



# METAPHORICAL INVESTIGATION ON EMPAGLIFLOZINE IN STEIN LEVENTHAL SYNDROME BY MOLECULAR DOCKING FOR CDK2 AND SGLT

<sup>1</sup>M.C. Kavin Kumar, <sup>2</sup>B. Periyanyaki, <sup>3</sup>J. Mohamed Harsath, <sup>4</sup>K. Ayyappan,

<sup>1</sup> Assistant Professor, Department of Pharmaceutical Chemistry, Vellalar college of pharmacy, Erode, Tamil Nadu.

<sup>2,3,4</sup> Student, Department of pharmaceutical Chemistry, Vellalar College of Pharmacy, Erode, Tamil Nadu.

## CORRESPONDING AUTHOR:

KAVINKUMAR M.C

ASSISTANT PROFESSOR

DEPARTMENT OF PHARMACEUTICAL CHEMISTRY,

VELLALAR COLLEGE OF PHARMACY,

THINDAL, ERODE.

TAMIL NADU.

## ABSTRACT:

Polycystic ovarian syndrome (PCOS) is a common hyperandrogenic condition characterized by continuous ovulation, polycystic ovaries, and reproductive complications in women. Infertility is a prominent feature of this syndrome. This study aimed to compare the binding effects of empagliflozin with two protein molecules, CDK2 and SGLT2, implicated in PCOS. Molecular docking simulations were performed to assess the binding affinity of empagliflozin for CDK2 and SGLT2. The results demonstrated that empagliflozin exhibited a higher binding propensity towards CDK2 compared to SGLT2. However, it is noteworthy that SGLT2 still possessed a recognized binding site for empagliflozin. These findings shed light on the potential of empagliflozin as a therapeutic agent in PCOS, suggesting its ability to interact with key protein targets involved in the syndrome. Further research is needed to elucidate the precise mechanisms underlying empagliflozin's effects on CDK2 and SGLT2 in the context of PCOS

**KEYWORDS:** PCOS, Autodock, Empagliflozine, CDK2, SGLT2

## I. INTRODUCTION:

PCOS (polycystic ovarian syndrome) is a frustrating endocrine ailment that has a substantial impact on a woman's fertility, metabolism, democratic behaviours, and physical attributes during her gestation period, as well as during puberty and after menopause. PCOS hampers 6 to 18% of women, according to estimates <sup>[1]</sup>. Numerous variables, including chronic inflammation, endothelial dysfunction, endocrine, metabolic, and dyslipidemia disorders, are linked to PCOS. Due to its complex pathology and unknown aetiology, PCOS is a heterogeneous disease that poses difficult clinical problems in gynaecology and endocrinology. It is a complex ailment characterised by a high amount of circulating testosterone, a protracted menstrual cycle, oligo or anovulation, hirsutism, polycystic ovarian disease, and polycystic ovarian hyperandrogenism <sup>[2],[3]</sup>. It is usually linked to insulin resistance and obesity. Infertility has been linked to PCOS, which affects 5–15% of women during their reproductive years <sup>[4],[5]</sup>.

By concentrating primarily on the most promising compounds, computer-aided drug design or computer-assisted drug design reduces the need for synthetic and biological testing. It describes the molecular underpinnings of therapeutic efficacy and speculates on potential compounds that could enhance activity. The method of molecular docking is used to foretell how proteins (enzymes) will interact with drugs and ligands <sup>[6],[7],[8]</sup>.

ChemDraw 16.0 was used to study the 2D structures (.mol) of all the compounds <sup>[9]</sup>. Chem3D 16.0 was used to transform all of the chemicals (EMPAGLIFLOZIN) into a 3D structure (.pdb). Chem3D was used to load the 3D coordinates of each compound (.pdb) for energy reduction. The docking receptor (PDB codes CDK2, SGLT2) is a protein target from the RCSB Protein Data Bank that can be used to prepare macromolecules. The active site of the receptor was cleared of all bound ligands and water molecules <sup>[10]</sup>. Implementing the software for molecular docking appraisal (Auto-Dock Vina), the pharmaceutical agent empagliflozin was docked contrary to the active site of the protein. The Auto-Dock Tool Kits serve to perform post-docking scrutiny <sup>[11]</sup>.

CDK2 (cyclin dependent kinase 2) is additionally acknowledged as CDKN2, p33 (CDK2), and CDKN2. This gene encodes an integral component of the serine/threonine protein kinase family, which has an impact on the control of cell cycle progression. The protein that has been encoded constitutes the catalytic subunit of the cyclin-dependent protein kinase complex, primarily sets forth cell cycle progression. This protein's exertion is extremely crucial whilst on the G1-to-S phase transition. This protein interacts with and is manipulated by numerous additional complex members, which might include cyclin A or E, the CDK inhibitor p21Cip1 (CDKN1A), and p27Kip1 (CDKN1B) <sup>[12],[13]</sup>.

SGLT stands for sodium glucose cotransporter. Up-to-date knowledge on the therapeutic potential of SGLT-2 blockers in PCOS Selective SGLT-2 is a medicine that facilitates the regulation of the metabolism of glucose by drastically reducing glucose levels via an insulin-independent mechanism, curbing hyperinsulinism, and enhancing insulin susceptibility <sup>[14],[15]</sup>.

Empagliflozin serves as a sodium-glucose co-transporter-2 (SGLT2) antagonist. SGLT2 transporters are predominantly responsible for regulating glucose reabsorption in the urinary tract. It is utilised in medical practises as a supplement to a healthy diet and physical activity, more frequently in conjunction with complementary pharmacological protocols. By autodocking, empagliflozin had the capacity to bind with CDK2 protein molecules and compare its affinity with SGLT2 <sup>[16],[17]</sup>.

## II. MATERIALS AND METHOD:

The *IN-SILICO* studies necessitate the subsequent stages. They entail protein molecule preparation, ligand molecule preparation, binding site prediction, molecular docking studies, and docking study evaluation.

### 1. FABRICATION OF PROTEIN MOLECULES

The fabrication of protein molecules involves implementing many online resources, with a favouritism for the National Centre for Biotechnology Information Protein Information Bank.

#### 1.1 BY NCBI:

Protein molecules must meet the criteria required for docking investigations. The protein molecules have been extracted from the National Centre for Biotechnology Information's server. [<https://www.ncbi.nlm.nih.gov>] [The National Centre for Biotechnology Information (NCBI) is an organ of the National Institutes of Health's (NIH) National Library of Medicine. The United States government endorsed it. The NCBI has its administrative offices in Bethesda, Maryland, and was inaugurated in 1988 as an aftermath of legislation sponsored by US Congressman Claude Pepper. For the extraction of the protein molecules, consult an abundance of publications at the NCBI. Protein molecules must have distinctive features that enable PCOS to be manipulated. According to NCBI, the retrieved protein molecules are CDK2<sup>[18],[19]</sup>.

#### 1.2 BY PDB:

Then, for docking studies, protein molecules such as Cyclin-dependent kinase [CDK2] were targeted. The structure must be downloaded from the "Protein Data Bank" [PDB]. Browse for the protein data bank on Google, then proceed to the RCSB PDB homepage [<http://www.rcsb.org>]. Then, on Homo sapiens, select the protein molecules on CDK2. Download the file in PDB format as well. The PDB format of CDK2 protein molecules has been altered on NOTEPAD++. Protein molecules must have a titer, a header, an atome, a master, and a terminator. Other sequences should be removed and saved in a different folder <sup>[20],[21]</sup>.

### 2.FABRICATION OF THE LIGAND MOLECULE:

The ligand creation was also fundamental for the docking study. The ligand compounds have been gathered by researching multiple scientific publications. The ligand molecule was retrieved via the Pubchem server [<https://pubchem.ncbi.nlm.nih.gov>]. PubChem is a National Institutes of Health (NIH) open resource for the study of chemistry to obtain information. "Open" reflects that your research information is capable of being entered into PubChem for consumption by anyone. As a result of its inception in 2004, PubChem has expanded to establish itself as an indispensable resource of chemical knowledge for researchers, students, and the general public <sup>[22]</sup>.

Open the pubchem server's home page. Then enter the ligand molecule name "empagliflozin" and press the search button. Pubchem provided the best matching pair, selected the best matching compound, and downloaded it when searching. The ligand molecule empagliflozin was then downloaded in.spd format. However, the downloaded file format must be in. mol2 format. To alter the file format, use open babel. [Open Babel is a community-oriented scientific project that offers both users and developers a cross-platform software

and library constituted to help facilitate molecular modelling, chemistry, and numerous other applications, including the format of files and data interconversion.]<sup>[23]</sup>.

In the open Babel app, provide input as the downloaded.sdf format of the ligand empagliflozin. Then select the type of format required for the output. So select the.mol2 format. Further, click the convert button. Open Babel provided the appropriate format as.mol2 format. This converted.mol2 format ligand, empagliflozin, and already obtained protein file had been stored in the same folder but not saved on the desktop.

### 3.DETERMINATION OF BINDING SITE:

The therapeutic activity of a drug has depended upon the interaction of the protein and ligand molecule. Because of the existence of binding molecules in the protein, the interaction between the protein and ligand molecules. The proteins binding site had been used to provide activity for the ligand molecule. The high affinity for ligand and protein molecules was due to the huge number of binding sites. We must first discover the binding location.

#### 3.1 BY CASTp:

We used the CASTp server to find the binding location of the protein molecules. The Computed Atlas of Protein Surface Topography (CASTp) is a complimentary internet-based tool for identifying, defining, and quantifying concave surface portions on 3-D protein structures. <http://cast.engr.uic.edu>] . CASTp (Computer Atlas of Surface Topography of Proteins) seeks to furnish a thorough and detailed statistical description of protein topographic characteristics. CASTp has been demonstrated to be a reliable tool for a broad spectrum of studies, encompassing signalling receptor studies, drug mechanism knowledge, immune instability disease studies, protein-nanoparticle interactions, protein function inference, and the development of high-throughput computational tools. Jie Liang's lab at the University of Illinois at Chicago maintains this site <sup>[24]</sup>.

Open the homepage on the CASTp server and select the calculation. Then, in the pdb file, include the generated protein molecule. As an output, pocID, chain, sequenceID, amino acid, and atoms have been provided, as well as the area and volume of the binding site. In that case, we must record the chain and sequence ID of the amino acids topmost binding site.

### 4.PROTEIN LIGAND DOCKING:

We must advocate the auto-docking strategy for docking investigations. They needed protein preparation in a pdb file, ligand preparation in an a.mol2 format file, a configuration text document, and a vina application file for auto-docking. These four have been saved in the same folder. They needed the autodock tool application and the Cygwin application to dock.

#### 4.1 IN THE AUTO-DOCK TOOL APPLICATION:

Auto-docking is a technique that anticipates the ideal position of a single molecule to a subsequent position when the ligand and a target are attached to one another to create a complex that remains stable. [autodock.scripps.edu (AutoDock)] [autodock.vina.scripps.edu (Auto Dock Vina)].

Open the autodock tool application homepage, click the file, select the read molecules, and provide the protein pdb format file as an input. Then click "Edit" and add the hydrogen (polar) and the kollaman charges and add assign radii atom. Further, they were saved in protein.pdbqt format.

Then select the ligand and choose input as the already extracted ligand molecules in mol2 format. Then click the torsion tree and select both choose root and detect root one by one. Further, the output has been saved in ligand.pdbqt format.

Choose the select from string, then input the data for chain and residue obtained from the CASTp server. Then click add. Further select the grid box and insert the XYZ values obtained from the protein PDB file. Then adjust the XYZ values for the binding site. Then note down the output of the XYZ value. Then open the config document, change protein and ligand, and upload the XYZ value obtained from adjusting the grid box further they have saved <sup>[25],[26]</sup>.

#### 4.2 IN CYGWIN APPLICATION:

Open the Cygwin application and provide the appropriate site of the folders that contain ligand.pdbqt, protein.pdbqt, config, and vina. Then type  `['./vina -config config.txt & ]` code for running the programme. Finally, Cygwin provided the output as ligand output <sup>[27]</sup>.

#### 5. VISUALISATION OF PROTEIN LIGAND INTERACTION:

For the visualisation, the ligand output file obtained from the Cygwin application had been intermixed with the protein pdbqt file, and the intermixing of protein and ligand output had been provided as the input for the Discovery Studio application. Software designated Discovery Studio can be used for modelling both tiny molecule and large molecule systems. Dassault Systemes BIOVIA is the company that creates and sells it. Open the Biovia Discovery Studio and choose the file to open the intermixed protein ligand output as the input. Then check the interaction <sup>[28],[29]</sup>.

### III. RESULT:

Comparing the interactions between two protein molecules, such as CDK2 and SGLT2, and the ligand molecule empagliflozine enabled the comparison research. Here, the comparison was based on the degree of affinity between the CDK2 and empagliflozine-binding proteins and the SGLT2 and empagliflozine-binding proteins. With the aid of numerous servers, they have further determined The docking methodology was primarily involved in the comparison.

## 1.IN FABRICATION OF PROTEIN MOLECULE:

The results of the search for protein molecules in the protein data library were precisely the right protein molecules. Large, intricate molecules known as proteins are vital to maintaining every aspect of the human body's normal functioning. They are crucial for the regulation of tissues and organs throughout the body, in addition to their structure and function. The hundreds of smaller components that are referred to as amino acids that make up proteins are joined together by peptide bonds to form a lengthy chain. the results of a search for the protein molecules CDK2 and SGLT2 in the protein data bank. The protein data bank was searched for the right protein molecules in order to retrieve them.

Then CDK2 contains the amino acid sequence as

MENFQKVEKIGEGTYGVVYKARNKLTGEVVALKKIRXDTETEGVPSTAIRESLLKELNHPNIVKLLDV  
IHTENKLYLVFEFLHQDLKKFMDASALTGIPLPLIKSYLFQLLQGLAFCHSHRVLHRDLKPQNLLINTEG  
AIKLADFGGLARAFGVVPTYTHEVVTLWYRAPEILLGCKYYSTAVDIWSLGCIFAEMVTRRALFPGDS  
EIDQLFRIFRTLGTPEVVWPGVTSMPDYKPSFKWARQDFSKVVPPLDEDGRSLLSQMLHYDPNKRIS  
AKAALAHPPFQDVTKPVPHRLR

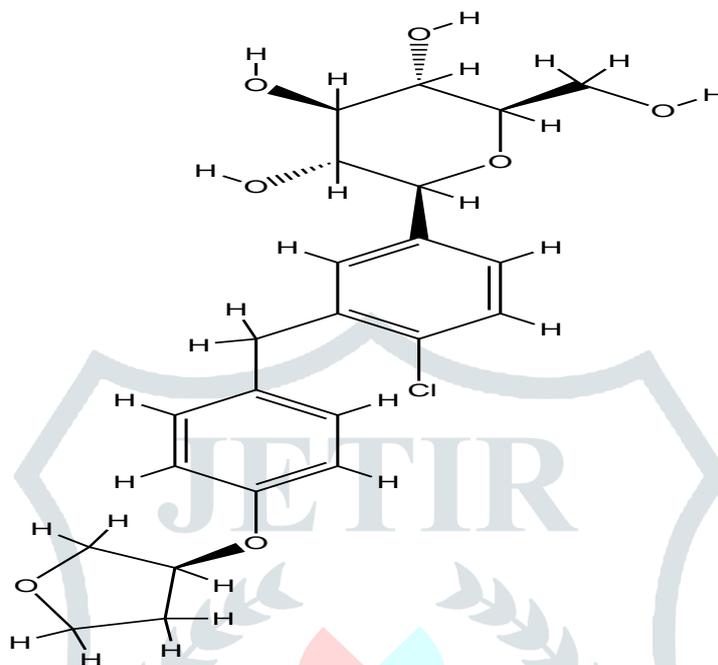
Then SGLT2 contain amino acid sequence as the

MEEHTEAGSAPEMGAQKALIDNPADILVIAAYFLLVIGVGLWSMCRNTRGTVGGYFLAGRSMVWWP  
VGASLFASNIGSGHFVGLAGTGAASGLAVAGFEWNALFVVLLLGWLFAPVYLTAGVITMPQYLRKRF  
GRRIRLYLSVLSFLYIFTKISVDMFSGAVFIQQALGWNIYASVIALLGITMIYTVTGGLAALMYTDTV  
QTFVILGGACILMGYAFHEVGGYSGLFDKYLGAATSLTVSEDPVGNISFCYRPRPDSYHLLRHPVTG  
DLPWPALLLGLTIVSGWYWCSDQVIVQRCLAGKSLTHIKAGCILCGYLKLTMPFLMVMMPGMISRILYP  
DEVACVVPEVCRRVCGTEVGCNSIAYPRLVVKLMPNGLRGLMLAVMLAALMSSLASIFNSSSTLFTM  
DIYTRLRPRAGDRELLLVGRLWVVFIVVSVAWLPVVQAAQGGQLFDYIQA VSSYLAPPVSAVFLAL  
FVPRVNEQGAFWGLIGLLMGLARLIPEFSFGSGSCVQPSACPAFLCGVHYLYFAIVLFFCSGLLTLTVS  
LCTAPIPRKHLHRLVFSLRHSKEEREDLDADEQQGSSLPVQNGCPESAMEMNEPQAPAPSLFRQCLLW  
FCGMSRGGVGSPPPLTQEEAAAAARRLEDISEDPWARVVNLNALLMMAVAVFLWGFY

## 2.IN FABRICATION OF LIGAND MOLECULE:

Further, both protein molecules had been downloaded in the PDB format. Then the pdb format file has been edited in Notepad++ by removing unnecessary sequences, and the resultant file contains TITLE, HEADER, ATOM, MASTER, and TERMINATOR The individual atoms of each protein are listed in these files, along with their three-dimensional positions in spacetime. The sequence and a lengthy list of the atoms and their coordinates are usually found after the huge "header" portion of text that summarises the protein, the citation information, and the specifics of the structural solution. The order of each chain of standard or modified linear, covalently linked amino acids or nucleotides is included in this listing. Other residues that are connected to the polymer's common backbone may also be present. Not included here are chemical elements or groups that are covalently bonded to side chains (in peptides) or sugars and/or bases (in nucleic acid polymers).

By searching for ligand molecules archived on the Pubchem portal, we were able to extract the protein molecules empagliflozin during the ligand synthesis. Each of the contributors to PubChem has contributed chemical information to the substance database, and the compound database has extracted specific chemical structures from the substance database. The structure retrieved from Pubchem is seen in **Figure 1**.



**Figure 1**

The extracted file is in SDF format. So this file had been converted in the open bebbel as the output of mol2 format.

### 3. DETERMINATION OF BNDING SITE:

#### 3.1 IN CASTp:

All surface pockets, interior cavities, and cross channels can be determined by CASTp, which also proposes the explained delineation of all atoms involved in their formation, as well as statistics of the area and volume of a pocket or void, as well as the quantity of mouth openings of a specific pocket ID, by solvent accessible surface models (Richards' surface) and by molecular surface models (Connolly surface). These statistics are all evaluated analytically. The basic technique aids in discovering a pocket or hollow large enough to hold a solvent with a 1.4 diameter.

A protein's active site is typically found at the protein's focal point and is essential to its performance. The identification of active sites on the protein surface and a precise description of their characteristics and limits constitute the initial step. These criteria serve as essential inputs for predicting or comparing a target's druggability in the future. The majority of active site detection algorithms are based on calculations of energetic characteristics or geometric modelling.

The chemical properties of compounds that bind to particular targets may be evaluated; therefore, the binding efficiency will depend on how effectively the compounds bind to the chemical space.

A protein's cross channels, inner cavities, and surface pockets are all identified by the CASTp server, which also offers a complete breakdown of all the atoms involved in their synthesis. It also determines their precise volumes and regions, as well as, if any mouth holes are present, their diameters. Both the molecular surface model (Connolly's surface) and the solvent-accessible surface model (Richards' surface) are used in the analytical calculation of these metrics. The CASTp server also offers topographic feature impressions.

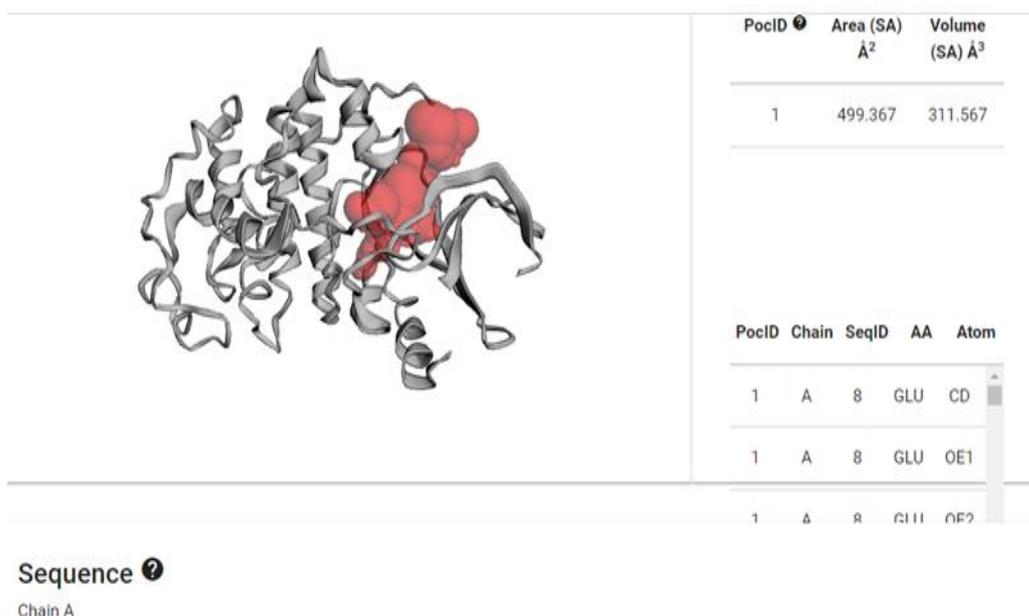


Figure 2

Figure 2 shows a picture of the CDK2 protein molecule's binding location. They also provide the area and volume that the proteins occupy in addition to the binding site. The area was determined to be 499.367 square amstrong, and the volume was determined to be 311.567 cubic amstrong. Additionally, they include the amino acid's sequencing ID and chain residue in a priority order.

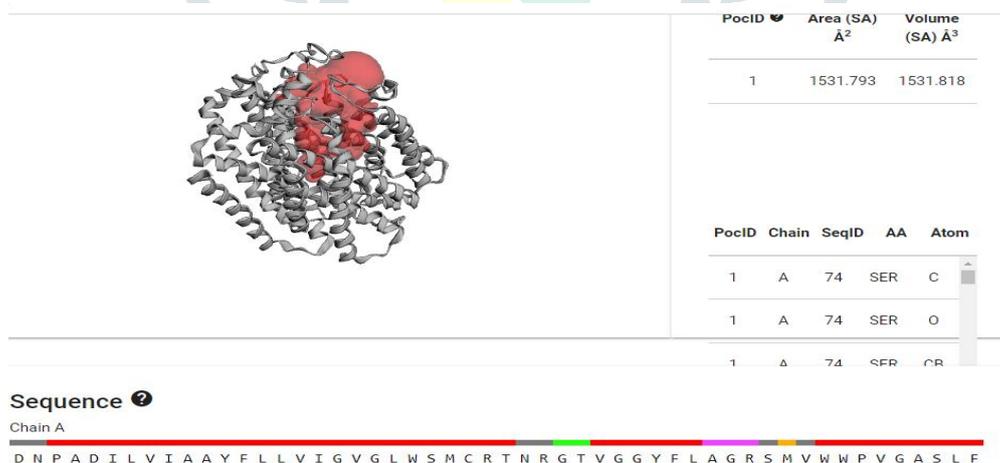


Figure 3

In figure 3 the binding site of the SGLT2 protein molecule had been viewed. Along with the binding site they contain the area and volume occupies by the proteins. The area had been calculated as 1531. 793sq.amstrong and volume had been calculated as 1531. 818cubic.amstrong. Further they contain the amino acid its sequenced id, chain residue on the priority manner.

#### 4. IN PROTEIN LIGAND DOCKING:

In the auto-docking, they required protein preparation in a pdb file and ligand preparation in a .mol2 format file, along with a config text document and a vina application file.

#### 4.1 IN THE AUTO-DOCK TOOL:

It soon became clear that in order to make Auto Dock a widely used tool, the activities of coordinate preparation, experiment design, and analysis required an effective graphical user interface. To meet this demand, Auto Dock Tools was developed. With a collection of functions that walk the user through protonation, computing charges, and identifying rotatable bonds in the ligand and the protein (explained below), Auto Dock Tools makes it easier to format input molecular files. It enables the user to identify the active site and estimate visually the volume of space searched in the docking simulation, simplifying the planning and preparation of docking experiments. Other techniques help the user enter search criteria and start docking computations. Finally, Auto Dock Tools offers a range of cutting-edge techniques for grouping, visualising, and evaluating the outcomes of docking trials.

Further the ligand molecules CDK2 as the input for the auto docking and follow that adding of hydrogen molecules, charges, and atoms to the ligand molecule CDK2, then they have been saved as the PDBQT format. Further the adding of ligand molecules along with the choosing torroison tree and also ligand molecules saved as the PDBQT format, and the output is shown in the figure 4.

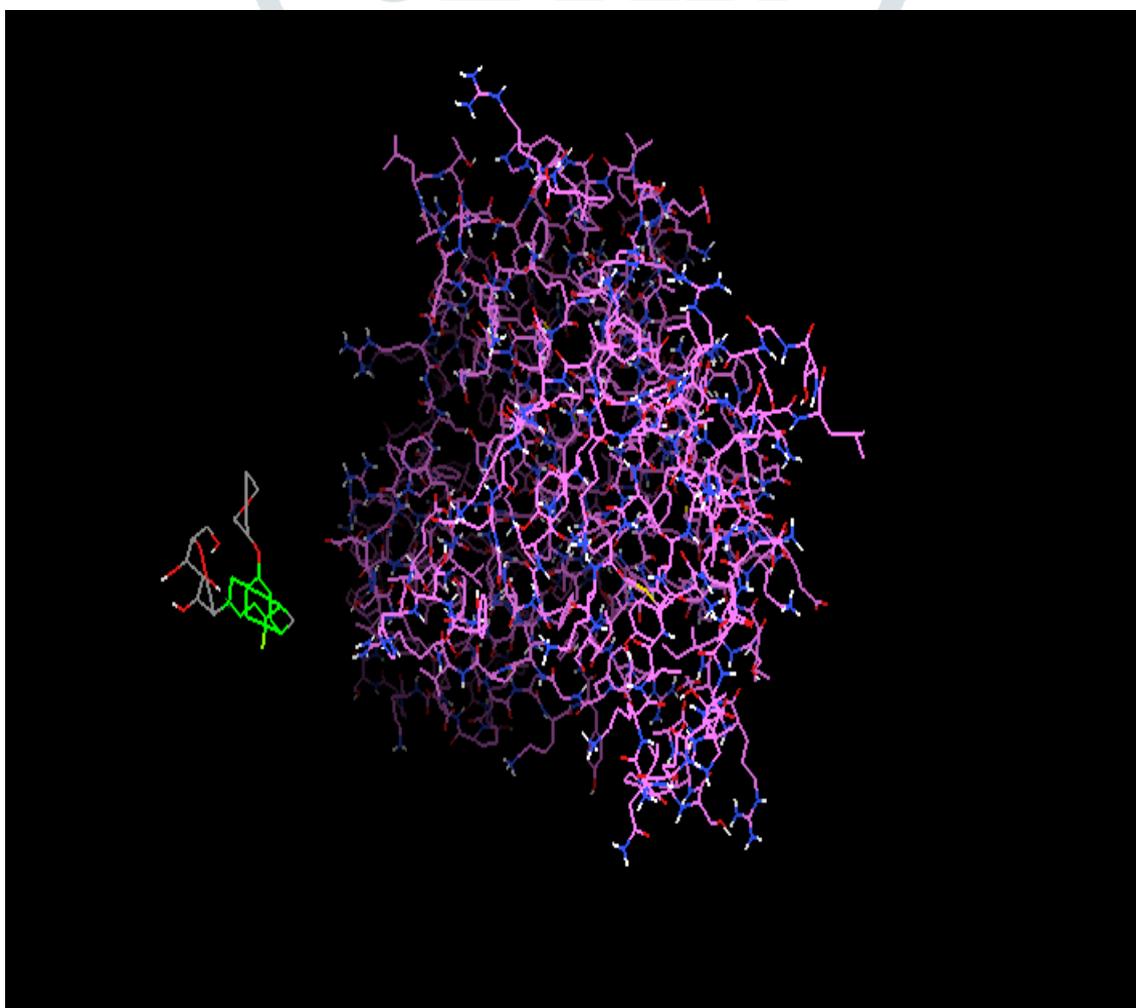
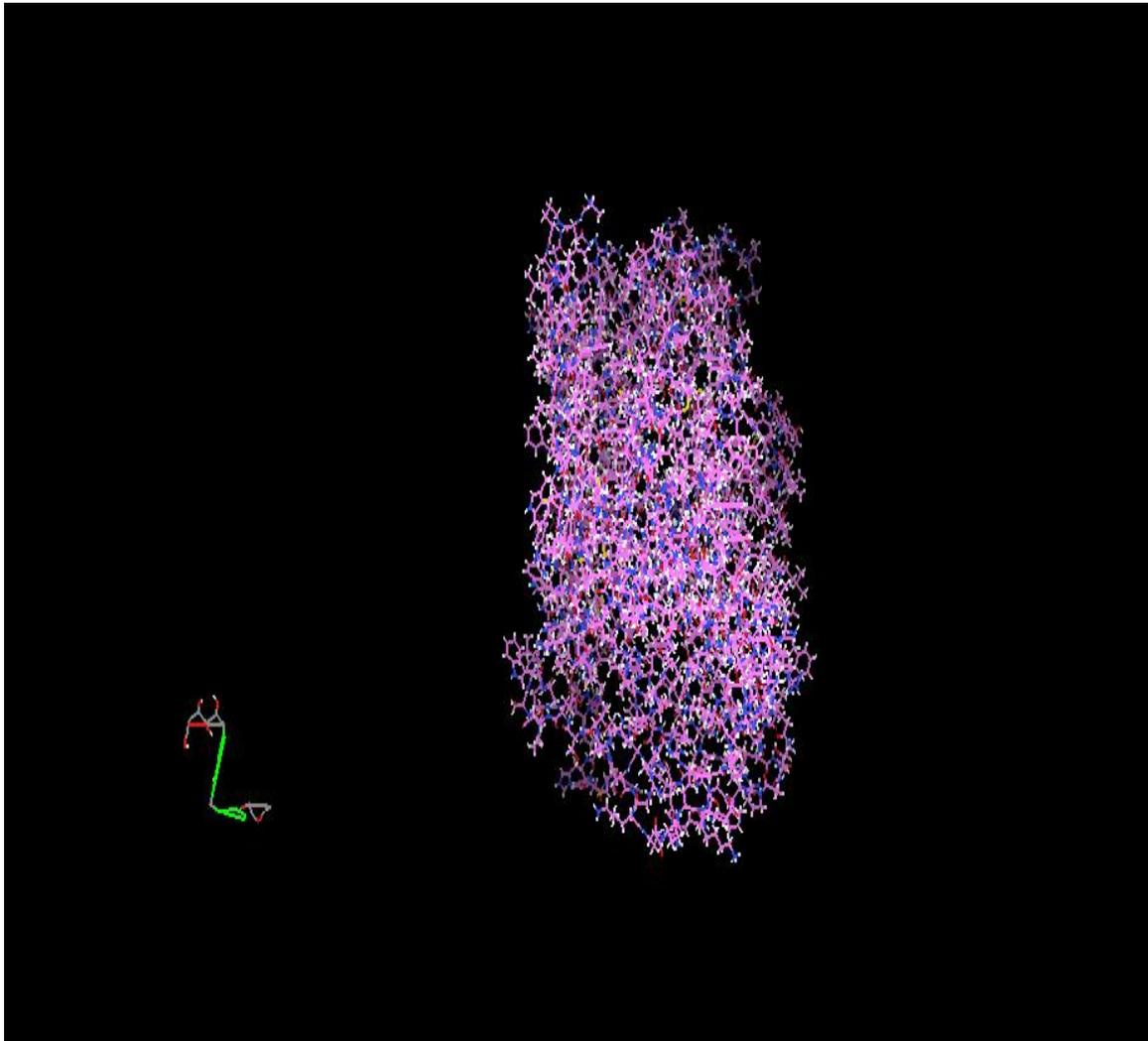


Figure 4

Further, the ligand molecules SGLT2 were used as the input for the auto docking, and after adding hydrogen molecules, charges, and atoms to the ligand molecules, they were saved in the PDBQT format additionally, additional ligand molecules were added along with the selected torroison tree, and additional ligand molecules were saved in the.pdbqt format. The results are depicted in the figure. 5



**Figure 5**

After getting these two structures, they have been saved as a pdbqt format file on the same file that contains the protein and ligand molecules, config, and Vina application. When you select the grid again, the binding site appears.

Then apply the grid box and input the pre-noted amino-acid in the castp server along with the amino acid's sequence ID and chain residue. output had for CDK2 and SGLT2. Further adjusting the grid box, they provided the values for XYZ coordinates. These coordinates had been uploaded to the config text document. Then they completed these processes in the auto-dock tool, and the processes continued in the Cygwin application. In Cygwin, the application uploaded the appropriate location of the file by using code and provided the [./vina -config config.txt &] and output shown as the binding affinity for protein molecules and also the ligand output like a file.

## **5.VISUALISATION OF THE PROTEIN-LIGAND INTERACTION:**

The output obtained from Cygwin had been edited with the protein file and allowed for interaction. The edited protein file had been input for the discovery studio and reported as figure 25 for the interaction of protein CDK2 with empagliflozin and figure 6 for the interaction of protein SGLT2 with empagliflozin.

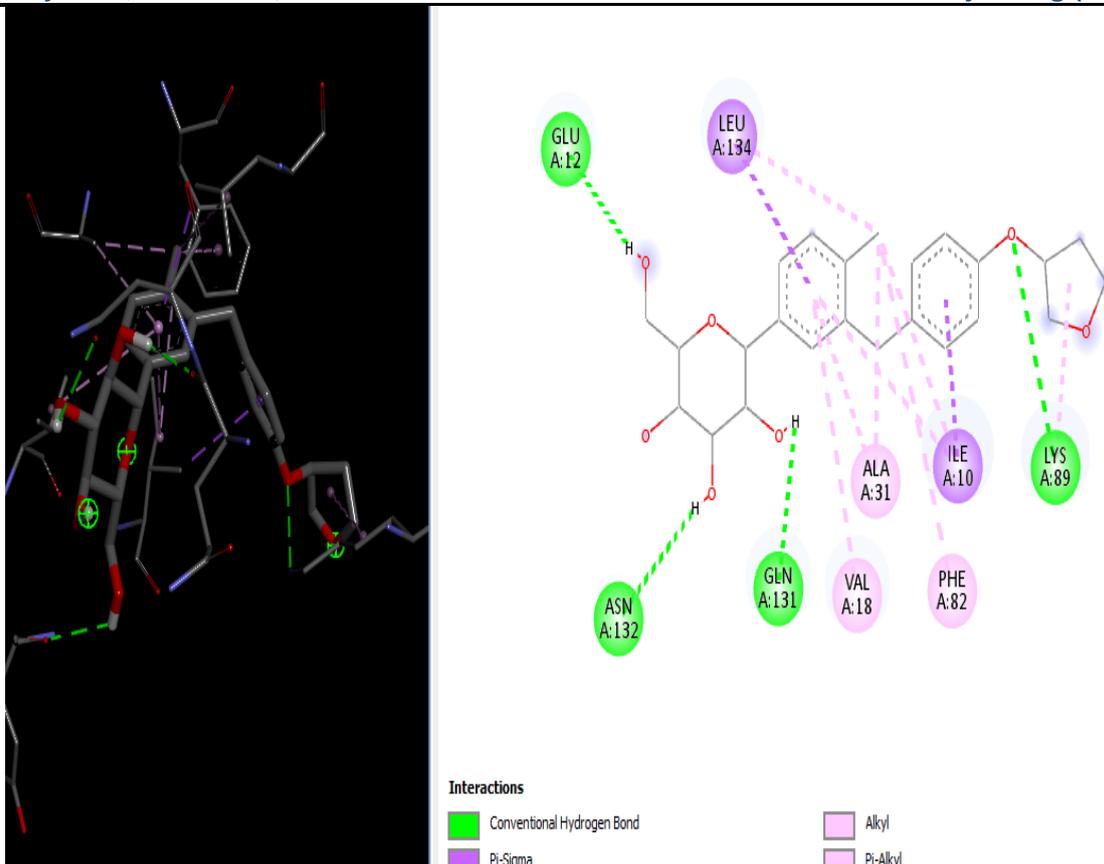


Figure 6

The interaction of the protein CDK2 and the drug empagliflozin is depicted in **Figure 6**. They were able to get both the two- and three-dimensional structures from that. Furthermore, the structure reveals 8 alkyl bonds, 2 pi-sigma bonds, and 4 hydrogen bonds.

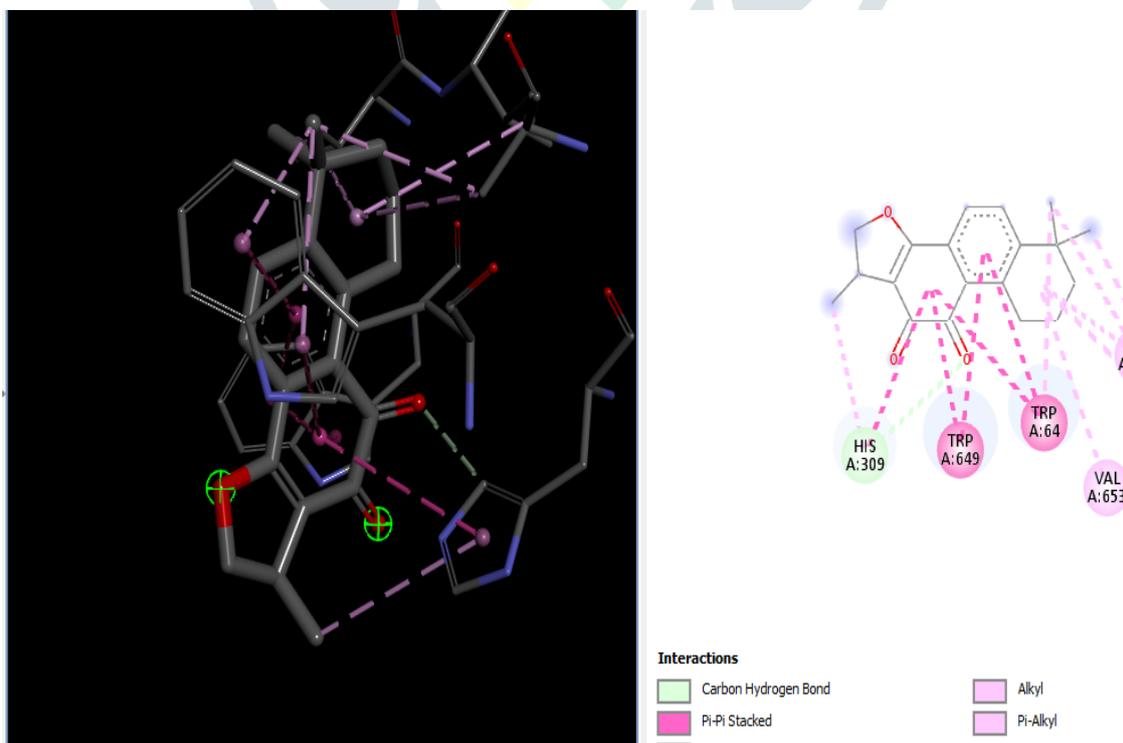


Figure 7

**Figure 7** depicts the relationship between the protein SGLT2 and the drug empagliflozin. They were able to get both the two- and three-dimensional structures from that. The structure also reveals that there is only one hydrogen bond.

#### IV. DISCUSSION:

In this study, we utilised bio-informatic techniques to assess the efficacy of empagliflozin against PCOS for drug interaction, drug similarity, quantum chemical properties, and binding energy calculation. One of these techniques was auto-dock Vina for molecular docking. There was a contrast between CDK2 and SGLT2 for empagliflozin. That investigation found that the CDK2 binding affinity of empagliflozin has a bigger binding site than the SGLT2 binding affinity. Due to the fact that SGLT2 only has one hydrogen bond as opposed to CDK2's four hydrogen bonds, it is less stable than CDK2. PCOS is a complex disorder that causes hyperandrogenism. The present in silico investigation found that CDK2 protein molecules interact with empagliflozin. Due to empagliflozine's strong affinity for CDK2, women with PCOS who have hyperandrogenemia may respond therapeutically to SGLT2 inhibitors differently. Therefore, only empagliflozine had been strongly chosen for treating PCOS to reverse insulin resistance and hyperandrogenism. In addition, it has been suggested that the stability of both proteins is important for pharmacokinetic and toxicological study. However, more in vitro and in vivo POS tests are recommended to corroborate the existing theories.

#### V. REFERENCE:

1. McShane A, Mole SE. Sex bias and omission exists in Batten disease research: Systematic review of the use of animal disease models. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*. 2022 Jul 14;166489.
2. Khan MS, Kim HS, Kim R, Yoon SH, Kim SG. Dysregulated Liver Metabolism and Polycystic Ovarian Syndrome. *International Journal of Molecular Sciences*. 2023 Apr 18;24(8):7454.
3. Mahalingaiah S, Diamanti-Kandarakis E. Targets to treat metabolic syndrome in polycystic ovary syndrome. *Expert opinion on therapeutic targets*. 2015 Nov 2;19(11):1561-74.
4. Gambineri A, Pelusi C, Vicennati V, Pagotto U, Pasquali R. Obesity and the polycystic ovary syndrome. *International journal of obesity*. 2002 Jul;26(7):883-96.
5. Diamanti-Kandarakis E, Dunaif A. Insulin resistance and the polycystic ovary syndrome revisited: an update on mechanisms and implications. *Endocrine reviews*. 2012 Dec 1;33(6):981-1030.
6. Kleinberg ML, Wanke LA. New approaches and technologies in drug design and discovery. *American Journal of Health System Pharmacy*. 1995 Jun 15;52(12):1323-36.
7. Basak SC, Vračko M, editors. *Big Data Analytics in Chemoinformatics and Bioinformatics: With Applications to Computer-Aided Drug Design, Cancer Biology, Emerging Pathogens and Computational Toxicology*. Elsevier; 2022 Dec 6.
8. Rodrigues JF, Florea L, de Oliveira MC, Diamond D, Oliveira ON. Big data and machine learning for materials science. *Discover Materials*. 2021 Dec;1:1-27.

9. Madlung A. Digital chemical intelligence. *Science*. 1999 Sep 17;285(5435):1866-7.
10. Kellenberger E, Muller P, Schalon C, Bret G, Foata N, Rognan D. sc-PDB: an annotated database of druggable binding sites from the Protein Data Bank. *Journal of chemical information and modeling*. 2006 Mar 27;46(2):717-27.
11. Morris GM, Huey R, Olson AJ. Using autodock for ligand-receptor docking. *Current protocols in bioinformatics*. 2008 Dec;24(1):8-14.
12. Lu J, Xuan F, Chen A, Jin R, Zhou W, Ye Y, Ren Y. MicroRNA-646 inhibits the proliferation of ovarian granulosa cells via insulin-like growth factor 1 (IGF-1) in polycystic ovarian syndrome (PCOS). *Endokrynologia Polska*. 2023 May 8.
13. Chen AX, Jin RY, Zhou WM, Ye YJ, Lu JL, Ren YF, Xuan FL. CircRNA circ\_0043533 facilitates cell growth in polycystic ovary syndrome by targeting miR-1179. *Reproductive Biology*. 2022 Jun 1;22(2):100637.
14. Glendining KA, Campbell RE. Recent advances in emerging PCOS therapies. *Current Opinion in Pharmacology*. 2023 Feb 1;68:102345.
15. Cerón HI. Relationship between Polycystic Ovary Syndrome and Metabolic Syndrome. *The Journal of Reproduction*. 2023 Apr 26;2(1).
16. Rakic D, Jakovljevic V, Jovic N, Bicanin Ilic M, Dimitrijevic A, Vulovic T, Arsenijevic P, Sretenovic J, Nikolic M, Petrovich Fisenko V, Bolevich S. The Potential of SGLT-2 Inhibitors in the Treatment of Polycystic Ovary Syndrome: The Current Status and Future Perspectives. *Biomedicines*. 2023 Mar 23;11(4):998.
17. Kourtidou C, Tziomalos K. Pharmacological Management of Obesity in Patients with Polycystic Ovary Syndrome. *Biomedicines*. 2023 Feb 8;11(2):496.
18. Masys DR, Benson DA. Don Lindberg and the creation of the National Center for Biotechnology Information. *Information Services & Use*. 2022 May 10(Preprint):1-9.
19. <https://www.ncbi.nlm.nih.gov>
20. Burley SK, Bhikadiya C, Bi C, Bittrich S, Chao H, Chen L, Craig PA, Crichlow GV, Dalenberg K, Duarte JM, Dutta S. RCSB Protein Data Bank (RCSB. org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. *Nucleic Acids Research*. 2023 Jan 6;51(D1):D488-508.
21. <http://www.rcsb.org/pdb/>
22. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, Zaslavsky L. PubChem 2023 update. *Nucleic Acids Research*. 2023 Jan 6;51(D1):D1373-80.
23. Hutchison GR, Morley C, James C, Swain C, De Winter H, Vandermeersch T, O'Boyle NM. Open Babel Documentation.
24. Tian W, Chen C, Liang J. CASTp 3.0: computed atlas of surface topography of proteins and beyond. *Biophysical Journal*. 2018 Feb 2;114(3):50a.
25. Huey R, Morris GM, Forli S. Using AutoDock 4 and AutoDock vina with AutoDockTools: a tutorial. *The Scripps Research Institute Molecular Graphics Laboratory*. 2012 Oct 26;10550(92037):1000.
26. Seeliger D, de Groot BL. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *Journal of computer-aided molecular design*. 2010 May;24(5):417-22.
27. Jarillo JS. Cygwin. *Mundo Linux: Sólo programadores Linux*. 2007(97):62-5.

28. Jejurikar BL, Rohane SH. Drug designing in discovery studio.

29. Pawar SS, Rohane SH. Review on discovery

