**Distribution Agreement**

In presenting this thesis as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of medial, now or hereafter known, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Signature

_____ _____
Qi Shi                                                    Date

# Design, Development and Application of a Novel *In-Silico* High-throughput Screening Approach

By

Qi Shi

Master of Science

Chemistry

_____

Dennis C. Liotta

Advisor

_____

James P. Snyder

Advisor

_____

Vince Conticello

Committee Member

_____

Huw Davies

Committee Member

Accepted:

_____

Lisa A. Tedesco, Ph.D.

Dean of the James T. Laney School of Graduate Studies

_____

Date

# Design, Development and Application of a Novel *In-Silico* High-throughput Screening Approach

By

Qi Shi

B.S., Peking University, 2008

Advisor:

Dr. Dennis C. Liotta

Dr. James P. Snyder

An abstract of

A thesis submitted to the Faculty of the Graduate School of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science

in Chemistry

2012

**Abstract**


# Design, Development and Application of a Novel *In-Silico* High-throughput Screening Approach


By Qi Shi


Computer-assisted molecular design is currently a routine component of medicinal chemistry and drug discovery. In this thesis, a novel *in-silico* high-throughput screening approach is developed and applied to two drug discovery projects. The approach, termed FRESH (Fragment-based Exploitation of modular Synthesis by vHTS), is designed to balance multiple factors in addition to potency, namely synthetic feasibility, octane-water partition coefficient, cell-permeability, blood-brain-barrier penetration, metabolism and toxicity. The method combines rapid vHTS (virtual High Throughput Screening), pharmacological property prioritizing and 3D QSAR (Quantitative Structure-activity Relationship) construction. In the first applied project, the approach has provided an initial list of potentially selective NET/SERT (Norepinephrine Transporter/Serotonin Transporter) antagonists with different potency ratios based on existing homology models. In the second application project, the approach has provided a group of novel KCN1 analogs with predicted improvements in both potency and pharmacological properties by employing a two-site binding model for the p300/HIF-1α interaction. The third project on curcumin analogs is still attempting to build a suitable QSAR model. However, once the latter is developed, FRESH can be applied to assist identifying novel curcumin analogs as well.

# Design, Development and Application of a Novel *In-Silico* High-throughput Screening Approach

By

Qi Shi

B.S., Peking University, 2008

Advisor:

Dr. James P. Snyder

Dr. Dennis C. Liotta

A thesis submitted to the faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science in Chemistry

2012

**Acknowledgments**

# Table of Contents

# List of Figures

# List of Tables

# List of Abbreviations

ADMET    adsorption, distribution, metabolism, elimination, toxicity

BBB    blood-brain-barrier

BPBD    binding pocket burial degree

CNS    central nerve system

FRESH    fragment-based exploitation of modular synthesis by vHTS

HBD    hydrogen bonding degree

HIF    hypoxia inducible factor

HTS    high-throughput screening

FDC    fixed dose combination

GBSA    generalized Born/surface area

GI    gastric intestine

GST    glutathione S-transferase

GSU    Georgia State University

MD    molecular dynamics

MM    molecular mechanics

MTL    multi-target ligand

MW    molecular weight

NET    norepinephrine transporter

PD    pharmacodynamics

PK    pharmacokinetics

PSA    polar surface area

QM    quantum mechanics

QSAR       quantitative structure-activity relationship

ROC         receiver operating curve

SERT        serotonin transporter

SNRI        serotonin norepinephrine reuptake inhibitor

SPR          surface plasmon resonance

SSRI         selective serotonin reuptake inhibitor

WCI          Winship Cancer Institute

# Chapter 1: Introduction

This thesis reports initial stages of the development of a computational procedure that is capable of predicting novel, potent and property/ADME-adjusted structures based on a synthetic scheme devised by the practicing chemist: FRESH (Fragment-based Exploitation of modular Synthesis by vHTS). Once implemented it is best applied by close collaboration between computational and synthetic chemists. The following introductory sections lay the groundwork for the construction of important features of the program that hopefully permit accurate and reasoned structure predictions.

## 1.1. Requirements for a molecule to qualify as a drug

### 1.1.1 Chemical space

Modern drug discovery projects often start with the identification of therapeutic targets. Most targets are cellular proteins either derived from the human genome or belong to pathogenic organisms like viruses, bacteria or fungi. Bio-assays are subsequently developed to assess the activity of chemicals against the therapeutic targets. Consequently, various chemicals, either natural products or synthetic analogs, are screened in the corresponding bio-assays to identify active compounds, followed by further optimization on multiple pharmacological properties to provide clinical candidates.

With the development of high-throughput screening (HTS) technology, thousands of compounds can be screened against a biological target within modest time-frame. Although HTS has significantly accelerated the screening process compared to the past, the total chemical space remains vastly unexplored. According to the projection by Bohacek et al., the estimated number of possible molecules with a molecular weight (MW) limit of 500 is $\sim 10^{60}$, which far exceeds the total number of atoms on the earth and unattainable for any available

screening method.[1] Therefore, identifying and focusing on small and discrete compound islands with enriched potential drug candidates is a feasible and more efficient option rather than simply searching hopefully in the vast ocean of the entire chemical space.

### 1.1.2. Drug likeness

Potency is definitely a crucial restriction on chemical space. However, for a potent compound to qualify as a drug, it also has to possess the desired pharmacological properties which insure delivery to the specific site and minimize side effects. Ideally, after administration, a drug molecule will be transferred from the site of administration to the specific target, then remain at an effective concentration within the therapeutic window for a period of time to achieve the desired efficacy and minimize potential side effects.

Various retrospective studies based on existing drugs or clinical candidates have revealed several physical/ADMET (Adsorption, Distribution, Metabolism, Elimination and Toxicity) factors contributing to the pharmacological space. The most widely accepted "Rule of Five" for assessing oral availability by Lipinski was derived from ~2,200 compounds that have at least entered phase II clinical trials.[2] The study showed that 90% of the compounds had no more than 5 hydrogen-bond donors (defined by total numbers of N-H and O-H), 10 hydrogen-bond acceptors (numbers of N and O), MW within 500 and logP (a measurement of lipophilicity) values of less than 5. Jorgensen's "Rule of Three" is based on calculation results of ~1,700 known neutral drugs and also addressed part of the ADME issue.[3] The result revealed property predictions for 90% of the drugs: logS (S for solubility) no less than -6, Caco-cell permeability greater than 30 nm/s and number of primary metabolites no more than 6. Morelli et al. has analyzed ~40 small molecule inhibitors of protein-protein interaction.[4] His conclusion is somewhat in the opposite direction of the Lipinski "Rule of Five": MW larger than 400, logP larger than 4, number of rings larger than 4 and more than 4 H-bond acceptors.

A recent review has summarized the trends of properties for the current medicinal chemistry literature.[5] It demonstrated increasing trends on MW and lipophilicity. For example, a list of 1,680 compounds from the recent literature has a mean cLogP of 4.0 and MW of 435.[6] Another research has revealed that more than 50% of compounds with high potency have cLogP values > 4.25 and MW > 425.[7] Enhanced skills and technologies in organic synthesis, development and drug formulation are likely the factors that contribute to the increased size of the molecules. Although it appears that compounds produced more recently tend to deviate from the chemical space occupied by historical drugs, it is nonetheless still reasonable to adopt the traditional rules, because MW and lipophilicity usually decline as the compounds progress through later stages, probably due to developmental selection pressures: Larger and more lipophilic molecules have augmented risks for bioavailability, solubility, toxicity, synthesis and formulation, which decreases the chance of success as the project proceeds.

Of course, exceptions to these rules exist: Many natural products and their analogs often violate the Lipinski "Rules of Five", but there are successful drugs like Taxol. Thus, rather than accurately defining the boundary between "drug" and "non-drug", the rules mentioned previously should be used as "rules of thumb". Within the vast chemical space, exploitation in these regions is more likely to result a drug. In other words, these rules assist the medicinal chemist to prioritize synthetic candidates and focus on particular areas rather than random probes in the whole chemical space, especially if the drug discovery project has originated from target-based HTS.

### 1.1.3. CNS drug likeness

Compared to peripheral drugs, the ones acting on the central nerve system (CNS) generally have more stringent requirements. One major difference between an orally active CNS drug and a peripheral drug is that besides metabolic stability in the intestine and liver, the CNS

drug also has to penetrate the blood-brain-barrier (BBB) to reach the target. The BBB consists of endothelial cells that control the transfer of drugs and other materials from the blood stream into the cell and also the reverse direction. These epithelial cells form tight junctions, possess few pinocytotic vesicles and lack fenestration. Also, even if the molecule has crossed this layer of endothelial cells, it can be pumped back into the blood by an active transfer process, mainly through p-glycoprotein. Thus, it is more challenging for a molecule to penetrate the BBB than the gastric intestine (GI) membrane. Designing a CNS drug requires more attention to physical/ADMET properties.

From the broader perspective, CNS drugs tend to be smaller (in terms of MW), more lipophilic, have less flexibility (measured by the number of rotatable bonds) and polarity (measured by polar surface area). Levin has suggested a MW cutoff of 400 for CNS drugs in 1980,[8] and in 2002, another study by van de Waterbeemd et al. suggested 450.[9] In a comparison study in 2004, Leeson and Davis analyzed ~330 drugs with ~70 acting on the CNS.[10] The average percent of polar surface area (PSA) for CNS drugs is 16.3, while the value for all drugs is 21. The average number of rotational bond for CNS drugs is 4.7, compared to 6.4 for all drugs. Kelder et al. also discovered that CNS drugs have lower PSA (cutoff is 70 $Å^2$) than non-CNS drugs (120 $Å^2$).[11] Hansch and Leo's quantitative analysis has revealed that cLogP correlates nicely with LogBB (log ratio of drug concentration in the brain to the one in the blood), however, as stated in the last section, more lipophilicity can reduce the chance of success as the project proceeds.[12] The more stringent requirement for CNS drugs on physical and ADME properties definitely adds to the difficulties in the CNS drug discovery process.

## 1.2. Computer assisted drug discovery

### 1.2.1. *In-silico* estimation of the ligand-protein interaction

The development of computational strategies has allowed the screening of large virtual libraries of molecules for the activity against certain protein receptors. Approaches like molecular mechanics (MM), quantum mechanics (QM), hybrid QM/MM and molecular dynamics (MD) are employed to assess the ligand-protein interaction. QM and MD can provide more accurate results and dynamic information, while MM, usually less computationally expensive, can provide excellent molecular geometries and reliable conformational energies. The hybrid QM/MM, which applies QM approach to protein regions around the binding pocket and MM to the rest, can be regarded as a balanced approach between accuracy and computational cost.

Various programs, like Glide, FlexX, GOLD, ICM and DOCK, provide quantitative and semi-quantitative estimation of the ligand-protein interaction. To evaluate the accuracy of such methods, several comparison studies have been carried out. Ligands were obtained from the already available cocrystallized PDB complexes, and the evaluations were based on reproducing the co-crystallized structures. These comparison studies have revealed that the Glide program, which was developed by Schrodinger Inc., remains the most accurate method. [13,14,15,16] In terms of overall performance, Glide had the highest portion of top-ranking poses within 2Å of the original crystal structures. In the more detailed analysis, Glide also showed the best enrichment factors, the minimum decrease on accuracy with increased ligand flexibility and less sensitivity to hydrogen bonding degree (HBD) and binding pocket burial degree (BPBD).[13,15]

Glide adopts a hierarchical protocol for predicting possible poses of ligands. To perform a docking, the protein receptors are first prepared by removing water and assigning amino-acid residues with appropriate protonation state. A docking grid which

specifies the docking site of the ligand is subsequently selected by users. On the ligand side, Glide initially completes a moderately extensive conformational search based on force-fields like OPLS2005. With a thorough exploration of possible positions and orientations over the designated binding site of the receptor, the ligand is placed, followed by the energy minimization on pre-computed OPLS Vander Waals and electrostatic grids. Poses with recognized hydrophobic, hydrogen bond or metal-ligation interactions are favored and the ones with unfavorable steric clashes are penalized. The best poses are selected by the Glidescore, an energy based scoring function.

Several features are attributed to the better performance of Glide. To accommodate the flexibility of protein side-chains and possible induced-fit, Glide can adopt reduced vdW scaling parameters, which softens the penalty for steric repulsive interactions and consequently increases the likelihood of discovering active ligands. In addition, the higher conformational space coverage implemented in Glide as mentioned above may also attribute to the better performance. The post-refinements by OPLS-based energy-minimization also benefit the performance of Glide relative to ICM and GOLD as Perola et al. has revealed.[13]

While the Glide docking program and its scoring function are able to reproduce poses of X-ray crystal structures and valuable information in the docking of novel ligands, the package still needs improvement when it comes to predicting binding affinities. The rescoring by physics-based methods like MM-GBSA (Generalized Born/Surface Area, included in the Prime program in Maestro package) can assist circumventing limitations of the Glide scoring functions, especially the ones associated with desolvation, and entropy penalties for the ligands upon binding. The comparison study by Guimaraes and

Cardozo on several types of protein receptors has revealed that MM-GBSA can produce remarkable correlations between calculated results and experimental data and perform better than Glidescore.[17] In the following sections, Glide was chosen to generate the pose of ligands docked in the receptors, and Prime MM-GBSA was applied to estimate the non-covalent interactions between ligands and receptors.

### 1.2.2. *In-silico* estimation of physical/ADMET properties

As stated in the previous section, favorable physical/ADMET properties are also crucial components for a successful drug. In the early 90s, poor bioavailability problems led to 40% clinical failure.[18] Since then, assessments of ADMET have gradually become routine, and the failure rate dropped to 11%, with the major causes now shifted to lack of efficacy and toxicity.[19] Various computer programs are designed to prioritize molecules based on predicted physical/ADMET properties.

Schrodinger Inc. also provides a program named "Qikprop" for estimation of some physical/ADMET properties in the Maestro package. The input molecule requires a 3D structure with explicit hydrogens. Some properties like MW, number of H-bond donors/acceptors, number of primary metabolites and number of rotatable bonds can be obtained directly from the structure. Others are estimated based on correlations of various 2D or 3D descriptors, like the predicted octanol/water partition coefficient (LogP), solubility (LogS), blood-brain-barrier penetration (logBB), Cell permeability (Pcaco, MDCK), hERG blockage and serum protein binding. QikProp also provides property ranges for any test molecule based on those for 95% of known drugs, which is useful reference for evaluation with a new molecule. It also flags about 30 reactive functional groups which may trigger false positive results in screening tests. Two calculation modes, normal and fast mode are available, with a speed difference of about 30 fold. Fast mode skips the PM3 calculation

which produces the dipole moment, ionization potential and electron affinity. As a result, some predictions may use different sets of descriptors and, thus, results may be different from the normal mode.[20]

Similar physical/ADMET prediction programs are also available from other companies. In the Pipeline Pilot software package developed by Accelrys, besides simple properties like MW and the number of H-bond donors/acceptors, components for estimating octanol/water partition coefficient, blood-brain-barrier penetration, cytochrome P450 2D6 inhibition, hepatotoxicity and plasma protein binding are also available.[21] It also allows users to supply additional data to parameterize for increased accuracy. ACD lab also provides software packages for estimation of the octanol/water partition coefficient, pKa, solubility, blood-brain-barrier penetration, cytochrome P450 inhibition and hERG inhibition.[22] User-added training set data is also a feature. Some other online programs are listed at http://www.vcclab.org/online.html. The latter software programs provide the FRESH approach with additional physical/ADMET parameters for compound target optimization.

## 1.3. Challenges in the lead optimization step

### 1.3.1. Multi-target/site therapies

Many drugs are designed selectively for a single target. Nevertheless, many diseases remain ineffectively treated by the single target drug. Since modulating multiple targets simultaneously can potentially enhance the efficacy compared to single target drugs, it is currently attracting increasing attention of medicinal chemists. The curcumin analogs to be discussed in Chapter 5 are such examples in which multiple kinases are likely to be targeted.

In the current pharmaceutical industry, several approaches are available for developing multi-target drugs. The "cocktail therapy", which combines various therapeutic mechanisms

and involves two or more individual tablets, are widely accepted in a number of diseases, such as the AIDS. While it is effective, the method can suffer from poor patient compliance, particularly if the drug is used for life-improvement or treatment of asymptomatic diseases like hypertension. The fixed dose combination (FDC) approach is thus introduced in which two or more agents are formulated in a single tablet to improve patient compliance. Vytorin, which combines ezetimibe (cholesterol absorption inhibitor) and simvastatin (a type of statin, HMG-CoA reductase inhibitors), belongs to the successful FDC drug category. An additional benefit for the cocktail therapy and FDC drug approaches is that patent life of old drugs can be prolonged. However, for both the cocktail therapy and the FDC applications, variations in the pharmacokinetics/pharmacodynamics (PK/PD) profiles of multiple components requires more extensive clinical studies, let alone the discrepancies of relative rates of metabolism between patients which increase complexity of PK/PD relationships. Furthermore, drug-drug interaction is always a concern when more than one component is administered simultaneously. For the FDC approach, there's also commercial uncertainty that clinicians might still prefer the "cocktail therapy" which offers greater dose flexibility and lower cost in the case of generic drugs.[23]

The advantage of using a single compound against multiple targets is obvious. Like the FDC drugs, patient compliance will be improved relative to cocktail therapies. Additionally, at the research & development stage, the single chemical component is free of complications such as complex PK/PD correlations or drug-drug interactions. Some knowledge-based method, like framework combination (either by a linker or completely merged together; see below), has offered interesting drug leads. As illustrated in **Figure 1**, the neuro-protective agent Ladostigil, which combines the framework of Rivastigmine (acetylcholinesterase inhibitor) and Rasagiline (monoamine oxidase inhibitor), demonstrates dual inhibition against acetylcholinesterase and monoamine oxidase.[24]
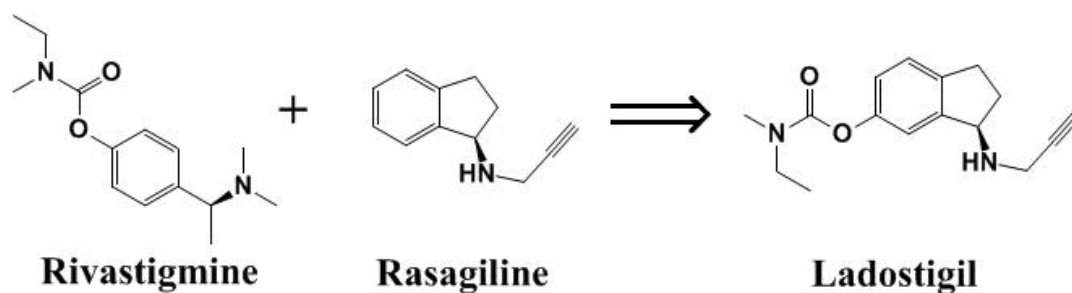
**Figure 1.** Ladostigil is derived from the framework combination of Rivastigmine and Rasagiline

## 1.3.2. Balancing multiple factors for multi-target drugs

The major drawback for developing multi-target ligand (MTL) drugs is that it is considerably more complicated to gain acceptable activity simultaneously at two or more targets/sites. Besides, there's always the question of whether an optimal potency ratio exists, and adjusting the ratio is even more difficult. Meanwhile, as stated in the sections above, other factors like variable lipophilicity, solubility, metabolic stability and BBB permeability (for CNS drugs) also have to be considered. Some of these factors may act against each other so the drug has to stay in an "operating window" in which the "conflicting" factors are all satisfied. At this stage, synthetic feasibility may be a concern as well, as synthetic processes generally consume considerable time and resources. Thus, synthetic chemists may be reluctant to invest large amounts of time and money pursuing just one compound that may eventually fail. Although the increased complexity in the design and optimization of MTL occurs at an earlier and less expensive stage, balancing multiple factors is nonetheless a challenging task. In this thesis, the author has developed a novel approach which combines knowledge-based rational design and high-throughput screening intended to address the multi-dimensional optimization issue of MTL.

# Chapter 2: Design and development of a novel high-throughput virtual screening approach for MTL

## 2.1. Overall design strategy

The knowledge-based rational design approach can provide some interesting candidates, but the limitations are obvious. As it's named, it requires information from previous studies. In addition, the knowledge-based approach alone generally addresses only one aspect of the requirement, like biological activity, metabolic stability or MW. When attempting to balance multiple factors, it will hardly work. During the lead optimization step, synthetic chemists will usually investigate the chemical space around the lead molecule and hopefully come up with at least one clinical candidate with both improved activity and ADMET properties. However, due to the relatively slow rate of organic synthesis, the coverage of chemical space by actual organic synthesis is very limited. On one side, a molecule has to meet various requirements to become a drug. On the other side, medicinal chemists face numerous choices. The chemical space around a drug lead, which is just a tiny portion of the total space, still represents possible synthetic candidates of astronomical numbers. The basic idea for the methodology presented here is thus generated: First, construct a diverse and tailored virtual compound library. Second, apply a set of prioritizing criteria to identify molecules with improvements in potency and physical/ADMET properties.

## 2.2. Screening software program selection

The initial combinatorial library will contain numerous compounds requiring virtual high throughput screening (vHTS). All the software programs involved should support batch calculations. For constructing the initial library, Pipeline Pilot and CombiGlide are able to perform combinatorial library enumeration. The advantage for Pipeline Pilot is huge and it

was selected to construct the initial virtual library: First, it is a visualized programming software which provides great user flexibility. Second, the working protocol of Pipeline Pilot shows "real-time" numbers reflecting the process of the job, which provides crucial information like calculation speed and the size of the library. In fact, these numbers are critical parameters for making protocol refinement decisions and optimizing the algorithm. The overall speed is much faster than the CombiGlide, which probably results from the optimized algorithm for internal-memory monitoring. Since results from other computational packages can be easily incorporated into the FRESH approach, Pipeline Pilot has been employed to construct the main protocol and several junctions are reserved for capturing output from such packages.

As discussed in the previous chapter, Glide performs best when generating ligand-receptor poses, while Prime MM-GBSA is better at estimating the energy of ligand-protein interaction. In addition, both Glide and Prime MM-GBSA support batch processing, so these methods were chosen to evaluate ligand-protein interactions. For physical and ADMET property estimations, both Pipeline Pilot and Qikprop are designed to do batch-mode calculations. Thus, they were employed in the approach. In addition, since all of these programs support the .sdf file format, this common file format serves as the "communication language" between Pipeline Pilot and other involved software programs throughout the entire procedure.

## 2.3. Algorithm design and optimization

The chemical space around a drug lead, though not as large as $10^{60}$, can still reach an astronomical scale, mainly due to combinatorial explosion. Hence, the construction of the initial screening library still requires informed selection. During lead-optimization, synthetic chemistry is frequently the rate-limiting step. Therefore, the approach was based on the existing synthetic route which is familiar to the chemists and primarily focused on the

exploitation of early synthetic analogs. The process of building the initial screening library can be treated as a virtual mimic of the actual organic synthesis. First, large commercial libraries are screened for synthetic building blocks to construct focused libraries. Each building block is consequently marked at the connection point representing the corresponding fragment. The fragments are then pieced together according to the synthetic reaction scheme. After exploitation of all possible combinations of the fragments, the initial screening combinatorial library is constructed and subjected to various selection criteria. The library is then subject to various selection criteria. As final output, the molecules are predicted to show either equal or improved activity, together with favorable physical/ADMET properties required for a drug-like substance. **Figure 2** illustrates the basic principles discussed herein. The list of molecules that passes all the selections is submitted for actual synthesis, bioactivity evaluation and various extra tests.
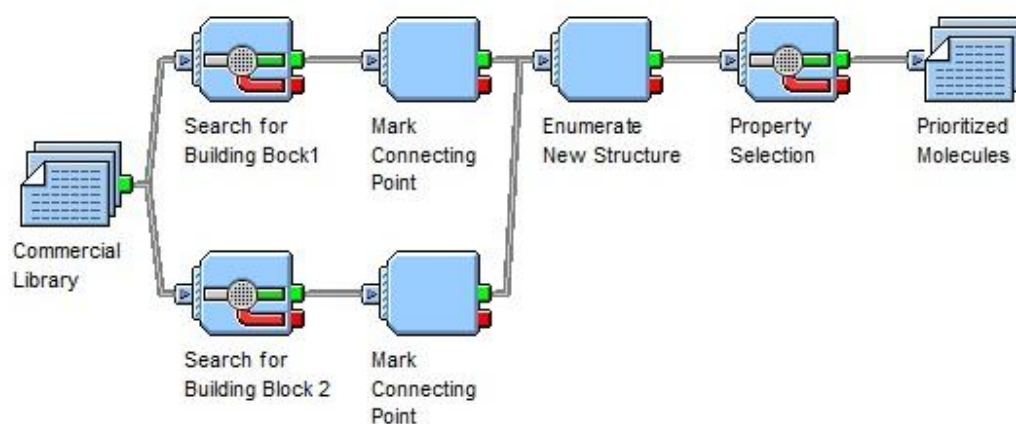


**Figure 2.** A simple demo protocol for the FRESH approach

The starting commercial library used in the following projects is the Chem-Navigator library which contains ~7 million compounds. For frequently-used building blocks like aldehydes, amines and halogens, the number can exceed 10,000. Assuming that the molecule is composed of three building fragments, and each fragment is represented by 10,000

intermediates, the total number of possible combinations will then reach 1 trillion. For a modest-sized computer, the time it consumes to process them can be unacceptably long, i.e. ~1,000 days. The possible threat of a size effect originating from the "combinatorial explosion" requires a significant effort to optimize the processing algorithm.

Several strategies were utilized to ensure that the entire screening cascade can handle extremely large libraries. First, reducing the numbers of intermediate fragments can effectively lower the burden of enumerating the initial library and the following screening process, so efficient selection of favorable fragments remains a major task for algorithm optimization. Identification of a list of functional groups and substructures that are reactive, unstable in plasma, prone to cause false-positives in screening assays or induce toxicity constitutes the first checklist for the fragments.[25] Second, the "Rule of Three" for fragments was also employed, which can require a MW threshold lower than 300 depending on the specific project requirement.[26] The third round of fragment selection is based on synthetic route. Functional groups or substructures that are capable of creating region/steric selectivity issues are assigned a lower priority.

The speed of the programs involved in the entire approach varies. Pipeline Pilot can process ~ 1 billion structures per day, while the speed for Qikprop is around several million, Glide docking followed by Prime MM-GBSA rescoring can only handle several thousand daily. Therefore, in the screening scheme, Pipeline Pilot is placed first, followed by Qikprop and subsequently Glide + Prime. The number of potential structures is so large that actions must be taken to ensure the computer can easily identify each individual molecule and possible conformers. Several numerical identification systems are employed along the processing pathway, and filters for removal of duplicate structures are applied at various points to avoid unnecessary repeated calculations.

The advantages of this novel approach are obvious. Compared to traditional lead-optimization strategies, this method can cover a much larger chemical space within an acceptable time period and at lower cost due to the enhanced speed. With months of effort spent combating technical difficulties, bugs, defects and optimization of the prioritizing scheme, the approach can now process up to several billion structures in a single run. Instead of producing only a single molecule, a complementary set of structures is generated simultaneously, each of which can potentially serve as backup to the others. The approach is also synthesis friendly, as the library construction originates from the chemists' synthetic routes, and building blocks are obtained directly from a commercial database. The problem of multi-dimensional optimization mentioned in the last chapter is also addressed, since bioactivity and drug-like physical/ADMET properties are incorporated into the output structures.

To avoid the drawback of missing possible interesting fragments not found in commercial databases, the package has been constructed to permit user selection of additional fragments deemed important for the project team to explore. In addition, self-designed sub-protocols for incorporation of fragments obtained from shape and electro-similarity match by the isosteric replacement BROOD software and from a phenyl-ring scanning approach are also available.

# Chapter 3: Application I: Composing novel SNRIs

## 3.1. Project background

Chronic pain originated from disease or injured nerves remains a problem in the US. A statistical study based on ~30,000 adults revealed that the prevalence of chronic pain is more than 30%.[27] Among these, half experienced daily pain with up to 30% reported severe pain intensity. The most common chronic main is the lower back pain, followed by osteoarthritis pain. The pain obviously reduces life quality and sometimes even leads to a debilitating life-style and loss of ability to work. In the US, lost productivity from chronic pain costs an estimated \$61.2 billion per year, and the additional annual cost for treating patients with chronic pain is ~\$6,000.[28,29] Therefore, chronic pain remains a large health, social and economic burden on the US. Effective therapies will have an enormous impact on the whole society.

The damaged nerve itself continues to be a challenging condition. There is currently no definite clinical treatment to repair the damaged tissue.[30] Alternative medications block the corresponding pain signaling pathway and alleviate severe pain. For example, opioids, which inhibit the release of neurotransmitters, have been adopted and remain an effective therapy against pain. Nonetheless, the major problem for opioids is addiction and drug abuse. There's also the potential threat of obtaining prescription opioids from medical prescription for illegal drug use. Therefore, devising alternative chronic and severe pain therapies besides opioids apart from opioids remains an attractive goal.

Monoamine reuptake inhibitors, which pump the monoamine neurotransmitters in the synaptic clefts back to the presynaptic neurons, are potential therapeutic targets. By blocking the reuptake inhibitors, the amount of neurotransmitters in the synaptic clefts increases and

stronger signals are transmitted to the postsynaptic neurons and subsequently suppress the pain pathways.[31] Previous studies have revealed that norepinephrine transporter (NET) inhibitors are effective against neuro-pain. While the SSRIs (selective serotonin reuptake inhibitor) are not effective against pain, the inhibition of the serotonin transporter (SERT) along with NET by SNRI (serotonin norepinephrine reuptake inhibitor) appears to enhance the efficacy.[32] Administering separate drugs for NET and SERT, as discussed under cocktail therapy in Chapter 1, can be problematic due to different PK/PD properties of each component and the potential threat from drug-drug interactions. Therefore, an MTL (multi-target ligand) drug, which targets both NET and SERT, would be more favorable. Several dual NET/SERT inhibitors have demonstrated efficacy against chronic pain, one of which is milnacipran (**Figure 3**).[33] However, whether an optimal ratio of NET to SERT potencies exists has not been determined yet. In this project, a series of arylcyclopropylamine compounds (**Table 1**), analogs of milnacipran, has demonstrated significant potency for both NET and SERT, and we intend to use them as guides to synthesis of additional analogs to fine-tune the therapeutic ratio of dual NET/SERT inhibitors. My task is to provide synthetic collaborators with specific suggestions for arylcyclopropylamines analogs with suitable variations of the potency ratio of NET/SERT while retaining overall potency.
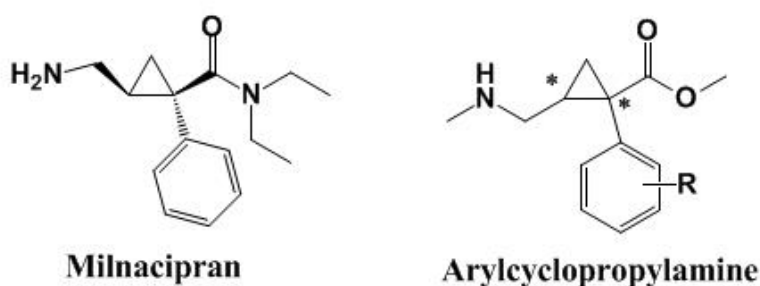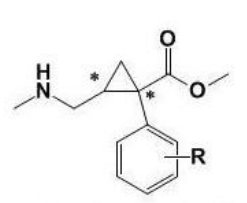


**Figure 3.** Milnacipran and arylcyclopropylamine compounds

## 3.2. Challenges in the lead optimization step

The arylcyclopropylamine analogs listed in **Table 1** stay well within the threshold of "drug-likeness". For example, Compound 1 has a MW of 298 and logP value of 2.2. Therefore, lack of possible modification options is less a concern. However, the main focus of this project at the moment is the potency ratio, which means a quantitative evaluation; the model has to be accurate for estimating the potency. Since a knowledge-based method is unlikely to be capable of providing quantitative estimates, systematic and focused screening suggests itself as a useful strategy to obtain new analogs. The FRESH approach implements vHTS, implying it is a candidate for addressing the goals of this project.
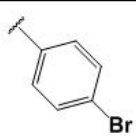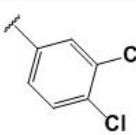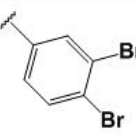
| R group | Name | Chirality | MW | cLogP | NET Ki (nM) | SERT Ki (nM) | SERT to NET ratio |
|---|---|---|---|---|---|---|---|
| | 1 | SS | 298 | 2.2 | 29.9 | 630 | 21.1 |
| | 2 | RR | 298 | 2.2 | 1.66 | 31 | 18.7 |
| | 3 | SS | 288 | 2.8 | 5.3 | 27 | 5.1 |
| | 4 | RR | 288 | 2.8 | 0.2 | 1.6 | 8.0 |
| | 5 | SS | 377 | 3.0 | 3.3 | 13 | 3.9 |
| | 6 | RR | 377 | 3.0 | 0.45 | 1.22 | 2.7 |

**Table 1.** Six arylcyclopropylamine compounds with ClogP and $IC_{50}$ values

Another concern lies in the bioavailability of the compounds. Although intrathecal administration (IT, injection into the spaces surrounding the spinal cord or brain) remains the current administration route for these compounds, it would certainly be advantageous to have the drug administered orally. However, two moieties in the current lead candidates probably need further modification before the compounds qualify as potential oral drugs. The ester

bond can be hydrolyzed, and the resulting carboxylic acid with a low capacity to cross the BBB can result in reduced efficacy and increased threat of toxicity. The amine moiety, which is basic, also poses a threat for the BBB penetration. Both the ester and amine have to be replaced or shielded (example: isosteric replacement or pro-drug approach, respectively) for enhanced bio-availability to the brain. The BROOD isosteric replacement software incorporated in the FRESH approach is able to perform shape and electrostatic tanimoto matching, which can possibly help to alleviate the problems.

## 3.3. Homology models of NET and SERT

Since the project requires assessment of ligand activity against protein, to apply the FRESH approach, an *in-silico* QSAR model needs to be constructed. Thus, suitable receptor structures (SERT and NET) combined with Glide docking followed by Prime MM-GBSA rescoring are required.  However, no crystal structure of NET or SERT is currently available. An alternative strategy is to employ homology models. Fortunately, a previous study by Ravna et al. has established such models for both NET and SERT based on a Leucine transporter.[34] The sequence similarity between NET and the Leucine transporter is 39% and for SERT 35%. These two homology models, together with the preliminary results listed in **Table 1**, provide a good starting point for the modeling work. Preliminary efforts along these lines were made by Dr. Spandan Chennamadhavuni in the spring of 2011 as he brought his Ph.D. studies to an end.  However, the modeling work did not proceed to the development of QSAR models.[35]

According to the study by Ravna et al, two possible binding sites (Site 1 and Site 2) exist on both the NET and SERT homology models. Site 1 can accommodate small molecules like cocaine, while larger tricyclic compounds tend to populate Site 2.  Thus, Site 1 was chosen to

initiate the work.[34] However, a protein pocket prediction check by the online program Castp (http://sts.bioengr.uic.edu/castp/calculation.php) revealed a problem: the pockets on both NET and SERT are too crowded (volume ~360 $Å^3$ by Castp estimation) to fit the ligands (volume ~900 $Å^3$ by Qikprop), and attempts to dock the ligands within the frozen receptor failed to yield any docking pose. Induced-fit docking was then considered to provide the protein receptor side chains with some flexibility. However, induced-fit docking is a computationally expensive method, which is not suitable for vHTS and thus less than ideal for incorporating into the FRESH method. An additional technical problem also exists in that that the induced-fit docking output pose file contains only one receptor and one ligand, making it impossible for MM-GBSA to perform batch calculations required in the FRESH vHTS module. A decision was made that induced-fit docking would only be used to generate docking receptors followed by subsequent performance of standard Glide docking. This two-part strategy captures a degree of protein flexibility while not sacrificing library processing speed. Among the ligands in **Table 1**, the two largest ones, Compounds 5 and 6 were chosen for induced-fit docking, and for each compound, only one pose was requested. The receptors derived from the resulting pose were extracted and subsequently used as the receptor for the normal Glide docking.

After the Glide docking followed by Prime MM-GBSA rescoring, the top pose for each ligand was submitted to QSAR analysis. A linear regression was attempted based on the equation $\Delta G = -RT\ln K$. For this particular project, the MM-GBSA value (E) was employed to estimate the free energy change, and the experimental $K_i$ value was used for K (which will result in a positive slope).

**Figures 4 to 7** illustrate the Compound 5 and 6 correlations for NET and SERT with all six ligands to define the best protein template. By comparing **Figure 4** with **5**, and **Figure 6**

with **7**, the conclusion can be drawn that docking receptors generated by Compound 6 deliver better correlations. It is worth mentioning that these correlations were based on preliminary results in which only six ligands are involved. Nevertheless, they provide a good starting point for generating prospective compounds for further synthesis.
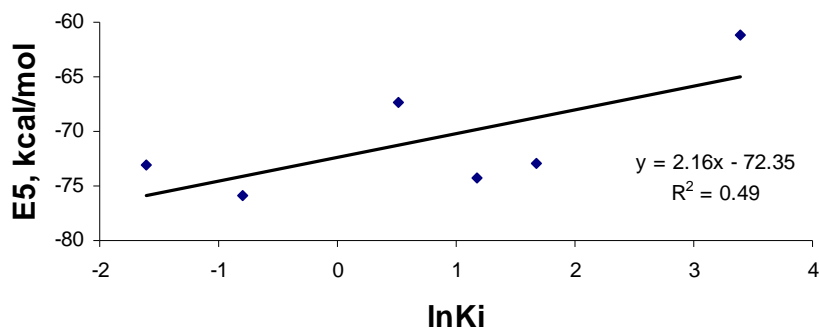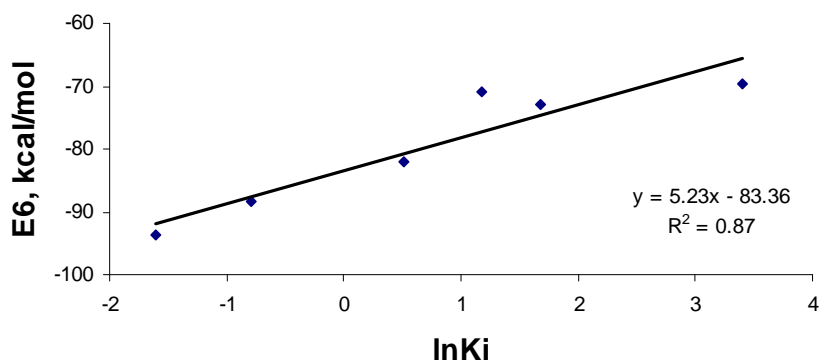


**Figure 4.** Correlations of estimated binding NET affinity with experimental Ki for six ligands. Docking receptor was generated from Compound 5 induced fitting.



**Figure 5.** Correlations of estimated binding NET affinity with experimental Ki for six ligands. Docking receptor was generated from Compound 6 induced fitting.
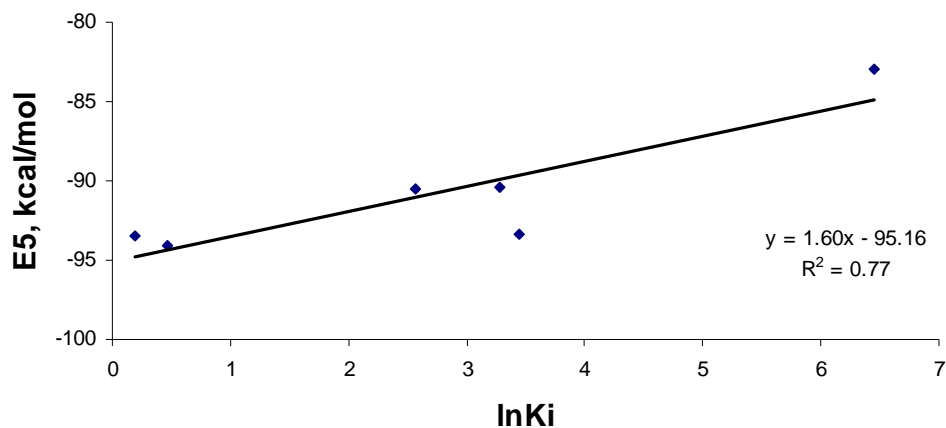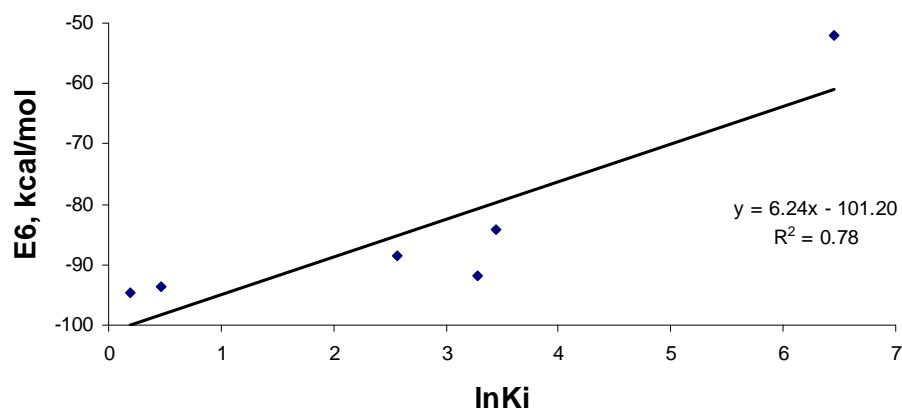
**Figure 6.** Correlations of estimated binding SERT affinity with experimental Ki for six ligands. Docking receptor was generated from Compound 5 induced fitting.



**Figure 7.** Correlations of estimated binding SERT affinity with experimental Ki for six ligands. Docking receptor was generated from Compound 6 induced fitting.

## 3.4. Application of the FRESH approach

### 3.4.1. Protocol design for the FRESH approach

The protocol for the FRESH approach was constructed in accordance with the synthetic route in **Figure 8**. The $R_1$ and $R_2$ groups originate from a carbene precursor which contains an aromatic ring ($R_1$) and an ester group ($R_2$), while the $R_3$ group is from the amine building

block. Because the six ligands in **Table 1** have the same $R_1$ and $R_3$ groups (both are methyls), any informative SAR, $R_2$ (only substituted phenyl and pyridine ring are allowed) was chosen to be the moiety screened by the commercial library. From **Table 3**, the RR ligands show better activity. Thus, these analogs were assigned a higher priority and investigated first.
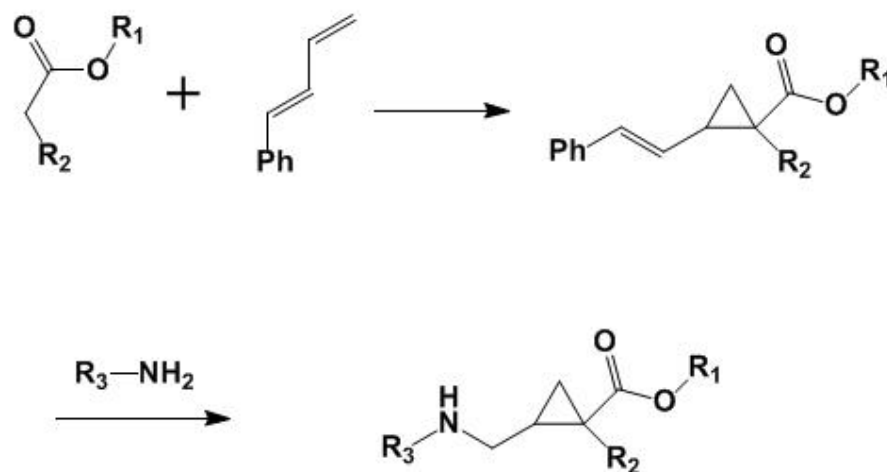


**Figure 8.** Synthetic route for the arylcyclopropylamine analogs

The FRESH protocol for $R_2$ evaluation was thus designed as shown in **Figure 9**. Starting from the ChemNavigator commercial library, the corresponding building block was identified, and the $R_2$ fragment, extracted. A sub-protocol was employed which prioritizes fragments by identifying unfavorable functional groups or structures that are known "trouble makers" for a drug molecule or pose synthetic issues.[25] The $K_i$ values for potential NET and SERT antagonists were predicted using the correlations in **Figure 5** and **7**. Other property/potency selection criteria incorporated into the protocol are the following: [2,3,20]

1. Pipeline Pilot: MW < 501, Number of O and N $\leq$ 10, Number of H-donors $\leq$ 5, LogP $\leq$ 5, PSA $\leq$ 90, LogS $\geq$ -6.0, Percent of $sp^3$ carbon $\geq$ 40

2. Qikprop: LogBB $\geq$ 0, Number of Primary Metabolites $\leq$ 3, Pcaco $\geq$ 500, PMDCK $\geq$ 500, oral absportion $\geq$ 0.8

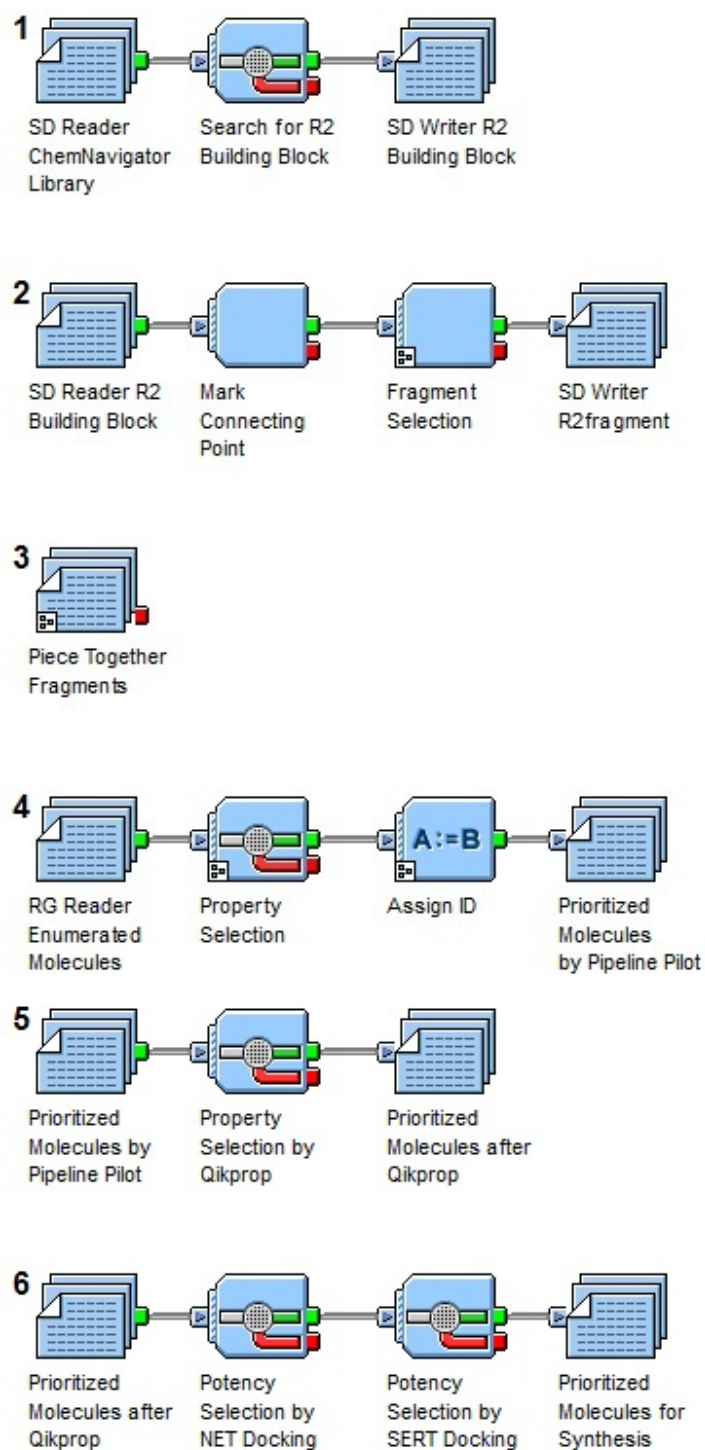3. Docking (potency): Predicted NET $K_i \leq 50$ nM, Predicted SERT $K_i \leq 50$ nM



**Figure 9.** The FRESH protocol (main interface) for prioritizing arylcyclopropylamine analogs. Some sub-protocol components are not shown.

### 3.4.2. Resulting structures

The FRESH approach has identified ~30 structures with the corresponding predicted ratios. **Table 2** depicts the ones selected for synthesis according to the SERT/NET ratio. The experimental $K_i$ values for NET were obtained and are also included. From the table, we can conclude that the experimental $K_i$ values for two compounds are within 10 fold of the experimental values, with one (Compound 69) essentially identical. The other four are clearly false positives. Additional work is required to eliminate them.
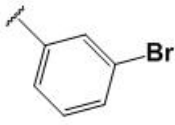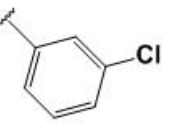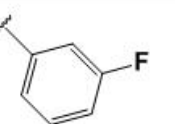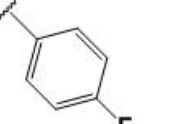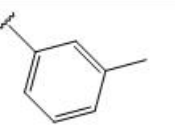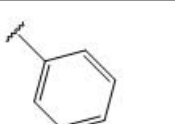
| R group | Name | Predicted NET Ki (nM) | Predicted SERT Ki (nM) | Predicted SERT to NET ratio | Experimental NET Ki (nM) |
|---|---|---|---|---|---|
| (phenyl)—Br | 8 | 1.2 | 2.5 | 2 | 15.4 |
| (phenyl)—Cl | 28 | 1.2 | 5.0 | 4 | 72.6 |
| (phenyl)—F | 68 | 2.0 | 3.7 | 2 | 13.7 |
| (phenyl)—F | 69 | 1.5 | 6.0 | 4 | 2.9 |
| (phenyl)—CH₃ | 140 | 2.7 | 8.1 | 3 | 34.7 |
| (phenyl) | 155 | 2.2 | 12.9 | 6 | 48.3 |

**Table 2.** Result of the FRESH approach for arylcyclopropylamine analogs

# Chapter 4: Application II: Identification of novel KCN1 analogs to block the p300/KCN1 interaction

## 4.1. Project background

Hypoxia is a condition in which tissues suffer from insufficient oxygen supply. In solid tumor tissues, hypoxia is prevalent, due to inadequate development of a vascular blood supply.[36] The hypoxia inducible factor (HIF) pathway is found to be crucial for tumor cell growth under hypoxic conditions, one consequence of which is that hypoxic tumors are more resistant to chemo- and radio-therapies.[37,38] Along the pathway, the HIF-1α subunit associates with the CH1 domain of cofactors p300 or CBP to form a functional transcription factor. Thus, the p300/HIF-1α complex has become an attractive target for cancer therapy. Reid-Mooring et al. have synthesized a series of arylsulfonamides and analogs and they've demonstrated sub-micromolar potency against the p300/HIF-1α interaction, with KCN1 (**Figure 3**) as the lead compound.[39] My task in this project is to use computational methodology to perform lead optimization and provide a list of molecules with predicted improvements in potency against the p300/HIF-1α interaction. Since one of the targeted cancers is glioblastoma, the molecules also have to demonstrate favorable and physical/ADMET properties required for CNS drugs.
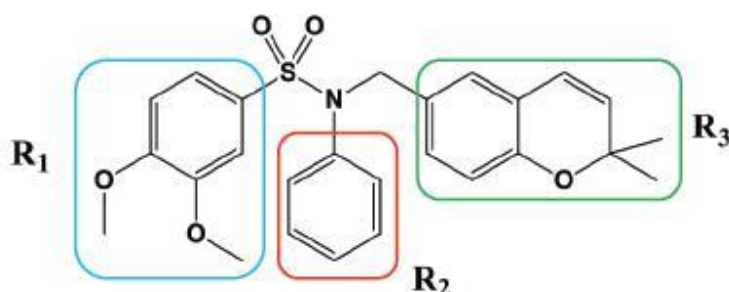


**Figure 10.** Structure of KCN1 with three highlighted substituent groups

## 4.2. Challenge in the lead optimization step

Although KCN1 is the lead molecule in the project, if evaluated by the "Rules of lead-likeness", it is clearly a violator. The MW (465) is already close to the threshold of 500 defined by the Lipinsky Rule of 5, which is significantly larger than the suggested MW cutoff of 350 for lead-likeness, let alone the general trend for CNS drugs to possess a smaller MW.[40] Therefore, in terms of ligand efficiency, KCN1 is not a lead with high efficiency. The experimental $IC_{50}$ value for KCN1 is ~650 nM.[39] If the intended $IC_{50}$ value is 5 nM or less after optimization, it requires an improvement of ~100 fold. Considering the 500 MW threshold, there are few options for additional heavy atoms. Although for certain small molecules the ligands efficiency can reach ~1.5 kcal/mol, as the ligand size increases, the likelihood of encountering one "magic methyl" decreases rapidly.[41] Thus, it is a remote possibility that by adding extra groups, an analog can be found with ~100 fold improvement while still within the 500 MW threshold. In addition, cLogP value also tends to increase during the lead optimization process.[40] A desired cLogP cutoff is therefore suggested at 3 to ensure the final clinical candidate can still stay within the threshold of 5.[40] Since the cLogP value for KCN1 is already close to 5, this parameter also requires attention in the lead optimization process. In fact, my previous attempts in this project by using structure-based methods alone to provide molecules with predicted improvements in both potency and physical properties while still employing the same synthetic route were not successful. In response to this situation, a revised FRESH approach has been devised.

## 4.3. Two-site modeling of KCN1 for the p300 transcription factor

### 4.3.1. Binding receptor selection

In this project, the potency of the lead compound along with its physical properties needs improvement. To apply the FRESH approach, a quantitative estimation of the ligand binding energy is desirable. Conceptually, KCN1 can interrupt the p300/HIF-1α interaction by binding separately to p300 or HIF-1α, or by associating with the complex to attenuate its function. An NMR study by De Guzman et al. has revealed that while p300 is able to maintain its 3D-architecture even in the absence of HIF-1α, the C terminal transactivation domain (CAD) of the latter is disordered when uncomplexed with p300.[42] Therefore, a reasonable hypothesis is that KCN1 binds to p300 and thereby blocks induced fit by HIF-1α.

A series of experiments by collaborators at the Winship Cancer Institute (WCI) at Emory and Georgia State University (GSU) have been conducted to test the hypothesis.[43] **Figure 11** demonstrates the results from affinity pull-down analysis (Dr. Erwin van Meir and colleagues; WCI). The input lane and the – lane were control groups: The former used a fraction of the cell extract before pull down to verify protein expression, and the latter employed uncoupled beads for non-specific binding (- lanes). The figure illustrates that while KCN1 can pull down a fraction of p300, no detectable interaction between KCN1 and HIF-1α is observed. A radio labeling experiment has also been performed. Recombinant fusion peptides, which contain Glutathione S-transferase (GST) and the CH1 domain of p300 (GST-p300-CH1) were incubated with $^{14}$C-KCN1. The size of GST-p300-CH1 was verified by Coomassie stained gel as shown in **Figure 12** (right). In **Figure 12** (left)**,** the bound activity for GST-p300-CH1 is shown to be significantly greater than that obtained with GST-only peptides. Another experiment based on surface plasmon resonance measurements was also performed by the GSU leg of the team under the guidance of Dr. Binghe Wang. KCN1 was attached covalently to a gold surface (**Figure 13**). The p300-CH1 peptides were streamed over the surface in a series of concentrations as illustrated by

**Figure 14**. SPR signals show response to the p300-CH1 peptides and changes with variations of concentration. Analysis of the curve shapes and concentration dependence results in a $K_d$ value of ~345 nM for KCN1 binding to p300-CH1 which is comparable to the bio-assay value.[39] Altogether, these findings have supported the hypothesis that KCN1 can bind to p300-CH1. Based on these results, a binding model of the p300-CH1/KCN1 complex was initiated. The model will be applied for estimating potency and incorporated into the FRESH approach for generation of new analogs for synthesis and bio-activity testing.
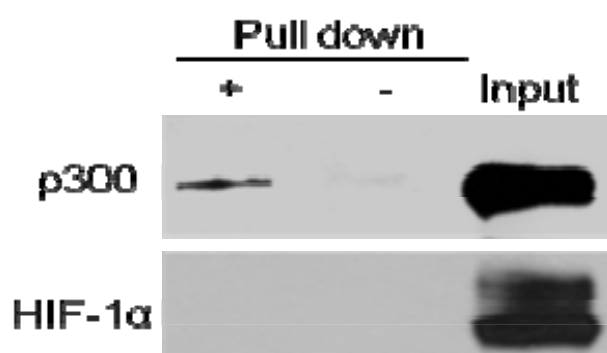


**Figure 11.** Affinity pull down analysis of p300 and HIF-1α proteins using KCN1-coupled agarose beads.
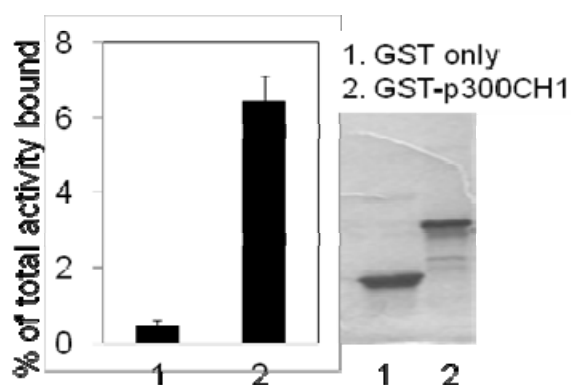


**Figure 12.** $^{14}$C-KCN1 binding experiment result.

**Figure 13.** KCN1 attached to the gold surface.



**Figure 14.** SPR sensorgrams for KCN1 binding to p300

### 4.3.2. Binding site selection

The above-mentioned experiments do not provide information on the exact binding site for KCN1 on the CH1 domain of p300. In addition, no crystal structure of a small molecule complexed with p300 is available. Therefore, binding site information was obtained from the structure of p300/ HIF-1α complex (PDB code: 1L3E) and a previously reported mutagenesis study.[44,45]

**Figure 15** illustrates the structure of the p300-CH1 domain extracted from the p300/HIF-1α complex. As shown, the p300-CH1 structure consists of three major helices nearly perpendicular to each other. According to a random mutagenesis study, four residues on p300 (Leu344, Leu345, Cys388 and Cys393) were determined to be crucial for complexation with HIF-1α.[45] Cys388 and Cys393 coordinate to a zinc ion, explaining that disruption of the p300 zinc finger interrupts the p300 structure and consequently the p300/HIF-1α interaction. Leu344 and Leu345 are located in the region where the three helices are adjacent to one another (**Figure 16**). Four possible clefts, at which KCN1 can interact with at least 2 helices, were investigated **(Figure 17**, left**).** Each was subjected to KCN1 Glide docking followed by Prime MM-GBSA rescoring. **Figure 17** (right) depicts the top two sites with high predicted binding energy values.
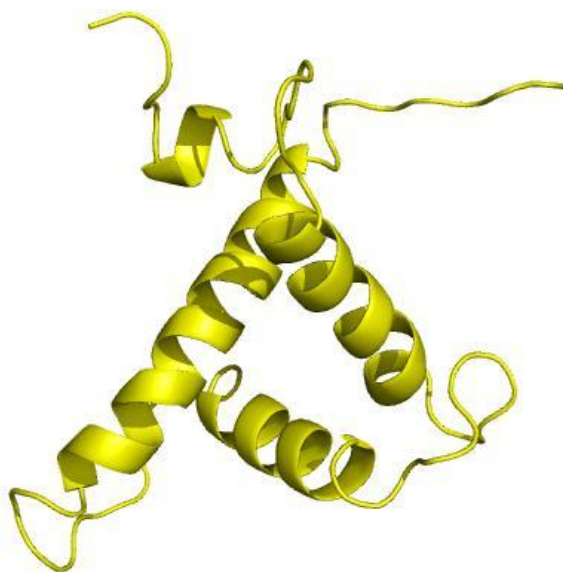


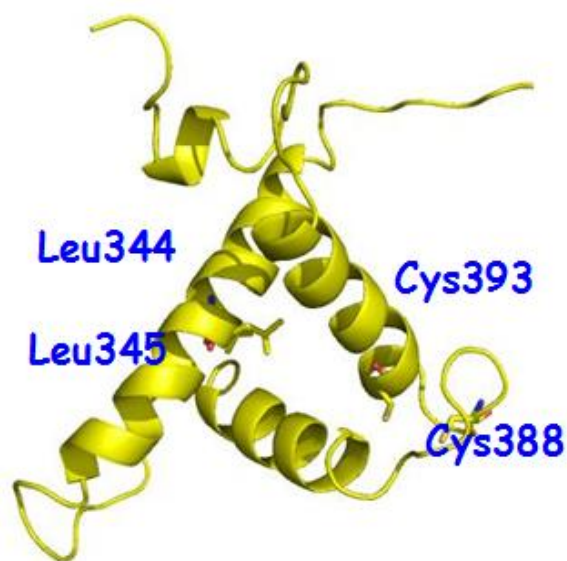**Figure 15.** p300-CH1 extracted from the complex.

**Figure 16.** Crucial residues Leu344, Leu345, Cys388 and Cys393 on p300 CH1. Leu344 is hidden under the helix behind Leu345.
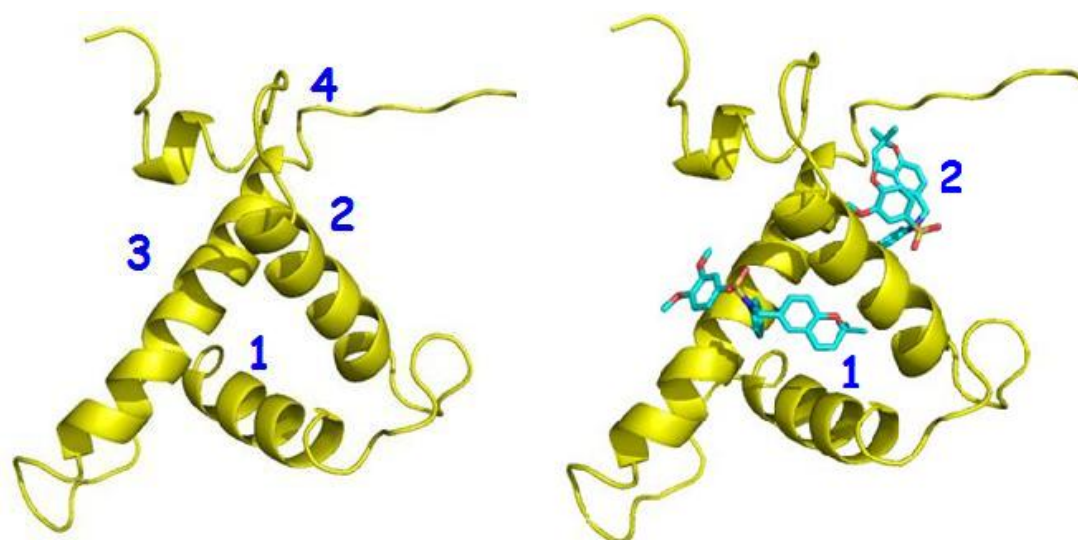


**Figure 17.** Four clefts chosen for the docking sites(left)  The top two sites with docked KCN1 (right)

Remarkably, the top two best-scoring sites (Sites 1 and 2) are coincident with the binding

locus of the two HIF-1α helices (**Figure 18**). Furthermore, the mutagenesis study revealed

four residues on HIF-1α to be crucial for the p300/HIF-1α interaction, namely Leu795,

Cys800, Leu818 and Leu822.[45] Coincidently, Leu795 and Cys800 are located on helix B,

while Leu818 and Leu822 are found on helix A (**Figure 19**). This outcome provides further

confidence that KCN1 is likely to bind to p300-CH1 at these two sites. Accordingly, these two centers have been employed in the following docking study for developing preliminary first-pass QSARs for the KCN1 analogs.
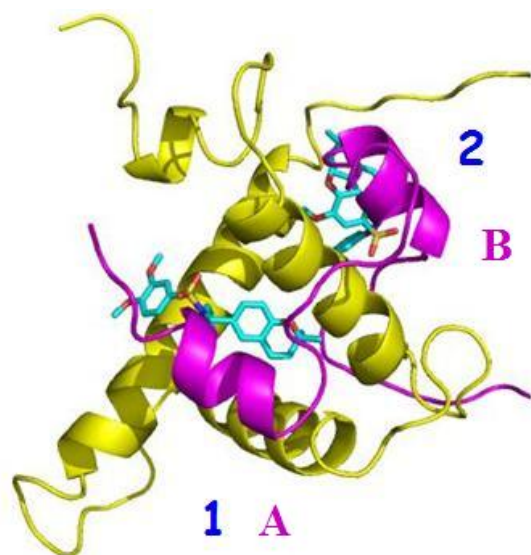


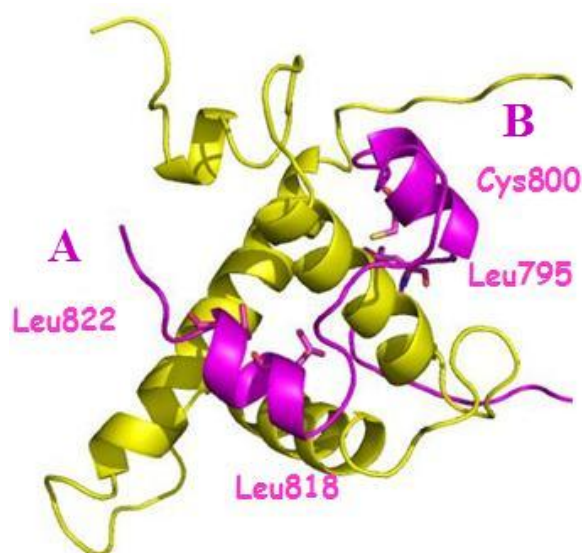**Figure 18.** Two helices on HIF-1α (purple) superimpose with docked KCN1 at Site 1 and Site 2



**Figure 19.** Crucial residues on HIF-1α, namely Leu795, Cys800, Leu818 and Leu822

### 4.3.3. Linear regression attempts for KCN1 analogs

Reid-Mooring has synthesized ~30 KCN1 analogs, and the corresponding $IC_{50}$ values (structures and $IC_{50}$ values are included in the **Appendix I**) were measured.[39] Although in general the structure-activity relationship of these compounds appears "flat", nevertheless they may provide preliminary guidance for development of analogs with more significant enhancement. Thus, attempts were made to construct a useful QSAR. Each analog were subjected to Glide docking at each of the two favored binding sites (**Figure 17,** Site 1 and Site 2) followed by energy rescoring with Prime MM-GBSA to obtain estimates of binding free energy ($\Delta E$).[46,47] Most of the experimental $IC_{50}$ values are from a single run and no average value is available. To minimize possible biological variability from measurements made over time, an $IC_{50}$ value for each analog scaled to the corresponding KCN1 $IC_{50}$ reference determined in the same experiment was employed for the correlations. Thus, the ratio of the $IC_{50}$ value of the analog and the corresponding $IC_{50}$ value of KCN1 was used as a measurement of activity ($\Delta IC_{50} \sim K$) and applied to the similar QSAR analysis as described in the previous chapter.

Initially, all 30 compounds were employed to generate a single correlation line. However, no satisfactory correlation was obtained. Since all the compounds with $IC_{50}$ values lower than KCN1 (**Table 3**) carry the same $R_1$ group as KCN1 (Group II compounds in **Appendix I**), these were selected to see if they would provide acceptable correlations. **Figures 20** and **21** illustrate the results at Site 1 and Site 2. The points on **Figure 21** appear more "clustered" than those on **Figure 20**. However, even for Site 2, several outliers exist. Simply removing the outliers and applying the corresponding correlation will generate skepticism. Alternative approaches are required to support the hypothesis that the predicted binding energy values can be used to prioritize the structures in the FRESH protocol.

| Name | No. of repeats | IC$_{50}$ (nM) |
|------|----------------|-----------------|
| KCN1 | 42 | 648 |
| 1601 | 6 | 306 |
| 1604 | 6 | 478 |
| 1606 | 6 | 378 |
| 2609 | 30 | 280 |

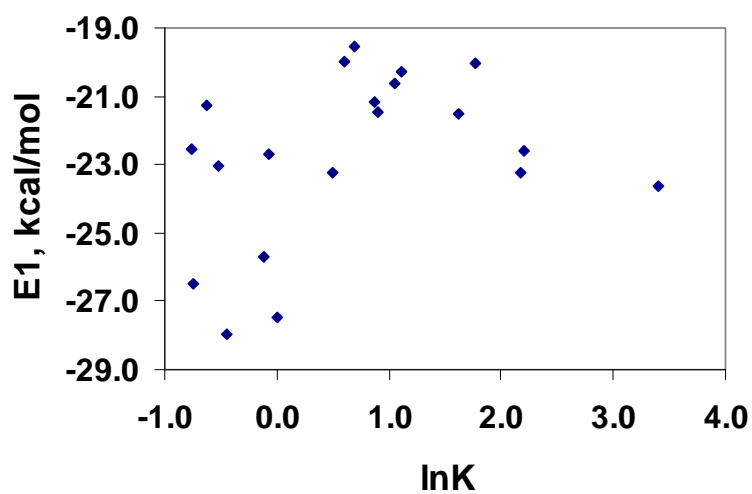**Table 3.** KCN1 and active analogs with multiple repeats.



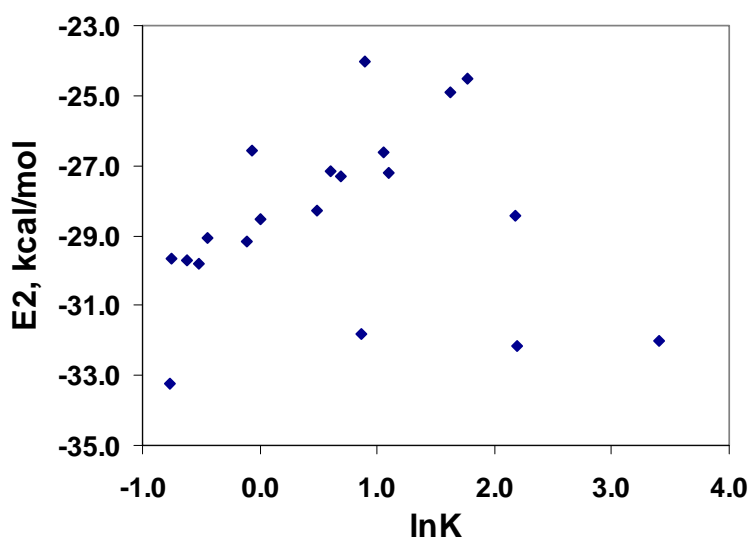**Figure 20.** Linear regression attempt at Site 1.



**Figure 21.** Linear regression attempt at Site 2.

### 4.3.4. Evaluation of calculated energy values by ROC

While a quantitative regression model with a high correlation $r^2$ value would definitely be highly valuable in this project, a model with a poor $r^2$ value can still be useful, if it can enrich the true positives from the false positives. Since the FRESH protocol incorporates a vHTS component, a more stringent selection scheme is allowed which trades false negatives (sensitivity) for better selectivity against false positives. In addition, at the lead-optimization stage, the synthetic step is rate-limiting. As a result, avoiding false positives is considered more valuable than losing possible false negatives. Therefore, the ROC (receiver operating curve) was employed to assess whether the estimated binding energy values at the two sites can be useful for modeling enrichment of the active compounds. The compounds listed in **Table 3** defined as active with true positive rates (sensitivity) were plotted against false positive rates (selectivity) with the thresholds of estimated binding energy values decreasing. **Figures 22** and **23** illustrate the ROCs at Sites 1 and 2.
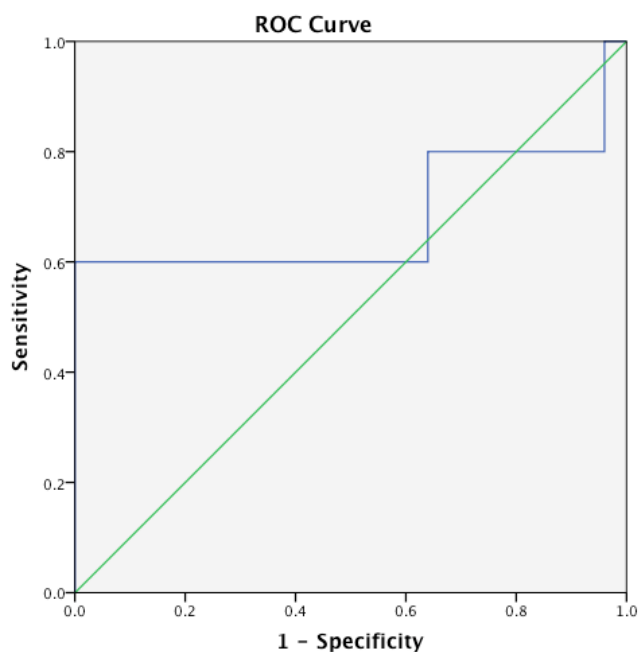


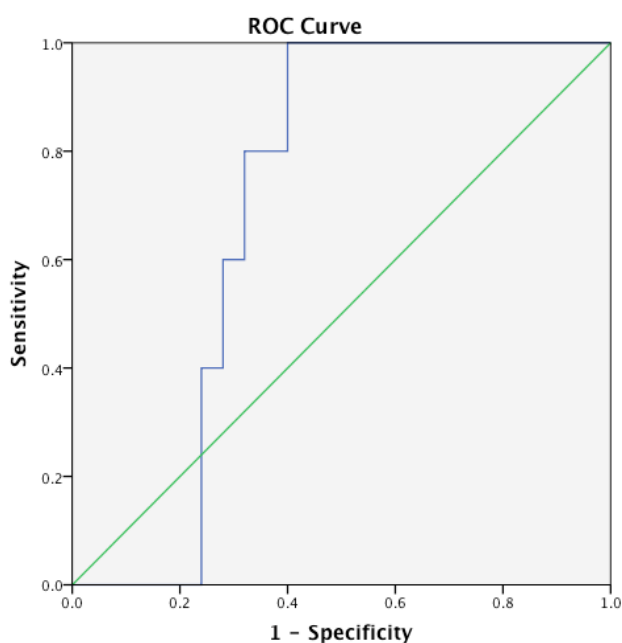**Figure 22.** ROC at Site 1. AUC = 0.68

**Figure 23.** ROC at Site 2. AUC = 0.70

The AUCs (area under curve) for both sites are around 0.7, indicating that both the estimated energy values at Sites 1 and 2 can differentiate active compounds from inactive analogs to some extent. Therefore, the intended plan is to incorporate the energy values at both sites (E1 and E2) for prioritizing. There will be two lists of prioritized structures after this round of selection, and the structures that appear on both lists will be assigned a higher priority. A test for this plan was carried out on the existing 30 compounds, with the MM-GBSA cutoffs for E1 and E2 at -26.5 kcal/mol and -27.5 kcal/mol. Only 3 compounds remained: KCN1, 1601 and 2609, all of which are true positives and included in **Table 3**. It is noteworthy that 1601 and 2609 are the two most potent analogs among all the test molecules. This result further supports the plan to incorporate both E1 and E2.

## 4.4. Application of the FRESH approach

### 4.4.1. Protocol design for the FRESH approach

The FRESH screening approach was applied according to the synthetic route in **Figure 24**. The $R_1$ group originates from a sulfonyl chloride building block, the $R_2$ group from an amine and the $R_3$ group from an aldehyde. The protocol is able to handle variations on all three R groups and enumeration of structures in the billions. However, since the most potent ones are from the Group II analogs, the $R_1$ group was kept constant and $R_3$ only varied a little, while only $R_2$ were screened from the commercial library.
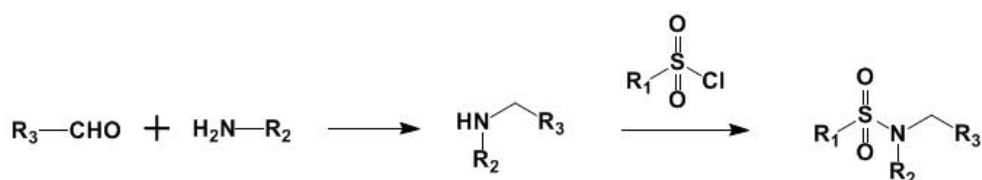


**Figure 24.** Synthetic route for KCN1 and its analogs

Within the context describe above, the FRESH protocol for additional Group II KCN1 analogs was thus designed as shown in **Figure 25**. Starting from the ChemNavigator commercial library, the corresponding building blocks for $R_2$ (amine) were collected. Additional modules were incorporated in the Pipeline Pilot flow to selecting amide and duplicated structures. Like the FRESH protocol employed in the previous chapter, fragments were prioritized by filtering the ones with functional groups or structures that are known to be reactive, toxic, unstable in serum/plasma, prone to hydrolysis, cause false positives in screening or introduce a region-selectivity issue.[25] Other property/potency selection criteria incorporate in the protocol are the following: [2,3,20]

1. Pipeline Pilot: MW < 501, Number of O and N $\leq$ 10, Number of H-donors $\leq$ 5, LogP $\leq$ 5, PSA $\leq$ 90, LogS $\geq$ -6.0

2. Qikprop: LogBB $\geq$ -0.5, Number of Primary Metabolites $\leq$ 6, Pcaco $\geq$ 500, PMDCK $\geq$ 500

3. Docking: Site 2 MM-GBSA value $\leq$ -27.5 kcal/mol, Site 1 MM-GBSA value $\leq$ -26.5 kcal/mol
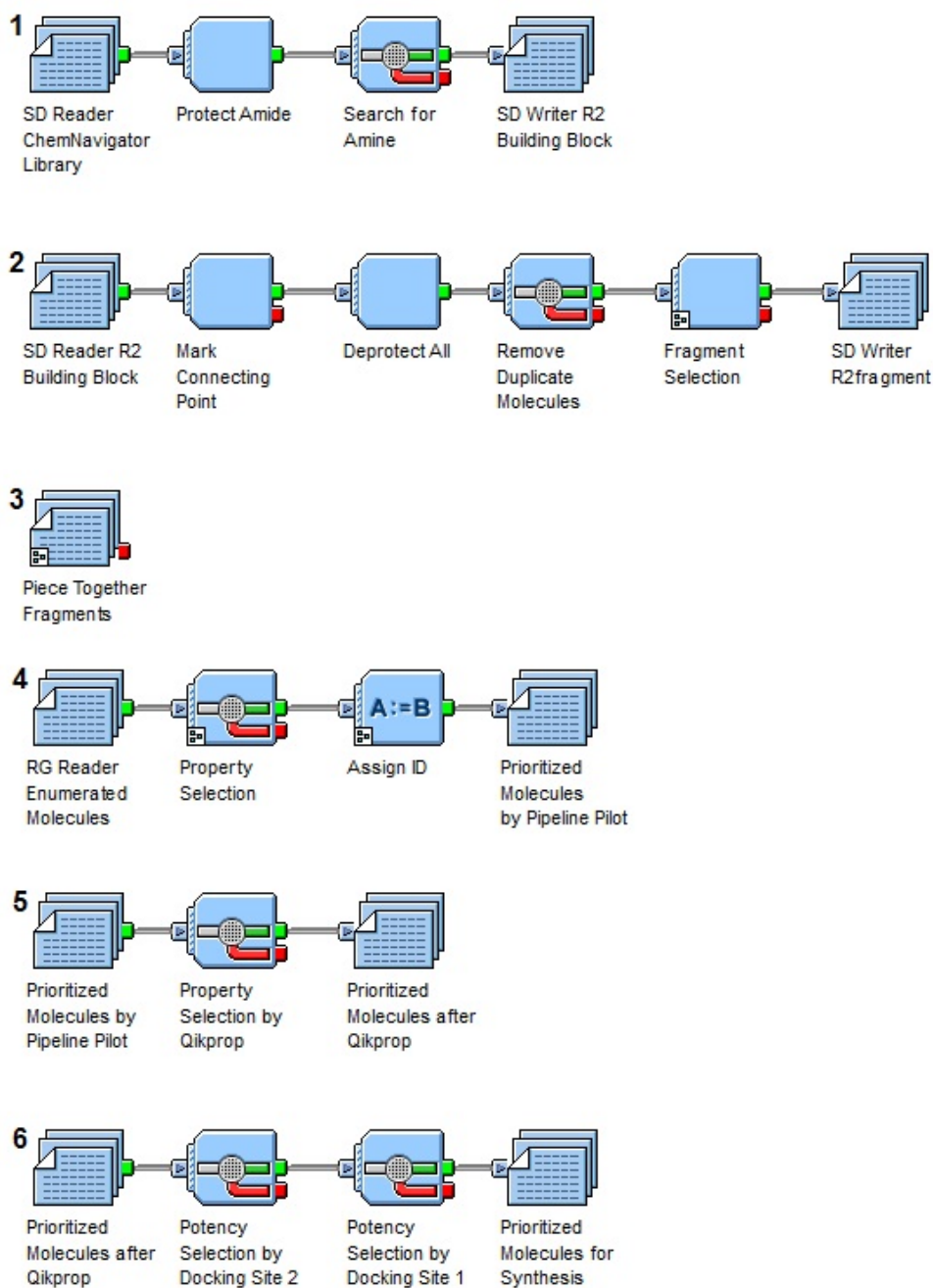


**Figure 25.** The FRESH protocol (main interface) for prioritizing KCN1 analogs. Some sub-protocol components are not shown.

**4.4.2. Resulting structures**

**Table 4** illustrates the final list of FRESH-generated structures. All six compounds are chiral except Compound 270. Upon further investigation on the Internet for the building block information, suitable intermediates for Compounds 95 and 98 are available both as racemates and optically-pure enantiomers. Based on cost, the decision was made to work with racemic mixtures first. **Figure 26** shows the initial bio-test results for Compound 95, the single predicted structure so far tested. Compared to a smooth curve of 2609, the compound demonstrates an abnormal wave-shaped curve. This compound is likely to be a false positive. Further work on modeling of ligand-protein interaction is still required.
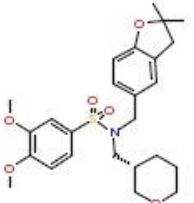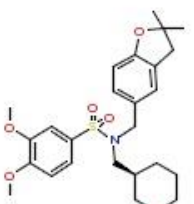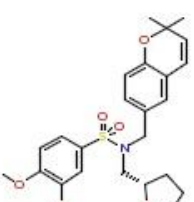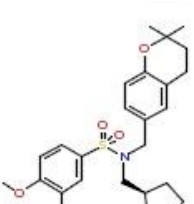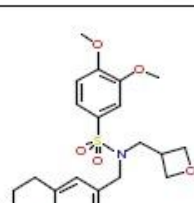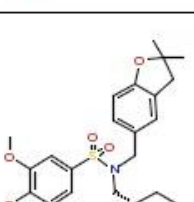
| Molecule | Name | MMGBSA_Site1 | MMGBSA_Site2 | predicted IC50 / nM |
|----------|------|--------------|--------------|---------------------|
|  | 93 | -28.7 | -27.9 | 834 |
|  | 94 | -28.8 | -30.2 | 241 |
|  | 95 | -29.9 | -33.3 | 48 |
|  | 98 | -28.2 | -30.3 | 237 |
|  | 270 | -27.8 | -31.4 | 130 |
|  | 280 | -30.3 | -33.0 | 57 |

**Table 4.** Result of the FRESH approach for KCN1 analogs
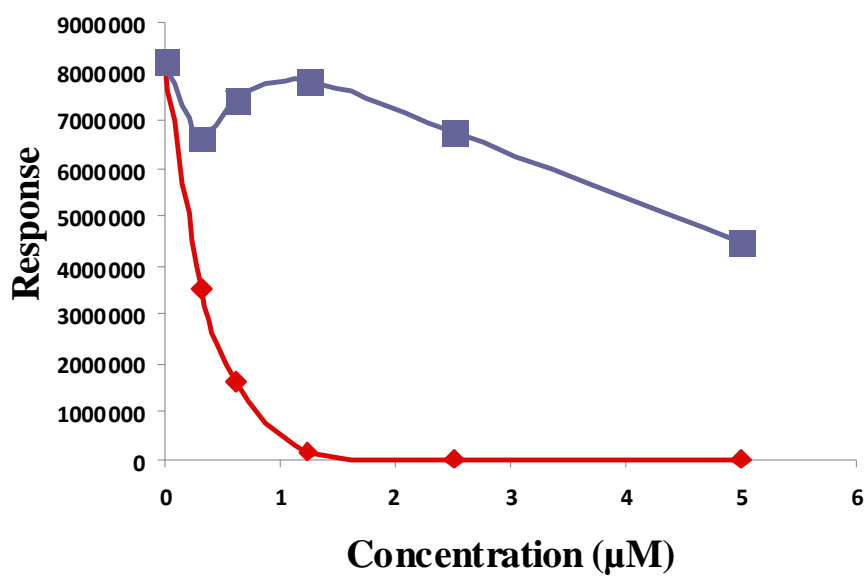
**Figure 26.** The initial bio-test results for Compound 95 (blue). 2609 (red) was used as a reference.

# Chapter 5: Curcumin analogs as Pleiotropic Kinase Blockers

## 5.1. Project background

Curcumin is the principle component of the root powder of *Curcuma longa*. It is consumed by a vast population in India as a spicy flavor ingredient. In addition, its yellow to curry color also makes it a candidate for food coloring. Curcumin plays a crucial role in traditional medicine for its therapeutic effects against rheumatoid arthritis, indigestion, liver disease like jaundice and insect bites.[48] Curcumin has attracted the attention of medicinal chemists recently due to its anti-tumor activity and relative low toxicity. However, low potency, poor bioavailability and fast metabolism have restricted its clinical application.

Attempts have been made to improve the potency and ADME properties of curcumin. Various analogs have been synthesized at Dr. Liotta's lab, with EF24 as the lead compound.[48] The analog has demonstrated enhanced anti-tumor activity relative to curcumin while remains tolerated to the cell. Nevertheless, the detailed mechanism of action remains unexplored. In the project, a list of curcumin analogs are evaluated against a series of kinases that appears crucial for various tumor cell signaling pathways. My task in this project is using molecular modeling to rationalize the activity data. Since EF24 and other analogs still demonstrate bioavailability issues, improvements on the ADME properties are still needed. The FRESH approach has incorporated ADME selection scheme, so it will be applied for screening alternative analogs.

## 5.2. Modeling of curcumin analog against several kinases

A series of curcumin analogs were synthesized by Dr. Terry Moore (Emory University) and screened against a variety of kinases by Andrew Brown (WCI Ph.D. candidate). **Table 5** displays the experimental IC$_{50}$ values, with structures shown in **Figure 27**. AKT-1, AKT-2,

ERK2, NEK1 and VEGFR2 demonstrate some of the most potent $IC_{50}$ values as reflections of binding. Since crystal structures are available for AKT-1, AKT-2 and VEGFR2, the corresponding ligand-protein complexes were investigated first.

| Kinase | EF 24 | EF 31 | UBS 109 | SEF 31 |
|---|---|---|---|---|
| AKT1 (PKB α) | 0.783 | 0.023 | 2.6 | 3.6 |
| AKT-2 | 0.721 | 0.019 | 1.9 | 3.3 |
| IKBKB (IKKβ) | 95 | 4.6 | 20 | 35 |
| RPS6K1 | 19.5 | 1.4 | 6 | 11.9 |
| AMPK (A1/B1/G1) | 104 | 2.1 | 9 | 12.8 |
| RAF1 *cascade | 24 | 1.6 | 6.5 | 8.5 |
| MEK1 *cascade | 28 | 2 | 9 | 13 |
| ERK2 | 13.2 | 0.824 | 1.8 | 2.2 |
| NEK1 | 77 | 0.504 | 3.5 | 13.8 |
| KDR (VEGFR2) | 0.779 | 0.664 | 1.3 | 6.5 |
| MAPK14 (p38α) | 92 | 35 | 131 | n/a |
| MAPKAPK2 | 27 | 6.7 | 9.8 | 15.6 |
| SRC | 39 | 9.7 | | |

**Table 5.** $IC_{50}$ values (unit: μM) of a series of EF24 analogs against various kinases.
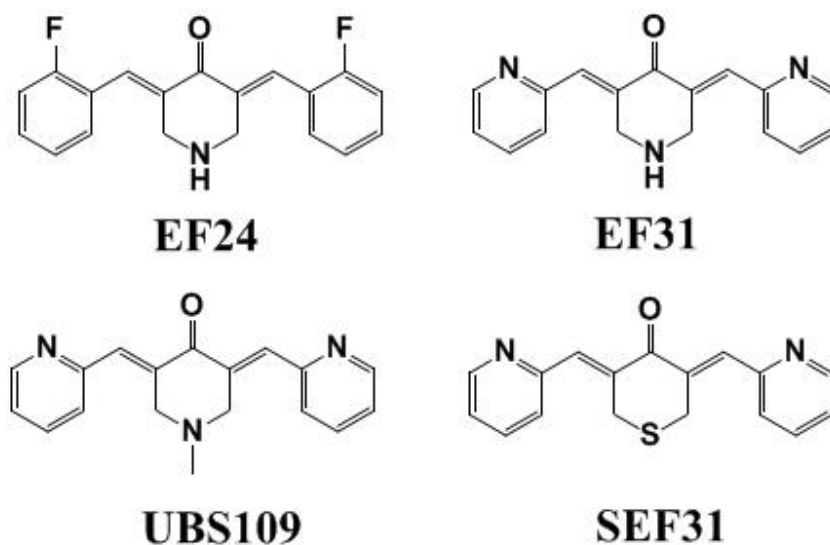


**Figure 27.** Structures of curcumin analogs: EF24, EF31, UBS109 and SEF31

### 5.2.1. Modeling of AKT-1 and AKT-2

The sequences around the ATP pockets of AKT-1 and AKT-2 were compared by BLAST. The result demonstrates that the kinase domains of the two proteins possess 82% identity and 93% similarity. **Figure 28** also reveals that the residues surrounding the ATP pockets (squared in black) are identical. It is therefore not surprising that the $IC_{50}$ values of AKT-1 and AKT-2 are likewise identical.



**Figure 28.** Sequences of aligned AKT-1(lower row) and AKT-2 (upper row with residue numbers). The residues around the ATP pocket are squared in black

To rationalize the IC50 values of the ligands, the curcumin analog ligands were modeled within the ATP binding sites. The choice of the binding site was made based upon kinetic experiments being conducted by Andrew Brown that a strong component of ligand action derives from ATP competition. Since AKT-1 and AKT-2 are biologically and structurally coincident, AKT-2 was selected for ligand comparison. First, we investigated whether a competitive mechanism alone can rationalize the $IC_{50}$ data. The docking procedure is similar to those described in Chapters 3 and 4. The protein receptor was extracted from the crystal structure (PDB code: 3E88) and processed by Maestro Protein Preparation Wizard.[49] The Docking grids were generated around the ATP binding pocket, and ligand-docking was

accomplished with Glide. The top 20 poses for each ligand were subsequently submitted to Prime MM-GBSA rescoring.[46,47] **Figure 29** illustrates the top pose re-ranked by MM-GBSA for EF31 in AKT-2. One of the pyridine nitrogen atoms forms a nice hydrogen bond with Thr292. The nitrogen atom on the central six-membered-ring is positioned close to Glu236 and Glu279 which is capable of forming favorable electrostatic interactions with the NH groups. The carbonyl group is directed toward Lys181 which can also contribute to coulombic association. EF24's top pose is very similar to that of EF31. The only structural difference between EF31 and EF24 is the replacement of *ortho*-N on the pyridine rings with C-F. As **Figure 30** illustrates, the fluorine atom can also form a hydrogen bond with Thr292, but the interaction is weaker and the angle is less favorable than that for EF31. In addition, the distance between the fluorine on EF24 and the oxygen on Thr292 is shorter than the corresponding sum of Van der Waals radii indicating an unfavorable steric clash. The weaker H-bond interaction and steric clashes of EF24 with AKT-2 probably contribute to the higher $IC_{50}$ values of EF24 compared to EF31.
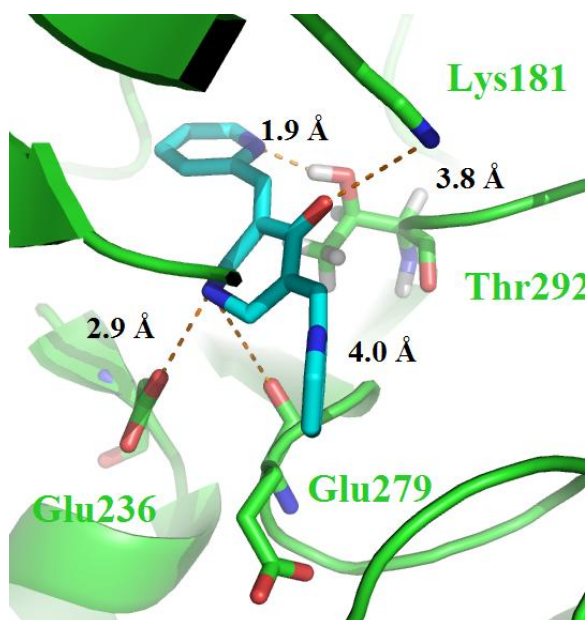


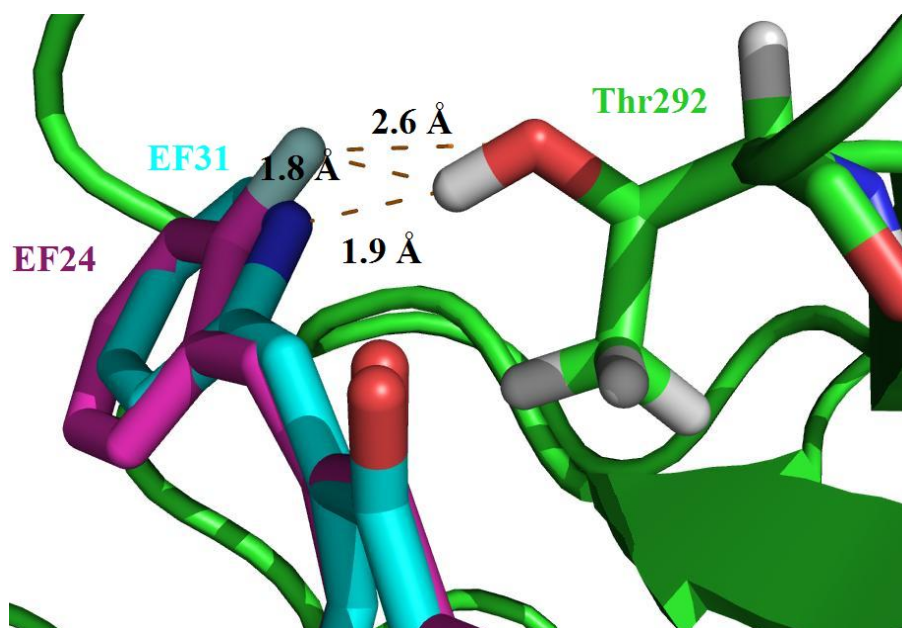**Figure 29.** Top pose of EF31 on AKT-2.

**Figure 30.** Top pose of EF31 and EF24 on AKT-2.

For UBS109 and SEF31, the bulky N-methyl group and sulfur atom, respectively, have

forced the molecules into different poses as illustrated in **Figures 31** and **32**, in which the

ligands lose favorable H-bond and electrostatic interactions. Since these curcumin analogs

have pka values around 6.5, it is possible that some protonated forms exist. **Figures 33** and

**34** demonstrate the top docking poses for protonated UBS109 with N-Me groups in axial and

equatorial conformations.  The axial conformer adopts a pose similar to EF31, which forms

favorable electrostatic interactions with the two glutamic acids, but it exists in an

energetically unfavorable conformer. The force-field-based energy calculations reveal that

for charged UBS109, the energy difference between the axial and equatorial conformers is

5.0 kcal/mol.  Since the predicted MM-GBSA binding energy for the axial conformer is -48.5

kcal/mol, 6.7 kcal/mol more stable in complex than  the equatorial conformer, the net

stability for the axial form is 1.7 kcal/mole (6.7 – 5.0 kcal/mol). The calculation demonstrates

that under protonation conditions, the N-Me group can adopt the axial conformation upon

binding to AKT-2. However, compared to the corresponding -51.5 kcal/mol binding energy

for EF31, it is considerably less tightly bound. Thus, the reduced activities of UBS-109

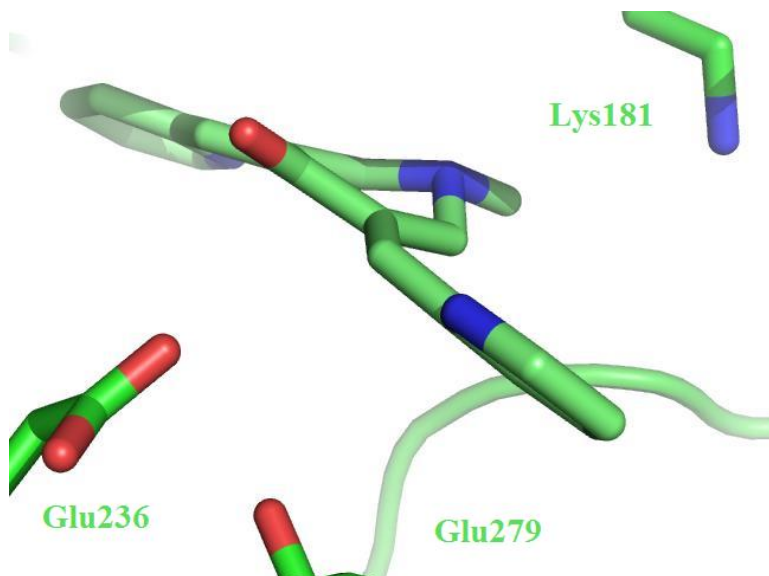relative to EF31 and EF24 can be rationalized.



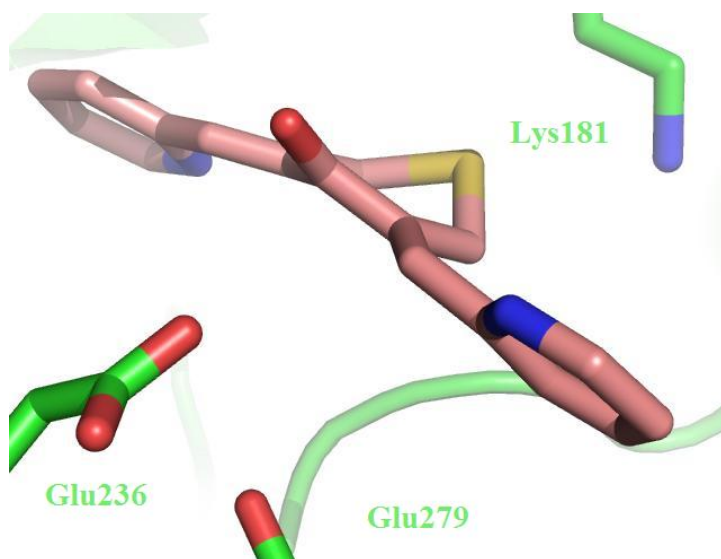**Figure 31.** Top pose of UBS109 (unprotonated) on AKT-2



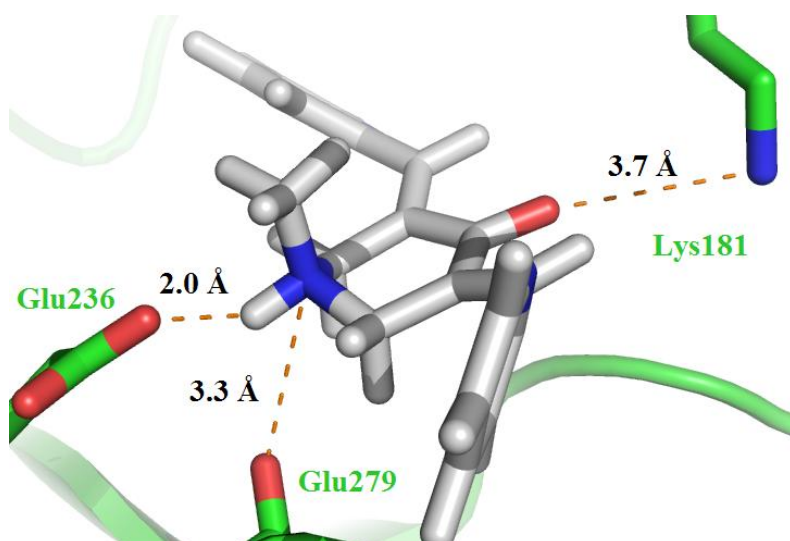**Figure 32.** Top pose of SEF31 on AKT-2

**Figure 33.** The top docking pose for protonated UBS109 with the N-Me group in the axial conformation
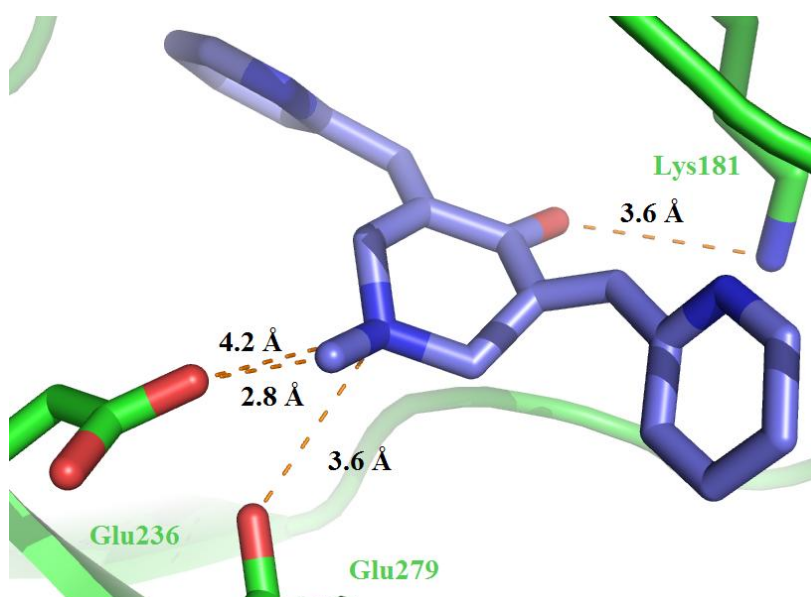


**Figure 34.** The top docking pose for protonated UBS109 with N-Me groups in the equatorial conformation

Overall, the docking results can provide a qualitative explanation of the variation of $IC_{50}$ values in AKT-2. However, a linear regression of predicted MM-GBSA binding free energies and experimental $lnIC_{50}$ values provided no QSAR correlation. Since all the curcumin analogs investigated herein carry the α,β-unsaturated ketone substructures, it is possible that

free cysteines on the AKT kinases form a covalent-bond with the analogs by Michael

addition. In this way the mixed-inhibition observed by collaborator Andrew Brown can be

understood. For both AKT-1 and AKT-2, a cysteine (Cys310 for AKT-1 and Cys311 for

AKT-2, **Figure 35**) is located near the substrate binding site. The principle of the assay used

in this project can be described as follows: A FRET (fluorescence resonance energy transfer)

pair is placed on the substrate of the kinases. In addition to the kinase, another enzyme that is

capable of cleaving the substrate and breaking the FRET pairs apart is also added to the

solution. If only ATP exists in the buffer solution, the kinase will phosphorylate the substrate

and protect it from the cleavage of the other enzyme. But if a competitive inhibitor occupies

the ATP pocket, the substrate won't be phosphorylated and the other enzyme will cleave it,

breaking the FRET pairs apart and generating a fluorescence signal. Thus, it is reasonable to

speculate that the formation of covalent bond with the cysteine will interrupt substrate

binding, preventing the cleavage of the substrate and consequently the separation of the

FRET pairs. This can also lead to an apparent high $IC_{50}$ value, but, unfortunately, it hinders

quantification of the non-covalent binding affinity of the ATP competitive inhibitor.
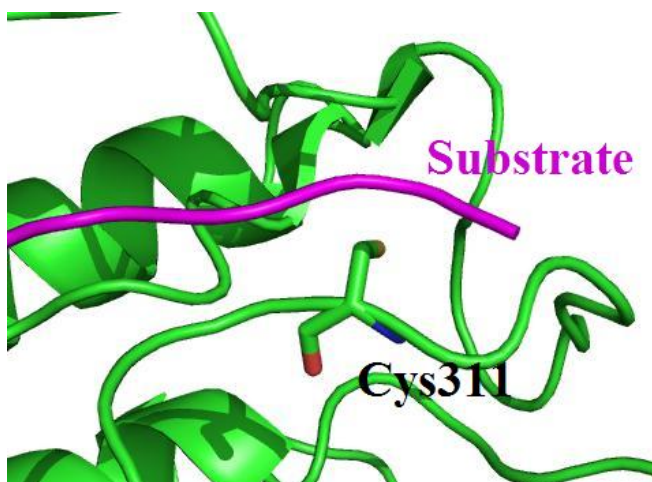


**Figure 35.** Cys311 in the cleft where the subtrate binds

It is worth mentioning that UBS109 routinely demonstrates more potency in the cell-based assays compared to EF24 and EF31. However, the $IC_{50}$ values of EF24 and EF31 in **Table 5** are generally lower than that of UBS109. One possible explanation is that other proteins are also involved in the cellular environment, and UBS109 is a more potent inhibitor for them. Antagonizing AKT-1 and AKT-2 may not play a major role in the anti-tumor effect of the curcumin analogs. Such defining experiments have not yet been carried out on cells treated with the analogs.

### 5.2.2. Modeling of VEGFR2

It is noteworthy that the IC50 values for the four ligands on VEGFR2 are much more similar than observed for the AKT kinases. The four structures were docked into VEGFR2 (PDB code: 2XIR) as described above for AKT-2.[50] The attempt to devise a QSAR correlation between estimated binding affinity by MM-GBSA and experimental $IC_{50}$ values failed again. A close examination of the docking pose of EF24 on VEGFR2 reveals two cysteines inside the ATP pocket. One of these, Cys1045, adopts an orientation that favors covalent Michael addition to the α,β-unsaturated ketone moiety in all the compounds (**Figure 36**). It is highly possible that covalent binding dominates the binding affinities of the ligands for this kinase and thereby rationalizes and can probably rationalize the relatively close $IC_{50}$ values of the analogs.
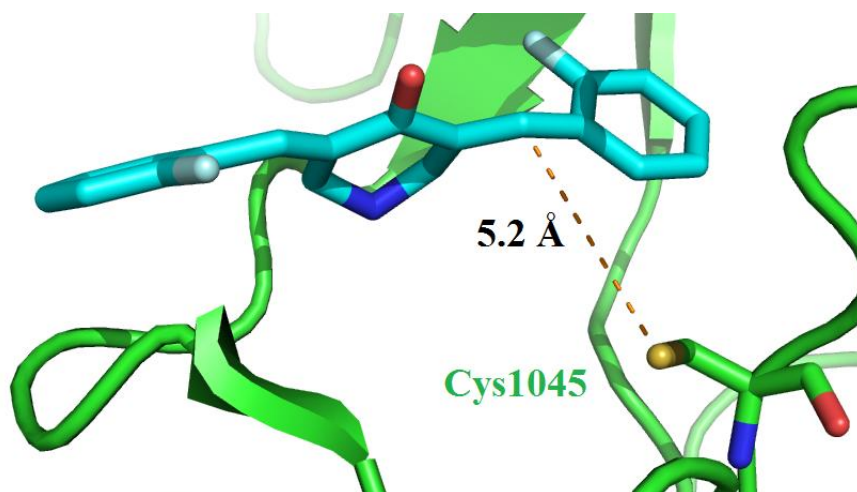
**Figure 36.** Top pose of EF24 on VEGFR2

## 5.3. Anticipated FRESH analysis

Although the developing of QSAR model is still in progress, the design for other components in the FRESH approach has already been considered. The synthetic route for the curcumin analogs is shown in **Figure 37**. The $R_1$ group comes from an aldehyde. In addition to fragments from building blocks in the commercial library, the phenyl ring scan method can also product fragments compatible with the route. The central core (ketone) can also be varied. The protocol for exploring additional analogs is illustrated in **Figure 38**.
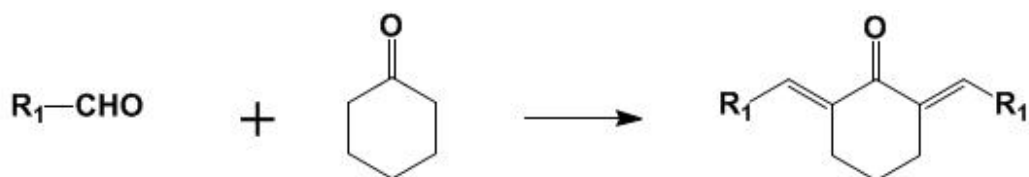


**Figure 37.** The synthetic route for additional curcumin analogs
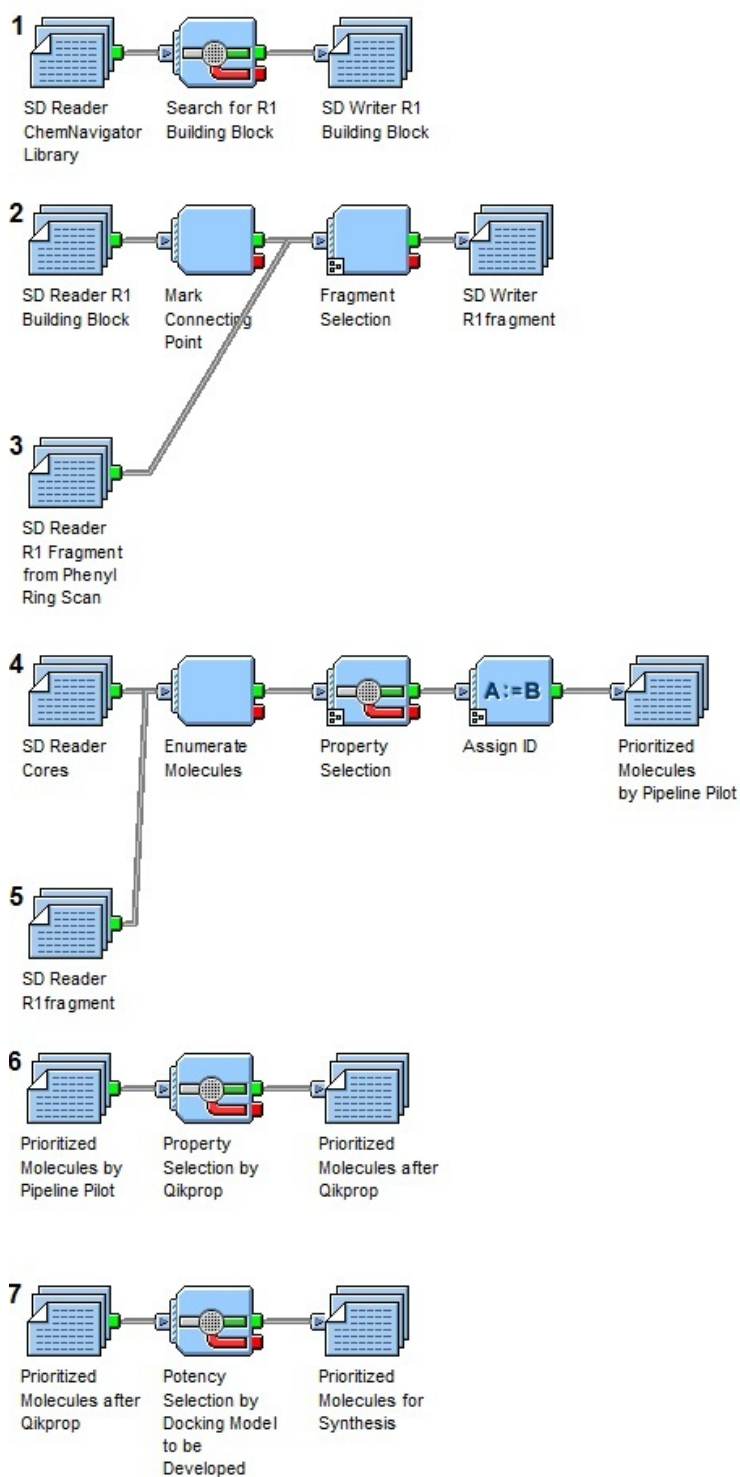
**Figure 38.** The FRESH protocol (main interface) for prioritizing curcumin analogs. Some sub-protocol components are not shown.

# Chapter 6: Conclusions and Future Work

Chapter 3 and Chapter 4 describe how the FRESH approach is able to provide a list of structures within a given project framework which satisfy multiple requirements including potency, physical properties and ADME aspects. The method appears to be valuable for multi-dimensional optimization during the drug discovery process. It can balance various factors contributing to the pharmacological properties of the molecule, while remaining synthesis-friendly for the synthetic chemists.

For the NET/SERT dual inhibitor project in Chapter 3, additional work on the potency prioritizing scheme is needed to better eliminate false positives. In addition to the structure-based methods described in this chapter, some ligand-based approaches using various 2D/3D descriptors and statistical methods (example: decision trees) can be useful for developing a QSAR as an alternative to linear regression. 2D ligand-based methods, like those involving Hammett sigma constants, are worth investigating, especially given the fact that the Ki values are sensitive to the substitution patterns on the phenyl ring. The FRESH approach employs a vHTS component which permits low sensitivity analysis. As a result, a scheme combining the QSARs of structure and ligand-based methods, which may have reduced sensitivity for true positives, can also be applicable if better selectivity against false negatives is achieved. Further modifications are also required to avoid potential problems in the context of oral administration, mainly substituting the ester with alternative groups which are less prone to hydrolysis, but capable of passing the BBB. The BROOD program will be applied to identify alternatives. The amine moiety will also be explored for better BBB penetration possibly by examining pro-drugs. Since the $K_i$ values are sensitive to the substitution pattern on the phenyl ring, the "phenyl ring scan" approach can also be adopted for additional $R_2$ fragments. The predicted hERG binding affinity was not included in the selection scheme. However, as
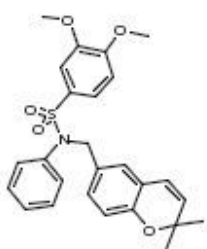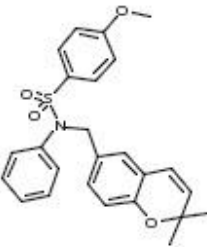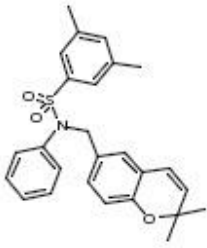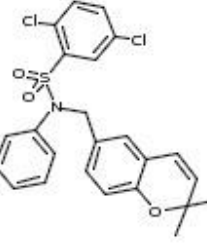
the potential drug is not used against an acute and life-threatening condition, the toxicity issue raised hERG binding is definitely a concern, and the predicted hERG binding affinity will be incorporated in future.
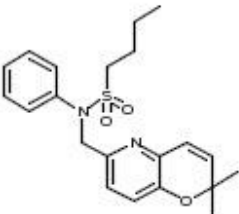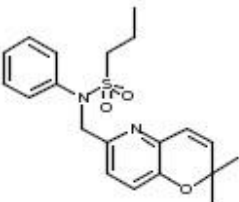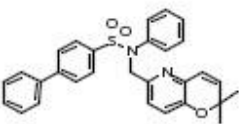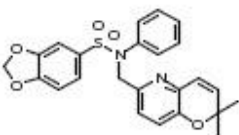
For the KCN1 analog project in Chapter 4, further optimization of the prioritizing scheme is still in progress. Like stated in the last paragraph, various 2D/3D ligand-based methods, with possible combinations to the current method, are also alternatives. Since the lead compound KCN1 originated from a focused-library screening, this can serve as additional data for developing the prioritizing scheme. Besides, the target ranges for MW are MW< 450 and LogBB >0, which is yet to achieve. The BROOD isosteric replacement approach will be used to find shape and electrostatic alternatives to the benzopyran group, the dimethoxylphenyl ring and the sulfonyl amide. The predicted hERG binding affinity was not included in the selection scheme. However, if the project requires this, the predicted hERG affinity will be incorporated to avoid the "danger zone" ($IC_{50} > 10$ μM).
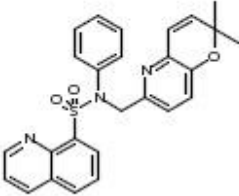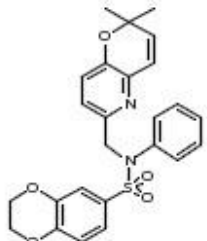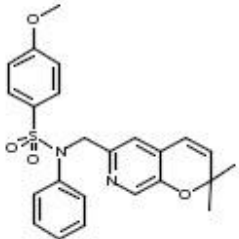
The model in Chapter 5 is able provide a qualitative rationale for the variations of $IC_{50}$ values of curcumin analogs, but further investigation is required to obtain a QSAR before application of the FRESH approach.

# Appendix I: A series of KCN1 analog with experimental IC$_{50}$ values and predicted MM-GBSA values
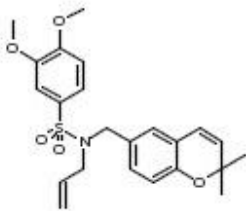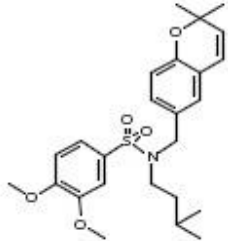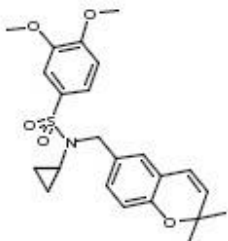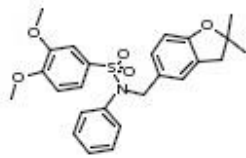
Unit for MMGBSA: kcal/mol, IC$_{50}$: $\mu$ M   KCN1 is Compound 1.
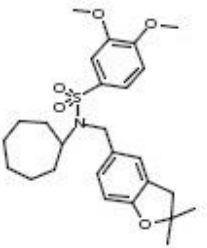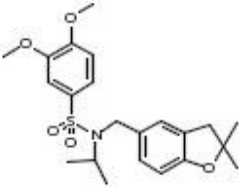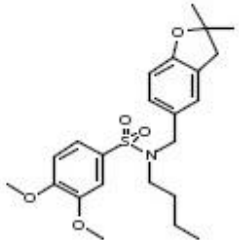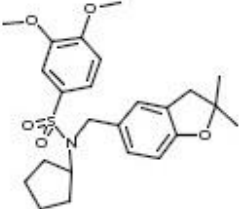
| Molecule | Name | MMGBSA_Site1 | MMGBSA_Site2 | IC50 | IC50KCN1 | Group |
|---|---|---|---|---|---|---|
|  | 1 | −27.5 | −28.5 | 0.7 | 0.7 | I |
|  | 510 | −23.4 | −29.7 | 0.6 | 1.2 | I |
|  | 511 | −22.3 | −27.3 | 0.5 | 1.2 | I |
|  | 512 | −24.9 | −28.6 | 2.1 | 1.7 | I |

| Molecule | Name | MMGBSA_Site1 | MMGBSA_Site2 | IC50 | IC50KCN1 | Group |
|----------|------|--------------|--------------|------|----------|-------|
|  | 2614 | -24.2 | -26.5 | 5 | 0.54 | I |
|  | 2615 | -20.6 | -24.1 | 6.4 | 0.54 | I |
|  | 2617 | -26.1 | -24.7 | 3.4 | 0.54 | I |
|  | 2618 | -25.8 | -25.6 | 6.5 | 0.54 | I |

| Molecule | Name | MMGBSA_Site1 | MMGBSA_Site2 | IC50 | IC50KCN1 | Group |
|----------|------|--------------|--------------|------|----------|-------|
|  | 2619 | -18.8 | -23.7 | 0.9 | 0.3 | I |
|  | 2620 | -19.7 | -24.1 | 0.9 | 0.3 | I |
|  | 3601 | -20.5 | -25.9 | 1.4 | 0.52 | I |

| Molecule | Name | MMGBSA_Site1 | MMGBSA_Site2 | IC50 | IC50KCN1 | Group |
|---|---|---|---|---|---|---|
|  | 1 | -27.5 | -28.5 | 0.7 | 0.7 | II |
|  | 501 | -20.0 | -24.5 | 3.1 | 0.53 | II |
|  | 502 | -21.4 | -24.0 | 1.3 | 0.53 | II |
|  | 504 | -20.0 | -27.2 | 3.5 | 1.9 | II |

| Molecule | Name | MMGBSA_Site1 | MMGBSA_Site2 | IC50 | IC50KCN1 | Group |
|----------|------|--------------|--------------|------|----------|-------|
|  | 505 | -23.3 | -28.3 | 3.4 | 1.9 | II |
|  | 506 | -20.3 | -27.2 | 1.6 | 0.53 | II |
|  | 508 | -20.6 | -26.6 | 1.5 | 0.53 | II |
|  | 1601 | -28.0 | -29.1 | 0.5 | 1.2 | II |

| Molecule | Name | MMGBSA_Site1 | MMGBSA_Site2 | IC50 | IC50KCN1 | Group |
|---|---|---|---|---|---|---|
|  | 1602 | -23.6 | -32.0 | 9.1 | 0.3 | II |
|  | 1603 | -21.5 | -24.9 | 1.5 | 0.3 | II |
|  | 1604 | -19.5 | -27.3 | 0.6 | 0.3 | II |
|  | 1606 | -21.3 | -29.7 | 0.4 | 0.52 | II |

| Molecule | Name | MMGBSA_Site1 | MMGBSA_Site2 | IC50 | IC50KCN1 | Group |
|----------|------|--------------|--------------|------|----------|-------|
|  | 2601 | -25.7 | -29.2 | 1.3 | 1.4 | II |
|  | 2602 | -22.7 | -26.6 | 0.9 | 1.4 | II |
|  | 2604 | -22.5 | -33.2 | 0.6 | 1.4 | II |
|  | 2605 | -23.0 | -29.8 | 0.8 | 1.35 | II |

| Molecule | Name | MMGBSA_Site1 | MMGBSA_Site2 | IC50 | IC50KCN1 | Group |
|----------|------|--------------|--------------|------|----------|-------|
|  | 2606 | -23.2 | -28.5 | 6.2 | 0.7 | II |
|  | 2607 | -22.6 | -32.1 | 6.6 | 0.7 | II |
|  | 2608 | -21.1 | -31.8 | 0.7 | 0.4 | II |
|  | 2609 | -26.5 | -29.7 | 0.25 | 0.59 | II |
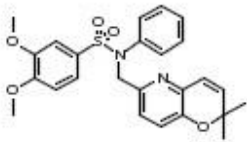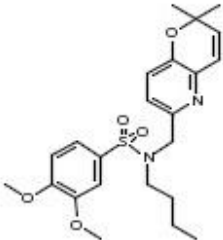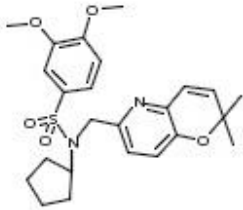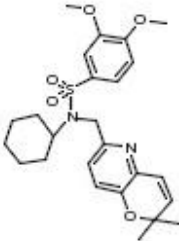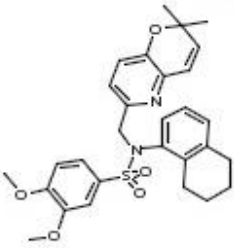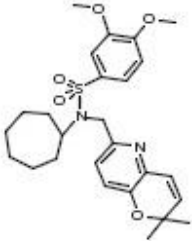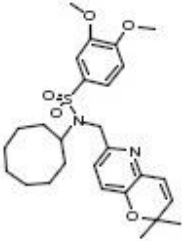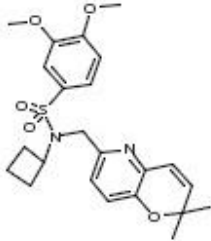
# References

(1) Bohacek R. S.; McMartin C.; Guida W. C. The art and practice of structure-based drug design *Med. Res. Rev.* **1996**, 16, 3–50

(2) Lipinski C. A.; Lombardo F.;Dominy B. W.; Feeney P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **1997**, 23, 3–25

(3) Jorgensen W. L. Efficient drug lead discovery and optimization *Acc. Chem. Res.* **2009**, 42, 724-733

(4) Morelli X.; Bourgeas R.; Roche P. Chemical and structural lessons from recent successes in protein-protein interaction inhibition (2P2I) *Curr. Opin. Chem. Biol.* **2011**, 15, 475-481

(5) Leeson P. D.; Springthorpe B. The influence of drug-like concepts on decision-making in medicinal chemistry *Nature Reviews Drug Discovery* **2007**, 6, 881-890

(6) Morphy R. The influence of target family and functional activity on the physicochemical properties of pre-clinical compounds *J.Med.Chem* **2006**, 49, 2969-2978

(7) Oprea T. I. Current trends in lead discovery: are we looking for the appropriate properties? *J. Comp Aided Mol. Design* **2002**, 16, 325–334

(8) Levin V. A. Relationship of octanol/water partition coefficient and molecular weight to rat brain capillary permeability *J.Med.Chem* **1980**, 23, 682–684

(9) Atkinson F.; Cole S.; Green C.; van de Waterbeemd H. Lipophilicity and other parameters affecting brain penetration *Curr. Med. Chem-Central Nervous System Agents* **2002**, 2, 229-240

(10) Leeson P. D.; Davis A. M. Time-related differences in the physical property profiles of oral drugs *J.Med.Chem* **2004**, 47, 6338–6348

(11) Kelder J.; Grootenhuis P. D. J.; Bayada D. M.; Delbressine L. P. C.; Ploemen J. P. Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs *Pharm. Res.* **1999**, 16, 1514-1519.

(12) Hansch C.; Leo A. J. *Substituent constant for correlation analysis in chemistry and biology.* New York: Wiley, **1979**

(13) Perola, E.; Walters W. P.; Charifson P. S. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *PROTEINS* **2004**, 56, 235-249

(14) Friesner R. A.; Banks J. L.; Murphy R. B.; Halgren T. A.; Klicic J. J.; Mainz D. T.; Repasky M. P.; Knoll E. H.; Shelley M.; Perry J. K.; Shaw D. E.; Francis P.; Shenkin P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *J.Med.Chem.* **2004**, 47, 1739-1749

(15) Halgren T. A.; Murphy R. B.; Friesner R. A.; Beard H. S.; Frye L. L.; Pollard W. T.; Banks J. L. Glide: a new approach for rapid, accurate docking and scoring. 2. enrichment factors in database screening. *J.Med.Chem.* **2004**, 47, 1750-1759

(16) Zhou Z.; Felts A. K.; Friesner R. A.; Levy R. M. Comparative performance of several flexible docking programs and scoring functions: enrichment studies for a diverse set of pharmaceutically relevant targets. *J. Chem. Inf. Model.* **2007,** *47,* 1599-1608

(17) Guimaraes C. R. W.; Cardozo M. MM-GB/SA Rescoring of Docking Poses in Structure-Based Lead Optimization *J. Chem. Inf. Model.* **2008**, 48, 958–970

(18) Kubinyi, H. Opinion: Drug Research: Myths, Hype and Reality. *Nat. Rev. Drug Discovery* **2003**, *2*, 665−668.

(19) Kola I.; Landis J. Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discovery* **2003**, 3, 711-715.

(20) QikProp, version 3.2, Schrödinger, LLC, New York, NY, **2009**

(21) http://accelrys.com/products/pipeline-pilot/component-collections/adme-tox.html

(22) http://www.acdlabs.com/products/pc_admet/admetox.php

(23) Herrick T. M.; Million R. P. From the analyst's couch: tapping the potential of fixed-dose combinations *Nat. Rev. Drug. Discov.* **2007**, 6, 513 –514

(24) Sterling J.;Herzig Y.; Goren T.; Finkelstein N.; Lerner D.; Goldenberg W.; Miskolczi I.; Molnar S.; Rantal F.; Tamas T.; Toth G.; Zagyva A.; Zekany A.; Finberg J.; Lavian G.; Gross A.; Friedman R.; Razin

M.; Huang W.; Krais B.; Chorev M.; Youdim M. B.; Weinstock M. Novel dual inhibitors of AChE and MAO derived from hydroxy aminoindan and phenethylamine as potential treatment for Alzheimer's disease *J. Med. Chem.* **2002**, 45, 5260–5270

(25) Rishton G. M. Reactive compounds and in vitro false positives in HTS *Drug Discov. Today* **1997**, 2, 382-384

(26) Congreve M.; Carr R.; Murray C.; Jhoti H. A 'rule of three' for fragment-based lead discovery *Drug Discov. Today* **2003**, 8, 876-877

(27) Johannes C. B.; Le T. K.; Zhou X.; Johnston J. A.; Dworkin R. H. The prevalence of chronic pain in United States adults: results of an Internet-based survey *J. Pain.* **2010**, 11, 1230-1239

(28) Stewart W. F.; Ricci J. A.; Chee E.; Morganstein D.; Lipton R. Lost productive time and cost due to common pain conditions in the US workforce. *J. Am. Med. Assoc.* **2003**, 290, 2443-2454.

(29) Ritzwoller D. P.; Ellis J. L.; Korner E. J.; Hartsfield C. L.; Sadosky A. Comorbidities, healthcare service utilization and costs for patients identified with painful DPN in a managed-care setting *Curr. Med. Res. Opin.* **2009**, 25, 1319-1328

(30) Bruehl S. An update on the pathophysiology of complex regional pain syndrome. *Anesthesiol* **2010**, 113, 713-725

(31) Lee Y. C.; Chen P. P. A review of SSRIs and SNRIs in neuropathic pain *Expert Opin Pharmacother.* **2010**, 11, 2813-2825

(32) Mochizucki D. Serotonin and noradrenaline reuptake inhibitors in animal models of pain *Hum. Psychopharm. Clin.* **2004**, 19, S15-S19

(33) Ikeda T.; Ishida Y.; Naono R.; Takeda R.; Abe H.; Nakamura T.; Nishimori T. Effects of intrathecal administration of newer antidepressants on mechanical allodynia in rat models of neuropathic pain Neurosci. Res. **2009**, 63, 42-46

(34) Ravna A. W.; Sylte I.; Dahl S. G. Structure and localisation of drug binding sites on neurotransmitter transporters *J. Mol. Model.* **2009**, 15, 1155-1164

(35) Dr. Spandan Chennamadhavuni, Ph. D. Thesis, **2011**

(36) Brat D. J.; Kaur B.; Van Meir E. G.; Genetic modulation of hypoxia-induced gene expression and vascular proliferation: relevance to brain tumors *Front Biosci* **2002**. 8, d100-116

(37) Lara, P. C.; Lloret, M,; Clavo, B.; Apolinario, R. M.; Henríquez-Hernández, L. A.; Bordón, E.; Fontes, F.; Rey, A. Severe hypoxia induces chemo-resistance in clinical cervical tumors through MVP over-expression *Radiat Oncol.* **2009**, 4:29

(38) Semenza, G. L. Hypoxia-inducible factors in physiology and medicine *Cell* **2012**, 148, 399-408

(39) Reid-Mooring, S..; Jin, H.; Devi, N. S.; Jabbar, A. A.; Kaluz, S.; Liu, Y.; Van Meir E. G.; Wang, B. Design and synthesis of novel small-molecule inhibitors of the hypoxia inducible factor pathway *J. Med. Chem.* **2011**, 54, 8471−8489

(40) Teague S. J.;Davis A. M.;Leeson P. D.;Oprea T. The design of leadlike combinatorial libraries *Angew. Chem. Int. Ed.* **1999**, 38, 3743–3748

(41) Kuntz I. D.; Chen K.; Sharp K. A.; Kollman P. A. The maximal affinity of ligands *Proc. Natl. Acad. Sci. USA* **1999**, 96, 9997–10002

(42) De Guzman, R. N.; Wojciak, J. M.; Martinez-Yamout, M. A.; Dyson, H. J.; Wright, P. E. CBP/p300 TAZ1 Domain Forms a Structured Scaffold for Ligand Binding *Biochemistry* **2005**, 44, 490-497

(43) Shi Q.; Yin S.; Kaluz S.; Ni N.; Devi N. S.; Mun J.; Wang D.; Damera K.; Chen W.; Burroughs S.; Reid-Mooring S.; Goodman M. M.; Van Meir E. G.; Wang B.; Snyder J. P. Binding Model for the Interaction of Anti-cancer Arylsulfonamides with the p300 Transcription co-factor  *submitted manuscript*

(44) Freedman S. J.; Sun Z. Y.; Poy F.; Kung A. L.; Livingston D. M.; Wagner G.; Eck M. J. Structural basis for recruitment of CBP/p300 by hypoxia-inducible factor-1 alpha *Proc.Natl.Acad.Sci.* **2002**, 99, 5367-5372

(45) Gu J.; Milligan J.; Huang L. E. Molecular Mechanism of Hypoxia-inducible Factor 1α-p300 Interaction *J. Biol. Chem.* **2001**, 276, 3550-3554

(46) Glide, version 5.5, Schrödinger, LLC, New York, NY, **2009**

(47) Prime, version 2.1, Schrödinger, LLC, New York, NY, **2009**

(48) Thomas S. L.; Zhong D.; Zhou W.; Malik S.; Liotta D.; Snyder J. P.; Hamel E.; Giannakakou P. EF24, a novel curcumin analog, disrupts the microtubule cytoskeleton and inhibits HIF-1 *Cell Cycle.* **2008**, 15, 24409-2417

(49) Rouse M. B.; Seefeld M. A.; Leber J. D.; McNulty K. C.; Sun, L.; Miller W. H.; Zhang S.; Minthorn E. A.; Concha N. O.; Choudhry A. E.; Schaber M. D.; Heerding D. A. *Bioorg. Med. Chem. Lett*. **2009**, 19, 1508-1511

(50) Mctigue M.; Wickersham J.; Marrone T. Discovery of the Selective Vegfr Inhibitor Pf- 00337210 *To be published*