Computational Insights into Molecular Docking Strategies for Non-Small Cell Lung Cancer Drug Discovery

By

Name: Reshita Das
Student ID: 20146043

A thesis submitted to the School of Pharmacy in partial fulfillment of the requirements for the degree of Bachelors in Pharmacy

School of Pharmacy
Brac University
February, 2024

Declaration

It is hereby declared that

1. The thesis submitted is my own original work while completing my degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material that has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. I have acknowledged all main sources of help.

Student's Full Name & Signature:

Reshita Das
20146043

Approval

The thesis titled "Computational Insights into Molecular Docking Strategies for Non-Small Cell Lung Cancer Drug Discovery" submitted by Reshita Das (20146043) of Spring, 2020 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Pharmacy (Hons.) on February 2024.

Examining Committee:

Supervised By:

Dr. Humair Bin MD Omer
Assistant Professor
School of Pharmacy
BRAC University

Professor Dr. Eva Rahman Kabir
Dean
School of Pharmacy
BRAC University

Approved By:

Program Director**:**

Professor Dr. Hasina Yasmin
Program Director and Assistant Dean
School of Pharmacy
BRAC University

Dean**:**

Professor Dr. Eva Rahman Kabir
Dean
School of Pharmacy
BRAC University

Ethics Statement

The study does not involve any human or animal trial.

Abstract

This work addresses computational approaches to non-small cell lung cancer (NSCLC) drug discovery, highlighting the vital part that molecular docking plays in the early stages of drug discovery. It demonstrates the effectiveness of molecular docking and structure-based drug design (SBDD) in hit identification while navigating through conventional obstacles. With an emphasis on adenocarcinoma and squamous cell carcinoma, the research sheds light on the genetic complexity of NSCLC. Potential associations between NSCLC, diabetes, hypertension, and cholesterol levels are investigated. The binding affinities of several drug classes with the target protein are examined in a critical investigation that highlights non-covalent interactions such as salt bridges compared to Rapamycin (binding affinity = 19.4 kcal/mol). By providing opportunities for more potentially effective and tailored treatments for NSCLC and other diseases, the significance of the research is that it applies the field of computational drug design in the discovery of new, potential ligands against NSCLC.

Keywords: Docking, Rapamycin, NSCLC, 5GPG, Salt Bridges, Vina.

Dedication

To my parents and elder brother, who constantly motivated me to improve and taught me to remain positive in the face of challenges in life.

Acknowledgment

To begin with, I want to express my gratitude to my parents for helping me to reach this point in life. The people I would like to thank the most are my supervisor, Honorable Assistant Professor Dr. Humair Bin MD Omer, Honorable Dean and Professor Dr. Eva Rahman Kabir and Honorable Lecturer Nashrah Mustafa, whose knowledge, inspiration, and direction helped me to get the idea from crude to life.

I would like to express my heartfelt gratitude to the BRAC University School of Pharmacy as well as all of my course teachers for their guidance and support over my four years as a student. This effort is founded on the collaboration and support that have been given, fostering a friendly environment for research.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Drug Design and Discovery and Computational Biology

Pharmaceutical innovation is characterized by novel techniques to finding new drugs by utilizing biological target knowledge. Essentially, what this process involves is building molecules that are in harmony with the particular molecular target they interact with and attach to, both in terms of form and electrical charge. Consequently, the process of discovering and producing novel therapeutic entities for a class of illnesses or a particular disease utilizing a combination of computational, experimental, translational, and clinical models is known as drug design and discovery (1). When biological targets, such as proteins, enzymes, genes, or receptors, are connected to a specific biological process that is assumed to be non-functional in a patient with a disorder, such as Parkinson's disease, design and development operations may commence. The process of creating and discovering new medications encompasses a number of stages, such as identifying the disease and target, verifying the target, altering the hit, locating a lead molecule, and optimizing the structure-activity relationship lead. After completing all of these stages, the probable drugs candidate is sent for pre-clinical and clinical study before getting regulatory body approval to be commercialized. Every one of these steps has a purpose. Each step resulted in the identification of probable therapeutic candidates, but there was still a significant lot of ambiguity about the medicine's ability to treat the condition for which it was being produced.

The first and most crucial step in beginning the drug design and development process is identifying the target ailment for which the medication is to be designed. When developing a new medication, the pharmaceutical company's initial objective is to ascertain whether the product's marketing profit will outweigh the costs of development and testing. Therefore, these

questions need to be answered in order to select the target disease. Is the illness common? Is there a clinically unmet need? Is the sickness affecting the developed world? Is an innovative treatment likely to have a market advantage? Determining the drug target is also essential to make the assumption that the molecules being targeted are druggable targets. As was previously said, druggable targets include proteins, enzymes, receptors, and genes. Therefore, it is conceivable that these treatable targets do not include the disease-causing molecule. Only 10 percent of genes contain proteins that are effectively the focus of a medication intervention that is authorized by law (2). There are different ways for finding targets, such as RNA screening and phenotypic assay. A key component of medicine design and discovery is target validation. But since it is a risk assessment tool, the likelihood of progressing in drug discovery falls with increased validation. After these methods comes the hit identification process, which is essentially the selection of a compound from a screen based on its ability to meet certain parameters (usually binding affinity) and be used for chemical follow-up. A variety of strategies exist for identifying hits, one of which is structure-based drug design (SBDD), which is addressed in greater detail in the paper's later parts. Following the generation of hits, leads are retrieved from them and optimized. The pharmacophore, target interaction, and pharmacokinetics properties are all tuned. After optimization, prospective pharmaceutical candidates are identified and exposed to preclinical and clinical trials. If a medicine is determined to be effective in treating unmet clinical requirements during clinical trials, further measures are performed to commercialize the medication.

However, it takes nearly 12 years, often even longer, from target identification to final approval (3). Millions of dollars are needed for this procedure in addition to a significant quantity of medications. But even after spending a significant amount of time and money, there is no guarantee that the optimized lead will perform well in clinical trials. Following all of these procedures, only 15% of clinical trials are successful (4) and the majority of central nervous

system medication instances have an even lower success rate and speed which is 8.2% (3). Furthermore, small molecular entities require nearly 2.6 million development cycles on average. The time and cost estimates for creating novel pharmaceutical substances are displayed in the diagram below -

	Target to Hit	Hit to Lead	Lead Optim	Non-Clinical	Phase 1	Phase 2	Phase 3	Sub to Launch
# per Launch	24.3	19.4	14.6	12.4	8.6	4.6	1.6	1.1
P(TS)	80%	75%	85%	69%	54%	34%	70%	91%
Cycle time (yrs)	1.0	1.5	2.0	1.0	1.5	2.5	2.5	1.5
Cost/launch ($mil)	$94	$166	$414	$150	$273	$319	$314	$48
P(TS)AD[1]					28%	8%	1.8%	100%

Figure 1: Cost and Time Estimation of Developing New Drug Entities (5, 6)

Substantial progress has been achieved in the field of computational methods aimed at enhancing the affinity, selectivity, and stability of these protein-based therapeutics. In the era of extensive data, bioinformatics tools and computer modeling approaches are commonly, but not always, applied in drug design. The creation and commercialization of novel therapeutic entities has been performed with significantly less time, money, and failure rate thanks to computational biology, commonly known as computer-aided drug design (CADD). Notable advancements in computational biology include the finding of binding/active sites, the decoding of the molecular mechanisms driving ligand binding, and the improvement of the structural aspects of ligand-target binding poses. Most of these approaches underline how vital it is to pinpoint the binding and active areas on the target protein correctly. Beyond tiny molecules, biopharmaceuticals—particularly therapeutic antibodies—are becoming acknowledged as a significant class of therapeutics. Considerable breakthroughs have been

achieved in the domain of computational approaches aimed to increase the stability, selectivity, and affinity of these protein-based therapeutics.

As was previously noted, one of the most critical phases in discovering possible medication candidates is hit identification. The most prevalent method for this is high throughput screening (HTS), which uses automated technology to quickly examine hundreds to millions of samples for biological activity at the molecular, cellular, pathway, or model organism levels. The drug development process currently incorporates this automated technology as regular practice, which enables a complete understanding of the interactions between target molecules and biological systems (7). High Throughput Screening (HTS) is confined to research programs able to screen vast chemical libraries due to its slow development rate. Using computational biology, or CADD, one can substantially cut down on the amount of chemicals required for screening while still finding lead compounds. By removing compounds that are anticipated to be inactive and giving preference to those that are, this strategy decreases the expenditure and time associated with completing a complete HTS screen without reducing lead finding (8). A great approach to illustrate this is to use the example of Pharmacia researchers, now part of Pfizer, searching for an inhibitor of the tyrosine phosphatase-1B enzyme, which is expressed in people with diabetes. Out of 365 compounds, the virtual screening method revealed that about 127 compounds had an inhibitory effect, yielding a hit rate of nearly 35%. The usual high throughput technique was utilized to search for similar inhibitory effects, and the hit rate was only 0.021% (9). Moreover, the traditional method was only able to identify 81 compounds that were able to show an inhibitory effect (9). Early examples of drugs discovered with the use of CADD approaches are the ACE inhibitor captopril, which was licensed in 1981 for the treatment of hypertension, and the carbonic anhydrase inhibitor dorzolamide (10). Three HIV therapeutics—saquinavir, ritonavir, and indinavir—were approved in 1996 (11), along with tirofiban, a fibrinogen antagonist, approved in 1998 (12). In addition, the 2014 West African

Ebola epidemic took over 11,000 lives, underlining the requirement of successful drug development to prevent similar epidemics in the future. In the article "Combating Ebola with Repurposed Therapeutics Using the CANDO Platform" (13), the research team introduced the computational analysis of novel drug opportunities (CANDO) platform. This strategy is predicated on the theory that drugs interact with several protein targets, producing a molecular signature that facilitates expedient therapeutic repurposing. Using CANDO, they matched the top-ranked drugs candidates from in vitro studies with those from potential treatments for the Ebola virus disease. It will take significantly less time, money, and resources to find effective medicines for future Ebola epidemics since computational forecasts and in vitro data have enabled the selection and prioritization of drugs for further testing. (1). In this way, computational biology or computer-aided drug design plays a crucial role in drug design and discovery.

1.2 Molecular Docking

As previously stated, structure-based drug design (SBDD) is a technique utilized in the process of target identification. The process of protein crystallization in the presence of a ligand (fragment/compound) reveals the manner in which the ligand bonds to particular amino acids, thereby highlighting the interactions between chemical groups. This facilitates predictions regarding the development of compounds with enhanced binding capabilities. Pharmaceutical companies are actively investigating the mechanism by which ligands recognize and bind to macromolecules, as this knowledge holds significant potential for expediting the development of novel medications while reducing expenses. SBDD is a cyclical procedure that commences with the identification of potential ligands via in-silico analysis of a known target structure. Then, the most promising molecules are produced (14). Following that, many platforms for experimentation evaluate biological attributes such as potency, affinity, and effectiveness (15). The three-dimensional structure of the ligand-receptor complex is ascertained if active

compounds are discovered. Important elements facilitating molecular recognition may be observed with the help of this structure. Understanding binding conformations, intermolecular interactions, unidentified binding sites, mechanistic investigations, and ligand-induced conformational changes is made easier with the aid of descriptions of these complexes (16). Following the identification of a ligand-receptor complex, biological activity data and structural details are co-related (17).

Molecular docking is one of the most often used methods that is used in SBDD. In the early 1980s the molecular docking study first arose to differentiate between the small and large molecules. As the years went by, it became an essential tool in the various stages drug discovery that include hit identification and optimization, drug repositioning and multi target ligand design. It assists in foreseeing interactions between molecules and biological targets (18). Initially, the process involves predicting how a ligand aligns within a receptor, followed by evaluating their compatibility using a scoring function (18). The increased availability of ligand and target information, together with improved algorithms, have made the docking process an important tool for drug discovery. Large-scale screening techniques now use docking due to its enhanced speed and predictive capabilities. Finding (i) potential ligand binding sites in proteins, (19) (ii) new molecular targets for established ligands, (20) (iii) potential adverse drug reactions (ADRs) (21), and (iv) ligands with new chemical structures that are efficient against a particular target or a set of desirable targets are some examples of what this entails (22).

This molecular docking study is now being used extensively for developing medicines for various orphan diseases, cancers, HIV and many more. The success rate of molecular coupling study over the years is nearly 2-30%. Moreover, they considerably reduce the cost of developing a new drug and therefore the researchers are taking the help of such in-silico studies. One of such examples is Raltegravir and its analogue, an HIV-1 integrase inhibitor.

1.3 Non-Small Lung Cancer (NSCLC)

The human body maintains a precise balance between cell division and cell death in the complex strut of life. Cell division is a complex process that allows cells to develop and multiply continuously during this harmonious process. But as they live longer, cells can sustain damage and eventually die according to a predetermined time, which allows for the emergence of new, healthy cells. Within the organism, this coordinated functions of cells are a basic and continuous occurrence. However, unregulated cell multiplication and a decrease in programmed cell death can upset this delicate equilibrium. Tumors, which are aberrant tissue growths that may or may not be malignant, can result from such disruptions. These tumors set off a series of events if they turn out to be malignant over time. They first affect adjacent tissues by a process called invasion, and finally, they go to other regions of the body, where they spread via a process called metastasis. Cancers arise from a variety of reasons, including genetic inheritance, mistakes in cell division, and damage to DNA that cannot be detected up by cell-cycle checkpoints. There is a wide spectrum of malignancies, and each has specific qualities and symptoms. Of all of them, breast cancer is the most common, with 300,209 new cases projected in the US in 2023 (23). Following closely are other prevalent malignancies, including prostate cancer and lung cancer. This diverse landscape underscores the importance of understanding the complex biology of cancer and devising targeted strategies for prevention, diagnosis, and treatment.

As previously mentioned, the second most frequent cancer globally is lung cancer. In addition to other concerns, smoking is the leading cause of it ranking second. Smoking is connected to all forms of lung cancer, however the association is stronger among smokers, notably for squamous cell carcinoma and small cell lung cancer (SCLC). Conversely, nonsmokers have a larger risk of acquiring a kind of cancer called adenocarcinoma (24). The most prevalent type of lung cancer among the different subtypes is non-small lung cancer. Including a wide range

of clinical circumstances and treatment options, 30% of newly diagnosed cases of non-small NSCLC are found to be locally progressed (25). NSCLC usually shows symptoms such as loss of appetite, chest pain, breathing difficulty, wheezing, hoarseness, and breathing issues. EGFR, ERBB2, and KRAS are the primary gene changes responsible for NSCLC. 63% of the 256 persons studied had multiple mutations involving two or more of these genes (26). Importantly, one patient in the subset of patients with two gene mutations exhibited co-mutations for both KRAS and EGFR. Additionally, mutations in EGFR and KRAS were frequently proven to be mutually exclusive. This shows the multiple genetic changes that add to the disease's heterogeneity and illuminates the intricate landscape of gene mutations in NSCLC.

1.4 Molecular Pathology of NSCLC

Adenocarcinoma and squamous cell carcinoma are the two forms of NSCLC. Despite their great genetic similarities, malignancies can be distinguished from one another based on distinctions in molecular changes. Because the molecular changes connected to NSCLC may differ based on criteria such as gender, race, and ethnicity, it is necessary to examine specific patient groups or datasets in order to gather reliable information. General ranges and averages are useful in determining the overall frequency of distinct alterations (27).

- **Adenocarcinoma:** A range of somatic changes, including insertions, deletions, substitutions, and splice site mutations, lead to the formation of adenocarcinoma (ADC). While all NSCLC subtypes share mutations in TP53 and LRP1B, ADC differs from lung squamous cell carcinoma (SCC) in that it has higher incidence of somatic mutations in certain genes, such as KRAS, EGFR, KEAP1, STK11, MET, and BRAF (28),(29). These alterations predominantly interfere with key pathways, including RAS-MEK-ERK and PIK3CA-MTOR. Furthermore, lung ADC may exhibit a substantially lower incidence of specific chromosomal gains and losses than lung SCC (30). In the

larger context of NSCLC, comprehensive knowledge of the distinct genetic landscape reveals the intricate molecular pathways underlying the initiation and progression of ADC.

- **Squamous Cell Carcinoma:** The second most prevalent subtype NSCLC is lung squamous cell carcinoma (SCC). It has a complex molecular background and has similar molecular characteristics with lung adenocarcinoma (ADC). However, it preserves a discrete set of alterations that allow some molecular differentiation. It is noteworthy for having several recurrent somatic mutations, most of which are related to TP53, LRP1B, CDKN2A, MLL2, KEAP1, NFE2L2, HLA-A, NOTCH1, and PDYN (31). Given that these changes connect with the mutational profile documented in several earlier investigations, they underscore the constancy of these alterations in the molecular landscape of lung SCC.

In addition to these pathways, the mTOR pathway is another critically crucial mechanism that may lead to cancer. The mTOR protein is critical for controlling cell division and proliferation under normal circumstances. However, aberrant mTOR activation within tumor cells triggers the transmission of signals that stimulate the growth, metastasis, and invasion of surrounding healthy tissues by cancer cells. The primary initiator of mTOR activation is the PI3K/phosphate and fungal homology deleted on chromosome 10 (PTEN)/AKT/TSC pathway (32). Mutations in the genes along this pathway can cause malignant tumors to arise because it is a crucial intersection in the intricate network of cellular control. The abnormal start of mTOR emphasizes its critical significance along the course of tumor formation, offering light on the many interactions among cellular signaling pathways and their possible impact on the development of NSCLC in the complex cellular microenvironment.

1.5 Available Treatment Options for NSCLC

In cases when surgery is not viable and the patient is clinically defined as being in stage I or II, definitive radiation therapy (RT) or stereotactic body radiation therapy (SBRT) would be the new course of treatment. Clinical stage III patients would require an integrated therapy strategy that comprised discussions with radiation oncology, thoracic surgery, and medical oncology in order to find the most successful combination method for controlling the condition. In tissue biopsies, targetable mutation analysis is critical, especially for stage IV NSCLC. If the biopsy results demonstrate that the epidermal growth factor receptor (EGFR) is present, therapeutic options include tyrosine kinase inhibitors like afatinib or gefitinib become possible (33). The recommended course of therapy when anaplastic lymphoma kinase (ALK) fusion oncogene is present is to use ALK tyrosine kinase inhibitors, such as Alexinib, Ceritinib, or Brigatinib (33).

Moreover, the potent mTOR inhibitor rapamycin is used in the treatment of NSCLC. Although effective, prolong use of this medicine may lead to the development of resistance. Both mTOR and FKBP12 mutations may be the cause of this. Rapamycin resistance can also be caused by alterations or mutations in the downstream effectors of mTOR, including ribosomal protein S6 kinase (S6K), eIF4E binding protein 1 (4E-BP1), the regulator of eukaryotic initiation factor 4E (eIF4E), and the cyclin-dependent kinase inhibitor (CKI) p27kip1 (34). The large spectrum of potential resistance mechanisms underscores how crucial it is to keep checking out for successful treatments for NSCLC. The current molecular docking experiment seeks either to develop an analogue resistant to MDR1 gene expression suppression or maximize the therapeutic efficacy of the prescribed medicine by preventing the development of resistance. In order to overcome the obstacles provided by drug resistance and progress the field of NS, it may be necessary to explore for novel techniques in this molecular context.

1.6 Association of Various Disease with NSCLC

In individuals with NSCLC, increased amounts of both total cholesterol (TC) and low-density lipoprotein cholesterol (LDL-C) increase the risk of metastasis. This is especially true when both LDL-C and TC levels are increased at the same time. As a result, regulating serum lipid levels—more notably, LDL-C and TC—becomes evident as a potentially effective technique for minimizing the probability of metastasis in people with NSCLC. This stresses how vital it is to take lipid profiles into account and carry out tailored treatments as a preventative step to stop the spread of NSCLC patients' cancer (35).

Moreover, NSCLC is more likely to arise in patients with chronic obstructive pulmonary disease. In 2023, Kim and colleagues did a study which revealed that patients diagnosed with type-2 diabetes had an elevated risk of having NSCLC in the future (36). Similarly, in one of the article titled "High glucose promotes tumor cell proliferation and migration in lung adenocarcinoma via the RAGE-NOXs pathway" Yuan and his team indicated that heightened blood sugar levels, or hyperglycemia, exert a stimulating influence on the growth of lung cancer (37). Moreover, an association has been established between increasing expression levels of insulin receptors and the progression of lung cancer. Among them, the IGF-1/IGF-1R pathway has been recognized as an important factor in the biology of lung cancer (38). These results emphasize the complicated relationships among insulin signaling, glucose metabolism, and the IGF-1/IGF-1R pathway that drive the dynamics of lung cancer development.

There may be a connection between hypertension and NSCLC in addition to cholesterol and diabetes. Although the specific molecular mechanism by which hypertension promotes cancer is yet unknown, it is most likely linked to reactive oxygen species that are formed as a result of prolonged oxygen deprivation and incorrect lipid peroxidation. Investigating these processes

is vital to comprehending the connection between increased blood pressure and the onset of cancer (39).

This hypothesis predicts that the patient's likelihood of developing rapamycin resistance will be greatly lowered if they take any of those drugs. Because practically all of the aforementioned diseases at some time in their lives contribute to the development of non-small cancer. There can also be other underlying reasons.

Chapter 2

Materials and Method

2.1 Introduction to AutoDock Vina

Predicting a molecule's preferred orientation in regard to another is crucial in the field of molecular simulations, especially when a ligand and a target form a stable complex. Aligning the spatial and kinetic properties across molecules is a complicated task. Attaching high-affinity molecules to target proteins is a frequent strategy employed in the quest for structure-based medicines to deactivate proteins linked to pathogenicity. The typical method of molecular virtual screening includes creating a large number of complexes by thoroughly docking every minuscule molecule in the repertoire with different proteins. The binding energy is then calculated for every small molecule to achieve a score. In the end, scores are awarded to the small molecules, and the compounds with the highest scores are selected for additional investigation in pharmaceutical research.

Dr. Pleg Tortt created the open-ended program AutoDock vina, which is one of the docking engines in the AutoDock Suite, with AutoDock4 (AD4), AutoDockGPU4, AutoDockFR5, and AutoDockCrankPe (40). AutoDock Vina utilizes force field center coordinates, box dimensions, and the PDBQT molecular structure file format to verify that ligands and receptors are inside the box. It determines the initial energy depending on position and then applies the BFGS algorithm to minimize energy to achieve the final score. The Monte Carlo tree search method is used to locate acceptable poses, apply random disturbances, and quantify energy. The algorithm finds the best poses, and the acceptability of each position is decided by its acceptance probability. The model generates N ideal postures and uses the least binding energy while classifying the remaining poses based on root-mean-square deviations (RMSDs). Vina is one of the most often used docking engines because of its simplicity of usage and speed

when compared to other docking engines. It is a command-line tool that you may use from a Linux terminal or a Windows command prompt. It can also be used with Autodock Tools via a graphical user interface, while utilizing systems with around 20 movable keys. AutoDock Vina offers consistent docking results while accelerating searches and enhancing the average accuracy of binding pattern predictions through the use of a more simplified scoring mechanism. Furthermore, AutoDock Vina can be used to adjust the structural modifications made to ligands and proteins. Grid map precomputes and atom type selection are deleted with AutoDock Vina. Rather, it swiftly determines the suitable internal grids for the essential atom shapes.

For this study, version 1.2.5 of AutoDock vina has been used.

2.2 RCSB Protein Download

For users of all colors, the RCSB PDB offers unrestricted access to a large range of fundamental research data and instructional resources. The PDB's structural entries are thoroughly annotated and confirmed according to community requirements, making it a complete chemical research archive (41). By employing the tools and resources given by RCSB PDB, users are able to investigate the complexity of chemical interactions between small molecules and huge macromolecules, learning about their minute characteristics. Moreover, the platform makes it easier to examine more general structural and functional ideas in the field of biochemistry.

Similarly, the co-crystal structure of the protein necessary for this investigation was made available by the RCSB Protein Data Bank. The essential macromolecules and proteins must be downloaded as a 3-D PDB file from the RCSB Protein Data Bank in order to start the molecular docking process. As indicated in section 2.4, this downloaded protein is changed further and converted into a PDBQT file. Two programs are used to process the protein downloaded from RCSB. PyMOL is one, and AutoDock Vina is the other.

2.3 Ligand Preparation

Generally, all essential ligands, including experimental ligands, are easily available through PubChem. Accessible and containing both tiny and large substances is PubChem's database. It provides all the details, such as toxicity, stability, structure, and many other physical and chemical qualities

The fact that the ligand needs to be downloaded from the 3-D SDG form and converted into PDB and PDBQT is vital to know. The experimental ligand, rapamycin, was not available in the 3-D SDG structure for this work. Consequently, 5GPG, the ligand that was isolated from the protein. With the exception of the experimental ligand, all of the chains, water molecules, and small molecules were removed using PyMOL (version 2.5.7). This is the ligand's PDB file. With AutoDock Vina (version 1.2.5), this PDB file is converted into a PDBQT file.

2.4 Protein Preparation

The RCSB Protein Data Bank (PDB ID:5GPG) provided the co-crystal structure of the protein, which comprises the FRM domain of human mTor, Rapamycin, and the FK506 binding region of human FKBP25. In PyMOL (Version 2.5.7), the protein was curated to increase interaction. Water molecules, small molecules, and cofactors were deleted as they didn't seem to be needed for docking. Water needs to be eliminated since the molecular docking tool interprets the structure as being stiff and unable to differentiate between the atoms. It has no knowledge which atom is a member of the receptor and which is a part of the bulk water. Moreover, water molecules are usually not involved in the binding of drugs. Therefore, by clearing the binding site, water molecules are deleted from the protein sequence to stop them from interfering with the search. A and B, two distinct chains observed in the RCSB protein, were present. The curated protein is then given polar hydrogen and Kollman charge, then AutoDock Vina (Version 1.2.5) is used to turn it into a PDBQT file. The macromolecule and the tested

compounds can establish hydrogen bonds when polar hydrogens are present. Docking methods account for these intermolecular forces; otherwise, computations can result in an incorrect binding energy to the target. Charges are absent from crystal structure data, although many methods require charges to perform effectively.

The protein, 5GPG, employed in this experiment has a total atom count of 2009 and a molecular weight of 25.31 kDa. It is expressed in E.coli (42)

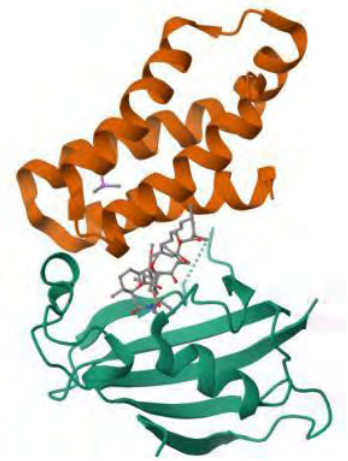

Figure 2: RCSB Protein, 5GPG (42)

2.5 Grid Box Preparation

The position of the grid box represents the area of the protein where docking will occur (Figure 3). During docking, no area outside the box will be examined.

For this study, there were 18 grid points total (x×y×z), or the grid box dimension. The center of the grid box (xyz coordinate) was -16.314, -1.338, and 7.320. There was a 1.00 Å grid point spacing. Once the initial docking box containing the bound ligand has been constructed, its size is arbitrarily increased to ensure that the minimum length of each dimension is at least 18 Å.

Table 1: Grid Box Preparation

	x	y	z
Box Dimension	18	18	18
Center	-16.314	-1.338	7.320

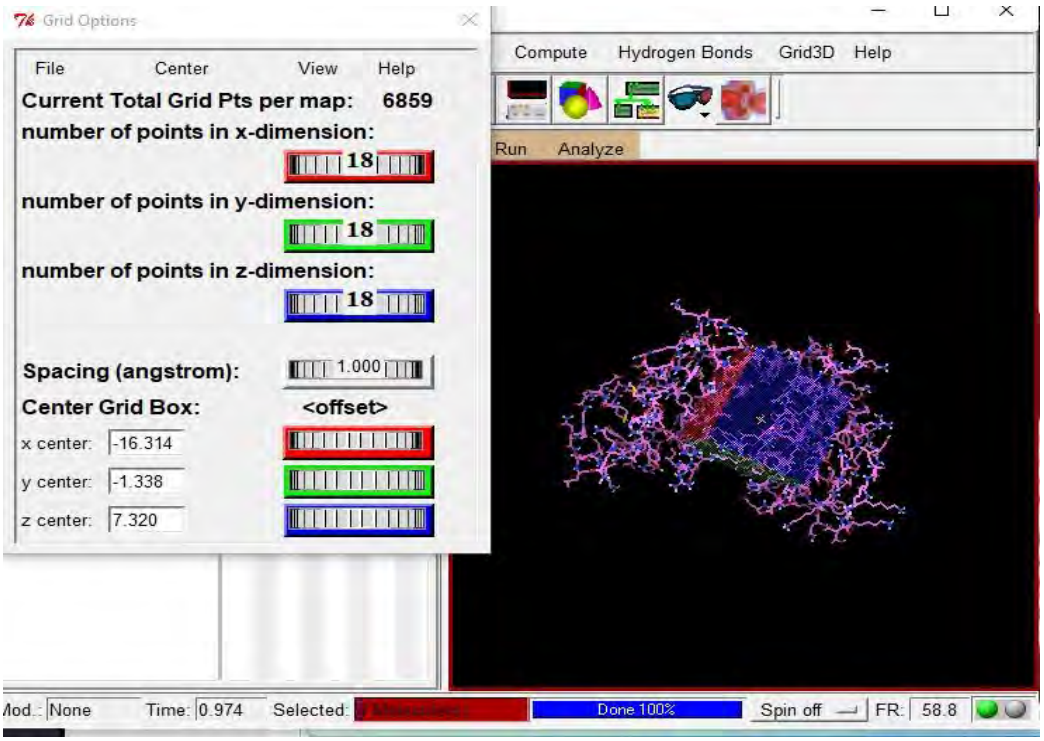

Figure 3: Visual Representation Region of the Protein where the Docking will Take Place

2.6 Data Collection

To help with understanding, Figure 4 displays the steps that took place in this procedure. The experimental protein was downloaded from the RCSB Protein Data Bank, and as shown in figure 1, the ligand was synthesized from a co-crystallized structure. After curation, AutoDock was used to convert the protein structure to PDBQT. As was previously observed, before the protein was transformed to the PDBQT format, the polar hydrogen and Kollman charge were added. Similar alterations were made to the ligand to PDBQT. The ligand and protein are subsequently bound using AutoDock Vina. The AutoDock Vina requires a rather tight setting

in order to dock successfully. For each ligand, a folder must be created. The folders should contain vina split extension, grid maps (grid boxes) for all sorts of atoms, ligand in PDBQT form, and a docking parameter file with the files and parameters required for the docking computation (configuration file). The grid box and configuration file must be in txt (text) format, and the protein and ligand must be in PDBQT form.

Once the protein and ligand have been created, altered, and placed in the proper position, the molecular docking process is accomplished. An AutoDock Vina has been used for docking in this study. Using the Windows prompt (ctrl + R), the Vina tool was launched. The path's directory was altered by the provided folder location. After that, a code was given to ascertain the ligand's affinity for binding to the protein. The following should be the directive: "C:\Program Files (x86)\The Scripps Research Institute\Vina\vina.exe" --output.pdbqt --log log.txt --config config.txt --ligand ligand.pdbqt (Figure: 4).

Computation timelines for binding affinity can vary based on the size of the grid box. In other words, computation times will increase with bigger grid box dimensions and decrease with lower grid box dimensions. But in both circumstances, the binding affinity will be the same.

Validation is essential once docking is complete. Docking is confirmed by RMSD measurements, which indicate the similarities between the two three-dimensional protein structures. In addition, following docking using AutoDock Vina, the binding relationship between the protein and ligand was evaluated using Discovery Studio (version 17.2.0.16349).

The binding affinity of protein 5GPG to Rapamycin was 19.4 kcal/mol.

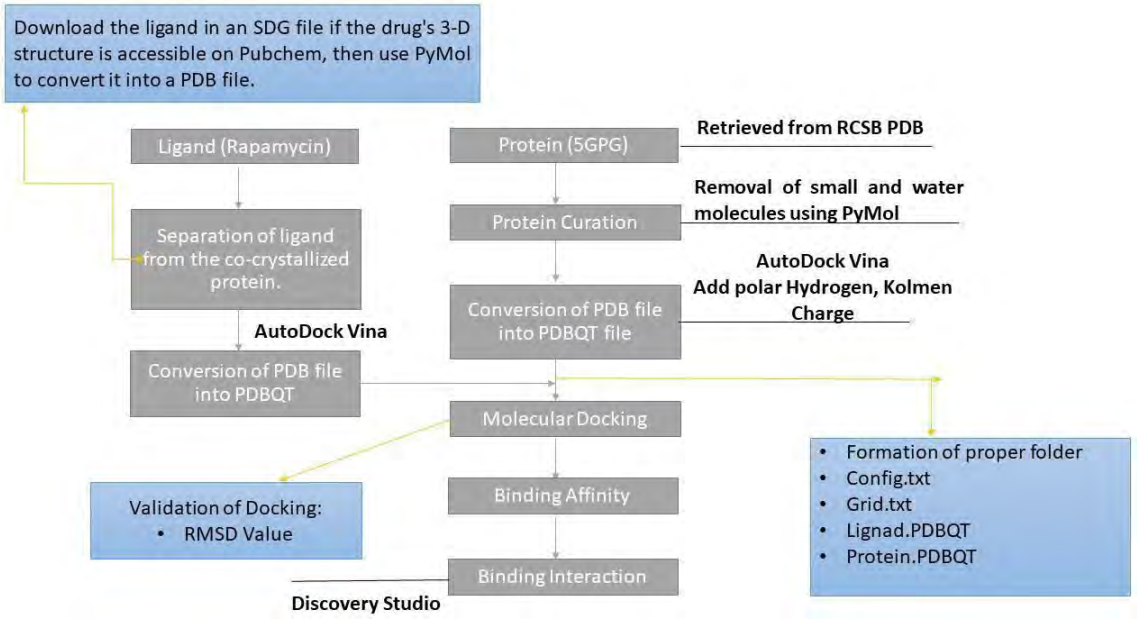

Figure 4: Schematic Representation of Steps Involved in the Molecular Docking

2.7 Protein Structure Validation

RMSD values are considered to be reliable measures of variability when applied to very similar proteins, such as several conformations of the same protein (43). By taking the inhibitor—in this case, Rapamycin—out of the complex, docking again, and figuring out the root mean square deviation (RMSD), the docking process was validated (44). In general, an RMSD value of less than 3 Å is generally regarded as satisfactory. As the RMSD value lowers, so does the degree of similarity between the two structures (co-crystallized ligand and experimental ligand).

Additionally, it is better to have a smaller grid box dimension. The docking's borders are defined by a lower grid box dimension. If the dimension is larger or the entire surface is addressed, a ligand may target the complete protein or macromolecule. Furthermore, if a binding pocket is not previously known, it may bind anywhere (45). It is known as blind docking. However, blind docking may not produce a sufficient result and may require additional processing time if the protein is complicated.

Table 2 demonstrates the relationship between RMSD values and grid box sizes, thereby verifying the co-crystallized Rapamycin and experimental Rapamycin docking technique. An RMSD value of 0.9262 demonstrated that the two structures were more similar than not, which led to the prediction of a close-match docked position.

Table 2: Different Grid Box Dimension and their RMSD Values

Grid Box Dimension	RMSD Values
Grid Box-40, Exhaustiveness-8	1.3456
Grid Box-18, Exhaustiveness-8	0.9262
Grid Box-22, Exhaustiveness-8	1.3466
Grid Box-30, Exhaustiveness-8	1.3367
Grid Box-40, Exhaustiveness-32	1.3475
Grid Box-18, Exhaustiveness-32	0.9307
Grid Box-22, Exhaustiveness-32	1.345
Grid Box-30 Exhaustiveness-32	1.3455

Here, exhaustiveness-8 and grid box-18 have the lowest RMSD values (0.9262). One of the most important metrics for assessing prediction effectiveness in the field of protein structure prediction is the root mean square deviation (RMSD) value between the experimental and predicted structures. Only when the RMSD is similar to that of closely analogous proteins—typically less than 3Å regarded successful (46). For this particular study, since the value is much lower when taken in a smaller grid box dimension it can be said that the experimental and predicted protein has almost a similar structure.

Chapter 3

Results

3.1 Docking Result of Experimental Rapamycin with 5GPG Protein

On the other side, 5GPG protein is the FRM domain of human mTOR, Rapamycin, and the FK506 binding domain of human FKBP25. Rapamycin is one of the FDA-approved medications for the treatment of NSCLC. Following the synthesis of the protein and ligand, a particular binding affinity was discovered. The robustness of the binding connection between a single biomolecule, like a protein or DNA, and its matching ligand or binding partner, like a drug or inhibitor, is referred to as binding affinity (47). The phrase describes the strength of this molecular connection and is vital to grasping and interpreting biological process dynamics and prospective therapeutic actions. Table 3 exhibits Rapamycin's binding affinity for the protein 5GPG. In this situation, the binding affinities were averaged after three observations. Binding affinity measurements are sometimes performed three times or more in scientific research in order to maximize the reliability, precision, and statistical significance of the obtained data.

Table 3: Binding Affinity of Rapamycin with 5GPG

Drug Name	Binding Affinity-1	Binding Affinity-2	Binding Affinity-3	Average Binding Affinity
Rapamycin	19.3	19.2	19.1	19.2

The binding affinity is particularly high in this situation. High-affinity ligand binding is often explained by the existence of larger attractive forces between the ligand and its matching receptor. On the other hand, low-affinity ligand binding is identified by a reduced degree of

attractive force in the ligand-receptor interaction. This basic notion shows the intricate dynamics that influence the molecular interactions between 5GPG and rapamycin. Furthermore, the total affinity and specificity of ligand-receptor interactions inside the body are dictated by the strength of binding.

Figure-5 displays the binding interaction between rapamycin and 5GPG highlighting the main amino acid involved:

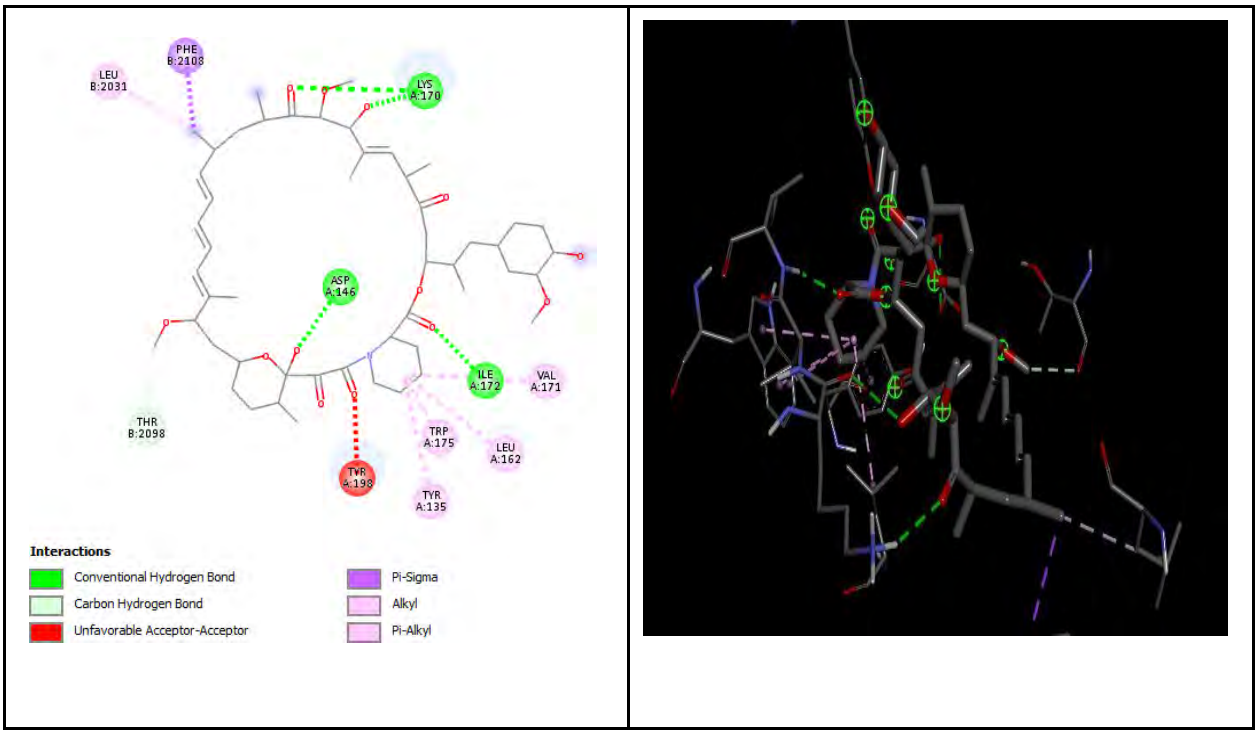

Figure 5: Binding Interaction between Rapamycin and 5GPG

3.2 Docking Results of Different Classes of Drugs with 5GPG Protein

Binding affinities of different classes are checked with 5GPG and these various pharmacological classes include beta blockers, alpha blockers, ARBs, phosphodiesterase-4 inhibitors, respiratory stimulants, antihypertensive, statins, and DPP4 inhibitors. To serve the purpose total 69 ligands were docked with 5GPG protein. Among these the best three binding afinnities are reported in Table 4.

Table 4: Binding Affinity of different classes of drugs with 5GPG

SL No	Name of Drug	Class of Drug	Binding Affinity-1	Binding Affinity -2	Binding Affinity -3	Average Binding Affinity
1	Dihydroergotamine	Analgesic	-13.2	-13.2	-13.2	-13.2
2	Bezitramide	Analgesic	-12.6	-12.3	-12.3	-12.4
3	Pitavastatin	Anti-lipidemic Agents	9.5	9.7	9.6	9.60

Chapter 4

Discussion

Determining if these three pharmacological classes interact with the same pocket of the target protein as rapamycin is the major purpose of testing their binding affinities with it. The assessment also aims to discover whether these relationships happen with similar or higher affinities. This is especially essential because it influences the drug's selectivity for a certain process. A higher binding affinity is connected with better activity and selectivity. On the other hand, determining if these pharmacological classes correlate with the experimental rapamycin has a corresponding purpose. However, even if a drug fits the experimental ligand precisely, its binding affinity could still be substantially less than that of the reference drug. A probable rationale for this mismatch is that the medication binds to the same pocket as the experimental ligand but is unable to maintain its presence, which leads to a loss in binding affinity and, eventually, a reduction in selectivity. Alternatively, more ligand spaces need to be searched and docked in order to locate better ligands than rapamycin.

4.1 Binding Affinity and Salt Bridge

It was revealed during this experiment that rapamycin has a higher affinity for binding the 5GPG protein. However, the resulting binding affinities were substantially lower than those of rapamycin when other medicines from different classes were docked with 5GPG. The greatest binding affinity ever measured, following a full docking of numerous drugs from these three groups, was 11.6 kcal/mol, which is significantly less than the binding affinity of rapamycin. It is significant to notice that certain medications, even when they achieved perfect superimposition with the experimental ligand, were unable to develop a strong binding association with the receptor site. This nuanced study underscores the complicated interactions

between medicines and protein targets and the significance of measuring binding affinity and spatial alignment to acquire a thorough picture.

It was discovered through researching the non-covalent interaction between 5GPG and rapamycin that binding interactions are improved when a salt bridge (electrostatics) occurs within the binding site. These interactions involve multiple amino acid residues and effect both chains of the 5GPG molecule. Van der Waals forces, electrostatic forces, and typically hydrophobic and hydrophilic connections are examples of non-covalent interactions. It is noteworthy that non-covalent interactions have a substantial effect in affecting the binding affinity between macromolecules or ligands, as proven by previous studies (47).

A salt bridge, which is a complicated ionic interaction, is generated when opposing charges within chemical groups and atoms come into contact. Electrostatic attraction and hydrogen bonding are combined in this interaction. In this symbiotic interaction, hydrogen is frequently stronger than it is in conventional hydrogen bonding. The domain of positively and negatively charged amino acid residues is where these salt bridges often form. The amino acids glutamic acid and aspartic acid are involved in negative charge situations, while lysine and arginine are involved in positive charge situations. Analyzing the specifics of the current experiment revealed a noteworthy discovery: the complex binding site of the 5GPG protein forms a salt bridge. The 5GPG protein consists of two chains of amino acids. A salt bridge is an intricate connection that runs between the two chains. Chain-A is the principal builder, generating seven salt bridges, although Chain-B also makes a substantial contribution, constructing three salt bridges. A thorough review of all the precise facts on the amino acids implicated in these interactions can be found in Table 5.

Table 5: Amino Acid Residues involved in Salt Bridge Formation

Amino Acid between which the Salt Bridge is formed	Type of Non-Bonding Interaction	Types
A:LYS113:HZ3 - A:GLU186:OE2	Hydrogen Bond; Electrostatic	Salt Bridge; Attractive Charge
A:ARG173:HH21 - A:GLU177:OE2	Hydrogen Bond; Electrostatic	Salt Bridge; Attractive Charge
A:LYS184:HZ3 - A:ASP222:OD1	Hydrogen Bond; Electrostatic	Salt Bridge; Attractive Charge
A:LYS187:HZ1 - A:GLU217:OE1	Hydrogen Bond; Electrostatic	Salt Bridge; Attractive Charge
A:ARG189:HH12 - A:GLU191:OE2	Hydrogen Bond; Electrostatic	Salt Bridge; Attractive Charge
A:LYS213:HZ2 - A:GLU191:OE1	Hydrogen Bond; Electrostatic	Salt Bridge; Attractive Charge
A:LYS213:HZ3 - A:GLU191:OE1	Hydrogen Bond; Electrostatic	Salt Bridge; Attractive Charge
B:ARG2036:HH12 - B:GLU2041:OE2	Hydrogen Bond; Electrostatic	Salt Bridge; Attractive Charge
B:ARG2036:HH21 - B:GLU2033:OE2	Hydrogen Bond; Electrostatic	Salt Bridge; Attractive Charge
B:ARG2036:HH22 - B:GLU2041:OE2	Hydrogen Bond; Electrostatic	Salt Bridge; Attractive Charge
B:ARG2076:HH22 - B:GLU2080:OE2	Hydrogen Bond; Electrostatic	Salt Bridge; Attractive Charge

B:ARG2086:HH11 - B:GLU2083:OE2	Hydrogen Bond; Electrostatic	Salt Bridge; Attractive Charge
B:LYS2087:HZ2 - B:ASP2096:OD1	Hydrogen Bond; Electrostatic	Salt Bridge; Attractive Charge

With reference to other pharmacological classes, no such interaction has been seen. As previously noted, Table 3 reveals that rapamycin generally has a very high binding affinity. The binding affinities of other pharmacological classes (Table 4) with the same protein are very low in compared to 5GPG. When the binding connections across the many pharmaceutical classes are explored, it has been revealed that none of them contain such an inhibitor. Table 6 reports the non-bonding inhibition of pitavastatin.

Table 6: Non-bonding Interaction of Pitavastatin

Amino Acid Residue Involved in Non-Bonding Interaction	**Types of Non-Bonding Interaction**	**Types**
A:LYS170:NZ - B:GLU2032:OE2	Electrostatic	Attractive Charge
:UNL1:H - B:SER2035:O	Hydrogen Bond	Conventional Hydrogen Bond
:UNL1:H - B:TYR2104:O	Hydrogen Bond	Conventional Hydrogen Bond
A:GLN203:HE21 - B:PHE2039:O	Hydrogen Bond	Conventional Hydrogen Bond

B:TYR2038:HH - A:ASP205:OD2	Hydrogen Bond	Conventional Hydrogen Bond
B:GLY2040:HN - :UNL1:O	Hydrogen Bond	Conventional Hydrogen Bond
B:ARG2042:HH11 - A:GLY202:O	Hydrogen Bond	Conventional Hydrogen Bond
B:ARG2042:HH21 - A:GLY202:O	Hydrogen Bond	Conventional Hydrogen Bond
B:TYR2088:HH - A:ASP205:O	Hydrogen Bond	Conventional Hydrogen Bond
B:TYR2105:HH - A:SER163:O	Hydrogen Bond	Conventional Hydrogen Bond
B:ARG2109:HH11 - A:SER163:OG	Hydrogen Bond	Conventional Hydrogen Bond
B:ARG2109:HH12 - A:PRO161:O	Hydrogen Bond	Conventional Hydrogen Bond
B:ARG2109:HD2 - A:SER163:OG	Hydrogen Bond	Carbon Hydrogen Bond
UNL1:C - B:TYR2105	Hydrophobic	Pi-Sigma
A:ALA206 - B:VAL2094	Hydrophobic	Alkyl
B:PHE2108 :UNL1	Hydrophobic	Pi-Alkyl

4.2 Docking Validation

4.2.1 Molecular Dynamic Simulation

Not only does the molecular docking approach accomplish its original objective of researching molecular interactions, but it is also a valuable and extensively validated tool. Effective usage of molecular dynamics modeling tools helps this difficult validation procedure. The usage of this simulation technology contributes greatly by delivering full insights into the detailed position and configuration of rapamycin and 5GPG. (48). A great deal of information is disclosed in the realm of molecular simulations. This comprehensive study addresses the subject of conformational alterations that proteins and ligands exhibit. Furthermore, the dynamics of ligand-protein binding are investigated in detail, providing a more sophisticated understanding of their complex interactions. In addition, the molecular modeling approach reveals potential instances of protein misfolding, shedding light on structural abnormalities that could have important consequences (48). One of the most useful applications of molecular dynamics simulations is the study of the relative locations of atoms in the ligand and the protein. This microscopic perspective gives light on the molecular movement that controls their interactions with one another. Above all, the simulation method provides a unique opportunity to virtually explore the chemical environment and understand how ligands interact with other molecules. The flexibility of molecular dynamic simulation techniques is illustrated by their capacity to adapt to many experimental ligand types. These could include cutting-edge techniques such as NMR (Nuclear Magnetic Resonance), cryo-electron microscopy, and other experimental processes (48). This versatility expands the range of molecular simulations and provides a trustworthy, multi-dimensional analysis of molecular interactions, leading to a greater grasp of the intricate molecular landscapes.

Understanding the detailed atom-to-atom movement of a bimolecular entity, such as a protein submerged in water or possibly surrounded by the fluid folds of a lipid bilayer, is a crucial idea

at the core of an MD simulation. Finding and studying each atom's positional coordinates with great care is vital to uncovering the deep interactions that exist between them. Following intensive computing, the forces exerted by the counterparts of each atom become obvious (49). In essence, the simulation offers a captivating exploration of Newtonian principles. At each point where paths intersect, the precisely balanced forces governing atomic motion are at play. These forces adeptly manage the trajectories determining the destiny of each atom, akin to imperceptible directors. This ongoing narrative provides an initial glimpse into the spatial evolution of individual particles and serves as a testament to the enduring relevance of Newton's laws of motion. This computational endeavor transcends mere mathematical exercise, unfolding as a dynamic narrative of motion and interaction (50). The bimolecular system's repetitive motion is depicted as a choreography of computations, with each brief interruption providing a glimpse into its dynamic alterations. The MD simulation is essentially a journey into the center of molecular complexity, where the forces powering the ensemble form a sophisticated interaction that uses an unusual computational approach to unveil the intricate pattern of the bimolecular motion. These simulations are extremely successful for a number of reasons. Firstly, they can correctly capture the spatial organization and dynamic trajectories of each individual atom over the time range, an accomplishment that is tough to do with any sort of experiment. As a result, the simulation environment has rigorous accuracy and controllability. The starting conformation of the protein, the specific ligands that bind to it, the likelihood of mutations or post-translational modifications, and the composition of molecules in the surrounding environment, the protonation state, room temperature, membrane voltage, and a host of other variables are all intricately understood in this context. This careful tailoring of the simulation parameters gives an ideal setting for determining the impacts of diverse chemical perturbations. A range of results from distinct chemical modifications may be demonstrated by comparing simulations done under varied settings. The versatility and

thoroughness of this technology offer a unique viewpoint that permits the detection and investigation of sophisticated molecular processes under a range of contextual complexity (51).

4.2.2 Molecular Mechanics

The Quantum Mechanics/Molecular Mechanics (QM/MM) approach is a potent tool that is commonly utilized to decipher the complexity contained in many chemical systems. This unique method separates the system into two domains: the Molecular Mechanics (MM) zone, which contains the greater chemical environment around the molecules of primary interest, and the Quantum Mechanics (QM) region, which focusses on the individual molecules of interest. This duality provides a complete comprehension of the outer influences influencing the chemical landscape in addition to a concentrated and meticulous exploration of the most minute molecular details (52).

The innermost molecule is treated with the highest precision in the QM/MM framework. Here, an advanced strategy based on quantum mechanics is used to carefully navigate the complex quantum landscape with unmatched precision. Concurrently, the external molecules that comprise the bottom layer are examined using the skilled and effective universal force field (MM) technique (53). The nuances of intermolecular interactions are caught by this flexible molecular mechanics approach, providing a full analysis of the surrounding chemical matrix. The implementation of the QEQ formalism, a sophisticated technique that determines Molecular Mechanics (MM) charges, is key to this smooth integration. These charges, typical of the complicated interactions in the MM layer, are easily integrated into the electronic structure of quantum mechanics (QM). This perfect combination guarantees a precise portrayal of the electrostatic interactions in the framework of QM/MM, simplifying the intricate dance of charges and energy (54). Essentially, the combining of accurate quantum mechanics with good molecular mechanics simulations allows for a full investigation of chemical occurrences.

This comprehensive approach establishes the Quantum Mechanics/Molecular Mechanics method as an indispensable instrument in the field of chemical research and understanding, guaranteeing the accuracy of quantum treatments for targeted molecules while also capturing the dynamic interplay of forces and interactions within the larger chemical context (55).

5. Significance of the Project

The issue of cancer treatment resistance is a serious impediment to medical study, as was discussed previously. When cancer cells acquire resistance to treatment, the treatment loses its effectiveness and patient care becomes more complicated. The purpose of this study is to evaluate the possibility of rapamycin—an FDA-approved treatment NSCLC—in combination with other drugs to prevent or delay the emergence of resistance. This strategy is validated by the fact that additional medications, from various classes, might improve rapamycin's activity and effectiveness, preventing the issue of resistance. For these other drugs to perform well with rapamycin, they must connect to the same target protein, 5 GPG, with an affinity similar to rapamycin. While rapamycin is more likely than other medications to bind to 5 GPG, these other therapies also need to bind well in order to assure therapeutic success. This effort attempts to discover lead compounds from many pharmacological classes that display optimal binding to 5 GPG, potentially rendering them possible adjuncts to rapamycin in the treatment NSCLC (56). This is accomplished by implementing a thorough screening strategy that initiates with high throughput screening (HTS) of 69 medicinal compounds against the 5 GPG target. The objective of this preliminary screening phase is to identify prospective lead compounds with exceptional chemical and pharmacological features that could be beneficial in the treatment of NSCLC. After conducting High-Throughput Screening (HTS), the process of lead optimization occurs. During this stage, the outcomes of the HTS and relevant analogs are meticulously assessed through a range of tests and compared with similar compounds. Lead optimization necessitates a thorough analysis of high-throughput screening (HTS) hits and derivative analogs to determine their appropriateness as rapamycin adjuncts. This procedure involves evaluating these compounds in various assay formats and determining their binding affinities towards different counterparts and related family members of the 5 GPG protein. The objective of this extensive inquiry is to discover lead compounds that possess exceptional

pharmacological and chemical characteristics for treating NSCLC and exhibit strong binding to 5 GPG. The primary objective of this endeavor is to address the problem of resistance to cancer therapy by investigating innovative therapeutic strategies. Scientists seek to improve the efficacy of NSCLC treatment and maybe postpone the development of resistance by identifying lead compounds that demonstrate optimum binding to the 5 GPG target protein and augment the effects of rapamycin. This interdisciplinary approach underscores the challenging nature of cancer treatment and the paramount importance of employing innovative thinking to enhance patient outcomes.

6. Conclusion

Drug development via molecular docking is a challenging yet potentially lucrative process, taking into account all relevant factors. By employing bioinformatics tools and computational approaches, researchers explore the complicated interactions that occur between drugs and biological targets. Molecular docking holds enormous potential for pharmaceutical innovation as it may identify possible lead compounds and predict how they will interact with the body. Through intensive investigation and testing, scientists navigate the various paths of disease to identify novel treatment approaches. Molecular docking not only expedites drug discovery but also gives light on the underlying molecular pathways of diseases such as NSCLC. Additionally, the combination of molecular docking with other fields like as structural biology and medicinal chemistry opens up new avenues for drug development. This multidisciplinary approach, which addresses everything from identifying potential drug candidates to determining drug resistance mechanisms, makes customized medicine and targeted therapy realistic.

References

1. Zhou SF, Zhong WZ. Drug Design and Discovery: Principles and Applications. Mol J Synth Chem Nat Prod Chem. 2017 Feb 13;22(2):279.

2. Hingorani AD, Kuan V, Finan C, Kruger FA, Gaulton A, Chopade S, et al. Improving the odds of drug development success through human genomics: modelling study. Sci Rep. 2019 Dec 11;9:18911.

3. Mohs RC, Greig NH. Drug discovery and development: Role of basic biological research. Alzheimers Dement Transl Res Clin Interv. 2017 Nov 11;3(4):651–7.

4. Zhong F, Xing J, Li X, Liu X, Fu Z, Xiong Z, et al. Artificial intelligence in drug design. Sci China Life Sci. 2018 Oct;61(10):1191–204.

5. Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nat Rev Drug Discov. 2010 Mar;9(3):203–14.

6. Cummings JL, Morstorf T, Zhong K. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. Alzheimers Res Ther. 2014;6(4):37.

7. Perera N, Hikkaduwa Koralege RS. High throughput screening. In: Wexler P, editor. Encyclopedia of Toxicology (Fourth Edition) [Internet]. Oxford: Academic Press; 2024 [cited 2024 Jan 8]. p. 297–301. Available from: https://www.sciencedirect.com/science/article/pii/B9780128243152010356

8. Sliwoski G, Kothiwale S, Meiler J, Lowe EW. Computational Methods in Drug Discovery. Pharmacol Rev. 2014 Jan;66(1):334–95.

9. Durant JL, Leland BA, Henry DR, Nourse JG. Reoptimization of MDL keys for use in drug discovery. J Chem Inf Comput Sci. 2002;42(6):1273–80.

10. Talele TT, Khedkar SA, Rigby AC. Successful applications of computer aided drug discovery: moving drugs from concept to the clinic. Curr Top Med Chem. 2010;10(1):127–41.

11. Van Drie JH. Computer-aided drug design: the next 20 years. J Comput Aided Mol Des. 2007;21(10–11):591–601.

12. Hartman GD, Egbertson MS, Halczenko W, Laswell WL, Duggan ME, Smith RL, et al. Non-peptide fibrinogen receptor antagonists. 1. Discovery and design of exosite inhibitors. J Med Chem. 1992 Nov 27;35(24):4640–2.

13. Chopra G, Kaushik S, Elkin PL, Samudrala R. Combating Ebola with Repurposed Therapeutics Using the CANDO Platform. Mol Basel Switz. 2016 Nov 25;21(12):1537.

14. Wilson GL, Lill MA. Integrating structure-based and ligand-based approaches for computational drug design. Future Med Chem. 2011 Apr;3(6):735–50.

15. Fang Y. Ligand-receptor interaction platforms and their applications for drug discovery. Expert Opin Drug Discov. 2012 Oct;7(10):969–88.

16. Kahsai AW, Xiao K, Rajagopal S, Ahn S, Shukla AK, Sun J, et al. Multiple ligand-specific conformations of the β2-adrenergic receptor. Nat Chem Biol. 2011 Aug 21;7(10):692–700.

17. Shoichet BK, Kobilka BK. Structure-based drug screening for G-protein-coupled receptors. Trends Pharmacol Sci. 2012 May;33(5):268–72.

18. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov. 2004 Nov;3(11):935–49.

19. Hazarika RR, Sostaric N, Sun Y, van Noort V. Large-scale docking predicts that sORF-encoded peptides may function through protein-peptide interactions in Arabidopsis thaliana. PloS One. 2018;13(10):e0205179.

20. Li H, Gao Z, Kang L, Zhang H, Yang K, Yu K, et al. TarFisDock: a web server for identifying drug targets with docking approach. Nucleic Acids Res. 2006 Jul 1;34(Web Server issue):W219-224.

21. LaBute MX, Zhang X, Lenderman J, Bennion BJ, Wong SE, Lightstone FC. Adverse Drug Reaction Prediction Using Scores Produced by Large-Scale Drug-Protein Target Docking on High-Performance Computing Machines. PLoS ONE. 2014 Sep 5;9(9):e106298.

22. Lyu J, Wang S, Balius TE, Singh I, Levit A, Moroz YS, et al. Ultra-large library docking for discovering new chemotypes. Nature. 2019 Feb;566(7743):224–9.

23. What Is Cancer? - NCI [Internet]. 2007 [cited 2024 Jan 8]. Available from: https://www.cancer.gov/about-cancer/understanding/what-is-cancer

24. Dela Cruz CS, Tanoue LT, Matthay RA. Lung Cancer: Epidemiology, Etiology, and Prevention. Clin Chest Med. 2011 Dec;32(4):10.1016/j.ccm.2011.09.001.

25. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021 May;71(3):209–49.

26. Chen X, Xu B, Li Q, Xu X, Li X, You X, et al. Genetic profile of non-small cell lung cancer (NSCLC): A hospital-based survey in Jinhua. Mol Genet Genomic Med. 2020 Jul 12;8(9):e1398.

27. Worldwide Frequency of Commonly Detected EGFR Mutations | Archives of Pathology & Laboratory Medicine [Internet]. [cited 2024 Jan 8]. Available from: https://meridian.allenpress.com/aplm/article/142/2/163/65795/Worldwide-Frequency-of-Commonly-Detected-EGFR

28. Chen X, Xu B, Li Q, Xu X, Li X, You X, et al. Genetic profile of non-small cell lung cancer (NSCLC): A hospital-based survey in Jinhua. Mol Genet Genomic Med. 2020 Jul 12;8(9):e1398.

29. Seo JS, Ju YS, Lee WC, Shin JY, Lee JK, Bleazard T, et al. The transcriptional landscape and mutational profile of lung adenocarcinoma. Genome Res. 2012 Nov 1;22(11):2109–19.

30. Suster DI, Mino-Kenudson M. Molecular Pathology of Primary Non-small Cell Lung Cancer. Arch Med Res. 2020 Nov;51(8):784–98.

31. Comprehensive genomic characterization of squamous cell lung cancers. Nature. 2012 Sep 27;489(7417):519–25.

32. Nathan N, Keppler-Noreuil KM, Biesecker LG, Moss J, Darling TN. Mosaic Disorders of the PI3K/PTEN/AKT/TSC/mTORC1 Signaling Pathway. Dermatol Clin. 2017 Jan;35(1):51–60.

33. Lindeman NI, Cagle PT, Beasley MB, Chitale DA, Dacic S, Giaccone G, et al. Molecular Testing Guideline for Selection of Lung Cancer Patients for EGFR and ALK Tyrosine Kinase Inhibitors. Arch Pathol Lab Med. 2013 Jun;137(6):828–60.

34. Gruppuso PA, Boylan JM, Sanders JA. The physiology and pathophysiology of rapamycin resistance. Cell Cycle. 2011 Apr 1;10(7):1050–8.

35. Li R, Liu B, Liu Y, Liu Y, He Y, Wang D, et al. Elevated serum lipid level can serve as early signal for metastasis for Non-Small Cell Lung Cancer patients: A retrospective nested case-control study. J Cancer. 2020 Oct 17;11(23):7023–31.

36. Kim NE, Kang EH, Ha E, Lee JY, Lee JH. Association of type 2 diabetes mellitus with lung cancer in patients with chronic obstructive pulmonary disease. Front Med [Internet]. 2023 [cited 2024 Jan 8];10. Available from: https://www.frontiersin.org/articles/10.3389/fmed.2023.1118863

37. Liao YF, Yin S, Chen ZQ, Li F, Zhao B. High glucose promotes tumor cell proliferation and migration in lung adenocarcinoma via the RAGE-NOXs pathway. Mol Med Rep. 2018 Jun;17(6):8536–41.

38. Leiter A, Charokopos A, Bailey S, Gallagher EJ, Hirsch FR, LeRoith D, et al. Assessing the association of diabetes with lung cancer risk. Transl Lung Cancer Res. 2021 Nov;10(11):4200–8.

39. Gago-Dominguez M, Castelao JE, Yuan JM, Ross RK, Yu MC. Lipid peroxidation: a novel and unifying concept of the etiology of renal cell carcinoma (United States). Cancer Causes Control CCC. 2002 Apr;13(3):287–93.

40. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: new docking methods, expanded force field, and Python bindings. J Chem Inf Model. 2021 Aug 23;61(8):3891–8.

41. Zardecki C, Dutta S, Goodsell DS, Voigt M, Burley SK. RCSB Protein Data Bank: A Resource for Chemical, Biochemical, and Structural Explorations of Large and Small Biomolecules. J Chem Educ. 2016 Mar 8;93(3):569–75.

42. Bank RPD. RCSB PDB - 5GPG: Co-crystal structure of the FK506 binding domain of human FKBP25, Rapamycin and the FRB domain of human mTOR [Internet]. [cited 2023 Dec 22]. Available from: https://www.rcsb.org/structure/5GPG

43. Carugo O, Pongor S. A normalized root-mean-spuare distance for comparing protein three-dimensional structures. Protein Sci Publ Protein Soc. 2001 Jul;10(7):1470–3.

44. Bhojwani HR, Joshi UJ. Pharmacophore and Docking Guided Virtual Screening Study for Discovery of Type I Inhibitors of VEGFR-2 Kinase. Curr Comput Aided Drug Des. 2017;13(3):186–207.

45. Hassan NM, Alhossary AA, Mu Y, Kwoh CK. Protein-Ligand Blind Docking Using QuickVina-W With Inter-Process Spatio-Temporal Integration. Sci Rep. 2017 Nov 13;7:15451.

46. Reva BA, Finkelstein AV, Skolnick J. What is the probability of a chance prediction of a protein structure with an rmsd of 6 å? Fold Des. 1998 Apr 1;3(2):141–7.

47. Binding Affinity | Malvern Panalytical [Internet]. [cited 2024 Jan 8]. Available from: https://www.malvernpanalytical.com/en/products/measurement-type/binding-affinity

48. Hollingsworth SA, Dror RO. Molecular dynamics simulation for all. Neuron. 2018 Sep 19;99(6):1129–43.

49. Salo-Ahen OMH, Alanko I, Bhadane R, Bonvin AMJJ, Honorato RV, Hossain S, et al. Molecular Dynamics Simulations in Drug Discovery and Pharmaceutical Development. Processes. 2021 Jan;9(1):71.

50. Monje-Galvan V, Warburton L, Klauda JB. Setting Up All-Atom Molecular Dynamics Simulations to Study the Interactions of Peripheral Membrane Proteins with Model Lipid Bilayers. Methods Mol Biol Clifton NJ. 2019;1949:325–39.

51. Schlick T. Molecular Modeling and Simulation: An Interdisciplinary Guide: An Interdisciplinary Guide [Internet]. New York, NY: Springer New York; 2010 [cited 2024 Jan 12]. (Interdisciplinary Applied Mathematics; vol. 21). Available from: https://link.springer.com/10.1007/978-1-4419-6351-2

52. Application of molecular docking and ONIOM methods for the description of interactions between anti-quorum sensing active (AHL) analogues and the Pseudomonas aeruginosa LasR binding site - Molecular BioSystems (RSC Publishing) [Internet]. [cited 2024 Jan 12]. Available from: https://pubs.rsc.org/en/content/articlelanding/2014/MB/c3mb70181f

53. QM and ONIOM studies on thermally activated delayed fluorescence of copper(i) complexes in gas phase, solution, and crystal - Physical Chemistry Chemical Physics

(RSC Publishing) [Internet]. [cited 2024 Jan 12]. Available from: https://pubs.rsc.org/en/content/articlelanding/2018/cp/c8cp03657h

54. Shen L, Yang W. Molecular Dynamics Simulations with Quantum Mechanics/Molecular Mechanics and Adaptive Neural Networks. J Chem Theory Comput. 2018 Mar 13;14(3):1442–55.

55. Groenhof G. Introduction to QM/MM Simulations. In: Monticelli L, Salonen E, editors. Biomolecular Simulations [Internet]. Totowa, NJ: Humana Press; 2013 [cited 2024 Jan 12]. p. 43–66. (Methods in Molecular Biology; vol. 924). Available from: https://link.springer.com/10.1007/978-1-62703-017-5_3

56. Chen J, Gopalakrishnan S, Lake MR, Bianchi BR, Locklear J, Reilly RM. Chapter Seventeen - Application of Large-Scale Transient Transfection to Cell-Based Functional Assays for Ion Channels and GPCRs. In: Conn PM, editor. Methods in Enzymology [Internet]. Academic Press; 2010 [cited 2024 Jan 28]. p. 293–309. (Constitutive Activity in Receptors and Other Proteins, Part B; vol. 485). Available from: https://www.sciencedirect.com/science/article/pii/B9780123812964000178