

Texas Medical Center Library
DigitalCommons@The Texas Medical Center

UT SBMI Dissertations (Open Access)

School of Biomedical Informatics

Spring 5-3-2012

Advanced Protein Modeling Method: Benchmarking and Applications in Computer- Aided Drug Discovery

Sharangdhar S. Phatak

University of Texas Health Science Center at Houston

Follow this and additional works at: http://digitalcommons.library.tmc.edu/uthshis_dissertations

 Part of the [Medicine and Health Sciences Commons](#)

Recommended Citation

Phatak, Sharangdhar S., "Advanced Protein Modeling Method: Benchmarking and Applications in Computer-Aided Drug Discovery" (2012). *UT SBMI Dissertations (Open Access)*. Paper 25.

This is brought to you for free and open access by the School of Biomedical Informatics at DigitalCommons@The Texas Medical Center. It has been accepted for inclusion in UT SBMI Dissertations (Open Access) by an authorized administrator of DigitalCommons@The Texas Medical Center. For more information, please contact laurel.sanders@library.tmc.edu.

Dissertation

Advanced Protein Modeling Method: Benchmarking and Applications in Computer-Aided
Drug Discovery

By

Sharangdhar S. Phatak M.S.

Graduation Date

May 3rd 2012

APPROVED:

Jack W. Smith M.D. PhD. Chair

Jiajie Zhang Ph.D.

Craig Johnson Ph.D.

Patricia Dolan Mullen Ph.D.

Advanced Protein Modeling Method: Benchmarking and Applications in Computer-Aided
Drug Discovery

A
DISSERTATION

Presented to the Faculty of
The University of Texas
School of Biomedical Informatics
at Houston
in Partial Fulfillment
of the Requirements
for the Degree of
Doctor of Philosophy

by

Sharangdhar S. Phatak M.S.

Committee Members:

Jack W. Smith M.D. Ph.D.¹
Jiajie Zhang Ph.D.¹
Craig Johnson Ph.D.¹
Patricia Dolan Mullen Ph.D.²

¹The School of Biomedical Informatics. ²The School of Public Health

Acknowledgments

This work was supported by the UTHHealth Innovation for Cancer Prevention Research Pre-Doctoral Fellowship, The University of Texas Health School of Public Health – CPRIT grant # RP10153, from January 1, 2011 – May 3, 2012.

Dedication

GOD and FAMILY

A Ph.D. diploma documents the untold hours of reading, research, and scientific labor of the person pursuing and the committee overseeing the process. What it doesn't document is the person's behind the scenes so to speak. I have been well and truly blessed to have an extremely strong support cast that stood steadfastly behind me through these five and a half years. My parents, Mrs. Vrushali Phatak and Mr. Shivanand Phatak, deserve as much as or in fact more credit than I do for reaching this stage. They imbued in me the raw courage to face life sans fear and to transcend any adversities that may come my way. I cannot emphasize enough on the sacrifices they made to advance my life and career. I hope I have made you proud. To my sister Shalmali, I am no longer in graduate school.

My wife, Bhavana Bakshi nee Manasvi Phatak, who had a front row seat to all the highs and lows of my graduate school experiences, deserves a special mention. She never lost the sight of the ultimate goal, risked and sacrificed many of her personal and professional goals in order to support me to reach mine. She patiently put up with my erratic behavior through these years and repeatedly helped me regain my focus. I love you from the bottom of my heart. Now, I need to increase that love for our daughter, who we both are eager to welcome in our world.

My extended family, (Late) Mrs. Pushpa Bakshi, Mr. Bhupendra Bakshi, Mrs. Shweta Sane, Mr. Piyush Sane, Mrs. Priyanka Wadhvani, Mr. Devesh Wadhvani and all the wonderful relatives from Manasvi's side of the family, the Bakshi's, the Joshi's and the Musalgaonkar's. I am glad to finally tell you that it's over. Thank you for all your wishes.

To my committee members, Drs. Jack Smith, Jiajie Zhang, Craig Johnson, Patricia Dolan Mullen, and in addition Dr. Robert W. Vogler, a big thank you for supporting me through extraordinary circumstances. I couldn't have crossed the finishing line without you. Thank you, Dr. Willy Wriggers, for giving me the opportunity to begin my Ph.D. at what was then The School of Health Information Sciences. Dr. Claudio Cavasotto, thank you. I learnt a great deal from you.

To my ex-labmates, Drs. Narender Mann, Arturas Ziemys, Victor Anisimov, and Martin *el Maestro* Indarte, my friends Dr. Indrani, Devadatta, Manuel, Ning (Sunny), Christy and Ram, thanks for all the wonderful memories in and outside the lab.

Last but not the least, Mrs. Jaime N. Hargrave, you will always have a special place in my heart. Jaime, I don't have to search for angels, because I already met you.

Abstract

Development of homology modeling methods will remain an area of active research. These methods aim to develop and model increasingly accurate three-dimensional structures of yet uncrystallized therapeutically relevant proteins e.g. Class A G-Protein Coupled Receptors. Incorporating protein flexibility is one way to achieve this goal. Here, I will discuss the enhancement and validation of the ligand-steered modeling, originally developed by Dr. Claudio Cavasotto, via cross modeling of the newly crystallized GPCR structures. This method uses known ligands and known experimental information to optimize relevant protein binding sites by incorporating protein flexibility. The ligand-steered models were able to model, reasonably reproduce binding sites and the co-crystallized native ligand poses of the β_2 adrenergic and Adenosine 2A receptors using a single template structure. They also performed better than the choice of template, and crude models in a small scale high-throughput docking experiments and compound selectivity studies. Next, the application of this method to develop high-quality homology models of Cannabinoid Receptor 2, an emerging non-psychotic pain management target, is discussed. These models were validated by their ability to rationalize structure activity relationship data of two, inverse agonist and agonist, series of compounds. The method was also applied to improve the virtual screening performance of the β_2 adrenergic crystal structure by optimizing the binding site using β_2 specific compounds. These results show the feasibility of optimizing only the pharmacologically relevant protein binding sites and applicability to structure-based drug design projects.

Table of Contents

| | |
|---|----|
| Chapter 1 | 1 |
| Drug Discovery Process..... | 1 |
| Experimental Drug Discovery Process and Methods..... | 3 |
| Computer-Aided Drug Discovery Process and Methods | 5 |
| Ligand-based <i>in silico</i> Methods | 6 |
| Structure-based <i>in silico</i> Methods | 8 |
| Need for Developing Protein Models | 10 |
| <i>De novo</i> Protein Modeling Methods | 12 |
| Homology based Protein Modeling Method | 12 |
| Protein Flexibility..... | 16 |
| Chapter 2 | 22 |
| Survey of GPCR Modeling Methods | 22 |
| Class A G Protein Coupled Receptors..... | 22 |
| GPCR Specific Homology Modeling Methods | 23 |
| Chapter 3 | 26 |
| Benchmarking the Ligand-steered Modeling Method | 26 |
| Motivation | 26 |
| Ligand-Steered Modeling Method | 28 |
| Results and Discussion | 31 |
| Case 1 Modeling Results | 32 |
| Case 2 Modeling Results..... | 34 |
| Case 3 Modeling Results | 35 |
| Case 4 Modeling Results..... | 37 |

| | |
|--|----|
| Performance of Ligand-Steered Models in the High Throughput Docking Studies. .38 | 38 |
| Compound Library Preparation for HTD and Selectivity Studies..... 39 | 39 |
| High Throughput Docking Protocol..... 40 | 40 |
| Results of HTD Studies.....41 | 41 |
| Performance of Ligand-Steered Models in Compound Selectivity Studies.....42 | 42 |
| Methods43 | 43 |
| Results..... 43 | 43 |
| Conclusions.....46 | 46 |
| Future Work47 | 47 |
| Chapter 449 | 49 |
| Applications of Ligand-Steered Modeling to Drug Discovery Problems.....49 | 49 |
| Cannabinoid Receptor 2 and its Therapeutic Relevance.....50 | 50 |
| Cannabinoid Receptor 2 Inactive State Modeling52 | 52 |
| Method52 | 52 |
| Rationalization of CB2 Inverse Agonist Structure Activity Data56 | 56 |
| Cannabinoid Receptor 2 Active State Modeling59 | 59 |
| Method61 | 61 |
| Rationalization of CB2 Agonist Structure Activity Data62 | 62 |
| Optimization of β_2 Adrenergic Crystal Structure and Docking Studies68 | 68 |
| Method68 | 68 |
| Results and Discussion69 | 69 |
| Chapter 5 71 | 71 |
| Modeling Studies for p65 using Open Source Libraries71 | 71 |
| Method71 | 71 |

| | |
|--|-----|
| Biology..... | 73 |
| Proposed Results of Modeling Studies | 74 |
| Chapter 6 | 77 |
| Research Projects Summary and Future Directions..... | 77 |
| Data, Information and Knowledge. | 81 |
| References | 82. |

List of Illustrations

| | |
|--|----|
| Figure 1: Ligand-based computational methods..... | 6 |
| Figure 2: Percentage difference (2001-2008) in the annual citations of HTS (blue bars) and virtual screening (red bars) papers indexed in google scholar. | 8 |
| Figure 3: Schematic for a virtual screening / high-throughput docking protocol | 10 |
| Figure 4: Outline of the homology modeling process. | 12 |
| Figure 5: General 3-D visualization of β_2 adrenergic (pdb code: 2rh1) receptor | 22 |
| Figure 6: Flowchart of the ligand-steered modeling method | 28 |
| Figure 7: Top ranked ligand-steered model of β_2 using bRho as the template | 32 |
| Figure 8: Top ranked ligand-steered model of β_2 using A2A as the template | 34 |
| Figure 9: Case 3 ligand-steered modeling result. Ligand-steered model of A2A | 36 |
| Figure 10: Case 4 modeling study result. | 38 |
| Figure 11: Case 1 selectivity results | 44 |
| Figure 12: Case 2 selectivity results | 44 |
| Figure 13: Case 3 selectivity results | 45 |
| Figure 14: Flowchart of the CB2 modeling protocol | 52 |
| Figure 15: Isatin acylhydrazone scaffold | 54 |
| Figure 16: Proposed ligand-steered model of inverse agonist bound CB2 complex | 55 |
| Figure 17: Proposed model of CB2 complexed with compound 33 of the agonist benzofuran series. | 62 |
| Figure 18: Benzofuran scaffold for CB2 agonist compounds | 63 |
| Figure 19: Proposed binding mode of 3-FC with p65 DNA-binding region | 74 |
| Figure 20: RMSD vs estimated binding energy of Case 1 studies | 79 |

List of Tables

| | |
|--|----|
| Table 1: Examples of applications of homology modeling in the drug discovery process.... | 15 |
| Table 2: Cases for the ligand-steered modeling and benchmarking studies..... | 27 |
| Table 3: Ligand-steered modeling results. Heavy atom rmsd between ligand-steered modeled and native cocrystallized ligands. | 31 |
| Table 4: Physiochemical properties of compounds and decoys for β_2 and A2A HTD..... | 39 |
| Table 5: HTD enrichment results. | 41 |
| Table 6: CB2 inverse agonist assay data..... | 57 |
| Table 7: CB2 agonist compound assay data..... | 63 |

Chapter 1: Drug Discovery Process.

The drug discovery process may be defined as the discovery and design of chemical entities to favorably alter the activity of aberrant biological molecules e.g. proteins. This process can be broadly classified into six steps a: target identification and validation; b: hit identification; c: compound optimization d: animal model testing; e: multi-step human clinical trials and f: Federal Drug Administration (FDA) approval and subsequent marketing(1). The biology oriented target identification stage involves identifying and associating biological targets e.g. a gene or a protein, to a disease under consideration. Subsequently using an array of experimental cellular and animal models the role of the target in controlling the disease is conclusively defined. Currently available therapeutics target a mere ~250 protein targets (207 of these are derived from the human genome)(2). This number pales in comparison to the ~25000 human genes sequenced by the National Human Genome Research Institute. Moreover greater than 50% of these genes have unknown functions(3). It is clear that biology will continue to present validated drug targets and the drug discovery process will continue to need novel methods and technologies for discovering new therapeutics. Given a validated drug target millions of compounds are experimentally or computationally screened and tested using methods such as high-throughput screening (HTS) and virtual screening (VS). Promising compounds (~ 5000 - 10000) from the HTS stage are subjected to an iterative cycle of medicinal chemistry and pharmacology methods to develop potent lead compounds for animal model testing. Computational structure-based methods that use crystallized or computationally modeled protein:compound complexes to elucidate protein:compound interactions are routinely used in the optimization process(1). Roughly 3% compounds from the optimization stage will pass the rigorous safety profiles and efficacy of animal model testing(4). Approximately 5 compounds

are then introduced in the three-stage human clinical trials. Only after determining a statistically significant safety, efficacy and therapeutic profile a drug candidate may be approved by the FDA for subsequent marketing. Despite the poor success rates and varied multi-dimensional challenges, ~1200 drugs targeting 266 human genome therapeutic targets have been approved by the FDA to date(2).

Several studies have documented the time and costs required to market a drug as ~12 -15 years and upwards of one billion US Dollars(5). Approximately 50% of the time and scientific research coupled with 40% of the costs are associated with the preclinical discovery and development stages(6). Despite the development of newer models that aim to make clinical trials effective and cost-effective currently they still follow rigid protocols(7). Thus, the development of novel scientific and technological methods for all areas in the preclinical discovery and design phase is an important area of research. In this thesis I will present an enhanced protein modeling method that can accurately model protein structures and its applications to structure-based drug discovery problems: specifically elucidation and rationalization of the biological activities of compounds and improvement of virtual screening success rates. This chapter will first provide a literature review on the developments in the drug discovery process and a primer of computational drug discovery, specifically structure-based discovery. Subsequently one of the major problems of structure-based drug discovery i.e. modeling three-dimensional protein structures of proteins (homology modeling) will be discussed. Next, I will discuss Class-A G-Protein Coupled Receptors, an important class of therapeutic proteins with very few crystallized structures, and GPCR modeling methods. The next chapters will contain published work performed using the ligand-steered homology modeling method.

Experimental Drug Discovery Process and Methods:

Serendipity played an important role in drug discovery during the early 20th century(8, 9).

Insulin for diabetes, and penicillin an important antibiotic are some examples of serendipitously discovered drugs. This approach is impractical as it is unable to discover target-specific drugs and target-specific biological activity. With advances in chemistry (e.g. combinatorial chemistry)(10) and industrialization (e.g. robotics, hardware control and data processing software, sensitive detectors and automated liquid handling devices), pharmaceutical companies developed the high-throughput screening (HTS) platforms to experimentally and consistently test several thousands of compounds against specific target assays(11-14). HTS has four main stages a: assay development, b: compound library development and selection, c: screening technology and d: readouts and result analysis.

High quality biological assays (e.g. cell-free or cell-based) are the primary requirement for generating robust and reproducible output off the screening exercise(15, 16). Historically HTS examined the effect of a single compound acting on a single target and at a single concentration to determine biological activity. Technical advancements in the last two decades, particularly miniaturization methods, have led to over four-fold improvement in number of assays / HTS plate (e.g. from 96 to 3456 wells / plate). Corresponding improvements in robotics and hardware increased screening rates, decreased usage of expensive reagents and made possible to obtain over 100,000 data points / day(1). Advanced screening methods e.g. fluorescence resonance energy transfer (FRET)(17), homogenous time-resolved fluorescence (HTRF)(18), biology-based high-content screening methods that permit multi-parameter readouts provide researchers with high-quality and reproducible data for biological targets under consideration. Improved data processing software have improved the ability to detect *true actives*, compare

diverse assay protocols and automate most of data analysis work. Unlike computational methods in HTS the protein / ligand flexibility is inherently considered and accuracy of experimental binding data is often better than computational scores.

Corresponding developments in compound library creation and management e.g. combinatorial chemistry, a technology that systematically combines multiple chemical blocks to develop large and possibly diverse sets of compounds, now provide several millions of compounds for HTS(19). However in order to further improve screening success rates and avoid costs associated with screening random compounds, a combination of computational and empirical models have been developed to identify smaller compound libraries with wanted characteristics such as drug-like, lead like- target-focused and eliminate compounds with unwanted physiochemical / toxic properties(20). The underlying principle of focused libraries is that target specific compounds are restricted to finite chemical and pharmacological profiles. In addition focused libraries potentially reduce costs associated with synthesis, inventory management and screening without compromising on the chemical diversity of compounds presented for screening.

HTS remains an important source of lead compounds and is the progenitor of over 100 compounds currently under clinical trials(21, 22). Despite these improvements and historical results, the ratio of results to financial investments for HTS remains poor. Developing target-based assays for all therapeutically relevant genes (e.g. ion channels) in high-throughput format is challenging and so is developing assays for over 25000 genes from the human genome project(23, 24). The nearly infinite chemical space (10^{60} - 10^{100}) is impossible to cover and prohibitive costs associated with compound management result in a repetitive screening of finite and limited chemotype diversity compound libraries(25, 26). Considering the paucity of

new drugs entering the market, loss of revenues associated with patent expiration of blockbuster drugs (e.g. Lipitor® for high cholesterol, Effexor® for depression, Enbrel® for arthritis, and Topamax® for seizures), and failures associated with late-stage drug discovery (e.g. Avastan® for colon cancer, Vioxx® for osteoarthritis and Rolofylline® for heart conditions)(1), researchers are increasingly using complementary computational methods with the aim of improving the success rates, productivity and most importantly limiting costs associated with the complex drug discovery process.

Computer-Aided Drug Discovery Process and Methods:

Computer-aided (*in silico*) drug discovery consists of the development of computational methods based on established biophysical and chemistry knowledge and their application in the drug discovery process. Rapid advancement in cheaper, faster computational power and accessibility coupled with continual improvement in biophysical modeling algorithms are the primary reasons that *in silico* methods are now firmly established as complementary methods to experimental drug discovery(1). *In silico* methods are routinely applied to develop high-quality three dimensional protein models(27), identify hits using docking or virtual screening methods, optimize lead compounds, predict biological activities using compound-based modeling methods (e.g. Quantitative Structure Activity Relationship, QSAR methods)(28), predict Absorption Distribution Metabolizing Excretion and Toxicity (ADMET) profiles of compounds(29), manage and process in-house compound data, and design focused chemical libraries for experimental screening design(20). *In silico* methods primarily aim to rationalize, increase the efficiency, speed and cost-effectiveness of the drug discovery process. Some successful applications of *in silico* methods in drug discovery include identification of hits against variety of target families (e.g. G-Protein coupled receptors, protein kinase, nuclear

hormone receptors, and proteases)(30) elucidation structure-activity relationships of compounds(31, 32), and lead optimization(33-35). *In silico* methods are classified as ligand or structure-based methods.

Ligand-based *in silico* Methods:

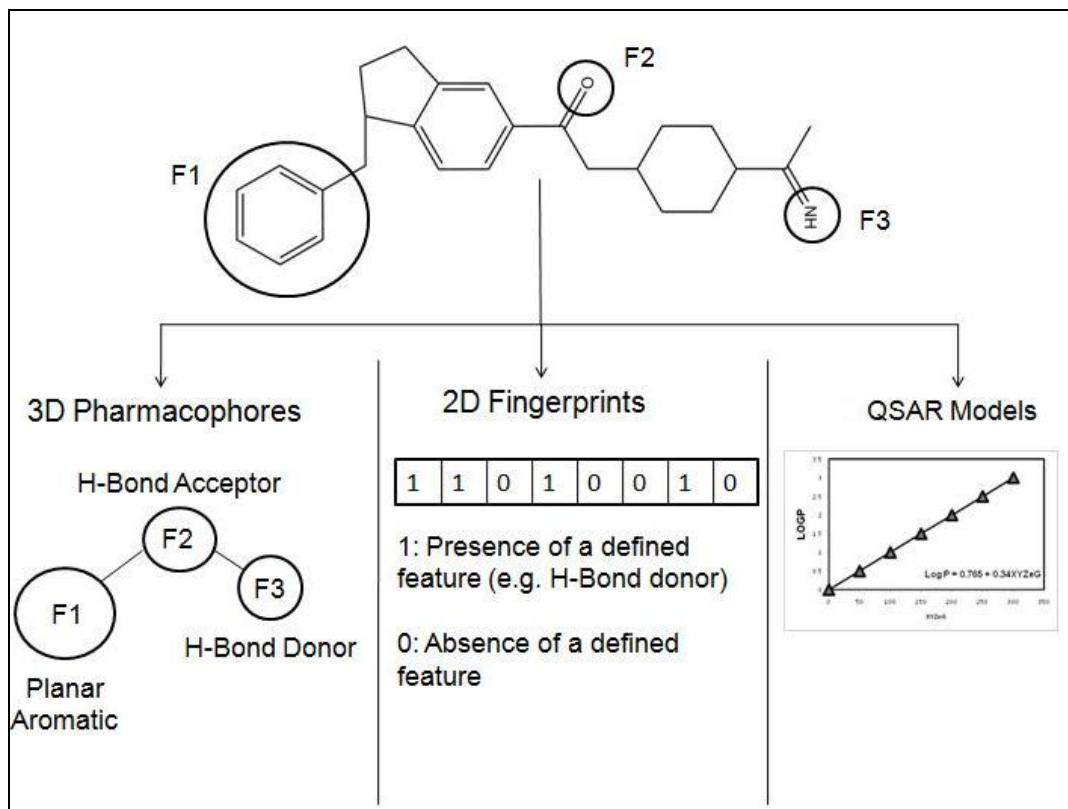


Figure 1: Ligand-based computational methods(1).

The underlying principle for ligand-based *in silico* methods is that compounds with similar chemistry will have similar biological profiles(36, 37). These methods are chemistry aware, relatively easier to develop and aim to identify compounds based on a small set or even one single active compound. They are further classified into two or three-dimensional methods.

Molecules are represented as graphs where nodes correspond to atoms and edges to bonds or in a linear notation such as SMILES. To identify similar compounds by in two dimensional space

graph theory methods e.g. *subgraph isomerism* may be used(38). Another efficient method is the fingerprint method where molecules are represented as a fixed-length vector of chemical substructure attributes. The similarity coefficient is calculated using metrics such as Jaccard coefficient, Euclidean or Manhattan distance etc.

$$\text{Jaccard Coefficient } S_{AB} = c / (a + b - c)$$

$$\text{Euclidean Distance } D_{AB} = [a + b - 2c]^{0.5}$$

Where a = number of bits on in molecule A, b = number of bits on in molecule B and c = number of bits on in both molecules. Fingerprint based methods are computationally efficient as compared to graph-based methods.

Quantitative structure-activity relationship models aim to correlate compound structural features to its biological activity by a mathematical model. This relationship follows the general equation where v is the activity and p are molecular descriptors and f is the function.

$$v = f(p) + k$$

A specific example is the original Hansch equation(39) that relates the biological activity of a compound to its electronic and hydrophobic characteristics.

$$\log(1/C) = k_1 \log P - k_2 (\log P)^2 + k_3 \sigma + k_4$$

Where C is the concentration of the compound required to produce a standard response in a given time, logP is the logarithm of the partition coefficient of the compound between 1-octanol and water, σ is the Hammett substituent parameter and k_1 - k_4 are constants.

Three-dimensional methods such as pharmacophore-based approaches use a three-dimensional representation of a compound and relevant pharmacophoric features (required steric or electronic features for optimal protein:compound interaction and subsequent biological activity) to develop a computational model that specifies spatial relationships between the pharmacophoric features. Complementary features from the protein may also be incorporated in the model building process. However in case of pharmacophore models the conformational space of compounds must be determined and a biologically relevant conformation must be selected. All three approaches are illustrated in the figure 1. Ligand-based methods are easier to develop and implement, but fall short in the identification of novel compounds for drug discovery or rationalizing biomolecular interactions. The second type of computer-aided drug design i.e. structure-based drug design offers several methods to complement ligand-based computational methods.

Structure-Based *In Silico* Methods:

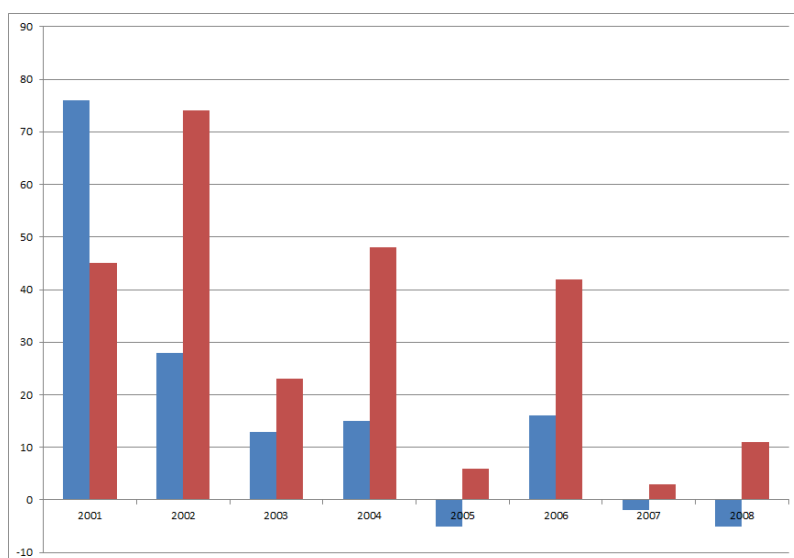


Figure 2: Percentage difference (2001 – 2008) in the annual citations of HTS (blue bars) and virtual screening (red bars) papers indexed in google scholar(1).

Structure-based drug design is a computational process in drug discovery that relies on the availability of three dimensional protein structures or models. Molecular docking or high throughput docking (HTD), the computational equivalent of high throughput screening, is one of the most important structure-based drug design method. Figure 2 represents the yearly change in number of citations of HTS (blue bars) and *in silico* or virtual screening (red bars) from 2001-2008 and shows the growth of HTD publications against HTS publications . Broadly this process consists of i) identifying potential binding sites where the small molecules may interact with the target; ii) virtual chemical library selection; iii) the docking process where compounds are positioned in the binding site and iv) scoring compounds that estimates the likelihood of binding to the target (scoring and ranking)(1). Figure 3 is a schematic for the high-throughput docking process. Molecular docking aims to prioritize a reduced set of compounds that have a higher probability of being active in subsequent experimental analysis. Molecular docking has been successfully applied in several drug discovery stages e.g. development of target-specific compound libraries, hit identification, and lead optimization. Docking applications(40) have been benefited by the continual development of docking algorithms(41), cheaper and faster computational resources and an increasing availability of high quality protein structures from Structural Genomics projects(42).

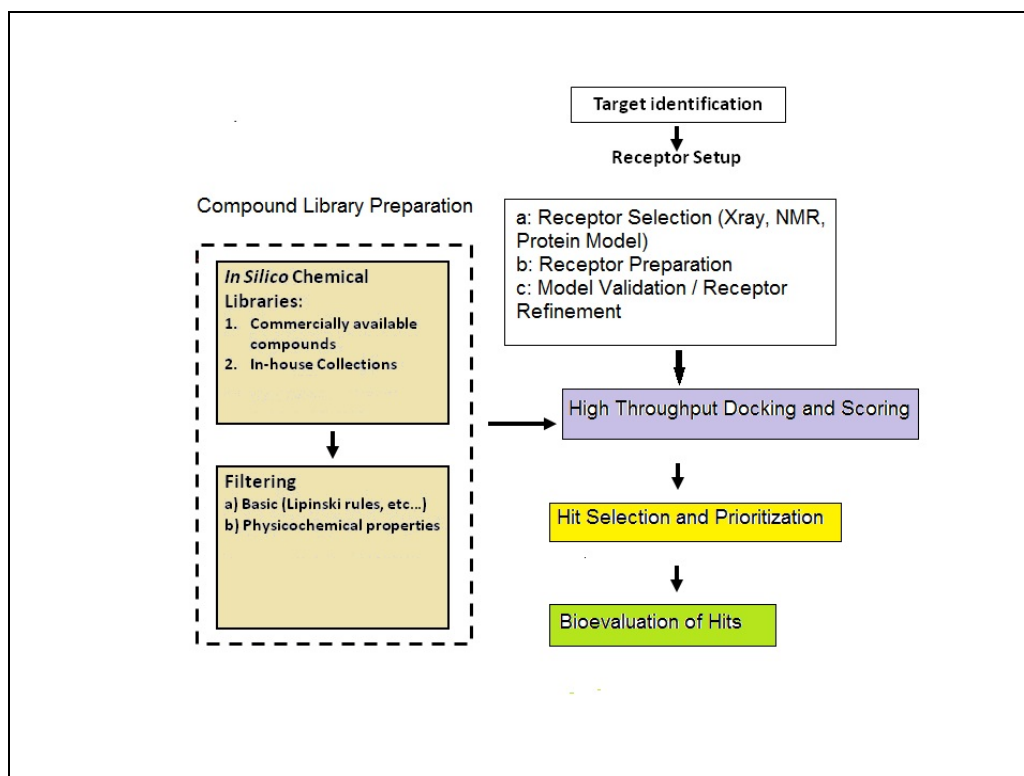


Figure 3: Schematic for a virtual screening / high-throughput docking protocol.

However, molecular docking and in-turn structure-based drug design has two major problems.

A: Three-dimensional crystal structures of many therapeutic proteins are currently unavailable(27).

B: Receptor flexibility is poorly handled in docking algorithms(43).

Need for Developing Protein Models.

Three-dimensional structural information of biological macromolecules is available in the Protein Data Bank (<http://www.pdb.org>). As of January 2012, the PDB contained ~79000 structures that can be grouped into ~3500 families consisting of over 1100 unique folds. In the past several years Structural Genomics projects have spurred the technical developments

associated with X-ray crystallography and NMR techniques(44-46) and have been fundamental in the growth of novel structures in the PDB (defined as <30% sequence identity). SG projects aim to solve at least one representative structure from protein families with no experimental crystal structures. Despite these developments the novelty of proteins deposited in the PDB (defined as <25% sequence identity between two structures) has remained constant since 1992(47). Technical problems associated with structure elucidation e.g. protein purification; crystallization techniques limit the availability of protein structures. However the gap between the number of annotated protein sequences (~408000) and the available crystal structures i.e. ~79000 including redundant proteins, is likely to remain in the near future.

In the absence of experimental protein crystal structures of GPCRs and other therapeutic proteins, computational methods are used to develop three-dimensional protein models. Computationally modeled proteins may be considered reasonable substitutes until a corresponding protein structure is available. Protein models may provide structural and biochemical functional insights to biologists / experimentalists for various biomedical research problems. Several publicly available repositories e.g. The SWISS-MODEL (<http://swissmodel.expasy.org/SWISS-MODEL>), Protein Model Portal (<http://proteinmodelportal.org>) and Modbase (<http://modbase.compbio.ucsf.edu>) contain protein models generated using automated methods. However models developed by automated methods have limited accuracy and applicability for drug discovery projects. Thus the development of newer and more accurate protein modeling methods is an important problem and an area of active research in the drug discovery domain.

***De novo* Protein Modeling Methods.**

Protein modeling can be broadly divided into two categories, template free modeling and template-based modeling. In the former *de novo* methods using knowledge and/or physics based potentials are used to build protein models. Some examples include Rosetta, that uses knowledge-based potentials to build crude models and all-atom physics based methods to refine the crude models. Threading assembling and refinement (TASSER) by Wu et al. uses knowledge-based approaches for the entire modeling process(48). *De novo* methods are independent of any pre-existing templates thus in principle can model novel folds of protein structures. Presently these methods have limited applicability in the drug discovery process as they can either reasonably model smaller proteins (~150 AA) or the models have poor resolution (> 2Å) to be useful(49).

Homology based Protein Modeling Method:

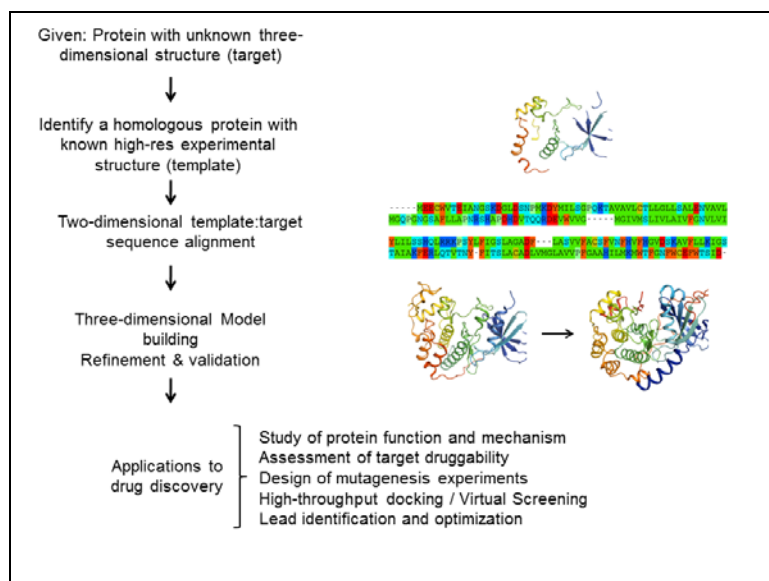


Figure 4: Outline of the homology modeling process(27).

An outline of the homology modeling method is depicted in Figure 4. The fundamental principle of homology modeling is that similar proteins will share similar three-dimensional structures. Homology modeling consists of four steps a: identification of a homologous (same family) protein preferably with a high-resolution protein crystal structure. This protein structure serves as the template, b: Sequence alignment stage: sequence alignment is primarily a bioinformatics based method in which two or more protein sequences are rearranged such that similar motifs (regions) / protein residues are aligned in a columnar representation. Sequence alignment is particularly important as it serves as the input and basis of homology model building process. The degree of homology and quality of alignment often play an important role in the quality of the model structure. Generally, if the sequence identity between template:target exceeds 50% highest quality models are generated and if it is less than 25% such models are speculative at best(50). Next, the three-dimensional Cartesian coordinates of the template are transferred to the corresponding residues of the target sequence obtained from the sequence alignment stage. Then subject to restraints satisfaction a crude model is developed. Spatial restraints expressed as probability density functions (pdf) as used in the package MODELLER (used extensively to develop preliminary homology models in this thesis) were developed from extensive statistical analysis of protein families that quantify relationships between $C\alpha$ - $C\alpha$ distances between backbone residues and corresponding main chain dihedral angles(51). The probabilities are calculated from the range of distances and angles of a particular residue type, conformation and sequence similarities between proteins present in the protein families training set. The function itself is represented as $p(X_1 | a,b, \dots,c)$ which can be used to predict X_1 the side-chain dihedral angle from the variables residue_type, phi, and psi angles. The complete details are available from the original paper by Sali et al(51). Thus, better

the crystal structure quality and resolution, larger numbers of spatial features are specified for restraint determination. These individual spatial features are assumed to be independent and combined into a molecular pdf that is the product of all the feature pdf's given by the equation:

$$P = \prod_i p^F(f_i).$$

Crude models are built to maximize the molecular pdf given the sequence alignment between the target and template proteins. The crude model is then iteratively optimized to a low energy state to refine the positions of all heavy (non-hydrogen) atoms. Energy minimization using conjugate gradient method is used to refine the crude models(51). Finally, the models are assessed using the discrete optimized protein energy method(52). DOPE is a statistical potential empirically derived using observed residue-residue contact frequencies among proteins with known structures. They assign a probability or energy score to each possible pairwise interaction between protein amino acids and combine them into a single score for a given model. Typically, several models are built and ranked before selection for advanced refinement procedures or direct applications (provided very high-sequence identities between the template and the target proteins).

Homology modeling has been successfully applied in various stages of the drug discovery process. Examples include, the application to elucidate biological function of BC0371 an enolase family member(53), M antigen homology model for function prediction in histoplamosis(54), RDH12 gene for Leber congenital amaurosis(55), and to elucidate Nod-like receptors role in immune response(56). Homology models have been applied to discover novel leads for MCHR-1 receptors for anti-obesity(57), alpha-glucosidase for diabetes(58), Cdc25 phosphatase for anti-cancer therapeutics(59), CK1-delta for Alzheimer's(60) amongst other. In

some other examples, homology models have been developed to investigate the role of certain residues in the biological function of the protein via mutagenesis experiments(61, 62). Models of Adenosine 2A(63), Alpha-1-A adrenoreceptor(64), Adenosine 3 receptor(65) were developed to elucidate and investigate compound:protein interactions and the models for Src kinase(66), PKC theta(67) and GPR109A(68) were developed to optimize lead compounds obtained from experimental screening exercises.

| Protein Target | Application | Ref. |
|--|--|-------------------|
| BC0371 (member of the enolase superfamily) | Study of protein function | 53 |
| M Antigen | Study of protein function | 54 |
| RDH12 | Assess the biological role of mutations in the binding site and their effect on in the function ofRDH12. | 55 |
| Nod-Like Receptors | Understand the protein mechanism implicated in the immune response. | 56 |
| MCH-R1 | Structure-based lead discovery for anti-obesity drugs. | 25 |
| Cannabinoid Receptor 2, Human Adenosine A2A Receptor, Alpha-1-A-adrenoreceptor, Adenosine 3 receptor | Binding mode prediction and elucidation of ligand:protein interactions | 101,102, 63,64,65 |
| Peptide CGP38560 in complex with renin, Src Kinase, PKC theta, GPR109A | Lead or compound optimization | 66,67,68 |
| Cdc25 phosphatases | Structure-based virtual screening for anti-cancer lead discovery | 59 |
| CK1 δ | Structure-based virtual screening to identify inhibitors for Alzheimer's disease | 60 |

Table 1: Examples of applications of homology modeling in the drug discovery process.

Homology models by definition are an abstraction and hence are likely to contain errors. Generally models built with over 50% sequence identity are reasonably accurate and applicable for drug discovery applications. However high overall sequence identity may mask dissimilarities in crucial flexible loop regions (e.g. enzymes) and adversely affect model quality and applicability. Likewise errors from template-sequence alignments affect model quality. Overall each step of the modeling process, template identification, alignments and final refinement methods contribute to the errors in the final model. Depending on the degree of sequence identity and the quality of alignment their accuracy as compared to the crystal structure can be up to $\sim 2\text{\AA}$ $C\alpha$ atom RMSD(69, 70). Errors of this magnitude, particularly in the binding site, can result in wrong side chain orientations thus misrepresenting receptor:ligand complementarity. This can potentially negate the applicability of the model in a drug discovery project. Thus it is important to develop and apply methods that incorporate receptor flexibility in the modeling process to generate accurate side-chain orientations to improve quality of protein models.

Protein Flexibility.

Here, I will introduce the second major challenge for structure-based drug discovery: the inadequate incorporation of protein flexibility.

Proteins are flexible biomolecules. They retain a well-defined overall fold but are present in multiple low-energy conformations on a rugged energy landscape. These low energy states represent local energy minima and are separated by barriers surmountable by thermal activation. Their flexibility is likely to be related to the binding of a ligand. Holo / complexed proteins may exhibit structural disorder e.g. change in protein side-chain, conformational

changes in the ligand associated with binding, or a combination of protein backbone, side chain and ligand conformational change upon binding. Two phenomena are associated with these changes. A: Kinetic Regulation: Here, the energy barrier between two conformational states is transcended during the process of binding and B: Thermodynamic regulation: Free energy of the other protein conformation is lowered prior to binding and the ligand chooses the new low energy state(71). Unlike experimental HTS where protein flexibility is inherently incorporated, modeling protein flexibility remains a challenging and important problem in SBDD(1). The shortcomings of a rigid receptor approach in virtual screening / docking studies are well documented(72). However unlike ligands which have a tractable 6-12 degrees of freedom to model flexibility, modeling protein flexibility is a high-dimensional problem. A simple representation of only the protein binding site with < 20 side chain residues results in multiple dozens of torsion angles. Including backbone atoms of residues and the potential multiple changes associated with a flexible backbone can result in order of magnitude higher degrees of freedom. Compounding the problem is the energy-based receptor conformation selection as these functions are approximate and can result in incorrect conformation selection(73). Thus it is essential to have computationally tractable and time effective modeling methods to reasonably estimate protein or at least binding site flexibility of proteins for applications in structure-based drug discovery.

Next, I will explain in brief some methods used to model protein flexibility with an emphasis on docking / virtual screening applications as these are the focus of my thesis. The assumption is that multiple receptor or ensemble based docking approaches offers a practical way to incorporate receptor flexibility in SBDD.

Soft-docking Approach: A larger tolerance for protein side chain and ligand atom clashes is the basis for soft scoring functions in docking(74, 75). This is an efficient method that uses a rigid receptor with variations only in the scoring function. However this approach does not account for backbone movements, structural rearrangement or larger side chain movement.

Rotamer exploration: Rotamers are side-chain conformations generated by systematic conformational searches. Rotamer libraries are generated by altering each rotatable bond by a range of bond type specific dihedral angles and subsequent scoring to produce diverse conformations(76-78). These pre-determined rotamer libraries can mimic receptor side chain flexibility; however they are inadequate to simulate ligand-dependent changes in a receptor.

Molecular Dynamics: MD methods are considered to be the most rigorous methods for generating receptor conformations. Using Newton's second law of motion, a time dependent trajectory of differential atom positions of the receptor is generated. Depending on time scales, side chain variations or even small structural changes can be observed. However this method is computationally intensive and often unrealistically short time scales miss larger structural changes associated with proteins. Nevertheless this physics-based method has proved successful in applications such as exploring receptor flexibility before docking; simulating induced fit effects and refinement of docked structures for binding energy calculations(79-82).

Normal mode analysis: NMA is a computationally tractable method to identify flexible regions in the protein. The frequency of normal modes together with the displacement of individual atoms is determined by the Eigen vectors of the Hessian matrix of the potential energy. Use of low frequency normal modes or normal modes associated with protein pockets has led to

reasonably accurate representations of protein flexibility and promising results in structure-based drug discovery(83-85).

Monte Carlo Sampling Methods: Monte Carlo simulations generate multiple conformations of a molecular system by randomly sampling positional changes of the constituents of the system. E.g. consider an atom represented in Cartesian coordinates x, y and z. A random number generator produces a random number ξ in the range 0-1. δ is the displacement and δr_{\max} is the maximum displacement. Then from Leach(86).

$$X_{\text{new}} = x_{\text{old}} + (2\xi - 1)\delta r_{\max}$$

$$Y_{\text{new}} = y_{\text{old}} + (2\xi - 1)\delta r_{\max}$$

$$Z_{\text{new}} = z_{\text{old}} + (2\xi - 1)\delta r_{\max}$$

The energy of the new conformation is calculated. If the new energy is lower than the prior energy, this state is retained for the next iteration. If the new energy is higher than the prior then the Boltzmann factor, $\exp(-\Delta v(\mathbf{r}^N) / k_B T)$, is compared to a random number between 0-1. If the Boltzmann factor is greater than the random number, the conformation is accepted and retained for next iteration. This can be represented as

$$\text{rand}(0,1) \leq \exp(-\Delta v(\mathbf{r}^N) / k_B T).$$

In case of molecules their orientations and their positions need to be varied in a Monte Carlo simulation. Translations are handled by varying the center of mass of a molecule. For rotations, the molecule is rotated by a randomly chosen angle δw limited by the maximum variation δw_{\max} .

However Monte Carlo simulations are difficult to perform successfully for flexible molecules. Small displacements in one region may result in large increase in another region accompanied by an unfavorable energy increase. This is likely in case of proteins, where changes introduced in the backbone regions may cause large movements down the chain / secondary structure. In these cases the hard degrees of freedom e.g. bond lengths and bond angles, particularly backbone atoms are held fixed.

Monte Carlo simulations can be performed at exact temperatures and pressure unlike MD methods. The intrinsic nature to make non-physical moves may permit Monte Carlo methods to explore a larger phase space for biomolecules. This is unlike MD based methods that may find the high energy barriers present between two low energy conformational states of a given biomolecule difficult to transcend. Monte Carlo methods have been applied for a variety of cases in drug discovery applications and offer specific advantages for modeling G-Protein Coupled Receptors.

Chapter 2: Survey of GPCR Modeling Methods.

Class A-G Protein Coupled Receptors.

Class A G-Protein Coupled receptors (GPCRs) are an important class of membrane proteins for drug discovery. There are ~1000 GPCRs that mediate nearly 80% of signal transduction processes across cellular boundaries(87, 88). GPCRs interact with diverse chemotype ligands to modulate downstream effectors (e.g. enzymes)(89) and subsequent physiological functions (e.g. cardiovascular, nervous, endocrine, and immune)(90). It is estimated that roughly ~50% of currently available drugs target this class of proteins(91). However, this number is misleading as drug discovery programs target only ~3% known GPCR family members(92). Historically, these proteins were not amenable for crystallography methods due to isolation problems, their membrane-bound nature, and other technological problems(93). Recent advances in membrane-bound protein crystallization methods have resulted in the availability of several Class A GPCR crystal structures (e.g. β_2 / β_1 adrenergic(94, 95), Adenosine 2A(96), opsin(97), CXCR4 Chemokine(98), Dopamine 3(99), and Histamine 1 receptors(100)). However, crystal structures of over 90% of Class A-GPCRs remain unsolved. The availability of these recently solved crystal structures served as the motivation for the first part of this thesis i.e. enhancement, validation and benchmarking of the ligand-steered modeling method in Class A-GPCRs. The second part of this thesis is devoted to the development of high-quality homology models of Cannabinoid Receptor 2 (CB2), an important Class A-GPCR for non-psychotic pain management therapeutic use(101, 102). These models were developed to mimic two biological states of CB2 receptors (inactive and active) and applied to rationalize the activities of two novel series of CB2-specific compounds.

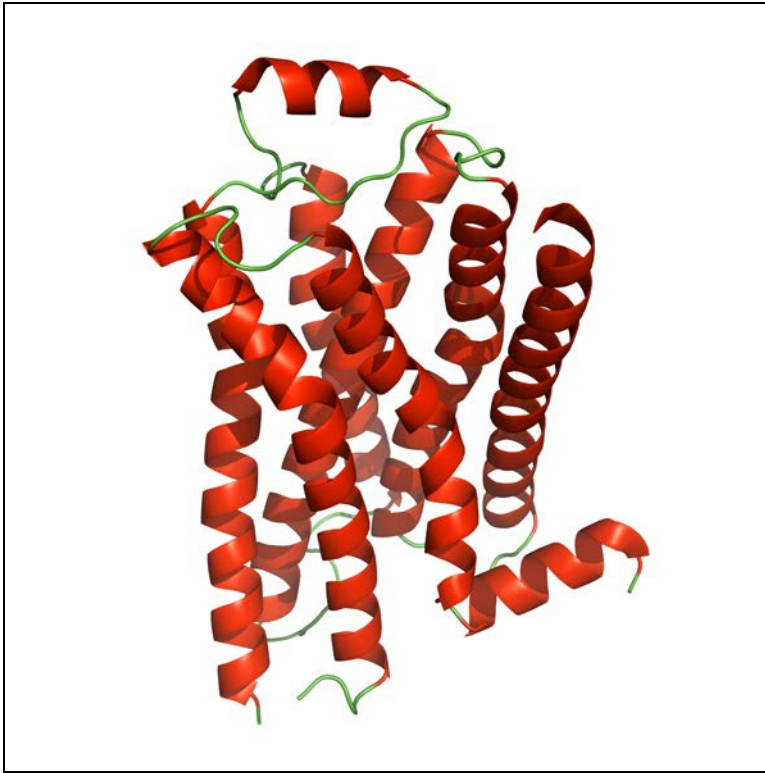


Figure 5: General 3-D visualization of β_2 adrenergic (pdb code: 2 rh1) receptor. Figure was prepared using Pymol (www.pymol.org).

Class A-GPCRs are a special case of homology modeling as they possess low sequence identities (~20%). However this drawback is compensated by the conserved 7 transmembrane (TM) structural characteristic share by all Class A-GPCRs (Figure 5). Each of these 7 transmembrane's share specific structural signatures that can be exploited for model building as demonstrated by Ballesteros and Weinstein. These include, the conserved Asn in TM1, LeuAlaXAlaAspLeu motif in TM2, AspArgTyr motif in TM3, Pro in TMs 4 and 5, CysTrpXPro in TM6 and AsnProXXTyr motif in TM7(103, 104). The most conserved residue is denoted as number X.50 (X denotes the TM number). Other residues in each membrane are serially numbered in increasing or decreasing order depending on the direction of the

membrane. In addition, the conserved disulfide bond present between two conserved Cysteine residues in the extracellular loop 2 is maintained for the model building process. GPCR sequences are aligned devoid of any gaps in the transmembrane regions. Sequence alignment of all GPCR models mentioned in this work is performed using this approach.

GPCR Specific Homology Modeling Methods.

GPCR modeling is an actively researched area. Highlighted below are some examples that have included receptor flexibility in the GPCR modeling process.

Klebe and colleagues developed the MOBILE (modeling binding sites using ligand information explicitly) approach using the neurokinin-1 (NK1) class A GPCR, that incorporated ligand-protein restraints obtained from docking methods to produce optimal geometries of the NK1 binding sites(105, 106). Sequences of NK1 and bRho were aligned and 100 crude homology models were generated using MODELLER. For each of the 100 models, backbone coordinates are kept similar to the template structure, but side chains conformational space is exhaustively sampled to explore the variability of residues that do not match in the alignment phase. To optimize the binding site specifically for the known antagonist CP-96345, its bioactive conformation was used in a rigid-docking based protocol on all the 100 initial models. Post visual inspection, protein:ligand data from 4 models that reproduced known experimental data was used as restraints in the subsequent refinement stage. Using the newly derived restraints, 400 homology models were regenerated, scored (DrugScore) based on the observed interactions with CP-96435 and the best individual solutions were assembled to generate a final model. This approach was successfully applied to discover novel inhibitors for NK1.

In another approach termed ligand-based modeling, Moro et al. developed a protocol to account for protein reorganization in Class A GPCRs(65). Here known ligands are docked to the crude model of the adenosine receptor to generate an ensemble of ligand poses. An approximate receptor definition was used i.e. the van der Waals radii of each atom was scaled down by 25% and an increased Coulomb-van der Waals cutoff was used to mimic receptor flexibility / alternate side chain positions with reduced computational overhead. The spatial information obtained from the ligand poses are subsequently used to generate alternate conformations of binding site residues using an approach based on rotamer exploration. Top scoring complexes are identified and are subjected to local energy minimization of the ligand and side chains. This method modeled significant steric changes within the binding pocket and explained the activities of diverse chemotype inhibitors of human Adenosine 3 receptor.

In contrast to the semi-automatic approaches explained above researchers have used validated experimental protein:ligand information to manually optimize side-chain orientations.

Radestock et al. manually optimized the side chains in the metabotropic glutamate receptor subtype 5 (mGluR5) to mimic experimental binding information(107). Costanzi S. altered the side chain conformation of Phe290 from trans to gauche+ state in the Beta 2 adrenergic receptor model and obtained near-native poses for the inverse agonist Carazolol(108).

However, manual optimization approaches are tedious and generally impractical.

Kimura et al. modeled and optimized the binding region of chemokine receptor-2 by using pressure-based steered molecular dynamics(109). Multiple small-radii Lennard-Jones particles were placed on the grid and tethered to four nearest atoms of the receptor via weak harmonic bonds. The backbone protein dihedral angles on the 7-transmembrane region are restrained to maintain the structural integrity of the model. The system is subjected to molecular dynamics

simulations wherein the radii of the Lennard-Jones particles are gradually increased to reproduce increased pressure. Models were validated by their ability to redock three known CCR2 antagonists and reproduce known binding information. This method does not use any ligand information, is potentially widely applicable but may result in higher false positives, as the binding regions are potentially unspecific.

The ligand-steered homology modeling method developed by Cavasotto et al. used known ligands of the melanin concentrating hormone receptor 1 (MCHR1) to reshape and optimize the MCHR1 binding pocket by incorporating full receptor and ligand flexibility stochastic docking protocol in the internal coordinate space(57). Unlike other methods, the models of MCHR1 were validated by their ability to discriminate native ligands in a small-scale high-throughput docking protocol. MCHR1 models developed using the LSHM methods were successfully applied in the discovery to novel chemistry anti-obesity inhibitors. In addition and further corroborating the premise of structure-based screening, the performance of the high-throughput docking was 10 fold better than the traditional HTS on the same target. This promising method was developed before the explosion the GPCR structure availability. Thus, the developer assessed the model quality via HTD simulations and the ability of the model to discover inhibitors in an actual drug discovery project. With the availability of high-resolution GPCR crystal structures, the first part of this work aims to enhance, benchmark and validate the modeling protocol using three GPCR structures (bovine rhodopsin, human β_2 adrenergic and human adenosine 2A) and established structure-based drug discovery protocols and metrics.

Chapter 3: Benchmarking the Ligand-Steered Modeling Method.

Motivation. The original ligand-steered modeling method was developed at a time when only one crystal structure of GPCR was available for modeling. Thus, it was impossible to investigate the quality of the generated models based on the established structure-based drug design metrics e.g. the ability to reproduce crystallographic pose of co-crystallized ligands. With the availability of multiple high-quality crystal structures (as of 2009) of Class A G-Protein Coupled Receptors, it is now possible to cross-model these structures using other available structures. The goal of this project was to develop high-quality homology models of GPCRs using one crystal structure and by incorporating biomolecular flexibility in the modeling process. In addition, the applicability of these developed models in simulated real-case scenario HTD process was assessed. To summarize, this part aims to enhance the ligand-steered modeling method and answer the following questions related to the modeling process:

A: Is the ligand-steered method capable to generate near-native models of known GPCR crystal structures by using one template and incorporating protein flexibility?

Near-native models have low root mean square deviation between the modeled binding site and the modeled cocrystallized ligand when compared to the original crystal structure.

B: Can the ligand-steered models deliver similar or better performances in HTD studies as compared to the crystal structure and unrefined / crude homology model?

Comparable performances against crystal structures will support the application of optimized ligand-steered models in actual drug discovery process in the absence of crystal structures.

Better performances than crude homology models will underline the importance of refining and incorporating receptor flexibility for Class A G-Protein Coupled Receptors.

C: Is there a correlation between the best models in terms of cocrystallized ligand rmsd and top models from the HTD experiment?

The hypothesis is that models with low ligand and binding site rmsd will have a superior performance in HTD studies.

D: Can optimized ligand-steered model discriminate receptor specific ligands from a set of decoy compounds?

Compound selectivity is an important criterion in the drug discovery process. Highly selective homology models will be able to reduce the percentage of false positive compounds in virtual screening studies.

Here, two class A GPCRs i.e. human β_2 adrenergic (β_2) and human adenosine 2A (A2A) receptor were cross-modeled in four case scenarios. The cases studied were bRho \rightarrow β_2 , A2A \rightarrow β_2 , bRho \rightarrow A2A and β_2 \rightarrow A2A are represented in Table 2.

| | Target | Template (pdb code) |
|--------|-----------|---------------------|
| Case 1 | β_2 | bRho (1gzm) |
| Case 2 | β_2 | A2A (3eml) |
| Case 3 | A2A | bRho (1gzm) |
| Case 4 | A2A | β_2 (2rh1) |

Table 2: 4 Cases for the ligand-steered modeling and benchmarking studies.

Model qualities were assessed using native ligand and binding site rmsd. For each case, the top ranking models were used in a HTD experiment. The performance of these models was measured by their ability to discern receptor specific ligands at 2% (EF2%) of the screened

database and compared against the performance of the corresponding crystal and unrefined receptor structures. For selectivity studies, the ligands of each receptor were merged together with the decoy compound library. Selectivity was measured at 5% of the database screening size.

Ligand-steered Modeling Method.

The LSHM method is outlined in Figure 6.

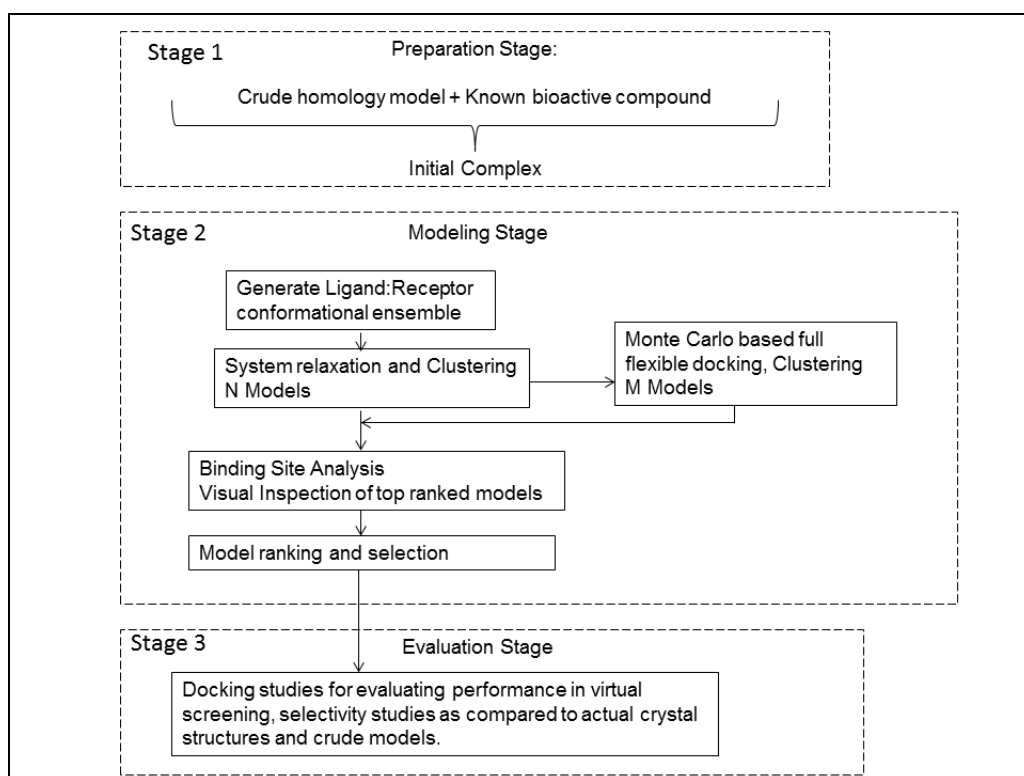


Figure 6: Flowchart of the ligand-steered modeling method.

Crude models of both the targets are developed using template crystal structures indicated in Table 2 as described before in the GPCR homology modeling section. Disulfide bonds in the extracellular loop 2 is included in the crude model building stage. For each case, Modeler9v6(51) was used to generate the crude model and residues in the TM region were

numbered per the Ballesteros and Weinstein numbering scheme described earlier. The crude model is subjected to a restrained local energy minimization to relieve structural strain and prevent the collapse of the 7TM region. Cocrystallized ligand from each target case is seeded within the general binding pocket of Class A GPCRs that is surrounded by TMs 3,4,5,6 and 7. Experimentally validated protein:ligand interactions (Carazolol's charged amine and D3.32 for cases 1 and 2, and the exocyclic N15 atom of ZM241385 and the side chain carbonyl atom of N6.65 in Cases 3 and 4) were represented in the form of quadratic distance restraints. Next, an ensemble(110) of 1000 structures was generated by randomizing the three positional and 3 rotational coordinates of the seeded ligand, followed by a multistep energy minimization where the van der Waals interaction is gradually turned from soft to full(57, 111). The entire system is described using the ECEPP/3 force field(112),charges for the ligand were obtained from the MMFF force field(113) and represented in ICM(114)The complexes are ranked using a crude binding energy calculation that included the van der Waals, electrostatic, hydrogen bonding and torsional energy terms. The solvation energy in the exposed binding site is accounted by atomic solvation parameters(115).Complexes in the top 10kcal window are clustered and variable number of cluster centers / case are subjected to a full flexible ligand: flexible receptor docking based on a Monte Carlo-based global energy minimization(116) as implemented in the ICM software platform(117-120). The minimization is performed in the torsional space. The binding site residues within 6Å of the ligand position are held flexible with the exception of their backbone. The six rigid coordinates and dihedral angles of the ligand are set free. Similar to stage 1, the complexes are scored and clustered. The two sets of cluster centers are thoroughly investigated for their ability to maintain crucial experimental mutagenesis data, to eliminate clearly wrong models (ligands moving away from the known pocket) and selected for the HTD

stage. In the case of drug discovery, only a small-set of models showing superior performance in HTD based validation may be used for actual compound screening exercises.

Clustering protocol.

The number of clusters was calculated using the data mining toolkit Rattle within the statistical package R (<http://www.r-project.org>). The binding energies of complexes in the top 10 kcal window were used as the input. The data set is partitioned into training and test sets, iterated through multiple clusters and calculates the within-cluster of sum squares. The number of clusters was chosen based on the plot where the minimum WCSS was observed. Using K-means clustering and the number of clusters, the complexes were clustered within ICM(114). Each cluster center was chosen as the complex for subsequent calculations as described in the LSHM flowchart.

Results and Discussion:

The overall results for the top 5 models of the ligand-steered cross modeling protocol are tabulated in table 3.

| Model | Case 1 bRho \rightarrow β_2 | | Case 2 A2A \rightarrow β_2 | | Case 3 bRho \rightarrow A2A | | Case 4 $\beta_2 \rightarrow$ A2A | |
|-------|--|---------------------------|---------------------------------------|---------------------------|----------------------------------|---------------------------|-------------------------------------|---------------------------|
| | Ligand ^a | Binding site ^b | Ligand ^a | Binding site ^b | Ligand ^{a, c} | Binding site ^b | Ligand ^{a, c} | Binding site ^b |
| 1 | 1.3 | 2.2 | 1.4 | 2.1 | 5.9 / 4.0 | 2.8 | 3.0 / 2.9 | 2.8 |
| 2 | 1.4 | 2.2 | 1.3 | 2.5 | 6.0 / 4.6 | 1.7 | 2.9 / 2.8 | 2.8 |
| 3 | 1.6 | 2.3 | 3.1 | 2.4 | 8.7 / 5.6 | 3.3 | 4.7 / 4.2 | 2.9 |
| 4 | 3.2 | 2.2 | 1.0 | 1.9 | 3.1 / 3.0 | 2.9 | 2.9 / 2.9 | 2.8 |
| 5 | 2.1 | 2.0 | 0.9 | 2.4 | 3.7 / 3.4 | 2.7 | 7.2 / 5.4 | 3.3 |

Table 3: ^a Heavy atom rmsd (\AA) between the ligand-steered modeled and native cocrystallized ligand (carazolol for β_2 and ZM241385 for A2A). ^b Binding site rmsd (\AA) between the ligand-steered models and the corresponding crystallographic structures (pdb codes 2rh1 for β_2 and 3eml for A2A). ^c first value corresponds to the complete (heavy atoms) of ZM241385 and the second value ignores the high B-factor phenoxy moiety of ZM241385. ^d ECL2 was not considered in this calculation.

Case-wise modeling and results for question A: Is the ligand-steered method capable to generate near-native models of known GPCR crystal structures by using one template and incorporating protein flexibility

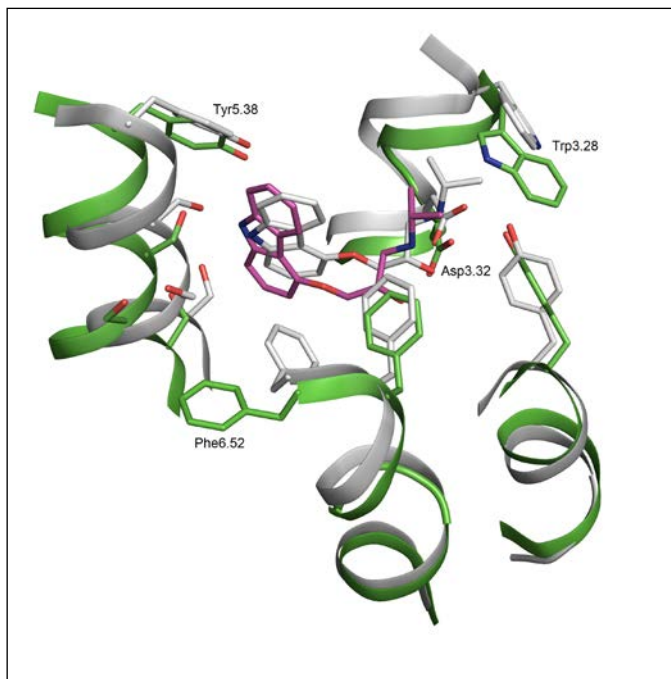
Case 1 Modeling Results.

For case 1, the structure of β_2 adrenergic receptor was developed using the bRho crystal structure as the template. The experimentally determined disulfide bond between C3.25 and Cys191 in the extracellular loop two and TM3 was retained in the modeling process. As the aim was to optimize the binding pocket of β_2 , its native co-crystallized ligand carazolol, was used for the ligand-steered optimization method. A distance restraint between the charged amine group of carazolol and D3.32 was used in the modeling process(121). 265 models from step 1 and another 314 models from the second stage were clustered to shortlist the final 5 models. For this and case 2, the final models were selected based on the following criteria

A: interaction between charged amine of carazolol and D3.32.

B: the overall location of the ligand should be between TMs 3,5,6 and 7.

Figure 7: Top ranked ligand-steered model of β_2 using bRho as the template, superimposed with the β_2 adrenergic crystal structure (pdb code 2rh1). Color code: green TMs and carbon atoms, ligand-steered model; magenta carbon atoms, modeled carazolol structure; white TMs and carbon atoms, crystal structure of β_2 ; white carbon atoms, cocrystallized carazolol. Figure prepared using Pymol (www.pymol.org).



In four of the 5 models, the ligand-steered modeling method was able to identify near-native ligand poses (defined as rmsd < 2.5Å). The best model possessed an extremely accurate (1.3Å) rmsd of carazolol. All 5 models also possessed reasonable binding site residue accuracies and this performance was similar to other publicly available data(108, 122, 123). Accurate modeling of binding site residues is important for optimal performance in docking studies. Figure 7 represents the molecular representation for the best model of Case 1. The carbazol moiety is oriented towards the serines (5.42, 5.43 and 5.46) in TM5. The hydrogen-bond interaction between carazolol's charged amine and the negatively charged carboxylate group of D3.32 is maintained in all 5 models. The ligand-steered method was able to model the correct orientation of the *chi1* side chain of F6.52 that is essential for optimal ligand pose. However, these models were not selected in the top five due to our choice of minimal manual intervention, ligand interaction energy miscalculations due to choice of force fields and /or the clustering method.

Comparison to other studies: We compared our results with modeling studies that used only one crystal structure for β_2 modeling. Mobarec et al. obtained a near native rmsd of 2.55Å for carazolol(123). Costanzi developed a β_2 model with a very low ligand rmsd of 1.7Å but he used manual optimization of the F6.52 residue to improve pose prediction(108).

Case 2 Modeling Results.

Case 2: In this case, the β_2 model was developed using human Adenosine 2A receptor.

Disulfide bonds and the charged amine:D3.32 interaction was maintained in the modeling process. A total of 18 ligand-steered models were developed.

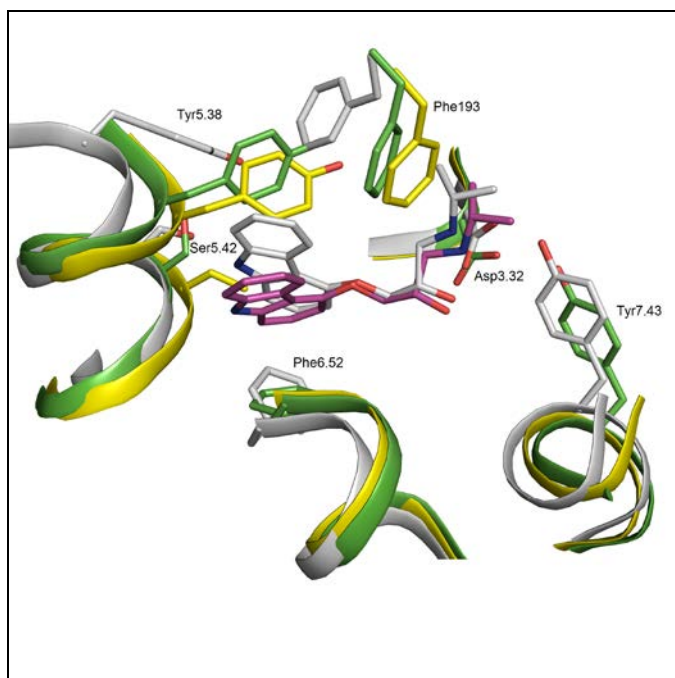


Figure 8: Top ranked ligand-steered model of β_2 (Green TMs and carbon atoms, modeled carazolol in magenta) using A2A crystal structure as a template superimposed with the β_2 adrenergic crystal structure (White TMs and carbon atoms including carazolol) and the crude homology model (Yellow TMs and carbon atoms).

The importance of the flexible ligand: flexible receptor Monte Carlo based modeling process is highlighted by the significant improvement in the optimized models as compared to the crude

models. From Figure 8 it is evident that a near native pose of carazolol in the crude model would be impossible to achieve due to the steric clashes caused by residues Y5.38, S5.42 and F193 from the extracellular loop 2 of β_2 . The incorrect orientation of the residues reduces the binding site volume. The LSHM optimized the orientation of these side chains, modeled the correct gauche+ orientation of F6.52 and F193 in the ECL2. The method increased the binding site volume encompassed by TMs 3,5,6 and 7. As a result highly accurate β_2 models (native ligand rmds < 1.0Å) were developed. However, similar to other reported literature the top models failed to identify a correct orientation of W3.28 and Y5.38(108). The reason for the incorrect orientation of Y5.38 is the bulge in the corresponding A2A template that is inherently difficult to model due to limitation of the homology modeling process.

Case 3 Modeling Results.

Case 3. In this case, the tertiary structure of human A2A was modeled using bRho.

This and case 4 are different as the rigid co-crystallized ligand of A2A, ZM241385, could not be accommodated perpendicular to the TMs of the models. First, a potential binding site located near and parallel to TMs 6 and 7 was identified in the crude model using the pocket finder algorithm implemented in the ICM platform(114). The overall geometry of the binding site supports the hypothesis of a parallel orientation of the ligand. To avoid any bias in the modeling process ZM241385 was seeded in two opposite orientations with the high B factor methoxy group facing the extracellular and intracellular regions of the TMs. The documented hydrogen bond interaction between the exocyclic N15 atom of ZM241385 and the side chain carbonyl atom of N6.55 was used as a distance restraint(124). The ECL2 was modeled de novo with only the disulfide bond restraint (C3.25 and Cys166) to accommodate the size of

ZM241385. Twice (2000) as many initial conformations were generated to accommodate for the two starting positions for ZM241385. Post the two modeling stages, 13 ligand steered models were generated. Analogous to Case 1 here the hydrogen bond interaction that was used as the restraint and using experimental information that strongly suggested interactions with residues in TMs 6 and 7, 5 models were selected and ranked based on their calculated binding energies.

The high Bfactor methoxy group was ignored in all ligand rmsd calculations as described in the recent A2A structure prediction studies published by in Nature(125). Two of the top 5 LSHM had ligand rmds of 3.1 and 3.7Å respectively. The best model by energy was slightly worse with 4Å rmsd of the native pose. In all cases the orientation of ZM241385 was correct with the methoxy group directed towards the extracellular region of the TMs. All 5 LSHM models returned a fairly reasonable binding site RMSD of 4Å.

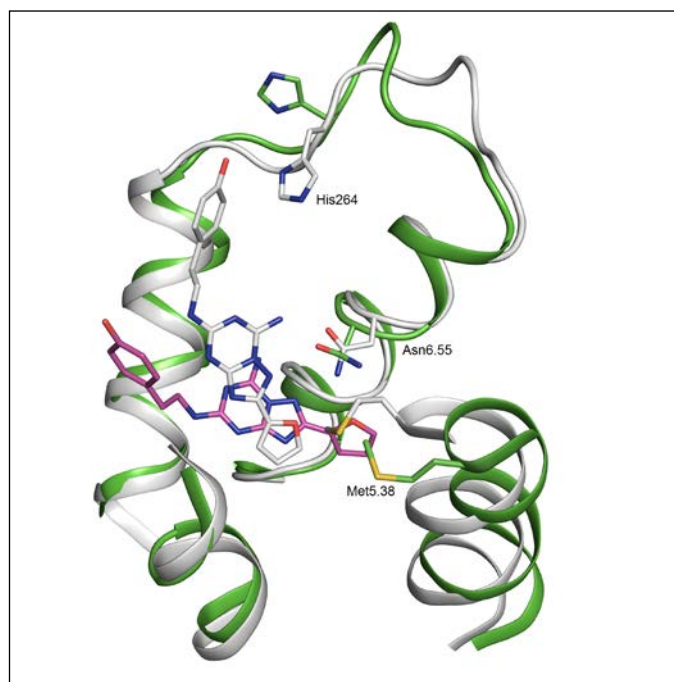


Figure 9: Case 3 modeling study results. Ligand-steered model of A2A (green TMs and carbon atoms with ZM241385 represented in magenta carbon atoms) is superimposed with the A2A crystal structure (white TMs and carbon atoms including the native pose of ZM241385).

The results of Case 3 (Figure 9) in general are slightly worse than the first two cases. Upon further investigation it is evident that bRho is not a suitable template for A2A modeling. It lacks the characteristics of A2A i.e. rotation of TM5, variability in the ECL3 which plays an important role in the positioning of ZM241385 and the signature orientation of H264 in ECL3(125). Here, the ECL2 of bRho is probably ill-suited to model GPCRs with larger native ligands. This modeling protocol was not developed to model the ECL2, thus most of the models have incorrect tertiary structure for that region of the GPCR.

Case 4 Modeling Results.

Case 4: In this final case A2A was modeled using β_2 as the template structure. Gaps in the TMs were avoided by introducing them in the ECL2 instead. As in case 3 the disulfide bond in ECL2 was maintained but unlike case 3, the ECL2 structure of β_2 was used to build the ECL2.

Similar to case 3, ZM241385 was oriented quasiparallel to the TMs 3,5,6 and 7 during the modeling stage. 16 LSHM were obtained and the top 5 were identified based on the strict visual inspection criteria outlined in Case 3 and binding energies.

The best two models produced near-native ligand poses (2.9 and 2.8Å) of the co-crystallized ligand. The distance between the side chain carbonyl atom of N6.55 and the exocyclic atom N15 of the ligand was found to be 3.7Å, comparable to other published literature and the crystal structure(124, 125). Similar to the crystal structure the top LSHM reproduced the aromatic stacking interactions between F193 and the bicyclic triazolotriazine core of the ligand structure. All models reproduced the quasiparallel orientation of the ligand when compared with the TMs. However, similar to other published literature the modeling protocol failed to identify the extended bulge of TM5 and the inward orientation of M5.38 as observed in the

crystal structure. GPCRs in general have unique structural characteristics that may not be reproducible by modeling protocols.

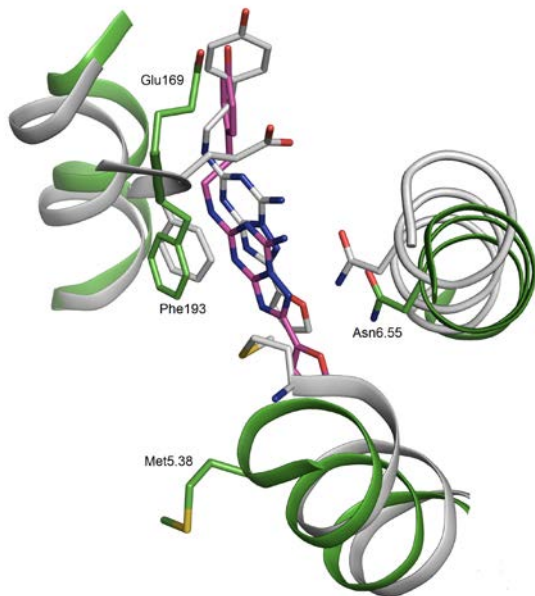


Figure 10: Case 4 modeling study results. Ligand-steered model of A2A (green TMs and carbon atoms with ZM241385 represented in magenta carbon atoms) is superimposed with the A2A crystal structure (white TMs and carbon atoms including the native pose of ZM241385).

Performance of the Ligand-steered Models in the High Throughput Docking

Experiments.

Homology models have been successfully applied to discover novel inhibitors of various therapeutic targets via virtual screening studies(27, 49). The goal of any virtual screening process is identifying a higher percentage of true positives. The hypothesis is that optimizing the binding pocket will improve the performance of the ligand-steered models as compared to the crude models and be similar to the crystal structure. Thus, in order to assess the quality of

the ligand-steered models their performance in a small-scale high throughput docking was compared against the respective crude homology model and the crystal structure.

Compound Library Preparation for HTD and Selectivity Studies.

29 β_2 antagonists and 51 A2A antagonists were assembled for the HTD and selectivity studies. ~3450 decoy compounds compatible with charged properties (+1 for β_2 and 0 for A2A) were assembled from the publicly available ZINC database(126). The physiochemical properties for both sets of libraries are tabulated in table 4. All compounds were *prepared* for HTD using the LigPrep module of the Schrodinger software suite. *Preparation* includes generating protonation states and tautomers at pH 7.0.

| Physiochemical Properties | Beta 2 compounds | Beta 2 decoy library | A2A Compounds | A2A decoy library |
|---------------------------|------------------|----------------------|---------------|-------------------|
| Molecular Weight | 318 \pm 63 | 312 \pm 27 | 300 \pm 52 | 303 \pm 32 |
| Rotatable Bonds | 7.5 \pm 2.1 | 5.1 \pm 1.8 | 4.0 \pm 1.9 | 4.5 \pm 1.7 |
| Log P | 2.4 \pm 1.2 | 2.4 \pm 1.2 | 1.5 \pm 1.2 | 1.7 \pm 0.6 |
| Hydrogen-bond Acceptors | 3.2 \pm 1.1 | 2.7 \pm 1.2 | 4.3 \pm 1.2 | 3.9 \pm 1.3 |
| Hydrogen-bond Donors | 3.9 \pm 0.9 | 2.6 \pm 1.0 | 1.8 \pm 1.1 | 1.3 \pm 1.2 |

Table 4: Physiochemical properties of compounds and decoys for β_2 and A2A high-throughput docking studies.

High Throughput Docking Protocol:

Glide, a commercially available docking program from Schrodinger was used for HTD.

Proteins were prepared using the Maestro Suite (www.schrodinger.com). The program was used in the standard precision mode without any additional constraints. Grids for the docking program were designed to cover the GPCR binding pocket. A 20 x 20 x 20 Å³ grid was developed and was centered between W6.48 and V3.33 for Cases 1 and 2 and between W6.48 and T3.36 for cases 3 and 4. The enrichment factor 2% was calculated as follows

$$EF(2\%) = (\text{hits}_{\text{sampled}}/\text{N}_{\text{sample}})/(\text{hits}_{\text{total}}/\text{N}_{\text{total}})$$

Hits_{sampled} = number of known active compounds in the top 2% of the compound database

N_{sampled} = subset of 2% of the compound database

Hits_{total} = total number of active compounds in the database

N_{total} = total number of compounds in the database.

EF2% was selected as the cutoff for performance calculation as typically in an experimental screening process the top 2% of compounds are selected for subsequent stages.

Results of HTD studies:

Table 5: High-throughput docking enrichment for crystal structures, crude homology model and top 5 ligand-steered models for all 4 cases.

| | EF(2) ^a Case 1 (bRho → β ₂) | EF(2) ^a Case 2 (A2A → β ₂) | EF(2) ^a Case 3 (bRho → A2A) | EF(2) ^a Case 4 (β ₂ →, A2A) |
|--------------------------|--|---|--|--|
| Crystal Structure | 31.0 | 31.0 | 8.8 | 12.7 |
| Crude Model | 22.4 | 1.7 | 2.0 | 4.9 |
| Model 1 | 20.6 | 24.1 | 0.0 | 7.8 |
| Model 2 | 17.2 | 18.9 | 0.9 | 3.9 |
| Model 3 | 20.6 | 22.4 | 3.9 | 4.9 |
| Model 4 | 18.9 | 18.9 | 0.0 | 5.8 |
| Model 5 | 17.2 | 18.9 | 0.9 | 6.8 |

HTD on the crystal structures of β₂ and A2A: To validate the docking protocol native ligands of the two crystal structures were redocked using the Glide software in SP mode. For β₂, the top ranked pose of carazolol had an rmsd of 0.6Å and for A2A the top ranked pose of ZM241385 was 1.6Å (excluding the high B factor methoxy group). RMSDs below 2Å are considered optimal for docking calculations. The EF(2) of β₂ and A2A was 24.1 and 8.8 respectively.

High-Throughput Docking Results.

Case 1: Contrary to our hypothesis, the ligand-steered models showed equal or marginally inferior performance in the HTD studies as illustrated in Table 5. However, these results were comparable to similar studies in the public domain. It should be noted that the ECL2 was not included in our model building stage and may have played an important role in identifying

higher false positives. The importance of modeling ECL2 is highlighted in the results of Case 2.

Case 2: Here, the crude model had an extremely low EF(2) at 1.7. The inaccuracies in the crude model i.e. steric hindrances caused due to incorrect modeling of the side chain residues of the binding pocket and smaller volume of the binding pocket. All 5 ligand-steered models showed improved performance in the HTD studies. The 2nd best ligand-steered model returned a marginally improved performance as compared to the top model (17.2 vs 15.5). However, none of the ligand-steered models were as accurate as the crystal structure of A2A.

Case 3: Case 3 was one of the most challenging cases due to the lack of ECL2 in the modeling process. Both, the crude model and all 5 ligand-steered models performed poorly in the HTD studies. The results confirm to similar studies by other researchers(125, 127). The results of Case 3 undermine the future use of using bRho as a template for GPCR modeling.

Case 4: Here, the ECL2 was modeled and included in the HTD calculations. Subsequently, three of the five ligand-steered models showed improved performance as compared to the crude A2A model. The top two models that reproduced good native ligand rmsds also had the best performance in HTD. In addition, the performance of model 2 was comparable to that of the A2A crystal structure. Other similar studies by McRobb et al. obtained EF(2) of β_2 that is in the same range as our models(122).

Performance of Ligand-Steered Models in Compound Selectivity Studies.

The identification of highly selective compounds for a given protein target over other homologous proteins is important in a drug discovery process. This may reduce side effects, improved therapeutic effects and prevent costs associated with medicinal chemistry efforts and

experimental testing. This is highly relevant for compounds of Class A GPCRs as they share similar binding pockets, and critical residues for binding (e.g. Asp 3.32) and are thus likely to interact with multiple compounds. Thus, in addition to the modeling and HTD performance the ability of ligand-steered models to identify selective compounds was tested. Our hypothesis was that the optimized ligand-steered models will be able to discriminate between target specific compounds.

Methods.

Both, β_2 and A2A compounds were merged into the decoy database to generate a library size of 3500 compounds. The fraction of β_2 and A2A compounds returned at 5% of the database was calculated. Case 3 was excluded from selectivity calculations as it had a poor HTD performance. The docking grids are identical to those used in the HTD calculations. The ligand-steered model with the best performance in HTD studies was used for selectivity calculations.

Results. Figures 11,12 and 13 depict the results of the selectivity studies on Cases 1,2 and 4. The red and gray bars indicate the percentage of β_2 and A2A inhibitors recovered at 5% of the screening database. Crude model refers to the initial model as depicted in the ligand-steered modeling flowchart. Model 1 refers to the top-ranking ligand-steered model as illustrated in Table 3.

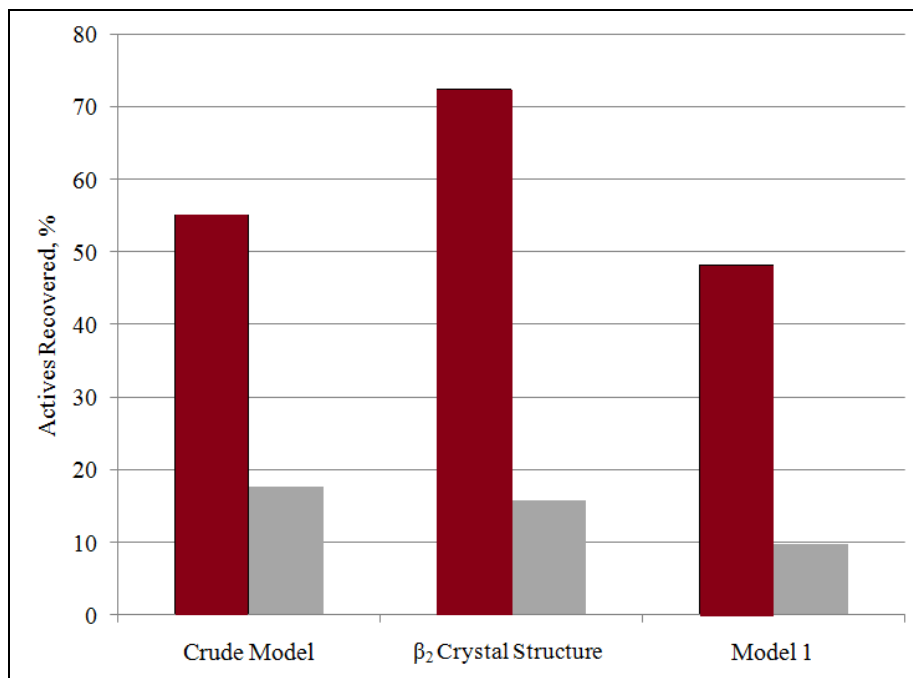


Figure 11: Case 1 selectivity study results.

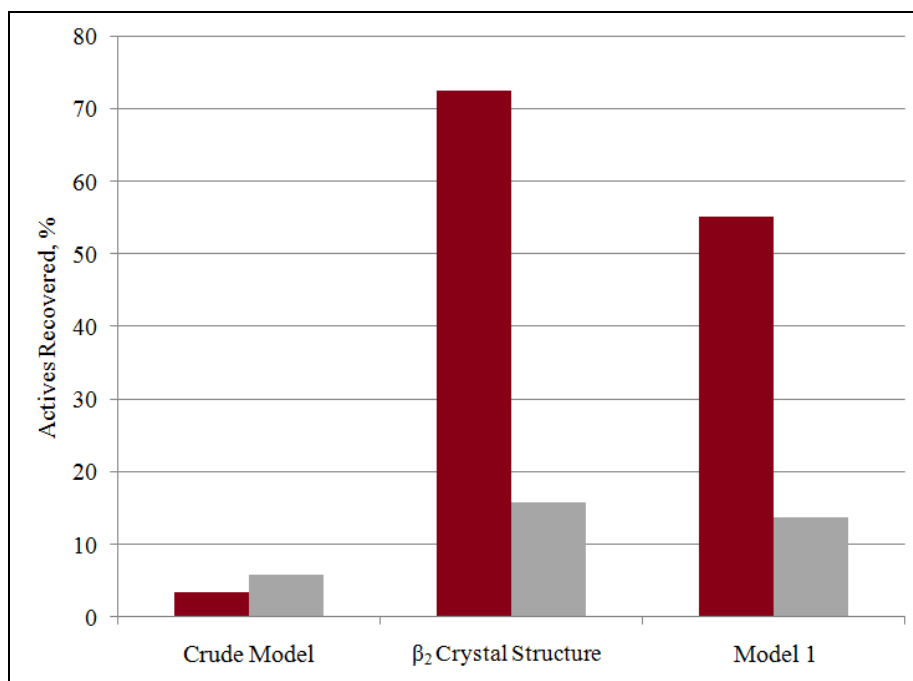


Figure 12: Case 2 selectivity results.

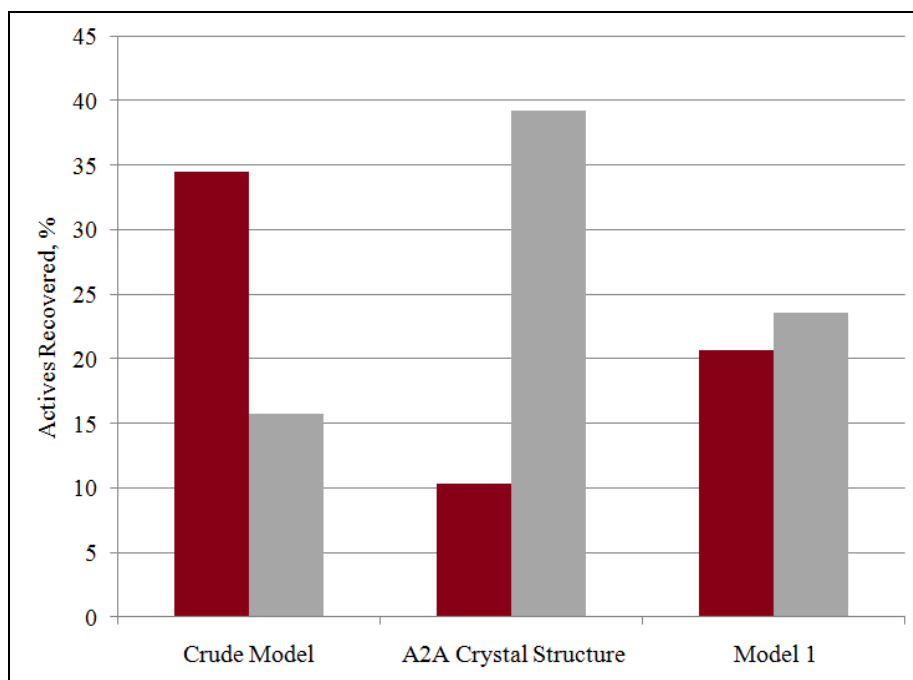


Figure 13: Case 4 selectivity results.

In all the three cases the respective crystal structures outperformed the best ligand-steered and crude homology models. In Case 1, the best ligand-steered model recovered roughly similar percentage of β_2 compounds as compared to the crude model (41 vs 43), but was able to reduce the percentage of A2A actives by 10. The biggest improvement was in Case 2, where the ligand-steered model identified ~35% more β_2 compounds while showing an insignificant increase in identifying A2A compounds as compared to the crude model (8 vs 5%). In Case 4, the optimized ligand-steered model clearly inverted the poor selectivity profile of the crude model. It identified ~12% more A2A ligands and reduced the selection of β_2 compounds by 50%. In fact, the ligand-steered models performance for identifying A2A ligands was roughly similar to that of the crystal structure (32 vs 35%) though it identified 7% more β_2 compounds. Thus overall the optimized ligand-steered models improved the selectivity profiles of native

compounds as compared to the crude models but in their selectivity profiles were inferior to the crystal structure, which is line with the overall hypothesis of this study.

Conclusions.

A: Is the ligand-steered method capable to generate near-native models of known GPCR crystal structures by using one template and incorporating protein flexibility

In all the four cases ligand-steered homology models reproduced near-native (low rmsd) poses of co-crystallized ligands. In cases 1 and 2 the poses may be classified as excellent ($< 2\text{\AA}$ rmsd). For cases 3 and 4 the results can be classified as acceptable ($\sim 3\text{\AA}$ rmsd) and are comparable to other published studies.

B: Can the ligand-steered models deliver similar or better performances in HTD studies as compared to the crystal structure and unrefined model?

As expected the crystal structure outperformed all the crude and ligand-steered models. Except Case 3, the performance of the top ranked ligand-steered model in the HTD studies was satisfactory as compared to the respective crystal structures. The top ranked ligand-steered models of Case 1 and 3 showed similar or worse performance as compared to the crude homology model. I hypothesize that the non-inclusion of ECL2 may have played an important role in these cases. It is also likely that the crystal structure of bRho might be a poor template for modeling GPCRs, a conclusion corroborated by similar studies in the literature. However, in cases 2 and 4 where the ECL2 was included in the modeling studies, the top ranked ligand-steered model performed better than the crude models and random selection. This improved performance may also suggest the applicability of the β_2 crystal structure for modeling other Class A GPCRs.

C: Is there a correlation between the best models in terms of cocrystallized ligand rmsd's and top models from the HTD experiment.

In cases 1,2 and 4 the top ranked models from the HTD experiments possessed low native ligand rmsd's as obtained from the modeling process. However this result is not generally applicable as models with higher native ligand rmsd from the modeling stages were found to be competitive in identifying native inhibitors in the HTD experiments.

D: Can optimized ligand-steered model discriminate receptor specific ligands from a set of decoy compounds?

The selectivity of the top-ranked ligand-steered models in Cases 1 and 2 was comparable to their respective crystal structures. The top-ranked ligand-steered model of Case 2 significantly improved the selectivity profile of β_2 ligands. In Case 4, the ligand-steered model clearly improved the selectivity profile when compared with the crude model.

Future Work.

Here I point out three shortcomings. The mediocre results of Cases 1 and 3 clearly highlight the importance of accurately modeling the ECL2 in Class A GPCRs. The ECL2 region of β_2 and A2A is 26 and 32AA long respectively. This loop has a well-documented role in small molecule binding(128) and is also the focus of modeling studies, particularly *de novo* methods. As the focus of this study was homology modeling and binding-site optimization using a full receptor and ligand flexibility, accurate ECL2 loop modeling was not considered. It is important to note that based on the fundamentals of homology modeling the accurate modeling of ECL2 remains extremely challenging. I hypothesize that a combination of homology

modeling and *de novo* loop modeling may offer an avenue to develop better quality homology models of Class A GPCRs.

Second, incorporating experimental information (e.g. via restraints) to predict the correct fold of the 7TMs prior to building the crude model may offset errors associated with the choice of template. Modeling the unique folds on TM5 of A2A receptors proved to be challenging using standard homology modeling methods. This resulted in incorrect side-chain orientations and subsequently affected the quality of models. Incorporating structural characteristics from multiple templates offer another way to improve model quality as shown by Filizola et al(123). Multi-template homology modeling is included in the development of the active state models of Cannabinoid 2 later in this dissertation. In general, consideration of experimental and / or structural evidence from other homologous GPCRs may limit the inherent errors associated with any protein modeling process.

Finally, modeling and energy calculations are dependent of the availability of the force field. Upon retrospection it was found that good quality models in Case 1 (defined as models with low ligand rmsd and with a correct gauche+ conformation of F6.52) were assigned poor energies thus precluding them from further calculations. The use of a force field specifically parameterized for proteins could eliminate such errors.

Chapter 4. Applications of Ligand-Steered Modeling to Drug Discovery Problems.

In the previous chapter I presented the enhanced ligand-steered homology modeling method. This method incorporates experimental and ligand information to model the pharmacologically relevant binding sites in Class A GPCRs explicitly using receptor flexibility. The retrospective benchmarking studies on then currently available crystal structures of two Class A GPCRs proved its accuracy and potential applicability to drug discovery problems. In this chapter I will describe the application of the ligand-steered modeling method to a: the development of Cannabinoid 2 homology models, b: rationalize activities of Cannabinoid receptor 2 compounds (agonists and inverse agonists) and propose their potential binding modes and c: to optimize the crystal structure of the β_2 adrenergic receptor to improve its performance in virtual screening studies.

The importance of superior quality three-dimensional protein structures in the modern drug discovery process cannot be overstated. Structure-based drug discovery programs routinely use such protein structures to probe structure-function relationships, assess target druggability and potential binding sites, identify hits via virtual screening, investigate the protein:ligand interactions at the molecular level and compound optimization(129, 130). Binding mode prediction serves as the basis to characterize protein:ligand interactions that in turn may be used for optimizing lead structures(130). Molecular docking, ligand-based methods like QSAR(28) and CoMFA(131) are routinely applied to predict protein:ligand interactions when either a: a high quality protein structure is available or b: a large set of binders is available and c: both, crystal structure and a set of binders is available. However scarce information about binding site side chain conformations, limited and / or the lack of receptor flexibility in docking programs often lead to inaccurate or inconclusive predictions. In addition when experimental

protein structures are unavailable inherent inaccuracies of homology models augment errors in prediction of protein:ligand interactions. Inclusion of validated experimental information and receptor flexibility in the docking process may reduce errors associated with protein:ligand binding pose modeling and characterization. The ligand-steered modeling protocol features concurrent optimization of both, the ligand and binding site side chain conformations to generate reasonably accurate binding modes and binding sites. Subsequent sections describe the binding mode prediction and validation through compound activity rationalization of two different classes of Cannabinoid Receptor 2 compounds by applying the ligand-steered modeling method.

Cannabinoid Receptor 2 and its Therapeutic Relevance.

The complex endogenous cannabinoid system consists of two Class A GPCR Cannabinoid Receptors (CB1 and CB2), 7 endocannabinoids and several proteins that regulate endocannabinoid metabolic pathways(101). This system regulates pain, emotion motivation and cognition by modulating neurotransmission at relevant inhibitory and excitatory synapses in the human brain(132). On a molecular level endocannabinoids are lipid-signaling molecules that interact with cannabinoid receptor subtypes(133). Thus their psychoactive pharmacological profile is similar to marijuana and Delta9-tetrahydrocannabinol(133). Endocannabinoids are implicated in several disease conditions e.g. obesity, pain management, immune response, inflammation, cardiac conditions and cancer amongst others(101, 102). Currently the cannabinoid system is considered as a promising therapeutic target and is under active investigation.

Cannabinoids activate two distinct Class A G Protein Coupled Receptors, Cannabinoid Receptor 1 and Cannabinoid Receptor 2. Overall these two receptor subtypes share 44% sequence similarity that increases to 68% in the 7 transmembrane region(133). Upon activation they inhibit adenylyl cyclase activity by attaching to the alpha subunit of the G protein of the Gi1,2 and 3 and Go 1 and 2 family(133). However differences exist in terms of localization and function. CB1 receptors are predominantly found in the central nervous system particularly in high density in the hippocampus, cerebellum, amygdala, basal ganglia, segments of the globus pallidus and striatum(101). CB1 receptor activation causes increased potassium and decreased calcium conductance that is associated with suppressed neuronal excitability and neurotransmitter release. CB1 distribution and function is linked to cognitive disorders and psychoactivity(101). Thus therapeutics developed against CB1 suffer from severe psychiatric side effects resulting in withdrawals and termination of CB1 specific drug discovery programs(134).

CB2 receptors are expressed primarily in the immune system – spleen, tonsils, mast cells, B and T lymphocytes, microglial cells and monocytes(135, 136). Contradictory reports exist about its presence in the CNS(101, 102). Increased CB2 expression is observed in different inflammatory conditions e.g. rat bladder after acute / chronic inflammation, dorsal root ganglia and spinal cord of animals post spinal nerve ligation, in women post endometrial inflammation(101, 102). Moreover increased CB2 expression after injury or inflammation coupled with the absence of immunomodulation by cannabinoids in CB2 absent mice suggests the importance of CB2 as a therapeutic target for immunomodulation(137, 138). Potential absence of CNS side-effects upon CB2 modulation differentiates it from the documented disadvantages of targeting CB1 receptor(139, 140).

Cannabinoid Receptor 2 Inactive State Modeling and Rationalization of 6-methoxy-N-alkyl Isatin acylhydrazone Compound Activities.

A novel series of 6-methoxy-N-alkyl isatin acylhydrazone derivative CB2 specific inverse agonists was developed by Diaz et al(101). This experimental study was unable to explain the diverse range of activities of those compounds and thus required molecular modeling studies to study a: the putative binding mode when complexed with CB2, b: identify protein:ligand interactions and c: subsequently rationalize compound activities. Here, I used the crystal structure of β_2 adrenergic receptor as the template to construct the ligand-steered model of CB2 in the inactive state.

Method.

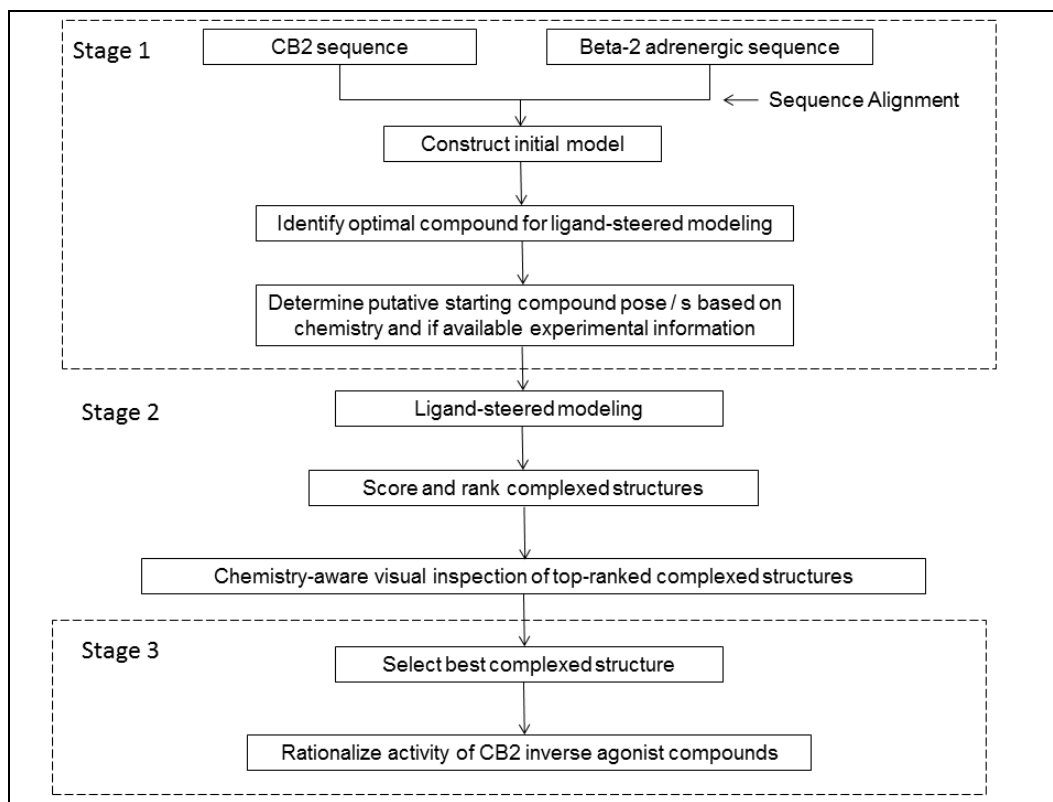


Figure 14: Flowchart of the CB2 modeling protocol.

The CB2 and β_2 sequences were aligned using the standard GPCR specific alignment methods i.e. using known conserved residue patterns and eliminating any gaps in the transmembrane regions. The second highly conserved Tyrosine residue of TM5 was used for the alignment of the TM5 region as CB2 lacks the normally conserved proline residue. The N- and C-terminus residues (AA1-26 and AA316-360 respectively) of CB2 and the artificially introduced T4L residues connecting helices V and VI of β_2 adrenergic receptor were ignored from the calculations as they are not pharmacologically relevant. This alignment was used as an input to Modeller9V4 to develop a crude model. A restraint based on experimental evidence was used to maintain a disulfide bond between Cysteines 174 and 179 in the extracellular loop 2 regions of the CB2 receptor. Subsequently, a restraint minimization was carried to relieve any structural strain caused by the non-conserved residues in the crude model building process. This model served as the input to the ligand-steered modeling protocol.

Incorporating experimental information e.g. mutagenesis data is one way to improve the quality of homology models. The ligand-steered modeling method relies on using such information to shape and optimize binding pockets. At the time of this work limited CB2 modeling, structural information and experimental data was available in the public domain(141-145). The available data suggested the presence of a hydrophobic pocket surrounded by TMs 3,5,6 and 7. The dominant hydrophobic residues F3.36, W6.48 and W5.43 play an important role in ligand recognition. Gouldson et al. reported potential hydrogen-binding interactions of SR144528 with Ser161 and 165 in TM4 regions of CB2(145). Another study by Montero et al. used a flexible docking protocol to predict the binding mode of the same compound and obtained similar results(143). In both cases the CB2 model was developed using the crystal structure of bRho. I have already described the limitations of using bRho as a template for Class A GPCR modeling

in the previous chapter of this thesis. In addition considering the limited experimental data related to hydrogen bonding the ligand-steered protocol was modified such that the restraints restricted the position of the ligand within the 7TM binding pocket.

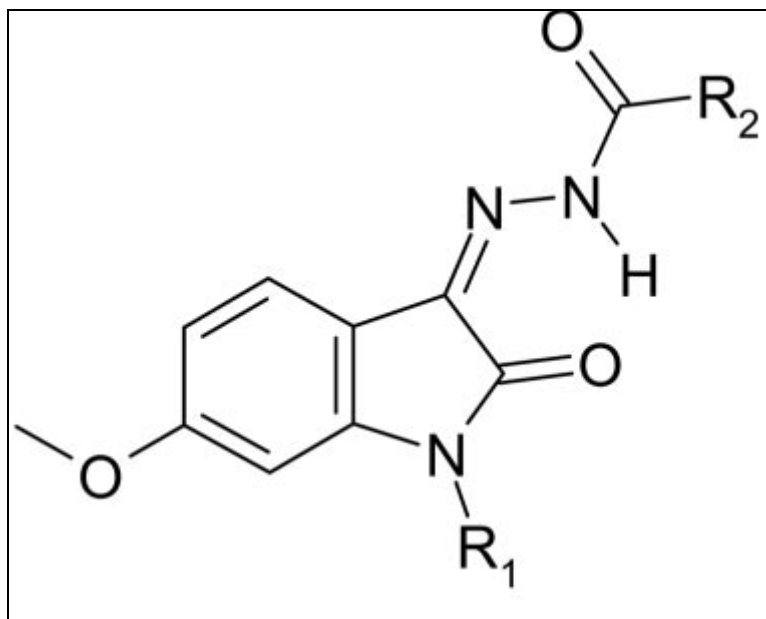


Figure 15: Isatin acylhydrazone scaffold.

Compound 18 was selected for the modeling process because it's median activity. Using the experimental binding data I hypothesized that the sensitivity of EC_{50} values was dependent on the substitutions at the R1 position as shown in Figure 15. Also, as the extracellular part of GPCRs is solvent exposed it was assumed that the R1 group is positioned toward the dominant hydrophobic intracellular part of the TMs. Two initial seed poses were selected such that the R2 group occupied the pockets facing TMs 2,3 and TMs 3,4 and 5. This was done because mutagenesis data indicated possible hydrogen bond interactions with the serines in TM4 and lysine in TM3. Moreover the experimental data indicated lack of affinity of those isatin derivative compounds without the carbonyl group. The two opposite poses made the ligand-steered modeling unbiased. The protocol involved generating an ensemble of 200 structures by

randomizing the position and orientation of the two initial seed positions and multi-step energy minimization where the van der Waals interaction was switched from soft to full interactions as described in the benchmarking study. The binding energy was estimated as the ligand-receptor interaction energy where the van der Waals, electrostatic, hydrogen bonding and torsional terms were considered. Ten structures ranked by a crude binding energy estimation were subjected to a full flexible-ligand:flexible-receptor side chain Monte Carlo-based global energy optimization. Side chains within 6Å of the compounds were considered free, the backbone was kept fixed to maintain the structural integrity of the 7TM structure. Two representative complexes with best binding energy estimates were selected for visual inspection. The final complex was selected where a: the R1 group was oriented towards the hydrophobic binding core, b: putative hydrogen bond interaction of the R2 carbonyl group to a charged side chain residue.

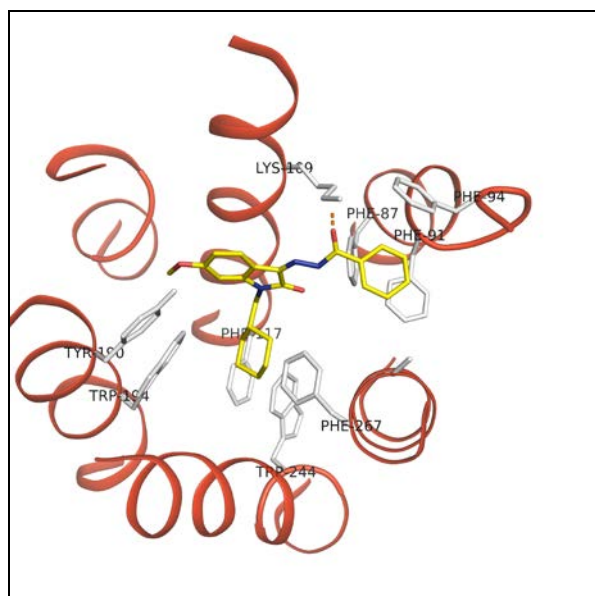


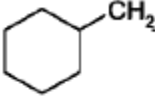
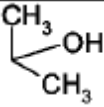
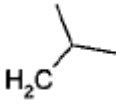
Figure 16: The proposed ligand-steered model of CB2 receptor complexed with compound #18 of the isatin series. Putative hydrogen bond with Lysine 3.28 (109) is represented with dashed orange lines. Π -

Π stacking is observed with F2.64 (94) and F2.61(91). The backbone of CB2 receptor is colored red.

Figure prepared with Pymol (www.pymol.org)

Rationalization of CB2 Inverse Agonists Structure-Activity Data.

I evaluated the accuracy of the representative model using the existing structure-activity data as at the time of development there was limited experimental data and no crystal structure for the cannabinoid receptor 2. In the representative model with compound 18 distances between the carbonyl group and the side chain polar hydrogen of K3.28 (109) and the methoxy group and Y5.39 (190) seem to favor potential hydrogen bonding interactions. The cyclohexyl moiety of compound 18 is ideally located to form van der Waals interactions in the deep hydrophobic pocket surrounded by residues W6.48(244), F3.36 (117) and W5.43(194). These observations suggest the orientation of the compounds. For rationalizing the structure-activity data we use respective substituents at R groups without changing the orientation of the scaffold (Please refer to Table 6). Based on the CB2 model and binding data it is clear that the optimal R1 substituent moiety is a n-pentyl group. Compound 16 the R2 phenyl group is located in the overall lipophilic pocket surrounded by residues F2.57(87), F2.61(91), F2.64 (94) and F7.35 (267) and is likely to form favorable aromatic stacking interactions with F2.61 and F2.64. The variation in the R1 group substituent lengths of compounds 15 and 17 may disrupt the aromatic stacking interactions of the R2 phenyl group explaining their reduced potency as compared to compound 16. Introduction of a polar oxygen atom in the R1 group substituent of compound 21 is poorly tolerated in the dominant interior hydrophobic core. Introduction of an additional alkyl group between the cyclohexyl ring and the nitrogen atom at the R1 group in compound 20 is suspected to cause steric clashes with W6.48 and F3.36 explaining the drastic loss of activity.

| Compound | R1 | R2 | EC ₅₀ (nM ± SEM) | E _{max} % |
|----------|---|---|-----------------------------|--------------------|
| 7 | | | 8.66 ± 1 | 100 |
| 15 | CH ₃ (CH ₂) ₅ | phenyl | 88.2 ± 1.5 | -92 |
| 16 | CH ₃ (CH ₂) ₄ | phenyl | 5.8 ± 8.8 | -108 |
| 17 | CH ₃ (CH ₂) ₃ | phenyl | 85 ± 1.6 | -95 |
| 18 |  | phenyl | 18 ± 1.2 | -82 |
| 19 | benzyl | phenyl | 102 ± 1.5 | -96 |
| 20 | 2-cyclohexylethyl | phenyl | 540 ± 1.3 | -104 |
| 21 | CH ₃ O(CH ₂) ₂ | phenyl | 492 ± 1.2 | -60 |
| 22 | CH ₃ (CH ₂) ₄ | cyclohexyl | 28.7 ± 1.2 | -105 |
| 23 | CH ₃ (CH ₂) ₄ |  | 14.7 ± 1.7 | -99 |
| 24 | CH ₃ (CH ₂) ₄ | -O-tBu | 15.9 ± 1.3 | -97 |
| 25 | CH ₃ (CH ₂) ₄ |  | 11.7 ± 1.4 | -78 |

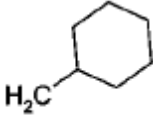
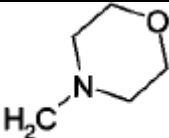
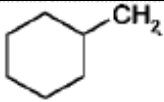
| | | | | |
|----|---|---|------------|-----|
| 26 | CH ₃ (CH ₂) ₄ |  | 43 ± 1.4 | -88 |
| 27 | CH ₃ (CH ₂) ₄ |  | 13.3 ± 1.5 | -80 |
| 28 | CH ₃ (CH ₂) ₄ | <i>Tert</i> -butyl | 10 ± 1.1 | -78 |
| 29 |  | cyclohexyl | 71.6 ± 1.4 | -69 |

Table 6: CB2 assay data. Compound 7 is the reference compound. Further details can be found here.

The large lipophilic pocket surrounding the R2 group of compound 16 was used to explore various R2 substituents while maintaining the optimal n-pentyl group for R1. The cyclohexyl and methoxycyclohexyl moieties of compounds 22 and 26 respectively are unable to maintain the optimal aromatic stacking interactions of compound 16 resulting in lower CB2 functional activity. However, branched alkyl moieties in compounds 23-25 and 28 were better tolerated. For compound 27 the methylmorpholine R2 substituent resulted in 4-fold increase in CB2 functional activity as compared to compound 26. It was hypothesized that due to the chair conformation of the morpholine ring its nitrogen atom may form a hydrogen bond with the hydrazoic moiety. In addition the additional CH₂ group may solvent expose the morpholine ring increasing its tolerability. The reduced activity of compound 29 as compared to compound 18 is attributed to the loss of stacking interactions of the cyclohexyl R2 group.

This was the first reported homology and molecular modeling study of the CB2 receptor that used the β_2 adrenergic crystal structure as the template. The adapted ligand-steered homology method proposed the binding mode for Isatin derivatives and also successfully rationalized their structure-activity data. This study suggests that the lipophilic pocket surrounded by residues W6.48, F3.36, V6.51, W5.43 and W5.46 is likely to be involved in van der Waals interactions with the R1 group substituents of the Isatin derivatives. The molecular modeling studies also suggest that potent compounds in this series either form van der Waals or Π - Π stacking interactions with F2.57, F2.61 and F2.64 residues. In addition the proposed model suggests that residues K3.28 and Y5.39 may form hydrogen bond interaction with the carbonyl and methoxy groups of the compounds respectively.

Cannabinoid Receptor 2 Active State Modeling.

CB2 receptor is an emergent target for the treatment of neuropathic pain(133, 146). The development of CB2 specific agonist compounds is an area of active research as evident by the spurt in patents filed over the last three years(102). In this section of my thesis, I will explain the development of a multi-template based ligand-steered model of the CB2 receptor in its active state and its application to explain putative ligand:receptor interactions for the benzofuran derivatives CB2 agonist compounds.

The quality and accuracy of homology models depends on the choice of the template structure(27). Usually one template structure is chosen for the modeling process but if the structure does not represent a physiologically relevant conformation the resultant model is likely to be inaccurate. One way of such errors is by incorporating two or more templates by a process known as multi-template modeling(147, 148). Multi template modeling has been

successfully applied to improve the accuracy of homology models in several cases(147, 148). In the case of Class A GPCRs including CB2 experimental / mutagenesis data indicates the characteristics of an active state GPCR include rotational movement of TM3 and / or TM6, translational movement of TM5 towards TM6 and breakage of the *ionic lock* between R3.50 and D/E6.30 residues amongst other(102, 142). At the time of this work there was not a single crystal structure for the active state CB2 receptor and any other holo active state GPCR. However, the apo crystal structure of opsin captured the characteristics associated with TMs 5 and 6 and the breakage of the ionic lock that is important for agonist binding(97, 149). The opsin structure however failed to characterize the movement of the ECL2, which plays an important role in capturing diffusible ligands by providing them access to the intracellular parts of the GPCR(150-152). As described in the earlier chapter ECL2 plays an important role in GPCR ligand binding highlighting the need to model it as accurately as possible. The last chapter also described the limitations of homology modeling to model major TM shifts that subsequently introduced errors in two of the four cases I investigated. This movement of the ECL2 is captured in the crystal structure of the β_2 adrenergic receptor which in turn is complexed with an inverse agonist compound rendering it doubtful for active state GPCR modeling by itself. Considering both crystal structure were inadequate to model the active state structure of CB2, I incorporated specific features from both e.g. TMs 1-4 and ECL2 from the β_2 adrenergic receptor and TMs 5-7 from the ligand-free opsin crystal structure to develop a putative active state model of the CB2 receptor.

Method.

Sequence alignment: The sequences of β_2 adrenergic 2RH1 and ligand-free opsin 3CAP were aligned using the conserved residue patterns as described before. This sequence was provided as the input for Modeller9v4(51). The N- and C-terminus residues (AA 1-26 and AA 316-360) of CB2, the T4L residues of β_2 , TMs 5-7 of β_2 and TMs1-4 of ligand-free opsin were disregarded in the crude model building process. The disulfide bond between Cys 174 and Cys 179 was maintained in the ECL2(143, 145). After the crude model was developed and based on the validated experimental information related to TM 3 rotation, the helix 3 of CB2 was manually rotated counter clockwise (as seen from the extracellular side) by 60 degrees. The model was restrained minimized as described before in the CB2 inverse agonist modeling.

Experimental information for initial ligand placement and ligand-steered modeling.

Prior structure-activity relationship and mutagenesis data highlighted the importance of the aromatic pocket surrounded by residues Y5.39, F5.46, W5.43 and W6.48 and hydrogen bond interactions with residues S3.31, T3.35, Y5.39 and N7.45(141, 153-157). Compound 33 of the benzofuran series of CB2 specific agonists was seeded in two initial poses such that the R1 group was oriented towards the hydrophobic pocket located in the interior of the transmembrane domain. The R2 group was oriented such that the carbonyl group could possibly form hydrogen bond interactions with S3.31 or Y5.39 per the available mutagenesis data(154). The ligand-steered modeling was performed as described before and the final model was chosen such that a: the R1 group was oriented towards the interior of the transmembrane regions and b: the carbonyl group was positioned to form an h-bond interaction.

Rationalization of CB2 Agonist Structure Activity Data.

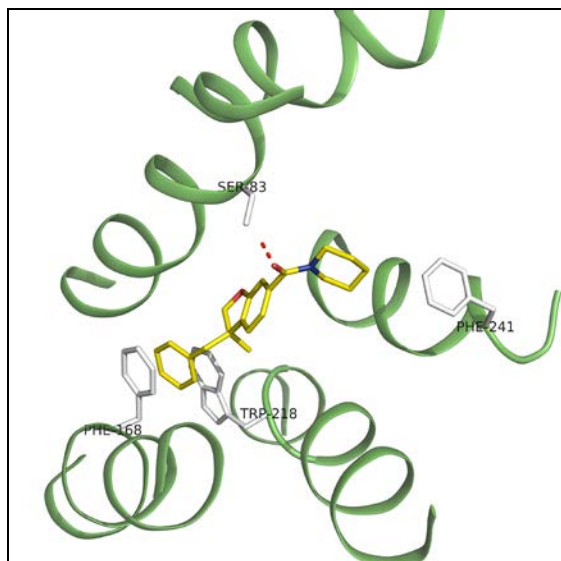


Figure 17: Proposed model of CB2 complexed with compound 33 of the agonist benzofuran series. The transmembrane regions are colored green, carbon atoms of side chains are colored white and carbon atoms of compound 33 are colored yellow. Putative hydrogen bond between Ser3.31 (83) and compound 33 is represented in red dashes. Figure was prepared using Pymol (www.pymol.org)

The modeling results were validated by the ability to rationalize structure-activity binding data of the benzofuran series of compounds. The carbonyl group of compound 33 is located at 1.8Å from the side chain of S3.31 (83) replicating a putative hydrogen-bond interaction as suggested in other published work. The absence of this hydrogen-bond results in loss of affinity / activity of CB2 agonists. The phenyl substituent at the R1 position forms van der Waals interaction with the aromatic residues F5.36 (168) and W6.48 (218). 1-piperidyl substituent at the R2 group is oriented towards the F7.35 of the 7th TM and this overall conformation suggests a putative binding mode for benzofuran derivatives CB2 agonist compounds. Next, I will explain the variation of activities associated with various substituents at the R1 group.

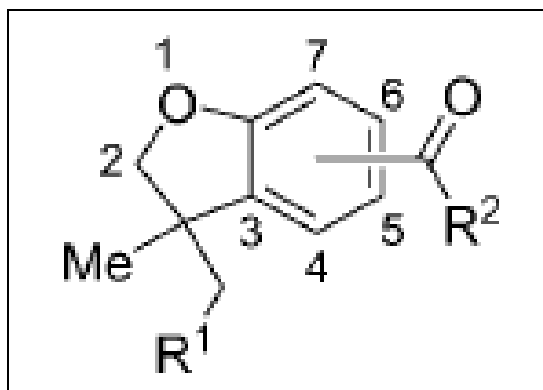
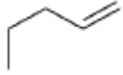
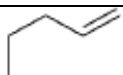
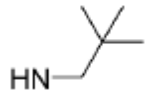
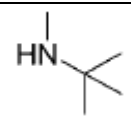
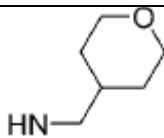


Figure 18: Benzofuran scaffold for CB2 agonist compounds.

| Compound | R1 | R2 | R2 position | EC ₅₀ (nM ± SEM) | E _{max} % |
|----------|------------------------|------------------|-------------|--------------------------------|--------------------|
| 13 | phenyl | N-(2-iodophenyl) | 5 | >10000 | ND |
| 14 | phenyl | N-cyclohexyl | 5 | 406 ± 1.4 | 47.1 |
| 15 | phenyl | 1-piperidyl | 5 | >10000 | ND |
| 16 | phenyl | N-(2-iodophenyl) | 6 | >10000 | ND |
| 17 | phenyl | N-cyclohexyl | 6 | 478 ± 1.3 | 57.3 |
| 18 | phenyl | 1-piperidyl | 6 | 128 ± 32 | 88.3 |
| 33 | S enantiomer (1) of 18 | | 6 | 108.2 | 86 |

| | | | | | |
|----|---|---|---|------------|------|
| 34 | R enantiomer (2) of 18 | | 6 | 960.89 | 43.3 |
| 19 |  | 1-piperidyl | 6 | 48.9 ± 1.4 | 97.1 |
| 20 |  | N-cyclohexyl | 6 | 839 ± 5.5 | 105 |
| 21 | phenyl |  | 6 | 246 ± 1.3 | 48.7 |
| 22 | phenyl | morpholine | 6 | 5583 ± 4.4 | 93 |
| 23 | phenyl |  | 6 | 95.3 ± 1.8 | 91 |
| 24 | phenyl |  | 6 | 659 ± 1.4 | 49 |
| 27 | 1-naphthyl | N-cyclohexyl | 6 | >10000 | ND |
| 28 | 1-naphthyl | 1-piperidyl | 6 | >10000 | ND |
| 29 | 2-naphthyl | 1-piperidyl | 6 | 875 ± 1.4 | 112 |
| 30 | 4-chlorphenyl | 1-piperidyl | 6 | 56.2 ± 1.2 | 102 |
| 31 | 4-methoxyphenyl | 1-piperidyl | 6 | 234.68 | 77.1 |

| | | | | | |
|----|------------|-------------|---|---------|------|
| 32 | 4-pyridine | 1-piperidyl | 6 | 2801.44 | 62.5 |
|----|------------|-------------|---|---------|------|

Table 7: CB2 agonist compound assay data. Further details can be obtained here.

Introduction of a chlorine (compound 30) as against the polar moieties methoxyphenyl (compound 31), 4-pyridine (compound 32) in the aromatic pocket increased CB2 functional activity. It is clear that polar moieties are not tolerated as R1 substituents. The bulky 1-naphthyl group at R1 (compound 28) caused complete loss of activity as it potentially causes steric clashes with side chains in TM3. On the other hand, smaller alkenyl group (compound 19) is accommodated within the pocket with less optimal hydrophobic interactions. Similar to compound 28, the 2-naphthyl group of compound 29 does not form the pi-stacking interactions with W6.48 resulting in less optimal affinity as compared to compound 18. Two potential causes, steric clashes of 1-naphthyl and disruption of the crucial hydrogen-bond interaction with S3.31, explain the complete loss of functional activity of compound 27. For compound 33 the *R* enantiomer of compound 33 flips the benzofuran core decreasing its activity. This modeling study for benzofuran series of compounds indicated that only a phenyl or alkene groups may be tolerated in the interior aromatic pockets of CB2 receptors.

Next, the effect of R2 group substituents is investigated with this CB2 agonist state model. Replacing the piperidine ring of compound 19 resulted in direct steric clashes with F7.35 decreasing CB2 functional activity. Moving the carboxamide moiety to position 5 in compounds 13-15 resulted in loss of the important hydrogen-bond interaction with S3.31 that explained the significant loss of activity. This explanation hold true for compound 16 as well. Compounds 21 and 23 with neopentylamine and *N*-tert-butylmethylamine moieties respectively

were nearly equipotent with compound 18. The marginal loss may be attributed to the lower lipophilicity of the neopentylamine nitrogen as compared to the trisubstituted piperidine nitrogen of compound 18.

In the previous chapter I highlighted few shortcomings of homology models. Choice of template/s is crucial to develop a better quality initial homology model. Homology modeling by principle is unable to model large structure movements e.g. the TM5 of Adenosine 2A receptor. Errors introduced by the wrong choice of template are generally impossible to correct. In the above two case studies of CB2 receptors I have shown that choosing a better quality template can potentially reduce some of the errors. As shown in previous chapter, the β_2 adrenergic crystal structure is definitely a better template choice. The CB2 inverse agonist model is developed using the β_2 adrenergic crystal structure as compared to the previous studies where the bRho template was used (primarily due to the unavailability of any other GPCR crystal structure). Multi-template modeling is another way to reduce errors associated with choice of a single template. Multi-template modeling often compensates for the structural deficiencies of one template. It is thus possible to include structural variability that would be extremely difficult to model otherwise. By using experimentally proven information and features within two templates that potentially corroborate those features, an agonist state model of CB2 was developed.

Like all modeling methods, homology modeling is an abstraction and a majority of crude models need refinement prior to any meaningful drug discovery applications. Accurate characterization of at-least the binding sites, determining exact side chain conformations are particularly important as even minor inaccuracies may lead to incorrect results. The validation studies of other GPCR crystal structures using the ligand-steered modeling method proved that

it is possible to generate reasonable binding poses of co-crystallized ligands along with the reasonable characterization of the binding site side chains. Known experimental information and a full flexible-ligand: flexible-receptor approach was used to shape and optimize the binding pockets of both CB2 models. In either case unavailability of the actual crystal structure of the CB2 receptor precluded any retrospective validation. Thus, in both cases the large experimental structure-activity data of inverse agonists and agonist compounds was used to validate the models. The two models were successful in rationalizing the activities of all the compounds provided by our collaborators at the MD Anderson Cancer Center. The models may find potential applications in inhibitor discovery by virtual screening studies.

In the benchmarking study, the crystal structures undoubtedly had the best performances in the small scale HTD experiments. The goal then was to check the retrieval rates of inhibitors. The results though good can potentially be improved as receptor flexibility was not considered. Moreover as shown in Case 2 of the CB2 modeling studies, agonist compounds too are useful for therapeutic applications. However, active state GPCR crystal structures did not exist at the time of this work. It was but natural to use known agonist compounds to optimize existing crystal structures and evaluate their performance in a virtual screening study. The next question attempted was to check if optimization of known GPCR crystal structures using both, inverse agonist and agonist compounds, will lead to better performances in structure-based virtual screening studies. The crystal structure of β_2 -adrenergic crystal structure was used as a case study. The underlying hypothesis is that ligand-steered modeling can optimize any type of protein structure by modeling side chain conformation variability thus increasing the retrieval rates in HTD studies.

Optimization of β_2 Adrenergic Crystal Structure and Docking Studies.

Background: The crystal structures of GPCRs clearly outperformed the best ligand-steered homology models in the HTD of the ligand-steered benchmarking and another study by Costanzi et al(108). In my study the receptor flexibility of the crystal structures of both, the β_2 adrenergic and the Adenosine 2A receptors, was not considered. Thus, theoretically it is possible to improve the performance of the HTD experiments by incorporating receptor flexibility. This section describes a collaborative study (primarily conducted by Costanzi et al.) where a receptor-ensemble docking protocol was developed to account for receptor flexibility and applied to improve HTD performance of the β_2 -adrenergic crystal structure(158). I optimized the β_2 adrenergic crystal structure by the ligand-steered modeling protocol in the presence of two beta-blockers, the co-crystallized carazolol and carvedilol, and one agonist ritodrine. Ritodrine was used to optimize β_2 crystal structure as there is no agonist / active state β_2 -adrenergic crystal structure and one of the aims of this study was to investigate the agonist only retrieval performance of β_2 when optimized with an agonist compound.

Method.

The details of compound selection, protein crystal structure preparation, and docking protocols is explained in the original paper. I chose carazolol and carvedilol for the inactive state ligand-steered modeling stage because a: carazolol is the co-crystallized ligand and b: carvedilol ranked poorly in the benchmarking docking with the crystal structure. The structure of ritodrine was used for the active state ligand-steered modeling because it was one of the top ranked agonist compounds in the initial HTD study. The β_2 crystal structure was optimized using the ligand-steered modeling method described earlier in this thesis. The initial conformations of

carvedilol and ritodrine were chosen such that they showed maximum overlap with the co-crystallized carazolol conformation. A quadratic restraint was maintained between the charged amine of each compound and D3.32. In this study the three positional and three orientational coordinates of the each compound, torsional coordinates of the compounds and the side-chains within 6Å radius of the ligand position were considered free. The backbone kept was kept rigid in the modeling process. HTD results of the crystal structure, the top ranked ligand-steered optimized structure and a combination of the crystal and modeled structures were compared.

Results and Discussion.

For the carazolol optimized structure the enrichment rates (EF10) improved by 10% and 5% as compared to the crystal structure using the high-throughput virtual screening and standard precision scoring functions respectively(158). A combination of the carazolol optimized and crystal structure results showed similar improvement. In the case of carvedilol optimized β_2 adrenergic structure the performance was worse by 14% and marginally lower by 4% while using the HTVS and SP scoring functions. However a combination of carvedilol-optimized structure and the actual crystal structure returned comparable results in the HTVS case and ~10% lower in the SP case. In the case of carvedilol the poor results are not unexpected.

Carvedilol is a non-selective low efficacy β_2 / β_1 inhibitor. In addition the additional degrees of freedom inherent in carvedilol's structure as compared to carazolol are likely to have had an adverse effect on the ligand-steered modeling. This may have resulted in incorrect side-chain orientations of the binding pocket residues causing subpar virtual screening results. Moreover carvedilol was ranked as a poor binder in the HTD studies against the β_2 adrenergic crystal structure. This case shows the importance of the choice of ligands to be selected for the ligand-steered modeling.

Reversing the results of carvedilol optimized β_2 structure, the ritodrine optimized putative agonist state β_2 adrenergic structure performed comparably / better in the HTVS / SP docking protocols.

Chapter 5: Modeling Studies for p65 using Open Source Libraries.

Protein modeling and incorporation of target flexibility is an important aspect of the structure-based drug design methodology. Commercial software, often in a black-box mode, offer easy to use access to modeling algorithms and methods. However, in order to widely disseminate these methods in the scientific field it is imperative to port such methods into open source modules. The ligand-steered method was developed within the framework of commercial molecular modeling software. However for wider applicability I investigated the possibility of using the comprehensive open source libraries from the Rosetta suite of programs(159) and retaining the essence of the original method i.e. manipulating side chain conformations using ligand information. The method was applied to explain the possible lack of binding of p65, a subunit of the NF-KB heterotrimer and DNA in the presence of 3-formylchromone. This study aimed to elucidate the possible relation between 3-formylchromone, an anti-tumor agent and the NF-KB pathway, which plays an important role in tumorigenesis and inflammation(160).

Method. The Rosetta suite of programs offers a wide range of utilities to develop custom molecular modeling algorithms with Rosetta sampling and scoring algorithms(161). Recently the development of PyRosetta offers a python-based scripting interface for rapid prototyping and implementation(159). By using the functions available within PyRosetta and retaining some aspects of the ligand-steered modeling method, I developed a small custom platform for the p65 case study. The results suggest possible rearrangement of the flexible loop region at the interface of p65-DNA binding region that likely precluded the binding of these two biological entities. Mutation of a crucial Cysteine residue in the loop region to serine abolished this effect and the lack of any conformational change in subsequent modeling studies seemed to support the biological data.

To enable optimization of the binding pocket it is imperative that a starting protein:ligand complex either via crystallization or docking method or manual positioning of the ligand in the binding pocket is available. p65 (pdb code 1VKX) does not have a well-defined binding pocket as in the case of Class A GPCRs. In addition, the complexed structure of p65-3FC is unavailable. However, a literature survey implicated residues Arg³³, Arg³⁵, Tyr³⁶, Glu³⁹ and Arg¹⁸⁷ in the binding with DNA(162). Thus in order to generate an unbiased initial position of 3FC in complex with p65, a soft-docking study was carried out. The assumption here is that soft docking will compensate for limited receptor flexibility and permit docking poses that would be unrealistic with a rigid receptor approach. A docking grid was calculated using the above-mentioned residues and soft-docking studies were carried out using the Glide commercial docking software. The top-ranking pose was used as a starting point for the optimization process.

Though the essence of the ligand-steered modeling method is retained, this method differs in some aspects. Here side chain flexibility is incorporated as repacking of side-chain low energy rotamers as compared to the full flexible approach described earlier. Rotamers offer a time-effective method for modeling most likely side-chain conformations with the assumption that the low energy rotamers are correct. The Rosetta rotamer library is developed using high-quality protein structures in the PDB and is used in this implementation using python bindings(159). Next, using the starting pose from Step 1, residues within 6.5Å were identified for side-chain rotamer and backbone packing. The backbone atoms of residues within a well-defined secondary structure were kept fixed. An ensemble of 100 complexes was generated using a Monte-Carlo minimization method. For every iteration, the ligand 3FC was randomly rotated and translated (1.0Å and 1.5Å) from the starting pose obtained earlier. These limits

were determined such that the ligand remains roughly in the binding pocket. Here, no restraints were incorporated, as there was no evidence in the literature about any potential ligand receptor interactions. For each positional change of the ligand the receptor residues were optimized using low-energy side chain rotamers. The resulting complex was scored using the all-atom Rosetta ligand scoring function and retained-rejected based on the Metropolis criterion. The resultant complexes were ranked per their energy values, visual inspection (checked for clashes or movement of 3FC beyond the binding pocket) and the Ramachandran plot to discard complexes that have side-chain conformations located in theoretically unfavorable regions. 5 top scoring complexes were selected for the final docking process. Here, the Glide software in the extra-precision mode (www.schrodinger.com) was used to re-dock 3FC to each of these 5 complexes. The complex with the best-scored pose was retained for subsequent analysis. This method was repeated for the Cys³⁸Ser mutation in p65.

Biology. Dysregulated inflammatory pathways are the root of most chronic diseases including cancer. Molecular agents that can suppress pro-inflammatory pathways have potential applications as anti-cancer or chemopreventive therapeutics. NF-KB pathway has a critical role in inflammation and tumorigenesis(163). It is present in the cytoplasm as an inactive heterotrimer consisting of p50, p65 and IK β kinase (IKK) subunits. Upon activation by carcinogens, tumor promoters or pro-inflammatory agents, IKK is phosphorylated, ubiquitinated and degraded(164). The p50/p65 subunits are released and translocate to the nucleus. Here upon binding to specific DNA sequences results in activation of over 500 genes linked with inflammation, cellular transformation, tumor cell survival, proliferation, invasion, angiogenesis and metastasis(165). Most tumor cells express constitutive NF-KB and hence this pathway is likely to be important for cancer prevention and treatment studies.

3-formylchromone has been associated with anti-cancer potential via an unknown mechanism. In this study it was hypothesized that 3-FC may mediate its effects via the modulation of the critical NF-KB activation pathway(160). The study was designed to test this hypothesis using a series of biological assays on NF-KB regulated gene products. NF-KB activation was induced mostly by TNF- α per solid literature evidence. More details on the biological assays can be found here(160).

Proposed Results of Modeling Studies.

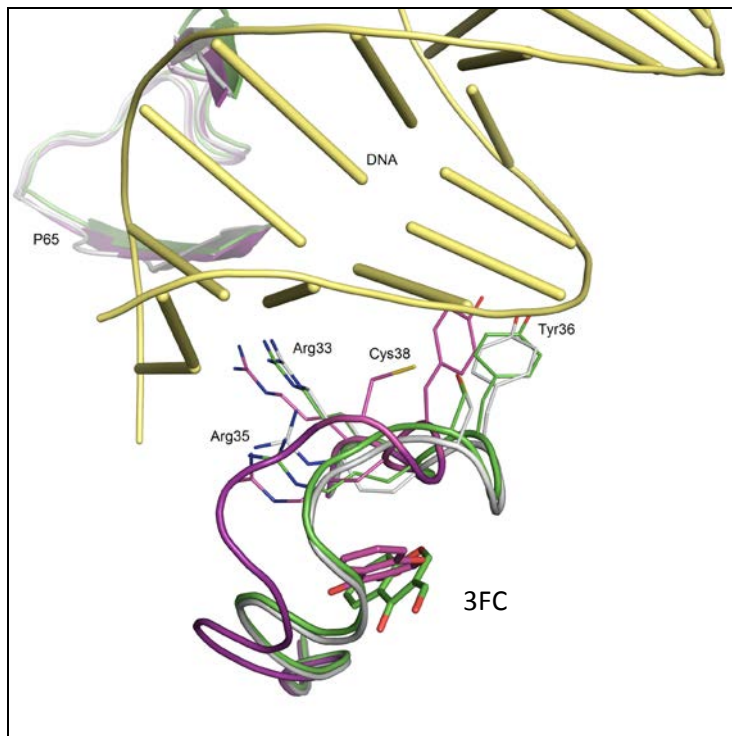


Figure 19: Possible binding mode of 3-FC with p65 DNA-binding region. The original crystal structure (Protein Data Bank code 1VKX) is superimposed with the modeled structures of the wild-type and C38S mutant proteins upon 3-FC binding. Gray, original crystal structure; purple, modeled wild-type structure upon 3-FC binding; green, modeled C38S mutant upon 3-FC binding. The final docked pose

for 3-FC is depicted in purple and green sticks for the wild-type and C38S mutant structures, respectively. DNA is represented as yellow tubes.

Extensive biological assay studies suggest that 3-FC a: directly interacted with p65 subunit of NF- κ B and b: it likely suppressed TNF- α induced IKK activation. 3-FC targets IKK to suppress the TNF- α induced phosphorylation and degradation of I κ B α that was concomitant with the inhibition of nuclear translocation and phosphorylation of p65. 3-FC down regulated the expression of NF- κ B regulated gene products such as survivin, Bcl-xL, Bcl-2 and cIAP-1, all anti-apoptotic agents(160). Other inhibitory effects of NF- κ B activation due to inflammatory stimuli or by tumor promoters suggest a common mode of action for 3FC. Prior studies indicate the critical role of IKK complex in NF- κ B activation suggesting that inhibition of IKK activity is the common step for the inhibition of NF- κ B by 3-FC. Further studies lead to the hypothesis that the reversal of effects of the 3-FC by a reducing agent is likely due to the modification of a Cysteine residue in p65. Cys³⁸ was identified in the p65 subunit of NF- κ B that is likely to play an important role in DNA binding. A mutagenesis experiment involving Cys³⁸Ser mutation was also designed to investigate this hypothesis. This mutation failed to inhibit the DNA binding ability of p65 and thus validating the potential role of Cys³⁸ in the p65-DNA binding process.

Biological studies indicated that as 3FC inhibits the binding of reconstituted p65 to the DNA *in vitro*, p65 is the direct target of 3-FC. Furthermore, with the Cys³⁸Ser mutation in p65, 3-FC failed to inhibit the DNA binding ability of p65. From the molecular modeling studies I observed that 3-FC docked just below the L1 DNA-binding loop region consisting of residues Arg³³, Cys³⁸ and Arg⁴¹. Post ligand-guided optimization the L1 loop region showed a minor conformational change and moved upward as compared to the crystal structure. As a result the

distance between the Cys³⁸ and Cys¹²⁰ changed from 7.7Å in the crystal structure to 10.2Å in the model. This loop movement altered the *Chi*1 angle of Tyr³⁶ from 66.3 to -67.5. This change resulted in the positioning of the Tyr³⁶ phenol moiety towards the DNA binding region. The modeling seems to suggest that the conformational change presents a steric hindrance to the DNA thus precluding its binding to p65. On the contrary changed introduced by the Cys³⁸Ser mutation failed to elicit any significant conformational change thus possibly maintaining the DNA-p65 binding. The modeling study hypothesizes that loop movements and the corresponding side chain rearrangement may explain the biology behind of p65-DNA binding.

Conclusions: The modeling protocol involved here represents an initial step of developing an open-source ligand-guided binding site optimization. It still relies heavily on the PyRosetta suite and does not explicitly model side chain flexibility. Though incorporating restraints is a possibility within the current libraries, it was not done for this case study. I envision a greater review of options to either develop such an open source protocol from the ground up or optimize any other publicly available resources to incorporate such a method.

Chapter 6: Research Projects Summary and Future Directions.

Identifying potent therapeutics to manage and treat physiological disorders will remain an active area of research far into the future. Experimental methods such as high throughput screening have played and will continue to play an important role in the discovery phase of a drug development cycle. Despite rapid advances in technology and improvement in process efficiency it is likely that HTS may never fulfill the hypothesized success projections. The significant costs associated with HTS, doubts over the benefits associated with random screening of corporate compound collections, and recent spate of failures of candidate drug molecules underline the need to incorporate different methods in the drug discovery cycle. Computational methods have shown promise in improving the productivity and decreasing costs associated with HTS programs. *In silico* methods, both structure and ligand based have proved their success in identifying enriched and smaller subset of compounds for exhaustive experimental processes. Increasing availability of structural and experimental biological data, development of better algorithms and access to cheaper yet faster computational hardware resources have made *in silico* methods a complement of the experimental HTS. Several examples suggest that a combination of experimental and computational methods may help offset their inherent drawbacks and can be applied to navigate the complex problems associated with drug discovery. However, akin to HTS methods, *in silico* methods are not without any drawbacks e.g. unavailability of high quality three-dimensional protein structures of biologically relevant proteins, and inadequate consideration of protein flexibility for computational structure-based drug discovery(1). This work attempts to present solutions for the above-mentioned problems by enhancing, benchmarking and applying modeling methods to relevant drug discovery problems.

Despite rapid growth in protein structure determination projects e.g. structural genomics, the number of novel proteins deposited in the Protein Data Bank has remained constant for several years. It is also likely that the massive sequence to structure gap of the human genome will not be overcome in the near future. In this scenario homology modeling plays an important role in the computational structure-based drug discovery and design. Homology modeling provides a cost-effective method to generate tertiary structures of proteins that may be applied to drug discovery problems. The literature is annotated with successful examples of homology modeling thus suggesting continual development and applications in the future. However, models by definition are an abstraction and will most likely contain errors. Of particular importance are those errors that occur in the pharmacologically relevant sites of a protein i.e. compound / ligand binding site which may negate the applicability of a protein model. It is thus imperative to develop protocols to minimize such modeling errors(27).

In this thesis I have primarily concentrated on homology modeling of Class A GPCRs. This 7 transmembrane serpentine family of proteins is the target of ~30 – 40% of drug discovery projects. Until 2008 there was only 1 high resolution crystal structure of Class A GPCR amongst a family of ~1000. Though novel technological breakthroughs during the time of this work (2007-2012) resulted in a spurt of several new crystal structures they still account for ~2% of all Class A GPCRs. It is clear that modeling will help offset this discrepancy associated with three-dimensional structure availability of this important therapeutic protein class. In this thesis, I developed optimized models for Cannabinoid Receptor 2, β_2 and Adenosine 2 A receptors and applied them to rationalize compound activities and in several case studies to highlight their improved performances in structure-based drug design methods such as molecular high-throughput docking(166).

As shown in the enhancement and benchmarking studies, the development of optimized Class A GPCR models using a single template and including receptor flexibility has a positive impact in compound pose prediction, high-throughput docking and compound selectivity studies. However like all protein ensemble docking studies, identifying the most relevant set of protein structures remains a challenging task. Our study indicates the shortcomings of energy functions to consistently rank models per their quality.

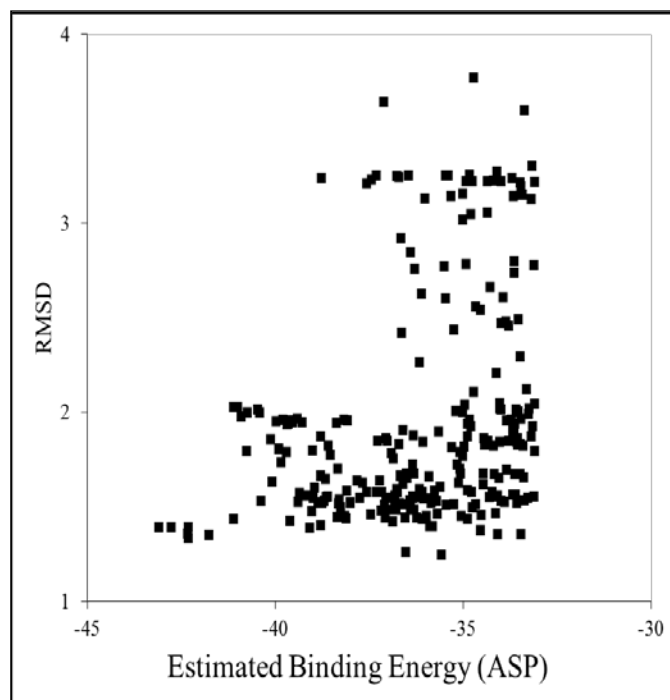


Figure 20: This plot represents the RMSD vs estimated binding energy of Case 1 studies.

From Figure 20 it is clear that models with good (i.e. lower RMSD) of co-crystallized ligands have been assigned poor energies. This shortcoming is likely due to force field limitations offered in the computational software used to develop the protocol. Further studies may be performed to investigate the effect of different force fields on the modeling protocol.

Second, as shown in the studies modeling of the extracellular loop 2 (ECL2) is an important factor in determining the overall accuracy of Class A GPCR models. Homology modeling methods are limited in their ability to determine completely novel folds. Using *de novo* based methods to generate multiple loop conformations and then incorporating them into homology

based models is one way of potentially addressing this situation. If conclusive experimental information is available it may be included as restraints.

Thirdly incorrect side chain conformations in the crude model may prevent the generation of optimal ligand conformations and thus binding pocket residues. Refining the crude model via molecular dynamics prior to ligand-based optimization may correct side chain conformation errors and also result in a larger binding pocket for initial ligand poses. This needs to be investigated in detail as GPCR's are membrane bound and likely need long simulation time for meaningful results. At least the benchmarked method offers clues to optimize the binding site of GPCRs using a fully flexible ligand and receptor approach.

In subsequent applications I developed protein models for Cannabinoid Receptor 2, a member of the Class A –GPCR family. CB2 receptors are investigated for their therapeutic use in non-psychotic pain management. The CB2 models developed here, to the best of our knowledge, are the first models developed using β_2 adrenergic receptor and a multi-template approach for an agonist state. Both the models were validated by their ability to rationalize the structure-activity relationship data of two series of compounds, the isatin series of inverse agonists and the benzofuran series of agonist compounds. However, further validation studies are required. Validation of these models may be performed using the metrics of the benchmarking studies i.e. ability to discriminate CB2 specific binders from a decoy library and by their ability to identify inhibitors in a virtual screening study(101, 102). A small-scale study involving CB2 agonist model is planned where molecular docking will be applied to rationally identify a small set of compounds for further experimental testing.

Data, Information and Knowledge.

In line with the training at the School of Biomedical Informatics the homology modeling process can be explained in terms of data, information and knowledge. Protein sequence may be considered as data. A protein sequence by itself carries little or no information on its three-dimensional structure or about how / why a bioactive compound would interact with this given protein. The next step in the modeling process is identifying a suitable template by means of sequence alignment. Once one or several homologous protein sequences are identified and sequences aligned a researcher now has information that can be used to build three-dimensional models. Here, the data now has meaning e.g. residues that may form well-defined tertiary structures are identified. In the cases explained in this thesis, information from experimental studies is incorporated in the modeling process to improve upon the accuracy of such crude models. This validated experimental information e.g. compound “x” interacts with protein “y” via residue “z” is included as e.g. restraints, to model the protein active site. Finally, the information is validated, and justified using means such as performance in high-throughput docking, rationalization of compound activities to provide knowledge about protein structure, functionality and its interaction with relevant biomolecules.

Sharangdhar S. Phatak, M.S.

References:

1. Phatak, S. S., C. C. Stephan, and C. N. Cavasotto. 2009. High-throughput and *in silico* screenings in drug discovery. *Expert Opin. Drug Disc.* 4:947-959.
2. Overington, J. P., B. Al-Lazikani, and A. L. Hopkins. 2006. How many drug targets are there? *Nat Rev Drug Discov* 5:993-996.
3. 2004. Finishing the euchromatic sequence of the human genome. *Nature* 431:931-945.
4. <http://www.alzdiscovery.org/index.php/alzheimers-disease/hope-through-drugs/drug-discovery-process/>
5. Tamimi, N. A., and P. Ellis. 2009. Drug development: from concept to marketing! *Nephron Clin Pract* 113:c125-131.
6. DiMasi, J. A., R. W. Hansen, and H. G. Grabowski. 2003. The price of innovation: new estimates of drug development costs. *J Health Econ* 22:151-185.
7. Lang, T. Adaptive trial design: could we use this approach to improve clinical trials in the field of global health? *Am J Trop Med Hyg* 85:967-970.
8. Drews, J. 2000. Drug discovery: a historical perspective. *Science* 287:1960-1964.
9. Macarron, R. 2006. Critical review of the role of HTS in drug discovery. *Drug Discov Today* 11:277-279.
10. Mayr, L. M., and P. Fuerst. 2008. The future of high-throughput screening. *J Biomol Screen* 13:443-448.
11. Entzeroth, M. 2003. Emerging trends in high-throughput screening. *Curr Opin Pharmacol* 3:522-529.
12. Keseru, G. M., and G. M. Makara. 2006. Hit discovery and hit-to-lead approaches. *Drug Discov Today* 11:741-748.

13. Keseru, G. M., and G. M. Makara. 2009. The influence of lead discovery strategies on the properties of drug candidates. *Nat Rev Drug Discov* 8:203-212.
14. Mahapatra, A. 2008. Screening for success. *ACS Chem Biol* 3:449-450.
15. Inglese, J., R. L. Johnson, A. Simeonov, M. Xia, W. Zheng, C. P. Austin, and D. S. Auld. 2007. High-throughput screening assays for the identification of chemical probes. *Nat Chem Biol* 3:466-479.
16. Inglese, J., C. E. Shamu, and R. K. Guy. 2007. Reporting data from high-throughput screening of small-molecule libraries. *Nat Chem Biol* 3:438-441.
17. Ha, T. 2001. Single-molecule fluorescence resonance energy transfer. *Methods* 25:78-86.
18. Degorce, F., A. Card, S. Soh, E. Trinquet, G. P. Knapik, and B. Xie. 2009. HTRF: A technology tailored for drug discovery - a review of theoretical aspects and recent applications. *Curr Chem Genomics* 3:22-32.
19. Shelat, A. A., and R. K. Guy. 2007. Scaffold composition and biological relevance of screening libraries. *Nat Chem Biol* 3:442-446.
20. Cavasotto, C. N., and S. S. Phatak. Docking methods for structure-based library design. *Methods Mol Biol* 685:155-174.
21. Fox, S., S. Farr-Jones, L. Sopchak, A. Boggs, H. W. Nicely, and R. Khoury. 2005. High Throughput Screening: New Users, More Cell-Based Assays, and a Host of New Tools. *HighTech Business Decisions*, Moraga, CA.
22. Fox, S., S. Farr-Jones, L. Sopchak, A. Boggs, H. W. Nicely, R. Khoury, and M. Biros. 2006. High-throughput screening: update on practices and success. *J Biomol Screen* 11:864-869.

23. Bajorath, J. 2002. Integration of virtual and high-throughput screening. *Nat Rev Drug Discov* 1:882-894.
24. Lundqvist, T. 2005. The devil is still in the details--driving early drug discovery forward with biophysical experimental methods. *Curr Opin Drug Discov Devel* 8:513-519.
25. Schneider, G., and U. Fechner. 2005. Computer-based de novo design of drug-like molecules. *Nat Rev Drug Discov* 4:649-663.
26. Bleicher, K. H., H. J. Bohm, K. Muller, and A. I. Alanine. 2003. Hit and lead generation: beyond high-throughput screening. *Nat Rev Drug Discov* 2:369-378.
27. Cavasotto, C. N., and S. S. Phatak. 2009. Homology modeling in drug discovery: current trends and applications. *Drug Discov Today* 14:676-683.
28. Guha, R. 2008. On the interpretation and interpretability of quantitative structure-activity relationship models. *J Comput Aided Mol Des* 22:857-871.
29. Vedani, A., and M. Smiesko. 2009. In silico toxicology in drug discovery - concepts based on three-dimensional models. *Altern Lab Anim* 37:477-496.
30. Stoermer, M. J. 2006. Current status of virtual screening as analysed by target class. *Med Chem* 2:89-112.
31. Sharma, H., X. Cheng, and J. K. Buolamwini. Homology Model-Guided 3D-QSAR Studies of HIV-1 Integrase Inhibitors. *J Chem Inf Model* 52:515-544.
32. Edink, E., A. Akdemir, C. Jansen, R. Elk, O. Zuiderveld, F. J. de Kanter, J. E. van Muijlwijk-Koezen, A. B. Smit, R. Leurs, and I. J. de Esch. Structure-based design, synthesis and structure-activity relationships of dibenzosuberyl- and benzoate-

- substituted tropines as ligands for acetylcholine-binding protein. *Bioorg Med Chem Lett* 22:1448-1454.
33. Nguyen, H. N., H. Bregman, J. L. Buchanan, B. Du, E. Feric, L. Huang, X. Li, J. Ligutti, D. Liu, A. B. Malmberg, D. J. Matson, J. S. McDermott, V. F. Patel, B. Wilenkin, A. Zou, S. I. McDonough, and E. F. Dimauro. Discovery and optimization of aminopyrimidinones as potent and state-dependent Nav1.7 antagonists. *Bioorg Med Chem Lett* 22:1055-1060.
34. Poulsen, A., M. Williams, H. M. Nagaraj, A. D. William, H. Wang, C. K. Soh, Z. C. Xiong, and B. Dymock. Structure-based optimization of morpholino-triazines as PI3K and mTOR inhibitors. *Bioorg Med Chem Lett* 22:1009-1013.
35. Reader, J. C., T. P. Matthews, S. Klair, K. M. Cheung, J. Scanlon, N. Proisy, G. Addison, J. Ellard, N. Piton, S. Taylor, M. Cherry, M. Fisher, K. Boxall, S. Burns, M. I. Walton, I. M. Westwood, A. Hayes, P. Eve, M. Valenti, A. de Haven Brandon, G. Box, R. L. van Montfort, D. H. Williams, G. W. Aherne, F. I. Raynaud, S. A. Eccles, M. D. Garrett, and I. Collins. Structure-guided evolution of potent and selective CHK1 inhibitors through scaffold morphing. *J Med Chem* 54:8328-8342.
36. Bacilieri, M., and S. Moro. 2006. Ligand-based drug design methodologies in drug discovery process: an overview. *Curr Drug Discov Technol* 3:155-165.
37. Guido, R. V., G. Oliva, and A. D. Andricopulo. 2008. Virtual screening and its integration with modern drug design technologies. *Curr Med Chem* 15:37-46.
38. Hattori, M., N. Tanaka, M. Kanehisa, and S. Goto. SIMCOMP/SUBCOMP: chemical structure search servers for network analyses. *Nucleic Acids Res* 38:W652-656.
39. Hansch, C. 1976. On the structure of medicinal chemistry. *J Med Chem* 19:1-6.

40. Klebe, G. 2006. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov Today* 11:580-594.
41. Leach, A. R., B. K. Shoichet, and C. E. Peishoff. 2006. Prediction of protein-ligand interactions. Docking and scoring: successes and gaps. *J Med Chem* 49:5851-5855.
42. Lundstrom, K. 2007. Structural genomics and drug discovery. *J Cell Mol Med* 11:224-238.
43. B-Rao, C., J. Subramanian, and S. D. Sharma. 2009. Managing protein flexibility in docking and its applications. *Drug Discov Today* 14:394-400.
44. Villar, H. O., J. Yan, and M. R. Hansen. 2004. Using NMR for ligand discovery and optimization. *Curr Opin Chem Biol* 8:387-391.
45. Mooij, W. T., M. J. Hartshorn, I. J. Tickle, A. J. Sharff, M. L. Verdonk, and H. Jhoti. 2006. Automated protein-ligand crystallography for structure-based drug design. *ChemMedChem* 1:827-838.
46. Manjasetty, B. A., A. P. Turnbull, S. Panjikar, K. Bussow, and M. R. Chance. 2008. Automated technologies and novel techniques to accelerate protein crystallography for structural genomics. *Proteomics* 8:612-625.
47. Levitt, M. 2007. Growth of novel protein structural data. *Proc Natl Acad Sci U S A* 104:3183-3188.
48. Wu, S., J. Skolnick, and Y. Zhang. 2007. Ab initio modeling of small proteins by iterative TASSER simulations. *BMC Biol* 5:17.
49. Phatak, S. S., H. T. Tran, and S. Zhang. 2011. Novel computational biology methods and their applications to drug discovery. *Frontiers in Biology* 6.

50. Hillisch, A., L. F. Pineda, and R. Hilgenfeld. 2004. Utility of homology models in the drug discovery process. *Drug Discov Today* 9:659-669.
51. Sali, A., and T. L. Blundell. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234:779-815.
52. Shen, M. Y., and A. Sali. 2006. Statistical potential for assessment and prediction of protein structures. *Protein Sci* 15:2507-2524.
53. Song, L., C. Kalyanaraman, A. A. Fedorov, E. V. Fedorov, M. E. Glasner, S. Brown, H. J. Imker, P. C. Babbitt, S. C. Almo, M. P. Jacobson, and J. A. Gerlt. 2007. Prediction and assignment of function for a divergent N-succinyl amino acid racemase. *Nat Chem Biol* 3:486-491.
54. Guimaraes, A. J., A. J. Hamilton, M. G. H. L. de, J. D. Nosanchuk, and R. M. Zancoppe-Oliveira. 2008. Biological function and molecular mapping of M antigen in yeast phase of *Histoplasma capsulatum*. *PLoS ONE* 3:e3449.
55. Sun, W., C. Gerth, A. Maeda, D. T. Lodowski, L. Van Der Kraak, D. A. Saperstein, E. Heon, and K. Palczewski. 2007. Novel RDH12 mutations associated with Leber congenital amaurosis and cone-rod dystrophy: biochemical and clinical evaluations. *Vision Res* 47:2055-2066.
56. Proell, M., S. J. Riedl, J. H. Fritz, A. M. Rojas, and R. Schwarzenbacher. 2008. The Nod-like receptor (NLR) family: a tale of similarities and differences. *PLoS ONE* 3:e2119.
57. Cavasotto, C. N., A. J. Orry, N. J. Murgolo, M. F. Czarniecki, S. A. Kocsi, B. E. Hawes, K. A. O'Neill, H. Hine, M. S. Burton, J. H. Voigt, R. A. Abagyan, M. L. Bayne, and F. J. Monsma, Jr. 2008. Discovery of novel chemotypes to a G-protein-coupled receptor

- through ligand-steered homology modeling and structure-based virtual screening. *J Med Chem* 51:581-588.
58. Park, H., K. Y. Hwang, K. H. Oh, Y. H. Kim, J. Y. Lee, and K. Kim. 2008. Discovery of novel alpha-glucosidase inhibitors based on the virtual screening with the homology-modeled protein structure. *Bioorg Med Chem* 16:284-292.
59. Park, H., Y. J. Bahn, S. K. Jung, D. G. Jeong, S. H. Lee, I. Seo, T. S. Yoon, S. J. Kim, and S. E. Ryu. 2008. Discovery of novel Cdc25 phosphatase inhibitors with micromolar activity based on the structure-based virtual screening. *J Med Chem* 51:5533-5541.
60. Cozza, G., A. Gianoncelli, M. Montopoli, L. Caparrotta, A. Venerando, F. Meggio, L. A. Pinna, G. Zagotto, and S. Moro. 2008. Identification of novel protein kinase CK1 delta (CK1delta) inhibitors through structure-based virtual screening. *Bioorg Med Chem Lett* 18:5672-5675.
61. Luther, K. B., H. Schindelin, and R. S. Haltiwanger. 2008. Structural and mechanistic insights into lunatic fringe from a kinetic analysis of enzyme mutants. *J Biol Chem*.
62. Gagnidze, K., Sachchidanand, R. Rozenfeld, M. Mezei, M. M. Zhou, and L. A. Devi. 2008. Homology modeling and site-directed mutagenesis to identify selective inhibitors of endothelin-converting enzyme-2. *J Med Chem* 51:3378-3387.
63. Michielan, L., M. Bacilieri, A. Schiesaro, C. Bolcato, G. Pastorin, G. Spalluto, B. Cacciari, K. N. Klotz, C. Kaseda, and S. Moro. 2008. Linear and nonlinear 3D-QSAR approaches in tandem with ligand-based homology modeling as a computational strategy to depict the pyrazolo-triazolo-pyrimidine antagonists binding site of the human adenosine A2A receptor. *J Chem Inf Model* 48:350-363.

64. Li, M., H. Fang, L. Du, L. Xia, and B. Wang. 2008. Computational studies of the binding site of alpha1A-adrenoceptor antagonists. *J Mol Model* 14:957-966.
65. Moro, S., F. Deflorian, M. Bacilieri, and G. Spalluto. 2006. Ligand-based homology modeling as attractive tool to inspect GPCR structural plasticity. *Curr Pharm Des* 12:2175-2185.
66. Noronha, G., K. Barrett, A. Boccia, T. Brodhag, J. Cao, C. P. Chow, E. Dneprovskaja, J. Doukas, R. Fine, X. Gong, C. Gritzen, H. Gu, E. Hanna, J. D. Hood, S. Hu, X. Kang, J. Key, B. Klebansky, A. Kousba, G. Li, D. Lohse, C. C. Mak, A. McPherson, M. S. Palanki, V. P. Pathak, J. Renick, F. Shi, R. Soll, U. Splittgerber, S. Stoughton, S. Tang, S. Yee, B. Zeng, N. Zhao, and H. Zhu. 2007. Discovery of [7-(2,6-dichlorophenyl)-5-methylbenzo [1,2,4]triazin-3-yl]-[4-(2-pyrrolidin-1-ylethoxy)phenyl]amine--a potent, orally active Src kinase inhibitor with anti-tumor activity in preclinical assays. *Bioorg Med Chem Lett* 17:602-608.
67. Cywin, C. L., G. Dahmann, A. S. Prokopowicz, 3rd, E. R. Young, R. L. Magolda, M. G. Cardozo, D. A. Cogan, D. Disalvo, J. D. Ginn, M. A. Kashem, J. P. Wolak, C. A. Homon, T. M. Farrell, H. Grbic, H. Hu, P. V. Kaplita, L. H. Liu, D. M. Spero, D. D. Jeanfavre, K. M. O'Shea, D. M. White, J. R. Woska, Jr., and M. L. Brown. 2007. Discovery of potent and selective PKC-theta inhibitors. *Bioorg Med Chem Lett* 17:225-230.
68. Deng, Q., J. L. Frie, D. M. Marley, R. T. Beresis, N. Ren, T. Q. Cai, A. K. Taggart, K. Cheng, E. Carballo-Jane, J. Wang, X. Tong, M. G. Waters, J. R. Tata, and S. L. Colletti. 2008. Molecular modeling aided design of nicotinic acid receptor GPR109A agonists. *Bioorg Med Chem Lett* 18:4963-4967.

69. Burley, S. K., A. Joachimiak, G. T. Montelione, and I. A. Wilson. 2008. Contributions to the NIH-NIGMS Protein Structure Initiative from the PSI Production Centers. *Structure* 16:5-11.
70. Ginalski, K. 2006. Comparative modeling for protein structure prediction. *Curr Opin Struct Biol* 16:172-177.
71. Verkhivker, G. M., D. Bouzida, D. K. Gehlhaar, P. A. Rejto, L. Schaffer, S. Arthurs, A. B. Colson, S. T. Freer, V. Larson, B. A. Luty, T. Marrone, and P. W. Rose. 2001. Monte Carlo simulations of HIV-1 protease binding dynamics and thermodynamics with ensembles of protein conformations: incorporating protein flexibility in deciphering mechanisms of molecular recognition. In *Theoretical Biochemistry- Processes and Properties of Biological Systems*. L. A. Eriksson, editor. Elsevier Science BV. 289-340.
72. Yuriev, E., M. Agostino, and P. A. Ramsland. Challenges and advances in computational docking: 2009 in review. *J Mol Recognit* 24:149-164.
73. Totrov, M., and R. Abagyan. 2008. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr Opin Struct Biol* 18:178-184.
74. Jiang, F., and S. H. Kim. 1991. "Soft docking": matching of molecular surface cubes. *J Mol Biol* 219:79-102.
75. Claussen, H., C. Buning, M. Rarey, and T. Lengauer. 2001. FlexE: efficient molecular docking considering protein structure variations. *J Mol Biol* 308:377-395.
76. Alberts, I. L., N. P. Todorov, and P. M. Dean. 2005. Receptor flexibility in de novo ligand design and docking. *J Med Chem* 48:6585-6596.

77. Carlson, H. A. 2002. Protein flexibility and drug design: how to hit a moving target. *Curr Opin Chem Biol* 6:447-452.
78. Leach, A. R. 1994. Ligand docking to proteins with discrete side-chain flexibility. *J Mol Biol* 235:345-356.
79. Cozzini, P., G. E. Kellogg, F. Spyrakis, D. J. Abraham, G. Costantino, A. Emerson, F. Fanelli, H. Gohlke, L. A. Kuhn, G. M. Morris, M. Orozco, T. A. Pertinhez, M. Rizzi, and C. A. Sotriffer. 2008. Target flexibility: an emerging consideration in drug discovery and design. *J Med Chem* 51:6237-6255.
80. Amaro, R. E., R. Baron, and J. A. McCammon. 2008. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J Comput Aided Mol Des* 22:693-705.
81. Marco, E., and F. Gago. 2007. Overcoming the inadequacies or limitations of experimental structures as drug targets by using computational modeling tools and molecular dynamics simulations. *ChemMedChem* 2:1388-1401.
82. Alonso, H., A. A. Bliznyuk, and J. E. Gready. 2006. Combining docking and molecular dynamic simulations in drug design. *Med Res Rev* 26:531-568.
83. Bahar, I., T. R. Lezon, A. Bakan, and I. H. Shrivastava. Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. *Chem Rev* 110:1463-1497.
84. Cavasotto, C. N., J. A. Kovacs, and R. A. Abagyan. 2005. Representing receptor flexibility in ligand docking through relevant normal modes. *J Am Chem Soc* 127:9632-9640.

85. Ma, J. 2005. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure* 13:373-380.
86. Leach, A. R. 2001. *Molecular Modelling: Principles and Applications*. Prentice Hall.
87. Millar, R. P., and C. L. Newton. The year in G protein-coupled receptor research. *Mol Endocrinol* 24:261-274.
88. Rosenbaum, D. M., S. G. Rasmussen, and B. K. Kobilka. 2009. The structure and function of G-protein-coupled receptors. *Nature* 459:356-363.
89. Jalink, K., and W. H. Moolenaar. G protein-coupled receptors: the inside story. *Bioessays* 32:13-16.
90. Lundstrom, K. 2009. An overview on GPCRs and drug discovery: structure-based drug design and structural biology on GPCRs. *Methods Mol Biol* 552:51-66.
91. Wolf, S., M. Bockmann, U. Howeler, J. Schlitter, and K. Gerwert. 2008. Simulations of a G protein-coupled receptor homology model predict dynamic features and a ligand binding site. *FEBS Lett* 582:3335-3342.
92. Schlyer, S., and R. Horuk. 2006. I want a new drug: G-protein-coupled receptors in drug development. *Drug Discov Today* 11:481-493.
93. Alkhalfioui, F., T. Magnin, and R. Wagner. 2009. From purified GPCRs to drug discovery: the promise of protein-based methodologies. *Curr Opin Pharmacol* 9:629-635.
94. Cherezov, V., D. M. Rosenbaum, M. A. Hanson, S. G. Rasmussen, F. S. Thian, T. S. Kobilka, H. J. Choi, P. Kuhn, W. I. Weis, B. K. Kobilka, and R. C. Stevens. 2007. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* 318:1258-1265.

95. Warne, T., M. J. Serrano-Vega, J. G. Baker, R. Moukhametzianov, P. C. Edwards, R. Henderson, A. G. Leslie, C. G. Tate, and G. F. Schertler. 2008. Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature* 454:486-491.
96. Jaakola, V. P., M. T. Griffith, M. A. Hanson, V. Cherezov, E. Y. Chien, J. R. Lane, A. P. Ijzerman, and R. C. Stevens. 2008. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* 322:1211-1217.
97. Park, J. H., P. Scheerer, K. P. Hofmann, H. W. Choe, and O. P. Ernst. 2008. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* 454:183-187.
98. Wu, B., E. Y. Chien, C. D. Mol, G. Fenalti, W. Liu, V. Katritch, R. Abagyan, A. Brooun, P. Wells, F. C. Bi, D. J. Hamel, P. Kuhn, T. M. Handel, V. Cherezov, and R. C. Stevens. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* 330:1066-1071.
99. Chien, E. Y., W. Liu, Q. Zhao, V. Katritch, G. W. Han, M. A. Hanson, L. Shi, A. H. Newman, J. A. Javitch, V. Cherezov, and R. C. Stevens. Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* 330:1091-1095.
100. Shimamura, T., M. Shiroishi, S. Weyand, H. Tsujimoto, G. Winter, V. Katritch, R. Abagyan, V. Cherezov, W. Liu, G. W. Han, T. Kobayashi, R. C. Stevens, and S. Iwata. Structure of the human histamine H1 receptor complex with doxepin. *Nature* 475:65-70.
101. Diaz, P., S. S. Phatak, J. Xu, F. Astruc-Diaz, C. N. Cavasotto, and M. Naguib. 2009. 6-Methoxy-N-alkyl isatin acylhydrazone derivatives as a novel series of potent selective cannabinoid receptor 2 inverse agonists: design, synthesis, and binding mode prediction. *J Med Chem* 52:433-444.

102. Diaz, P., S. S. Phatak, J. Xu, F. R. Fronczek, F. Astruc-Diaz, C. M. Thompson, C. N. Cavasotto, and M. Naguib. 2009. 2,3-Dihydro-1-Benzofuran Derivatives as a Series of Potent Selective Cannabinoid Receptor 2 Agonists: Design, Synthesis, and Binding Mode Prediction through Ligand-Steered Modeling. *ChemMedChem*.
103. Mirzadegan, T., G. Benko, S. Filipek, and K. Palczewski. 2003. Sequence analysis of G-Protein-coupled receptors: Similarities to rhodopsin. *Biochemistry* 42:2759-2767.
104. Ballesteros, J., and H. Weinstein. 1995. Integrated methods for the construction of three-dimensional models of structure-function relations in G protein-coupled receptors. *Methods Neurosci.* 25:366-428.
105. Evers, A., H. Gohlke, and G. Klebe. 2003. Ligand-supported homology modelling of protein binding-sites using knowledge-based potentials. *J Mol Biol* 334:327-345.
106. Evers, A., and G. Klebe. 2004. Ligand-supported homology modeling of g-protein-coupled receptor sites: models sufficient for successful virtual screening. *Angew Chem Int Ed Engl* 43:248-251.
107. Radestock, S., T. Weil, and S. Renner. 2008. Homology model-based virtual screening for GPCR ligands using docking and target-biased scoring. *J Chem Inf Model* 48:1104-1117.
108. Costanzi, S. 2008. On the applicability of GPCR homology models to computer-aided drug discovery: a comparison between in silico and crystal structures of the beta2-adrenergic receptor. *J Med Chem* 51:2907-2914.
109. Kimura, S. R., A. J. Tebben, and D. R. Langley. 2008. Expanding GPCR homology model binding sites via a balloon potential: A molecular dynamics refinement approach. *Proteins* 71:1919-1929.

110. Abagyan, R., and P. Argos. 1992. Optimal protocol and trajectory visualization for conformational searches of peptides and proteins. *J. Mol. Biol.* 225:519-532.
111. Caflich, A., A. Miranker, and M. Karplus. 1993. Multiple copy simultaneous search and construction of ligands in binding sites: application to inhibitors of HIV-1 aspartic proteinase. *J Med Chem* 36:2142-2167.
112. Nemethy, G., K. D. Gibson, K. A. Palmer, C. N. Yoon, M. G. Paterlini, A. Zagari, S. Rumsey, and H. A. Scheraga. 1992. Energy Parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. *J. Phys. Chem* 96:6472-6484.
113. Halgren, T. 1995. Merck molecular force field I-V. *J. Comput. Chem* 17:490-641.
114. ICM Version 3.6 Molsoft LLC: La Jolla CA 2009.
115. Wesson, L., and D. Eisenberg. 1992. Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Sci* 1:227-235.
116. Li, Z., and H. A. Scheraga. 1987. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc Natl Acad Sci U S A* 84:6611-6615.
117. Bisson, W. H., R. Abagyan, and C. N. Cavasotto. 2008. Molecular basis of agonicity and antagonicity in the androgen receptor studied by molecular dynamics simulations. *J Mol Graph Model* 27:452-458.
118. Monti, M. C., A. Casapullo, C. N. Cavasotto, A. Tosco, F. Dal Piaz, A. Ziemys, L. Margarucci, and R. Riccio. 2009. The binding mode of petrosaspongiolide M to the human group IIA phospholipase A(2): exploring the role of covalent and noncovalent interactions in the inhibition process. *Chemistry* 15:1155-1163.

119. Monti, M. C., A. Casapullo, C. N. Cavasotto, A. Napolitano, and R. Riccio. 2007. Scalaradial, a dialdehyde-containing marine metabolite that causes an unexpected noncovalent PLA2 Inactivation. *Chembiochem* 8:1585-1591.
120. Cavasotto, C. N., and R. A. Abagyan. 2004. Protein flexibility in ligand docking and virtual screening to protein kinases. *J Mol Biol* 337:209-225.
121. Kolb, P., D. M. Rosenbaum, J. J. Irwin, J. J. Fung, B. K. Kobilka, and B. K. Shoichet. 2009. Structure-based discovery of beta2-adrenergic receptor ligands. *Proc Natl Acad Sci U S A* 106:6843-6848.
122. McRobb, F. M., B. Capuano, I. T. Crosby, D. K. Chalmers, and E. Yuriev. Homology modeling and docking evaluation of aminergic G protein-coupled receptors. *J Chem Inf Model* 50:626-637.
123. Mobarec, J. C., R. Sanchez, and M. Filizola. 2009. Modern homology modeling of G-protein coupled receptors: which structural template to use? *J Med Chem* 52:5207-5216.
124. Ivanov, A. A., D. Barak, and K. A. Jacobson. 2009. Evaluation of homology modeling of G-protein-coupled receptors in light of the A(2A) adenosine receptor crystallographic structure. *J Med Chem* 52:3284-3292.
125. Michino, M., E. Abola, C. L. Brooks, 3rd, J. S. Dixon, J. Moulton, and R. C. Stevens. 2009. Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008. *Nat Rev Drug Discov* 8:455-463.
126. Irwin, J. J., and B. K. Shoichet. 2005. ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model* 45:177-182.

127. Yuzlenko, O., and K. Kiec-Kononowicz. 2009. Molecular modeling of A1 and A2A adenosine receptors: comparison of rhodopsin- and beta2-adrenergic-based homology models through the docking studies. *J Comput Chem* 30:14-32.
128. Goldfeld, D. A., K. Zhu, T. Beuming, and R. A. Friesner. Successful prediction of the intra- and extracellular loops of four G-protein-coupled receptors. *Proc Natl Acad Sci U S A* 108:8275-8280.
129. Chen, L., J. K. Morrow, H. T. Tran, S. S. Phatak, L. Du-Cuny, and S. Zhang. From Laptop to Benchtop to Bedside: Structure-based Drug Design on Protein Targets. *Curr Pharm Des*.
130. Zhang, S. In silico lead identification and Optimization for Drug Discovery. *Curr Pharm Des*.
131. Zhang, L., K. C. Tsai, L. Du, H. Fang, M. Li, and W. Xu. How to generate reliable and predictive CoMFA models. *Curr Med Chem* 18:923-930.
132. Marco, E. M., M. S. Garcia-Gutierrez, F. J. Bermudez-Silva, F. A. Moreira, F. Guimaraes, J. Manzanares, and M. P. Viveros. Endocannabinoid system and psychiatry: in search of a neurobiological basis for detrimental and potential therapeutic effects. *Front Behav Neurosci* 5:63.
133. Guindon, J., and A. G. Hohmann. 2009. The endocannabinoid system and pain. *CNS Neurol Disord Drug Targets* 8:403-421.
134. Lazary, J., G. Juhasz, L. Hunyady, and G. Bagdy. Personalized medicine can pave the way for the safe use of CB receptor antagonists. *Trends Pharmacol Sci* 32:270-280.
135. Munro, S., K. L. Thomas, and M. Abu-Shaar. 1993. Molecular characterization of a peripheral receptor for cannabinoids. *Nature* 365:61-65.

136. Facci, L., R. Dal Toso, S. Romanello, A. Buriani, S. D. Skaper, and A. Leon. 1995. Mast cells express a peripheral cannabinoid receptor with differential sensitivity to anandamide and palmitoylethanolamide. *Proc Natl Acad Sci U S A* 92:3376-3380.
137. Lunn, C. A., E. P. Reich, J. S. Fine, B. Lavey, J. A. Kozlowski, R. W. Hipkin, D. J. Lundell, and L. Bober. 2008. Biology and therapeutic potential of cannabinoid CB2 receptor inverse agonists. *Br J Pharmacol* 153:226-239.
138. Lunn, C. A., E. P. Reich, and L. Bober. 2006. Targeting the CB2 receptor for immune modulation. *Expert Opin Ther Targets* 10:653-663.
139. Yao, B. B., G. Hsieh, A. V. Daza, Y. Fan, G. K. Grayson, T. R. Garrison, O. El Kouhen, B. A. Hooker, M. Pai, E. J. Wensink, A. K. Salyers, P. Chandran, C. Z. Zhu, C. Zhong, K. Ryther, M. E. Gallagher, C. L. Chin, A. E. Tovcimak, V. P. Hradil, G. B. Fox, M. J. Dart, P. Honore, and M. D. Meyer. 2009. Characterization of a cannabinoid CB2 receptor-selective agonist, A-836339 [2,2,3,3-tetramethyl-cyclopropanecarboxylic acid [3-(2-methoxy-ethyl)-4,5-dimethyl-3H-thiazol-(2Z)-ylidene]-amide], using in vitro pharmacological assays, in vivo pain models, and pharmacological magnetic resonance imaging. *J Pharmacol Exp Ther* 328:141-151.
140. Naguib, M., P. Diaz, J. J. Xu, F. Astruc-Diaz, S. Craig, P. Vivas-Mejia, and D. L. Brown. 2008. MDA7: a novel selective agonist for CB2 receptors that prevents allodynia in rat neuropathic pain models. *Br J Pharmacol* 155:1104-1116.
141. Poso, A., and J. W. Huffman. 2008. Targeting the cannabinoid CB2 receptor: modelling and structural determinants of CB2 selective ligands. *Br J Pharmacol* 153:335-346.

142. Tuccinardi, T., P. L. Ferrarini, C. Manera, G. Ortore, G. Saccomanni, and A. Martinelli. 2006. Cannabinoid CB2/CB1 selectivity. Receptor modeling and automated docking analysis. *J Med Chem* 49:984-994.
143. Montero, C., N. E. Campillo, P. Goya, and J. A. Paez. 2005. Homology models of the cannabinoid CB1 and CB2 receptors. A docking analysis study. *Eur J Med Chem* 40:75-83.
144. Xie, X. Q., J. Z. Chen, and E. M. Billings. 2003. 3D structural model of the G-protein-coupled cannabinoid CB2 receptor. *Proteins* 53:307-319.
145. Gouldson, P., B. Calandra, P. Legoux, A. Kerneis, M. Rinaldi-Carmona, F. Barth, G. Le Fur, P. Ferrara, and D. Shire. 2000. Mutational analysis and molecular modelling of the antagonist SR 144528 binding site on the human cannabinoid CB(2) receptor. *Eur J Pharmacol* 401:17-25.
146. Valenzano, K. J., L. Tafesse, G. Lee, J. E. Harrison, J. M. Boulet, S. L. Gottshall, L. Mark, M. S. Pearson, W. Miller, S. Shan, L. Rabadi, Y. Rotshteyn, S. M. Chaffer, P. I. Turchin, D. A. Elsemore, M. Toth, L. Koetzner, and G. T. Whiteside. 2005. Pharmacological and pharmacokinetic characterization of the cannabinoid receptor 2 agonist, GW405833, utilizing rodent models of acute and chronic pain, anxiety, ataxia and catalepsy. *Neuropharmacology* 48:658-672.
147. Cheng, J. 2008. A multi-template combination algorithm for protein comparative modeling. *BMC Struct Biol* 8:18.
148. Hazai, E., and Z. Bikadi. 2008. Homology modeling of breast cancer resistance protein (ABCG2). *J Struct Biol* 162:63-74.

149. Scheerer, P., J. H. Park, P. W. Hildebrand, Y. J. Kim, N. Krauss, H. W. Choe, K. P. Hofmann, and O. P. Ernst. 2008. Crystal structure of opsin in its G-protein-interacting conformation. *Nature* 455:497-502.
150. Ahuja, S., V. Hornak, E. C. Yan, N. Syrett, J. A. Goncalves, A. Hirshfeld, M. Ziliox, T. P. Sakmar, M. Sheves, P. J. Reeves, S. O. Smith, and M. Eilers. 2009. Helix movement is coupled to displacement of the second extracellular loop in rhodopsin activation. *Nat Struct Mol Biol* 16:168-175.
151. Weis, W. I., and B. K. Kobilka. 2008. Structural insights into G-protein-coupled receptor activation. *Curr Opin Struct Biol* 18:734-740.
152. Kobilka, B., and G. F. Schertler. 2008. New G-protein-coupled receptor crystal structures: insights and limitations. *Trends Pharmacol Sci* 29:79-83.
153. Pei, Y., R. W. Mercier, J. K. Anday, G. A. Thakur, A. M. Zvonok, D. Hurst, P. H. Reggio, D. R. Janero, and A. Makriyannis. 2008. Ligand-binding architecture of human CB2 cannabinoid receptor: evidence for receptor subtype-specific binding motif and modeling GPCR activation. *Chem Biol* 15:1207-1219.
154. Ashton, J. C., J. L. Wright, J. M. McPartland, and J. D. Tyndall. 2008. Cannabinoid CB1 and CB2 receptor ligand specificity and the development of CB2-selective agonists. *Curr Med Chem* 15:1428-1443.
155. McAllister, S. D., Q. Tao, J. Barnett-Norris, K. Buehner, D. P. Hurst, F. Guarnieri, P. H. Reggio, K. W. Nowell Harmon, G. A. Cabral, and M. E. Abood. 2002. A critical role for a tyrosine residue in the cannabinoid receptors for ligand recognition. *Biochem Pharmacol* 63:2121-2136.

156. Chin, C. N., J. W. Murphy, J. W. Huffman, and D. A. Kendall. 1999. The third transmembrane helix of the cannabinoid receptor plays a role in the selectivity of aminoalkylindoles for CB2, peripheral cannabinoid receptor. *J Pharmacol Exp Ther* 291:837-844.
157. Song, Z. H., C. A. Slowey, D. P. Hurst, and P. H. Reggio. 1999. The difference between the CB(1) and CB(2) cannabinoid receptors at position 5.46 is crucial for the selectivity of WIN55212-2 for CB(2). *Mol Pharmacol* 56:834-840.
158. Vilar, S., G. Ferino, S. S. Phatak, B. Berk, C. N. Cavasotto, and S. Costanzi. Docking-based virtual screening for ligands of G protein-coupled receptors: not only crystal structures but also in silico models. *J Mol Graph Model* 29:614-623.
159. Chaudhury, S., S. Lyskov, and J. J. Gray. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* 26:689-691.
160. Yadav, V. R., S. Prasad, S. C. Gupta, B. Sung, S. S. Phatak, S. Zhang, and B. B. Aggarwal. 3-Formylchromone interacts with cysteine 38 in p65 protein and with cysteine 179 in I κ B kinase, leading to down-regulation of nuclear factor- κ B (NF- κ B)-regulated gene products and sensitization of tumor cells. *J Biol Chem* 287:245-256.
161. Leaver-Fay, A., M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper, A. Treuille, D. J. Mandell, F. Richter, Y. E. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popovic, J. J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J. J. Gray, B. Kuhlman, D. Baker, and P. Bradley. ROSETTA3: an

object-oriented software suite for the simulation and design of macromolecules.

Methods Enzymol 487:545-574.

162. Garcia-Pineres, A. J., V. Castro, G. Mora, T. J. Schmidt, E. Strunck, H. L. Pahl, and I. Merfort. 2001. Cysteine 38 in p65/NF-kappaB plays a crucial role in DNA binding inhibition by sesquiterpene lactones. *J Biol Chem* 276:39713-39720.
163. Haddad, J. J., and N. E. Abdel-Karim. NF-kappaB cellular and molecular regulatory mechanisms and pathways: therapeutic pattern or pseudoregulation? *Cell Immunol* 271:5-14.
164. Rothwarf, D. M., and M. Karin. 1999. The NF-kappa B activation pathway: a paradigm in information transfer from membrane to nucleus. *Sci STKE* 1999:RE1.
165. Huxford, T., A. Hoffmann, and G. Ghosh. Understanding the logic of IkappaB:NF-kappaB regulation in structural terms. *Curr Top Microbiol Immunol* 349:1-24.
166. Phatak, S. S., E. A. Gatica, and C. N. Cavasotto. Ligand-steered modeling and docking: A benchmarking study in class A G-protein-coupled receptors. *J Chem Inf Model* 50:2119-2128.

Vita

Sharangdhar S. Phatak is a native of Pune, India. In 2001, he graduated from The University of Pune with a Bachelor of Engineering degree in Electronics Engineering. He earned his Master of Science degree in Biomedical Engineering in 2004 from The Virginia Commonwealth University at Richmond, Virginia, where he focused on computational ligand-based drug discovery methods.

During his Master's program Sharang interned with Incyte Corporation, Delaware, in their computational chemistry program. He later accepted a full time position of an Applications Scientist with Chemical Computing Group AG and moved to Koln, Germany and later to Bengaluru, India.

In 2007, he came to The University of Texas Health Science Center to pursue a Doctor of Philosophy degree in Biomedical Informatics and the School of Biomedical Informatics. While in school he primarily worked in computational structure-based drug design methods to develop and apply high-quality protein models for drug discovery applications. He received three competitive academic awards at UT-Health and is a pre-doctoral fellow of The Cancer Prevention Research Institute of Texas. His research is documented in several peer reviewed publications.

Sharang's parents are Mrs. Vrushali S. Phatak and Mr. Shivanand P. Phatak. He has one sister, Ms. Shalmali S. Phatak. Sharang is married to Bhavana Bakshi (nee Manasvi Phatak) and is set to welcome their first baby, a daughter, in May 2012. At the time of this writing identifying her name remains an area of active research.