MOLECULAR DOCKING STUDIES OF SMALL MOLECULE INHIBITORS TARGETING EGFR/MAPK PATHWAY

Dissertation submitted to the University of Kerala in partial fulfillment of the requirement for the award of the Degree of Master of Science in Analytical chemistry

M.Sc. Analytical Chemistry

2020 – 2022

CONTENT

Sl. No.	Content	Page No.
	Abstract	3
1	Chapter 1 - Introduction	4
	1.1 Epidermal Growth Factor Receptor	4
	1.2 Mitogen Activated Protien Kinase	6
	1.3 Computer aided drug design	8
	1.3.1 Structure Based Drug Design	10
	1.3.2 Ligand Based Drug Design	12
	1.3.3 Benefits of CAAD	13
	1.3.4 Limitations of CAAD	13
	1.4 Molecular docking	14
	1.4.1 Theory of docking	16
	1.4.2 Experimental docking procedures	18
2	AIM AND SCOPE OF STUDY	21
3	Chapter 2 - Literature review	22
4	Chapter 3 - Materials and methods	26
	3.1 Marvin Sketch	26
	3.2 Protein data bank	26
	3.3 Maestro 12.3	26
	3.4 Biovia Discovery Studio	27
	3.5 Auto dock	27
	3.6 Cygwin	28
	3.7 5D41	28
	3.8 2Y9Q	31
	3.9 Ligand structures	35
	3.10 Methodology	40
5	Chapter 4 - Result and Discussion	41
6	Chapter 5 - Conclusion	48
7	Reference	49

ABSTRACT

Epidermal growth factor is considered as a valid target in the clinical trials of anticancer therapy and MAPk pathway has a critical role in stimulating proliferation and suppressing apoptosis. The degree of structure and other information available about the target (enzyme/receptor/protein) and ligands affects computer-aided drug design (CADD). Molecular docking is becoming a more important method in drug development. The differences between available docking software and their performance are also explored. Flexible receptor molecular docking approaches, particularly those including receptor backbone flexibility, present a challenge for currently available docking methods.

The study of molecular docking using Autodock is conducted using two proteins **5D41** and **2Y9Q**. Lower the binding energy greater its affinity to bind with the protein. Each ligand shows different binding energy due to presence of various functional group. Binding energy calculations revealed that **ligand 17** for **5D41** protein and **ligand 15** for **2Y9Q** protein has obtained the least binding energy from the results obtained after conducting docking with the provided ligands and **ligand 39** and **ligand 40** gives high binding energy for **5D41** and **2Y9Q** proteins respectively.

CHAPTER 1

INTRODUCTION

One of the most important signaling pathways in cancer is the mitogen- activated protein kinase (MAPK)[also known as extracellular signal regulated kinase(ERK)] pathway, which has a critical role in both stimulating proliferation and suppressing apoptosis. Understanding MAPK regulation helps to rationally design new antiproliferative drugs and other therapies.[1] The members of the epidermal growth factor receptor (EGFR) family are potent regulators of MAPK in both transformed and normal epithelial cells.[2] Both the mutations amplified expressions of the gene encoding this receptor are associated with poor prognosis in cancer,[3,4] and the targeting it are effective in subsets of cancers.[5] It is not yet clear that is this is due to secondary mutations or to altered abundance of either EGFR-MAPK pathway proteins.

1.1. Epidermal Growth Factor Receptor

The epidermal growth factor receptor (EGFR; ErbB-1; HER1 in humans) is a trans membrane protein that serves as a receptor for extracellular protein ligands from the epidermal growth factor family (EGF family).[6] EGFR (ErbB-1), HER2/neu (ErbB-2), Her 3 (ErbB-3) and Her 4 (ErbB-4) are members of the ErbB family of receptors, a subfamily of four closely related receptor tyrosine kinases (ErbB-4). Mutations altering EGFR expression or activity have been linked to malignancy in a variety of cancer types.[7] Stanley Cohen of Vanderbilt University discovered epidermal growth factor and its receptor. Cohen and Rita Levi-Montalcini earned the Nobel Prize in Medicine in 1986 for

their discovery of growth factors. In humans, faulty EGFR and other receptor tyrosine kinase signaling is linked to cancer.

The epidermal growth factor receptor (EGFR) is a trans membrane protein that is activated by epidermal growth factor and transforming growth factor (TGF) binding.[7] ErbB2 has no known direct activation ligand, however it may be activated by itself or through hetero dimerization with other members of the ErbB family, such as EGFR. EGFR changes from an inactive monomeric state to an active homodimer after being activated by its growth factor ligands[8] meanwhile, there is evidence that premade inactive dimers exist prior to ligand interaction.[9] EGFR can link with another member of the ErbB receptor family, such as ErbB2/Her2/neu, to produce an active heterodimer in addition to producing homodimers after ligand engagement. There is also evidence that clusters of activated EGFRs form, albeit it is unclear whether this clustering is necessary for activation or occurs after individual dimers have been activated. There is also evidence that clusters of activated EGFRs form, albeit it is unclear whether this clustering is necessary for activation or occurs after individual dimers have been activated EGFe.

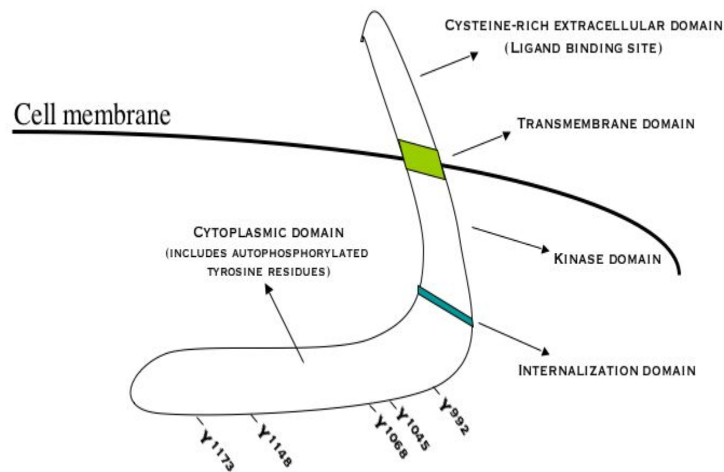


Figure 1: EGF Receptor

1.2. Mitogen Activated Protien Kinase

A serine/threonine-specific protein kinase (MAPK or MAP kinase) is a type of mitogen-activated protein kinase that is selective for the amino acids serine and threonine. MAPKs control how cells respond to a wide range of stimuli, including mitogens, osmotic stress, heat shock, and pro inflammatory cytokines. Proliferation, gene expression, differentiation, mitosis, cell survival, and apoptosis are all regulated by them.[10] MAP kinases are only present in eukaryotes, however they are found in a wide range of mammals, fungi, and plants, as well as a variety of unicellular eukaryotes. MAPKs are kinases that belong to the CMGC (CDK/MAPK/GSK3/CLK) kinase family. The cyclin-dependent kinases are MAPKs' closest relatives (CDKs).[11]

In their natural state, mitogen-activated protein kinases are catalytically inactive. They require (possibly numerous) phosphorylation events in their activation loops to become active. Specialized enzymes from the STE protein kinase group are responsible for this. Long-range allostery can be used to cause a conformational change in the structure of a protein in this way.

The activation loop of conventional MAP kinases has a TxY (threonine-x-tyrosine) motif (TEY in mammalian ERK1 and ERK2, TDY in ERK5, TPY in JNKs, TGY in p38 kinases) that must be phosphorylated on both the threonine and tyrosine residues to lock the kinase domain in a cascade Tyrosine phosphorylation frequently precedes phosphorylation of serine in vivo and in vitro. Members of the Ste7 protein kinase family, also known as MAP2 kinases, phosphorylate the tandem activation loop (which has been postulated to be either distributive or processive depending on the cellular context). Phosphorylation of MAP2 kinases by a variety of upstream serine-threonine kinases also activates them (MAP3 kinases). Classic MAPK pathways are multi-tiered, yet rather linear, because MAP2 kinases show very limited action on substrates other than its cognate MAPK. These pathways can transport stimuli from the cell membrane (where numerous MAP3Ks are activated) to the nucleus (where only MAPKs can enter) or to a variety of additional intracellular locations. Some atypical MAP kinases appear to have a more ancient, two-tiered system than the three-tiered traditional MAPK pathways. PAK kinases have recently been found to directly phosphorylate and activate ERK3 (MAPK6) and ERK4 (MAPK4) (related to other MAP3 kinases).[12] Unlike traditional MAP kinases, these atypical MAPKs only require phosphorylation of a single position in their activation loops. The activation of NLK and ERK7 (MAPK15) is still a mystery.

A number of phosphatases are involved in the inactivation of MAPKs. The MAP kinase phosphatases (MKPs), a subclass of dual-specificity phosphatases, are a relatively conserved family of specialized phosphatases (DUSPs).[13] These enzymes may hydrolyze the phosphate from both phosphotyrosine and phosphothreonine residues, as their name implies. Some tyrosine phosphatases are also implicated in inactivating MAP kinases since removing either phosphate group dramatically reduces MAPK activity, thus eliminating signalling (e.g. the phosphatases HePTP, STEP and PTPRR in mammals).

1.3. Computer Aided Drug Design (CADD)

The degree of structure and other information available about the target (enzyme/receptor/protein) and ligands affects computer-aided drug design (CADD).[14] CADD is based on molecular mechanics, quantum mechanics, molecular dynamics, structure-based drug design (SBDD), ligand-based drug design (LBDD), homology modeling, ligplot analysis, molecular docking, de novo drug design, pharmacophore modeling and mapping, virtual screening (VS), quantitative structure-activity relationships (QSARs)[15], in silico ADMET[16] (absorption, distribution, metabolism, excretion. The CADD Center was established to encourage biologists, biophysicists, structural biologists, and computational scientists to collaborate on research. The CADD center's main purpose is to start these collaborations, which will lead to the creation of research projects aimed at discovering novel chemical entities with the potential to be turned into new medicinal medicines. Drug discovery and development is a lengthy, difficult, expensive, and high-risk process with few parallels in the commercial sector. This is why, in order to speed up the process, computer-aided drug design (CADD)

technologies are frequently used in the pharmaceutical business. Using computational techniques in the lead optimization phase of drug development saves a lot of money.

On average, it takes 10-15 years and $500-800 million to bring a medicine to market, with lead analogue synthesis and testing accounting for a significant portion of that cost. As a result, using computational methods in hit-to-lead optimization to cover a larger chemical space while minimizing the number of compounds that must be manufactured and tested in vitro can be useful. The assumption behind computational drug design is that pharmacologically active drugs interact with macromolecular targets, primarily proteins or nucleic acids. Surfaces of molecules, electrostatic force, hydrophobic contact, and hydrogen bond formation are all important components in such interactions. These parameters are mostly evaluated during the study and prediction of two-molecule interactions. Computer-aided drug design (CADD) is a type of computer technology that allows you to create a product while also documenting the process. CADD can help with the production process by transferring precise schematics of a product's materials, procedures, tolerances, and measurements in accordance with the product's specific conventions.[17] It can create two-dimensional or three-dimensional diagrams that may be rotated to be viewed from any angle, including from the inside looking out. Disease selection, target selection, lead compound identification, lead optimization, pre-clinical trial testing, clinical trial testing, and pharmacogenomic optimization are the seven basic processes in the drug discovery process from concept to market. In practice, the last five steps had to be repeated several times. Compounds for testing might be derived from natural sources (plants, animals, and microbes) or synthesized chemically. These compounds may be ruled out as potentials due to a lack of or poor activity, the presence

of toxicity or carcinogenicity, synthesis complexity, insufficient efficiency, and so on. As a result, just one out of every 100,000 compounds studied makes it to market, and the average cost of developing a new medication has risen to $800 million. The last steps of the process are becoming less time-consuming and expensive. Due to tight governmental standards on its implementation, a reduction in the time and cost of the final phases of drug testing is improbable. As a result, the majority of efforts to improve the efficiency of drug development are focused on the stages of ligand identification and optimization.[18] Bioinformatics technologies can assist find drug targets using in silico methodologies. They can also be used to assess the target structure for potential binding/active sites, produce candidate molecules, check for similarity, dock these molecules with the target, rank them by binding affinities, and further tune the molecule to improve binding characteristics. Computers and computational approaches are now used in virtually every area of drug discovery, and they are at the heart of (a) structure-based drug design and (b) ligand-based drug design.[19]

1.3.1. Structure Based Drug Design (SBDD)

The technique to be employed in drug design is structure-based drug design. The identification of novel medications is aided by structure-based drug design.[20] The following points are included in the structure-based drug design process:

- Crystallography should be used to determine the structure of the chosen target, which should be prepared in a solution form
- In order to determine the binding sites, a thorough examination of the structure is required

• Different compounds from databases should be docked at the binding site and their affinity for the site rated

• The compounds having the best affinity for the site are chosen

• Biochemical assays involve the use of Leads and Tests that bind to specific target locations

a) If the lead is found to be posing as an inhibitor at the site, then it should be analyzed by crystallography regarding its structure

b) It should be further tested for potency and bioavailability in order to launch it

c) In Structure Based Drug Design, the action of the leads can be modified or optimized which would ensure higher success rates

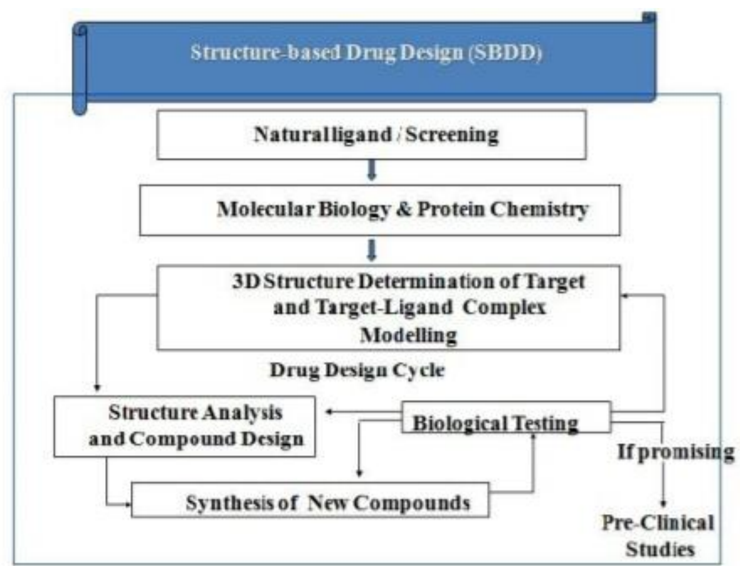


Figure 2: Structure Based Drug Design

1.3.2. Ligand Based Drug Design (LBDD)

The examination of ligands known to interact with a target is part of the ligand-based drug design strategy. These approaches analyze the 2D or 3D structure of a set of reference structures obtained from substances known to interact with the target of interest.[21] In some circumstances, such as when data on the 3D structure of a target protein is unavailable, drug design can instead be based on a procedure that starts with the known ligands of a target protein. This method is referred to as "ligand-based drug design".[22]

a) Ligand-Based Drug Design involves knowing which molecules bind to the desired target site

b) These compounds could be utilized to construct a pharmacophore model

c) A pharmacophore model is a molecule that has the structural capabilities to bind to a specific target location

d) Once the Pharmacophore has been identified, it is tested to see if it is compatible with the receptor; if not, the Pharmacophore is further tweaked to make it a possible medication

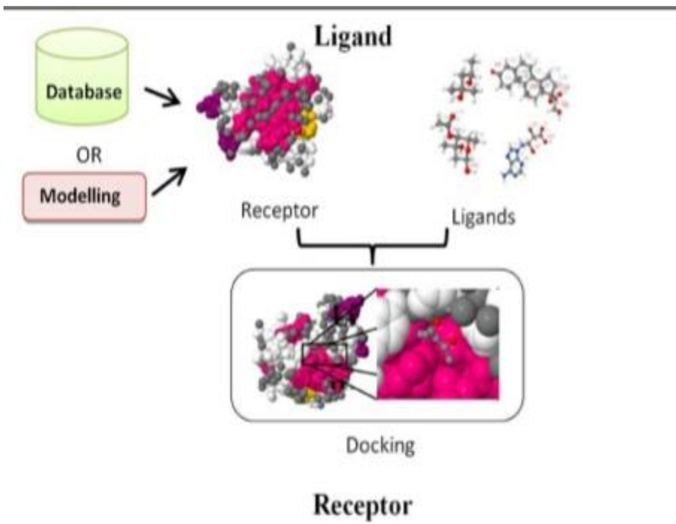


Figure 3: Ligand Based Drug Design

1.3.3. Benefits of CAAD

Cost-cutting: CADD is used by many biopharmaceutical businesses to cut costs. CADD has replaced traditional experimentation that required animal and human models, saving both time and money. It is believed that, in the case of diseases such as influenza, Computational Drug Designing will play a key role in minimizing drug resistance and, as a result, will lead to the development of lead compounds that target the causative cause. Potent hits can be obtained in a matter of weeks by utilizing computational approaches. CADD has also resulted in the development of high-quality datasets and libraries that can be optimized for molecular diversity and similarity.

1.3.4. Limitations of CAAD

CADD's advancement is hampered by a lack of accurate experimental data.

Some computer-aided drug design techniques take a long time, especially when seeking for the right lead component.

1.4. Molecular Docking

Molecular docking is becoming a more important method in drug development. The differences between available docking software and their performance are also explored. Flexible receptor molecular docking approaches, particularly those including receptor backbone flexibility, present a challenge for currently available docking methods. A new Local Move Monte Carlo (LMMC)-based technique is presented as a possible solution to flexible receptor docking difficulties. Three examples of drug discovery applications using molecular docking techniques are shown. With the completion of the human genome project, a growing number of new therapeutic targets for drug development have emerged. Simultaneously, high-throughput protein purification, crystallography, and nuclear magnetic resonance spectroscopy techniques have been developed, allowing researchers to learn more about the structural features of proteins and protein–ligand complexes.[23]

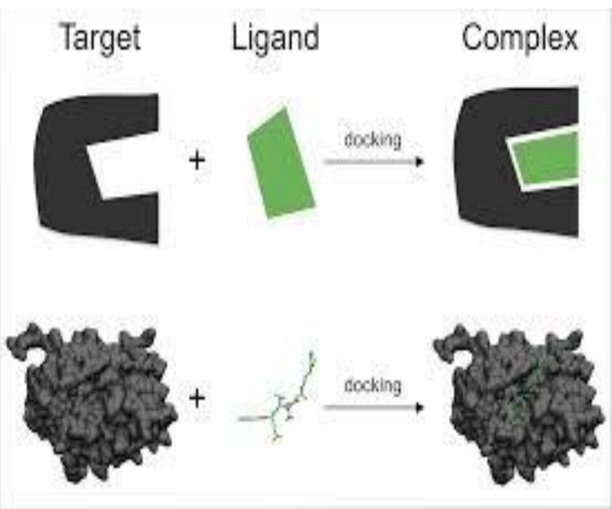


Figure 4: Schematic Diagram of Molecular Docking

These advancements have allowed computational methodologies to pervade many parts of drug development today, including virtual screening (VS) approaches for hit detection and lead optimization procedures. VS is a more direct and rational drug discovery strategy than traditional experimental high-throughput screening (HTS), with the added benefit of cheap cost and effective screening. VS can be divided into two types: ligand-based and structure-based. When a set of active ligand molecules is known but little or no structural information about targets is available, ligand-based approaches like pharmacophore modelling and quantitative structure activity relationship (QSAR) methodologies can be used. When it comes to structure-based drug design, molecular docking has been the most used strategy since the early 1980s. To undertake molecular docking investigations, programmes based on various algorithms have been created, making docking an increasingly significant tool in pharmaceutical research. Several great evaluations on docking have been written in the past, as well as numerous comparative studies to assess the programmes' relative performance. The molecular docking approach can be used to represent the atomic level interaction between a small molecule and a protein, allowing us to characterise small molecule behaviour in target protein binding sites as well as elucidate key biochemical processes. The docking procedure consists of two main steps: predicting the ligand structure as well as its position and orientation within these sites (known as pose) and determining the binding affinity. These two steps have to do with sample methods and scoring schemes, which will be covered in the theoretical section. Knowing where the binding site is before starting the docking process improves docking efficiency dramatically. In many cases, the binding site is identified before ligands are docked into it. Also, by comparing the target protein to a family of proteins with

comparable functions or proteins co-crystallized with other ligands, one can learn more about the locations. Cavity detection tools or web servers, such as GRID, POCKET, Surf Net, PASS, and MMC, can be used to locate probable active sites within proteins when the binding sites are unknown. Blind docking is the process of docking without making any assumptions about the binding location. The lock-and-key theory provided by Fischer, in which the ligand fits into the receptor like a lock and key, was the first explication of the ligand-receptor binding mechanism. The initial docking approaches were based on this principle, and the ligand and receptor were both considered as rigid entities. . The "induced-fit" idea proposed by Koshland extends the lock-and-key theory by claiming that as ligands engage with the protein, the active region of the protein is constantly altered by interactions with the ligands. This idea implies that during docking, the ligand and receptor should be viewed as malleable. As a result, it may be able to more correctly represent binding events than the rigid treatment.

Because of the lack of computer resources, docking has long been done with a flexible ligand and a rigid receptor, and it is still the most popular technique. Many efforts have recently been made to cope with receptor flexibility; yet, flexible receptor docking, particularly backbone flexibility in receptors, remains a key difficulty for currently available docking approaches. In this paper, we suggest the use of a Local Move Monte Carlo (LMMC) approach to solve flexible receptor docking difficulties.

1.4.1. Theory Of Docking

The goal of molecular docking is to use computer methods to anticipate the structure of the ligand-receptor complex. Docking is accomplished in two steps: first, sampling ligand

conformations in the active site of the protein, and then ranking these conformations using a scoring function. In an ideal world, sampling algorithms would be able to replicate the experimental binding mode, and the scoring function would give it the highest score out of all created conformations. We present a brief introduction of basic docking theory from these two approaches. It's difficult to model the interaction of two molecules. Hydrophobic, van der Waals, or stacking interactions between aromatic amino acids, hydrogen bonding, and electrostatic forces all play a role in intermolecular association. Because there are many degrees of freedom and little knowledge of the effect of solvent on the binding relationship, modelling the intermolecular interactions in a ligand-protein complex is difficult. The docking of a ligand to a binding site attempts to emulate the natural course of interaction between the ligand and its receptor by taking the least energy path. There are straightforward ways for docking rigid ligands with rigid receptors and flexible ligands with rigid receptors, but generic docking methods for conformationally flexible ligands and receptors are difficult to come by. The search method should generate as many combinations as possible that incorporate the empirically discovered binding modesDespite the fact that a thorough search algorithm would examine all conceivable binding modes between the two molecules, such a search would be unfeasible due to the breadth of the search space and the time it would take to finish. As a result, only a small portion of the overall conformational space may be sampled, necessitating a trade-off between computing cost and the amount of search space covered. Molecular dynamics, Monte Carlo methods, genetic algorithms, fragment-based, point complementary and distance geometry approaches, Tabu, and systematic searches are some of the most common searching algorithms. The scoring function, on the other hand, is a collection of

mathematical methods for predicting the strength of a non-covalent interaction known as binding affinity. . One of the most difficult problems in all computational techniques is the construction of an energy scoring function that can precisely and quickly explain the interaction between the protein and the ligand. In the ligand, there are several reviews on scoring. The docking technique for proteins and ligands is usually separated into two parts: rigid body docking and flexible docking.

Rigid Docking: In this approach, both the ligand and the receptor are treated as rigid, and only six degrees of translational and rotational freedom are explored, effectively eliminating any flexibility. The rigid body docking process is used as a first step in most docking suites.

Flexible Docking: A more popular method is to represent ligand flexibility while assuming a stiff protein receptor, focusing primarily on the ligand's conformational space. Protein flexibility, on the other hand, should ideally be considered, and various techniques have been developed in this regard. To deal with ligand flexibility, there are three types of algorithms: systematic approaches, random or stochastic methods, and simulation methods.[22] Protein flexibility may be the most difficult challenge in molecular docking due to their vast size and various degrees of freedom. Soft docking, side-chain flexibility, molecular relaxation, and protein ensemble docking are some of the strategies used to address protein flexibility.

1.4.2. Experimental Docking Procedures

There are several great evaluations of molecular docking methods available, as well as numerous studies comparing the performance of various molecular docking tools.

Following that, we'll go over the four-step process used to accomplish molecular docking in this work.

1.4.2.1. Choosing a target

In an ideal world, the target structure would be determined experimentally using X-ray crystallography or nuclear magnetic resonance, both of which may be downloaded from the PDB; however, docking has proven to be more successful than homology models or threading. The model must be of high quality. Validation tools such as Molprobity[23] can be used to test it. After choosing a model, it must be prepared by removing water molecules from the cavity, stabilising charges, filling in missing residues, and constructing side chains using the supplied parameters. At this time, the receptor should be biologically active and stable.

1.4.2.2. Ligand Selection And Preparation

The sort of docking ligands chosen will be determined by the objective. It can be retrieved from a variety chemical databases, such as ZINC or PubChem, or sketched using the Chemsketch tool.[24] Filters are frequently used to minimise the amount of molecules that must be docked. Net charge, molecular weight, polar surface area, solubility, commercial availability, similarity thresholds, pharmacophores, synthetic accessibility, and absorption, distribution, metabolism, excretion, and toxicological qualities are all examples of attributes. Many times, researchers create their own molecules, such as those created by us in the example given in section 5 of this paper.

1.4.2.3. Docking

The ligand gets docked onto the receptor and the interactions are verified in this final stage. The scoring function calculates a score based on the best ligand selected.

1.4.2.4. Evaluating Docking Results

The root-mean-square deviation (RMSD) between the experimentally observed heavy-atom locations of the ligands and the one(s) anticipated by the algorithm is commonly used to assess the efficacy of docking algorithms in predicting ligand binding poses. The system's flexibility presents a significant difficulty in the search for the ideal stance. The amount of degrees of freedom incorporated in the conformational search is a key factor in determining the effectiveness of the search.[25] When the RMSD is less than 2Amstrong, it is considered to be a good performance.

1.4.2.5. Docking Software Description

There are a plethora of algorithms for assessing and rationalising ligand-protein or protein-protein interactions, and the number is growing all the time. In docking approaches, speed and accuracy are critical elements for achieving successful results. Several algorithms have similar approaches with novel expansions aimed at achieving a fast method with the highest possible accuracy. AutoDock[26], DOCK, FlexX, GOLD, ICM, ADAM, DARWIN, DIVALI, and DockVision are some of the most popular docking programmes.

AIM AND SCOPE OF THE STUDY

Epidermal growth factor is considered as a valid target in the clinical trials of anticancer therapy and MAPk pathway has a critical role in stimulating proliferation and suppressing apoptosis. In present work, 50 ligands were selected and treated with **5D41**and **279Q** proteins respectively and docking was carried out computationally to find out a better ligand which binds the protein and thus it can be used in cancer treatments. Higher the negative binding energy, higher its affinity to bind with the protein.

CHAPTER 2

LITERATURE REVIEW

Epidermal growth factor has a valid target in the clinical trials of anticancer therapy and MAPk pathway has a critical role in stimulating proliferation and expressing apoptosis.

In the article “Epidemal Growth factor receptor inhibition strategies in Oncology” by P M Harari, explains that in contrast to traditional chemotherapy and radiotherapy, molecular targeted techniques for cancer therapy have the potential to offer greater tumour selectivity. The epidermal growth factor receptor was one molecular target in oncology with great potential (EGFR). In many epithelial malignancies, the EGFR was overexpressed, dysregulated, or mutated, and EGFR activation seemed crucial for the development and growth of tumours. Our knowledge of the precise contributions of EGFR signaling networks to cancer behavior was continually being refined by developments in signal transduction biology. Monoclonal antibodies (mAbs) that target the extracellular domain of EGFR, like cetuximab (Erbitux), and small molecule tyrosine kinase inhibitors (TKIs) that target the receptor catalytic domain of EGFR, like gefitinib (Iressa) and erlotinib, are the two main classes of EGFR inhibitors that have been developed (Tarceva). In preclinical model systems, EGFR inhibitors' mechanisms of action have been studied. Clinical trials have evaluated safety, activity, pharmacokinetics, and pharmacodynamics. The toxicity profiles of anti-EGFR mAbs and TKIs somewhat overlap, but they have different delivery methods, serum half-lives, and consequently dose regimens. Since the FDA recently approved cetuximab and gefitinib for use in the treatment of lung and colorectal cancers, respectively, both classes of drugs have

demonstrated unmistakable anticancer effect. To identify the tumours and patients who will respond predictably to EGFR inhibitor approaches was a major challenge in anti-EGFR oncology therapeutics, which was highlighted by the absence of a survival benefit for EGFR TKIs in combination with chemotherapy in large-scale phase III lung cancer trials in 2003. EGFR TKI sensitivity appeared conferred by newly discovered mutations in the EGFR catalytic domain, which bodes well for future research into response prediction. In order to advance, it would be necessary to increase our molecular understanding of the overall EGFR signaling network and our methodologies for determining how much each tumour depends on the EGFR signaling pathways for its ability to develop. A survival advantage for the use of EGFR inhibitors in combination with high-dose radiation in the treatment of head and neck cancer and refractory lung cancer, respectively, has been confirmed by results from recently reported phase III trials in 2004. In the upcoming years, it would be likely that EGFR inhibitors (together with other logically created molecular growth inhibitors) contribute significantly to the treatment of cancer.

The article "The BAD protein integrates survival signaling by EGFR/MAPK and P13K/Akt kinase pathways in PTEN-deficient tumour cells by Qing-Bai She, David B Solit, Qing Ye, Kathryn E O'Reilly, joseLobo and Nael Rosen explained that Mutant PTEN causes the proliferation of tumour cells to occur independently of EGFR. PTEN induction made cells more susceptible to EGFR inhibition, and the two together promote synergistic apoptosis. Synergy was caused by two parallel pathways that phosphorylate the proapoptotic protein BAD at different locations being inhibited. Serine 136 phosphorylation was dependent on PI3K/Akt, whereas serine 112 phosphorylation was

dependent on EGFR/MEK/MAPK. It took only one of two phosphorylations to sequester BAD in 14-3-3. Only after both serines were dephosphorylated in response to the blockage of both routes does BAD get released and apoptosis was brought about. In response to pathway obstruction, RNA interference-mediated BAD expression reduction inhibits apoptosis. BAD thereby combines both pathways' antiapoptotic actions. Combining EGFR and PI3K signaling suppression could be a beneficial therapeutic approach.

The article "Conservation of protein abundance patterns reveals the regulatory architecture of the EGFR/MAPK pathway by Tujin shi,Mario Niepel, Jason E mcdermott, Yuqian Gao, Carrie D N icora, William B Chrisler, Lye M Markillie,Vladislav A petyuk, Richard D smith, Karin D Rodland, Peter K Sorger, Wei-Jun Qian and H. Steven Wiley explained that Mutant PTEN causes the proliferation of tumour cells to occur independently of EGFR. PTEN induction made cells more susceptible to EGFR inhibition, and the two together promote synergistic apoptosis. Synergy was caused by two parallel pathways that phosphorylate the proapoptotic protein BAD at different locations being inhibited. Serine 136 phosphorylation was dependent on PI3K/Akt, whereas serine 112 phosphorylation was dependent on EGFR/MEK/MAPK. It took only one of two phosphorylations to sequester BAD in 14-3-3. Only after both serines were dephosphorylated in response to the blockage of both routes does BAD get released and apoptosis brought about. In response to pathway obstruction, RNA interference-mediated BAD expression reduction inhibits apoptosis. BAD thereby combines both pathways' antiapoptotic actions. Combining EGFR and PI3K signaling suppression could be a beneficial therapeutic approach. Contrarily, the presence of EGFR and transcriptionally

regulated feedback regulators was very varied. Most core proteins had an absolute abundance of 50,000–70,000 copies per cell, whereas the adaptors SOS1, SOS2, and GAB1 were present in much lower concentrations (2000 to 5000 copies per cell). All cells between 3000 and 10,000 occupied EGFRs displayed MAPK signaling saturation, supporting the notion that adaptors restrict signaling. Our findings imply that the relative stoichiometry of the key proteins in the MAPK pathway is remarkably consistent across different cell types, with cell-specific variations mostly confined to the varying concentrations of feedback regulators and receptors. Previous findings that just a portion of the total cell surface EGFR is expressed could be explained by the low abundance of adaptors relative to EGFR.

CHAPTER 3

MATERIALS AND METHODS

Ligands were selected and its binding capability towards proteins 5D41 and 2Y9Q were observed using software's and there binding energy can be calculated. The following tools used for present work

3.1. Marvin sketch

Marwin sketch features an extensive set of functionalities to enable the fast and accurate drawing of chemicals compounds, reactions, Markus structures and query molecules. Furthermore Marwin sketch has built-in-structure and valence checkers to provide guidance, and integrated property calculators to pull live results-upon your request.

3.2. Protein data bank

The Protein data bank is the single worldwide archive of structural data of biological macromolecules. The Protein data bank (PDB) was established at Brookhaven National Laboratories (BNL) (1) in 1971 as an archive for biological macromolecular crystal structures. In the beginning the archive held seven structures, and with each year a hand full more were deposited. In the 1980s the number of deposited structures began to increase dramatically.

3.3. Maestro 12.3

Maestro is a major component of transformative drug discovery. It is used to optimize the design-make –test- analyse cycle that leads to better quality compounds faster. Better

quality molecules can be made in less time. It is a stream lined portal for structural visualization. It helps in target validation and structure enablement, Hit Discovery, Lead optimization.

DOCKING

3.4. Biovia Discovery Studio

BIOVIA Discovey studio brings together over 30 years of peer-reviewed research and world class Insilco techniques such as molecular mechanics, free energy calculations, bio therapeutics develop ability and more into a common environment. It provides researchers with a complete toolset to explore the nuances of protein chemistry and catalyze discovery of small and large molecule therapeutics Target ID to Lead Optimization.

3.5. Auto dock

Auto dock is a suite of automated docking tools. It is designed to predict how small molecules, such as substrate or drug candidates, bind to a receptor of known 3D structure. Current distributions of Auto dock consist of two generations of software: Auto Dock 4 and Auto Dock Vina. Auto Dock 4 actually consists of two main programs: Auto dock performs the docking of the ligand to a set of grids describing the target protein; autogrid precalculates these grids. In addition to using them for docking, this can help, for example, to guide organic synthetic chemists design better binders. Auto Dock Vina does not require choosing atom types and pre-calculating grid maps for them. Instead, it calculates the grids internally, for the atom types that are needed, and it does not virtually instantly.

3.6. Cygwin

Cygwin is a useful free tool for the stimulation of Linux like environment for Windows. It is a free software that provides Unix-like environment and software tool set to uses of any modern x85 32-bit and 64-bit versions of MS-Windows.

3.7. 5D41 (EGFR KINASE DOMAIN IN COMPLEX WITH MUTANT SELECTIVE ALLOSTERIC INHIBITOR)

5D41 is a 2-chain structure with sequence from Human. One of the receptor tyrosine kinases (RTKs), known as the epidermal growth factor receptor (EGFR or HER1), ErbB1 (HER1 or EGFR), ErbB2 (HER2 or Neu), ErbB3 (HER3), and ErbB4 (HER4). They have comparable fundamental structures. Such as a -helix transmembrane, an extracellular ligand binding, or a cytoplasmic tyrosine Except for ErbB3, kinases also have carboxy-terminal signalling domains. The growth elements are split into three classes, including the particular ligand of ErbB1or EGFR, those that bind to ErbB1 and ErbB4 (such as EGF, TGF-, and amphiregulin), (such as betacellulin, epiregulin, and heparin-binding EGF), and ErbB3/ErbB4 (such as (Neuregulins, Heregulins). But unlike the others, ErbB2 lacks a particular ligand. Any ligand that resembles the one that can activate it will cause it to bind and create a dimer. With other ErbB family members, the ligand-binding mechanism produces homodimerization and heterodimerization.has tyrosine kinase domain membership and activation. This intracellular pathway is activated by autophosphorylation and allows it to communicate with signalling components.to signalling pathways downstream. Additionally, this was done to promote cell

proliferation via the RASRAF-MEK-MAPK, PI3KPTEN-AKT, and STAT pathways also prevent apoptosis.

In the majority of cases of lung cancer, EGFR signalling was seen. This receptor is overexpressed in 40 to 89% of NSCLC cases, with squamous tumours and adenocarcinomas having the highest and lowest occurrence rates, respectively. The amount of EGFR activation is associated with a poor prognosis and a low rate of tumour regression in addition to its high rate in NSCLC. The EGFR mutation and overexpression, which are frequently detected in tumour cells, are two ways that might cause phosphorylation activity in the kinase domain of EGFR to be uncontrollably increased. This suggests that incorrect tyrosine kinase activation promotes tumour growth and prevents cell death. Numerous investigations for mutant EGFR that caused autophosphorylation and lacked some of the residues in the extracellular domain found spontaneous oligomerization as well. In order to change and increase cell adhesion, motility, and invasion as well as induce metastases, EGFR also interacts with the integrin pathway and activates matrix metalloproteinases. Additionally, several NSCLC instances have overexpressed or activated EGFR mutations, which results in constitutive TK activity. Due to this, it is a logical target for therapeutic intervention, and it also encourages the creation of new anticancer drugs that target.

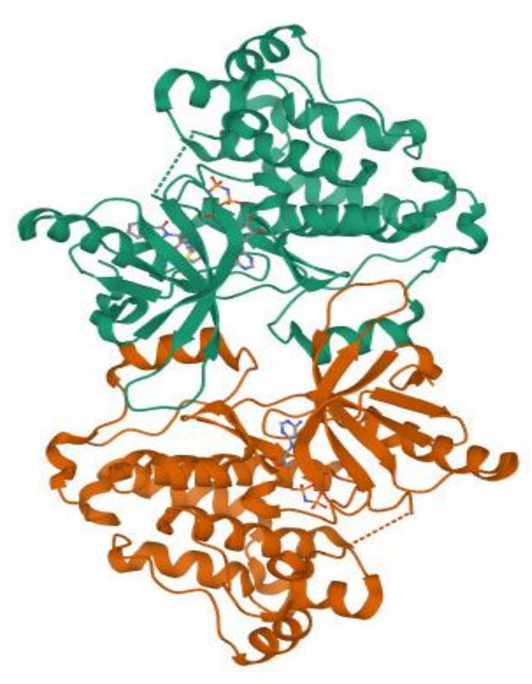


Figure 5 : 5D41 Protein

Lung cancer (LNCR) is linked to EGFR mutations. A typical cancer that affects lung tissues is LNCR. Non-small cell lung cancer (NSCLC), the most prevalent kind of lung cancer, can be further broken down into three major histologic subtypes: squamous cell carcinoma, adenocarcinoma, and large cell lung cancer. NSCLC frequently has a bad prognosis and is discovered at an advanced stage. Receptor tyrosine kinase binds to EGF family ligands and initiates a number of signalling cascades to transform extracellular signals into the proper cellular reactions. EGF, TGFA/TGF-alpha, amphiregulin, epigen/EPGN, BTC/betacellulin, epiregulin/EREG, and HBEGF/heparin-binding EGF are examples of known ligands. Receptor homo- and/or heterodimerization and autophosphorylation on significant cytoplasmic residues are triggered by ligand interaction. Adapter proteins like GRB2 are attracted by the phosphorylated receptor, which then initiates intricate downstream signalling cascades. Activates the RAS-RAF-

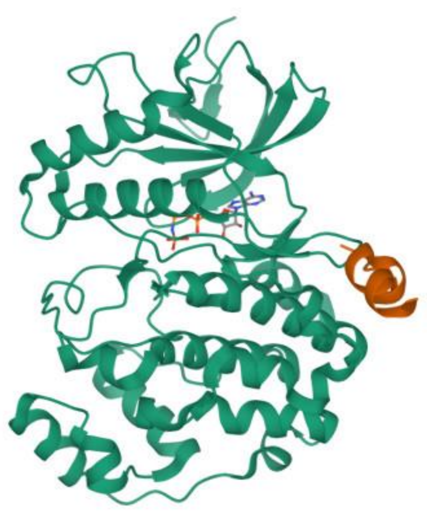
MEK-ERK, PI3 kinase-AKT, PLCgamma-PKC, and STATs modules, as well as at least four other important downstream signalling cascades. NF-kappa-B signalling may also be activated. Additionally, it directly phosphorylates other proteins, such as RGS16, stimulating its GTPase activity and likely tying the EGF receptor signal to the G protein-coupled receptor signal. Additionally enhances MUC1's interaction with SRC and CTNNB1/beta-catenin by phosphorylating it Isoform 2 might inhibit the effects of EGF.

3.8. 2Y9Q (Crystal structure of human ERK2 complexed with a MAPK docking peptide)

2Y9Q is a 2-chain structure with sequence from Human. Mitogen-activated protein kinases (MAPKs) regulate diverse aspects of cellular life such as cell division, differentiation, or apoptosis and interact with proteins through linear binding motifs. MAPKs are part of multitiered kinase cascades in which extracellular signal–regulated kinase (ERK) 1/2, c-Jun N-terminal kinase (JNK), and p38 are activated by dedicated mitogenactivated protein kinase kinases (MAP2Ks) MKK1/2, MKK4/7, and MKK3/ 4/6, respectively. These upstream activators bind to the same surface on their cognate MAPK as downstream MAPK substrates, inactivating phosphatases and protein scaffolds by their independently evolved linear motifs. Within the diverse repertoire of physical links observed in signaling networks. The simplest method to improve the specificity of promiscuous active sites is to use docking (D) motifs, which are linear motifs that directly attach to the catalytic domains of signalling enzymes. Similar to many signalling enzymes, MAPKs have active sites that are promiscuous. As a result, new protein interaction sites are crucial for figuring out how MAPK pathways are wired. The binding

surface and active site are separate during this form of contact, known as docking for protein kinases and phosphatases. For instance, MAPKs bind to D motifs, which are 8 to 12 amino acid-long segments satisfying a loose consensus sequence, and phosphorylate serine and threonine residues that are followed by proline (referred to as a S/TP target site or motif). This allows these enzymes to: interacts other signalling proteins by physical interactions.

The high degree of similarity between the docking grooves of paralogous MAPKs is probably due to the fact that each major group of MAPKs in animals—typical MAPKs like ERK1/2, JNKs, p38s, and ERK5 and atypical MAPKs like ERK3, ERK4, and ERK7—emerged through whole genome or individual gene duplication events and those three members of this family—ERK2, p38a, and JNK1



.

Figure 6 : 2Y9Q Protein

A crucial part of the MAP kinase signal transduction pathway is the serine/threonine kinase. The two MAPKs that are crucial to the MAPK/ERK cascade are MAPK1/ERK2 and MAPK3/ERK1. Additionally, they take part in a signalling cascade that is started by KIT and KITLG/SCF activation. The MAPK/ERK cascade controls transcription, translation, and cytoskeletal rearrangements to mediate a variety of biological processes, including cell growth, adhesion, survival, and differentiation, depending on the cellular context. By phosphorylating a number of transcription factors, the MAPK/ERK cascade also contributes to the start and control of meiosis, mitosis, and postmitotic processes in differentiated cells. For ERKs, about 160 substrates have previously been identified. Numerous of these substrates are found in the nucleus and appear to be involved in the control of transcription in response to stimulus. The cytosol and other cellular organelles include additional substrates, though, and those are in charge of procedures like translation, mitosis, and death. Additionally, the MAPK/ERK cascade is essential for the regulation of endosomal dynamics, including lysosome processing, endosome cycling via the PNRC, and Golgi apparatus fragmentation during mitosis. The molecules that make up the substrates include transcription factors (like ATF2, BCL6, ELK1, ERF, FOS, HSF4 or SPZ1), cytoskeletal components (like CANX, CTTN, GJA1, MAP2, MAPT, PXN, SORBS3 or STMN1), regulators of apoptosis (like BAD, BTG2, CASP9, DAPK1, IER3, MCL1 or PPARG), regulators of translation (like ARHGEF2, DCC, FRS2 or GRB10). Other substrates that allow the MAPK/ERK signal to spread to other cytosolic and nuclear destinations, increasing the specificity of the cascade, include protein kinases (such as RAF1, RPS6KA1/RSK1, RPS6KA3/RSK2, RPS6KA2/RSK3, RPS6KA6/RSK4, SYK, MKNK1/MNK1, MKNK2/MNK2, RPS6KA5/MSK1,

RPS6KA4/MS Possible contribution to the spindle assembly checkpoint plays the role of a transcriptional repressor. binds to the consensus sequence [GC]AAA[GC]. stifle the expression of the genes that interferon gamma induces. seems to bind to the promoters of STAT1, LAMP3, OAS1, OAS2, and IFIH1, IFITM1, CCL5, DMP1, and IFIH1. Kinase activity does not affect transcriptional activity. Possibly affect how the body reacts to cytokines and environmental stress. appears to control translation by phosphorylating EIF4E, enhancing its affinity for the mRNA cap containing 7-methylguanosine.

<u>3.9. STRUCTURE OF LIGANDS</u>

Ligand 1	**Ligand 2**
Ligand 3	**Ligand 4**
Ligand 5	**Ligand 6**
Ligand 7	**Ligand 8**
Ligand 9	**Ligand 10**

Ligand 11	**Ligand 12**
Ligand 13	**Ligand 14**
Ligand 15	**Ligand 16**
Ligand 17	**Ligand 18**
Ligand 19	**Ligand 20**

Ligand 21	**Ligand 22**
Ligand 23	**Ligand 24**
Ligand 25	**Ligand 26**
Ligand 27	**Ligand 28**
Ligand 29	**Ligand 30**

Ligand 31	**Ligand 32**
Ligand 33	**Ligand 34**
Ligand 35	**Ligand 36**
Ligand 37	**Ligand 38**
Ligand 39	**Ligand 40**

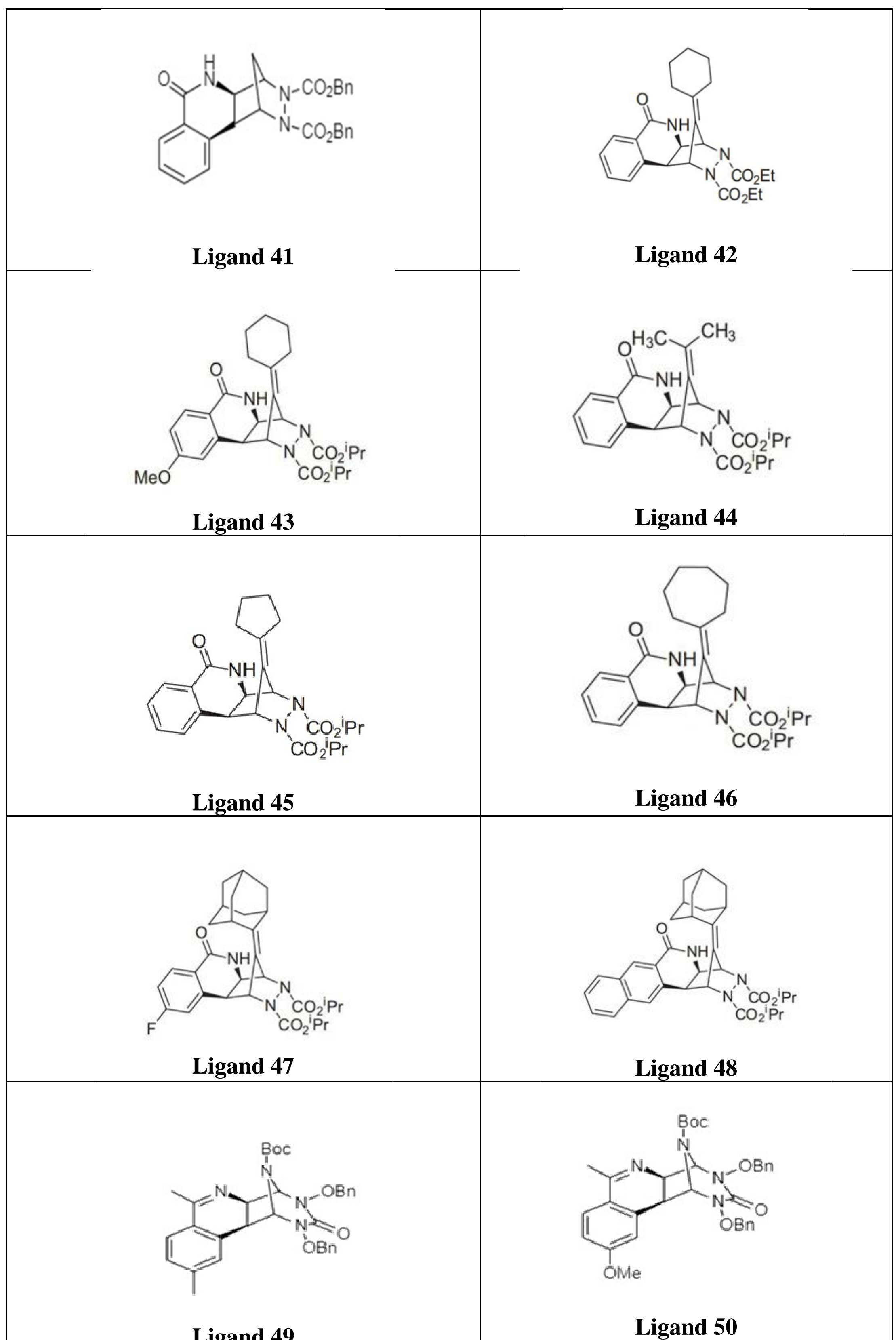


Table 1. Structure of ligands

3.10. METHODOLOGY

Proteins associated with EGFR and MAPK pathway; 5D41 and 2Y9Q was retrieved from protein data bank. The protein was downloaded in PDB format. Then using Maestro 12.3 the protein was splitted into water, ligands etc., and the ligands and proteins are saved as .pdb format. This protein was cleared in Discovery studio and all hetero atoms and unwanted ligands were deleted in this and saved as protein prepared .pdb. The above saved protein prepared .pdb is converted into pdbqt. Similarly, our ligands (given structure) which was drawn using Marvin sketch are cleaned in 2D and then in 3D. Both proteins and ligands were saved in pdbqt format using Autodock. The active site residue for 5D41 protein is taken as GLN791 and for 2Y9Q is taken as GLU109 which are selected by reading various research papers. The grid box was generated with coordinates (-32.253,26.975,11.503) for 5D41 protein and (62.226, 37.597, 10.337) for 2Y9Q with dimensions 60 x 60 x 60 A^o. Then are saved as .gpf format. Similarly docking is done with 50 GENETIC algorithm runs and saved as .dpf format. Using Cygwin, the commands were given and docking was completed for 50 runs. 2D interaction diagram was observed for further investigation.

CHAPTER 4

RESULTS AND DISCUSSION

Molecular docking was carried out using a software Auto dock. The two proteins were selected (**5D41** and **2Y9Q**), and natural ligand was isolated. The 50 ligands were used for the docking process. In docking the active site of proteins was blocked using the set of ligands. Blocking will inhibit further cell division and growth by preventing the mechanism. Auto dock provide more information about the binding energy in which the ligand perfectly fit to the protein. The binding energy obtained with the Auto dock software is given in table 1 and 2 respectively.

From the table we can say that different ligands obtained different binding energy in Auto dock. Using Auto dock calculations of protein **5D41**, **ligand 17** (-12.05 kcal/mol) has minimum amount of binding energy. **Ligand 39** has maximum binding energy (-5.19 kcal/mol) and hence minimum stability among the ligands.

Ligand Name	Binding Energy (kcal/mol)	Ligand Name	Binding Energy (kcal/mol)
Ligand 1	-10.34	Ligand 26	-10.99
Ligand 2	-11.06	Ligand 27	-11.49
Ligand 3	-11.40	Ligand 28	-6.41
Ligand 4	-11.63	Ligand 29	-7.89
Ligand 5	-10.78	Ligand 30	-6.49
Ligand 6	-11.09	Ligand 31	-8.15
Ligand 7	-10.60	Ligand 32	-6.21
Ligand 8	-10.54	Ligand 33	-7.72
Ligand 9	-11.12	Ligand 34	-6.61
Ligand 10	-10.53	Ligand 35	-8.32
Ligand 11	-10.85	Ligand 36	-6.29
Ligand 12	-7.83	Ligand 37	-6.77
Ligand 13	-8.48	Ligand 38	-5.51
Ligand 14	-10.46	Ligand 39	-5.19
Ligand 15	-11.27	Ligand 40	-5.68
Ligand 16	-11.23	Ligand 41	-7.39
Ligand 17	**-12.05**	Ligand 42	-7.44
Ligand 18	-11.62	Ligand 43	-7.53
Ligand 19	-11.27	Ligand 44	-7.40
Ligand 20	-11.60	Ligand 45	-7.32
Ligand 21	-10.68	Ligand 46	-7.55
Ligand 22	-10.98	Ligand 47	-6.18
Ligand 23	-11.07	Ligand 48	-6.69
Ligand 24	-11.56	Ligand 49	-8.78
Ligand 25	-11.21	Ligand 50	-5.85

Table 2: Binding energy obtained from Auto dock with **5D41** protein

Ligand 17 (**Figure 7**) shows various interactions like Van der waals interaction, conventional hydrogen bond, carbon hydrogen bond, π-donor hydrogen bond, pi-sulfur, alkyl, π-alkyl interactions. Van der waals interactions are represented by Gly796, Leu718, Leu792, Ile744, Ile789, Leu788, Thr854, Asp855, Phe723, Asp837, Ala722, Gly721, Asp800. Protein form conventional hydrogen bonding with Lys745, Met793, Cys797 and carbon hydrogen bond is formed with Gln791, Arg841, Asn842. π-sulfur bond with Met790, Alkyl and π-alkyl bond with Leu844, Val726, Ala743. The amount of binding energy obtained is -12.05 kcal/mol.

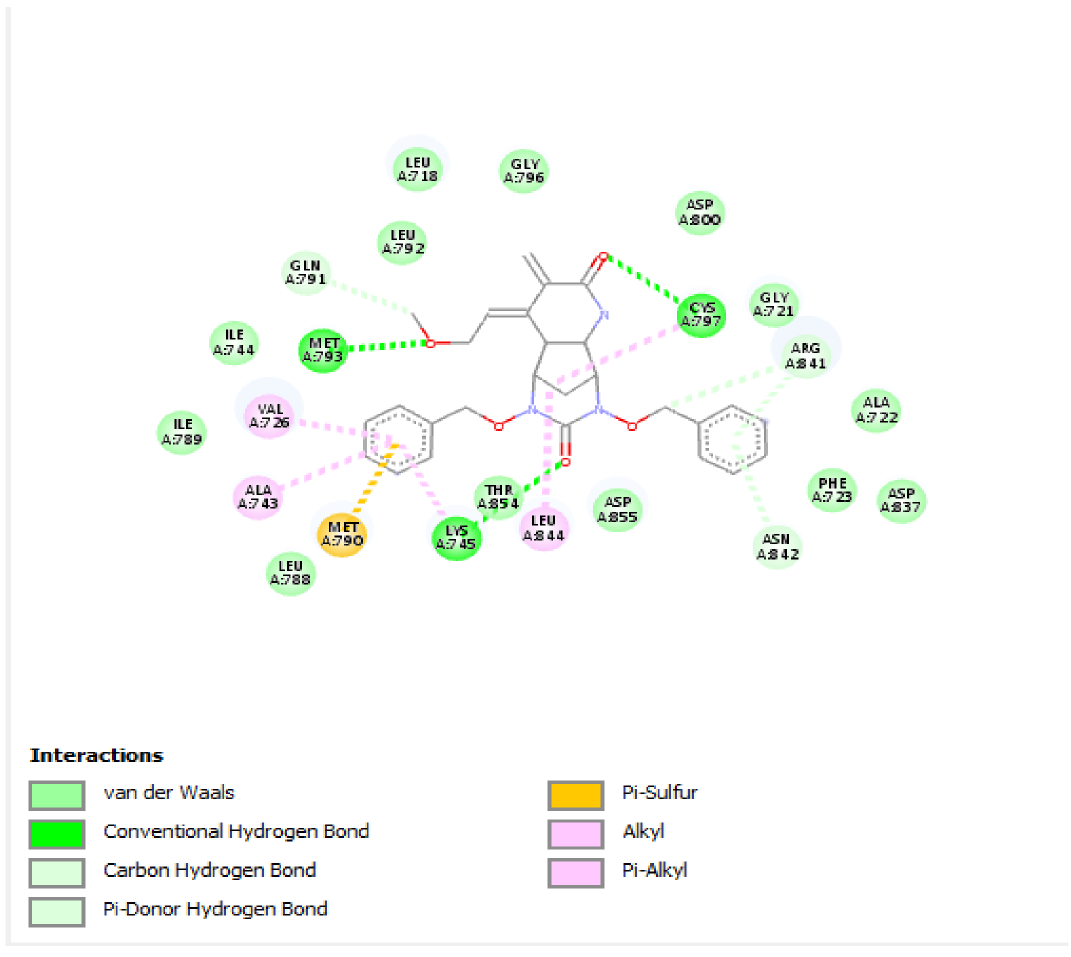


Figure 7 : 2D interaction of protein **5D41** with **ligand 17**

Ligand 39 (**Figure 8**) shows various interactions like Van der waals interaction, conventional hydrogen bond, halogen (fluorine) bond, alkyl bond and an unfavorable bump. Van der waals interactions are shown by Val774, Cys775, Lys852, Lys846, Glu1005, Asn771, Asp770, Ala767 and conventional hydrogen bond with Arg776, Gln791. Leu703, Leu778 shows alkyl interaction and Val769 shows a halogen(fluorine) bond. Here an unfavorable bump is shown by Unk0. The amount of binding energy obtained is -5.19 kcal/mol.

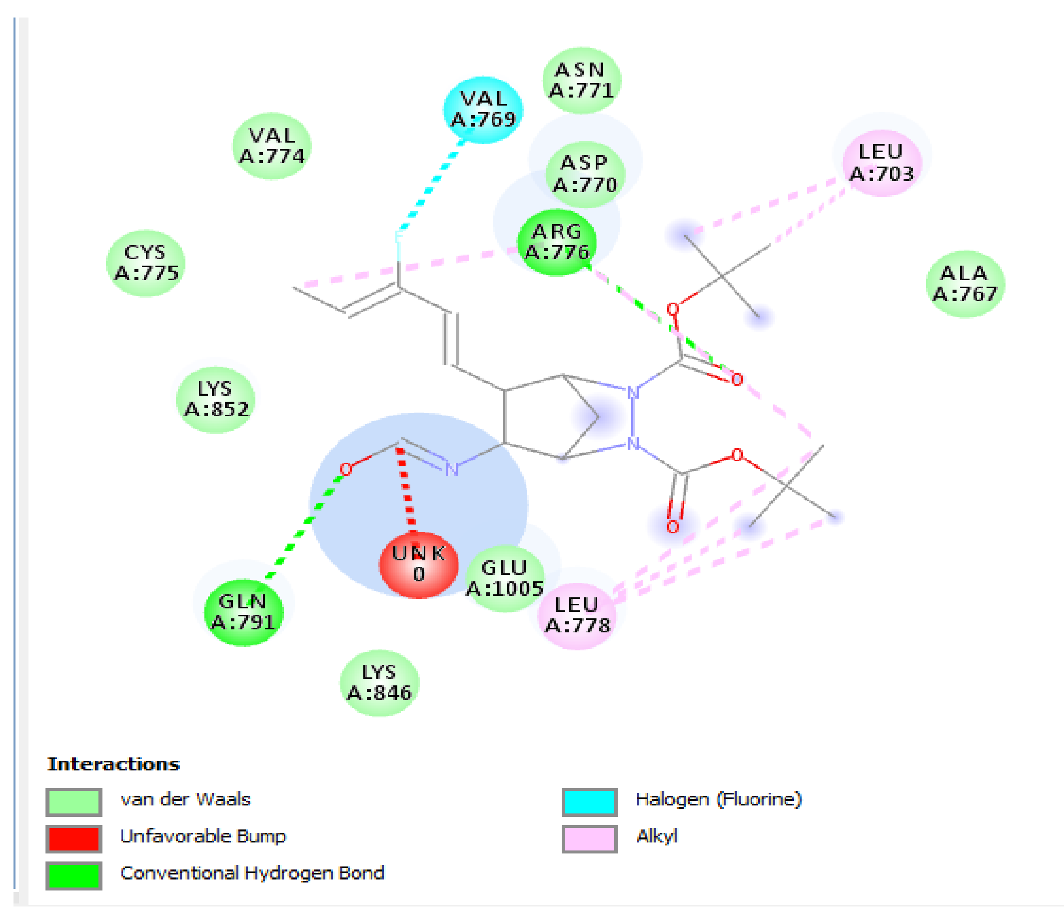


Figure 8 : 2D interaction of protein **5D41** with **ligand 39**

Ligand Name	**Binding Energy (kcal/mol)**	**Ligand Name**	**Binding Energy (kcal/mol)**
Ligand 1	-7.99	Ligand 26	-9.25
Ligand 2	-7.98	Ligand 27	-8.01
Ligand 3	-8.38	Ligand 28	-6.24
Ligand 4	-8.08	Ligand 29	-7.02
Ligand 5	-8.32	Ligand 30	-6.20
Ligand 6	-8.24	Ligand 31	-6.92
Ligand 7	-8.26	Ligand 32	-6.12
Ligand 8	-8.09	Ligand 33	-6.55
Ligand 9	-8.37	Ligand 34	-6.36
Ligand 10	-8.24	Ligand 35	-6.69
Ligand 11	-8.53	Ligand 36	-5.96
Ligand 12	-6.88	Ligand 37	-7.22
Ligand 13	-6.77	Ligand 38	-6.02
Ligand 14	-8.04	Ligand 39	-5.83
Ligand 15	**-9.50**	Ligand 40	-5.28
Ligand 16	-7.88	Ligand 41	-7.56
Ligand 17	-7.53	Ligand 42	-6.73
Ligand 18	-8.55	Ligand 43	-7.19
Ligand 19	-7.77	Ligand 44	-7.12
Ligand 20	-7.58	Ligand 45	-7.49
Ligand 21	-8.78	Ligand 46	-7.44
Ligand 22	-8.82	Ligand 47	-7.08
Ligand 23	-8.37	Ligand 48	-7.85
Ligand 24	-9.16	Ligand 49	-7.92
Ligand 25	-8.36	Ligand 50	-7.68

Table 3: Binding Table obtained from Auto dock with **2Y9Q** protein

Using Auto dock calculations of protein 2Y9Q, **ligand 15** (-9.50 kcal/mol) has minimum amount of binding energy. **Ligand 40** has maximum binding energy (-5.28 kcal/mol) and hence minimum stability among the ligands.

Ligand 15 (**Figure 9**) shows various interactions like van der waals interaction, conventional Hydrogen Bond, π-cation bond and π-alkyl bond. Van der waals interactions are represented by Arg447, Arg443, Tyr139, His80, Asp106, Gly85, Ile84, Ile83, Glu81. Protein form conventional hydrogen bond with Lys164 and π-alkyl bond with Arg79, Pro 323. The amount of binding energy obtained is -9.50 kcal/mol.

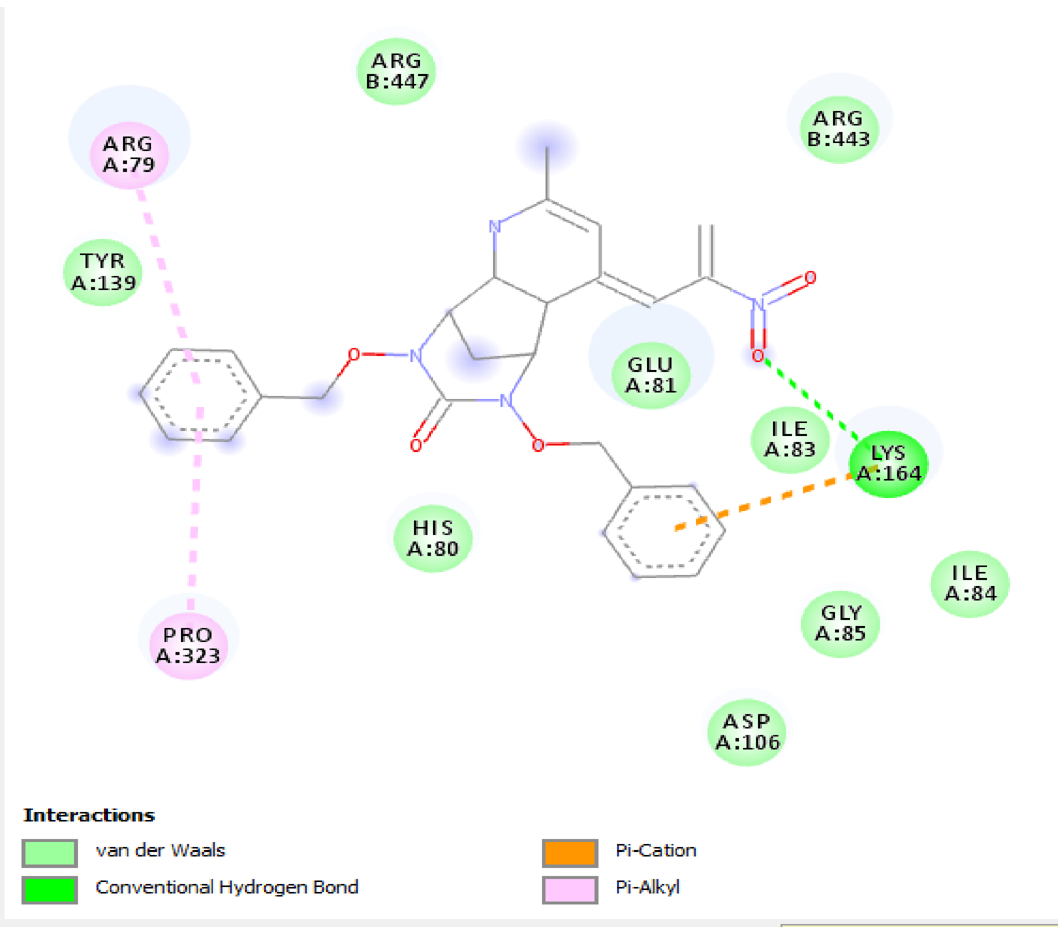


Figure 9 : 2D interaction of protein **2Y9Q** with **ligand 15**

Ligand 40 (**Figure 10**) show interactions like Van der waals interaction, conventional hydrogen bond, alkyl bond and an unfavorable bump. Protein shows van der waals interaction with Glu81, His80, Gly85, Ile84, Ile83, Phe78, Asn87 and conventional hydrogen bond with Lys164, Asp106 and an alkyl bond with Arg79. An unfavorable bump is shown by Unk0. The amount of binding energy obtained is -5.28 kcal/mol.

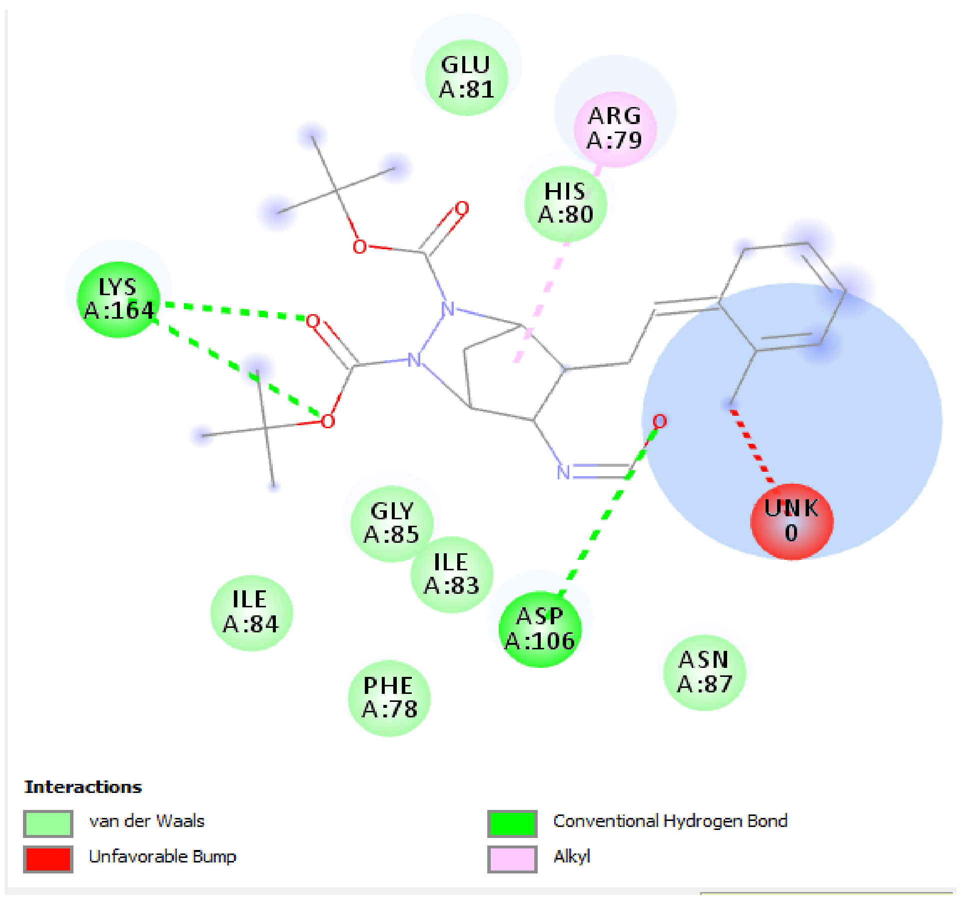


Figure 10 : 2D interaction of protein **2Y9Q** with **ligand 40**

The instability of protein with **ligand 40** and **ligand 39** is due to the presence of an un favorable bump. The bump increases the binding energy and thus leads to the decrease in stability. The common interaction seen in **ligand 15** and **17** which has minimum binding energies are conventional hydrogen bond, π-alkyl bond, Van der waals forces. These forces help the ligand to bind with the proteins. The binding energies of two ligands are -12.05 and -9.50 kcal/mol respectively.

CHAPTER 5

CONCLUSION

The study of molecular docking using Autodock is conducted using two proteins **5D41** and **2Y9Q**. Lower the binding energy greater its affinity to bind with the protein. Each ligand shows different binding energy due to presence of various functional group. Binding energy calculations revealed that **ligand 17** for **5D41** protein and **ligand 15** for **2Y9Q** protein has obtained the least binding energy from the results obtained after conducting docking with the provided ligands and **ligand 39** and **ligand 40** gives high binding energy for respective proteins. The presence of an unfavorable bump decreases the instability of proteins with respective ligands.

REFERENCE

1. P. J. Roberts, C. J. Der, Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer, *Oncogene 26*, 3291–3310 (**2007**).
2. A. Citri, Y. Yarden, EGF–ERBB signalling: Towards the systems level, *Nat. Rev. Mol. Cell Biol.* 7, 505–516 (**2006**).
3. R. I. Nicholson, J. M. W. Gee, M. E. Harper, EGFR and cancer prognosis, *Eur. J. Cancer* 37 (Suppl. 4), S9–S15 (**2001**).
4. D. J. Slamon, G. M. Clark, S. G. Wong, W. J. Levin, A. Ullrich, W. L. McGuire, Human breast cancer: Correlation of relapse and survival with amplification of the HER-2/neu oncogene, *Science* 235, 177–182 (**1987**).
5. W. Pao, J. Chmielecki, Rational, biologically based treatment of EGFR-mutant non-small cell lung cancer, *Nat. Rev. Cancer*, 10, 760–774 (**2010**).
6. Herbst R S,"Review of epidermal growth factor receptor biology", *International Journal of Radiation Ongology, Biology, physics*. **59** (2 Suppl), 21–6 (**2004)**
7. Zhang H, Berezov A, Wang Q, Zhang G, Drebin J, Murali R, Greene M I, *The Journal of Clinical Investigation*, 117 (8), 2051–8 (**2007**).
8. Yarden Y, Schlessinger J, "Epidermal growth factor induces rapid, reversible aggregation of the purified epidermal growth factor receptor", *Biochemistry*, 26(5), 1443–51 (**1987**).
9. Maruyama I N, "Mechanisms of activation of receptor tyrosine kinases: monomers or dimers", *Cells,* 3(2), 304–30, (**2014**).

10. Pearson G, Robinson F, Beers Gibson T, Xu B E, Karandikar M, Berman K, Cobb MH, "Mitogen-activated protein (MAP) kinase pathways: regulation and physiological functions", *Endocrine Reviews*, 22(2), 153–83 (**2001**).
11. Manning G, Whyte D B, Martinez R, Hunter T, Sudarsanam S, "The protein kinase complement of the human genome", *Science*, 298(5600), 1912–34(**2002**).
12. Deleris P, Trost M, Topisirovic I, Tanguay P L, Borden K L, Thibault P, Meloche S, "Activation loop phosphorylation of ERK3/ERK4 by group I p21-activated kinases (PAKs) defines a novel PAK-ERK3/4-MAPK-activated protein kinase 5 signaling pathway", *The Journal of Biological Chemistry*, 286(8), 6470–8 (**2011**).
13. Theodosiou A, Ashworth A, *"MAP* kinase phosphatases*"*, *Genome Biology*, 3(7), reviews3009, 10 (**2002**).
14. Sehgal, V. K, Das, S, Vardhan, A, Computer aided drug designing, DOI:10.19056/ijmdsjssmes/**2017**/v6i1/125571.
15. Borman, S, New QSAR Techniques Eyed for Environmental Assessments, *Chem Eng News*, 68,20-3, (**1990**).
16. Anderson, A. C, The process of Structure Based Drug Design. *J. Chembiol.,* 10,7 87-97(**2003**).
17. Yun, T. Weiliang. Z, Kaixian, C. Hualiang J, New technologies in computer-aided drug design: Toward target identification and new chemical entity discovery; *Drug discovery today: technologies*, 3(3), 307-313 (**2006**).

18. Pranita, P. K, Madhavi, M. M, Rishikesh, V. A, Rajesh, J. O, Sandip, S. K, Computer-aided drug design: an innovative tool for modeling; *Journal of medicinal chemistry*, 2, 139-148 (**2012**).

19. Pugazhendhi, D, Umamaheswari, T. S, In silico methods in drug discovery–a review; International journal of advanced research in computer science and software engineering, 3(5), 680-683 (**2013**).

20. http://www.slideshare.net/adammbbs/structure-based-drug-design-48389641/12; Structure-based drug design (Docking and de novo drug design); Authored by Adam Shahul Hameed; Anna University Chennai; Available online from **2015**.

21. Satyajit, D, Sovan, S, Kapil, S, Computer-aided drug design - a new approach in drug design and discover, 4(3), 146-151(**2010**).

22. Choudhary, L. K, Shukla, A, Zade, S, Charde, R, (C.A.D.D)-a new modern software-based approach in drug design and discovery, *International journal of pharmaceutical chemistry*, 1(1) 10-20 (**2011**).

23. Nataraj, S. pagadala, Khajamohiddin, syed, Jack, Tuszynski, Software for molecular docking Biophysical reviews, 9 (2), 91-102(**2017**).

24. Garrett, M. Morris, Margueritta Lim wilby, Molecular docking, molecular modelling of proteins, 365-382(**2008**).

25. Xuan-Yu, meng, Hong-xing, zhang, Mihaly, mezei, Meing cui, Molecular docking: a powerful approach for structure based drug discovery, current computer aided drug design, 7(2), 146-157(**2011**).

26. Juan Pablo Arcon, Carlos P Modenutti, Demian Avendano, Elias, D. Lopez, Lucas A Defelipe, Francesca Alessandra Ambrosio, Adrian G Turjanski, Stefano Forli, Marcelo A Marti, Autodock Bias, improving binding mode prediction and virtual screening using known protein-ligand interactions, Bioinformatics, 35(19), 3836-3838(**2019**).

****