


8-2017

LBPI: A WEB INTERFACE FOR THE IDENTIFICATION OF ALLOSTERIC LIGAND BINDING SITES

Nabina Paudyal

Follow this and additional works at: https://digitalcommons.library.tmc.edu/utgsbs_dissertations

 Part of the [Biochemistry, Biophysics, and Structural Biology Commons](#)

Recommended Citation

Paudyal, Nabina, "LBPI: A WEB INTERFACE FOR THE IDENTIFICATION OF ALLOSTERIC LIGAND BINDING SITES" (2017). *The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access)*. 830.
https://digitalcommons.library.tmc.edu/utgsbs_dissertations/830

This Thesis (MS) is brought to you for free and open access by the The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences at DigitalCommons@TMC. It has been accepted for inclusion in The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact digitalcommons@library.tmc.edu.

LBPI: A WEB INTERFACE FOR THE IDENTIFICATION OF ALLOSTERIC
LIGAND BINDING SITES

by

NABINA PAUDYAL, MS

APPROVED

Alemayehu Gorfe, Ph.D.
Advisory Professor

Jeffrey Chang, PhD

Xiaodong Cheng, PhD

Vasanthi Jayaraman, PhD

Shuxing Zhang, PhD

Michael Zhu, PhD

APPROVED:

Dean, The University of Texas
MD Anderson Cancer Center UTHHealth Graduate School of Biomedical Sciences

LBPI: A WEB INTERFACE FOR THE IDENTIFICATION OF ALLOSTERIC
LIGAND BINDING SITES

A

THESIS

Presented to the Faculty of

The University of Texas

MD Anderson Cancer Center UTHealth

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

by

Nabina Paudyal, MS
Houston, Texas

August, 2017

ACKNOWLEDGEMENTS

First of all, I would like to thank you Dr. Gorfe for believing in me and providing me a wonderful opportunity to work in this project. I would also like to thank him for his consistent support and guidance throughout. His way of mentoring as an advisor was very different, encouraging, enlightening and also equally caring. He always allowed me to work independently and taught me how as a researcher one should be critical in thinking and working in the field of biomedical research. His mentorship helped me to have a smooth transition from physics to biomedical science. I am sincerely thankful for having such a great mentor in my academic career.

I would also like to thank my previous advisor Dr. Jutta Luettmmer Strathmann who has been a role model for me as a person, mentor and a scientist and has always helped me to introduce a new learning system totally different than my home country. I really appreciate her patience and her generosity for redefining my career path.

To my committee members Jeffrey Chang, Xiaodong Cheng, Vasanthi Jayaraman, Shuxing Zhang and Michael Zhu, I would like to thank you for your insightful questions, and unique perspective towards the completion of my work. I would also like to thank you for your valuable time you gave to me for evaluating my performance.

To the members of Gorfe lab, Priyanka Srivastava, Suparna Sarkar-Banerjee, Amit Gupta, Xubo Lin, Cynthia Pagba, Abdallah Sayyed-Ahmed, Zuhail, and Vinay, thank you for your thorough discussions regarding my project and thank you for providing me wonderful

environment and time to work together. Special thanks to my friend Michael McCarthy from my lab who has always been my mini-mentor and has guided me in every possible way he can.

Finally, I would like to dedicate my thesis to my parents, my brother, my dear husband and my close friends who have always been there with me through my thick and thin situations and have always believed in me for my accomplishment.

Thank you!

LBPI: A WEB INTERFACE FOR THE IDENTIFICATION OF ALLOSTERIC LIGAND BINDING SITES

Nabina Paudyal (MS)

Supervisory Professor: Alemayehu Gorfe, PhD

The development of efficient tools for allosteric ligand binding site identification in potential drug targets is an important step for computational drug design. Ligand binding specificity analysis (LIBSA) is one of the protocols that utilizes filtering algorithms to assess the propensity of a site on a target structure or structures to bind a ligand. However, LIBSA requires expert skills to be properly executed. Thus, a *Web interface*, LBPI (Ligand Binding Pocket Identification) has been developed using Django, a Python based web framework. A Python *Wrapper* has also been developed to streamline pre-existing algorithms of LIBSA. The *Wrapper* helps in the preparation of files, execution of individual programs and generation of appropriate results. LBPI provides an ideal platform for making complex binding site identification protocols readily available for non-expert users to submit jobs and monitor the results. The goal of LBPI is to integrate available algorithms in a systematic way and make it easily available for both experts and non-experts.

TABLE OF CONTENTS

APPROVAL PAGE	i
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	viii
LIST OF FIGURES	ix
CHAPTER 1 INTRODUCTION TO STRUCTURE-BASED DRUG DISCOVERY	1
1.1 Computational ligand design	1
1.2 Structure based computer-aided drug design	3
1.3 Molecular docking and scoring function	5
1.4 Targeting of an active site versus allosteric site	7
1.5 Approaches for allosteric ligand binding site identification	7
CHAPTER 2 LIGAND BINDING SPECIFICITY ANALYSIS	9
CHAPTER 3 LBPI: A WEB INTERFACE	14
3.1 Overview	14
3.2 Python programming and its modules used as a <i>Wrapper</i>	15
3.3 Django and its work structure used for web design	15
CHAPTER 4 IMPLEMENTATION OF LBPI	18
4.1 Overview	18
4.2 Implementation of blind docking and filters (option 1):	20
4.2.1 Processing coordinate files	20
4.2.2 Docking of the probe on the entire surface of the target protein	22
4.2.3 LIBSA filters	26
4.2.4 Ensemble-based SNR	29
4.3 Implementation of filters (option 2)	30
4.3.1 File preparation	30
4.3.2 LIBSA filters	31
4.4 Summary	31
CHAPTER 5 APPLICATIONS	33
5.1 Re-docking with LBPI	33
5.2 Cross-docking with LBPI	36
5.3 Identification of binding sites of test ligands with LBPI	43
5.4 Identification of the binding preferences of estrogens with LBPI	46
5.5 Identification of allosteric ligand binding site for sorafenib on p38 through LBPI	51
CHAPTER 6 CONCLUSIONS AND FUTURE DIRECTIONS	53

REFERENCES	56
VITA.....	66

LIST OF TABLES

Table 1. Summary of re-docking results with LBPI.....	36
Table 2. Relative binding affinity for estrogen compounds from literature and summary of results obtained from LBPI.....	48

LIST OF FIGURES

Figure 1.1 Timeline for drug discovery	1
Figure 1.2. Workflow of SB-CADD	3
Figure 2.1. Schematic outline of the LIBSA workflow, with the filters for the quantification of binding site preference.....	9
Figure 2.2. Blind docking of a ligand 0QV ((4-hydroxypiperidin-1-yl)(1H-indol-3-yl) methanethione) on the entire surface of GDP-bound K-Ras structure from PDB id 4EPW. Dense black dots represent poses of the probe ligand that might represent a probable binding site. The sparsely occupied region represented by fewer black dots might represent docking noise.	10
Figure 3.1. The workflow of the Django framework.....	16
Figure 4.1. Front-page of LBPI with two options of use highlighted in box 1. Manual is made available in the page through the link highlighted in box 2.....	18
Figure 4.2. A schematic outline of the workflow of LBPI.	19
Figure 4.3. Choices for submitting coordinate files that are for both the target receptor and the probe ligand as in the red boxes 1 and 3. Red box 2 represents a list of PDB chain identifiers and co-crystals that exist in the target structure returned to the user for selection. For docking, the blue box highlights example files provided for demonstration purposes....	20
Figure 4.4. Python snippet from Django Views representing a function which accepts and returns user requests for submitting files via the Web interface.....	21
Figure 4.5. A Python snippet from the “Wrapper” for extracting chain identifiers and residue names to facilitate returning of a list of chain identifiers and co-crystals to the user.....	22
Figure 4.6. Parameters required for blind docking shown in the red box. Blue box highlights downloadable ligand coordinate files generated by LBPI.	23
Figure 4.7. Python snippet from Django Views showing the method used to accept input parameters.	24

Figure 4.8. Python snippet from the “Wrapper” representing system arguments for passing parameter values retrieved from the framework shown in Figure 4.7.	24
Figure 4.9. Python snippet from the “Wrapper” showing the passing of parameter values through system arguments to invoke Python scripts present in MGL Tools of AutoDock for preparing parameter files needed for docking.	24
Figure 4.10. Grid box mapping the surface of the protein with its center overlapping the center of the protein.	25
Figure 4.11. LBPI facilitates selection of LIBSA filters with parameters for corresponding selected filters as shown in red boxes. Blue box 1 represents processed log files available for download, which can be viewed with any graphical interface. Histograms provided in blue box 1 represent affinity and contact spectra obtained from docking without utilizing filters. Histograms in blue box 2 represent affinity and contact spectra after applying LIBSA filters. SNR values are provided along with the residues representing the region of interest.	27
Figure 4.12. Python snippet of scripts written for preparing NumPy array used by Matplotlib as pl for preparing histogram.	28
Figure 4.13. Python snippet from Django framework for selecting filters.	29
Figure 4.14. Uploading multiple conformations to calculate SNR by averaging over an ensemble.	30
Figure 4.15. Options for submitting files of the receptor and log files from docking is highlighted in the red box with example files highlighted in the blue box.....	31
Figure 5.1. Chemical structures of ligand 0QV, SB2, and EST from PDB 4EPW, 1QKT and 1A9U respectively.	34
Figure 5.2. (A-C) Contact histograms after applying affinity and highpass filter to remove docking noise (left). The target structure shown in grey surface with the co-crystallized ligand is represented in red and the docked poses in black dots; SNR for the region of interest is shown in the right panel and is highlighted in green on the protein surface, which coincides with the co-crystal site.	35
Figure 5.3. Contact histogram after affinity and high pass filtering of docking results for NIL (3GP0) to protein kinase P38 (1A9U).	37

Figure 5.4. Chemical structures of ligands 0QR, 0QW, 0QX and BZI from PDB 4EPX, 4EPT, 4EPV and 4DSU respectively.....	38
Figure 5.5. Docked poses obtained from AutoDock for 0QR (blue), 0QW (yellow), 0QX (red) and BZI (green). Regions 1 and 2 are represented by pink and cyan color respectively and indicate the binding regions of interest for calculating SNR for each ligand.	39
Figure 5.6. Panels A, B, C, and D represent contact histograms for 0QR, 0QW, 0QX and BZI after docking on K-Ras (4EPW) and applying both affinity and high pass filters. SNR computed for region 1 is represented in pink and SNR computed for region 2 is represented in cyan.	40
Figure 5.7. Contact histograms for 0QR and 0QW obtained from docking these ligands to the parent protein X-Ray structure. SNR for region 1 and region 2 is represented as pink and blue respectively.	41
Figure 5.8. Contact histograms after blind docking of 0QR on five K-Ras conformers (A to E) and a single averaged contact histogram (F).....	42
Figure 5.9. Contact histograms after blind docking of 0QW on five K-Ras conformers (A to E) and a single averaged contact histogram (F).....	43
Figure 5.10. Contact histograms for three test ligands promazine (A), E22 (B) and V2 (C) docked on X-Ray structure 4EPW of K-Ras.	44
Figure 5.11. Contact histograms after blind docking of E22 on five K-Ras conformers (A to E) and a single averaged contact histogram (F).....	45
Figure 5.12. Contact histograms after blind docking of V2 on five K-Ras conformers (A to E) and a single averaged contact histogram (F).....	46
Figure 5.13. Chemical structure of estrogen compounds	49
Figure 5.14. Contact histograms obtained after applying affinity and high-pass filters to remove docking noise for estrogen compounds to estrogen receptor (1QKT).....	50
Figure 5.15. Contact histogram after affinity and high pass filtering of docking results for sorafenib to its protein kinase P38 (3HEG) shown in left panel, structure of P38 with docked poses in yellow, three identified region from LBPI, co-crystal (sorafenib) in red, and	

chemical structure of sorafenib in right panel.....	52
---	----

CHAPTER 1

INTRODUCTION TO STRUCTURE-BASED DRUG DISCOVERY

The process of structure-based drug discovery, as outlined in Figure 1.1 starts with basic research towards investigating the effect of potential drug candidates on an identified druggable target. First, basic research is performed to identify appropriate disease-causing targets or druggable targets such as proteins. Compounds are then screened to identify potential hits and eventually identify a drug lead, which is further subjected to pre-clinical and clinical trials, and finally filing for approval by the Food and Drug Administration. The new drug introduced to the market is administered to patients ensuring its safety, potency, and bioavailability .

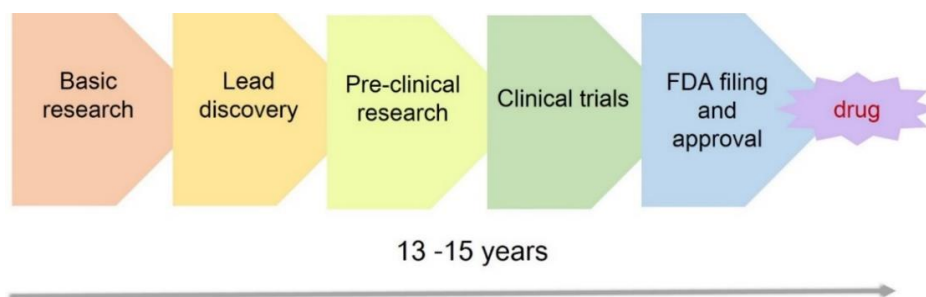


Figure 1.1 Timeline for drug discovery

1.1 Computational ligand design

Experimental screening approaches can be used to identify compounds for druggable targets but they can often be expensive and time-consuming (2). In addition, the number of druggable targets has increased enormously over the past 50 years (3). In this period, computational biology has made tremendous advances in facilitating drug discovery by

expediting the process and reducing cost. Various *in silico* applications have been introduced that are collectively referred to as computer aided drug design method (CADD) (4). FDA approved drugs required billions of dollars using traditional research and methodologies, and thus discovery of new drugs has been declining since 1950 (5), which can be partly reversed through the use of CADD. The introduction of CADD also provided opportunities for developing novel techniques to enhance better understanding of drug mechanisms and improve efficiency (6).

There are two major techniques used in CADD: a) ligand-based computer aided drug design (LB-CADD) and b) structure-based computer aided drug design (SB-CADD) (7). LB-CADD is a statistical approach that relies on structure-activity relationship information between existing ligands and their targets. This technique can also be used as an indirect drug design method that does not directly use knowledge of the target of interest and thus becomes more challenging if there is no known ligand for the target of interest. In this case, SB-CADD is typically a better approach (8). SB-CADD requires the 3D structure of the target for the identification of new chemical compounds. In 1973, Beddell and Goodford reported the first structure-based drug design method for designing compounds that emulate the function of 2-3 diphosphoglycerate in human hemoglobin (9-11). Compounds that can modulate the function of the target by binding with high affinity and selectivity have a higher chance of becoming hits (7, 12, 13). A key aspect of SB-CADD is the development of efficient strategies for identifying potential hits through high throughput virtual screening (HTVS) involving molecular docking and calculation of scoring functions. SB-CADD also allow us for a detailed

understanding of the interaction in ligand-target complexes, and for addressing conformational changes of both compound and target (14, 15). The identified hits are further subjected to experimental procedures like biochemical or biophysical assays and undergo optimization before they become drug leads (Figure 1.2).

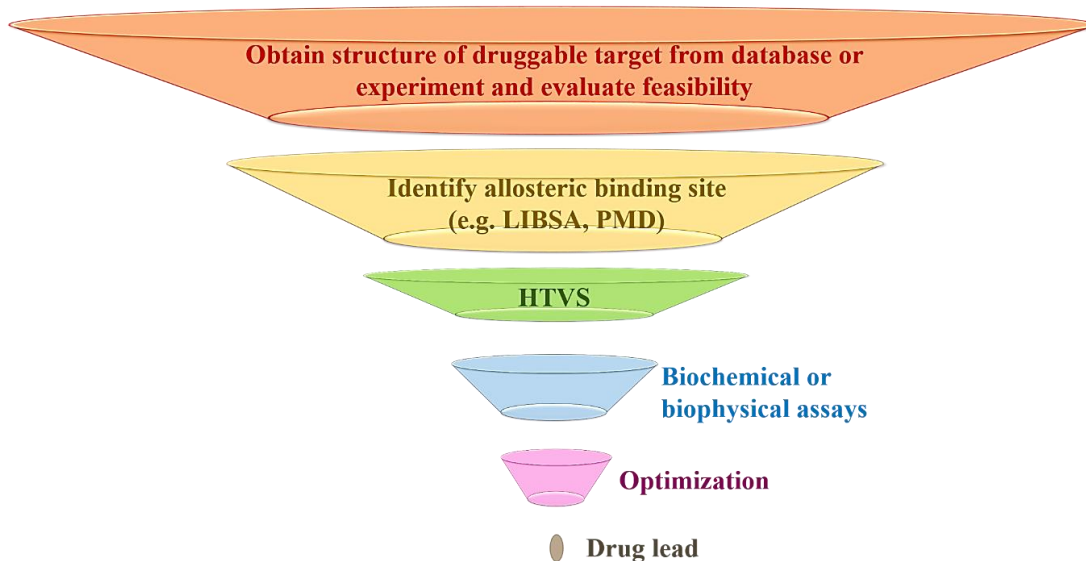


Figure 1.2. Workflow of SB-CADD

1.2 Structure based computer-aided drug design

Application of SB-CADD (or any other similar technique) requires a 3D structure of the target macromolecule, which can be obtained from X-ray crystallography or NMR spectroscopy (16, 17). X-ray crystallography offers detailed yet static information about the target structure while NMR provides an ensemble of structures for relatively small proteins (17). The structures of macromolecules obtained from these and other methods are deposited in the protein data bank (PDB) (18). PDB, the most commonly used database for protein

structure, was established in 1971 at the Brookhaven National Laboratory and the data deposited were maintained in Brookhaven, Tokyo, and Cambridge. At present PDB is managed by members of the Research Collaboration for Structural Bioinformatics (RCSB) consortium based at Rutgers and UC San Diego (16, 18). The first reported 3D protein structure was X-ray structure of sperm whale myoglobin published in 1958, and deposited to the PDB as PDB-id 1MBN in 1973 by Kendrew and Watson (11, 16, 19, 20). Statistics of yearly growth of total structure show an exponential growth with more than 125,000 biological macromolecules currently available in the PDB. These structures are solved by X-Ray crystallography or NMR. Comparative or homology modeling may also help in providing information about large macromolecules whose structure remains unidentified by experimental methods (21). While some experimental methods might be able to account for receptor flexibility, it is expensive to generate multiple structures for each macromolecule. Molecular dynamics simulations can be used to generate large conformations of target structures to account for its dynamic nature (22). Representative structures of these large number of conformations from MD can be generated through various methods, such as RMSD clustering and principal component analysis (23, 24).

Availability of ligand libraries is another important component of SB-CADD. Several repositories of chemical compounds are available. Some of the compounds in these libraries possess drug-like properties satisfying Lipinski's Rule of 5: a drug-like compound should possess less than five hydrogen bond donors, less than ten hydrogen bond acceptors, should have a molecular weight of 500 Daltons or less and a logP (calculated log of octanol-water partition coefficient) less than 5 (25, 26). Examples of chemical databases include PubChem,

a database of 25 million unique compounds with annotations about their biological activities (27). Others include ZINC (28) and ChemDB (29), containing about 21 million and 5 million compounds respectively, with Vendor's catalog (12, 28, 29).

1.3 Molecular docking and scoring function

Molecular docking is widely used in structure-based drug design for the identification of compounds that interact with the target of interest by characterizing its binding mode and calculating its binding energy. The earliest reported molecular docking was in 1982, on two systems: interaction of heme with myoglobin, and binding of thyroid hormone analogs to prealbumin. At that time, both the macromolecule and the ligand were treated as a rigid body during docking (11, 30, 31). Molecular docking can be used for a) site-directed docking, which refers to docking of molecules at a specific site of the target for virtual screening of chemical compounds to identify potential hits or b) blind docking, which refers to docking of probe molecules or ligands on the entire surface of the target to identify an appropriate binding site. The most common molecular docking programs currently used are AutoDock (32, 33), GOLD (34), FlexX (35), DOCK (36). Binding energy typically obtained from molecular docking is calculated using a physics-based scoring function (33).

$$\begin{aligned}
\Delta G_{binding} = & \Delta W_{vdW} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \Delta W_{hbond} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) \\
& + \Delta W_{elec} \sum_{i,j} \frac{q_i q_j}{r_{ij} \epsilon(r_{ij})} + \Delta W_{tor} N_{tor} \\
& + \Delta W_{sol} \sum_{i,j} (S_i V_j + S_j V_i) e^{\left(\frac{r_{ij}^2}{2\sigma^2} \right)}
\end{aligned}$$

Where $\Delta G_{binding}$ refers to the binding energy or the scoring function that depends on a 12-6 Lennard-Jones potential (vdW) for non-polar interactions, a modified 12-10 Lennard-Jones potential for hydrogen bonding interactions (hbond), a coulomb term for electrostatic interactions (elec), as well as torsional (tor) and solvation energy (sol). The ΔW coefficient for each interaction term is calculated by linear regression analysis of known protein-ligand complexes. A_{ij} and B_{ij} represent constants for the Lennard-Jones potential, which depend on the strength of van der Waals interaction between two atoms i and j with equilibrium distance r_{ij} (37). Whereas C_{ij} and D_{ij} represent constants for a modified Lennard-Jones potential with their value depending on the interaction between atoms i and j through H-bonding (33). $E(t)$ represents directionality based on hydrogen bonding angle. The charges on atoms i and j are represented by q_i and q_j with ϵ being the dielectric constant of the medium. The number of torsions of the ligand is represented by N_{tor} and depends on the number of its sp^3 -hybridized bonds. Finally, the solvation term is atom specific and is represented as S_i for an atom i with fragmental volume V_i . The constant term σ in the calculation of solvation energy is a Gaussian distance of value is 3.5 Å (33).

1.4 Targeting of an active site versus allosteric site

Referring back to the workflow of SB-CADD shown in Figure 1.2, before the identification of potential hits it is very important to identify a feasible binding site on an appropriate druggable target, without which identification of a drug lead is very challenging (38, 39). Druggable targets, which are typically proteins, bind to other macromolecules like proteins or small molecules to execute their functions (17). Molecules binding to the active site can have high specificity and usually bind with high affinity (17). However, many drugs designed to target the active site and inhibit function might produce toxic effects by binding to other proteins with a structurally similar active site (40). An alternative approach is to alter the function of the target protein via allosteric modulation (40, 41). Allosteric sites are distinct binding sites from the active site and can be feasible sites for developing novel drugs. When an allosteric ligand binds to the allosteric site it could modify the active site and thereby alter the function (42, 43). The mechanism for allosteric binding is well elucidated with the analysis of the available 3D structures of proteins (44). Drugs targeting allosteric sites are generally less toxic and have greater selectivity, and hence allosteric binding site identification is an important endeavor in current SB-CADD. (41).

1.5 Approaches for allosteric ligand binding site identification

FTMAP (45), MDpocket (46), Sitemap (47), pMD (48), and LIBSA (49) are some of the approaches commonly used for allosteric ligand binding site identification. FTMAP is based on a fast Fourier transform (FFT) correlation approach for sampling consensus site on

the entire surface of the protein using small organic probe molecules. The speed with which it solves detailed energy expressions used for scoring protein-probe interactions makes this approach unique (45). MDpocket is a geometry based cavity detection approach based on the f-pocket algorithm. It helps in the identification and characterization of binding cavities that might be transiently present in MD trajectories or ensembles of protein conformations (46). On the other hand, Sitemap helps in the prediction and classification of target druggability based on ligand binding efficiency. In this method, site points are located on the protein's surface, followed by the preparation of contours representing the nature of the site and hence prediction and classification of sites (47). pMD-membrane (48) and LIBSA (49) techniques developed previously in our lab utilize MD ensembles for identifying allosteric binding pockets. In pMD-membrane, MD is performed in the presence of small molecular probes to assess the propensity of a site on a soluble or membrane bound protein to bind a ligand. In this method, probe occupancy is used for evaluating druggability of a binding site, and modified force field parameters are introduced between selected probe and membrane atoms to mitigate potential effects of the probe molecules on the membrane (48). LIBSA utilizes three steps to identify binding sites: a) blind docking of the ligand onto a target receptor, b) application of filters on docked poses, and c) use of signal to noise ratio for quantification of binding site preferences (49).

Since LIBSA is the primary tool used for allosteric ligand bind site identification in this project, the next chapter provides a detailed description of this technique.

CHAPTER 2

LIGAND BINDING SPECIFICITY ANALYSIS

Ligand binding specificity analysis (LIBSA) (49) is a novel approach for addressing the challenge of ranking blind docked results designed to identify preferred binding site(s). The general workflow of LIBSA starts with the processing of coordinate files of a target receptor (a single or multiple conformations) and a probe ligand followed by blind docking using AutoDock or any other docking algorithm (the original paper utilizes AutoDock for blind docking). Filtering algorithms were introduced (affinity filter and high pass filter) to remove docking noise and to prioritize poses, followed by quantification of binding site preferences. The flowchart in Figure 2.1 provides a schematic outline of LIBSA.

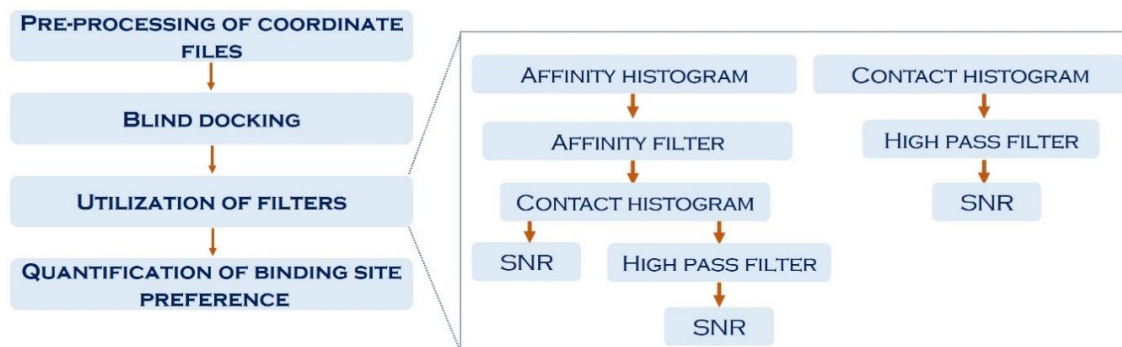


Figure 2.1. Schematic outline of the LIBSA workflow, with the filters for the quantification of binding site preference.

Blind docking, which is the key component of LIBSA, refers to the identification of probable binding sites by docking probe molecules on the entire surface of the target. AutoDock, the most cited docking program in the literature, was first used for blind docking by Csaba Hetényi and David van der Spoel (50, 51). The basic procedure for blind docking

using AutoDock involves processing structure files for the probe ligand and the target molecule by adding Gasteiger charges and hydrogen atoms, and removing water molecules, ions and metals (if any). A grid, which is important for defining the search space, is built on a three-dimensional box covering the entire protein with the center of the box overlapping with the center of the protein. This is done by means of AutoGrid. Lamarckian Genetic Algorithm (LGA), a common and efficient search method in AutoDock, determines the termination criteria based on the number of energy evaluations or the number of generations, whichever comes first (32, 33). Ultimately, conformations for the best binding mode of the ligand are obtained and their corresponding binding free energy is returned to the user as a log file.

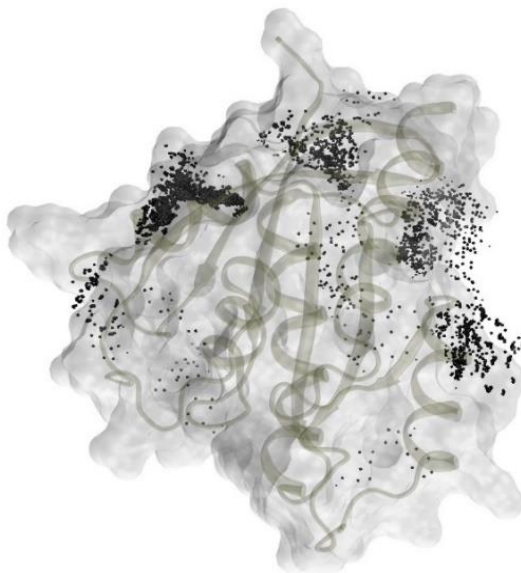


Figure 2.2. Blind docking of a ligand 0QV ((4-hydroxypiperidin-1-yl)(1H-indol-3-yl) methanethione) on the entire surface of GDP-bound K-Ras structure from PDB id 4EPW. Dense black dots represent poses of the probe ligand that might represent a probable binding site. The sparsely occupied region represented by fewer black dots might represent docking noise.

Two filters, namely affinity filter and high pass filter, were developed in order to facilitate analysis of blind docking results. The filters use the log file (dlg format) from the docking program to remove docking noise. Filtering signal from noise is the most novel aspect of LIBSA (49). Each filter works in a different way but serves the same purpose: remove docking noise. Affinity filter reduces docking noise by utilizing docking free energy values. Since energy values calculated through docking programs typically entail an error of 2kcal/mol (32, 52), accounting for binding preference based on the energy alone is not effective. Affinity filter addresses this challenge by considering the frequency of occurrence of docking scores. Firstly, an affinity histogram is prepared based on the distribution of binding scores, and a cutoff value is introduced to focus on the most prominent peaks. Peaks with frequency of occurrence more than or equal to cutoff value are identified as explicit peaks. Then for further analysis, peaks within a certain sampling window of the explicit peaks are included as auxiliary peaks. The docked conformations corresponding to the explicit and auxiliary peaks represent filtered poses based on affinity filtering. The second filter is a high-pass filter that utilizes the concept of digital signal processing for removing noise. It is based on the frequency of contact between heavy atoms of the probe molecule and heavy atoms of residues of the protein that lie within a certain distance. This histogram is presented as input for the filter. A cutoff for peak value is once again used to focus on the residue that interact most frequently with the ligand. The flowchart for utilization of LIBSA filters is shown in the right panel of Figure 2.1. The two filters can be used independently or together for reducing docking noise.

Calculating the signal-to-noise (SNR) ratio is the final step of LIBSA. SNR helps to quantify binding site preference. With the calculation of SNR, a best binding site out of probable multiple binding sites can be identified, or ligands can be ranked for their preference to bind a site based on their SNR values. A contact histogram is required for computing SNR, and is prepared by considering the frequency of contacts between heavy atoms of the probe molecule and heavy atoms of residues of the protein that lie within a certain cutoff distance. With the application of LIBSA filters, filtered poses are utilized for preparing contact histogram from which the identified region of interest is considered as signal and other regions are considered as noise. SNR is calculated as,

$$SNR = \log_{10} \left[\frac{N^{noise} \sum u_i}{N^{signal} \sum u_j} \right]$$

where u_i is the height of the peak of residue i in the region of interest (considered as signal), N^{signal} is the total number of residues that lie within the same patch, u_j represents the height of the peak of residue j that is outside the signal region and N^{noise} represents the total number of residues that lie outside the signal patch.

Ensembles-based SNR involves calculation of SNR over multiple representative conformations of a target of interest. Identification of a preferred binding site of an ensemble by utilizing LIBSA involves independent docking on each target conformation, followed by application of filters to each independent molecular dockings and averaging of the results to produce a single contact spectrum. The averaged spectrum is then utilized for calculating SNR as described above.

LIBSA was applied to several sets of known protein-ligand complexes for validation. Few sets of probe-ligands were used on Ras protein conformations obtained from the PDB or MD simulation for re-identification of binding sites to test the robustness of LIBSA. In addition, andrographolide derivatives were used as ligands with K-Ras ensemble for quantification of preferred binding sites, and the ligands were ranked based on their binding preference of those ligands to a site. Hence, LIBSA can be applied in any challenging drug target for the identification of binding site or can be applied with diverse ligands for ranking them based on their preference to bind to a site.

However, application of LIBSA, which entails processing of files, performing blind docking, applying filters to reduce docking noise, and quantifying binding site preferences by calculating SNR, requires expert skills to be properly executed. The minimum skills required for the execution of LIBSA include, processing of structure files in appropriate file formats, setting grid and docking parameters, executing LIBSA algorithms written in C++ programming language, preparing histograms and providing values to the algorithms for calculation of SNR. Thus, LIBSA is not easily accessible to non-experts. Hence, to reduce the significant learning curve and make accessible to both experts and non-experts, we hypothesized that, *a web interface for streamlining algorithms of LIBSA and automating its disparate protocols will facilitate efficient identification of allosteric binding pockets.*

CHAPTER 3

LBPI: A WEB INTERFACE

3.1 Overview

Ligand Binding Pocket Identification (LBPI) is a *Web interface* designed for the identification of allosteric ligand binding sites by employing the protocols of LIBSA. The protocols include processing of coordinate files followed by molecular docking using AutoDock, and applying LIBSA filters for removing docking noise and quantifying binding site preference. LBPI is built on a freely-available Python-based web framework, Django (53), accompanied by a set of scripts written in the Python programming language that serves as a *Wrapper*. The main function of the web framework is to handle requests for uploading or downloading files, accept modules and parameters that have been passed through the web page, and return a response to URL to view the contents of the web page. It also stores information about file location. The *Wrapper*, on the other hand, helps to process coordinate files, automate the application of existing algorithms, pass variables to the relevant algorithms, execute programs and prepare histograms. The platforms to upload or download files, select program modules, and input user defined parameters and monitor results, are facilitated through the *Web interface* accessed via a web browser. The goals of the LBPI *Web interface* are to make LIBSA easily accessible to both experts and non-experts and to reduce the significant learning curve of the technical details needed to perform research independently by non-experts.

3.2 Python programming and its modules used as a *Wrapper*

Python (54, 55) is a freely available high-level object oriented programming language rich in a wide variety of libraries, including scientific libraries extending from bioinformatics to structural biology. Python programming is chosen for preparing a *Wrapper* as it requires fewer line scripts compared to other object-oriented languages such as C++, or C. Hence, it can be written with fewer errors and is easily understood (56, 57). Python is available in two versions: Python 2.X and Python 3.X. Since Python 3.X has more updated libraries and features, it is used in the development of our *Wrapper*.

Python modules such as sys, re, subprocess, NumPy (58), Matplotlib (59), etc. are used in the *Wrapper* for executing existing algorithms, and preparing results. The sys module is used as a command line argument for accessing the variables used, while the re module, commonly known as regular expression syntax, is used to match strings and implement simultaneous jobs designed for each matched string(s). Similarly, the subprocess module helps to access system commands and is used in executing new processes, followed by piping inputs and outputs. On the other hand, NumPy and Matplotlib are standard libraries of Python used for scientific computation and preparing 2D plots. Arrays of data are created by NumPy for data normalization and plotting of histograms, simple plots or bar plots utilizing Matplotlib.

3.3 Django and its work structure used for web design

Django (53, 60), a freely available Python-based web framework, facilitates developing, testing and maintaining web applications. Django is very well documented and its

active community of users continuously update the Django project. Therefore, Django was chosen as a web framework for LBPI. Some of the scientific projects built on Django web framework are CHARMMing (61), TargetNet (62), ChemDes (63), and DEPTH (64). The working of Django starts with creating a project consisting of several sets of codes defined as applications. Applications of Django include Views, Models, Forms, URLs, and templates. The workflow of Django is as shown in Figure 3.1.

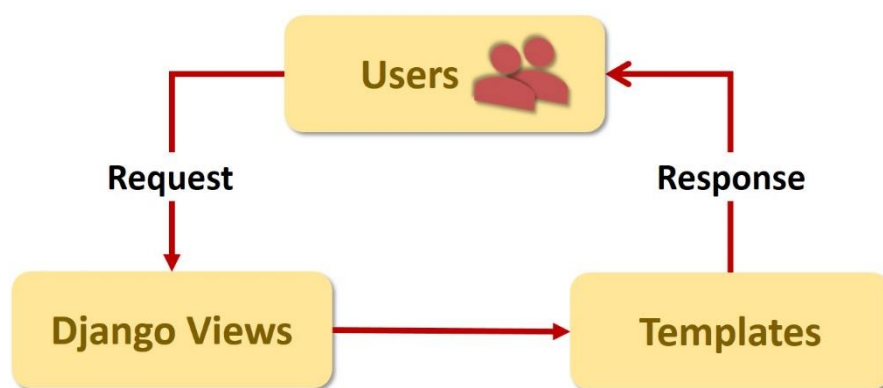


Figure 3.1. The workflow of the Django framework.

A Django template consists of HTML contents, tags (if, else, extend, for etc.) and variables (which change with the results) for viewing web applications. A Django-powered site is viewed on a web browser through Django templates containing a specific URL (Uniform Resource Locator), a web address written in a certain pattern mapped with its Django Views. Each view in Django application is written as Python function assigning specific tasks, and hence renders requests and returns responses as Django templates. Django Models cover essential fields and behaviors of the data to be stored, and derive essential information from the data. The file field attribute of the model helps in setting folder and file names for the data

file to be stored. Django Forms are used to accept user inputs as files or data, validate the inputs and convert them to Python objects.

The Django framework, the Python *Wrapper* and *Web interface* templates written in HTML, complement each other to automate LIBSA protocols, execute jobs and return the results to the user for the identification of druggable binding sites and quantification of binding site preferences.

CHAPTER 4

IMPLEMENTATION OF LBPI

4.1 Overview

As discussed in previous chapters, the goals of LBPI are to make LIBSA easily accessible to both experts and non-experts and thereby reduce the significant learning curve needed to use the tool independently. In this chapter we present the implementation of LBPI in two ways, option 1 and option 2, as shown in Figure 4.1:

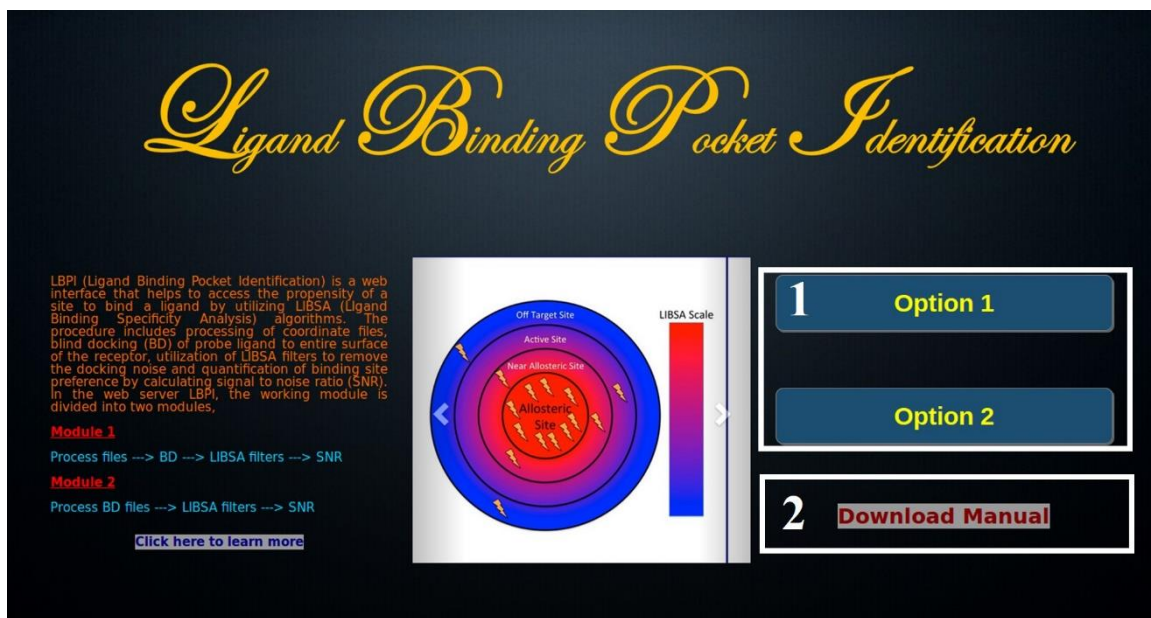


Figure 4.1. Front-page of LBPI with two options of use highlighted in box 1. Manual is made available in the page through the link highlighted in box 2.

Option 1: Process files → perform blind docking → apply LIBSA filters → calculate SNR

This option starts with processing coordinate files for docking, followed by blind docking and executing LIBSA as illustrated in Figure 4.2.

Option 2: Process docking files → apply LIBSA filters → calculate SNR

This option starts with processing of a target structure file used for docking and the docking log files. These files are converted to LIBSA readable formats if needed and serve as input to LIBSA. The flowchart shown in Figure 4.2 holds true for this application as well, except that the module for blind docking is skipped and processing of coordinate files is replaced by processing of docking output files.

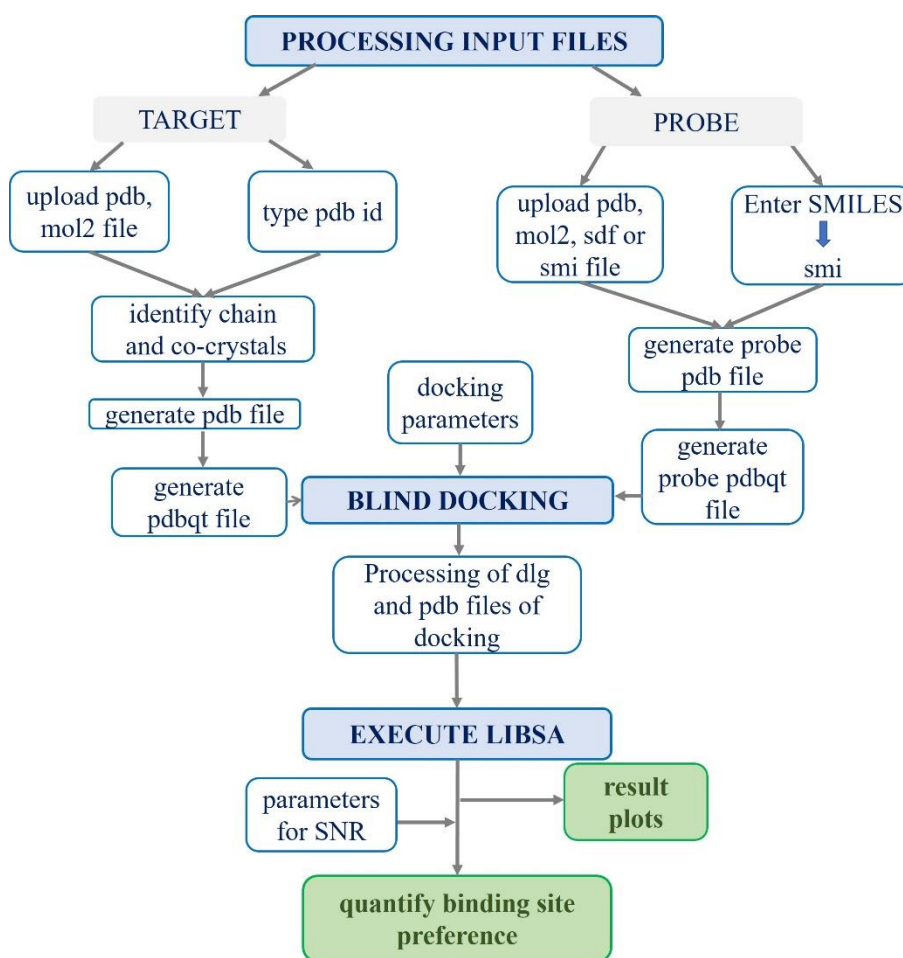


Figure 4.2. A schematic outline of the workflow of LBPI.

4.2 Implementation of blind docking and filters (option 1):

4.2.1 Processing coordinate files

LBPI offers an option for submitting and processing structure files along with example files for demonstration purposes. Figure 4.3 shows a snapshot of the *Web interface* for processing structure files.

LBPI

Example files

Schematic outline

- Step 1: Process files
- Step 2: Blind docking
- Step 3: LIBSA filters

Process coordinate files

Binding site identification is one of the important step for structure based drug design. To begin with, appropriate structure formats for both target receptor (mainly proteins) and probe ligand is required.

For target receptor, mol2 and pdb file format can be uploaded or appropriate pdb id can be used for downloading the structure from protein data bank. For probe ligand, sdf, mol2, pdb and smi file format can be used or appropriate SMILES string can be used, which is converted to 3D structure using Open Babel.

The files are processed by converting files (other than pdb files) to pdb format using Open Babel followed by adding hydrogen and Gasteiger charges, removing ions and water molecules by using MGL tools.

Processing of receptor 1

- ☐ Upload your receptor file (.mol2, .pdb format)
- ☐ Download from RCSB

Processing of ligand 3

- ☐ Upload your ligand file (.smi, .sdf, .mol2, .pdb format)
- ☐ SMILES

For 4epw.pdb downloaded include - 2

Note: The information about 4epw.pdb can be found [here](#)

- ☒ Chain A
- ☒ Include co-crystals:
- ☐ OQV ☒ GDP

Copyright © 2017 Ocular Research group at UThweth, Medical School

Figure 4.3. Choices for submitting coordinate files that are for both the target receptor and the probe ligand as in the red boxes 1 and 3. Red box 2 represents a list of PDB chain identifiers and co-crystals that exist in the target structure returned to the user for selection. For docking, the blue box highlights example files provided for demonstration purposes.

The process of submission and preparation of the target structure for molecular docking is done automatically. Submission of a structure or structures can be done in two ways: a)

uploading a file in a pdb or mol2 format or b) providing a valid 4 letter PDB id for retrieving the structure from the protein data bank. The Python snippet shown in Figure 4.4 highlights a function in Django Views that handles the user request for submitting the structure via the POST method. Document and Document Form used in the script are defined as functions in Django Models and Django Forms for storing the file in a defined folder, and handling file submission, respectively. A structure submitted by providing valid pdb id is handled through the GET method.

```
def Libsa(request):
    if request.method == 'POST':
        form = DocumentForm(request.POST, request.FILES)
        if form.is_valid():
            newdoc = Document(docfile=request.FILES['docfile'])
            newdoc.save()
    else:
        form = DocumentForm() # An empty, unbound form

        documents = Document.objects.all()

    return render ( request, 'libsa/preparefile.html',{'documents':
documents, 'form': form})
```

Figure 4.4. Python snippet from Django Views representing a function which accepts and returns user requests for submitting files via the Web interface.

The *Wrapper* invoked through Django Views after submission of the target file assists in the conversion of a mol2 file format to a pdb file format by means of Open Babel (65). Figure 4.5 shows a snippet from the *Wrapper* used for extracting chain identifiers and co-crystals that might be present in the coordinate file. A list is then returned to the user for selection of chains and co-crystals to be included in the target structure for docking. The prepared structure is then processed automatically with tasks such as renumbering of residue

id, addition of atomic Gasteiger charges and polar hydrogens, removal of ions and water molecules by invoking Python scripts in MGL Tools of AutoDock, via the *Wrapper*.

```
if re.match('ATOM',line[0:6]):  
    res_lists.append(line[17:20])  
    chain.append(line[21:22])
```

Figure 4.5. A Python snippet from the “Wrapper” for extracting chain identifiers and residue names to facilitate returning of a list of chain identifiers and co-crystals to the user.

Similarly, the probe ligand can be submitted in pdb, sdf, smi or mol2 file formats. It can also be submitted as line notation entry termed as SMILES (simplified molecular-input line-entry system) (66). The *Wrapper* converts SMILES strings to a 3D coordinate file in pdb format utilizing Open Babel. The ligand coordinate file is then made available for download through the *Web interface* to confirm if the desired structure has been generated. The working of Django framework for handling submitted structure for ligand and processing of prepared ligand structure is similar to that explained for processing of the target structure.

4.2.2 Docking of the probe on the entire surface of the target protein

Following the preparation of coordinates of the probe and the target molecule, the *Wrapper* calculates the size of the target molecule to map a cubical grid on the entire surface of the target to define the search space for docking. The parameters required for blind docking can be provided through the *Web interface* as shown in Figure 4.6, which include:

Spacing is the distance in Å between the centers of two grid points, with the same value used in every dimension for a cubic box. The default (recommended) value depends on the size of the target and ranges from 0.1 Å to 1 Å.

Energy evaluations is the maximum number of energy evaluations to be performed during docking. Its value ranges from 250,000 to 25,000,000 depending on the number of rotatable bonds present in the probe ligand, with the default being 250,000.

Number of generations is the maximum number of generations a docking should attain to terminate the run. The default is 27,000, which is sufficient for ligands with less than 10 torsions.

Number of runs refers to the number of independent runs to be requested for docking. Its maximum value is 256 and is used as default.

The screenshot shows the LBPI web interface. On the left, a 'Schematic outline' shows three steps: 'Step 1: Process files', 'Step 2: Blind docking' (highlighted with a red dotted line), and 'Step 3: LIBSA filters'. The main content area is titled 'View uploaded ligand or re-upload' and contains a 'Blind docking' section. This section includes a description of blind docking and a 'Parameters for blind docking' form. The form is divided into two columns: 'Parameters for autogrid' and 'Parameters for autodock'. The 'Parameters for autogrid' column has a 'Spacing' input field with a value of 0.437. The 'Parameters for autodock' column has three input fields: 'Energy evals' (250000), 'No of gens' (27000), and 'No of runs' (10). A 'Submit Values' button is at the bottom left of the form. A blue box highlights the 'View uploaded ligand or re-upload' section, which contains a 'Ligand file' download link and a link to upload another ligand file. A red box highlights the 'Parameters for blind docking' section, which contains the parameter input fields.

Figure 4.6. Parameters required for blind docking shown in the red box. Blue box highlights downloadable ligand coordinate files generated by LBPI.

The parameters submitted through the *Web interface* are handled by Django Views by the code snippet shown in Figure 4.7, which is then passed to the *Wrapper* through system

arguments as shown in Figure 4.8, for executing specific tasks assigned as shown in Figure 4.9.

```
if request.method == "GET":
    ag1 = request.GET.get('ag1') # parameter for spacing in autogrid
    ad1 = request.GET.get('ad1') # parameter for energy evaluations
    ad2 = request.GET.get('ad2') # parameter for no of generations
    ad3 = request.GET.get('ad3') # parameter for no of runs
```

Figure 4.7. Python snippet from Django Views showing the method used to accept input parameters.

```
dirr = sys.argv[1]
space = sys.argv[2]
points = sys.argv[3]
evals = sys.argv[4]
gens = sys.argv[5]
run = sys.argv[6]
```

Figure 4.8. Python snippet from the “Wrapper” representing system arguments for passing parameter values retrieved from the framework shown in Figure 4.7.

```
sb.call([shellpath, '{}prepare_dpf4.py'.format(pythonpath), '-l',
'{}ligand.pdbqt'.format(dirr), '-r', '{}protein.pdbqt'.format(dirr),
'-p', evals, '-p', gens, '-p', run, '-o', '{}ligand_protein.dpf'.format(dirr)])
```

Figure 4.9. Python snippet from the “Wrapper” showing the passing of parameter values through system arguments to invoke Python scripts present in MGL Tools of AutoDock for preparing parameter files needed for docking.

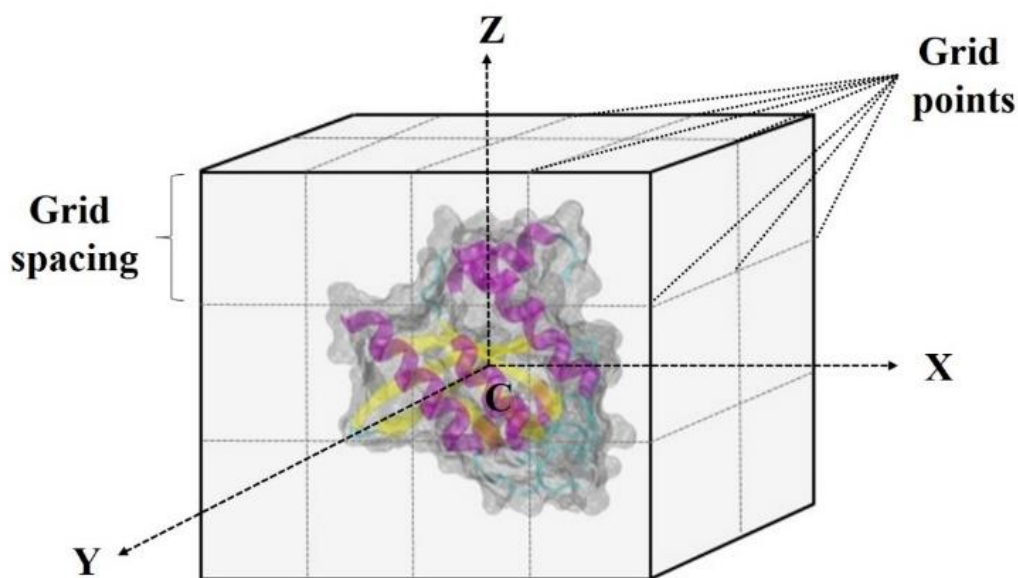


Figure 4.10. Grid box mapping the surface of the protein with its center overlapping the center of the protein.

The maximum length of the target protein structure in each direction represents the length of the cubic grid box after adding 10 Å in each direction to accommodate the ligand in the search space. For a cubic box of length l , the recommended spacing (S_{rec}) for efficient calculation of interaction energies is $S_{\text{rec}} = \frac{l}{126} \text{ Å}$, where 126 is the maximum total number of grid points allowed in AutoDock. For $S_{\text{rec}} \geq 1 \text{ Å}$, S_{rec} is set to 1 Å as AutoDock has a spacing (S) limit of $(0.10 \leq S \leq 1) \text{ Å}$. If a user provides a spacing value through the *Web interface*, the number of grid points (N) is calculated as,

$$N = \begin{cases} \frac{l}{S} & \text{if } N \leq 126 \\ 126 & \text{if } N > 126 \end{cases}$$

With the adjustment of the grid box around the protein surface with its center located at the center of the protein, and using the parameters provided through the *Web interface*, the *Wrapper* invokes corresponding AutoDock Tools for blind docking. The target file used for docking and the docking log file obtained from docking are then converted to LIBSA readable formats to execute LIBSA.

4.2.3 LIBSA filters

The module for LIBSA filters in the *Web interface* is as shown in Figure 4.11. Histograms for contact and affinity spectra from docking are displayed through the *Web interface* to help the user decide about the filter to be used for removing docking noise. Histograms are prepared with the help of the *Wrapper* using NumPy and Matplotlib library as shown in Figure 4.12, which is then returned to the user through the *Web interface*.

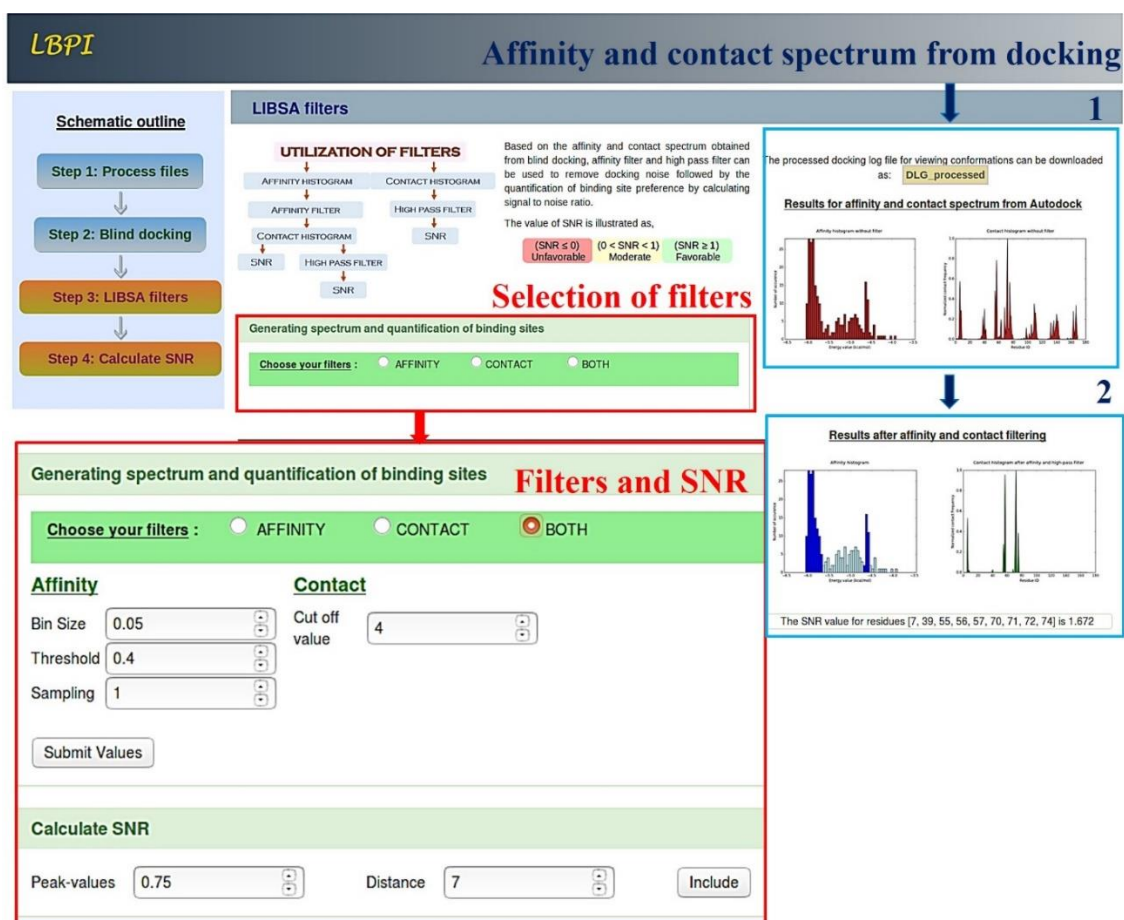


Figure 4.11. LBPI facilitates selection of LIBSA filters with parameters for corresponding selected filters as shown in red boxes. Blue box 1 represents processed log files available for download, which can be viewed with any graphical interface. Histograms provided in blue box 1 represent affinity and contact spectra obtained from docking without utilizing filters. Histograms in blue box 2 represent affinity and contact spectra after applying LIBSA filters. SNR values are provided along with the residues representing the region of interest.


```

data = np.loadtxt(hist_dat)
xdata = data[:,0]
ydata = data[:,1]
pl.xlabel("Residue ID")
pl.ylabel("Normalized contact frequency")
pl.plot(xdata,ydata,color='black',alpha=0.7)

```

Figure 4.12. Python snippet of scripts written for preparing NumPy array used by Matplotlib as pl for preparing histogram.

If LIBSA filter is selected, the *Web interface* allows the user to input the appropriate parameters. These parameters include **bin size**, which is the width of bins to be used when calculating histograms; **threshold**, which refers to the cut-off for the peak height ranging from 0 to 1 due to normalization, peak values less than the defined threshold are excluded; **sampling bin size** 1 to 100, the number of bins to the left and right of the selected explicit peak; and **cut-off**, the distance between any heavy atom of the ligand and the target in Å for calculating the number of contacts. The default value for each parameter is provided based on the recommendations in the original Journal article (49). When the parameter values are submitted through LBPI, the *Wrapper* automates the execution of LIBSA and returns the results to the user through LBPI. Figure 4.13 represents a snippet of a function based on the selection of data through LBPI. A function in Django Views assigned for the corresponding filter passes the parameter values and executes the LIBSA algorithm for affinity filtering. Finally, the results are returned to the user as an HTML page in LBPI.

```
def affinity(request):
    if request.method == "GET":
        energy_steps = request.GET.get('aff1')
        percentchange = request.GET.get('aff3')
        aux_peak = request.GET.get('aff4')
        cutoff = request.GET.get('con1')

        sb.call(['python', '{}libs.py'.format(wrappers), current_dir,
        LIBSA, 'affinity_only', energy_steps, percentchange, aux_peak, cutoff])

    return render(request, 'personal/filters/filters_affinity.html')
```

Figure 4.13. Python snippet from Django framework for selecting filters.

Based on the selection of LIBSA filters through LBPI, a contact histogram is calculated and returned to the user for them to decide on parameters to be provided for calculating SNR. These parameters include **peak value**, which refers to the minimum value of the dominant peaks in the contact spectrum to be considered for locating residues representing the center of region of interest. Then neighboring residues within a certain **distance** of these central residues that satisfy the peak histogram criterion are selected to define a binding region of interest. This region serves as a signal to calculate SNR through LBPI.

4.2.4 Ensemble-based SNR

LBPI facilitates uploading of multiple receptor conformations through the ‘**For ensembles**’ button provided in the *processing of receptor section*. All conformations should consist of an equal number of atoms and should be in the pdb file format. When uploaded, these conformations are saved in separate folders and are processed separately with the help of the *Wrapper*. At least 2 target conformations should be uploaded. A probe ligand is then uploaded, processed and sent to each folder for docking, application of filters and obtaining

contact and affinity spectra independently. Next, by means of the *Wrapper*, a single contact spectrum and affinity spectrum is produced by averaging the corresponding spectra obtained for each target conformation. Finally, a binding region of interest is identified and SNR is calculated for the identified region of interest to get a best binding site for the ensemble.

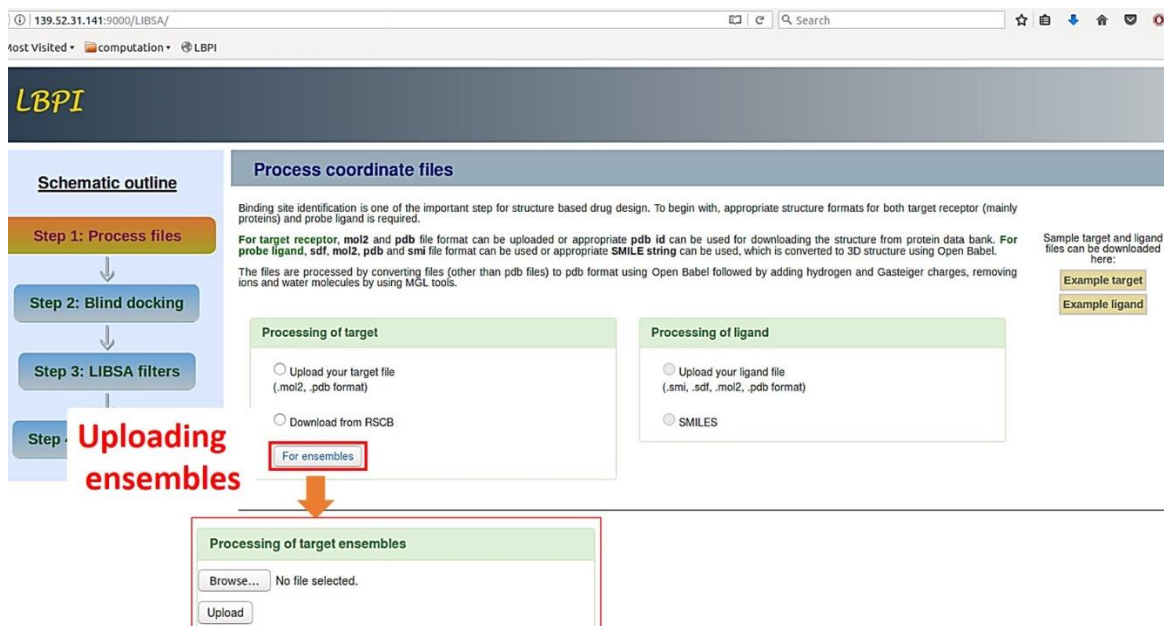


Figure 4.14. Uploading multiple conformations to calculate SNR by averaging over an ensemble.

4.3 Implementation of filters (option 2)

4.3.1 File preparation

In this section, a target pdb or mol2 file used in molecular docking or obtained in a pdbqt format from AutoDock is accepted and processed as a LIBSA readable format. For the probe, a dlq file from AutoDock is accepted and similarly processed. The processed dlq file includes conformation of the probe from docking with its corresponding binding energy.

Figure 4.15 shows a snapshot of this section from the *Web interface*. This section then proceeds to the utilization of filters as the module for blind docking is skipped in this section.

The screenshot displays the LBPI web interface. On the left, a 'Schematic outline' shows 'Step 1: Process files' and 'Step 2: Blind docking'. The main content area is titled 'Process docking files' and includes a description: 'The key working file for LIBSA is the receptor file used for docking and docking log file which consists of ligand conformations of number of independent runs performed during docking. Both of the files are then processed as LIBSA working file format for executing LIBSA.' Below this is a section for 'Uploading files for LIBSA' with two rows: 'Receptor file (.pdb format)' and 'Docking file (.dlg format)'. Each row has a 'Browse...' button, the text 'No file selected.', and an 'Upload' button. A red box highlights this entire section. To the right, an 'Example files' section contains a blue box with a downward arrow and text: 'Sample DLG and receptor files can be downloaded here:'. Below this text are two links: 'Example DLG' and 'Example receptor'. A dashed red arrow points from the 'Example DLG' link to the 'Docking file (.dlg format)' field in the red box.

Figure 4.15. Options for submitting files of the receptor and log files from docking is highlighted in the red box with example files highlighted in the blue box.

4.3.2 LIBSA filters

LIBSA filters in this section is similar to the one described in section 4.2.3 and shown in Figure 4.11. The filters selected, parameters provided and histograms prepared through the *Web interface* in this section are as explained in section 4.2.3 for the identification of appropriate binding site and quantification of its binding site preference.

4.4 Summary

We have shown that LBPI has been implemented in 2 ways to enable flexibility. Option 1 is mainly targeted to non-experts who would like to focus on the identification of binding sites using LIBSA protocols without the knowing technical details of molecular docking or LIBSA filters. However, the advantage of choosing option 2 over option 1 is that users

(especially experts) can directly focus on the application of LIBSA filters by skipping molecular docking. Also, users can choose a wide range of customizable parameters for molecular docking to obtain log files required for LIBSA filters instead of conducting docking with AutoDock via LBPI. However, this section is feasible to experts who are acquainted with molecular docking. Our implementation of LBPI thus allows users the flexibility necessary to select the modules on the basis of their preference.

CHAPTER 5

APPLICATIONS

5.1 Re-docking with LBPI

We applied LBPI on three protein-ligand complexes (mentioned as cases in Table 1 and throughout this section) taken from the protein data bank for re-identifying the known binding site of their co-crystal ligand. For each case, first, we used LBPI for extracting the protein structure from the PDB and then uploaded the corresponding ligand coordinate extracted from the complex structure utilizing Visual Molecular Dynamics (VMD), a graphical interface (67). The information about proteins and ligands used are shown in Table 1 and Figure 5.1. Next, we provided parameters to LBPI for molecular docking (mentioned in Table 1) and LIBSA filters (default parameters were used). The submission of these parameters to LBPI executed molecular docking and LIBSA filters automatically. The results obtained and the site identified via LBPI are shown in Table 1 and Figure 5.2. For cases 1 and 2, we identified a single binding region of interest (green region) for the corresponding ligands and compared with it with the known binding site (the region with the ligand shown in red). The identified region agrees well with the known binding site. Also, the value of the computed SNR for these regions was greater than 1 (Table 1, Figure 5.2), illustrating that LBPI successfully identified the known binding sites as preferred binding sites for cases 1 and 2. For case 3, we identified two potential binding sites (green and blue regions in panel C, Figure 5.2) and compared their SNR values. As the SNR for the green region is higher than the blue region,

we conclude that the green colored binding region (Figure 5.2, panel C) is the preferred binding site, which remarkably coincided with the known binding site.

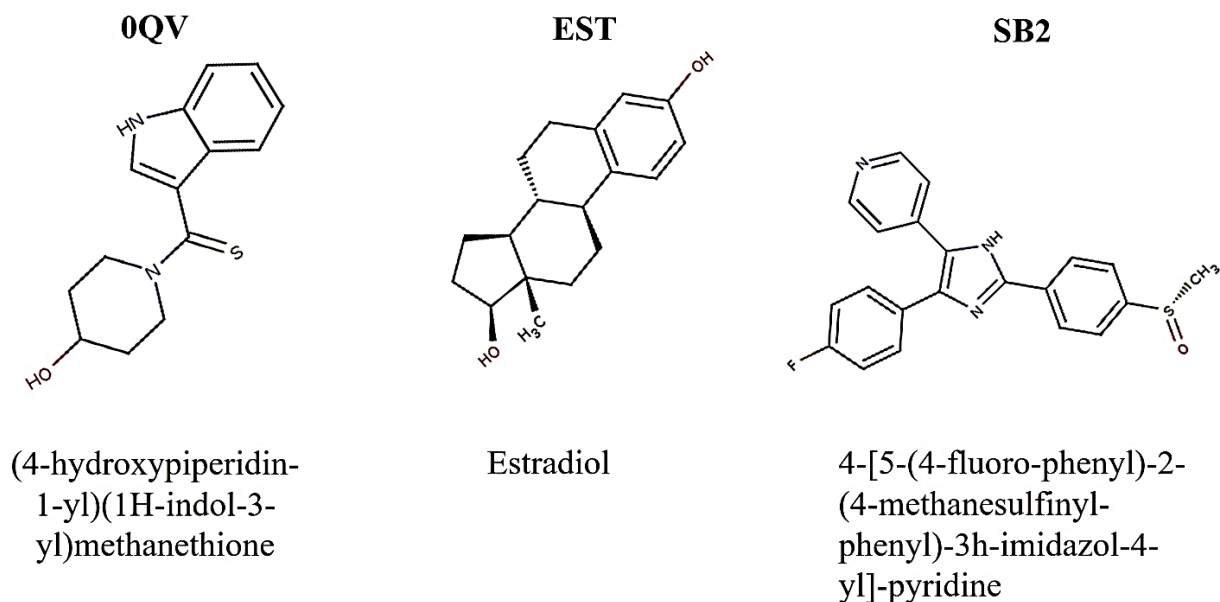


Figure 5.1. Chemical structures of ligand 0QV, SB2, and EST from PDB 4EPW, 1QKT and 1A9U respectively.

To put it briefly, by directly submitting coordinate files of the proteins and known co-crystallized ligands and other parameters to LBPI, we were able to identify the correct binding site for all three cases without reverting to manual procedures or knowing the technical details of the molecular docking protocol or LIBSA filters.

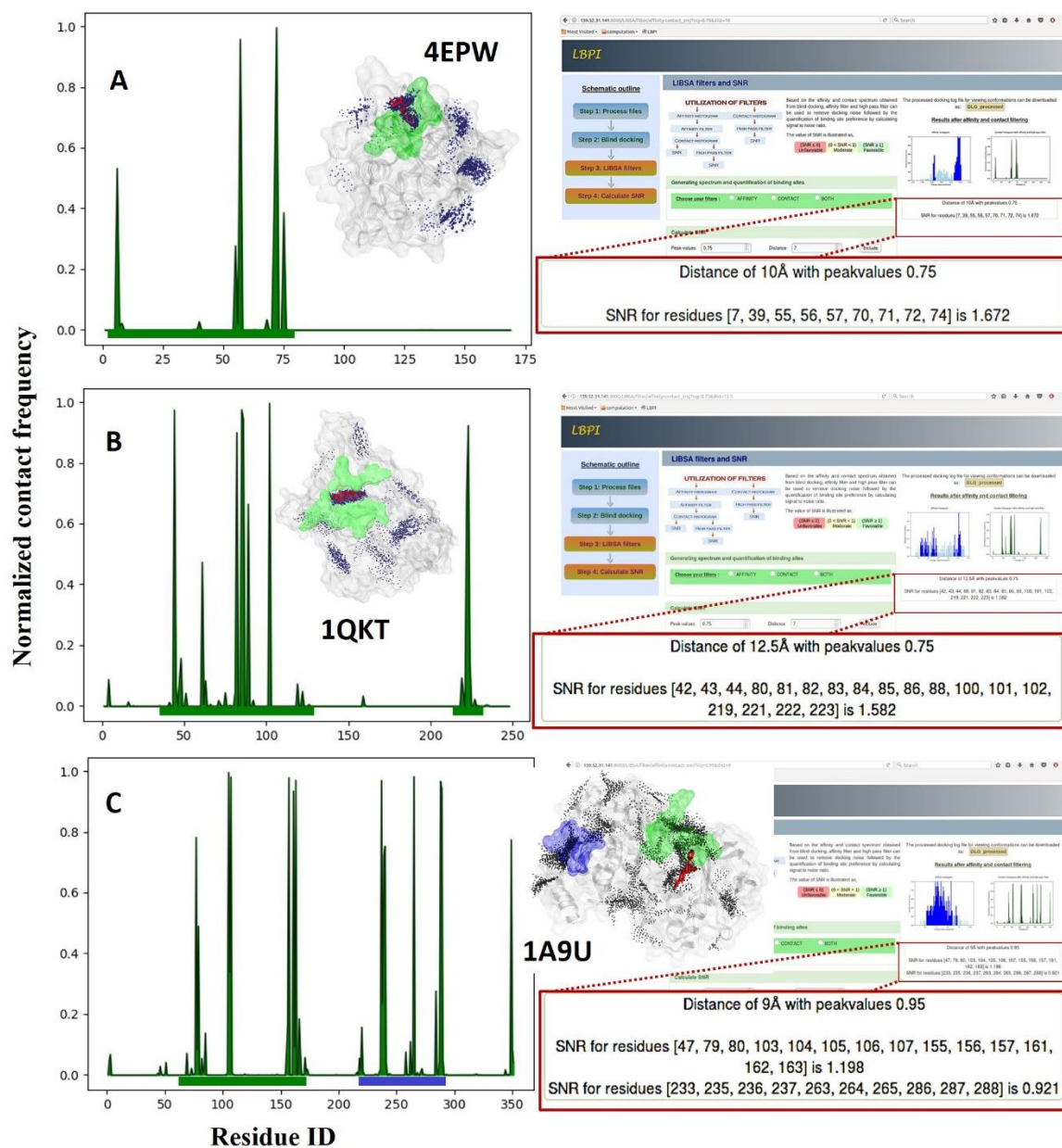


Figure 5.2. (A-C) Contact histograms after applying affinity and highpass filter to remove docking noise (left). The target structure shown in grey surface with the co-crystallized ligand is represented in red and the docked poses in black dots; SNR for the region of interest is shown in the right panel and is highlighted in green on the protein surface, which coincides with the co-crystal site.

Table 1. Summary of re-docking results with LBPI

Case	PDB id	Co-crystallized ligand	Protein	Docking parameters (E.E., N.G.)*	SNR (predicted binding site residues)
1	4EPW	0QV (4-hydroxypiperidin-1-yl)(1H-indol-3-yl)methanethione	K-Ras	2.5*10 ⁶ 2.7*10 ⁴	1.672 (7, 39, 55, 56, 57, 70, 71, 72, 74)
2	1QKT	EST (estradiol)	Nuclear-estrogen receptor	2.5*10 ⁵ 2.7*10 ⁴	1.582 (42, 43, 44, 80, 81, 82, 83, 84, 86, 85, 88, 100, 101, 102, 219, 221, 223)
3	1A9U	SB2 4-[5-(4-fluorophenyl)-2-(4-methanesulfinylphenyl)-3h-imidazol-4-yl]-pyridine	Kinase protein (P38)	2.5*10 ⁵ 2.7*10 ⁴	1.198 (47, 79, 80, 103, 104, 105, 106, 107, 155, 156, 157, 161, 162, 163) 0.921 (233, 235, 236, 237, 263, 264, 265, 286, 287, 288)

* E.E. = energy evaluations and N.G. = number of generations

5.2 Cross-docking with LBPI

We demonstrated an additional use of LBPI by performing cross-blind docking of ligand NIL (obtained from 3GP0) to a kinase-protein, P38, obtained after removing the co-crystals in 1A9U. The coordinate files were submitted to LBPI as mentioned earlier, and parameter values were provided to LBPI in a similar way as case 2, shown in Table 1. The result obtained after executing molecular docking and application of LIBSA filters through LBPI is shown in Figure 5.3. The calculation identified two regions as binding regions of interest, shown in green and blue regions on the protein structure in Figure 5.3, along with the

SNR. Since SNR for the green region is higher, it is considered a preferred binding site as compared to the blue region.

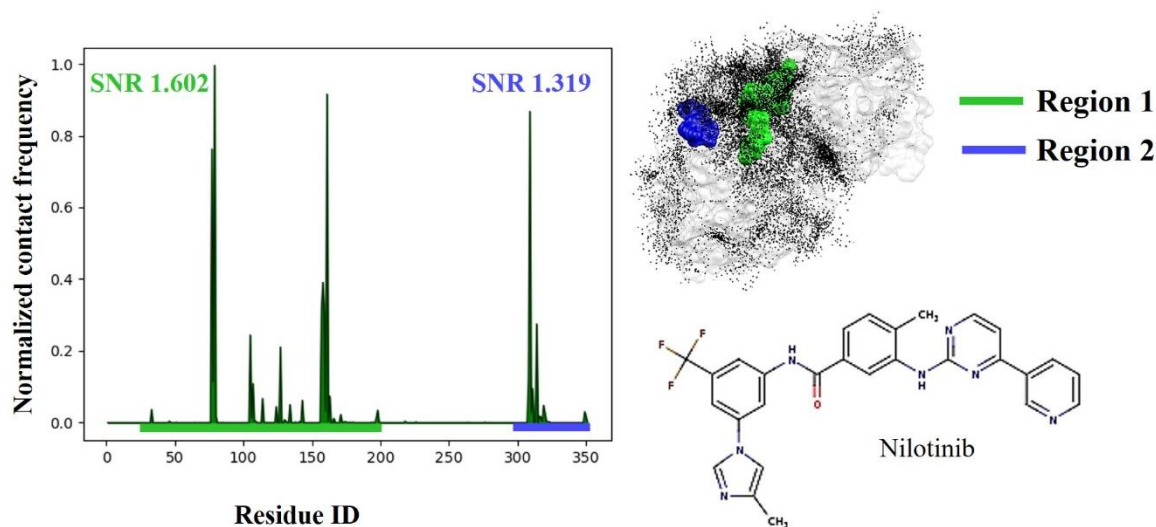


Figure 5.3. Contact histogram after affinity and high pass filtering of docking results for NIL (3GP0) to protein kinase P38 (1A9U).

The crystal binding site of ligand NIL in 3GP0 is similar to the crystal binding site of ligand SB2 in 1A9U (68, 69). On comparing the preferred binding region obtained from LBPI with the known crystal binding site, we could say that LBPI was able to locate the correct binding site from cross docking of NIL on P38.

Further, for a rigorous test, we took four different ligands 0QR (N-(6-aminopyridin-2-yl)-4-fluorobenzenesulfonamide), 0QW ((2-hydroxyphenyl)(pyrrolidin-1-yl) methanethione), 0QX (2-(1H-indol-3-ylmethyl)-1H-imidazo[4,5-c] pyridine) and BZI (benzimidazole) from PDB- 4EPX, 4EPT, 4EPV and 4DSU respectively (Figure 5.4), and cross-docked them to GDP-bound K-Ras using 4EPW crystal structure, after removing all co-crystals. As discussed

in earlier sections, the coordinate files and parameters were provided to LBPI. The parameters values used were similar to case 1 shown in Table 1.

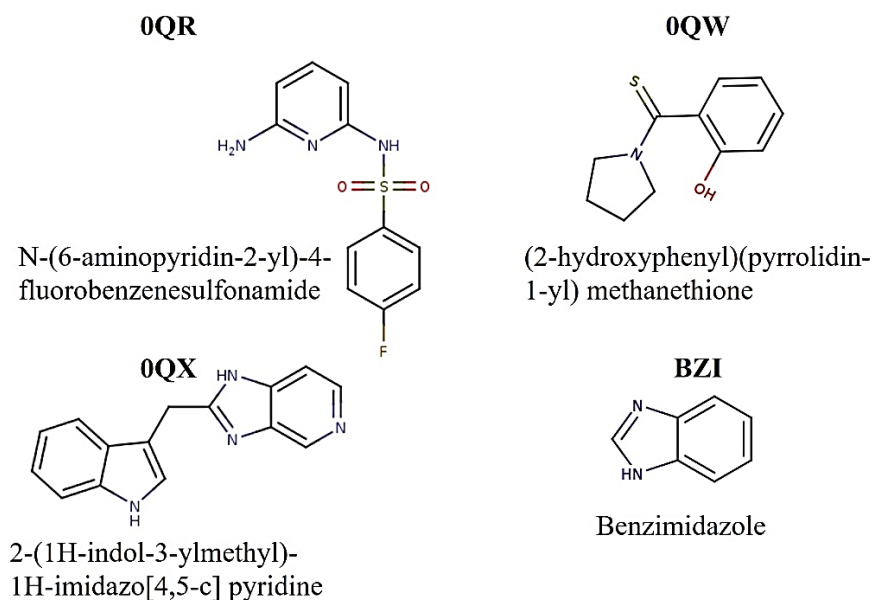


Figure 5.4. Chemical structures of ligands 0QR, 0QW, 0QX and BZI from PDB 4EPX, 4EPT, 4EPV and 4DSU respectively

Docked poses obtained from docking through LBPI were rendered with VMD for each ligand and shown in Figure 5.5. The histograms prepared by LBPI after executing LIBSA filters are shown in Figure 5.6. The ligands we have chosen for cross docking are known to bind at the same site as the co-crystallized ligand 0QV in 4EPW (70, 71). When cross-docked, LBPI identified two regions of interest, region 1 (pink) and region 2 (cyan) for each ligand, except for BZI, which has only one predicted site, region 1 (pink). After computing the SNR values (Figure 5.6) for the identified regions through LBPI and comparing them, the ligands could be ranked as BZI > 0QX > 0QW > 0QR for their binding preference to region 1, and 0QW > 0QX > 0QR > BZI for their binding preference to region 2. LBPI correctly identified

region 1 as the preferred binding site for 0QX and BZI but failed for ligands 0QR and 0QW, as the SNR for region 2 was higher than the SNR for region 1 for these ligands.

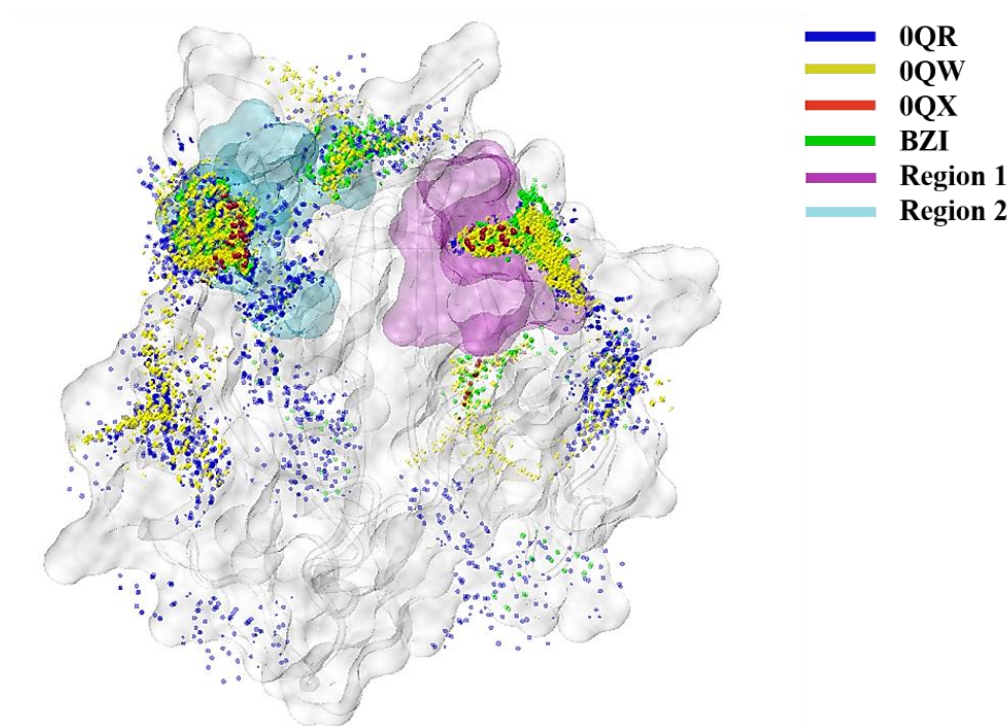


Figure 5.5. Docked poses obtained from AutoDock for 0QR (blue), 0QW (yellow), 0QX (red) and BZI (green). Regions 1 and 2 are represented by pink and cyan color respectively and indicate the binding regions of interest for calculating SNR for each ligand.

Since the ligands used for cross docking were selected from PDB files of the same protein but of different conformations, each protein conformation might differ in terms of the size and volume of the correct binding site. Thus we checked if re-docking of 0QW and 0QR to their parent protein structure through LBPI could identify the appropriate binding site. The procedure for submitting coordinate structures and parameter values to LBPI was the same as mentioned before. Our results (Figure 5.7) show that only one binding site was identified for 0QW, whereas, once again, two possible sites were identified for 0QR. The computed SNR for

the predicted regions of interest for both 0QR and 0QW suggest that the preferred binding site identified by LBPI coincides with the known binding site for both ligands.

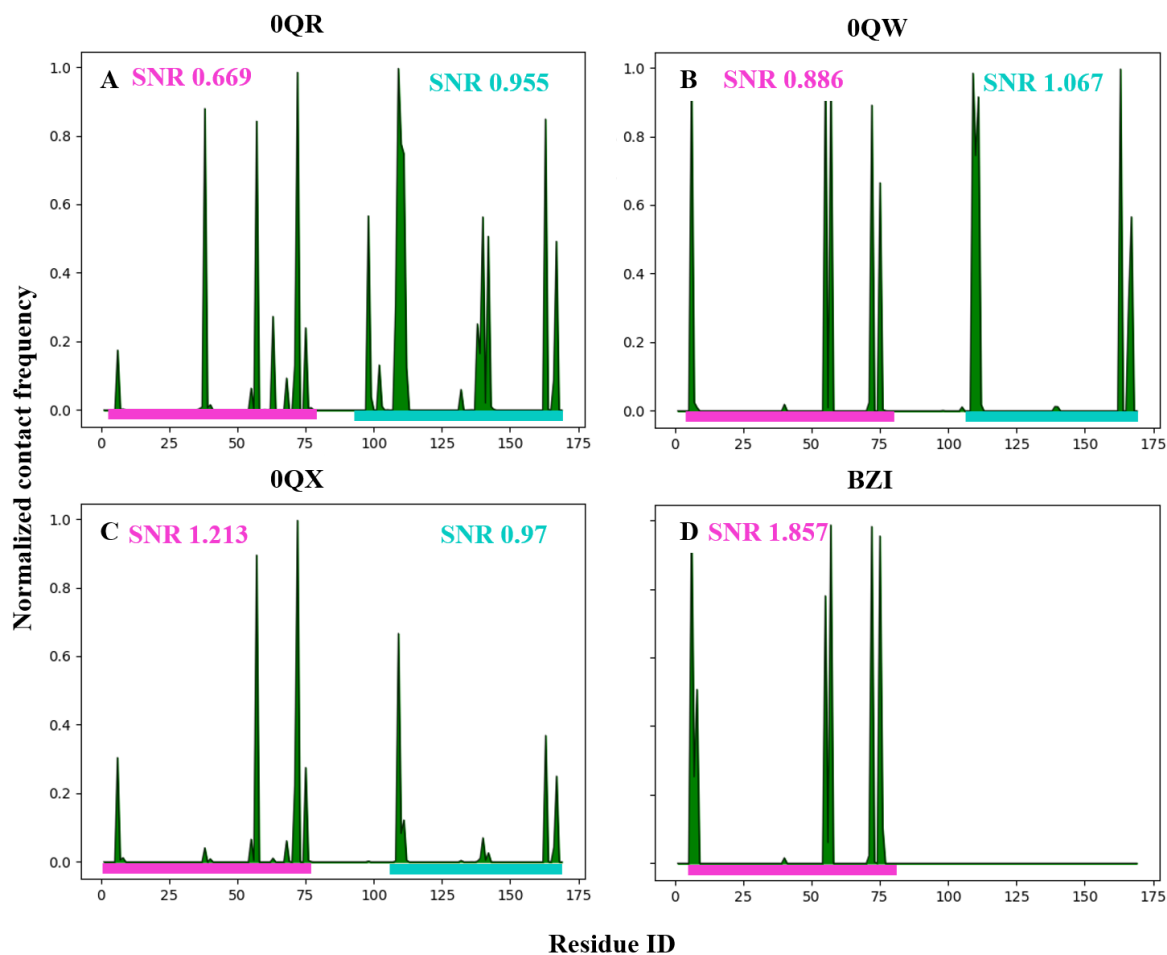


Figure 5.6. Panels A, B, C, and D represent contact histograms for 0QR, 0QW, 0QX and BZI after docking on K-Ras (4EPW) and applying both affinity and high pass filters. SNR computed for region 1 is represented in pink and SNR computed for region 2 is represented in cyan.

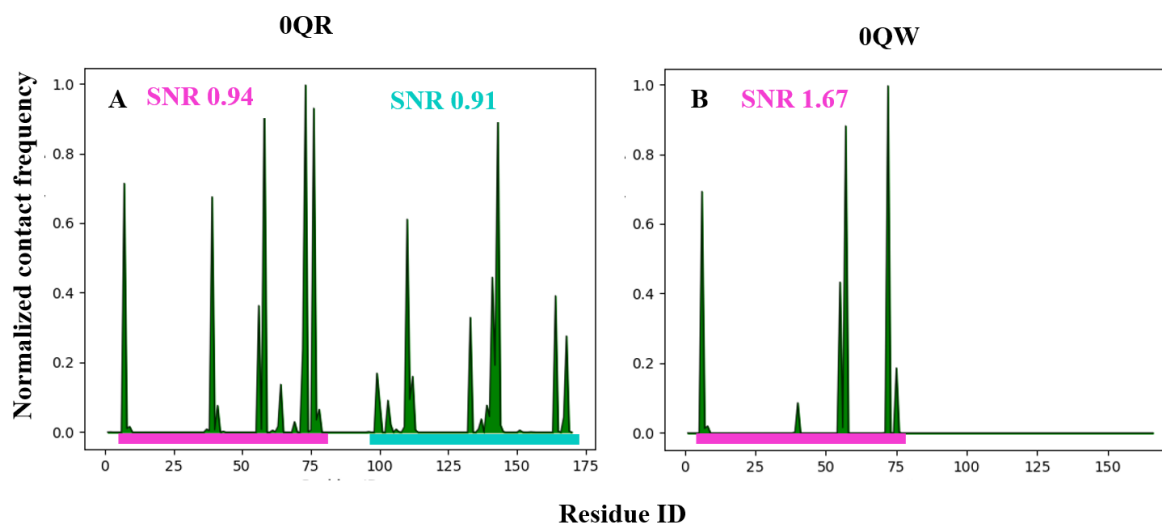


Figure 5.7. Contact histograms for 0QR and 0QW obtained from docking these ligands to the parent protein X-Ray structure. SNR for region 1 and region 2 is represented as pink and blue respectively.

However, the SNR computed for 0QR when docked on a single K-Ras structure of PDB id 4EPW or 4EPX is still less than 1. I believe this is because the docking protocol did not account for the flexibility of the target protein. With the incorporation of receptor flexibility, the conformational space sampled gets improved (72, 73) and thereby the probability of finding the right binding site with better SNR values. To check this, we selected 5 different conformations of K-Ras from an MD simulation for ensemble blind docking of 0QR and 0QW. We submitted the five conformations of K-Ras and that of the ligand to LBPI. Using the default parameters, we executed docking and LIBSA filters for each conformation and averaged the results to obtain a single spectrum. Figure 5.8 and Figure 5.9 show contact histograms for each conformation as well as the single average contact histogram obtained after averaging the contact histograms of each conformation. Visual analysis of the results

shows two probable binding regions in some of the conformations for 0QR but just one for 0Qw. However, the result of the average spectrum suggests a single binding region with SNR > 1 for both 0QW and 0QR. The region agrees well with the known binding site in the X-Ray structure. Hence, we conclude that LBPI can successfully identify the preferred binding site of a protein with single conformation or multiple conformations and a diverse set of ligands without manually performing molecular docking and LIBSA analysis.

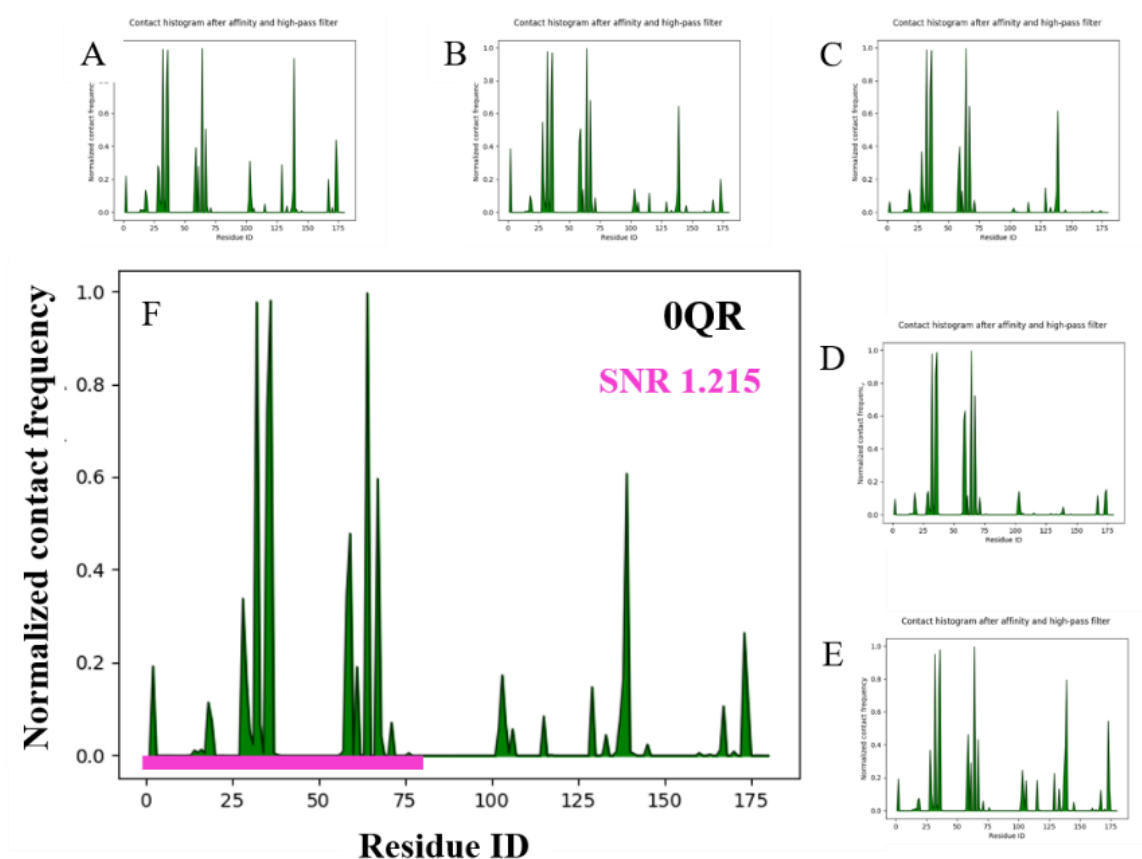


Figure 5.8. Contact histograms after blind docking of 0QR on five K-Ras conformers (A to E) and a single averaged contact histogram (F).

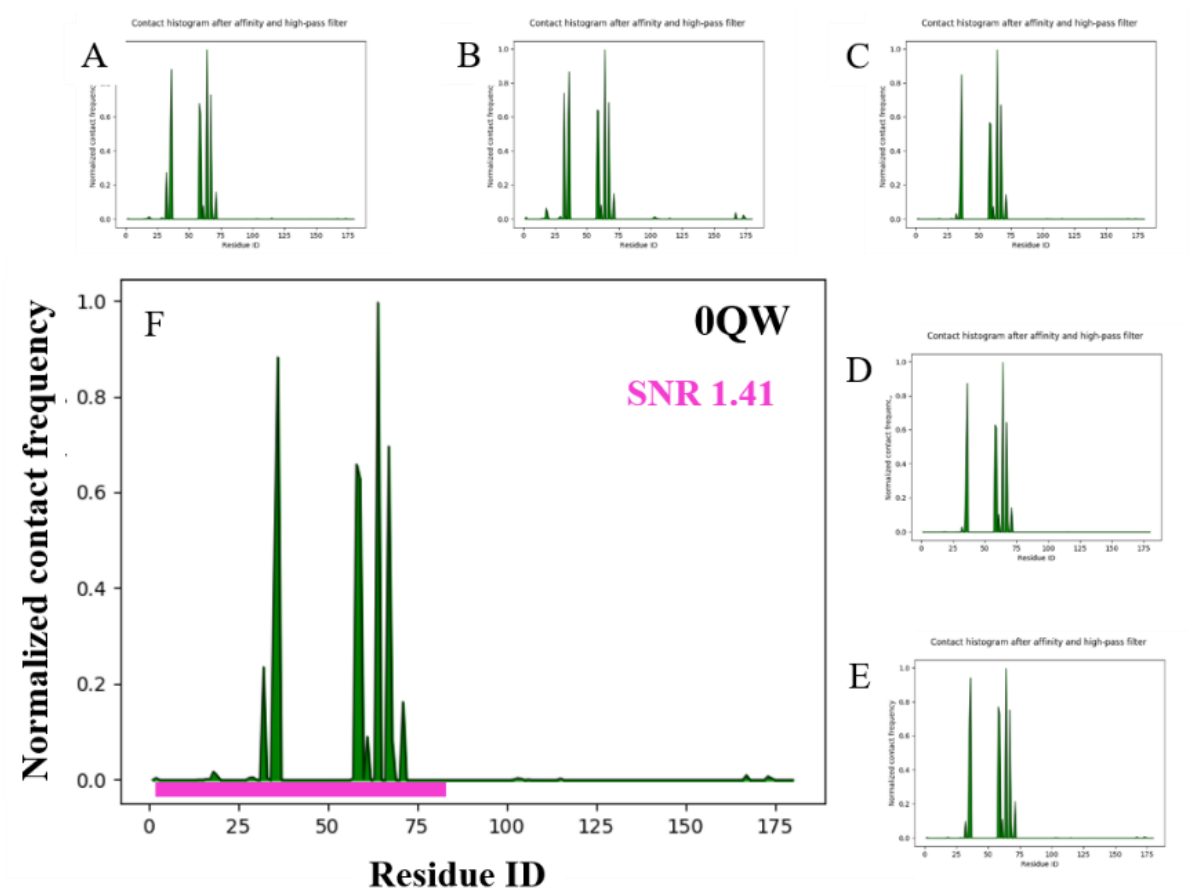


Figure 5.9. Contact histograms after blind docking of 0QW on five K-Ras conformers (A to E) and a single averaged contact histogram (F).

5.3 Identification of binding sites of test ligands with LBPI

We used LBPI on three test ligands promazine, E22, and V2, for the identification of their potential preferred binding site on the K-Ras protein structure PDB id 4EPW. The experimental analysis in our lab suggested that all three ligands may bind to the same site that coincides with a known binding site for ligand 0QV in PDB 4EPW. We utilized LBPI to extract K-Ras structure from 4EPW and submitted our test ligands through the *Web interface*. Default parameters were provided to LBPI to conduct docking and LIBSA analysis. The results (Figure

5.10) show that the predicted binding site of promazine agrees well with our experimental analysis whereas the predicted binding site for E22 and V2 differs from the expected site. This could be due to the limitation of using a single X-Ray structure, as discussed in the previous section. To check if using multiple conformations would help, we utilized five conformations of K-Ras obtained from an MD simulation. The coordinate files were uploaded and default parameters were submitted via LBPI for executing molecular docking and application of LIBSA filters.

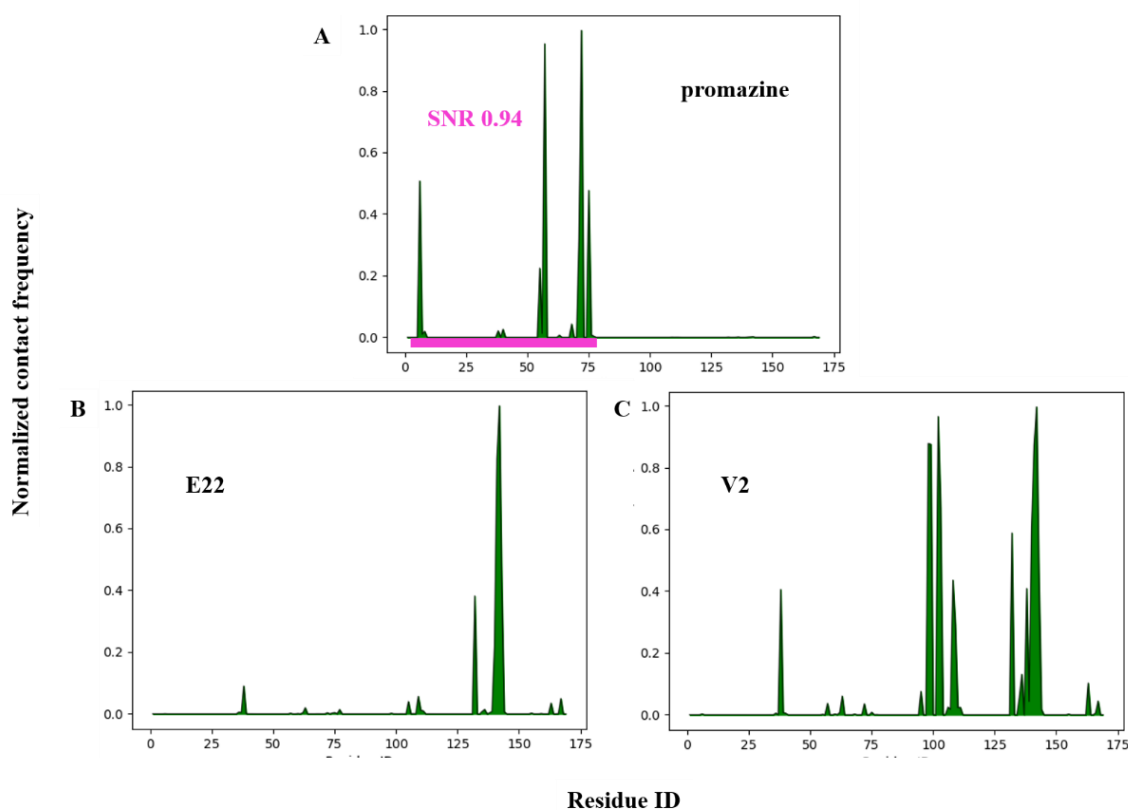


Figure 5.10. Contact histograms for three test ligands promazine (A), E22 (B) and V2 (C) docked on X-Ray structure 4EPW of K-Ras.

Visual analysis of the results for each conformation (Figure 5.11 and Figure 5.12) indicated two distinct binding sites for both E22 and V2, depending on the conformation of K-Ras used for docking. However, when averaged, a single region was identified as a preferred binding site for E22 and V2 with a computed SNR of 1.165 and 1.339, respectively. The SNR >1 indicates that this is the potential binding site for both ligands, which agrees well with our experimental results. In summary, LBPI was able to identify potential binding sites for our test ligands, particularly if multiple target conformations are used.

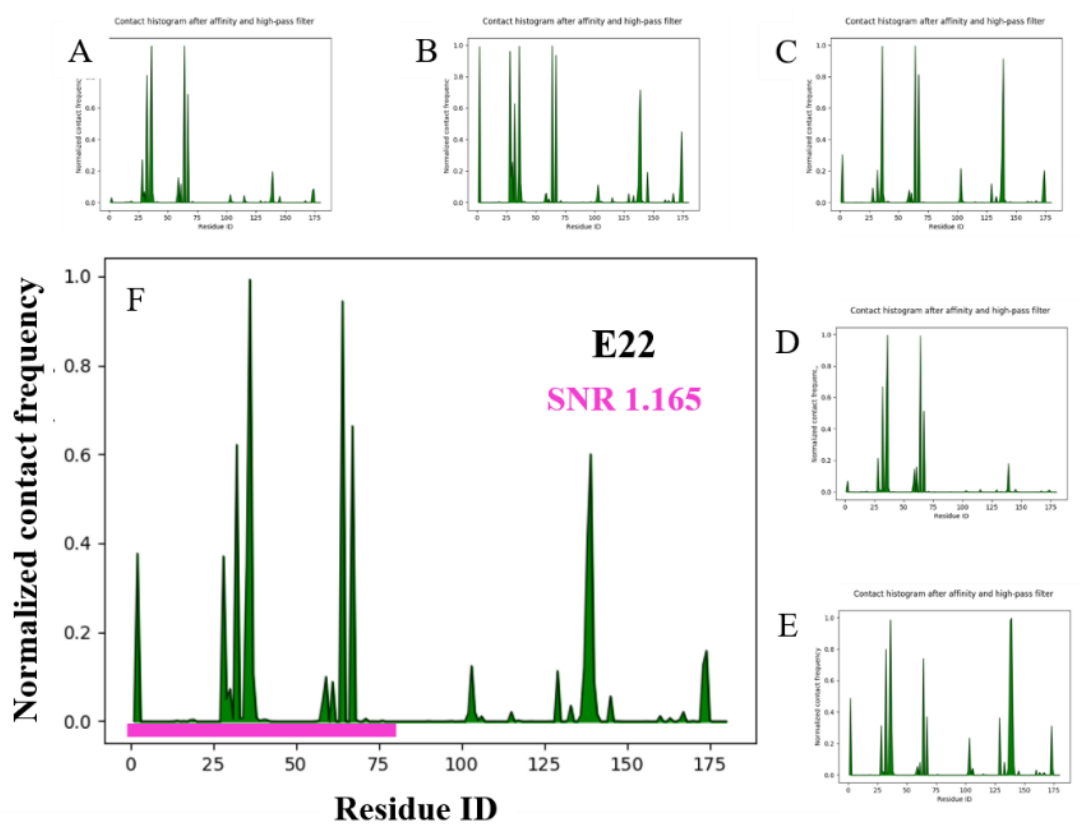


Figure 5.11. Contact histograms after blind docking of E22 on five K-Ras conformers (A to E) and a single averaged contact histogram (F).

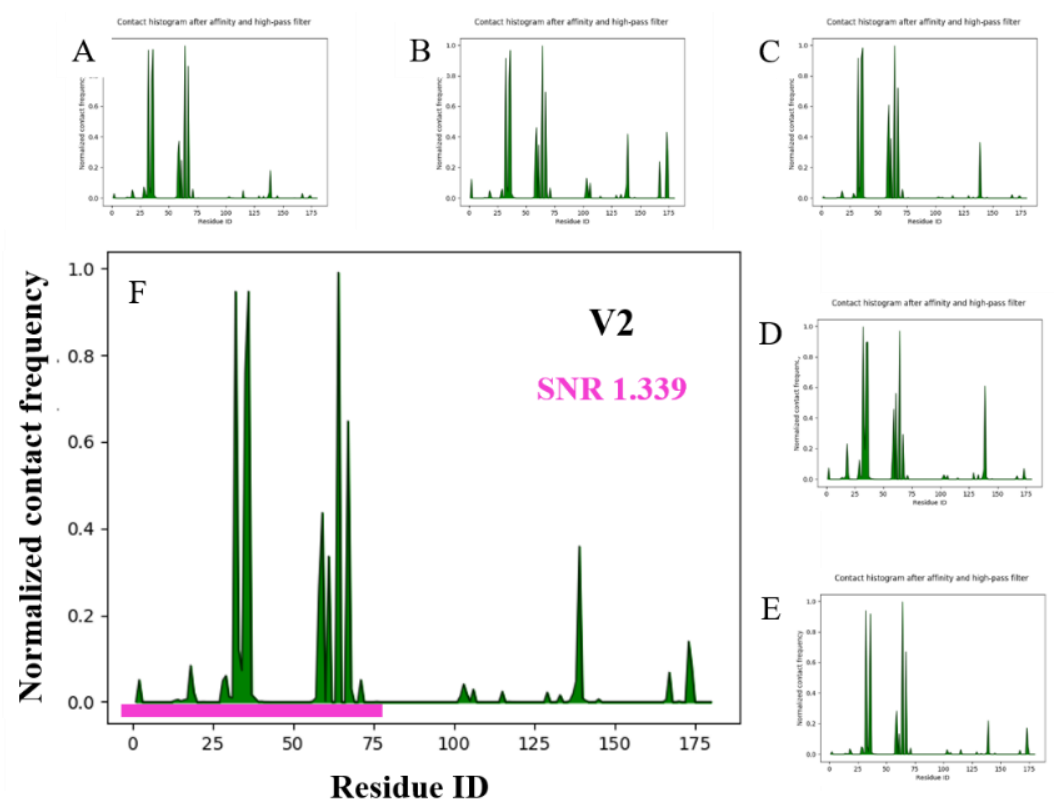


Figure 5.12. Contact histograms after blind docking of V2 on five K-Ras conformers (A to E) and a single averaged contact histogram (F).

5.4 Identification of the binding preferences of estrogens with LBPI

Estrogens are hormones responsible for the development and balance of the female reproductive system (74). Estradiol, estrone, and estriol are the most common naturally occurring estrogens (75). Some of the other estrogen compounds including 17 alpha estradiol, ethinyl estradiol and, estradiol valerate are either naturally occurring or artificially synthesized. All of these estrogenic compounds are clinically used as steroidal estrogens (75). These compounds bind to estrogen receptor proteins and induce conformational changes in the protein. From the data available in the literature, the relative binding affinity of these

compounds (Table 2) for estrogen receptor alpha is ethinyl estradiol > estradiol > estrone > 17-alpha estradiol > estriol > estradiol valerate (74-77) To study the predicted binding preference of these estrogen compounds using LBPI, we docked these compounds on the structure of estrogen receptor protein extracted from pdb id 1QKT, after removing its co-crystal through LBPI followed by the use of LIBSA filters to identify a potential binding site for each compound. 2.5×10^6 energy evaluations and 27000 number of generations were used for docking and default parameters were used for LIBSA filters. The results are shown in Figure 5.14 and Table 2. Estradiol and 17-alpha estradiol are stereoisomers and differ in their binding affinity by two folds, with estradiol binding more strongly to estrogen receptor (78). However, the results obtained from LBPI show minor differences in binding affinity of these compounds (Table 2). Similarly, for other estrogenic compounds, identical binding pattern was seen with minor differences due to the change in very small group present in the 17th position of estradiol, except for estradiol valerate. Estradiol valerate, which possesses valerate group was larger than any groups present at the 17th position of estradiol and has a higher number of rotatable bonds and thus the difference observed in its binding affinity was notable.

Table 2. Relative binding affinity for estrogen compounds from literature and summary of results obtained from LBPI.

Compounds	Relative binding affinity	Affinity from LBPI	Residues identified by LBPI	SNR from LBPI
Estradiol	1	-9.6	42, 43, 44, 83, 84, 85, 86, 88, 100, 101, 102	1.581
Estradiol 17-alpha	0.58	-9.2	42, 43, 44, 80, 81, 83, 84, 85, 86, 88, 100, 101, 102	1.420
Estradiol valerate	0.02	-6.8	2, 3, 4, 59, 60, 61, 62, 63 219, 221, 222, 223, 224	1.115 1.136
Estriol	0.14	-9.3	42, 43, 44, 80, 81, 82, 83, 84, 85, 86, 88, 100, 101, 102, 219, 220, 221, 222, 223	1.733
Ethinyl estradiol	2.33	-9.9	42, 43, 44, 80, 81, 83, 84, 85, 86, 88, 100, 101, 102, 220, 221, 222	1.332
Estrone	0.6	-10.15	42, 43, 44, 80, 81, 83, 84, 85, 86, 88, 100, 101, 102	1.455

The results (Table 2, Figure 5.14) obtained from this study suggests that through LBPI energy values obtained were not significantly different for estrogen compounds. This is because the compounds are structurally very similar. On the other hand, estradiol valerate has larger valerate group attached at the 17th position of estradiol and thus the difference observed was significant. This demonstrates the limitation of the docking program and suggests that to study substantial changes in binding affinity of a compound and its derivatives, there should be a significant difference in their structures. However, LBPI was able to identify the correct binding site for estrogen compounds except for estradiol valerate (Table 2). For estradiol

valerate, two binding sites were observed out of which the most preferred binding site based on calculated SNR through LBPI was able to capture some of the surrounding residues of the correct binding site.

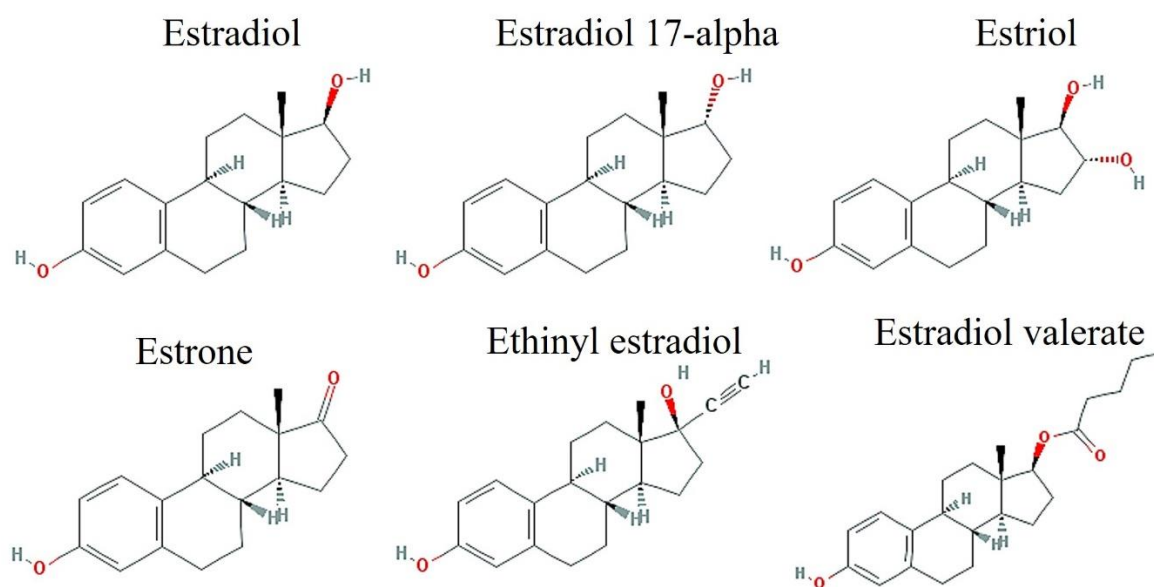


Figure 5.13. Chemical structure of estrogen compounds

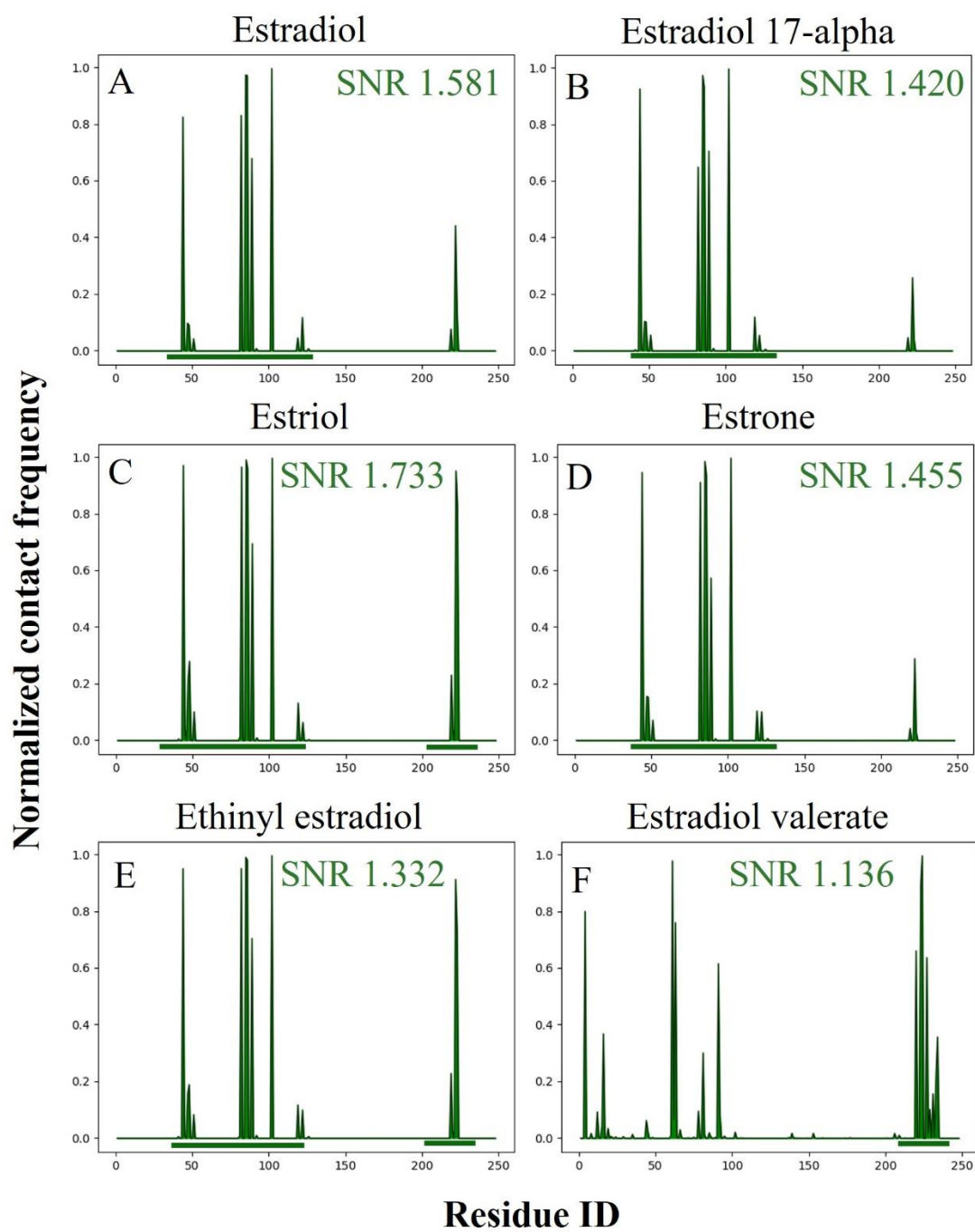


Figure 5.14. Contact histograms obtained after applying affinity and high-pass filters to remove docking noise for estrogen compounds to estrogen receptor (1QKT).

5.5 Identification of allosteric ligand binding site for sorafenib on p38 through LBPI

Compounds targeting the ATP binding site for modulating the function of kinase proteins are identified as inhibitors e.g. binding of SB2 to P38 as shown in Figure 5.2 case 3 (69). Research has also been performed in investigating compounds that alter the function of proteins through allosteric modulation without directly binding to the ATP binding region. Sorafenib has been identified as a notable allosteric inhibitor that binds to the region immediately next to the ATP binding region of kinase protein and modulates the function (79, 80). To identify a potential binding site for sorafenib in kinase protein (P38) through LBPI, we extracted the structure of P38 from pdb id 3HEG and performed docking with its co-crystal sorafenib (BAX) utilizing LBPI. 2.5×10^6 energy evaluations and 27000 number of generations were used for docking and default parameters were used for LIBSA filters. Three binding sites were identified through LBPI represented as blue, green and pink (Figure 5.15), of which the blue region coincides with the known binding site and it is the favorable binding site as it has SNR greater than 1. Hence, LBPI was able to identify the correct binding site as the preferred binding site for sorafenib.

When we compare the results obtained from LBPI for identifying potential binding sites for SB2 on P38 (Figure 5.2, case 3) and sorafenib on P38 (Figure 5.15), LBPI was able to correctly identify the catalytic binding site for SB2 and allosteric binding site for sorafenib respectively. We can thus conclude that, in the long run LBPI can be used for identifying potential binding sites be it catalytic or allosteric for any drug targets.

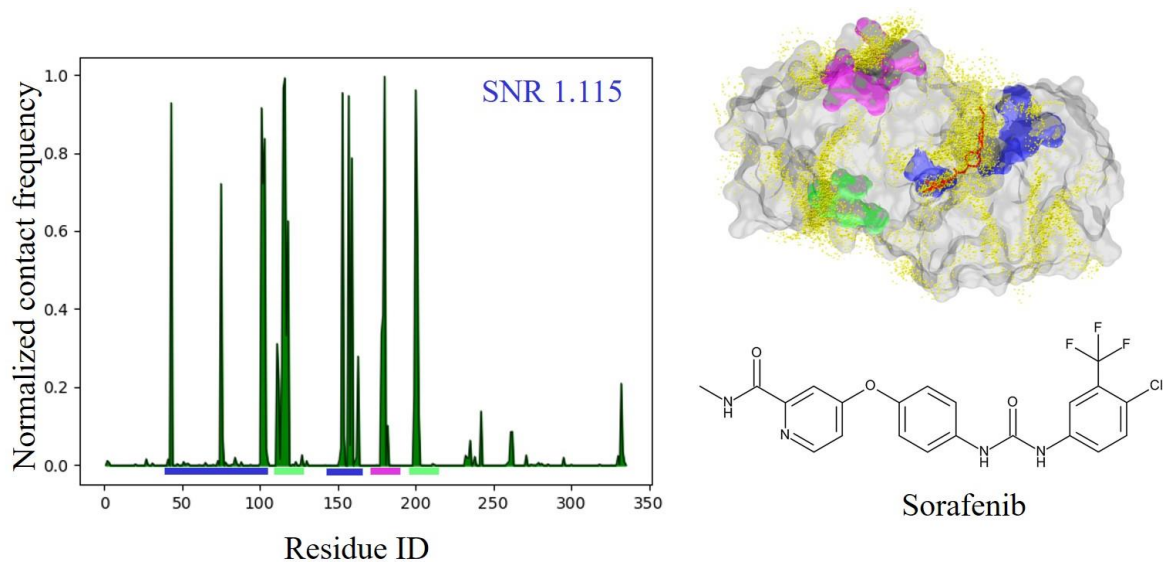


Figure 5.15. Contact histogram after affinity and high pass filtering of docking results for sorafenib to its protein kinase P38 (3HEG) shown in left panel, structure of P38 with docked poses in yellow, three identified region from LBPI, co-crystal (sorafenib) in red, and chemical structure of sorafenib in right panel.

CHAPTER 6

CONCLUSIONS AND FUTURE DIRECTIONS

Computational approaches introduced for identifying allosteric ligand binding sites e.g. LIBSA requires expert skills to be properly executed and hence limits its users. Building a computational pipeline as a web interface by automating protocols will bridge the gap and help users focus on the biological applications. We have presented a *Web interface*, LBPI, which is built on Django, a Python framework, supplemented by a Python *Wrapper*, for automating the protocols of LIBSA, executing it and monitoring the results. LBPI is implemented in two ways to allow users flexibility. Option 1 is mainly targeted to inexperienced users for applying LIBSA protocols without undergoing technical details for docking and LIBSA filters. On the other hand, option 2 is mainly targeted to users who are expert in molecular docking and can directly focus on utilizing LIBSA filters by skipping molecular docking through the *Web interface*. The user manual and example files provided through the *Web interface* will help users to get better acquainted with the workings of LBPI. Through LBPI, by submitting structure files for target and probe ligand of appropriate file format, providing few customizable parameters and selecting modules for LIBSA filters will help in the effective identification of binding sites without understanding technical details.

For demonstration purposes of LBPI, we have used three known protein-ligand complexes, and directly submitted coordinate files of those proteins and their co-crystallized ligands and other parameter values to LBPI to successfully identify their correct binding sites without manual procedures. Further, from cross docking of a diverse set of ligands on different

target conformations obtained from the protein databank, we showed that there are cases where LBPI fails to identify the correct binding site. This was largely due to the limitation of using a single X-Ray structure for the target, which might be unsuitable in terms of the size and volume of the binding pocket. This suggests that for effective identification of a binding site, a single structure may not be sufficient. Hence, we have demonstrated that LBPI was able to correctly identify predicted binding sites for each ligand when multiple target conformations from MD simulation are used. We have also shown that LBPI helps rank ligands based on SNR calculated for a preferred site. We have also shown the use of LBPI for a successful identification of preferred binding sites (based on experimental analysis done in our lab) for test ligands on a single X-Ray structure or multiple target conformations. Further, we utilized LBPI to study the predicted binding preference of different estrogenic compounds. It is due to the limitation of docking program which failed to show the substantial change in the binding affinity as the estrogenic compounds used in LBPI had very minor change in the structures. However, it was able to locate the correct binding site for each estrogen compounds, which is the major goal of LBPI. Finally, LBPI was able to correctly identify catalytic and allosteric site for catalytic and allosteric inhibitor respectively, through which we can conclude that LBPI is capable of identifying potential binding site of a ligand based on preference of binding to a site.

We have demonstrated the successful use of LIBSA through LBPI for diverse set of drugs but only a few target structures. However, Hocker et al. in the original paper of LIBSA have demonstrated its application for a more diverse set of druggable targets (49). Thus we

expect that LBPI will be utilized on even more challenging drug targets for the identification of potential binding site(s). The time required for executing LIBSA filters and performing docking through LBPI is comparable to standalone LIBSA. However, the efforts required to prepare structure files, understand technical details for performing docking and executing LIBSA algorithms is reduced significantly through LBPI. The demonstration for the performance of LBPI on a set of examples was done locally in our lab. Our next goal is to work on making it available through the Internet to make it easily accessible to both experts and non-experts and reduce the significant learning curve for understanding the technical details and focus more on biological applications. We can facilitate the calculation of SNR based on residues provided by the users. Next, we can add the ensemble approach in option 2 for focusing the use of LIBSA filters for ensemble. Furthermore, we have a clear vision of adding pMD protocols (developed in our lab) to LBPI, which consist of MD simulation in presence of probe(s) to assess the propensity of a site in a wide variety soluble or membrane proteins (48).

REFERENCES

1. JP Hughes, S Rees, SB Kalindjian, KL Philpott. Principles of early drug discovery. *British journal of pharmacology*. 2011;162(6):1239-49.
2. LM Mayr, D Bojanic. Novel trends in high-throughput screening. *Current opinion in pharmacology*. 2009;9(5):580-8.
3. J Drews. Drug discovery: a historical perspective. *Science*. 2000;287(5460):1960-4.
4. T Langer, SD Bryant. Chapter 10 - In Silico Screening: Hit Finding from Database Mining A2 - Wermuth, Camille Georges. *The Practice of Medicinal Chemistry (Third Edition)*. New York: Academic Press; 2008. p. 210-27.
5. JW Scannell, A Blanckley, H Boldon, B Warrington. Diagnosing the decline in pharmaceutical R&D efficiency. *Nature reviews Drug discovery*. 2012;11(3):191-200.
6. RA Hodos, BA Kidd, K Shameer, BP Readhead, JT Dudley. In silico methods for drug repurposing and pharmacology. *Wiley interdisciplinary reviews Systems biology and medicine*. 2016;8(3):186-210.
7. C Liao, M Sitzmann, A Pugliese, MC Nicklaus. Software and resources for computational medicinal chemistry. *Future medicinal chemistry*. 2011;3(8):1057-85.
8. N Singh, G Cheve, DM Ferguson, CR McCurdy. A combined ligand-based and target-based drug design approach for G-protein coupled receptors: application to salvinorin A, a selective kappa opioid receptor agonist. *Journal of computer-aided molecular design*. 2006;20(7-8):471-93.

9. CR Beddell, PJ Goodford, FE Norrington, S Wilkinson, R Wootton. Compounds designed to fit a site of known structure in human haemoglobin. *British journal of pharmacology*. 1976;57(2):201-9.
10. H Kubinyi. Structure-based design of enzyme inhibitors and receptor ligands. *Current opinion in drug discovery & development*. 1998;1(1):4-15.
11. J Blaney. A very short history of structure-based design: how did we get here and where do we need to go? *Journal of computer-aided molecular design*. 2012;26:13-4.
12. G Sliwoski, S Kothiwale, J Meiler, EW Lowe, Jr. Computational methods in drug discovery. *Pharmacological reviews*. 2014;66(1):334-95.
13. HJ Huang, HW Yu, CY Chen, CH Hsu, HY Chen, KJ Lee, FJ Tsai, and CY Chen. Current developments of computer-aided drug design. *Journal of the Taiwan Institute of Chemical Engineers*. November 2010;41(6):623-5.
14. S Costanzi, IG Tikhonova, TK Harden, KA Jacobson. Ligand and structure-based methodologies for the prediction of the activity of G protein-coupled receptor ligands. *Journal of computer-aided molecular design*. 2009;23(11):747-54.
15. V Lounnas, T Ritschel, J Kelder, R McGuire, RP Bywater, N Foloppe. Current progress in Structure-Based Rational Drug Design marks a new mindset in drug discovery. *Computational and structural biotechnology journal*. 2013;5:e201302011.
16. HM Berman, J Westbrook, Z Feng, G Gilliland, TN Bhat, H Weissig, IN Shindyalov, PE Bourne. The Protein Data Bank. *Nucleic acids research*. 2000;28(1):235-42.

17. AB Harvey Lodish, Paul Matsudaira, Chris A. Kaiser, Monty Krieger, Matthew P. Scott, S. Lawrence Zipursky, James Darnell. *Molecular Cell Biology*. New York: Macmillan; 2004.
18. FC Bernstein, TF Koetzle, GJ Williams, EF Meyer, Jr., MD Brice, JR Rodgers, O Kennard, T Shimanouchi, M Tasumi. The Protein Data Bank. A computer-based archival file for macromolecular structures. *European journal of biochemistry*. 1977;80(2):319-24.
19. Gb John C Kendrew, Howard M Dintzis, RG Parrish, Harold Wyckoff, David C Phillips. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*. 1958;181:662-6.
20. H Watson. The stereochemistry of the protein myoglobin. *Prog Stereochem*. 1969;4:5.
21. D Baker, A Sali. Protein structure prediction and structural genomics. *Science*. 2001;294(5540):93-6.
22. JD Durrant, JA McCammon. Molecular dynamics simulations and drug discovery. *BMC biology*. 2011;9:71.
23. A Wolf, KN Kirschner. Principal component and clustering analysis on molecular dynamics data of the ribosomal L11.23S subdomain. *Journal of molecular modeling*. 2013;19(2):539-49.
24. M Seeber, M Cecchini, F Rao, G Settanni, A Caflisch. Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics*. 2007;23(19):2625-7.
25. CA Lipinski, F Lombardo, BW Dominy, PJ Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*. 2001;46(1-3):3-26.

26. P Leeson. Drug discovery: Chemical beauty contest. *Nature*. 2012;481(7382):455-6.
27. Y Wang, J Xiao, TO Suzek, J Zhang, J Wang, SH Bryant. PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic acids research*. 2009;37(Web Server issue):W623-33.
28. JJ Irwin, BK Shoichet. ZINC--a free database of commercially available compounds for virtual screening. *Journal of chemical information and modeling*. 2005;45(1):177-82.
29. J Chen, SJ Swamidass, Y Dou, J Bruand, P Baldi. ChemDB: a public database of small molecules and related chemoinformatics resources. *Bioinformatics*. 2005;21(22):4133-9.
30. ID Kuntz, JM Blaney, SJ Oatley, R Langridge, TE Ferrin. A geometric approach to macromolecule-ligand interactions. *Journal of molecular biology*. 1982;161(2):269-88.
31. XY Meng, HX Zhang, M Mezei, M Cui. Molecular docking: a powerful approach for structure-based drug discovery. *Current computer-aided drug design*. 2011;7(2):146-57.
32. GM Morris, R Huey, W Lindstrom, MF Sanner, RK Belew, DS Goodsell, AJ Olson. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry*. 2009;30(16):2785-91.
33. GM Morris, DS Goodsell, RS Halliday, R Huey, WE Hart, RK Belew, AJ Olson. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of computational chemistry*. 1998;19(14):1639-62.
34. G Jones, P Willett, RC Glen, AR Leach and R Taylor. Development and validation of a genetic algorithm for flexible docking. *Journal of molecular biology*. 1997;267:727-48.
35. M Rarey, B Kramer, T Lengauer, and G Klebe. A fast flexible docking method using an incremental construction algorithm. *Journal of molecular biology*. 1996;261:470-89.

36. TJ Ewing, S Makino, AG Skillman, ID Kuntz. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *Journal of computer-aided molecular design*. 2001;15(5):411-28.
37. T Schlick. *Molecular modeling and simulation: an interdisciplinary guide: an interdisciplinary guide*: Springer science & business media; 2010.
38. M McCarthy, P Prakash, AA Gorfe. Computational allosteric ligand binding site identification on Ras proteins. *Acta biochimica et biophysica Sinica*. 2016;48(1):3-10.
39. RS Bohacek, C McMartin, WC Guida. The art and practice of structure-based drug design: a molecular modeling perspective. *Medicinal research reviews*. 1996;16(1):3-50.
40. AK Grover. Use of allosteric targets in the discovery of safer drugs. *Medical Principles and Practice*. 2013;22:418-26.
41. A Christopoulos. Allosteric binding sites on cell-surface receptors: novel targets for drug discovery. *Nature reviews Drug discovery*. 2002;1(3):198-210.
42. J Monod, J Wyman, JP Changeux. On the Nature of Allosteric Transitions: A Plausible Model. *Journal of molecular biology*. 1965;12:88-118.
43. BJ Grant, S Lukman, HJ Hocker, J Sayyah, JH Brown, JA McCammon, AA Gorfe. Novel allosteric sites on Ras for lead generation. *PloS one*. 2011;6(10):e25711.
44. RA Laskowski, F Gerick, JM Thornton. The structural basis of allosteric regulation in proteins. *FEBS letters*. 2009;583(11):1692-8.
45. R Brenke, D Kozakov, GY Chuang, D Beglov, D Hall, MR Landon, C Mattos, S Vajda. Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics*. 2009;25(5):621-7.

46. P Schmidtke, A Bidon-Chanal, FJ Luque, X Barril. MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics*. 2011;27(23):3276-85.
47. TA Halgren. Identifying and characterizing binding sites and assessing druggability. *Journal of chemical information and modeling*. 2009;49(2):377-89.
48. P Prakash, A Sayyed-Ahmad, AA Gorfe. pMD-Membrane: A Method for Ligand Binding Site Identification in Membrane-Bound Proteins. *PLoS computational biology*. 2015;11(10):e1004469.
49. HJ Hocker, N Rambahal, AA Gorfe. LIBSA--a method for the determination of ligand-binding preference to allosteric sites on receptor ensembles. *Journal of chemical information and modeling*. 2014;54(2):530-8.
50. C Hetényi, and DV Spoel. Efficient docking of peptides to proteins without prior knowledge of the binding site. *Protein science*. 2002;11:1729-37.
51. SF Sousa, PA Fernandes, and MJ Ramos. Protein--ligand docking: current status and future challenges. *Proteins: Structure, Function, and Bioinformatics*. 2006;65:15-26.
52. VY Tanchuk, VO Tanin, AI Vovk, G Poda. A New, Improved Hybrid Scoring Function for Molecular Docking and Scoring Based on AutoDock and AutoDock Vina. *Chemical biology & drug design*. 2016;87(4):618-25.
53. CA Lipinski. Chapter 22 - Compound Properties and Drug Quality A2 - Wermuth, Camille Georges. *The Practice of Medicinal Chemistry (Third Edition)*. New York: Academic Press; 2008. p. 481-90.

54. H van de Waterbeemd, S Rose. Chapter 23 - Quantitative Approaches to Structure–Activity Relationships A2 - Wermuth, Camille Georges. *The Practice of Medicinal Chemistry* (Third Edition). New York: Academic Press; 2008. p. 491-513.
55. Python software foundation. Python programming language.
56. B Ekmekci, CE McAnany, C Mura. An Introduction to Programming for Bioscientists: A Python-Based Primer. *PLoS computational biology*. 2016;12(6):e1004867.
57. K Hinsen. The molecular modeling toolkit: A new approach to molecular simulations. *Journal of Computational Chemistry*. 2000;21(2):79-85.
58. SV Walt, SC Colbert, and G Varoquaux. The NumPy array: a structure for efficient numerical computation. *Computing in science and engineering*. 2011;13:22-30.
59. JD Hunter. Matplotlib: A 2D graphics environment. *Computing in science and engineering*. 2007;9:90-5.
60. Django Project. Django
61. BT Miller, RP Singh, JB Klauda, M Hodošček, BR Brooks, HL Woodcock. CHARMMing: A New, Flexible Web Portal for CHARMM. *Journal of chemical information and modeling*. 2008;48(9):1920-9.
62. ZJ Yao, J Dong, YJ Che, MF Zhu, M Wen, NN Wang, S Wang, AP Lu, DS Cao. TargetNet: a web service for predicting potential drug-target interaction profiling via multi-target SAR models. *Journal of computer-aided molecular design*. 2016;30(5):413-24.
63. J Dong, DS Cao, HY Miao, S Liu, BC Deng, YH Yun, NN Wang, AP Lu, WB Zeng, AF Chen. ChemDes: an integrated web-based platform for molecular descriptor and fingerprint computation. *Journal of cheminformatics*. 2015;7:60.

64. KP Tan, R Varadarajan, MS Madhusudhan. DEPTH: a web server to compute depth and predict small-molecule binding cavities in proteins. *Nucleic acids research*. 2011;39(Web Server issue):W242-8.
65. NM O'Boyle, M Banck, CA James, C Morley, T Vandermeersch, GR Hutchison. Open Babel: An open chemical toolbox. *Journal of cheminformatics*. 2011;3:33.
66. D Weininger. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*. 1988;28:31-6.
67. W Humphrey, A Dalke, K Schulten. VMD: visual molecular dynamics. *Journal of molecular graphics*. 1996;14(1):33-8, 27-8.
68. AB Filippakopoulos P, O Fedorov, T Keates, M Soundararajan, J Elkins, E Salah, N Burgess-Brown, E Ugochukwu, ACW Pike, J Muniz, A Roos, A Chaikuad, FV Delft CH Arrowsmith, AM Edwards, J Weigelt, C Bountra, S Knapp. Crystal Structure of Human Mitogen Activated Protein Kinase 11 (p38 beta) in Complex with Nilotinib. *Structural Genomics Consortium (SGC)*. 2009.
69. Z Wang, BJ Canagarajah, JC Boehm, S Kassisa, MH Cobb, PR Young, S Abdel-Meguid, JL Adams, EJ Goldsmith. Structural basis of inhibitor selectivity in MAP kinases. *Structure*. 1998;6(9):1117-28.
70. Q Sun, JP Burke, J Phan, MC Burns, ET Olejniczak, AG Waterson, T Lee, OW Rossanese, SW Fesik. Discovery of small molecules that bind to K-Ras and inhibit Sos-mediated activation. *Angewandte Chemie*. 2012;51(25):6140-3.

71. T Maurer, LS Garrenton, A Oh, K Pitts, DJ Anderson, NJ Skelton, BP Fauber, B Pan, S Malek, D Stokoe, MJ Ludlam, KK Bowman, J Wu, AM Giannetti, MA Starovasnik, I Mellman, PK Jackson, J Rudolph, W Wang, G Fang. Small-molecule ligands bind to a distinct pocket in Ras and inhibit SOS-mediated nucleotide exchange activity. *Proceedings of the National Academy of Sciences of the United States of America*. 2012;109(14):5299-304.
72. G Bottegoni, W Rocchia, M Rueda, R Abagyan, A Cavalli. Systematic exploitation of multiple receptor conformations for virtual ligand screening. *PloS one*. 2011;6(5):e18845.
73. X Barril, X Fradera. Incorporating protein flexibility into docking and structure-based drug design. *Expert opinion on drug discovery*. 2006;1(4):335-49.
74. J Gustafsson. What pharmacologists can learn from recent advances in estrogen signalling. *Trends in pharmacological sciences*. 2003:479--85.
75. GG Kuiper, B Carlsson, K Grandien, E Enmark, J Haggblad, S Nilsson, JA Gustafsson. Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors alpha and beta. *Endocrinology*. 1997;138(3):863-70.
76. N Heldring, A Pike, S Andersson, J Matthews, G Cheng, J Hartman, M Tujague, A Ström, E Treuter, M Warner. Estrogen receptors: how do they signal and what are their targets. *Physiological reviews*. 2007;87(3):905-31.
77. B Dusterberg, Y Nishino. Pharmacokinetic and pharmacological features of oestradiol valerate. *Maturitas*. 1982;4(4):315-24.
78. WH Moos, JA Dykens, N Howell. 17 α -estradiol: a less-feminizing estrogen. *Drug Development Research*. 2008;69(4):177-84.

79. HV Namboodiri, M Bukhtiyarova, J Ramcharan, M Karpusas, Y Lee, EB Springman. Analysis of imatinib and sorafenib binding to p38 α compared with c-Abl and b-Raf provides structural insights for understanding the selectivity of inhibitors targeting the DFG-out form of protein kinases. *Biochemistry*. 2010;49(17):3611-8.
80. MA Bogoyevitch, DP Fairlie. A new paradigm for protein kinase inhibition: blocking phosphorylation without directly targeting ATP binding. *Drug discovery today*. 2007;12(15-16):622-33.

VITA

Nabina Paudyal was born in Kathmandu, Nepal. She did her undergraduate in physics from Tribhuvan University, Nepal and moved to USA in 2012 for pursuing her MS in physics from University of Akron, Ohio. In August 2015, she joined GSBS, UT Health at Houston for her career in biomedical sciences. She looks forward to implementing her knowledge of physics in biomedical research.