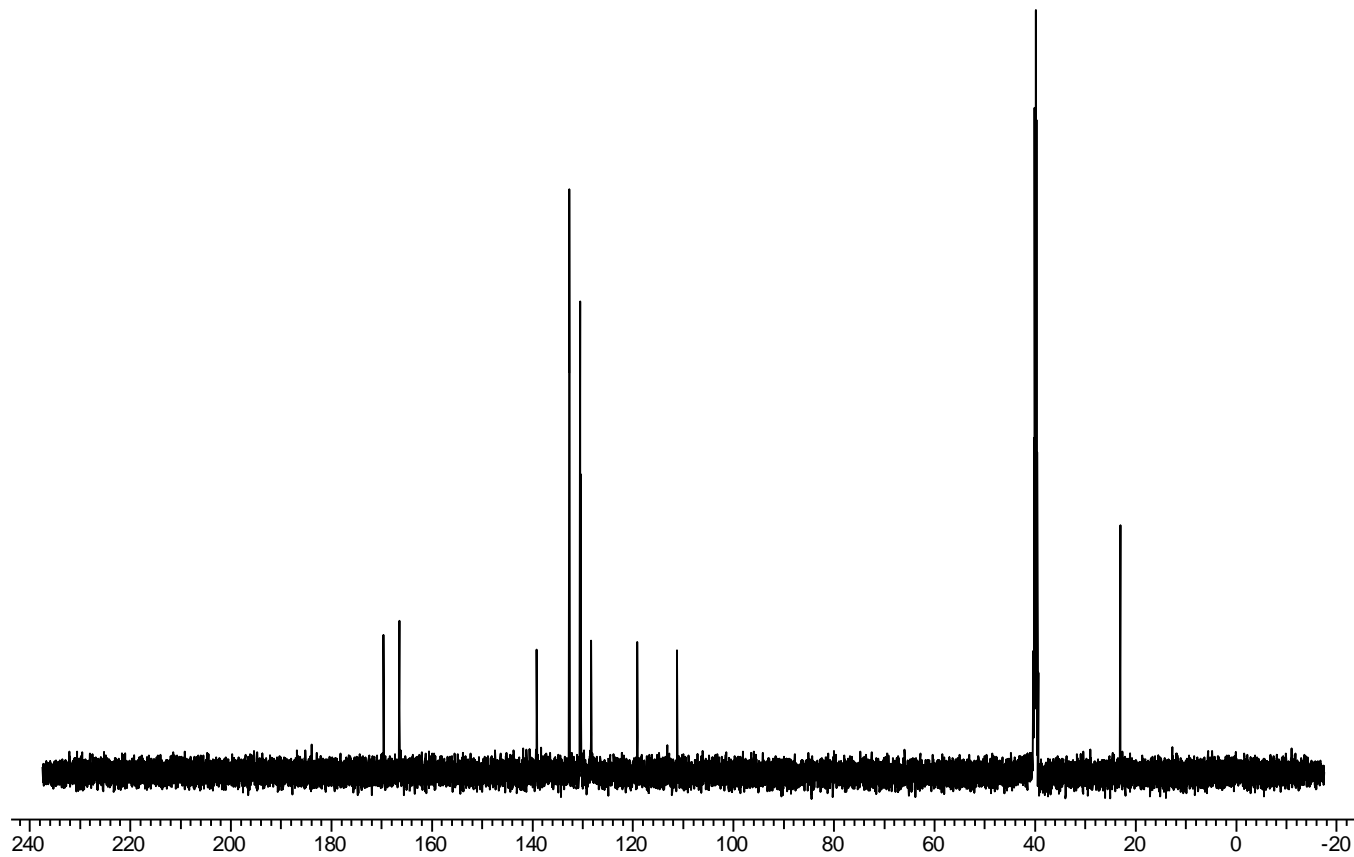
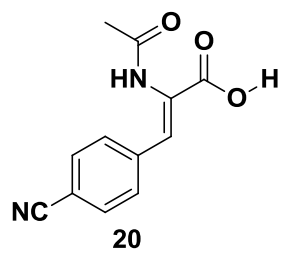
19:  $^1\text{H}$  NMR



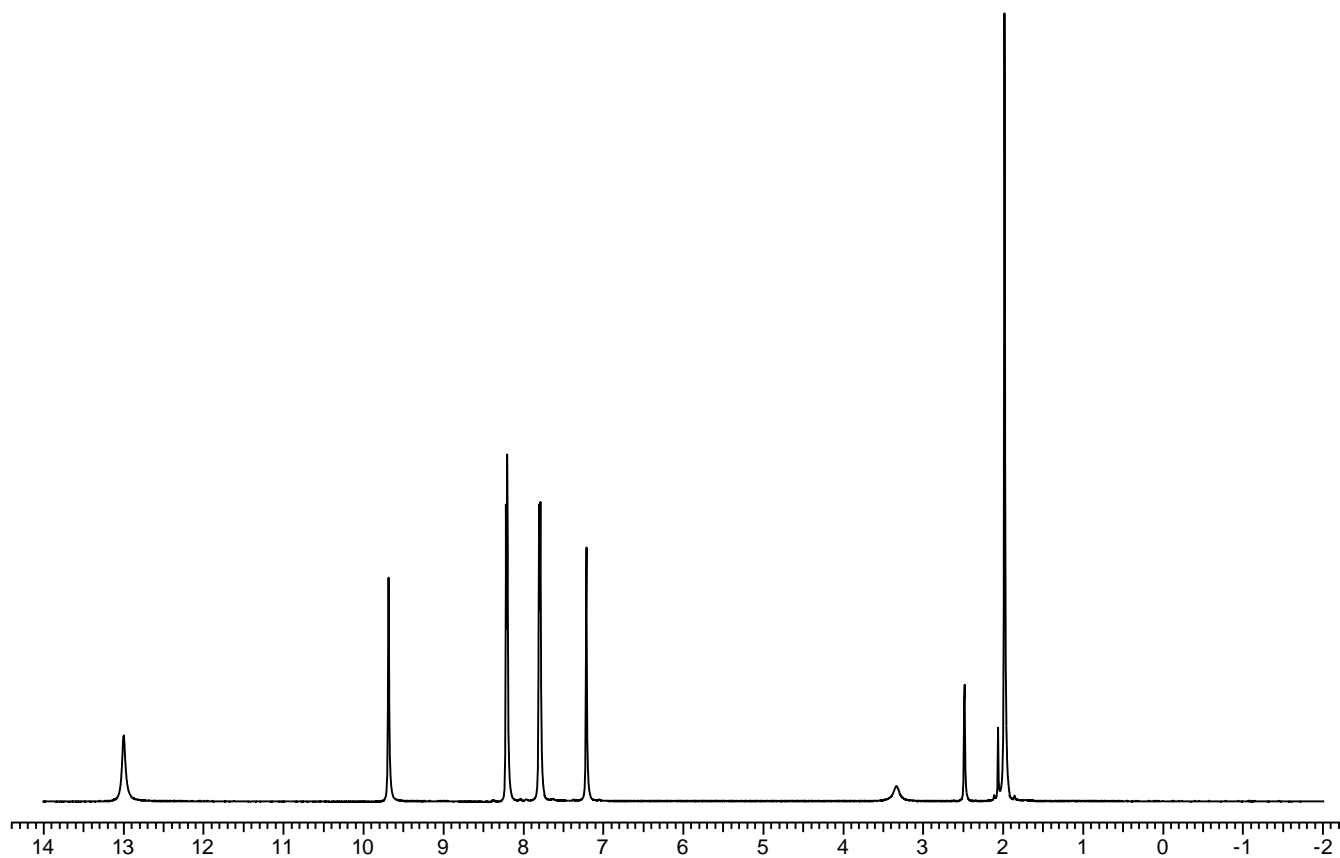
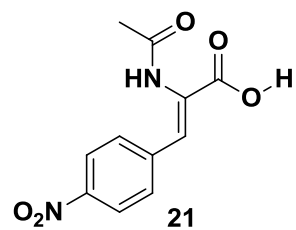
19:  $^{13}\text{C}$  NMR



20: <sup>1</sup>H NMR

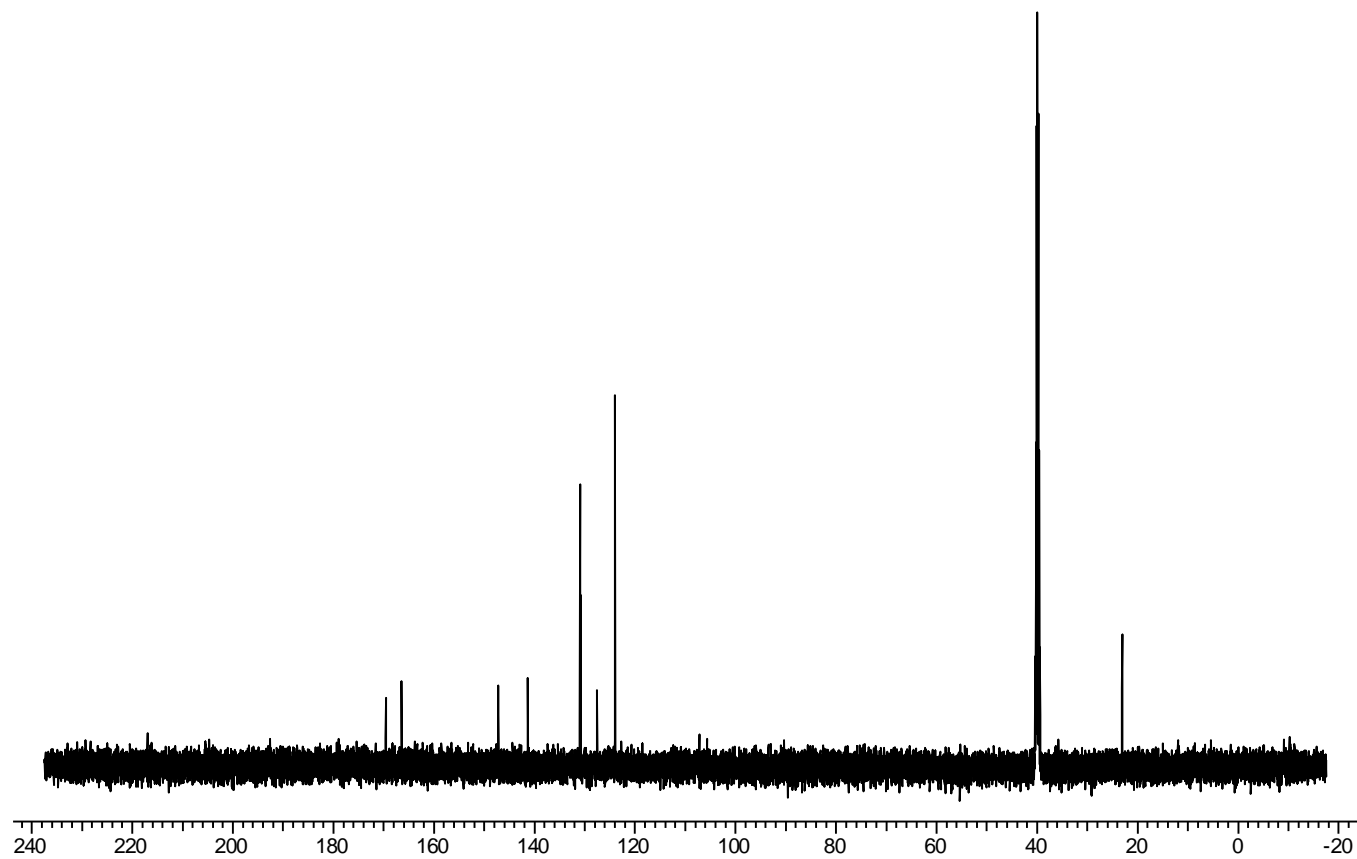
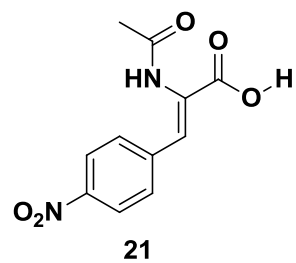


20:  $^{13}\text{C}$  NMR

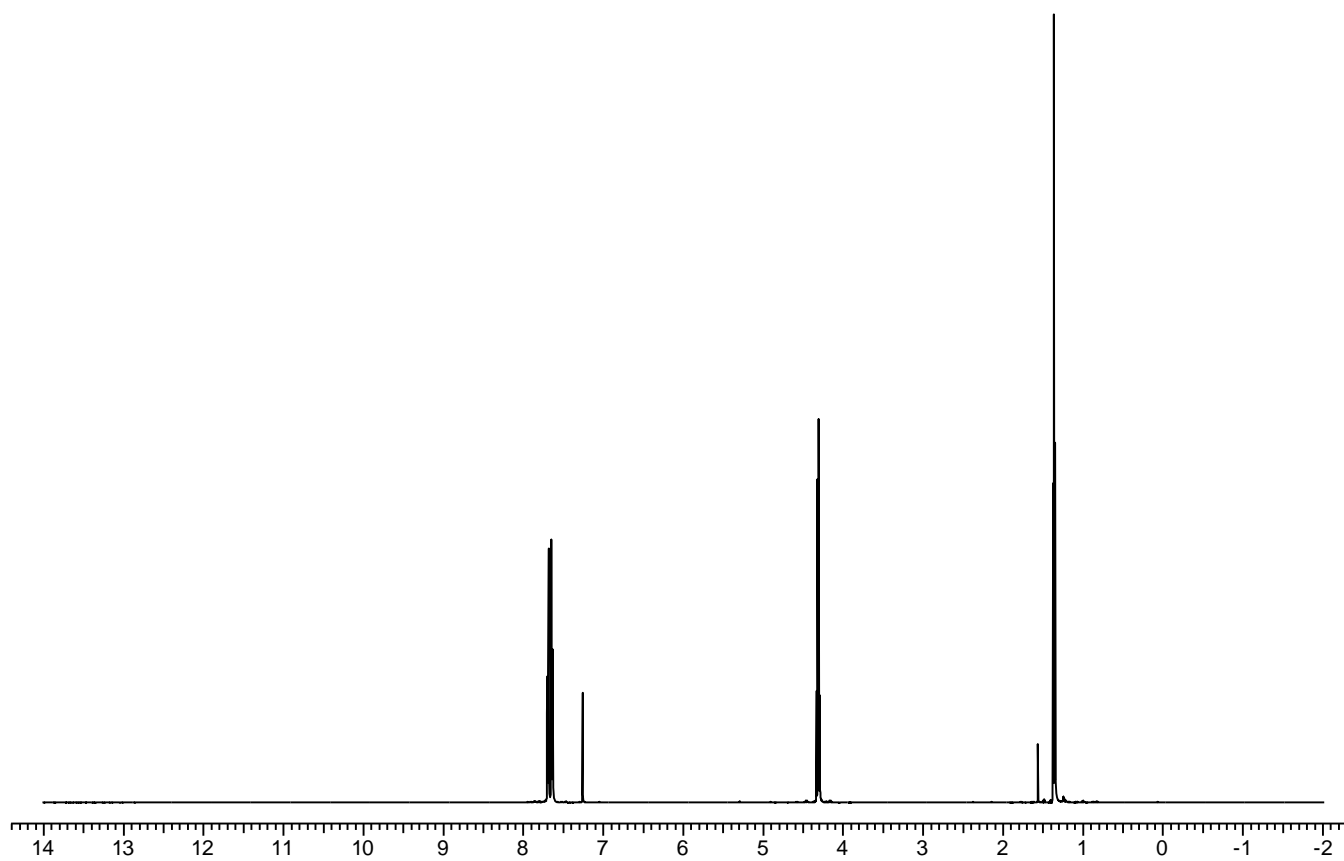
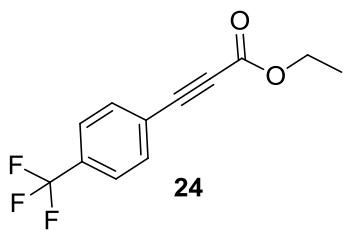


21: <sup>1</sup>H NMR

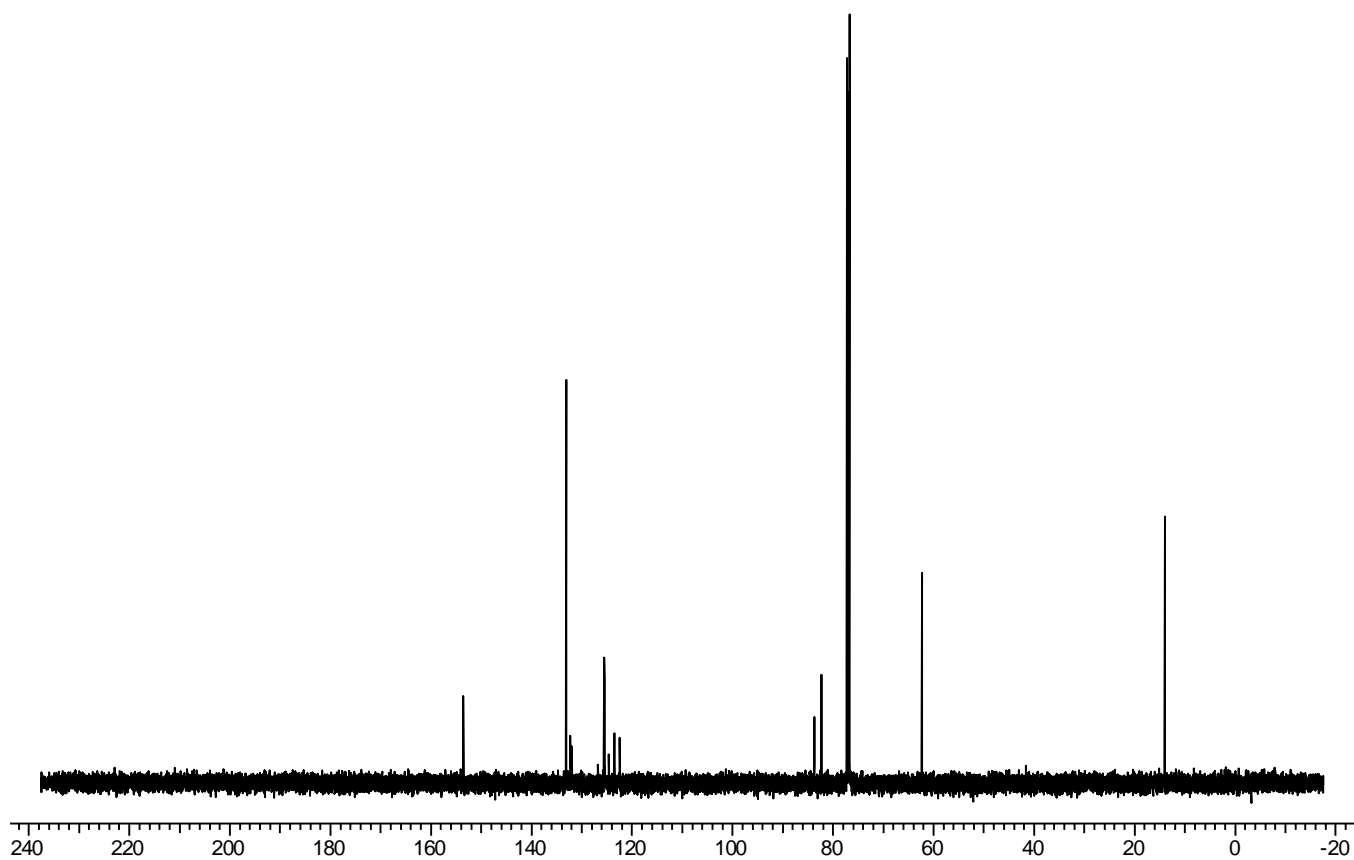
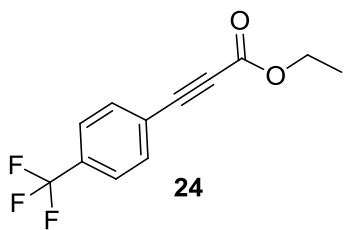


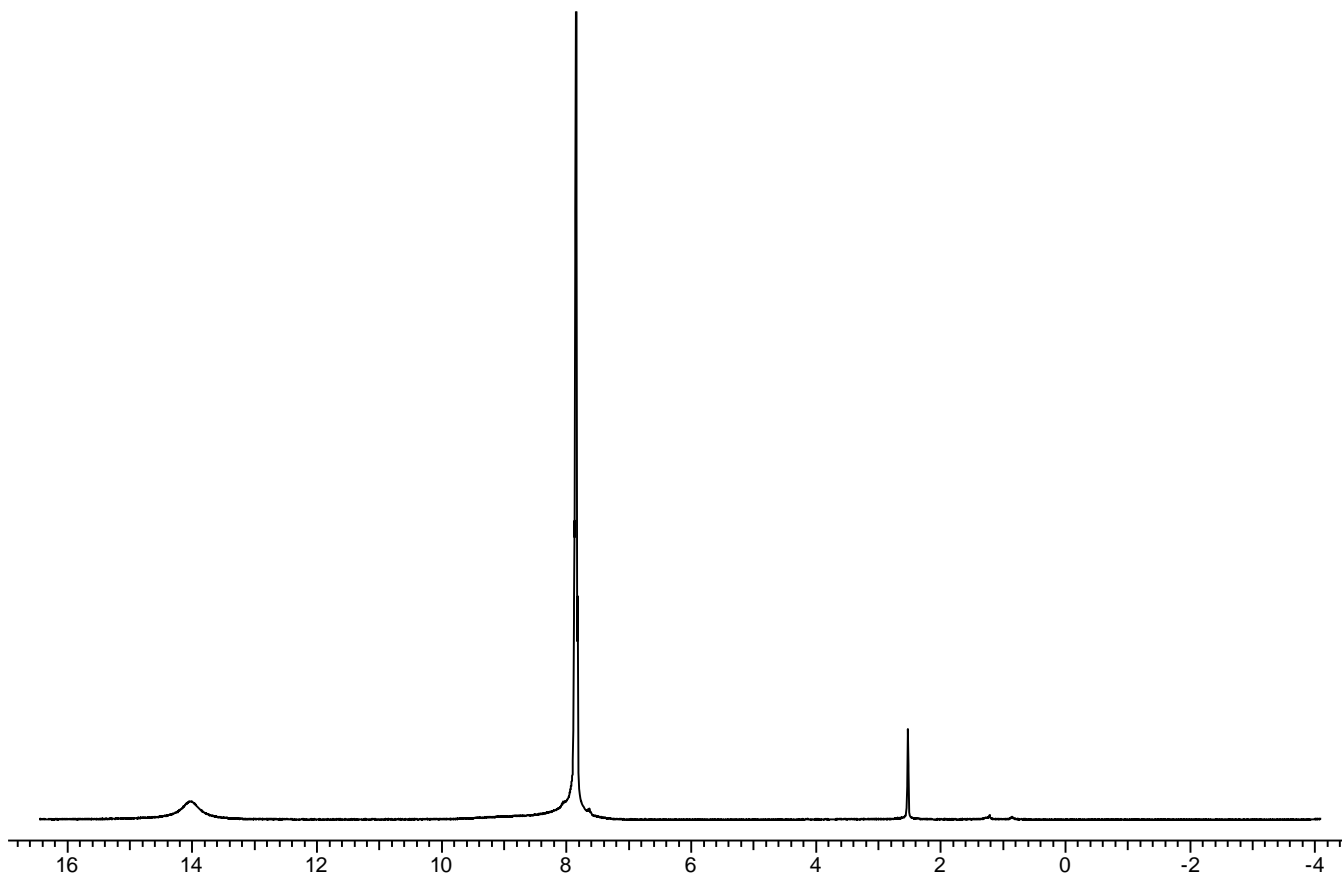
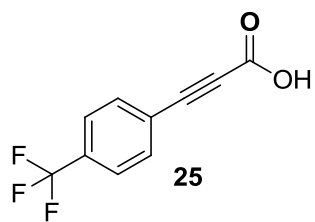


21:  $^{13}\text{C}$  NMR

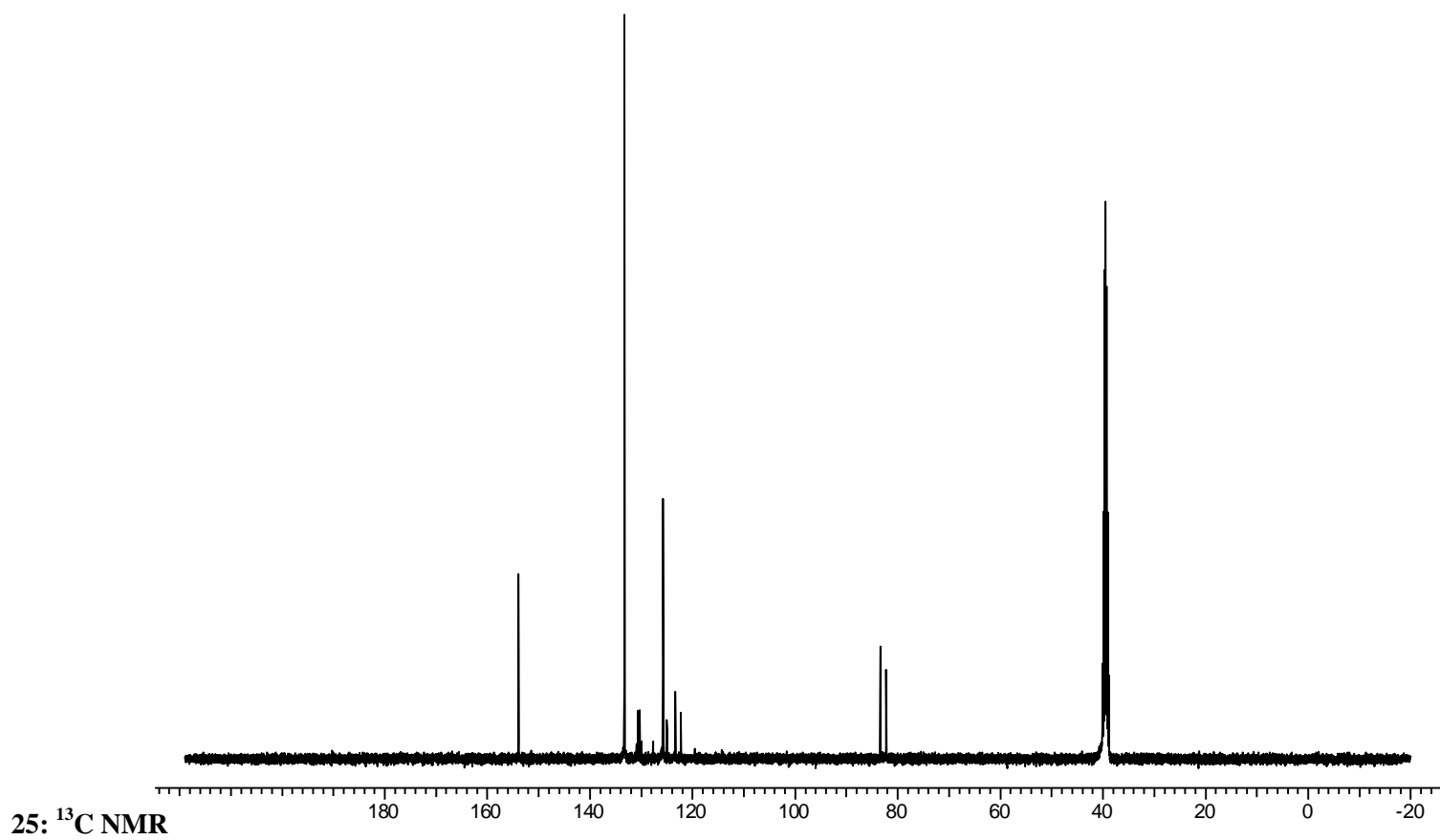
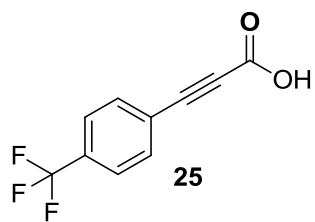


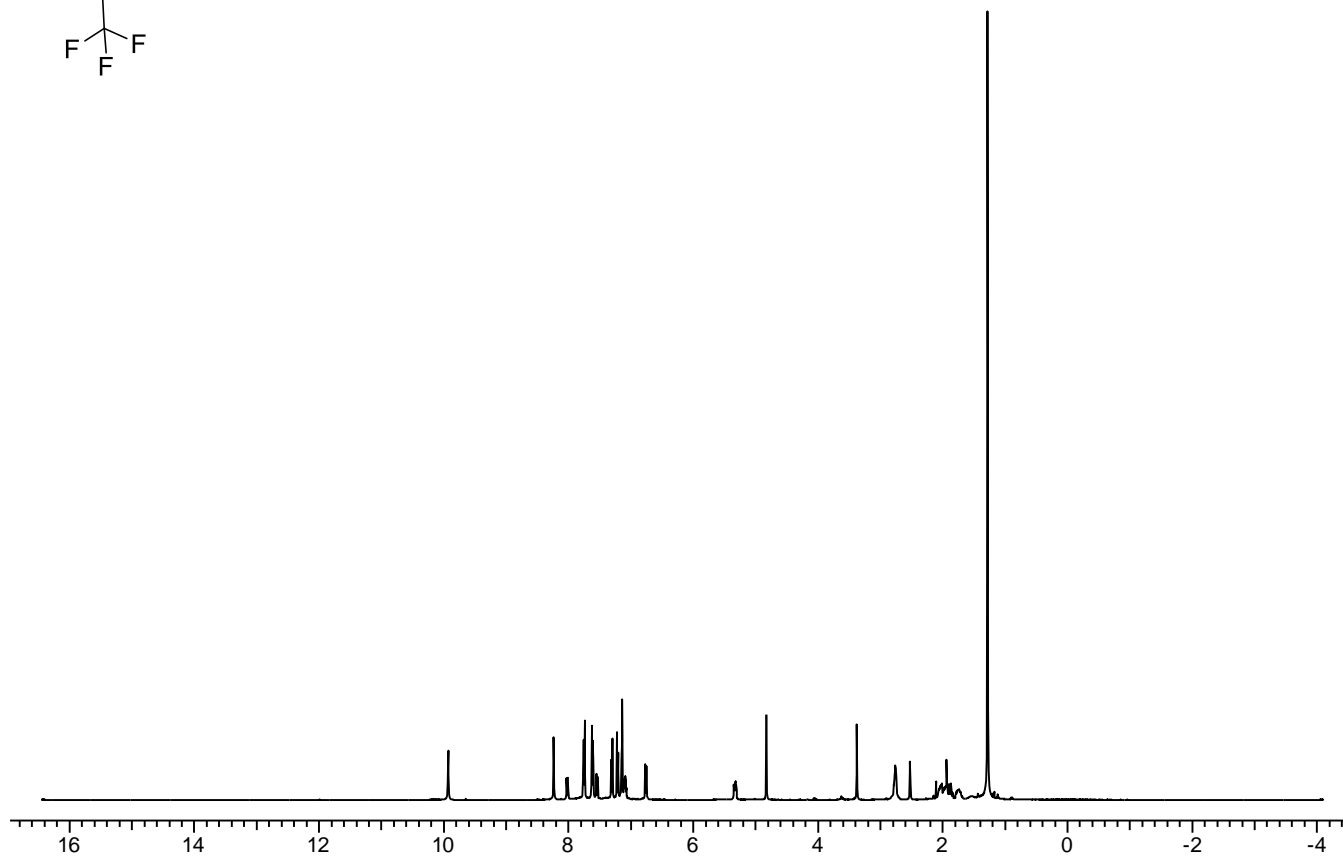
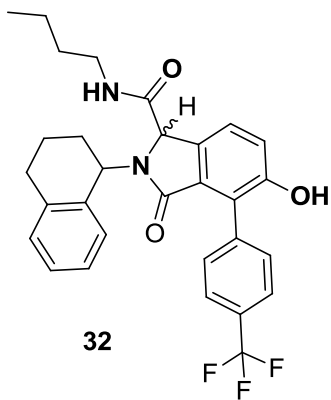
24: <sup>1</sup>H NMR



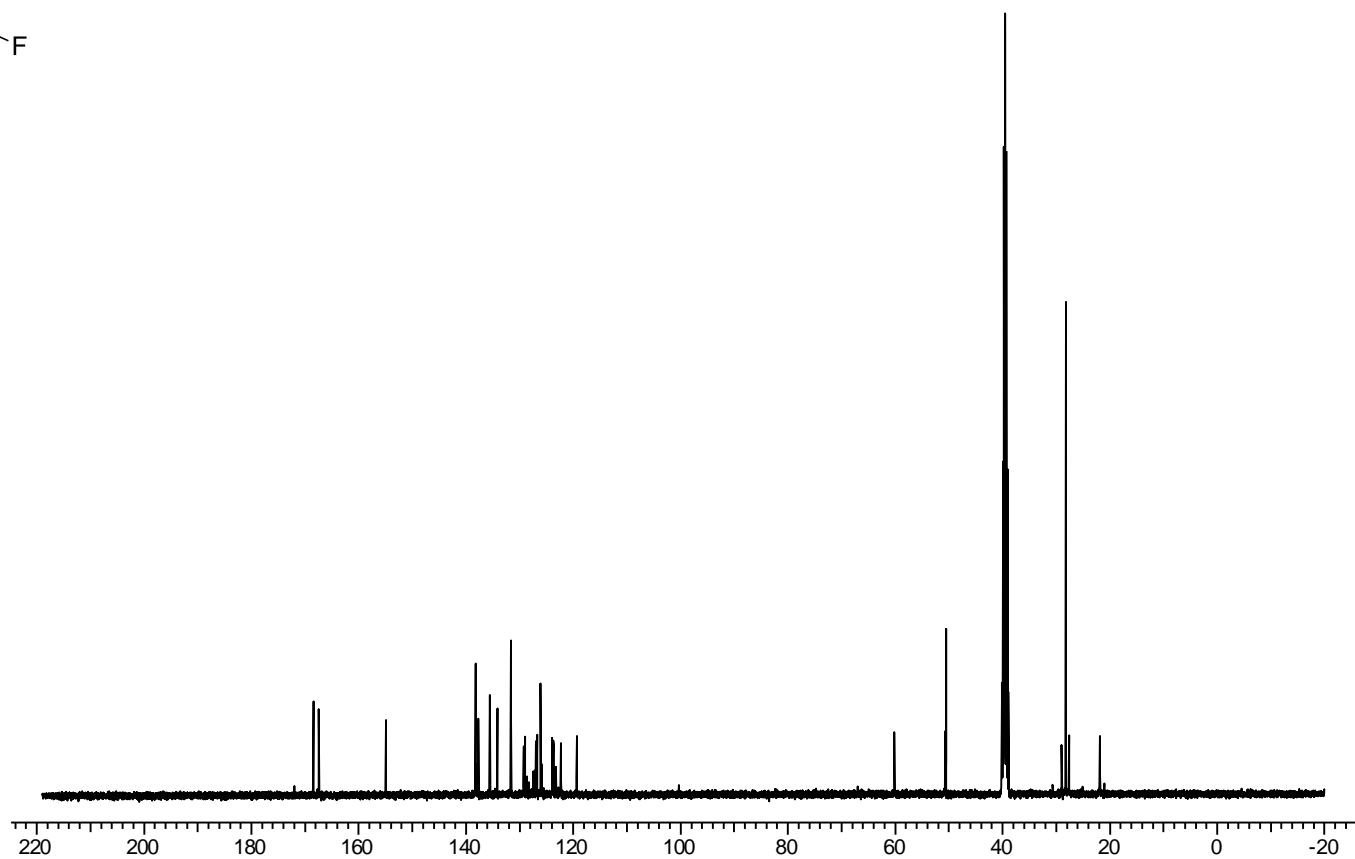
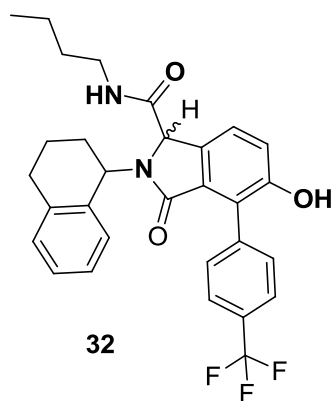


25:  $^1\text{H}$  NMR

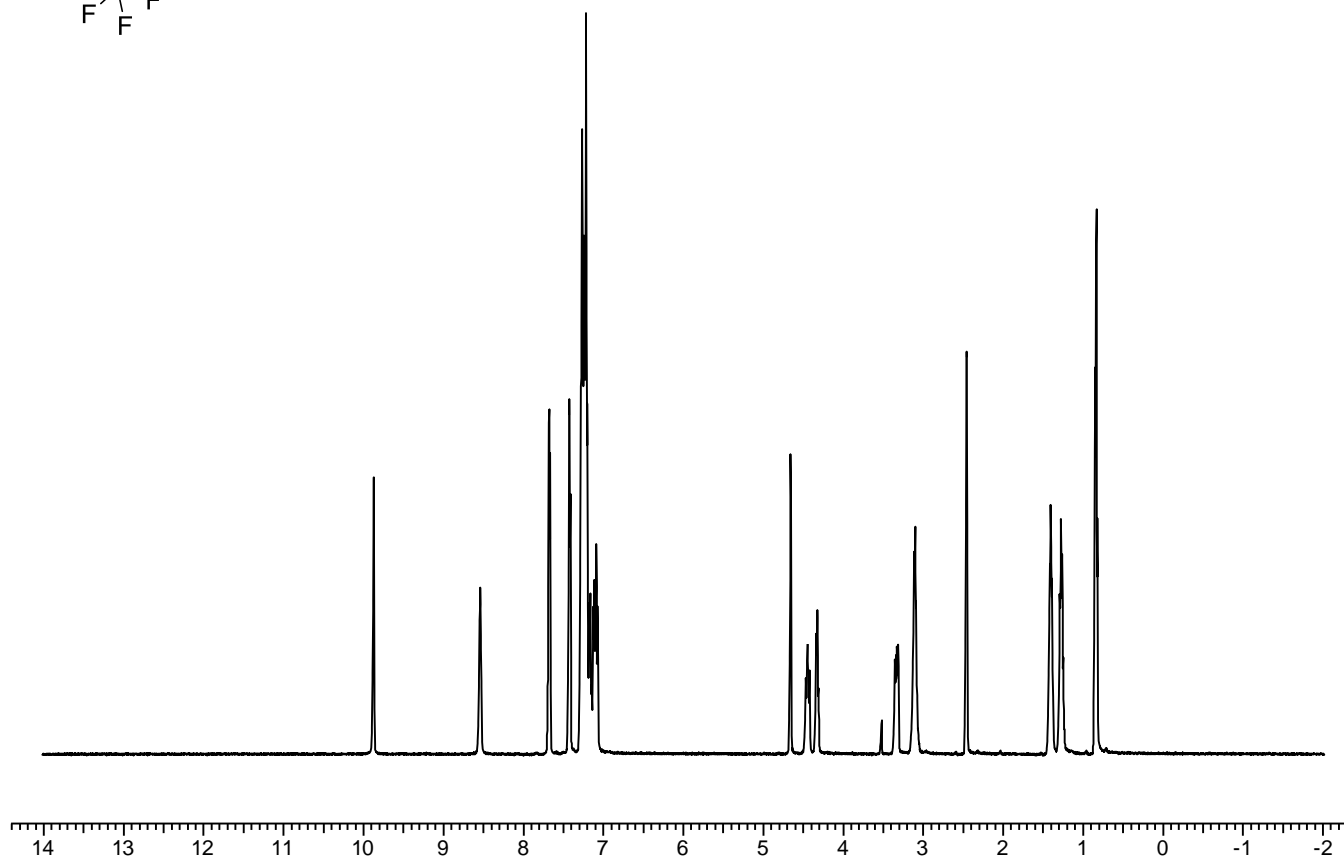
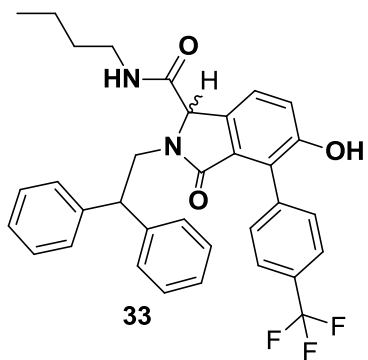




32:  $^1\text{H}$  NMR

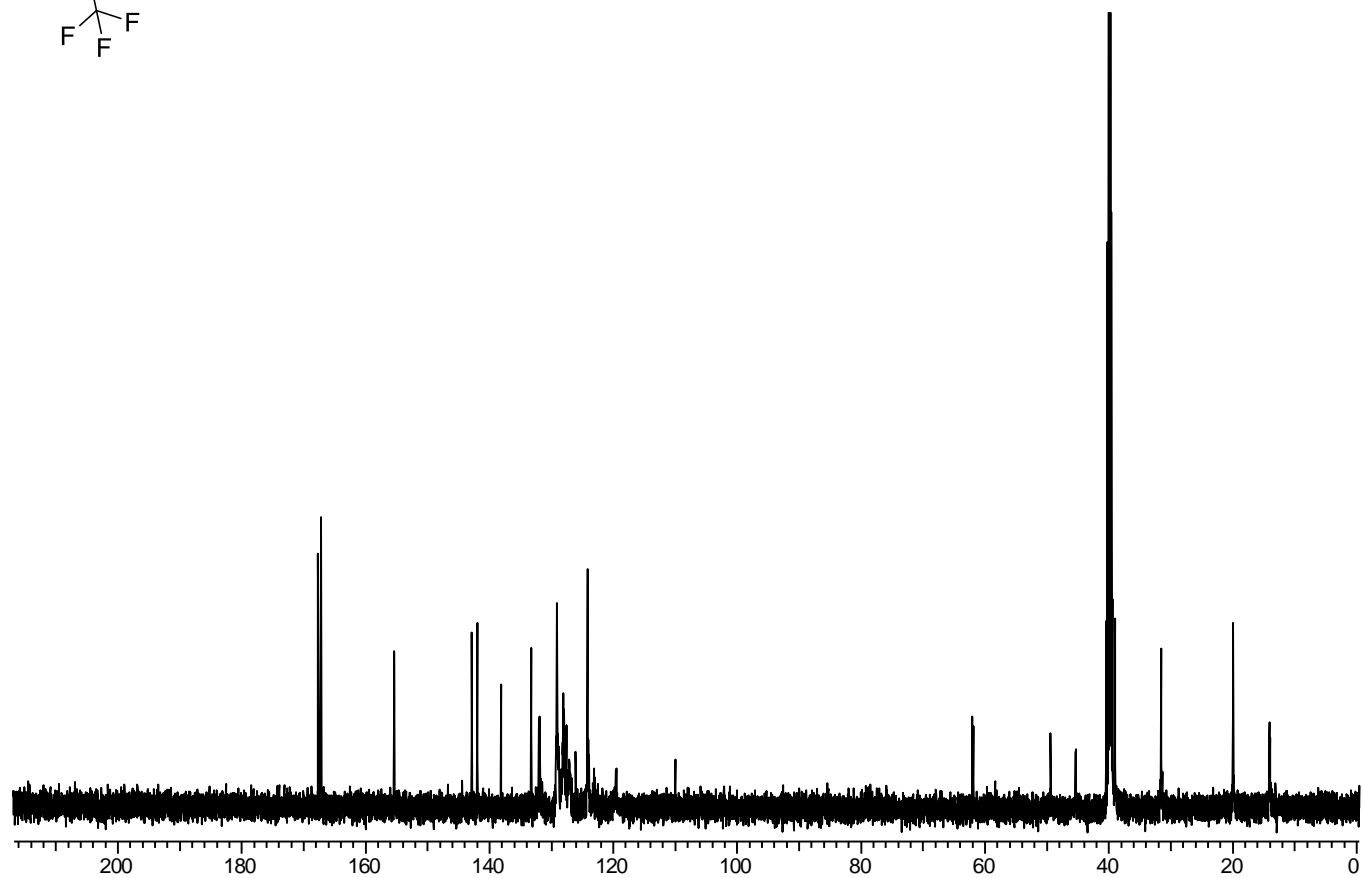
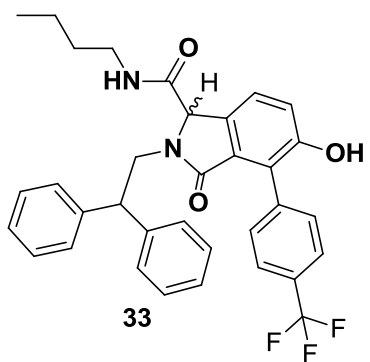


32:  $^{13}\text{C}$  NMR

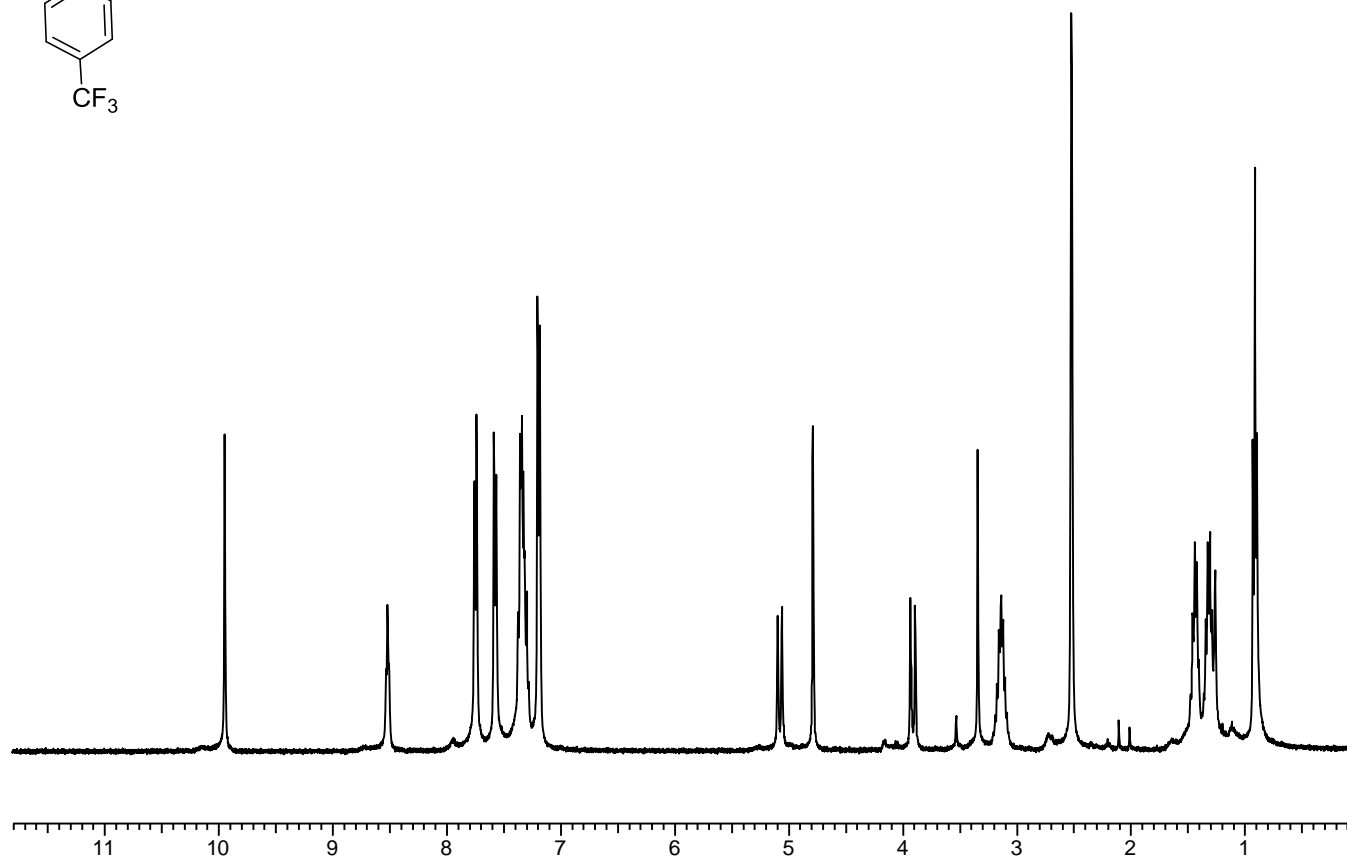
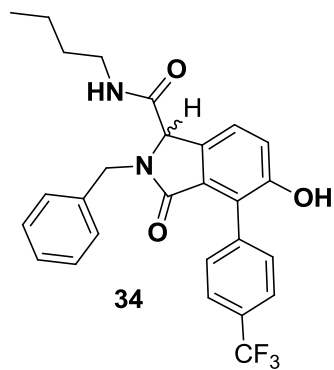


33:  $^1\text{H}$  NMR

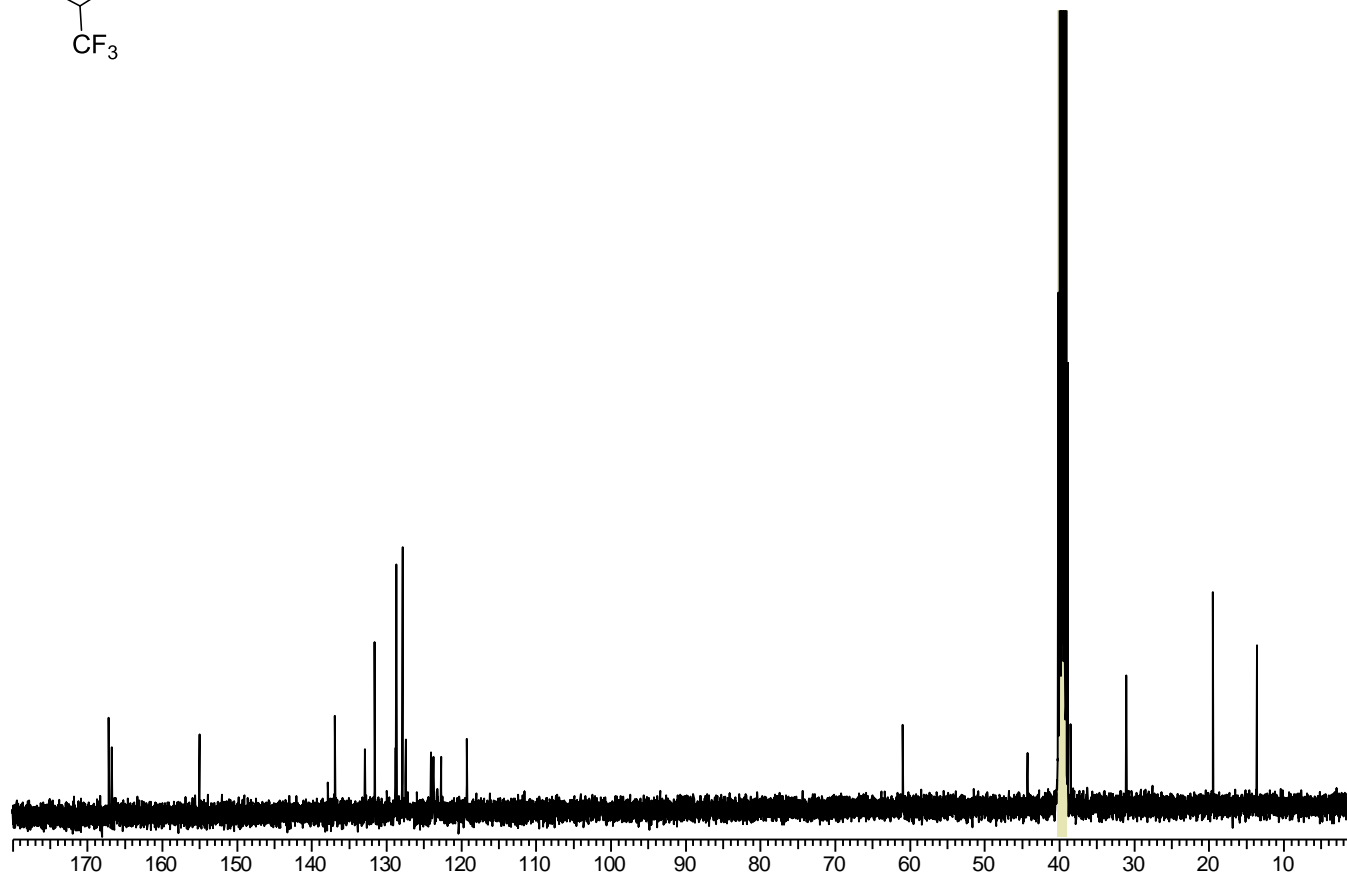
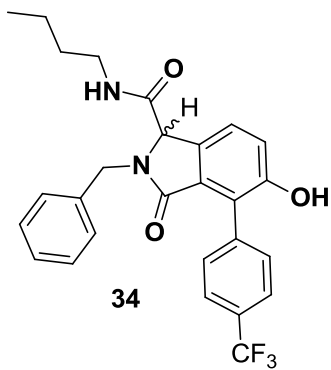




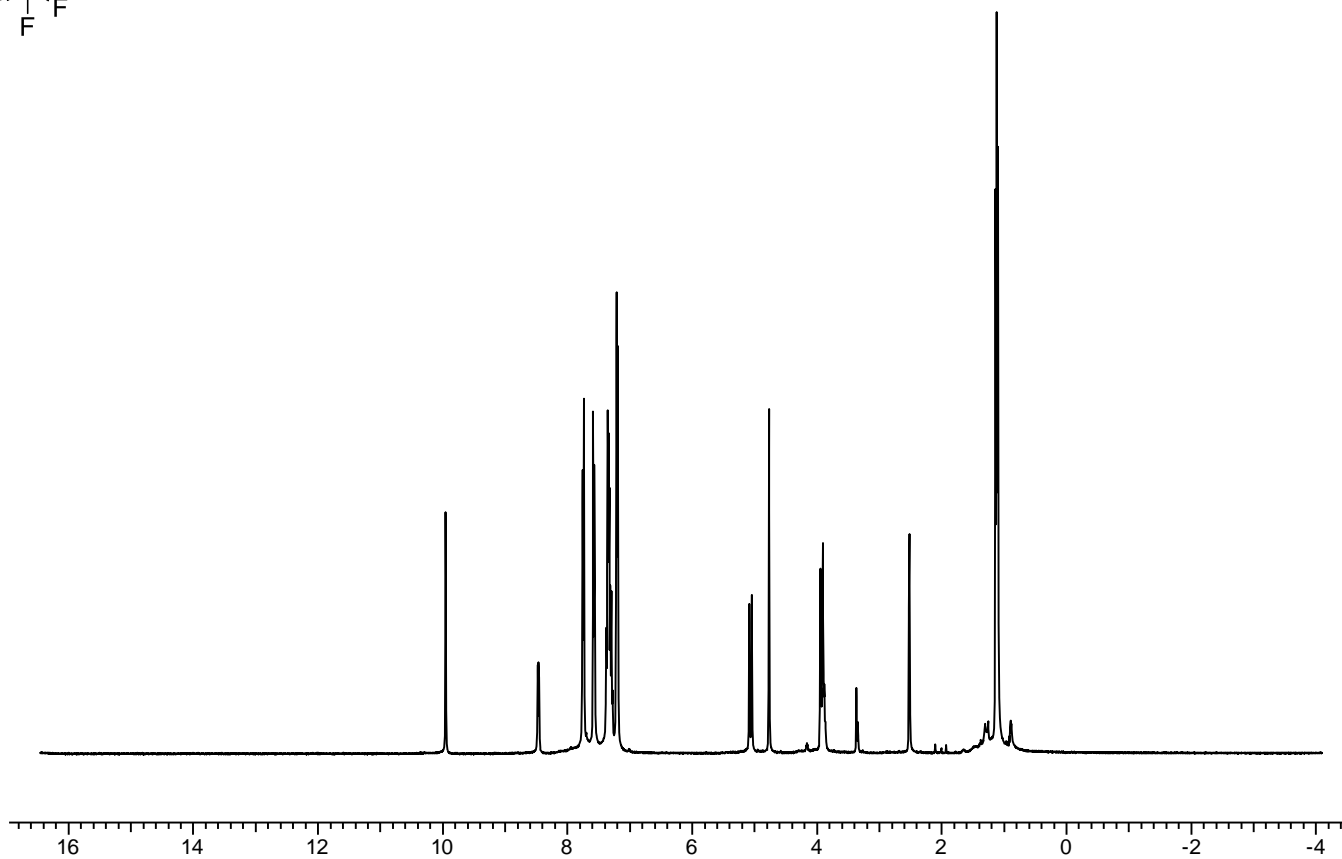
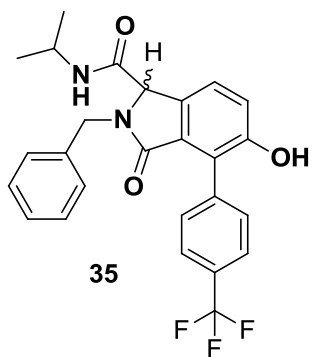
33: <sup>13</sup>C NMR



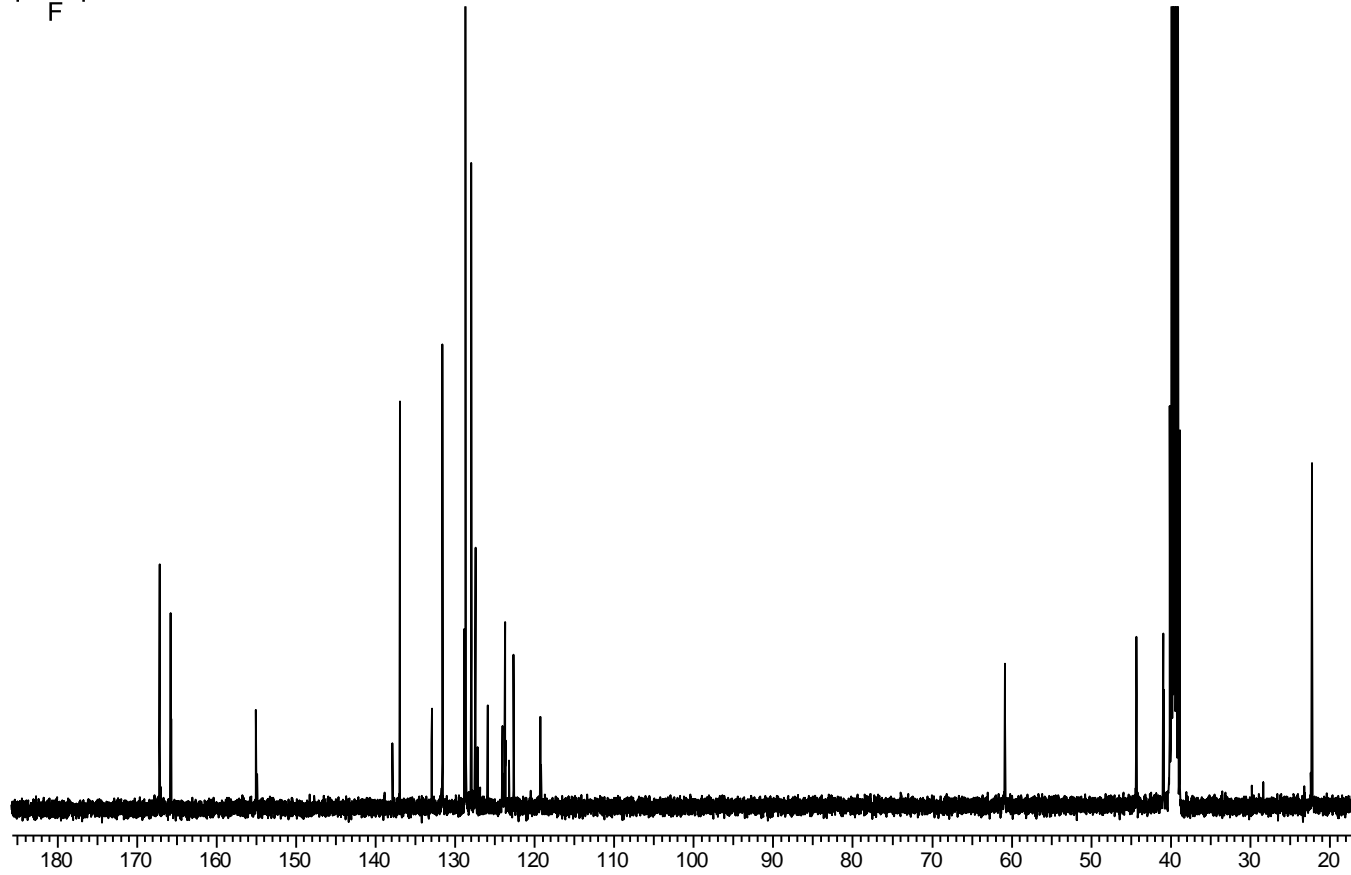
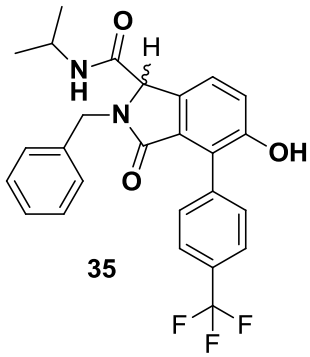
34: <sup>1</sup>H NMR



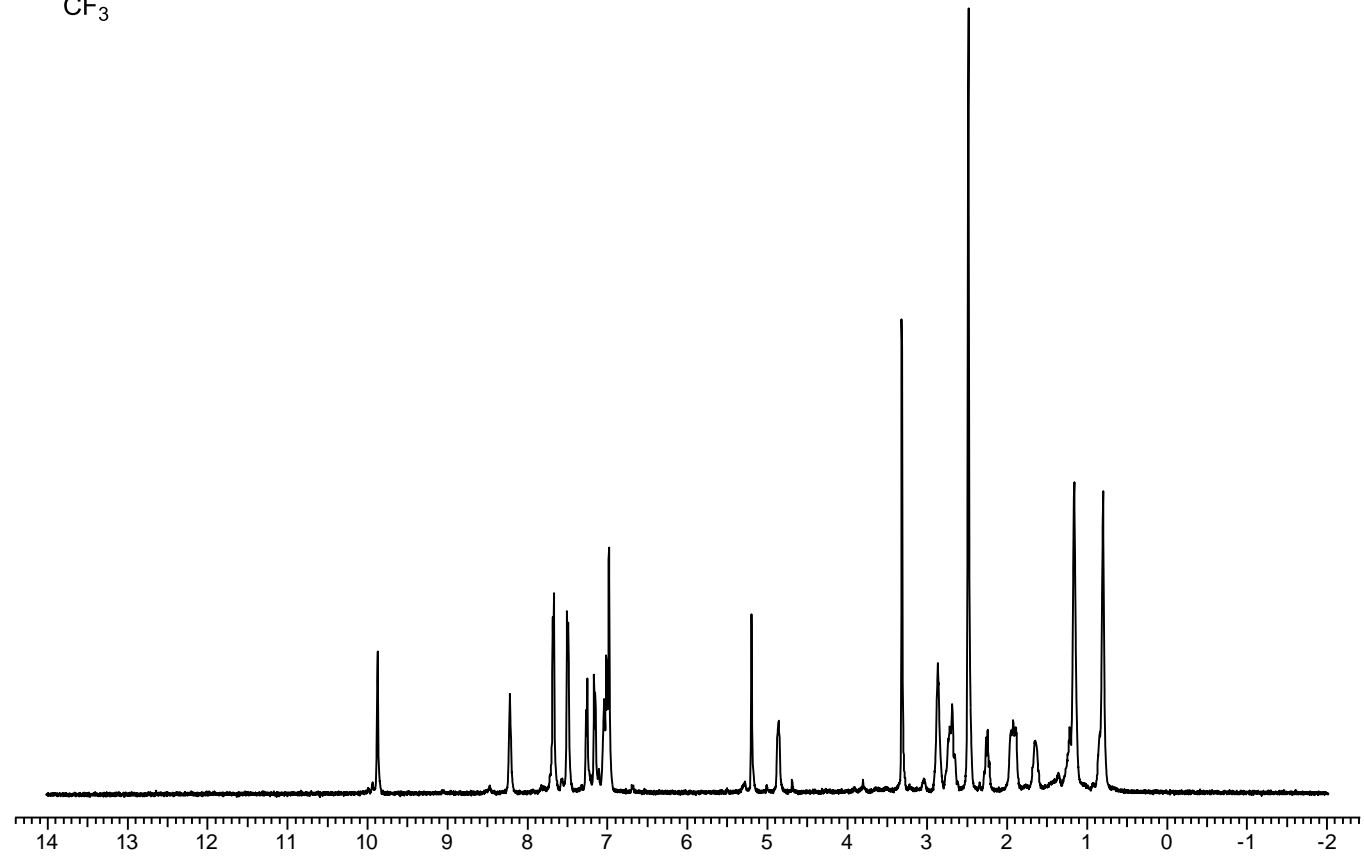
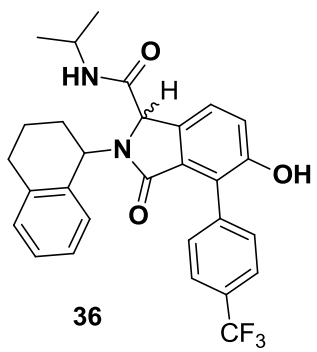
34:  $^{13}\text{C}$  NMR



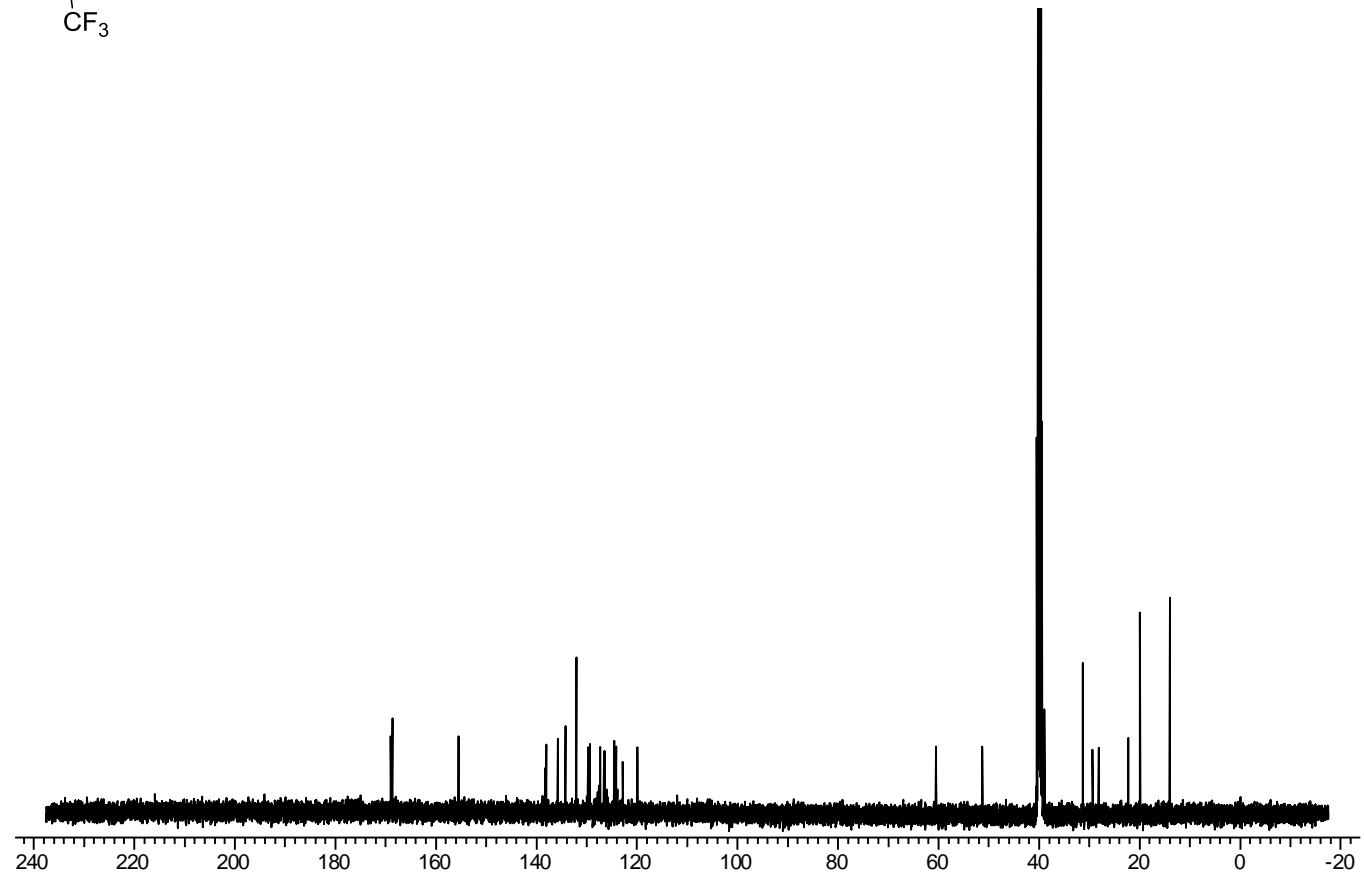
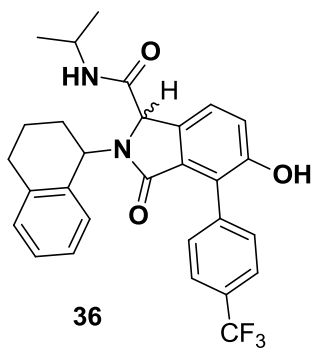
35:  $^1\text{H}$  NMR



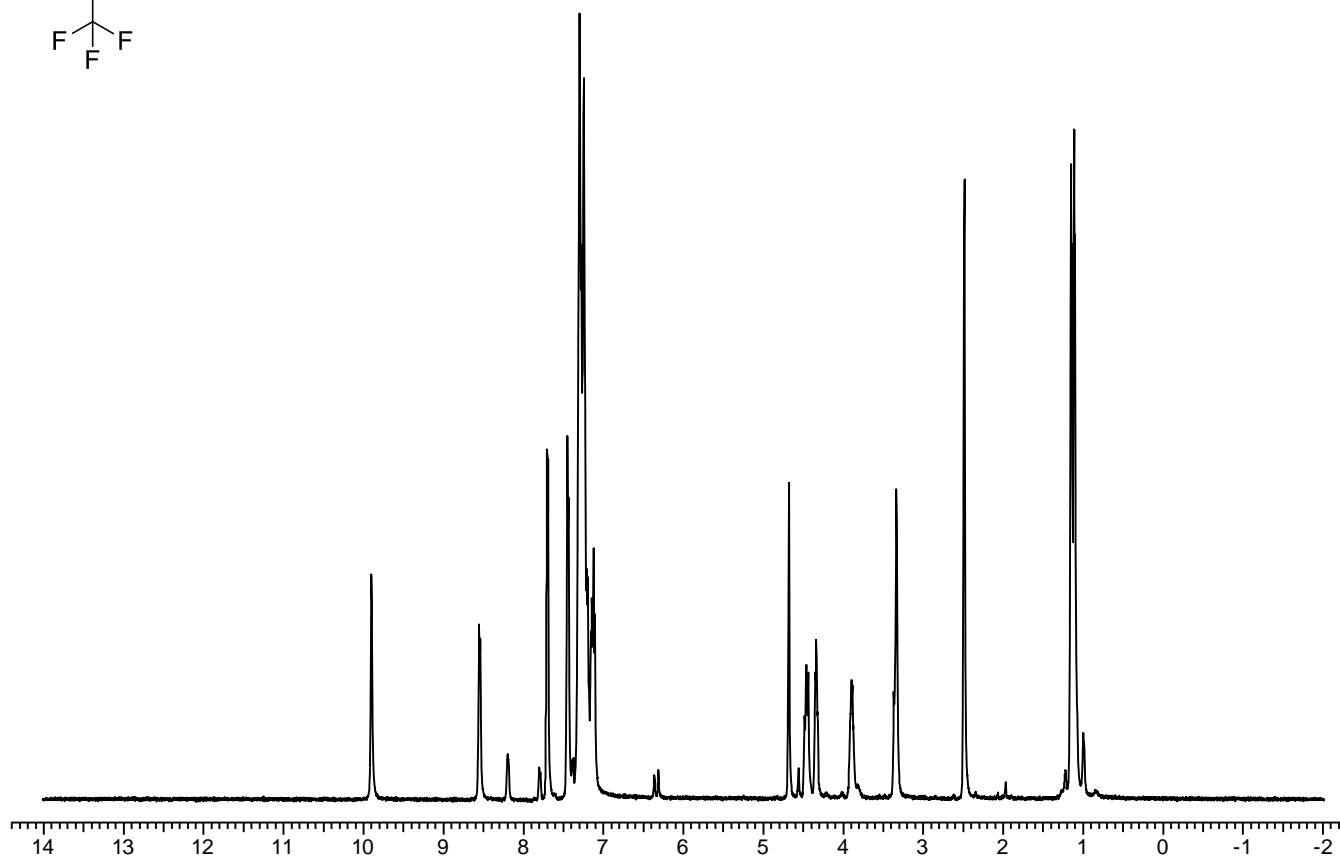
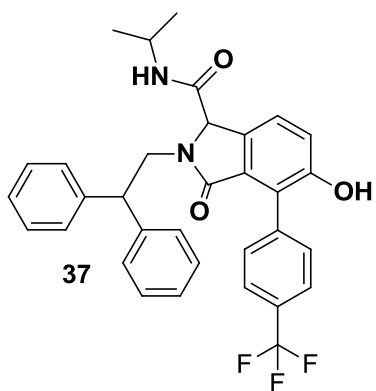
35:  $^{13}\text{C}$  NMR



36: <sup>1</sup>H NMR

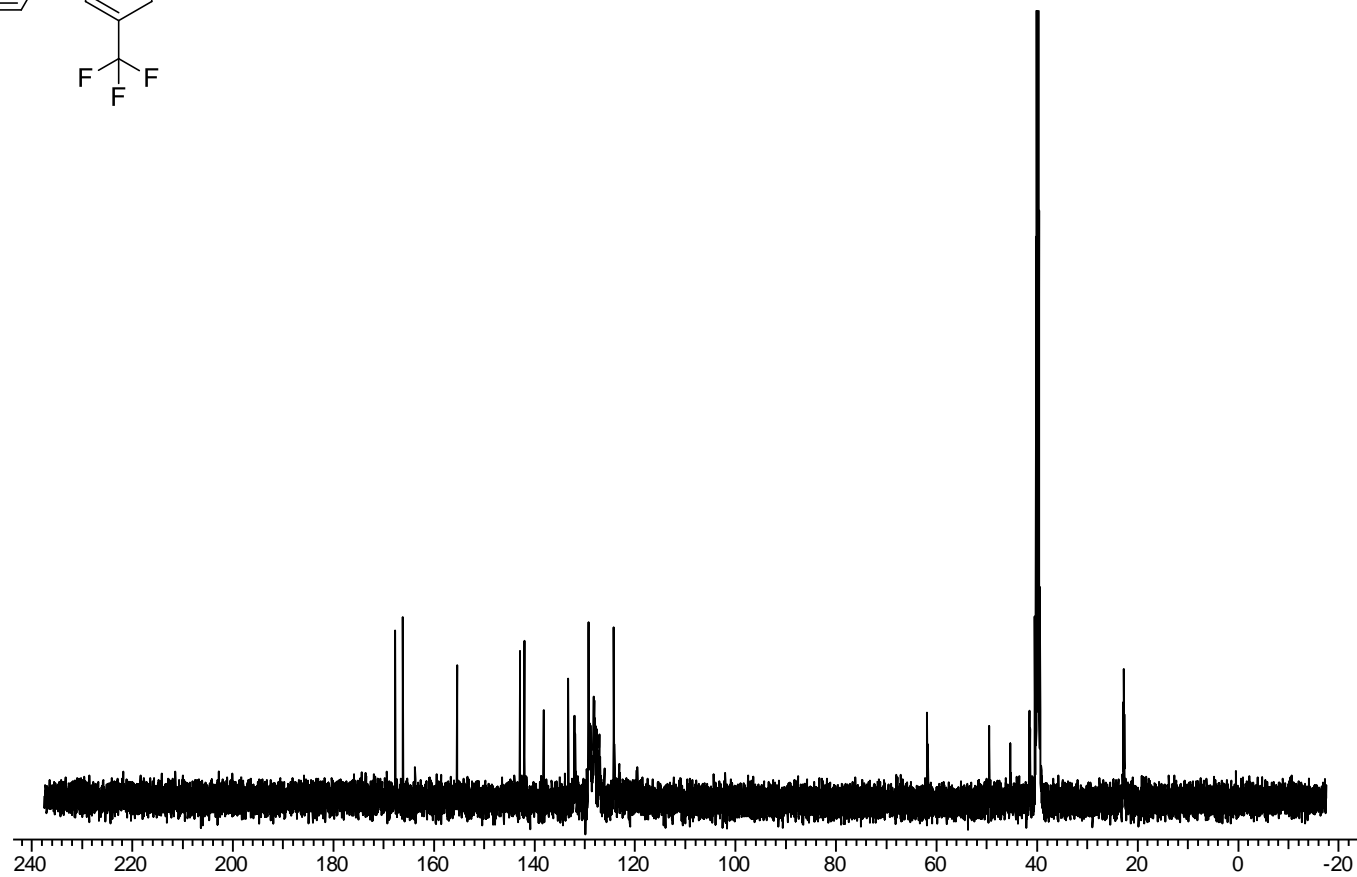
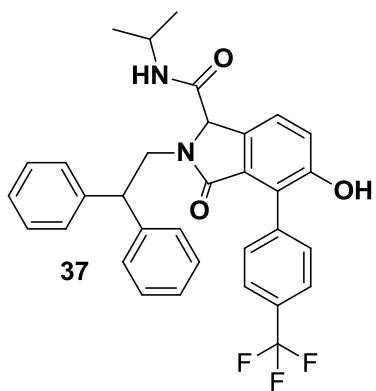


36: <sup>13</sup>C NMR

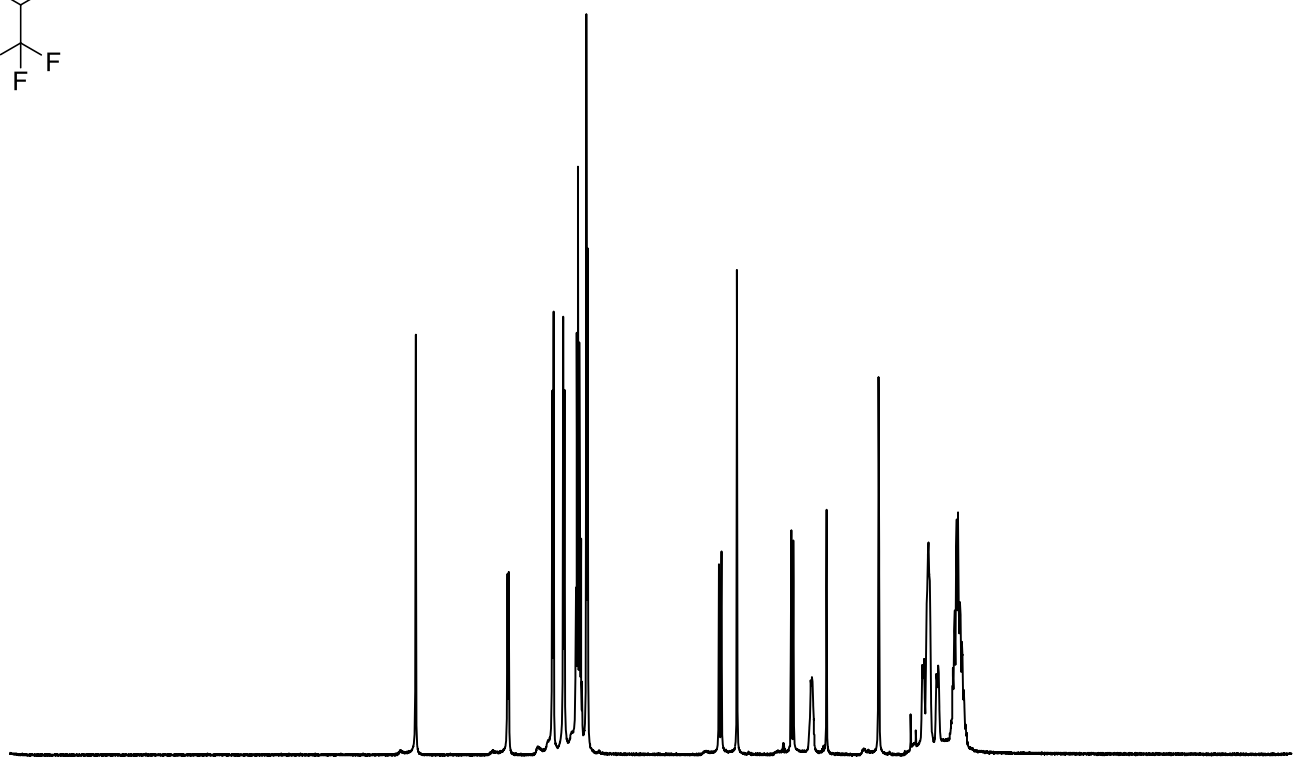
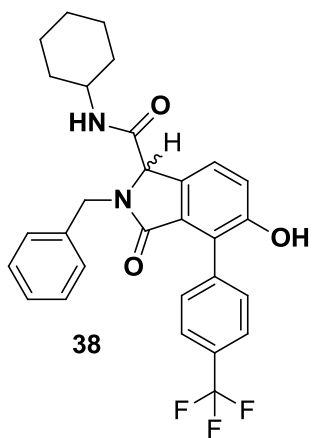


37:  $^1\text{H}$  NMR

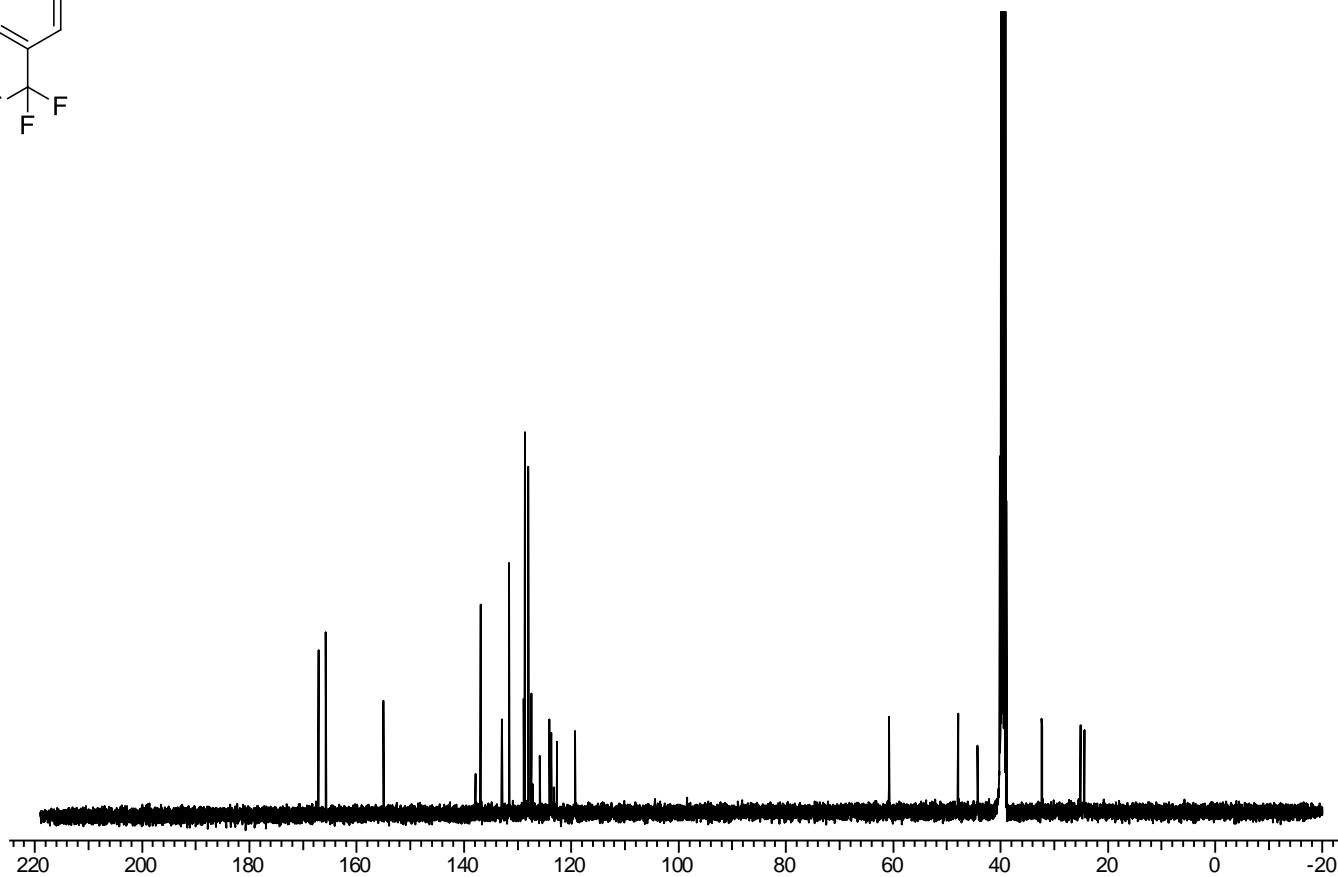
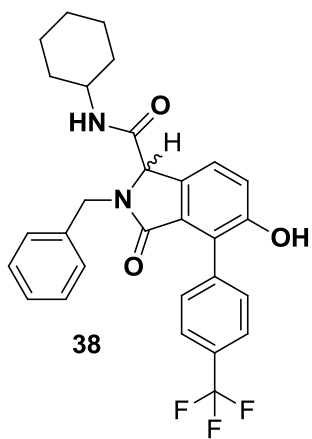




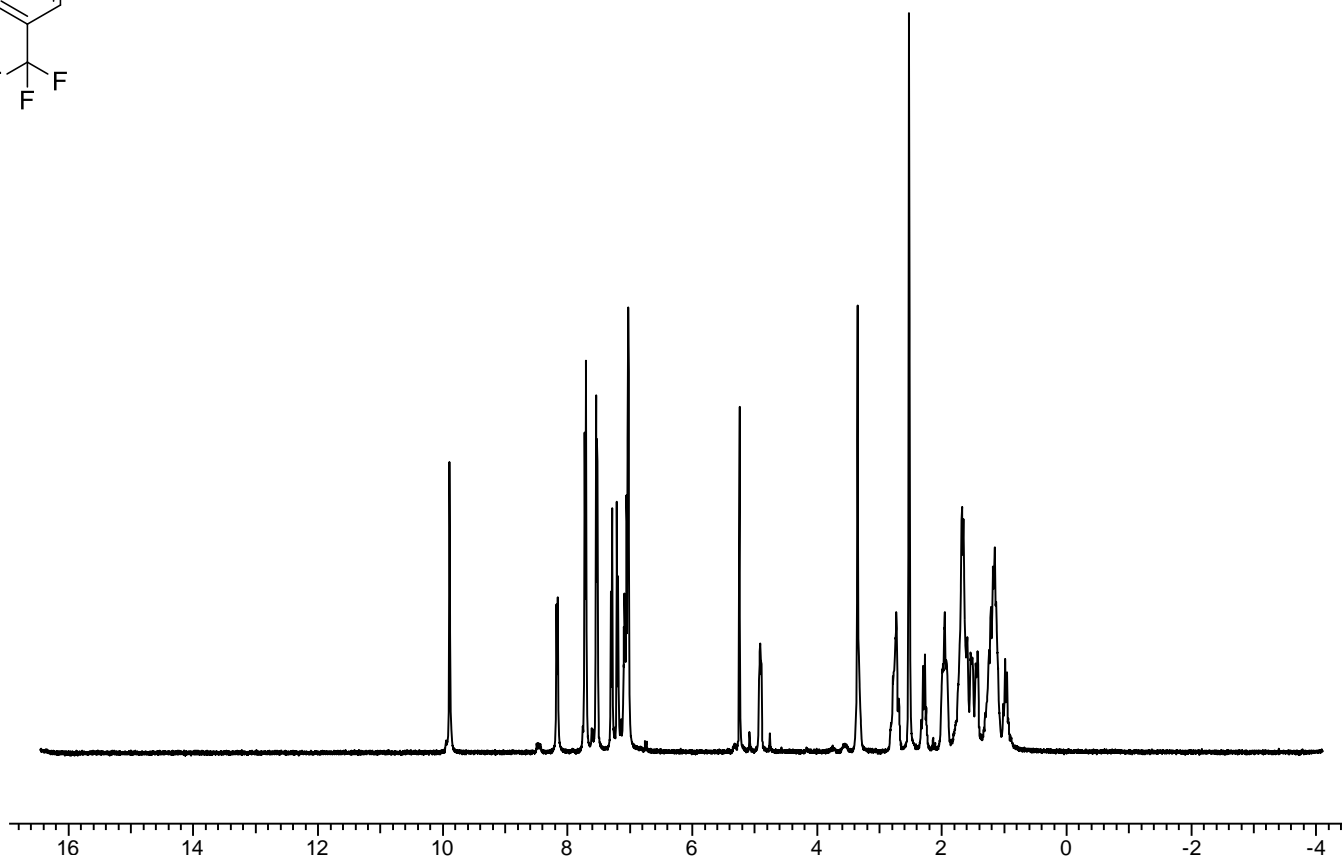
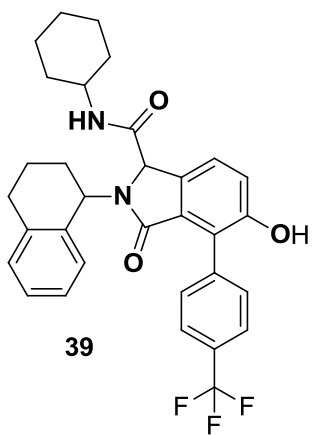
37:  $^{13}\text{C}$  NMR



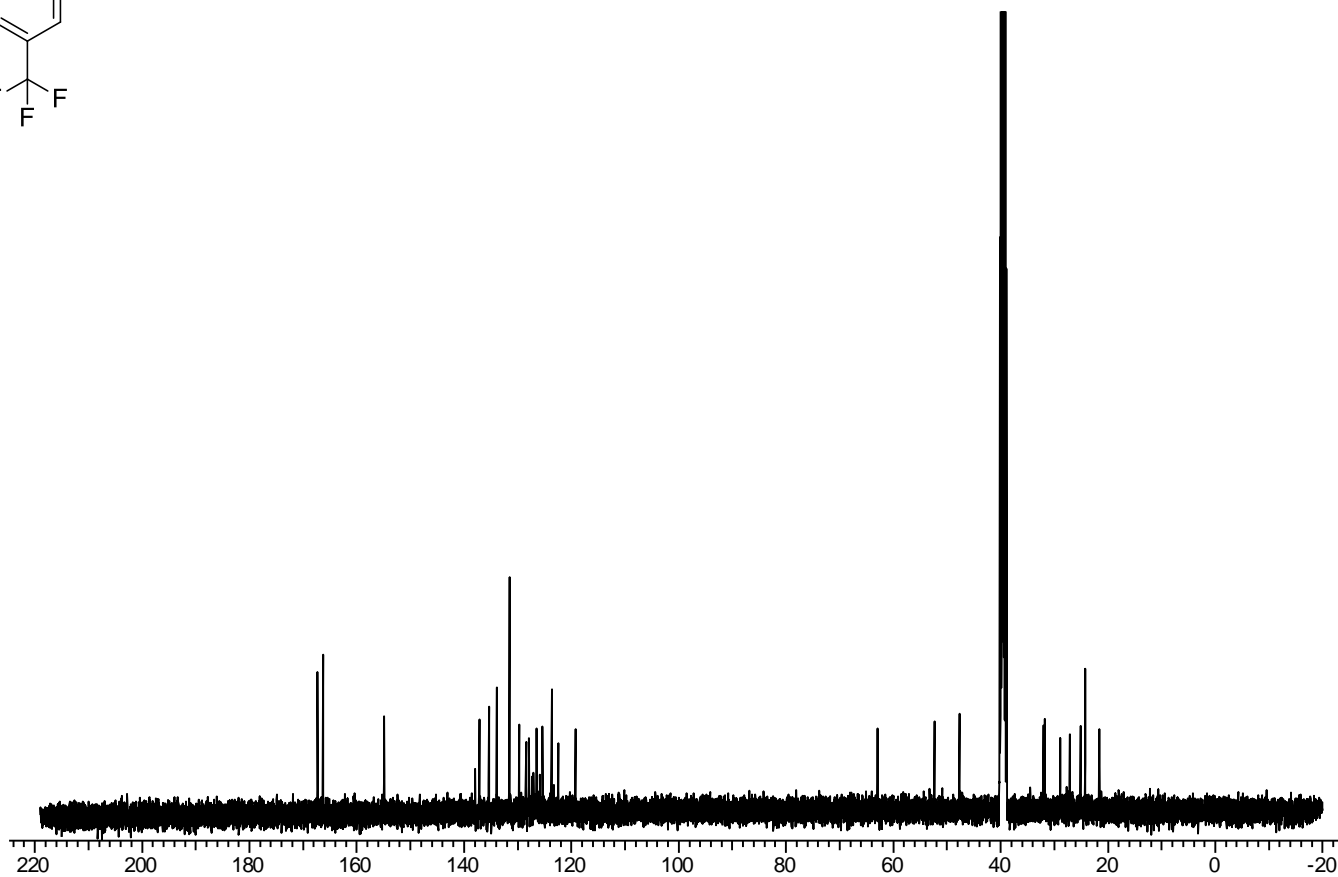
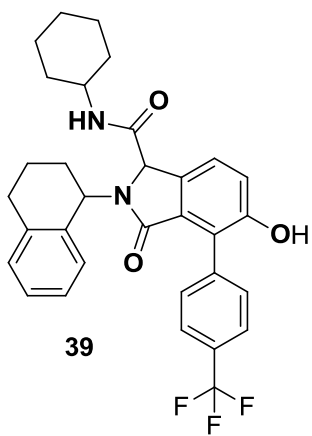
38:  $^1\text{H}$  NMR



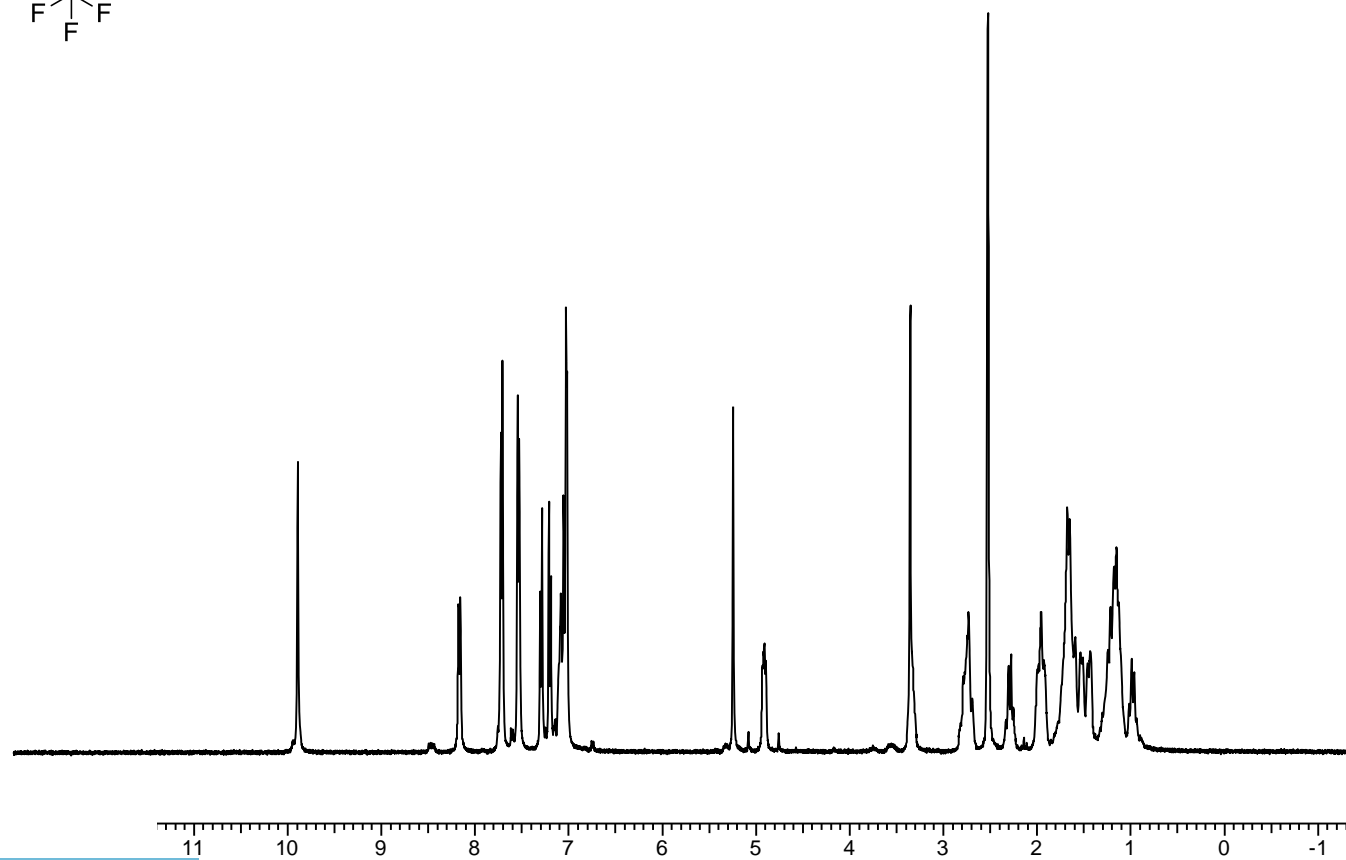
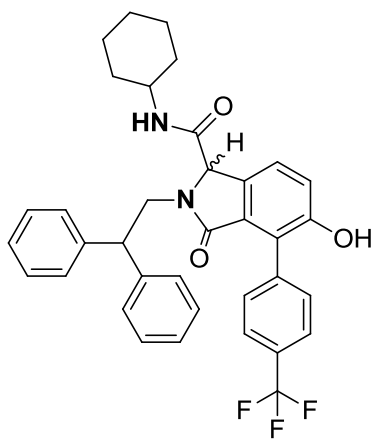
38:  $^{13}\text{C}$  NMR



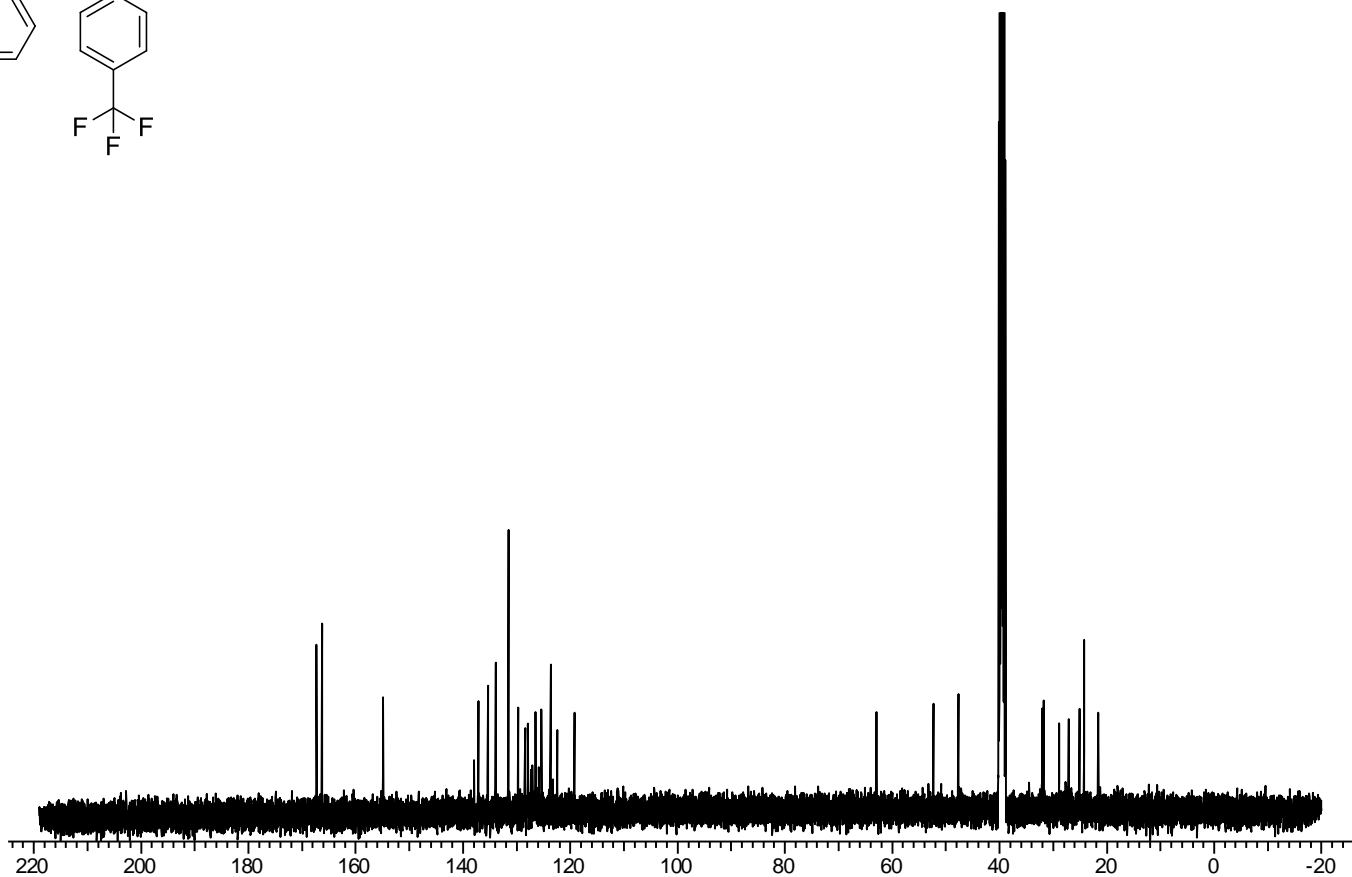
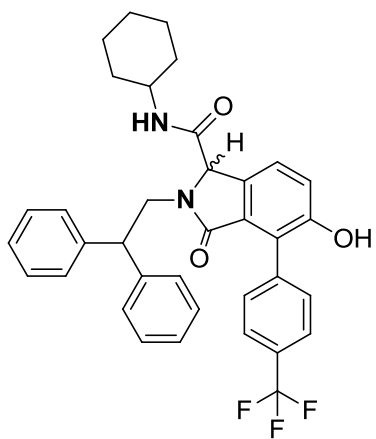
39: <sup>1</sup>H NMR



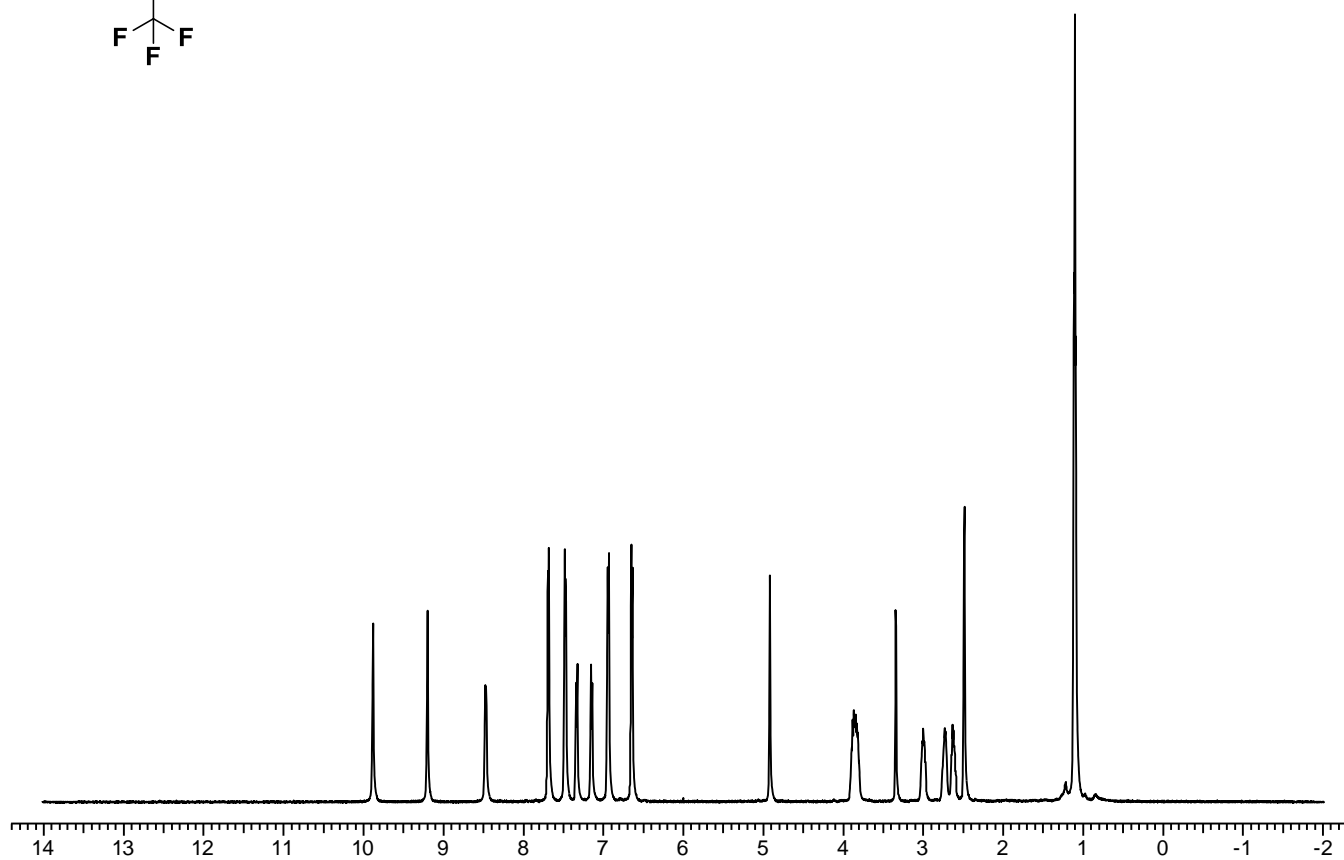
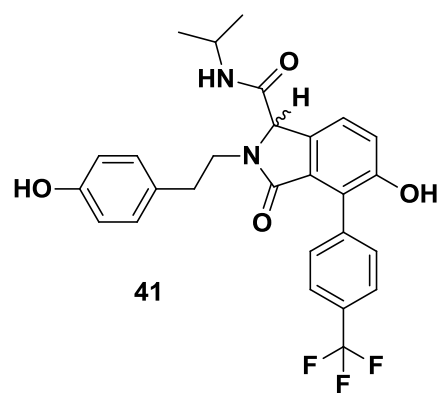
39:  $^{13}\text{C}$  NMR



40:  $^1\text{H}$  NMR

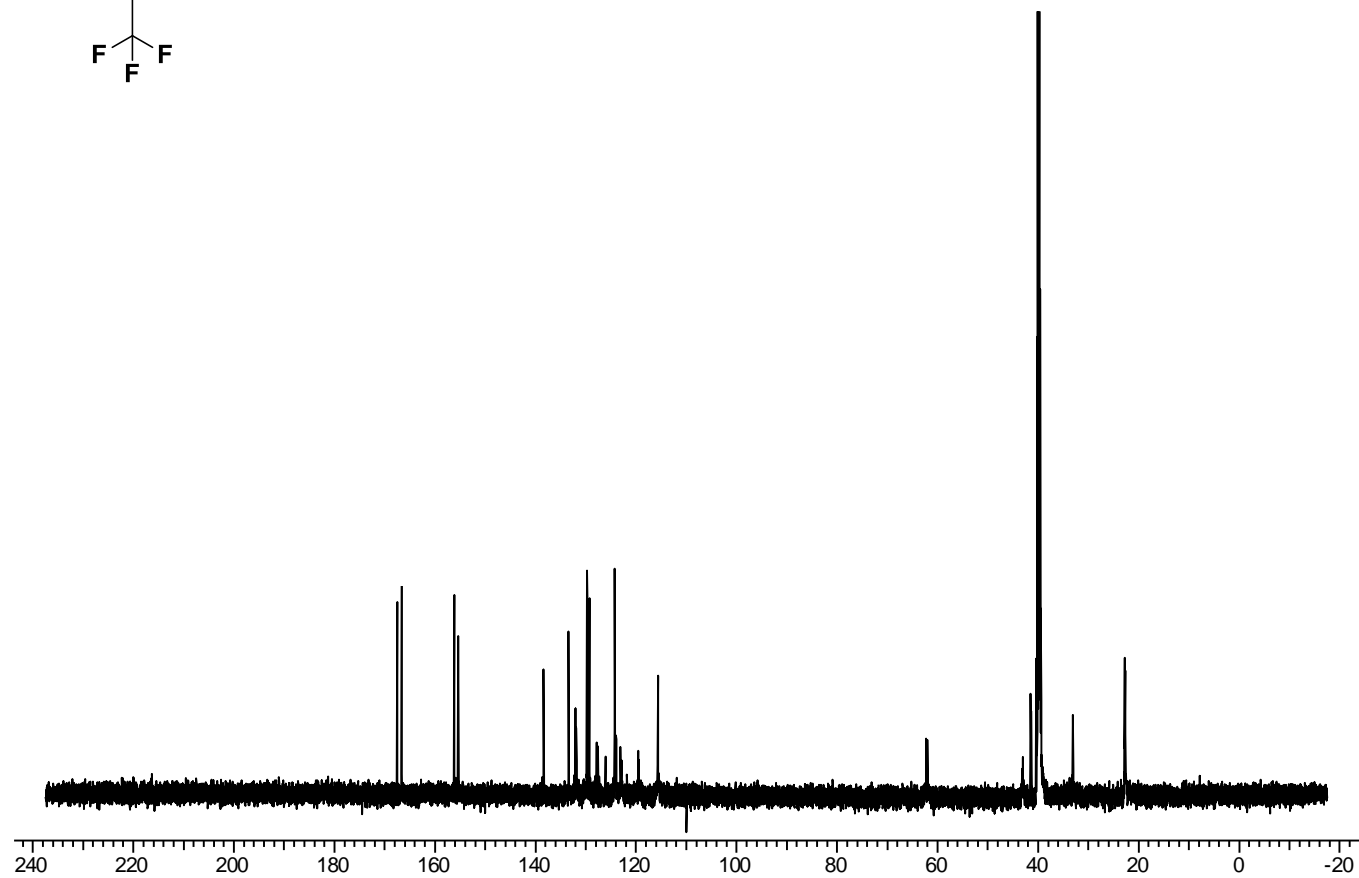
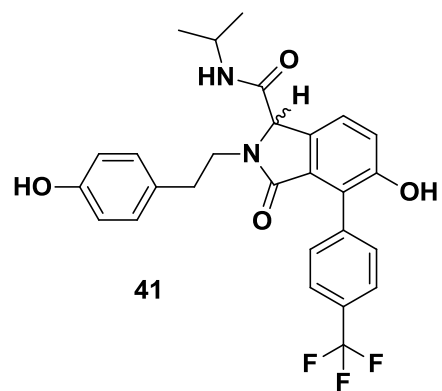


40: <sup>13</sup>C NMR

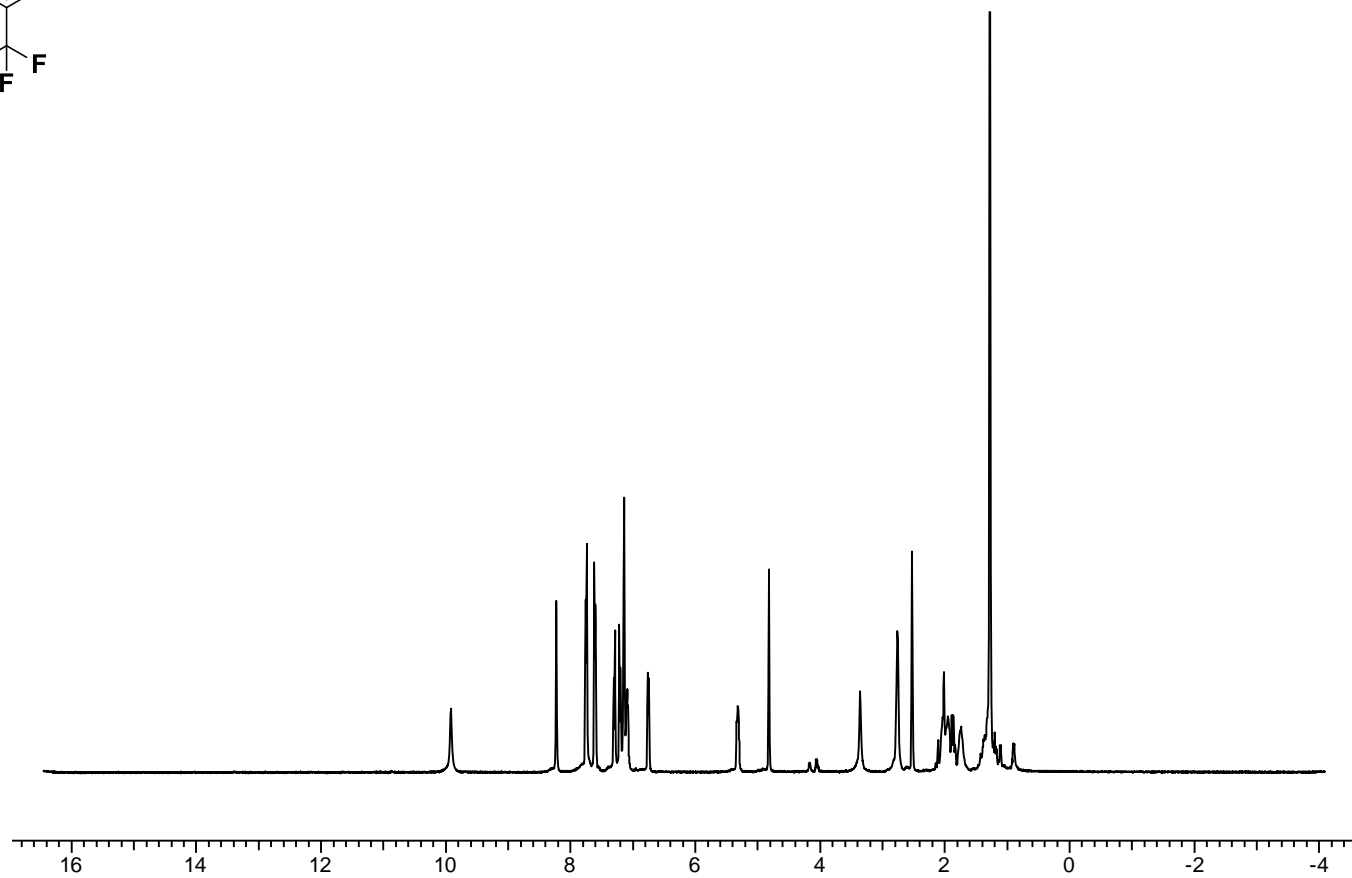
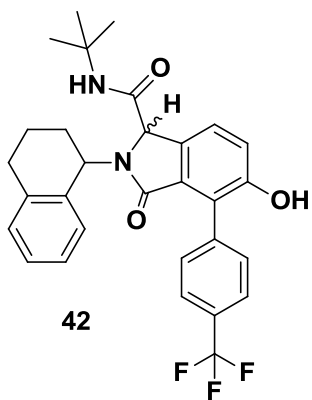


41:  $^1\text{H}$  NMR

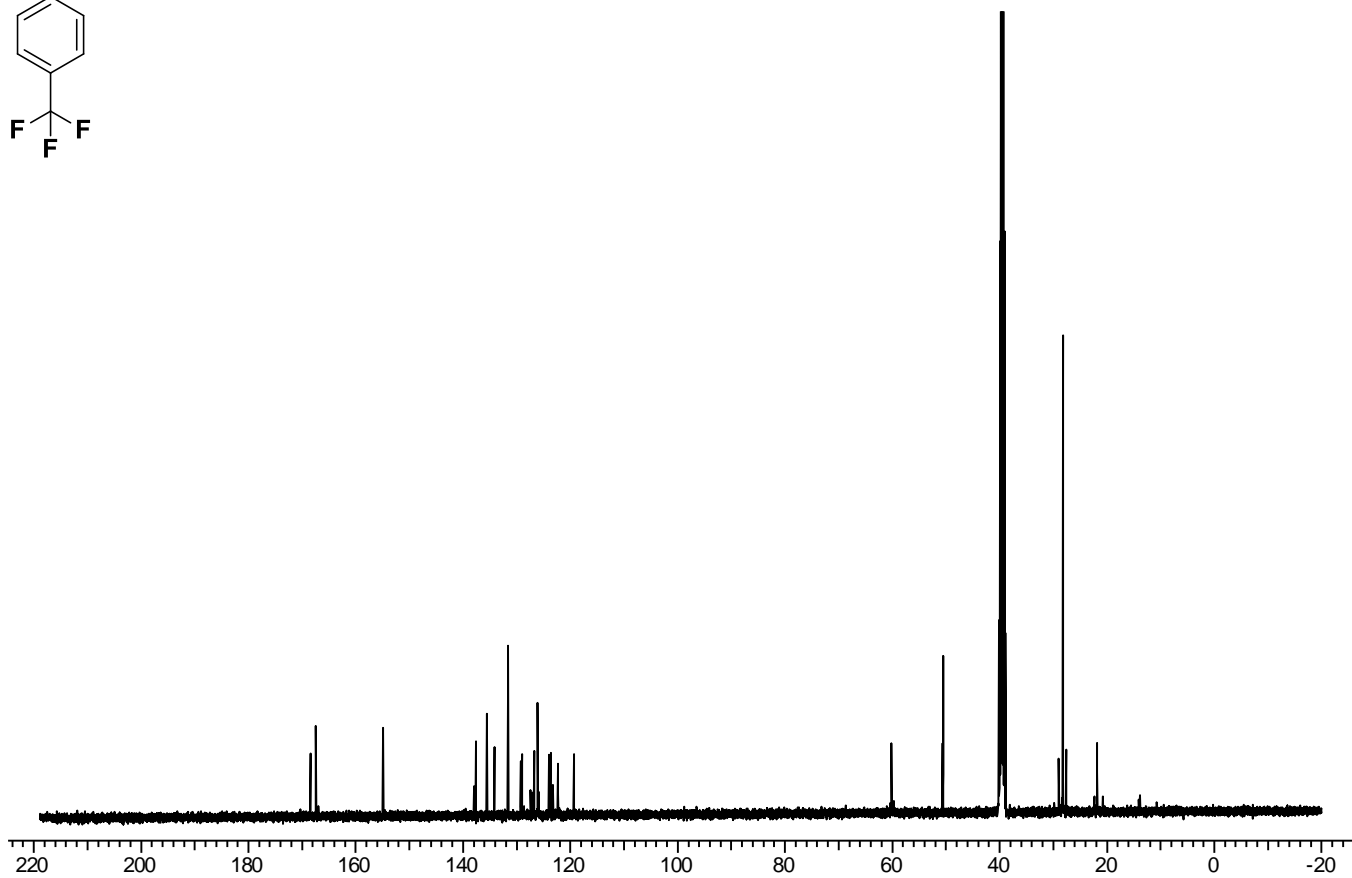
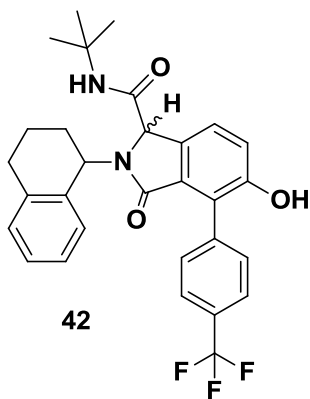




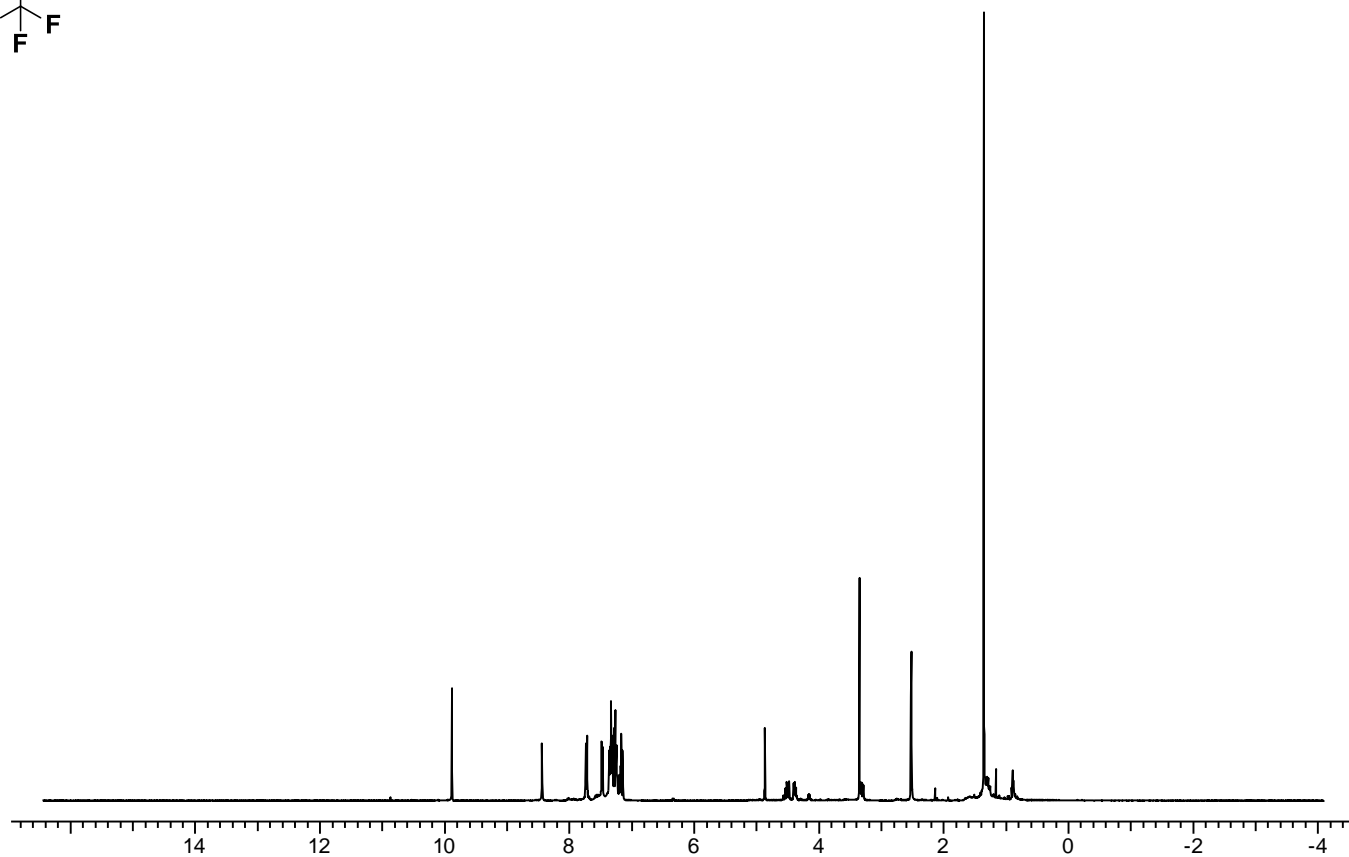
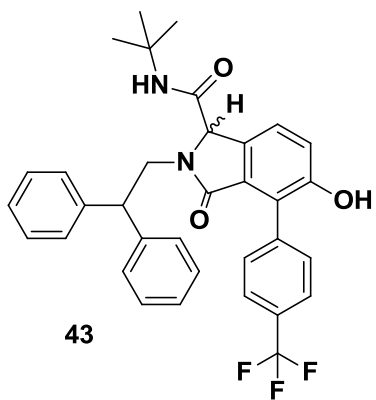
41: <sup>13</sup>C NMR



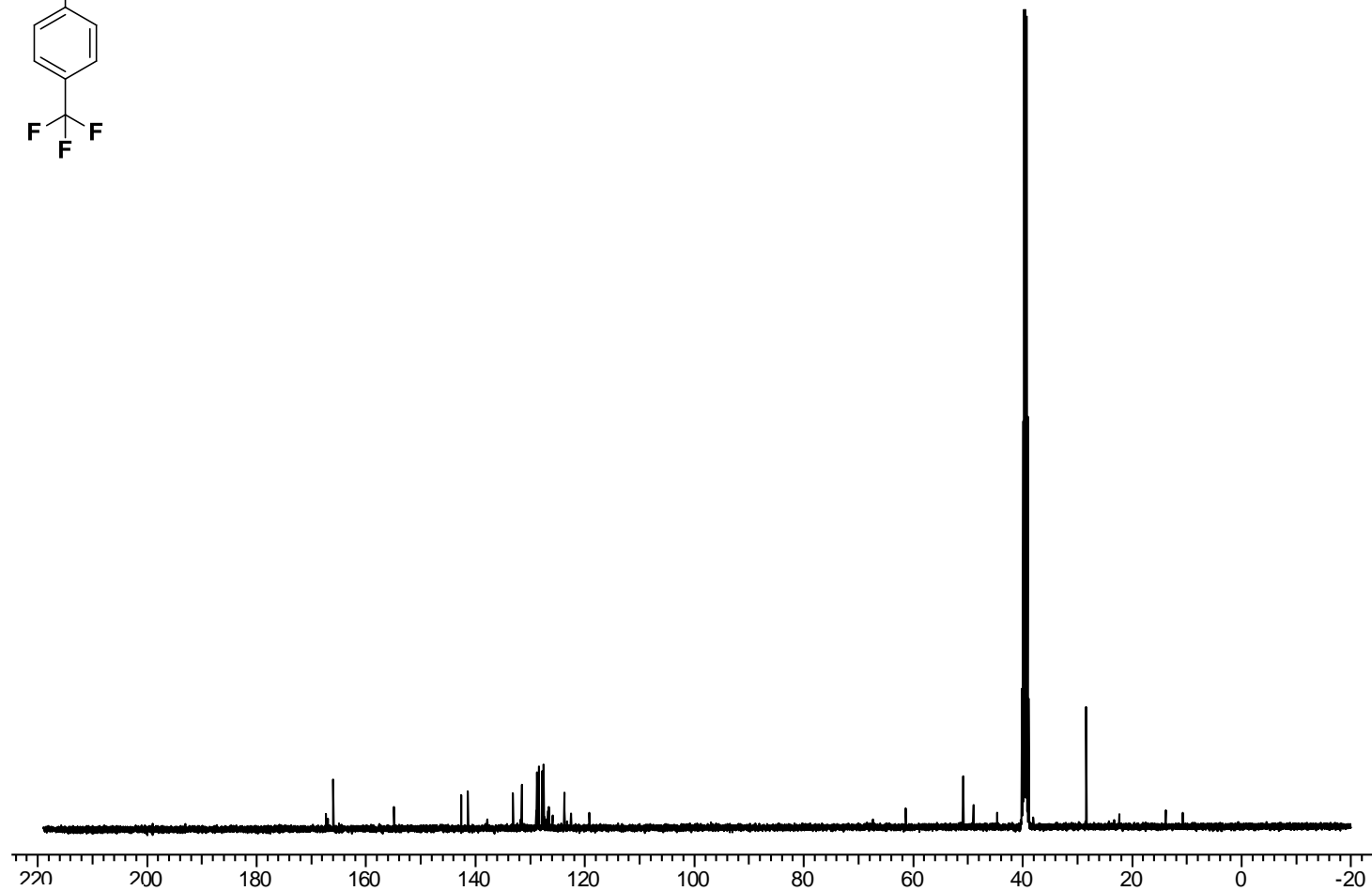
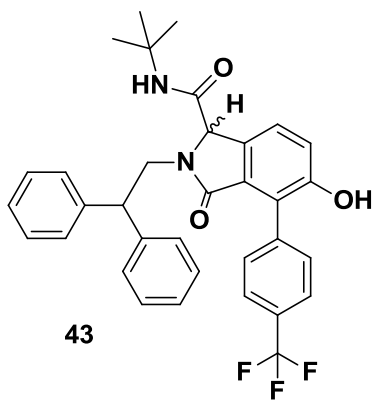
42:  $^1\text{H}$  NMR



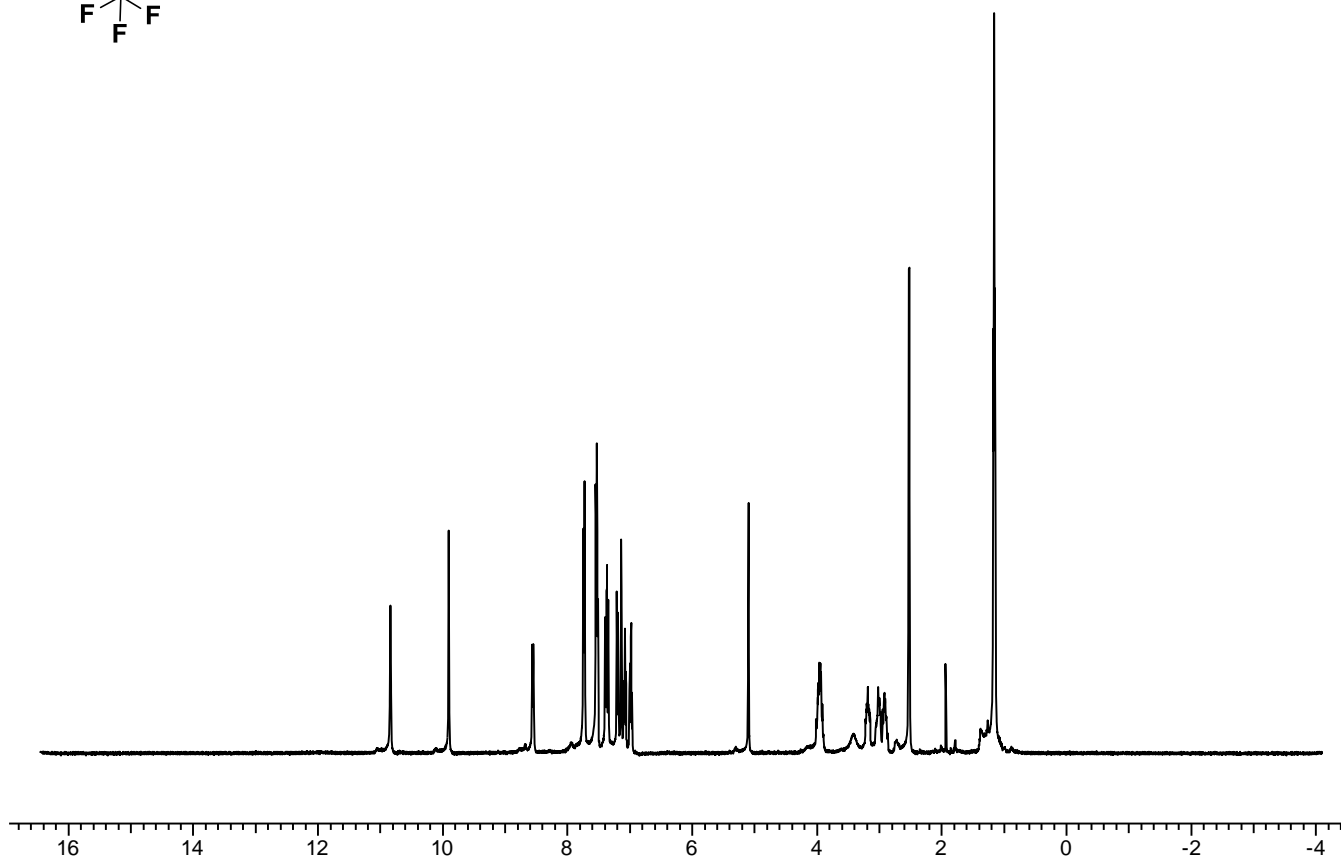
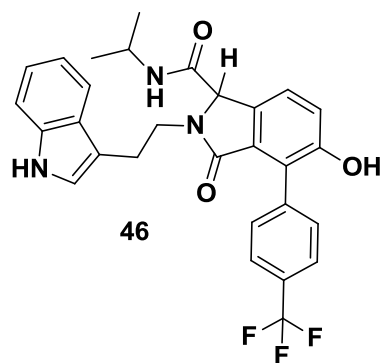
42:  $^{13}\text{C}$  NMR



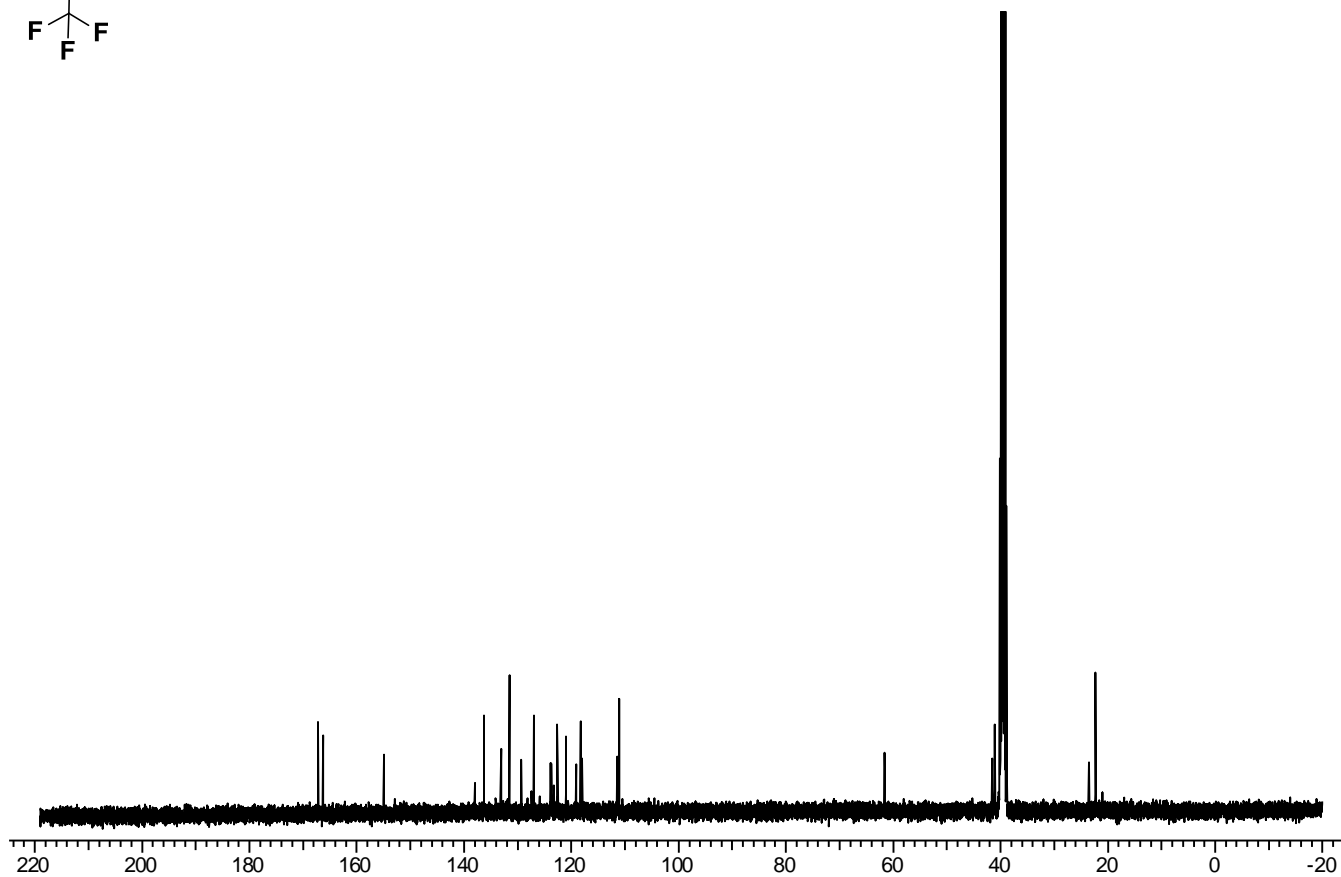
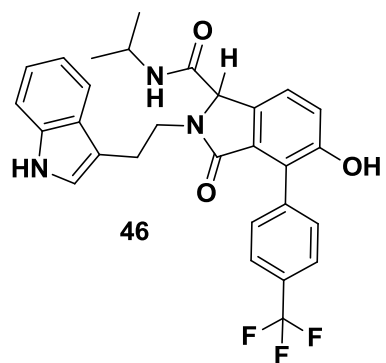
43:  $^1\text{H}$  NMR



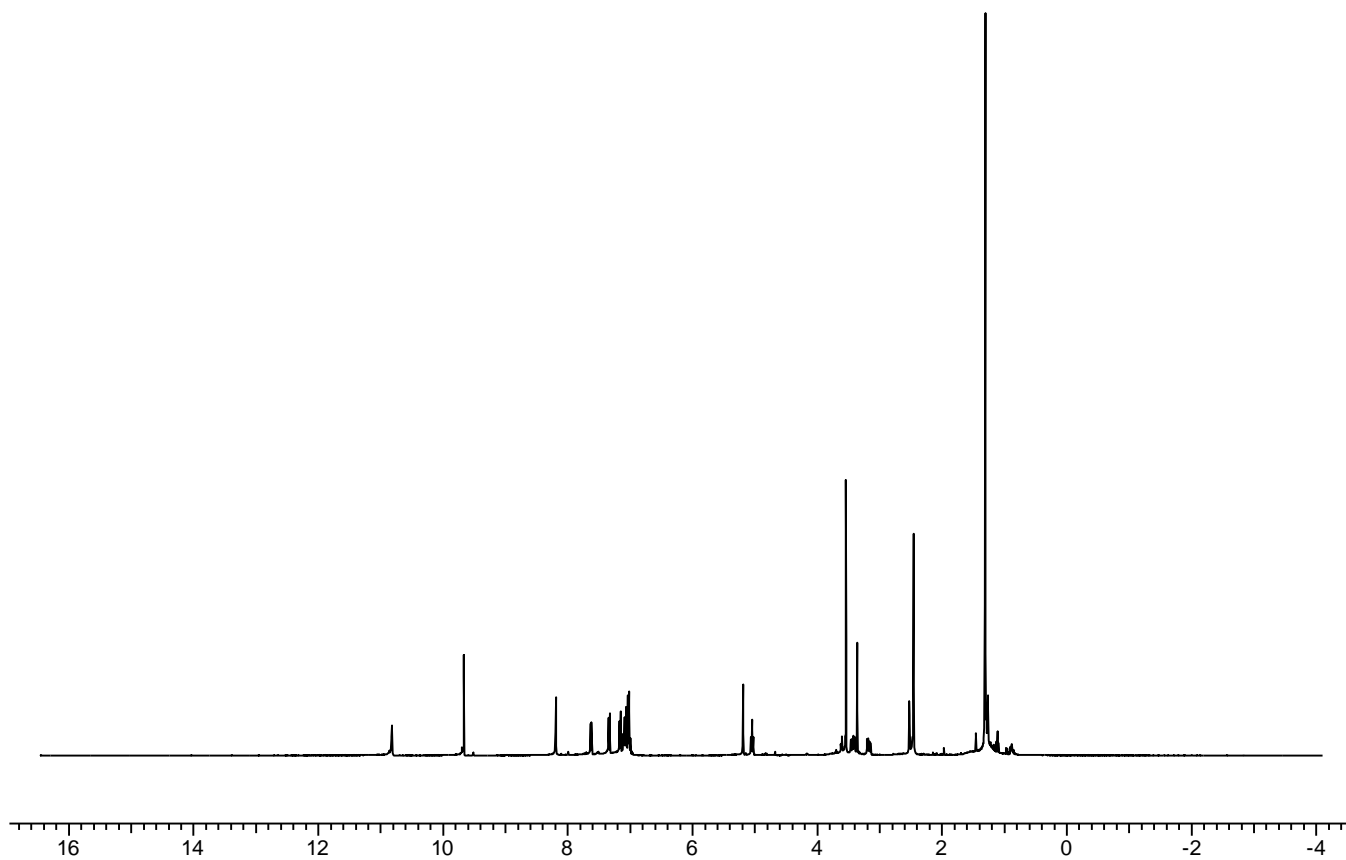
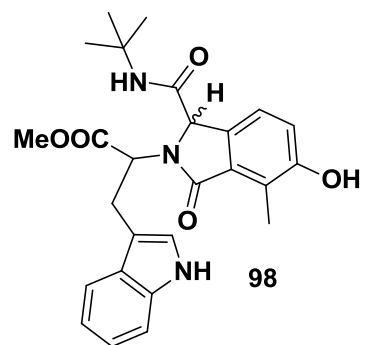
43:  $^{13}\text{C}$  NMR



46:  $^1\text{H}$  NMR

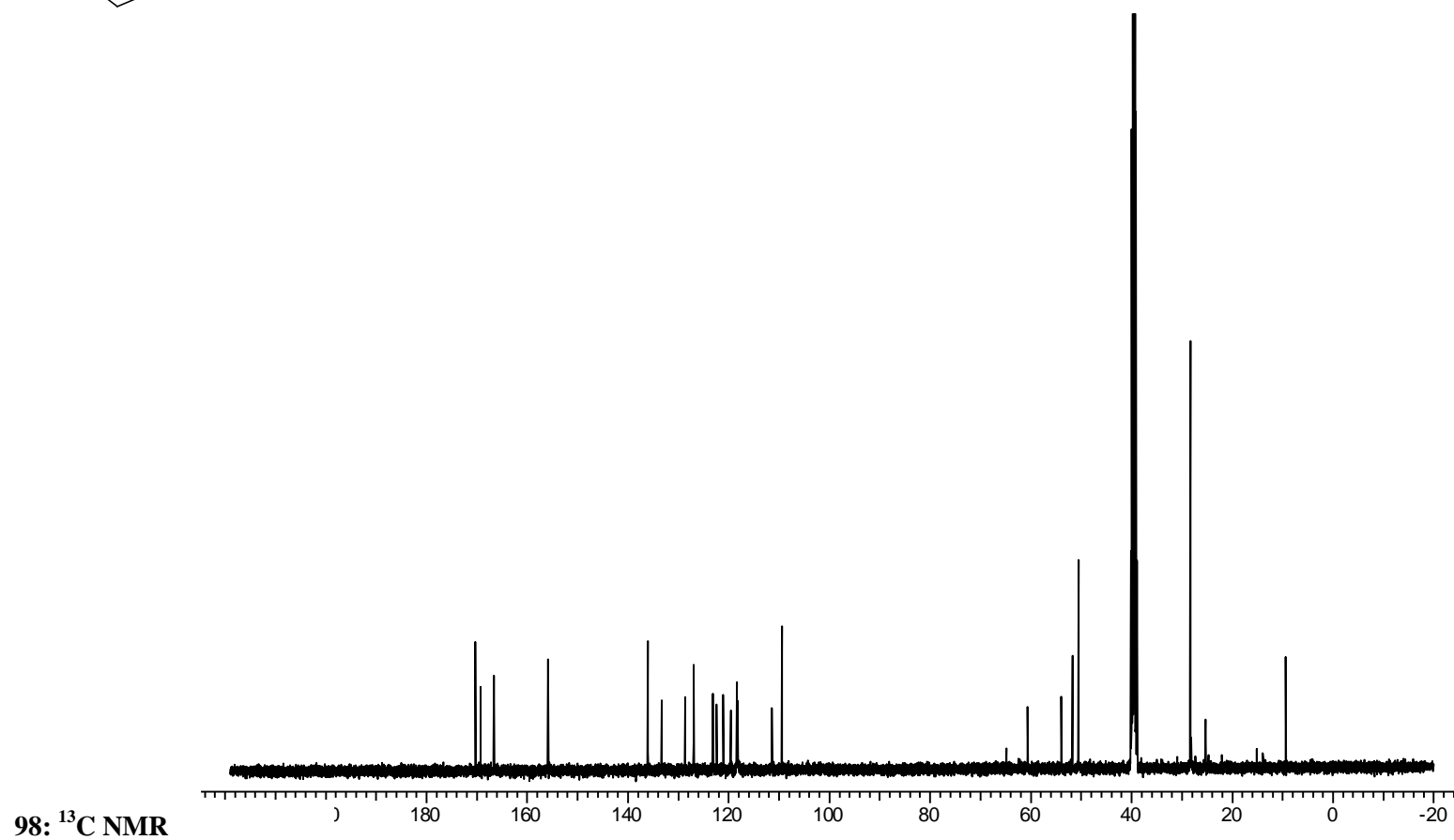
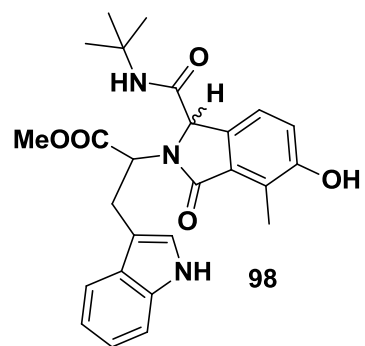


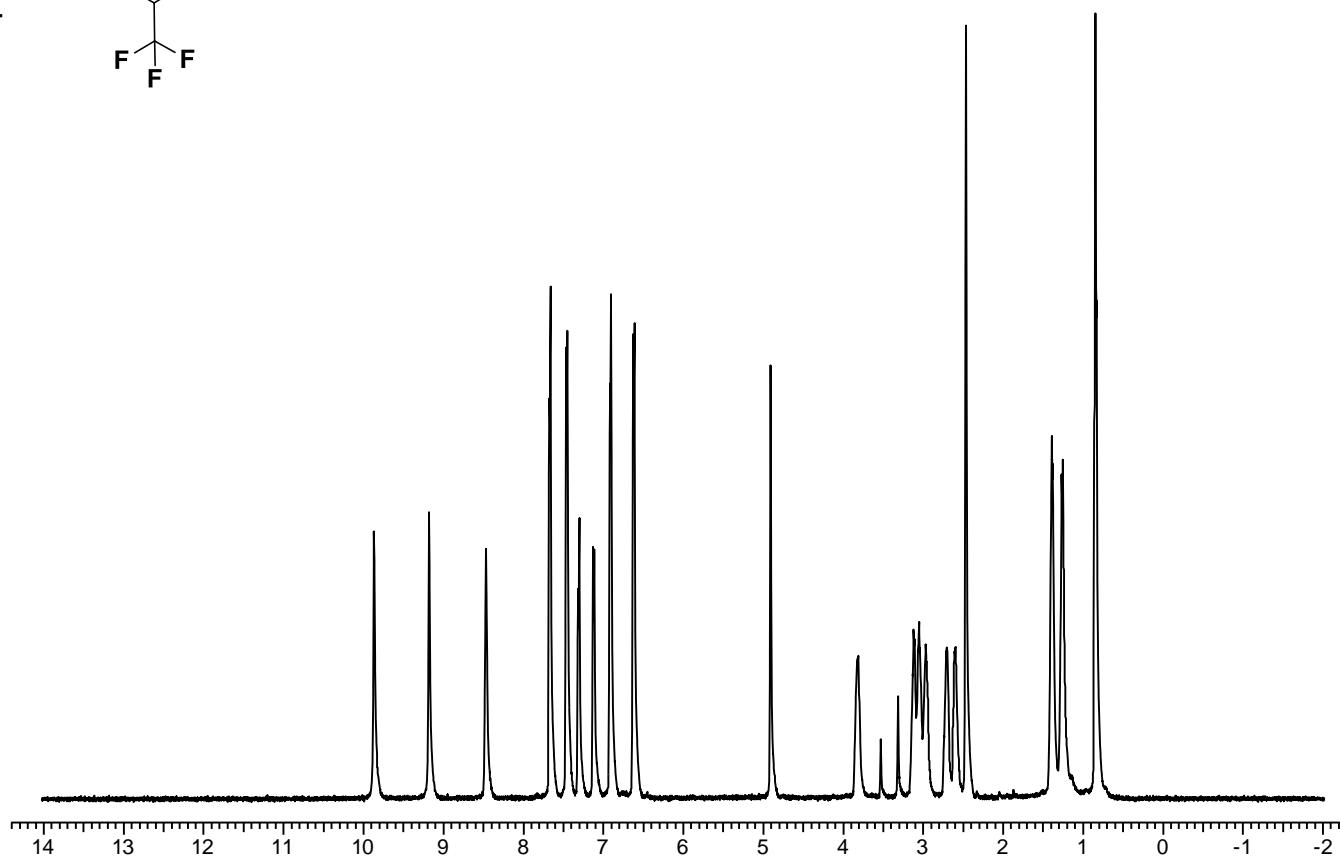
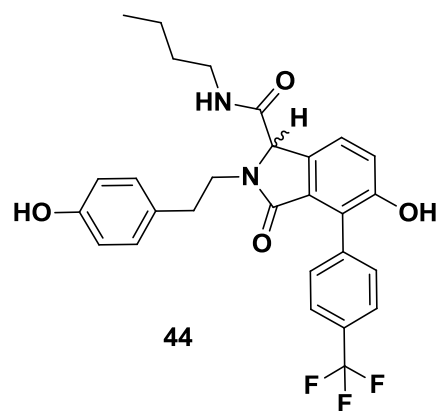
46:  $^{13}\text{C}$  NMR



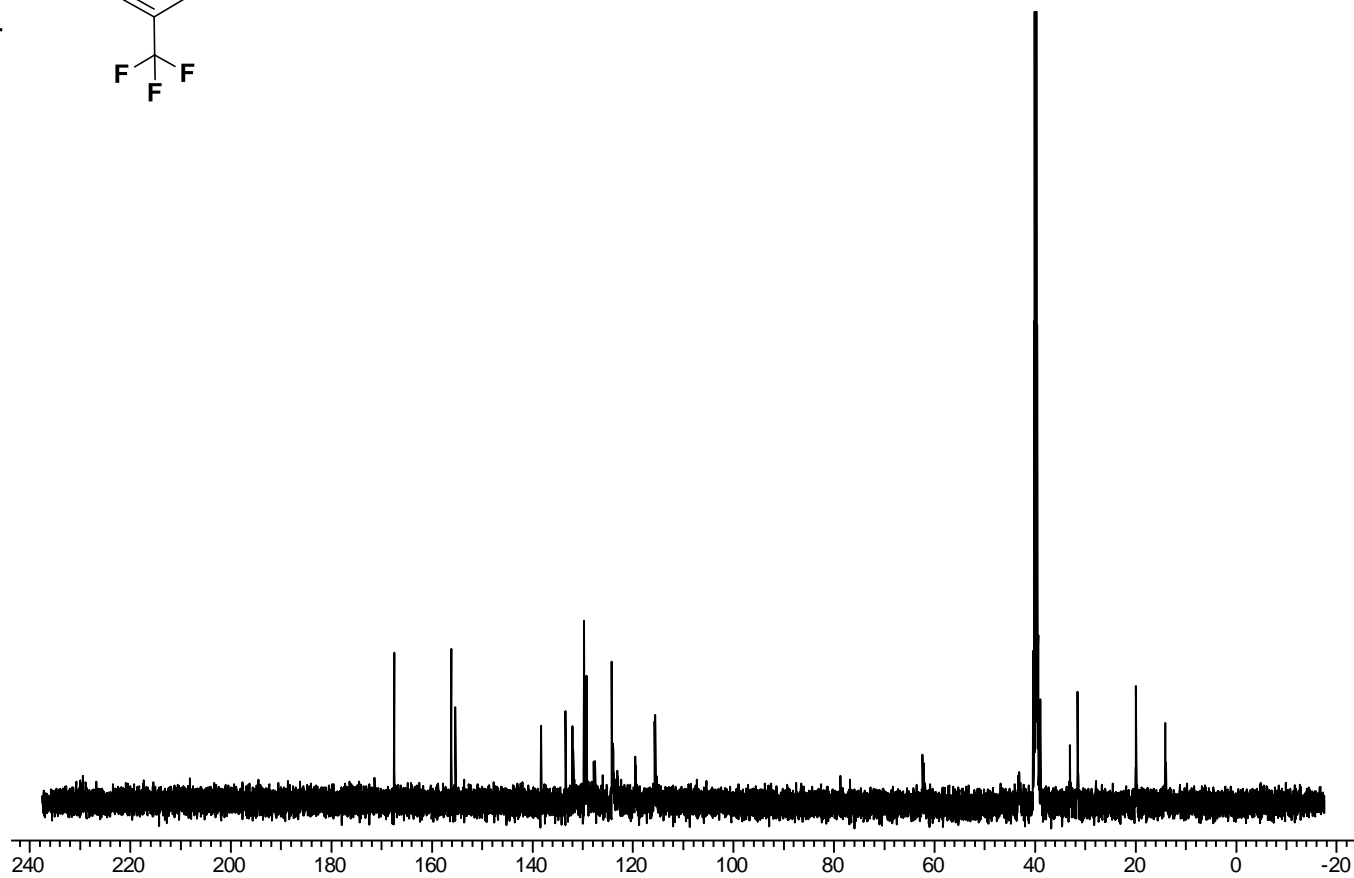
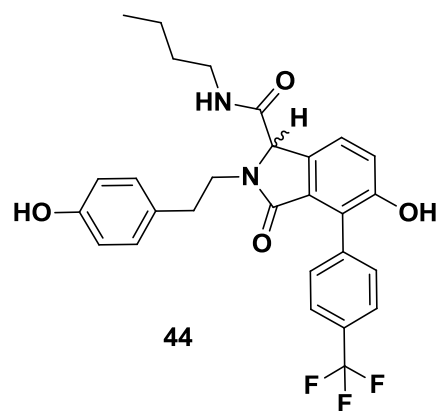
98:  $^1\text{H}$  NMR



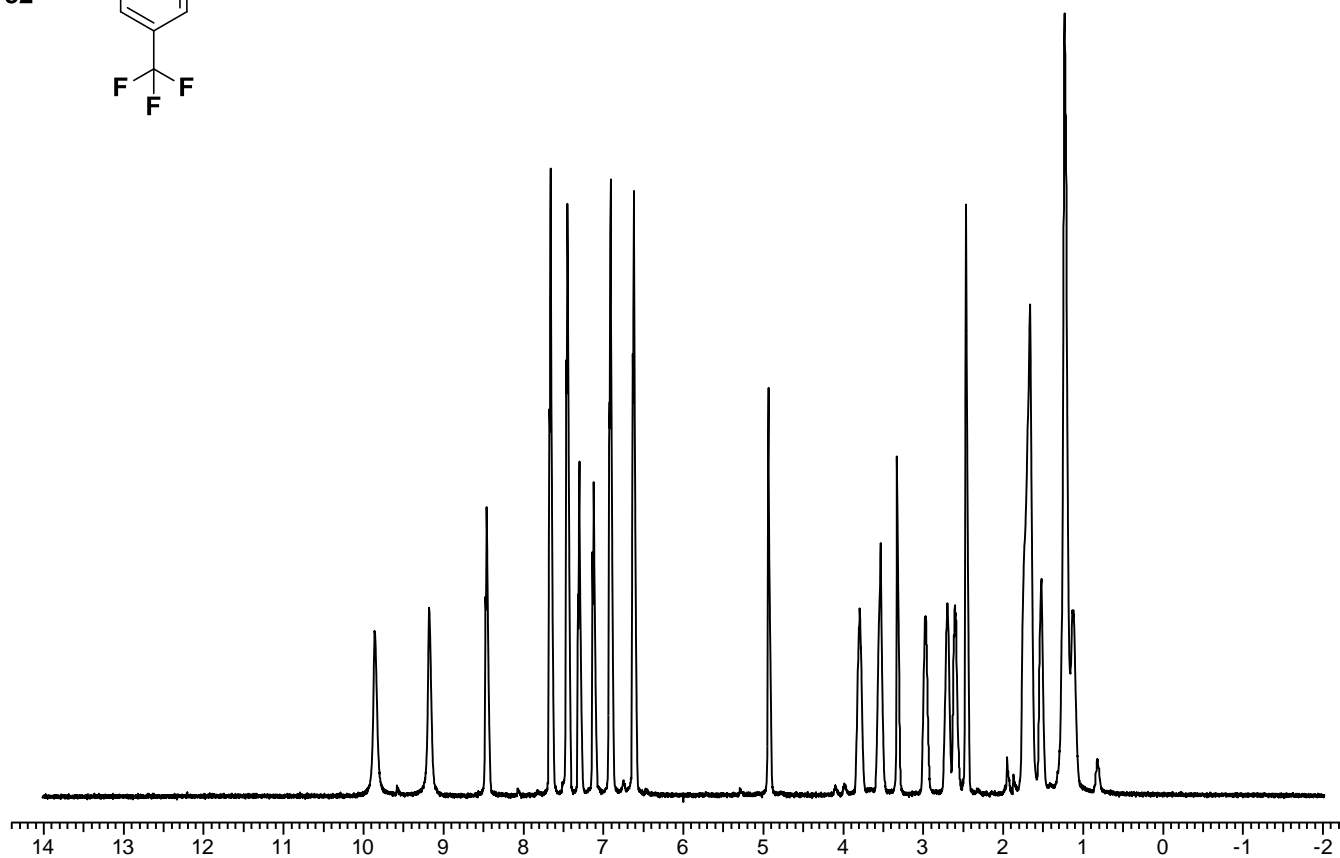
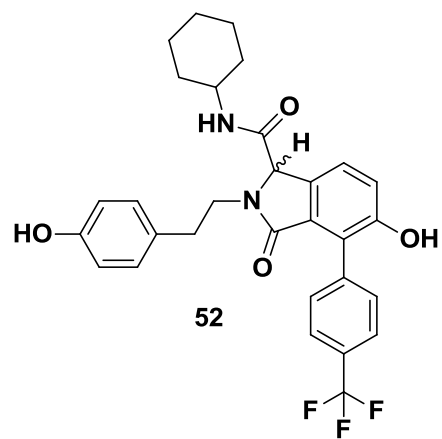




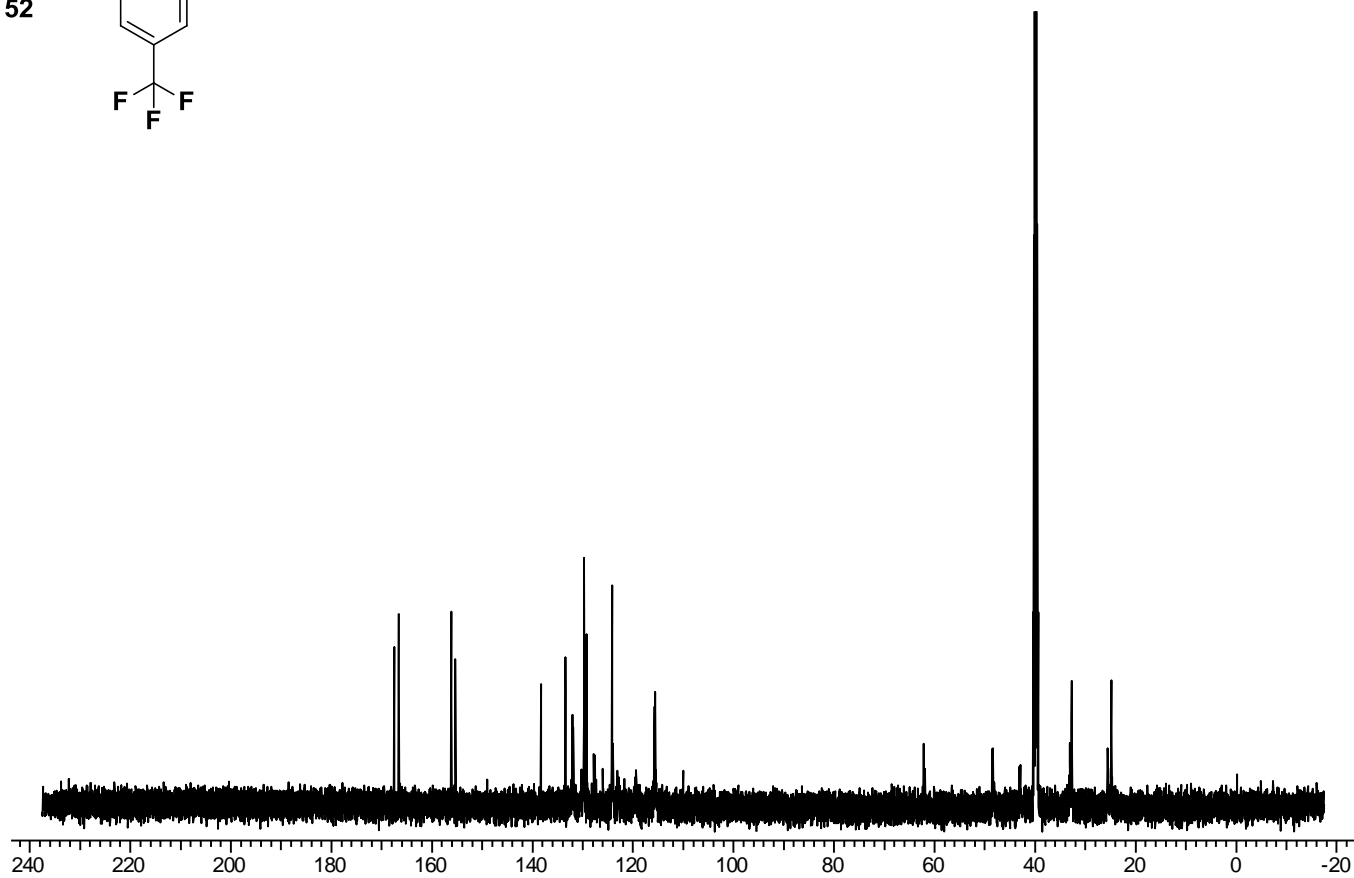
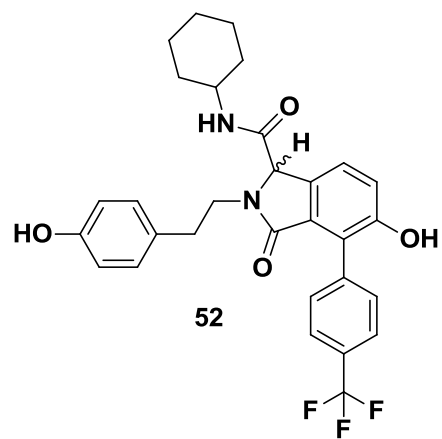
44:  $^1\text{H}$  NMR



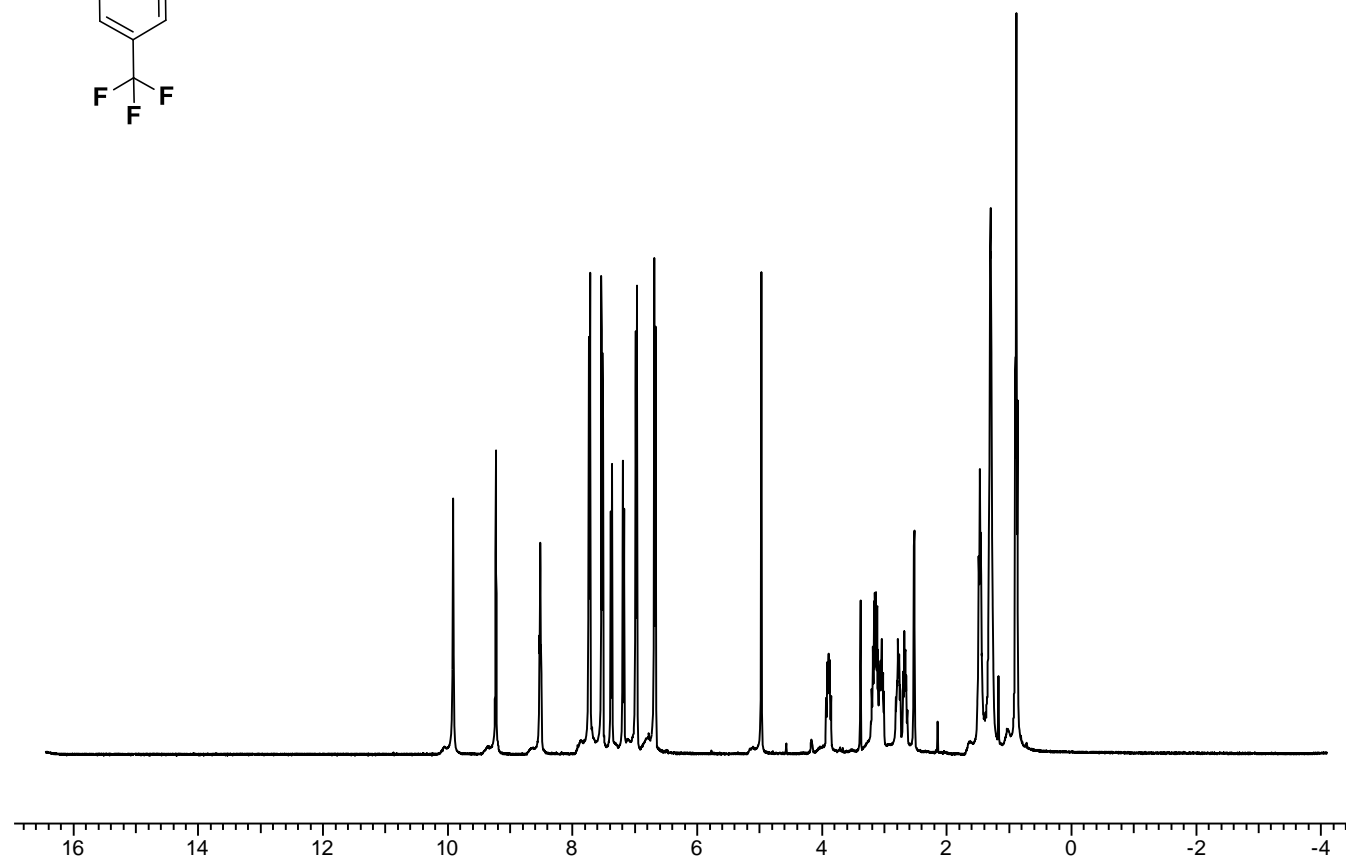
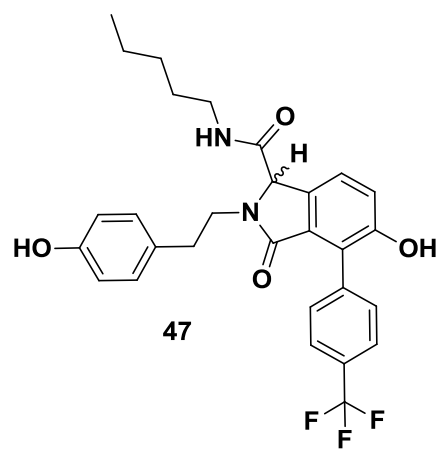
44:  $^{13}\text{C}$  NMR



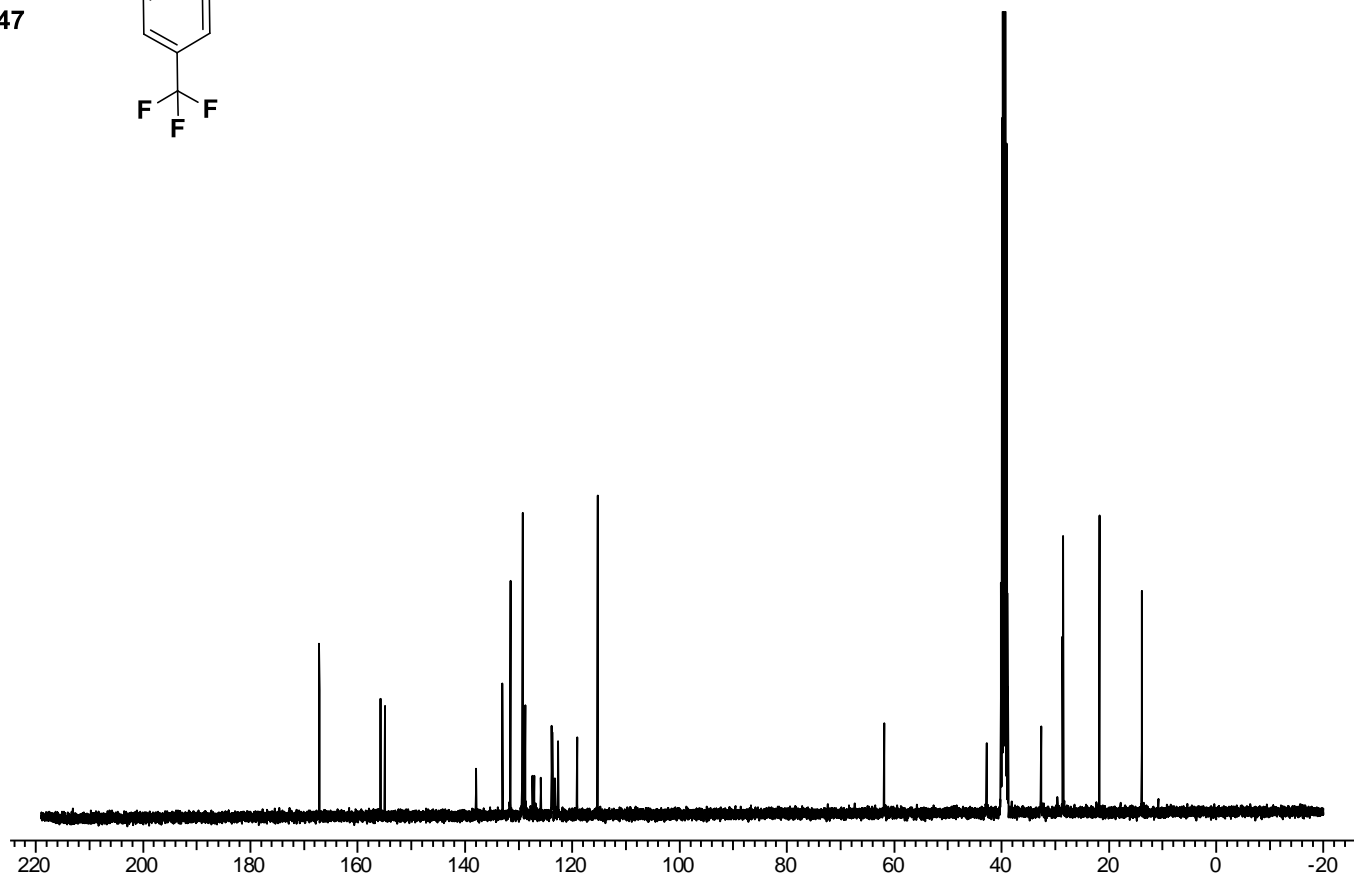
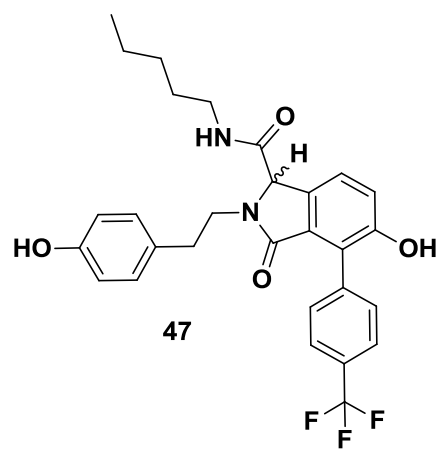
52: <sup>1</sup>H NMR



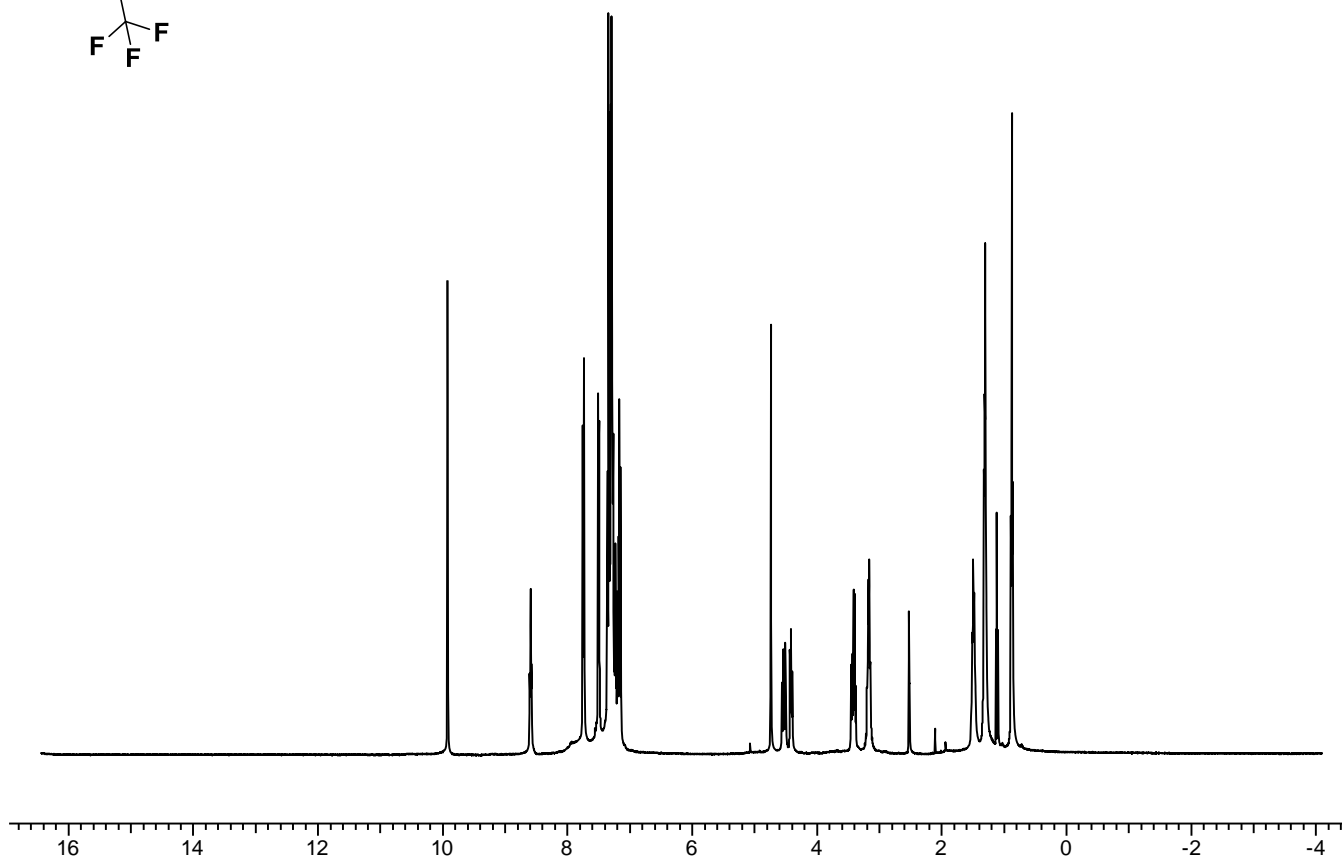
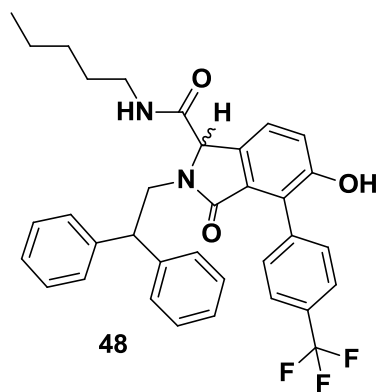
52: <sup>13</sup>C NMR



47: <sup>1</sup>H NMR

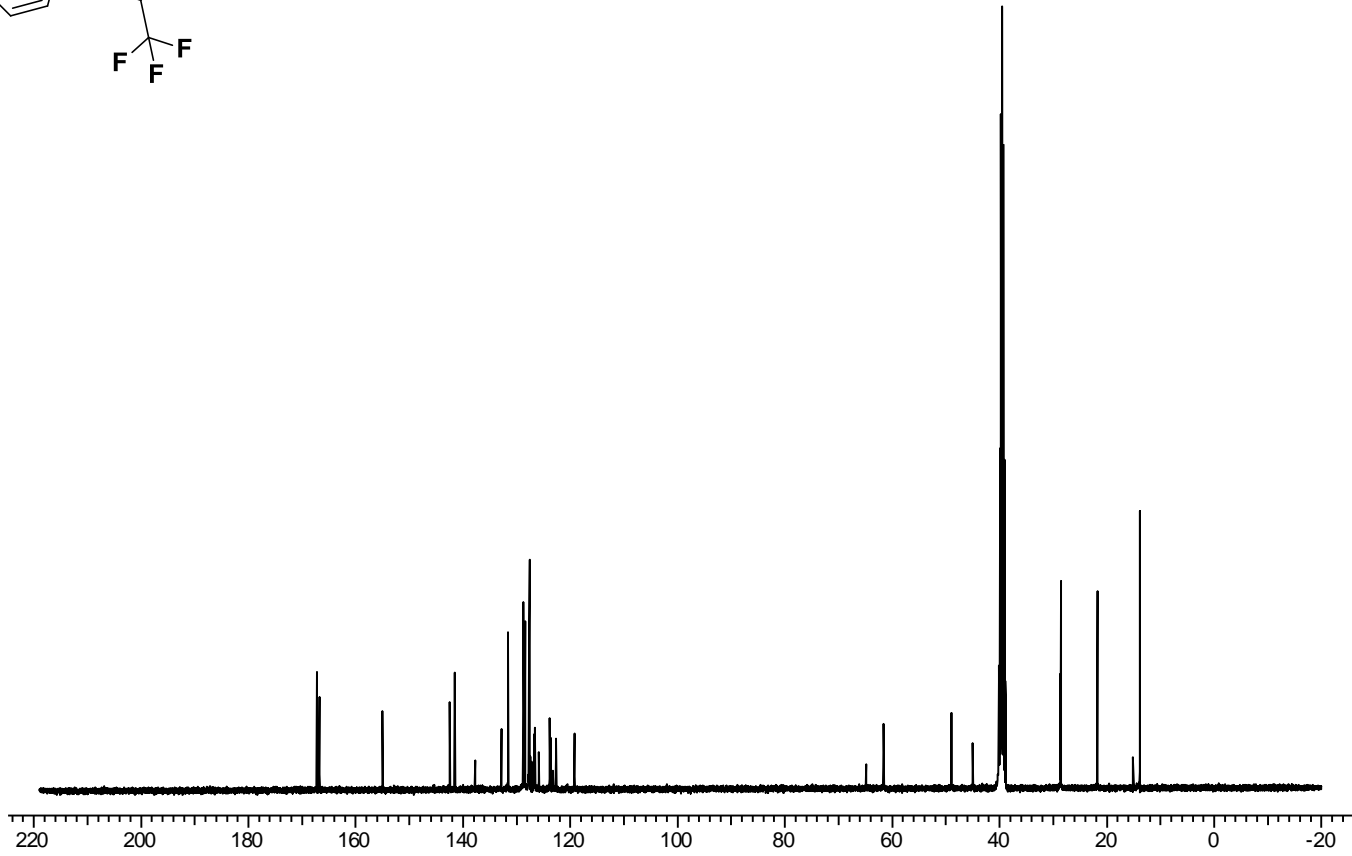
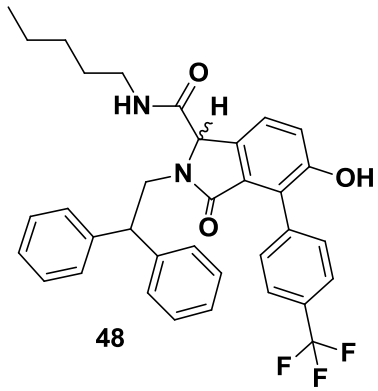


47: <sup>13</sup>C NMR

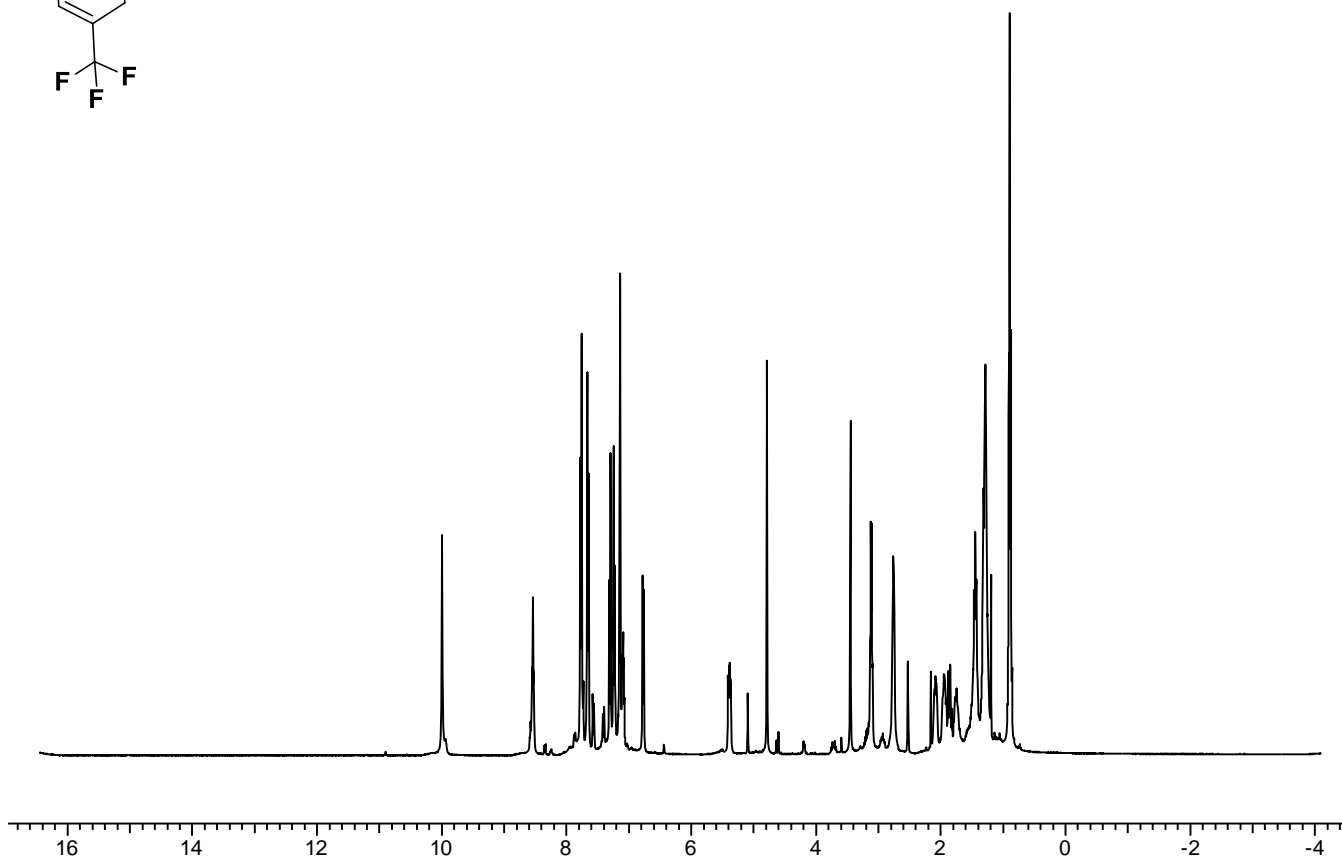
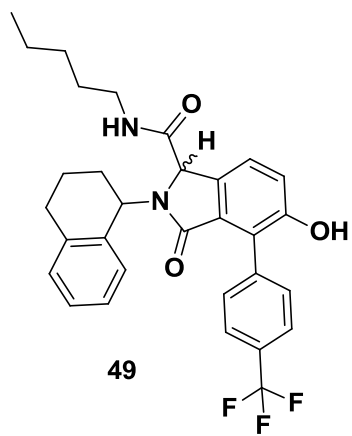


48:  $^1\text{H}$  NMR

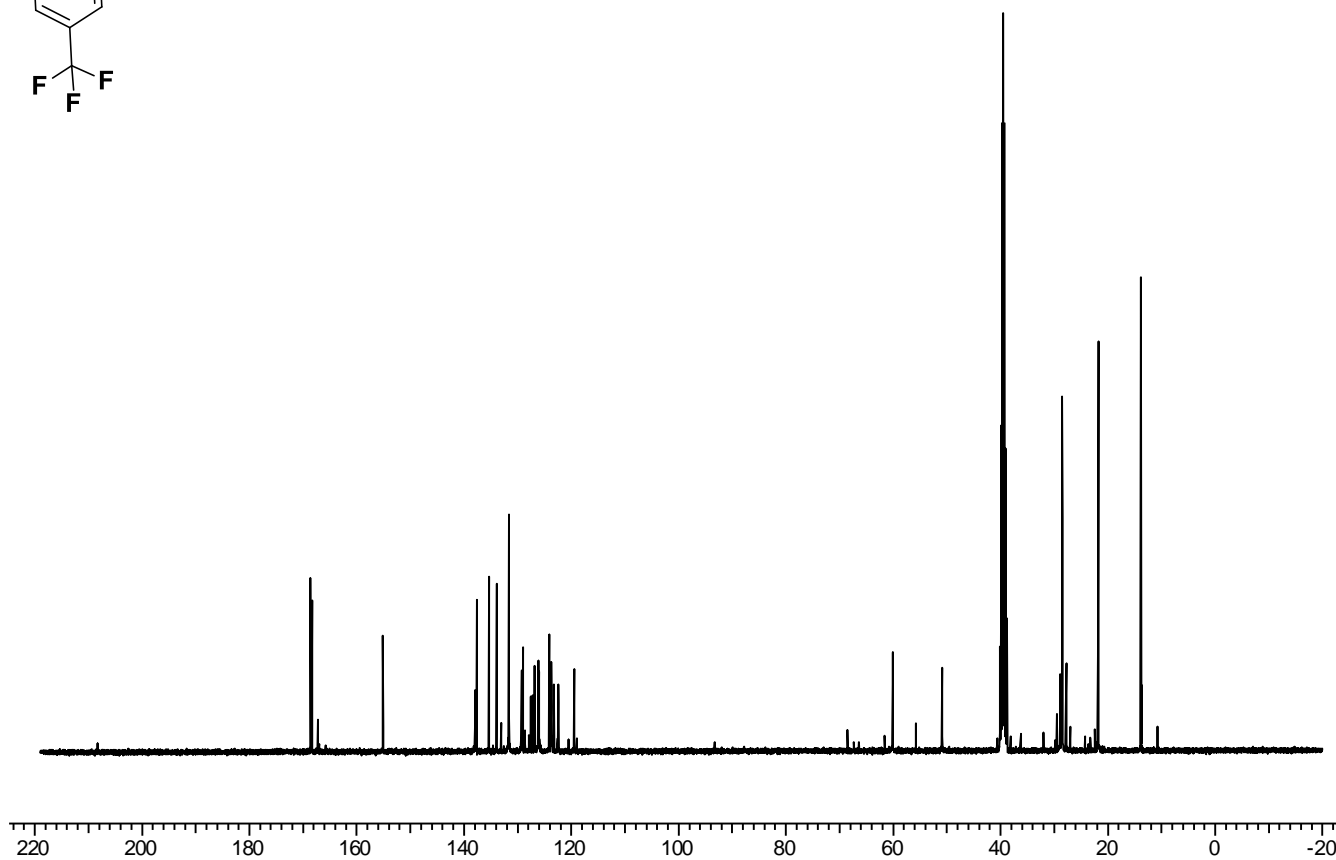
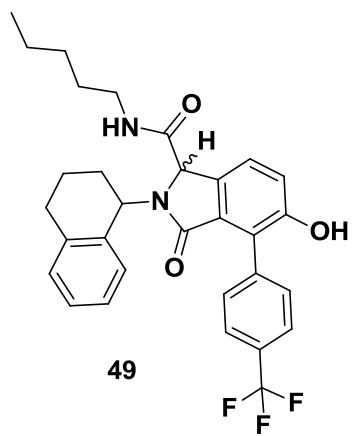




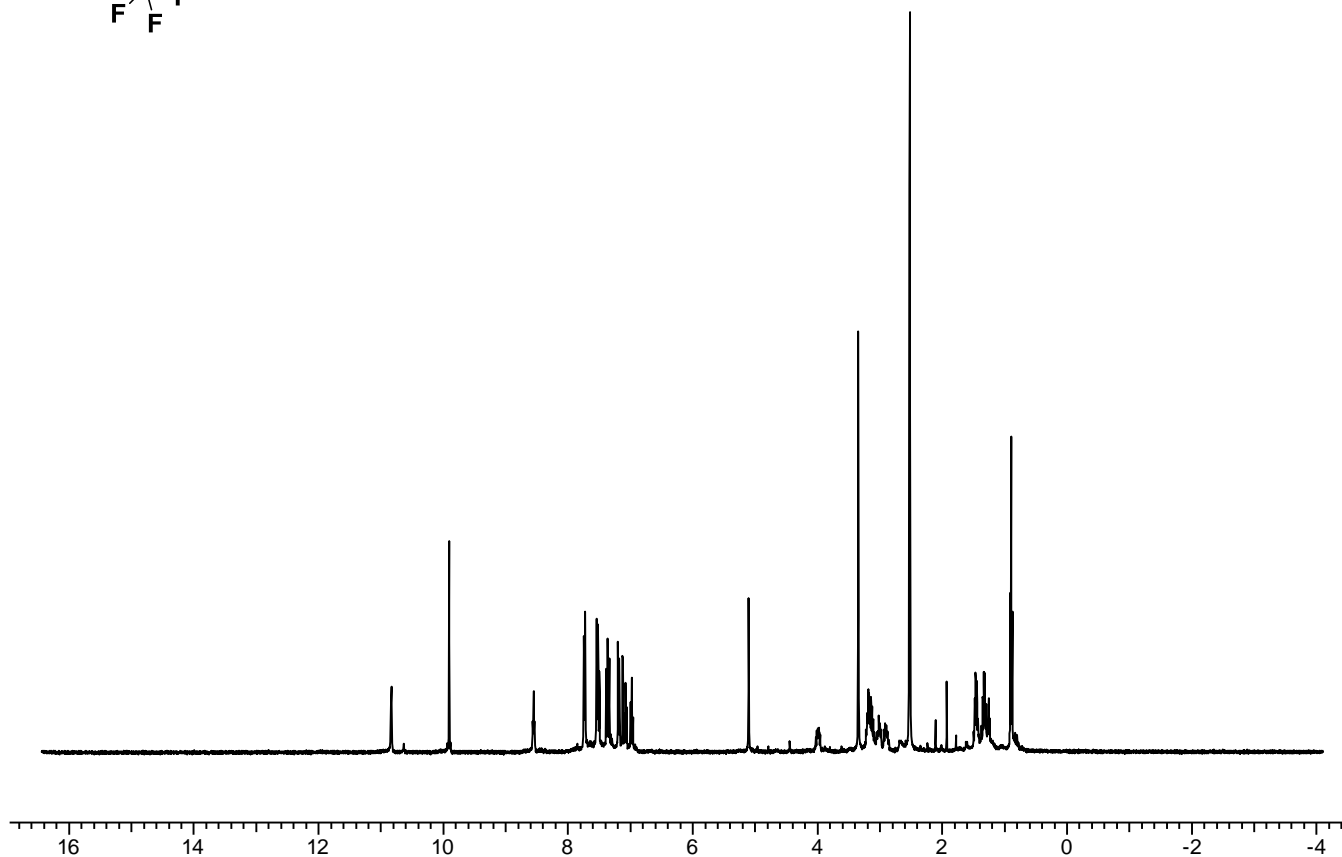
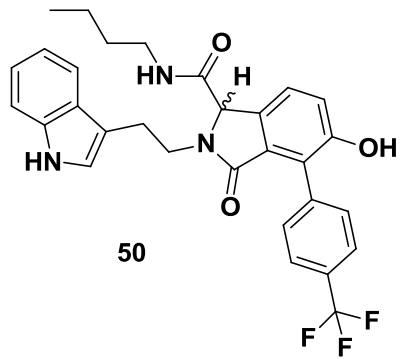
48:  $^{13}\text{C}$  NMR



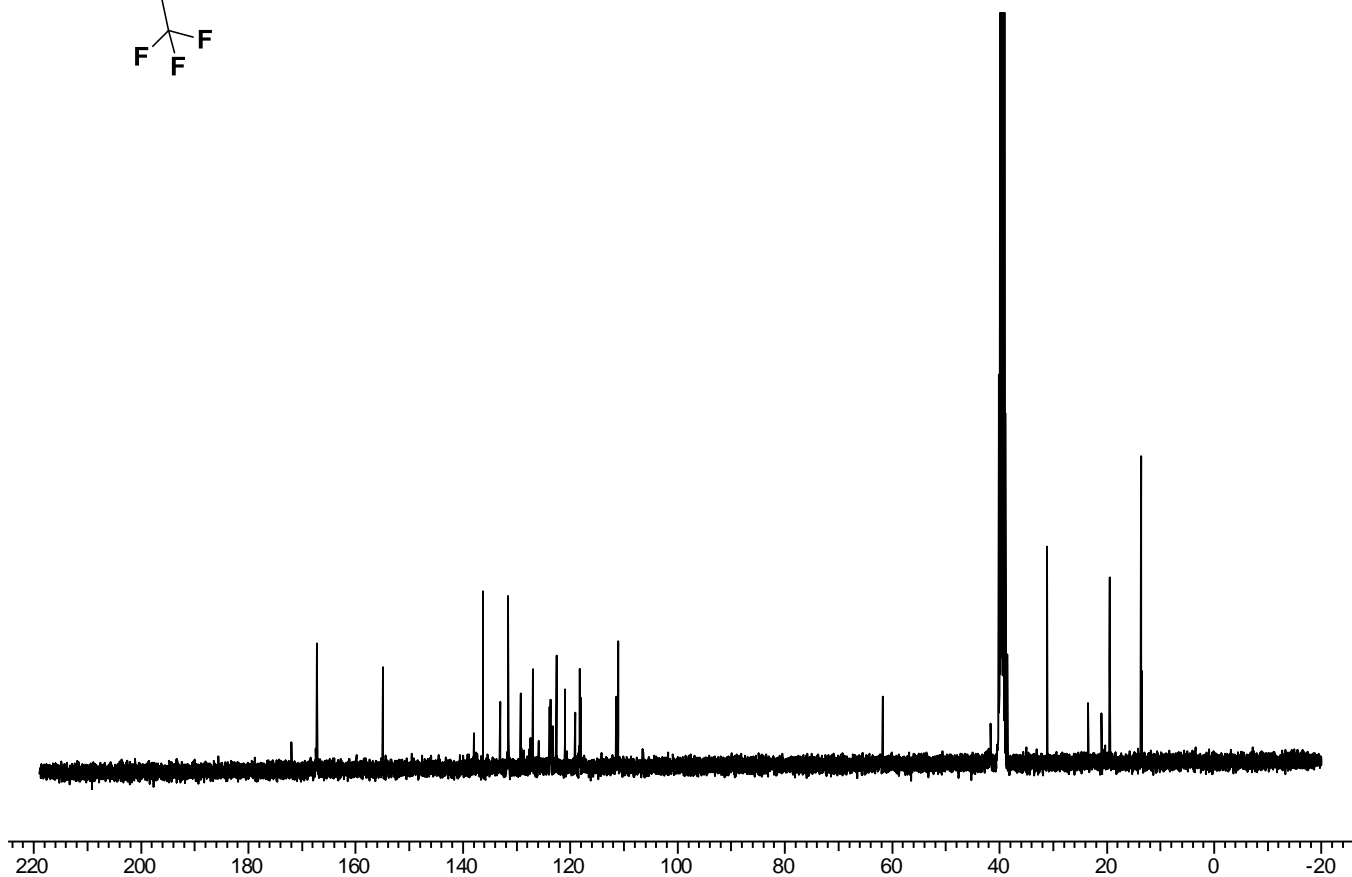
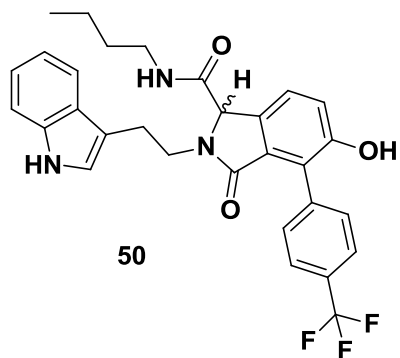
49:  $^1\text{H}$  NMR



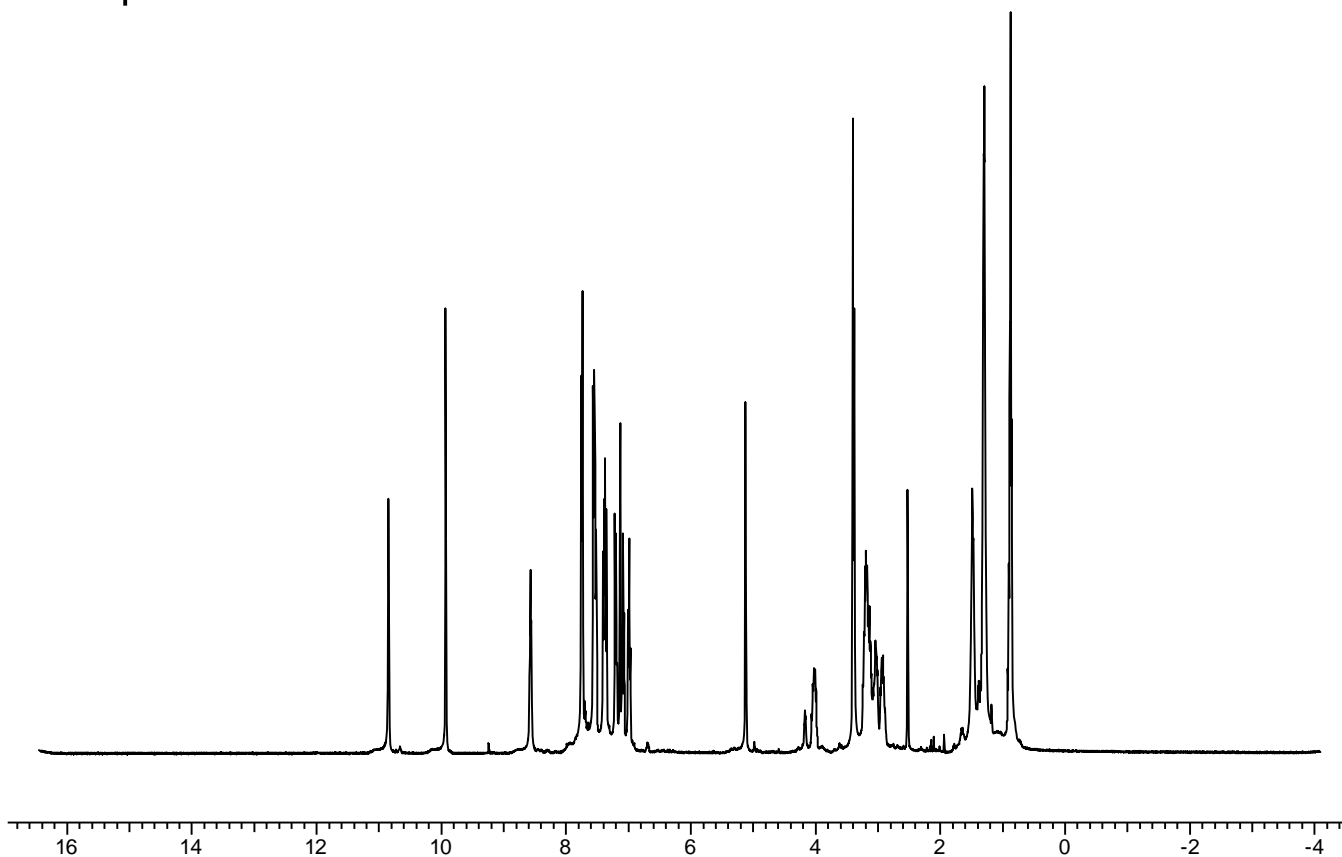
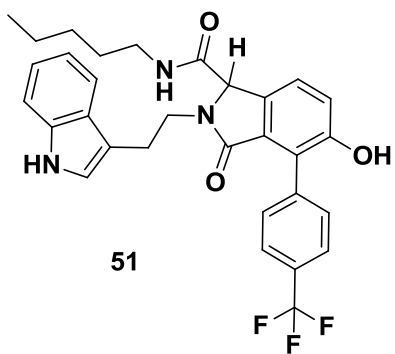
49:  $^{13}\text{C}$  NMR



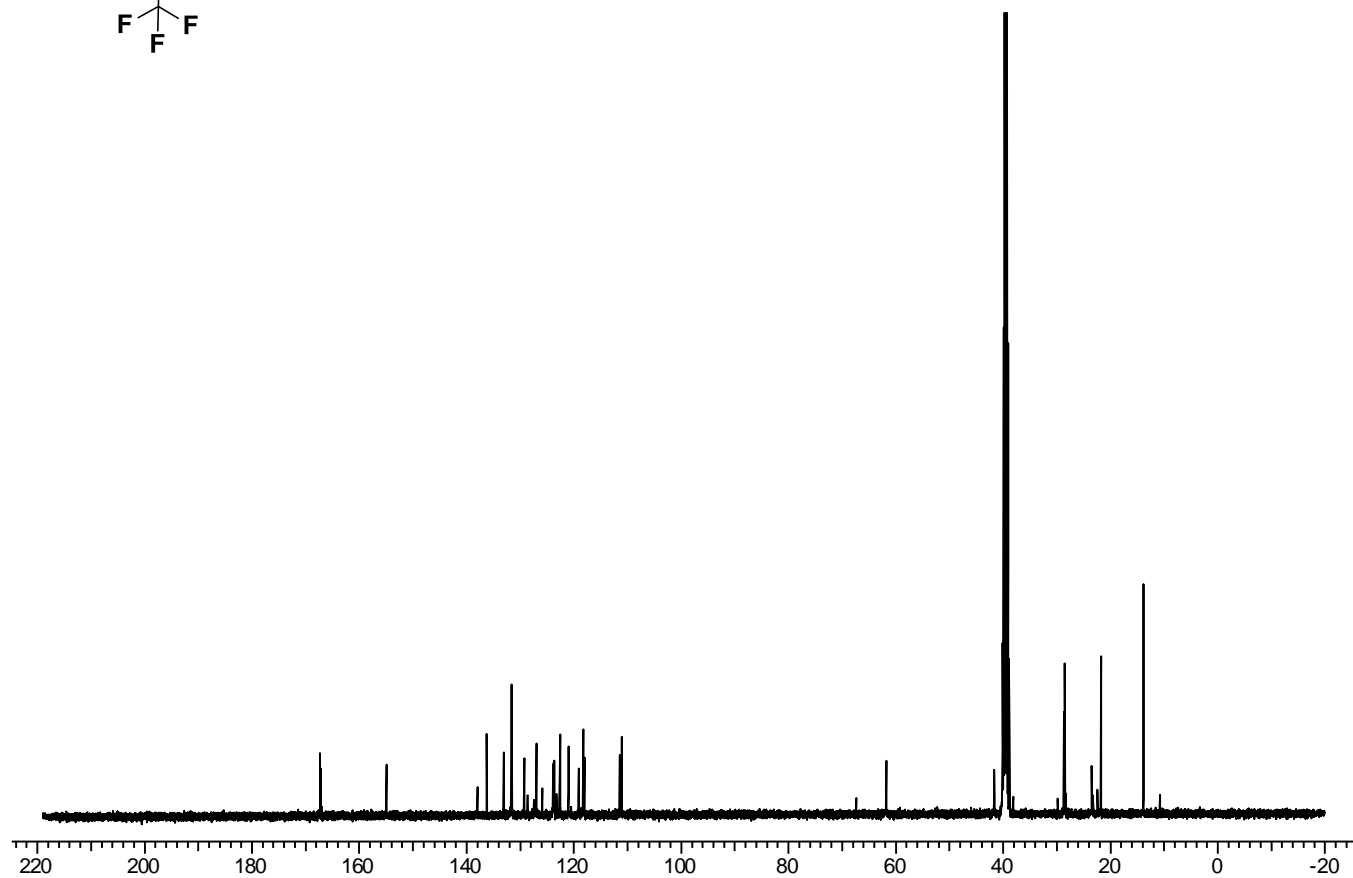
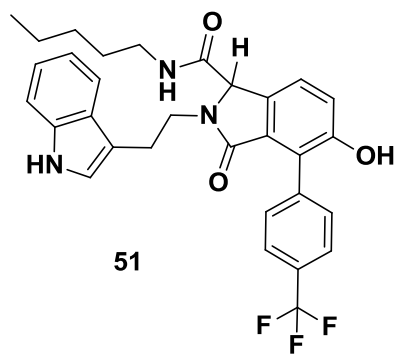
50:  $^1\text{H}$  NMR



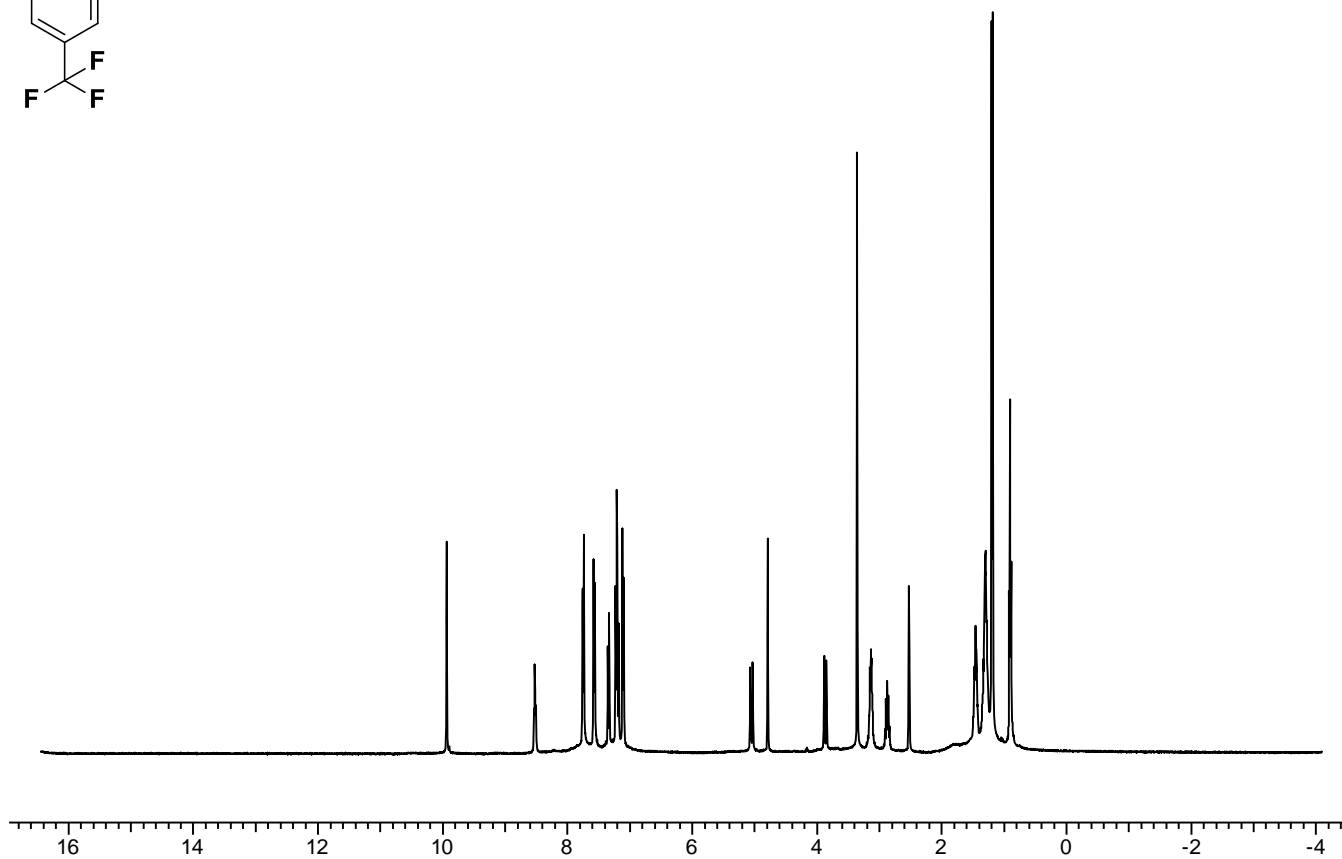
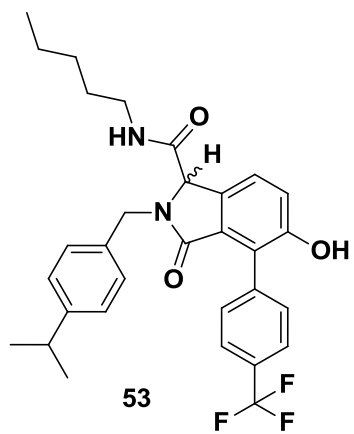
50:  $^{13}\text{C}$  NMR



51:  $^1\text{H}$  NMR

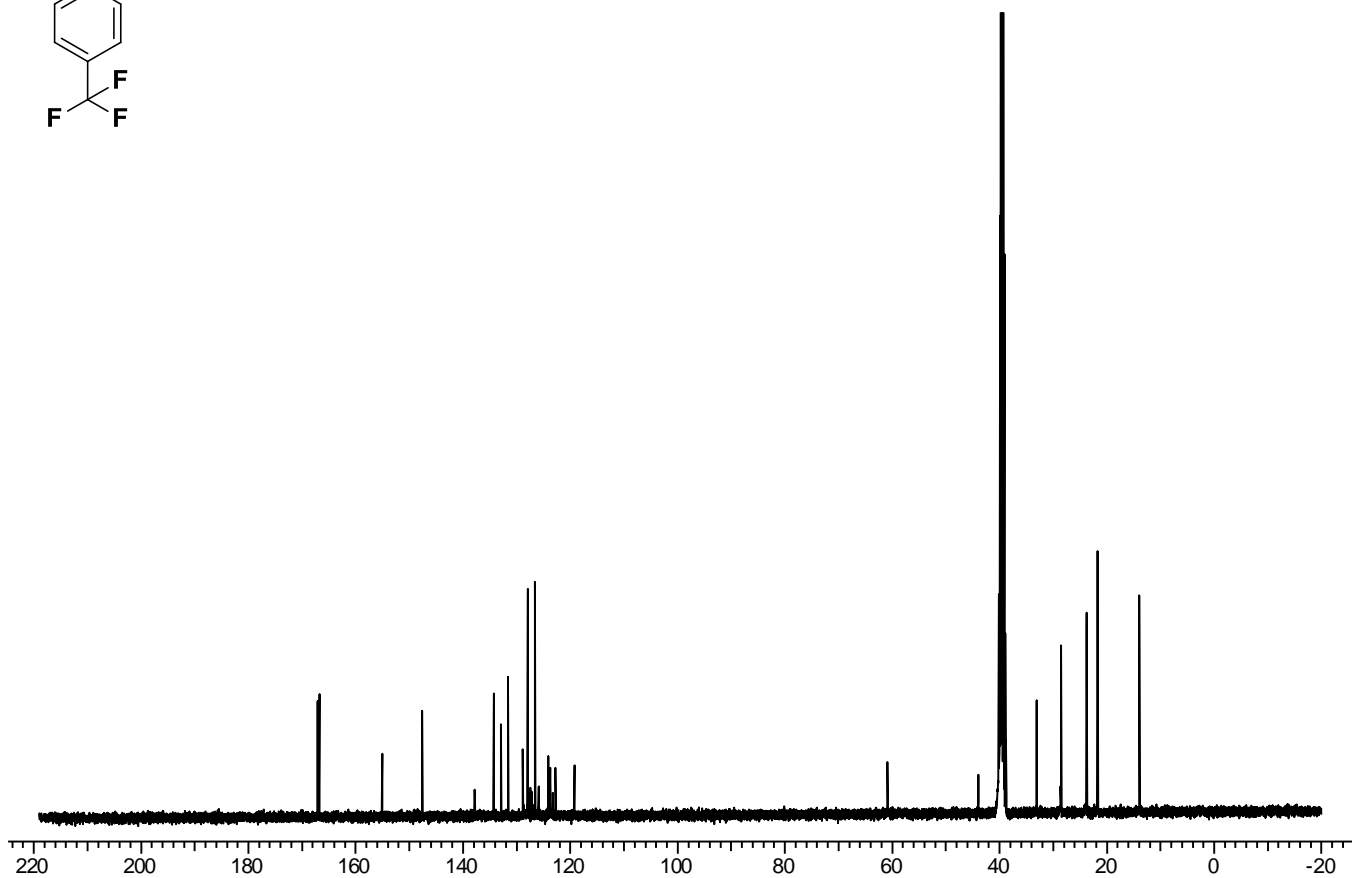
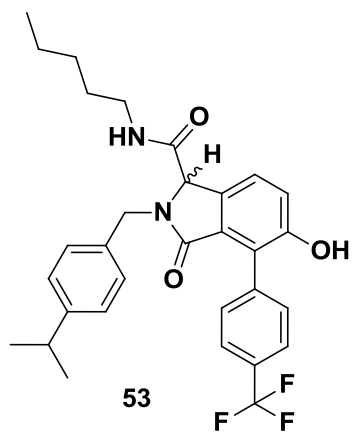


51:  $^{13}\text{C}$  NMR

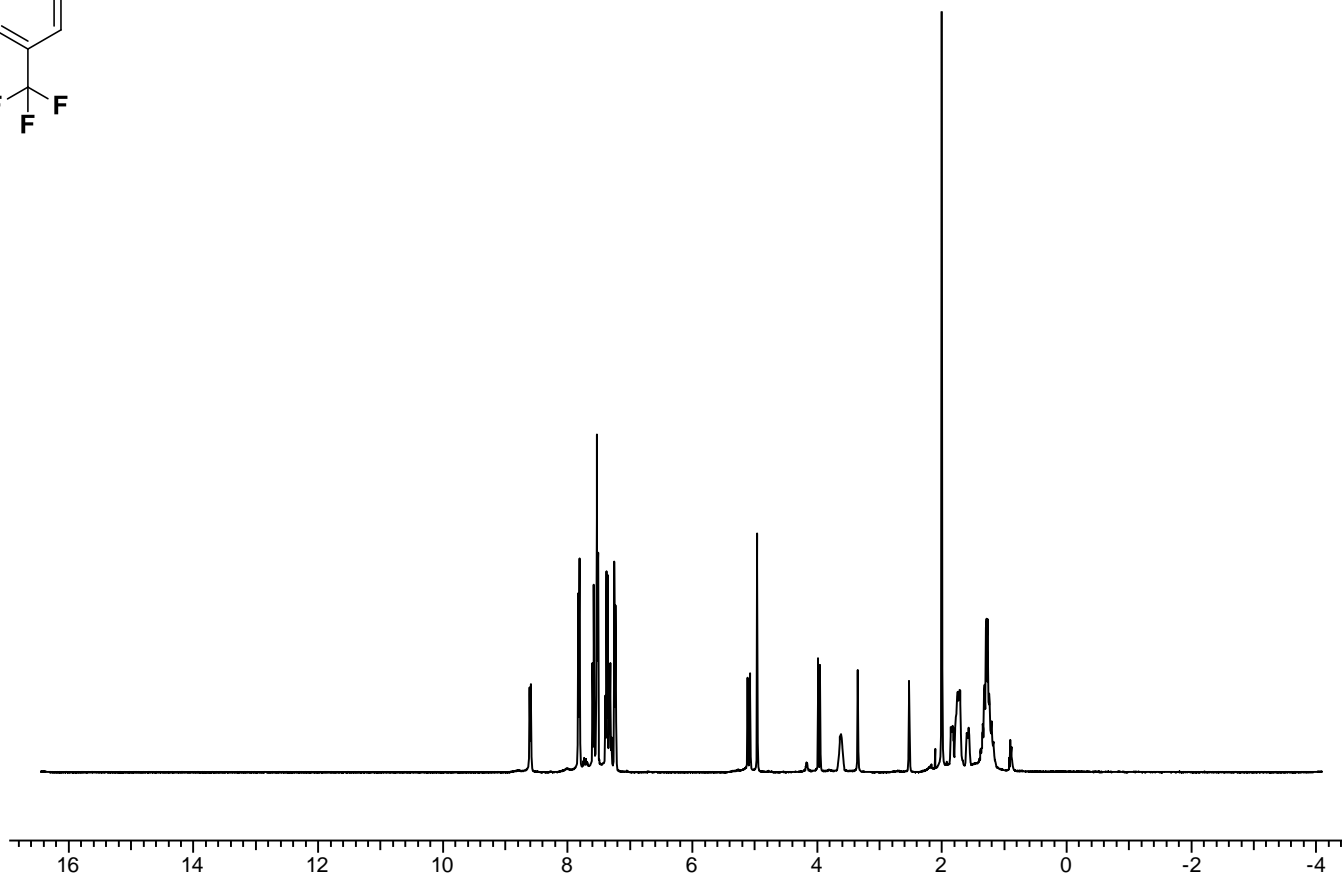
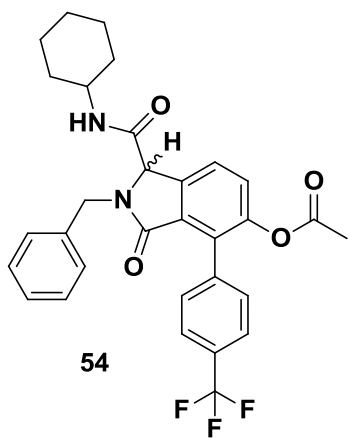


51:  $^1\text{H}$  NMR

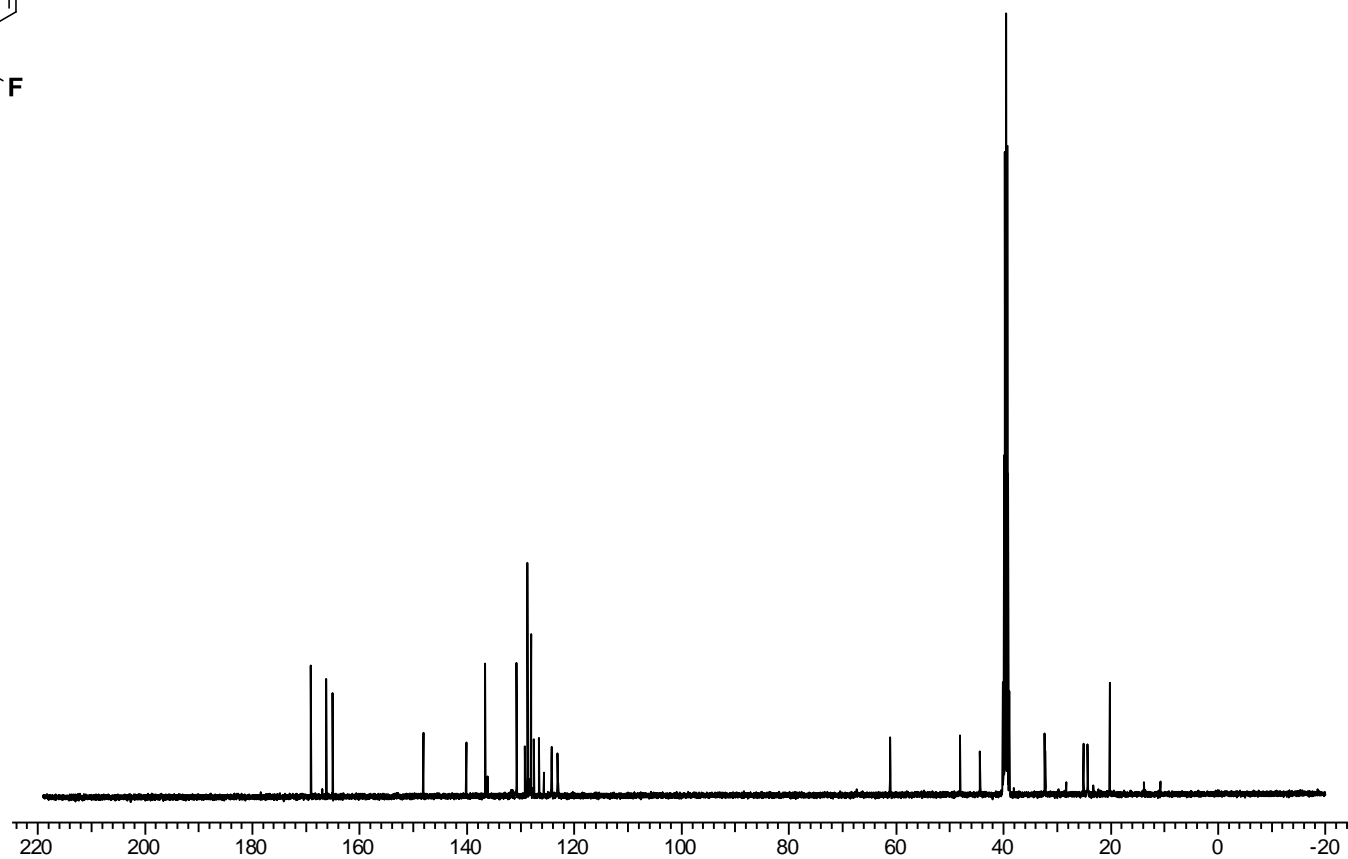
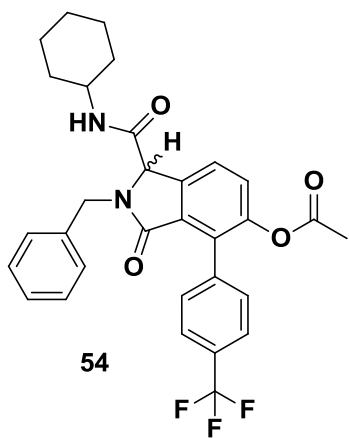




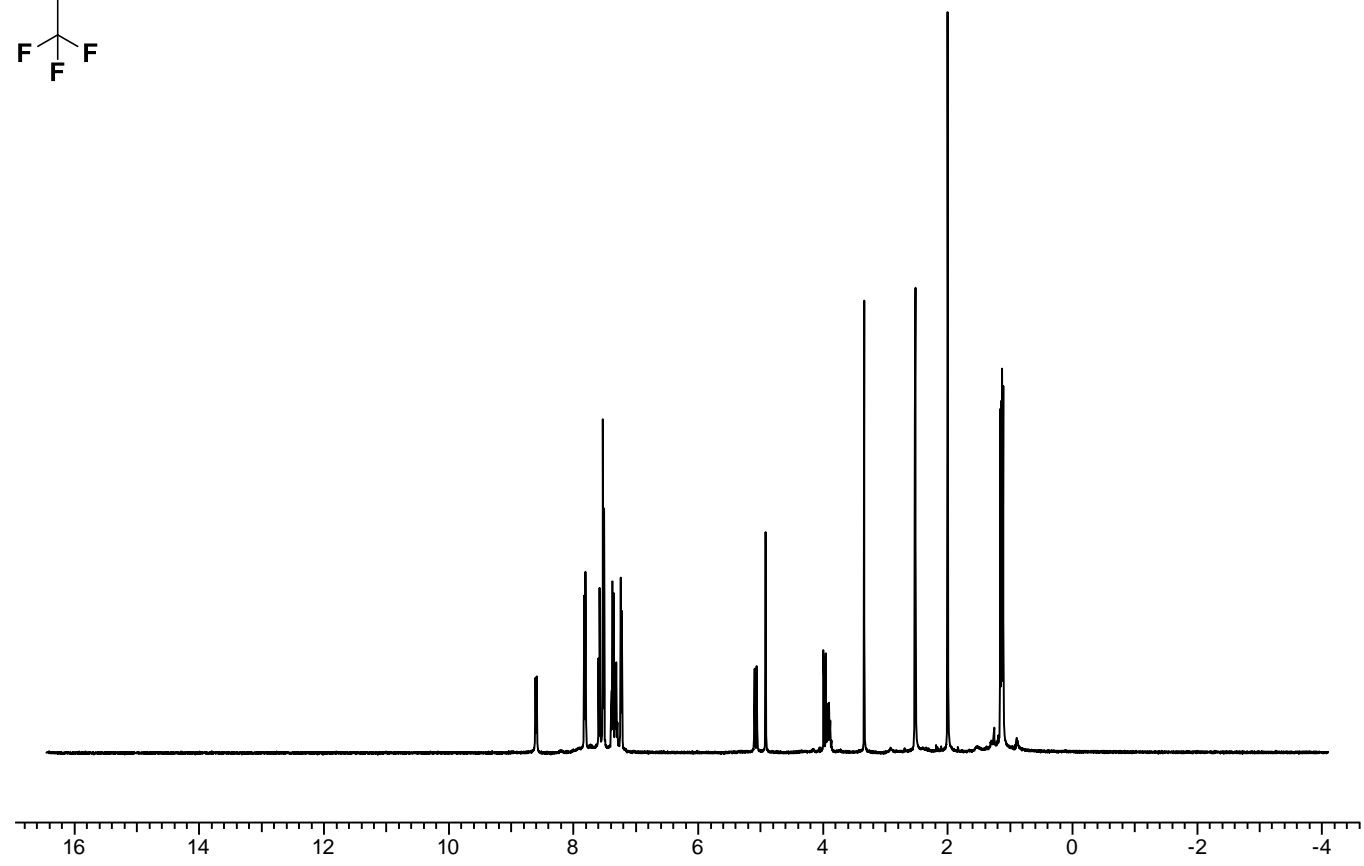
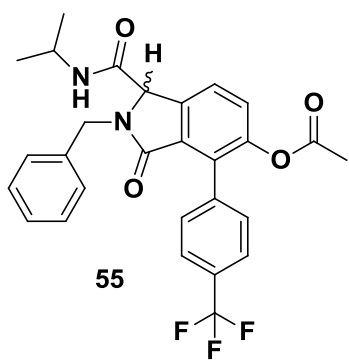
51: <sup>13</sup>C NMR



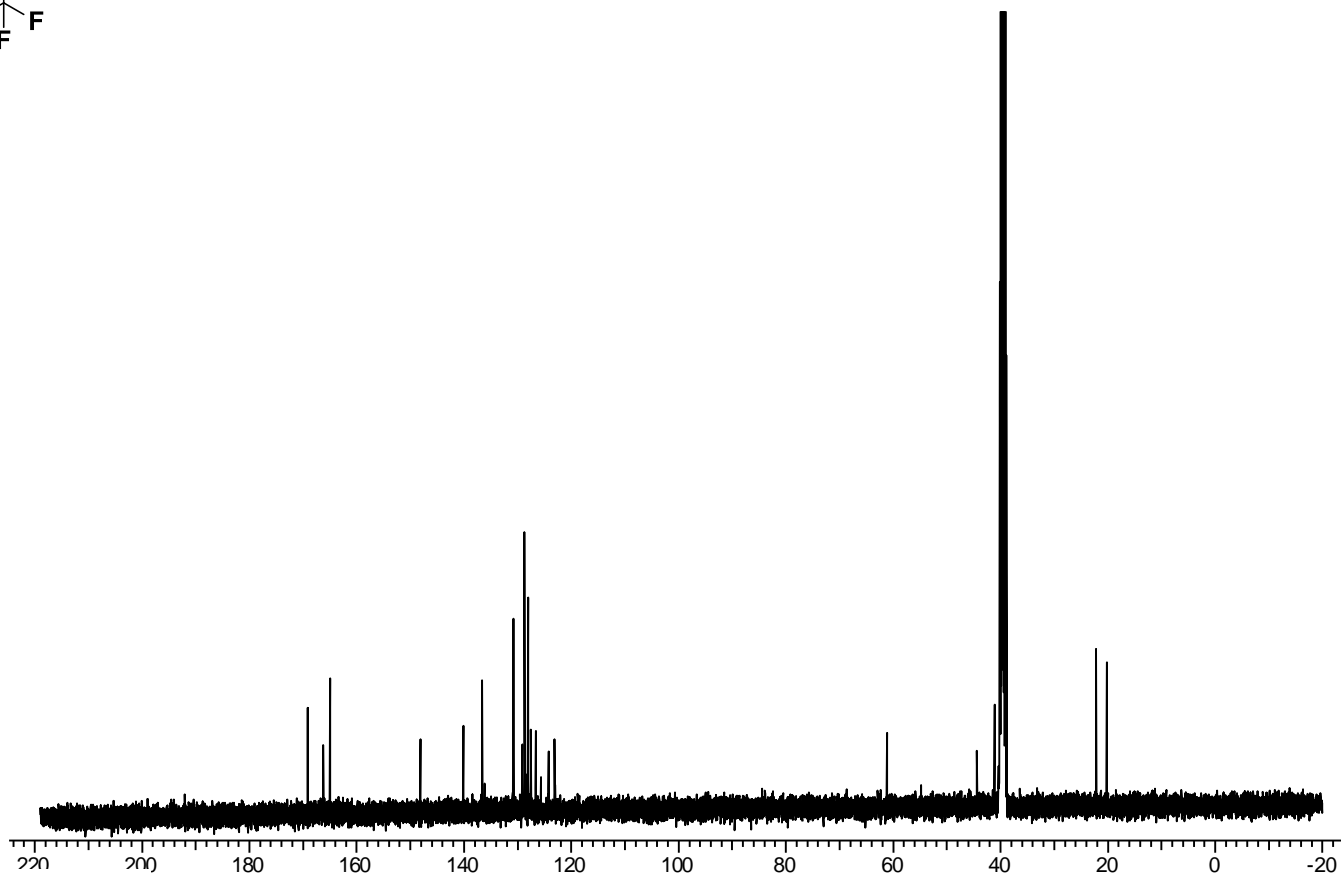
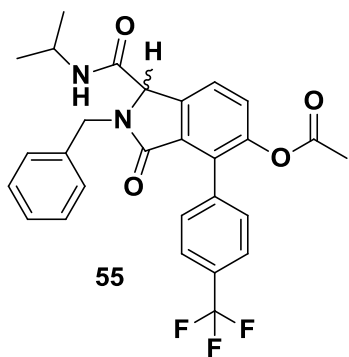
54:  $^1\text{H}$  NMR



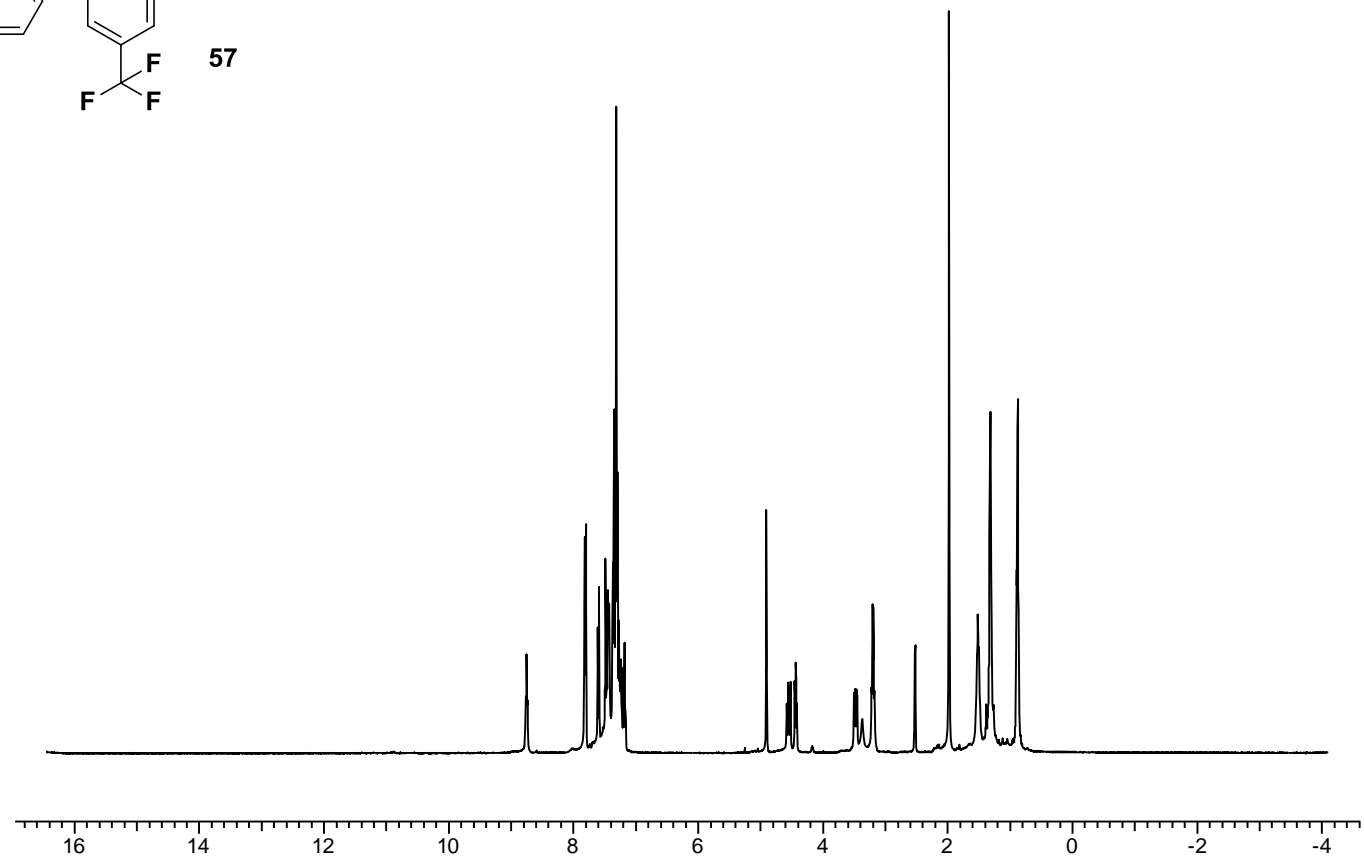
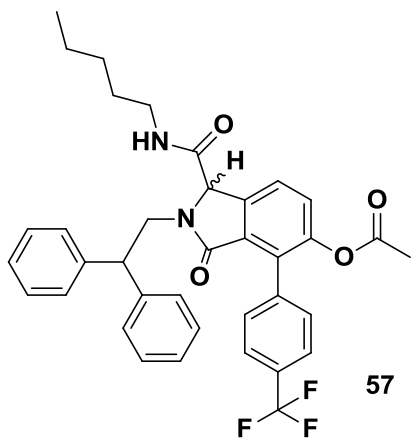
54:  $^{13}\text{C}$  NMR



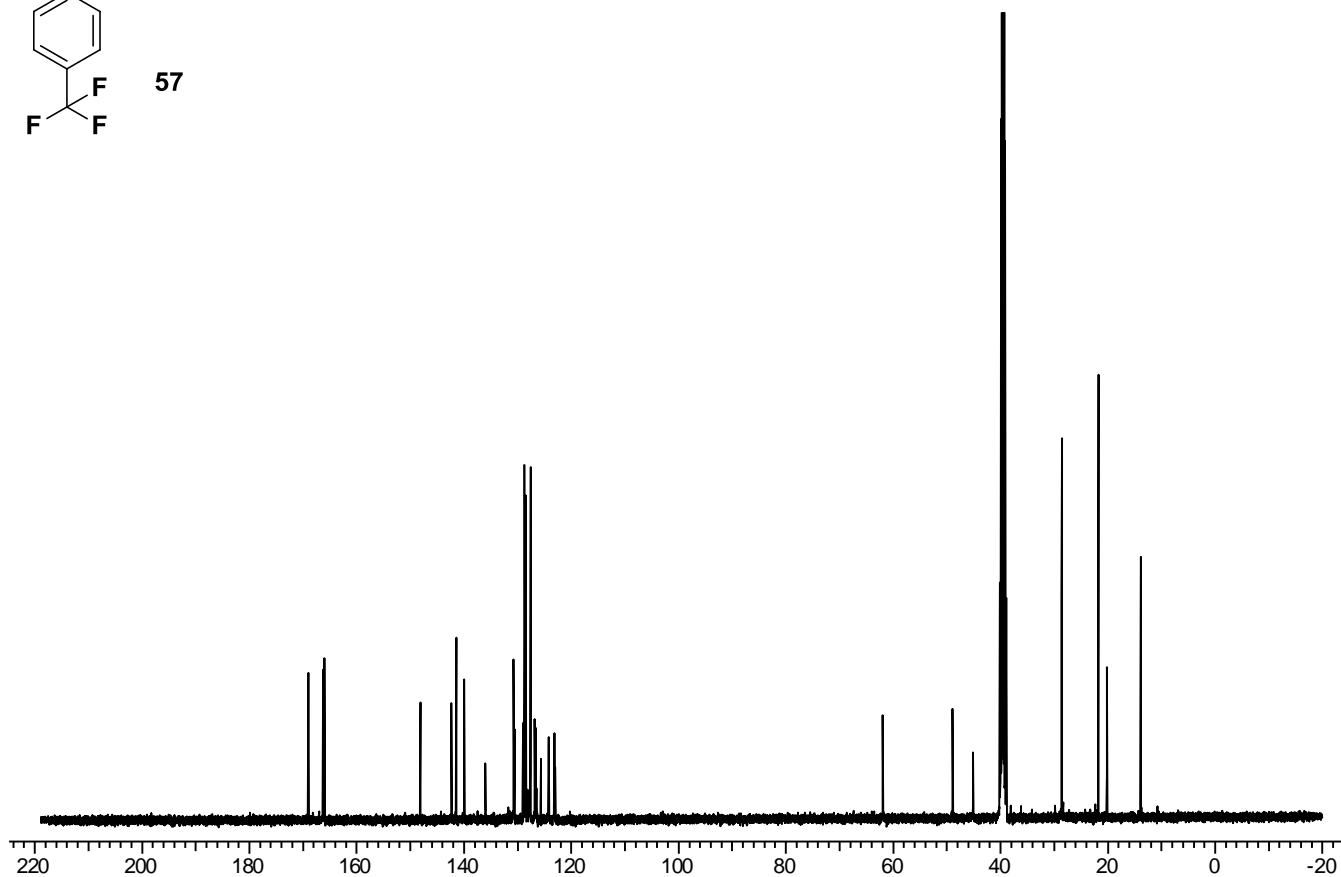
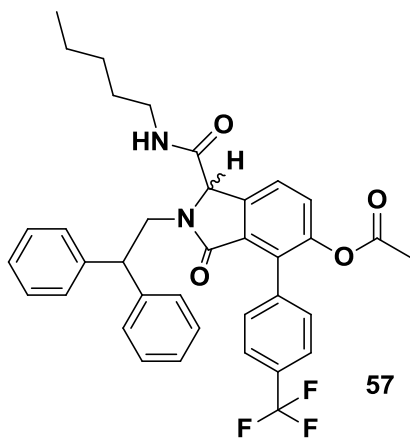
55: <sup>1</sup>H NMR



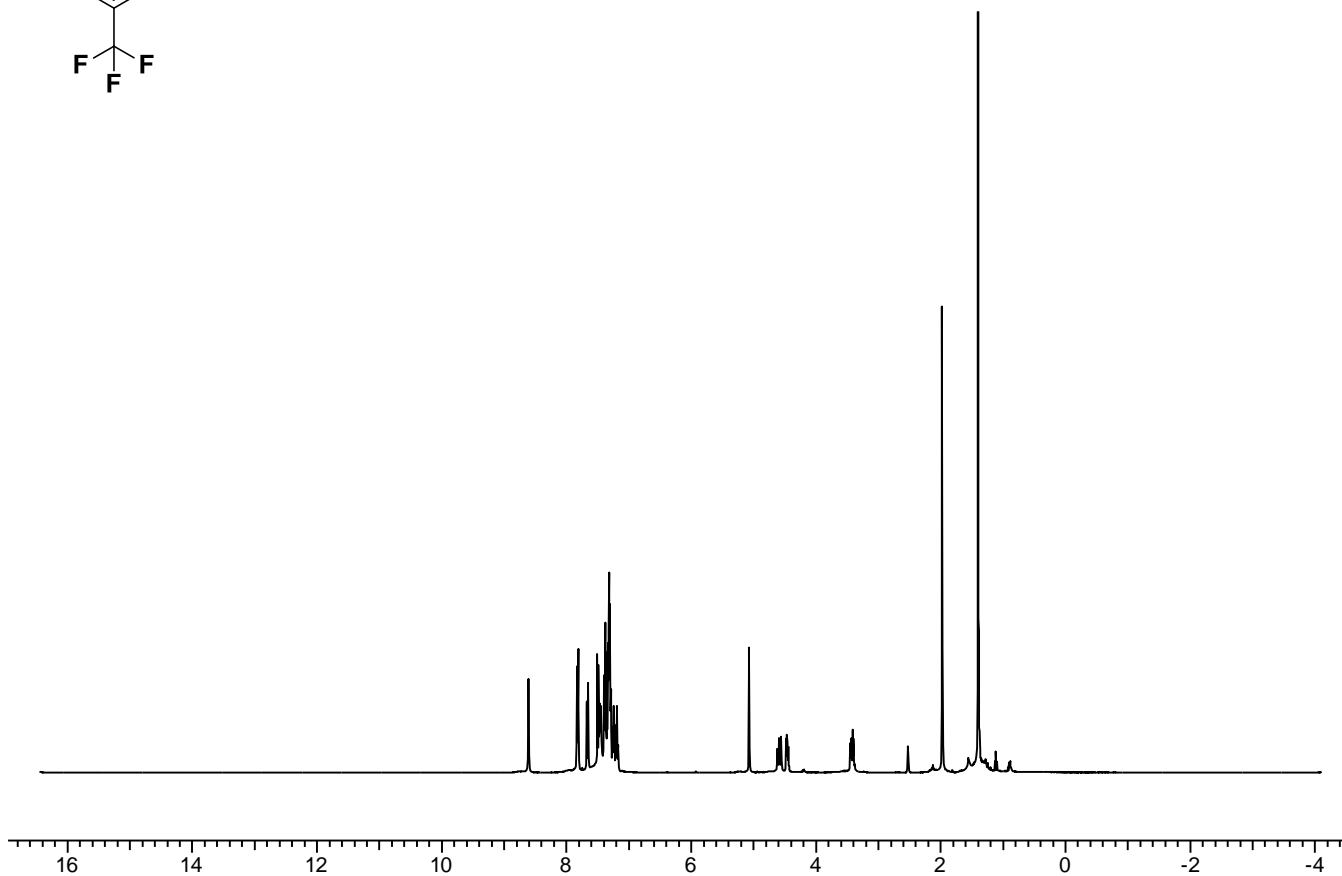
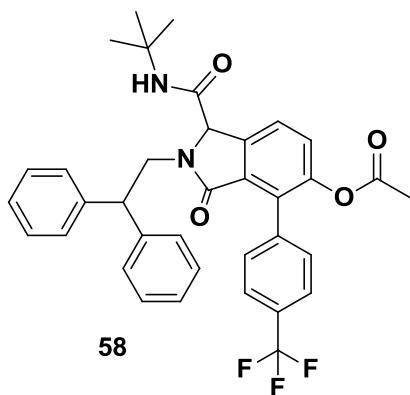
55: <sup>13</sup>C NMR



57: <sup>1</sup>H NMR

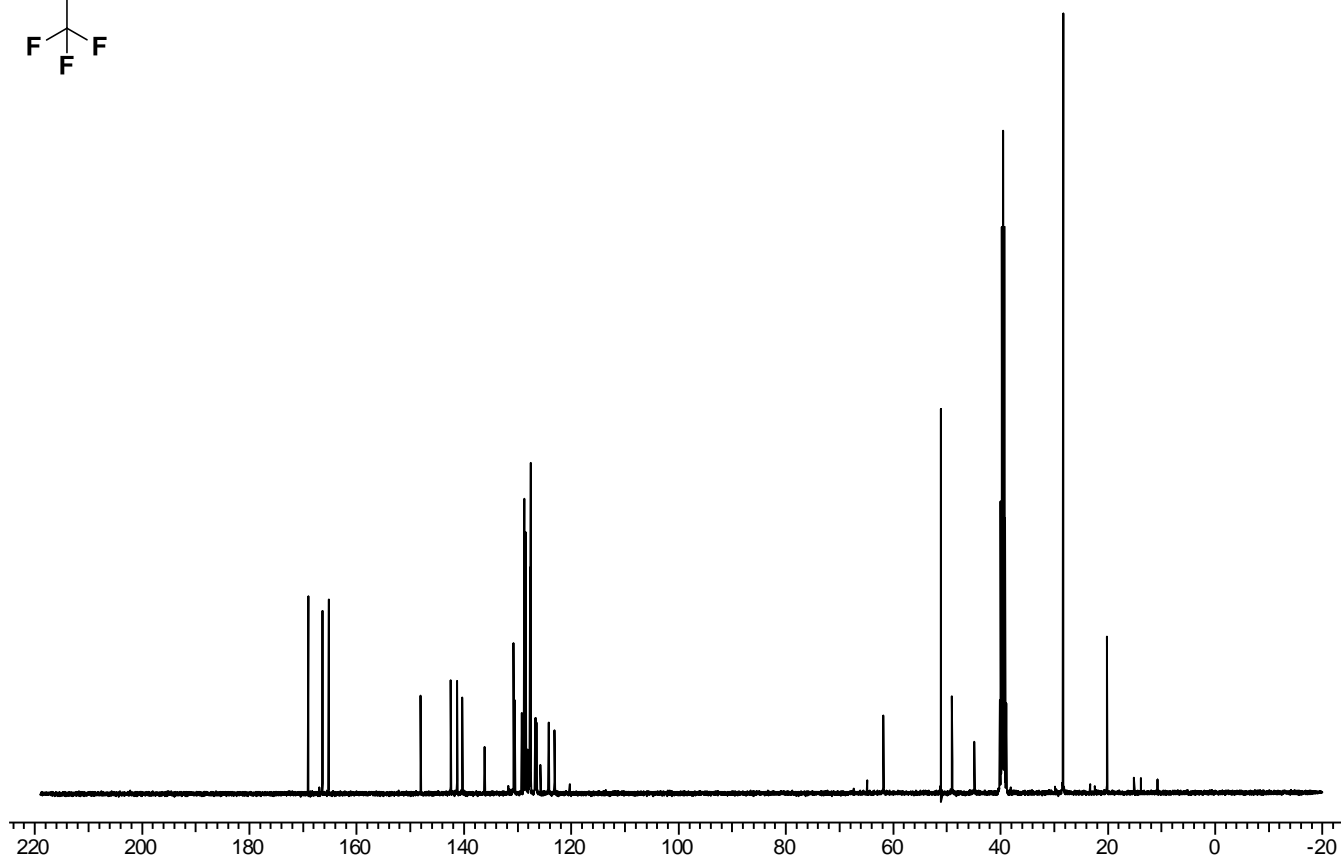
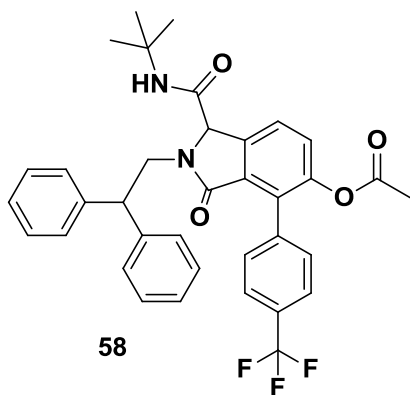


57: <sup>13</sup>C NMR

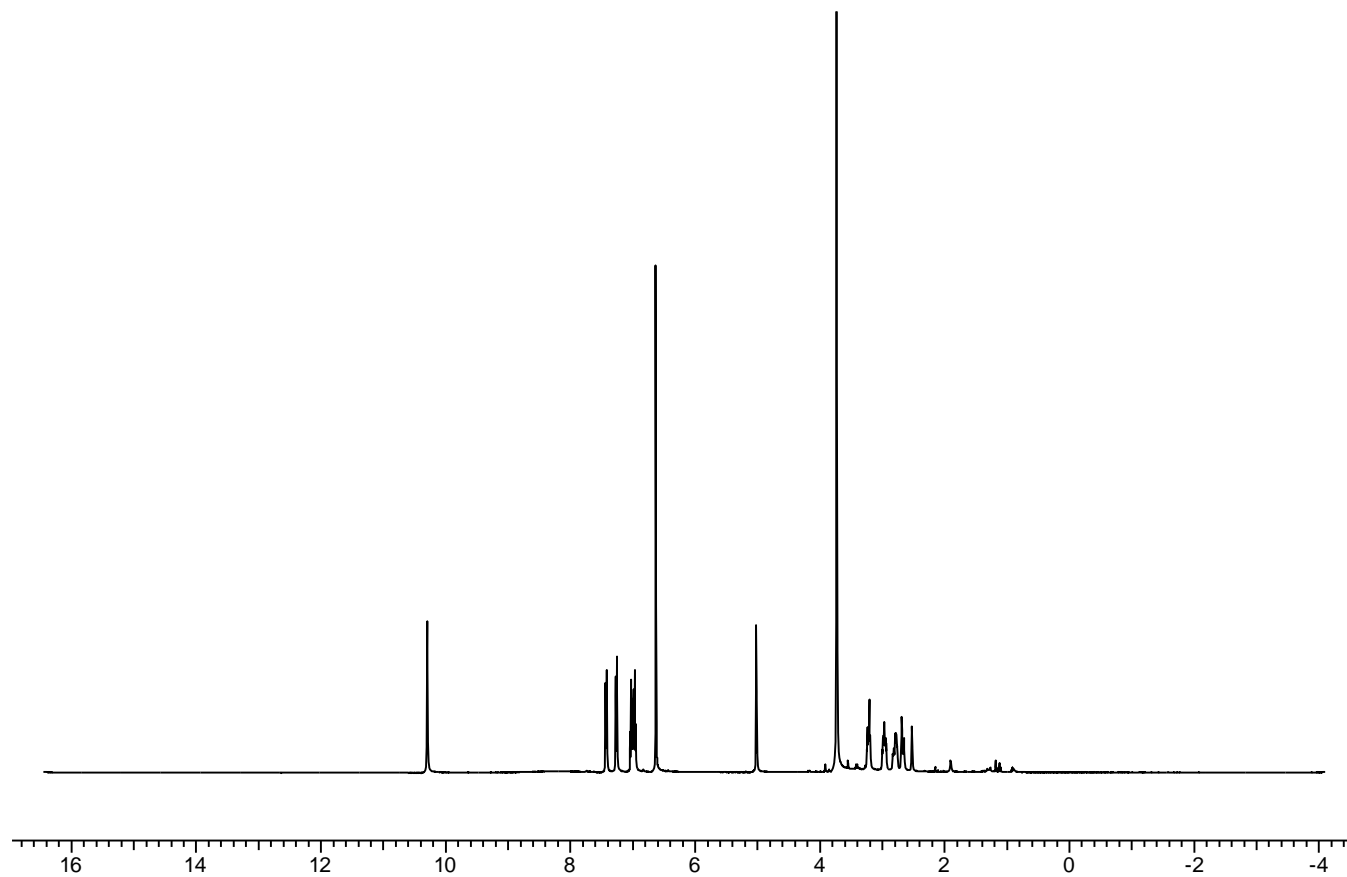
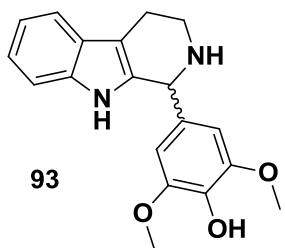


58:  $^1\text{H}$  NMR

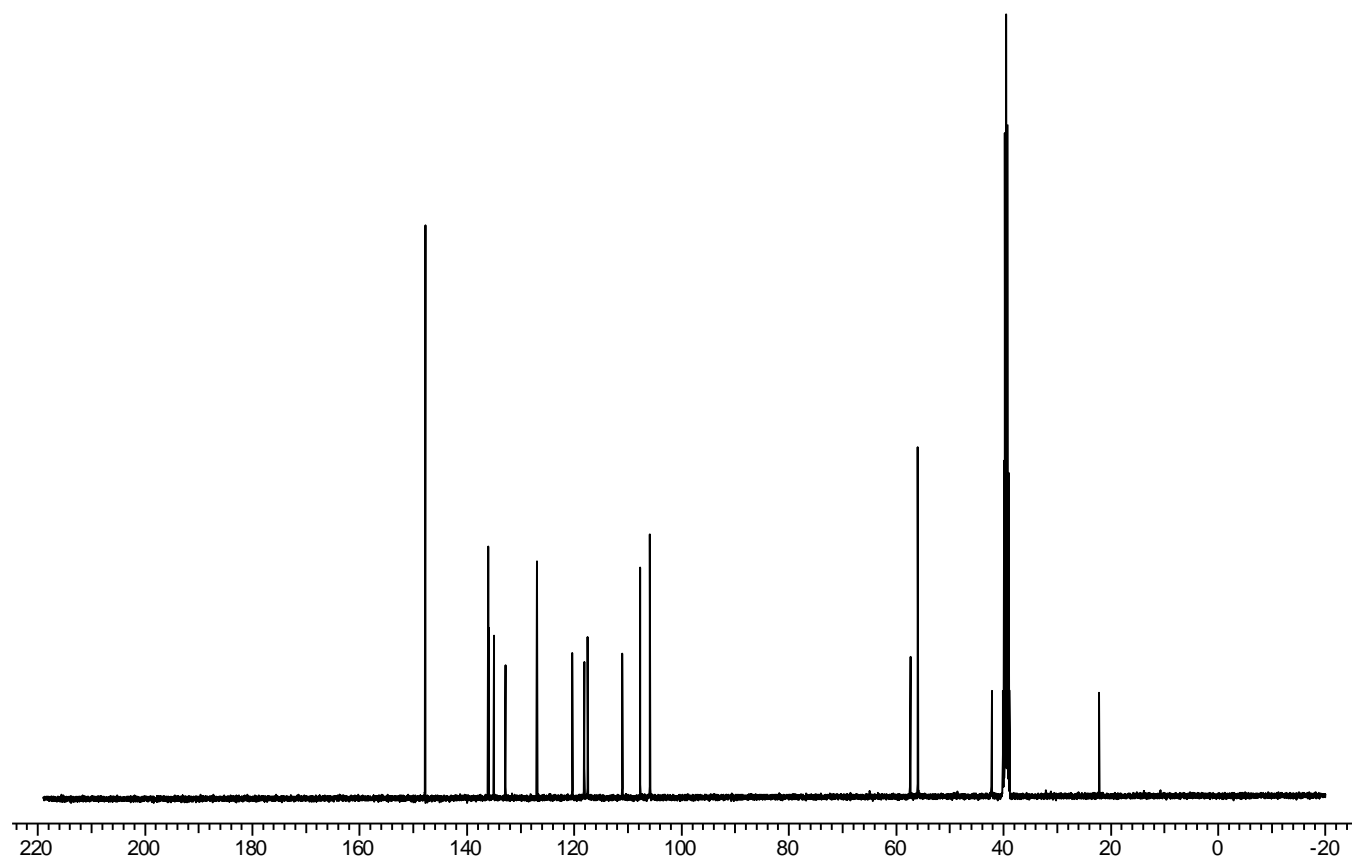
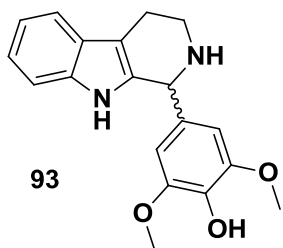




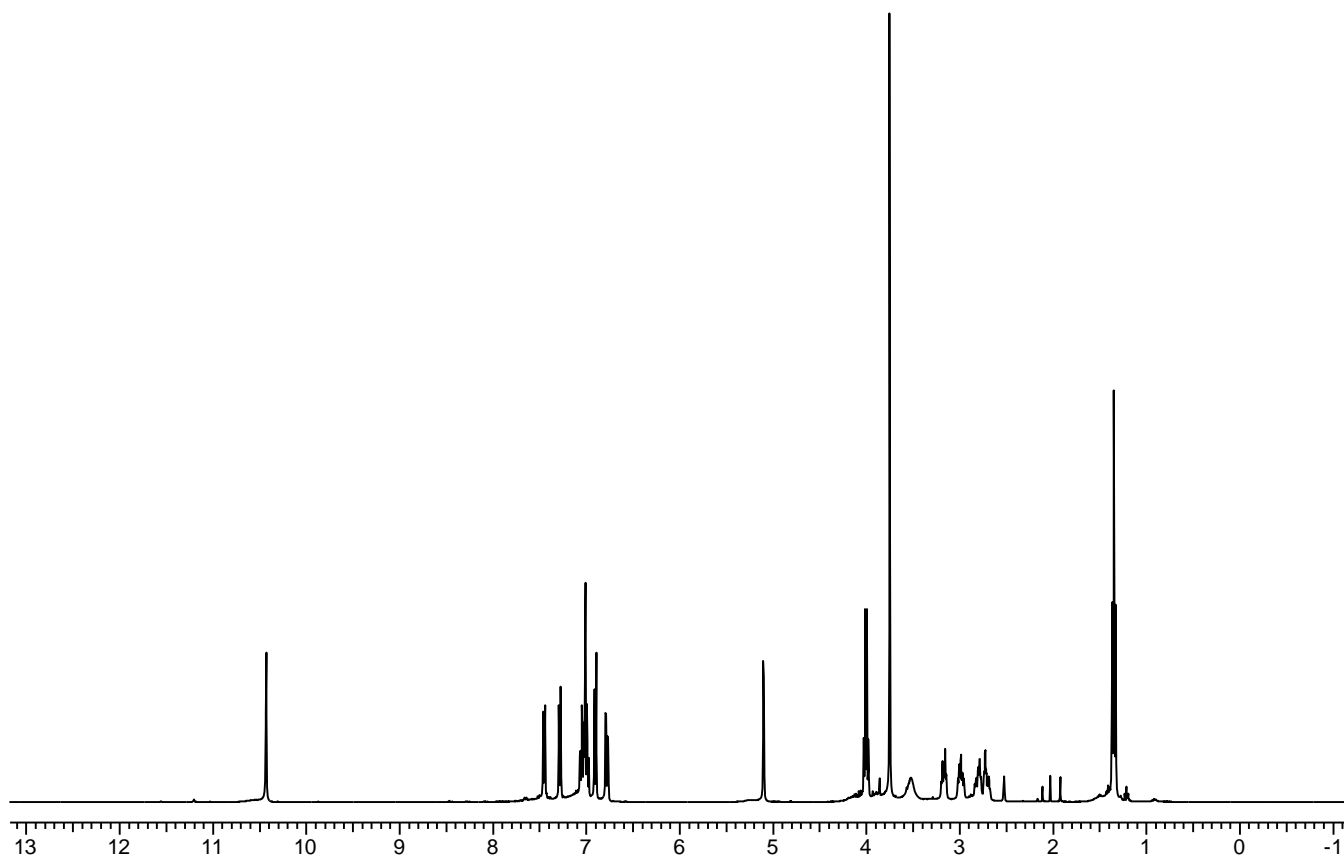
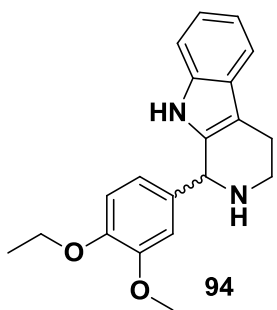
58:  $^{13}\text{C}$  NMR



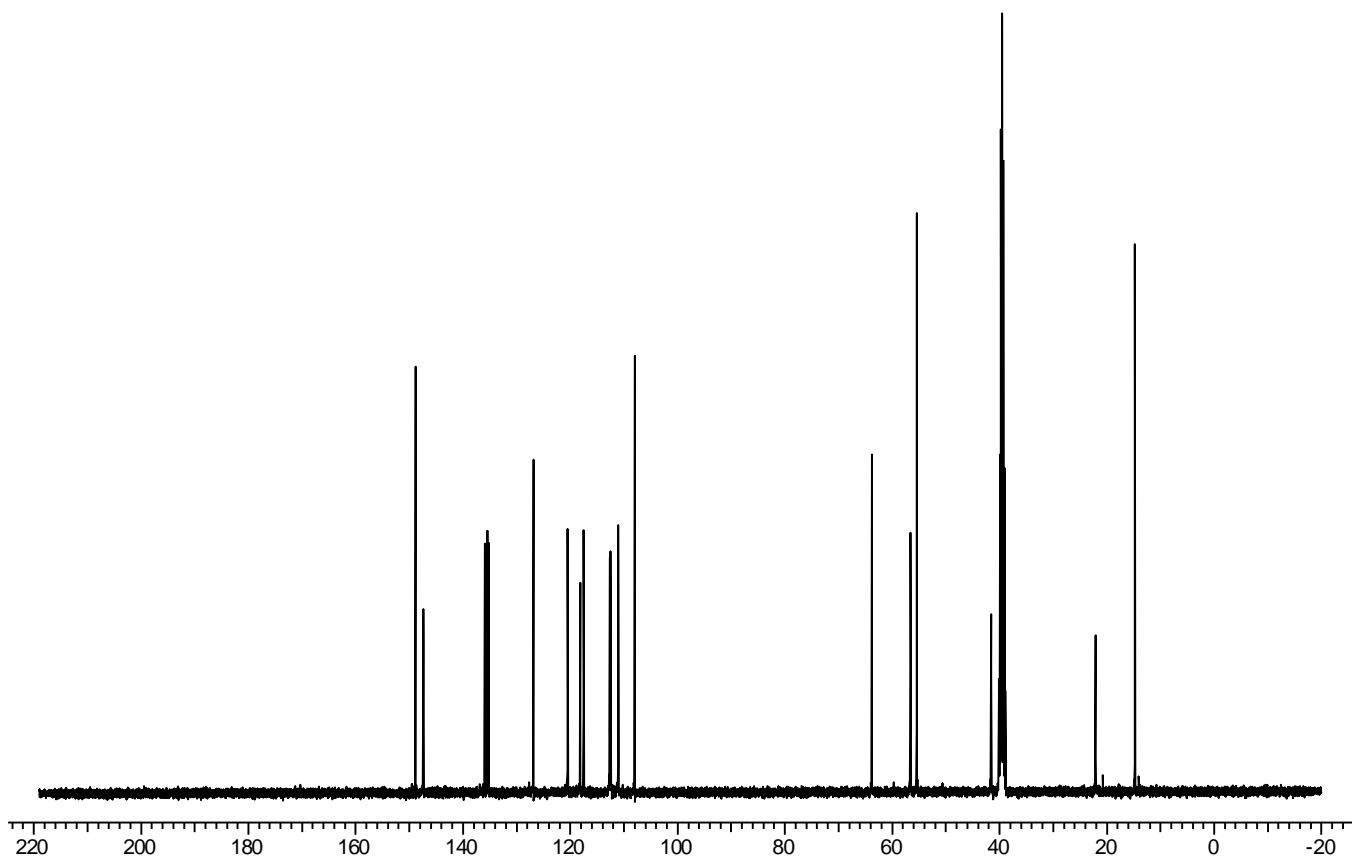
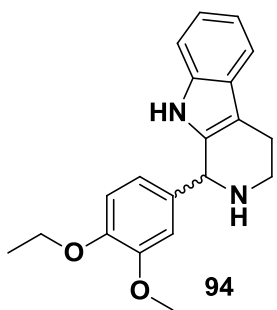
93: <sup>1</sup>H NMR



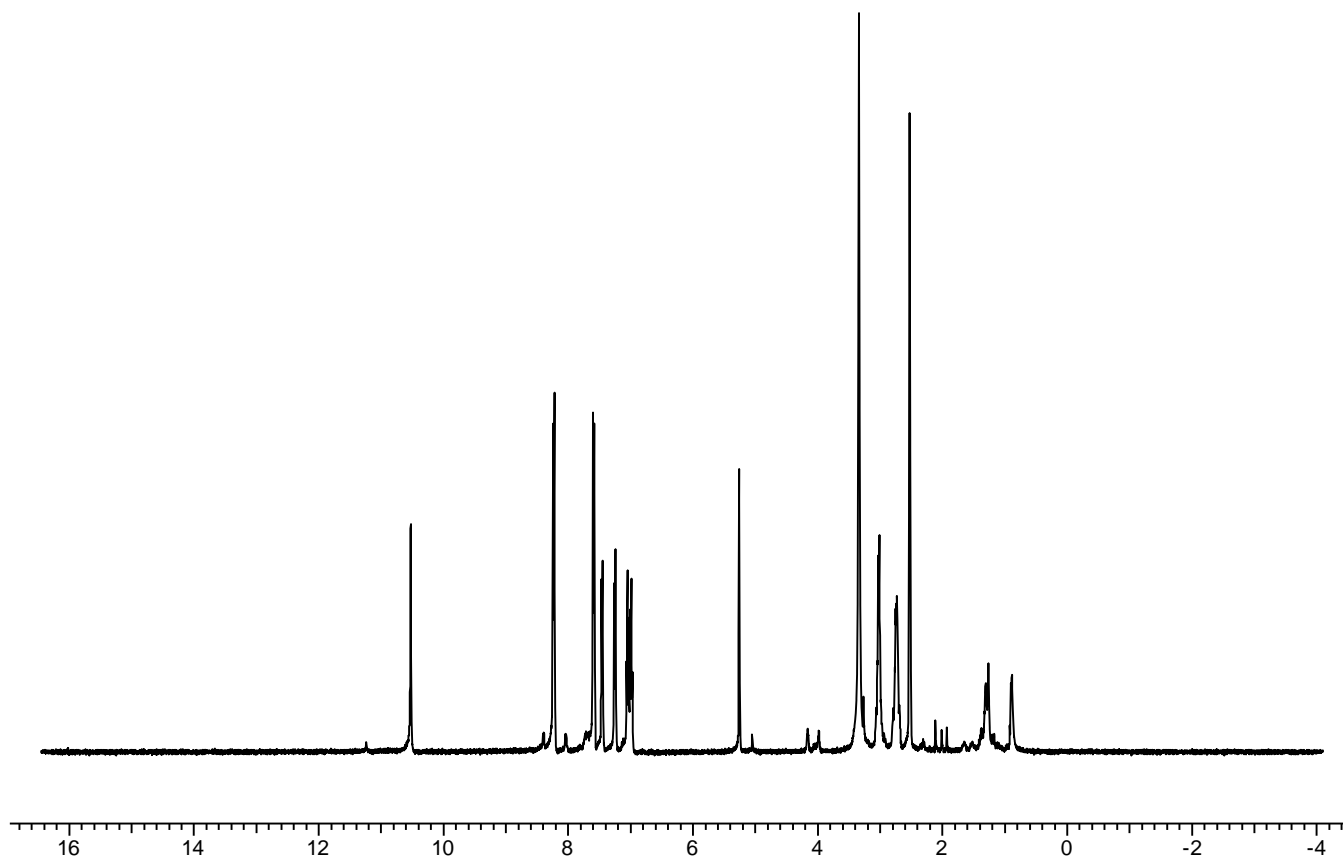
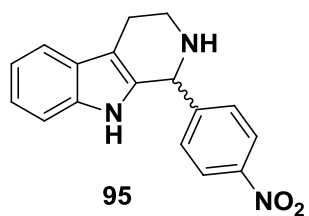
93:  $^{13}\text{C}$  NMR



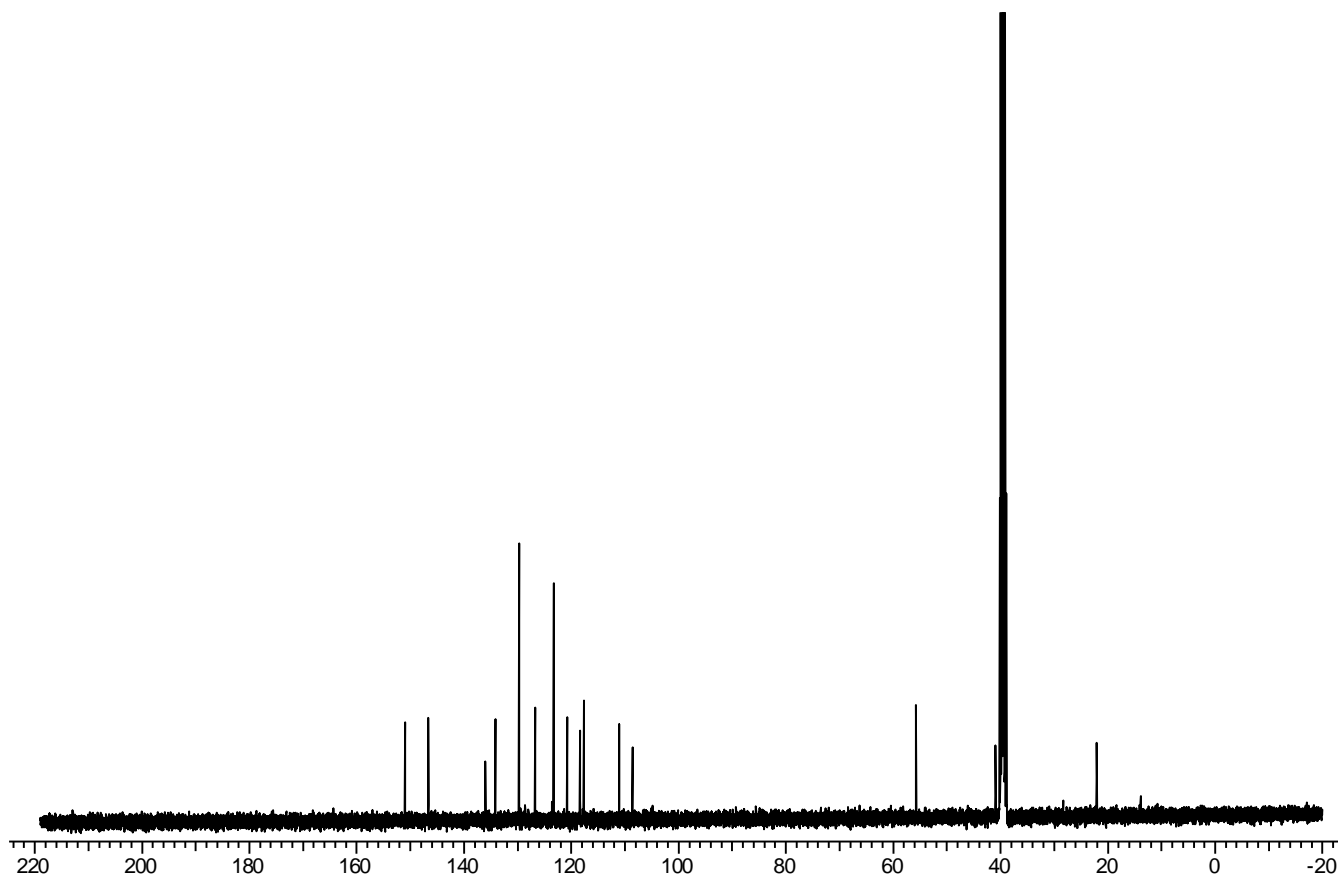
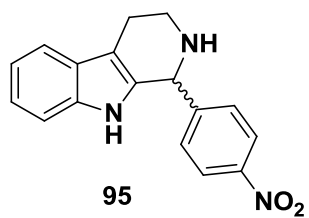
94:  $^1\text{H}$  NMR



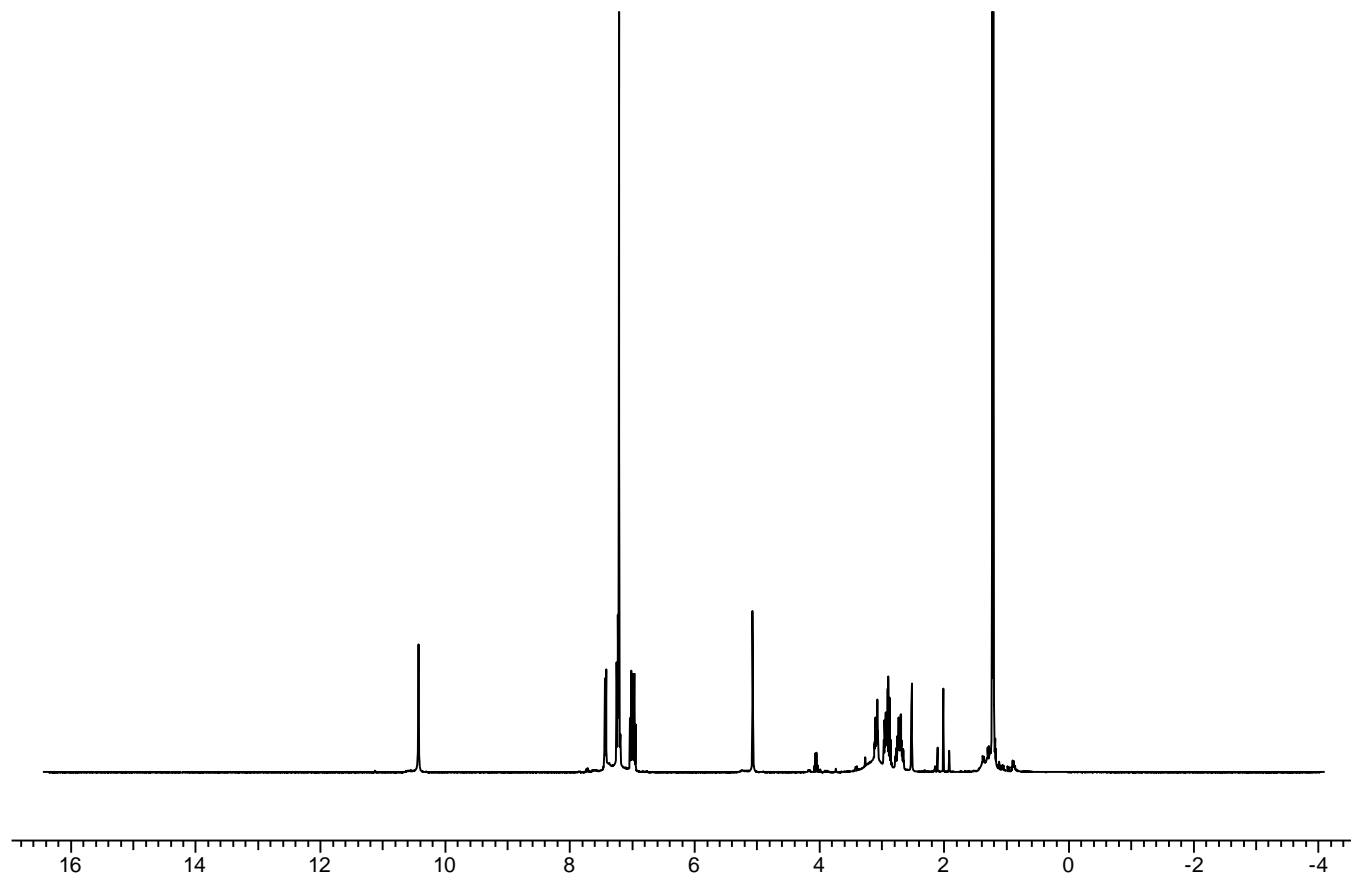
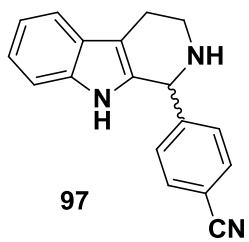
94:  $^{13}\text{C}$  NMR



95: <sup>1</sup>H NMR

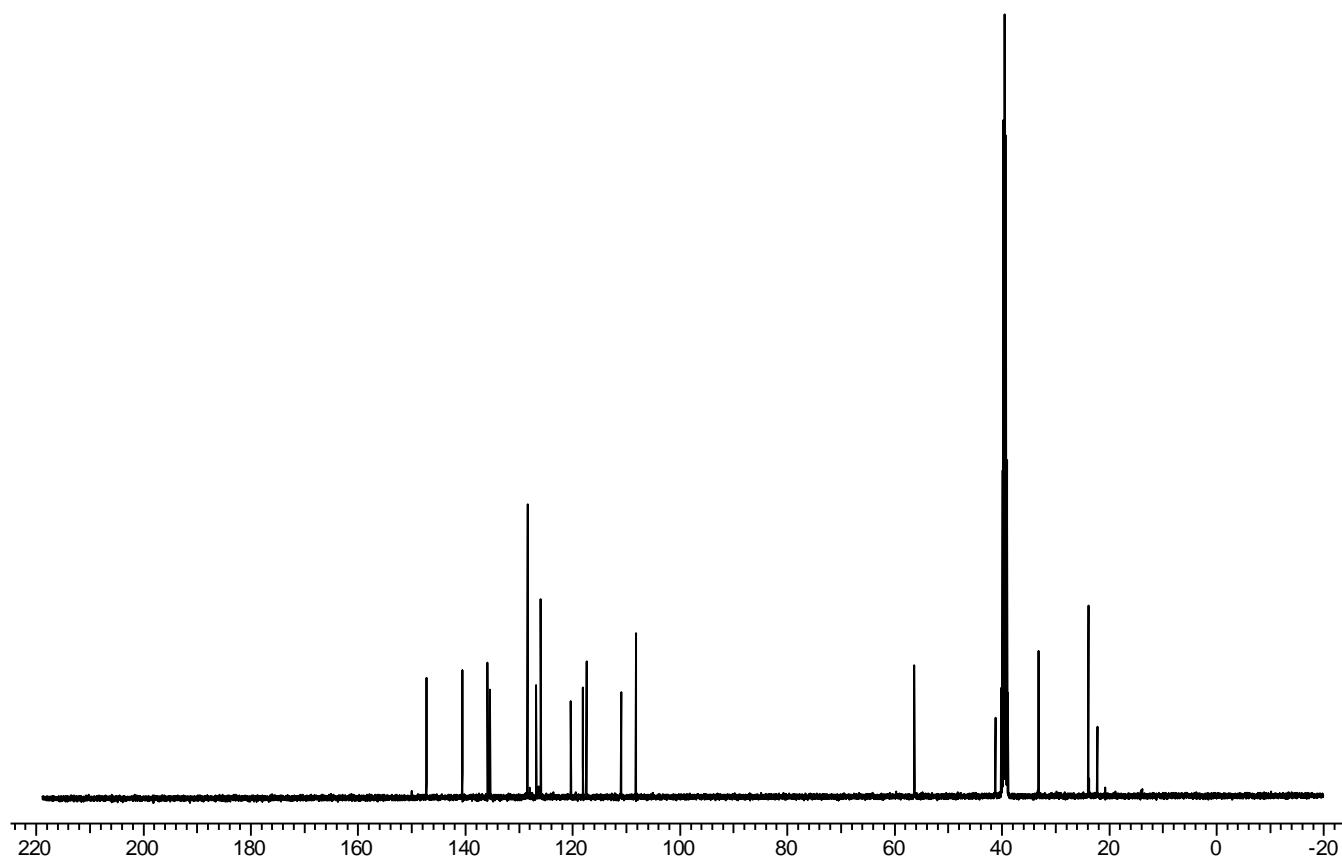
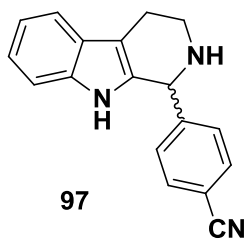


95:  $^{13}\text{C}$  NMR

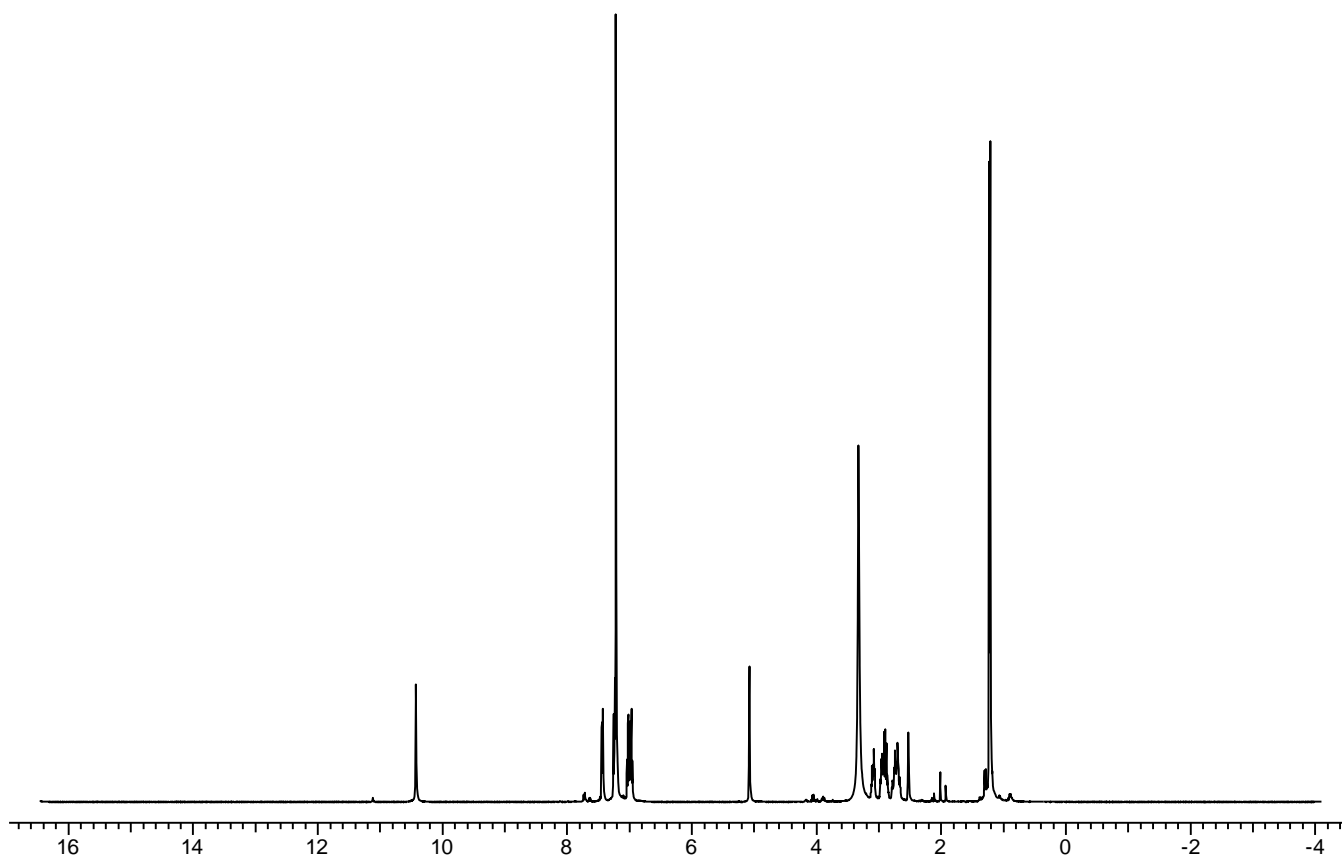
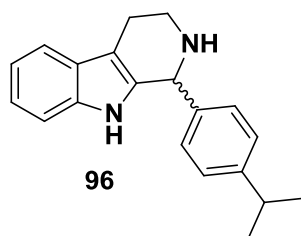


97:  $^1\text{H}$  NMR

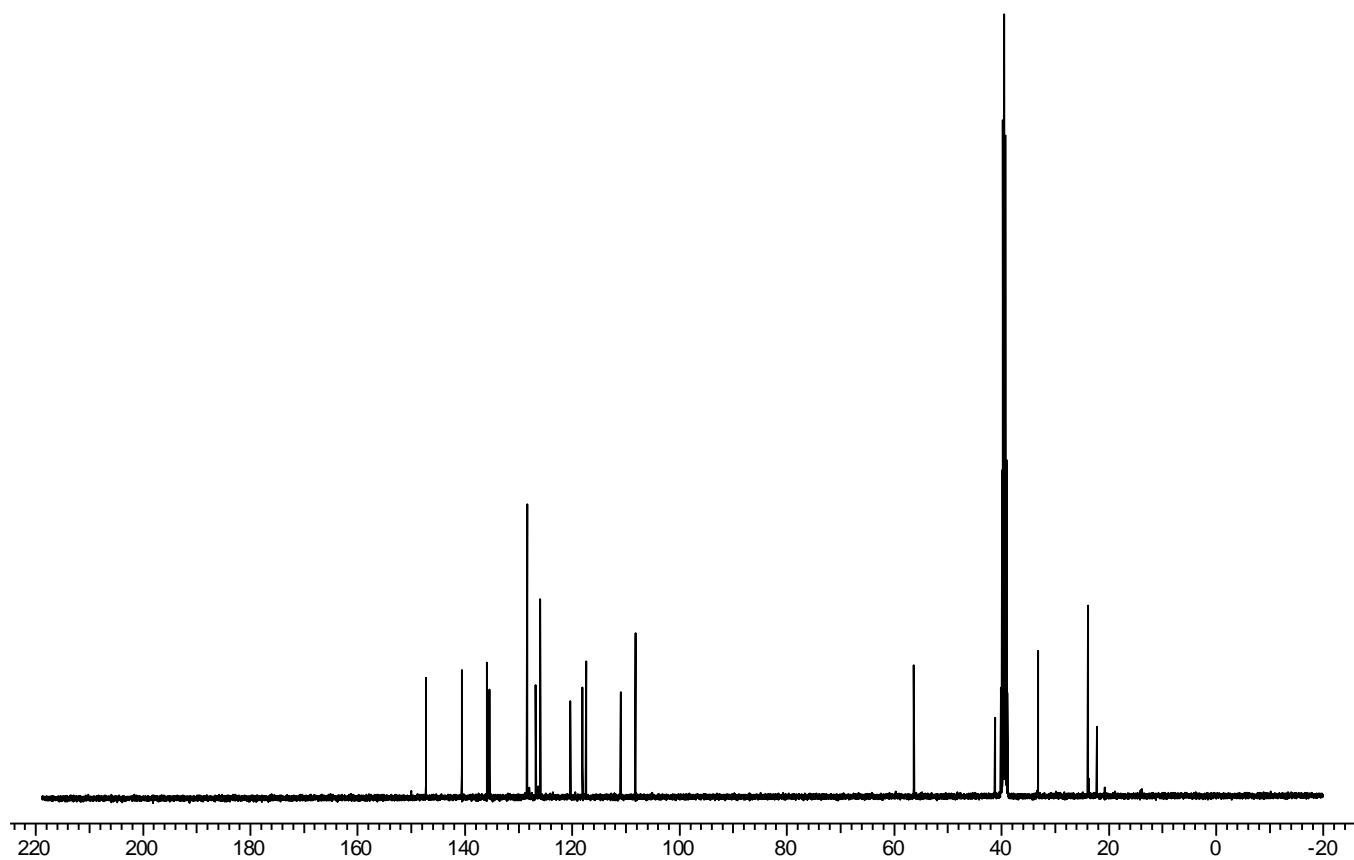
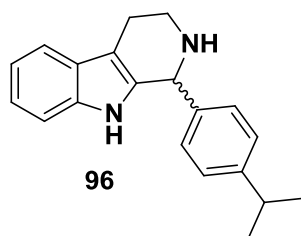




97:  $^{13}\text{C}$  NMR



96:  $^1\text{H}$  NMR



96:  $^{13}\text{C}$  NMR