# DESIGN AND SYNTHESIS OF CYSMETHYNIL ANALOGUES AS INHIBITORS OF ISOPRENYLCYSTEINE CARBOXYL METHYLTRANSFERASE (ICMT)

## **LEOW** JO LENE

*(B.Sc.(Pharmacy)(Hons.), NUS)*

A THESIS SUBMITTED FOR

THE DEGREE OF DOCTOR OF PHILOSOPHY IN SCIENCE

(MEDICINAL CHEMISTRY)

DEPARTMENT OF PHARMACY

NATIONAL UNIVERSITY OF SINGAPORE

2009

# ACKNOWLEDGEMENT

First and foremost, I would like to express my heartfelt gratitude to my supervisor, A/Prof Go Mei Lin, for her constant guidance and advice throughout the course of my research and writing of this thesis, without which this work would not be possible. Her continuous encouragements, support and patience are also deeply appreciated.

My sincere thanks also goes to my co-supervisor, Prof Patrick J. Casey, of Duke-NUS Graduate Medical School, for his continuous guidance, encouragements and immense knowledge, and the opportunity to work along side experienced research fellows in the Casey lab at Duke University Medical Center, Durham, NC.

I would also like to thank Dr Mei Wang for her support, advice and motivations and the opportunity to carry out my research work at the Casey lab at Duke-NUS Graduate Medical School.

It is also my pleasure to thank the following research fellows who made this thesis possible: Dr Rudi Baron, for his assistance in performing the biological assay on the initial chemical library; Dr Suresh Kumar Gorla, for his assistance in completing my chemical library and his advice in my synthetic chemistry work; Dr Yuri Karl Peterson, for preparing the enzymes necessary for my work and his guidance in my biological work; and Dr Andreas Peter Schüller, for sharing his immense knowledge and assistance in my computational work.

I am also indebted to the all the laboratory managers, technologists and assistants, for their invaluable technical assistance, and providing me with the necessary reagents and equipments. Special thanks to Mdm Oh Tang Booy, Ms Ng Sek Eng, Ms Tee Hui Wearn, Ms Audrey Chan, from the Department of Pharmacy, NUS; Mr Lee Peng Chou, Mr Heng Joo Seng, Ms Ho Pei Leng, Ms Kavitha

# CONFERENCES AND PUBLICATIONS

**Conference presentation**

1. 3$^{rd}$ Pharmaceutical Sciences World Congress, Amsterdam, The Netherlands (22-25 April 2007): Optimising drug Therapy: An Imperative for World Health. Poster presentation titled "Quantitative Structure-Activity Relationship (QSAR) of indoloacetamides as inhibitors of human isoprenylcysteine carboxyl methyltransferase".

2. Experimental Biology 2009, New Orleans, Louisiana (18 – 22 April 2009): Today's Research: Tomorrow's Health. Poster presentation titled "Analogues of cysmethynil demonstrate improved isoprenylcysteine carboxyl methyltransferase (Icmt) inhibition activity and antiproliferative activity in PC3 prostate cancer and MDA-MB-231 breast cancer cells".

**Publications**

1. Leow JL, Baron R, Casey PJ, Go ML. Quantitative structure-activity relationship (QSAR) of indoloacetamides as inhibitors of human isoprenylcysteine carboxyl methyltransferase. *Bioorg Med Chem Lett.* **2007**, 17, 1025-32.

2. Wang M, Tan W, Zhou J, Leow J, Go M, Lee HS, Casey PJ. A small molecule inhibitor of isoprenylcysteine carboxymethyltransferase induces autophagic cell death in PC3 prostate cancer cells. *J Biol Chem.* **2008**, 283, 18678-84.

**Manuscripts in preparation**

1. Leow JL, Gorla SK, Go ML, Wang M, Casey PJ. Analogues of the Isoprenylcysteine Carboxylmethyltransferase Inhibitor Cysmethynil With Improved Antiproliferative Activity Against Breast and Prostate Cancer Cells

2. Go ML, Leow JL, Gorla SK, Schüller AP, Wang M, Casey PJ. Amino derivatives of indole as potent inhibitors of isoprenylcysteine carboxyl methyltransferase (Icmt).

# TABLE OF CONTENTS

# SUMMARY

The objective of this thesis was to investigate the chemotherapeutic potential of a group of compounds with an indole core structure as isoprenylcysteine carboxyl methyltransferase (Icmt) inhibitors. Ras proteins contain a C-terminal CaaX motif which direct the proteins through a three-step post-translational process termed prenylation, in which Icmt catalyzes the last step methylation of the C-terminal prenylcysteine. Inhibition of Icmt has multiple impacts on cellular signaling processes which ultimately leads to cell death.

The lead Icmt inhibitor, cysmethynil, was identified from a screen of a diverse chemical library of ~10,000 compounds. A structure activity relationship (SAR) study was first carried out on the indole core compounds discovered from the library. Different methods were employed in this study, namely the (i) principal component anaylsis (PCA) and partial least squares projection to latent structures (PLS), (ii) multiple linear regression and (iii) comparative molecular field analysis (CoMFA). All three approaches complement each other, and identified the steric factor to be important for activity. Improved activity was predicted by incorporation of a bulky side chain at position 1 of the indole ring while a smaller substituent at the position 5 phenyl ring was preferred for Icmt inhibition activity.

Lead optimization efforts guided by the SAR results were performed together with conventional analogue design approach. The compounds synthesized were classified according to (i) substitution on position 1 (Series 1), (ii) alteration at position 5 (Series 2), and (iii) modification of the acetamide side chain at position 3 (Series 3: tertiary amides, Series 4: amines and Series 5: homologues and bioisosteres). A 6th series consisted of compounds with modification made at more than one position (1, 3 and/or 5). The synthesized compounds were evaluated for Icmt

inhibition activity and effects on the viability of human breast cancer cells. Based Icmt inhibitory activity, it was found that the position 3 subsituent was important for activity, as demonstrated by the complete loss of activity in compounds lacking a substituent at this position. Further evaluation of other variation at this position elucidated the importance of a H bond acceptor property at position 3. Optimal length and flexibility is also necessary for activity, as seen in the loss of activity for the shortened homologue. Although presence of substituents at position 1 and 5 was also important for activity, omission of either substituent only resulted in a drop in activity rather than a complete loss of activity. The need for a bulky substituent at position 5 as discovered in the previous QSAR study was verified, as seen by the drop in activity when a shorter side chain such as the 5-carbon isoprenyl moiety was incorporated.

Good correlation was observed between the Icmt inhibition and antiproliferative activities. The SAR of the antiproliferative activity observed was the same as that of the Icmt inhibitory activity, with the exception that the amine analogues (Series 4) were significantly more potent than the other series.

Computational methods were employed to analyze the results from the biological evaluation, specifically the Icmt inhibitory activity, on the 47 compounds synthesized. A pharmacophore model was first developed from the Icmt inhibition data, identifying aromatic and hydrophobic groups as the main features of the indole based Icmt inhibitors, and a H bond acceptor group as the only polar feature in the model. Other reported Icmt inhibitors were screened against the model and 5 out of the 14 screened were found to comply. Further validation could be carried out by actual determination of the biological activity of prospective library screen hits. QSAR analysis was also carried out on the same set of biological data using multiple linear regression, MOE's AutoQsar and Sybyl's CoMFA/CoMSIA. The models

complement each other and also agreed with the pharmacophore model, where size and hydrophobicity was determined to be the main contributor to activity. Visualization of the steric field from CoMFA identified that substituent at position 1 and 5 provides the bulk important for activity.

To better understand the consequences of Icmt inhibition, the amino analogues identified as the most potent inhibitors thus far, were selected for further detailed biological evaluation. Prior studies have demonstrated that inhibition of Icmt by cysmethynil has resulted in various downstream effects such as cell cycle arrest and induction of autophagy, resulting in cell death. Investigation of these biological consequences were conducted on the selected amine analogues and they were found to be able to demonstrate the same effects, with some analogues showing activity more prominently and at much lower doses.

In conclusion, the indole analogues are effective as Icmt inhibitors and are potentially useful as chemotherapeutic agents. The lead optimization effort has elucidated the amine analogues as Icmt inhibitors with significantly improved activity in cell-based studies. These compounds may be used for further investigation of the outcome of Icmt inhibition and subsequent pre-clinical studies. However, there are some limitations in these compounds due to their high lipophilic nature (ClogP > 5). Issues may arise in pre-clinical studies due to difficulties in formulating the compound into a dosage form that can be administered into the test animals. One possible approach to circumvent this problem is to identify new leads with different scaffolds that maybe less lipophilic and more drug-like. The pharmacophore model generated using the synthesized library of compounds may be useful in this aspect, where it can be used to screen against database of lead-like or drug-like compounds

that can be readily obtained for biological evaluation. A new lead compound may also

assist in our quest to further understand the consequences of Icmt inhibition.

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF SCHEMES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| **$^{13}$C NMR** | Carbon-13 nuclear magnetic resonance spectrum |
| **$^{1}$H NMR** | Proton nuclear magnetic resonance spectrum |
| **3D-QSAR** | Three dimensional quantitative structure activity relationship |
| **AdoHcy** | S-adenosylhomocysteine (also SAH) |
| **AdoMet** | S-adenosylmethionine (also SAM) |
| **APCI** | Atmospheric pressure chemical ionization |
| **BFC** | Biotin-S-farnesylcysteine |
| **CD$_3$OD** | Deuterated methanol |
| **CDCl$_3$** | Deuterated chloroform |
| **CDK** | Cyclin-dependent kinase |
| **ClogP** | Calculated log octanol/water partition coefficient |
| **CoMFA** | Comparative molecular field analysis |
| **CoMSIA** | Comparative molecular similarity indices analysis |
| **DMEM** | Dulbecco's Modified Eagle's Medium |
| **DME** | Dimethoxyethane |
| **DMF** | *N,N*-Dimethylformamide |
| **DMSO-*d$_6$*** | Deuterated dimethylsulfoxide |
| **EDTA** | Ethylenediaminetetraacetic acid |
| **ESI** | Electron spray ionization |
| **FACS** | Fluorescence activated cell sorter analysis |
| **FBS** | Fetal bovine serum |
| **FN** | False negative |
| **FP** | False positive |
| **FTase** | Farmesyltransferase |
| **GAP** | GTPase activating protein |

| | |
|---|---|
| **GAPDH** | Glyceraldehyde-3-phosphate dehydrogenase |
| **GDP** | Guanosine 5'-diphosphate |
| **GEF** | Guanosine nucleotide exchange factor |
| **GGTase-I** | Geranylgeranyltransferase type I |
| **GTP** | Guanosine 5'-triphosphate |
| **Hepes** | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| **HOMO** | Highest occupied molecular orbital |
| **HPLC** | High performance liquid chromatography |
| **HRMS** | High resolution mass spectrum |
| **Icmt** | Isoprenylcysteine carboxyl methyltransferase |
| **$K_d$** | Dissociation constant |
| **LC3** | Microtubule-associated protein 1 light chain 3 (also MAP1LC3) |
| **LUMO** | Lowest unoccupied molecular orbital |
| **MDA-MB-231** | Human breast cancer cell line |
| **MLR** | Multiple linear regression |
| **MOE** | Molecular Operating Environment |
| **mTOR** | Mammalian target of rapamycin |
| **MTS** | 3-(4,5-Dimethylthiazol-2-yl)-5-(3-carboxymethyoxyphenyl)-2-(4-sulfonphenyl)-2$H$-tetrazolium |
| **PBS** | Phosphate-buffered saline |
| **PC3** | Human prostate cancer cell line |
| **PCA** | Principal component analysis |
| **PLS** | Partial least squares projection to latent structures |
| **PVDF** | Polyvinylidene difluoride |
| **Rce1** | Ras converting enzyme |
| **RIPA** | Radioimmunoprecipitation assay |
| **SAH** | See **AdoHcy** |

| | |
|---|---|
| **SAM** | See **AdoMet** |
| **SDS-PAGE** | Sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| **SEE** | Standard error of estimate |
| **SEP** | Standard error of prediction |
| **THF** | Tetrahydrofuran |
| **TMS** | Tetramethylsilane |
| **TN** | True negative |
| **TP** | True positive |

# CHAPTER 1 INTRODUCTION

## 1.1. OVERVIEW OF POST-TRANSLATIONAL PRENYLATION OF PROTEINS

Many proteins are subjected to a range of chemical modifications after they are synthesized. These modifications are termed post-translation modifications and may involve the attachment of other functional groups (by acylation, phosphorylation, glycosylation, glycation, prenylation among others) or peptides (eg. ubiquitination) to the original protein, a change in the chemical nature of a specific amino acid (for example deamination of arginine to give citrulline) or changes in structure like the formation of disulfide bridges between cysteine residues. These modifications can have a significant impact on the biological activity of the protein. Over the past decade, post-translational modification of proteins by isoprenoid lipids, a process termed prenylation, has been widely investigated due in part to its central role in the subcellular localization and function of a group of proteins characterized by a carboxy-terminal tetrapeptide (CaaX) motif, where "C" is cysteine, "a" is an aliphatic amino acid and "X" represents an unspecified amino acid.[1-2] The presence of this motif on a protein triggers three sequential post-translational modifications which are depicted in **Figure 1-1**.
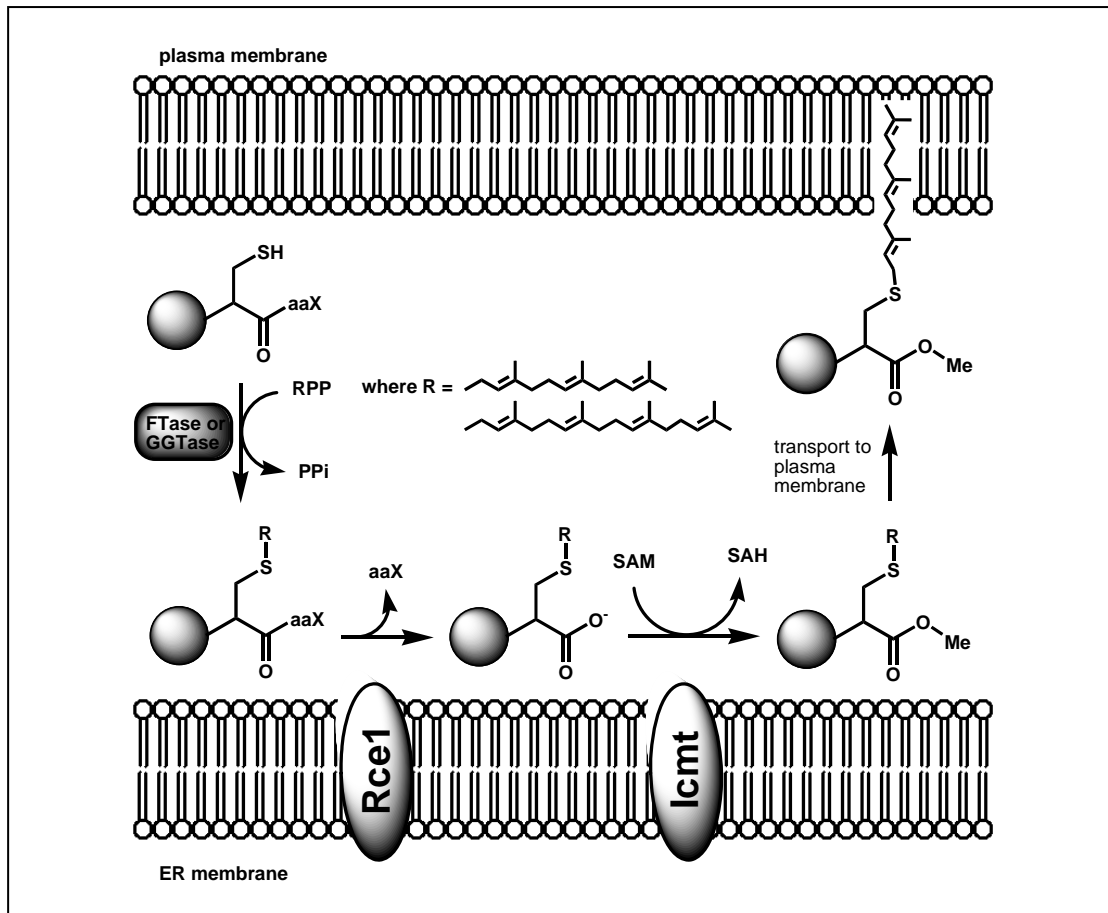
**Figure 1-1. Post-translational modification of CaaX proteins.[2-3]**

In the first step of CaaX protein processing, an isoprenoid residue, farnesyl (15-carbon) or geranylgeranyl (20-carbon), is attached by a thioether linkage to the cysteine of the CaaX sequence.[3] This reaction is catalyzed by protein farnesyltransferase (FTase) or protein geranylgeranyltransferase type I (GGTase-I). Generally, when X is methionine, serine, glutamine or alanine, the protein is a substrate of FTase, whereas if X is leucine, it is processed by GGTase-I.[2, 4] The nature of the X residue does provide some guidance as to whether the CaaX sequence is farnesylated or geranylgeranylated but it is still best to rely on experimental confirmation.[5] There are some CaaX proteins that are substrates of both FTase and GGTase-I, for example, K-Ras, RhoB and RhoH.[4] After prenylation, the modified CaaX protein moves from the cytosol to the endoplasmic reticulum where it

associates with the cellular membrane presumably by inserting its lipophilic isoprenoid side chain into the phospholipid bilayer. It is then recognized by an endoprotease termed Ras converting enzyme 1 (Rce1), an integral ER membrane protein, that removes the last three amino acids (-aaX) from the C-terminus.[6] This brings the newly exposed α-carboxylate anion of the prenylcysteine residue in close proximity to the negatively charged phospholipids in the membrane. Likely due in part to minimize the electrostatic repulsion, the terminal carboxylate group is converted to a methyl ester in a reaction catalyzed by isoprenylcysteine carboxyl methyltransferase (Icmt), which like Rce1 is a membrane bound protein located in the endoplasmic reticulum.[7-8] Unlike the preceding steps, methylation by Icmt is potentially reversible[9] but a specific methylesterase catalyzing this step has yet to be identified. The sequence of reactions involved in the post-translational prenylation of CaaX proteins enhances the lipophilic character of otherwise hydrophilic proteins and promotes their ability to associate with membranes and to engage in specific protein-protein interactions.[3]

About 280 potential CaaX proteins were identified during the sequencing of the human and mouse genomes but less than half of these proteins were proposed to undergo post-translational prenylation.[10] It appears that the processing enzymes do not recognize a particular CaaX sequence or fail to gain access to certain terminal carboxyl groups.[11] An important class of CaaX proteins that are processed by prenylation are the Ras proteins, which play essential roles in controlling several crucial signaling pathways that regulate normal cellular proliferation. Ras proteins function as binary molecular switches that alternate between two states, an inactive state in which Ras is in complex with guanosine 5'-diphosphate (GDP) and an activated state in which it is bound to guanosine 5'-triphosphate (GTP).[12] The

equilibrium between these states is regulated by two proteins: guanine nucleotide exchange factors (GEFs) which promote activation by enhancing the release of GDP and the binding of GTP, and GTPase activating proteins (GAPs) which promote inactivation by accelerating the intrinsic GTPase activity of the Ras proteins.[13] In the activated state, Ras proteins trigger an array of cellular messengers that transduce key signals regulating cell proliferation,[14] migration[15] and apoptosis.[16] Mutations in *Ras* genes frequently involve the loss of intrinsic Ras GTPase activity and insensitivity to Ras-GAPs.[17] The aberrant Ras proteins retain constitutive activity and participate in unregulated signaling that result in the malignant transformation of cells.[18-19] Ras mutations are found in approximately 30% of all human tumors, notably pancreatic, colon, lung and hematological malignancies.[17, 20] In these tumors, the activated Ras protein contributes significantly to several aspects of the malignant phenotype including the deregulation of tumor-cell growth, angiogenesis, programmed cell death and invasiveness.[21] For this reason, targeting Ras proteins and its network of signalling pathways is viewed as a viable and promising strategy for developing therapies against cancer. Unfortunately targeting Ras per se, with the aim of blocking its activity, has not yet been a feasible option.[22] A more tenable approach that has been undertaken in recent years is to divert Ras from its subcellular locations (plasma membrane, endoplasmic reticulum, Golgi apparatus) by interfering with the enzymes involved in the post-translational modification of their carboxy terminal tetrapeptide (CaaX) residues.[11]

## 1.2. CAAX PROCESSING ENZYMES AS TARGETS IN ONCOGENESIS

## 1.2.1. Inhibitors of FTase

Of the enzymes involved in post-translational prenylation, FTase is the preferred target for therapeutic intervention. The primary reasons for targeting FTase are that the enzyme catalyzed the first and rate-limiting step in CaaX processing, it is soluble and can be readily purified, and FTase processes fewer substrates compared to Rce1 or Icmt, which can translate to less toxicity from an FTase inhibitor.[23] In addition, the major Ras proteins implicated in oncogenesis, H-, K- and N-Ras, are all substrates of FTase. Several FTase inhibitors have been developed[24] and some members have advanced to clinical trials.[25-26] Unfortunately, the overall results have not been encouraging, possibly because under conditions of FTase inhibition, N-Ras and K-Ras which are the isoforms most commonly involved in human cancers, are substrates of GGTase-I and the geranylgeranylated Ras proteins retained biological activity.[27] Nonetheless, the FTase inhibitor lonafarnib (**Figure 1-2a**) has showed good activity against hematological cancers[28] and there is evidence to suggest that another inhibitor tibifarnib (**Figure 1-2b**) affects targets that are downstream of Ras.[29]
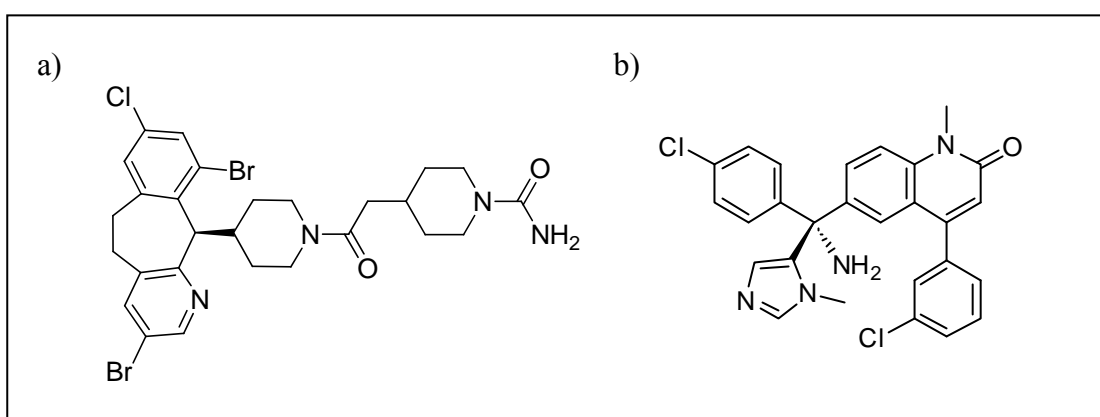


**Figure 1-2. Structures of selected FTIs.** (a) Lonafarnib (SCH66336)[28], (b) Tipifarnib (R15777)[29]

### 1.2.2. GGTase-I inhibitors

The finding of alternate prenylation of Ras proteins has led to increased attention being paid to GGTase-I inhibitors, along with growing evidence that geranylgeranylated proteins are involved in oncogenesis and other disease states. Several GGTase-I inhibitors have since been described (**Figure 1-3**).[30-34] Recent findings that the majority of Rho proteins are modified by FTase and not GGTase-I[5] may mean that GGTase-I inhibitors can be used in combination with other anti-cancer drugs in order to maximize their biological effects.
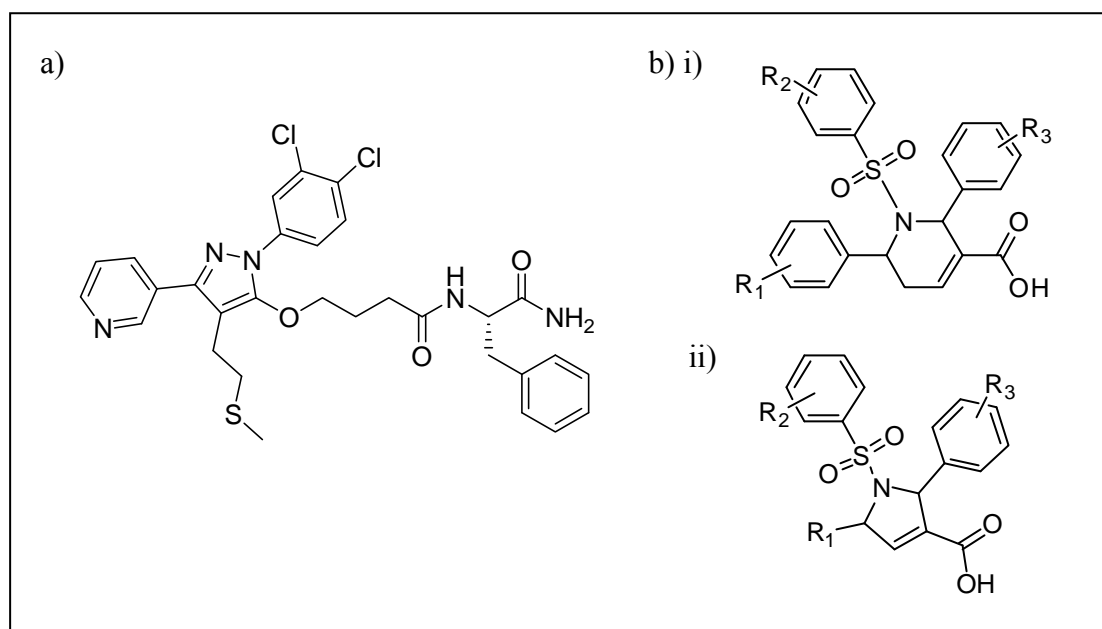


**Figure 1-3. Structures of small molecule GGTIs.** a) Structure of GGTI-DU40, optimized by Pharmaceutical Product Development (PPD) discovery from a novel inhibitor identified from a diverse chemical library[30] b) Core structures of GGTIs identified from a library of allenoate-derived compounds[34] i) tetrahydropyridine ring core, ii) dihydropyrrole ring core.

### 1.2.3 Inhibitors of the post-prenylation enzymes: Rce1 and Icmt

Interest in the inhibitors of the post-prenylation enzymes Rce1 and Icmt have only recently intensified.[11] Initially, there have been concerns regarding the effectiveness of inhibiting these enzymes since essentially all FTase and GGTase-I substrates are substrates of Rce1 and Icmt. Thus, inhibition of Rce1 and Icmt may

have widespread effects on cellular functions and may cause significant toxicity in normal cells. Indeed, disruption of either the *Rce1* or *Icmt* genes in mice has resulted in embryonic lethality.[35-36] Additionally, Ras proteins that do not undergo post prenylation processing retain partial localization and function, calling into question whether targeting these enzymes will have sufficient impact on the function of oncogenic CaaX proteins to impair cancer growth.[37-38] Mammalian genomes encode only one member of the Icmt class of methyltransferases, and Icmt lacks homology compared to any other protein methyltransferases.[39] Additionally only a single gene has been identified in encoding a Rce1-like protein in vertebrates, which implies that Rce1 is the only enzyme involved in the processing of prenylated substrates.[40-41] Both enzymes are distinct in their structure and mechanism from other proteases and methyltransferases, a point that should favour the development of specific inhibitors. Despite concerns with the effectiveness and safety issues in targeting these enzymes, with their unique properties and critical roles they have emerged as promising drug targets in cancer chemotherapy.

*1.2.3.1. Rce1 inhibitors*

Rce1 inhibitors have not been widely investigated, but some interesting compounds have emerged from recent studies. A small number of non-peptidic, non-prenylic compounds have been investigated as inhibitors of Rce1.[42] These compounds are regarded as product analogues of the farnesyltransferase reaction as they mimick the structure of the farnesylated CaaX peptide. This has led to the proposal that they may function as substrate-based inhibitors of Rce1. As it turned out, most of the compounds inhibited FTase but not all inhibited Rce1. The more promising compounds could serve as lead structures for the further development of Rce1 inhibitors with anti-cancer potential. One such member is shown in **Figure 1-4a**.

Several halomethylketones which are classic protease affinity labeling agents have been identified as non-specific Rce1 inhibitors.[43] The introduction of a S-farnesylcysteine moiety greatly enhanced specificity for Rce1 as seen in BFCCMK (**Figure 1-4b**)[41] and further functionalization yielded peptidyl acyloxymethyketones (AOMKS) that were dual inhibitors of yeast Rce1 and Icmt.[44] A screen of compounds in the National Cancer Institute (NCI) Developmental Therapeutics Program Diversity Set Compound Library for small molecule inhibitors of yeast and human Rce1 yielded nine members of which five inhibited both Rce1 and Icmt while the remaining four were fairly specific inhibitors of Rce1.[45] Interestingly, the most specific Rce1 inhibitor (**Figure 1-4c**) is prone to colloid aggregation which is normally observed among promiscuous inhibitors.[46] It remains to be determined if this interferes with its Rce1 inhibitory activity. Potential Rce1 inhibitors should be screened for effects on cardiac function as it is known that deletion of Rce1 caused mortality in mice due to cardiomyopathy.[47]
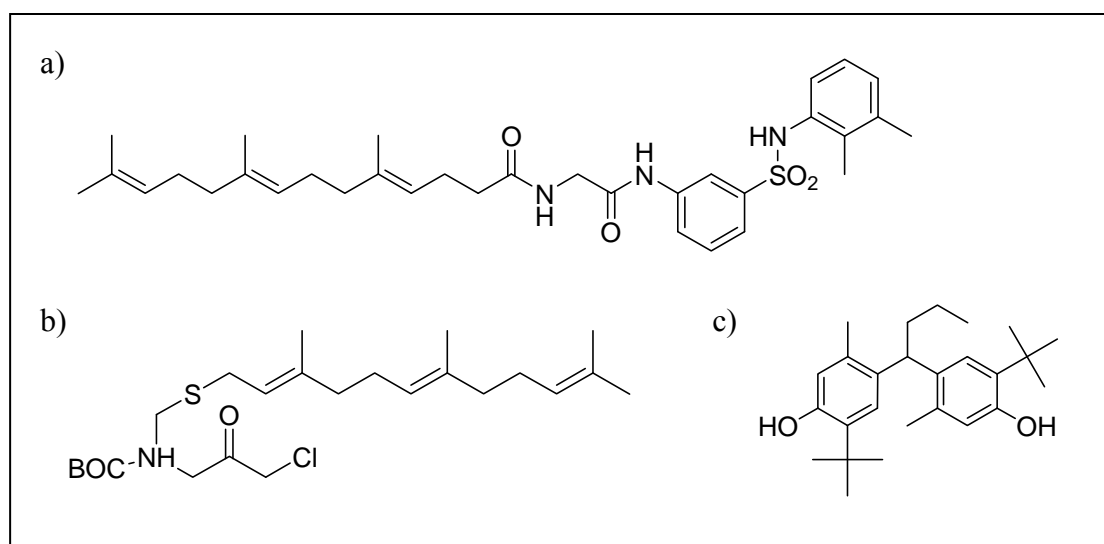


**Figure 1-4. Structures of selected Rce1 inhibitors.** a) An example of a non-peptidic, non-prenylic Rce1 inhibitor[42] b) BFCCMK,[41] c) Most promising compound reported by Manadhar et al[45]

*1.2.3.2. Icmt inhibitors*

Three classes of Icmt inhibitors have been reported to date. The first are the product-based inhibitors typified by S-adenosylhomocysteine (SAH) and compounds that increase intracellular SAH.[48-49] Icmt catalyzes the transfer of a methyl group from the endogenous methyl donor S-adenosylmethionine (SAM) to the isoprenylated CaaX proteins (**Figure 1-5**). In the process, SAM is converted to SAH which binds to and functions as a feedback competitive inhibitor of Icmt.[50] SAH hydrolase catalyzes the hydrolysis of SAH to adenosine and homocysteine. This reaction is reversible and the reverse reaction to give SAH is energetically more favourable. Product-based inhibitors are likely to be non-specific as other cellular methyltransferases are also inhibited by SAH.[51] However, the recent finding that methotrexate, one of the most established drugs in cancer chemotherapy, targets Icmt through an elevation of SAH is of particular interest.[52] Methotrexate inhibits nucleotide biosynthesis by its antifolate activity. Depletion of 5-methyltetrahydrofolate by methotrexate blocks the methylation of homocysteine to form methionine. The accumulated homocysteine reacts with adenosine to give SAH which in turn inhibits cellular methyltransferases. Investigations showed that methotrexate inhibits Icmt with an $IC_{50}$ value of 7.5 μM, causes the mislocalization of Ras proteins, reduces signalling through Ras pathways and has no effect on *Icmt* -/- cells.[52] These findings lend support to the view that Icmt is a critical additional target of methotrexate.
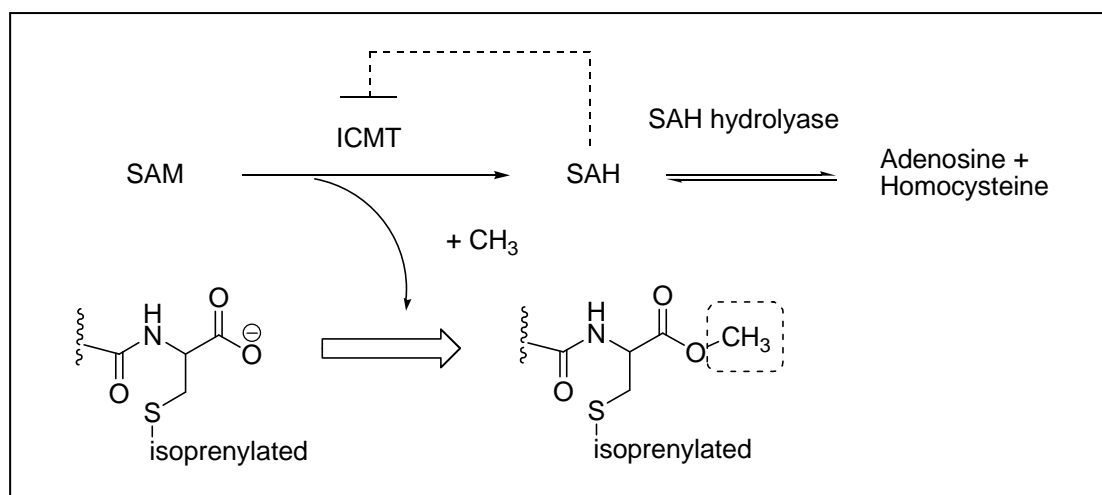
**Figure 1-5. Inhibition of Icmt by SAH and compounds that increase intracellular SAH.**

The second class of inhibitors is derivatives of the substrate prenylcysteine (*N*-acetyl-*S*-farnesyl-L-cysteine, AFC or *N*-acetyl-*S*-geranylgeranyl-L-cysteine, AGGC) which can function as competitive inhibitors of Icmt. A well-studied example of substrate-based inhibitor is the functionalized derivatives of AFC (**Figure 1-6**).[53-55] As these compounds are structural mimics of the C-terminal prenylcysteine of processed CaaX proteins, and are processed by Icmt to yield their methylated versions, they are likely to have pleiotropic effects.[56]
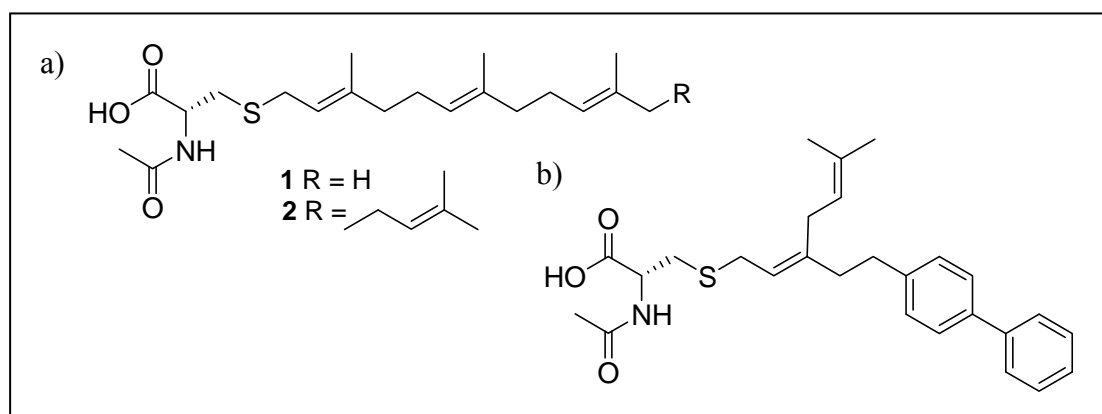


**Figure 1-6. Structures of (a) minimal substrate of Icmt and (b) substrate based inhibitors.**[53] a) *N*-acetyl-*S*-farnesyl-L-cysteine (AFC) (**1**) and *N*-acetyl-*S*-geranylgeranyl-L-cysteine (AGGC) (**2**) b) An isobutenyl biphenyl derivative of AFC.

The third class of inhibitors is a miscellaneous class of small molecules that have been identified by high throughput screens. The screening of marine sponges and plants have yielded promising compounds with *in vitro* Icmt inhibitory IC$_{50}$ values of 2 -30 μM) (**Figure 1-7**).[57-59] These compounds have yet to be evaluated on cell-based assays but their varied scaffolds may be promising leads for drug discovery.
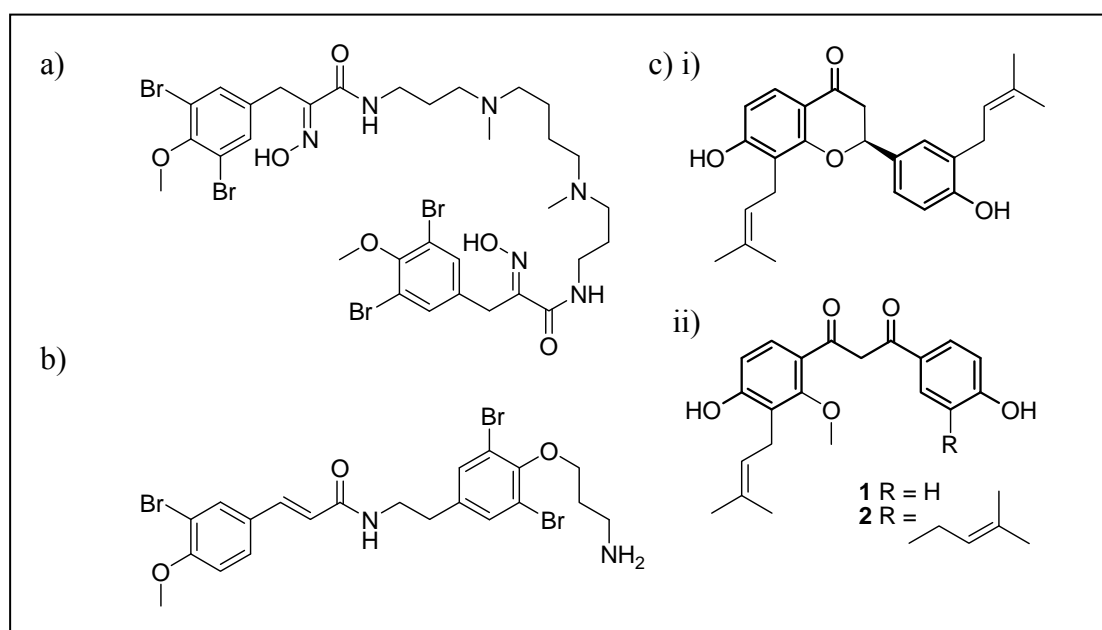


**Figure 1-7. Structures of reported natural product inhibitors.** a) Spermatinamine from marine sponge *Pseudoceratina* sp.[57]; b) Aplysamine 6 from marine sponge *Pseudoceratina* sp.[59]; c) *Hovea parvicalyx* extracts[58]

Thus far, cysmethynil (2-(1-octyl-5-*m*-tolyl-1*H*-indol-3-yl)acetamide) is the most potent and extensively investigated Icmt inhibitor (**Figure 1-8**). This compound was discovered during a screen of a chemically diverse library comprising of over 10,000 compounds.[60] It is a competitive inhibitor with respect to the isoprenylated cysteine substrate but a non-competitive inhibitor with respect to the methyl donor SAM.[61] Inhibition has been shown to be time-dependent, involving an initial rapid association with the enzyme followed by a slow isomerization to give a tighter enzyme-inhibitor complex.[61] Cancer cells treated with cysmethynil caused

mislocalization of Ras, impaired growth factor signaling and blocked anchorage-independent growth of human colon cancer cells which was reversed by over-expression of Icmt.[60] Subsequent investigations showed that cysmethynil is able to control tumor growth in a xenograph mouse model of prostate cancer, to trigger G1 arrest in PC3 prostate cancer cells and to induce autophagic cell death.[62] The dual effects of cysmethynil on cell cycle and atutophagic cell death may arise from its ability to reduce mTOR signaling due to reduced Ras and Rheb activities.[62]
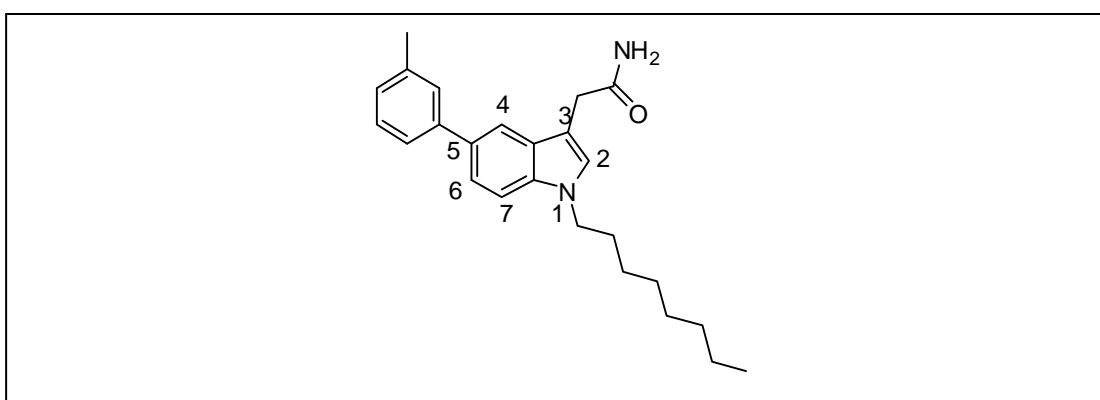


**Figure 1-8. Structure of Cysmethynil with numbering system used in this thesis.**

## 1.3 STATEMENT OF PURPOSE

Of the various CaaX processing enzymes, there is consensus that Icmt is a promising druggable target with good potential in drug discovery. There are various reasons to support this viewpoint. The post-prenylation processing enzymes Rce1 and Icmt may be more attractive targets than the prenylation enzymes (FTase, GGTase-I) because they are less likely to disrupt the biological activities of the prenylated CaaX proteins, although as noted above this may also impact their effectiveness. However, there is increasing evidence that attenuation but not elimination of CaaX protein function can offer a more attractive means of therapeutic intervention.[63-64] In addition, the biological consequences arising from Icmt inhibition have exceeded that due to

blocking Rce1 activity. Inactivation of Icmt inhibits K-Ras induced myeloproliferative diseases in mice[65] while inactivation of Rce1 accelerates the progression of myeloproliferative disease caused by oncogenic K-Ras.[66] These findings provide support for Icmt as a preferred therapeutic target compared to Rce1. Additionally, the inhibition of Icmt by cysmethynil has led to widespread downstream effects (cell cycle arrest, autophagy) that contributed to cell death.[62] The unexpected finding that methotrexate, one of the most effective anticancer drugs available, also inhibited Icmt[52] thus further strengthens the rationale for targeting Icmt.

As mentioned in **Section 1.2.3.2**, cysmethynil is the most effective and widely investigated Icmt inhibitor to date. There is, however, very limited information on the structural features of cysmethynil that are associated with its inhibitory activity. Without this information, it would be difficult to improve or optimize the activity or to propose novel lead compounds that are superior to cysmethynil. Thus, one objective of this thesis is to carry out a detailed analysis of structure-activity relationships relating to the inhibitory activity of cysmethynil.

Cysmethynil was identified through the screening of a diverse chemical library comprising of over 70 subfamilies derived from different scaffolds.[60] The sub-family to which it belongs consists of some 70 indole derivatives with different substituents on the 5-phenyl ring and the indole nitrogen (position 1) but retained the same acetamide side chain (position 3) as cysmethynil. These compounds have been evaluated for *in vitro* Icmt inhibition[67] and they constitute a valuable database from which important clues may be derived relating to the optimal substitutions at positions 1 and 5 of the indole ring. As such an analysis has not been carried out, it is pursued here with the aim of determining the critical features for activity at these positions. The inhibitory activities of the compounds in this indole library range from $IC_{50}$

values 2 μM to more than 50 μM.[67] As these variations did not result in more potent inhibitors (nanomolar $IC_{50}$ values), it may be that positions 1 and 5 are not key sites affecting inhibitory activity. The acetamide side chain at position 3 which is retained in all the compounds and has not been altered thus far, and hence is considered as a promising site for modification. To investigate this hypothesis, modifications to the acetamide side chain will be explored in this investigation. Conventional analogue design strategies such as homologation, bioisosteric replacement and variation of substituents[68] will be employed for this purpose. The option of structure-based design, a design strategy based on the 3D structure of the biological target obtained by X-ray or NMR, cannot be pursued because the structure of the membrane bound Icmt has not been elucidated. Thus, a series of indole derivatives will be synthesized with modifications at (i) positions 1 and 5. The study is guided by the analysis of SAR from the previously evaluated indole library and (ii) position 3 based on analogue-based changes of the acetamide side chain. The compounds have been evaluated for in vitro Icmt inhibitory activity and effect on the viability of a cancer cell line. With this information on hand, it became possible to derive a pharmacophore model (which is the ensemble of steric and electronic features necessary for biological response) for Icmt activity, and to analyze SAR in descriptive and quantitative terms.

Cysmethynil has poor water solubility and has been noted to have strong plasma protein binding properties. These are features that can compromise its pharmacokinetic profile. When cysmethynil was assessed for its drug-likeness by established guidelines like the Lipinski Rules[69] or the additional rules introduced by Veber and co-workers,[70] non-compliance was found for certain properties, namely lipophilicity (too high) and flexibility (large number of rotatable bonds). Hence, an important objective of the design strategy is to incorporate functionalities that would

reduce lipophilicity, enhance polar surface area and limit flexibility. A reduction in lipophilicity and an increase in polar area are common design strategies aimed at reducing binding to plasma proteins.[71] It is of interest to determine how far this could be achieved without compromising biological activity.

Post prenylation is a process that affects a broad range of CaaX proteins and the biological consequences of Icmt inhibition remain to be fully understood. Cysmethynil has been shown, at least in prostate cancer cells, to arrest the cell cycle at the G1 phase, to have no effect on apoptosis and to activate the authophagic pathway leading to cell death.[62] Thus, another objective of this thesis is to determine if other Icmt inhibitors (besides cysmethynil) affect cancer cells in the same way.

In summary, it is hypothesized that the anticancer activity of cysmethynil can be enhanced by modifying its acetamide side chain and by incorporating functionalities with "drug-like" or "lead-like" features into the molecule. The hypothesis will be investigated by synthesis of target compounds, evaluation of their Icmt inhibitory activities and effect on cell viability, and computational analysis of the results to aid future lead design approaches.

.

# CHAPTER 2: QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP (QSAR) OF INDOLOACETAMIDES AS INHIBITORS OF ICMT

## 2.1. INTRODUCTION

In this chapter, the Icmt inhibitory activities of the indole library from which cysmethynil was identified were analyzed for a better understanding of the structural requirements for activity at positions 1 (indole N) and 5 (phenyl ring) of the indole template. Three levels of analysis were undertaken. First, the results were analyzed by projection methods, namely principal component analysis (PCA) and partial least squares projection to latent structures (PLS). The purpose was to identify the physicochemical descriptors that were important determinants of activity. Second, multiple linear regression was applied to derive an appropriate equation relating activity to physicochemical descriptors. Third, a comparative molecular field analysis (CoMFA) was carried out to provide a visual means of interpreting the structure-activity relationships (SAR) derived from the above mentioned approaches.

## 2.2. METHODS

The indole library consisted of 72 compounds and each compound was drawn and energy minimized using the Sybyl 7.0 standard Tripos force field (Tripos Inc, St Louis, MO, USA). Several descriptors representing size, electronic and lipophilic characteristics were collected from forcefield minimized structures of the compounds. The following descriptors were determined with Sybyl 7.0 (ClogP/CMR and QSAR with CoMFA modules): ClogP, area, volume, polar surface area, polar volume, molar refractivity and the Hansch $\pi$ values. $\pi_{N\ substituent}$ was determined from $ClogP_{compound} - ClogP_{compound\ without\ N\ substituent}$ and $\pi_{phenyl}$ was determined from $ClogP_{compound} - ClogP_{compound\ without\ phenyl\ ring}$. The following descriptors were determined using Molecular

Modeling Pro Plus Version 6.2.3 (Chemsoftware). Sterimol parameters (L1,B1,B5), highest occupied molecular orbital (HOMO), lowest unoccupied molecular orbital (LUMO), dipole moment, Hammett constant ($\sigma_{aromatic}$) of the group on the 5-phenyl ring, inductive constant ($\sigma^*$) of the group attached to the indole nitrogen. In total, 20 descriptors were collected for each compound.

PCA and PLS were carried out on SIMCA-P+ version 11 (2005) (Umetrics AB, Umea, Sweden) with default settings. Descriptors were not transformed to logarithmic values for the analysis. Multiple linear regression and Pearson correlation analysis were carried out on SPSS for Windows 14.0 (SPSS Inc., Chicago, IL, USA).

CoMFA was performed with the Sybyl 7.0 molecular modeling software (Tripos Inc., St. Louis, MO). The compounds were built using fragments in the Sybyl database and geometry-optimized with the standard Tripos force field with a distance dependent dielectric function until a root mean square (rms) deviation of 0.001 kcal/mol Å was achieved. The partial atomic charges required for the electrostatic interaction were computed using the Gasteiger-Huckel method. The most active compound (cysmethynil, **1D**) was chosen as the template on which other molecules were aligned based on the shared indole ring. CoMFA steric and electrostatic interaction fields were calculated at each lattice intersection point of a regularly spaced grid of 2.0 Å. The grid pattern, generated automatically by the Sybyl/CoMFA routine, extended 4.0 Å units in X, Y and Z directions beyond the dimensions of each molecule. The steric term, which represented van der Waals (Lennard-Jones) interaction, and the coulombic term, which represented electrostatic interactions, were calculated with the standard Tripos force field. An sp$^3$ carbon atom with a van der Waals radius of 1.52 Å and +1.0 charge was used as the probe to calculate the steric

and electrostatic fields. Values of the steric and electrostatic fields were truncated at 30 kcal/mol.

## 2.3. RESULTS AND DISCUSSION

### 2.3.1. The indole database

The structures of the 72 compounds in the indole database are in **Table 2-1**. In this series, modifications centred on the substituted 5-phenyl ring and the side chain attached to the indole N. Eight different substituents were present on the 5-phenyl ring: *m*-methyl, *p*-methyl, *m*-fluoro, *o*-methoxy, *m*-ethoxy, *m*-chloro-*p*-fluoro, *m,m*-bistrifluoromethyl and *p*-phenoxy. Nine different side chains were attached to position 1: isobutyl, cyclopropylmethyl, *n*-hexyl, *n*-octyl, benzyl, 3'-trifluoromethylbenzyl, 2'-naphthylmethyl, 3'-phenoxypropyl and 4'-*tert*-butylbenzyl. There were also compounds with no subsituent on the indole N ($R_2 = H$).

**Table 2-1. Structures and experimental[a] $pIC_{50}$ values of compounds in database.**
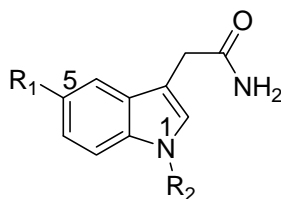


| No | $R_1$ | $R_2$ | $pIC_{50}$ |
|---|---|---|---|
| **1A1** | *m*-tolyl | isobutyl | 4.80 |
| **1B1** | *m*-tolyl | cyclopropylmethyl | 5.04 |
| **1C1** | *m*-tolyl | *n*-hexyl | 5.13 |
| **1D[b]** | *m*-tolyl | *n*-octyl | 5.68 |
| **1E** | *m*-tolyl | benzyl | 5.19 |
| **1G1** | *m*-tolyl | 3'-trifluoromethylbenzyl | 5.10 |
| **1H1** | *m*-tolyl | 2'-naphthylmethyl | 4.76 |
| **1I1** | *m*-tolyl | phenoxypropyl | 4.00 |
| **1J1** | *m*-tolyl | H | 4.00 |

**Table 2-1. (cont'd)**

| No | R$_1$ | R$_2$ | pIC$_{50}$ |
|---|---|---|---|
| **2A1** | *m*-fluorophenyl | isobutyl | 5.43 |
| **2B1** | *m*-fluorophenyl | cyclopropylmethyl | 4.83 |
| **2C1** | *m*-fluorophenyl | *n*-hexyl | 4.89 |
| **2D** | *m*-fluorophenyl | *n*-octyl | 5.62 |
| **2E** | *m*-fluorophenyl | benzyl | 4.97 |
| **2F1** | *m*-fluorophenyl | 4'-*tert*-butylbenzyl | 5.30 |
| **2G1** | *m*-fluorophenyl | 3'-trifluoromethylbenzyl | 5.24 |
| **2I1** | *m*-fluorophenyl | phenoxypropyl | 4.00 |
| **2J1** | *m*-fluorophenyl | H | 4.40 |
| **3A1** | *p*-tolyl | isobutyl | 5.01 |
| **3B1** | *p*-tolyl | cyclopropylmethyl | 4.95 |
| **3C1** | *p*-tolyl | *n*-hexyl | 5.13 |
| **3D** | *p*-tolyl | *n*-octyl | 4.97 |
| **3G1** | *p*-tolyl | 3'-trifluoromethylbenzyl | 5.19 |
| **3H1** | *p*-tolyl | 2'-naphthylmethyl | 4.71 |
| **3J1** | *p*-tolyl | H | 4.00 |
| **4A1** | *m*-ethoxyphenyl | isobutyl | 5.21 |
| **4B1** | *m*-ethoxyphenyl | cyclopropylmethyl | 5.19 |
| **4C1** | *m*-ethoxyphenyl | *n*-hexyl | 5.18 |
| **4D** | *m*-ethoxyphenyl | *n*-octyl | 5.66 |
| **4E** | *m*-ethoxyphenyl | benzyl | 5.11 |
| **4F1** | *m*-ethoxyphenyl | 4'-*tert*-butylbenzyl | 4.90 |
| **4G1** | *m*-ethoxyphenyl | 3'-trifluoromethylbenzyl | 5.31 |
| **4H1** | *m*-ethoxyphenyl | 2'-naphthylmethyl | 5.08 |
| **4I1** | *m*-ethoxyphenyl | phenoxypropyl | 4.00 |
| **4J1** | *m*-ethoxyphenyl | H | 4.00 |
| **5A1** | *o*-methoxyphenyl | isobutyl | 4.66 |
| **5B1** | *o*-methoxyphenyl | cyclopropylmethyl | 4.79 |
| **5C1** | *o*-methoxyphenyl | *n*-hexyl | 5.09 |
| **5D** | *o*-methoxyphenyl | *n*-octyl | 5.22 |
| **5E** | *o*-methoxyphenyl | benzyl | 4.88 |
| **5G1** | *o*-methoxyphenyl | 3'-trifluoromethylbenzyl | 5.21 |
| **5H1** | *o*-methoxyphenyl | 2'-naphthylmethyl | 4.53 |

**Table 2-1. (cont'd)**

| No | R$_1$ | R$_2$ | pIC$_{50}$ |
|---|---|---|---|
| **5I1** | *o*-methoxyphenyl | phenoxypropyl | 4.00 |
| **5J1** | *o*-methoxyphenyl | H | 4.00 |
| **6A1** | *m*-chloro-*p*-fluorophenyl | isobutyl | 5.04 |
| **6B1** | *m*-chloro-*p*-fluorophenyl | cyclopropylmethyl | 5.24 |
| **6C1** | *m*-chloro-*p*-fluorophenyl | *n*-hexyl | 5.03 |
| **6D** | *m*-chloro-*p*-fluorophenyl | *n*-octyl | 5.15 |
| **6E** | *m*-chloro-*p*-fluorophenyl | benzyl | 4.98 |
| **6F1** | *m*-chloro-*p*-fluorophenyl | 4'-*tert*-butylbenzyl | 5.12 |
| **6G1** | *m*-chloro-*p*-fluorophenyl | 3'-trifluoromethylbenzyl | 5.21 |
| **6I1** | *m*-chloro-*p*-fluorophenyl | phenoxypropyl | 4.00 |
| **6J1** | *m*-chloro-*p*-fluorophenyl | H | 4.00 |
| **7A1** | *m,m*-bis(trifluoromethyl)phenyl | isobutyl | 4.59 |
| **7B1** | *m,m*-bis(trifluoromethyl)phenyl | cyclopropylmethyl | 4.00 |
| **7D** | *m,m*-bis(trifluoromethyl)phenyl | *n*-octyl | 4.43 |
| **7E** | *m,m*-bis(trifluoromethyl)phenyl | benzyl | 4.48 |
| **7F1** | *m,m*-bis(trifluoromethyl)phenyl | 4'-*tert*-butylbenzyl | 4.53 |
| **7G1** | *m,m*-bis(trifluoromethyl)phenyl | 3'-trifluoromethylbenzyl | 4.00 |
| **7H1** | *m, m*-bis(trifluoromethyl)phenyl | 2'-naphthylmethyl | 4.00 |
| **7I1** | *m,m*-bis(trifluoromethyl)phenyl | phenoxypropyl | 4.00 |
| **7J1** | *m,m*-bis(trifluoromethyl)phenyl | H | 4.00 |
| **8A1** | *p*-phenoxyphenyl | isobutyl | 5.23 |
| **8B1** | *p*-phenoxyphenyl | cyclopropylmethyl | 4.00 |
| **8C1** | *p*-phenoxyphenyl | *n*-hexyl | 4.00 |
| **8E** | *p*-phenoxyphenyl | benzyl | 5.03 |
| **8G1** | *p*-phenoxyphenyl | 3'-trifluoromethylbenzyl | 4.00 |
| **8H1** | *p*-phenoxyphenyl | 2'-naphthylmethyl | 4.00 |
| **8I1** | *p*-phenoxyphenyl | phenoxypropyl | 4.00 |
| **8J1** | *p*-phenoxyphenyl | H | 4.00 |

[a]Determination of pIC$_{50}$ (-log IC$_{50}$ where IC$_{50}$ is the concentration required to inhibit Icmt activity by 50%) was based on the method described by Baron et al.[67]

[b]Cysmethynil

Descriptors for size, lipophilicity, electronic and hydrogen bonding capacity were collected for each compound. Size parameters were given by area, volume, polar volume and Sterimol parameters. The Sterimol parameter L1 measured the substituent length along the axis formed by the bond between the substituent and the atom to which it was attached. B1 was the shortest radius perpendicular to the L1 axis and B5 was the longest possible radius perpendicular to L1.[72] Lipophilicity was reflected by ClogP and Hansch $\pi$ values for the phenyl substituent(s) ($\pi_{Ph}$) and the N substitutent ($\pi_N$). HOMO, LUMO, dipole moment, Hammett constants $\sigma_{aromatic}$ and $\sigma^*$ and molar refractivity were electronic properties. Molar refractivity was also indicative of the size of the molecule. Polar surface area provided an indication of hydrogen (H) bonding capacity. The descriptors for each compound are listed in **Table A1-1 of Appendix 1.**

The pIC$_{50}$ values for Icmt inhibition have been determined [67] and are provided in **Table 2-1**. In the original report, compounds with IC$_{50}$ values that exceeded 50 $\mu$M were not accurately determined. For the purpose of SAR analysis, three IC$_{50}$ values (50, 100 or 200 $\mu$M) were assigned in turn to these weak inhibitors and for each IC$_{50}$ value, a PCA/PLS model was generated. The best model evaluated on the basis of regression coefficient $R^2Y$ and cross-validated correlation coefficient ($Q^2$) was obtained when the IC$_{50}$ value was 100 $\mu$M. Therefore, this value was assigned to compounds that had IC$_{50}$ values greater than 50 $\mu$M.

### 2.3.2. Analysis by Projection Methods PCA and PLS

Projection methods are widely employed for establishing structure-activity relationships (SAR). The most commonly employed methods are PCA and PLS. In PCA, a large set of descriptors are reduced to a smaller set derived from the linear

combination of the original descriptors. These are called principal components (PC) and they summarize information residing in the original data so that relationships between compounds and descriptors may be analyzed.[73] PLS is a regression extension of PCA and serves to identify the principal components that concurrently explain the variance in the descriptor set and maximize their correlation to the dependent variable (biological activity). The advantage of PLS is that it still works in circumstances where the descriptors are highly correlated to each other or where there are missing descriptor values.

When the data set of 72 compounds and their descriptors were analyzed by PCA, a significant three-component model was obtained. The properties captured by these components explained 72 % ($R^2Y$ =0.72) and predicted 54 % ($Q^2$ = 0.54) of the structural differences in the compounds. The loading plot of the components was next examined to identify descriptors that contributed to each component. In terms of relevance to the model, the $1^{st}$ component is the most important, followed by the $2^{nd}$ and $3^{rd}$ component. Thus only the loading plot of the $1^{st}$ and $2^{nd}$ PC was examined (**Figure 2-1**). In this plot, descriptors located at the ends of the axes contributed significantly to the respective PC while those found near the origin made lesser contributions. In **Figure 2-1**, size (volume, area, molar refractivity, Sterimol parameters of the N-substituent) and lipophilicity (ClogP) were important contributors to the $1^{st}$ PC whereas HOMO and several characteristics of the substituted 5-phenyl ring, namely $\sigma_{Ph}$, $\pi_{Ph}$ and the Sterimol width parameter of the ring (PB1) contributed significantly to the $2^{nd}$ PC. Interestingly, the $1^{st}$ component encoded properties of the N-substituent whereas the $2^{nd}$ component received inputs from the substituted 5-phenyl ring. This may reflect the relative influence of the functionalities at the two

positions, with the N-substituent making a bigger contribution to the overall physicochemical profile of the compounds.
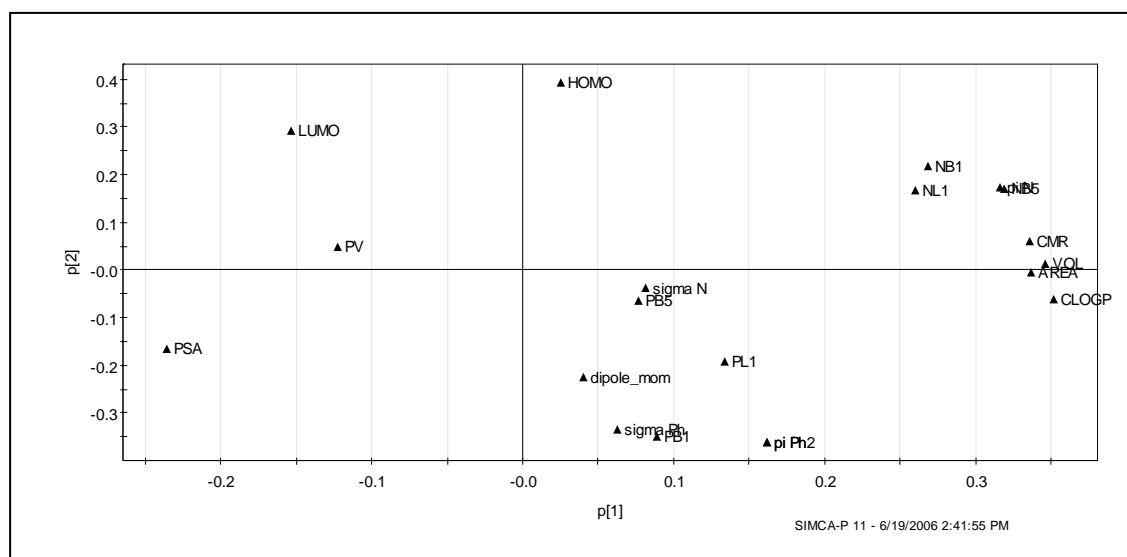


**Figure 2-1. Loading plot of first and second principal components (p[1], p[2]) of 72 compounds and 20 descriptors.**

Next the score plot of the two components was examined to see how the compounds were distributed with respect to the 1st and 2nd components (**Figure 2-2**). An obvious feature of this plot was the location of compounds without an N-substituent (codes ending in "**J1**") at the lower left quadrant of the plot (**Figure 2-2**, Box **I**), with many members found outside the Hotelling ellipse which demarcated the 95% confidence region for the model. The separation of the "**J1**" compounds from the rest implied that their properties were clearly different from the other compounds. Furthermore since compounds without an N-substituent were also the least active in their respective classes, it may imply that the properties captured by these principal components were relevant for Icmt inhibitory activity. Other than the "**J1**" compounds, the other compounds did not show obvious clustering based on the type of N-substituent. However, some compounds were found grouped together based on the substituted 5-phenyl ring, namely compounds with prefixes "**5**" and "**7**" (**Figure**

**2-2**, Box **II** and **III**). There was no obvious reason why they were found together. The

compounds prefixed with **7** had two trifluoromethyl ($CF_3$) groups on the phenyl ring

and they were among the weakest compounds in the series. Coincidentally they were

also found near the edge of the Hotelling ellipse. The compounds prefixed with "**5**"

had an *o*-methoxy group on the phenyl ring and were distinguished by being the only

compounds in the library with an ortho substituent on the phenyl ring.
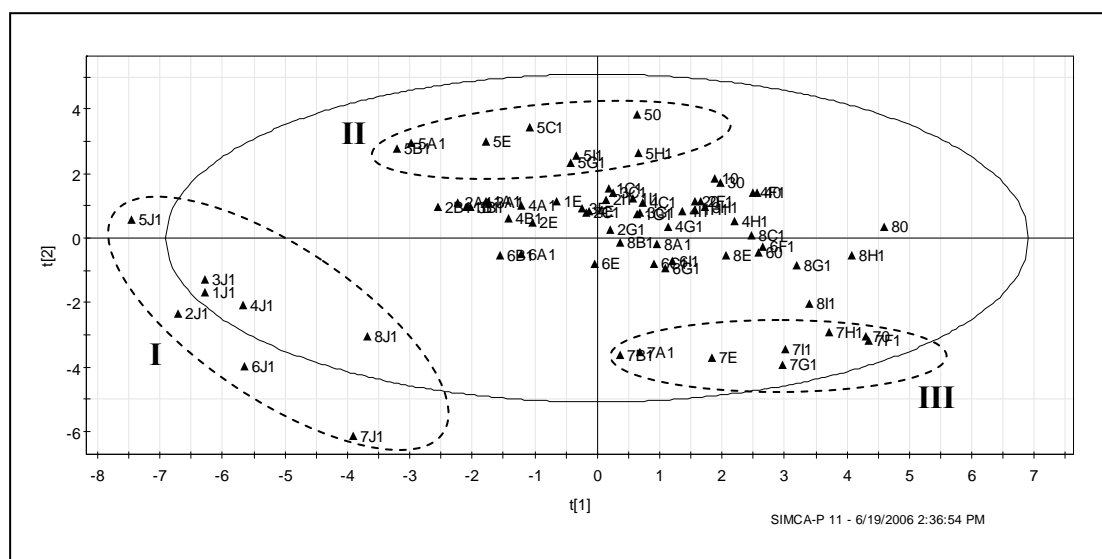


**Figure 2-2. Score plot of principal components t1 versus t2 for compounds (n = 72, 20 descriptors).** Compounds are identified in Table 1. The solid line ellipse corresponds to the confidence region based on Hotelling $T^2$ (0.05). I) Compounds identified by '5' have *o*-methoxyphenyl group on the phenyl ring, II) compounds identified by '7' have *m,m*-bis(trifluoromethyl) groups on the phenyl ring, and III) compounds identified by 'J1' have no substituent on the indole N. These sets of compounds are isolated from the rest.

Next, PLS was used to develop a projection model for predicting biological

activity from the principal components. First, a model was generated with all

compounds and parameters. It was only satisfactory, with $R^2Y = 0.68$ and $Q^2 = 0.49$.

The model was improved by removing 2 compounds that had large differences

between the observed and predicted values (compounds **3E** and **8D**) and descriptors

that did not make significant contributions to activity (dipole moment, $\sigma_{nitrogen}$,

$\sigma_{phenyl}$) or were duplicated by existing descriptors (volume, surface area, $\pi_{Ph}$). The resulting model (A) had $R^2Y$ of 0.72 and $Q^2$ of 0.60. A test set of 12 compounds was arbitrarily selected from the 70 compounds and used to predict the activities of the remaining compounds. The level of predictability, as measured by the root mean square of prediction (RMSEP), was satisfactory at 0.48. The coefficient plot of the model was then examined (**Figure 2-3**). This plot provided information on the descriptors that made the largest contribution to activity (reflected by the length of the bar) as well as how it affected activity (directly correlated or inversely correlated). As seen from **Figure 2-3**, the model identified the following descriptors (in order of decreasing importance) as contributors to activity: polar surface area, polar volume > lipophilicity of phenyl ring ($\pi_{Ph}$) > width parameters B1, B5 of the N substituent and $\pi_N$ > Sterimol L1 of N substituent. The plot also illustrated the contrasting requirements of the N and phenyl ring substituents, namely that for good Icmt inhibitory activity, the N-substitutent must be lipophilic and bulky (large width dimensions) whereas the substituted phenyl ring should have substituents that were less lipophilic and had sterically smaller dimensions.
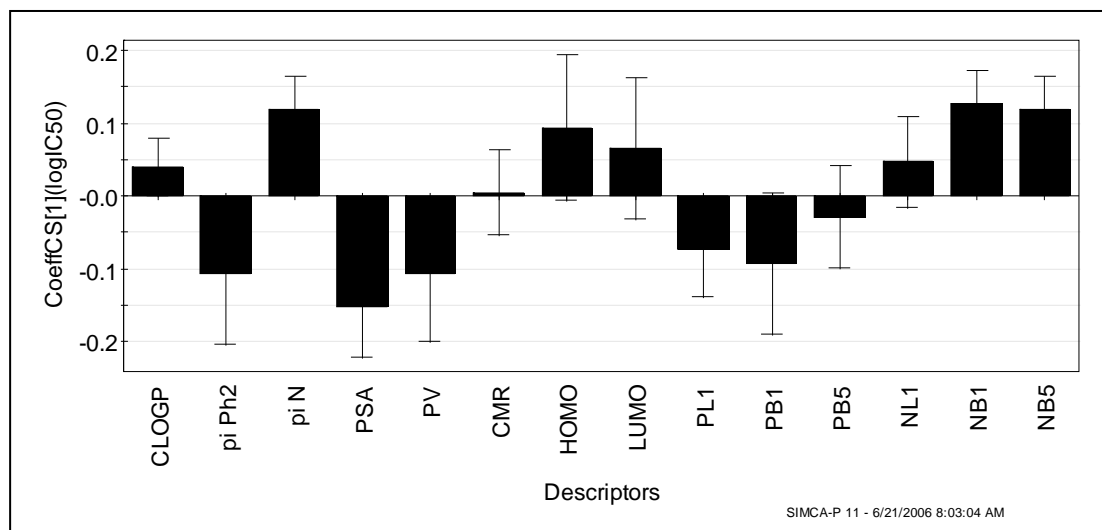
**Figure 2-3. Coefficient plot for PLS model derived from 70 compounds and 14 descriptors, based on the first component.** Descriptors with positive coefficients are directly related to activity, while those with negative coefficients are inversely related. Magnitude of parameter indicates its relative contribution to the model. CLOGP = computed log P; pi Ph2 = Hansch constant $\pi$ of substituted phenyl ring; pi N = Hansch constant $\pi$ of N-substituent; PSA = polar surface area, PV = polar volume; CMR = molar refractivity; HOMO = highest occupied molecular orbital; LUMO = lowest unoccupied molecular orbital; PL1, PB1, PB5 = sterimol parameters of phenyl ring; NL1, NB1, NB5 = sterimol parameters of N-substituent.

### 2.3.3. Analysis by Multiple Linear Regression

The dataset used to derive the PLS model was analyzed by stepwise multiple linear regression (MLR)[72] to determine which descriptors would be identified as the main contributors to activity and if there was any overlap with those determined by PLS. The analysis yielded **Equation 2-1** which identified polar surface area (PSA), polar volume (PV), Sterimol parameter B1 of the substituted phenyl ring and lipophilicity contribution of the substituted phenyl ring ($\pi_{Ph}$) as important contributors to activity.

$$pIC_{50} = 10.23 \,(\pm 0.55) - 0.016 \,(\pm 0.003)* \, PSA - 0.010 \,(\pm 0.002)* \, PV \quad \ldots \textbf{(2-1)}$$
$$- 0.835 \,(\pm 0.195) * PB1 - 0.190 \,(\pm 0.056) * \pi_{Ph}.$$

$n = 70$, $r = 0.811$, $r^2_{adjusted} = 0.636$, standard error of estimate = 0.32, F = 30.69.

When the coefficients in **Equation 2-1** were standardized, the relative contribution of each parameter was PSA (most important) > PV > PB1 > $\pi_{Ph}$. The 1$^{st}$ three parameters were significant at $p < 0.001$ whereas $\pi_{Ph}$ was significant at $p = 0.001$.

When the models derived from MLR and PLS were compared, the following observations were made. First, both models identified polar surface area, polar volume and $\pi_{Ph}$ as important contributors to activity. That these properties were inversely correlated to activity was also recognized in both models. Second, the Sterimol and lipophilicity parameters of the N-substituent were conspicuously absent from the MLR equation although they were important in the PLS model. A possible explanation may be that these Sterimol parameters were captured by polar surface area (PSA) to which they were significantly correlated as seen from the Pearson correlation coefficients (**Table 2-2**). Third, the breadth parameter of the phenyl ring (PB1) was recognized in **Equation 2-1** but not the PLS model. PB1 was significantly correlated to HOMO and LUMO ($p < 0.0001$) but these descriptors were not identified by the PLS model.

**Table 2-2. Pearson Correlation Coefficients of Parameters employed in developing Equation (2-1)**

| | | PSA | PV | πN | πPh | NL1 | NB1 | NB5 | PL1 | PB1 | PB5 | HOMO | LUMO | CLOGP | CMR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PSA | Pearson Correlation | 1.000 | | | | | | | | | | | | | |
| | Sig. (1-tailed) | 0.000 | | | | | | | | | | | | | |
| PV | Pearson Correlation | 0.309** | 1.000 | | | | | | | | | | | | |
| | Sig. (1-tailed) | 0.005 | 0.000 | | | | | | | | | | | | |
| πN | Pearson Correlation | -0.713** | -0.335** | 1.000 | | | | | | | | | | | |
| | Sig. (1-tailed) | 0.000 | 0.002 | 0.000 | | | | | | | | | | | |
| πPh | Pearson Correlation | 0.039 | -0.253* | -0.030 | 1.000 | | | | | | | | | | |
| | Sig. (1-tailed) | 0.374 | 0.017 | 0.402 | 0.000 | | | | | | | | | | |
| NL1 | Pearson Correlation | -0.473** | -0.288** | 0.781** | -0.048 | 1.000 | | | | | | | | | |
| | Sig. (1-tailed) | 0.000 | 0.008 | 0.000 | 0.347 | 0.000 | | | | | | | | | |
| NB1 | Pearson Correlation | -0.918** | -0.185 | 0.734** | -0.010 | 0.631** | 1.000 | | | | | | | | |
| | Sig. (1-tailed) | 0.000 | 0.062 | 0.000 | 0.469 | 0.000 | 0.000 | | | | | | | | |
| NB5 | Pearson Correlation | -0.802** | -0.259* | 0.959** | -0.009 | 0.658** | 0.803** | 1.000 | | | | | | | |
| | Sig. (1-tailed) | 0.000 | 0.015 | 0.000 | 0.471 | 0.000 | 0.000 | 0.000 | | | | | | | |
| PL1 | Pearson Correlation | 0.143 | -0.125 | -0.039 | 0.813** | -0.008 | -0.002 | -0.026 | 1.000 | | | | | | |
| | Sig. (1-tailed) | 0.119 | 0.152 | 0.375 | 0.000 | 0.472 | 0.494 | 0.414 | 0.000 | | | | | | |
| PB1 | Pearson Correlation | -0.145 | -0.241* | 0.026 | 0.499** | 0.003 | 0.001 | 0.034 | 0.029 | 1.000 | | | | | |
| | Sig. (1-tailed) | 0.116 | 0.022 | 0.416 | 0.000 | 0.489 | 0.498 | 0.391 | 0.406 | 0.000 | | | | | |
| PB5 | Pearson Correlation | 0.220* | 0.179 | -0.003 | 0.281** | -0.041 | 0.009 | 0.021 | 0.271* | -0.097 | 1.000 | | | | |
| | Sig. (1-tailed) | 0.034 | 0.069 | 0.492 | 0.009 | 0.370 | 0.469 | 0.433 | 0.012 | 0.213 | 0.000 | | | | |
| HOMO | Pearson Correlation | -0.217* | 0.125 | 0.191 | -0.350** | 0.179 | 0.381** | 0.206* | 0.011 | -0.616** | 0.072 | 1.000 | | | |
| | Sig. (1-tailed) | 0.037 | 0.153 | 0.058 | 0.002 | 0.071 | 0.001 | 0.045 | 0.463 | 0.000 | 0.278 | 0.000 | | | |
| LUMO | Pearson Correlation | 0.220* | 0.179 | -0.263* | -0.442** | -0.100 | -0.084 | -0.309** | -0.118 | -0.613** | 0.019 | 0.574** | 1.000 | | |
| | Sig. (1-tailed) | 0.034 | 0.071 | 0.014 | 0.000 | 0.208 | 0.245 | 0.005 | 0.167 | 0.000 | 0.437 | 0.000 | 0.000 | | |
| CLOGP | Pearson Correlation | -0.578** | -0.425** | 0.824** | 0.542** | 0.630** | 0.613** | 0.802** | 0.428** | 0.305** | 0.158 | -0.038 | -0.471** | 1.000 | |
| | Sig. (1-tailed) | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.005 | 0.096 | 0.379 | 0.000 | 0.000 | |
| CMR | Pearson Correlation | -0.479** | -0.117 | 0.779** | 0.332** | 0.625** | 0.638** | 0.818** | 0.371** | 0.044 | 0.321** | 0.202* | -0.350** | 0.843** | 1.000 |
| | Sig. (1-tailed) | 0.000 | 0.168 | 0.000 | 0.002 | 0.000 | 0.000 | 0.000 | 0.001 | 0.359 | 0.003 | 0.048 | 0.002 | 0.000 | 0.000 |

Correlation is significant at the 0.01 (**) and 0.05 (*) levels (1-tailed) respectively.

Taken together, the results of the PLS and MLR analyses supported the requirement for a small polar surface area for activity. Polar surface area was strongly correlated to the dimensions (NL1, NB1, NB5) and lipophilicity ($\pi_N$) of the N-substituent as seen from their Pearson correlation coefficients in **Table 2-2**. In contrast, the properties (PL1, PB1, PB5, $\pi_{Ph}$) of the group on 5-phenyl were poorly correlated to polar surface area. Thus, compounds with small polar surface areas were likely to be those with lipophilic N-substituents. As for the dimensions of the N-substituent, **Figure 2-3** suggests that the breadth dimensions (NB1, NB5) of the substituent had a greater role than its length (NL1). This would explain the consistently poor activity of compounds with a phenoxypropyl side chain attached to the indole N. This side chain was as long as the *n*-octyl side chain but lacked its conformational flexibility. Thus substituents that were conformationally restricted may fare less well than those of similar length but greater flexibility.

The PLS and MLR analyses also supported the requirement for a less lipophilic and sterically more compact substituted phenyl ring at position 5. It was noted that phenyl rings substituted with *m,m*-bistrifluoromethyl (compounds prefixed with "**7**") and *p*-phenoxy (compounds prefixed with "**8**") were generally poor inhibitors possibly because these subsituents were more lipophilic and bulkier than substituents like methoxy, methyl, fluoro or ethoxy.

### 2.3.4. Comparative Molecular Field Analysis (CoMFA)

From the preceding sections, it was deduced that there were different requirements for activity at position 1 (indole N) and the substituted 5-phenyl ring. At position 1, the substituent should be lipophilic and conformationally flexible whereas the phenyl ring at position 5 should be substituted with less lipophilic and

less bulky subsituents. Support for this finding was also found from comparative molecular field analysis (CoMFA).
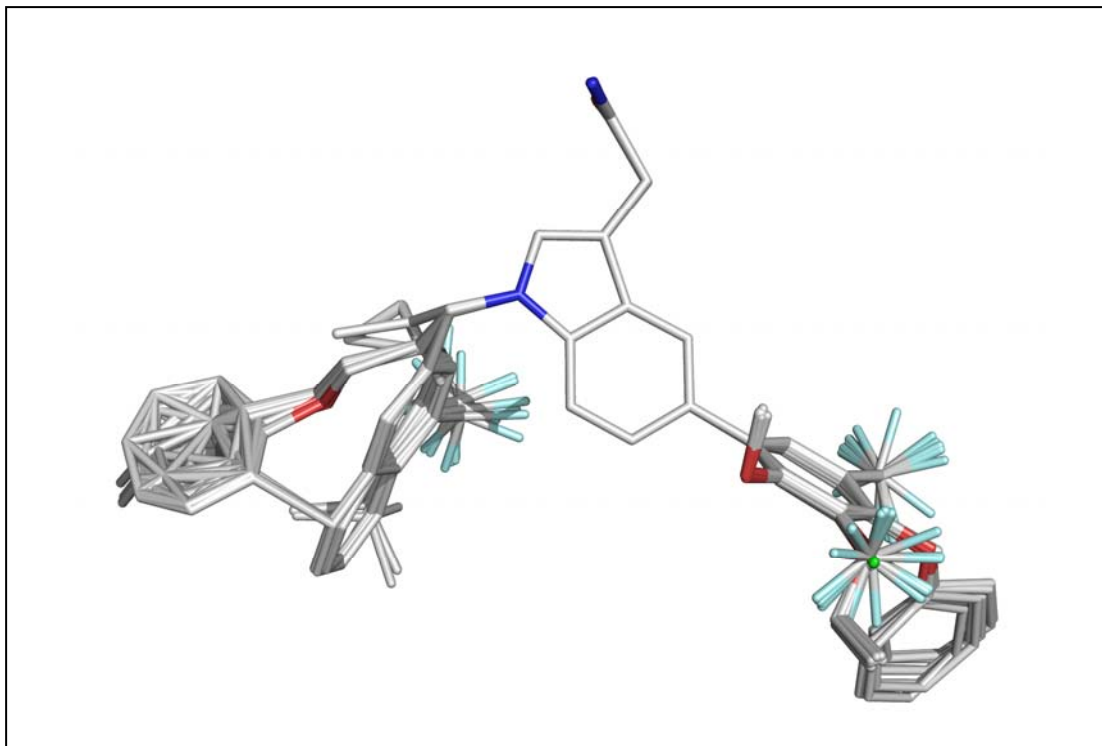


**Figure 2-4. Alignment of the 72 compounds.** Figure created with PyMOL (DeLano Scientific, Palo Alto, CA, USA).

CoMFA is a widely employed tool for the analysis of 3-dimensional quantitative structure-activity relationships (3D QSAR).[74] 3D-QSAR is based on the premise that the properties of a molecule are related to its 3D structure, namely its overall size, shape and electronic properties. In conventional QSAR, the contributions of individual moieties or substituents are summed up to give the final property "value" of the compound. In CoMFA, steric and electrostatic interaction fields are measured around a series of aligned structures (**Figure 2-4**) and these are related to the activity of the compounds. By comparing the steric and electrostatic fields for different molecules against their activity, it is possible to define interactions that are desirable or undesirable and to visualize these interactions. Several caveats should be

noted when employing CoMFA. With reference to the present exercise, it was assumed that the forcefield minimized conformation was the active conformation, since the latter had not been established for the compounds under investigation. Second, the molecules were aligned based on the indole core which was common to all compounds. The preferred method is to use the pharmacophore as a basis of alignment, but since this has not been established, it could not be applied.

For alignment, three compounds (**8A1**, **3E** ,**8D**) were omitted because of their large residual values. The remaining 69 compounds were divided into a training set (n = 56) and a test set (n = 13). Compounds in these sets covered the same range of activity and were well represented in terms of the type of substituents on the indole nitrogen and phenyl rings. The statistical parameters of the best CoMFA model derived from the training set is given in **Table 2-3**. The model had a favorable cross-validated $q^2$ of 0.646 (a $q^2$ value greater than 0.3 is considered significant) and a non cross-validated $r^2$ of 0.868 with a low standard error of estimate (SEE) of 0.209. The training set predicted the activity of the test set compounds with a satsifactory $r^2_{pred}$ of 0.601.

**Table 2-3. Summary of CoMFA Analysis of Training Set (n = 56)**

| | |
|---|---|
| Cross validated correlation coefficient, $q^2$ | 0.646 |
| Standard error of prediction (SEP) | 0.342 |
| Number of Components[a] | 7 |
| Non cross-validated correlation coefficient, $r^2$ | 0.868 |
| Standard error of estimate (SEE) | 0.209 |
| F value | 44.91 |
| Field Contributions | |
|     Steric | 0.68 |
|     Electrostatic | 0.32 |
| $r^2_{pred}$ [b] | 0.601 |

[a] Optimum number of components obtained from cross-validated PLS analysis. This number of components is also used in the final non cross-validated analysis.

[b] Predictive $r^2$ is based only on the 13 compounds not included in the training set. It is determined from $r^2_{pred} = [SD - PRESS] / SD$ where SD = sum of the squared deviations between $IC_{50}$ values of compounds in test set and mean $IC_{50}$ values of training set molecules, and PRESS is sum of the squared deviations between predicted and actual $IC_{50}$ values for every member in the test set.

In CoMFA, the steric contours were color-coded green and yellow. Green indicated regions where steric bulk was favoured for activity whereas yellow indicated regions where steric bulk was not favoured for activity. The steric contours of the model displayed a swathe of green at the proximal and distal ends of the N substituent. A yellow patch was interspersed between these green zones (**Figure 2-5**). The inference was that there were two optimal lengths of the N substituent corresponding to a short group that would occupy the proximal green patch and a longer group that filled the distal green patch. The 3'-trifluromethylbenzyl and *n*-octyl groups were examples of optimal short and long substituents. On the other hand, the phenoxypropyl group extended beyond the proximal green zone and occupied the sterically disfavoured yellow zone. Interestingly, the presence of two optimal lengths for the N-substituent was revealed only in CoMFA and not the other methods.
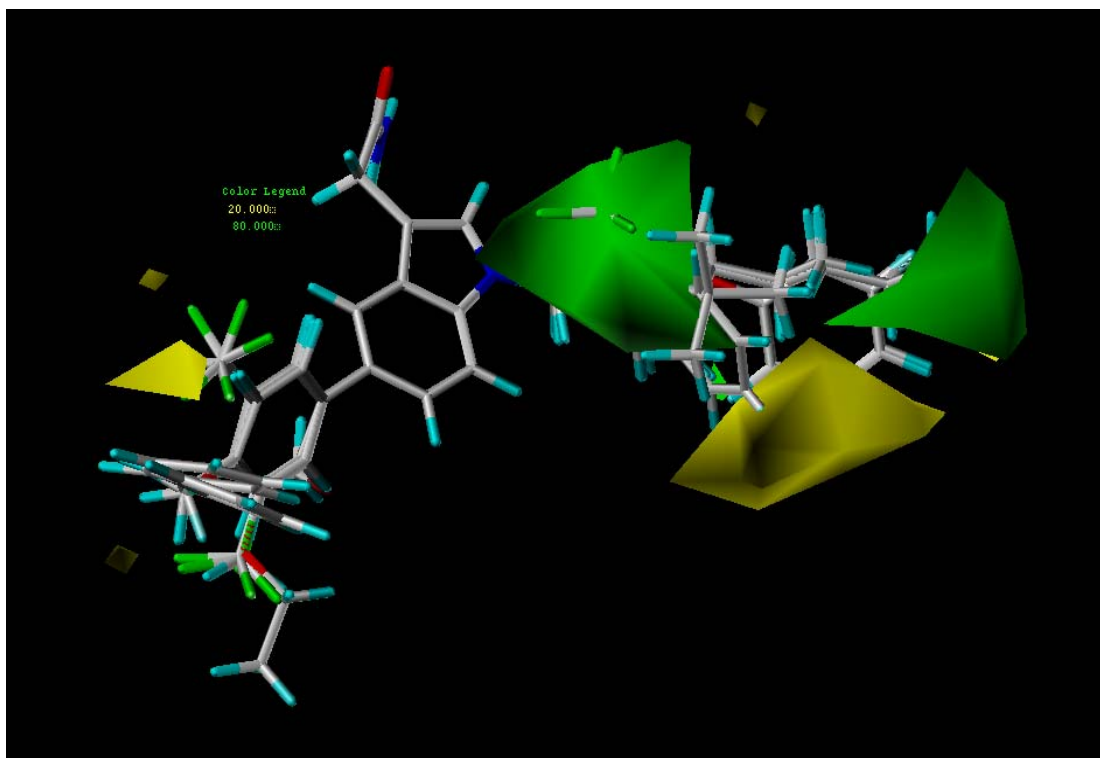
**Figure 2-5. Steric map from the CoMFA model showing the alignment based on indole ring as shown in Figure 2-4.** Green contours represent areas where steric bulk will enhance activity, and yellow contours highlight areas which should be kept unoccupied for increased activity.

As for the 5-phenyl ring, only sparse yellow contours were found in its vicinity. The inference was that the steric nature of the substituted phenyl ring was of lesser importance, in spite of deductions from PLS and MLR that less bulk was preferred at this position.

The electrostatic field contribution (32 %) to model B was far less than the steric field contribution (68 %) (**Table 2-3**), which agreed with earlier SAR findings that proposed a greater steric influence on activity. Not unexpectedly, the electrostatic field contours of model B were sparsely distributed at both positions (**Figure 2-6**). Electrostatic field contours were color-coded red (more negative charge was favoured) and blue (more positive charge was favoured).
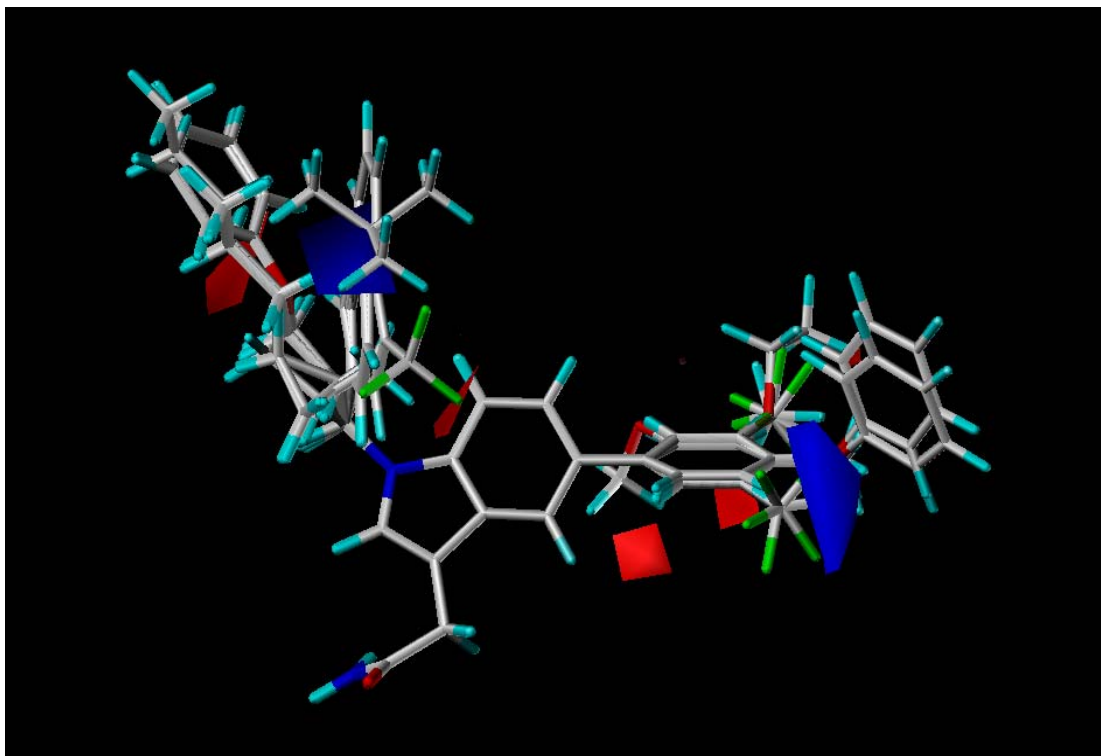
**Figure 2-6. Electrostatic map from the CoMFA model showing the same alignment as in Figure 2-4.** Blue contours represent regions where an increase in positive charge will enhance activity, and red contours highlight areas where more negative charge is favoured.

## 2.4. CONCLUSIONS

Distinct approaches were used to analyze the structure-activity relationship of the library of indoleacetamides from which cysmethynil was identified. The approaches were complementary and served to give a definitive picture of the structural requirements at the indole N and the substituted phenyl ring. The main findings were:

(i) Steric factors played a greater role than electrostatic factors in determining activity.

(ii) There were contrasting steric and lipophilic requirements at the two positions. The side chain attached to the indole N should be lipophilic and possess bulk as implied from the Sterimol width parameters (B1 and B5). The latter also implied that flexible side chains with a greater "sweep" of conformational space were

preferred to those that were less flexible. In contrast, the group(s) on the 5-phenyl ring should be of limited lipophilicity and bulkiness.

(iii) There were two optimal lengths for the side chain attached to the indole N: a short chain typified by 3'-trifluoromethylbenzyl and a longer chain that corresponded to *n*-octyl. Side chains that fell between these two optimal lengths were likely to give rise to poor activity.

These findings contribute to the definition of the pharmacophore for Icmt inhibitory activity but in its present state, but the study was incomplete because only modifications at two positions could be considered due to availability of data. A more comprehensive picture would emerge when the SAR of the acetamide side chain was known (**Chapter 4**). In any case, the study provided useful guidelines for structural modifications at the indole N and the substituted 5-phenyl ring.

# CHAPTER 3: DESIGN AND SYNTHESIS OF CYSMETHYNIL ANALOGUES

## 3.1. INTRODUCTION

The design and synthesis of cysmethynil analogues are described in this chapter. Modifications to the indole N and the 5-phenyl ring were directed by the SAR described in **Chapter 2**. Modifications to the acetamide side chain at position 3 of cysmethynil were guided by conventional analogue design approaches which are elaborated in the following paragraphs. The routes for the synthesis of target compounds are also described in this chapter.

## 3.2. RATIONALE OF DESIGN

### 3.2.1. Drug-like character of synthesized compounds

The term "drug-like" refers to those properties of a target compound that will favour its transition to a successful therapeutic agent. These are properties related to the pharmacokinetic (absorption, distribution, metabolism, excretion) profile of the compound and its toxicological potential. Awareness of the role played by drug-likeness in the development of new chemical entities was aroused by the seminal publication of Lipinski and coworkers[69] which laid the basis for the Lipinski's Rule of Five (Ro5). Briefly, these rules stated that a target compound was likely to have poor absorption or permeation if it had more than 5 hydrogen (H) bond donors (expressed as the sum of all OH and NH groups), more than 10 H bond acceptors (expressed as the sum of all O and N atoms), Log P (where P is partition coefficient) that exceeded 5 and a molecular weight that was greater than 500. These rules did not apply to compounds that were substrates of biological transporters such as P-glycoprotein. Moreover, compounds that violated one or more conditions need not necessarily be

penalized with poor absorption because the number of rules broken and the degree to which they crossed threshold values should be taken into account. Other guidelines for deducing drug-like properties of a compound followed in the wake of the Lipinski Rules. One of the better known additions was that proposed by Veber and co-workers.[70] Their guidelines, based on more than 1100 drug candidates tested in rats, proposed that molecular flexibility ($\leq$ 10 rotatable bonds) and polar surface area ($\leq$ 140 $\text{\AA}^2$ or $\leq$ 12 total H bond acceptors and donors) were important determinants of oral bioavailabilty. Unlike the Lipinski Rules, there was no mention of molecular weight.

The drug-likeness of cysmethynil was assessed by the guidelines given by Lipinski and Veber. Cysmethynil has a molecular weight of 376.53, 2 H bond donor groups, 3 H bond acceptor groups, 10 rotatable bonds and a ClogP of 7.0 (**Figure 3-1**). The main concern here was the high partition coefficient (ClogP = 7) and number of rotatable bonds that was close to the stated limit. However, the Lipinski Rules applied a simple counting of number of O and N atoms as a means of identifying H bond acceptors but based on chemical principles, the indole and amide N atoms should not be recognized as H bond acceptors because their lone pairs are delocalized and not available for H bonding. Hence cysmethynil has very limited H bonding capacity and is likely to have poor water solubility. Thus, one objective of this investigation was to propose less lipophilic/more soluble cysmethynil analogues for synthesis and to determine how far this could be achieved without compromising biological activity.
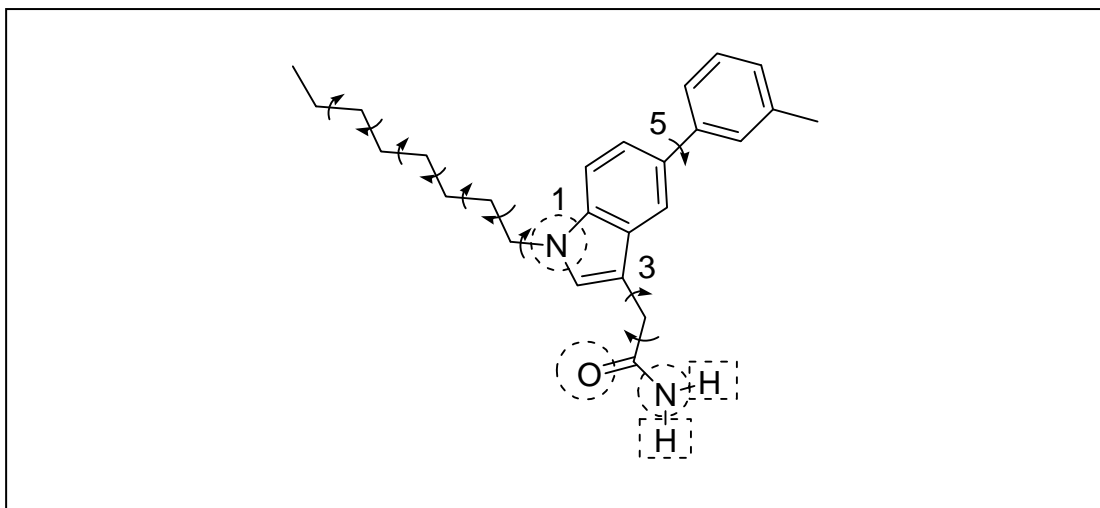
**Figure 3-1. Evaluation of drug-like character of cysmethynil.** Arrow = rotatable bond; circle = H bond acceptor; box = H bond donor. ClogP = 7.0 as determined by ChemDraw Ultra, Version 10, Cambridgesoft.


### 3.2.2. Modifications at position 5

In the original indole library, the 5-phenyl ring was substituted with 8 different groups (*m*-methyl, *p*-methyl, *m*-fluoro, *o*-methoxy, *m*-ethoxy, *m*-chloro-*p*-fluoro, *m,m*-bistrifluoromethyl, *p*-phenoxy) and the SAR (**Chapter 2**) proposed that good activity was associated with small and not overly lipophilic groups. With this in mind, the following modifications were proposed for the 5- phenyl ring:

i) Introduction of a *m*-methoxy substituent: This compound had a predicted activity, $pIC_{50}$ value of 5.57 based on equation (**1**) (**Section 2.3.3**) and 5.52 based on the CoMFA model (**Section 2.3.4**) which were comparable to cysmethynil ($pIC_{50}$ value of 5.72 and 5.32 respectively). Moreover, compounds with *o*-methoxy and *m*-ethoxy on the 5-phenyl ring were already represented in the indole library and found to have good activities.

ii) Synthesis of *o*, *m* and *p*-methyl derivatives: The methyl substituent had the desired features of lipophilicity and size that were required for activity. Except for *o*-methyl, the regiosomeric *m*- and *p*-methyl were represented in the indole library

(**Table 2-1** of **Chapter 2**) and they were generally of equivalent activity. The inclusion of the *o*-isomer would complete the picture of how substitution at different positions of 5-phenyl affected activity.

iii) Omission of the phenyl ring: The omission of the 5-phenyl ring of cysmethynil reduced ClogP to 4.6. In comparison, omission of the *n*-octyl side chain of cysmethynil caused a more drastic reduction in ClogP to 2.8. As seen from the indole library, compounds without the *n*-octyl side chain were inactive regardless of the substituted 5-phenyl ring. The removal of the *n*-octyl side chain might have caused too large a fall in lipophilicity, resulting in complete loss of activity. A lesser decrease in lipophilicity brought about by omitting the phenyl ring could have a more favourable effect on activity.

### 3.2.3. Modifications at position 1

The SAR described in **Chapter 2** indicated that the side chain at this position should be lipophilic and flexible. Two optimal lengths were proposed, corresponding to the *m*-trifluoromethylbenzyl and *n*-octyl moieties. Measurement of the distance from the indole N to the most distal non-H atom in these side chains showed that they corresponded to 6.7 Å and 10.5 Å respectively (**Figure 3-2**). The equivalent distance in the *p*-phenoxyphenyl side chain (associated with low activity) was 8.6 Å and this led to the proposal that side chains of intermediate lengths (like *p*-phenoxyphenyl) would be less active. In order to verify these claims, two side chains were introduced at the indole N of cysmethynil: isoprenyl (3-methylbut-2-ene) and geranyl (3,7-dimethylocta-2,6-diene) (**Figure 3-2**).
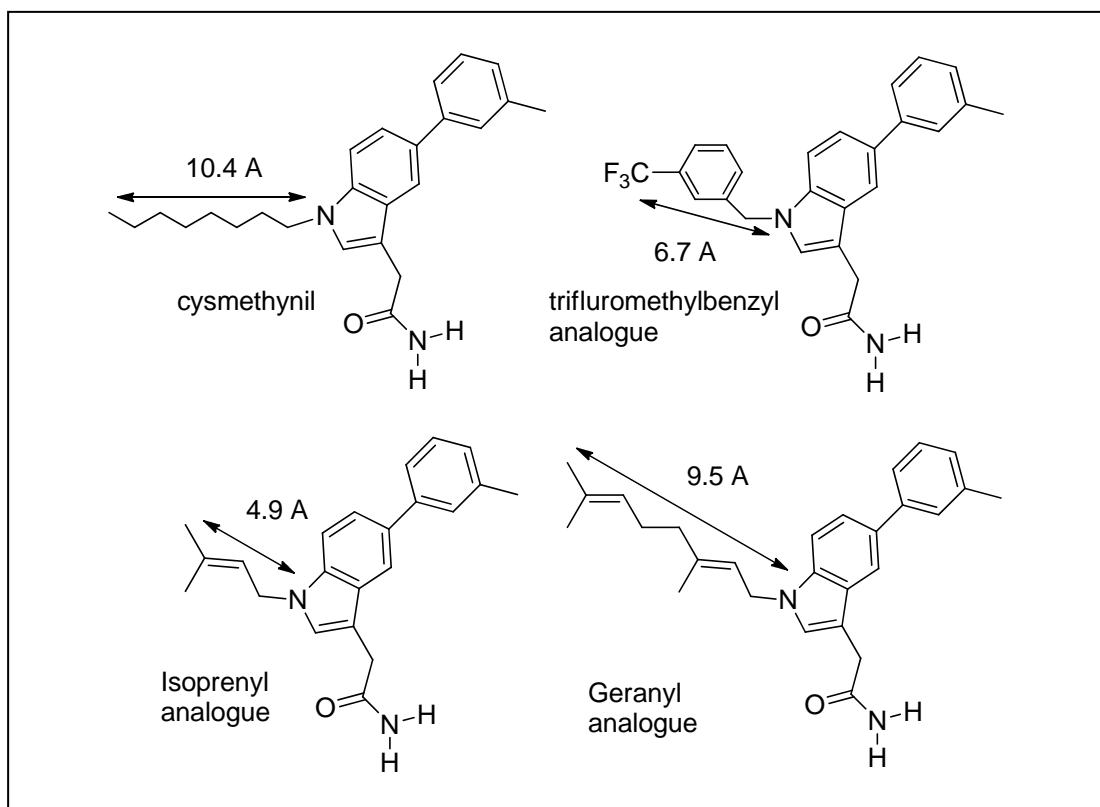
**Figure 3-2. Length of side chains at position 1 in cysmethynil and its m-trifluromethylbenzyl, isoprenyl and geranyl analogues.** Measurements were made on forcefield minimized structures on MOE (Molecular Operating Environment 2008.10, Chemical Computing Group).

If the SAR held true, then the estimated chain length of the isoprenyl analogue suggested it should be active. The outcome of the geranyl analogue was uncertain as its length was between that of *n*-octyl and *p*-phenoxyphenyl. The synthesis and evaluation of these compounds would provide the necessary verification.

### 3.2.4. Modification of the acetamide side chain at position 3

No modification has been reported at position 3 of cysmethynil. As the crystal structure of Icmt was not available, a structure-based approach to SAR could not be pursued. Thus modifications were based on conventional analogue-based methods as discussed in the following paragraph.

i) Omission of the acetamide side chain: The acetamide side chain was the only polar functionality in cysmethynil. It provided 2 H bond donor groups (-NH$_2$) and one H bond acceptor group (carbonyl oxygen of amide). Removing the acetamide side chain would give a more lipophilic compound (ClogP of 8.7) with no H bond capability. Its activity would provide useful information on the importance of these features for activity.

ii) Homologation: The carbon side chain linking the primary amide moiety to the indole ring was lengthened or shortened by one carbon atom. These changes affected lipophilicity to a small extent but had a greater effect on flexibility, especially that of the shorter homologue. As seen in **Figure 3-3**, the primary amide is now directly linked to the ring and electron delocalization from the indole N to the amide carbonyl would restrict the flexibility of the side chain.
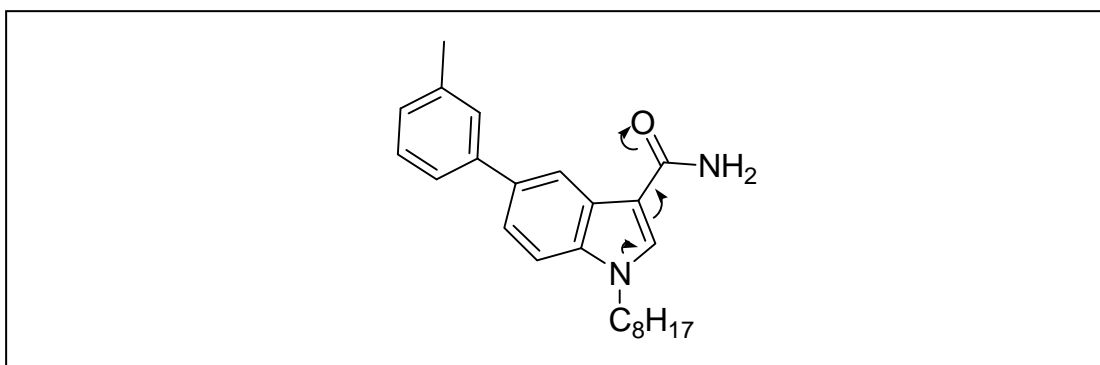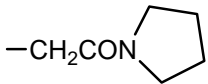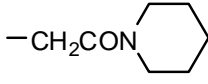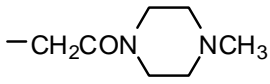


**Figure 3-3. Restricted flexibility of the C-3 side chain in the shortened homologue of cysmethynil.**

iii) Conversion to tertiary and secondary amides: The purpose of this modification was to assess the role of H bond donor groups for activity. The tertiary amide had no H bond donor capability while the secondary amide retained one H bond donor group. Lipophilicity would be expected to increase as predicted from the ClogP values, with the likely exception of the *N*-methylpiperazine analogue, as the non-amide N of this side chain would be protonated at physiological pH (**Table 3-1**).

**Table 3-1. ClogP[a] of secondary and tertiary amide analogues of cysmethynil[b]**

| Side chain at C-3 | ClogP | Side Chain at C-3 | ClogP |
|---|---|---|---|
| -CH$_2$CONH$_2$ (Cysmethynil) | 7.0 | —CH$_2$CON (pyrrolidine) | 9.7 |
| -CH$_2$CONHCH$_3$ | 7.2 | | |
| -CH$_2$CON(CH$_3$)$_2$ | 7.5 | —CH$_2$CON (piperidine) | 10.3 |
| -CH$_2$CON(C$_2$H$_5$)$_2$ | 8.6 | | |
| | | —CH$_2$CON NCH$_3$ | 9.3[c] |

[a] Determined with ChemDraw Ultra Version 10.0, CambridgeSoft.
[b] Sidechain attached to position 3, **Figure 3-1**
[c] N-CH$_3$ of ring is protonated at physiological pH.

iv) Conversion to amines: Replacing the acetamide functionality with an amine had the effect of reducing the lipophilic character of the compound because the amino group would be ionized at pH 7.4. In its ionized state, the amino group would not be a H bond acceptor. If the amine is tertiary in nature, H bond donor properties would be absent as well. The amino side chains targeted for synthesis are given in **Figure 3-4**.
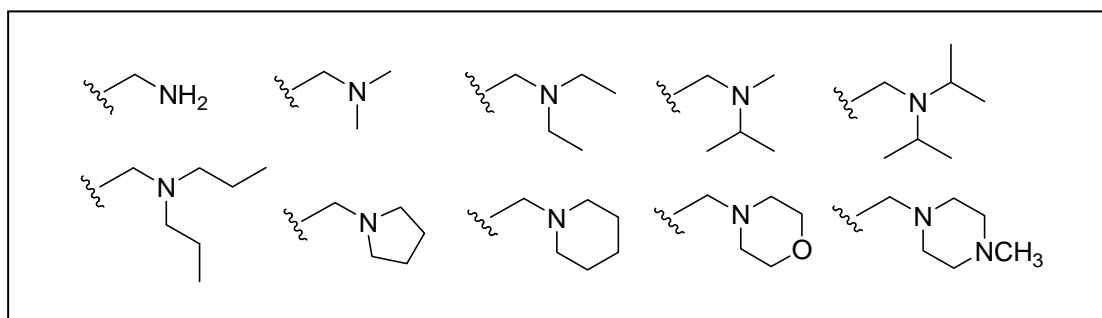


**Figure 3-4. Amino analogues of cysmethynil targeted for synthesis (side chain at position 3)**

v) Isosoteric replacement: Isosteres are groups or functionalities that share chemical and physical similarities. These similarities may result in isosteres having broadly similar biological effects, in which case, they would be called bioisosteres.[75]

The most common reason to replace an amide/carboxamide moiety with an isosteric group was its pharmacokinetic liability. Many examples of isosteric transformation of the carboxamide were found in the area of peptidomimetic design.[76] Thus there is a wide choice of isosteres for the carboxamide moiety and the following functionalities were proposed in the present study (**Figure 3-5**).



Methyl ester     Ethyl ester     Retroamide     Sulphonamide
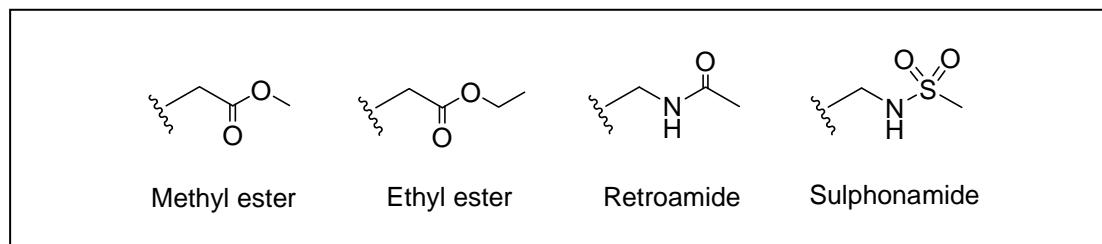
**Figure 3-5. Isosteric replacement of the acetamide side chain of cysmethynil.**

The methyl and ethyl esters mimicked both the substrate and product of Icmt catalysis. The minimal substrate of Icmt was identified to be *N*-acetyl-*S*-farnesyl-L-cysteine.[53] In the proposed ester analogues of cysmethynil, the *n*-octyl side chain would resemble the prenyl side chain of the substrate whereas the ester moiety mimicked the product of Icmt catalysis (methyl ester of cysteine). The combined features of substrate and product found in the ester analogues might enhance their affinity for Icmt. In the retro (reverse) amide and sulphonamide analogues, the H bond donor NH group was now located nearer the indole ring rather than at the distal end of the side chain whereas the H bond acceptor (carbonyl or sulphonyl oxygen atoms) group was shifted further from the ring. If the amide moiety of cysmethynil was involved in H bonding interactions, the displacement of these groups in the retroamide and sulphonamide could result in weaker interactions and inhibition.

### 3.2.5. Classification of target compounds

The target compounds were classified into 6 series according to the modifications made at positions 1, 3 and 5. Series 1 consisted of compounds with only variations at the 5-phenyl ring while retaining the *n*-octyl and acetamide sidechains (**Table 3-2**). In **1-7** and **1-8**, the 5-phenyl ring was omitted and replaced with either H or F. The ClogP values were determined with ChemDraw Ultra Version 10.0 (CambridgeSoft UK).

**Table 3-2. Structures and ClogP values of Series 1 compounds.**
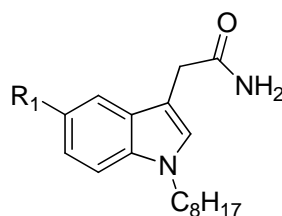


| Compound | R$_1$ | ClogP | Compound | R$_1$ | ClogP |
|----------|-------|-------|----------|-------|-------|
| **1-1** | *m*-tolyl | 7.0 | **1-5** | 3-ethoxyphenyl | 7.0 |
| **1-2** | *o*-tolyl | 6.7 | **1-6** | phenyl | 6.5 |
| **1-3** | *p*-tolyl | 7.0 | **1-7** | -H | 4.6 |
| **1-4** | 3-methoxyphenyl | 6.4 | **1-8** | -F | 4.8 |

In Series 2, the *n*-octyl side chain was omitted or replaced by other side chains. No change was made to the rest of the molecule (**Table 3-3**).

**Table 3-3. Structures and ClogP values of Series 2 compounds.**
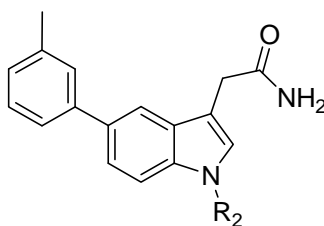


| Compound | R$_2$ | ClogP | Compound | R$_2$ | ClogP |
|----------|-------|-------|----------|-------|-------|
| **2-1** | -H | 2.8 | **2-3** | | 5.0 |
| **2-2** | | 6.4 | **2-4** | | 7.0 |
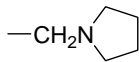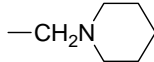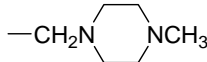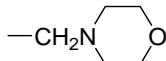
Series 3 consisted of compounds that had a secondary or tertiary amide side chain at position 3 while retaining *n*-octyl and *m*-tolyl at positions 1 and 5 respectively (**Table 3-4**).

**Table 3-4. Structures and ClogP values of Series 3 compounds.**
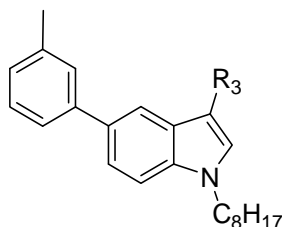


| Compound | R$_3$ | ClogP | Compound | R$_3$ | ClogP |
|----------|-------|-------|----------|-------|-------|
| **3-1** | -H | 8.7 | **3-5** | $-CH_2CON$ | 9.7 |
| **3-2** | -CH$_2$CONHCH$_3$ | 7.2 | | | |
| **3-3** | -CH$_2$CON(CH$_3$)$_2$ | 7.5 | **3-6** | $-CH_2CON$ | 10.3 |
| **3-4** | -CH$_2$CON(C$_2$H$_5$)$_2$ | 8.6 | **3-7** | $-CH_2CON\quad NCH_3$ | 9.3 |

Series 4 is similar to Series 3 except that the acetamide side chain was replaced by a secondary or tertiary amino function (**Table 3-5**). ClogP values were for the non-ionized (free base, non-protonated) state of the amino side chains.

**Table 3-5. Structures and ClogP values of Series 4 compounds.**

| Compound | $R_3$ | ClogP | Compound | $R_3$ | ClogP |
|---|---|---|---|---|---|
| **4-1** | -CH$_2$NH$_2$ | 7.6 | **4-7** | —CH$_2$N⬠ | 9.2 |
| **4-2** | -CH$_2$N(CH$_3$)$_2$ | 8.5 | | | |
| **4-3** | -CH$_2$N(C$_2$H$_5$)$_2$ | 9.6 | **4-8** | —CH$_2$N⬡ | 9.7 |
| **4-4** | -CH$_2$N(i-C$_3$H$_7$)$_2$ | 10.2 | **4-9** | —CH$_2$N⬡NCH$_3$ | 8.9 |
| **4-5** | -CH$_2$N(n-C$_3$H$_7$)$_2$ | 10.6 | | | |
| **4-6** | -CH$_2$NCH$_3$(i-C$_3$H$_7$) | 9.4 | **4-10** | —CH$_2$N⬡O | 8.4 |

Like Series 3 and 4, only modifications were made to the acetamide side chain in Series 5 (**Table 3-6**). These were homologation and isosteric replacement.

**Table 3-6. Structures and ClogP values of Series 5 compounds.**

| Compound | $R_3$ | ClogP | Compound | $R_3$ | ClogP |
|---|---|---|---|---|---|
| **5-1** | -CONH$_2$ | 7.4 | **5-4** | -CH$_2$COOC$_2$H$_5$ | 8.9 |
| **5-2** | -CH$_2$CH$_2$CONH$_2$ | 7.8 | **5-5** | -CH$_2$NHCOCH$_3$ | 7.5 |
| **5-3** | -CH$_2$COOCH$_3$ | 8.4 | **5-6** | -CH$_2$NHSO$_2$CH$_3$ | 7.4 |

The final Series 6 comprised compounds that were modified at two or more positions (**Table 3-7**). They were divided broadly into those with *n*-octyl and isoprenyl side chains at position 1. At position 3, the acetamide side chain was retained in a few members but most had amino functionalities. At position 5, the phenyl ring was substituted with methyl (most were *m*-methyl) or replaced with 5-fluoro.

**Table 3-7. Structures and ClogP values of Series 6 compounds.**
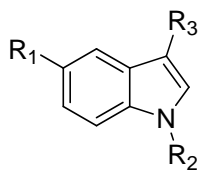


| Compound | $R_1$ | $R_2$ | $R_3$ | ClogP |
|----------|-------|-------|-------|-------|
| 6-1 | o-tolyl | n-C$_8$H$_{17}$ | -CH$_2$N(C$_2$H$_5$)$_2$ | 9.3 |
| 6-2 | p-tolyl | n-C$_8$H$_{17}$ | -CH$_2$N(C$_2$H$_5$)$_2$ | 9.6 |
| 6-3 | m-tolyl |  | -H | 6.7 |
| 6-4 | m-tolyl |  | -CH$_2$N(C$_2$H$_5$)$_2$ | 7.6 |
| 6-5 | m-tolyl |  | —CH$_2$N(piperidine) | 7.7 |
| 6-6 | m-tolyl |  | —CH$_2$N(morpholine) | 6.8 |
| 6-7 | -H |  | -CH$_2$N(C$_2$H$_5$)$_2$ | 5.2 |
| 6-8 | -F |  | -CH$_2$CONH$_2$ | 2.8 |
| 6-9 | -F | n-C$_8$H$_{17}$ | -CH$_2$N(C$_2$H$_5$)$_2$ | 7.4 |
| 6-10 | -F |  | -CH$_2$N(C$_2$H$_5$)$_2$ | 5.4 |
| 6-11 | 3-methoxyphenyl | n-C$_8$H$_{17}$ | -CH$_2$CON(C$_2$H$_5$)$_2$ | 8.0 |
| 6-12 | 3-ethoxyphenyl | n-C$_8$H$_{17}$ | -CH$_2$CON(C$_2$H$_5$)$_2$ | 8.5 |

[a]No phenyl ring, substituent is directly attached to indole C-5

Except for cysmethynil (**1-1**), the other 46 compounds were novel. Half of the compounds were synthesized by the candidate: these were the compounds in Series 1 (except **1-8**), Series 2, Series 3 (except **3-2**, **3-7**), **4-1** and **4-3** in Series 4, Series 5 (except **5-6**) and **6-11**, **6-12** in Series 6. The remaining compounds were synthesized by a post-doctoral fellow in the supervisor's laboratory, Dr Suresh Kumar Gorla.

## 3.3 CHEMICAL CONSIDERATIONS

### 3.3.1 Series 1 and 2

The following reaction scheme (**Scheme 3-1**) was adopted for the synthesis of compounds in Series 1 (with substituted 5-phenyl ring) and Series 2. For **1-8**, the same reactions were carried out on 5-fluoro-1*H*-indole-3-carbaldehyde but without Step (c).
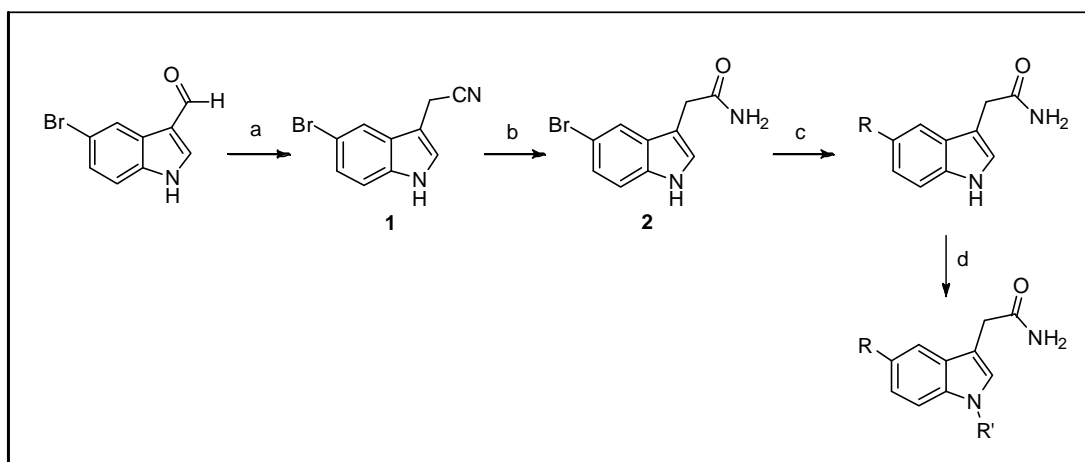


**Scheme 3-1. Synthesis of Series 1 and 2 compounds[60].** Reagents and conditions: (a) (i) NaBH$_4$, NH$_2$CHO-MeOH, rt (ii) KCN, 100 °C; (b) KOH, t-BuOH, reflux; (c) substituted phenylboronic acid, Pd(PPh$_3$)$_4$, NaHCO$_3$, EtOH/toluene, reflux; (d) R'-X (X = halogen), NaH, DMF, rt → 53-58°C.

In the first step, the aldehydic group in 5-bromo-1*H*-indole-3-carbaldehyde was converted to the nitrile **1** in a stepwise process that involved reduction of the aldehyde to an alcohol by sodium borohydride (NaBH$_4$), followed by displacement of the alcoholic OH by the cyanide ion. The reduction of the aldehydic carbonyl group by NaBH$_4$ proceeded by the transfer of a H atom together with the pair of electrons from the B-H bond to the electron deficient carbon atom of the C=O group (**Figure 3-6**). The resulting oxyanion helped to stabilize the electron deficient BH$_3$ by adding to its empty p orbital to generate a tetravalent boron anion again. The latter could transfer a second hydrogen (with pair of electrons) to another molecule of aldehyde. Theoretically, all 4 H atoms of NaBH$_4$ could be transferred to the carbonyl groups but

this rarely happened. The alcohol was not isolated but reacted *in situ* with the nucleophilic cyanide anion. The reaction occured via the more polarizable carbon atom in an $S_N2$ reaction to give nitrile **1** as the product.
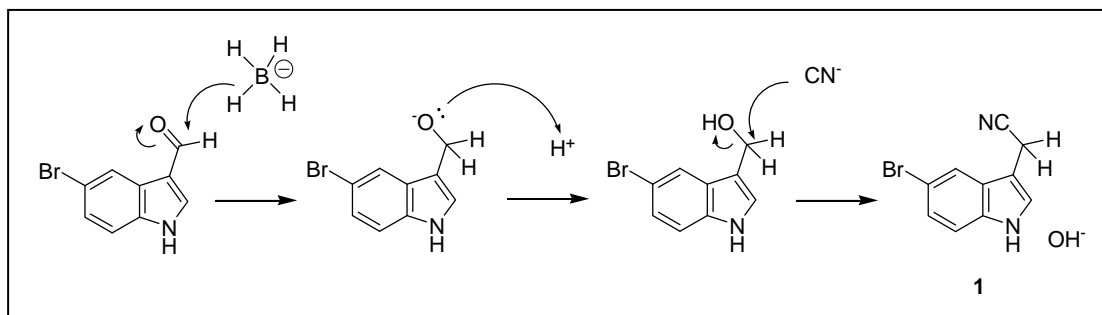


**Figure 3-6. Reaction mechanisms involved in the formation of the nitrile 1.**

Nitrile **1** was hydrolyzed under basic conditions to give the carboxamide **2**. Mechanistically, the reaction proceeded by a nucleophilic attack by the hydroxide anion on the carbon atom of the nitrile which was positively charged due to the greater electronegative character of N. A hydroxyimine anion (strong base) was formed and it deprotonated a water molecule. The hydroxyimine tautomerized to give the more stable amide **2**.
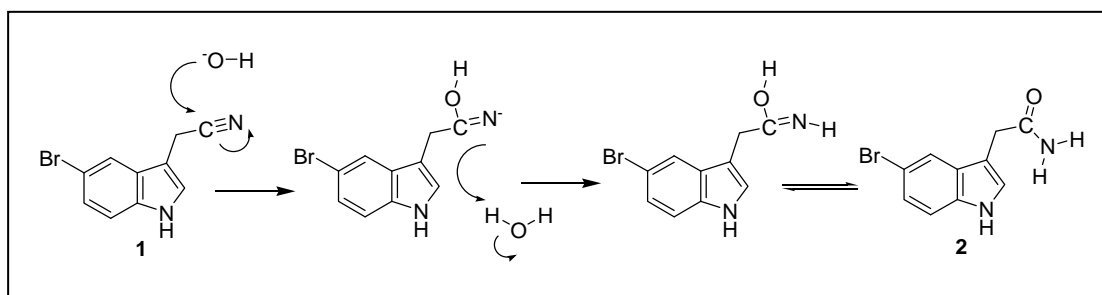


**Figure 3-7. Reaction mechanisms involved in the hydrolysis of nitrile 1 to give the amide 2.**

The preparation of 2-(5-substituted phenyl-1*H*-indol-3-yl)acetamides involved the Suzuki coupling reaction between **2** and an appropriately substituted phenylboronic acid in the presence of a palladium catalyst. The palladium catalyst used in this reaction was tetrakis(triphenylphosphine)palladium (0) in which

palladium existed in its lower and nominally electron rich oxidation state (represented as Pd(0)L$_n$). The catalytic cycle began with the rate limiting oxidative addition of the bromoindole **2** to Pd(0) to give the Pd(II) intermediate (**Figure 3-8**). Reaction of the phenylboronic acid with base led to the formation of borate which was transmetallated with the Pd(II) intermediate to afford the Pd(Ar)(Ar') complex. The latter underwent reductive elimination to form the C-C bond with the concomitant regeneration of the Pd(0) catalyst.
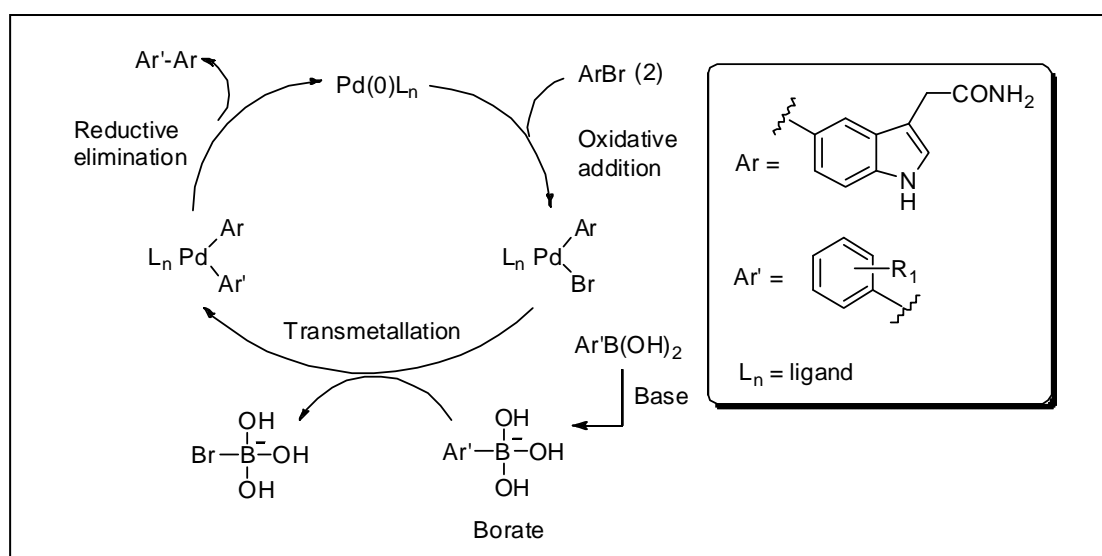


**Figure 3-8. Mechanism of Suzuki Coupling.**[77]

The mechanism of transmetallation had been investigated in detail and evidence suggested two different pathways for this process (**Figure 3-9**). The first involved exchange of hydroxide with bromide to generate palladium hydroxide that was transmetallated with the phenylboronic acid through a 4-centred transition state. The alternative pathway was when the borate complex reacted with the cationic Pd(II) intermediate formed when bromide dissociated from the oxidative addition complex.
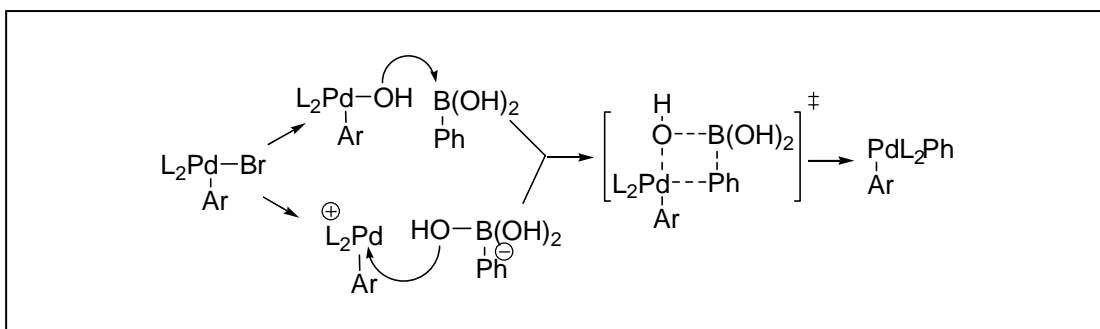
**Figure 3-9. Mechanism of transmetallation.**[77]

In the final step, the indole N was *N*-alkylated in the presence of sodium hydride which deprotonated the indole NH to generate a indolyl anion. The latter was a nucleophile and displaced the halide to give the N-substituted indole.
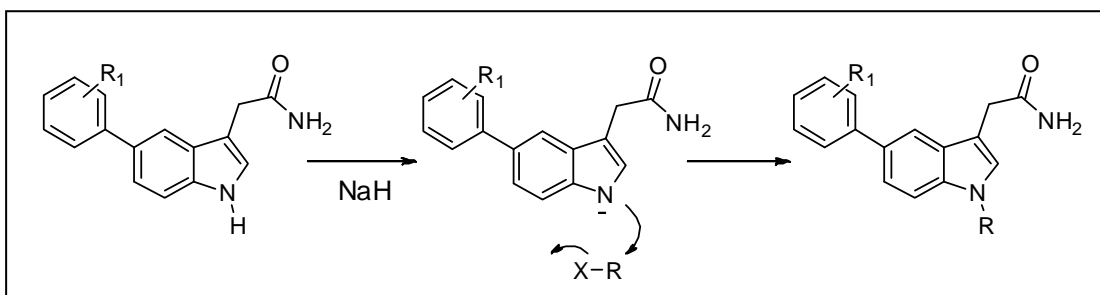


**Figure 3-10. Mechanism of N-alkylation of 2-(5-substituted phenyl-1H-indol-3-yl) acetamides.**

### 3.3.2. Series 3

The tertiary amides of Series 3 were synthesized by *N*-alkylating the indole N of 5-bromo-1*H*-indole-3-acetic acid to give 2-(5-bromo-1-octyl-1*H*-indol-3-yl)acetic acid **3**. The carboxylic acid was converted to the acid chloride with thionyl chloride (**Scheme 3-2**). The acid chloride was reacted in situ with the secondary amine or heterocyclic amine to give the corresponding amide. Subsequently, the 5-*m*-tolyl functionality was added by the Suzuki coupling reaction. The conversion of a carboxylic acid to acid chloride with thionyl chloride resulted from the strongly electrophilic sulphur atom in the latter. An unstable and highly electrophilic intermediate was formed which on protonation with HCl collapsed to give the acid

chloride and gaseous products. The conversion of the acid chloride to amide involved a nucleophilic substitution at the carbonyl group by the amine. A similar route was followed for the secondary amide **3-2**.
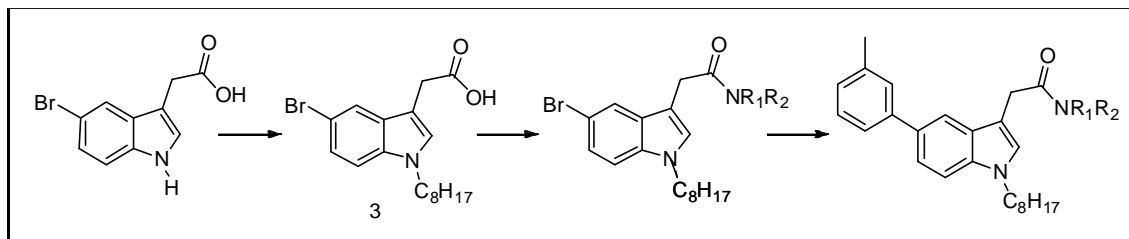


**Scheme 3-2. Synthesis of Series 3 compounds.** Reagents and conditions: (a) $C_8H_{17}Br$, NaH, DMF, rt → 53-58°C (b) (i) $SOCl_2$, benzene, reflux, (ii) 2° amine/heterocyclic amine, THF; (c) *m*-tolyl boronic acid $Pd(PPh_3)_4$, $NaHCO_3$, EtOH/toluene, reflux.

### 3.3.3. Series 4

**Scheme 3-3** outlines the synthesis of the tertiary amine analogues of Series 4.
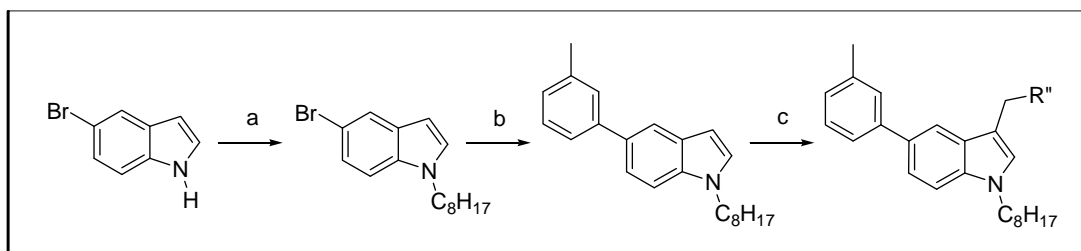


**Scheme 3-3. Synthesis of Series 4 compounds.** Reagents and conditions: (a) R-Br, NaH, DMSO; (b) substituted phenylboronic acid, DME, $Pd(PPh_3)_4$, $K_2CO_3$; (c) (i) aq. $CH_2O$, 2° amine (R"H), $ZnCl_2$, EtOH, rt, 3 h.

The amino side chain was introduced in the final step (c) by a Mannich reaction. Formaldehyde and the secondary amine reacted to form an iminium cation. The π-excessive indole ring reacted with the iminium ion as a nucleophile at its position 3 to give the desired product (**Figure 3-11**).
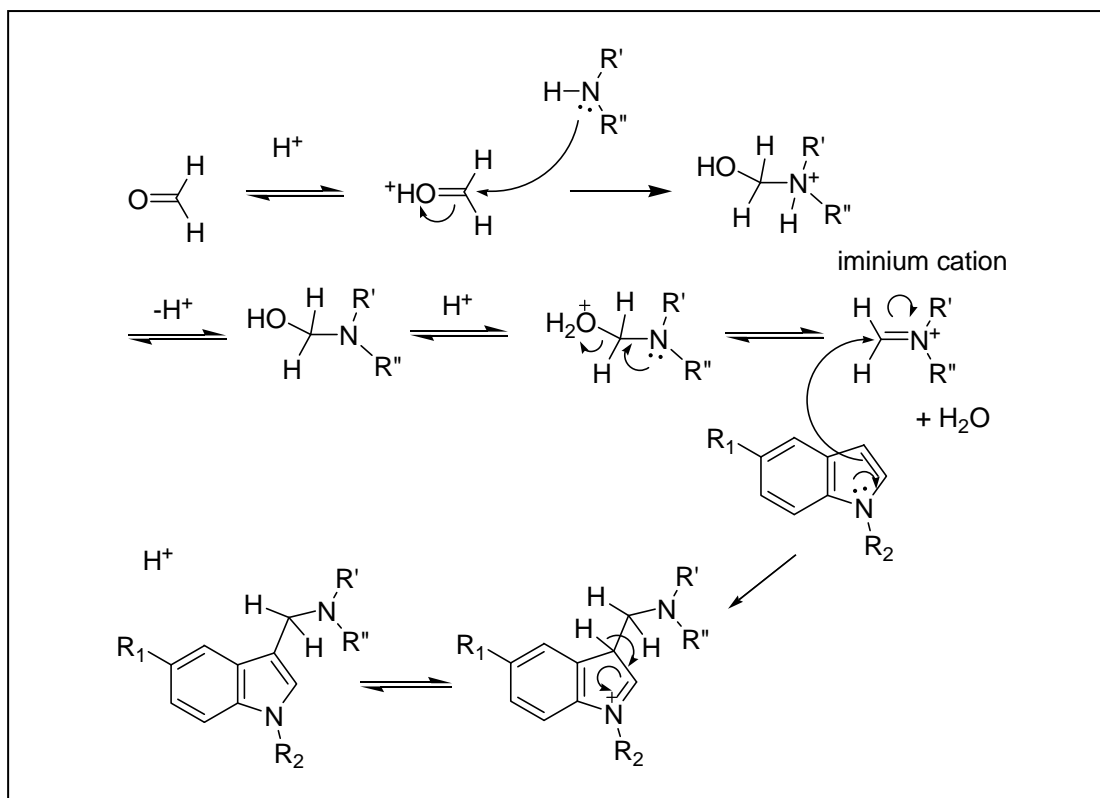
**Figure 3-11. Mechanism of the Mannich reaction.**

The primary amine **4-1** was prepared by another route, similar to that used to prepare the retroamide **5-5** and sulphonamide **5-6**. It is discussed in **Section 3.3.4.**

### 3.3.4. Series 5

**Scheme 3-4** outlined the synthesis of the retroamide **5-5**, sulphonamide **5-6** and the short chain homologue **5-1** of cysmethynil. The primary amine **4-1** was involved as an intermediate for the synthesis of **5-5** and **5-6**. Starting from 5-bromo-1*H*-indole-3-carbaldehyde, sequential functionalization at positions 5 and 1 were carried out as described earlier (**Section 3.3.1**). The carbonyl group of 1-octyl-5-(*m*-tolyl)-1*H*-indole-3-carbaldehyde was reacted with hydroxylamine to give a mixture of E and Z oximes which were not separated. The oxime was then reduced to give the primary amine **4-1**. The reduction was initially attempted with sodium borohydride

but the yield of **4-1** was improved when nickel chloride was added, possibly due to the greater electrophilicity of the carbonyl carbon in the presence of nickel chloride. **4-1** was acetylated with acetylchloride in the presence of triethylamine to give the retroamide **5-5** or sulphonylated with methane sulphonyl chloride in the presence of triethylamine to give the sulphonamide **5-6**.
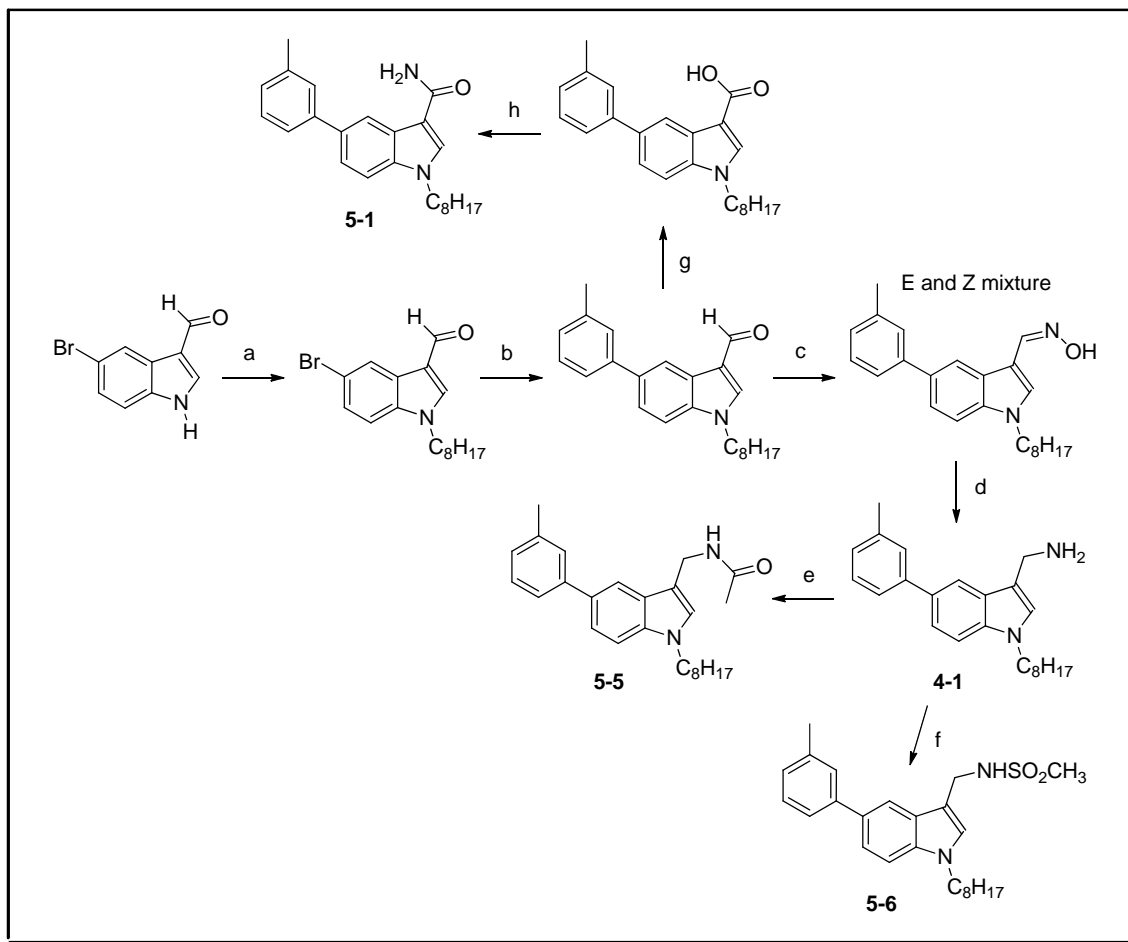


**Scheme 3-4. Synthesis of compounds 5-1, 5-5 and 5-6.** Reagents and conditions: (a) $C_8H_{17}Br$, NaH, DMF; (b) *m*-tolylboronic acid, EtOH/DME, $Pd(PPh_3)_4$, $K_2CO_3$; (c) $NH_2OH.HCl$, Py, EtOH; (d) $NiCl_2.6H_2O$, MeOH, $NaBH_4$; (e) $CH_3COCl$, $CH_2Cl_2$, $Et_3N$; (f) $CH_3SO_2Cl$, THF, $Et_3N$; (g) $KMnO_4$, acetone; (h) (i) $SOCl_2$, benzene, reflux, (ii) $NH_3$ (g), THF.

Starting from 1-octyl-5-(*m*-tolyl)-1*H*-indole-3-carbaldehyde, the short homologue of cysmethynil (**5-1**) was prepared by oxidation of the aldehyde to a

carboxylic acid which was then converted to an acid chloride with thionyl chloride, followed by reaction with gaseous ammonia to give the desired product.

The synthesis of the long chain homologue of cysmethynil **5-2** is shown in **Scheme 3-5**. The reaction started with 2-(5-bromo-1*H*-indol-3-yl) acetic acid which was converted to its methyl ester and then reduced with lithium aluminium hydride to give the alcohol. The alcoholic OH group was displaced by cyanide and hydrolyzed to give the primary amide. One difference was that the alcoholic OH had to be activated by conversion to a methane sulphonyl ester to enhance its leaving group ability. It was then reacted with KCN to give the nitrile. The conversion of the nitrile to amide and functionalization at positions 5 and 1 were carried out as described in **Section 3.3.1.**
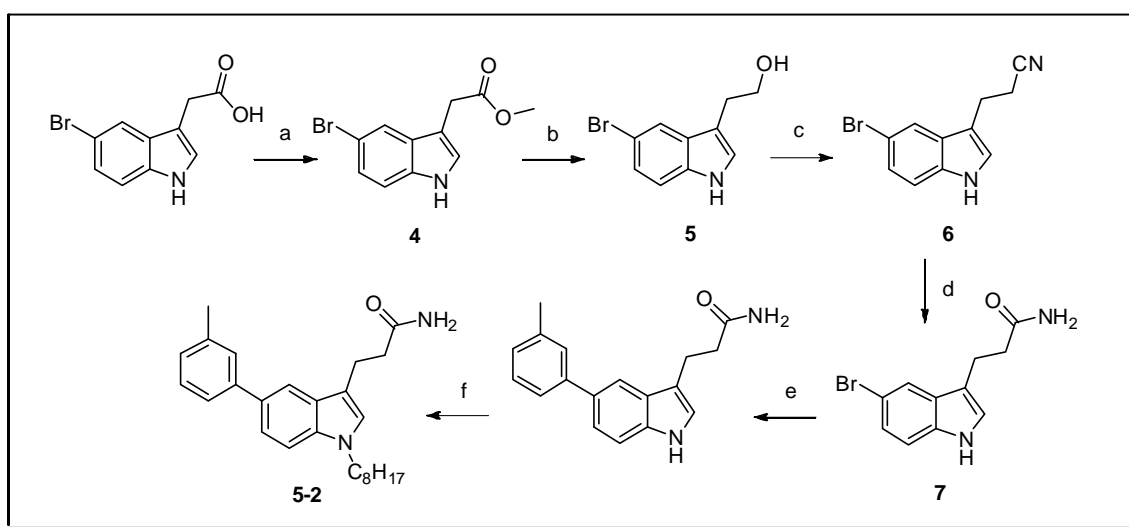


**Scheme 3-5. Synthesis of compound 5-2.** Reagents and conditions: (a) MeOH, $H_2SO_4$, reflux (1 h), (b) LiAlH$_4$, THF, rt (30 min), (c) (i) TEA, MsCl, CH$_2$Cl$_2$, 0 °C (30 min), (ii) KCN, DMSO, 100 °C (d) KOH, *t*-BuOH, reflux (3 h) (e) *m*-tolylboronic acid, Pd(PPh$_3$)$_4$, NaHCO$_3$, EtOH/toluene, reflux (1-6 h) (f) bromooctane, NaH, DMF, rt (1.5 h) 53-58 °C (3-6 h)

The bioisosteric esters of cysmethynil (**5-3**, **5-4**) were also prepared starting from 2-(5-bromo-1*H*-indol-3-yl) acetic acid. The latter was functionalized at positions 1 and 5 and then converted to the methyl or ethyl ester by an acid catalyzed esterification reaction (**Scheme 3-6**).
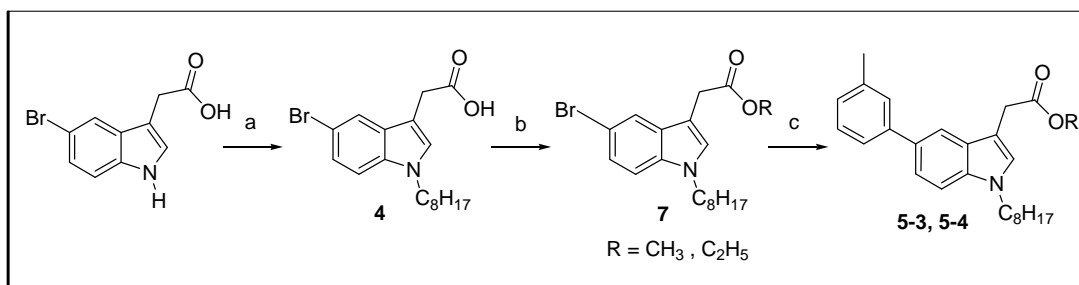
**Scheme 3-6. Synthesis of compounds 5-3 and 5-4.** Reagents and conditions: (a) $C_8H_{17}Br$, NaH, DMSO; (b) *m*-tolylboronic acid in EtOH, DME, $Pd(PPh_3)_4$, $K_2CO_3$; (c) R-OH, cat. $H_2SO_4$.

### 3.3.5. Series 6

The compounds in this Series were synthesized following earlier procedures. The amines **6-1**, **6-2**, **6-4** to **6-7**, **6-9** and **6-10** were prepared by the Mannich reaction starting from 5-bromo-1*H*-indole (**Scheme 3-3**), indole or 5-fluoro-1*H*-indole with appropriate functionalization at positions 1 and 5. **Scheme 3-1** was used for the synthesis of the primary amide **6-8** starting with 5-fluoro-1*H*-indole-3-carbaldehyde. The tertiary amides **6-11** and **6-12** were synthesized following the reaction sequences described in **Scheme 3-2**, except that Suzuki coupling was carried out with 3-methoxyphenylboronic acid and 3-ethoxyphenylboronic acid, instead of 3-methylphenylboronic acid. **6-3** was synthesized from 5-bromo-1*H*-indole by *N*-alkylation with 1-chloro-3-methylbut-2-ene to introduce the isoprenyl side chain, followed by Suzuki coupling at position 5 to introduce the *m*-tolyl residue.

## 3.4. EXPERIMENTAL

### 3.4.1. General Details

Reagents (synthetic grade or better) were obtained from commercial suppliers (Sigma-Aldrich Chemical Company Inc, Singapore; Alfa Aesar, MA, USA) and used without further purification. $^1$H (300 MHz) and $^{13}$C (75 MHz) spectra were measured on a Bruker Spectrospin 300 Ultrashield spectrometer magnetic resonance spectrometer. Chemical shifts were reported in ppm and referenced to TMS or residual deuterated solvents (CDCl$_3$ $\delta_H$ 7.26 ppm, DMSO-$d_6$  $\delta_H$  2.50 ppm; CD$_3$OD $\delta_H$  3.31 ppm) for $^1$H spectra and residual CDCl$_3$ ($\delta_C$ 77.16 ppm) and DMSO-d$_6$  ($\delta_C$ 39.52 ppm) for proton decoupled $^{13}$C NMR spectra.  Coupling constants (J) were reported in Hertz (Hz). Reactions were routinely monitored by thin layer chromatography (TLC) on pre-coated plates (Silica Gel 60 F254, Merck) with ultraviolet light as visualizing agent. Column chromatography was carried out with Silica Gel 60 (0.04–0.063 mm).  Nominal mass spectra were captured on an LCQ Finnigan MAT equipped with an atmospheric pressure chemical ionization (APCI) probe and m/z values for the molecular ion were reported. Accurate mass was determined using a Finnigan Mat 95/XL-T spectrometer equipped with an electron spray ionization probe (ESI). Purity of final compounds were verified by combustion analysis (C and H) on a Perkin Elmer PRE-2400 Elemental Analyzer or by reverse phase high performance liquid chromatography (HPLC) carried out on a Waters Delta 600 liquid chromatography system, using an Agilent Zorbax Eclipse XDB-C$_{18}$ column (4.6 mm × 150 mm, 5 μm) on two different solvent systems (isocratic mode), details of which are given in **Table A2-1**. Compounds synthesized by the candidate are reported in this section while those prepared by Dr Suresh are described in **Appendix 2**.

### 3.4.2. Synthesis of Series 1 and 2 compounds
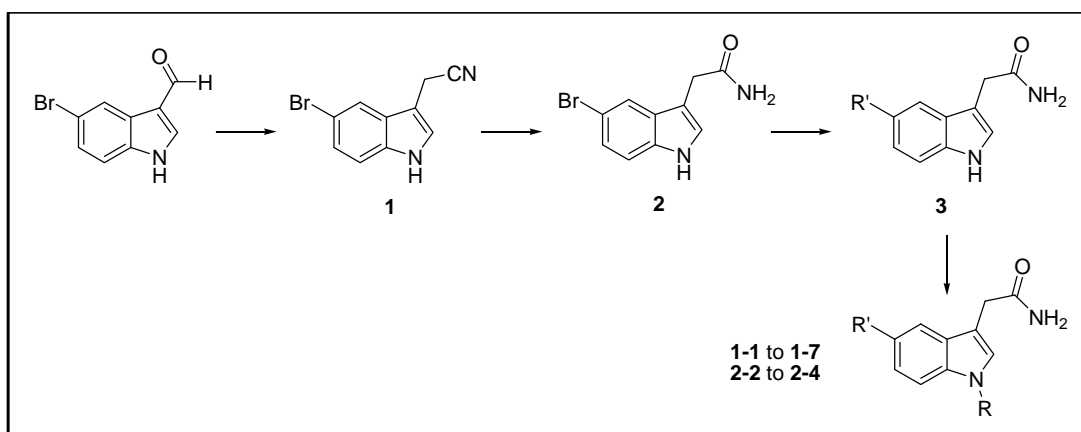


**2-(5-Bromo-1*H*-indol-3-yl) acetonitrile (1).**[60] To a solution of 5-bromo-1*H*-indole-3-carbaldehyde (4.5 mmol, 1 equiv.) in formamide-methanol (NH$_2$CHO-MeOH; 1:1, v/v; 200 mL) was added sodium borohydride (NaBH$_4$; 13.5 mmol, 3 equiv.) and the mixture was stirred for 0.5 h. To the reaction mixture was added potassium cyanide (KCN; 45 mmol, 10 equiv.) and the whole mixture was refluxed on an oil bath at 100 °C for 2.5 h with stirring. After cooling to room temperature, brine was added and the mixture was extracted with chloroform (CHCl$_3$), washed with brine, dried over anhydrous sodium sulphate (Na$_2$SO$_4$), and evaporated under reduced pressure to give the nitrile **1**. Recrystallization from ethanol (EtOH)/water gave needle-like white crystals. Yield: 81 %; $^1$H NMR (300 MHz, CD$_3$OD): $\delta$ 3.73 (s, 2 H), 7.19-7.33 (m, 3 H), 7.69 (s, 1 H); MS (APCI): *m/z* 236.1 [M + H]$^+$.

**2-(5-Bromo-1*H*-indol-3-yl)acetamide (2).**[60] Nitrile **1** (3.5 mmol, 1 equiv.), was refluxed in tertiary butanol (*t*-BuOH;10 mL) containing finely powdered 85% potassium hydroxide (KOH; 28 mmol, 8 equiv.) for 3 h. The reaction mixture was cooled to room temperature, diluted with water, and acidified with 1 M hydrochloric acid (HCl). The resulting suspension was filtered under reduced pressure. The filter cake was washed with water then dried *in vacuo*. The product was isolated as an off-

white solid. Yield: 89 %; $^{1}$H NMR (300 MHz, DMSO-$d_6$): $\delta$ 3.44 (s, 2 H), 7.15-7.38 (m, 4 H), 7.73 (s, 1 H), 11.07 (s, 1 H); MS (APCI): $m/z$ 254.1 [M + H]$^{+}$.

**General procedure for preparation of 2-(5-(substituted phenyl)-1*H*-indole-3-yl)acetamides (3).** To a suspension of bromoindole **2** (2 mmol, 1 equiv.) in anhydrous toluene (40 mL) in an ice bath was added tetrakis(triphenylphosphine)palladium (0) (Pd(PPh$_3$)$_4$; 5.7 mol%). The bright yellow suspension was stirred for 0.5 h, after which was added in one portion a solution of substituted phenylboronic acid (3 mmol, 1.5 equiv) in absolute EtOH (10 ml) followed immediately by a saturated aqueous solution of sodium bicarbonate (NaHCO$_3$; 25 mL). After refluxing for 1-6 h, the biphasic mixture was cooled to room temperature and then poured into a solution of brine. The organic phase was separated and the aqueous phase was extracted with ethyl acetate (EtOAc). The organic extracts were combined, dried (anhydrous Na$_2$SO$_4$), the drying agent was removed by filtration and the filtrate was concentrated under reduced pressure to give the crude product **3**, which was used without further purification for the *N*-alkylation reaction.

**2-(5-*m*-Tolyl-1*H*-indol-3-yl)acetamide (2-1).**[60] Off-white solid (purified by column chromatography using EtOAc/ dichloromethane (CH$_2$Cl$_2$) as eluate); Yield: 80 %; $^{1}$H NMR (300 MHz, CDCl$_3$): $\delta$ 2.44 (s, 3 H), 3.79 (s, 2 H), 7.14 (d, *J* 6 Hz, 1 H), 7.20 (s, 1 H), 7.31 (t, *J* = 15 Hz, 1 H), 7.43-7.52 (m, 4 H), 7.78 (s, 1 H), 8.25 (s, 1 H); $^{13}$C NMR (75 MHz, CDCl$_3$): $\delta$ 21.5, 32.9, 109.7, 111.6, 117.1, 122.6, 124.2, 124.5, 127.3, 127.4, 128.1, 128.6, 134.0, 135.8, 138.3, 142.0, 174.1; MS (APCI): $m/z$ 265.2 [M + H]$^{+}$.

**General procedure for preparation of 2-(1-substituted-5-(substituted phenyl)-1*H*-indol-3-yl)acetamides (1-1 to 1-7 and 2-2 to 2-4).** To a stirred suspension of sodium hydride (NaH, 60% dispersion in mineral oil; 3 mmol, 1.5

equiv.) in anhydrous dimethylformamide (DMF; 5 ml) in ice bath was added dropwise a solution of the crude acetamide **3** (2 mmol, 1 equiv.) in anhydrous DMF (10 mL) over a period of 10 min. After stirring at room temperature for 1.5 h, the alkyl or arylalkyl halide (1-bromooctane for **1-1** to **1-7**, trifluromethylbenzylbromide for **2-2**, 1-chloro-3-methylbut-2-ene for **2-3** and geranyl bromide for **2-4**; 6 mmol, 3 equiv.) was added dropwise over 5 min. The reaction mixture was then heated in an oil bath at 53-58°C for 3-6 h, after which it was cooled to room temperature and poured into ice water. The suspension was stirred 10 min and extracted with diethyl ether (Et$_2$O). The organic extracts were combined, washed with brine, and dried (anhydrous Na$_2$SO$_4$). The drying agent was removed by filtration and the filtrate was concentrated under reduced pressure to give the crude alkylated product. The crude product was purified by flash silica gel column chromatography by elution with EtOAc/CH$_2$Cl$_2$ to give compounds **1-1** to **1-7** and **2-2** to **2-4**.

**2-(1-Octyl-5-*m*-tolyl-1*H*-indol-3-yl)acetamide (Cysmethynil, 1-1).[60]** Off-white solid, mp 62-65 °C; Yield: 77 %; ^{1}H NMR (300 MHz, CD$_3$OD,): $\delta$ 0.92 (t, $J$ =15 Hz, 3 H), 1.29-1.35 (m, 10 H), 1.88 (t, 2 H, $J$ = 15 Hz), 2.42 (s, 3 H), 3.70 (s, 2 H), 4.20 (t, $J$ = 15 Hz, 2 H), 7.12 (d, $J$ = 9 Hz, 1 H), 7.21-7.32 (m, 2 H), 7.37-7.48 (m, 4 H), 7.80 (s, 1 H); ^{13}C NMR (75 MHz, CD$_3$OD): $\delta$ 13.6, 20.8, 22.8, 27.2, 29.5, 29.6, 30.6, 32.1, 32.6, 46.3, 108.6, 110.0, 117.3, 121.5, 124.5, 127.1, 127.6, 128.2, 128.7, 131.1, 133.1, 136.6, 138.4, 143.0, 177.0. MS (APCI): *m/z* 377.5 [M+H]$^+$.

**2-(1-Octyl-5-*o*-tolyl-1*H*-indol-3-yl)acetamide (1-2).** Off-white solid; Yield: 63 %; ^{1}H NMR (300 MHz, CDCl$_3$): $\delta$ 0.85 (t, $J$ = 15 Hz , 3 H), 1.27-1.34 (m, 10 H), 1.85 (t, $J$ = 15 Hz, 2 H), 2.30 (s, 3 H), 3.72 (s, 2 H), 4.09 (t, $J$ = 15 Hz, 2 H), 7.12 (d, $J$ 9 Hz, 1 H), 7.21-7.32 (m, 5 H), 7.37-7.48 (m, 3 H), 7.80 (s, 1 H) ^{13}C NMR (75 MHz, CDCl$_3$): $\delta$ 14.1, 20.6, 22.6, 27.0, 29.1, 29.1, 30.3, 31.7, 32.9, 46.5, 107.7, 109.1,

119.2, 123.8, 125.7, 126.8, 127.3, 127.6, 130.3, 130.3, 133.5, 135.5, 135.6, 142.6, 174.3; MS (APCI): *m/z* 377.3 [M + H]$^+$.

**2-(1-Octyl-5-*p*-tolyl-1*H*-indol-3-yl)acetamide (1-3).** Off-white solid, mp 123-128 ˚C; Yield: 75 %; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 0.85 (t, *J* = 15 Hz, 3 H), 1.26-1.33 (m, 10 H), 1.83 (t, *J* = 12 Hz, 2 H), 2.40 (s, 3 H), 3.76 (s, 2 H), 4.09 (t, *J* = 12 Hz, 2 H), 7.07, (s, 1 H) , 7.24 (d, *J* = 6 Hz, 2 H), 7.38 (d, *J* 6 Hz, 1 H), 7.48-7.55 (m, 3 H), 7.75 (s, 1 H) $^{13}$C NMR (75 MHz, CDCl$_3$): $\delta$ 14.0, 21.0, 22.6, 27.0, 29.1, 29.1, 30.2, 31.7, 32.9, 46.5, 107.9, 109.9, 117.0, 121.8, 127.2, 127.2, 127.8, 129.4, 129.4, 133.2, 135.9, 136.2, 139.2, 174.3; MS (APCI): *m/z* 377.3 [M + H]$^+$.

**2-(5-(3-Methoxyphenyl)-1-octyl-1*H*-indol-3-yl)acetamide (1-4).** Off-white solid, mp 89-91 ˚C; Yield: 26 %; $^1$H NMR (300 MHz , CD$_3$OD): $\delta$ 0.91 (t, *J* = 12 Hz, 3 H), 1.29-1.34 (m, 10 H), 1.88 (t, *J* = 12 Hz, 2 H), 3.71 (s, 2 H), 3.87 (s, 3 H), 4.20 (t, *J* = 12 Hz, 2 H), 6.87 (d, *J* = 6 Hz, 1 H), 7.17-7.44 (m, 6 H), 7.82 (s, 1 H). $^{13}$C NMR (75 MHz, DMSO-*d$_6$*): $\delta$ 14.9, 15.7, 23.0, 27.3, 29.6, 29.7, 30.9, 32.2, 33.3, 46.5, 56.1, 109.9, 111.0, 112.6, 113.4, 118.3, 120.1, 121.4, 129.0, 129.1, 130.9, 131.8, 136.6, 144.4, 160.7, 174.2; MS (APCI): *m/z* 393.5 [M + H]$^+$.

**2-(5-(3-Ethoxyphenyl)-1-octyl-1*H*-indol-3-yl)acetamide (1-5).** Off-white solid, mp 74-76 ˚C; Yield: 74 %; $^1$H NMR (300 MHz, DMSO-*d$_6$*): $\delta$ 0.90 (t, *J* = 12 Hz, 3 H), 1.05-1.15 (m, 6 H), 1.27-1.31 (m, 10 H), 1.84 (t, *J* = 9 Hz, 2 H), 3.40-3.48 (m, 4 H), 3.87 (s, 3 H), 3.88 (s, 2 H), 4.19 (t, *J* = 15 Hz, 2 H), 6.87 (d, *J* = 6 Hz, 1 H), 7.14-7.44 (m, 6 H), 7.83 (s, 1 H); $^{13}$C NMR (75 MHz, DMSO-*d$_6$*): $\delta$ 15.0, 15.8, 23.1, 27.4, 29.7, 29.8, 31.0, 32.2, 33.3, 46.5, 64.0, 110.1, 111.0, 113.0, 114.0, 118.4, 120.0, 121.4, 128.9, 129.2, 130.8, 131.8, 136.6, 144.4, 160.0, 173.8; MS (APCI): *m/z* 407.4 [M + H]$^+$.

**2-(1-Octyl-5-phenyl-1*H*-indol-3-yl)acetamide (1-6).** Off-white solid, mp 74-76 °C; Yield: 70 %; ^{1}H NMR (300 MHz, CDCl$_3$): $\delta$ 0.85 (t, $J$ = 15 Hz, 3 H), 1.26-1.31 (m, 10 H), 1.84 (t, $J$ = 12 Hz, 2 H), 3.77 (s, 2 H), 4.10 (t, $J$ = 15 Hz, 2 H), 7.07 (d, $J$ = 6 Hz, 1 H), 7.28-7.34 (m, 2 H), 7.40-7.52 (m, 3 H), 7.63 (d, $J$ 6 Hz, 2 H), 7.77 (s, 1 H); ^{13}C NMR (75 MHz , CDCl$_3$): $\delta$ 14.1, 22.6, 27.0, 29.1, 29.1, 30.3, 31.7, 32.9, 46.5, 108.0, 110.0, 117.3, 121.9, 124.9, 126.5, 127.3, 127.3, 128.7, 128.7, 133.2, 136.0, 142.1, 174.3; MS (APCI): *m/z* 363.6 [M + H]$^+$.

**2-(1-Octyl-1*H*-indol-3-yl)acetamide (1-7).** Off-white solid, mp 74-76 °C; Yield: 65 %; ^{1}H NMR (300 MHz, CDCl$_3$): $\delta$ 0.85 (t, $J$ 12 Hz, 3 H), 1.25-1.31 (m, 10 H), 1.81 (t, $J$ = 15 Hz, 2 H), 3.73 (s, 2 H), 4.07 (t, $J$ = 15 Hz, 2 H), 7.06 (s, 1 H), 7.12 (t, $J$ = 15 Hz, 1 H), 7.22 (t, $J$ = 12 Hz, 1 H), 7.34 (d, $J$ = 9 Hz, 1 H), 7.57 (d, $J$ = 6 Hz, 1 H); ^{13}C NMR (75 MHz, CDCl$_3$): $\delta$ 14.0, 22.6, 27.0, 29.1, 29.1, 30.2, 31.7, 32.9, 46.4, 107.6, 109.7, 118.9, 119.5, 122.0, 127.2, 127.4, 136.5, 174.3; MS (APCI): *m/z* 287.5 [M + H]$^+$.

**2-(5-*m*-Tolyl-1-(3-(trifluoromethyl)benzyl)-1*H*-indol-3-yl)acetamide (2-2).** Off-white solid; Yield: 35 %; ^{1}H NMR (300 MHz , CDCl$_3$): $\delta$ 2.43 (s, 3 H), 3.79 (s, 2 H), 5.38 (s, 2 H), 7.13 (d, $J$ = 9 Hz, 2 H), 7.30 (t, $J$ = 18 Hz, 2 H), 7.41-7.50 (m, 5 H), 7.54 (d, $J$ = 6 Hz, 2 H) 7.79 (s, 1 H); ^{13}C NMR (75 MHz , CDCl$_3$): $\delta$ 28.7, 37.8, 51.5, 99.6, 100.3, 106.3, 110.5, 111.2, 112.0, 112.2, 114.4, 114.8, 115.0, 115.4, 116.1, 116.5, 119.7, 121.5, 123.0, 123.2, 126.0, 152.1; MS (APCI): *m/z* 423.4 [M + H]$^+$.

**2-(1-(3-Methylbut-2-enyl)-5-*m*-tolyl-1*H*-indol-3-yl)acetamide (2-3).** Off-white solid; Yield: 62 %; ^{1}H NMR (300 MHz, CDCl$_3$): $\delta$ 1.79 (s, 3 H), 1.84 (s, 3 H), 2.43 (s, 3 H), 3.75 (s, 2 H), 4.68 (d, $J$ = 6 Hz, 2 H), 5.37 (t, $J$ = 15 Hz, 1 H), 7.05-7.15 (m, 2 H), 7.30-7.51 (m, 5 H), 7.75 (s, 1 H) ^{13}C NMR (75 MHz, CDCl$_3$): $\delta$ 18.0, 21.6,

25.7, 33.0, 44.2, 108.0, 110.0, 117.3, 119.5, 121.9, 124.5, 127.2, 127.4, 128.0, 128.1, 128.6, 133.4, 136.0, 136.9, 138.3, 142.1, 174.4; MS (APCI): $m/z$ 333.4 $[M + H]^+$.

**(E/Z)-2-(1-(3,7-dimethylocta-2,6-dienyl)-5-*m*-tolyl-1*H*-indol-3-yl)acetamide (2-4).** Amber oil; Yield: 45 %; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 1.59 (s, 3 H), 1.66 (s, 3 H), 1.80 (s, 3 H), 2.10 (t, $J = 21$ Hz, 4 H), 2.43 (s, 3 H), 3.65 (s, 2 H), 4.70 (d, $J = 6$ Hz, 2 H), 5.04 (t, $J = 9$ Hz, 1 H), 5.39 (t, $J = 6$ Hz, 1 H), 7.05-7.15 (m, 2 H), 7.30-7.51 (m, 5 H), 7.75 (s, 1 H); $^{13}$C NMR (75 MHz, CDCl$_3$): $\delta$ 16.4, 17.7, 21.5, 25.7, 26.2, 32.9, 39.4, 44.1, 107.9, 110.1, 117.2, 119.3, 121.9, 123.6, 124.4, 127.2, 127.4, 128.1, 128.6, 131.9, 133.4, 136.0, 138.3, 140.4, 140.7, 142.8, 174.5. MS (APCI): $m/z$ 401.5 $[M + H]^+$.

### 3.4.3. Synthesis of Series 3 compounds

**1-Octyl-5-*m*-tolyl-1*H*-indole (3-1).** Synthesis of **3-1** is described in **Section 3.4.4.**
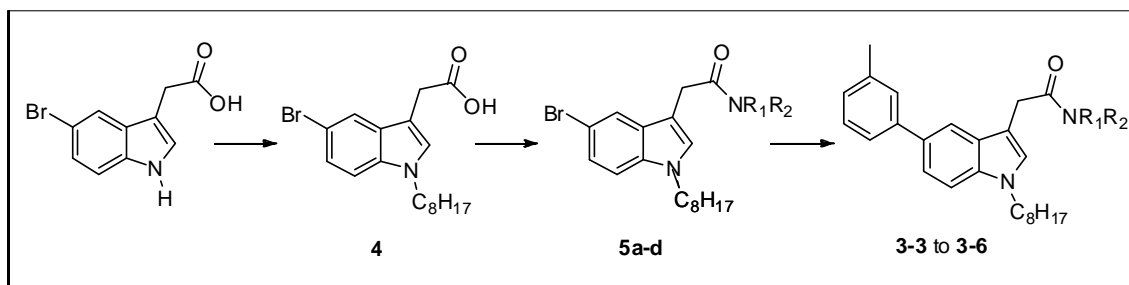


**2-(5-Bromo-1-octyl-1*H*-indol-3-yl)acetic acid (4).** Compound **4** was synthesized according to the method described in Roy et al.[78] To a stirred suspension of NaH (60% dispersion in mineral oil, 138 mmol, 5 equiv.) in tetrahydrofuran (THF; 100 mL) at 0 °C was added a solution of 2-(5-bromo-1*H*-indol-3-yl)acetic acid (27.7 mmol, 1 equiv.) in THF (50 mL). The mixture was stirred for 30 min at 0 °C after which a solution of 1-bromooctane (83.1 mmol, 3 equiv.) in THF (50 mL) was added dropwise. The mixture was allowed to slowly warm up to room temperature after

which it was stirred for another 4 h. The reaction mixture was then cooled to 0 °C and excess NaH was carefully destroyed by the slow addition of MeOH with vigorous stirring, followed by cold water until a clear yellow solution was obtained. Et$_2$O (100 mL) was added, the aqueous phase was separated, acidified with 6 M HCl, and extracted with sufficient CH$_2$Cl$_2$. The combined organic extracts were dried (anhydrous Na$_2$SO$_4$) and concentrated in vacuo. The crude product was purified by column chromatography with hexane/EtOAc as a . Yield: 64 %; $^1$H NMR (300 MHz, CDCl$_3$): δ 7.71 (s, 1 H), 7.26 (d, $J$ = 2.1 Hz, 1 H), 7.16 (d, $J$ = 8.7 Hz, 1H), 7.07 (s, 1 H), 4.03 (t, $J$ = 7.2 Hz, 2 H), 3.73 (s, 2 H), 1.79 (t, $J$ = 9.9 Hz), 1.28-1.24 (m, 10 H), 0.86 (t, $J$ = 6.3 Hz). $^{13}$C NMR (75 MHz, CDCl$_3$): δ 177.3, 134.8, 129.2, 128.0, 124.5, 121.5, 112.5, 110.9, 105.6, 46.5, 31.7, 30.7, 30.1, 29.1, 29.0, 26.9, 22.5, 14.0.

**General procedure for the synthesis of 2-(5-bromo-1-octyl-1*H*-indol-3-yl)-*N*-substituted acetamides (5).** A mixture of acetic acid **4** (1 mmol, 1 equiv.) and thionyl chloride (SOCl$_2$; 2 mL, 27 equiv.) were refluxed for 4 h in dry benzene (5 mL), after which excess SOCl$_2$ and benzene were removed by distillation in vacuo to give the corresponding acid chloride. The acid chloride was dissolved in dry THF (4 ml) and added dropwise to a stirred solution of the amine in dry THF at 0-5 ºC. The reaction mixture was stirred at the same temperature for 1 h after which THF was removed in vacuo and the residue extracted with CH$_2$Cl$_2$ and dried (anhydrous Na$_2$SO$_4$). Removal of the solvent under reduced pressure gave the desired amide which was purified by column chromatography with EtOAc/hexane as eluting solvents.

**2-(5-bromo-1-octyl-1*H*-indol-3-yl)-*N*,*N*-dimethylacetamide (5a).** Yield: 38%; $^1$H NMR (300 MHz, CDCl$_3$): δ 7.70 (s, 1 H), 7.25 (d, $J$ = 8.1 Hz, 1 H), 7.15 (d, $J$ =
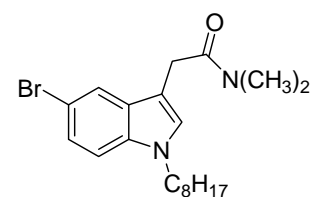
8.7 Hz, 1 H), 7.05 (s, 1 H), 4.01 (t, $J$ = 7.2 Hz, 2 H), 3.72 (s, 2 H), 3.40 (s, 6 H), 1.77 (t, $J$ = 6.9 Hz, 2 H), 1.25-1.23 (m, 10 H), 0.86 (t, $J$ = 6.6 Hz, 3 H); $^{13}$C NMR (75 MHz, CDCl$_3$): $\delta$ 171.2, 134.8, 129.3, 127.4, 124.3, 121.4, 112.3, 110.8, 107.2, 46.4, 37.8, 35.6, 31.7, 30.6, 30.1, 29.1 (2C), 26.9, 22.5, 14.0.

**2-(5-bromo-1-octyl-1$H$-indol-3-yl)-$N$,$N$-diethylacetamide (5b).** Yield: 38 %; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 7.70 (s, 1 H), 7.25 (d, $J$ = 8.1 Hz, 1 H), 7.15 (d, $J$ = 8.7 Hz, 1 H), 7.05 (s, 1 H), 4.01 (t, $J$ = 7.2 Hz, 2 H), 3.72 (s, 2 H), 3.40 (q, $J$ = 6.9 Hz, 2 H), 3.34 (q, $J$ = 7.5 Hz, 2 H), 1.77 (t, $J$ = 6.9 Hz, 2 H), 1.25-1.23 (m, 10 H), 1.13 (t, $J$ = 6.9 Hz, 3 H), 1.11 (t, 3 H, $J$ = 6.9 Hz), 0.86 (t, $J$ = 6.6 Hz, 3 H); $^{13}$C NMR (75 MHz, CDCl$_3$): $\delta$ 170.3, 134.8, 129.3, 127.3, 124.2, 121.3, 112.2, 110.8, 107.7, 46.4, 42.4, 40.2, 31.7, 30.5, 30.1, 29.12, 29.10, 26.9, 22.5, 14.3, 14.0, 12.9.
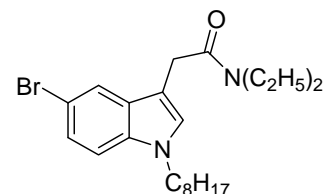
**2-(5-Bromo-1-octyl-1$H$-indol-3-yl)-1-(pyrrolidin-1-yl)ethanone (5c).** This compound was not characterized.
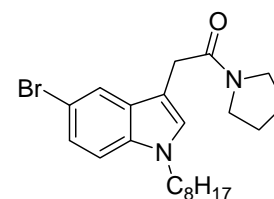
**2-(5-bromo-1-octyl-1$H$-indol-3-yl)-1-(piperidin-1-yl)ethanone (5d).** Yield: 38 %; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 0.86 (t, $J$ = 6.9 Hz, 3 H), 1.23-1.25 (m, 10 H), 1.36 (t, $J$ = 6.9 Hz, 2 H), 1.55 (t, $J$ = 6.9 Hz, 4 H), 1.77 (t, $J$ = 6.9 Hz, 2 H), 3.39 (t, $J$ = 5.4 Hz, 2 H), 3.58 (t, $J$ = 5.4 Hz, 2 H), 3.75 (s, 2 H), 4.02 (t, $J$ = 6.9 Hz, 2 H), 7.01 (s, 1 H), 7.16 (d, $J$ = 8.7 Hz, 1 H), 7.26 (d, $J$ = 8.7 Hz, 1 H), 7.71 (s, 1 H); $^{13}$C NMR (75 MHz, CDCl$_3$): $\delta$ 14.0, 22.5, 24.4, 25.5, 26.3, 26.9, 29.11, 29.14, 30.1, 30.8, 31.7, 42.8, 46.4, 47.3, 107.6, 110.8, 112.3, 121.4, 124.3, 127.2, 129.2, 134.8, 169.4.
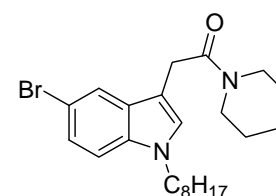
**General procedure for the synthesis of the tertiary amides (3-3 to 3-6).**
The tertiary amides (**3-3** to **3-6**) was synthesized according to the method described in

Li et al.[79] To a solution of the acetamide **5** (1 mmol, 1 equiv.) in 4 ml dimethoxyethane (DME) was added Pd(PPh$_3$)$_4$ (0.05 mmol, 0.05 equiv.) followed by stirring under argon for 15 min. A solution of *m*-tolylboronic acid (1 mmol, 1 equiv.) in 1.5 ml EtOH was added and stirring was continued for another 15 min. A solution of 2 M aqueous sodium carbonate (Na$_2$CO$_3$; 4 mL) was added and the mixture was refluxed for another 5 h under argon. On cooling, the organic solvent was removed under reduced pressure, the resulting suspension was extracted with CH$_2$Cl$_2$ and dried (anhydrous Na$_2$SO$_4$). The residue obtained on removal of the solvent was purified by column chromatography on silica gel with EtOAc/hexane as eluting solvents.

**N,N-dimethyl-2-(1-octyl-5-*m*-tolyl-1*H*-indol-3-yl)acetamide (3-3).** Amber oil; Yield: 48 %; $^1$H NMR (300 MHz, CD$_3$OD): $\delta$ 0.91 (t, $J$ = 12 Hz, 3 H), 1.27-1.31 (m, 10 H), 1.86 (t, $J$ = 9 Hz, 2 H), 2.42 (s, 3 H), 3.10 (s, 6 H), 3.90 (s, 2 H), 4.18 (t, $J$ = 15 Hz, 2 H), 7.09-7.13 (m, 2 H), 7.29 (t, 1 H), 7.40-7.48 (m, 4 H), 7.83 (s, 1 H); $^{13}$C NMR (75 MHz, CD$_3$OD): $\delta$ 13.6, 20.8, 22.8, 27.1, 29.5, 29.6, 30.6, 32.1, 32.5, 35.2, 37.6, 46.3, 107.8, 110.0, 117.5, 121.5, 124.5, 127.1, 127.7, 128.0, 128.7, 128.9, 133.0, 136.6, 138.4, 143.1, 172.0; HRMS (ESI): *m/z* 405.2943 (calcd for [C$_{27}$H$_{36}$N$_2$O + H]$^+$: 405.2828)

**N,N-diethyl-2-(1-octyl-5-*m*-tolyl-1*H*-indol-3-yl)acetamide (3-4).** Amber oil; Yield: 38 %; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 0.89 (t, $J$ = 12 Hz, 3 H), 1.13-1.15 (m, 6 H), 1.26-1.30 (m, 10 H), 1.85 (t, $J$ = 12 Hz, 2 H), 2.44 (s, 3 H), 3.33-3.44 (m, 4 H), 3.84 (s, 2 H), 4.11 (t, $J$ = 15 Hz, 2 H), 7.08-7.14 (m, 2 H), 7.26-7.48 (m, 5 H), 7.78 (s, 1 H); $^{13}$C NMR (75 MHz, CDCl$_3$): $\delta$ 14.1, 14.3, 20.8, 21.6, 22.6, 27.0, 29.3, 29.4, 30.8, 31.8, 32.8, 40.1, 42.7, 46.4, 108.4, 109.6, 117.2, 121.4, 124.5, 126.7, 127.0, 128.1, 128.2, 128.5, 132.7, 135.7, 138.1, 142.6, 170.9; HRMS (ESI): *m/z* 433.3194 (calcd for [C$_{29}$H$_{40}$N$_2$O + H]$^+$: 433.3141)

**2-(1-octyl-5-*m*-tolyl-1*H*-indol-3-yl)-1-(pyrrolidin-1-yl)ethanone** (**3-5**).

Amber oil; Yield: 46 %; $^1$H NMR (300 MHz, CD$_3$OD): $\delta$ 0.91 (t, $J$ = 15 Hz, 3 H), 1.27-1.31 (m, 10 H), 1.84-1.96 (m, 6 H), 2.42 (s, 3 H), 3.49 (t, $J$ = 15 Hz, 2 H), 3.60 (t, $J$ = 15 Hz, 2 H), 3.84 (s, 2 H), 4.19 (t, $J$ = 12 Hz, 2 H), 7.09-7.15 (m, 2 H), 7.32 (t, $J$ = 12 Hz, 1 H), 7.40-7.48 (m, 4 H), 7.83 (s, 1 H); $^{13}$C NMR (75 MHz, CD$_3$OD): $\delta$ 13.6, 20.8, 22.8, 24.5, 26.2, 27.1, 29.5 (2C), 30.6, 32.1, 32.5, 46.4, 50.2, 52.0, 107.8, 110.0, 117.5, 121.5, 124.5, 127.1, 127.7, 128.0, 128.7, 128.9, 133.1, 136.6, 138.4, 143.1, 172.0; MS (APCI): $m/z$ 431.4 [M + H]$^+$.

**2-(1-octyl-5-*m*-tolyl-1*H*-indol-3-yl)-1-(piperidin-1-yl)ethanone** (**3-6**).

Amber oil; Yield: 28 %; $^1$H NMR (300 MHz , CD$_3$OD): $\delta$ 0.90 (t, $J$ = 12 Hz, 3 H), 1.26-1.29 (m, 10 H), 1.47-1.57 (m, 6 H), 1.84 (t, $J$ = 12 Hz , 2 H), 2.42 (s, 3 H), 3.51-3.58 (m, 4 H), 3.90 (s, 2 H), 4.18 (t, $J$ = 15 Hz, 2 H), 7.09-7.12 (m, 2 H), 7.32 (t, $J$ = 15 Hz, 1 H), 7.40-7.48 (m, 4 H), 7.85 (s, 1 H); $^{13}$C NMR (75 MHz, CD$_3$OD): $\delta$ 13.6, 20.8, 22.8, 24.5, 25.9, 26.4, 27.1, 29.5, 29.6, 30.7, 31.6, 32.1, 43.3, 43.4, 46.2, 108.4, 110.0, 117.4, 121.5, 124.5, 127.1, 127.4, 128.0, 128.6, 128.7, 133.1, 136.6, 138.4, 143.0, 171.6; HRMS (ESI): $m/z$ 445.3351 (calcd for [C$_{30}$H$_{40}$N$_2$O + H]$^+$: 445.3141).

### 3.4.4. Synthesis of Series 4 compounds (4-1 and 4-3)

**(1-Octyl-5-*m*-tolyl-1*H*-indol-3-yl)methanamine (4-1).** Synthesis of **4-1** is described in **Section 3.4.5**.
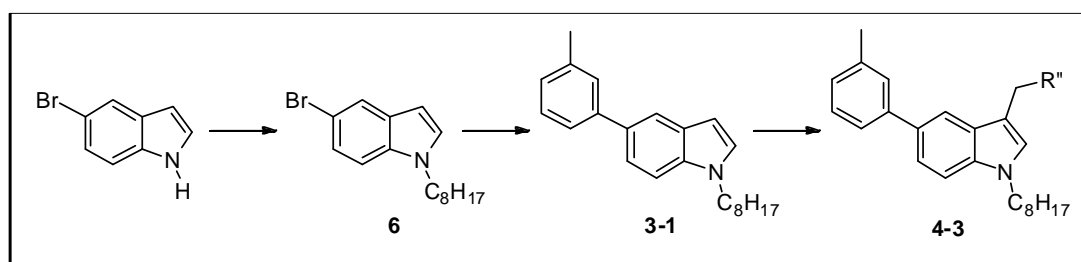
**5-Bromo-1-octyl-1*H*-indole (6).** Compound **6** was synthesized according to the method described in Na et al.[80] NaH (60% in mineral oil dispersion, 6 mmol, 1.2 equiv.) was added to a stirred solution of 5-bromo-1*H*-indole (5 mmol, 1 equiv.) in anhydrous dimethylsulfoxide (DMSO) at room temperature. After 1 h of stirring at the same temperature, 1-bromooctane was added and the mixture stirred for another 3 h. Distilled water was added to the reaction mixture which was then extracted with $CH_2Cl_2$. The organic layer was washed with brine, dried (anhydrous $Na_2SO_4$) and evaporated under reduced pressure. The residue was purified by column chromatography with EtOAc/hexane as eluting solvents. Yield: 92 %; $^1$H NMR (300 MHz, $CDCl_3$): $\delta$ 0.82 (t, $J$ = 6.9 Hz, 3 H), 1.20-1.24 (m, 10 H), 1.76 (t, $J$ = 6.9 Hz, 2 H), 4.03 (t, $J$ = 7.2 Hz, 2 H), 6.37 (d, $J$ = 3 Hz, 1 H), 7.04 (d, $J$ = 3 Hz, 1H), 7.17 (s, 1 H), 7.21 (d, $J$ = 1.2 Hz, 1 H), 7.69 (d, $J$ = 1.5 Hz, 1 H); $^{13}$C NMR (75 MHz, $CDCl_3$) $\delta$ 14.0, 22.5, 26.9, 29.1(2C), 30.1, 31.7, 46.5, 100.4, 110.7, 112.4, 123.3, 124.0, 128.8, 130.1, 134.6.

**1-Octyl-5-*m*-tolyl-1*H*-indole (3-1).** Indole **6** was reacted with *m*-tolylboronic acid as described for the synthesis of amide **3-3** (**Section 3.4.3**). The product was isolated as an amber oil. Yield: 85 %; $^1$H NMR (300 MHz, $CD_3OD$): $\delta$ 0.91 (t, $J$ = 18 Hz, 3 H), 1.20-1.30 (m, 10 H), 1.85 (t, $J$ = 15 Hz, 2 H), 2.41 (s, 3 H), 4.17 (t, $J$ = 12 Hz, 2 H), 6.48 (d, $J$ = 3 Hz, 1 H), 7.10 (d, $J$ = 6 Hz, 1 H), 7.20 (d, $J$ = 3 Hz, 1 H), 7.31 (t, $J$ = 15 Hz, 1 H), 7.40-7.45 (m, 4 H), 7.76 (s, 1 H); $^{13}$C NMR (75MHz, $CD_3OD$): $\delta$ 13.6, 20.8, 22.8, 27.1, 29.4, 29.5, 30.6, 32.1, 46.4, 101.3, 109.8, 119.2, 121.1, 124.4, 127.0, 128.0, 128.7, 128.9, 129.8, 133.0, 136.2, 138.4, 143.2; MS (ESI): *m/z* 320.3 [M + H]$^+$.

***N*-ethyl-*N*-((1-octyl-5-*m*-tolyl-1*H*-indol-3-yl)methyl)ethanamine (4-3).** Diethylamine (1 mmol, 1equiv.), zinc chloride (ZnCl$_2$; 1.5 mmol, 1.5 equiv.),

formaldehyde (HCHO; 1 mmol, 36% aq., 1 equiv.), indole **3-1** (1 mmol, 1equiv.) and EtOH (3 mL) were stirred together in a round bottom flask for 10 h at room temperature. Distilled water was added to the mixture which was made alkaline by the addition of 4 M sodium hydroxide (NaOH). The mixture was extracted with EtOAc, the solvent was removed under reduced pressure and the residue purified by column chromatography with $CH_2Cl_2$/MeOH or EtOAc/hexane as eluting solvents. The product was isolated as an amber oil. Yield: 68 %; $^1H$ NMR (300 MHz, CDCl$_3$): $\delta$ 0.84 (t, $J$ = 4.8 Hz, 3 H), 1.08 (t, $J$ = 7.2 Hz, 6 H), 1.28-1.23 (m, 10 H), 1.81 (t, $J$ = 6.9 Hz, 2 H), 2.41 (s, 3 H), 2.56 (q, $J$ = 6.9 Hz, 4 H), 3.81 (s, 2 H), 4.06 (t, $J$ = 7.2 Hz, 2 H), 7.04 (s, 1 H), 7.09 (d, $J$ = 6.9 Hz, 1 H), 7.23 (s, 1 H), 7.30 (s, 1 H), 7.44-7.39 (m, 3 H), 7.88 (s, 1 H); $^{13}C$ NMR (75 MHz, CDCl$_3$): $\delta$ 142.8, 138.0, 135.7, 132.4, 129.1, 128.5, 128.2, 127.7, 126.9, 124.5, 121.1, 118.1, 112.3, 109.4, 47.8, 46.6, 46.3, 31.7, 30.2, 29.2, 29.2, 27.0, 22.6, 21.6, 14.0, 12.1; MS (APCI): $m/z$ 404.9 [M + H]$^+$. Found (calcd. for $C_{28}H_{48}N_2$) C 83.12 (83.11), H 9.59 (9.96).

### 3.4.5. Synthesis of Series 5 compounds
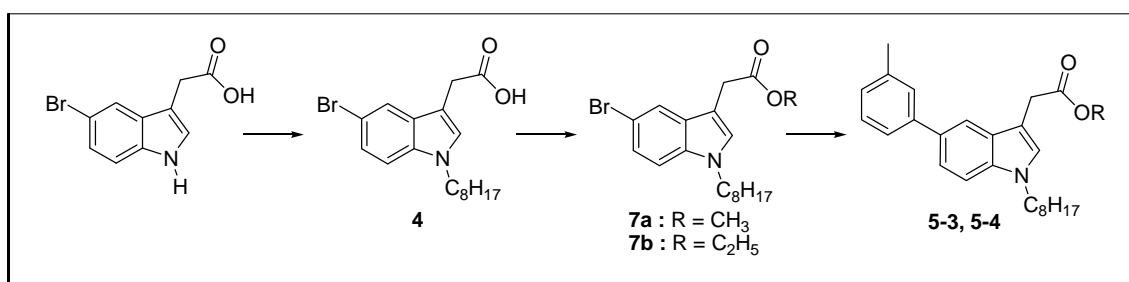


**General method for the preparation of alkyl esters of 2-(5-bromo-1-octyl-1*H*-indol-3-yl) acetic acid (7).** To a solution of acid **4** (4 mmol, 1 equiv.) in 20 mL of the required alcohol (MeOH or EtOH) was added 10 drops of concentrated sulphuric acid ($H_2SO_4$). The mixture was heated under reflux with continuous stirring for 1-1.5 h. The reaction mixture was then cooled to room temperature and evaporated in

vacuo. The residue was treated with a saturated solution of NaHCO$_3$ (50 mL), extracted with EtOAc and dried with anhydrous Na$_2$SO$_4$. Removal of the solvent in vacuo gave a residue that was purified by column chromatography with CH$_2$Cl$_2$ as eluting solvent.

**Methyl-2-(5-bromo-1-octyl-1*H*-indol-3-yl)acetate (7a).** Yield: 95 %; ^{1}H NMR (300 MHz, CDCl$_3$): $\delta$ 0.86 (t, $J$ = 6 Hz, 3 H), 1.24-1.28 (m, 10 H), 1.78 (t, $J$ = 6.6 Hz, 2 H), 3.70 (s, 2 H), 3.71 (s, 3 H), 4.01 (t, $J$ = 6.9Hz, 2 H), 7.07 (s, 1 H), 7.16 (d, $J$ = 8.7 Hz,1H), 7.26 (d, $J$ = 8.7 Hz, 1H), 7.70 (s, 1 H); ^{13}C NMR (75 MHz, CDCl$_3$): $\delta$ 14.0, 22.5, 26.9, 29.1, 29.14, 30.1, 30.8, 31.7, 46.4, 52.0, 106.3, 110.9, 112.4, 121.5, 124.4, 127.8, 129.3, 134.8, 172.2.

**Ethyl-2-(5-bromo-1-octyl-1*H*-indol-3-yl)acetate (7b).** Yield: 95 %; ^{1}H NMR (300 MHz, CDCl$_3$): $\delta$ 0.86 (t, $J$ = 6.9 Hz, 3 H), 1.24-1.27 (m, 13 H), 1.78 (t, $J$ = 6.6 Hz, 2 H), 3.68 (s, 2 H),4.01 (t, $J$ =6.9 Hz ,2 H),4.16 (q, $J$ = 6.9 Hz, 2H), 7.07 (s, 1 H), 7.15 (d, $J$ = 8.7 Hz,1H), 7.25 (d, $J$ = 8.7 Hz, 1H), 7.72 (s, 1 H); ^{13}C NMR (75 MHz, CDCl$_3$): $\delta$ 14.0, 14.1, 22.5, 26.9, 29.11, 29.15, 30.1, 31.1, 31.7, 46.4, 60.8, 106.5, 110.8, 112.3, 121.7, 124.3, 127.8, 129.3, 134.8, 171.7.

**Methyl 2-(1-octyl-5-*m*-tolyl-1*H*-indol-3-yl)acetate (5-3).** The ester **7a** was reacted with *m*-tolylboronic acid as described for the synthesis of amide **3-3** (**Section 3.4.3**). The product was isolated as an amber oil. Yield: 47 %; ^{1}H NMR (300 MHz, CDCl$_3$,): $\delta$ 0.85 (t, $J$ = 12 Hz, 3 H), 1.26-1.32 (m, 10 H), 1.82 (t, $J$ = 6 Hz, 2 H), 2.44 (s, 3 H), 3.70 (s, 2 H), 3.81 (s, 3 H), 4.06 (t, $J$ = 15 Hz, 2 H), 7.11 (d, $J$ = 3 Hz, 2 H), 7.32-7.37 (m, 2 H), 7.44 (d, $J$ = 6 Hz, 3 H), 7.78 (s, 1 H); ^{13}C NMR (75 MHz, CDCl$_3$): $\delta$ 14.1, 21.6, 22.6, 27.0, 29.1, 29.2, 30.3, 31.1, 31.8, 46.5, 52.0, 107.0, 109.6, 117.5, 121.5, 124.5, 127.0, 127.3, 128.1, 128.2, 128.5, 132.8, 135.7, 138.1, 142.6, 172.6; MS (APCI): *m/z* 392.5 [M +H]$^+$.

**Ethyl 2-(1-octyl-5-*m*-tolyl-1*H*-indol-3-yl)acetate (5-4).** The ester **7b** was reacted with *m*-tolylboronic acid as described for **3-3** (**Section 3.4.3**). The product was isolated as an amber oil. Yield: 38 %; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 0.85 (t, *J* = 12 Hz, 3 H), 1.19-1.31 (m, 13 H), 1.82 (t, *J* = 6 Hz, 2 H), 2.43 (s, 3 H), 3.79 (s, 2 H), 4.06 (t, *J* = 9 Hz, 2 H), 4.13 (q, *J* = 21 Hz, 2 H), 7.12 (d, *J* = 3 Hz, 2 H), 7.32-7.36 (m, 2 H), 7.44 (d, *J* = 6 Hz, 3 H), 7.79 (s, 1 H); $^{13}$C NMR (75 MHz, CDCl$_3$): $\delta$ 14.0, 14.2, 21.6, 22.6, 27.0, 29.1, 29.1, 30.3, 31.3, 31.7, 46.5, 60.7, 107.2, 109.6, 117.6, 121.4, 124.5, 127.0, 127.3, 128.2, 128.5, 128.7, 132.7, 135.7, 138.1, 142.6, 172.1; HRMS (APCI): *m/z* 406.2882 (calcd for [C$_{27}$H$_{35}$NO$_2$ + H]$^+$: 406.2668).
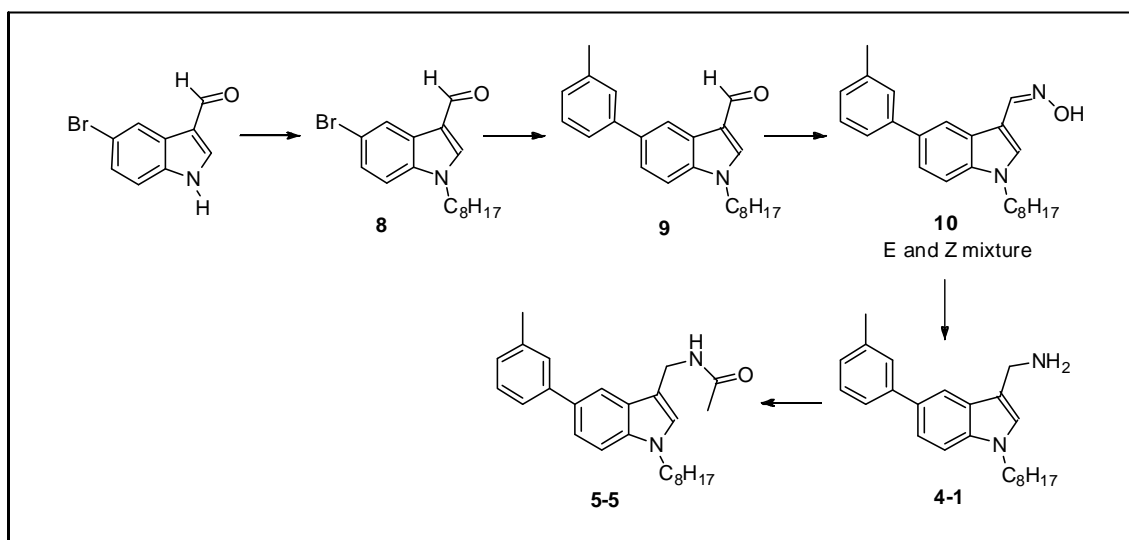


**1-Octyl-1*H*-indole-3-carbaldehyde (8).** 5-Bromo-1*H*-indole-3-carbaldehye was reacted with 1-bromooctane as described for the synthesis of indole **6**. Yield: 85 %; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 9.91 (s, 1 H), 8.43 (s, 1 H), 7.66 (s, 1 H), 7.36 (d, *J* =7.5 Hz, 1 H), 7.20 (d, *J* = 8.7, 1 H), 4.11 (t, *J* = 7.2 Hz, 2 H), 1.85 (t, *J* = 6.3 Hz, 2 H), 1.30-1.24 (m,10 H), 0.86 (t, J = 6 Hz, 3 H); $^{13}$C NMR (75 MHz, CDCl$_3$): $\delta$ 184.1, 138.7, 135.8, 126.8, 126.8, 124.7, 117.3, 116.4, 111.4, 47.4, 32.7, 31.6, 29.6, 29.0, 26.7, 22.5, 14.0.

**1-Octyl-5-*m*-tolyl-1*H*-indole-3-carbaldehyde (9).** Aldehyde **8** was reacted with *m*-tolylboronic acid as described for the synthesis of **3-3** (**Section 3.4.3**). The product was isolated as an amber oil. Yield: 62 %; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 9.98 (s, 1 H), 8.53 (s, 1 H), 7.66 (s, 1 H), 7.57-7.46 (m, 3 H), 7.39-7.29 (m, 2 H), 7.13 (d, *J* = 7.2 H, 1 H), 4.1 (t, *J* = 8.7 Hz, 2 H), 2.42 (s, 3 H), 1.86 (t, J = 3 H), 1.29-1.24 (m, 10 H), 0.86 (t, J = 6.6 Hz, 3 H); $^{13}$C NMR (75 MHz, CDCl$_3$): $\delta$ 184.8, 143.7, 141.6, 138.9, 135.4, 135.1, 129.1, 128.9, 127.1, 119.3, 119.2, 118.5, 116.8, 111.6, 47.4, 32.7, 31.6, 29.6, 29.0, 26.7, 22.5, 14.0.

**(E & Z)-1-Octyl-5-*m*-tolyl-1*H*-indole-3-carbaldehyde oxime (10).** Compound **10** was synthesized according to the method described in Jansen et al.[81] Aldehyde **8** (1.03 mmol, 1equiv.) was refluxed with hydroxylamine hydrochloride (NH$_2$OH.HCl; 2.57 mmol, 2.5 equiv.) and pyridine (2.67 mmol, 2.6 equiv.) in EtOH (20 mL). The reaction was monitored by TLC for the disappearance of **9**, which occurred after 2 h. The mixture was cooled to room temperature, diluted with water, acidified with 10% HCl and extracted with Et$_2$O. The ethereal layer was washed successively with 10% HCl and water, dried (anhydrous Na$_2$SO$_4$) and the solvent removed under reduced pressure. The residue was triturated with petroleum ether and filtered to give **10** in 74% crude yield. It was used without further purification for the next step. $^1$H NMR (300 MHz, DMSO-$d_6$): $\delta$ 0.81 (t, *J* = 9 Hz, 3H), 1.22-1.26 (m, 10H), 1.76 (t, *J* = 6 Hz, 2 H), 2.39 (s, 3 H), 4.12 (t, *J* = 15 Hz, 2 H), 7.12 (d, *J* = 9 Hz, 1 H), 7.32 (t, *J* = 15 Hz, 1 H), 7.40-7.44 (m, 2 H), 7.48 (d, *J* = 12 Hz, 1 H), 7.57 (d, *J* = 9 Hz, 1 H), 7.69 (s, 1 H,), 8.24 (m, *J* = 12 Hz, 2 H), 10.60 (s, 1 H);

**1-Octyl-5-*m*-tolyl-1*H*-indol-3-yl)methanamine (4-1).** To a stirred solution of nickel chloride (NiCl$_2$.6H$_2$O; 0.76 mmol, 1 equiv.) in MeOH (12 mL) was added **10** (0.76 mmol, 1 equiv.) and NaBH$_4$ (4.16 mmol, 5.5 equiv.) in one portion. After 5 min

of stirring, the black precipitate was removed by vacuum filtration and the filtrate concentrated *in vacuo* to approximately a third of its original volume. It was then poured into 20 mL of water containing 3 mL of ammonia solution, extracted with EtOAc and dried (anhydrous $Na_2SO_4$). The solvent was removed under reduced pressure to give **4-1** as a viscous, dark colored oil. Purification using flash silica gel column chromatography by elution with $CH_2Cl_2$/MeOH yielded the corresponding amine **4-1** as an amber oil. Yield: 54 %; $^1$H NMR (300 MHz, DMSO-$d_6$): $\delta$ 0.82 (t, $J$ = 9 Hz, 3H), 1.22-1.30 (m, 10H), 1.72 (t, $J$ = 9 Hz, 2 H), 2.39 (s, 3 H), 3.93 (s, 2 H), 4.12 (t, $J$ = 12 Hz, 2 H), 7.10-7.49 (m, 7 H), 7.87 (s, 1 H,); $^{13}$C NMR (75 MHz, DMSO-$d_6$): $\delta$ 14.0, 21.5, 22.6, 27.0, 29.1, 29.2, 30.3, 31.7, 46.4, 57.6 109.7, 117.2, 121.5, 124.4, 126.6, 127.1, 127.3, 128.1, 128.5, 132.8, 135.9, 138.1, 142.4; MS (APCI): $m/z$ 348.3 $[M + H]^+$.

**N-((1-octyl-5-*m*-tolyl-1*H*-indol-3-yl)methyl)acetamide** **(5-5).** Acetyl chloride (0.0.628 mmol, 1 equiv.) was added to a solution of amine **4-1** (0.628 mmol, 1 equiv.) and triethylamine (0.942 mmol, 1.5 equiv.) in THF (4 mL) at 0 °C. The reaction mixture was stirred at room temperature (27°C) for 1 h after which the precipitate that formed was removed by vacuum filtration, the filtrate was concentrated under reduced pressure and the residue purified by column chromatography with $CH_2Cl_2$ as eluting solvent. The product was isolated as an amber oil. Yield: 65 %; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 7.71 (s, 1 H), 7.49-7.05 (m, 6 H ), 7.01 (s, 1 H), 5.73 (bs, 1 H), 4.49 (d, 2 H, $J$ = 4.8 Hz), 4.04 (t, 2 H, $J$ = 6.9 Hz), 2.43 (s, 3 H ), 1.80 (t, 2 H, $J$ = 6.3 Hz), 1.95 (s, 3 H), 1.28-1.24 (m, 10 H), 0.86 (t, 3 H, $J$ = 7.2 Hz); $^{13}$CNMR (75 MHz, CDCl$_3$) $\delta$ 169.9, 142.2, 138.2, 135.9, 133.0, 128.6, 128.1, 127.5, 127.1, 124.6, 124.4, 121.7, 117.3, 111.3, 109.8, 46.4, 35.1, 31.7, 30.3, 29.2, 29.1, 27.0, 23.1, 22.6, 21.5, 14.0; LC-MS (APCI), $m/z$ = 391.1 $[M + H]^+$.
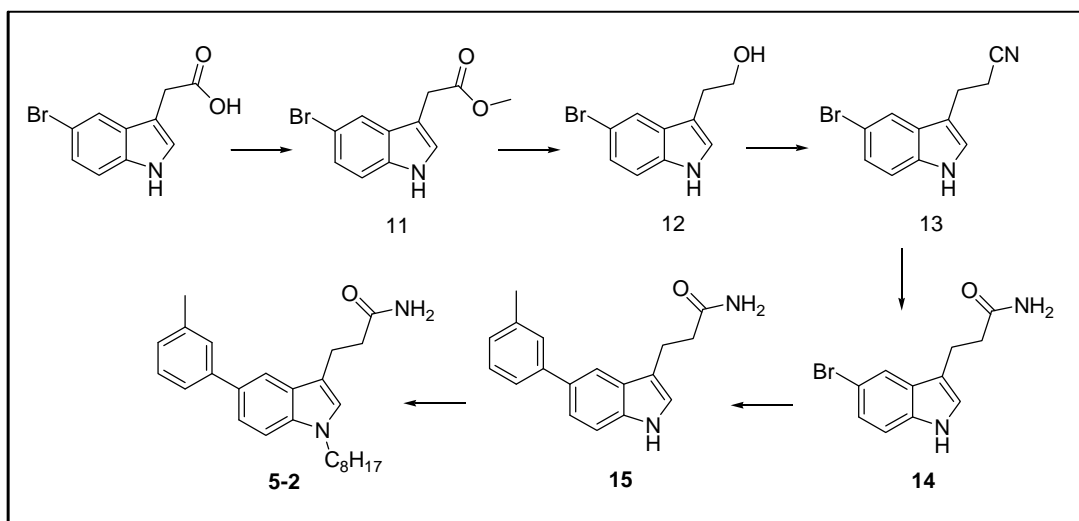
**Methyl-2-(5-bromo-1*H*-indol-3-yl)acetate (11).** To a solution of 2-(5-bromo-1*H*-indol-3-yl)acetic acid (4 mmol) in 20 mL of MeOH was added 10 drops of concentrated sulphuric acid. The mixture was heated under reflux with continuous stirring for a period of 1-1.5 h. Following reflux, the reaction mixture was cooled to room temperature and evaporated in vacuo. The residue was added with a saturated solution of NaHCO$_3$ (50 mL) to neutralize acid, and the resulting alkaline aqueous mixture was extracted with EtOAc and dried over Na$_2$SO$_4$. Evaporation of the solvent gave a residue which was purified by flash silica gel column chromatography. Elution with CH$_2$Cl$_2$ yielded the corresponding ester **11** as an off-white solid. Yield: 89 %; ^{1}H NMR (300 MHz, CDCl$_3$): $\delta$ 3.66 (s, 3 H), 3.68 (s, 2 H), 7.18-7.21 (m, 4 H); MS (APCI), *m/z* 269.2 [M + H][+].

**2-(5-Bromo-1*H*-indol-3-yl)ethanol (12).** Compound **12** and **13** were prepared according to method described in Bascop et al.[82] To a solution of ester **11** (4 mmol, 1 equiv) in anhydrous THF (50 mL) was added lithium aluminium hydride (LiAlH$_4$; 16 mmol, 4 equiv) in portions with the temperature maintained at 0 °C. The reaction mixture was brought to room temperature and stirred for 30 min. Excess LiAlH$_4$ was destroyed by careful addition of a saturated aqueous solution of Na$_2$SO$_4$ with the temperature maintained at about 0 °C. The mixture was filtered and the filtrate

washed with THF. The combined filtrates were concentrated under reduced pressure, the residue was acidified to pH 6 with 10 % HCl and extracted with CHCl$_3$. The combined organic layer was dried (anhydrous Na$_2$SO$_4$), filtered, evaporated to dryness and purified using flash column chromatography by elution with EtOAc/hexane to give the alcohol **12** as an amber oil. Yield: 83 %; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 2.97 (t, $J$ = 18 Hz, 2 H), 3.86-3.93 (m, 2 H), 4.09 (t, $J$ = 21 Hz,1 H), 7.09 (d, $J$ = 3 Hz, 1 H), 7.22 (d, $J$ = 9 Hz, 1 H), 7.75 (s, 1 H), 8.10 (s, 1 H); MS (APCI): $m/z$ 241.2 [M + H]$^+$.

**2-(5-Bromo-1*H*-indol-3-yl)propanenitrile (13).** Methane sulphonyl chloride (MsCl, 0.55mL, 2 equiv) was added to a stirred solution of **12** (3.2 mmol, 1 equiv), triethylamine (0.9 mL, 2 equiv) and anhydrous CH$_2$Cl$_2$ (20 mL) maintained at 0 °C. The reaction mixture was then stirred for 30 min at 0 °C under N$_2$, after which it was treated with 2M NaOH and extracted with CH$_2$Cl$_2$. The combined organic layer was washed with water, dried (anhydrous Na$_2$SO$_4$) and evaporated to dryness under reduced pressure. The residue was dissolved in anhydrous DMSO (20 mL), KCN (3.25 mmol, 3 equiv) was added and the reaction mixture was heated at 100 °C in an oil bath for 1 h. It was diluted with a water-ice mixture, extracted with chloroform, the combined organic layer was washed with water, dried (anhydrous Na$_2$SO$_4$), and evaporated to dryness under reduced pressure. The residue was purified by column chromatography with CHCl$_3$ as eluting solvent and **13** was obtained as an amber oil. Yield: 75 %; $^1$H NMR (300 MHz, DMSO-$d_6$): $\delta$ 2.78 (t, $J$ = 15 Hz, 2 H), 3.58 (t, $J$ = 12 Hz, 2 H), 7.14-7.17 (m, 2 H), 7.20 (d, $J$ = 27 Hz, 1 H), 7.68 (s, 1 H), 11.01 (s, 1 H); MS (APCI): $m/z$ 261.3 [M + H]$^+$.

**2-(5-Bromo-1*H*-indol-3-yl)propanamide (14).** Amide **14** was prepared from the nitrile **13** according to the procedure described for amide **2** (**Section 4.3.2**). Yield:

82 %; ^{1}H NMR (300 MHz, CDCl$_3$): $\delta$ 2.44 (t, $J$ = 15 Hz, 2 H), 2.96 (t, $J$ = 15 Hz, 2 H), 7.09-7.58 (m, 4 H); MS (APCI): $m/z$ 268.1 [M + H]$^+$.

**3-(5-*m*-Tolyl-1*H*-indol-3-yl)propanamide (15).** Amide **15** was prepared from the amide **14** according to the procedure described for amide **3** (**Section 3.4.2**). Yield: 50 %; ^{1}H NMR (300 MHz, CDCl$_3$): $\delta$ 2.38 (s, 3 H), 2.49 (t, $J$ = 18 Hz, 2 H), 3.02 (t, $J$ = 15 Hz, 2 H), 7.12-7.61 (m, 8 H); MS (APCI), $m/z$ 279.4 [M + H]$^+$

**3-(1-Octyl-5-*m*-tolyl-1*H*-indol-3-yl)propanamide (5-2)**. Amide **5-2** was prepared from the amide **15** according to the general procedure described for the synthesis of **1-1** (**Section 3.4.2**). Yield: 77 %; ^{1}H NMR (300 MHz, CDCl$_3$): $\delta$ 0.78 (t, $J$ = 12 Hz, 3 H), 1.18-1.21 (m, 10 H), 1.69 (t, $J$ = 9 Hz, 2 H), 2.35 (s, 3 H), 2.49 (t, $J$ = 15 Hz, 2 H), 3.01 (t, $J$ = 15 Hz, 2 H), 3.92 (t, $J$ =15 Hz, 2 H), 6.84 (s, 1 H), 7.02 (d, $J$ = 9 Hz, 1 H), 7.22 (d, $J$ = 12 Hz, 2 H), 7.35-7.38 (m, 3 H), 7.68 (s, 1 H); ^{13}C NMR (75 MHz, CDCl$_3$): $\delta$ 14.1, 21.0, 21.6, 22.6, 27.0, 29.2, 29.2, 30.3, 31.8, 36.7, 46.3, 109.7, 113.6, 117.3, 121.4, 124.5, 126.1, 127.1, 128.0, 128.2, 128.6, 132.4, 135.9, 138.2, 142.6, 175.4; MS (APCI), $m/z$ = 391.5 [M + H]$^+$.

### 3.4.6. Synthesis of compounds in Series 6 (6-11, 6-12)

**N,N-Diethyl-2-(5-(3-methoxyphenyl)-1-octyl-1*H*-indol-3-yl)acetamide (6-11).** Acetamide **5** was reacted with 3-methoxyphenylboronic acid as described for the synthesis of **3-3** (**Section 3.4.3**). The product was isolated as an amber oil. Yield: 80 %; ^{1}H NMR (300 MHz, CD$_3$OD): $\delta$ 0.90 (t, $J$ = 12 Hz, 3 H), 1.05-1.15 (m, 6 H), 1.27-1.31 (m, 10 H), 1.84 (t, $J$ = 9 Hz, 2 H), 3.40-3.48 (m, 4 H), 3.87 (s, 3 H), 3.88 (s, 2 H), 4.19 (t, $J$ = 15 Hz, 2 H), 6.87 (d, $J$ = 6 Hz, 1 H), 7.14-7.44 (m, 6 H), 7.83 (s, 1 H); ^{13}C NMR (75 MHz, CD$_3$OD): $\delta$ 12.4, 13.5, 13.6, 15.7, 22.8, 27.1, 29.5, 29.6, 30.6, 31.4, 32.1, 40.8, 43.2, 46.3, 54.9, 108.6, 110.0, 111.8, 113.0, 117.5, 119.9., 121.5, 127.5,

128.7, 129.8, 132.8, 136.7, 144.6, 160.7, 172.7; HRMS (ESI): *m/z* 449.3157 (calcd for

$[C_{29}H_{40}N_2O_2 + H]^+$: 449.309).

**2-(5-(3-Ethoxyphenyl)-1-octyl-1*H*-indol-3-yl)-*N,N*-diethylacetamide (6-12).** Acetamide **5** was reacted with 3-ethoxyphenylboronic acid as described for the synthesis of **3-3** (**Section 3.4.3**). The product was isolated as an amber oil. Yield: 74 %; $^1$H NMR (300 MHz, CD$_3$OD,): $\delta$ 0.91 (t, *J* = 15 Hz, 3 H), 1.05-1.15 (m, 6 H), 1.27-1.31 (m, 10 H), 1.46 (t, *J* 15 Hz, 3 H), 1.85 (t, *J* = 9 Hz, 2 H), 3.38-3.49 (m, 4 H), 3.89 (s, 2 H), 4.09-4.19 (m, 4 H), 6.86 (d, *J* 9 Hz, 1 H), 7.14-7.43 (m, 6 H), 7.83 (s, 1 H); $^{13}$C NMR (75 MHz, CD$_3$OD): $\delta$ 12.5, 13.5, 13.6, 14.4, 22.8, 27.1, 29.5, 29.6, 30.6, 31.4, 32.1, 40.8, 43.2, 46.3, 63.7, 108.6, 110.0, 112.4, 113.7, 117.5, 119.8, 121.5, 127.6, 128.7, 129.8, 132.8, 136.7, 144.6, 169.9, 172.8; MS (APCI): *m/z* 463.4 [M + H]$^+$.

## 3.5. SUMMARY

The synthesis, characterization and purification (at least 95 % purity based on combustion analysis or reverse phase HPLC) were successfully achieved by several reported methods. The compounds were novel and have not been previously reported.

The compounds synthesized by Dr Suresh were also novel. Compounds **1-8**, **3-7** , **4-10**, **6-8** and **6-10** were less than 95 % pure (based on HPLC analysis).

.

# CHAPTER 4: INVESTIGATIONS INTO THE ICMT INHIBITORY AND ANTIPROLIFERATIVE ACTIVITIES OF SYNTHESIZED COMPOUNDS

## 4.1. INTRODUCTION

This chapter describes the evaluation of the compounds described in Chapter 3 for inhibition of Icmt. The compounds were also evaluated for their effects on the viability of a human breast cancer cell line. Inhibitory activities were expressed in terms of $IC_{50}$ values (concentration required to reduce enzyme activity or proliferation of treated cells by 50%) and the results were analyzed to (i) reveal structure activity relationships (SAR) in each case, and (ii) ascertain if there was overlap in the structural requirements for the two activities. The significance of the findings is also discussed.

## 4.2. EXPERIMENTAL METHODS

### 4.2.1. Materials for biological assay

Sf9 (*Spodoptera frugiperda* ovarian) membranes containing recombinant Icmt were prepared as described previously.[52] The human breast cancer cell line, MDA-MB-231, was purchased from American Type Culture Collection (ATCC, Rockville, MD). Streptavidin sepharose beads were purchased from Amersham Biosciences (GE Healthcare, UK), S-adenosylmethionine *p*-toluene sulfonate from Sigma-Aldrich (Singapore) and [3H] S-adenosylmethionine (SAM or AdoMet) from Perkin Elmer (Waltham, MA). Biotin S-farnesylcysteine (BFC) was synthesized by the Duke Small Molecule Synthesis Facility [67] and its purity was assessed by mass spectrometry and NMR. CellTiter ® 96 AQueous One Solution was obtained from Promega (Madison USA). Suppliers for other reagents were: fetal bovine serum (Hyclone Laboratories,

Logan, UT), penicillin and streptomycin (Gibco, Invitrogen, Carlsbad, CA), DMEM (Sigma Aldrich, Singapore).

### 4.2.2. Outline of the Icmt inhibition assay

Briefly, the assay involved measuring the rate at which Icmt (expressed on Sf9 membranes) transfered a methyl group from the methyl donor S-adenosyl methionine (AdoMet or SAM) to the substrate biotin-farnesyl-L-cysteine (BFC). BFC corresponds to the C-terminus of a farnesylated, Rce-1-protetolyzed substrate of Icmt and exhibits a Michaelis-Menten constant ($K_m$) of 2.1 μM.[67] The attached biotin moiety facilitated the isolation of the methylated substrate from the reaction mixture with the aid of commercially available streptavidin beads. The methyl group that is donated to the substrate by AdoMet is radiolabelled with $^3$H so that its presence in the substrate can be measured. In the assay, the enzyme and test compound were first incubated to allow binding of the inhibitor to the enzyme. Thereafter, the substrate cocktail of BFC and AdoMet was added to effect the Icmt-catalyzed methylation of BFC. If the test compound inhibited Icmt, the methylation of BFC would be interrupted. The reaction was terminated by adding a surfactant (Tween 20) to break up the membrane and release the methylated BFC which was then bound to streptavidin beads due to the strong affinity of streptavidin for biotin. Once the binding reaction reached an equilibrium (at least 8 hours), the beads were washed to remove any unbound radioactive AdoMet. Only the radioactivity emitted from methylated BFC bound to streptavidin was measured.

### 4.2.3. Measurement of Icmt activity.

The method described by Winter-Vann et al[60] was followed with some modifications. Briefly, the standard reaction mixture contained recombinant Icmt (1.0 μg of Sf9 membrane protein) in a final total volume of 50 μL buffer A (100 mM Hepes buffer pH 7.4, 5 mM $MgCl_2$ and 100 μM EDTA). The test compound (dissolved in DMSO, final concentration of DMSO in the assay mixture was 2 % v/v) was added and the mixture was incubated at 37 °C for 20 min. After this time, the substrate mixture which comprised of 5 μM BFC and 2.5 μM [$^3$H] AdoMet (15 Ci/mmol) in buffer A was added to initiate the reaction (37 °C, 20 min). The final concentrations of BFC and [$^3$H] AdoMet in the mixture were 1 μM and 0.5 μM respectively. The reaction was terminated by the addition of 10% Tween 20 (5 μL) and streptavidin beads (10 μL of packed beads) in 500 μL of buffer B which consisted of 20 mM $NaH_2PO_4$ (pH 7.4) and 150 mM NaCl. The interaction between biotin and streptavidin beads was allowed to proceed overnight at 4 °C with gentle agitation. The beads were harvested by centrifugation (1000 x g, 5 min) on a tabletop microcentrifuge and washed 3x with 500 μL buffer B. The beads were then suspended in 500 μL of the same buffer and transferred to scintillation vials (each with 4.5 ml scintillation fluid) for the determination of radioactivity. Radioactivity was measured on the LS6500 multipurpose scintillation counter (Beckmann Coulter Inc, Fullerton, CA) and data captured by the LS6000 data capture/network software version 2.11 (Beckman Coulter).

### 4.2.4. Outline of the cell viability assay

Colorimetric assays based on tetrazolium compounds are commonly employed to measure the viability of cells. These assays are based on the ability of

mitochondrial reductases of viable cells to reduce the tetrazolium dye to a purple product called formazan. Cells use NADPH or NADH produced from normal cellular activity to achieve this reduction. When cells lose their viability, the reduction fails to occur due to insufficient NADP/NAD. The most common tetrazolium dye employed in viability assays is 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT). The purple formazan crystals that are formed from the reduction of MTT are insoluble and must be dissolved in DMSO for absorbance measurements. In the present investigation, a commercial kit (CellTiter 96$^®$ AQueous One Solution) that contained another tetrazolium compound [3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2$H$-tetrazolium, MTS] was used. MTS is provided together with another reagent phenazine ethosulfate (PES) in the kit. PES has enhanced chemical stability which allows it to be combined with MTS to form a stable solution. When MTS was reduced to its formazan, it is soluble in the culture media. Thus, DMSO need not be added to dissolve the formazan and the assay procedure was reduced to an "add-incubate-measure" format. Moreover, there was greater flexibility as the plate could be read and returned to the incubator for further color development if a longer incubation period was necessary.

### 4.2.5. Cell Viability Assay

MDA-MB-231 human breast cancer cells were grown in DMEM supplemented with 10% fetal bovine serum, 50 U/ml penicillin and 50 μg/ml streptomycin at 37 °C, 5 % $CO_2$. Cells were subcultured at 80-90 % confluency and used within 10-25 passages for the assays. For the assay, cells were seeded at 2400 cells per well in DMEM containing 5 % FBS in 96-well plates for 24 h. Aliquots of test compounds (various concentrations, prepared in DMSO stock solutions) were

then added to each well (final concentration of DMSO maintained at 0.5 % v/v per well) and incubated for 72 h. Control wells contained cells and DMSO at the same concentration as those in test wells. At the end of the incubation period, 20 μL of CellTiter® 96 AQueous One Solution was added to each well and the plates were incubated in the dark for 2 h at 37 °C before absorbance readings at 490 nm were taken. Background absorbance readings from blank wells that contained only media and DMSO were subtracted from each well. Each condition was determined in triplicate. Cell viability was expressed as % of absorbance at 490 nm (treated cells) over that of control (untreated) cells. The concentration of test compound required to reduce cell growth by 50 % ($IC_{50}$ value) was determined from the sigmoidal curve obtained by plotting % cell viability versus concentration with GraphPad Prism (Version 4.0, GraphPad Software, San Diego, CA). $IC_{50}$ values were the mean of at least two separate determinations.

## 4.3. RESULTS

**Table 4-1** summarizes the $IC_{50}$ values for Icmt inhibition and antiproliferative activity for the test compounds (Series 1-6).

**Table 4-1. IC$_{50}$ values for Icmt inhibition and antiproliferative activities of test compounds.** Standard deviations (± SD) are provided for compounds that had more than two separate determinations. Values in parentheses are Hill slopes determined from dose-response curves constructed with GraphPad Prism.
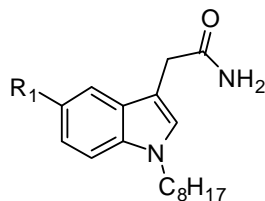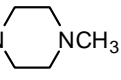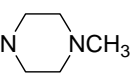


| Compound | R$_1$ | IC$_{50}$ (µM) | |
| --- | --- | --- | --- |
| | | Icmt inhibition [a] | Antiproliferative activity [a] |
| **1-1** | *m*-tolyl | 1.5 ±0.2 [2.3] | 21.8 ±0.8 [9.4] |
| **1-2** | *o*-tolyl | 1.0 [1.8] | 21.2 ±1.1 [9.1] |
| **1-3** | *p*-tolyl | 1.0 [1.0] | NA[b] |
| **1-4** | 3-methoxyphenyl | 1.9 [2.7] | 22.4 ±1.2 [7.4] |
| **1-5** | 3-ethoxyphenyl | 1.3 [2.2] | 18.7 ±1.0 [7.6] |
| **1-6** | phenyl | 1.8 [2.3] | 26.0 ±1.5 [8.6] |
| **1-7** | -H | 6.5 [1.3] | 50 ±4 [4.0] |
| **1-8** | -F | 7.0 [1.0] | 70 ±20 [3.1] |



| Compound | R$_2$ | IC$_{50}$ (µM) | |
| --- | --- | --- | --- |
| | | Icmt inhibition [a] | Antiproliferative activity [a] |
| **2-1** | -H | 33 [NA] | > 100 [NA] |
| **2-2** |  | 2.5 [2.4] | 17.2 ±2.3 [3.6] |
| **2-3** |  | 7.7 [2.4] | 28.5 [3.7] |
| **2-4** |  | 1.1 [2.5] | 10.6 [8.0] |

| Compound | R$_3$ | IC$_{50}$ (µM) | |
| :---: | :---: | :---: | :---: |
| | | Icmt inhibition [a] | Antiproliferative activity [a] |
| **3-1** | -H | > 100 [NA] | > 100 [NA] |
| **3-2** | -CH$_2$CONHCH$_3$ | 2.1 [1.8] | 4.7 [8.7] |
| **3-3** | -CH$_2$CON(CH$_3$)$_2$ | 1.5 [1.3] | 31 [6.5] |
| **3-4** | -CH$_2$CON(C$_2$H$_5$)$_2$ | 1.4 ±0.8 [0.7] | 27.0 ±6.7 [2.4] |
| **3-5** | —CH$_2$CON⟨pyrrolidine⟩ | 1.2 [2.0] | 15.6 ± 2.7 [3.0] |
| **3-6** | —CH$_2$CON⟨piperidine⟩ | 1.8 ± 0.8 [0.8] | 28.9 ±5.2 [2.3] |
| **3-7** | —CH$_2$CON⟨NCH$_3$⟩ | 1.7 [1.8] | 19.7 ±0.9 [16.4] |
| **4-1** | -CH$_2$NH$_2$ | 0.7 [2.2] | 2.9 [12.7] |
| **4-2** | -CH$_2$N(CH$_3$)$_2$ | 1.3 [1.5] | 6.8 [31.9] |
| **4-3** | -CH$_2$N(C$_2$H$_5$)$_2$ | 0.7 ±0.1 [2.1] | 3.6 ±0.1 [13.0] |
| **4-4** | -CH$_2$N($i$-C$_3$H$_7$)$_2$ | 1.7 [1.0] | 20.1 [2.7] |
| **4-5** | -CH$_2$N($n$-C$_3$H$_7$)$_2$ | 0.8 [1.2] | 13.0 [6.7] |
| **4-6** | -CH$_2$NCH$_3$($i$-C$_3$H$_7$) | 0.9 [1.8] | 5.3 [12.5] |
| **4-7** | —CH$_2$N⟨pyrrolidine⟩ | 0.5 [1.5] | 5.1 ± 1.2 [4.5] |
| **4-8** | —CH$_2$N⟨piperidine⟩ | 0.7 [2.3] | 4.9 ±0.2 [9.2] |
| **4-9** | —CH$_2$N⟨NCH$_3$⟩ | 0.9 [1.8] | 6.1 [19.1] |
| **4-10** | —CH$_2$N⟨O⟩ | 2.7 [1.2] | 24.5 ±2.3 [6.4] |
| **5-1** | -CONH$_2$ | 30 [NA] | >100 [NA] |
| **5-2** | -CH$_2$CH$_2$CONH$_2$ | 1.2 [1.6] | 20.7 ± 0.8 [11.1] |

| | | | |
|---|---|---|---|
| **5-3** | $-CH_2COOCH_3$ | 76 [NA] | >100 [NA] |
| **5-4** | $-CH_2COOC_2H_5$ | >100 [NA] | >100 [NA] |
| **5-5** | $-CH_2NHCOCH_3$ | 1.8 [1.6] | 30 ±3 [3.0] |
| **5-6** | $-CH_2NHSO_2CH_3$ | 1.2 [1.7] | 17.5 ± 0.9 [9.7] |



| Compound | R$_1$ | R$_2$ | R$_3$ | IC$_{50}$ (μM) | |
|---|---|---|---|---|---|
| | | | | Icmt inhibition [a] | Antiproliferative activity [a] |
| **6-1** | o-tolyl | $n$-C$_8$H$_{17}$ | $-CH_2N(C_2H_5)_2$ | 0.6 ±0.2 [1.5] | 5.5 ± 0.3 [9.9] |
| **6-2** | p-tolyl | $n$-C$_8$H$_{17}$ | $-CH_2N(C_2H_5)_2$ | 0.6 ±0.2 [1.7] | 3.4 ±0.1 [11.8] |
| **6-3** | m-tolyl | | -H | >100 [NA] | >100 [NA] |
| **6-4** | m-tolyl | | $-CH_2N(C_2H_5)_2$ | 2.4 ±0.3 [1.6] | 3.8 ± 0.3 [7.9] |
| **6-5** | m-tolyl | | $-CH_2N$⟨piperidine⟩ | 2.1 ±0.7 [1.4] | 3.9 ± 0.4 [6.1] |
| **6-6** | m-tolyl | | $-CH_2N$⟨morpholine⟩ | 2.0 [1.5] | 34 [3.9] |
| **6-7** | -H | | $-CH_2N(C_2H_5)_2$ | 67 [NA] | 74 [5.9] |
| **6-8** | -F | | $-CH_2CONH_2$ | 69 [NA] | >100 [NA] |
| **6-9** | -F | $n$-C$_8$H$_{17}$ | $-CH_2N(C_2H_5)_2$ | 4.1 [1.6] | 7.2 [11.2] |
| **6-10** | -F | | $-CH_2N(C_2H_5)_2$ | 35 ±6 [NA] | 32 [6.1] |
| **6-11** | 3-methoxy phenyl | $n$-C$_8$H$_{17}$ | $-CH_2CON(C_2H_5)_2$ | 0.8 [0.6] | 52 [1.0] |
| **6-12** | 3-ethoxy phenyl | $n$-C$_8$H$_{17}$ | $-CH_2CON(C_2H_5)_2$ | 1.0 [0.5] | 26.8 [1.0] |
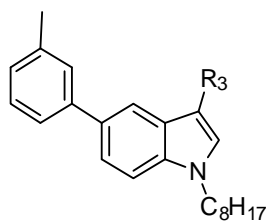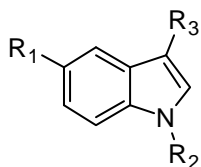
[a]Mean of two or more separate determinations; The results where SD are not provided were from two separate determinations where the difference between independent results and the mean were 10-40% (average: 27.3%) for the Icmt inhibition assay and 2-25% (average: 9.3%) for the antiproliferative assay

[b]Not determined due to poor solubility

### 4.3.1. Effect of structural modifcations on Icmt inhibitory activity: Position 5

Position 5 of cysmethynil (**1-1**) was occupied by a *m*-tolyl ring. Changes in the substitution pattern of the 5-phenyl ring, namely switching the position of the *m*-methyl to *o*- and *p*- positions (**1-2, 1-3**), replacing *m*-methyl with *m*-methoxy (**1-4**) or m-ethoxy (**1-5**), were generally well-tolerated and caused only incremental changes in inhibitory activity. An unsubstituted 5-phenyl ring (**1-6**) fared just as well but omitting the ring altogether (**1-7**) caused a modest 4-fold drop in activity. Introducing 5-fluoro (**1-8**) in place of 5-phenyl caused a similar loss of activity. The good agreement between the predicted ($\approx$ 3 $\mu$M) and experimentally determined (2 $\mu$M) IC$_{50}$ values of the methoxy analogue (**1-4**) provided some support for the SAR derived in **Chapter 2**.

### 4.3.2. Effect of structural modifcations on Icmt inhibitory activity: Position 1

The compounds in Series 2 exemplified the modifications made at position 1. The most notable observation was the 20-fold drop in activity when the *n*-octyl chain was removed (**2-1**). This was significantly greater than the decrease in activity brought about by removing the 5-phenyl ring (**1-7**). In **Section 3.2.3**, it was proposed that the isoprenyl analogue **2-3** should be active because its length was comparable to the *m*-trifluoromethylbenzyl side chain of **2-2**. On the other hand, the activity of the geranyl side chain analogue (**2-4**) was in doubt because of its questionable flexibility and length. As it turned out, the geranyl analogue **2-4** was almost as active as cysmethynil while the isoprenyl analogue **2-3** was about 5-fold weaker. Thus, it may be that the measured lengths (determined in silico from energy minimized structures) were not accurate or that length of side chain was not the main determinant of activity. Noting that lipophilicity at this position was highlighted in the SAR

described in **Chapter 2**, the lower activity of the isoprenyl analogue **2-3** could be due to its lower lipophilic character (100 fold less than cysmethynil based on ClogP values) while the more lipophilic geranyl analogue **2-4** (ClogP of 7) was as potent as cysmethynil.

### 4.3.3. Effect of structural modifcations on Icmt inhibitory activity: Acetamide side chain at position 3

The removal of the acetamide side chain resulted in a compound (**3-1**) with no measurable Icmt inhibitory. This was observed in spite of the significantly greater lipophilic character of **3-1** (ClogP 8.7). Thus, the poor Icmt inhibitory activity of **3-1** must stem from the intrinsic loss of the acetamide functionality and not to lipophilicity changes. That the latter did not play a significant role was further seen in compounds **3-1** and **3-4** which had very different inhibitory activities ($IC_{50}$ value of > 100 μM versus 1.4 μM) but comparable lipophilicities (ClogP of 8.7 versus 8.6).

As mentioned in **Section 3.2.4**, the acetamide moiety can function as H bond donor and H bond acceptor. The importance of the H bond donor capability of the side chain was investigated in the Series 3 compounds. Conversion of the primary amide to a secondary amide (**3-2**) and various tertiary amides (**3-3** to **3-7**) reduced the number of H bond donor groups. These changes did not dramatically affect inhibitory activity which remained within a narrow range of $IC_{50}$ values (1.2 to 2.1 μM) in spite of an almost 1000-fold variation in lipophilicity (ClogP varied from 7.2 to 10.3) among the compounds. These observations suggested that the H bond donor property of the acetamide side chain did not contribute to Icmt inhibition, and that changes in the lipophilic character of this functionality had little influence on activity.

Next, the role of the side chain as a H bond acceptor was investigated by removing the carbonyl functionality to give the amines of Series 4. Most of the

amines were tertiary amines (**4-2** to **4-10**) but there was one primary amine (**4-1**). Most notably, only Series 4 yielded compounds that were more potent inhibitors than cysmethynil. The increase in potency was modest (no more than 3-fold) and was observed for most of the amines (**4-1**, **4-3**, **4-5** to **4-9**). Branching of alkyl groups at the basic nitrogen was not a desirable feature was seen from the poor activity of the di(isopropyl) amino analogue **4-4** as compare to the di(*n*-propyl)amino (**4-3**) and the *N*-methyl-*N*-isopropylamino (**4-6**) analogues. The good activities of the heterocyclic amines (**4-7** to **4-9**) in which the substituents on N were cyclized further emphasized the need for a "sterically compact" basic centre. The exception was the morpholino analogue **4-10** and this may be due to other factors like its lower basicity owing to the presence of the electron withdrawing morpholino oxygen. Taken together, these findings showed that omitting the amide carbonyl to give amine analogues of cysmethynil generally improved or maintained inhibitory activity. Whether this meant that the H bond acceptor properties of the carbonyl oxygen was not required for activity remained questionable because the resulting amine could function as a H bond acceptor if it was not protonated. Going by the basic character of the amines present in the side chain, the protonated form was likely to predominate at physiological pH but this need not necessarily mean that the protonates species was also the "active" species interacting with the enzyme.

The compounds in Series 5 were homologues or isosteres of cysmethynil. A significant observation was the poor activity of the short chain homologue **5-1** compared to the longer chain homologue **5-2**. A possible explanation may be that the side chain at position 3 must have a minimum length (probably equivalent to that found in cysmethynil) to interact with a putative residue on the Icmt active site. Flexibility may also be an important factor as the side chain of **5-1** had limited

flexibility (**Section 3.2.4**). Thus the poor activity of the short chain homologue may be attributed to its less than optimal length and limited flexibility, features that were not found in the longer chain homologue.

Of the isosteres, the esters fared poorly but the retroamide (**5-5**) and sulphonamide (**5-6**) were as active as cysmethynil. The esters were proposed as inhibitors because they resembled both the substrate and product of Icmt. Their poor inhibitory activities suggested that this proposal is not tenable. It may be that the esters were hydrolyzed under the experimental conditions to give a common (and inactive) carboxylic acid intermediate. This would explain the broadly similar $IC_{50}$ values of the two compounds for both Icmt inhibition and antiproliferative activities.

The good activities of the retroamide and sulphonamide derivatives showed that shifting the H bond acceptor carbonyl or sulphonyl moieties further away from the indole ring did not adversely affect activity. In fact, these two side chains were similar to the side chain of the longer homologue **5-2** of cysmethynil in terms of length and relative positions of the electron rich carbonyl/sulphonyl moieties.

### 4.3.4. Effect of structural modifcations on Icmt inhibitory activity: Series 6 compounds

Unlike Series 1 to 5, the compounds in Series 6 were modified at two or more positions of the indole ring.  A convenient way of discussing this series is to divide it into 2 categories, namely those with *n*-octyl or isoprenyl at position 1. Of the compounds with *n*-octyl, modifications were made at the 5-phenyl ring (position of methyl group, replacement with *m*-methoxy or *m*-ethoxy, 5-fluoro in place of 5-phenyl) and the 3-position (tertiary amide or amine in place of acetamide). The inhibitory activities of the resulting compounds closely reflected the trends observed in cysmethynil. Thus, **6-1** and **6-2** with diethylamino side chains ($IC_{50}$ value of 0.6

μM) were almost as potent as the amines in Series 4 whereas the inhibitory activities of the tertiary amides **6-11** and **6-12** fell within the range of the amide analogues of cysmethynil (Series 3). The activity of **6-9** ($IC_{50}$ value of 4.1 μM) was instructive as it showed that even with a "preferred" amine side chain at position 3, replacing 5-phenyl with 5-fluoro reduced activity to almost the same degree as that observed when cysmethnil was similarly modified to give **1-8** ( $IC_{50}$ value of 7.0 μM). It may be that the *n*-octyl side chain strongly influenced inhibitory activities in spite of modifications at the other two positions. This would explain why activity trends closely mirrored those observed for cysmethynil for the octyl analogues. If this was true, then compounds with isoprenyl at position 1 should exhibit wider variations in their activities. The findings with the isoprenyl analogues are summarized in **Figure 4-1**.
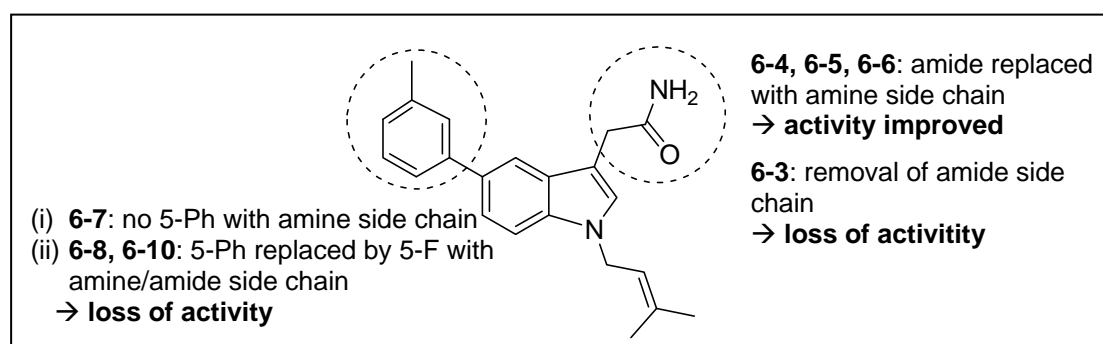


**Figure 4-1. Series 6 compounds with isoprenyl side chain at position 1.** Effect of modifications on Icmt inhibitory activity.

Starting with **2-3** (structure shown in **Figure 4-1**) which had an $IC_{50}$ value of 7.7 μM, replacing the acetamide side chain with various amino side chains (diethylaminomethyl, **6-4**, *N*-piperidinylmethyl, **6-5**, *N*-morpholinomethyl, **6-6**) resulted in a 3 fold improvement in activity but these amine analogues ($IC_{50}$ value of 2.0 to 2.4 μM) were generally not as potent as the Series 4 compounds ($IC_{50}$ value of 0.5 to 2.7 μM). Removing the acetamide side caused the expected large loss in

activity (**6-3**, $IC_{50}$ value of $> 100$ μM). Removing the 5-phenyl ring and replacing it with 5-fluoro while keeping the acetamide and isoprenyl sidechains intact gave **6-8** which was largely inactive as an Icmt inhibitor ($IC_{50}$ value of 69 μM). This loss was significantly greater than that observed when the 5-phenyl ring of cysmethynil was similarly altered to give **1-8** ($IC_{50}$ value of 7.0 μM). The significantly better activity of **1-8** may be attributed to the presence of the *n*-octyl side chain. The role of *n*-octyl in retaining good activity was further observed in **6-9** and **6-10**. **6-9** ($IC_{50}$ value of 4.1 μM) with an *n*-octyl side chain was 9 times stronger as an inhibitor compared to **6-10** ($IC_{50}$ value of 35 μM) which had an isoprenyl side chain.

### 4.3.5. Effect on cell growth/proliferation.

The two activities of Icmt inhibition and cell growth inhibition were significantly correlated to each other when analyzed by Spearman non-parametric correlation analysis (Spearman's rho = 0.707, significant at 0.001 level, 2-tailed). Thus, those compounds with good Icmt inhibitory activities also curtailed cell growth and proliferation. However, two differences were noted. First, anti-proliferative $IC_{50}$ values were generally an order magnitude higher than Icmt $IC_{50}$ values. Second, a wider spread was observed for Icmt $IC_{50}$ values (approximately 100-fold spread in $IC_{50}$ values) compared to antiproliferative $IC_{50}$ values (about 30-fold spread). These differences may be traced to the way the assays were carried out. Icmt activity was evaluated by an *in vitro* assay using Sf9 membranes enriched with the enzyme, whereas effect on cell viability involved incubating test compounds for several days with human cancer cells. In the latter, compounds must move across the cell membrane to gain access to the large number of biological targets in the cell interior. Activity would thus be influenced to a greater degree by the permeability

characteristics of the test compounds. Moreover, the media for cell culture contained components, most notably serum albumin, that could sequester the test compounds by adsorption or binding and this could further influence the results.

### 4.3.6. Hill slopes of dose response curves for determination of $IC_{50}$ values

Dose response curves were constructed for the determination of $IC_{50}$ values for Icmt inhibition and antiproliferative activity. An important feature of the curve was the steepness of the slope from which the $IC_{50}$ value was derived. The steepness of the slope was quantified by the Hill slope (Hill coefficient). Briefly, the Hill slope is used to measure the cooperativity in a binding process. A value greater than 1 was indicative of positive cooperativity where the binding of one ligand facilitated the binding of subsequent ligands on the receptor or enzyme complex. Practically, a large Hill slope signified that the inhibitory activity of the compound increased much more quickly with concentration than would be expected for a classical single site inhibitor. For the latter, inhibition may increase from 10% to 90% over a 80-fold concentration range. For compounds with steep dose response curves, this increase would occur over a narrower 10-fold concentration range. In this investigation, the Hill slopes of the test compounds had a wide range, namely 1.0 to 2.7 for Icmt inhibition and 1.0 to 33 for antiproliferative activity (**Table 4-1**). Representative dose response curves are given in **Figure 4-2** and **4-3**.
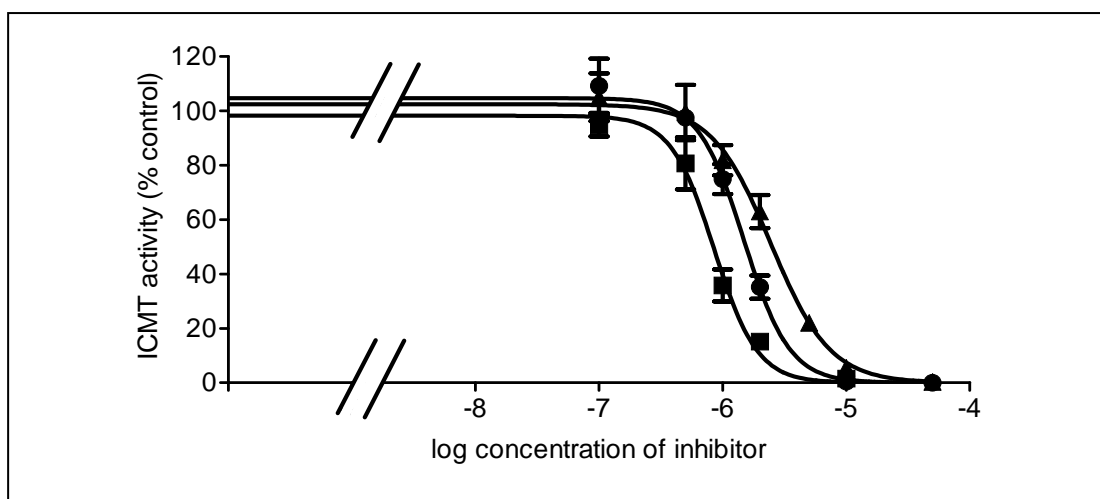
**Figure 4-2. Inhibition of Icmt by cysmethynil and analogues.** Icmt activity was measured as the incorporation of [³H] from [³H]AdoMet into the Icmt substrate BFC. The assay was performed in the presence of various inhibitors (● − cysmethynil, ■ − 4-3, ▲ − 6-4).
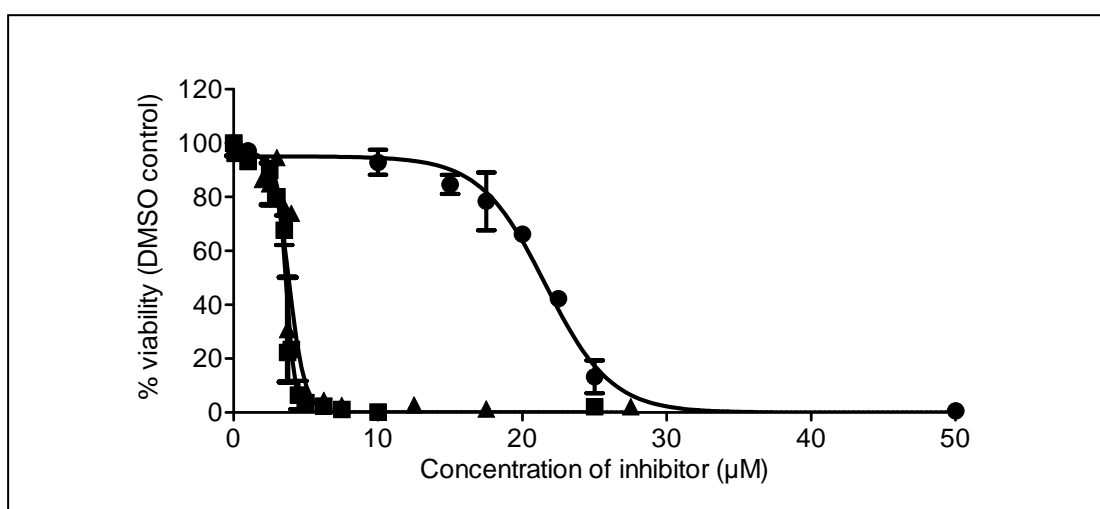


**Figure 4-3. Antiproliferative activity of cysmethynil and analogues.** Cells were seeded at 10% confluency in DMEM containing 5% FBS in 96-well plates for 24hr prior to treatment with the various inhibitors (● − cysmethynil, ■ − 4-3, ▲ − 6-4) at indicated concentrations for 72 hours.

The phenomenon of steep dose response curves and their significance have been discussed at length[46, 83-84]. Briefly, the appearance of steep dose response curves should alert the investigator to 3 possibilities: (i) the presence of multisite binding on the target protein, or a multimeric protein complex; that is, several inhibitor molecules may bind to one target protein/enzyme complex; (ii) the inhibitor may form aggregates (in the range of 50 nm to over 400 nm) and these particles adsorb or

absorb the target protein/enzyme thereby inhibiting them; (iii) the enzyme concentration is greater than the $K_d$ (equilibrium dissociation constant) value of the inhibitor. In this case, inhibition will be observed only when the concentration of the inhibitor approaches that of the enzyme. A steep rise would be observed because of the very high inhibitor concentration (> $K_d$).

Of these 3 possibilities, the second possibility (colloidal aggregation followed by enzyme sequestration) would present the greatest concern because it would mean that the inhibition was non-specific. However, there were several reasons to exclude this possibility for the present test compounds. The Icmt assay of cysmethynil was investigated with different enzyme concentrations varying from 0.5 to 5 μg protein. The change in enzyme concentration did not abolish the steep dose response curves as it should have done if there was non-specific inhibition. Moreover, cysmethynil and **4-3** were investigated for their ability to form micelles and it was found that the critical micelle concentrations of these compounds were in the range of 10 to 25 μM, which exceeded the $IC_{50 \text{ Icmt}}$ values by at least 10 times. If there was non-specific inhibition, a good correlation between Icmt inhibition and antiproliferative activities would be unlikely. Furthermore, recent reports that cysmethynil disrupted the cell cycle and caused cell death, and that overexpression of Icmt would protect cells from these consequences, argue against the operation of a non-specific mechanism.[62] Also noteworthy in this regard are recent findings that Icmt exists in membranes as a homodimer, and the two subunits of the dimer interact such that perturbation of one impacts the other (C. Hrycyna, Purdue Univ, personal communication).

As seen from **Table 4-1** and **Figure 4-2**, the Hill slopes obtained from the dose response curves constructed for antiproliferative activities were much larger than those observed for Icmt inhibition. The mechanisms proposed by Shoichet and co-

workers were directed at steep curves obtained in the course of investigating enzyme inhibition and may not apply to cell based assays[46]. A notable observation was that increasing the fetal bovine serum concentration in the culture media significantly affected the steepness of the dose response curves of cysmethynil, **4-3** and **6-4** suggesting that binding to serum binding may be a factor contributing to this phenomenon, ie. the concentration of inhibitor available for cell entry is less than the bulk concentration as a fixed component is bound to serum proteins.

### 4.3.7. SAR for antiproliferative activity

A comparison of the SAR for Icmt and anti-proliferative activities is described in the following paragraphs:

(i) Omitting each of the three key functionalities (5-*m*-tolyl, *n*-octyl, acetamide) of cysmethynil to give **1-6**, **2-1** and **3-1** reduced antiproliferative activities. Of the three "deletions", removing the 5-*m*-tolyl residue had the least effect on activity ($IC_{50}$ value of 50 μM) while removing the other two functionalities resulted in almost the same degree of loss ($IC_{50}$ value of >100 μM). In contrast, the extent of loss in Icmt inhibitory activity was of the order acetamide (**3-1**, greatest loss of activity $IC_{50}$ value of > 100 μM) > *n*-octyl (**2-1**, $IC_{50}$ value of 33 μM) > 5-*m*-tolyl (**1-6**, $IC_{50}$ value of 6.5 μM).

(ii) The Series 1 compounds that had changes made to the substitution of the 5-phenyl ring retained the same level of antiproliferative activity as cysmethynil ($IC_{50}$ value of 21.6 μM). Omitting 5-phenyl or replacing it with 5-fluoro caused a larger loss in antiproliferative activity. These changes closely paralleled those observed for Icmt inhibitory activities.

(iii) When *n*-octyl was replaced with *m*-trifluoromethylbenzyl, isoprenyl or geranyl, Icmt inhibitory activities were decreased (**2-2**, **2-3**) or maintained (**2-4**). The antiproliferative activities of **2-2** and **2-3** were broadly comparable to cysmethynil but the geranyl analogue **2-4** ($IC_{50}$ value of 10.6 μM) was now twice as potent. Lipophilicity was unlikely to have contributed to the improved antiproliferative activity of **2-4.**

(iv) The tertiary amides of Series 3 had comparable antiproliferative activities ($IC_{50}$ value of 15.6 to 31 μM) to cysmethynil. The secondary amide **3-2** is of particular interest because its antiproliferative $IC_{50}$ value (4.7 μM) was lower than expected, given that its Icmt $IC_{50}$ value (2.1 μM) was close to that of cysmethynil.

(v) The amines of Series 4 had significantly more potent effects on cell proliferation ($IC_{50}$ value of 2.9 to 13.0 μM) compared to cysmethynil ($IC_{50}$ value of 21.8 μM). The least potent amines **4-4** and **4-10** were also those with the weakest Icmt inhibitory activities.

(vi) The antiproliferative activities of the cysmethynil homologues and isosteres in Series 5 closely paralleled their Icmt inhibitory activities. Thus, the esters (**5-3**, **5-4**) and short chain homologue (**5-1**) had negligible effect on cell viability while the long chain homologue **5-2**, retroamide **5-5** and sulphonamide **5-6** were almost as potent as cysmethynil.

(vii) Good agreement was observed between the two activities (Icmt inhibition, antiproliferative activity) of the compounds in Series 6. Exceptions were found in two cases. First, the antiproliferative activity of **6-11** was lower than expected given its good Icmt inhibitory activity. Second, the amines **6-4** and **6-5** with isoprenyl side chains had better than expected antiproliferative activities ($IC_{50}$ value

of approximately 4 µM) although they inhibited Icmt to almost the same extent as cysmethynil.

## 4.4. DISCUSSION

One of the objectives of this thesis was to investigate modifications at position 3 of cysmethynil and to determine if these changes would result in more potent inhibitors of Icmt. The compounds in Series 3, 4 and 5 were synthesized and evaluated to answer this question. A comparison of the activities of **1-6**, **2-1** and **3-1** clearly showed that the acetamide side chain made a significant contribution to Icmt inhibition, which exceeded that of the *n*-octyl and 5-phenyl functionalities. Modifying the acetamide to the retroamide, sulphonamide and various tertiary amides retained activity. Varying the length of the primary amide side chain showed that the shorter homologue was inactive whereas the longer homologue was as active as cysmethynil. The ester analogues (**5-3**, **5-4**) were without activity. Of the many modifications made to the acetamide side chain, only the amino analogues resulted in improved Icmt inhibitory activities, albeit to a modest degree (approximately 2-3 fold). The good inhibitory activities of the amino analogues raised the question as to whether they interacted with Icmt in their protonated or non-protonated states. There was little doubt that the amines were largely protonated at physiological pH because they were moderate to strong bases with pKa values that exceeded 7.0. But the "active" species interacting with the enzyme need not be the protonated form. Icmt is a membrane bound enzyme found in the hydrophobic environment of the endoplasmic reticulum. The protonated amine analogue may lose a proton in this environment and the species interacting with the enzyme could be the non-protonated amine. If this was so, then a H bond acceptor moiety would be a required feature of the side chain.

Hence, a tentative SAR profile was proposed, namely that (i) a H bond donor was not essential for activity (tertiary amides were not H bond donors but were as active as cysmethynil); (ii) a H bond acceptor moiety was required for activity (provided by the carbonyl oxygen of cysmethynil, the tertiary amides, retroamide **5-5**; sulphonyl oxygen of sulphonamide **5-6** and the nitrogen lone pair of the amine analogues in Series 4), (iii) a longer or more flexible side chain at position 3 was preferred to a shorter side chain (the short chain homologue **5-1** had poor activity).

Based on the Icmt $IC_{50}$ values of **1-6**, **2-1** and **3-1**, the *n*-octyl side chain was ranked after the acetamide side chain in terms of importance for inhibition. The contribution of this functionality was closely linked to its lipophilic character as replacing *n*-octyl with less lipophilic residues (*m*-trifluoromethylbenzyl **2-2** and isoprenyl **2-3**) reduced activity. The geranyl analogue **2-4** which was as lipophilic as cysmethynil (in terms of ClogP) had comparable inhibitory activity. It may be that the *n*-octyl side chain was required to anchor the molecule in the endoplasmic reticulum and to bring it close to the membrane bound Icmt. As a related objective of this thesis was to seek out more drug-like substitutes to cysmethynil with lower lipophilicities and equal or greater potencies, geranyl analogues would not be the preferred choice. Rather, isoprenyl analogues would be better alternatives in spite of the relatively weak Icmt inhibitory activity of **2-3** ($IC_{50}$ value of 7.7 μM). Indeed, the amino analogues with isoprenyl side chains (**6-4**, **6-5**) turned out to be among the more promising "drug-like" substitutes to cysmethynil, combining good Icmt inhibitory activity (comparable to cysmethynil) with even better antiproliferative activity. Compared to selected amine analogues in Series 4 which constituted the other class of drug like substitutes, **6-4** and **6-5** were less potent Icmt inhibitors ($IC_{50 \; Icmt}$ value ≈ 2μM) but

had comparable antiproliferative activities. Compounds in Series 4, **6-4** and **6-5** had lower lipophilicities than cysmethynil because of their protonated amino side chains.

While lipophilicity played a role in influencing inhibitory activity, the relationship is not as clear cut as it may appear. Notably, compounds that were more lipophilic than cysmethynil did not always turn out to be better inhibitors. This was well illustrated by the tertiary amides (**3-4** to **3-7**) which had ClogP values in the range of 8 to 10 (10-1000 times more lipophilic than cysmethynil) but were not better Icmt inhibitors than cysmethynil. There were also instances where compounds with the same ClogP values (for example **3-1** and **3-4**) had very different Icmt inhibitory activity.

The SAR of the 5-*m*-tolyl ring was investigated in **Chapter 2** and the few modifications undertaken in this investigation were made with these findings in mind. The most striking observation was the relatively small effect on activities when the 5-phenyl ring was removed altogether (**1-7**, $IC_{50 \text{ Icmt}}$ value of 6.5 μM, $IC_{50 \text{ antiproliferative}}$ value of 50 μM), emphasizing the lesser role of this position in influencing activities. From the drug design perspective, this finding could be used as a means of modulating drug-like properties without incurring significant loss of activity.

In general, antiproliferative activities of the test compounds moved in parallel with Icmt inhibitory activities. There were a few exceptions, namely **2-4, 3-2, 6-4, 6-5** (more antiproliferative activity than expected based on $IC_{50 \text{ Icmt}}$ values) and **6-11** (less antiproliferative activity than expected based on $IC_{50 \text{ Icmt}}$ values). Differences in the magnitude and range of $IC_{50}$ values for these two activities may be attributed to the nature of these assays (in vitro versus cell-based). In spite of the good correlation between the two activities, it is not known if Icmt is the only cellular target of the test compounds and if the arrest of cell growth stemmed solely from its inhibition.

Moreover, the mechanism by which Icmt affected cell viability was not well understood. Recently, different concentrations of cysmethynil were found to arrest cell proliferation (cytostatic) by interfering with the cell cycle at the G1 phase, or to kill cells (cytotoxic) by autophagy.[62] Thus, while the correlation between $IC_{50}$ values observed here was encouraging, more work needs to be done to understand the connection between these two activities.

## 4.5. CONCLUSION

The main findings of this chapter are as follows:

(i) Loss of the acetamide side chain of cysmethynil resulted in a significant decrease in Icmt inhibitory activity which was greater than that observed when the *n*-octyl side chain or the 5-phenyl ring was omitted from the molecule. Thus, the acetamide side chain of cysmethynil made a significant contribution to its Icmt inhibitory activity.

(ii) The H bond acceptor property and length/flexibility were important features of the acetamide side chain for Icmt inhibition.

(iii) The most important contribution of the *n*-octyl side chain to Icmt inhibition may be its lipophilic character. It was proposed that this side chain localized the molecule in the endoplasmic reticulum, in proximity to the membrane bound Icmt enzyme. Thus for compounds that differed only at this position (*n*-octyl versus isoprenyl), better Icmt inhibitory activity was generally found for the *n*-octyl analogue (for example **4-3** and **6-4**; **4-8** and **6-5**; **1-8** and **6-8**).

(iv) Modification to the substituted 5-phenyl ring caused only incremental changes to Icmt inhibitory activity. It may be a useful position to introduce drug-like features to the molecule without causing marked changes to inhibitory activity.

(v) Several amines in Series 4 (**4-3**, **4-7**, **4-8**) and Series 6 (**6-4**, **6-5**) were more potent than cysmethynil in terms of inhibition of Icmt and antiproliferative activities. They were also less lipophilic than cysmethynil. Thus, these compounds may be considered as the more drug-like alternatives to cysmethynil.

(vi) There was a good correlation between Icmt inhibitory and antiproliferative activities, but this did not necessarily mean that the compounds only targeted Icmt to affect cell viability. A few compounds affected cell growth to a greater extent than expected from their Icmt inhibitory activities. It is possible that these compounds acted on other targets beside Icmt to arrest cell proliferation.

# CHAPTER 5: PHARMACOPHORE MODELING AND QSAR

## 5.1. INTRODUCTION

This chapter describes attempts to rationalize the Icmt inhibitory activities of the test compounds described in **Chapter 3**. The first approach was to develop a three-dimensional (3D) pharmacophore model for Icmt inhibition based on the 47 compounds. A 3D pharmacophore represents the relative position of important binding groups in space and disregards the molecular structure that holds them in place. The 3D pharmacophore for a particular binding site should be common to all ligands that bind to it. Hence, it could be used as a screening tool to identify structures that share the same pharmacophoric features and thus stand a good chance of inhibiting Icmt. The second approach was to identify and quantify physicochemical properties that influenced Icmt inhibitory activity. A diverse series of physicochemical descriptors were collected for the test compounds and these were investigated for correlations to Icmt inhibition by stepwise multiple linear regression (MLR) and PLS (partial least squares projection to latent structures) regression. 3D quantitative structure activity relationship (QSAR) using comparative molecular field analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA) were also attempted. In **Chapter 4**, the $pIC_{50\ Icmt}$ and $pIC_{50\ antiproliferation}$ values were found to be highly correlated (Spearman rho coefficient of 0.707, p <0.001). A decision was made to model the pharmacophore and SAR based on Icmt inhibitory activities rather than cell viability effects based on the reasoning that the effects on cell viability might be due to other targets besides Icmt, in spite of the good correlation between the two activities. It is well known that the cellular phenotype is complex as it represents the summation of the activity states of several pathways.[85]

Hence, there is a "black-box" quality in cell-based screens that cannot be avoided. For this reason, it was decided that the pharmacophore model and SAR would be analyzed with respect to a specific target, namely Icmt.

## 5.2 MATERIALS AND METHODS

### 5.2.1. General

The compound database comprised 47 functionalized indoles described in Chapters 3 and 4. The structures of the test compounds were drawn and energy minimized using the MMFF94x force field (gradient setting of 0.1 kcal mol$^{-1}$ Å$^{-1}$)[86-87] on the software Molecular Operating Environment (MOE, version 2008.1001, Chemical Computing Group, Montreal, Canada). Partial atomic charges were calculated on the same force field. Eighteen compounds (ten in Series 4 and eight in Series 6) had basic amino groups at position 3. These compounds were energy minimized in both their protonated and non-protonated states, giving rise to 2 databases of the 47 compounds, one comprising of the 18 compounds in the non-protonated states and another with the same compounds represented in their protonated states. Several descriptors were collected for these two databases by MOE. They are divided broadly into 2D descriptors calculated from connection tables and are independent of conformation and 3D descriptors that were dependent on conformation. A total of 41 descriptors were used for MLR and PLS regression analyses (**Table 5-1**). **Tables A4-1** to **A4-4** in **Appendix 4** list the descriptors values for the two databases.

**Table 5-1. Descriptors used for MLR and PLS Regression Analyses** [a]

**2D Descriptors**

| | |
|---|---|
| 1. Physical Properties | Molar refractivity (MR), log P(o/w), SlogP (considers state of protonation), van der Waals (vdw) area and volume, polar surface area (TPSA), log S (S= aqueous solubility), molecular weight |
| 2. Atom and Bond Counts | Number of bonds (b_count), rotatable bonds (b_rotN), rotatable single bonds (b_1rotN); Fraction of rotatable single bonds (b_1rotR) and rotatable bonds (b_rotR).Lipinski's druglike features (lip_druglike) and number of violations of Lipinski's rules (lip_violation). |
| 3. Pharmacophore feature descriptors | Number of H bond acceptors (a_acc), donors (a_don), hydrophobic atoms (a_hyd), basic atoms (a_base). vdw surface areas of H bond acceptors (vsa_acc), donors (vsa_don), hydrophobic atoms (vsa_hyd), polar atoms (vsa_pol) |

**3D Descriptors**

| | |
|---|---|
| 1. MOPAC descriptors | Properties calculated using the PM3 Hamiltonian (MOPAC): energy (PM3_E), electronic energy (PM3_Eele), PM3_HOMO, PM3_LUMO, ionization potential (PM3_IP), heat of formation (PM3_HF), dipole moment (PM3_dipole) |
| 2. Potential energy descriptors | Potential energy E, solvation energy Esol, bond stretch potential energy (Estr), local strain energy (E strain), torsion potential energy (E tor), vdw component of potential energy (E vdw) |
| 3. Surface area, volume and shape descriptors | Water accessible surface area (ASA), vdw volume (volume), vdw surface area (VSA) |
| 4. Conformation dependent charge descriptors | Water accessible surface areas of hydrophobic atoms (ASA_H) and polar atoms (ASA_P); dipole moment (dipole) |

[a] Abbreviations are given in parentheses. Information relating to these descriptors as given by MOE are given in the **Appendix 4**.

### 5.2.2. Generation of the Pharmacophore Model

The database of non-protonated test compounds was imported into a molecular database in MOE. Up to 500 conformers of each compound were generated with a modified version of the Conformation Import module in MOE. Conformation Import follows a fragment-based approach to rapidly generate conformers[88]. Briefly, input molecules were dissected into overlapping fragments for which a number of conformations were generated by means of stochastic conformational search. The results of this search were stored in a fragment conformer library. The fragments of this library were then reassembled in a systematic way to yield the final conformers. The geometry of final conformer was optimised by the MOE MMFF94x force field with a gradient setting of 1.0 kcal mol$^{-1}$ Å$^{-1}$. The flexible *n*-octyl side chain at position 1 posed a considerable challenge for conformer generation. Therefore, the "Conformation Import" method was modified to restrict conformational freedom of this group. With the aid of scripts implemented in the MOE scripting language SVL, the fragment conformer library was restricted to trans-fragments of the *n*-octyl residue. Butyl fragments (derived from the *n*-octyl residue) with a dihedral angle of $\theta < 175°$ or $\theta > 185°$ were removed. Furthermore, indole fragments were modified such that the N1 side chains were aligned on the same side of the indole ring. This was done by monitoring the respective dihedral angle. All negative C2′–C1′–N1–C2 dihedral angles ('prime' atoms referring to the indole N1 residue atoms, 'non-prime' atoms referring to the indole ring) were flipped by 180°. Conformers generated in this way had limited conformational freedom for the N1 residues and unrestricted conformational freedom for the other residues.

The pharmacophore model was generated with this conformer library with the Pharmacophore Elucidator module in MOE. First, pharmacophoric features were

assigned to the test compounds, namely 'aromatic $OR$ $\pi$ ring', 'hydrophobic', 'H-bond donor projection point', 'H-bond acceptor projection point', 'cation', 'anion', 'carboxyl bioisostere', and 'amidinium bioisostere'. A maximum number of 3 hydrophobic features and a minimum number of 1 H-bond acceptor projection point were specified. These restrictions were imposed to avoid generation of pharmacophore models composed of purely hydrophobic/aromatic features. Pharmacophoric features were treated as solid spheres with a radius of 1.4 Å. Next, an exhaustive search of all pharmacophore queries that induced a good overlay of most of the active molecules and separate actives from inactives were performed. Thus 3D alignments of selected conformers of the input molecules were generated. The plausibility of a pharmacophore was measured by the overlay of actives (goodness of 3D alignment) and the relationship to activity was measured by its "classification accuracy" (ability to separate actives from inactives). Molecules were considered actives if $IC_{50 \, Icmt}$ value of < 5 µM. The results were potential 4-point and 5-point pharmacophore models that included at least 90% of the actives. The generated pharmacophore hypotheses (734 models) were then sorted by their overlay scores and classification accuracies. One model was chosen by manual inspection of the top-scoring hypotheses according to the abovementioned quality measures (overlay and accuracy). This model was subsequently subjected to manual refinement with MOE's Pharmacophore Query Editor. Briefly, the radii of pharmacophoric spheres were reduced to 1.2 Å, and their positions were optimized to center on the respective pharmacophoric points. The pharmacophore type for the indole ring was also set to aromatic.

### 5.2.3. Multiple linear regression (MLR) and Spearman correlation analysis

Multiple linear regression was carried out separately on the two databases of non-protonated and protonated test compounds, with descriptors in **Table 5-1** as independent variables and $pIC_{50\ Icmt}$ value as the dependent variable. In each case, stepwise regression was carried out in which descriptors were sequentially added to or deleted from the equation until the best fit relationships were obtained. A Spearman correlation analysis was similarly carried out on $pIC_{50\ Icmt}$ value and descriptors. Spearman rho values are given in **Table A4-5 and A4-6 of Appendix 4**. Both MLR and correlation analysis were carried out on SPSS for Windows 17.0 (SPSS Inc., Chicago, IL, USA).

### 5.2.4. PLS Analysis

PLS analysis was carried out separately on the non-protonated and protonated compounds using the AutoQSAR module in MOE. PLS identifies the principal components (derived by "summarizing" the large number of descriptors into a smaller number that represent the main variation in the data set) that can concurrently explain the variance in the descriptor set as well as maximize their correlation to biological activity. The option to restrict descriptors to the three most representative members was selected in favour of the "Auto" select option which would generate a large number of descriptors. The z factor (threshold value beyond which compounds are identified as outliers) was set at the default value of 2.5.

### 5.2.5. CoMFA Analysis

The 3D-QSAR model was developed with CoMFA/CoMSIA on the modeling software Sybyl 8.0 (Tripos, St Louis, Missouri). Force field minimized structures (of

non-protonated members) from MOE were aligned using the best pharmacophore model described in **Section 5.2.2**. Steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor potential fields were calculated at each lattice intersection of a regularly spaced grid of 2.0 Å. A probe atom with + 1 charge, radius of 1.0 Å and + 1 hydrophobicity was used for this purpose. A Gaussian type distance dependence was considered between the grid point in the lattice and each atom of molecule. The attenuation factor ($\alpha$) was set at the default value of 0.3. The PLS analysis was used to construct and validate the models. Column filtering was set to 2.0 kcal/mol to speed up the analyses and reduce background noise. The CoMSIA descriptors served as independent variables, while the $pIC_{50\ Icmt}$ values served as dependent variable. The performance of the models were tested using the leave-one-out (LOO) cross-validation method. The optimum number of components was used to derive the non-cross validated model, leading to the highest cross-validated correlation coefficient $q^2$ and the lowest standard error of estimate (SEE). The predictive power of the 3D QSAR model was determined using a test set of 9 compounds (**1-4, 1-8, 3-4, 4-5, 4-9, 6-1, 6-3, 6-5, 6-9**) for external validation. The optimization, alignment and all other steps of these test set molecules were the same as that of the training set, and their activities were predicted using the model derived from the training set.

## 5.3. RESULTS AND DISCUSSION

### 5.3.1. Establishing a pharmacophore model for Icmt inhibition
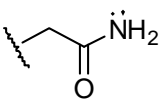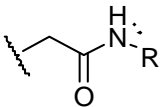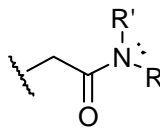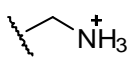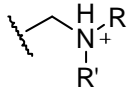
A pharmacophore is the ensemble of steric and electronic features necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response[89]. It does not represent a real molecule or a real association of functional groups. Rather it should be viewed as
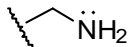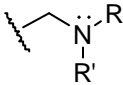
an abstract concept that embraces the common molecular interaction capacities of a group of compounds for their target structure. The pharmacophore can be viewed as the largest common denominator shared by a set of active molecules[89]. Pharmacophore modelling is generally applied to achieve the following objectives. First, it is used to define relevant pharmacophoric features in a compound that are necessary to achieve a desired biological effect. This would help to establish a definitive SAR. Second, a pharmacophore model can be used to detect molecules with different scaffolds by virtually screening large compound libraries by a technique called scaffold hopping. Third, it can be used together with other pharmacophore based screens to predict the pharmacological profiles of lead structures *in silico*. This requires deriving pharmacophore fingerprints for different activities. For instance, an overlap between an activity fingerprint and a toxicity fingerprint would reduce the attractiveness of a particular structure or scaffold.  In this section, pharmacophore modelling was used for the elucidation of SAR[90-92].

The pharmacophore model was based on the 47 test compounds and was built using the Pharmacophore Elucidator module in MOE. As this module considered binary activity class labels (active vs. inactive), an arbitrary activity threshold of 5 μM ($IC_{50 \text{ Icmt}}$ value) was used to classify the compounds. By this criterion, 35 compounds (or 74 %) were classified as actives. Conformations were generated and they were aligned by the software to give various pharmacophore models. However, the generated pharmacophore hypotheses consisted of only hydrophobic and aromatic features. These were rejected because they failed to include features related to the critical polar side chain at position 3. Furthermore, the computation was carried out with settings that assumed an aqueous environment of pH 7.0. Under these conditions, the molecules, in particular the 18 compounds with amino groups, would be

protonated by the software. As discussed in **Section 4.4**, while this was physiologically "correct", it did not necessarily mean that these compounds interacted with Icmt in their protonated states. Loss of a proton could occur as the molecule approached the active site of the enzyme and the final interacting species might be the non-protonated state.     Furthermore, if it was assumed that the amide and amine derivatives shared the same binding mode to the enzyme, there would be no common pharmacophoric feature in their side chains at position 3, if the amino group was protonated. The tertiary amides of Series 3 were only H bond acceptors, not H bond donors while the protonated amines of Series 4 and 6 were H bond donors and not H bond acceptors. On the other hand, if interaction involved the non-protonated amine, a common H bond acceptor feature would be present in both amide and amino side chains (**Table 5-2**). With these considerations in mind, it was decided that the test molecules would be aligned in their non-protonated states. Furthermore, to avoid generating purely hydrophobic/aromatic models, the presence of at least one hydrogen bond acceptor feature in the putative model was specified.

**Table 5-2. Hydrogen-bond donating and accepting features in representative side chains at position 3 of the indole scaffold.**

|  | Primary | Secondary | Tertiary |
|---|---|---|---|
| **Amide** |  |  |  |
| H bond donor | Yes | Yes | No |
| H bond acceptor | Yes | Yes | Yes |
| **Protonated amines** |  | |  |
| H bond donor | Yes | | Yes |
| H bond acceptor | No | | No |

| Non-protonated amines |  |  |
| --- | --- | --- |
| H bond donor | Yes | No |
| H bond acceptor | Yes | Yes |

However, the 3D alignment of the compounds in the pharmacophore models generated this way was poor. It was reasoned that the poor outcomes could arise from insufficient conformational sampling of the input molecules contributed to the poor 3D alignment. Apparently, most conformers focused on geometry variations of the flexible *n*-octyl side chain (seven rotatable bonds) that was present in 36 of the 47 input molecules (77 %) at the expense of conformational variations of the polar side chain at position 3. Therefore, a method was developed that would restrict the conformational freedom of the *n*-octyl side chain. The 'Conformation Import' tool in MOE was modified to allow only trans-isomers of the *n*-octyl side chain. Furthermore, conformations were restricted to those that had the *n*-octyl side chain oriented towards the same "direction" relative to the indole ring. Conformers generated in this way had restricted conformational freedom for the *n*-octyl sidechain but not for other residues on the indole scaffold. In this way, up to 500 conformers were generated for each molecule resulting in a total of 23,474 conformers (ca. 499 per molecule). With this approach, pharmacophore models with good 3D alignments and high accuracy were obtained. In total, 734 four-point and five-point pharmacophore models were generated by MOE and these were sorted by their overlap scores and classification accuracies. The accuracy of the pharmacophore model was defined by **Equation 5-1**.

$$\% Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \cdot 100$$

... **(5-1)**

In this equation, TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives. Based on the threshold of $IC_{50}$ value of < 5 μM for actives, 35 molecules were actives while 12 molecules were inactives. The most accurate models (93.6 % accuracy) correctly identified all 35 molecules as actives (TP = 35, FN = 0). However, the indole ring of **6-9** (the amine with a fluoro group in place of the 5-phenyl) did not superimpose with the indole rings of the remaining actives, possibly due to the missing phenyl residue on position 5. Hence, the next best models (91.4 % accuracy; TP = 34, FN = 1) was selected, where **6-9** was regarded as a false negative. Of the remaining molecules, 9 were identified as true inactives (TN) but 3 were deemed to be actives in spite of having $IC_{50\ Icmt}$ values that exceeded 5 μM (FP = 3). These were the isoprenyl analogue **2-3** and the esters **5-3**, **5-4**. There was no good reason why the isoprenyl analogue **2-3** ($IC_{50\ Icmt}$ value of 7.7 μM) was classified as a false positive. Re-determination of its $IC_{50}$ value may be necessary. As for the esters **5-3** and **5-4**, they may be hydrolyzed to give inactive carboxylic acids under the conditions of the assay, although the esters were not found to be unstable during their syntheses and characterization. A carboxylic acid derivative should be synthesized and evaluated for inhibitory activity to provide the needed confirmation.

The final pharmacophore model consisted of five pharmacophoric features (**Figure 5-1**). These were (i) a hydrophobic feature at position 1, the location of which was guided by the shorter isoprenyl and trifluoromethylbenzyl residues, (ii) an aromatic feature for a substituent at position 5 of the indole ring, (iii) two aromatic features on the fused indole ring, and (iv) H bond acceptor projection point where the topological distance was 2 bonds from the ring for amine analogues and 3 to 4 bonds from the ring for amide and sulphonamide analogues. Atom exclusion volumes were

placed above and below the plane of the indole ring to ensure planarity of the aromatic rings (not depicted in **Figure 5-1**). In some of the proposed models, the side chains at position 3 were twisted by 90 degrees which placed the H bond acceptor projection point in front of the indole ring at position 2. However, these models were not considered due to their lower accuracies. For example, with the H bond acceptor feature at this position, the shorter cysmethynil homologue **5-1** ($IC_{50}$ value of 30 μM) would be classified as an active compound.
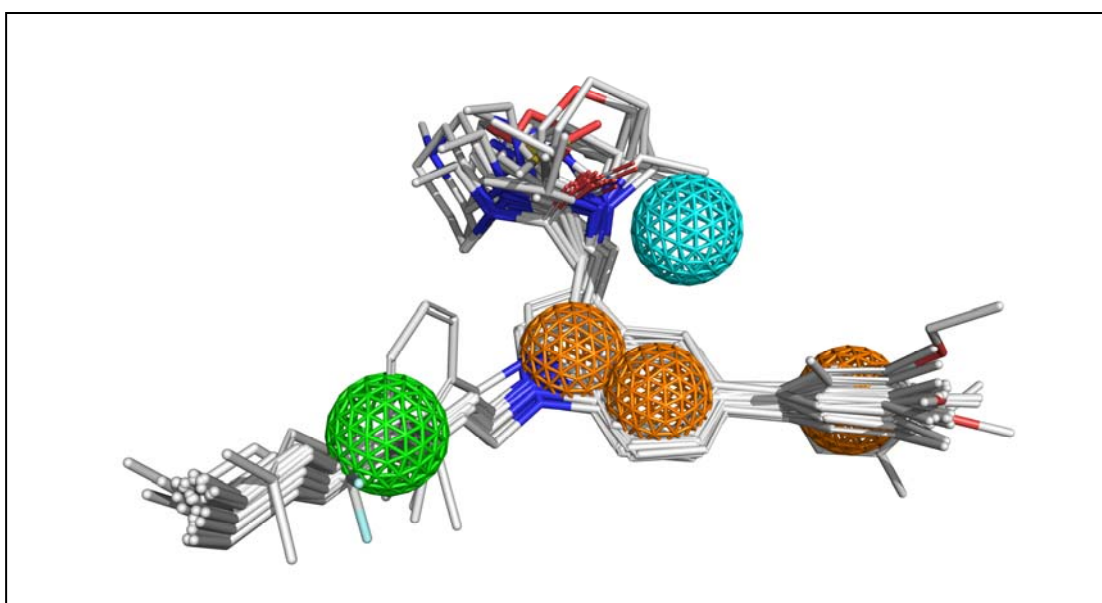


**Figure 5-1. Pharmacophore model.** A 3D alignment of 34 analogues with Icmt $IC_{50}$ value of < 5 μM is shown. Pharmacophoric features are depicted as spheres. Green: hydrophobic, orange: aromatic *OR* π ring, cyan: H-bond acceptor projection point. Feature radius: 1.2 Å. Figure created with PyMOL (DeLano Scientific, Palo Alto, CA, USA).

Next, 14 compounds that were known Icmt inhibitors and not structurally related to cysmethynil were tested on the pharmacophore to determine if they had the same pharmacophoric features as the model. Five compounds were found to comply. The features present were the hydrophobic, aromatic (two), and H bond acceptor (one or two) (**Figure 5-1**). Some flexibility was permitted in this screening process in that ligands that had only one of the two aromatic features on the indole ring were allowed to pass through. Ligands that had more than one H-bond acceptor atom could also

potentially match the pharmacophoric feature. As for the remaining 9 Icmt inhibitors not identified by the model, this showed that the present model required further refinement with additional compounds of greater structural diversity. It could also imply that these compounds did not interact with Icmt in the same manner as cysmethynil and its analogues, and hence would not be expected to fit into the model.[93] Ultimately, a true validation of the predictive capabilities of the model could only be addressed by prospective library screening to short-list hit compounds and to follow up on their inhibitory activities. It should be noted that the present model was generated to elucidate the structure-activity relationships of cysmethynil analogues. As such it would be more descriptive than predictive in nature.
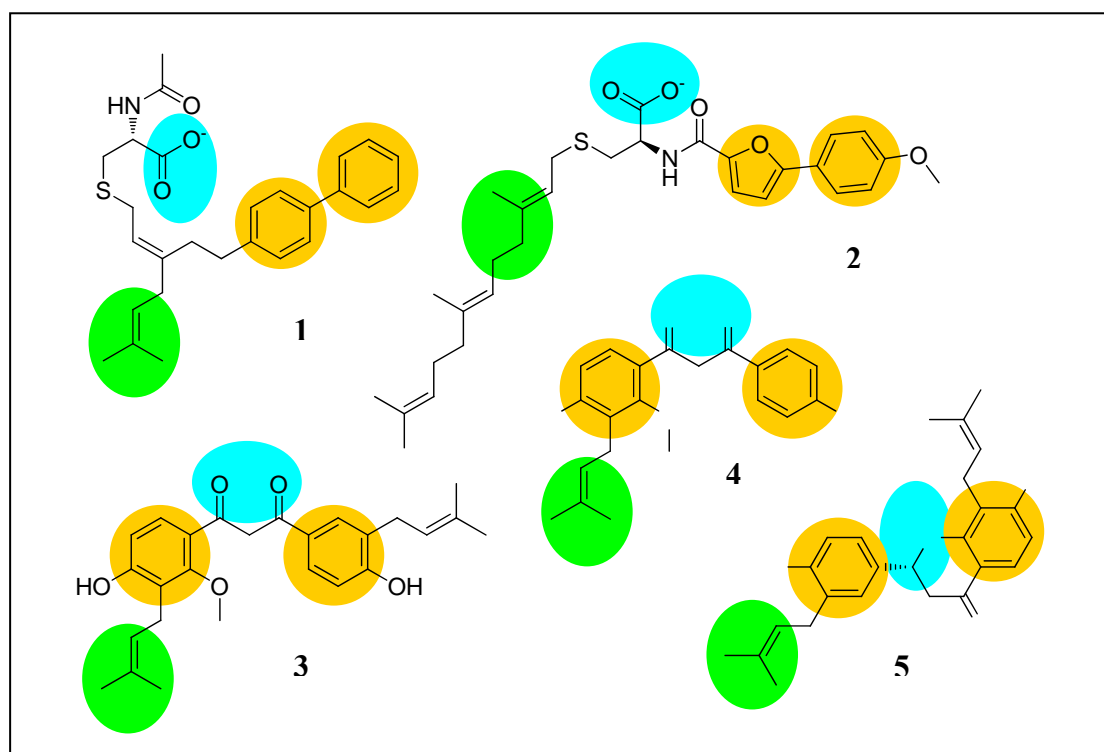


**Figure 5-2. Known Icmt inhibitors with pharmacophoric features that match the proposed model. 1**: Anderson[53], **2**: Donelson[55]; **3**, **4**, and **5**: Buchanan[58]. Green: hydrophobic, Orange: aromatic OR π ring, Blue: H-bond acceptor.

### 5.3.2. Multiple linear regression and correlation analysis

As mentioned in **Section 5.2.1**, the test compounds were prepared in their non-protonated and protonated states and descriptors were collected for both states. Only the 18 compounds with amino side chains would be expected to have different descriptors. Even then, not all of the descriptors were affected by protonation of the basic amino group. An inspection of the descriptors (**Tables A4-1** to **A4-4**) showed that the protonated compounds were less lipophilic (SlogP values were about 2 units lower than the non-protonated states), more polar (dipole moments of greater magnitude) and had larger polar surface areas (captured by the 3D descriptor water accessible surface area for polar atoms ASA_P but not van der Waals polar surface area of polar atoms). In addition, solvation (Esol) and electrostatic (Eele) energies of the protonated species were lower indicating the ease of solvation and electrostatic stabilities of these species. Several energy terms calculated by the PM3 Hamiltonian function (MOPAC) such as heats of formation, ionization potential, HOMO and LUMO were also different. On the other hand, descriptors like area, volume and pharmacophore features, were the same regardless of whether the compound was protonated or not.

MLR was carried out separately on the protonated and non-protonated databases to determine if different descriptors would be recognized in each instance. If this was observed, and more importantly, if the key descriptors identified for the protonated species are among those mentioned in the earlier paragraph, this would lend support to the view that the compounds were interacting with Icmt in their protonated states. But the results should be interpreted with caveats in mind. First, only 18 compounds (37 %) could exist in the protonated state and not withstanding their strong Icmt inhibitory activities, their input to the MLR equation might be

overshadowed by the remaining compounds. Second, the descriptors for H bonding used in this analysis did not discriminate adequately between the protonated and non-protonated states of 18 compounds. For example, the van der Waals surface areas of H bond donors and acceptors (vsa_don, vsa_acc) were set at zero for these compounds in both databases. This was because the software defined vsa_acc as the "approximation of van der Waals surface areas of pure H bond acceptors, not counting acidic atoms and atoms that are both H bond donors and acceptors such as OH". Similarly the number of H bond donors (a_don) did not include the protonated $NH^+$ because of the software definition for a_don where cations are excluded as donors. Hence, protonated and non-protonated members had similar a_don values. The exception was the primary amine **4-1** which had a_don = 1 when non-protonated state and a_don=0 when protonated. Meanwhile, another interesting finding is that the non-protonated amino function had vsa_acc of zero because the van der Waals surface area of tertiary amines was approximately calculated to be zero. This prevented discrimination of the protonated and non-protonated species.

MLR with stepwise selection of descriptors yielded 11 models when applied to the non-protonated compounds but only 3 models for the protonated compounds. In the latter case, the $3^{rd}$ model (with 3 descriptors) was selected. To facilitate discussion and to avoid over-fitting, the model with 3 descriptors ($3^{rd}$ model) was also selected for the non-protonated compounds. The MLR equations and statistics are given as follows:

(i) Equation for non-protonated compounds:

$$pIC_{50 \text{ Icmt}} = 1.232 \ (0.481) + 0.066 \ (0.007) * \text{b\_count} + 0.816 \qquad \dots \textbf{(5-2)}$$
$$(0.174) * \text{a\_don} - 0.029 \ (0.010) * \text{vsa\_acc}$$

$R^2 = 0.662$, Standard error of estimate (SEE) = 0.420, F = 28.102 (significant at $p < 0.001$). The standard error of each coefficient in the equation is given in the bracket.

When the coefficients were standardized, the largest contribution was made by b_count (1.00), followed by a_donor (0.633) and vsa_acc (0.318).

(ii) Equation for protonated compounds:

$$pIC_{50\ Icmt} = -0.692(0.703) + 0.009\ (0.001) * VSA + \qquad \ldots \textbf{(5-3)}$$
$$0.045\ (0.009) * vsa\_donor + 0.173\ (0.046) * PM3\_IP$$

$R^2 = 0.689$, SEE = 0.402, F = 31.78 (significant at $p < 0.001$). When the coefficients were standardized, the largest contribution was made by VSA (0.899), followed by vsa_don (0.551) and PM3_IP (0.372).

Both equations had correlation coefficients ($R^2$) of approximately 0.7 which meant that the activities of approximately 70 % of the compounds could be accounted by these equations. The F values showed that the models were significant at the level of $p < 0.001$.

**Equation 5-2** identified b_count (number of bonds) as the main contributor to activity. The Spearman correlation showed that b_count was significantly correlated to several size and lipophilicity parameters, namely molecular weight, van der Waals area and volume, water accessible surface area for hydrophobic atoms, SlogP and logP. The other two descriptors in **Equation 5-2** were a_donor (number of H bond donors) and vsa_acc (van der Waals surface area of H bond acceptors). Their correlations to activity might suggest that H bond donors were more important than H bond acceptors. As mentioned earlier, vsa_acc was defined in such a way that it is unable to discrimate the H bond acceptor properties of the tertiary amine analogues from non H bond acceptors. Thus vsa_acc may not reflect on the role of H bond acceptors per se. Rather it may emphasize the importance of non-polar residues for activity. jVsa_acc was significantly correlated to dipole moment and polar surface areas (ASA_P, vsa_pol) with Spearman rho values of 0.7-0.8 ($p < 0.01$), as compared to a rho value of 0.437 for number of H bond acceptors (a_acc)

For the protonated compounds (**Equation 5-3**), van der Waals surface area (VSA) was the main contributor to activity. It was highly correlated to b_count and other size and lipophilicity parameters. Thus, size and lipophilicity were found to be the most critical determinants of activity regardless of whether the compound was protonated or otherwise. Two other descriptors in **Equation 5-3** were van der Waals surface area of H bond donors (vsa_don) and ionization potential (PM3_IP). Of these, only ionization potential values differed for the protonated and non-protonated amine analogues. On protonation, the ionization potentials of the amine analogues increased by about 3 units while those of the other compounds remained the same. As for the van der Waals surface area of H bond donors (vsa_don), like vsa_acc, it should be interpreted with caution.

Taken together, the MLR described in **Equations 5-2** and **5-3** highlighted the following points. First, size and lipophilicity were the most important determinants of activity regardless of the protonation state of the compounds. Noting that the pharmacophore model had 3 aromatic residues and 1 hydrophobic feature, the presence of these descriptors in the equations were not unexpected. Second, H bond donor was identified in both equations (a_don in **Equation 5-2** and vsa_don in **Equation 5-3**) but the interpretation of its role in activity should be done with caution, not least because of the way the terms were defined. The presence of the H bond donor term need not necessarily imply that the protonated amino species of the 18 compounds were preferred for good activity. As the protonated ammonium group ($-NH^+$) was not recognized as a H bond donor in its definition, a_don values were actually the same for both protonated and non-protonated species (except for a_don of the primary amine **4-1**). Rather, the H bond donor term may reflect on the contribution of the amide NH (present in Series 1, the secondary amide **3-2**, the

reverse amide **5-5**), sulphonamide NH in **5-6**, and the NH groups in the primary amine **4-1**. Thirdly, ionization potential (PM3_IP) was the only "unique" parameter recognized in the MLR for the protonated species (**Equation 5-3**) but it carried the lowest weightage among the three descriptors. In spite of the inclusion of this term in **Equation 5-3**, it was uncertain if interaction with Icmt occurred preferably with the protonated species. It may be unreasonable to expect MLR to provide a satisfactory answer in view of the caveats mentioned in the earlier paragraph.

### 5.3.3. PLS regression analysis

The databases were analyzed by the AutoQSAR module in MOE which used PLS to perform the regression analysis. To avoid over-fitting and to have models that were "comparable" to those derived from MLR, the number of descriptors was restricted to just three. **Equations 5-4** and **5-5** were obtained for non-protonated and protonated compounds respectively.

(i) Equation for non-protonated compounds:

$$pIC_{50 \text{ Icmt}} = 1.46 - 0.00004 * PM3\_E + 0.0086 * PM3\_HF + \qquad \dots \textbf{(5-4)}$$
$$0.0013 * ASA\_P$$

$R^2 = 0.521$, cross validated $R^2 = 0.414$, RMSE = 0.478, cross validated RMSE = 0.533. Relative importance of descriptors were PM3_E (1.0) > PM3_HF (0.54) > ASA_P (0.106)

(ii) Equation for protonated compounds:

$$pIC_{50 \text{ Icmt}} = 1.49 + 0.0138 * VSA + 0.00003 * PM3\_E + \qquad \dots \textbf{(5-5)}$$
$$0.0062 * ASA\_P$$

$R^2 = 0.562$, cross validated $R^2 = 0.462$, RMSE = 0.457, cross validated RMSE = 0.512. Relative importance of descriptors were VSA (1.0), followed by PM3_E (0.431) and ASA_P (0.302).

The correlation coefficients $R^2$ were not as good as those obtained by MLR as the activities of only 50 % of the test compounds could be accounted for by the descriptors in either equation. Increasing the number of descriptors to 5 for each regression equation improved $R^2$ to $\approx 0.7$ but the two most important descriptors in each case (protonated and non-protonated species) were similar to those identified in **Equations 5-4** and **5-5**. Hence, it was decided that the analysis would be based on the latter equations.

**Equation 5-4** showed that the total energies calculated using the PM3 Hamiltonian (PM3_E) had the greatest effect on activity. Its relationship to the $pIC_{50}$ value was inverse, indicating that compounds with low total energies were associated with improved activities. The Spearman rho values showed that total energies were inversely correlated ($p < 0.01$) to size parameters, namely b_count, molecular weight and van der Waals area. Thus the inverse relationship between the $pIC_{50}$ value and total energy implied that better activity was found in larger and more lipophilic compounds. This agreed with the MLR-derived equation (**Equation 5-2**) where b_count was the main contributor to activity. The heats of formation measured by PM3 Hamiltonian (PM3_HF) ranked $2^{nd}$ in terms of contribution. The Spearman rho values were again scrutinized to give a better understanding of the significance of this parameter. The heats of formation were inversely correlated to dipole moments and van der Waals surface areas of polar atoms ($p < 0.01$). Thus, two terms (PM3_E and PM3_HF) in equation 3 emphasize on the importance of size, diminished polarity and lipophilicity for good activity. On the other hand, the presence of the ASA_P term (water accessible surface area of polar atoms) in **Equation 5-4** showed that the polar component was a small but important contributor to activity. Taken together, both the analysis by MLR (**Equation 5-2**) and PLS (**Equation 5-4**) identified size and

lipophilicity as important determinants of inhibitory activity. Both approaches gave a smaller role to polar surface areas which encoded the H bond capabilities of the test compounds.

Turning to **Equation 5-5** which was derived by PLS for the protonated species, it was striking that two descriptors in this equation (PM3_E and ASA_P) were related to polarity as compared to one descriptor (VSA) that favoured size and lipophilicity. VSA was, however, still the main contributor to Equation 5-5, in keeping with the other equations (**Equations 5-2, 5-3, 5-4**) where size and lipophilicity were the main determinants of activity. The presence of ASA-P in Equation 5-5 was notable as it was the only descriptor that was "unique" to the 18 protonated compounds in this database. Therefore, both approaches (MLR and PLS) identified unique descriptors for the protonated species in the regression equations (**Equations 5-3** and **5-5**). Whether this meant that interaction with Icmt occurred preferentially with the protonated states of the 18 amino analogues, most of which had exceptionally good Icmt inhibitory activities, remained an open question. But the results conveyed that a more representative analysis was obtained when the compounds were analyzed in their protonated states. **Equations 5-3** and **5-5** recognized the importance of the amino groups for activity by identifying parameters that were unique to these groups. Furthermore, in **Equation 5-5**, two of the three descriptors were related to polarity (rather than size and lipophilicity), possibly due to presence of the protonated compounds. In comparison, the equations (**5-2**, **5-4**) derived from non-protonated compounds identified generic descriptors like size and polarity that did not highlight the exceptional activities of the amine analogues.

**5.3.4. 3D QSAR**

When a 3D model of the receptor is available or a proper alignment of ligands has been achieved, QSAR models based on molecular descriptors obtained from 3D structures may be derived. Comparative Molecular Field Analysis (CoMFA) is commonly employed for this purpose and its application has been described in **Chapter 2**.

The success of a CoMFA model is critically dependent on the spatial alignment of the ligands. In this case, the pharmacophore model described in **Section 5.3.1** was used for aligning the test ligands which consisted of 38 compounds (training set). The remaining nine compounds (**1-4, 1-8, 3-4, 4-5, 4-9, 6-1, 6-3, 6-5, 6-9**) were reserved as the test set to validate the model derived from the training set. The main requirement of the test set was that it should be representative of the training set both in terms of structure and activity. In spite of the absence of compounds from Series 2 and 5 in the test set, this was made up by the presence of 4 members from Series 6 which had variations at two positions on the scaffold, many of which overlapped with those observed in the missing series. The range of activity in the test set (0.6 μM to > 100 μM) was in line with that of the training set. The alignment of the training set based on the pharmacophore model is depicted in **Figure 5-3**.
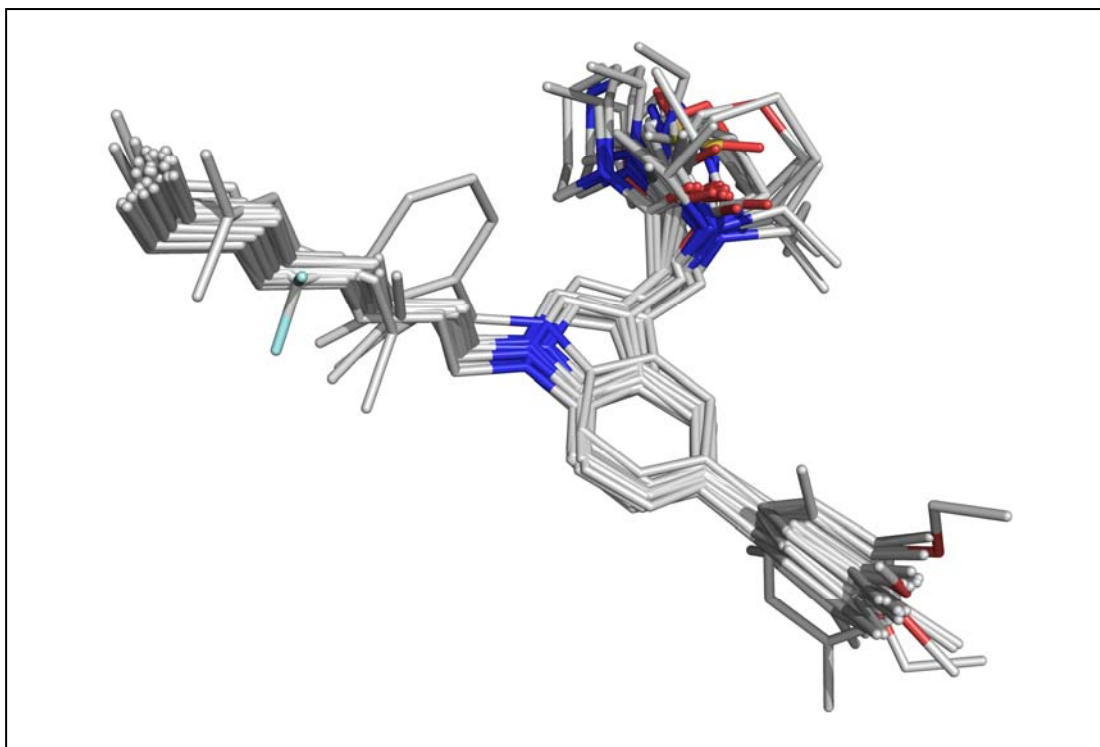
**Figure 5-3. Alignment of compounds based on pharmacophore model.** Figure created with PyMOL (DeLano Scientific, Palo Alto, CA, USA).

**Table 5-3. CoMFA and selected CoMSIA models.**

|  | CoMFA | CoMSIA |
|---|---|---|
| $q^2$ | 0.619 | 0.650 |
| SEP | 0.485 | 0.467 |
| Optimal no. of components | 3 | 4 |
| $r^2$ | 0.873 | 0.892 |
| SEE | 0.277 | 0.260 |
| F | 77.992 | 67.909 |
| Field contribution |  |  |
|    Steric | 0.480 | 0.176 |
|    Electrostatic | 0.520 | 0.301 |
|    Hydrophobic | - | 0.241 |
|    Donor | - | 0.171 |
|    Acceptor | - | 0.111 |
| $r^2_{pred}$ [a] | 0.824 | 0.796 |

[a] Predictive $r^2$ is based only on the 9 compounds not included in the training set. It is determined from $r^2_{pred} = (SD - PRESS)/SD$ where $SD$ = sum of the squared deviations between $IC_{50}$ values of compounds in test set and mean $IC_{50}$ values of training set molecules, and PRESS is sum of the squared deviation between predicted and actual $IC_{50}$ values for every member in the test set.

To visualize the information content of the CoMFA model, the steric and electrostatic contour maps were generated. **Figure 5-4** shows the stereoview of the steric contour maps generated for 7 representative compounds in the training set (**1-1, 2-2, 2-3, 3-5, 4-3, 5-1, 5-2**) of this model. The green and yellow polyhedra described regions of space around the molecules where increases in steric bulk enhance or diminish Icmt inhibitory activity respectively. Swathes of yellow and green were found at both positions 1 and 3. At position 1, the location of a green zone at the distal end of the side chain and a yellow zone in the "middle" of the side chain implied a preference for a long side chain, corresponding to *n*-octyl and geranyl, but not shorter side chains like isoprenyl. Unlike the observations made with an earlier CoMFA model (**Chapter 2**), the yellow zone was not sandwiched between two green zones, possibly because shorter side chains that might have fallen within the proximal green zone were not represented in this study.

Both green and yellow zones were observed at the side chain at position 3. These zones were on opposite sides of the side chain. A close examination of the zones suggested that there were different steric requirements for side chain amides (Series 3) and amines (Series 4), with bulk better tolerated by the shorter amines as compared to the longer amides.
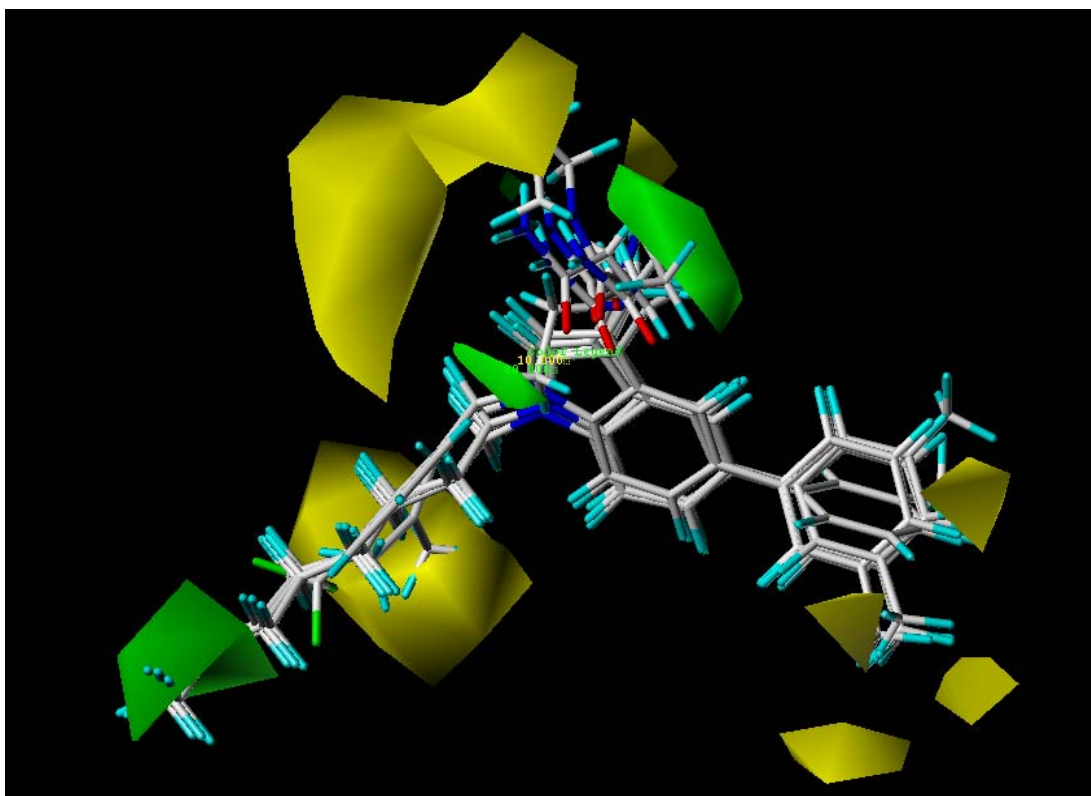
**Figure 5-4. CoMFA steric contour map.** Green contours represent areas where steric bulk improves activity, while yellow contours correspond to regions to be kept unoccupied for increased activity. Compounds **1-1, 2-2, 2-3, 3-5, 4-3, 5-1, 5-2** were used as reference ligands for the CoMFA/CoMSIA contour

     **Figure 5-5** depicts the electrostatic contours for the model. The blue contours represented regions where an increase in positive charge favours activity while the red contours represent regions where more negative charge is favoured. Electrostatic contours were found only at the side chain at position 3. The carbonyl oxygen of the amides and the lone pair of electrons on the amine nitrogen fell within the red zone. As these were H bond acceptors, the red zones corresponded to regions where H bond acceptors were sited. The blue zone was located at the opposite side of the side chain and it coincided with the amide nitrogen which was positively charged as a result of electron delocalization. The NH of the secondary amide **3-2** pointed towards the blue zone, but the NH of the primary amine **4-1** did not.
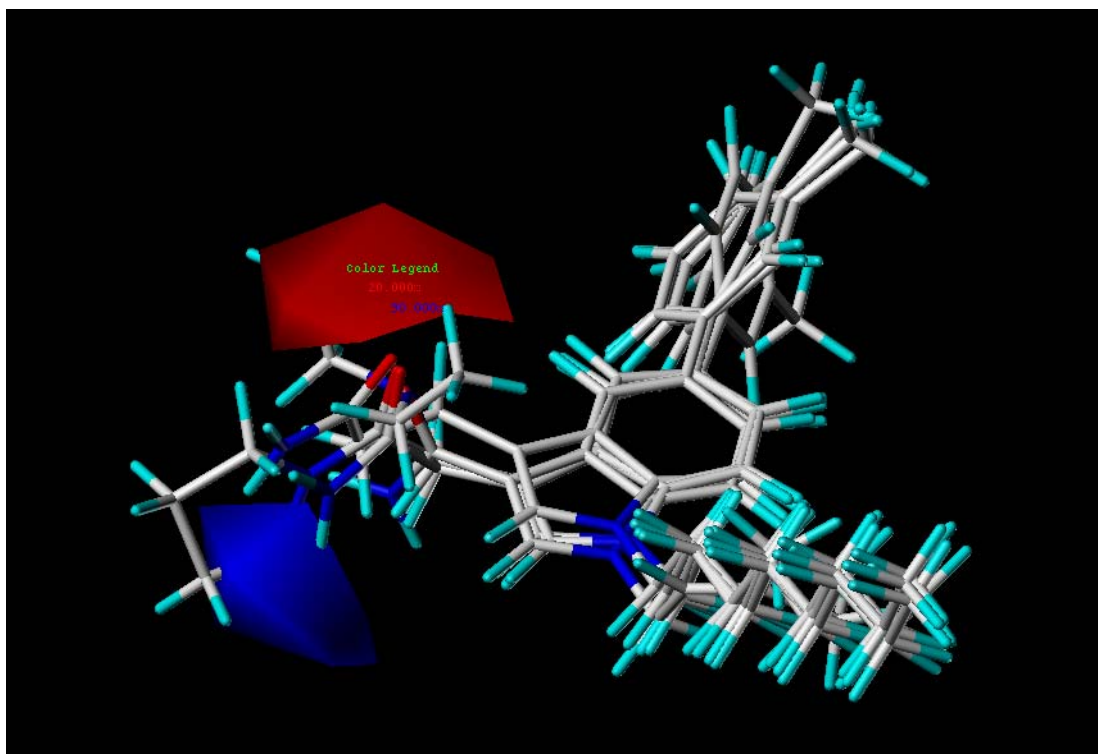
**Figure 5-5. CoMFA electrostatic contour map.** Blue contours represent regions where an increase in positive charge will enhance activity, while red contours correspond to areas where more negative charge is favoured.

3D QSAR could also be explored with a related technique, comparative molecular similarity analysis (CoMSIA).[94] CoMSIA has the advantage of addressing one of the key problems in CoMFA, namely small alignment shifts causing large changes in the molecular field at certain grid points. Instead of grid-based fields, CoMSIA is based on similarity indices obtained by using a functional form adapted from the SEAL algorithm.[95] The similarity indices were calculated by a Gaussian-type distance dependence function, instead of the Lennard-Jones and the Coulomb potential. By this approach, the more or less arbitrary application of cut-off values used in CoMFA was avoided. But CoMSIA considers additional property fields (hydrophobic, H bond donor, H bond acceptor) besides those found in CoMFA (steric and electrostatic). As H bond acceptor and hydrophobic features were recognized in the pharmacophore model (**Section 5.3.1**), the use of CoMSIA instead of CoMFA to

may be more appropriate. On the other hand, the five property fields in CoMSIA are known to be highly interdependent. For example, the H bond donor and acceptor properties are implicitly considered in electrostatic interactions. Nonetheless, a CoMSIA model was derived using the five property fields (steric, electrostatic, hydrophobic, H bond donor and H bond acceptor). As shown in **Table 5-3**, the CoMSIA model (5-component) had comparable statistics to the CoMFA model (3-component) but did not predict the activity of the test set ($r^2_{pred}$ of 0.8) better than the CoMFA model. The H bond donor, H bond acceptor and hydrophobic contours of this CoMSIA model are given in **Figures 5-6** and **5-7**.

Three zones were observed in **Figure 5-6**: a cyan zone where a H bond donor group was favoured, a purple zone where the H bond donor group was not favoured and a magenta zone where a H bond acceptor group was favoured. A close inspection showed that the magenta H bond acceptor zone coincided with the red contour in the CoMFA electrostatic map. The favoured H bond donor zone was located at nearly the same position as the blue contour in **Figure 5-5**, except for the difference in size of the two zones. There was no equivalent to a region where H bond donor was not favoured in the electrostatic contours.
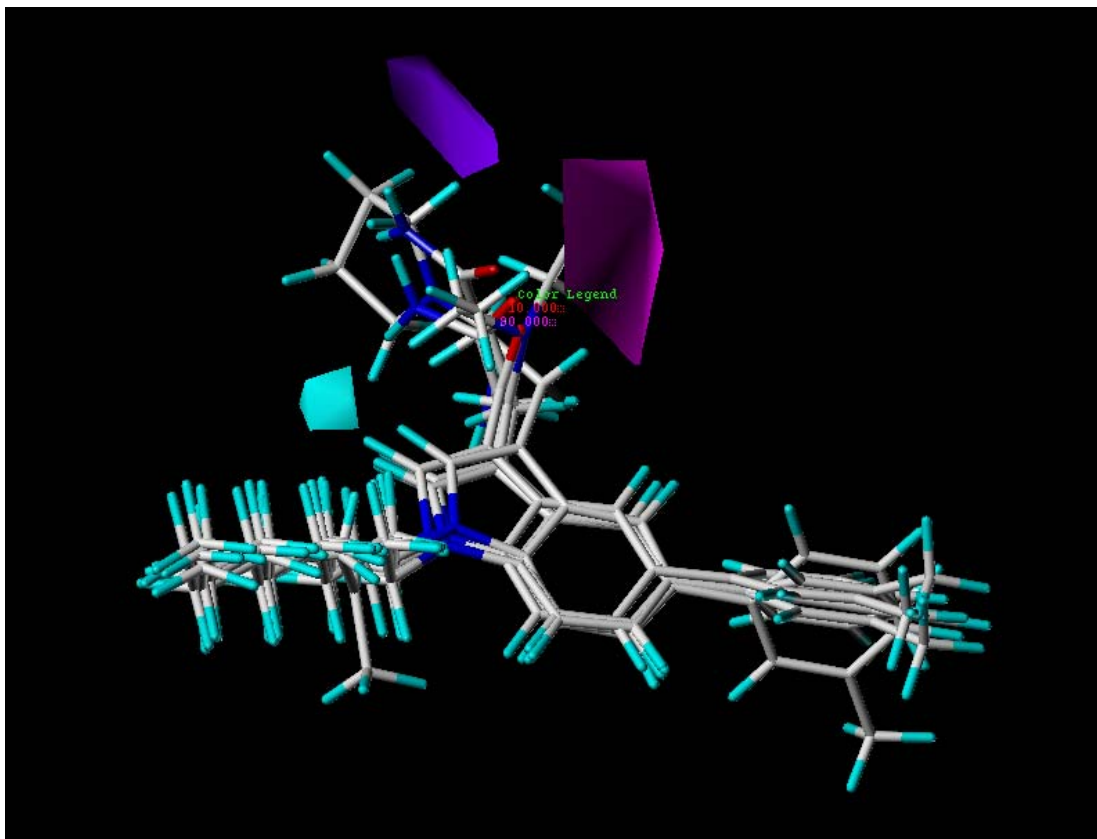
**Figure 5-6. Hydrogen bond donor/acceptor contour map.** Cyan contour represents region where presence of a hydrogen bond donor group would enhance activity while purple contour represents area where hydrogen bond donor group is disfavoured. On the other hand, magenta contour indicates preference for hydrogen bond acceptor group at that region.

As for the hydrophobic contours (**Figure 5-6**), an orange zone represented areas where hydrophobicity was favoured while the grey zone represented areas where hydrophilicity was favoured. The orange zone was located at the position of the 5-phenyl ring, implying the need to maintain a hydrophobic residue at this position. The presence of the orange zone at this position rather than at position 1 was surprising as the absence of a substituent at position 1 resulted in a far greater loss in activity than the omission of a group at position 5 (**Section 4.3.2**). This anomaly may have arisen because there were far more compounds without a 5-substituent (six) than without a 1-substituent (one), thus skewing the hydrophobic contours in favour of position 5. The grey zone at the side chain located at position 3 reinforced the role of

this side chain as a source of polarity and hydrophilicity in an otherwise hydrophobic scaffold.
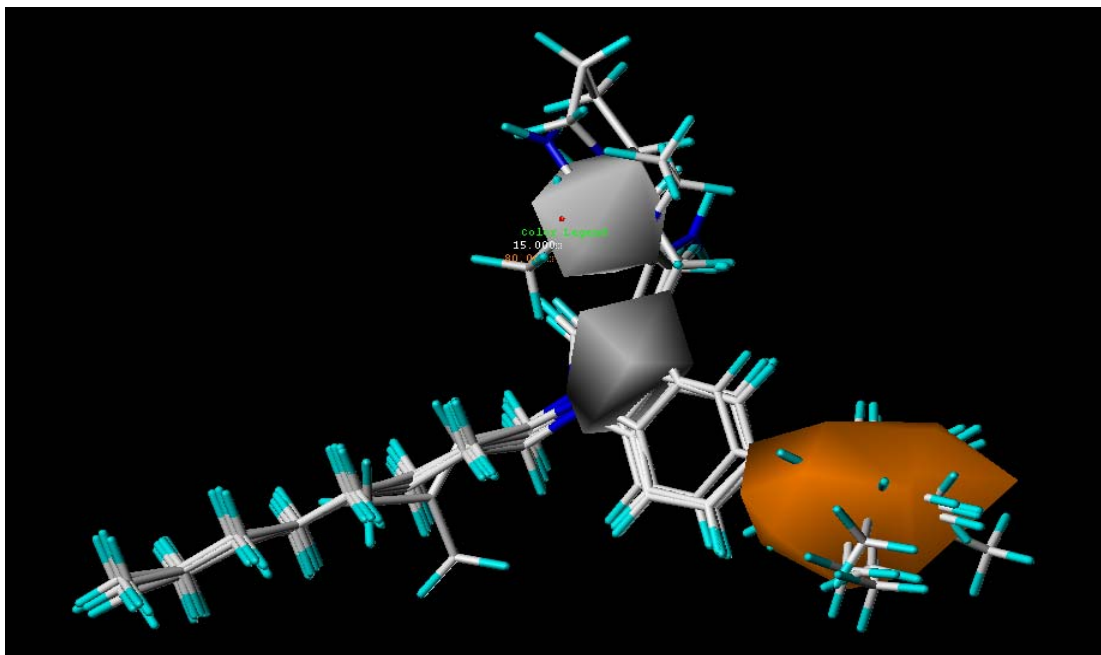


**Figure 5-7. CoMSIA hydrophobic contour map.** Orange contour represents region where hydrophobicity is favoured while grey contour represents area where hydrophilicity is preferred.

Taken together, the present findings showed that the additional information provided by CoMSIA was not as useful as anticipated and the same conclusions regarding the steric and electrostatic contributions could be obtained from the CoMFA model.

## 5.4. CONCLUSION

The objectives of this chapter were to establish a pharmacophore model for Icmt inhibition and to explain the structure-activity trends observed for the inhibitory activities of the indole analogues by quantitative means. In order to obtain the pharmacophore model, the test compounds were aligned in their non-protonated forms and with their side chains at position 1 restrained to a fixed conformation.

These conditions were imposed because the most active compounds had (i) amino groups in their side chains and these could interact with the enzyme in either the protonated or non-protonated state, and (ii) most of the compounds had highly flexible *n*-octyl side chains that could assume a large number of conformations. With these restrictions, a five point pharmacophore model was obtained. It comprised 3 aromatic features, 1 hydrophobic feature and a H bond acceptor feature. The model had good accuracy (91.4 %) and identified all but one of the actives ($IC_{50}$ value of < 5 μM) and most of the inactives (9 out of 12). It was able to accommodate 5 known Icmt inhibitors that were not structurally related to cysmethynil.

Three approaches were employed to analyze the SAR of the test compounds, namely multiple linear regression (MLR), partial least squares projection to latent structures (PLS) and a 3D approach based on CoMFA/CoMSIA. Both MLR and PLS identified size and hydrophobicity as the main determinants of activity. This was not unexpected given the presence of 3 aromatic and 1 hydrophobic feature in the pharmacophore model. Because some of the test compounds could interact with Icmt in their protonated or non-protonated states, MLR and PLS were carried out separately on protonated and non-protonated compounds. Regardless of the state of protonation, size and hydrophobicity were the main determinants of activity. However, more representative regression equations were derived when MLR and PLS were carried out on the protonated states, in that descriptors specific for the protonated amines were detected and their correlation to $pIC_{50}$ value reflected the better activities of these compounds. The CoMFA model showed that the side chains at positions 1 was primarily responsible for the steric contributions to activity while the side chain at 3 was the sole determinant of polarity.

# CHAPTER 6: INVESTIGATIONS INTO THE EFFECT OF SELECTED DERIVATIVES OF CYSMETHYNIL ON VIABILITY OF CANCER CELL LINES

## 6.1. INTRODUCTION

Several compounds have been proposed as promising substitutes of cysmethynil because they had comparable, if not better, Icmt inhibitory and cell growth inhibitory activities, besides having a more favourable physicochemical profile that might enhance its pharmacokinetic properties. The most notable examples were the amino analogues **4-3**, **6-4** and **6-5**. These compounds were protonated at physiological pH and would be expected to be less lipophilic than cysmethynil. This might be true for **6-4** and **6-5** but not **4-3** as seen from their estimated SlogP values (**Table 6-1**). Nonetheless, the impressive activity of **4-3** still made it a worthwhile candidate for consideration.

Previous investigations on cysmethynil showed that at moderate concentrations (20 μM, close to its $IC_{50}$ value), it inhibited the growth of human prostate cells but at higher concentrations (25, 30 μM), this cytostatic profile was augmented with cytotoxic activity[62]. Cysmethynil arrested the cell cycle at the G1 phase which might account for its cytostatic activity. Interesting, it also induced autophagic cell death. It was proposed that the inhibition of Icmt by cysmethynil led to a reduction of Ras and Rheb activity, which in turn reduced mTOR signaling, leading to cell cycle arrest and autophagy[62]. In this Chapter, **4-3** and **6-4** were investigated to determine if they mimicked cysmethynil in some of these activities, namely effects on the cell cycle and induction of autophagic cell death. If similar effects were observed, this would provide useful information on the functional consequences of Icmt inhibition in tumor cells.

**Table 6-1. Structures, IC50 values, ClogP and SlogP values of cysmethynil (1-1), 4-3, 6-4 and 6-5.**
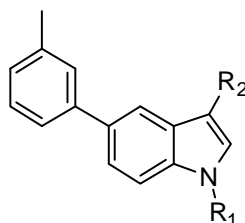


| Cpd | $R_1$ | $R_2$ | $IC_{50 \text{ Icmt}}$ [a] | $IC_{50 \text{ antiproliferative}}$ [a] | $ClogP$ [b] | $SlogP$ [c] |
|------|-------|-------|-----------|---------------------|--------|--------|
| **1-1** | $n$-$C_8H_{17}$ | -$CH_2CONH_2$ | 1.5 μM | 21.8 μM | 7.0 | 6.3 |
| **4-3** | $n$-$C_8H_{17}$ | -$CH_2N(C_2H_5)_2$ | 0.7 μM | 3.6 μM | 9.6 | 6.9 |
| **6-4** | | -$CH_2N(C_2H_5)_2$ | 2.4 μM | 3.8 μM | 7.6 | 5.5 |
| **6-5** | | —$CH_2N$◯ | 2.1 μM | 3.9 μM | 7.7 | 5.7 |

[a] From **Table 4-1**. $IC_{50 \text{ antiproliferative}}$ were determined on MDA-MB-231 cells.

[b] From **Table 3-2**, **3-5** and **3-7**. Log of octanol /water partition coefficient calculated for non-protonated species with ChemDraw 9.0.

[c] Log of octanol /water partition coefficient calculated from protonated form (except cysmethynil **1-1**) with MOE 2008.1001.


## 6.2 MATERIALS AND METHODS

### 6.2.1. Materials

Materials used were as described in **Chapter 4**. The PC3 prostate cancer cells were from American Type Cell Culture (ATCC, Rockville, MD). Antibodies directed against LC3 (MAP1LC3B) and GAPDH were purchased from Abgent (San Diego, CA) and Cell Signalling (Danvers, MA) respectively. Propidium iodide and ribonuclease A were obtained from Sigma Chemical Compamy, Singapore.

### 6.2.2. Cell Viability Assay

MDA-MB-231 human breast cancer cells and PC3 prostate cancer cells were grown in DMEM supplemented with 10 % fetal bovine serum, 50 U/ml penicillin and 50 µg/ml streptomycin at 37 °C with 5 % $CO_2$. Cells were subcultured at 80-90 % confluency and used within 10-25 passages for the cell-based assays.

Cell viability assay was carried out as described in **Chapter 4**. Briefly, cells were seeded at 2400 cells per well in DMEM containing 5 % FBS in 96-well plates for 24 h prior to treatment with various concentrations of test compound or vehicle (media) for 72 h. Final concentration of DMSO in the media was maintained at 0.5 % v/v per well. The relative number of the live cells was determined using the CellTiter® 96 AQueous One Solution Cell Proliferation Assay (Promega). Each condition was performed in triplicate, and data presented represent that obtained from at least two separate experiments.

### 6.2.3. Cell cycle analysis

MDA-MB-231 and PC3 cells were seeded at 50,000 cells/ml in 100mm dishes in DMEM containing 5 % FBS and incubated for 24 h prior to treatment with specific compounds or vehicle for 48 h. Cells were harvested by trypsinization, washed with PBS, pelleted by centrifugation at 300 x g for 5 mins and fixed in ice cold 70 % ethanol overnight. The cells were then stained by resuspension in PBS containing 40 µg/ml propidium iodide and 0.1 mg/ml ribonuclease A for 1 h at 37 °C. The stained cells were then analyzed for distribution in the G1, S and G2/M phases on a FC500 flow cytometer (Beckman Coulter, CA, USA) and analysed using the CXP analysis software.

## 6.2.4. Western blot analysis

MDA-MB-231 and PC3 cells were seeded at 50,000 cells/ml in 100mm dishes in DMEM containing 5 % FBS and incubated for 24 h prior to treatment with various concentrations of test compound or vehicle for 48 h. Cells were lysed in plate by the addition of RIPA (radioimmunoprecipitation assay) lysis buffer (Santa Cruz Biotechnology Inc., Santa Cruz, CA), transferred to 1.5 ml Eppendorf tubes and incubated on ice for 30 mins. The cell debris was removed by centrifugation (10,000 x g, 5 mins) and the supernatants were stored in aliquots at -20 °C. The protein content of the lysates was determined (BCA ® protein assay, ThermoScientific, Rockford, IL). The lysates were treated with Laemmli buffer (4x) and samples (10-20 μg protein) were resolved by 12 % SDS-PAGE gel electrophoresis at 110V for 90 mins. The resolved proteins were then transferred to polyvinylidene difluoride (PVDF) membranes (Amersham Hybond-P, GE Healthcare, UK) by wet transfer at 100 V for 80 mins. Membranes were then treated with 5% blocking buffer (comprising of phosphate buffered saline (PBS) with 5% non-fat milk and 0.1% Tween-20) and probed with a specific antibody (GAPDH or LC3) by incubating the membranes in blocking buffer containing the antibody at 4 °C overnight. After washing in PBS with 0.1% Tween 20 (PBST), the membranes previously probed with the primary antibody were reacted with secondary antibodies (horseradish peroxidase anti-mouse IgG) in blocking buffer at room temperature for 1 hour. They were then washed with PBST and developed using an enhanced chemiluminescence procedure (Amersham ECL Advanced Western Blotting Detection Kit, GE Healthcare, UK). Immunoblots were viewed using the ImageQuant RT-ECL imager (GE Healthcare) and the images captured with the ImageQuant TL software.

## 6.3 RESULTS

### 6.3.1. Effect of 4-3 and 6-4 on viability of human prostate (PC3) and breast (MDA-MB-231) cancer cells.

Both **4-3** and **6-4** were evaluated for their effects on the growth of MDA-MB-231 cells (**Section 4.3.5**) and found to be more potent ($\approx$ 6-fold) than cysmethynil in affecting cell viability although they were only comparable (**6-4**) or modestly more potent (**4-3**) than cysmethynil in terms of Icmt inhibitory activity. In this section, they were tested against another cancer cell line (PC3 prostate cancer cells) to confirm their effects on tumor cell growth. As seen from **Figure 6-1**, both compounds affected the viabilitiy of MDA-MB-231 and PC3 cells with comparable $IC_{50}$ values. A similar observation was made for cysmethynil while **3-1** (analogue of cysmethynil which is not an Icmt inhibitor) did not affect the viability of either cell line. As pointed out in an earlier report,[62] the results obtained with **3-1** provided strong evidence that the inhibition of Icmt contributed significantly to the effects on cell growth.
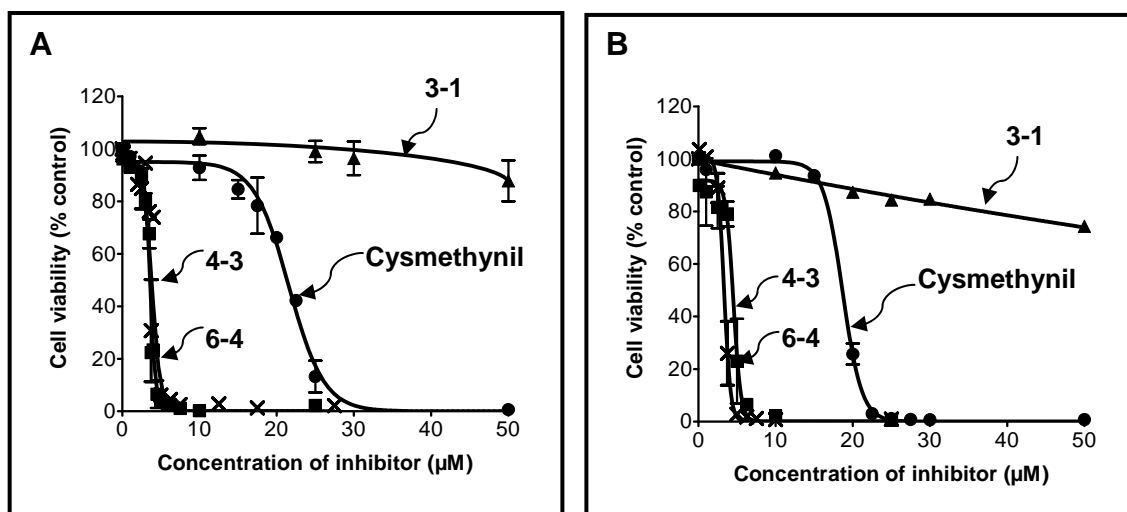


**Figure 6-1. Effect of cysmethynil (●), 3-1 (▲), 4-3 (■) and 6-4 (×) on cell viability (72 h) of (A) MDA-MB-231 cells and (B) PC3 cells.** $IC_{50\ PC3}$ values determined by the method described in Section 4.2.5 were 17.7 µM (cysmethynil), 4.0 µM (4-3) and 3.6 µM (6-4) compared to $IC_{50\ 231}$ values of 21.8 µM (cysmethynil), 3.6 µM (4-3) and 3.8 µM (6-4). Cell viability was expressed in terms of % DMSO control. Data represent mean and SD of triplicate determinations from 3 separate experiments.

The compounds might affect cell viability by killing them (cytotoxic mechanism) or inhibiting their growth and proliferation (cytostatic effect). This could be investigated by monitoring the viability of cells treated with varying concentrations of the test compound over several days. The results are given in **Figures 6-2** and **6-3**. Broadly similar outcomes were observed on both cell lines. When tested at a concentration that corresponded to their $IC_{50 \text{ Icmt}}$ value ($\approx$ 3.8 µM), **4-3** and **6-4** maintained cell viability for the first 2 days, and up to 3 days for MDA-MD-231 cells, but not PC3 cells. At higher concentrations (5 and 10 µM), both compounds showed concentration-dependent cytotoxic effects, with loss of cell viability observed after 24 hours at 10 µM. A similar profile was observed for cysmethynil on both cell lines which confirmed the earlier observation[62] that it had a biphasic effect on cell growth that was time and concentration-dependent.
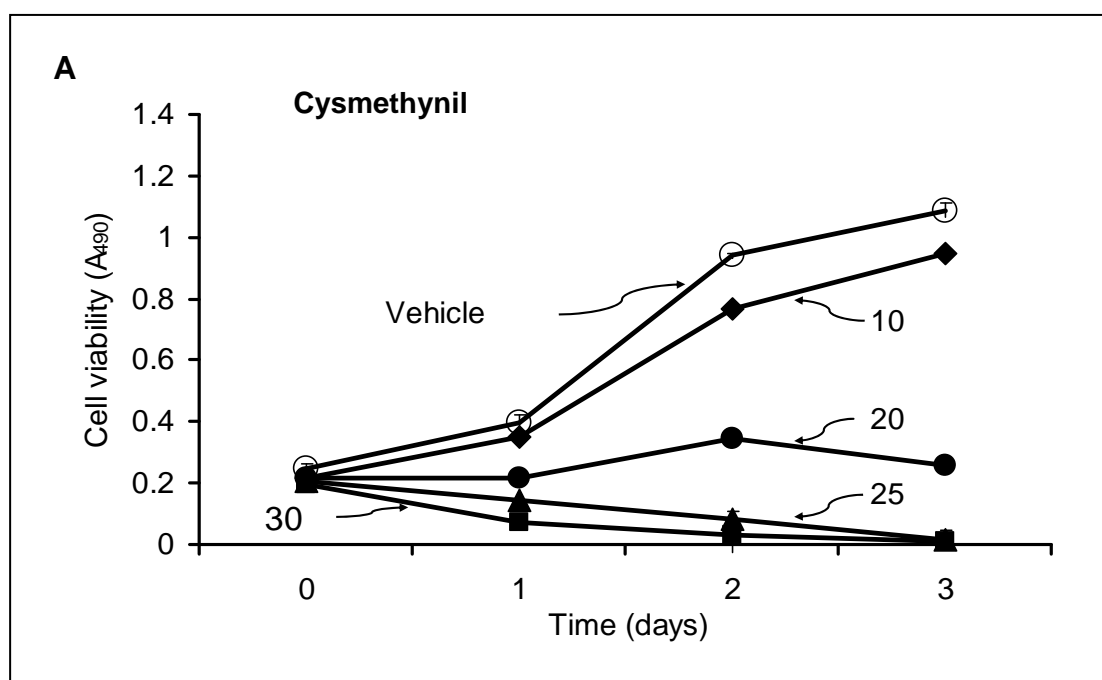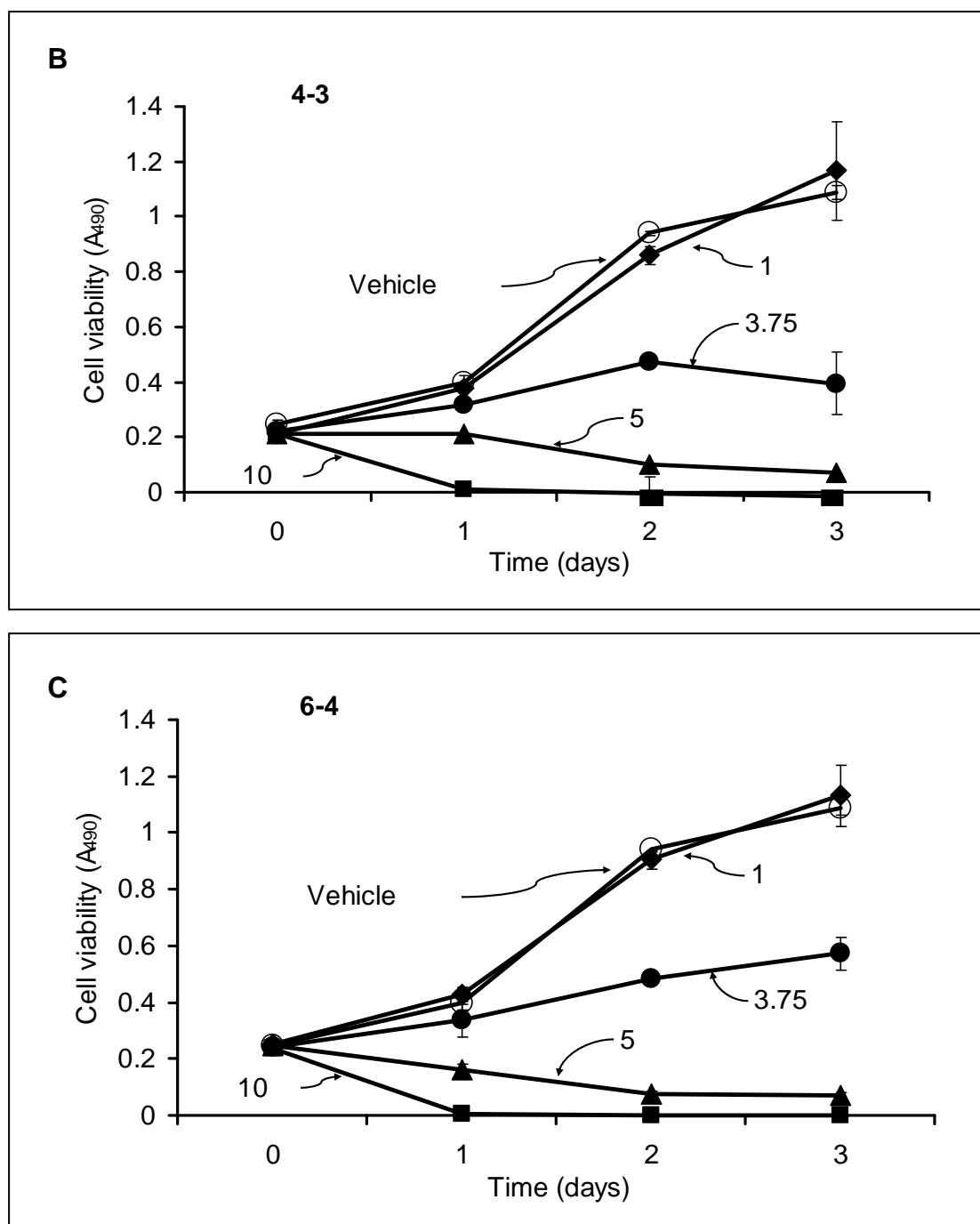
**Figure 6-2. Effect of (A) cysmethynil (0, 10, 20, 25, 30 µM), (B) 4-3 (0, 1, 3.75, 5, 10 µM), (C) 6-4 (0, 1, 3.75, 5, 10 µM) on viability of MDA-MB-231 cells over 3 days.** Cell viability was expressed in terms of absorbance readings at 490 nm. Data represent the mean of triplicate determinations from a single experiment that was repeated on 3 separate occasions.
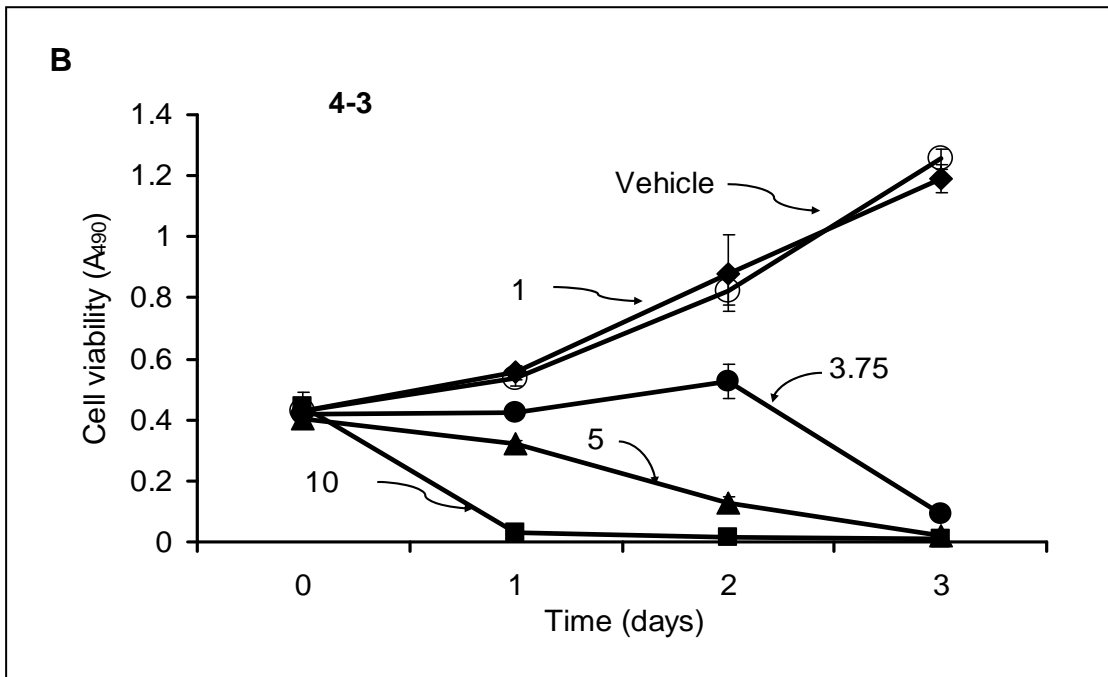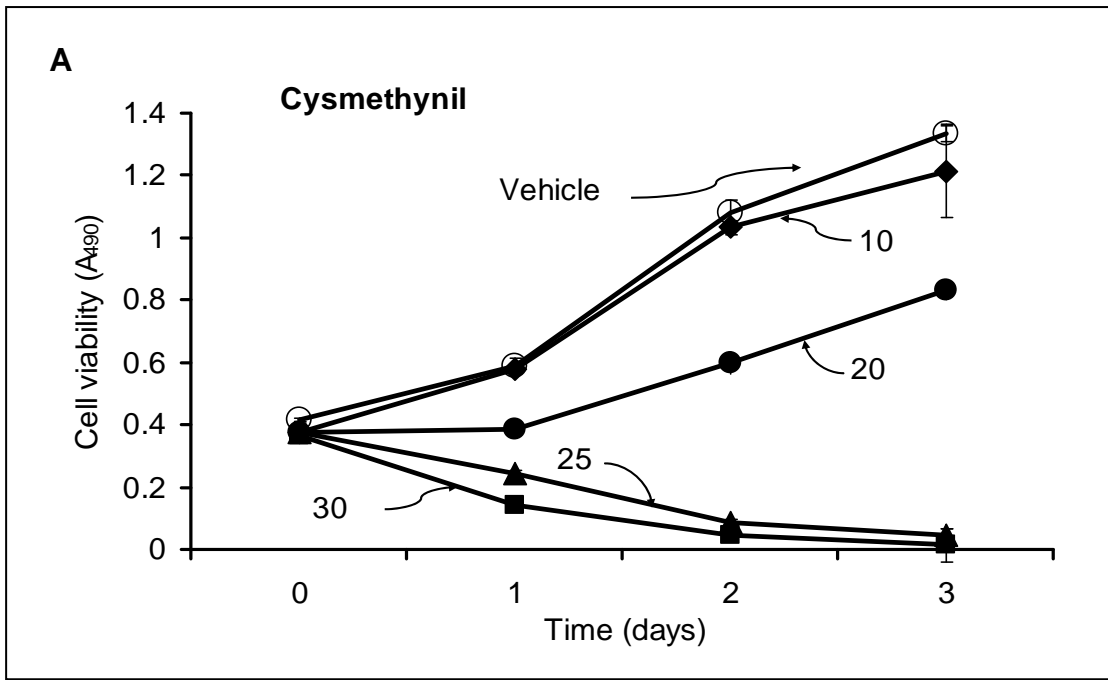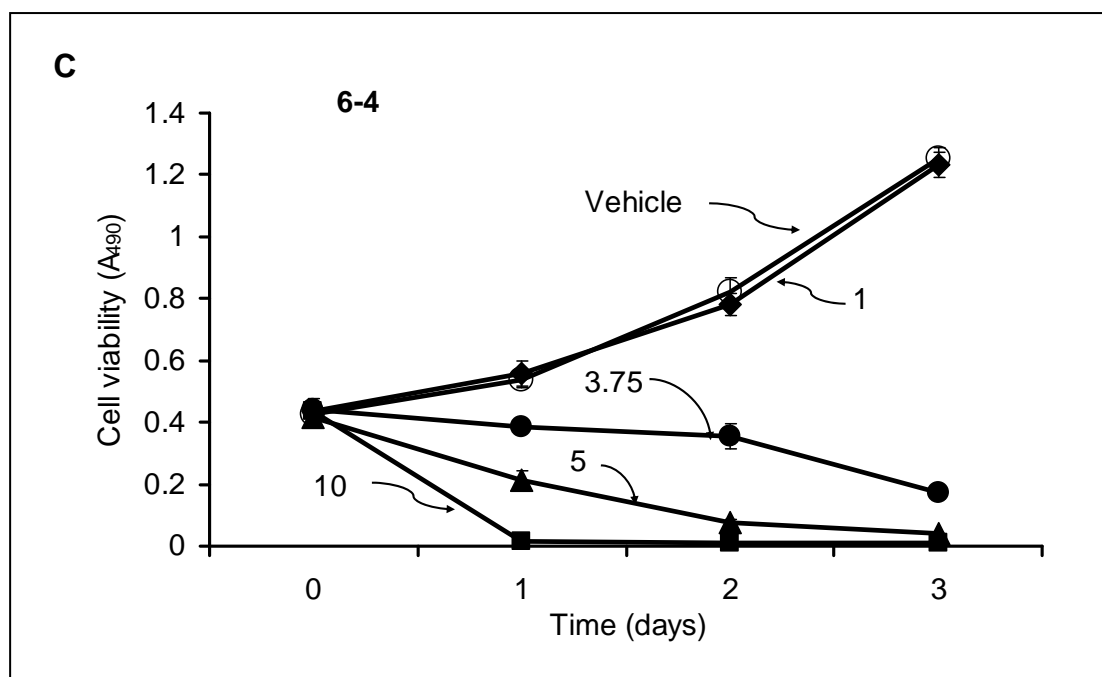
**Figure 6-3. Effect of (A) cysmethynil (0, 10, 20, 25, 30 μM), (B) 4-3 (0, 1, 3.75, 5, 10 μM), (C) 6-4 (0, 1, 3.75, 5, 10 μM) on viability of PC3 cells over time 3 days.** Cell viability was expressed in terms of absorbance readings at 490 nm. Data represent the mean of triplicate determinations from a single experiment that was repeated on 3 separate occasions.

## 6.3.2. Effects of 4-3 and 6-4 on cell cycle of MDA-MB-231 and PC3 cells

Cysmethynil was reported to arrest the cell cycle of PC3 cells at the G1 phase[62]. This finding was prompted by the observation that, at concentrations approximating to its $IC_{50}$ value, cysmethynil had a cystostatic effect. As **4-3** and **6-4** were also cytostatic at their $IC_{50}$ concentrations, this may suggest that they might also cause G1 arrest. Briefly, the G1 (gap 1) phase is one of the four phases that occur during cell growth and multiplication. At the G1 phase, the cell is actively growing and preparing to copy its DNA (2C) in response to various growth factors or internal signals. The next phase is the S (synthesis) phase where replication of DNA (4C) takes place. Once the chromosomes are copied, there is another gap or interval (G2, gap 2) during which the cell "checks" the copied DNA and repairs any damaged

copies. Finally, there is mitosis (M) where cell division takes place to produce two daughter cells, each with their complement of 2C DNA. Cells may re-enter the G1 phase to repeat the cycle or exit the G1 phase and enter into G0 to stop proliferation and if necessary, provide time for repair to take place, or for cell death to occur. Regulation of each step of the cycle is precise and complex, involving several proteins called cyclins, cylin dependent kinases (CDK) and CDK inhibitors.

The effects of **4-3**, **6-4** and **3-1** on the cell cycle were investigated by fluorescence activated cell sorter analysis (FACS) using flow cytometry. **Figure 6-4** shows the distribution of DNA/cells (MDA-MB-231) in the G1 and G2/M phases after incubation with the test compound for 24 hours. The first panel (i) shows the distribution of DNA in untreated cells. Cells treated with the inactive analogue **3-1** (Panel (v) of **Figure 6-4**) showed a similar distribution pattern, indicating that it did not interfere with the cell cycle transverse. In contrast, cells treated with cysmethynil, **4-3** and **6-4** had a higher proportion of DNA/cells in the G1 phase and a lesser proportion in the G2 phase, compared to untreated cells examined at the same time point (Panel (i) of **Figure 6-4**). This was indicative of the arrest of the cell cycle at the G1 phase. The same experiments were carried out on PC3 cells (**Figure 6-5**) and similar observations were made. However, it was noted that G1 arrest was more noticeable in the treated MDA-MB-231 cells, as seen from the larger proportion of DNA/cells in the G1 phase, particularly for **6-4** (71 % in G1 phase versus 52 % in the control for MDA-MB-231 cells, as compared to 47 % in G1 phase versus 38 % in the control for PC3 cells). **4-3** and **6-4** were clearly more potent than cysmethynil in inducing G1 arrest as they were tested at lower concentrations (4 μM) than cysmethynil (25 μM) but caused similar changes in the distribution of cells in the G1 and G2/M phases.
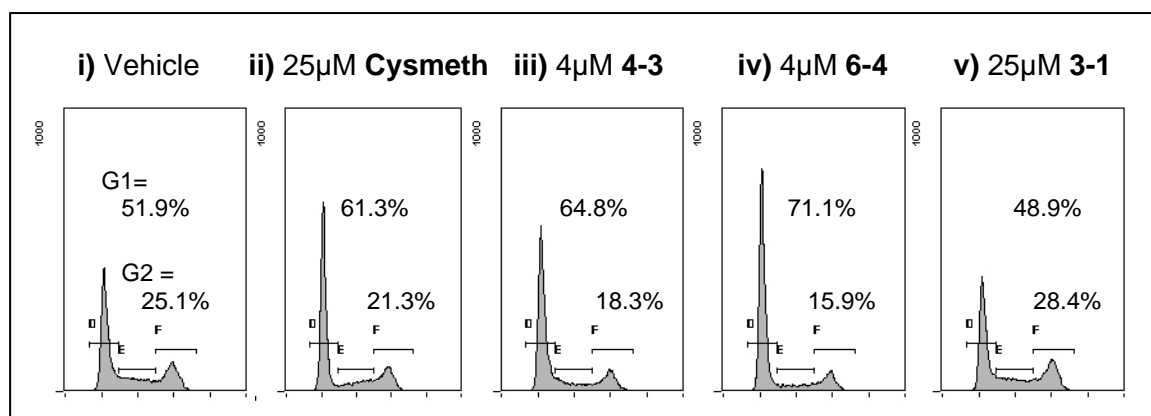
**Figure 6-4. DNA content analysis of MDA-MB-231 breast cancer cells, after 24 h incubation with (i) vehicle (DMSO), (ii) cysmethynil (25 μM); (iii) 4-3 (4 μM); (iv) 6-4 (4 μM) and (v) 3-1 (25 μM).** The % of cells in the G1 and G2 phases are indicated in each panel. Experiments were repeated on two separate occasions and (i)-(v) are representative diagrams.
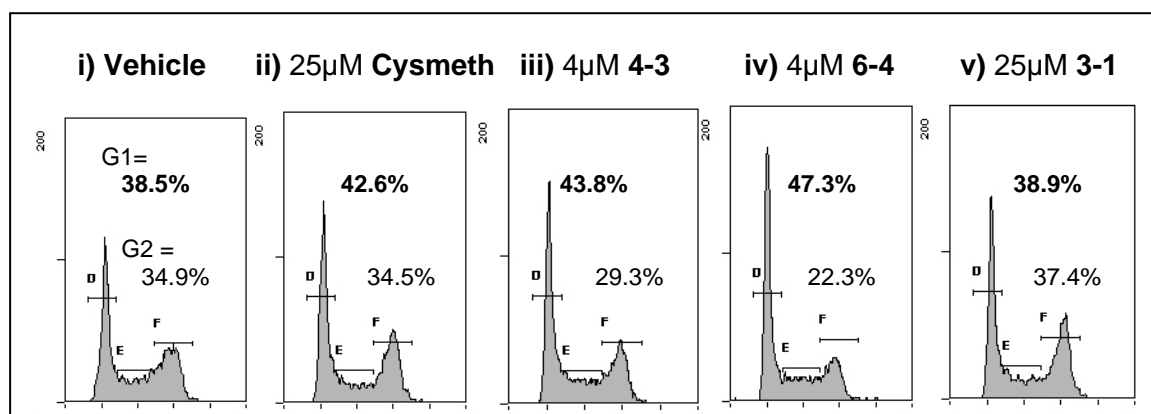


**Figure 6-5. DNA content analysis of PC3 prostate cancer cells after 24 h incubation with (i) vehicle (DMSO), (ii) cysmethynil (25 μM); (iii) 4-3 (4 μM); (iv) 6-4 (4 μM) and (v) 3-1 (25 μM).** The % of cells in the G1 and G2 phases are indicated in each panel. Experiments were repeated on two separate occasions and (i)-(v) are representative diagrams.

### 6.3.3. Effects of 4-3 and 6-4 on autophagic-induced cell death

Wang et al reported that cysmethynil also induced cell death in addition to arresting the cell cycle at the G1 phase[62]. Although inhibition of CaaX protein processing had been linked to apoptosis in many cell types[96], cysmethynil did not cause cell death by apoptosis even when PC3 cells were treated at concentrations that were cytotoxic to cells. Interestingly, cysmethynil was found to induce a non-apoptotic form of cell death called autophagy ("self-eating") which involved

lysosome-mediated degradation of proteins and cellular organelles, usually in response to starvation or stress.[97] In view of these findings, it was of interest to determine if **4-3** and **6-4** behaved in like manner. Both compounds were tested at two concentrations, namely 4 μM which corresponded to its $IC_{50}$ value and a lower concentration of 1 μM. A preliminary experiment was carried out at 5 μM as this concentration displayed prominent cytotoxic effects by 48 h (which was the incubation time employed for this investigation) (**Figures 6-2** and **6-3**). However, almost all the cells were killed at this concentration and this led to the use of 4 μM for the experiments. The marked difference on cell viability observed with such a small difference in concentration (4 μM vs 5 μM) may be related to the steep rising portion of the concentration-viability curves of **4-3** and **6-4** (**Figure 6-1**).

The microtubule-associated protein 1 light chain 3 (MAP1LC3, also known as LC3) was used as a marker for autophagy.[98] Briefly, autophagy begins with the production of double-membrane bound structures inside an intact cell. As these membrane structures elongate and mature, LC3 is recruited to the membrane. Prior to its recruitment, LC3 undergoes ubiquitin-like post-translational modifications, although involving phospholipid rather than peptide attachment, so that it could target the membrane structures. The processed form is called LC3-I and it resides in the cytosol. When autophagy is induced, a portion of the LC3-I is conjugated to phosphatidylethanolamine to give LC3-II, which associates tightly with the autophagosomal membrane. On SDS-PAGE, LC3-II migrates faster than LC3-I. Hence, immunoblotting of LC3 results in two bands, LC3-I and LC3-II. The amount of LC3-II or the ratio of LC3-II/LC3-I is correlated to the number of autophagosomes, and hence represents the autophagic activity in the cell. If **4-3** and **6-4** induced autophagy, then the LC3-II band should be more pronounced than LC3-I.

**Figure 6-6** and **6-7** shows the immunoblots of LC3-I and LC3-II obtained from MDA-MD-231 and PC3 cells that were exposed to test compounds. Cysmethynil was included as the positive control and the inactive analogue **3-1** was used as the negative control. Compounds **4-3** and **6-4** increased the content of LC3-II at 4 µM on both cell lines. The effect was concentration-dependent as seen from the less intense LC3-II bands detected at 1 µM of test compound. Compared to cysmethynil (25 µM), both **4-3** and **6-4** (at 4 µM) caused comparable or greater increases in LC3-II content. The inactive analogue **3-1** did not increase LC3-II levels at either 10 µM or 25 µM. These observations were supported by densitometric quantification of the blots to give the LC3-II/LC3-I ratios.
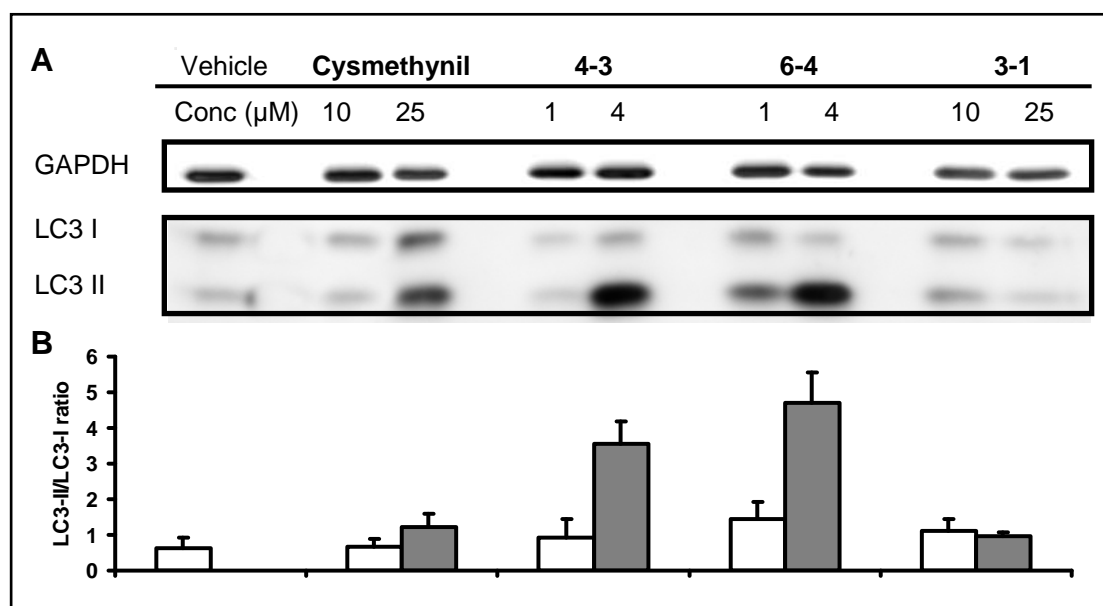


**Figure 6-6. Effect of cysmethynil, 4-3, 6-4 and 3-1 on LC3-1 and LC3-II protein levels in MDA-MB-231 breast cancer cell lysates, determined by western blot analysis after 48 h of incubation.** (A) Cell lysates (10-20 µg) were subjected to SDS-PAGE and probed with anti-LC3-I and anti-LC3-II antibodies. GADPH was used as the loading control. Immunoblots are representative results from a single experiment that was repeated at least three times. (B) These observations were supported by densitometric quantification of the blots to give the LC3-II/LC3-I ratios.
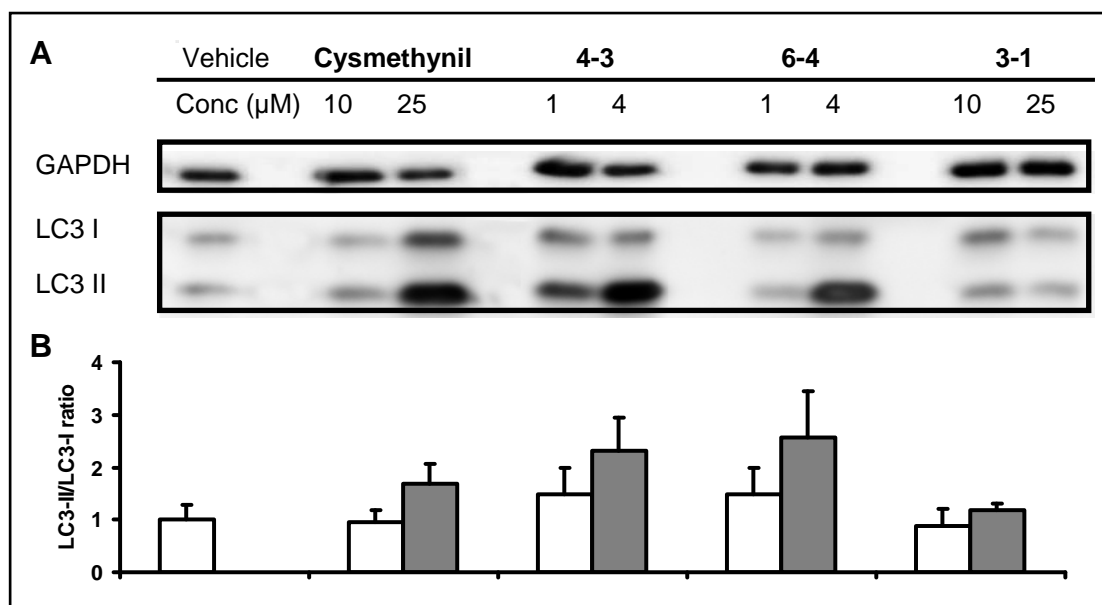
**Figure 6-7. Effect of cysmethynil, 4-3, 6-4 and 3-1 on LC3-1 and LC3-II protein levels in PC3 prostate cancer cell lysates, determined by western blot analysis after 48 h of incubation.** (A) Immunoblots are prepared as described in Figure 6-6. (B) Densitometric quantification of the blots to give the LC3-II/LC3-I ratios.

## 6.4. DISCUSSION

The purpose of this chapter was to determine if the Icmt inhibitors **4-3** and **6-4** affected the growth and viability of cancer cells in a similar fashion as cysmethynil. An earlier study has shown that cysmethynil had biphasic effects on the growth of prostate cancer cells (PC3), namely a cytostatic effect at concentrations approximating to its $IC_{50}$ value and a cytotoxic effect at higher concentrations. Cysmethynil was shown to interfere with the cell cycle transverse by arresting cells at the G1 phase in accord with its cytostatic action on cell growth. It enhanced autophagy and autophagic cell death. The findings in this chapter showed that **4-3** and **6-4** exhibited a dual cytostatic-cytotoxic profile that was dependent on concentration. They caused G1 arrest at their antiproliferative $IC_{50}$ values and increased the content of the autophagy biomarker (LC3-II) in PC3 and MDA-MD-231 cells in a concentration-dependent manner. While these effects were similar to that observed for

cysmethynil, **4-3** and **6-4** were evidently more potent in that they induced these effects at lower concentrations (4 μM for G1 arrest and increase in LC3-II content) compared to cysmethynil (25 μM for both effects). That the inhibition of Icmt was linked to these processes was inferred from **3-1** which neither inhibited Icmt nor interfered with the growth and viability of the PC3 and MDA-MD-231 cells. On the other hand, only **4-3** but not **6-4** was a more potent inhibitor of Icmt in vitro (**Table 6-1**), albeit over a narrow 3-fold range (0.7 to 2.3 μM) that may not significantly impact their Icmt inhibitory properties. Rather the improved potencies of **4-3** and **6-4** on these cell-based processes could be the result of structural or physicochemical features that favoured their access to the target site. These may be related to the protonated amino functionality present in both compounds which would influence properties (solubility, lipophilicity, protein binding characteristics) that are critical for passage across biological membranes.

Wang et al proposed a model to explain how the inhibition of Icmt by cysmethynil resulted in G1 cell cycle arrest and autophagy in PC3 cells[62]. In this model, it was proposed that inhibition of Ras- and possibly Rheb-methylation by Icmt activated the mTOR (mammalian target of rapamycin) pathway via a cascade of other regulatory proteins. mTOR had important effects on cell proliferation and autophagy. It promoted cell proliferation by reducing the levels of endogenous CDK2 inhibitors (such as p27$^{kip1}$) and inhibited autophagy. Thus when Icmt is inhibited, the defective Ras and/or Rheb diminished mTOR signalling which led to the induction of p27$^{kip1}$ and a block in the G1-S cell cycle progression. Inhibition of mTOR also led to the induction of autophagy.

Whether the induction of autophagy favours or restricts cancer cell growth has been widely discussed[97, 99-106]. That autophagy has an anticancer role comes from the

observations that oncogenes inhibited autophagy, tumor suppressors induced autophagy, and that a bone fide autophagy regulator (Beclin-1) was itself a tumor suppressor[107]. On the other hand, autophagy may keep tumor cells alive by supplying them with macromolecular precursors like amino acids, fatty acids and nucleotides, thus overriding nutrient deprivation and hypoxia caused by limited angiogenesis. Cancer cells are also known to undergo "protective autophagy" to protect themselves from the effects of anticancer therapies. In this context, the inhibition of autophagy may be preferred. Many anticancer agents have been reported to induce autophagy. These include tamoxifen, rapamycin, arsenic trioxide, and etoposide. But it remained to be seen if autophagy played a key role in their mechanism of action. Because autophagy occurs in tumor cells before they are killed, it does not necessarily mean that it contributed to cell death. It may be a mechanism by which the cells are striving to survive. However, the finding by Wang et al that PC3 cells could be protected from the cytotoxic effect of cysmethynil by inhibiting the autophagy pathway argues strongly that, in the case of autophagy induced by Icmt inhibition, the autophagic response contributes to the cell killing activity of cysmethynil.

## 6.5. CONCLUSION

Taken together, the findings in this chapter showed that Icmt inhibition resulted in the arrest of the cell cycle at the G1 phase and the induction of autophagic cell death. This was observed for the Icmt inhibitors **4-3** and **6-4** but not the inactive analogue **3-1**. **4-3** and **6-4** were more potent than cysmethynil possibly due to an improved physicochemical profile that facilated their passage across cellular membranes.

# CHAPTER 7 CONCLUSION AND FUTURE WORK

The aim of the thesis centred on the investigation of structure-activity relationships of the Icmt inhibitor cysmethynil. It was hypothesized that the anticancer activity of cysmethynil could be enhanced by undertaking structural changes at the primary amide side chain of the molecule, and by concurrently incorporating more drug or lead-like features in its structure. The focus on the primary amide side chain was prompted by the fact that this was the only position on cysmethynil that had not been previously investigated. Moreover, modifications at other pendant groups on the indole ring did not result in more potent compounds than cysmethynil. Although cysmethynil incorporated several "drug like" features (compliance to the number of H bond donors and acceptors, molecular weight), its estimated lipophilicity (ClogP >5) and rotatable bond count exceeded or bordered to the recommended threshold limits. Thus, one objective of this investigation was to propose less lipophilic/more soluble cysmethynil analogues for synthesis and to determine how far this could be achieved without compromising Icmt inhibitory activity.

To achieve this objective, forty six analogues of cysmethynil were synthesized. The design of these compounds, particularly modifications at positons 1 and 5 of the scaffold, was aided in part by the analysis of structure-activity trends derived from the original library of indoleacetamides which gave cysmethynil. The analysis which was carried out with multiple linear regression (MLR), projection methods and comparative molecular field analysis (CoMFA), showed that there were contrasting steric and lipophilic requirements at positions 1 and 5, with more bulk and lipophilicity favoured at position 1 than 5. Notwithstanding the requirement for bulkiness and lipophilicity at position 1, the steric CoMFA fields showed two optimal lengths at this position, with a long side chain corresponding to *n*-octyl and a shorter

one typified by a trifluoromethylbenzyl side chain. Modifications at the acetamide side chain were guided by conventional analogue-based approaches like isosteric replacement, homologation and functional group changes. In all, 46 novel indole derivatives were synthesized and screened for Icmt inhibitory activity and cell viability effects on a human breast cancer cell line (MDA-MB-231).

The Icmt inhibitory activity of cysmethynil was found to be critically dependent on the presence of the acetamide side chain. Omission of this side chain resulted in a significant loss of activity, far exceeding that observed when the 1-(*n*-octyl) or the 5-substituted phenyl ring was omitted from the molecule. An analysis of the structure-activity relationships (SAR) of the series suggested that a lipophilic residue should be present at positon 1, a hydrogen (H) bond acceptor entity at position 3 and a group equivalent to a phenyl ring at position 5. Modifications at the latter position resulted in mostly incremental changes in activity and could be explored in future work as a possible site for the introduction of functionalities to moderate drug-likeness of the final compound.

A significant correlation (Spearman rho value of 0.707, p< 0.01) was observed between $IC_{50}$ values for Icmt inhibition and cell viability effects. With few exceptions, similar structural trends were observed for the two activities. The most promising compounds were the amine analogues of cysmethynil in which the acetamide side chain was replaced by an amino group. Most of the amines (notably **4-3**) showed a modest improvement in Icmt inhibitory (2-fold) but were significantly more potent than cysmethynil (7-fold) in affecting cell viability. Other amines with isoprenyl sidechains (**6-4**, **6-5**) that had comparable Icmt inhibitory activities as cysmethynil also showed improved antiproliferative activities. These compounds are proposed as more drug-like alternatives to cysmethynil in view of their protonated states at

physiological pH which could translate to lower lipophilicities and improved solubilities. In fact, their improved potencies on the cell based viability assay which involved passage across biological membranes might have stemmed from a more favourable physicochemical profile.

Computational analysis of the Icmt inhibitory activities of the test compounds yielded a five point pharmacophore model comprising of 3 aromatic features (5-phenyl and indole rings), 1 hydrophobic feature (side chain at position 1) and a polar H bond acceptor feature (at position 3). This model required the amino analogues to be represented in their non-protonated rather than protonated states. It may be that the physiologically relevant form of the compound (protonated state) need not be the "active" entity that interacted with the enzyme. Noting that Icmt is a membrane bound enzyme located in the hydrophobic environment of the endoplasmic reticulum, it may be that the protonated amine analogue lost a proton in this environment and finally interacted with the enzyme in its non-protonated state. Notwithstanding this proposal, it was noted that a more informative model was derived when quantitative analysis of structure-activity trends were applied to compounds in their protonated states. The resulting equations derived from stepwise MLR and PLS regression identified prominent contributions from size and hydrophobicity to activity, and a lesser input by descriptors that were specific to the protonated states of the test compounds. The CoMFA steric and electrostatic fields closely reflected the pharmacophore model with sterically favourable green zones at the side chain at position 1, and prominent red zones (electron density is favoured) in the region of the electron lone pair of the amide carbonyl or amino nitrogen. Taken together, these varied but complementary computational tools supported a prominent steric/lipophilic role and a small but still essential polar contribution for good Icmt inhibitory activity.

The consequences of Icmt inhibition were investigated with the more promising Icmt inhibitors **4-3** and **6-4**. Both compounds were found to disrupt the G1 phase of the cell cycle and to increase the content of the autophagic biomarker LC3-II in human and prostate cancer cell lines. These effects closely mimicked that observed for cysmethynil, except that **4-3** and **6-4** were approximately 6-times more potent than cysmethynil in inducing these effects. The observation that **3-1** which had negligible Icmt inhibition, did not disrupt the cell cycle or induced autophagy gave added evidence of a link between these effects and Icmt inhibition.

The most promising compounds to emerge from this study were only modestly more potent than cysmethynil as Icmt inhibitors or antiproliferative agents. The improvement in their physicochemical profiles has not been rigorously tested in pharmacokinetic studies. The steep Hill slopes associated with their dose-response curves for Icmt inhibition and antiproliferative activity might imply that protein binding, a problem associated with cysmethynil, would still be present in the new compounds. Thus, there is still a need for Icmt inhibitors with improved physicochemical properties and greater potencies. The present findings are a useful starting point. In this context, the pharmacophore model constructed in this investigation, notwithstanding some limitations, would be a useful platform to screen against commercial databases of leadlike or druglike compounds. Promising leads would be tested for Icmt inhibitory activity and positive outcomes iteratively fed back to the model to improve it further. In this way, scaffolds other than the indole ring of cysmethynil might be discovered. Changes in the indole scaffold in cysmethynil have not been investigated to date.  This approach might yield novel leads with the desired characteristics of enhanced potency and drug-like features that would greatly aid in the quest for anticancer agents that act by the inhibition of Icmt.

# BIBLIOGRAPHY

1. Glomset, J. A.; Farnsworth, C. C. Role of protein modification reactions in programming interactions between ras-related GTPases and cell membranes. *Annu Rev Cell Biol* **1994,** 10, 181-205.

2. Zhang, F. L.; Casey, P. J. Protein prenylation: molecular mechanisms and functional consequences. *Annu Rev Biochem* **1996,** 65, 241-69.

3. Casey, P. J.; Seabra, M. C. Protein prenyltransferases. *J Biol Chem* **1996,** 271, 5289-92.

4. Clarke, S. Protein isoprenylation and methylation at carboxyl-terminal cysteine residues. *Annu Rev Biochem* **1992,** 61, 355-86.

5. Roberts, P. J.; Mitin, N.; Keller, P. J.; Chenette, E. J.; Madigan, J. P.; Currin, R. O.; Cox, A. D.; Wilson, O.; Kirschmeier, P.; Der, C. J. Rho Family GTPase modification and dependence on CAAX motif-signaled posttranslational modification. *J Biol Chem* **2008,** 283, 25150-63.

6. Boyartchuk, V. L.; Ashby, M. N.; Rine, J. Modulation of Ras and a-factor function by carboxyl-terminal proteolysis. *Science* **1997,** 275, 1796-800.

7. Ashby, M. N. CaaX converting enzymes. *Curr Opin Lipidol* **1998,** 9, 99-102.

8. Young, S. G.; Ambroziak, P.; Kim, E.; Clarke, S. In *The Enzymes*, 3rd ed.; Tamanoi, F.; Sigman, D. G., Eds. Academic: San Diego, 2001; Vol. 21, pp 156-213.

9. Chelsky, D.; Ruskin, B.; Koshland, D. E., Jr. Methyl-esterified proteins in a mammalian cell line. *Biochemistry* **1985,** 24, 6651-8.

10. Reid, T. S.; Terry, K. L.; Casey, P. J.; Beese, L. S. Crystallographic analysis of CaaX prenyltransferases complexed with substrates defines rules of protein substrate selectivity. *J Mol Biol* **2004,** 343, 417-33.

11. Winter-Vann, A. M.; Casey, P. J. Post-prenylation-processing enzymes as new targets in oncogenesis. *Nat Rev Cancer* **2005,** 5, 405-12.

12. Wennerberg, K.; Rossman, K. L.; Der, C. J. The Ras superfamily at a glance. *J Cell Sci* **2005,** 118, 843-6.

13. Bernards, A.; Settleman, J. GAP control: regulating the regulators of small GTPases. *Trends Cell Biol* **2004,** 14, 377-85.

14. Boguski, M. S.; McCormick, F. Proteins regulating Ras and its relatives. *Nature* **1993,** 366, 643-54.

15.     Joneson, T.; White, M. A.; Wigler, M. H.; Bar-Sagi, D. Stimulation of membrane ruffling and MAP kinase activation by distinct effectors of RAS. *Science* **1996,** 271, 810-2.

16.     Downward, J. Ras signalling and apoptosis. *Curr Opin Genet Dev* **1998,** 8, 49-54.

17.     Bos, J. L. ras oncogenes in human cancer: a review. *Cancer Res* **1989,** 49, 4682-9.

18.     Malumbres, M.; Barbacid, M. RAS oncogenes: the first 30 years. *Nat Rev Cancer* **2003,** 3, 459-65.

19.     Gschwind, A.; Fischer, O. M.; Ullrich, A. The discovery of receptor tyrosine kinases: targets for cancer therapy. *Nat Rev Cancer* **2004,** 4, 361-70.

20.     Rodenhuis, S. ras and human tumors. *Semin Cancer Biol* **1992,** 3, 241-7.

21.     Shields, J. M.; Pruitt, K.; McFall, A.; Shaub, A.; Der, C. J. Understanding Ras: 'it ain't over 'til it's over'. *Trends Cell Biol* **2000,** 10, 147-54.

22.     Downward, J. Targeting RAS signalling pathways in cancer therapy. *Nat Rev Cancer* **2003,** 3, 11-22.

23.     Wright, L. P.; Philips, M. R. Thematic review series: lipid posttranslational modifications. CAAX modification and membrane targeting of Ras. *J Lipid Res* **2006,** 47, 883-91.

24.     Basso, A. D.; Kirschmeier, P.; Bishop, W. R. Lipid posttranslational modifications. Farnesyl transferase inhibitors. *J Lipid Res* **2006,** 47, 15-31.

25.     Doll, R. J.; Kirschmeier, P.; Bishop, W. R. Farnesyltransferase inhibitors as anticancer agents: critical crossroads. *Curr Opin Drug Discov Devel* **2004,** 7, 478-86.

26.     Mazieres, J.; Pradines, A.; Favre, G. Perspectives on farnesyl transferase inhibitors in cancer therapy. *Cancer Lett* **2004,** 206, 159-67.

27.     Whyte, D. B.; Kirschmeier, P.; Hockenberry, T. N.; Nunez-Oliva, I.; James, L.; Catino, J. J.; Bishop, W. R.; Pai, J. K. K- and N-Ras are geranylgeranylated in cells treated with farnesyl protein transferase inhibitors. *J Biol Chem* **1997,** 272, 14459-64.

28.     Feldman, E. J.; Cortes, J.; DeAngelo, D. J.; Holyoake, T.; Simonsson, B.; O'Brien, S. G.; Reiffers, J.; Turner, A. R.; Roboz, G. J.; Lipton, J. H.; Maloisel, F.; Colombat, P.; Martinelli, G.; Nielsen, J. L.; Petersdorf, S.; Guilhot, F.; Barker, J.; Kirschmeier, P.; Frank, E.; Statkevich, P.; Zhu, Y.; Loechner, S.; List, A. On the use of lonafarnib in myelodysplastic syndrome and chronic myelomonocytic leukemia. *Leukemia* **2008,** 22, 1707-11.

29.     Martinelli, G.; Iacobucci, I.; Paolini, S.; Ottaviani, E. Farnesyltransferase inhibition in hematologic malignancies: the clinical experience with tipifarnib. *Clin Adv Hematol Oncol* **2008,** 6, 303-10.

30.     Peterson, Y. K.; Kelly, P.; Weinbaum, C. A.; Casey, P. J. A novel protein geranylgeranyltransferase-I inhibitor with high potency, selectivity, and cellular activity. *J Biol Chem* **2006,** 281, 12445-50.

31.     McGuire, T. F.; Qian, Y.; Vogt, A.; Hamilton, A. D.; Sebti, S. M. Platelet-derived growth factor receptor tyrosine phosphorylation requires protein geranylgeranylation but not farnesylation. *J Biol Chem* **1996,** 271, 27402-7.

32.     Vasudevan, A.; Qian, Y.; Vogt, A.; Blaskovich, M. A.; Ohkanda, J.; Sebti, S. M.; Hamilton, A. D. Potent, highly selective, and non-thiol inhibitors of protein geranylgeranyltransferase-I. *J Med Chem* **1999,** 42, 1333-40.

33.     Carrico, D.; Blaskovich, M. A.; Bucher, C. J.; Sebti, S. M.; Hamilton, A. D. Design, synthesis, and evaluation of potent and selective benzoyleneurea-based inhibitors of protein geranylgeranyltransferase-I. *Bioorg Med Chem* **2005,** 13, 677-88.

34.     Watanabe, M.; Fiji, H. D.; Guo, L.; Chan, L.; Kinderman, S. S.; Slamon, D. J.; Kwon, O.; Tamanoi, F. Inhibitors of protein geranylgeranyltransferase I and Rab geranylgeranyltransferase identified from a library of allenoate-derived compounds. *J Biol Chem* **2008,** 283, 9571-9.

35.     Kim, E.; Ambroziak, P.; Otto, J. C.; Taylor, B.; Ashby, M.; Shannon, K.; Casey, P. J.; Young, S. G. Disruption of the mouse Rce1 gene results in defective Ras processing and mislocalization of Ras within cells. *J Biol Chem* **1999,** 274, 8383-90.

36.     Bergo, M. O.; Leung, G. K.; Ambroziak, P.; Otto, J. C.; Casey, P. J.; Gomes, A. Q.; Seabra, M. C.; Young, S. G. Isoprenylcysteine carboxyl methyltransferase deficiency in mice. *J Biol Chem* **2001,** 276, 5841-5.

37.     Kato, K.; Cox, A. D.; Hisaka, M. M.; Graham, S. M.; Buss, J. E.; Der, C. J. Isoprenoid addition to Ras protein is the critical modification for its membrane association and transforming activity. *Proc Natl Acad Sci U S A* **1992,** 89, 6403-7.

38.     Michaelson, D.; Ali, W.; Chiu, V. K.; Bergo, M.; Silletti, J.; Wright, L.; Young, S. G.; Philips, M. Postprenylation CAAX processing is required for proper localization of Ras but not Rho GTPases. *Mol Biol Cell* **2005,** 16, 1606-16.

39.     Dai, Q.; Choy, E.; Chiu, V.; Romano, J.; Slivka, S. R.; Steitz, S. A.; Michaelis, S.; Philips, M. R. Mammalian prenylcysteine carboxyl methyltransferase is in the endoplasmic reticulum. *J Biol Chem* **1998,** 273, 15030-4.

40.     Jang, G. F.; Gelb, M. H. Substrate specificity of mammalian prenyl protein-specific endoprotease activity. *Biochemistry* **1998,** 37, 4473-81.

41.     Chen, Y.; Ma, Y. T.; Rando, R. R. Solubilization, partial purification, and affinity labeling of the membrane-bound isoprenylated protein endoprotease. *Biochemistry* **1996,** 35, 3227-37.

42.     Schlitzer, M.; Winter-Vann, A.; Casey, P. J. Non-peptidic, non-prenylic inhibitors of the prenyl protein-specific protease Rce1. *Bioorg Med Chem Lett* **2001,** 11, 425-7.

43.     Chen, Y. In *ASBMB/DBC-ACS Joint Meeting*, San Francisco, 1995; p Late Breaking Abstract No. LB 46.

44.     Porter, S. B.; Hildebrandt, E. R.; Breevoort, S. R.; Mokry, D. Z.; Dore, T. M.; Schmidt, W. K. Inhibition of the CaaX proteases Rce1p and Ste24p by peptidyl (acyloxy)methyl ketones. *Biochim Biophys Acta* **2007,** 1773, 853-62.

45.     Manandhar, S. P.; Hildebrandt, E. R.; Schmidt, W. K. Small-molecule inhibitors of the Rce1p CaaX protease. *J Biomol Screen* **2007,** 12, 983-93.

46.     Shoichet, B. K. Interpreting steep dose-response curves in early inhibitor discovery. *J Med Chem* **2006,** 49, 7274-7.

47.     Bergo, M. O.; Wahlstrom, A. M.; Fong, L. G.; Young, S. G. Genetic analyses of the role of RCE1 in RAS membrane association and transformation. *Methods Enzymol* **2008,** 438, 367-89.

48.     Shi, Y. Q.; Rando, R. R. Kinetic mechanism of isoprenylated protein methyltransferase. *J Biol Chem* **1992,** 267, 9547-51.

49.     Kramer, K.; Harrington, E. O.; Lu, Q.; Bellas, R.; Newton, J.; Sheahan, K. L.; Rounds, S. Isoprenylcysteine carboxyl methyltransferase activity modulates endothelial cell apoptosis. *Mol Biol Cell* **2003,** 14, 848-57.

50.     Wnuk, S. F.; Yuan, C. S.; Borchardt, R. T.; Balzarini, J.; De Clercq, E.; Robins, M. J. Anticancer and antiviral effects and inactivation of S-adenosyl-L-homocysteine hydrolase with 5'-carboxaldehydes and oximes synthesized from adenosine and sugar-modified analogues. *J Med Chem* **1997,** 40, 1608-18.

51.     Chiang, P. K.; Gordon, R. K.; Tal, J.; Zeng, G. C.; Doctor, B. P.; Pardhasaradhi, K.; McCann, P. P. S-Adenosylmethionine and methylation. *FASEB J* **1996,** 10, 471-80.

52.     Winter-Vann, A. M.; Kamen, B. A.; Bergo, M. O.; Young, S. G.; Melnyk, S.; James, S. J.; Casey, P. J. Targeting Ras signaling through inhibition of carboxyl methylation: an unexpected property of methotrexate. *Proc Natl Acad Sci U S A* **2003,** 100, 6529-34.

53.     Anderson, J. L.; Henriksen, B. S.; Gibbs, R. A.; Hrycyna, C. A. The isoprenoid substrate specificity of isoprenylcysteine carboxylmethyltransferase: development of novel inhibitors. *J Biol Chem* **2005,** 280, 29454-61.

54.     Henriksen, B. S.; Anderson, J. L.; Hrycyna, C. A.; Gibbs, R. A. Synthesis of desthio prenylcysteine analogs: sulfur is important for biological activity. *Bioorg Med Chem Lett* **2005,** 15, 5080-3.

55.     Donelson, J. L.; Hodges, H. B.; Macdougall, D. D.; Henriksen, B. S.; Hrycyna, C. A.; Gibbs, R. A. Amide-substituted farnesylcysteine analogs as inhibitors of human isoprenylcysteine carboxyl methyltransferase. *Bioorg Med Chem Lett* **2006,** 16, 4420-3.

56.     Ma, Y. T.; Shi, Y. Q.; Lim, Y. H.; McGrail, S. H.; Ware, J. A.; Rando, R. R. Mechanistic studies on human platelet isoprenylated protein methyltransferase: farnesylcysteine analogs block platelet aggregation without inhibiting the methyltransferase. *Biochemistry* **1994,** 33, 5414-20.

57.     Buchanan, M. S.; Carroll, A. R.; Fechner, G. A.; Boyle, A.; Simpson, M. M.; Addepalli, R.; Avery, V. M.; Hooper, J. N.; Su, N.; Chen, H.; Quinn, R. J. Spermatinamine, the first natural product inhibitor of isoprenylcysteine carboxyl methyltransferase, a new cancer target. *Bioorg Med Chem Lett* **2007,** 17, 6860-3.

58.     Buchanan, M. S.; Carroll, A. R.; Fechner, G. A.; Boyle, A.; Simpson, M.; Addepalli, R.; Avery, V. M.; Forster, P. I.; Guymer, G. P.; Cheung, T.; Chen, H.; Quinn, R. J. Small-molecule inhibitors of the cancer target, isoprenylcysteine carboxyl methyltransferase, from Hovea parvicalyx. *Phytochemistry* **2008,** 69, 1886-9.

59.     Buchanan, M. S.; Carroll, A. R.; Fechner, G. A.; Boyle, A.; Simpson, M.; Addepalli, R.; Avery, V. M.; Hooper, J. N.; Cheung, T.; Chen, H.; Quinn, R. J. Aplysamine 6, an alkaloidal inhibitor of Isoprenylcysteine carboxyl methyltransferase from the sponge Pseudoceratina sp. *J Nat Prod* **2008,** 71, 1066-7.

60.     Winter-Vann, A. M.; Baron, R. A.; Wong, W.; dela Cruz, J.; York, J. D.; Gooden, D. M.; Bergo, M. O.; Young, S. G.; Toone, E. J.; Casey, P. J. A small-molecule inhibitor of isoprenylcysteine carboxyl methyltransferase with antitumor activity in cancer cells. *Proc Natl Acad Sci U S A* **2005,** 102, 4336-41.

61.     Baron, R. A.; Peterson, Y. K.; Otto, J. C.; Rudolph, J.; Casey, P. J. Time-dependent inhibition of isoprenylcysteine carboxyl methyltransferase by indole-based small molecules. *Biochemistry* **2007,** 46, 554-60.

62.     Wang, M.; Tan, W.; Zhou, J.; Leow, J.; Go, M.; Lee, H. S.; Casey, P. J. A small molecule inhibitor of isoprenylcysteine carboxymethyltransferase induces autophagic cell death in PC3 prostate cancer cells. *J Biol Chem* **2008,** 283, 18678-84.

63.     Clarke, S.; Tamanoi, F. Fighting cancer by disrupting C-terminal methylation of signaling proteins. *J Clin Invest* **2004,** 113, 513-5.

64.     Philips, M. R. Methotrexate and Ras methylation: a new trick for an old drug? *Sci STKE* **2004,** 2004, pe13.

65.     Wahlstrom, A. M.; Cutts, B. A.; Liu, M.; Lindskog, A.; Karlsson, C.; Sjogren, A. K.; Andersson, K. M.; Young, S. G.; Bergo, M. O. Inactivating Icmt ameliorates K-RAS-induced myeloproliferative disease. *Blood* **2008,** 112, 1357-65.

66.     Wahlstrom, A. M.; Cutts, B. A.; Karlsson, C.; Andersson, K. M.; Liu, M.; Sjogren, A. K.; Swolin, B.; Young, S. G.; Bergo, M. O. Rce1 deficiency accelerates the development of K-RAS-induced myeloproliferative disease. *Blood* **2007,** 109, 763-8.

67.     Baron, R. A.; Casey, P. J. Analysis of the kinetic mechanism of recombinant human isoprenylcysteine carboxylmethyltransferase (Icmt). *BMC Biochem* **2004,** 5, 19.

68.     Cannon, J. G. Analog Design. In *Burger's Medicinal Chemistry and Drug Design*, 6th ed.; Abraham, D. J., Ed. Wiley: Hoboken, 2003; Vol. 1.

69.     Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discover and development settings. *Advanced Drug Delivery Reviews* **1997,** 23, 3-25.

70.     Veber, D. F.; Johnson, S. R.; Cheng, H. Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem* **2002,** 45, 2615-23.

71.     Kerns, E. H.; Di, L. *Drug-like Properties: Concepts, Structure Design and Methods*. Academic Press: Burlington, MA, 2008.

72.     Verloop, A.; Hoogenstraaten, W.; Tipker, J. In *Drug Design*, Ariens, E. J., Ed. Academic Press: New York, 1976; Vol. 7, pp 165-207.

73.     Livingstone, D. *Data analysis for chemist*. Oxford University Press: Oxford, 1995.

74.     Patrick, G. L. *An Introduction to Medicinal Chemistry*. 4th ed.; Oxford University Press: Oxford, 2001.

75.     Thornber, C. W. Isosterism and molecular modification in drug design. *Chem Soc Rev* **1979,** 8, 563-80.

76.     Ciapetti, P.; Giethlen, B. Molecular Variations Based on Isosteric Replacements. In *The Practice of Medicinal Chemistry*, 3rd ed.; Wermuth, C. G., Ed. Academic Press: Boston, 2008; pp 290-342.

77.     Wolfe, J. P.; Nakhla, J. S. The Suzuki Reaction. In *Name Reactions for Homologations*, J., L. J., Ed. Wiley: 2009; Vol. I.

78.     Roy, S.; Eastman, A.; Gribble, G. W. Synthesis of *N*-alkyl substituted bioactive indolocarbazoles related to Gö6976 *Tetrahedron* **2006,** 62, 7838-45.

79.    Li, Z.; Lucas, N. T.; Wang, Z.; Zhu, D. Facile synthesis of Janus "double-concave" tribenzo[a,g,m]coronenes. *J Org Chem* **2007,** 72, 3917-20.

80.    Na, Y. M.; Le Borgne, M.; Pagniez, F.; Le Baut, G.; Le Pape, P. Synthesis and antifungal activity of new 1-halogenobenzyl-3-imidazolylmethylindole derivatives. *Eur J Med Chem* **2003,** 38, 75-87.

81.    Jansen, M.; Potschka, H.; Brandt, C.; Loscher, W.; Dannhardt, G. Hydantoin-substituted 4,6-dichloroindole-2-carboxylic acids as ligands with high affinity for the glycine binding site of the NMDA receptor. *J Med Chem* **2003,** 46, 64-73.

82.    Bascop, S. I.; Laronze, J. Y.; Sapi, J. Synthetis of 2-aminopropyle-3-indole-acetic(propionic) acid derivatives. *ARKIVOC* **2003**, 46-61.

83.    McGovern, S. L.; Caselli, E.; Grigorieff, N.; Shoichet, B. K. A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *J Med Chem* **2002,** 45, 1712-22.

84.    McGovern, S. L.; Helfand, B. T.; Feng, B.; Shoichet, B. K. A specific mechanism of nonspecific inhibition. *J Med Chem* **2003,** 46, 4265-72.

85.    Feng, Y.; Mitchison, T. J.; Bender, A.; Young, D. W.; Tallarico, J. A. Multi-parameter phenotypic profiling: using cellular effects to characterize small-molecule compounds. *Nat Rev Drug Discov* **2009,** 8, 567-78.

86.    Halgren, T. A. MMFF VII. Characterization of MMFF94, MMFF94s, and Other Widely Available Force Fields for Conformational Energies ad for Intermolecular-Interaction Energies and Geometries. *J Comput Chem* **1999,** 20, 730-48.

87.    Halgren, T. A. MMFF VI. MMFF94s Option for Energy Minimization Studies. *J Comput Chem* **1999,** 20, 720-9.

88.    Chen, I. J.; Foloppe, N. Conformational sampling of druglike molecules with MOE and catalyst: implications for pharmacophore modeling and virtual screening. *J Chem Inf Model* **2008,** 48, 1773-91.

89.    Wermuth, C. G.; Ganellin, C. R.; Lindberg, P.; Mitscher, L. A. Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure & Appl. Chem.* **1998,** 70, 1129-43.

90.    Sun, H. Pharmacophore-based virtual screening. *Curr Med Chem* **2008,** 15, 1018-24.

91.    Mason, J. S.; Good, A. C.; Martin, E. J. 3-D pharmacophores in drug discovery. *Curr Pharm Des* **2001,** 7, 567-97.

92.    Khedkar, S. A.; Malde, A. K.; Coutinho, E. C.; Srivastava, S. Pharmacophore modeling in drug discovery and development: an overview. *Med Chem* **2007,** 3, 187-97.

93.     Sippl, W. Pharmacophore Identification and Pseudo-Receptor Modelling. In *The Practice of Medicinal Chemistry*, 3rd ed.; Wermuth, C. G., Ed. Academic Press: Boston, 2008; pp 572-86.

94.     Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J Med Chem* **1994,** 37, 4130-46.

95.     Kearsley, S. K.; Smith, G. M. An Alternative Method for the Alignment of Molecular Structures: Maximizing Electrostatic and Steric Overlap *Tetrahedron Computer Methodology* **1990,** 3, 615-33.

96.     Sebti, S. M. Protein farnesylation: implications for normal physiology, malignant transformation, and cancer therapy. *Cancer Cell* **2005,** 7, 297-300.

97.     Kondo, Y.; Kanzawa, T.; Sawaya, R.; Kondo, S. The role of autophagy in cancer development and response to therapy. *Nat Rev Cancer* **2005,** 5, 726-34.

98.     Mizushima, N. Methods for monitoring autophagy. *Int J Biochem Cell Biol* **2004,** 36, 2491-502.

99.     Ogier-Denis, E.; Codogno, P. Autophagy: a barrier or an adaptive response to cancer. *Biochim Biophys Acta* **2003,** 1603, 113-28.

100.    Furuta, S.; Hidaka, E.; Ogata, A.; Yokota, S.; Kamata, T. Ras is involved in the negative control of autophagy through the class I PI3-kinase. *Oncogene* **2004,** 23, 3898-904.

101.    Kanzawa, T.; Germano, I. M.; Komata, T.; Ito, H.; Kondo, Y.; Kondo, S. Role of autophagy in temozolomide-induced cytotoxicity for malignant glioma cells. *Cell Death Differ* **2004,** 11, 448-57.

102.    Opipari, A. W., Jr.; Tan, L.; Boitano, A. E.; Sorenson, D. R.; Aurora, A.; Liu, J. R. Resveratrol-induced autophagocytosis in ovarian cancer cells. *Cancer Res* **2004,** 64, 696-703.

103.    Scarlatti, F.; Bauvy, C.; Ventruti, A.; Sala, G.; Cluzeaud, F.; Vandewalle, A.; Ghidoni, R.; Codogno, P. Ceramide-mediated macroautophagy involves inhibition of protein kinase B and up-regulation of beclin 1. *J Biol Chem* **2004,** 279, 18384-91.

104.    Boya, P.; Gonzalez-Polo, R. A.; Casares, N.; Perfettini, J. L.; Dessen, P.; Larochette, N.; Metivier, D.; Meley, D.; Souquere, S.; Yoshimori, T.; Pierron, G.; Codogno, P.; Kroemer, G. Inhibition of macroautophagy triggers apoptosis. *Mol Cell Biol* **2005,** 25, 1025-40.

105.    Ellington, A. A.; Berhow, M.; Singletary, K. W. Induction of macroautophagy in human colon cancer cells by soybean B-group triterpenoid saponins. *Carcinogenesis* **2005,** 26, 159-67.

106.    Moretti, L.; Yang, E. S.; Kim, K. W.; Lu, B. Autophagy signaling in cancer and its potential as novel target to improve anticancer therapy. *Drug Resist Updat* **2007,** 10, 135-43.

107.    Hippert, M. M.; O'Toole, P. S.; Thorburn, A. Autophagy in cancer: good, bad, or both? *Cancer Res* **2006,** 66, 9349-51.

108.    Andreani, A.; Granaiola, M.; Leoni, A.; Locatelli, A.; Morigi, R.; Rambaldi, M.; Roda, A.; Guardigli, M.; Traniello, S.; Spisani, S. N-Benzyl-2-chloroindole-3-carboxylic acids as potential anti-inflammatory agents. Synthesis and screening for the effects on human neutrophil functions and on COX1/COX2 activity. *Eur J Med Chem* **2004,** 39, 785-91.

# APPENDIX 1. QSAR OF INDOLOACETAMIDES

## Table A1-1. Descriptors of compounds in database.

| Comp | σPh[1] | σN[1] | CLOGP[2] | πPh[3] | πPh₂[3] | πN[3] | AREA[4] | PSA[5] | PV[6] | VOL[7] | CMR[8] | DM[9] | HOMO | LUMO | PL1[10] | PB1[10] | PB5[10] | NL1[11] | NB1[11] | NB5[11] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1A1 | -0.06 | -0.015 | 4.75 | 0.50 | 2.38 | 1.92 | 647.537 | 106.277 | 161.757 | 1028.175 | 9.97 | 3.92 | -0.357 | 0.101 | 6.58 | 1.85 | 4.46 | 5.15 | 1.71 | 4.66 |
| 1B1 | -0.06 | 0.013 | 4.27 | 0.50 | 2.39 | 1.44 | 646.709 | 106.790 | 163.014 | 1003.580 | 9.83 | 4.30 | -0.358 | 0.104 | 6.58 | 1.85 | 4.46 | 5.03 | 1.70 | 4.70 |
| 1C1 | -0.06 | -0.032 | 5.94 | 0.50 | 2.38 | 3.11 | 722.312 | 106.094 | 149.055 | 1133.590 | 10.90 | 3.56 | -0.359 | 0.096 | 6.58 | 1.85 | 4.46 | 8.34 | 1.71 | 6.32 |
| 1D | -0.06 | -0.032 | 7.00 | 0.50 | 2.39 | 4.17 | 778.714 | 106.772 | 116.062 | 1206.793 | 11.82 | 3.56 | -0.359 | 0.096 | 6.58 | 1.85 | 4.46 | 10.43 | 1.70 | 7.82 |
| 1E | -0.06 | 0.058 | 5.07 | 0.50 | 2.39 | 2.24 | 666.768 | 102.688 | 140.091 | 1080.034 | 11.09 | 3.24 | -0.362 | 0.099 | 6.58 | 1.85 | 4.46 | 5.37 | 1.70 | 6.26 |
| 1G1 | -0.06 | 0.078 | 5.95 | 0.50 | 2.39 | 3.12 | 714.462 | 102.309 | 163.132 | 1188.415 | 11.60 | 4.59 | -0.365 | 0.088 | 6.58 | 1.85 | 4.46 | 5.47 | 1.70 | 7.61 |
| 1H1 | -0.06 | 0.078 | 6.24 | 0.50 | 2.39 | 3.41 | 720.039 | 102.892 | 159.963 | 1208.393 | 12.78 | 3.82 | -0.359 | 0.060 | 6.58 | 1.85 | 4.46 | 6.82 | 1.70 | 8.02 |
| 1I1 | -0.06 | 0.008 | 5.45 | 0.50 | 2.39 | 2.62 | 780.725 | 114.912 | 160.671 | 1200.047 | 12.17 | 4.50 | -0.362 | 0.102 | 6.58 | 1.85 | 4.46 | 10.15 | 1.70 | 5.85 |
| 1J1 | -0.06 | 0 | 2.83 | 0.50 | 2.39 | 0.00 | 566.562 | 148.468 | 163.555 | 835.525 | 8.11 | 4.15 | -0.372 | 0.097 | 6.58 | 1.85 | 4.46 | 2.03 | 1.00 | 1.00 |
| 2A1 | 0.34 | -0.015 | 4.40 | 0.15 | 2.03 | 1.93 | 627.338 | 105.700 | 155.554 | 983.208 | 9.52 | 2.90 | -0.362 | 0.091 | 6.37 | 1.78 | 3.83 | 5.15 | 1.71 | 4.66 |
| 2B1 | 0.34 | 0.013 | 3.91 | 0.14 | 2.03 | 1.44 | 628.078 | 106.460 | 161.470 | 974.093 | 9.38 | 3.20 | -0.363 | 0.095 | 6.37 | 1.78 | 3.83 | 5.03 | 1.70 | 4.70 |
| 2C1 | 0.34 | -0.032 | 5.59 | 0.15 | 2.03 | 3.12 | 702.381 | 106.772 | 146.134 | 1090.180 | 10.45 | 5.48 | -0.364 | 0.086 | 6.37 | 1.78 | 3.83 | 8.34 | 1.71 | 6.32 |
| 2D | 0.34 | -0.032 | 6.64 | 0.14 | 2.03 | 4.17 | 761.622 | 106.772 | 117.555 | 1184.861 | 11.38 | 5.47 | -0.364 | 0.087 | 6.37 | 1.78 | 3.83 | 10.43 | 1.70 | 7.82 |
| 2E | 0.34 | 0.058 | 4.71 | 0.14 | 2.03 | 2.24 | 644.240 | 102.827 | 143.865 | 1050.181 | 10.64 | 5.13 | -0.367 | 0.090 | 6.37 | 1.78 | 3.83 | 5.37 | 1.70 | 6.26 |
| 2F1 | 0.34 | 0.058 | 6.53 | 0.14 | 2.03 | 4.06 | 745.225 | 102.827 | 134.772 | 1220.352 | 12.49 | 5.27 | -0.363 | 0.088 | 6.37 | 1.78 | 3.83 | 6.66 | 1.70 | 8.80 |
| 2G1 | 0.34 | 0.078 | 5.59 | 0.14 | 2.03 | 3.12 | 691.327 | 102.825 | 151.545 | 1149.016 | 11.15 | 5.85 | -0.370 | 0.084 | 6.37 | 1.78 | 3.83 | 5.47 | 1.70 | 7.61 |
| 2I1 | 0.34 | 0.008 | 5.09 | 0.14 | 2.03 | 2.62 | 759.188 | 114.640 | 169.934 | 1182.821 | 11.72 | 3.84 | -0.365 | 0.090 | 6.37 | 1.78 | 3.83 | 10.15 | 1.70 | 5.85 |
| 2J1 | 0.34 | 0 | 2.47 | 0.14 | 2.03 | 0.00 | 547.048 | 148.778 | 171.570 | 793.653 | 7.66 | 5.99 | -0.377 | 0.089 | 6.37 | 1.78 | 3.83 | 2.03 | 1.00 | 1.00 |
| 3A1 | -0.12 | -0.015 | 4.75 | 0.50 | 2.38 | 1.92 | 655.479 | 106.411 | 157.771 | 1019.079 | 9.97 | 3.61 | -0.356 | 0.101 | 7.43 | 1.96 | 3.15 | 5.15 | 1.71 | 4.66 |
| 3B1 | -0.12 | 0.013 | 4.27 | 0.50 | 2.39 | 1.44 | 649.549 | 106.586 | 170.152 | 1016.847 | 9.83 | 4.03 | -0.355 | 0.101 | 7.43 | 1.96 | 3.15 | 5.03 | 1.70 | 4.70 |
| 3C1 | -0.12 | -0.032 | 5.94 | 0.50 | 2.38 | 3.11 | 732.607 | 106.068 | 152.426 | 1142.734 | 10.90 | 3.63 | -0.358 | 0.097 | 7.43 | 1.96 | 3.15 | 8.34 | 1.71 | 6.32 |
| 3D | -0.12 | -0.032 | 7.00 | 0.50 | 2.39 | 4.17 | 786.873 | 106.580 | 118.829 | 1205.092 | 11.82 | 3.63 | -0.357 | 0.094 | 7.43 | 1.96 | 3.15 | 10.43 | 1.70 | 7.82 |
| 3G1 | -0.12 | 0.078 | 5.95 | 0.50 | 2.39 | 3.12 | 714.954 | 102.535 | 146.326 | 1167.301 | 11.60 | 4.05 | -0.363 | 0.088 | 7.43 | 1.96 | 3.15 | 5.47 | 1.70 | 7.61 |
| 3H1 | -0.12 | 0.078 | 6.24 | 0.50 | 2.39 | 3.41 | 738.560 | 102.314 | 150.870 | 1217.364 | 12.78 | 3.57 | -0.356 | 0.061 | 7.43 | 1.96 | 3.15 | 6.82 | 1.70 | 8.02 |
| 3J1 | -0.12 | 0 | 2.83 | 0.50 | 2.39 | 0.00 | 571.403 | 148.827 | 165.521 | 831.725 | 8.11 | 3.31 | -0.365 | 0.098 | 7.43 | 1.96 | 3.15 | 2.03 | 1.00 | 1.00 |
| 4A1 | 0.145 | -0.015 | 4.70 | 0.45 | 2.33 | 1.92 | 697.719 | 115.598 | 164.406 | 1127.357 | 10.59 | 3.02 | -0.357 | 0.102 | 7.37 | 1.78 | 6.45 | 5.15 | 1.71 | 4.66 |
| 4B1 | 0.145 | 0.013 | 4.22 | 0.45 | 2.34 | 1.44 | 697.933 | 115.411 | 159.359 | 1115.983 | 10.45 | 4.38 | -0.358 | 0.105 | 7.37 | 1.78 | 6.45 | 5.03 | 1.70 | 4.70 |
| 4C1 | 0.145 | -0.032 | 5.89 | 0.45 | 2.33 | 3.11 | 771.976 | 115.602 | 162.731 | 1221.646 | 11.51 | 4.02 | -0.359 | 0.096 | 7.37 | 1.78 | 6.45 | 8.34 | 1.71 | 6.32 |
| 4D | 0.145 | -0.032 | 6.95 | 0.45 | 2.34 | 4.17 | 839.363 | 115.474 | 105.805 | 1277.910 | 12.44 | 3.99 | -0.359 | 0.096 | 7.37 | 1.78 | 6.45 | 10.43 | 1.70 | 7.82 |
| 4E | 0.145 | 0.058 | 5.01 | 0.44 | 2.33 | 2.23 | 726.759 | 112.452 | 185.304 | 1192.032 | 11.70 | 3.80 | -0.359 | 0.098 | 7.37 | 1.78 | 6.45 | 5.37 | 1.70 | 6.26 |
| 4F1 | 0.145 | 0.058 | 6.84 | 0.45 | 2.34 | 4.06 | 830.827 | 112.452 | 178.910 | 1369.584 | 13.56 | 3.84 | -0.361 | 0.102 | 7.37 | 1.78 | 6.45 | 6.66 | 1.70 | 8.80 |
| 4G1 | 0.145 | 0.078 | 5.90 | 0.45 | 2.34 | 3.12 | 760.459 | 111.913 | 194.929 | 1291.469 | 12.22 | 5.32 | -0.364 | 0.088 | 7.37 | 1.78 | 6.45 | 5.47 | 1.70 | 7.61 |
| 4H1 | 0.145 | 0.078 | 6.19 | 0.45 | 2.34 | 3.41 | 783.639 | 112.634 | 169.337 | 1310.430 | 13.39 | 4.04 | -0.359 | 0.061 | 7.37 | 1.78 | 6.45 | 6.82 | 1.70 | 8.02 |
| 4I1 | 0.145 | 0.008 | 5.40 | 0.45 | 2.34 | 2.62 | 833.325 | 123.496 | 157.568 | 1288.638 | 12.79 | 5.21 | -0.362 | 0.103 | 7.37 | 1.78 | 6.45 | 10.15 | 1.70 | 5.85 |
| 4J1 | 0.145 | 0 | 2.78 | 0.45 | 2.34 | 0.00 | 615.253 | 157.137 | 161.373 | 911.206 | 8.73 | 4.60 | -0.371 | 0.098 | 7.37 | 1.78 | 6.45 | 2.03 | 1.00 | 1.00 |

**Table A1-1. (cont'd)**

| Comp | σPh[1] | σN[1] | CLOGP[2] | πPh[3] | πPh2[3] | πN[3] | AREA[4] | PSA[5] | PV[6] | VOL[7] | CMR[8] | DM[9] | HOMO | LUMO | PL1[10] | PB1[10] | PB5[10] | NL1[11] | NB1[11] | NB5[11] |
|------|------|------|-------|------|------|------|---------|---------|---------|----------|-------|------|--------|-------|------|------|------|-------|------|------|
| 5A1 | 0 | -0.015 | 3.61 | -0.64 | 1.24 | 1.92 | 638.413 | 113.356 | 193.822 | 1049.959 | 10.12 | 3.69 | -0.352 | 0.112 | 4.77 | 1.78 | 5.18 | 5.15 | 1.71 | 4.66 |
| 5B1 | 0 | 0.013 | 3.13 | -0.64 | 1.25 | 1.44 | 635.850 | 113.738 | 194.288 | 1036.692 | 9.98 | 4.07 | -0.351 | 0.110 | 4.77 | 1.78 | 5.18 | 5.03 | 1.70 | 4.70 |
| 5C1 | 0 | -0.032 | 4.80 | -0.64 | 1.24 | 3.11 | 712.911 | 113.868 | 176.677 | 1138.920 | 11.05 | 3.04 | -0.355 | 0.112 | 4.77 | 1.78 | 5.18 | 8.34 | 1.71 | 6.32 |
| 5D | 0 | -0.032 | 5.86 | -0.64 | 1.25 | 4.17 | 773.329 | 113.538 | 156.597 | 1234.810 | 11.98 | 3.06 | -0.354 | 0.112 | 4.77 | 1.78 | 5.18 | 10.43 | 1.70 | 7.82 |
| 5E | 0 | 0.058 | 3.92 | -0.65 | 1.24 | 2.23 | 660.567 | 109.586 | 173.154 | 1113.361 | 11.24 | 2.96 | -0.355 | 0.106 | 4.77 | 1.78 | 5.18 | 5.37 | 1.70 | 6.26 |
| 5G1 | 0 | 0.078 | 4.81 | -0.64 | 1.25 | 3.12 | 706.159 | 109.855 | 183.797 | 1204.648 | 11.75 | 4.11 | -0.361 | 0.088 | 4.77 | 1.78 | 5.18 | 5.47 | 1.70 | 7.61 |
| 5H1 | 0 | 0.078 | 5.10 | -0.64 | 1.25 | 3.41 | 722.257 | 109.268 | 172.104 | 1245.609 | 12.93 | 3.69 | -0.351 | 0.062 | 4.77 | 1.78 | 5.18 | 6.82 | 1.70 | 8.02 |
| 5I1 | 0 | 0.008 | 4.31 | -0.64 | 1.25 | 2.62 | 770.949 | 121.659 | 190.050 | 1228.835 | 12.32 | 4.17 | -0.354 | 0.107 | 7.43 | 1.96 | 3.15 | 10.15 | 1.70 | 5.85 |
| 5J1 | 0 | 0 | 1.69 | -0.64 | 1.25 | 0.00 | 556.712 | 155.670 | 181.851 | 845.831 | 8.27 | 3.20 | -0.362 | 0.110 | 4.77 | 1.78 | 5.18 | 2.03 | 1.00 | 1.00 |
| 6A1 | 0.44 | -0.015 | 5.11 | 0.86 | 2.74 | 1.92 | 648.167 | 105.700 | 160.900 | 1025.737 | 10.01 | 3.90 | -0.364 | 0.083 | 7.10 | 1.93 | 4.59 | 5.15 | 1.71 | 4.66 |
| 6B1 | 0.44 | 0.013 | 4.63 | 0.86 | 2.75 | 1.44 | 647.486 | 106.460 | 170.254 | 1015.951 | 9.87 | 3.91 | -0.365 | 0.087 | 7.10 | 1.93 | 4.59 | 5.03 | 1.70 | 4.70 |
| 6C1 | 0.44 | -0.032 | 6.30 | 0.86 | 2.74 | 3.11 | 722.912 | 106.772 | 143.339 | 1136.505 | 10.94 | 6.73 | -0.366 | 0.079 | 7.10 | 1.93 | 4.59 | 8.34 | 1.71 | 6.32 |
| 6D | 0.44 | -0.032 | 7.36 | 0.86 | 2.75 | 4.17 | 781.941 | 106.772 | 121.883 | 1217.625 | 11.87 | 6.74 | -0.365 | 0.079 | 7.10 | 1.93 | 4.59 | 10.43 | 1.70 | 7.82 |
| 6E | 0.44 | 0.058 | 5.42 | 0.85 | 2.74 | 2.23 | 664.441 | 102.827 | 147.097 | 1086.489 | 11.13 | 5.12 | -0.367 | 0.080 | 7.10 | 1.93 | 4.59 | 5.37 | 1.70 | 6.26 |
| 6F1 | 0.44 | 0.058 | 7.25 | 0.86 | 2.75 | 4.06 | 765.726 | 102.827 | 130.091 | 1246.831 | 12.99 | 5.35 | -0.365 | 0.080 | 7.10 | 1.93 | 4.59 | 6.66 | 1.70 | 8.80 |
| 6G1 | 0.44 | 0.078 | 6.30 | 0.85 | 2.74 | 3.11 | 711.462 | 102.825 | 159.458 | 1185.140 | 11.64 | 5.16 | -0.374 | 0.080 | 7.10 | 1.93 | 4.59 | 5.47 | 1.70 | 7.61 |
| 6I1 | 0.44 | 0.008 | 5.80 | 0.85 | 2.74 | 2.61 | 779.260 | 114.640 | 163.019 | 1212.219 | 12.21 | 6.25 | -0.368 | 0.085 | 7.10 | 1.93 | 4.59 | 10.15 | 1.70 | 5.85 |
| 6J1 | 0.44 | 0 | 3.19 | 0.86 | 2.75 | 0.00 | 566.278 | 148.778 | 171.516 | 838.361 | 8.16 | 7.33 | -0.379 | 0.081 | 7.10 | 1.93 | 4.59 | 2.03 | 1.00 | 1.00 |
| 7A1 | 1.08 | -0.015 | 6.02 | 1.77 | 3.65 | 1.92 | 721.521 | 106.231 | 139.487 | 1157.136 | 10.53 | 4.20 | -0.372 | 0.066 | 7.10 | 2.50 | 5.08 | 5.15 | 1.71 | 4.66 |
| 7B1 | 1.08 | 0.013 | 5.54 | 1.77 | 3.66 | 1.44 | 719.401 | 106.238 | 147.715 | 1141.248 | 10.39 | 4.39 | -0.371 | 0.067 | 7.10 | 2.50 | 5.08 | 5.03 | 1.70 | 4.70 |
| 7D | 1.08 | -0.032 | 8.27 | 1.77 | 3.66 | 4.17 | 851.836 | 106.531 | 121.774 | 1349.768 | 12.38 | 5.31 | -0.373 | 0.062 | 7.10 | 2.50 | 5.08 | 10.43 | 1.70 | 7.82 |
| 7E | 1.08 | 0.058 | 6.33 | 1.76 | 3.65 | 2.23 | 738.974 | 102.531 | 134.787 | 1224.007 | 11.64 | 4.99 | -0.374 | 0.061 | 7.10 | 2.50 | 5.08 | 5.37 | 1.70 | 6.26 |
| 7F1 | 1.08 | 0.058 | 8.16 | 1.77 | 3.66 | 4.06 | 841.066 | 103.030 | 157.596 | 1417.819 | 13.50 | 5.10 | -0.377 | 0.068 | 7.10 | 2.50 | 5.08 | 6.66 | 1.70 | 8.80 |
| 7G1 | 1.08 | 0.078 | 7.22 | 1.77 | 3.66 | 3.12 | 784.656 | 102.510 | 148.107 | 1323.325 | 12.16 | 5.29 | -0.383 | 0.064 | 7.10 | 2.50 | 5.08 | 5.47 | 1.70 | 7.61 |
| 7H1 | 1.08 | 0.078 | 7.51 | 1.77 | 3.66 | 3.41 | 793.505 | 102.911 | 152.252 | 1359.324 | 13.33 | 4.10 | -0.369 | 0.057 | 7.10 | 2.50 | 5.08 | 6.82 | 1.70 | 8.02 |
| 7I1 | 1.08 | 0.008 | 6.72 | 1.77 | 3.66 | 2.62 | 859.179 | 115.329 | 146.925 | 1338.394 | 12.73 | 4.93 | -0.377 | 0.068 | 7.10 | 2.50 | 5.08 | 10.15 | 1.70 | 5.85 |
| 7J1 | 1.08 | 0 | 4.10 | 1.77 | 3.66 | 0.00 | 645.803 | 148.933 | 165.081 | 976.571 | 8.67 | 4.18 | -0.385 | 0.063 | 7.10 | 2.50 | 5.08 | 2.03 | 1.00 | 1.00 |
| 8A1 | -0.208 | -0.015 | 6.35 | 2.10 | 3.99 | 1.92 | 742.187 | 115.456 | 83.223 | 1172.695 | 12.17 | 2.77 | -0.351 | 0.107 | 9.78 | 1.78 | 5.85 | 5.15 | 1.71 | 4.66 |
| 8B1 | -0.208 | 0.013 | 5.87 | 2.10 | 3.99 | 1.44 | 752.242 | 118.249 | 172.005 | 1202.720 | 12.03 | 3.04 | -0.350 | 0.108 | 9.78 | 1.78 | 5.85 | 5.03 | 1.70 | 4.70 |
| 8C1 | -0.208 | -0.032 | 7.54 | 2.10 | 3.98 | 3.11 | 830.958 | 118.813 | 204.669 | 1343.243 | 13.10 | 4.06 | -0.353 | 0.104 | 9.78 | 1.78 | 5.85 | 8.34 | 1.71 | 6.32 |
| 8E | -0.208 | 0.058 | 6.66 | 2.09 | 3.98 | 2.23 | 781.580 | 114.735 | 127.490 | 1287.931 | 13.29 | 4.43 | -0.354 | 0.103 | 9.78 | 1.78 | 5.85 | 5.37 | 1.70 | 6.26 |
| 8G1 | -0.208 | 0.078 | 7.55 | 2.10 | 3.99 | 3.12 | 823.508 | 115.016 | 166.532 | 1368.260 | 13.80 | 5.21 | -0.357 | 0.088 | 9.78 | 1.78 | 5.85 | 5.47 | 1.70 | 7.61 |
| 8H1 | -0.208 | 0.078 | 7.84 | 2.10 | 3.99 | 3.41 | 837.961 | 116.160 | 183.067 | 1424.444 | 14.98 | 2.83 | -0.355 | 0.058 | 9.78 | 1.78 | 5.85 | 6.82 | 1.70 | 8.02 |
| 8I1 | -0.208 | 0.008 | 7.05 | 2.10 | 3.99 | 2.62 | 885.516 | 128.038 | 181.519 | 1437.821 | 14.37 | 4.93 | -0.377 | 0.068 | 9.78 | 1.78 | 5.85 | 10.15 | 1.70 | 5.85 |
| 8J1 | -0.131 | 0 | 4.43 | 2.10 | 3.99 | 0.00 | 674.083 | 160.013 | 170.528 | 1048.461 | 10.31 | 4.63 | -0.361 | 0.102 | 9.78 | 1.78 | 5.85 | 2.03 | 1.00 | 1.00 |

[1]Hammett sigma values of substituents on phenyl ring (σPh) and substituents on indole N (σN)

[2] log of the octanol/water partition coefficient, as calculated by the CLOGP program

[3]Hansch π values of substituents on phenyl (πPh), substituted phenyl ring (πPh$_2$) and indole N (πN)

[4]Total surface area

[5]Polar surface area (By default, the surface computation includes all O, N, and S atoms as well as hydrogen atoms covalently bonded to these atoms. The surface is computed as a solvent accessible surface, essentially a van der Waals surface with inflated radii. The surface is affected by the molecule's conformation, but is independent of atomic charges.)

[6]Polar volume (By default, the volume computation includes all O, N, and S atoms as well as hydrogen atoms covalently bonded to these atoms.)

[7]Total volume

[8]Molar refractivity

[9]Dipole moment

[10]Sterimol parameters of substituted phenyl ring

[11]Sterimol parameters of N-substituent

## APPENDIX 2. SYNTHESIS AND CHARACTERIZATION OF CYSMETHYNIL ANALOGUES BY DR SURESH KUMAR GORLA

### A2.1. Synthesis of 1-8 (Series 1)
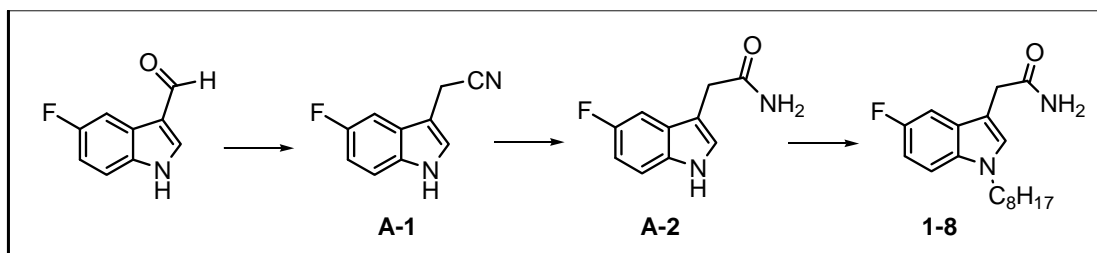


**2-(5-Fluoro-1*H*-indol-3-yl)acetonitrile (A-1).** Nitrile **A-1** was prepared from 5-fluoro-1*H*-indole-3-carbaldehyde following the method described for the synthesis of nitrile **1** (**Section 3.4.2**). Yield: 52 %; $^{1}$H NMR (300 MHz, CDCl$_3$): $\delta$ 3.66 (s, 2H), 6.90 (dt, $J_1$ = 9 Hz, $J_2$ = 2.1 Hz, 1 H), 7.08 (s, 1 H), 7.13-7.22 (m, 2 H), 8.4 (1H, NH); $^{13}$C NMR (75 MHz, CDCl$_3$) $\delta$ 14.2, 102.8, 104.2, 111.0, 112.4, 118.3, 124.8, 126.2, 132.8, 156.3.

**2-(5-Fluoro-1*H*-indol-3-yl)acetamide (A-2).** Amide **A-2** was obtained from nitrile **A-1** following the method described for the synthesis of amide **2** (**Section 3.4.2**). Yield: 61 %; $^{1}$H NMR (300 MHz, CDCl$_3$): $\delta$ 3.69 (s, 2 H), 5.29 (s, 2 H), 6.98 (dt, $J_1$ = 9 Hz, $J_2$ = 2.4 Hz, 1 H), 7.21 (s, 1 H), 7.25-7.33 (m, 2 H), 8.34 (s, 1 H).

**2-(5-Fluoro-1-octyl-1*H*-indol-3-yl)acetamide) (1-8).** Amide **A-2** reacted with bromooctane according to the general procedure described for the synthesis of indole **6** (**Section 3.4.4**). Yield: 70 %; $^{1}$H NMR (300 MHz, CDCl$_3$): $\delta$ 7.26 − 7.18 (m, 2 H), 7.09 (s, 1 H), 6.97 (dt, $J_1$ = 9 Hz, $J_2$ = 2.4 Hz, 1 H), 5.73 (br. s, 2 H), 4.06 (t, $J$ = 7.2 Hz, 2 H), 3.65 (s, 2 H), 1.83 (t, $J$ = 6.3 Hz, 2 H), 1.29-1.25 (m, 10 H), 0.86 (t, $J$ = 6.3 Hz, 3H); $^{13}$C NMR (75 MHz, CDCl$_3$) $\delta$ 174.10, 159.39, 156.27, 139.66, 133.03, 128.67, 127.68, 127.55, 110.66, 117.23, 103.88, 46.63, 32.81, 31.70, 30.17, 29.09 (2C), 26.94, 22.55, 14.02; MS (APCI): *m/z* 305.5 [M + H]$^{+}$

### A2.2. Synthesis of 3-2 and 3-7 (Series 3)
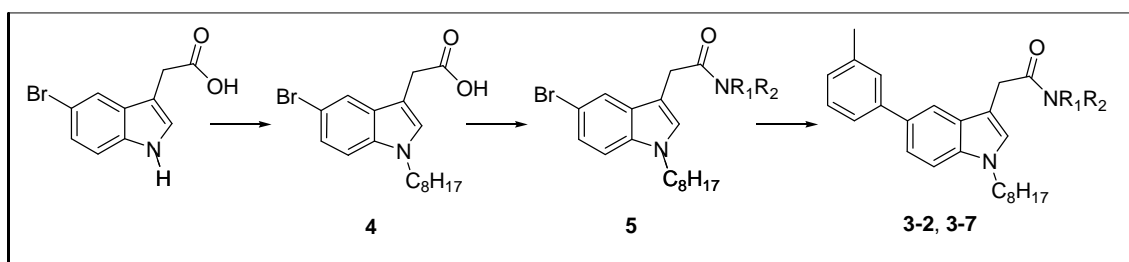


**2-(5-Bromo-1-octyl-1*H*-indol-3-yl)-*N*-methylacetamide (5e).** Acid **4** was reacted with SOCl$_2$ and methylamine as described in the general procedure for the synthesis of amide **5** (**Section 3.4.3**). Yield: 73 %; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 0.87 (t, $J$ = 6.9 Hz, 3 H), 1.26- 1.29 (m, 10 H), 1.82 (t, $J$ = 6.9 Hz, 2 H), 2.74 (d, $J$ = 4.8 Hz, 3 H), 3.66 (s, 2 H), 4.07 (t, $J$ = 7.2 Hz, 2 H), 7.04 (s, 1 H), 7.20- 7.34 ((m, 2 H), 7.67 (s, 1 H); $^{13}$C NMR  (75 MHz, CDCl$_3$) $\delta$ 14.0, 22.5, 26.4, 26.9, 29.1, 29.1, 30.1, 31.6, 32.9, 46.5, 107.0, 111.1, 112.8, 121.4, 124.8, 128.4, 129.2, 135.0, 171.7.
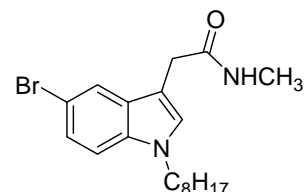
**2-(5-Bromo-1-octyl-1*H*-indol-3-yl)-1-(4-methylpiperazin-1-yl)ethanone (5f).** Acid **4** was reacted with SOCl$_2$ and *N*-methylpiperazine as described in the general procedure for the synthesis of amide **5** (**Section 3.4.3**). Yield: 58 %; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 0.85 (t, $J$ = 6.9 Hz, 3 H), 1.23–1.25 (m, 10 H), 1.76 (t, $J$ = 6.9 Hz, 2 H), 2.20 (t, $J$ = 6.3 Hz, 2 H), 2.23 (s, 3 H), 2.35 (t, $J$ = 4.8 Hz, 2 H), 3.46 (t, $J$ = 4.8 Hz, 2 H), 3.66 (t, $J$ = 4.5 Hz, 2 H), 3.74 (s, 2 H), 4.01 (t, $J$ = 7.2 Hz, 2 H), 7.0 (s, 1 H), 7.15 (d, $J$ = 8.7 Hz, 1 H), 7.25 (d, $J$ = 8.7 Hz, 1 H), 7.70 (s, 1 H); $^{13}$C NMR (75 MHz, CDCl$_3$): $\delta$ 14.0, 22.5, 26.8, 29.10, 29.12, 30.7, 31.7, 41.6, 45.9, 46.4, 54.5, 54.9, 107.2, 110.9, 112.3, 121.3, 124.4, 127.2, 129.1, 134.8, 169.6.
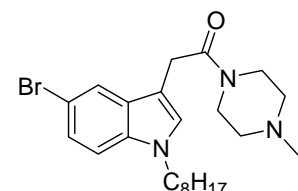
**N-Methyl-2-(1-octyl-5-*m*-tolyl-1*H*-indol-3-yl)acetamide  (3-2).** Amide **5e** was reacted with *m*-tolylboronic acid as described in the general procedure for the

synthesis of amide **3-3** (**Section 3.4.3**). Yield: 68 %; $^{1}$H NMR (300 MHz, CDCl$_3$): $\delta$ 7.72 (s, 1 H), 7.49-7.30 (m, 5 H), 7.12 (d, $J$ = 6.6 Hz, 1 H), 7.01 (s, 1 H), 5.89 (bs, 1 H), 4.06 (t, $J$ = 6.6 Hz, 2 H), 2.68 (d, $J$ = 4.5 Hz, 3 H), 2.41 (s, 3 H), 1.82 (t, $J$ = 6.8 Hz, 2 H), 1.39-1.25 (m, 10 H), 0.86 (t, $J$ = 6.3 Hz, 3 H); $^{13}$C NMR (75 MHz, CDCl$_3$) $\delta$ 174.10, 159.39, 156.27, 139.66, 133.03, 128.67, 127.68, 127.55, 110.66, 110.47, 110.31, 107.52, 107.46, 103.88, 103.57; MS (APCI): $m/z$ 391.8 [M + H]$^{+}$; Found (calcd. for C$_{26}$H$_{34}$N$_2$O) C 79.86 % (79.96); H 8.96 % (8.77).

**1-(4-Methylpiperazin-1-yl)-2-(1-octyl-5-*m*-tolyl-1*H*-indol-3-yl)ethanone**

**(3-7).** Amide **5f** was reacted with *m*-tolylboronic acid as described in the general procedure for the synthesis of amide **3-3** (**Section 3.4.3**). The product is isolated as an amber oil. Yield: 45 %; $^{1}$H NMR (300 MHz, CDCl$_3$): $\delta$ 7.75 (s, 1 H)7.46-7.29 (m, 5 H), 7.12 (d, $J$ = 6.9 Hz, 1 H), 7.02 (s, 1 H), 4.07 (t, $J$ = 6.9 Hz, 2 H), 3.85 (s, 2 H) 3.68 (t, $J$ = 6.9 Hz, 2 H), 3.48 (t, $J$ = 6.9 Hz, 2 H),  2.43 (s, 3 H), 2.36 (t, $J$ = 6.9 Hz, 2 H), 2.23 (s, 3 H), 2.20 (t, $J$ = 6.9 Hz, 2 H), 1.81 (t, $J$ = 6.9 Hz, 2 H), 1.29-1-25 (m, 10 H), 0.87 (t, $J$ = 6.9 Hz, 3 H); $^{13}$C NMR  (75 MHz, CDCl$_3$) $\delta$ 172.54, 142.54, 138.15, 135.60, 133.07, 132.13, 131.69, 128.41, 127.33, 127.06, 126.24, 124.52, 121.39, 114.06, 109.74, 57.2, 53.3, 53.0, 46.3, 46.0, 31.7, 30.2, 29.2 (2C), 27.0, 22.6, 21.6, 14.1; MS (APCI): $m/z$  460.5 [M + H]$^{+}$.

## A2.3.  Synthesis of Series 4 compounds (4-2, 4-4 to 4-10)
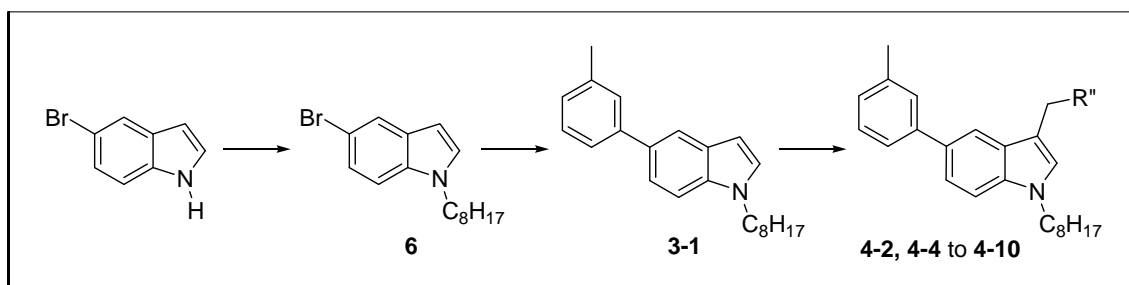
**N,N-dimethyl(1-octyl-5-*m*-tolyl-1*H*-indol-3-yl)methanamine** **(4-2).**

Compound **3-1** was reacted with dimethylamine following the method described for

the amine **4-3** (**Section 3.4.4**). The product is isolated as an amber oil. Yield: 61 %;

^{1}H NMR (300 MHz, CDCl₃): $\delta$ 7.84 (s, 1 H), 7.46-7.42 (m, 3 H), 7.36 (s, 1 H), 7.32

(d, 1 H $J$ = 4.5 Hz), 7.11 (d, 1 H $J$ = 7.5 Hz), 7.07 (s, 1 H), 4.09 (t, 2 H, $J$ = 7.2 Hz),

3.66 (s, 2 H), 2.43 (s, 3 H), 2.29 (s, 6 H), 1.84 (t, 2 H $J$ = 6.6 Hz), 1.31-1.25 (m, 10

H), 0.86 (t, 3 H $J$ = 5.7 Hz). ^{13}C NMR  (75 MHz, CDCl₃): $\delta$ 142.7, 138.0, 135.7,

132.6, 128.8, 128.4, 128.2, 127.9, 126.9, 125.5, 121.2, 117.8, 111.8, 109.4, 54.2, 46.3,

45.2, 31.7, 30.2, 29.1, 29.1, 26.9, 22.5, 21.5, 14.0; MS (APCI): *m/z*  377.3 [M + H][+].

**N-isopropyl-N-((1-octyl-5-*m*-tolyl-1*H*-indol-3-yl)methyl)propan-2-amine**

**(4-4).** Compound **3-1** was reacted with diisopropylamine following the method

described for the amine **4-3** (**Section 3.4.4**). The product is isolated as an amber oil.

Yield:  62 %; ^{1}H NMR (300 MHz, CDCl₃): $\delta$ 8.08 (s, 1 H), 7.46-7.30 (m, 4 H), 7.1 (d,

2 H $J$ = 5.4 Hz), 7.01 (s, 1 H), 4.05 (t, 2 H $J$ = 7.2 Hz), 3.85 (s, 2 H), 3.13 (p 2 H, $J$ =

6.6 Hz), 2.41 (s, 3 H), 1.80 (t, 2 H $J$ = 5.7 Hz), 1.28-1.24 (m, 10 H), 1.07 (d, 12 H, $J$

=6.3 Hz), 0.86 (t, 3 H, $J$ = 6.9 Hz); ^{13}C NMR  (75 MHz, CDCl₃) $\delta$ 142.9, 138.0,

136.1, 131.8, 128.5, 128.3, 128.1, 127.3, 126.7, 124.4, 120.9, 118.5, 115.7, 109.3,

47.3, 46.3, 40.4, 31.7, 30.3, 29.2, 29.2, 27.0, 22.6, 21.6, 20.6, 14.0; MS (APCI): *m/z*

433.0 [M + H][+].

**N-((1-Octyl-5-*m*-tolyl-1*H*-indol-3-yl)methyl)-N-propylpropan-1-amine  (4-**

**5).** Compound **3-1** was reacted with dipropylamine following the method described

for the amine **4-3** (**Section 3.4.4**). The product is isolated as an amber oil. Yield: 65

%; ^{1}H NMR (300 MHz, CDCl₃): $\delta$ 7.90 (s, 1 H), 7.46-7.41 (m, 3 H), 7.33 (d, 2 H, $J$ =

8.4 Hz), 7.11 (d, 1 H, $J$ = 7.5 Hz), 7.02 (s, 1 H), 4.07 (t, 2 H, $J$ = 7.2 Hz), 3.79 (s, 2

H), 2.42 (s, 3 H), 2.42 (t, 2 H, $J$ = 4.2 Hz), 1.82 (t, 2 H, $J$ = 5.7 Hz), 1.54 (q, 4 H, $J$ =

7.5 Hz), 1.28-1.24 (m, 10 H), 0.87 (t, 9 H, $J$ = 7.2 Hz); $^{13}$C NMR (75 MHz, CDCl$_3$) $\delta$ 142.7, 138.0, 135.8,132.2, 128.9, 128.4, 128.1, 127.6, 126.8, 124.4, 121.0, 118.2, 112.6, 109.3, 55.8, 49.2, 46.3, 31.7, 30.2, 29.1, 29.1, 27.0, 22.6, 21.5, 20.3, 14.0, 12.0; MS (APCI): $m/z$ 433.6 [M + H]$^+$.

**N-methyl-N-((1-octyl-5-m-tolyl-1H-indol-3-yl)methyl)propan-2-amine (4-6).** Compound **3-1** was reacted with N-methylpropan-2-amine following the method described for the amine **4-3** (**Section 3.4.4**). The product is isolated as an amber oil. Yield: 52%; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 7.87 (s, 1 H), 7.46-7.41 (m, 3 H), 7.32 (d, 2 H, $J$ = 8.4 Hz), 7.10 (d, 1 H, $J$ = 7.5 Hz), 7.07 (s, 1 H), 4.06 (t, 2 H, $J$ 6.9 Hz), 3.76 (s, 2 H), 2.98 (p, 1 H, $J$ = 6.6 Hz), 2.42 (s, 3 H), 2.23 (s, 3 H), 1.81 (t, 2 H, $J$ = 6.3 Hz), 1.30-1.24 (m, 10 H), 1.11 (d, 6 H, $J$ = 6.6 Hz), 0.86 (t, 3 H, $J$ = 6.9 Hz); $^{13}$C NMR (75 MHz, CDCl$_3$) $\delta$ 142.8, 138.0, 135.8, 132.5, 128.8, 128.5, 128.2, 127.8, 126.9, 124.5, 121.2, 117.9, 112.6, 109.5, 52.7, 48.3, 46.3, 36.8, 31.7, 30.3, 29.2, 29.2, 27.0, 22.6, 21.6, 18.0, 14.1; MS (APCI), $m/z$ = 405.5 [M + H]$^+$. Found (calcd. for C$_{28}$H$_{40}$N$_2$): C 83.12 (83.11); H 9.59 (9.96).

**1-Octyl-3-(pyrrolidin-1-ylmethyl)-5-m-tolyl-1H-indole (4-7).** Compound **3-1** was reacted with pyrrolidine following the method described for the amine **4-3** (**Section 3.4.4**). The product is isolated as an amber oil. Yield: 71 %; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 7.85 (s, 1 H), 7.46-7.41 (m, 3 H), 7.32 (d, 2 H, $J$ = 8.1 Hz), 7.10 (s, 1 H), 7.01 (d, 1 H, $J$ = 7.2 Hz), 4.04 (t, 2 H, $J$ = 7.2 Hz), 3.87 (s, 2 H), 2.61 (br. s, 4 H), 2.42 (s, 3 H), 1.78 (br. s, 6 H), 1.29-1.24 (m, 10 H), 0.86 (t, 3 H, $J$ = 6.3 Hz); $^{13}$C NMR (75 MHz, CDCl$_3$) $\delta$ 142.7, 138.1, 135.6, 132.6, 128.7, 128.5, 128.2, 127.9, 127.0, 124.6, 121.2, 117.7, 112.0, 109.5, 54.0, 50.1, 46.4, 31.8, 30.3, 29.26, 29.22, 27.0, 23.5, 22.6, 21.6, 14.1. MS (APCI): $m/z$ 403.5 [M + H]$^+$.

**1-Octyl-3-(piperidin-1-ylmethyl)-5-*m*-tolyl-1*H*-indole (4-8).** Compound **3-1** was reacted with piperidine following the method described for the amine **4-3** (**Section 3.4.4**). The product is isolated as an amber oil. Yield: 78%; ^{1}H NMR (300 MHz, CDCl$_3$): $\delta$ 7.87 (s, 1 H), 7.46-7.41 (m, 3 H), 7.34-7.32 (m, 2 H), 7.11 (d, 1 H, *J* = 7.2Hz), 7.07 (s, 1 H), 4.06 (t, 2 H, *J* = 6.9 Hz), 3.72 (s, 2 H), 2.45 (s, 3 H), 2.46-2.43 (m, 4 H), 1.84 (t, 2 H, *J* = 6.3 Hz), 1.58-1.55 (m, 4 H), 1.41-1.22 (m, 12 H), 0.86 (t, 3 H, *J* = 6.9 Hz); ^{13}C NMR (75 MHz, CDCl$_3$) $\delta$ 142.8, 138.1, 135.6, 132.5, 129.2, 128.5, 128.28, 128.21, 126.9, 124.5, 121.8, 118.0, 111.2, 109.4, 60.3, 54.3, 53.8, 46.3, 31.7, 30.2, 29.1, 27.0, 26.0, 24.4, 22.6, 21.6, 14.1. MS (APCI): *m/z* 417.5 [M + H]$^+$.

**3-((4-Methylpiperazin-1-yl)methyl)-1-octyl-5-*m*-tolyl-1*H*-indole (4-9).** Compound **3-1** was reacted with 1-methylpiperazine following the method described for the amine **4-3** (**Section 3.4.4**). The product is isolated as an amber oil. Yield: 87 %; ^{1}H NMR (300 MHz, CDCl$_3$): $\delta$ 7.89 (s, 1 H), 7.46-7.41 (m, 3 H), 7.33-7.29 (m, 2 H), 7.05 (s, 1 H), 7.01 (d, 1 H, *J* = 7.2 Hz), 4.04 (t, 2 H, *J* = 6.9 Hz), 3.73 (s, 2 H), 2.54-2.47 (m, 8 H), 2.42 (s, 3 H), 2.25 (s, 3 H), 1.80 (t, 2 H, *J* = 5.7 Hz), 1.28-1.24 (m, 10 H), 0.86 (t, 3 H, *J* = 5.4 Hz); ^{13}C NMR (75 MHz, CDCl$_3$) $\delta$ 142.7, 138.1, 135.8, 132.6, 129.0, 128.5, 128.2, 128.0, 126.9, 124.5, 121.3, 118.1, 111.2, 109.5, 55.2, 53.3, 53.0, 46.3, 46.0, 31.7, 30.2, 29.2(2C), 27.0, 22.6, 21.6, 14.1; MS (APCI): *m/z* 432.6 [M + H]$^+$.

**4-((1-Octyl-5-*m*-tolyl-1*H*-indol-3-yl)methyl)morpholine (4-10).** Compound **3-1** was reacted with morpholine following the method described for the amine **4-3** (**Section 3.4.4**). The product is isolated as an amber oil. Yield: 78 %; ^{1}H NMR (300 MHz, CDCl$_3$): $\delta$ 7.90 (s, 1 H), 7.46 (m, 3 H), 7.33-7.28 (m, 2 H), 7.09 (d, 1 H, *J* = 7.5 Hz), 7.01 (s, 1 H), 4.01 (t, 2 H, *J* = 6.9 Hz), 3.69 (s, 2 H), 3.68 (t, 4 H, *J* = 4.8 Hz), 2.48 (t, 4 H, *J* = 4.8 Hz), 2.41 (s, 3 H), 1.78 (t, 2 H, *J* = 6 Hz), 1.27-1.21 (m, 10 H),

0.86 (t, 3 H, $J$ = 6.9 Hz); $^{13}$C NMR (75 MHz, CDCl$_3$) $\delta$ 142.8, 138.1, 135.9, 132.7, 129.0, 128.6, 128.3, 128.1, 127.1, 124.6, 121.4, 118.2, 110.9, 109.6, 67.2, 54.0, 53.7, 46.4, 31.8, 30.3, 29.2, 29.2, 27.0, 22.7, 21.7, 14.2. MS (APCI): $m/z$ 419.3 [M + H]$^+$.

## A2.4. Synthesis of the 5-1 and 5-6 (Series 5)
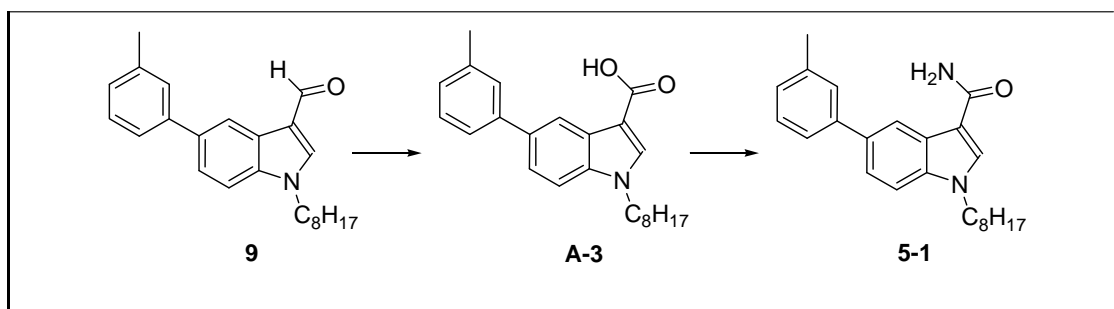


**1-Octyl-5-m-tolyl-1$H$-indole-3-carboxylic acid (A-3).** Compound **A-3** was synthesized according to the method described in Andreani et al[108]. Compound **9** (0.573 mmol, 1 equiv) was dissolved in 10 mL of acetone and treated with a solution of potassium permanaganate (KMnO$_4$, 1.5 mmol) in water (5 mL). The reaction mixture was stirred at room temperature for 5h, treated with 10 % H$_2$O$_2$ until the pink color of KMnO$_4$ was not observed and then filtered. The solvent was removed under reduced pressure and the resulting residue was filtered again (if necessary), acidified with 2M HCl and extracted with Et$_2$O. The organic solvent was removed in vacuo and the residue was purified by column chromatography with hexane/EtOAc as eluting solvents. Yield: 62 %; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 8.84 (s, 1 H), 7.92 (s, 1 H), 7.54-7.48 (m, 2 H), 7.41-7.29 (m, 3 H), 7.14 (d, $J$ = 6.6 Hz, 1 H), 4.12 (t, $J$ = 6.6 Hz, 2 H), 2.44 (s, 3 H), 1.87 (t, $J$ = 6.6 Hz, 3 H), 1.37-1.25 (m, 10 H), 0.86 (t, $J$ = 6.6 Hz, 3 H); $^{13}$C NMR (75 MHz, CDCl$_3$): $\delta$ 170.3, 142.3, 141.9, 138.4, 138.2, 136.1, 136.0, 135.8, 128.5, 128.3, 127.4, 124.6, 120.3, 110.3, 106.3, 47.2, 31.7, 31.1, 29.8, 29.1, 22.6, 21.5, 21.5, 14.0.

**1-Octyl-5-*m*-tolyl-1*H*-indole-3-carboxamide (5-1).** A mixture of **A-3** (0.4 mmol, 1 equiv.) and $SOCl_2$ (1 mL, 13 equiv.) in dry benzene (5 mL) were refluxed for 4 h. Excess $SOCl_2$ and benzene were removed by distillation under reduced pressure and the residue containing the acid chloride was dried in vacuo. It was dissolved in dry THF (4 ml) and ammonia gas was bubbled into the stirred solution for 30 min. THF was removed under vacuum, the residue was extracted with $CH_2Cl_2$, dried (anhydrous $Na_2SO_4$) and the solvent removed under reduced pressure to give carboxamide **5-1.** Yield: 74 %; $^1$H NMR (300 MHz, $CDCl_3$): $\delta$ 8.16 (s, 1 H), 7.74 (s, 1 H), 7.49-7.13 (m, 6 H), 5.99 (bs, 2 H), 4.09 (t, $J$ = 7.2 Hz), 2.42 (s, 3 H); 1.82 (t, $J$ = 6.8 Hz, 2 H), 1.28-1.24 (m, 10 H) 0.86 (t, $J$ = 6.3 Hz, 3 H); $^{13}$C NMR (75 MHz, $CDCl_3$) $\delta$ 167.32, 141.92, 138.25, 136.07, 135.31, 132.61, 128.58, 128.29, 127.47, 126.18, 124.60, 122.37, 118.94, 110.43, 110.00, 47.02, 31.71, 29.92, 29.11, 29.08, 26.86, 22.57, 21.53, 14.0; MS (APCI): *m/z* 363.5 $[M + H]^+$.
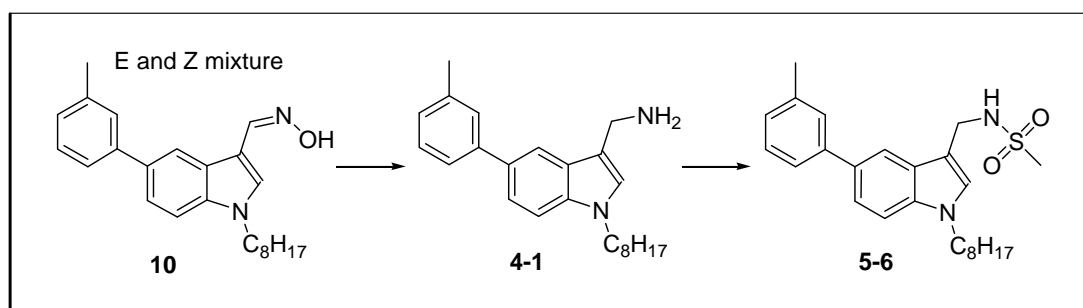


***N*-((1-Octyl-5-*m*-tolyl-1*H*-indol-3-yl)methyl)methanesulfonamide (5-6).** MsCl (0.434 mmol, 1 equiv.) was added to a solution of amine **4-1** (0.434 mmol, 1 equiv.) and triethylamine (0.651 mmol, 1.5 equiv.) in THF (4 mL) at 0 $^o$C. The reaction mixture was stirred at room temperature for 1 h. The precipitate was removed by filtration, the filtrate was concentrated under reduced pressure and the residue purified by column chromatography with $CH_2Cl_2$ as eluting solvent to give sulfonamide **5-6** as an amber oil. Yield: 61 %; $^1$H NMR (300 MHz, $CDCl_3$): $\delta$ 7.19 (s,

1 H), 7.50-7.31 (m, 7 H), 7.15 (s, 1 H), 4.54 (d, 2 H, $J$ = 4.8 Hz), 4.10 (t, 2 H, $J$ = 7.2

Hz), 2.85 (s, 3 H), 2.44 (s, 3 H), 1.84 (t, 2 H, $J$ = 6.3 Hz), 1.31-1.25 (m, 10 H), 0.87

(t, 3 H, $J$ = 7.2 Hz); $^{13}$C NMR (75 MHz, CDCl$_3$) $\delta$ 142.1, 138.2, 135.9, 133.3, 128.6,

128.1, 127.8, 127.26, 127.22, 124.4, 121.9, 117.1, 110.1, 109.9, 60.4, 46.5, 40.8, 38.8,

31.7, 30.2, 29.1, 26.9, 22.6, 21.5, 14.0; MS (APCI): $m/z$ 427.3 [M + H]$^+$.

## A2.5. Synthesis of the 6-1 to 6-10 (Series 6)
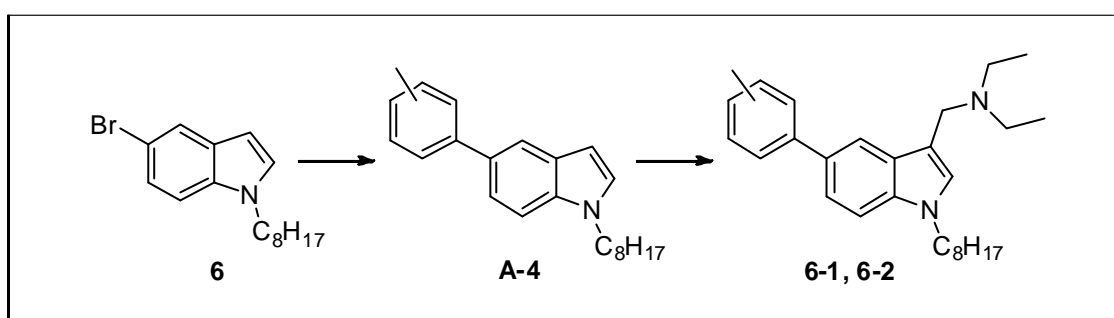


**1-Octyl-5-$o$-tolyl-1$H$-indole (A-4a).** Compound **6** was

reacted with $o$-tolylboronic acid as described for **3-3** (**Section**

**3.4.3**). Yield: 68 %; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 7.53 (d, $J$

= 6.9 Hz, 1 H), 7.35-7.11 (m, 7 H), 6.49 (d, $J$ = 2.7 Hz, 1 H),

4.11 (t, $J$ = 6.3 Hz, 2 H), 2.30 (s, 3 H), 1.85 (t, $J$ = 6.9 Hz, 2 H), 1.31-1.25 (m, 10

H), 0.86 (t, $J$ = 6.9 Hz, 3 H); $^{13}$C NMR (75 MHz, CDCl$_3$) $\delta$ 143.1, 135.7, 134.9, 133.0,

130.3, 130.1, 128.3, 128.2, 126.5, 125.5, 123.0, 121.2, 108.7, 100.9, 46.5, 31.7, 30.2,

29.1, 27.0, 22.6, 20.7, 14.0.
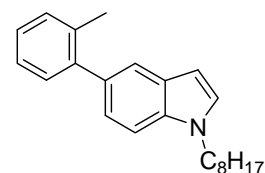
**1-Octyl-5-$p$-tolyl-1$H$-indole (A-4b).** Compound

**A-4b** was reacted with $p$-tolylboronic acid as described for

**3-3** (**Section 3.4.3**). Yield: 72 %; $^1$H NMR (300 MHz,

CDCl$_3$): $\delta$ 0.87 (t, $J$ = 6.3 Hz, 3 H), 1.26-1.30 (m, 10 H), 1.83 (t, $J$ = 6.3 Hz, 2 H),

2.39 (s, 3 H), 4.11 (t, $J$ = 7.2 Hz, 2 H), 6.52 (d, $J$ = 5.7 Hz, 1 H), 7.10 (d, $J$ = 2.7 Hz, 1
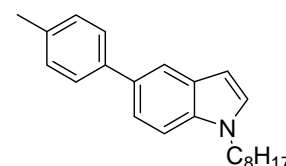
H), 7.24 (d, $J$ = 7.8 Hz, 2 H), 7.36-7.50 (m, 2 H), 7.55 (d, $J$ = 7.8 Hz, 2 H), 7.81 (s, 1

H); $^{13}$C NMR (75 MHz, CDCl$_3$) $\delta$ 14.0, 20.7, 22.6, 27.0, 29.1(2C), 30.2, 31.7, 46.5,

101.1, 109.5, 119.1, 121.1, 126.9, 127.1, 128.4, 128.6, 129.3, 132.6, 135.3, 135.7.

**$N$-ethyl-$N$-((1-octyl-5-$o$-tolyl-1$H$-indol-3-yl)methyl)ethanamine (6-1).**

Compound **A-4a** was reacted with diethylamine as described for the synthesis of **4-3**

(**Section 3.4.4**). The product is isolated as an amber oil. Yield: 72 %; $^1$H NMR (300

MHz, CDCl$_3$): $\delta$ 7.63 (s, 1 H,), 7.33-7.22 (m, 5 H, ), 7.16 (d, 1 H, $J$ = 8.4 Hz), 7.08 (s,

1 H), 4.09 (t, 2 H, $J$ = 7.2 Hz), 7.39 (s, 2 H), 2.56 (q, 4 H, $J$ = 7.2 Hz), 2.32 (s, 3 H),

1.85 (t, 2 H, $J$ = 6.9 Hz), 1.32-1.26 (m, 10 H), 1.08 (t, 6 H, $J$ = 7.2 Hz), 0.87 (t, 3 H, $J$

= 4.8 Hz); $^{13}$C NMR  (75 MHz, CDCl$_3$) $\delta$ 143.2, 135.7, 135.2, 132.6, 130.3, 130.1,

128.4, 127.7, 126.5, 125.5, 122.9, 119.9,111.5, 108.7, 47.7, 46.5, 46.3, 31.7, 30.2,

29.1, 29.1, 27.0, 22.6, 20.7, 14.0, 11.9; MS (APCI): $m/z$ 405.6 [M + H]$^+$. Found

(calcd. for C$_{28}$H$_{48}$N$_2$): C 83.01 (83.11), H 9.21 (9.96).

**$N$-Ethyl-$N$-((1-octyl-5-$p$-tolyl-1$H$-indol-3-yl)methyl)ethanamine (6-2).**

Compound **A-4b** was reacted with diethylamine as described for the synthesis of

amine **4-3** (**Section 3.4.4**). The product is isolated as an amber oil. Yield: 65 %; $^1$H

NMR (300 MHz, CDCl$_3$): $\delta$ 7.87 (s, 1 H), 7.55 (d, 2 H, $J$ = 7.8 Hz), 7.42 (d, 2H, $J$ =

8.1 Hz), 7.33 (d, 1 H, $J$ = 8.4 Hz), 7.24 (d, 2 H, $J$ = 7.8 Hz), 7.06 (s, 1 H), 4.08 (t, 2 H,

$J$ = 6.9 Hz), 3.81 (s, 2 H), 2.54 (q, 4 H, $J$ = 6.9 Hz), 2.39 (s, 2 H), 1.83 (t, 2 H, $J$ = 6.6

Hz), 1.31-1.29 (m, 10 H), 1.10 (t, 6 H, $J$ = 7.2 Hz), 0.86 (t, 3 H, $J$ = 6.9 Hz).  $^{13}$C

NMR  (75 MHz, CDCl$_3$) $\delta$ 139.9, 135.7, 135.5 132.2, 129.3, 129.0, 127.8, 127.2,

121.0, 117.7, 111.8, 109.4, 47.7, 46.8, 46.3, 31.7, 30.2, 29.1, 29.1,  26.9, 22.5, 21.0,

14.0, 11.9. MS (APCI): $m/z$ 405.3 [M + H]$^+$.
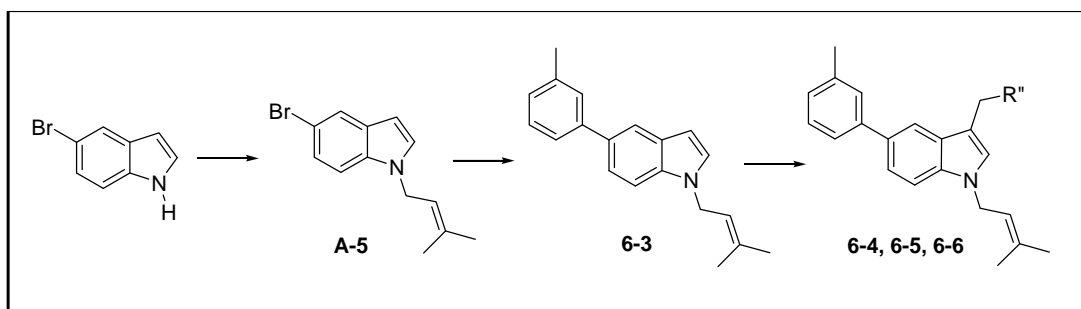
**5-Bromo-1-(3-methylbut-2-enyl)-1*H*-indole (A-5).** Indole **A-5** was prepared from 5-bromo-1*H*-indole and 1-chloro-3-methylbut-2-ene according to the method described for the synthesis of indole **6** (**Section 3.4.4**). Yield: 96 %; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 7.32 (d, J = 1.5 Hz, 1H), 7.23 (d, J = 1.5 Hz), 7.23 (d, J = 1.5 Hz, 1 H), 7.16 (s, 1 H), 7.06 (d, J = 3Hz, 1H), 6.39 (d, J = Hz, 1H), 5.32 (t, J = 6.9 Hz, 1 H), 4.61 (d, J = 6.9 Hz, 2 H), 1.78 (s, 3H), 1.74 (s, 3H); $^{13}$C NMR (75 MHz, CDCl$_3$) $\delta$ 136.7, 134.6, 130.4, 128.5, 124.1, 123.3, 119.5, 112.5, 110.9, 100.5, 44.3, 25.6, 18.0.

**1-(3-Methylbut-2-enyl)-5-*m*-tolyl-1*H*-indole (6-3).** Indole **A-5** was reacted with *m*-tolylboronic acid as described for the synthesis of amide **3-3** (**Section 3.4.4**). Yield: 75 %; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 7.87 (d, *J* = 0.9 Hz, 1 H), 7.45-7.28 (m, 5 H), 7.11-7.09 (m, 2 H), 6.57 (dd, *J$_1$* = 3.0 Hz, *J$_2$* = 0.6 Hz, 1 H), 5.46-5.43 (m, 1 H), 4.76 (d, *J* = 6.9 Hz, 2 H), 2.47 (s, 3 H), 1.89 (s, 3 H), 1.82 (s, 3 H); $^{13}$C NMR (75 MHz, CDCl$_3$) $\delta$ 142.70, 138.14, 136.34, 135.55, 132.98, 129.25, 128.57, 128.26, 128.06, 127.01, 124.54, 121.31, 119.99, 119.47, 109.71, 101.39, 44.30, 25.70, 21.66, 18.08. MS (APCI): *m/z* 276.5 [M + H]$^+$.

***N*-Ethyl-*N*-((1-(3-methylbut-2-enyl)-5-*m*-tolyl-1*H*-indol-3-yl)methyl)ethanamine (6-4).** Indole **6-3** was reacted with diethylamine, as described for the synthesis of amine **4-3** (**Section 3.4.4**). The product is isolated as an amber oil. Yield: 70%; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 7.88 (s, 1 H), 7.46-7.41 (m, 3 H), 7.34-7.29 (m, 2 H), 7.11 (d, 1 H, *J* = 7.2 Hz), 7.06 (s, 1 H), 5.38 (t, 1 H, *J* = 6.9 Hz), 4.67

(d, 2 H, $J$ = 6.6 Hz), 3.80 (s, 2 H,), 2.57 (q, 4 H, $J$ = 6.9 Hz), 2.43 (s, 3 H), 1.82 (s, 3 H), 1.76 (s, 3 H), 1.09 (t, 6 H, $J$ = 7.2 Hz). ^{13}C NMR (75 MHz, CDCl$_3$) $\delta$ 142.7, 138.1, 136.1, 135.6, 132.6, 129.2, 128.5, 128.2, 127.7, 126.9, 124.5, 121.2, 119.9, 117.8, 111.6, 109.6, 47.7, 46.5, 44.1, 25.6, 21.6, 18.0, 11.7; MS (APCI): $m/z$ 361.1 [M + H]$^+$; Found (calcd. for C$_{25}$H$_{32}$N$_2$) C 82.92% (83.28) H 8.85 (8.95).

**1-(3-Methylbut-2-enyl)-3-(piperidin-1-ylmethyl)-5-$m$-tolyl-1$H$-indole (6-5).** Compound **6-3** was reacted with piperidine as described for the synthesis of amine **4-3** (**Section 3.4.4**). The product is isolated as an amber oil. Yield: 71 %; ^{1}H NMR (300 MHz, CDCl$_3$): $\delta$ 7.86 (s, 1 H), 7.46-7.42 (m, 3 H), 7.34 (d, 2 H, $J$ = 7.2 Hz), 7.12 (d, 1 H, $J$ = 10.2 Hz), 7.03 (s, 1 H), 5.39 (t, 1 H, $J$ = 5.1 Hz), 4.68 (d, 2 H, $J$ = 6.6 Hz), 3.73 (s, 2 H), 2.49 (br. s, 4 H), 2.44 (s, 3 H), 1.83 (s, 3 H), 1.77 (s, 3 H), 1.42 (m, 2 H), 0.86 (m, 4 H). ^{13}C NMR (75 MHz, CDCl$_3$) $\delta$ 142.8, 138.1, 136.1, 135.6, 132.6, 129.4, 128.5, 128.2, 127.8, 126.9, 124.5, 121.1, 120.0, 118.0, 111.4, 109.5, 54.3, 53.8, 44.1, 26.0, 25.6, 24.4, 21.6, 18.0. MS (APCI): $m/z$ 373.3 [M + H]$^+$: Found (calcd. for C$_{26}$H$_{32}$N$_2$): C 83.62% (83.82) H 8.85 (8.66).

**4-((1-(3-methylbut-2-enyl)-5-$m$-tolyl-1$H$-indol-3-yl)methyl)morpholine (6-6).** The product is isolated as an amber oil. Compound **6-3** was reacted with morpholine as described for the synthesis of amine **4-3** (**Section 3.4.4**). Yield: 68 %; ^{1}H NMR (300 MHz, CDCl$_3$): $\delta$ 7.76 (s, 1 H), 7.44-7.40 (m, 3 H), 7.34-7.30 (m, 3 H), 7.10 (d, $J$ = 7.5 Hz, 1 H), 5.35 (t, $J$ = 6.6 Hz, 1 H), 4.67 ($J$ = 6.6 Hz, 2 H), 3.93 (s, 2 H), 2.68 (br. s, 2 H), 2.41 (s, 3 H), 1.80 (s, 3 H), 1.74 (s, 3 H), 1.71 (t, $J$ = 8.7 Hz, 4 H); ^{13}C NMR (75 MHz, CDCl$_3$) $\delta$ 142.45, 138.22, 136.60, 135.54, 133.33, 129.73, 129.26, 128.58, 128.22, 127.17, 124.53, 121.56, 119.62, 117.27, 110.27, 110.02, 100.61, 53.32, 44.41, 25.66, 24.69, 23.40, 21.59, 18.09. MS (APCI): $m/z$ 375.4 [M + H]$^+$; Found (calcd. for C$_{25}$H$_{30}$N$_2$O): C 79.92% (80.17); H 8.10 (8.07).
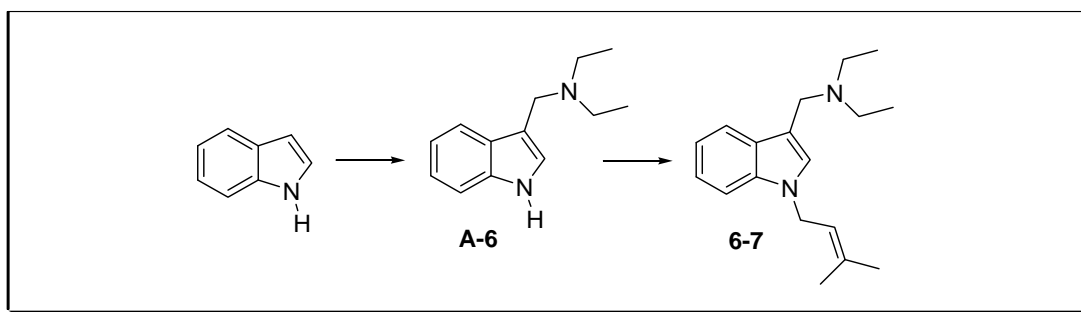
**N-((1H-Indol-3-yl)methyl)-N-ethylethanamine (A-6).**. Compound **A-6** was reacted with diethylamine as described for the synthesis of amine **4-3** (**Section 3.4.4**). Yield: 64 %; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 1.07 (t, $J$ = 6.0 Hz, 6 H), 2.55 (q, $J$ = 6.0 Hz, 4 H), 3.77 (s, 2 H), 6.95 (s, 1 H), 7.08- 7.16 (m, 2 H), 7.23 (d, $J$ = 6.9 Hz, 1 H), 7.67 (d, $J$ = 7.8 Hz, 1 H); $^{13}$C NMR (75 MHz, CDCl$_3$) $\delta$ 11.5, 46.4, 50.0, 111.2, 112.1, 119.0, 119.2, 121.6, 124.2, 128.1, 136.1.

**N-Ethyl-N-((1-(3-methylbut-2-enyl)-1H-indol-3-yl)methyl)ethanamine (6-7).** Amine **6-7** was prepared from indole **A-6** and 1-chloro-3-methylbut-2-ene according to the method described for the synthesis of indole **6** (**Section 3.4.4**). The product is isolated as an amber oil. Yield: 62 %; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 7.69 (d, $J$ = 7.5 Hz, 1 H), 7.28 (d, $J$ = 8.1 Hz, 1 H), 7.18 (t, $J$ = 6.9 Hz, 1 H), 7.10 (d, $J$ = 7.2 Hz, 1 H), 7.04 (s, 1 H), 5.36 (t, $J$ = 6.6 Hz, 1 H), 4.64 (d, $J$ = 6.6 Hz, 2 H), 3.77 (s, 2 H), 2.56 (q, $J$ = 6 Hz, 4 H), 1.80 (s, 3 H), 1.74 (s, 3 H), 1.08 (t, $J$ = 7.2 Hz, 6 H); $^{13}$C NMR (75 MHz, CDCl$_3$) $\delta$ 136.1, 135.9, 128.8, 126.8, 121.2, 120.1, 119.4, 118.8, 111.6, 109.3, 47.8, 46.5, 44.0, 40.9, 25.6, 18.0, 11.8; MS (APCI): $m/z$ 271.5 [M + H]$^+$.

**2-(5-Fluoro-1-(3-methylbut-2-enyl)-1H-indol-3-yl)acetamide (6-8).** Amide **6-8** was prepared from indole **A-6** and 1-chloro-3-methylbut-2-ene according to the method described for the synthesis of indole **6** (**Section 3.4.4**). The product is isolated as an amber oil. Yield: 67 %; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 7.26-7.23 (m, 2 H), 7.19

(s, 1 H), 6.98 (t, $J$ = 9Hz, 1 H), 5.72 (br s, 2 H), 5.36 (t, $J$ = 6 Hz, 1 H), 4.65 (d, $J$ = 6.6 Hz, 2 H), 3.65 (s, 2 H), 1.82 (s, 3 H), (1.79 s, 3 H); $^{13}$C NMR (75 MHz, CDCl$_3$) $\delta$ 174.16, 137.04, 133.04, 128.29, 119.27, 110.62, 110.45, 110.27, 107.54, 103.86, 103.55, 44.30, 32.85, 25.63, 18.00. MS (APCI): $m/z$ 261.3 [M + H]$^+$.
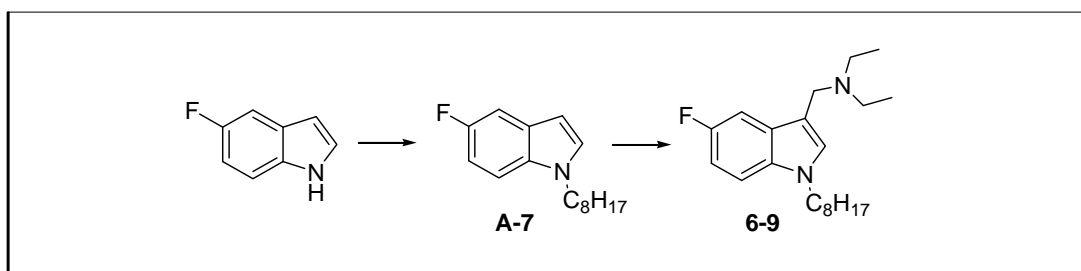


**5-Fluoro-1-octyl-1$H$-indole (A-7).** Indole **A-7** was prepared from 5-fluoro-1$H$-indole and 1-bromooctane according to the method described for the synthesis of indole **6** (**Section 3.4.4**). Yield: 92%; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 7.27-7.19 (m, 2 H), 7.10 (d, $J$ = 3 Hz, 1 H), 6.93 (dt, $J$ = 8.4 Hz, $J_2$ = 2.1 Hz), 6.41 (d, $J$ = 2.7 Hz, 1 H), 4.06 (t, $J$ = 6.9 Hz, 2 H), 1.79 (t, $J$ = 6.6 Hz, 2 H), 1.28-1.24 (m, 10 H), 0.86 (t, $J$ = 6 Hz, 3 H); $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 156.1, 132.6, 129.3, 109.8, 109.4, 105.6, 105.3, 100.7, 46.6, 31.7, 30.2, 29.1, 26.9, 22.6, 14.0.

***N*-Ethyl-*N*-[(5-fluoro-1-octyl-1*H*-indol-3-yl)methyl]ethanamine (6-9).** Amine **6-9** was reacted with diethylamine as described for the synthesis of amine **4-3** (**Section 3.4.4**). The product is isolated as an amber oil. Yield: 73 %; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 7.35 (dd, 1 H, $J_1$ = 10.5 Hz, J$_2$ = 0.9 Hz), 7.18 (dd 1 H, $J_1$= 9 Hz, J$_2$ = 4.2 Hz), 7.05 (s, 1 H), 6.91 (dt, 1 H, $J_1$ = 9Hz, $J_2$ = 2.4 Hz), 4.03 (t, 2 H, $J$ = 6.9 Hz), 3.70 (s, 2 H), 2.53 (q, 4 H, $J$ = 6.9 Hz), 1.79 (t, 2 H, $J$ = 6.9 Hz), 1.28-1.24 (m, 10 H), 1.07 (t, 6 H, $J$ = 7.2 Hz), 0.86 (t, 3 H, $J$ = 6.3 Hz). $^{13}$C NMR (75 MHz, CDCl$_3$) $\delta$ 159.0, 155.9, 132.8, 128.6, 111.9, 109.7, 109.3, 104.6, 47.9, 46.5, 46.4, 31.7, 30.1, 29.1, 29.1, 26.9, 22.5, 14.0, 11.8; MS (APCI): $m/z$ 333.0 [M + H]$^+$; Found (calcd. for C$_{21}$H$_{33}$FN$_2$): C 75.68% (75.86) H 9.85% (10.00).
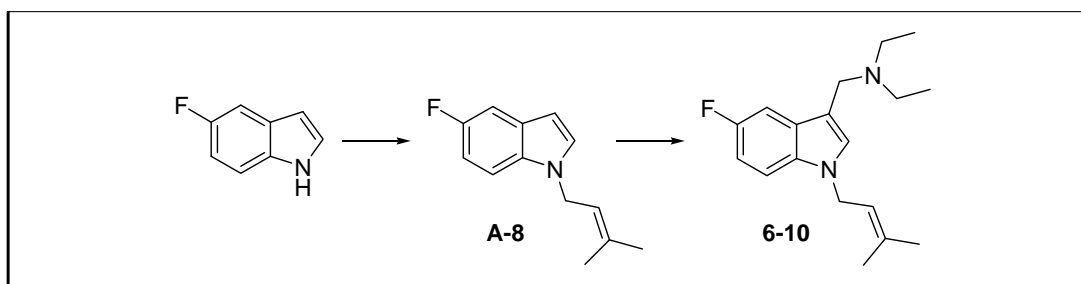
**5-Fluoro-1-(3-methylbut-2-enyl)-1*H*-indole (A-8).** Indole **A-8** was prepared

from 5-fluoro-1*H*-indole and 1-chloro-3-methylbut-2-ene according to the method

described for the synthesis of indole **6** (**Section 3.4.4**). Yield: 89%; $^1$H NMR (300

MHz, CDCl$_3$): $\delta$ 7.29-7.20 (m, 2 H), 7.13 (d, $J = 3.3$ Hz, 1 H), 6.94 (dt, $J_1 = 9.3$ Hz, $J_2$

$= 2.7$ Hz, 1 H), 6.44 (d, $J = 3.0$ Hz), 5.37 (t, $J = 6.6$ Hz, 1 H), 4.67 (d, $J = 6.6$ Hz, 2

H), 1.82 (s, 3 H), 1.77 (s, 3 H); $^{13}$C NMR (75 MHz, CDCl$_3$) $\delta$ 156.2, 136.5, 132.6,

128.9, 119.6, 110.0, 109.4, 105.6, 105.3, 100.7, 44.3, 25.6, 17.9.

*N*-Ethyl-*N*-((5-fluoro-1-(3-methylbut-2-enyl)-1*H*-indol-3-

**yl)methyl)ethanamine (6-10).** Indole **A-7** was reacted with diethylamine as described

for the synthesis of amine **4-3** (**Section 3.4.4**). The product is isolated as an amber oil.

Yield: 77 %; $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 7.34 (dd, 1 H, $J_1 = 9.6$ Hz, $J_2 = 2.1$ Hz),

7.17 (d, 1 H, $J = 3.6$ Hz), 7.09 (s, 1 H), 6.91 (dd, 1 H, $J_1 = 8.7$ Hz, $J_2 = 4.2$Hz), 5.34 (t,

1 H, $J = 6.6$ Hz), 4.64 (d, 2 H, $J = 6.9$ Hz),3.84 (s, 2 H), 2.66 (q, 4 H, $J = 7.2$ Hz),

1.81 (s, 3 H), 1.76 (s, 3 H), 1.16 (t, 6 H, $J = 6.9$ Hz); $^{13}$C NMR (75 MHz, CDCl$_3$) $\delta$

159.1, 156.0, 136.3, 132.8, 128.4, 119.7, 111.6, 111.5, 110.0, 109.8, 109.7, 109.4,

104.5, 104.2, 47.9, 46.4, 44.3, 25.6, 17.9, 11.7. MS (APCI): *m/z* 289.0 [M + H]$^+$.

**APPENDIX 3. HPLC PURITY DETERMINATION OF ALL SYNTHESIZED COMPOUNDS.**

| Comp | Mobile Phase A | | | Mobile Phase B | | |
|------|----------------|---|---|----------------|---|---|
| | Composition[a] | RT (min)[c] | Area (%)[d] | Composition[b] | RT (min)[c] | Area (%)[d] |
| **1-1** | A1 | 5.3 | 98.5 | B1 | 3.7 | 98.1 |
| **1-2** | A1 | 5.0 | 96.8 | B1 | 3.7 | 96.5 |
| **1-3** | A1 | 5.4 | 96.5 | B1 | 3.8 | 95.4 |
| **1-4** | A1 | 4.2 | 96.1 | B1 | 3.1 | 95.5 |
| **1-5** | A1 | 4.9 | 95.5 | B1 | 3.5 | 96.9 |
| **1-6** | A1 | 4.7 | 96.2 | B1 | 3.3 | 96.3 |
| **1-7** | A1 | 2.9 | 95.4 | B1 | 2.5 | 97.2 |
| **1-8** | A2 | 2.0 | 93.5 | B2 | 2.1 | 93.3 |
| **2-1** | A1 | 1.8 | 97.7 | B1 | 1.6 | 99.6 |
| **2-2** | A1 | 2.5 | 95.4 | B1 | 2.1 | 95.4 |
| **2-3** | A1 | 2.8 | 98.0 | B1 | 2.2 | 98.1 |
| **2-4** | A1 | 5.0 | 98.1 | B1 | 3.4 | 98.4 |
| **3-1** | A1 | 16.0 | 98.8 | B1 | 13.5 | 98.4 |
| **3-2** | Combustion Analysis[e] | | | Combustion Analysis[e] | | |
| **3-3** | A1 | 7.4 | 97.4 | B1 | 5.4 | 95.7 |
| **3-4** | A1 | 9.9 | 96.7 | B1 | 7.1 | 97.5 |
| **3-5** | A1 | 9.0 | 98.2 | B1 | 6.3 | 97.6 |
| **3-6** | A1 | 10.8 | 95.4 | B1 | 7.8 | 96.7 |
| **3-7** | A2 | 3.5 | 92.1 | B2 | 4.0 | 92.8 |
| **4-1** | A3 | 1.9 | 96.9 | B3 | 1.5 | 97.8 |
| **4-2** | A4 | 10.9 | 98.0 | B2 | 14.9 | 98.1 |
| **4-3** | A3 | 2.1 | 98.3 | B3 | 1.8 | 98.1 |
| **4-4** | A4 | 7.0 | 97.0 | B2 | 9.8 | 96.9 |
| **4-5** | A4 | 8.1 | 97.3 | B2 | 10.3 | 96.8 |
| **4-6** | Combustion Analysis[e] | | | Combustion Analysis[e] | | |
| **4-7** | A4 | 8.2 | 96.4 | B2 | 10.0 | 95.5 |
| **4-8** | A4 | 6.1 | 97.2 | B2 | 8.8 | 96.7 |
| **4-9** | A4 | 7.8 | 96.2 | B2 | 10.7 | 95.2 |
| **4-10** | A4 | 6.8 | 93.5 | B2 | 9.8 | 92.0 |
| **5-1** | A5 | 3.6 | 99.0 | B6 | 3.4 | 99.0 |
| **5-2** | A1 | 6.0 | 97.5 | B1 | 4.1 | 96.9 |
| **5-3** | A1 | 11.5 | 96.5 | B1 | 7.6 | 97.0 |
| **5-4** | A1 | 13.8 | 98.1 | B1 | 9.3 | 97.5 |
| **5-5** | A1 | 5.9 | 97.6 | B1 | 4.2 | 980 |
| **5-6** | A4 | 2.7 | 98.3 | B4 | 4.0 | 99.0 |
| **6-1** | A4 | 7.3 | 97.0 | B2 | 9.9 | 96.6 |
| **6-2** | A4 | 8.0 | 95.4 | B2 | 10.5 | 95.9 |
| **6-3** | A4 | 3.7 | 96.0 | B5 | 5.8 | 95.3 |
| **6-4** | Combustion Analysis[e] | | | Combustion Analysis[e] | | |
| **6-5** | Combustion Analysis[e] | | | Combustion Analysis[e] | | |
| **6-6** | Combustion Analysis[e] | | | Combustion Analysis[e] | | |
| **6-7** | Combustion Analysis[e] | | | Combustion Analysis[e] | | |
| **6-8** | A2 | 1.8 | 93.2 | B2 | 2.0 | 93.0 |
| **6-9** | Combustion Analysis[e] | | | Combustion Analysis[e] | | |
| **6-10** | A4 | 2.2 | 92.8 | B2 | 3.4 | 93.2 |
| **6-11** | A1 | 7.0 | 95.6 | B1 | 5.3 | 96.4 |
| **6-12** | A1 | 8.5 | 95.6 | B1 | 6.5 | 95.3 |

[a] Composition of Mobile Phase A: Methanol and Water
A1: 90% methanol
A2: 95% methanol + 0.1% triethylamine (TEA)
A3: 90% methanol + 0.1% TEA
A4: 92% methanol + 0.1% TEA
A5: 100 % methanol

[b] Composition of Mobile Phase B: Acetonitrile and Water
B1: 90% acetonitrile
B2: 95% acetonitrile + 0.1% triethylamine (TEA)
B3: 90% acetonitrile + 0.1% TEA
B4: 95% acetonitrile
B5: 92% acetonitrile + 0.1% TEA
B6: 100% acetonitrile

[c] Retention time of Major Peak in chromatogram. Chromatogram was run for at least 15 min for the detection of the major peak

[d] Area (%) of Major Peak = [Area of Major Peak / Total Area of All Peaks] x 100

[e] Combustion analysis for carbon were within 0.4 % of calculated values

**APPENDIX 4. QSAR OF SYNTHESIZED ANALOGUES**

**Table A4-1. 2D Descriptors of Synthesized Compounds (non-protonated)**

| Comp | mr | logP(o/w) | SlogP | vdw_area | vdw_vol | TPSA | logS | Weight | lip_druglike | lip_violation |
|------|------|-----------|-------|----------|---------|-------|--------|---------|--------------|---------------|
| 1-1 | 11.774 | 7.190 | 6.271 | 410.237 | 575.378 | 48.02 | -7.888 | 376.544 | 1 | 1 |
| 1-2 | 11.777 | 7.151 | 6.271 | 410.237 | 575.378 | 48.02 | -7.888 | 376.544 | 1 | 1 |
| 1-3 | 11.774 | 7.153 | 6.271 | 410.237 | 575.378 | 48.02 | -7.888 | 376.544 | 1 | 1 |
| 1-4 | 11.991 | 6.848 | 5.972 | 425.240 | 586.040 | 57.25 | -7.464 | 392.543 | 1 | 1 |
| 1-5 | 12.451 | 7.189 | 6.362 | 442.472 | 610.469 | 57.25 | -7.791 | 406.570 | 1 | 1 |
| 1-6 | 11.333 | 6.855 | 5.963 | 393.004 | 550.949 | 48.02 | -7.414 | 362.517 | 1 | 1 |
| 1-7 | 8.847 | 4.858 | 4.296 | 322.506 | 435.975 | 48.02 | -4.987 | 286.419 | 1 | 0 |
| 1-8 | 8.927 | 5.048 | 4.435 | 326.915 | 439.118 | 48.02 | -5.282 | 304.409 | 1 | 0 |
| 2-1 | 8.055 | 3.635 | 3.171 | 262.333 | 378.329 | 58.88 | -4.889 | 264.328 | 1 | 0 |
| 2-2 | 11.605 | 6.785 | 6.289 | 390.570 | 553.204 | 48.02 | -7.607 | 422.450 | 1 | 1 |
| 2-3 | 10.308 | 4.839 | 4.877 | 357.117 | 496.189 | 48.02 | -5.940 | 332.447 | 1 | 0 |
| 2-4 | 12.569 | 5.703 | 6.604 | 441.854 | 612.433 | 48.02 | -7.794 | 400.566 | 1 | 1 |
| 3-1 | 10.415 | 8.162 | 7.244 | 356.911 | 510.670 | 4.93 | -7.574 | 319.492 | 1 | 1 |
| 3-2 | 12.265 | 7.546 | 6.532 | 431.138 | 601.423 | 34.03 | -7.782 | 390.571 | 1 | 1 |
| 3-3 | 12.704 | 7.743 | 6.874 | 458.417 | 627.468 | 25.24 | -7.676 | 404.598 | 1 | 1 |
| 3-4 | 13.640 | 8.425 | 7.654 | 492.881 | 676.327 | 25.24 | -8.330 | 432.652 | 1 | 1 |
| 3-5 | 13.445 | 8.275 | 7.408 | 464.258 | 662.431 | 25.24 | -8.228 | 430.636 | 1 | 1 |
| 3-6 | 13.909 | 8.717 | 7.799 | 481.490 | 686.860 | 25.24 | -8.430 | 444.663 | 1 | 1 |
| 3-7 | 14.207 | 7.268 | 6.560 | 509.275 | 704.763 | 28.48 | -7.723 | 459.678 | 1 | 1 |
| 4-1 | 11.171 | 7.415 | 6.969 | 388.211 | 549.770 | 30.95 | -7.239 | 348.534 | 1 | 1 |
| 4-2 | 12.087 | 8.208 | 7.572 | 436.392 | 601.860 | 8.17 | -7.446 | 376.588 | 1 | 1 |
| 4-3 | 13.021 | 8.890 | 8.352 | 470.856 | 650.719 | 8.17 | -8.100 | 404.642 | 1 | 1 |
| 4-4 | 13.881 | 9.814 | 9.129 | 505.320 | 699.577 | 8.17 | -8.754 | 432.696 | 1 | 1 |
| 4-5 | 13.945 | 10.118 | 9.132 | 505.320 | 699.577 | 8.17 | -8.504 | 432.696 | 1 | 1 |
| 4-6 | 12.984 | 9.011 | 8.350 | 470.856 | 650.719 | 8.17 | -8.100 | 404.642 | 1 | 1 |
| 4-7 | 12.827 | 8.740 | 8.106 | 442.232 | 636.823 | 8.17 | -7.998 | 402.626 | 1 | 1 |
| 4-8 | 13.290 | 9.182 | 8.496 | 459.464 | 661.252 | 8.17 | -8.200 | 416.653 | 1 | 1 |
| 4-9 | 13.586 | 7.733 | 7.257 | 487.249 | 679.155 | 11.41 | -7.492 | 431.668 | 1 | 1 |
| 4-10 | 13.072 | 7.774 | 7.342 | 457.236 | 647.484 | 17.40 | -7.737 | 418.625 | 1 | 1 |
| 5-1 | 11.311 | 7.102 | 6.343 | 393.004 | 550.949 | 48.02 | -7.826 | 362.517 | 1 | 1 |
| 5-2 | 12.238 | 7.278 | 6.661 | 427.469 | 599.808 | 48.02 | -7.770 | 390.571 | 1 | 1 |
| 5-3 | 12.151 | 8.188 | 6.959 | 428.404 | 595.798 | 31.23 | -8.017 | 391.555 | 1 | 1 |
| 5-4 | 12.609 | 8.529 | 7.349 | 445.636 | 620.227 | 31.23 | -8.345 | 405.582 | 1 | 1 |
| 5-5 | 12.256 | 7.584 | 7.146 | 431.138 | 601.423 | 34.03 | -7.728 | 390.571 | 1 | 1 |
| 5-6 | 12.667 | 7.388 | 6.559 | 451.113 | 614.467 | 51.10 | -7.356 | 426.625 | 1 | 1 |
| 6-1 | 13.025 | 8.851 | 8.352 | 470.856 | 650.719 | 8.17 | -8.100 | 404.642 | 1 | 1 |
| 6-2 | 13.021 | 8.853 | 8.352 | 470.856 | 650.719 | 8.17 | -8.100 | 404.642 | 1 | 1 |
| 6-3 | 8.957 | 5.811 | 5.849 | 303.791 | 431.481 | 4.93 | -5.626 | 275.395 | 1 | 1 |
| 6-4 | 11.567 | 6.539 | 6.958 | 417.736 | 571.529 | 8.17 | -6.152 | 360.545 | 1 | 1 |
| 6-5 | 11.834 | 6.831 | 7.102 | 406.345 | 582.063 | 8.17 | -6.252 | 372.556 | 1 | 1 |
| 6-6 | 11.611 | 5.423 | 5.948 | 404.116 | 568.295 | 17.40 | -5.789 | 374.528 | 1 | 1 |
| 6-7 | 8.646 | 4.207 | 4.982 | 330.006 | 432.126 | 8.17 | -3.252 | 270.420 | 1 | 0 |
| 6-8 | 7.452 | 2.697 | 3.041 | 273.796 | 359.928 | 48.02 | -3.334 | 260.312 | 1 | 0 |
| 6-9 | 10.183 | 6.748 | 6.516 | 387.535 | 514.458 | 8.17 | -5.495 | 332.507 | 1 | 1 |
| 6-10 | 8.725 | 4.397 | 5.121 | 334.415 | 435.269 | 8.17 | -3.547 | 288.410 | 1 | 1 |
| 6-11 | 13.862 | 8.083 | 7.355 | 507.885 | 686.988 | 34.47 | -7.907 | 448.651 | 1 | 1 |
| 6-12 | 14.321 | 8.424 | 7.745 | 525.117 | 711.417 | 34.47 | -8.234 | 462.678 | 1 | 1 |

**Table A4-1 (cont'd)**

| Comp | b_count | b_rotN | b_1rotN | b_rotR | b_1rotR | a_acc | A_don | a_base | vsa_acc | vsa_don | vsa_hyd | vsa_pol |
|------|---------|--------|---------|--------|---------|-------|-------|--------|---------|---------|---------|---------|
| 1-1 | 62 | 10 | 10 | 0.333 | 0.333 | 1 | 1 | 0 | 13.567 | 17.742 | 338.477 | 31.309 |
| 1-2 | 62 | 10 | 10 | 0.333 | 0.333 | 1 | 1 | 0 | 13.567 | 17.742 | 338.477 | 31.309 |
| 1-3 | 62 | 10 | 10 | 0.333 | 0.333 | 1 | 1 | 0 | 13.567 | 17.742 | 338.477 | 31.309 |
| 1-4 | 63 | 11 | 11 | 0.355 | 0.355 | 2 | 1 | 0 | 16.071 | 17.742 | 350.977 | 33.813 |
| 1-5 | 66 | 12 | 12 | 0.375 | 0.375 | 2 | 1 | 0 | 16.071 | 17.742 | 366.786 | 33.813 |
| 1-6 | 59 | 10 | 10 | 0.345 | 0.345 | 1 | 1 | 0 | 13.567 | 17.742 | 322.669 | 31.309 |
| 1-7 | 48 | 9 | 9 | 0.409 | 0.409 | 1 | 1 | 0 | 13.567 | 17.742 | 255.018 | 31.309 |
| 1-8 | 48 | 9 | 9 | 0.391 | 0.391 | 1 | 1 | 0 | 13.567 | 17.742 | 260.139 | 31.309 |
| 2-1 | 38 | 3 | 3 | 0.136 | 0.136 | 1 | 2 | 0 | 13.567 | 23.425 | 193.918 | 36.992 |
| 2-2 | 55 | 6 | 5 | 0.176 | 0.147 | 1 | 1 | 0 | 13.567 | 17.742 | 326.640 | 31.309 |
| 2-3 | 51 | 5 | 5 | 0.185 | 0.185 | 1 | 1 | 0 | 13.567 | 17.742 | 291.052 | 31.309 |
| 2-4 | 64 | 8 | 8 | 0.250 | 0.250 | 1 | 1 | 0 | 13.567 | 17.742 | 370.094 | 31.309 |
| 3-1 | 55 | 8 | 8 | 0.308 | 0.308 | 0 | 0 | 0 | 0.000 | 0.000 | 336.269 | 0.000 |
| 3-2 | 65 | 11 | 10 | 0.355 | 0.323 | 1 | 1 | 0 | 13.567 | 5.683 | 372.377 | 19.249 |
| 3-3 | 68 | 11 | 11 | 0.344 | 0.344 | 1 | 0 | 0 | 13.567 | 0.000 | 406.276 | 13.567 |
| 3-4 | 74 | 13 | 13 | 0.382 | 0.382 | 1 | 0 | 0 | 13.567 | 0.000 | 437.893 | 13.567 |
| 3-5 | 73 | 11 | 11 | 0.314 | 0.314 | 1 | 0 | 0 | 13.567 | 0.000 | 410.693 | 13.567 |
| 3-6 | 76 | 11 | 11 | 0.306 | 0.306 | 1 | 0 | 0 | 13.567 | 0.000 | 426.502 | 13.567 |
| 3-7 | 78 | 11 | 11 | 0.297 | 0.297 | 2 | 0 | 0 | 13.567 | 0.000 | 453.575 | 13.567 |
| 4-1 | 60 | 9 | 9 | 0.321 | 0.321 | 1 | 1 | 0 | 0.000 | 0.000 | 342.968 | 17.742 |
| 4-2 | 66 | 10 | 10 | 0.333 | 0.333 | 1 | 0 | 0 | 0.000 | 0.000 | 410.767 | 0.000 |
| 4-3 | 72 | 12 | 12 | 0.375 | 0.375 | 1 | 0 | 0 | 0.000 | 0.000 | 442.384 | 0.000 |
| 4-4 | 78 | 12 | 12 | 0.353 | 0.353 | 1 | 0 | 0 | 0.000 | 0.000 | 474.001 | 0.000 |
| 4-5 | 78 | 14 | 14 | 0.412 | 0.412 | 1 | 0 | 0 | 0.000 | 0.000 | 474.001 | 0.000 |
| 4-6 | 72 | 11 | 11 | 0.344 | 0.344 | 1 | 0 | 0 | 0.000 | 0.000 | 442.384 | 0.000 |
| 4-7 | 71 | 10 | 10 | 0.303 | 0.303 | 1 | 0 | 0 | 0.000 | 0.000 | 415.184 | 0.000 |
| 4-8 | 74 | 10 | 10 | 0.294 | 0.294 | 1 | 0 | 0 | 0.000 | 0.000 | 430.993 | 0.000 |
| 4-9 | 76 | 10 | 10 | 0.286 | 0.286 | 2 | 0 | 0 | 0.000 | 0.000 | 458.066 | 0.000 |
| 4-10 | 72 | 10 | 10 | 0.294 | 0.294 | 2 | 0 | 0 | 2.504 | 0.000 | 427.684 | 2.504 |
| 5-1 | 59 | 9 | 9 | 0.310 | 0.310 | 1 | 1 | 0 | 13.567 | 17.742 | 322.669 | 31.309 |
| 5-2 | 65 | 11 | 11 | 0.355 | 0.355 | 1 | 1 | 0 | 13.567 | 17.742 | 354.286 | 31.309 |
| 5-3 | 64 | 11 | 10 | 0.355 | 0.323 | 1 | 0 | 0 | 13.567 | 0.000 | 374.136 | 13.567 |
| 5-4 | 67 | 12 | 11 | 0.375 | 0.344 | 1 | 0 | 0 | 13.567 | 0.000 | 389.944 | 13.567 |
| 5-5 | 65 | 11 | 10 | 0.355 | 0.323 | 1 | 1 | 0 | 13.567 | 5.683 | 372.377 | 19.249 |
| 5-6 | 66 | 11 | 11 | 0.344 | 0.344 | 2 | 1 | 0 | 32.017 | 8.075 | 374.353 | 40.092 |
| 6-1 | 72 | 12 | 12 | 0.375 | 0.375 | 1 | 0 | 0 | 0.000 | 0.000 | 442.384 | 0.000 |
| 6-2 | 72 | 12 | 12 | 0.375 | 0.375 | 1 | 0 | 0 | 0.000 | 0.000 | 442.384 | 0.000 |
| 6-3 | 44 | 3 | 3 | 0.130 | 0.130 | 0 | 0 | 0 | 0.000 | 0.000 | 288.843 | 0.000 |
| 6-4 | 61 | 7 | 7 | 0.241 | 0.241 | 1 | 0 | 0 | 0.000 | 0.000 | 394.959 | 0.000 |
| 6-5 | 63 | 5 | 5 | 0.161 | 0.161 | 1 | 0 | 0 | 0.000 | 0.000 | 383.567 | 0.000 |
| 6-6 | 61 | 5 | 5 | 0.161 | 0.161 | 2 | 0 | 0 | 2.504 | 0.000 | 380.259 | 2.504 |
| 6-7 | 47 | 6 | 6 | 0.286 | 0.286 | 1 | 0 | 0 | 0.000 | 0.000 | 311.499 | 0.000 |
| 6-8 | 37 | 4 | 4 | 0.200 | 0.200 | 1 | 1 | 0 | 13.567 | 17.742 | 212.713 | 31.309 |
| 6-9 | 58 | 11 | 11 | 0.440 | 0.440 | 1 | 0 | 0 | 0.000 | 0.000 | 364.045 | 0.000 |
| 6-10 | 47 | 6 | 6 | 0.273 | 0.273 | 1 | 0 | 0 | 0.000 | 0.000 | 316.620 | 0.000 |
| 6-11 | 75 | 14 | 14 | 0.400 | 0.400 | 2 | 0 | 0 | 16.071 | 0.000 | 450.393 | 16.071 |
| 6-12 | 78 | 15 | 15 | 0.417 | 0.417 | 2 | 0 | 0 | 16.071 | 0.000 | 466.201 | 16.071 |

**Table A4-2. 3D Descriptors for Synthesized Compounds (non-protonated)**

| Comp | PM3_E | PM3_Eele | PM3_HOMO | PM3_LUMO | PM3_IP | PM3_HF | PM3_dipole | VSA | vol |
|---|---|---|---|---|---|---|---|---|---|
| 1-1 | -93967.1 | -802835 | -8.401 | -0.132 | 8.401 | -27.163 | 5.744 | 466.566 | 413.375 |
| 1-2 | -93965.3 | -808437 | -8.463 | -0.107 | 8.463 | -25.382 | 6.105 | 467.341 | 414.625 |
| 1-3 | -93966.5 | -796794 | -8.301 | -0.069 | 8.301 | -26.600 | 5.614 | 466.244 | 414.875 |
| 1-4 | -100727.3 | -856201 | -8.426 | -0.150 | 8.426 | -55.325 | 6.919 | 474.798 | 421.000 |
| 1-5 | -104176.2 | -899924 | -8.403 | -0.122 | 8.403 | -61.285 | 5.277 | 497.788 | 438.875 |
| 1-6 | -90514.4 | -753086 | -8.394 | -0.139 | 8.394 | -17.570 | 6.007 | 445.458 | 397.750 |
| 1-7 | -72709.3 | -541570 | -8.428 | -0.036 | 8.428 | -42.255 | 5.947 | 368.404 | 315.250 |
| 1-8 | -82507.6 | -591475 | -8.618 | -0.318 | 8.618 | -77.748 | 7.445 | 370.609 | 315.750 |
| 2-1 | -66382.5 | -467564 | -8.528 | -0.175 | 8.528 | 14.831 | 4.942 | 303.260 | 269.625 |
| 2-2 | -120493.0 | -946854 | -8.495 | -0.812 | 8.495 | -96.640 | 4.209 | 436.559 | 393.125 |
| 2-3 | -82892.3 | -673906 | -8.284 | -0.033 | 8.284 | 11.263 | 6.039 | 397.244 | 356.500 |
| 2-4 | -99404.8 | -878781 | -8.303 | -0.038 | 8.303 | 30.541 | 6.275 | 491.387 | 438.125 |
| 3-1 | -76909.7 | -610810 | -8.270 | 0.073 | 8.270 | 24.992 | 2.411 | 411.543 | 364.875 |
| 3-2 | -97411.9 | -856358 | -8.390 | -0.124 | 8.390 | -28.084 | 5.431 | 490.140 | 431.125 |
| 3-3 | -100854.1 | -915406 | -8.287 | -0.022 | 8.287 | -28.180 | 5.107 | 504.560 | 451.000 |
| 3-4 | -107751.4 | -1022892 | -8.227 | 0.045 | 8.227 | -39.319 | 5.131 | 542.640 | 486.375 |
| 3-5 | -107038.1 | -1003048 | -8.251 | 0.020 | 8.251 | -33.207 | 5.340 | 526.654 | 475.625 |
| 3-6 | -110485.6 | -1072204 | -8.237 | 0.030 | 8.237 | -37.623 | 5.156 | 538.294 | 490.875 |
| 3-7 | -114575.9 | -1105719 | -8.256 | 0.029 | 8.256 | -26.657 | 3.212 | 562.900 | 506.375 |
| 4-1 | -84455.3 | -709901 | -8.201 | 0.090 | 8.201 | 46.321 | 3.378 | 449.144 | 393.125 |
| 4-2 | -91345.1 | -813941 | -8.171 | 0.109 | 8.171 | 17.070 | 2.654 | 486.505 | 431.250 |
| 4-3 | -98240.9 | -927323 | -8.171 | 0.113 | 8.171 | 3.215 | 2.787 | 524.443 | 467.750 |
| 4-4 | -105140.8 | -1049912 | -8.133 | 0.130 | 8.133 | -10.517 | 2.575 | 560.674 | 501.500 |
| 4-5 | -105139.7 | -1035346 | -8.154 | 0.117 | 8.154 | -9.181 | 2.933 | 565.826 | 502.000 |
| 4-6 | -98241.8 | -925836 | -8.139 | 0.128 | 8.139 | 2.576 | 2.496 | 521.719 | 465.750 |
| 4-7 | -97528.7 | -902286 | -8.153 | 0.116 | 8.153 | 8.277 | 2.979 | 507.478 | 455.500 |
| 4-8 | -100977.8 | -959677 | -8.161 | 0.115 | 8.161 | 2.252 | 2.725 | 525.390 | 472.500 |
| 4-9 | -105067.1 | -1014552 | -8.168 | 0.105 | 8.168 | 14.353 | 3.547 | 543.130 | 487.000 |
| 4-10 | -104292.3 | -966304 | -8.225 | 0.049 | 8.225 | -13.196 | 2.971 | 516.625 | 462.750 |
| 5-1 | -90517.8 | -748233 | -8.438 | -0.195 | 8.438 | -21.097 | 5.434 | 447.115 | 396.000 |
| 5-2 | -97412.8 | -846752 | -8.273 | -0.012 | 8.273 | -29.991 | 4.936 | 483.839 | 429.500 |
| 5-3 | -100083.7 | -854378 | -8.295 | -0.005 | 8.295 | -70.176 | 2.386 | 485.247 | 426.875 |
| 5-4 | -103532.1 | -904322 | -8.274 | 0.002 | 8.274 | -75.172 | 2.305 | 506.663 | 446.750 |
| 5-5 | -97413.1 | -839602 | -8.418 | -0.134 | 8.418 | -29.794 | 5.081 | 490.718 | 433.250 |
| 5-6 | -105733.7 | -916523 | -8.448 | -0.288 | 8.448 | -53.096 | 4.835 | 501.902 | 444.750 |
| 6-1 | -98240.9 | -936451 | -8.183 | 0.180 | 8.183 | 3.271 | 3.173 | 521.365 | 464.750 |
| 6-2 | -98243.3 | -923728 | -8.094 | 0.109 | 8.094 | 0.916 | 2.651 | 522.359 | 469.000 |
| 6-3 | -65835.5 | -497690 | -8.216 | 0.114 | 8.216 | 55.608 | 2.711 | 341.106 | 309.000 |
| 6-4 | -87168.8 | -797941 | -8.105 | 0.166 | 8.105 | 40.958 | 2.556 | 455.391 | 405.625 |
| 6-5 | -89902.1 | -830230 | -8.119 | 0.162 | 8.119 | 41.590 | 2.935 | 457.985 | 415.625 |
| 6-6 | -93217.8 | -829402 | -8.242 | 0.072 | 8.242 | 14.767 | 3.209 | 449.350 | 405.250 |
| 6-7 | -65911.1 | -528620 | -8.114 | 0.280 | 8.114 | 23.805 | 2.498 | 355.527 | 311.250 |
| 6-8 | -71433.4 | -476757 | -8.540 | -0.253 | 8.540 | -47.797 | 7.679 | 298.185 | 260.875 |
| 6-9 | -86782.1 | -703565 | -8.406 | -0.074 | 8.406 | -56.118 | 4.369 | 428.180 | 372.500 |
| 6-10 | -75710.0 | -580928 | -8.314 | -0.001 | 8.314 | -20.392 | 4.194 | 360.944 | 316.000 |
| 6-11 | -114512.2 | -1072357 | -8.309 | -0.059 | 8.309 | -68.195 | 3.823 | 558.791 | 494.875 |
| 6-12 | -117959.7 | -1142324 | -8.290 | -0.025 | 8.290 | -72.595 | 6.370 | 575.456 | 510.125 |

**Table A4-2 (cont'd)**

| Comp | E | E_ele | E_sol | E_str | E_strain | E_tor | E_vdw | ASA_H | ASA_P | dipole |
|------|------|-------|-------|-------|----------|-------|-------|-------|-------|--------|
| **1-1** | 43.566 | -4.315 | -13.525 | 5.896 | 0.011 | -5.624 | 44.092 | 622.977 | 121.307 | 1.218 |
| **1-2** | 45.897 | -4.042 | -13.901 | 6.022 | 0.011 | -3.030 | 43.289 | 622.756 | 123.193 | 1.276 |
| **1-3** | 44.157 | -3.666 | -15.197 | 5.977 | 0.350 | -6.214 | 44.309 | 633.692 | 125.401 | 1.173 |
| **1-4** | 50.162 | -8.041 | -14.157 | 5.743 | 0.021 | -4.910 | 47.609 | 610.340 | 161.182 | 1.623 |
| **1-5** | 50.112 | -7.510 | -12.690 | 5.898 | 0.024 | -7.624 | 47.552 | 653.792 | 143.977 | 1.013 |
| **1-6** | 41.930 | -3.426 | -15.528 | 5.824 | 0.324 | -6.157 | 42.195 | 605.441 | 124.591 | 1.238 |
| **1-7** | 18.116 | -3.931 | -14.014 | 3.240 | 0.008 | -10.042 | 25.679 | 470.366 | 122.647 | 1.277 |
| **1-8** | 15.202 | -6.547 | -12.565 | 3.103 | 0.033 | -10.137 | 25.522 | 471.807 | 123.243 | 1.664 |
| **2-1** | 43.865 | -4.433 | -16.053 | 4.805 | 0.006 | 8.142 | 33.133 | 367.345 | 153.500 | 1.127 |
| **2-2** | 81.016 | 8.487 | -14.714 | 7.349 | 0.083 | 8.458 | 52.672 | 475.295 | 222.117 | 1.124 |
| **2-3** | 55.776 | 1.767 | -15.471 | 5.821 | 0.013 | 1.566 | 41.853 | 506.115 | 141.792 | 1.159 |
| **2-4** | 64.185 | 5.249 | -15.517 | 6.974 | 0.028 | -4.570 | 49.594 | 634.276 | 152.122 | 1.140 |
| **3-1** | 41.546 | 1.098 | -5.162 | 5.325 | 0.007 | -8.723 | 41.583 | 652.854 | 24.301 | 0.462 |
| **3-2** | 43.610 | -1.977 | -12.804 | 6.083 | 0.029 | -9.479 | 45.892 | 693.507 | 95.245 | 1.238 |
| **3-3** | 55.086 | -1.458 | -10.796 | 6.562 | 0.011 | -8.452 | 51.707 | 701.565 | 90.740 | 1.255 |
| **3-4** | 51.845 | -0.608 | -10.500 | 7.315 | 0.013 | -14.286 | 52.945 | 805.667 | 51.638 | 1.227 |
| **3-5** | 55.985 | -0.307 | -10.867 | 6.780 | 0.017 | -9.988 | 49.081 | 771.090 | 64.038 | 1.241 |
| **3-6** | 52.907 | -1.311 | -10.440 | 7.284 | 0.033 | -16.558 | 56.280 | 798.704 | 54.494 | 1.257 |
| **3-7** | 80.662 | 17.616 | -10.346 | 7.656 | 2.367 | -13.409 | 59.318 | 794.142 | 101.258 | 0.953 |
| **4-1** | 39.390 | -1.847 | -7.418 | 5.512 | 0.082 | -9.763 | 43.185 | 647.422 | 87.045 | 0.650 |
| **4-2** | 51.755 | 0.297 | -3.652 | 6.048 | 0.028 | -8.590 | 49.688 | 717.162 | 59.866 | 0.589 |
| **4-3** | 48.823 | 0.314 | -1.933 | 6.481 | 0.027 | -16.340 | 52.428 | 782.858 | 30.968 | 0.558 |
| **4-4** | 54.540 | 0.437 | -1.292 | 7.388 | 0.022 | -22.646 | 57.565 | 838.761 | 18.205 | 0.505 |
| **4-5** | 49.930 | 0.610 | -2.603 | 6.647 | 0.051 | -22.169 | 53.898 | 848.766 | 22.942 | 0.477 |
| **4-6** | 53.565 | 0.328 | -2.288 | 7.151 | 0.091 | -16.321 | 55.920 | 774.548 | 37.202 | 0.556 |
| **4-7** | 52.768 | 0.295 | -2.625 | 6.312 | 0.017 | -9.837 | 49.884 | 774.691 | 35.402 | 0.594 |
| **4-8** | 47.978 | 0.358 | -2.301 | 6.686 | 0.019 | -21.133 | 54.723 | 787.231 | 30.944 | 0.549 |
| **4-9** | 80.389 | 20.128 | -6.517 | 7.546 | 0.015 | -14.845 | 58.522 | 777.595 | 67.659 | 0.840 |
| **4-10** | 74.127 | 17.305 | -5.523 | 6.945 | 0.073 | -12.696 | 54.870 | 739.395 | 74.161 | 0.698 |
| **5-1** | 37.102 | -10.259 | -12.297 | 6.142 | 0.025 | -8.443 | 45.676 | 605.770 | 122.013 | 1.254 |
| **5-2** | 36.536 | -12.880 | -15.700 | 6.032 | 0.027 | -8.292 | 46.834 | 668.690 | 114.895 | 1.080 |
| **5-3** | 49.389 | 1.770 | -5.894 | 5.908 | 0.047 | -9.321 | 47.205 | 693.361 | 92.840 | 0.490 |
| **5-4** | 47.459 | 1.969 | -5.532 | 6.022 | 0.030 | -12.600 | 47.566 | 752.831 | 65.676 | 0.501 |
| **5-5** | 33.058 | -12.511 | -12.776 | 5.705 | 0.048 | -9.612 | 46.556 | 715.633 | 77.608 | 1.189 |
| **5-6** | 16.267 | -30.746 | -13.208 | 5.743 | 0.021 | -5.577 | 43.572 | 703.050 | 114.033 | 1.140 |
| **6-1** | 50.019 | 0.340 | -2.500 | 6.546 | 0.071 | -17.963 | 52.229 | 771.363 | 27.058 | 0.598 |
| **6-2** | 48.069 | 0.256 | -2.582 | 6.389 | 0.128 | -19.611 | 52.913 | 785.700 | 28.148 | 0.533 |
| **6-3** | 53.119 | 6.316 | -6.588 | 5.168 | 0.017 | -1.073 | 39.286 | 538.742 | 39.287 | 0.480 |
| **6-4** | 58.956 | 5.602 | -2.959 | 6.501 | 0.011 | -10.965 | 50.780 | 658.618 | 41.381 | 0.567 |
| **6-5** | 56.191 | 5.473 | -2.443 | 6.454 | 0.014 | -14.184 | 52.142 | 681.750 | 42.498 | 0.567 |
| **6-6** | 81.967 | 22.145 | -2.984 | 6.770 | 0.024 | -6.153 | 52.690 | 628.508 | 85.162 | 0.637 |
| **6-7** | 33.824 | 6.625 | -5.723 | 3.884 | 0.009 | -15.791 | 32.401 | 517.812 | 42.234 | 0.583 |
| **6-8** | 27.045 | -0.683 | -14.730 | 2.922 | 0.042 | -2.930 | 23.255 | 358.993 | 138.971 | 1.612 |
| **6-9** | 20.393 | -1.816 | -2.702 | 3.805 | 0.021 | -21.622 | 34.158 | 645.261 | 28.621 | 0.999 |
| **6-10** | 30.571 | 3.421 | -3.423 | 3.761 | 0.009 | -15.769 | 32.439 | 526.147 | 43.929 | 0.996 |
| **6-11** | 59.690 | -3.336 | -12.851 | 6.962 | 1.104 | -12.841 | 56.279 | 769.096 | 96.012 | 1.143 |
| **6-12** | 55.742 | -4.803 | -9.833 | 7.017 | 0.019 | -16.046 | 56.763 | 829.927 | 64.147 | 1.628 |

**Table A4-3. 2D Descriptors of Synthesized Compounds (protonated)**

| Comp | mr | logP(o/w) | SlogP | vdw_area | vdw_vol | TPSA | logS | Weight | lip_druglike | lip_violation |
|------|------|------|------|------|------|------|------|------|------|------|
| 1-1 | 11.774 | 7.190 | 6.271 | 410.237 | 575.378 | 48.02 | -7.888 | 376.544 | 1 | 1 |
| 1-2 | 11.777 | 7.151 | 6.271 | 410.237 | 575.378 | 48.02 | -7.888 | 376.544 | 1 | 1 |
| 1-3 | 11.774 | 7.153 | 6.271 | 410.237 | 575.378 | 48.02 | -7.888 | 376.544 | 1 | 1 |
| 1-4 | 11.991 | 6.848 | 5.972 | 425.240 | 586.040 | 57.25 | -7.464 | 392.543 | 1 | 1 |
| 1-5 | 12.451 | 7.189 | 6.362 | 442.472 | 610.469 | 57.25 | -7.791 | 406.570 | 1 | 1 |
| 1-6 | 11.333 | 6.855 | 5.963 | 393.004 | 550.949 | 48.02 | -7.414 | 362.517 | 1 | 1 |
| 1-7 | 8.847 | 4.858 | 4.296 | 322.506 | 435.975 | 48.02 | -4.987 | 286.419 | 1 | 0 |
| 1-8 | 8.927 | 5.048 | 4.435 | 326.915 | 439.118 | 48.02 | -5.282 | 304.409 | 1 | 0 |
| 2-1 | 8.055 | 3.635 | 3.171 | 262.333 | 378.329 | 58.88 | -4.889 | 264.328 | 1 | 0 |
| 2-2 | 11.605 | 6.785 | 6.289 | 390.570 | 553.204 | 48.02 | -7.607 | 422.450 | 1 | 1 |
| 2-3 | 10.308 | 4.839 | 4.877 | 357.117 | 496.189 | 48.02 | -5.940 | 332.447 | 1 | 0 |
| 2-4 | 12.569 | 5.703 | 6.604 | 441.854 | 612.433 | 48.02 | -7.794 | 400.566 | 1 | 1 |
| 3-1 | 10.415 | 8.162 | 7.244 | 356.911 | 510.670 | 4.93 | -7.574 | 319.492 | 1 | 1 |
| 3-2 | 12.265 | 7.546 | 6.532 | 431.138 | 601.423 | 34.03 | -7.782 | 390.571 | 1 | 1 |
| 3-3 | 12.704 | 7.743 | 6.874 | 458.417 | 627.468 | 25.24 | -7.676 | 404.598 | 1 | 1 |
| 3-4 | 13.640 | 8.425 | 7.654 | 492.881 | 676.327 | 25.24 | -8.330 | 432.652 | 1 | 1 |
| 3-5 | 13.445 | 8.275 | 7.408 | 464.258 | 662.431 | 25.24 | -8.228 | 430.636 | 1 | 1 |
| 3-6 | 13.909 | 8.717 | 7.799 | 481.490 | 686.860 | 25.24 | -8.430 | 444.663 | 1 | 1 |
| 3-7 | 14.245 | 7.268 | 5.143 | 517.894 | 708.647 | 29.68 | -7.698 | 460.686 | 1 | 1 |
| 4-1 | 11.217 | 7.415 | 6.252 | 391.284 | 553.655 | 32.57 | -7.215 | 349.542 | 1 | 1 |
| 4-2 | 12.124 | 8.208 | 6.155 | 445.011 | 605.745 | 9.37 | -7.421 | 377.596 | 1 | 1 |
| 4-3 | 13.059 | 8.890 | 6.935 | 479.475 | 654.603 | 9.37 | -8.076 | 405.650 | 1 | 1 |
| 4-4 | 13.920 | 9.814 | 7.712 | 513.939 | 703.461 | 9.37 | -8.730 | 433.704 | 1 | 1 |
| 4-5 | 13.983 | 10.118 | 7.715 | 513.939 | 703.461 | 9.37 | -8.479 | 433.704 | 1 | 1 |
| 4-6 | 13.022 | 9.011 | 6.933 | 479.475 | 654.603 | 9.37 | -8.076 | 405.650 | 1 | 1 |
| 4-7 | 12.866 | 8.740 | 6.689 | 450.851 | 640.707 | 9.37 | -7.974 | 403.634 | 1 | 1 |
| 4-8 | 13.328 | 9.182 | 7.079 | 468.083 | 665.136 | 9.37 | -8.175 | 417.661 | 1 | 1 |
| 4-9 | 13.663 | 7.733 | 4.423 | 504.487 | 686.923 | 13.81 | -7.444 | 433.684 | 1 | 0 |
| 4-10 | 13.111 | 7.774 | 5.925 | 465.855 | 651.368 | 18.60 | -7.713 | 419.633 | 1 | 1 |
| 5-1 | 11.311 | 7.102 | 6.343 | 393.004 | 550.949 | 48.02 | -7.826 | 362.517 | 1 | 1 |
| 5-2 | 12.238 | 7.278 | 6.661 | 427.469 | 599.808 | 48.02 | -7.770 | 390.571 | 1 | 1 |
| 5-3 | 12.151 | 8.188 | 6.959 | 428.404 | 595.798 | 31.23 | -8.017 | 391.555 | 1 | 1 |
| 5-4 | 12.609 | 8.529 | 7.349 | 445.636 | 620.227 | 31.23 | -8.345 | 405.582 | 1 | 1 |
| 5-5 | 12.256 | 7.584 | 7.146 | 431.138 | 601.423 | 34.03 | -7.728 | 390.571 | 1 | 1 |
| 5-6 | 12.667 | 7.388 | 6.559 | 451.113 | 614.467 | 51.10 | -7.356 | 426.625 | 1 | 1 |
| 6-1 | 13.063 | 8.851 | 6.935 | 479.475 | 654.603 | 9.37 | -8.076 | 405.650 | 1 | 1 |
| 6-2 | 13.059 | 8.853 | 6.935 | 479.475 | 654.603 | 9.37 | -8.076 | 405.650 | 1 | 1 |
| 6-3 | 8.957 | 5.811 | 5.849 | 303.791 | 431.481 | 4.93 | -5.626 | 275.395 | 1 | 1 |
| 6-4 | 11.606 | 6.539 | 5.540 | 426.355 | 575.414 | 9.37 | -6.128 | 361.553 | 1 | 1 |
| 6-5 | 11.874 | 6.831 | 5.685 | 414.964 | 585.947 | 9.37 | -6.227 | 373.564 | 1 | 1 |
| 6-6 | 11.652 | 5.423 | 4.531 | 412.735 | 572.179 | 18.60 | -5.765 | 375.536 | 1 | 0 |
| 6-7 | 8.685 | 4.207 | 3.565 | 338.625 | 436.011 | 9.37 | -3.227 | 271.428 | 1 | 0 |
| 6-8 | 7.452 | 2.697 | 3.041 | 273.796 | 359.928 | 48.02 | -3.334 | 260.312 | 1 | 0 |
| 6-9 | 10.221 | 6.748 | 5.098 | 396.154 | 518.342 | 9.37 | -5.470 | 333.515 | 1 | 1 |
| 6-10 | 8.765 | 4.397 | 3.704 | 343.034 | 439.153 | 9.37 | -3.522 | 289.418 | 1 | 0 |
| 6-11 | 13.862 | 8.083 | 7.355 | 507.885 | 686.988 | 34.47 | -7.907 | 448.651 | 1 | 1 |
| 6-12 | 14.321 | 8.424 | 7.745 | 525.117 | 711.417 | 34.47 | -8.234 | 462.678 | 1 | 1 |

**Table A4-3 (cont'd)**

| Comp | b_count | b_rotN | b_1rotN | b_rotR | b_1rotR | a_acc | a_don | a_base | vsa_acc | vsa_don | vsa_hyd | vsa_pol |
|------|---------|--------|---------|--------|---------|-------|-------|--------|---------|---------|---------|---------|
| 1-1  | 62 | 10 | 10 | 0.333 | 0.333 | 1 | 1 | 0 | 13.567 | 17.742 | 338.477 | 31.309 |
| 1-2  | 62 | 10 | 10 | 0.333 | 0.333 | 1 | 1 | 0 | 13.567 | 17.742 | 338.477 | 31.309 |
| 1-3  | 62 | 10 | 10 | 0.333 | 0.333 | 1 | 1 | 0 | 13.567 | 17.742 | 338.477 | 31.309 |
| 1-4  | 63 | 11 | 11 | 0.355 | 0.355 | 2 | 1 | 0 | 16.071 | 17.742 | 350.977 | 33.813 |
| 1-5  | 66 | 12 | 12 | 0.375 | 0.375 | 2 | 1 | 0 | 16.071 | 17.742 | 366.786 | 33.813 |
| 1-6  | 59 | 10 | 10 | 0.345 | 0.345 | 1 | 1 | 0 | 13.567 | 17.742 | 322.669 | 31.309 |
| 1-7  | 48 | 9 | 9 | 0.409 | 0.409 | 1 | 1 | 0 | 13.567 | 17.742 | 255.018 | 31.309 |
| 1-8  | 48 | 9 | 9 | 0.391 | 0.391 | 1 | 1 | 0 | 13.567 | 17.742 | 260.139 | 31.309 |
| 2-1  | 38 | 3 | 3 | 0.136 | 0.136 | 1 | 2 | 0 | 13.567 | 23.425 | 193.918 | 36.992 |
| 2-2  | 55 | 6 | 5 | 0.176 | 0.147 | 1 | 1 | 0 | 13.567 | 17.742 | 326.640 | 31.309 |
| 2-3  | 51 | 5 | 5 | 0.185 | 0.185 | 1 | 1 | 0 | 13.567 | 17.742 | 291.052 | 31.309 |
| 2-4  | 64 | 8 | 8 | 0.250 | 0.250 | 1 | 1 | 0 | 13.567 | 17.742 | 370.094 | 31.309 |
| 3-1  | 55 | 8 | 8 | 0.308 | 0.308 | 0 | 0 | 0 | 0.000 | 0.000 | 336.269 | 0.000 |
| 3-2  | 65 | 11 | 10 | 0.355 | 0.323 | 1 | 1 | 0 | 13.567 | 5.683 | 372.377 | 19.249 |
| 3-3  | 68 | 11 | 11 | 0.344 | 0.344 | 1 | 0 | 0 | 13.567 | 0.000 | 406.276 | 13.567 |
| 3-4  | 74 | 13 | 13 | 0.382 | 0.382 | 1 | 0 | 0 | 13.567 | 0.000 | 437.893 | 13.567 |
| 3-5  | 73 | 11 | 11 | 0.314 | 0.314 | 1 | 0 | 0 | 13.567 | 0.000 | 410.693 | 13.567 |
| 3-6  | 76 | 11 | 11 | 0.306 | 0.306 | 1 | 0 | 0 | 13.567 | 0.000 | 426.502 | 13.567 |
| 3-7  | 79 | 11 | 11 | 0.297 | 0.297 | 1 | 0 | 1 | 13.567 | 0.000 | 453.575 | 13.567 |
| 4-1  | 61 | 9 | 9 | 0.321 | 0.321 | 0 | 0 | 1 | 0.000 | 17.742 | 342.968 | 17.742 |
| 4-2  | 67 | 10 | 10 | 0.333 | 0.333 | 0 | 0 | 1 | 0.000 | 0.000 | 410.767 | 0.000 |
| 4-3  | 73 | 12 | 12 | 0.375 | 0.375 | 0 | 0 | 1 | 0.000 | 0.000 | 442.384 | 0.000 |
| 4-4  | 79 | 12 | 12 | 0.353 | 0.353 | 0 | 0 | 1 | 0.000 | 0.000 | 474.001 | 0.000 |
| 4-5  | 79 | 14 | 14 | 0.412 | 0.412 | 0 | 0 | 1 | 0.000 | 0.000 | 474.001 | 0.000 |
| 4-6  | 73 | 11 | 11 | 0.344 | 0.344 | 0 | 0 | 1 | 0.000 | 0.000 | 442.384 | 0.000 |
| 4-7  | 72 | 10 | 10 | 0.303 | 0.303 | 0 | 0 | 1 | 0.000 | 0.000 | 415.184 | 0.000 |
| 4-8  | 75 | 10 | 10 | 0.294 | 0.294 | 0 | 0 | 1 | 0.000 | 0.000 | 430.993 | 0.000 |
| 4-9  | 78 | 10 | 10 | 0.286 | 0.286 | 0 | 0 | 2 | 0.000 | 0.000 | 458.066 | 0.000 |
| 4-10 | 73 | 10 | 10 | 0.294 | 0.294 | 1 | 0 | 1 | 2.504 | 0.000 | 427.684 | 2.504 |
| 5-1  | 59 | 9 | 9 | 0.310 | 0.310 | 1 | 1 | 0 | 13.567 | 17.742 | 322.669 | 31.309 |
| 5-2  | 65 | 11 | 11 | 0.355 | 0.355 | 1 | 1 | 0 | 13.567 | 17.742 | 354.286 | 31.309 |
| 5-3  | 64 | 11 | 10 | 0.355 | 0.323 | 1 | 0 | 0 | 13.567 | 0.000 | 374.136 | 13.567 |
| 5-4  | 67 | 12 | 11 | 0.375 | 0.344 | 1 | 0 | 0 | 13.567 | 0.000 | 389.944 | 13.567 |
| 5-5  | 65 | 11 | 10 | 0.355 | 0.323 | 1 | 1 | 0 | 13.567 | 5.683 | 372.377 | 19.249 |
| 5-6  | 66 | 11 | 11 | 0.344 | 0.344 | 2 | 1 | 0 | 32.017 | 8.075 | 374.353 | 40.092 |
| 6-1  | 73 | 12 | 12 | 0.375 | 0.375 | 0 | 0 | 1 | 0.000 | 0.000 | 442.384 | 0.000 |
| 6-2  | 73 | 12 | 12 | 0.375 | 0.375 | 0 | 0 | 1 | 0.000 | 0.000 | 442.384 | 0.000 |
| 6-3  | 44 | 3 | 3 | 0.130 | 0.130 | 0 | 0 | 0 | 0.000 | 0.000 | 288.843 | 0.000 |
| 6-4  | 62 | 7 | 7 | 0.241 | 0.241 | 0 | 0 | 1 | 0.000 | 0.000 | 394.959 | 0.000 |
| 6-5  | 64 | 5 | 5 | 0.161 | 0.161 | 0 | 0 | 1 | 0.000 | 0.000 | 383.567 | 0.000 |
| 6-6  | 62 | 5 | 5 | 0.161 | 0.161 | 1 | 0 | 1 | 2.504 | 0.000 | 380.259 | 2.504 |
| 6-7  | 48 | 6 | 6 | 0.286 | 0.286 | 0 | 0 | 1 | 0.000 | 0.000 | 311.499 | 0.000 |
| 6-8  | 37 | 4 | 4 | 0.200 | 0.200 | 1 | 1 | 0 | 13.567 | 17.742 | 212.713 | 31.309 |
| 6-9  | 59 | 11 | 11 | 0.440 | 0.440 | 0 | 0 | 1 | 0.000 | 0.000 | 364.045 | 0.000 |
| 6-10 | 48 | 6 | 6 | 0.273 | 0.273 | 0 | 0 | 1 | 0.000 | 0.000 | 316.620 | 0.000 |
| 6-11 | 75 | 14 | 14 | 0.400 | 0.400 | 2 | 0 | 0 | 16.071 | 0.000 | 450.393 | 16.071 |
| 6-12 | 78 | 15 | 15 | 0.417 | 0.417 | 2 | 0 | 0 | 16.071 | 0.000 | 466.201 | 16.071 |

**Table A4-4. 3D Descriptors for Synthesized Compounds (protonated)**

| Comp | PM3_E | PM3_Eele | PM3_HOMO | PM3_LUMO | PM3_IP | PM3_HF | PM3_dipole | VSA | vol |
|------|-------|----------|----------|----------|--------|--------|------------|-----|-----|
| 1-1 | -93967.0 | -802952 | -8.383 | -0.112 | 8.383 | -27.139 | 5.706 | 467.784 | 414.250 |
| 1-2 | -93965.3 | -809218 | -8.459 | -0.099 | 8.459 | -25.399 | 6.080 | 467.473 | 414.625 |
| 1-3 | -93966.7 | -797032 | -8.339 | -0.108 | 8.339 | -26.601 | 5.690 | 465.154 | 413.125 |
| 1-4 | -100727.4 | -856963 | -8.453 | -0.185 | 8.453 | -55.571 | 6.938 | 474.798 | 421.125 |
| 1-5 | -104176.3 | -900474 | -8.418 | -0.137 | 8.418 | -61.325 | 5.297 | 497.479 | 438.875 |
| 1-6 | -90514.4 | -753086 | -8.394 | -0.139 | 8.394 | -17.570 | 6.007 | 445.458 | 397.750 |
| 1-7 | -72709.3 | -541570 | -8.428 | -0.036 | 8.428 | -42.255 | 5.947 | 368.404 | 315.250 |
| 1-8 | -82507.6 | -591475 | -8.618 | -0.318 | 8.618 | -77.748 | 7.445 | 370.609 | 315.750 |
| 2-1 | -66382.6 | -467374 | -8.541 | -0.197 | 8.541 | 12.720 | 4.922 | 302.951 | 269.875 |
| 2-2 | -120493.1 | -946202 | -8.577 | -0.847 | 8.577 | -117.771 | 3.397 | 436.704 | 396.625 |
| 2-3 | -82892.1 | -674633 | -8.336 | -0.078 | 8.336 | 11.441 | 6.234 | 397.760 | 356.500 |
| 2-4 | -99404.8 | -878781 | -8.303 | -0.038 | 8.303 | 30.541 | 6.275 | 491.387 | 438.125 |
| 3-1 | -76909.7 | -610810 | -8.270 | 0.073 | 8.270 | 24.992 | 2.411 | 411.543 | 364.875 |
| 3-2 | -97411.9 | -856495 | -8.390 | -0.124 | 8.390 | -28.087 | 5.429 | 490.929 | 429.125 |
| 3-3 | -100854.1 | -915406 | -8.287 | -0.022 | 8.287 | -28.180 | 5.107 | 504.560 | 451.000 |
| 3-4 | -107751.4 | -1022892 | -8.227 | 0.045 | 8.227 | -39.319 | 5.131 | 542.640 | 486.375 |
| 3-5 | -107038.1 | -1003048 | -8.251 | 0.020 | 8.251 | -33.207 | 5.340 | 526.654 | 475.625 |
| 3-6 | -110485.6 | -1072204 | -8.237 | 0.030 | 8.237 | -37.623 | 5.156 | 538.294 | 490.875 |
| 3-7 | -114771.2 | -1133621 | -10.288 | -4.370 | 10.288 | 131.647 | 24.005 | 564.800 | 506.500 |
| 4-1 | -84667.0 | -719196 | -11.322 | -4.124 | 11.322 | 158.449 | 16.379 | 450.894 | 396.625 |
| 4-2 | -91552.4 | -823764 | -11.246 | -3.926 | 11.246 | 159.216 | 15.378 | 490.838 | 438.000 |
| 4-3 | -98453.5 | -934653 | -11.202 | -3.778 | 11.202 | 144.202 | 14.423 | 526.283 | 469.625 |
| 4-4 | -105353.1 | -1059673 | -11.170 | -3.657 | 11.170 | 130.799 | 13.040 | 562.641 | 506.000 |
| 4-5 | -105350.3 | -1047212 | -11.199 | -3.693 | 11.199 | 133.632 | 13.193 | 568.909 | 506.125 |
| 4-6 | -98453.1 | -934408 | -11.224 | -3.793 | 11.224 | 144.614 | 14.461 | 522.404 | 471.750 |
| 4-7 | -97738.1 | -913427 | -11.200 | -3.739 | 11.200 | 152.452 | 14.614 | 512.332 | 461.875 |
| 4-8 | -101187.0 | -970802 | -11.184 | -3.702 | 11.184 | 146.706 | 13.719 | 529.371 | 477.375 |
| 4-9 | -105389.1 | -1034314 | -13.013 | -8.010 | 13.013 | 399.347 | 36.845 | 549.357 | 493.125 |
| 4-10 | -104497.3 | -976611 | -11.253 | -3.942 | 11.253 | 125.254 | 11.898 | 520.442 | 467.250 |
| 5-1 | -90517.6 | -747965 | -8.422 | -0.203 | 8.422 | -20.964 | 5.605 | 443.873 | 397.500 |
| 5-2 | -97412.8 | -845614 | -8.264 | -0.011 | 8.264 | -3.638 | 4.968 | 484.599 | 429.250 |
| 5-3 | -100083.7 | -853449 | -8.284 | 0.001 | 8.284 | -60.770 | 2.497 | 484.458 | 426.125 |
| 5-4 | -103531.7 | -900712 | -8.277 | -0.015 | 8.277 | -74.661 | 2.046 | 506.346 | 448.000 |
| 5-5 | -97413.1 | -839488 | -8.416 | -0.135 | 8.416 | -29.789 | 5.050 | 490.902 | 433.250 |
| 5-6 | -105733.9 | -919890 | -8.443 | -0.303 | 8.443 | -58.097 | 5.116 | 502.294 | 444.125 |
| 6-1 | -98450.7 | -944231 | -11.388 | -3.781 | 11.388 | 179.831 | 14.534 | 522.424 | 469.875 |
| 6-2 | -98453.4 | -934728 | -11.085 | -3.775 | 11.085 | 144.366 | 14.290 | 523.619 | 468.250 |
| 6-3 | -65835.6 | -497851 | -8.224 | 0.111 | 8.224 | 55.589 | 2.685 | 341.066 | 309.000 |
| 6-4 | -87381.3 | -806787 | -11.166 | -3.710 | 11.166 | 180.092 | 12.016 | 457.592 | 412.125 |
| 6-5 | -90115.1 | -839803 | -11.183 | -3.652 | 11.183 | 201.271 | 11.515 | 461.249 | 417.125 |
| 6-6 | -93425.1 | -838924 | -11.210 | -3.887 | 11.210 | 161.124 | 9.840 | 454.149 | 408.625 |
| 6-7 | -66123.0 | -537143 | -11.494 | -3.771 | 11.494 | 165.473 | 9.550 | 356.926 | 315.875 |
| 6-8 | -71433.4 | -476757 | -8.540 | -0.253 | 8.540 | -47.797 | 7.679 | 298.185 | 260.875 |
| 6-9 | -86992.1 | -708446 | -11.769 | -3.934 | 11.769 | 87.494 | 14.542 | 434.137 | 375.625 |
| 6-10 | -75920.0 | -588872 | -11.679 | -3.859 | 11.679 | 125.574 | 10.238 | 360.883 | 318.000 |
| 6-11 | -114511.5 | -1068743 | -8.285 | -0.019 | 8.285 | -67.531 | 3.870 | 558.753 | 494.875 |
| 6-12 | -117959.7 | -1142324 | -8.290 | -0.025 | 8.290 | -72.595 | 6.370 | 575.456 | 510.125 |

**Table A4-4 (cont'd)**

| Comp | E | E_ele | E_sol | E_str | E_strain | E_tor | E_vdw | ASA_H | ASA_P | dipole |
|------|-----|-------|-------|-------|----------|-------|-------|-------|-------|--------|
| **1-1** | 43.567 | -4.281 | -13.432 | 5.975 | 0.012 | -5.616 | 43.976 | 624.497 | 122.644 | 1.215 |
| **1-2** | 45.896 | -4.044 | -13.906 | 6.018 | 0.010 | -3.031 | 43.294 | 622.653 | 123.193 | 1.276 |
| **1-3** | 44.127 | -3.690 | -15.306 | 5.941 | 0.320 | -5.999 | 44.258 | 628.010 | 124.202 | 1.189 |
| **1-4** | 50.162 | -8.041 | -14.157 | 5.742 | 0.021 | -4.909 | 47.617 | 610.340 | 161.182 | 1.623 |
| **1-5** | 50.110 | -7.506 | -12.702 | 5.885 | 0.022 | -7.617 | 47.548 | 653.593 | 144.289 | 1.013 |
| **1-6** | 41.930 | -3.426 | -15.528 | 5.824 | 0.324 | -6.157 | 42.195 | 605.441 | 124.591 | 1.238 |
| **1-7** | 18.116 | -3.931 | -14.014 | 3.240 | 0.008 | -10.042 | 25.679 | 470.366 | 122.647 | 1.277 |
| **1-8** | 15.202 | -6.547 | -12.565 | 3.103 | 0.033 | -10.137 | 25.522 | 471.807 | 123.243 | 1.664 |
| **2-1** | 43.863 | -4.434 | -16.016 | 4.858 | 0.005 | 8.141 | 33.076 | 368.081 | 153.443 | 1.128 |
| **2-2** | 80.941 | 8.088 | -13.243 | 7.158 | 0.008 | 9.125 | 52.510 | 479.013 | 221.820 | 1.073 |
| **2-3** | 55.774 | 1.776 | -15.521 | 5.798 | 0.012 | 1.539 | 41.893 | 506.330 | 141.186 | 1.159 |
| **2-4** | 64.185 | 5.249 | -15.517 | 6.974 | 0.028 | -4.570 | 49.594 | 634.276 | 152.122 | 1.140 |
| **3-1** | 41.546 | 1.098 | -5.162 | 5.325 | 0.007 | -8.723 | 41.583 | 652.854 | 24.301 | 0.462 |
| **3-2** | 43.606 | -1.970 | -12.886 | 6.070 | 0.025 | -9.432 | 45.853 | 685.860 | 94.114 | 1.236 |
| **3-3** | 55.086 | -1.458 | -10.796 | 6.562 | 0.011 | -8.452 | 51.707 | 701.565 | 90.740 | 1.255 |
| **3-4** | 51.845 | -0.608 | -10.500 | 7.315 | 0.013 | -14.286 | 52.945 | 805.667 | 51.638 | 1.227 |
| **3-5** | 55.985 | -0.307 | -10.867 | 6.780 | 0.017 | -9.988 | 49.081 | 771.090 | 64.038 | 1.241 |
| **3-6** | 52.907 | -1.311 | -10.440 | 7.284 | 0.033 | -16.558 | 56.280 | 798.704 | 54.494 | 1.257 |
| **3-7** | 87.284 | 23.471 | -74.117 | 8.531 | 0.746 | -13.009 | 58.875 | 767.975 | 102.009 | 6.096 |
| **4-1** | 35.414 | -6.906 | -67.250 | 5.403 | 0.014 | -8.411 | 42.916 | 642.116 | 94.387 | 4.387 |
| **4-2** | 44.536 | -5.648 | -54.115 | 6.180 | 0.014 | -8.445 | 48.230 | 710.013 | 68.247 | 4.195 |
| **4-3** | 45.366 | -5.754 | -51.548 | 6.787 | 0.030 | -11.844 | 50.974 | 788.044 | 33.703 | 3.936 |
| **4-4** | 56.223 | -5.782 | -50.293 | 8.039 | 0.011 | -12.409 | 56.095 | 836.266 | 19.137 | 3.421 |
| **4-5** | 46.713 | -5.976 | -51.543 | 6.921 | 0.014 | -17.045 | 53.022 | 855.194 | 27.008 | 3.440 |
| **4-6** | 51.277 | -5.812 | -51.291 | 7.384 | 0.014 | -10.276 | 54.030 | 770.896 | 40.602 | 3.825 |
| **4-7** | 45.864 | -5.728 | -51.799 | 6.374 | 0.017 | -9.022 | 47.741 | 776.036 | 37.203 | 3.903 |
| **4-8** | 43.351 | -5.760 | -51.154 | 6.652 | 0.013 | -16.491 | 52.649 | 785.835 | 38.396 | 3.630 |
| **4-9** | 108.923 | 49.221 | -198.431 | 8.590 | 0.025 | -12.449 | 54.296 | 764.379 | 83.369 | 9.272 |
| **4-10** | 73.540 | 17.541 | -57.783 | 7.091 | 0.030 | -10.927 | 52.956 | 735.537 | 82.540 | 3.265 |
| **5-1** | 37.089 | -10.248 | -12.428 | 6.113 | 0.012 | -8.469 | 45.688 | 599.832 | 122.635 | 1.252 |
| **5-2** | 36.537 | -12.884 | -15.701 | 6.042 | 0.027 | -8.299 | 46.835 | 668.778 | 114.895 | 1.079 |
| **5-3** | 49.354 | 1.795 | -5.985 | 5.954 | 0.012 | -9.448 | 47.178 | 693.921 | 91.897 | 0.498 |
| **5-4** | 47.080 | 1.867 | -5.869 | 6.013 | 0.021 | -12.217 | 47.532 | 756.173 | 64.303 | 0.486 |
| **5-5** | 33.058 | -12.511 | -12.775 | 5.706 | 0.048 | -9.613 | 46.558 | 715.633 | 77.608 | 1.189 |
| **5-6** | 16.280 | -30.701 | -13.417 | 5.817 | 0.034 | -5.686 | 43.593 | 707.643 | 113.606 | 1.142 |
| **6-1** | 47.372 | -5.447 | -53.227 | 6.961 | 0.014 | -12.189 | 50.571 | 776.521 | 33.706 | 3.861 |
| **6-2** | 45.452 | -5.122 | -54.968 | 6.442 | 0.008 | -13.144 | 50.613 | 790.898 | 31.882 | 3.829 |
| **6-3** | 53.118 | 6.316 | -6.587 | 5.156 | 0.017 | -1.073 | 39.290 | 538.468 | 39.287 | 0.480 |
| **6-4** | 56.060 | -0.522 | -51.958 | 6.637 | 0.018 | -5.388 | 49.084 | 668.614 | 43.798 | 3.485 |
| **6-5** | 52.843 | -0.668 | -50.428 | 6.473 | 0.012 | -9.015 | 50.203 | 674.083 | 45.256 | 3.313 |
| **6-6** | 82.852 | 22.273 | -53.680 | 6.886 | 0.012 | -2.823 | 50.182 | 622.059 | 89.335 | 2.636 |
| **6-7** | 31.001 | 0.478 | -58.153 | 4.060 | 0.008 | -10.246 | 30.690 | 521.912 | 44.557 | 3.025 |
| **6-8** | 27.045 | -0.683 | -14.730 | 2.922 | 0.042 | -2.930 | 23.255 | 358.993 | 138.971 | 1.612 |
| **6-9** | 18.143 | -6.653 | -59.132 | 4.174 | 0.014 | -17.230 | 32.734 | 638.146 | 31.875 | 3.950 |
| **6-10** | 28.849 | -1.593 | -58.322 | 4.027 | 0.008 | -10.239 | 30.577 | 528.260 | 44.245 | 3.136 |
| **6-11** | 59.690 | -3.336 | -12.847 | 6.965 | 1.104 | -12.842 | 56.281 | 769.096 | 96.012 | 1.143 |
| **6-12** | 55.742 | -4.803 | -9.833 | 7.017 | 0.019 | -16.046 | 56.763 | 829.927 | 64.147 | 1.628 |

**Table A4-5. Spearman Rho Correlation Coefficient of Descriptors employed in developing Equations 5-2 and 5-4 (non-protonated compounds)**

| | | ICMT | b_1rotN | b_1rotR | b_count | b_rotN | b_rotR | weight | lip_violation | E | E_ele | E_sol | E_str | E_strain | E_tor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| b_1rotN | Correlation Coefficient | **.601**\*\* | 1.000 | | | | | | | | | | | | |
| | Sig. (2-tailed) | 0.000 | . | | | | | | | | | | | | |
| b_1rotR | Correlation Coefficient | **.387**\*\* | **.801**\*\* | 1.000 | | | | | | | | | | | |
| | Sig. (2-tailed) | 0.007 | 0.000 | . | | | | | | | | | | | |
| b_count | Correlation Coefficient | **.672**\*\* | **.812**\*\* | **.394**\*\* | 1.000 | | | | | | | | | | |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.006 | . | | | | | | | | | | |
| b_rotN | Correlation Coefficient | **.528**\*\* | **.979**\*\* | **.790**\*\* | **.794**\*\* | 1.000 | | | | | | | | | |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.000 | 0.000 | . | | | | | | | | | |
| b_rotR | Correlation Coefficient | **.307**\* | **.778**\*\* | **.972**\*\* | **.387**\*\* | **.813**\*\* | 1.000 | | | | | | | | |
| | Sig. (2-tailed) | 0.036 | 0.000 | 0.000 | 0.007 | 0.000 | . | | | | | | | | |
| weight | Correlation Coefficient | **.580**\*\* | **.767**\*\* | **.337**\* | **.926**\*\* | **.758**\*\* | **.338**\* | 1.000 | | | | | | | |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.020 | 0.000 | 0.000 | 0.020 | . | | | | | | | |
| lip_violation | Correlation Coefficient | **.432**\*\* | **.460**\*\* | **0.148** | **.544**\*\* | **.465**\*\* | **0.148** | **.550**\*\* | 1.000 | | | | | | |
| | Sig. (2-tailed) | 0.002 | 0.001 | 0.320 | 0.000 | 0.001 | 0.320 | 0.000 | . | | | | | | |
| E | Correlation Coefficient | **0.238** | **0.103** | **-.303**\* | **.475**\*\* | **0.072** | **-.327**\* | **.533**\*\* | **.348**\* | 1.000 | | | | | |
| | Sig. (2-tailed) | 0.106 | 0.489 | 0.038 | 0.001 | 0.631 | 0.025 | 0.000 | 0.017 | . | | | | | |
| E_ele | Correlation Coefficient | **-0.171** | **-.299**\* | **-.545**\*\* | **0.069** | **-.294**\* | **-.520**\*\* | **0.070** | **0.089** | **.582**\*\* | 1.000 | | | | |
| | Sig. (2-tailed) | 0.251 | 0.041 | 0.000 | 0.646 | 0.045 | 0.000 | 0.640 | 0.551 | 0.000 | . | | | | |
| E_sol | Correlation Coefficient | **0.222** | **0.239** | **0.071** | **.393**\*\* | **0.228** | **0.072** | **0.215** | **.343**\* | **0.201** | **.503**\*\* | 1.000 | | | |
| | Sig. (2-tailed) | 0.134 | 0.105 | 0.634 | 0.006 | 0.124 | 0.632 | 0.146 | 0.018 | 0.175 | 0.000 | . | | | |
| E_str | Correlation Coefficient | **.503**\*\* | **.449**\*\* | **0.001** | **.788**\*\* | **.421**\*\* | **-0.023** | **.818**\*\* | **.522**\*\* | **.793**\*\* | **.377**\*\* | **.296**\* | 1.000 | | |
| | Sig. (2-tailed) | 0.000 | 0.002 | 0.997 | 0.000 | 0.003 | 0.880 | 0.000 | 0.000 | 0.000 | 0.009 | 0.043 | . | | |
| E_strain | Correlation Coefficient | **0.269** | **.318**\* | **0.225** | **.312**\* | **.357**\* | **0.264** | **.366**\* | **0.273** | **0.115** | **-0.004** | **0.003** | **0.276** | 1.000 | |
| | Sig. (2-tailed) | 0.067 | 0.030 | 0.129 | 0.033 | 0.014 | 0.073 | 0.011 | 0.064 | 0.442 | 0.978 | 0.986 | 0.061 | . | |
| E_tor | Correlation Coefficient | **-.332**\* | **-.541**\*\* | **-.407**\*\* | **-.588**\*\* | **-.525**\*\* | **-.416**\*\* | **-.419**\*\* | **-0.207** | **-0.046** | **-0.230** | **-.755**\*\* | **-.365**\* | **-0.127** | 1.000 |
| | Sig. (2-tailed) | 0.023 | 0.000 | 0.005 | 0.000 | 0.000 | 0.004 | 0.003 | 0.163 | 0.757 | 0.121 | 0.000 | 0.012 | 0.396 | . |

| | | ICMT | b_1rotN | b_1rotR | b_count | b_rotN | b_rotR | weight | lip_violation | E | E_ele | E_sol | E_str | E_strain | E_tor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| E_vdw | Correlation Coefficient | .554** | .565** | 0.115 | .873** | .539** | 0.098 | .859** | .555** | .748** | .336* | .443** | .936** | .328* | -.498** |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.441 | 0.000 | 0.000 | 0.514 | 0.000 | 0.000 | 0.000 | 0.021 | 0.002 | 0.000 | 0.024 | 0.000 |
| logS | Correlation Coefficient | -.545** | -.733** | -.413** | -.801** | -.754** | -.434** | -.770** | -.545** | -0.287 | 0.060 | -.288* | -.647** | -.307* | .431** |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.004 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 | 0.050 | 0.689 | 0.050 | 0.000 | 0.036 | 0.002 |
| mr | Correlation Coefficient | .654** | .791** | .353* | .989** | .777** | .352* | .951** | .555** | .519** | 0.087 | .336* | .817** | .325* | -.529** |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.015 | 0.000 | 0.000 | 0.015 | 0.000 | 0.000 | 0.000 | 0.563 | 0.021 | 0.000 | 0.026 | 0.000 |
| dipole | Correlation Coefficient | -0.047 | 0.004 | 0.178 | -0.161 | -0.017 | 0.154 | -0.062 | -0.287 | -0.189 | -.622** | -.716** | -0.142 | -0.069 | .364* |
| | Sig. (2-tailed) | 0.752 | 0.978 | 0.231 | 0.281 | 0.908 | 0.300 | 0.680 | 0.051 | 0.202 | 0.000 | 0.000 | 0.343 | 0.644 | 0.012 |
| a_acc | Correlation Coefficient | 0.270 | .351* | 0.164 | .408** | .315* | 0.138 | .499** | 0.134 | .347* | -0.074 | -0.099 | .336* | 0.202 | -0.046 |
| | Sig. (2-tailed) | 0.066 | 0.016 | 0.270 | 0.004 | 0.031 | 0.356 | 0.000 | 0.370 | 0.017 | 0.621 | 0.507 | 0.021 | 0.173 | 0.758 |
| a_don | Correlation Coefficient | -0.149 | -.327* | -0.042 | -.512** | -.303* | -0.042 | -.383** | -.371* | -.436** | -.610** | -.819** | -.463** | 0.030 | .699** |
| | Sig. (2-tailed) | 0.316 | 0.025 | 0.778 | 0.000 | 0.038 | 0.778 | 0.008 | 0.010 | 0.002 | 0.000 | 0.000 | 0.001 | 0.844 | 0.000 |
| vsa_acc | Correlation Coefficient | -0.108 | 0.183 | 0.219 | 0.003 | 0.216 | 0.245 | 0.206 | -0.118 | -0.054 | -.581** | -.777** | -0.048 | 0.104 | .501** |
| | Sig. (2-tailed) | 0.471 | 0.218 | 0.138 | 0.985 | 0.145 | 0.096 | 0.165 | 0.430 | 0.721 | 0.000 | 0.000 | 0.748 | 0.487 | 0.000 |
| vsa_don | Correlation Coefficient | -0.213 | -.313* | -0.031 | -.509** | -.308* | -0.052 | -.365* | -.416** | -.363* | -.560** | -.835** | -.412** | -0.031 | .710** |
| | Sig. (2-tailed) | 0.150 | 0.032 | 0.835 | 0.000 | 0.035 | 0.728 | 0.012 | 0.004 | 0.012 | 0.000 | 0.000 | 0.004 | 0.836 | 0.000 |
| vsa_hyd | Correlation Coefficient | .649** | .733** | .329* | .946** | .715** | .321* | .843** | .569** | .545** | 0.248 | .593** | .799** | .297* | -.673** |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.024 | 0.000 | 0.000 | 0.028 | 0.000 | 0.000 | 0.000 | 0.093 | 0.000 | 0.000 | 0.042 | 0.000 |
| vsa_pol | Correlation Coefficient | -0.152 | -0.115 | 0.068 | -.316* | -0.098 | 0.067 | -0.122 | -.297* | -0.262 | -.642** | -.903** | -.292* | 0.067 | .704** |
| | Sig. (2-tailed) | 0.308 | 0.440 | 0.650 | 0.030 | 0.513 | 0.654 | 0.414 | 0.043 | 0.075 | 0.000 | 0.000 | 0.046 | 0.653 | 0.000 |
| PM3_dipole | Correlation Coefficient | -0.012 | -0.060 | 0.139 | -0.229 | -0.090 | 0.103 | -0.129 | -.291* | -0.199 | -.596** | -.761** | -0.207 | -0.061 | .458** |
| | Sig. (2-tailed) | 0.935 | 0.690 | 0.352 | 0.121 | 0.549 | 0.490 | 0.387 | 0.047 | 0.180 | 0.000 | 0.000 | 0.163 | 0.684 | 0.001 |
| PM3_E | Correlation Coefficient | -.491** | -.710** | -.299* | -.844** | -.711** | -.310* | -.977** | -.531** | -.528** | -0.035 | -0.089 | -.785** | -.363* | .293* |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.041 | 0.000 | 0.000 | 0.034 | 0.000 | 0.000 | 0.000 | 0.814 | 0.553 | 0.000 | 0.012 | 0.045 |
| PM3_Eele | Correlation Coefficient | -.600** | -.746** | -.307* | -.941** | -.735** | -.309* | -.985** | -.545** | -.583** | -0.143 | -.296* | -.867** | -.361* | .489** |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.036 | 0.000 | 0.000 | 0.035 | 0.000 | 0.000 | 0.000 | 0.337 | 0.044 | 0.000 | 0.013 | 0.000 |

**Table A4-5 (cont'd)**

| | | ICMT | b_1rotN | b_1rotR | b_count | b_rotN | b_rotR | weight | lip_violation | E | E_ele | E_sol | E_str | E_strain | E_tor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PM3_HF | Correlation Coefficient | **0.115** | **-.423**\*\* | **-.509**\*\* | **-0.139** | **-.474**\*\* | **-.568**\*\* | **-.334**\* | **0.014** | **0.208** | **.470**\*\* | **.361**\* | **0.029** | **-0.237** | **-0.028** |
| | Sig. (2-tailed) | 0.440 | 0.003 | 0.000 | 0.353 | 0.001 | 0.000 | 0.022 | 0.925 | 0.160 | 0.001 | 0.013 | 0.845 | 0.108 | 0.850 |
| PM3_HOMO | Correlation Coefficient | **.369**\* | **0.194** | **-0.094** | **.471**\*\* | **0.154** | **-0.127** | **0.267** | **.291**\* | **.428**\*\* | **.577**\*\* | **.753**\*\* | **.459**\*\* | **0.008** | **-.601**\*\* |
| | Sig. (2-tailed) | 0.011 | 0.191 | 0.529 | 0.001 | 0.300 | 0.396 | 0.070 | 0.047 | 0.003 | 0.000 | 0.000 | 0.001 | 0.959 | 0.000 |
| PM3_IP | Correlation Coefficient | **-.369**\* | **-0.194** | **0.094** | **-.471**\*\* | **-0.154** | **0.127** | **-0.267** | **-.291**\* | **-.428**\*\* | **-.577**\*\* | **-.753**\*\* | **-.459**\*\* | **-0.008** | **.601**\*\* |
| | Sig. (2-tailed) | 0.011 | 0.191 | 0.529 | 0.001 | 0.300 | 0.396 | 0.070 | 0.047 | 0.003 | 0.000 | 0.000 | 0.001 | 0.959 | 0.000 |
| PM3_LUMO | Correlation Coefficient | **.289**\* | **0.132** | **-0.099** | **.381**\*\* | **0.094** | **-0.128** | **0.165** | **0.226** | **.361**\* | **.615**\*\* | **.805**\*\* | **.363**\* | **-0.096** | **-.618**\*\* |
| | Sig. (2-tailed) | 0.049 | 0.378 | 0.507 | 0.008 | 0.531 | 0.391 | 0.267 | 0.127 | 0.013 | 0.000 | 0.000 | 0.012 | 0.522 | 0.000 |
| SlogP | Correlation Coefficient | **.636**\*\* | **.669**\*\* | **.330**\* | **.845**\*\* | **.668**\*\* | **.339**\* | **.720**\*\* | **.578**\*\* | **.343**\* | **0.164** | **.649**\*\* | **.653**\*\* | **0.267** | **-.682**\*\* |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.024 | 0.000 | 0.000 | 0.020 | 0.000 | 0.000 | 0.018 | 0.270 | 0.000 | 0.000 | 0.069 | 0.000 |
| ASA_H | Correlation Coefficient | **.619**\*\* | **.763**\*\* | **.357**\* | **.953**\*\* | **.758**\*\* | **.369**\* | **.832**\*\* | **.569**\*\* | **.389**\*\* | **0.142** | **.568**\*\* | **.715**\*\* | **0.259** | **-.692**\*\* |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.014 | 0.000 | 0.000 | 0.011 | 0.000 | 0.000 | 0.007 | 0.341 | 0.000 | 0.000 | 0.078 | 0.000 |
| ASA_P | Correlation Coefficient | **-0.235** | **-.308**\* | **-0.144** | **-.403**\*\* | **-.293**\* | **-0.143** | **-0.200** | **-.334**\* | **-0.073** | **-.390**\*\* | **-.894**\*\* | **-0.237** | **0.092** | **.775**\*\* |
| | Sig. (2-tailed) | 0.112 | 0.035 | 0.335 | 0.005 | 0.046 | 0.337 | 0.177 | 0.022 | 0.626 | 0.007 | 0.000 | 0.109 | 0.540 | 0.000 |
| VSA | Correlation Coefficient | **.684**\*\* | **.820**\*\* | **.404**\*\* | **.994**\*\* | **.807**\*\* | **.404**\*\* | **.937**\*\* | **.545**\*\* | **.476**\*\* | **0.057** | **.362**\* | **.790**\*\* | **.316**\* | **-.564**\*\* |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.005 | 0.000 | 0.000 | 0.005 | 0.000 | 0.000 | 0.001 | 0.702 | 0.012 | 0.000 | 0.031 | 0.000 |
| TPSA | Correlation Coefficient | **-0.115** | **-0.067** | **0.096** | **-0.252** | **-0.048** | **0.099** | **-0.061** | **-.300**\* | **-0.213** | **-.614**\*\* | **-.880**\*\* | **-0.230** | **0.089** | **.646**\*\* |
| | Sig. (2-tailed) | 0.443 | 0.653 | 0.520 | 0.088 | 0.746 | 0.510 | 0.686 | 0.040 | 0.150 | 0.000 | 0.000 | 0.120 | 0.553 | 0.000 |
| vdw_area | Correlation Coefficient | **.669**\*\* | **.832**\*\* | **.411**\*\* | **.985**\*\* | **.817**\*\* | **.404**\*\* | **.934**\*\* | **.550**\*\* | **.493**\*\* | **0.072** | **.362**\* | **.801**\*\* | **.333**\* | **-.561**\*\* |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.004 | 0.000 | 0.000 | 0.005 | 0.000 | 0.000 | 0.000 | 0.631 | 0.012 | 0.000 | 0.022 | 0.000 |
| vdw_vol | Correlation Coefficient | **.674**\*\* | **.807**\*\* | **.384**\*\* | **.995**\*\* | **.790**\*\* | **.378**\*\* | **.944**\*\* | **.546**\*\* | **.505**\*\* | **0.078** | **.362**\* | **.814**\*\* | **.321**\* | **-.562**\*\* |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.008 | 0.000 | 0.000 | 0.009 | 0.000 | 0.000 | 0.000 | 0.602 | 0.012 | 0.000 | 0.028 | 0.000 |
| vol | Correlation Coefficient | **.675**\*\* | **.814**\*\* | **.387**\*\* | **.995**\*\* | **.797**\*\* | **.382**\*\* | **.938**\*\* | **.555**\*\* | **.490**\*\* | **0.068** | **.366**\* | **.800**\*\* | **.324**\* | **-.570**\*\* |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.007 | 0.000 | 0.000 | 0.008 | 0.000 | 0.000 | 0.000 | 0.650 | 0.011 | 0.000 | 0.026 | 0.000 |
| logP | Correlation Coefficient | **.607**\*\* | **.744**\*\* | **.439**\*\* | **.843**\*\* | **.755**\*\* | **.453**\*\* | **.731**\*\* | **.564**\*\* | **0.220** | **0.022** | **.548**\*\* | **.573**\*\* | **.306**\* | **-.592**\*\* |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.137 | 0.884 | 0.000 | 0.000 | 0.037 | 0.000 |

**Table A4-5 (cont'd)**

| | | E_vdw | logS | mr | dipole | a_acc | a_don | vsa_acc | vsa_don | vsa_hyd | vsa_pol | PM3_dipole | PM3_E | PM3_Eele |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| E_vdw | Correlation Coefficient | 1.000 | | | | | | | | | | | | |
| | Sig. (2-tailed) | . | | | | | | | | | | | | |
| logS | Correlation Coefficient | -.659** | 1.000 | | | | | | | | | | | |
| | Sig. (2-tailed) | 0.000 | . | | | | | | | | | | | |
| mr | Correlation Coefficient | .880** | -.813** | 1.000 | | | | | | | | | | |
| | Sig. (2-tailed) | 0.000 | 0.000 | . | | | | | | | | | | |
| dipole | Correlation Coefficient | -0.248 | 0.153 | -0.127 | 1.000 | | | | | | | | | |
| | Sig. (2-tailed) | 0.093 | 0.305 | 0.395 | . | | | | | | | | | |
| a_acc | Correlation Coefficient | .438** | -0.027 | .410** | 0.251 | 1.000 | | | | | | | | |
| | Sig. (2-tailed) | 0.002 | 0.855 | 0.004 | 0.089 | . | | | | | | | | |
| a_don | Correlation Coefficient | -.561** | .333* | -.485** | .576** | -0.002 | 1.000 | | | | | | | |
| | Sig. (2-tailed) | 0.000 | 0.022 | 0.001 | 0.000 | 0.990 | . | | | | | | | |
| vsa_acc | Correlation Coefficient | -0.107 | -0.060 | 0.063 | .722** | .437** | .559** | 1.000 | | | | | | |
| | Sig. (2-tailed) | 0.475 | 0.687 | 0.674 | 0.000 | 0.002 | 0.000 | . | | | | | | |
| vsa_don | Correlation Coefficient | -.525** | .301* | -.474** | .599** | 0.001 | .945** | .597** | 1.000 | | | | | |
| | Sig. (2-tailed) | 0.000 | 0.039 | 0.001 | 0.000 | 0.996 | 0.000 | 0.000 | . | | | | | |
| vsa_hyd | Correlation Coefficient | .907** | -.704** | .925** | -.337* | .378** | -.670** | -0.213 | -.673** | 1.000 | | | | |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.000 | 0.020 | 0.009 | 0.000 | 0.150 | 0.000 | . | | | | |
| vsa_pol | Correlation Coefficient | -.388** | 0.181 | -0.267 | .716** | 0.253 | .885** | .863** | .873** | -.524** | 1.000 | | | |
| | Sig. (2-tailed) | 0.007 | 0.224 | 0.069 | 0.000 | 0.086 | 0.000 | 0.000 | 0.000 | 0.000 | . | | | |
| PM3_dipole | Correlation Coefficient | -.321* | 0.171 | -0.186 | .932** | 0.200 | .689** | .682** | .721** | -.412** | .764** | 1.000 | | |
| | Sig. (2-tailed) | 0.028 | 0.249 | 0.212 | 0.000 | 0.179 | 0.000 | 0.000 | 0.000 | 0.004 | 0.000 | . | | |
| PM3_E | Correlation Coefficient | -.803** | .721** | -.883** | -0.034 | -.525** | 0.285 | -.336* | 0.260 | -.744** | -0.011 | 0.041 | 1.000 | |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.000 | 0.822 | 0.000 | 0.052 | 0.021 | 0.078 | 0.000 | 0.942 | 0.783 | . | |
| PM3_Eele | Correlation Coefficient | -.908** | .768** | -.965** | 0.092 | -.467** | .455** | -0.113 | .437** | -.889** | 0.217 | 0.169 | .950** | 1.000 |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.000 | 0.538 | 0.001 | 0.001 | 0.449 | 0.002 | 0.000 | 0.143 | 0.255 | 0.000 | . |

**Table A4-5 (cont'd)**

| | | E_vdw | logS | mr | dipole | a_acc | a_don | vsa_acc | vsa_don | vsa_hyd | vsa_pol | PM3_dipole | PM3_E | PM3_Eele |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PM3_HF | Correlation Coefficient | 0.010 | 0.216 | -0.162 | -.521** | -.323* | -0.208 | -.709** | -0.269 | 0.018 | -.486** | -.413** | .449** | 0.261 |
| | Sig. (2-tailed) | 0.947 | 0.144 | 0.276 | 0.000 | 0.027 | 0.162 | 0.000 | 0.067 | 0.906 | 0.001 | 0.004 | 0.002 | 0.077 |
| PM3_HOMO | Correlation Coefficient | .540** | -.322* | .426** | -.697** | -0.121 | -.735** | -.729** | -.760** | .616** | -.817** | -.716** | -0.128 | -.350* |
| | Sig. (2-tailed) | 0.000 | 0.028 | 0.003 | 0.000 | 0.418 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.393 | 0.016 |
| PM3_IP | Correlation Coefficient | -.540** | .322* | -.426** | .697** | 0.121 | .735** | .729** | .760** | -.616** | .817** | .716** | 0.128 | .350* |
| | Sig. (2-tailed) | 0.000 | 0.028 | 0.003 | 0.000 | 0.418 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.393 | 0.016 |
| PM3_LUMO | Correlation Coefficient | .444** | -0.254 | .335* | -.729** | -0.193 | -.772** | -.788** | -.787** | .544** | -.868** | -.748** | -0.017 | -0.255 |
| | Sig. (2-tailed) | 0.002 | 0.085 | 0.021 | 0.000 | 0.195 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.909 | 0.083 |
| SlogP | Correlation Coefficient | .755** | -.792** | .819** | -.454** | 0.019 | -.628** | -.357* | -.664** | .882** | -.581** | -.511** | -.612** | -.764** |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.000 | 0.001 | 0.898 | 0.000 | 0.014 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| ASA_H | Correlation Coefficient | .823** | -.780** | .931** | -.335* | 0.229 | -.637** | -0.198 | -.664** | .952** | -.503** | -.406** | -.726** | -.861** |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.000 | 0.021 | 0.121 | 0.000 | 0.183 | 0.000 | 0.000 | 0.000 | 0.005 | 0.000 | 0.000 |
| ASA_P | Correlation Coefficient | -.353* | .341* | -.356* | .664** | .299* | .819** | .769** | .841** | -.564** | .918** | .713** | 0.056 | 0.276 |
| | Sig. (2-tailed) | 0.015 | 0.019 | 0.014 | 0.000 | 0.041 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.710 | 0.060 |
| VSA | Correlation Coefficient | .864** | -.819** | .992** | -0.137 | .414** | -.485** | 0.041 | -.482** | .938** | -0.281 | -0.198 | -.863** | -.948** |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.000 | 0.358 | 0.004 | 0.001 | 0.782 | 0.001 | 0.000 | 0.056 | 0.183 | 0.000 | 0.000 |
| TPSA | Correlation Coefficient | -.314* | 0.145 | -0.211 | .715** | .346* | .850** | .871** | .858** | -.453** | .984** | .756** | -0.071 | 0.154 |
| | Sig. (2-tailed) | 0.031 | 0.329 | 0.155 | 0.000 | 0.017 | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.635 | 0.302 |
| vdw_area | Correlation Coefficient | .874** | -.794** | .983** | -0.140 | .432** | -.502** | 0.049 | -.490** | .944** | -.288* | -0.213 | -.862** | -.947** |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.000 | 0.348 | 0.002 | 0.000 | 0.744 | 0.000 | 0.000 | 0.050 | 0.151 | 0.000 | 0.000 |
| vdw_vol | Correlation Coefficient | .882** | -.810** | .995** | -0.135 | .411** | -.498** | 0.040 | -.484** | .941** | -.289* | -0.201 | -.872** | -.959** |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.000 | 0.365 | 0.004 | 0.000 | 0.791 | 0.001 | 0.000 | 0.049 | 0.176 | 0.000 | 0.000 |
| vol | Correlation Coefficient | .875** | -.808** | .993** | -0.139 | .414** | -.501** | 0.037 | -.489** | .940** | -.293* | -0.202 | -.864** | -.951** |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.000 | 0.351 | 0.004 | 0.000 | 0.804 | 0.000 | 0.000 | 0.046 | 0.174 | 0.000 | 0.000 |
| logP | Correlation Coefficient | .679** | -.861** | .817** | -.377** | 0.011 | -.532** | -0.222 | -.565** | .818** | -.452** | -.444** | -.636** | -.750** |
| | Sig. (2-tailed) | 0.000 | 0.000 | 0.000 | 0.009 | 0.942 | 0.000 | 0.134 | 0.000 | 0.000 | 0.001 | 0.002 | 0.000 | 0.000 |

**Table A4-5 (cont'd)**

|  |  | PM3_HF | PM3_HOMO | PM3_IP | PM3_LUMO | SlogP | ASA_H | ASA_P | VSA | TPSA | vdw_area | vdw_vol | vol | logP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PM3_HF | Correlation Coefficient | 1.000 |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Sig. (2-tailed) | . |  |  |  |  |  |  |  |  |  |  |  |  |
| PM3_HOMO | Correlation Coefficient | .610** | 1.000 |  |  |  |  |  |  |  |  |  |  |  |
|  | Sig. (2-tailed) | 0.000 | . |  |  |  |  |  |  |  |  |  |  |  |
| PM3_IP | Correlation Coefficient | -.610** | -1.000** | 1.000 |  |  |  |  |  |  |  |  |  |  |
|  | Sig. (2-tailed) | 0.000 | . | . |  |  |  |  |  |  |  |  |  |  |
| PM3_LUMO | Correlation Coefficient | .633** | .956** | -.956** | 1.000 |  |  |  |  |  |  |  |  |  |
|  | Sig. (2-tailed) | 0.000 | 0.000 | 0.000 | . |  |  |  |  |  |  |  |  |  |
| SlogP | Correlation Coefficient | 0.100 | .658** | -.658** | .601** | 1.000 |  |  |  |  |  |  |  |  |
|  | Sig. (2-tailed) | 0.503 | 0.000 | 0.000 | 0.000 | . |  |  |  |  |  |  |  |  |
| ASA_H | Correlation Coefficient | -0.013 | .605** | -.605** | .530** | .924** | 1.000 |  |  |  |  |  |  |  |
|  | Sig. (2-tailed) | 0.929 | 0.000 | 0.000 | 0.000 | 0.000 | . |  |  |  |  |  |  |  |
| ASA_P | Correlation Coefficient | -.375** | -.766** | .766** | -.816** | -.693** | -.617** | 1.000 |  |  |  |  |  |  |
|  | Sig. (2-tailed) | 0.009 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | . |  |  |  |  |  |  |
| VSA | Correlation Coefficient | -0.162 | .433** | -.433** | .343* | .833** | .941** | -.373** | 1.000 |  |  |  |  |  |
|  | Sig. (2-tailed) | 0.277 | 0.002 | 0.002 | 0.018 | 0.000 | 0.000 | 0.010 | . |  |  |  |  |  |
| TPSA | Correlation Coefficient | -.526** | -.799** | .799** | -.857** | -.550** | -.459** | .930** | -0.216 | 1.000 |  |  |  |  |
|  | Sig. (2-tailed) | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 | 0.145 | . |  |  |  |  |
| vdw_area | Correlation Coefficient | -0.161 | .439** | -.439** | .346* | .819** | .931** | -.369* | .991** | -0.221 | 1.000 |  |  |  |
|  | Sig. (2-tailed) | 0.278 | 0.002 | 0.002 | 0.017 | 0.000 | 0.000 | 0.011 | 0.000 | 0.135 | . |  |  |  |
| vdw_vol | Correlation Coefficient | -0.160 | .440** | -.440** | .349* | .833** | .938** | -.371* | .997** | -0.224 | .991** | 1.000 |  |  |
|  | Sig. (2-tailed) | 0.282 | 0.002 | 0.002 | 0.016 | 0.000 | 0.000 | 0.010 | 0.000 | 0.130 | 0.000 | . |  |  |
| vol | Correlation Coefficient | -0.155 | .444** | -.444** | .349* | .833** | .943** | -.378** | .997** | -0.228 | .992** | .998** | 1.000 |  |
|  | Sig. (2-tailed) | 0.298 | 0.002 | 0.002 | 0.016 | 0.000 | 0.000 | 0.009 | 0.000 | 0.123 | 0.000 | 0.000 | . |  |
| logP | Correlation Coefficient | -0.049 | .528** | -.528** | .468** | .939** | .898** | -.601** | .833** | -.429** | .812** | .824** | .824** | 1.000 |
|  | Sig. (2-tailed) | 0.742 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.003 | 0.000 | 0.000 | 0.000 | . |

**Table A4-6. Spearman Rho Correlation Coefficient of Descriptors employed in developing Equations 5-3 and 5-5 (protonated compounds)**

| | | ICMT | b_1rotN | b_1rotR | b_count | b_rotN | b_rotR | weight | lip_violation | E | E_ele | E_sol | E_str | E_strain | E_tor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| b_1rotN | Correlation Coefficient | .601** | 1 | | | | | | | | | | | | |
| | Sig. (2-tailed) | 0 | . | | | | | | | | | | | | |
| b_1rotR | Correlation Coefficient | .387** | .801** | 1 | | | | | | | | | | | |
| | Sig. (2-tailed) | 0.007 | 0 | . | | | | | | | | | | | |
| b_count | Correlation Coefficient | .670** | .798** | .373** | 1 | | | | | | | | | | |
| | Sig. (2-tailed) | 0 | 0 | 0.01 | . | | | | | | | | | | |
| b_rotN | Correlation Coefficient | .528** | .979** | .790** | .779** | 1 | | | | | | | | | |
| | Sig. (2-tailed) | 0 | 0 | 0 | 0 | . | | | | | | | | | |
| b_rotR | Correlation Coefficient | .307* | .778** | .972** | .366* | .813** | 1 | | | | | | | | |
| | Sig. (2-tailed) | 0.036 | 0 | 0 | 0.011 | 0 | . | | | | | | | | |
| weight | Correlation Coefficient | .600** | .767** | .337* | .919** | .754** | .334* | 1 | | | | | | | |
| | Sig. (2-tailed) | 0 | 0 | 0.021 | 0 | 0 | 0.022 | . | | | | | | | |
| lip_violation | Correlation Coefficient | .399** | .534** | .309* | .490** | .548** | .313* | .503** | 1 | | | | | | |
| | Sig. (2-tailed) | 0.006 | 0 | 0.034 | 0 | 0 | 0.032 | 0 | . | | | | | | |
| E | Correlation Coefficient | 0.178 | 0.154 | -0.257 | .488** | 0.135 | -0.27 | .563** | 0.191 | 1 | | | | | |
| | Sig. (2-tailed) | 0.232 | 0.302 | 0.081 | 0.001 | 0.367 | 0.067 | 0 | 0.198 | . | | | | | |
| E_ele | Correlation Coefficient | -.379** | -.386** | -.576** | -0.071 | -.361* | -.529** | -0.005 | -0.243 | .583** | 1 | | | | |
| | Sig. (2-tailed) | 0.009 | 0.007 | 0 | 0.636 | 0.013 | 0 | 0.974 | 0.1 | 0 | . | | | | |
| E_sol | Correlation Coefficient | -.302* | 0.088 | 0.151 | -0.086 | 0.167 | 0.222 | 0.07 | 0.239 | -0.01 | 0.033 | 1 | | | |
| | Sig. (2-tailed) | 0.039 | 0.556 | 0.31 | 0.567 | 0.263 | 0.134 | 0.64 | 0.105 | 0.946 | 0.825 | . | | | |
| E_str | Correlation Coefficient | .529** | .475** | 0.035 | .810** | .443** | 0.007 | .823** | .399** | .767** | 0.253 | -0.126 | 1 | | |
| | Sig. (2-tailed) | 0 | 0.001 | 0.816 | 0 | 0.002 | 0.963 | 0 | 0.006 | 0 | 0.086 | 0.399 | . | | |
| E_strain | Correlation Coefficient | 0.205 | .294* | 0.183 | .295* | .298* | 0.203 | .303* | 0.223 | 0.054 | -0.094 | 0.039 | 0.133 | 1 | |
| | Sig. (2-tailed) | 0.167 | 0.045 | 0.218 | 0.044 | 0.042 | 0.172 | 0.038 | 0.131 | 0.72 | 0.53 | 0.795 | 0.375 | . | |
| E_tor | Correlation Coefficient | -.302* | -.646** | -.505** | -.668** | -.647** | -.540** | -.536** | -0.175 | -0.019 | 0.108 | 0.147 | -.395** | -0.123 | 1 |
| | Sig. (2-tailed) | 0.039 | 0 | 0 | 0 | 0 | 0 | 0 | 0.238 | 0.899 | 0.47 | 0.325 | 0.006 | 0.41 | . |

**Table A4-6 (cont'd)**

| | | ICMT | b_1rotN | b_1rotR | b_count | b_rotN | b_rotR | weight | lip_violation | E | E_ele | E_sol | E_str | E_strain | E_tor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| E_vdw | Correlation Coefficient | .554** | .589** | 0.141 | .886** | .564** | 0.123 | .883** | .446** | .738** | 0.145 | -0.093 | .945** | 0.179 | -.490** |
| | Sig. (2-tailed) | 0 | 0 | 0.345 | 0 | 0 | 0.411 | 0 | 0.002 | 0 | 0.33 | 0.536 | 0 | 0.229 | 0 |
| logS | Correlation Coefficient | -.543** | -.734** | -.417** | -.778** | -.756** | -.439** | -.760** | -.634** | -.325* | 0.147 | -0.279 | -.648** | -0.11 | .506** |
| | Sig. (2-tailed) | 0 | 0 | 0.004 | 0 | 0 | 0.002 | 0 | 0 | 0.026 | 0.325 | 0.057 | 0 | 0.462 | 0 |
| mr | Correlation Coefficient | .655** | .790** | .352* | .986** | .776** | .350* | .950** | .514** | .537** | -0.02 | 0.011 | .826** | .320* | -.618** |
| | Sig. (2-tailed) | 0 | 0 | 0.015 | 0 | 0 | 0.016 | 0 | 0 | 0 | 0.893 | 0.941 | 0 | 0.028 | 0 |
| dipole | Correlation Coefficient | .442** | 0.177 | 0.107 | .351* | 0.098 | 0.041 | 0.141 | -0.072 | 0.034 | -0.156 | -.780** | .289* | 0.018 | -.459** |
| | Sig. (2-tailed) | 0.002 | 0.234 | 0.475 | 0.016 | 0.514 | 0.786 | 0.343 | 0.632 | 0.821 | 0.294 | 0 | 0.049 | 0.904 | 0.001 |
| a_acc | Correlation Coefficient | -0.119 | 0.167 | 0.187 | -0.025 | 0.195 | 0.21 | 0.207 | 0.02 | 0.1 | -0.023 | .554** | -0.05 | .361* | 0.268 |
| | Sig. (2-tailed) | 0.426 | 0.261 | 0.208 | 0.87 | 0.188 | 0.156 | 0.163 | 0.894 | 0.504 | 0.878 | 0 | 0.741 | 0.013 | 0.068 |
| a_base | Correlation Coefficient | .332* | 0.058 | -0.105 | .330* | 0.001 | -0.148 | 0.124 | -0.075 | 0.148 | 0.02 | -.856** | .328* | -0.124 | -.413** |
| | Sig. (2-tailed) | 0.023 | 0.699 | 0.484 | 0.023 | 0.996 | 0.32 | 0.405 | 0.619 | 0.322 | 0.893 | 0 | 0.024 | 0.407 | 0.004 |
| a_don | Correlation Coefficient | -0.21 | -.299* | -0.032 | -.510** | -0.275 | -0.032 | -.344* | -0.202 | -.354* | -0.278 | 0.262 | -.438** | 0.11 | .617** |
| | Sig. (2-tailed) | 0.157 | 0.041 | 0.832 | 0 | 0.062 | 0.832 | 0.018 | 0.173 | 0.015 | 0.059 | 0.075 | 0.002 | 0.462 | 0 |
| vsa_acc | Correlation Coefficient | -0.108 | 0.183 | 0.219 | -0.042 | 0.216 | 0.245 | 0.192 | 0.046 | 0.031 | -0.075 | .604** | -0.093 | .351* | 0.264 |
| | Sig. (2-tailed) | 0.471 | 0.218 | 0.138 | 0.781 | 0.145 | 0.096 | 0.196 | 0.761 | 0.837 | 0.614 | 0 | 0.534 | 0.016 | 0.073 |
| vsa_don | Correlation Coefficient | -0.15 | -.343* | -0.042 | -.564** | -.339* | -0.064 | -.405** | -0.206 | -.361* | -.296* | 0.172 | -.469** | 0.056 | .648** |
| | Sig. (2-tailed) | 0.315 | 0.018 | 0.777 | 0 | 0.02 | 0.669 | 0.005 | 0.164 | 0.013 | 0.043 | 0.249 | 0.001 | 0.71 | 0 |
| vsa_hyd | Correlation Coefficient | .649** | .733** | .329* | .959** | .715** | .321* | .849** | .459** | .529** | -0.012 | -0.206 | .826** | 0.221 | -.668** |
| | Sig. (2-tailed) | 0 | 0 | 0.024 | 0 | 0 | 0.028 | 0 | 0.001 | 0 | 0.938 | 0.166 | 0 | 0.135 | 0 |
| vsa_pol | Correlation Coefficient | -0.152 | -0.115 | 0.068 | -.355* | -0.098 | 0.067 | -0.13 | -0.105 | -0.194 | -0.203 | .414** | -.326* | 0.221 | .542** |
| | Sig. (2-tailed) | 0.308 | 0.44 | 0.65 | 0.014 | 0.513 | 0.654 | 0.383 | 0.481 | 0.192 | 0.171 | 0.004 | 0.026 | 0.135 | 0 |
| PM3_dipole | Correlation Coefficient | .449** | 0.098 | 0.039 | 0.282 | 0.012 | -0.037 | 0.085 | -0.104 | 0.052 | -0.139 | -.843** | 0.251 | 0.044 | -.332* |
| | Sig. (2-tailed) | 0.002 | 0.512 | 0.794 | 0.055 | 0.935 | 0.803 | 0.571 | 0.488 | 0.728 | 0.351 | 0 | 0.088 | 0.77 | 0.023 |
| PM3_E | Correlation Coefficient | -.492** | -.706** | -.294* | -.829** | -.707** | -.305* | -.973** | -.490** | -.569** | -0.051 | -0.175 | -.769** | -.304* | .434** |
| | Sig. (2-tailed) | 0 | 0 | 0.045 | 0 | 0 | 0.037 | 0 | 0 | 0 | 0.734 | 0.24 | 0 | 0.038 | 0.002 |
| PM3_Eele | Correlation Coefficient | -.604** | -.740** | -.297* | -.936** | -.725** | -.296* | -.986** | -.486** | -.605** | -0.047 | -0.013 | -.869** | -0.267 | .573** |
| | Sig. (2-tailed) | 0 | 0 | 0.043 | 0 | 0 | 0.044 | 0 | 0.001 | 0 | 0.752 | 0.932 | 0 | 0.069 | 0 |

**Table A4-6 (cont'd)**

| | | ICMT | b_1rotN | b_1rotR | b_count | b_rotN | b_rotR | weight | lip_violation | E | E_ele | E_sol | E_str | E_strain | E_tor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PM3_HF | Correlation Coefficient | **0.273** | **-0.185** | **-.322<sup>*</sup>** | **0.127** | **-0.252** | **-.384<sup>**</sup>** | **-0.131** | **-0.076** | **0.116** | **0.083** | **-.788<sup>**</sup>** | **0.217** | **-0.214** | **-0.12** |
| | Sig. (2-tailed) | 0.063 | 0.213 | 0.027 | 0.393 | 0.088 | 0.008 | 0.378 | 0.613 | 0.438 | 0.581 | 0 | 0.143 | 0.149 | 0.422 |
| PM3_HOMO | Correlation Coefficient | **-0.21** | **0.115** | **0.088** | **-0.022** | **0.166** | **0.134** | **0.097** | **0.283** | **0.107** | **0.154** | **.863<sup>**</sup>** | **-0.049** | **0.155** | **0.191** |
| | Sig. (2-tailed) | 0.157 | 0.441 | 0.558 | 0.883 | 0.264 | 0.368 | 0.515 | 0.054 | 0.474 | 0.301 | 0 | 0.743 | 0.299 | 0.199 |
| PM3_IP | Correlation Coefficient | **0.21** | **-0.115** | **-0.088** | **0.022** | **-0.166** | **-0.134** | **-0.097** | **-0.283** | **-0.107** | **-0.154** | **-.863<sup>**</sup>** | **0.049** | **-0.155** | **-0.19** |
| | Sig. (2-tailed) | 0.157 | 0.441 | 0.558 | 0.883 | 0.264 | 0.368 | 0.515 | 0.054 | 0.474 | 0.301 | 0 | 0.743 | 0.299 | 0.199 |
| PM3_LUMO | Correlation Coefficient | **-0.248** | **0.063** | **0.115** | **-0.113** | **0.116** | **0.161** | **-0.009** | **0.191** | **0.015** | **0.121** | **.887<sup>**</sup>** | **-0.153** | **0.002** | **0.184** |
| | Sig. (2-tailed) | 0.093 | 0.674 | 0.442 | 0.449 | 0.438 | 0.279 | 0.954 | 0.198 | 0.92 | 0.419 | 0 | 0.306 | 0.989 | 0.215 |
| SlogP | Correlation Coefficient | **.454<sup>**</sup>** | **.700<sup>**</sup>** | **.442<sup>**</sup>** | **.683<sup>**</sup>** | **.742<sup>**</sup>** | **.482<sup>**</sup>** | **.689<sup>**</sup>** | **.682<sup>**</sup>** | **0.204** | **-0.214** | **.460<sup>**</sup>** | **.498<sup>**</sup>** | **0.091** | **-.462<sup>**</sup>** |
| | Sig. (2-tailed) | 0.001 | 0 | 0.002 | 0 | 0 | 0.001 | 0 | 0 | 0.169 | 0.148 | 0.001 | 0 | 0.543 | 0.001 |
| ASA_H | Correlation Coefficient | **.639<sup>**</sup>** | **.784<sup>**</sup>** | **.396<sup>**</sup>** | **.944<sup>**</sup>** | **.782<sup>**</sup>** | **.407<sup>**</sup>** | **.821<sup>**</sup>** | **.554<sup>**</sup>** | **.351<sup>*</sup>** | **-0.129** | **-0.042** | **.710<sup>**</sup>** | **0.207** | **-.718<sup>**</sup>** |
| | Sig. (2-tailed) | 0 | 0 | 0.006 | 0 | 0 | 0.004 | 0 | 0 | 0.016 | 0.388 | 0.779 | 0 | 0.163 | 0 |
| ASA_P | Correlation Coefficient | **-0.2** | **-.309<sup>*</sup>** | **-0.148** | **-.416<sup>**</sup>** | **-.300<sup>*</sup>** | **-0.154** | **-0.199** | **-0.247** | **-0.021** | **0.039** | **0.224** | **-0.258** | **0.203** | **.634<sup>**</sup>** |
| | Sig. (2-tailed) | 0.178 | 0.034 | 0.322 | 0.004 | 0.04 | 0.301 | 0.181 | 0.094 | 0.889 | 0.793 | 0.129 | 0.08 | 0.172 | 0 |
| VSA | Correlation Coefficient | **.689<sup>**</sup>** | **.822<sup>**</sup>** | **.408<sup>**</sup>** | **.990<sup>**</sup>** | **.808<sup>**</sup>** | **.405<sup>**</sup>** | **.938<sup>**</sup>** | **.502<sup>**</sup>** | **.487<sup>**</sup>** | **-0.071** | **-0.019** | **.802<sup>**</sup>** | **.317<sup>*</sup>** | **-.648<sup>**</sup>** |
| | Sig. (2-tailed) | 0 | 0 | 0.004 | 0 | 0 | 0.005 | 0 | 0 | 0.001 | 0.636 | 0.897 | 0 | 0.03 | 0 |
| TPSA | Correlation Coefficient | **-0.105** | **-0.071** | **0.095** | **-.292<sup>*</sup>** | **-0.054** | **0.095** | **-0.073** | **-0.145** | **-0.146** | **-0.194** | **.347<sup>*</sup>** | **-0.263** | **0.241** | **.483<sup>**</sup>** |
| | Sig. (2-tailed) | 0.482 | 0.637 | 0.526 | 0.047 | 0.72 | 0.524 | 0.627 | 0.332 | 0.329 | 0.192 | 0.017 | 0.074 | 0.102 | 0.001 |
| vdw_area | Correlation Coefficient | **.663<sup>**</sup>** | **.816<sup>**</sup>** | **.396<sup>**</sup>** | **.989<sup>**</sup>** | **.799<sup>**</sup>** | **.389<sup>**</sup>** | **.924<sup>**</sup>** | **.487<sup>**</sup>** | **.509<sup>**</sup>** | **-0.056** | **-0.078** | **.820<sup>**</sup>** | **.311<sup>*</sup>** | **-.658<sup>**</sup>** |
| | Sig. (2-tailed) | 0 | 0 | 0.006 | 0 | 0 | 0.007 | 0 | 0.001 | 0 | 0.71 | 0.602 | 0 | 0.033 | 0 |
| vdw_vol | Correlation Coefficient | **.680<sup>**</sup>** | **.798<sup>**</sup>** | **.365<sup>*</sup>** | **.992<sup>**</sup>** | **.782<sup>**</sup>** | **.361<sup>*</sup>** | **.940<sup>**</sup>** | **.503<sup>**</sup>** | **.517<sup>**</sup>** | **-0.048** | **-0.045** | **.824<sup>**</sup>** | **.307<sup>*</sup>** | **-.644<sup>**</sup>** |
| | Sig. (2-tailed) | 0 | 0 | 0.012 | 0 | 0 | 0.013 | 0 | 0 | 0 | 0.747 | 0.765 | 0 | 0.036 | 0 |
| vol | Correlation Coefficient | **.679<sup>**</sup>** | **.809<sup>**</sup>** | **.378<sup>**</sup>** | **.992<sup>**</sup>** | **.790<sup>**</sup>** | **.370<sup>*</sup>** | **.939<sup>**</sup>** | **.502<sup>**</sup>** | **.505<sup>**</sup>** | **-0.061** | **-0.035** | **.818<sup>**</sup>** | **.308<sup>*</sup>** | **-.645<sup>**</sup>** |
| | Sig. (2-tailed) | 0 | 0 | 0.009 | 0 | 0 | 0.011 | 0 | 0 | 0 | 0.686 | 0.817 | 0 | 0.035 | 0 |
| logP | Correlation Coefficient | **.607<sup>**</sup>** | **.744<sup>**</sup>** | **.439<sup>**</sup>** | **.842<sup>**</sup>** | **.755<sup>**</sup>** | **.453<sup>**</sup>** | **.733<sup>**</sup>** | **.606<sup>**</sup>** | **0.212** | **-0.22** | **0.067** | **.602<sup>**</sup>** | **0.092** | **-.631<sup>**</sup>** |
| | Sig. (2-tailed) | 0 | 0 | 0.002 | 0 | 0 | 0.001 | 0 | 0 | 0.152 | 0.137 | 0.656 | 0 | 0.537 | 0 |

**Table A4-6 (cont'd)**

| | | E_vdw | logS | mr | dipole | a_acc | a_base | a_don | vsa_acc | vsa_don | vsa_hyd | vsa_pol | PM3_dipole | PM3_E | PM3_Eele |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| E_vdw | Correlation Coefficient | 1 | | | | | | | | | | | | | |
| | Sig. (2-tailed) | . | | | | | | | | | | | | | |
| logS | Correlation Coefficient | -.670** | 1 | | | | | | | | | | | | |
| | Sig. (2-tailed) | 0 | . | | | | | | | | | | | | |
| mr | Correlation Coefficient | .890** | -.805** | 1 | | | | | | | | | | | |
| | Sig. (2-tailed) | 0 | 0 | . | | | | | | | | | | | |
| dipole | Correlation Coefficient | .305* | 0.002 | 0.247 | 1 | | | | | | | | | | |
| | Sig. (2-tailed) | 0.037 | 0.99 | 0.094 | . | | | | | | | | | | |
| a_acc | Correlation Coefficient | -0.005 | -0.05 | 0.071 | -.578** | 1 | | | | | | | | | |
| | Sig. (2-tailed) | 0.973 | 0.75 | 0.633 | 0 | . | | | | | | | | | |
| a_base | Correlation Coefficient | .331* | 0.025 | 0.221 | .856** | -.754** | 1 | | | | | | | | |
| | Sig. (2-tailed) | 0.023 | 0.87 | 0.136 | 0 | 0 | . | | | | | | | | |
| a_don | Correlation Coefficient | -.504** | .295* | -.451** | -.497** | .569** | -.641** | 1 | | | | | | | |
| | Sig. (2-tailed) | 0 | 0.044 | 0.001 | 0 | 0 | 0 | . | | | | | | | |
| vsa_acc | Correlation Coefficient | -0.049 | -0.07 | 0.055 | -.607** | .981** | -.811** | .614** | 1 | | | | | | |
| | Sig. (2-tailed) | 0.743 | 0.658 | 0.715 | 0 | 0 | 0 | 0 | . | | | | | | |
| vsa_don | Correlation Coefficient | -.534** | .329* | -.516** | -.406** | .493** | -.571** | .938** | .537** | 1 | | | | | |
| | Sig. (2-tailed) | 0 | 0.024 | 0 | 0.005 | 0 | 0 | 0 | 0 | . | | | | | |
| vsa_hyd | Correlation Coefficient | .896** | -.699** | .929** | .468** | -0.18 | .492** | -.650** | -0.213 | -.693** | 1 | | | | |
| | Sig. (2-tailed) | 0 | 0 | 0 | 0.001 | 0.224 | 0 | 0 | 0.15 | 0 | . | | | | |
| vsa_pol | Correlation Coefficient | -.345* | 0.175 | -0.27 | -.569** | .835** | -.766** | .876** | .863** | .879** | -.524** | 1 | | | |
| | Sig. (2-tailed) | 0.018 | 0.239 | 0.062 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | . | | | |
| PM3_dipole | Correlation Coefficient | 0.238 | 0.045 | 0.187 | .952** | -.575** | .856** | -.388** | -.608** | -.288* | .396** | -.503** | 1 | | |
| | Sig. (2-tailed) | 0.108 | 0.763 | 0.208 | 0 | 0 | 0 | 0.007 | 0 | 0.05 | 0.006 | 0 | . | | |
| PM3_E | Correlation Coefficient | -.828** | .712** | -.877** | -0.01 | -.349* | 0.015 | 0.242 | -.336* | .305* | -.744** | -0.011 | 0.049 | 1 | |
| | Sig. (2-tailed) | 0 | 0 | 0 | 0.946 | 0.016 | 0.921 | 0.101 | 0.021 | 0.037 | 0 | 0.942 | 0.742 | . | |
| PM3_Eele | Correlation Coefficient | -.926** | .756** | -.962** | -0.22 | -0.13 | -0.205 | .424** | -0.102 | .480** | -.891** | 0.225 | -0.153 | .947** | 1 |
| | Sig. (2-tailed) | 0 | 0 | 0 | 0.132 | 0.403 | 0.167 | 0.003 | 0.495 | 0.001 | 0 | 0.128 | 0.305 | 0 | . |

**Table A4-6 (cont'd)**

| | | E_vdw | logS | mr | dipole | a_acc | a_base | a_don | vsa_acc | vsa_don | vsa_hyd | vsa_pol | PM3_dipole | PM3_E | PM3_Eele |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PM3_HF | Correlation Coefficient | **0.153** | **0.132** | **0.033** | **.678**** | **-.809**** | **.856**** | **-.482**** | **-.849**** | **-.411**** | **0.278** | **-.695**** | **.723**** | **0.285** | **0.042** |
| | Sig. (2-tailed) | 0.304 | 0.376 | 0.827 | 0 | 0 | 0 | 0.001 | 0 | 0.004 | 0.059 | 0 | 0 | 0.052 | 0.778 |
| PM3_HOMO | Correlation Coefficient | **-0.037** | **-0.28** | **0.08** | **-.801**** | **.577**** | **-.856**** | **0.271** | **.631**** | **0.203** | **-0.168** | **.450**** | **-.824**** | **-0.19** | **-0.047** |
| | Sig. (2-tailed) | 0.803 | 0.053 | 0.592 | 0 | 0 | 0 | 0.066 | 0 | 0.171 | 0.26 | 0.002 | 0 | 0.202 | 0.755 |
| PM3_IP | Correlation Coefficient | **0.037** | **0.284** | **-0.08** | **.801**** | **-.577**** | **.856**** | **-0.27** | **-.631**** | **-0.203** | **0.168** | **-.450**** | **.824**** | **0.19** | **0.047** |
| | Sig. (2-tailed) | 0.803 | 0.053 | 0.592 | 0 | 0 | 0 | 0.066 | 0 | 0.171 | 0.26 | 0.002 | 0 | 0.202 | 0.755 |
| PM3_LUMO | Correlation Coefficient | **-0.141** | **-0.25** | **-0.02** | **-.816**** | **.497**** | **-.856**** | **0.268** | **.557**** | **0.198** | **-0.247** | **.398**** | **-.848**** | **-0.077** | **0.051** |
| | Sig. (2-tailed) | 0.344 | 0.093 | 0.919 | 0 | 0 | 0 | 0.069 | 0 | 0.183 | 0.094 | 0.006 | 0 | 0.608 | 0.731 |
| SlogP | Correlation Coefficient | **.570**** | **-.894**** | **.719**** | **-0.18** | **0.078** | **-0.169** | **-.316*** | **0.109** | **-.363*** | **.618**** | **-0.168** | **-0.255** | **-.656**** | **-.677**** |
| | Sig. (2-tailed) | 0 | 0 | 0 | 0.233 | 0.603 | 0.256 | 0.031 | 0.466 | 0.012 | 0 | 0.259 | 0.084 | 0 | 0 |
| ASA_H | Correlation Coefficient | **.791**** | **-.808**** | **.922**** | **.344*** | **-0.19** | **.356*** | **-.620**** | **-0.197** | **-.668**** | **.940**** | **-.501**** | **0.257** | **-.710**** | **-.849**** |
| | Sig. (2-tailed) | 0 | 0 | 0 | 0.018 | 0.194 | 0.014 | 0 | 0.185 | 0 | 0 | 0 | 0.081 | 0 | 0 |
| ASA_P | Correlation Coefficient | **-.301*** | **.340*** | **-.351*** | **-.447**** | **.735**** | **-.606**** | **.810**** | **.745**** | **.842**** | **-.548**** | **.909**** | **-.356*** | **0.05** | **0.272** |
| | Sig. (2-tailed) | 0.04 | 0.019 | 0.015 | 0.002 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.014 | 0.736 | 0.064 |
| VSA | Correlation Coefficient | **.871**** | **-.813**** | **.993**** | **0.287** | **0.049** | **0.247** | **-.454**** | **0.039** | **-.508**** | **.937**** | **-0.281** | **0.226** | **-.860**** | **-.947**** |
| | Sig. (2-tailed) | 0 | 0 | 0 | 0.051 | 0.744 | 0.094 | 0.001 | 0.796 | 0 | 0 | 0.056 | 0.126 | 0 | 0 |
| TPSA | Correlation Coefficient | **-0.278** | **0.147** | **-0.22** | **-.499**** | **.840**** | **-.702**** | **.854**** | **.866**** | **.858**** | **-.456**** | **.985**** | **-.433**** | **-0.068** | **0.164** |
| | Sig. (2-tailed) | 0.058 | 0.324 | 0.138 | 0 | 0 | 0 | 0 | 0 | 0 | 0.001 | 0 | 0.002 | 0.652 | 0.27 |
| vdw_area | Correlation Coefficient | **.891**** | **-.765**** | **.982**** | **.338*** | **0.006** | **.315*** | **-.509**** | **-0.011** | **-.574**** | **.966**** | **-.346*** | **0.274** | **-.838**** | **-.939**** |
| | Sig. (2-tailed) | 0 | 0 | 0 | 0.02 | 0.969 | 0.031 | 0 | 0.941 | 0 | 0 | 0.017 | 0.062 | 0 | 0 |
| vdw_vol | Correlation Coefficient | **.895**** | **-.792**** | **.993**** | **.308*** | **0.02** | **0.281** | **-.489**** | **0.009** | **-.542**** | **.951**** | **-.316*** | **0.246** | **-.862**** | **-.954**** |
| | Sig. (2-tailed) | 0 | 0 | 0 | 0.035 | 0.893 | 0.056 | 0 | 0.951 | 0 | 0 | 0.03 | 0.096 | 0 | 0 |
| vol | Correlation Coefficient | **.887**** | **-.800**** | **.993**** | **.295*** | **0.034** | **0.267** | **-.474**** | **0.023** | **-.534**** | **.943**** | **-.305*** | **0.234** | **-.859**** | **-.949**** |
| | Sig. (2-tailed) | 0 | 0 | 0 | 0.044 | 0.82 | 0.07 | 0.001 | 0.876 | 0 | 0 | 0.037 | 0.113 | 0 | 0 |
| logP | Correlation Coefficient | **.675**** | **-.859**** | **.816**** | **0.251** | **-0.23** | **0.263** | **-.539**** | **-0.222** | **-.557**** | **.818**** | **-.452**** | **0.165** | **-.631**** | **-.746**** |
| | Sig. (2-tailed) | 0 | 0 | 0 | 0.089 | 0.122 | 0.074 | 0 | 0.134 | 0 | 0 | 0.001 | 0.269 | 0 | 0 |

| | | PM3_HF | PM3_HOMO | PM3_IP | PM3_LUMO | SlogP | ASA_H | ASA_P | VSA | TPSA | vdw_area | vdw_vol | vol | logP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PM3_HF | Correlation Coefficient | 1 | | | | | | | | | | | | |
| | Sig. (2-tailed) | . | | | | | | | | | | | | |
| PM3_HOMO | Correlation Coefficient | -.675** | 1 | | | | | | | | | | | |
| | Sig. (2-tailed) | 0 | . | | | | | | | | | | | |
| PM3_IP | Correlation Coefficient | .675** | -1.000** | 1 | | | | | | | | | | |
| | Sig. (2-tailed) | 0 | . | . | | | | | | | | | | |
| PM3_LUMO | Correlation Coefficient | -.662** | .953** | -.953** | 1 | | | | | | | | | |
| | Sig. (2-tailed) | 0 | 0 | 0 | . | | | | | | | | | |
| SlogP | Correlation Coefficient | -0.243 | .432** | -.432** | .416** | 1 | | | | | | | | |
| | Sig. (2-tailed) | 0.1 | 0.002 | 0.002 | 0.004 | . | | | | | | | | |
| ASA_H | Correlation Coefficient | 0.182 | -0.006 | 0.006 | -0.066 | .780** | 1 | | | | | | | |
| | Sig. (2-tailed) | 0.22 | 0.967 | 0.967 | 0.658 | 0 | . | | | | | | | |
| ASA_P | Correlation Coefficient | -.530** | 0.286 | -0.286 | 0.224 | -.403** | -.628** | 1 | | | | | | |
| | Sig. (2-tailed) | 0 | 0.051 | 0.051 | 0.13 | 0.005 | 0 | . | | | | | | |
| VSA | Correlation Coefficient | 0.044 | 0.045 | -0.045 | -0.047 | .718** | .931** | -.363* | 1 | | | | | |
| | Sig. (2-tailed) | 0.769 | 0.762 | 0.762 | 0.754 | 0 | 0 | 0.012 | . | | | | | |
| TPSA | Correlation Coefficient | -.673** | .381** | -.381** | .330* | -0.19 | -.464** | .924** | -0.22 | 1 | | | | |
| | Sig. (2-tailed) | 0 | 0.008 | 0.008 | 0.023 | 0.214 | 0.001 | 0 | 0.139 | . | | | | |
| vdw_area | Correlation Coefficient | 0.105 | -0.016 | 0.016 | -0.113 | .679** | .936** | -.410** | .987** | -0.28 | 1 | | | |
| | Sig. (2-tailed) | 0.483 | 0.914 | 0.914 | 0.45 | 0 | 0 | 0.004 | 0 | 0.056 | . | | | |
| vdw_vol | Correlation Coefficient | 0.077 | 0.021 | -0.021 | -0.073 | .706** | .934** | -.381** | .996** | -0.25 | .992** | 1 | | |
| | Sig. (2-tailed) | 0.605 | 0.89 | 0.89 | 0.624 | 0 | 0 | 0.008 | 0 | 0.086 | 0 | . | | |
| vol | Correlation Coefficient | 0.07 | 0.029 | -0.029 | -0.061 | .708** | .930** | -.373** | .996** | -0.24 | .991** | .997** | 1 | |
| | Sig. (2-tailed) | 0.638 | 0.846 | 0.846 | 0.685 | 0 | 0 | 0.01 | 0 | 0.102 | 0 | 0 | . | |
| logP | Correlation Coefficient | 0.108 | 0.028 | -0.028 | -0.007 | .838** | .921** | -.589** | .832** | -.432** | .814** | .823** | .824** | 1 |
| | Sig. (2-tailed) | 0.468 | 0.851 | 0.851 | 0.964 | 0 | 0 | 0 | 0 | 0.002 | 0 | 0 | 0 | . |