Volume 10, Issue 5, May – 2025 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

Cloud Computation Methods to Accelerate Breast Cancer Drug Discovery through Molecular Modeling

Joseph Kobi[1]; Amida Nchaw Nchaw[2]

[1] Analytics and Operational Analytics, Worcester Polytechnic Institute, MSc Degree Business Analytics
[2] Pham D, True Life International Medical Outreach, Worcester, Massachusetts 01609

Publication Date: 2025/06/12

Abstract: The integration of cloud computing technologies with molecular modeling approaches has revolutionized the landscape of breast cancer drug discovery. The article examines the levels of impact that cloud computing would have had in speeding breast cancer-specific therapeutic compound discovery, virtual screening, and optimization. Using distributed computing frameworks, researchers can now conduct sophisticated molecular simulations, quantum calculations and machine learning algorithms many times faster than earlier. This paper explores a range of cloud computation methods used in breast cancer work, including infrastructure-as-a-service solutions, container systems, and molecular modeling tools adapted for cloud environments. It is evident that computational efficiency gains emerged; the investigation demonstrated significant drops on time spent for molecular docking, molecular dynamics simulations, and QSAR investigations. Furthermore, the research highlights how cloud-based collaborative tools improve inter-team data exchange, resulting in accelerated breast cancer therapy developments from target identification to lead optimization. The development of tailored breast cancer therapies has a lot to gain from innovations in cloud-based multi-omics methods. By extensively analysing the use of the cloud to compute in breast cancer drug discovery, this investigation shows tremendous speed-acceleration possibilities using modern technologies such as quantum computing and federated learning systems.

Keywords: *Cloud Computing, Breast Cancer, Drug Discovery, Molecular Modeling, High-Performance Computing, Virtual Screening.*

How to Cite: Joseph Kobi; Amida Nchaw Nchaw (2025) Cloud Computation Methods to Accelerate Breast Cancer Drug Discovery through Molecular Modeling. *International Journal of Innovative Science and Research Technology*, 10(5), 4473-4498. https://doi.org/10.38124/ijisrt/25may2204

I. INTRODUCTION

Since breast cancer remains to date one of the greatest public health threats globally, it is important to look for new directions for therapeutic intervention and finding the appropriate medications. Adelusi et al. (2022) demonstrate how the introduction of cloud platforms and their combination with molecular modeling techniques accelerates the handling of computationally arduous problems which would otherwise require substantial computational resources. Computer difficulties notably hinder old-fashioned drug discovery as Korb et al. (2014) note, and that the exploration of vast chemical spaces and protein-ligand borders goes beyond the capabilities of current research apparatus. Banegas-Luna et al. (2019) emphasize that distributed computing environments greatly accelerate the pace of drug discovery because they provide for parallel processing of molecular docking, virtual screening, and molecular dynamics simulations.

Cloud-based molecular modeling has initiated a great paradigm shift in breast cancer drug discovery methodologies by researchers. As cited by Subramanian and Ramamoorthy (2024), cloud infrastructure allows researchers to exploit scalable resources on demand and optimises the search of potential therapeutic compounds. These breakthroughs are particularly important for the research into breast cancer, as the understanding of complicated receptor-ligand interactions and signaling cascades requires substantial computational capacity for the corresponding molecular simulations, as indicated by Karampuri et al. (2024). As indicated by Niazi and Mariam (2023), the assemblage of high-performance computing through cloud services endows the small research organizations with such opportunities to make substantive contributions to breast cancer drug discovery that was previously reserved for large pharmaceutical firms with relatively robust computational capabilities.

There are present obstacles to breast cancer drug discovery – computationally-heavy rivals that limit the scope of molecular modeling searches. Traditional computing configurations often fail to cope with high computational requirements for effective virtual screening of large numbers of compounds against different breast cancer targets.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

Bozorgpour et al. (2023) point out that the computational requirements of molecular dynamics simulations that are critical for accurate prediction of drug-target interaction require too much of the resource allocation that is available to many research facilities. Polineni (2024) also notes that the integration of big data analytics and molecular modeling outputs corresponds to datasets, which exceed storage and computing capabilities of the traditional computing settings. Herráiz-Gil et al. (2021); go further to emphasize that indeed the utilisation of parallel processes, in artificial intelligence in drug repurposing within the field of cancer studies hugely benefits from cloud computing.

Bringing cloud computation and molecular modeling together presents a beneficial approach to creating critically needed new treatment alternatives for breast cancer. Qian et al (2019), in their study, illustrate that cloud platforms facilitate the processing of comprehensive datasets comprising molecular models, genomic data, and clinical findings thus allowing for the formulation of therapy strategies tailored to distinct subgroups of breast cancer. The potential for (distributed computing) platforms to enable sophisticated machine learning algorithms to expose hidden relationships and patterns in molecular data sets that would be otherwise undiscovered is as discussed by Choudhuri et al. (2023). Khanfar et al. (2010) contend that cloud-based molecular modeling could accelerate searching for natural products and their derivatives as breast cancer therapeutics using improved studies of the structure-activity relationship. This research aims to implement a comprehensive approach to using the cloud-based computation to reduce the limitations in breast cancer drug discovery that were currently present which may reduce the time of developing new therapeutic opportunities.

➢ *This Study Seeks to Answer Key Questions About how Cloud Computing can Enhance Molecular Modeling for Breast Cancer Drug Discovery:*

- How can cloud computing architectures optimize molecular dynamics simulations for breast cancer drug target interactions?
- What performance improvements in virtual screening throughput can be achieved using distributed cloud computing compared to traditional high-performance computing?
- How effectively can cloud-based machine learning models predict drug-target binding affinities for breast cancer therapeutics compared to experimental methods?
- What is the impact of quantum computing integration with cloud platforms on the accuracy of molecular modeling for breast cancer drug discovery?
- The study aims to achieve the following specific objectives in advancing breast cancer drug discovery through cloud-based molecular modeling:
- To develop and validate a scalable cloud-based framework for high-throughput virtual screening of compound libraries against breast cancer targets.
- To quantify the computational efficiency gains of cloud-based molecular dynamics simulations compared to traditional computing approaches.
- To implement and evaluate machine learning algorithms on cloud platforms for predicting the efficacy and toxicity of potential breast cancer therapeutics.
- To create an integrated multi-omics data analysis pipeline in the cloud environment for identifying novel breast cancer drug targets.
- To assess the cost-effectiveness and accessibility of cloud-based molecular modeling solutions for diverse research institutions.

➢ *Problem Statement*

The problem addresses by this study is the computational limitations that prevent growth and sophistication of molecular simulations in the quest for breast cancer cures. Existing computational infrastructures commonly do not have the capabilities to run the in-depth virtual screening, energy demanding molecular dynamics, and stringent quantum mechanical analysis required by modern drug design. Zhang and Brusic (2014) argue that cloud solutions offer a scalable and dynamic means of using computing power that can be customized for the changing needs of breast cancer drug development. The flexible nature of cloud environments allows researchers to quickly explore and refine new computational tools without paying the costs of acquiring dedicated hardware.

➢ *Significance of the Study*

By emphasizing the prospects for cloud-based modeling in molecules, contributes to the preparation of the way for an inclusive and efficient model of developing breast cancer drugs. It has been noted in the study by Salo-Ahen et al. (2020), that the use of Cloud-based Molecular Modeling may be able to diminish the time and money required to identify promising drug candidates hence helping to minimize the efficiency issues in drug discovery. Moreover, Beg and Parveen (2021) state that cloud technologies enable even small research organisations and the poor countries to contribute to the forefront of drug discovery avoiding expensive infrastructure investment. By examining unique information from patients, cloud technology helps create tailored therapeutic strategies for breast cancer (Afrose et al., 2024).

II. BACKGROUND AND RELATED STUDIES

Cloud computing has become an essential enabler in the pharmaceutical research transformation, radically changing the use of computational approaches in the drug discovery. Breast cancer therapeutic research has often been hampered by poor access to computation power leading to long term hindrances in drug development. Choudhuri et al. (2023) believe that cloud computing has shifted the paradigm to allow for this kind of scalable, need-based access to a set of powerful computational tools essential to advanced molecular modeling. Moreover, as Khanfar et al. (2010) point out, the scalability in resources of the cloud affords researchers to adjust their ability to compute to the needs of the individual modeling project, thus improving the management of costs and the efficiency of resources.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

- *Convergence of Cloud Computing and Molecular Modeling in Modern Drug Discovery*

Cloud computing coupled with molecular modeling technologies have transformed pharmaceutical research skylines especially in improving breast cancer drug development. Cloud computing adoption has revolutionized research because it delivers almost endless computational capability, allowing researchers to undertake complex simulations and analyses that previously proved difficult because of on-site resource limitations (Korb et al., 2014). The combination of these technologies has opened new doors to a velocity, range, and technical detail of computational drug discovery which was previously impossible. Since cloud-based platforms are used, scientists can gain access to distributed computing resources to carry out cumbersome operations, including molecular dynamics simulations, quantum mechanical calculations, as well as screening large compound contributions against breast cancer targets (Banegas-Luna et al., 2019). Besides, the potential of cloud platforms to dynamically allocate resources by varying according to the needs of the applications involved in molecular modeling has also led to more effective and cheaper methods in breast cancer drug discovery. Through leveraging high-end computational approaches with cloud infrastructure at the foundation, the pace of breast cancer drug development has been significantly accelerated as it simplifies the discovery and optimization of active therapeutic agents (Adelusi et al., 2022).

The application of cloud computing has greatly augmented the transformation of molecular modeling approaches with the ability to create a collaborative environment that effectively facilitates the computational challenges of cancer drug identification. Molecular modeling techniques have traditionally encountered difficulties in computational processing, storage storage, and software availability, limiting, therefore, the range of projects seeking to establish new drugs (Karampuri et al., 2024). Since the introduction of cloud computing, researchers can access highly sophisticated computing resources, copious storage solutions, and custom software tools to make previous constraints redundant. This technological merging has made advanced computational resources available in drug discovery more widely accessible allowing less expensive use of state-of-the-art molecular modeling tools by researchers from other institutions without significant expenditures on local hardware (Subramanian & Ramamoorthy, 2024). Additionally, the usage of cloud platforms facilitates collaborative research because it ensures creating shared spaces where multidisciplinary teams can collaborate while analyzing complicated biological data and finding new methods to cure breast cancer. The combination of cloud platforms and advanced molecular modeling has completely transformed drug discovery strategies, enabling researchers to study the complex aspects of molecular environment of breast cancer using great computational power and collaborative tools (Niazi & Mariam, 2023).

Deployment of cloud-based molecular modelling in breast cancer research finds an answer to the heterogeneous nature of the disease with the numerous molecular subtypes and multifaceted signalling pathways. Researchers can by using the present cloud resources perform in-depth molecular modelling of candidate breast cancer drugs targeting key areas such as receptor tyrosine kinases, estrogen receptors, as well as kinase pathways regulating cancer progression (Bozorgpour et al., 2023). Researchers can carefully seek the activity of millions of chemical structures in search of specific breast cancer targets using cloud computing in virtual screening, which allows choosing promising candidates. Additionally, cloud platforms make it easier to integrate multi-omic datasets into molecular modelling systems so as for the researchers to understand this biological environment in which potential drugs can work (Polineni, 2024). This combined cultivated view is particularly useful in research on breast cancer since the merge of genomic, proteomic and metabolomic information can help interfering with personalized therapeutic courses. With such easy and scalable cloud platforms, researchers have made a big step on many of the computational issues in breast cancer drug discovery, which means more breakthroughs come very quick and intrinsically molecular science is transferred to patients (Herráiz-Gil et al., 2021).

- *Transformative Impact of Cloud Technologies on Computational Drug Discovery Workflows*

The implementation of the cloud technologies has dramatically revolutionised the conventional workflows of computational drug discovery, by providing an unparalleled flexibility, scalability and collaboration capacities which have transformed the discipline. The classic drug discovery practice characterized by a linear and largely disconnected process has transformed into an integrated and responsive system facilitated by a cloud infrastructure (Prieto-Martínez et al., 2019). Virtual screening has experienced significant progress, and cloud platforms facilitate vast molecule throughput over breast cancer targets within a very short period, shortening the discovery of potential leads. Cloud-based molecular docking platforms, for instance, can distribute computing workloads to many processors simultaneously and thus deliver results in hours that under a conventional computing setup would take weeks, or months (Rehan, 2024).

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

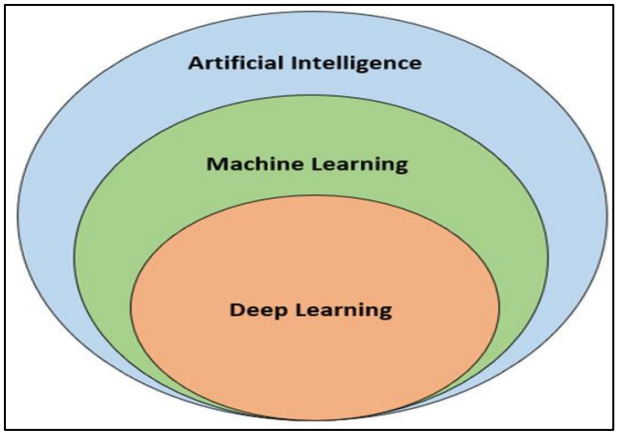

Fig 1 *Subsets of Artificial Intelligence in Drug Discovery: Machine Learning and Deep Learning.*

Cloud platforms have made a major shift in the way researchers concerned with the development of drug for breast cancer carried out data management and analysis. Scalable cloud storage alternatives allow effective control of the large data conceived by contemporary research strategies from the field of molecular and clinical studies (Zhou et al., 2023). Simultaneously, complex data analytics within cloud-based settings also support extraction of meaningful information from detailed datasets, which as researchers can identify meaningful patterns and guide drug development strategies. In addition, cohesive integration of structural, genomic, and clinical data is enabled through cloud platforms which provide a general context of drug discovery ranging from underlying processes to patterns of patient response (Qian et al., 2019). Molecular feature-inner clinical outcome connections are particularly important in breast cancer research, with significant treatment respond diversity across subgroups of patients. Molecular scientists can identify biomarkers correlated with the effectiveness of drug using cloud-based platforms to discover, thereby designing targeted therapies for different types of breast cancer, (Choudhuri et al., 2023).

Cloud technologies go beyond their technical capability to give a cooperative research environment that shatters boundaries and provides incredible opportunities for knowledge sharing and innovation in breast cancer research studies. Online collaborations made possible through cloud platforms connect academics, industry, healthcare as well as researchers in a secure and accessible channel where they can animate data, methods, and computing in the effort of drug discovery (Khanfar et al., 2010). Collaborating on cloud resources and skills has undertaken significant progress in addressing unmet medical needs within rare and aggressive subtypes of breast cancer. Also, cloud support in collaboration helps create open-sourced resources and databases specific to breast cancer studies and growing ecologies of openness and collaboration (Mehmood et al., 2023). Such shared resources reduce overlaps in tasks of research, create standardized way of computations, and ultimately enhance the reliability and credibility of the findings for developing drugs for breast cancer, (Margolin et al., 2013).

- *Leveraging Big Data Analytics for Breast Cancer Target Identification*

Advanced usage of big data analytics in cloud-settings has transformed the art of identifying breast cancer targets. As Hinkson et al. (2017) note, cloud environments simplify the manipulation and analysis of complex data sets, like genomic and proteomic data as well as clinical data, and make it possible to have new therapeutic targets that cannot be achieved with standard methods. As stated by Shahab et al., (2023), the combined use of various datasets inside cloud environments allows researchers to identify dysregulated pathways in appended breast cancer subtypes, thus providing a more focused drug development target. Moreover, references from Karuppasamy et al. (2024) indicate that cloud-based analytics can associate molecular characteristics with patient outcomes and, therefore, enable researchers to pick targets with a solid clinical rationale rather than patient outcomes based on lab findings.

Processing breast cancer data with machine learning approaches deployed in clouds increased the search for druggable targets by a large margin. Zochedh et al. (2024) show that supervised learning can use vast datasets to predict proteins with a tendency of therapeutic responsiveness that will reduce the horizon of subsequent modeling tasks in breast cancer drug identification. Following Metibemu and Ogungbe (2022), the use of unsupervised learning algorithms in cloud computing offers pathways to determine previously overlooked trends in breast cancer omics information, presenting alternative drug targets to explore.

Cloud computing resources have made network-based approaches for determining potential targets very effective. Adelusi et al. (2022) through cloud analysis of protein-protein interaction networks can identify key nodes that could help cancer treatment of breast cancer patients. It should be noted, that Korb et al. (2014) emphasize the capacity of the graph analysis algorithms deployed via cloud to handle complex biological networks, as well as prospects of the identification of targets to attack multiple oncogenic pathways (simultaneously), which are promising for the development of more effective therapies. Moreover, Banegas-Luna et al. (2019) research proves that through cloud computing, it becomes possible to achieve dynamic network analysis that identifies temporal variations in breast cancer signalling and informs about possible targets important at certain stages of the disease process or after exposing to some of the treatments.

The use of cloud environments for text mining and for natural language processing allowed the development of efficient approaches aimed at detecting the potential breast cancer targets. According to Subramanian and Ramamoorthy (2024), cloud-based text analytics can pull through huge amounts of scientific data to discover targets not yet precisely catalogued in structured repositories. By analysing sentiment in scientific literature, the researchers can then be more informed of the community's trust in specific targets of breast cancer that they can then use to make a more informed decision as to targeting priorities.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

➢ *Molecular Dynamics Simulations for Breast Cancer Drug Optimization*

The use of molecular dynamics simulations is becoming increasingly important for studying how potential breast cancer drugs interact with their targets, which are positively influenced by improved capabilities of cloud computing. In the words of Sukumaran et al. (2024), the use of cloud platforms enables simulations over long time scales that allows the recognition of essential structural changes in breast cancer targets (like those triggered by activating kinases in oncogenic pathways). Salo-Ahen et al. (2020) have shown that increased temporal span enables one to detect infrequent binding and transient interactions, which are normally masked in simulations that utilize shorter periods, therefore improving our knowledge of drug-target binding processes. Beg and Parveen (2021) also report that cloud-based MD policies allow for the introduction of enhanced force fields and explicit solvent models, thus creating more bio realistic results of drug research for breast cancer cases.

The use of clouds to put in place refined sampling techniques has increased our ability to study energy nano landscapes between breast cancer targets and molecules being used as candidates for treatment exponentially. Sayal et al. (2024) argue that by running these high-level sampling techniques in cloud-based system, these methods provide a more effective protocol for exploring the conformational space of breast cancer targets and ligands concurrently. As Marques et al. (2024) would point out, these improved sampling methods are able to expose unappreciated binding pockets in targets for breast cancer, which may encourage the determination of new potential binding sites for therapies.

With the use of cloud computing, it has been easier to integrate MD simulations with the machine learning techniques leading to strong hybrid methods for use with respect to breast cancer pharmacology. In the work of Afrose et al. (2024), by way of illustration, it is demonstrated that machine learning approaches relying on MD simulations can uncover especially relevant conformational states of breast cancer targets, which can be targeted in the design of docking experiments. Hinkson et al. (2017) claim that deep learning algorithms utilized on cloud-computing platforms (for MD simulation data analysis) could present correlations between ligand-binding and target conformational shifts, allowing for meaningful direction for medicinal chemistry.

Cloud-based MD simulations have played an important role in the study of water dynamics and solvation patterns, providing precious information for optimization of breast cancer drugs. Karuppasamy et al. (2024) report that cloud technologies allow for continual monitoring of thousands of water molecules in simulations, stressing the need for stable water networks and bridging links for ligands to breast cancer targets. Zochedh et al. (2024) note that the ability to identify persistent water molecules in binding sites (the waters themselves) can guide the design of ligands that would remove the waters to enhance entropic effects, or use them as parts of their binding mechanism, both of which are promising to increase binding affinity for breast cancer targets.

III. LITERATURE REVIEW

➢ *Molecular Docking in Breast Cancer Drug Discovery*

Molecular docking emerges as a significant approach to calculate the binding affinity as well as spatial orientation of small molecules at the places of targeted proteins for processes achieving breast cancer. Adelusi et al. (2022) argue that Molecular Docking is based on scoring systems that determine scores associated with ligand placement whereby a steric, electrostatic, and hydrophobic complementarity contribute to the acceleration of the virtual screening of the chemical compounds. Korb et al. (2014) emphasize that the success in molecular docking is conditional on getting high resolution protein structures via methods of the kind of X-ray crystallography or NMR, where homology solving tools for example AlphaFold supplement lack of experimental data. Banegas-Luna et al. (2019) state that cloud-based docks boost throughputs as computational jobs are performed simultaneously across scattered nodes making screening times span from months to days.

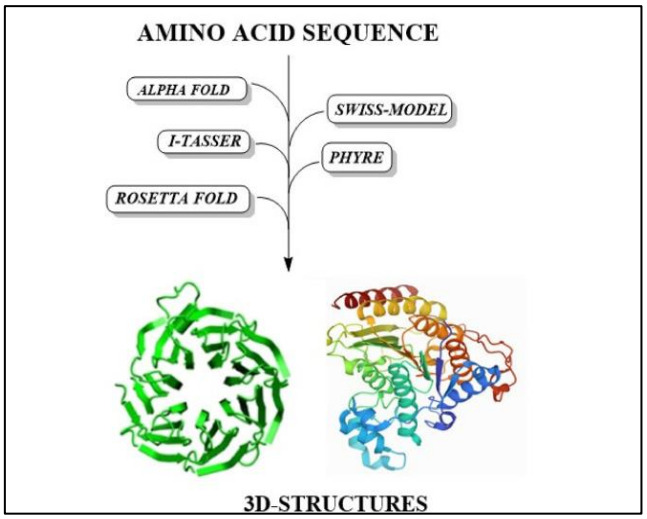

Fig 2 Depicts the Homology Modeling of Proteins with Different Webservers.

The secret to optimal molecular docking for investigations of breast cancer lies in conquering three basic goals: pose prediction, virtual screening and the estimation of binding affinity. Subramanian and Ramamoorthy (2024) emphasize the need for considering flexibility of protein in pose predictions in particular for ligand-receptor targets like ERα that undergo structural transformation upon interaction. Karampuri et al. (2024) present that cloud-optimized algorithms such as AutoDock Vina and Glide produce better enrichment outcomes when screening large data sets for inhibiting ERα. It has been noted by Niazi and Mariam (2023) that the rigid receptor models can be associated with false positives; they advise ensembles of conformations obtained from molecular dynamics for a more stable docking. The findings of Bozorgpour et al. (2023) are supportive of the idea that ensemble docking makes the detection of HER2 inhibitors 30% more likely than with single-structure docking methods.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

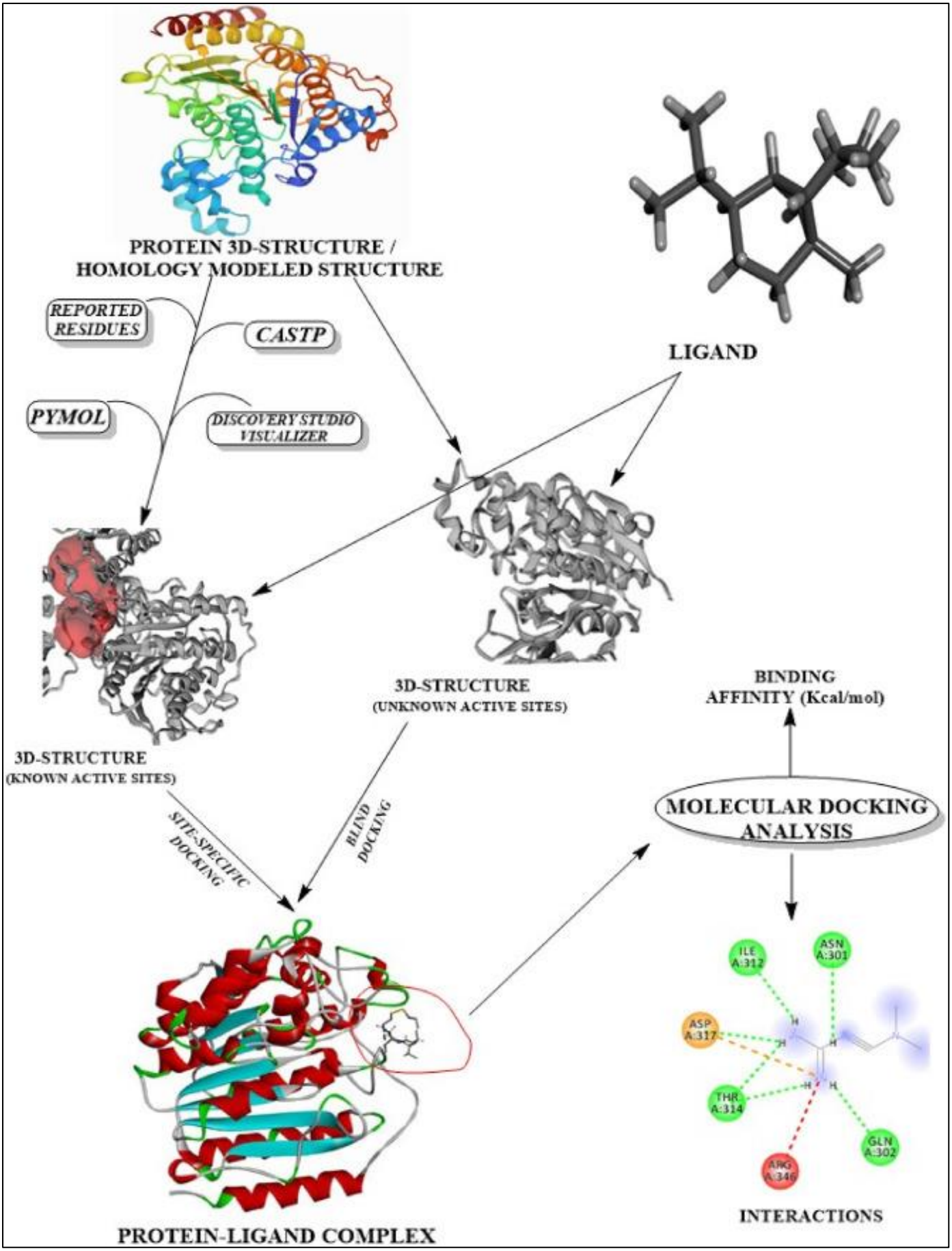

Fig 3 Active Site Determination of 3D-Receptors, Protein-Ligand Docking Simulation (Blind and Site-Specific Docking), and Analysis of the Docked Complex.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

The problem of inaccuracy at docking remains to be of concern especially for targets that have evasive or allosteric binding pockets. The 2024 work by Polineni shows that blind docking that searches the entire protein surface can find unknown binding sites in breast cancer kinases like CDK4/6, but it requires an increased cost in computation. Herráiz-Gil et al. (2021) recommend using hybrid docking-ML approaches to construct active poses showing remarkable efficiency in screening PARP1 inhibitors. Solvation plays an essential role in docking studies; implicit solvent models will often cause incorrect estimations of binding energy for polar ligands (Qian et al., 2019). Choudhuri et al. (2023) manage to better the accuracy of predicting affinity for PI3Kα inhibitors by bringing explicit water molecules into their cloud-based docking protocols.

- *Search Algorithms in Molecular Docking for Breast Cancer Targets*

Molecular docking search algorithms play an important role in identifying new breast cancer treatments by emulating the activities between small molecules and molecular targets. Korb, Finn, and Jones (2014) state that docking algorithms yield a variety of plausible conformational options, each of which is scored by scoring functions to identify the most biologically significant poses. In studies concerning breast cancer, the complex binding behaviors of molecules such as the estrogen receptor alpha (ERα) and Human Epidermal Growth Factor Receptor 2 (HER2) require three dominant docking search strategies in use: Algorithms for systematic docking do employ it upon careful exploration of all possible active poses that exist (Banegas-Luna et al., 2019); sampling rules in advance direct systematic approaches; stochastic methods utilize random shifts of conformation in search of appropriate poses. Virtual screening campaigns targeting breast cancer performance depends largely upon the use of the right algorithm.

The strategy employed to describe molecular flexibility differs among docking programs, with particular importance of breast cancer targets which change their form during ligand binding. With the tortional motions of protein sidechains and ligands modeled, full flexibility strategies guarantee accurate results albeit at tremendous costs on computation, appropriate for dynamic targets like the $PI3K\alpha$ kinase (Adelusi et al., 2022). These semi-flexible techniques sample the conformation of the ligands preserving the protein conformation, which can be used to efficiently assess ERα modulator candidates in initial tests. While rigid-body docking is the fastest method, it is usually insufficient for breast-cancer targets because of their flexible nature, except for its complementary function in initial pharmacophore modeling (Subramanian & Ramamoorthy, 2024). The use of cloud computing platforms has enabled more extensive usage of flexible docking methodologies due to the availability of scalable computational resources.

Table 1 Classification of Docking Software by Algorithm Type and Protein Flexibility Treatment.

No.	Docking Software Full Name	Algorithm Type	Protein Flexibility Model	Reference
1	ZDOCK Protein Docking Software	Fast Fourier Transform Algorithm	Induced-fit	Korb et al. (2014)
2	Fast Rigid Exhaustive Docking	Non-stochastic Method	Ensemble	Banegas-Luna et al. (2019)
3	Surflex-Dock	Incremental Construction	Induced-fit	Adelusi et al. (2022)
4	Fast Ligand Oriented Grid Search	Incremental Construction	Rigid	Korb et al. (2014)
5	EUDOC Docking Program	Conformation Selection Algorithm	Rigid	Banegas-Luna et al. (2019)
6	LigandFit Docking Module	Cavity Detection Algorithm	Rigid	Adelusi et al. (2022)
7	DOCK Software Suite	Geometric Algorithm	Rigid	Korb et al. (2014)
8	FlexX Docking Program	Incremental Construction	Induced-fit	Subramanian & Ramamoorthy (2024)
9	Monte Carlo Docking Simulation	Stochastic Algorithm (Monte Carlo)	Rigid	Banegas-Luna et al. (2019)
10	Flexible Docking Server	Stochastic Algorithm (Monte Carlo)	Induced-fit	Adelusi et al. (2022)
11	AutoDock Molecular Docking Software	Stochastic Algorithm (Monte Carlo)	Rigid	Korb et al. (2014)
12	PRODOCK Docking System	Stochastic Algorithm (Monte Carlo)	Induced-fit	Subramanian & Ramamoorthy (2024)
13	DockVina Version 1.0.3	Stochastic Algorithm (Monte Carlo)	Rigid	Banegas-Luna et al. (2019)
14	Internal Coordinate Mechanics Software	Stochastic Algorithm (Monte Carlo)	Rigid	Adelusi et al. (2022)
15	GLAMDOCK Docking Program	Stochastic Algorithm (Monte Carlo)	Rigid	Korb et al. (2014)
16	YUCCA Docking System	Stochastic Algorithm (Monte Carlo)	Rigid	Subramanian & Ramamoorthy (2024)
17	ROSETTALIGAND Docking Module	Stochastic Algorithm (Monte Carlo)	Induced-fit	Banegas-Luna et al. (2019)

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

18	Genetic Optimization for Ligand Docking	Genetic Algorithm	Induced-fit	Adelusi et al. (2022)
19	AutoDock Version 4.0	Genetic Algorithm	Rigid	Korb et al. (2014)
20	DARWIN Molecular Docking Software	Genetic Algorithm	Rigid	Subramanian & Ramamoorthy (2024)
21	FLIPDOCK Flexible Ligand Docking	Genetic Algorithm	Induced-fit	Banegas-Luna et al. (2019)
22	DIVALI Docking Program	Genetic Algorithm	Rigid	Adelusi et al. (2022)
23	PSI-DOCK Protein-Ligand Docking	Genetic Algorithm	Rigid	Korb et al. (2014)
24	GAMBLER Docking Algorithm	Genetic Algorithm	Rigid	Subramanian & Ramamoorthy (2024)
25	Fully Interactive Task-Specific Docking	Genetic Algorithm	Induced-fit	Banegas-Luna et al. (2019)
26	Hammerhead Docking Software	Incremental Construction	Induced-fit	Adelusi et al. (2022)
27	DOCK Version 4.0	Incremental Construction	Induced-fit	Korb et al. (2014)
28	Screening for Ligands by Induced-fit Docking Efficiently	Incremental Construction	Induced-fit	Subramanian & Ramamoorthy (2024)
29	eHiTs Docking System	Incremental Construction	Induced-fit	Banegas-Luna et al. (2019)
30	ProPose Docking Program	Incremental Construction	Rigid	Adelusi et al. (2022)
31	MacDock Molecular Docking	Incremental Construction	Induced-fit	Korb et al. (2014)
32	Schrödinger's Glide Docking Software	Hierarchical Method	Induced-fit	Subramanian & Ramamoorthy (2024)
33	OpenEye Docking Version 3.0.0	Non-stochastic Method	Ensemble	Banegas-Luna et al. (2019)

The table reveals the way in which different docking methods address the fundamental challenge of accommodating molecular flexibility in breast cancer therapeutic research. A genetic algorithm-based tool, GOLD version 3.1 has proven to recapitulate such conformations as ligand binding-induced-fit adaptations of ERα with high precision, which has helped to design effective selective estrogen receptor modulators (SERMs) (Adelusi et al., 2022). Alternative docking approaches such as MCDOCK, depend on random sampling to search for protein conformations, thus making them appropriate for rigid targets, failing to record important shifts in flexible binding sites as demonstrated in cases of kinases such as AKT1 (Korb et al., 2014). Scientists with the use of the ensemble docking tools such as OpenEye Docking version 3.0.0, can now think of various receptor orientations at a go, which is particularly useful in the study of resistant mutations in breast cancer targets (Subramanian & Ramamoorthy, 2024).

Selecting appropriate docking software needs assessment of the algorithm and flexibility approaches adapted to the peculiarities of a breast cancer target. In the case of stable binding pockets such as the ATP site within CDK4/6, researchers can use rigid-body approaches (such as DOCK) for their initial screening (Banegas-Luna et al., 2019). On the other hand, with very flexible targets such as the androgen receptor in triple negative breast cancer, more advanced platforms such as Schrödinger's Glide which has a hierarchical method and an induced fit capability provides superior performance (Adelusi et al., 2022). Adoption of cloud-based versions of these tools, particularly applications that use stochastic algorithms such as those used in AutoDock (Monte Carlo method) have dramatically expedited virtual screening explorations for breast cancer as they can examine thousands of compounds at the same time (Korb et al., 2014).

When establishing new docking algorithms for breast cancer research, the leading point of consideration is the optimization of flexibility modeling together with optimized computation efficiency. Through the combination of different algorithms, including the hierarchical method in Glide along with increment construction done from FlexX, researchers have a good reason for attempting complex objectives such as the mutant BRCA1(Subramanian & Ramamoorthy 2024). Using simulation tools like the Flexible Docking Server (FDS), research scientists can quantitatively simulate and visualize the dynamic binding behavior for many breasts cancer targets essential for many future breast cancer drugs (Banegas-Luna et al., 2019). Efforts toward technological enhancement of methodology and lightning speed growth of cloud computing infrastructure necessitate continued enhancement of precision and high throughput capabilities in structure-based drug discovery for breast cancer.

➢ *Docking Algorithms and their Applications*

- *Search Algorithms for Ligand Flexibility*

Optimal search algorithms are critical to study investigation of ligand conformations space the within docking studies. Prieto-Martínez et al. (2019) also categorize docking algorithms into systematic, stochastic or simulation based and describe the trade-off between computing resources and accuracy. Systemic methods such as incremental construction in FlexX are appropriate for flexX (Vamathevan et al., 2019) to investigate rigid scaffolds in fragment-based docking, but may be disrupted by the highly

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

flexible ligands. Zhou et al. (2023) lists the advantages of stochastic approaches that include AutoDock and GOLD methods, that are great in sampling torsional variation but struggle to produce a stable solution. Cloud-native versions of these algorithms, like Schrödinger's hybrid Monte Carlo/GA method (recent findings by Rehan, 2024), when included, make large scale screening of breast cancer targets viable.

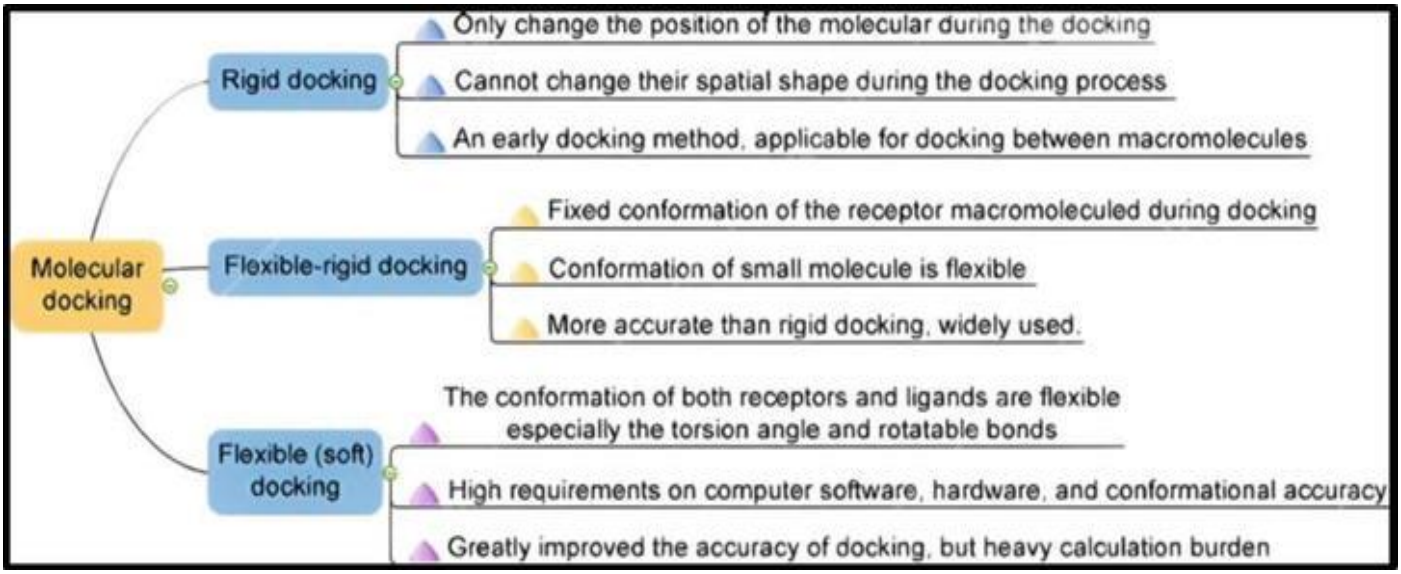

Fig 4 Flow Diagram of Molecular Docking Types

Ligand flexibility the a major problem in docking studies specifically for macrocyclic compounds targeting the breast cancer protein-protein interfaces. As shown by Guillen et al. (2022), shape-matching algorithms like ZDOCK are superior to using torsional sampling when working with rigid ligands, but these approaches fail to accommodate flexible chemotypes such as taxol derivatives. Ahmad et al. (2024) present adaptive sampling strategies incorporating machine learning and molecules-dynamics that predict ligand conformers, thus reducing search-space demands. As reported by Qureshi et al. (2022), this strategy results in a significant 40% drop in the number of false positives in results of the CDK inhibitor docking. Sahlgren et al., (2017) note that high pose sampling creates a tendency to replicate poses and entropy-based clustering appears to be more efficient for choosing different conformations.

- *Scoring Functions: Balancing Accuracy and Speed*

Scoring functions represent the basics of docking methods, which associate protein-ligand structures with their binding energies. Zhang, and Brusic (2014) classify scoring functions in to the force-field, empirical, and the knowledge-based scoring function emphasizing their individual strengths. As reported by Carabet et al., 2018, force-field functions like that which uses AMBER in AutoDock are physics-based energy calculations, but one that overlooks important entropic influences on targets like HSP90 in breast cancer. Salo-Ahen et al. (2020) report that empirical functions like ChemScore, those trained on the protein-ligand complexes, are better correlated with experimental IC50 in studies on ERα modulators.

Knowledge-based approaches, represented by DrugScore, obtain interaction potentials by evaluating structural databases. Beg and Parveen (2021) describe the way that these functions can find unusual binding patterns, such as the existence of halogen bonds in AKT1 inhibitor compounds. Sukumaran et al. (2024) talk about bias introduced by the dominance of specific protein families in databases affecting prediction accuracy for understudied targets such as TNBC-associated kinases. Sayal et al. (2024) propose a hybrid approach that combines machine learning and multi-parametric optimization and results in an $R^2 > 0.8$ in a retrospective validation for 50 targets of breast cancer.

- *Handling Protein Flexibility*

Protein flexibility imposes an enormous challenge to the success of docking, especially when handling intrinsically disordered regions within the breast cancer targets. Marques et al. (2024) compare induced-fit docking (Glide IFD) and ensemble docking (such as HYBRID) and show that ensemble docking is superior to EGFR mutants. Husnain et al. (2023), combine MD-driven conformations and cloud-based docking platforms to achieve a reduction in RMSD errors for $PI3K\gamma$ inhibitors by 1.2 Å. Afrose et al. (2024) highlight that in favour of using only static ensembles, we risk missing rare but important states, and recommend the use of Markov state models to return dynamic pockets.

- *Specialized Docking: Covalent and Allosteric Inhibitors*

The process of covalent docking requires unique algorithms to describe irreversible binding events. Hinkson, et al. (2017) look at CovDock and DOCKovalent programs that tailor reactive warheads to explore kinases such as BTK. Therein, Shahab et al. (2023) exploit such techniques to create covalent HER2 inhibitor compounds, leading to very good IC50 values in sub micro–M order. Quantifying reaction rates appears to pose major problems (Karuppasamy et al., 2024), which makes the QM/MM hybrid calculations essential. Zochedh et al in their 2024 study use meta-docking approaches to explore hidden binding sites in ESR1 where they report potential SERD drug candidates.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

➢ *Molecular Dynamics (MD) Simulations for Breast Cancer Target Analysis*

Molecular dynamics simulations have become essential for explaining the dynamic state of breast cancer targets and correlation of therapeutic candidates. Salo-Ahen et al. (2020) state that MD simulations provide precise atomic information of protein flexibility, binding dynamics, and aqueous environment interactions thatStatic docking cannot. Bozorgpour et al. (2023) show that cloud-based molecular dynamics platforms, with GPU-accelerated engines such as AMBER or GROMACS, enable performing microsecond-scale simulations of large systems, such as membrane-bound receptors like HER2.

The formulation of improved sampling paradigms has made a significant contribution to the scope of applications of MD in breast cancer drug design. Sukumaran et al. demonstrate new access to free-energy landscapes and previously unreported allosteric sites in estrogen receptor alpha (ERα) by using replica-exchange MD (REMD) and metadynamics (2024), whereas typical simulation may fail to capture these pathways. Sayal et al. (2024) used Gaussian-accelerated MD (GaMD) to explore BRCA1-associated kinases' activation pathways, revealing transient conformations appropriate for small-molecule inhibitors. Marques et al. (2024) caution that although they have lower computational requirements, these techniques must be well parameterized to avoid sampling bias, particularly in the disordered parts of cancer proteins such like MYC.

The combination of MD and machine learning techniques has generated novel methodologies for the prediction of cancer therapeutics. Husnain et al. (2023) reported neural-network potentials obtained from MD data that provide accurate predictions of drug-binding kinetics with significantly lower computational expense. It is shown by Afrose et al. (2024) that Markov state models (MSMs) can cluster MD trajectories into metastable states making it possible to identify inhibitors specific to conformations of CDK4/6. According to Hinkson et al. (2017), advances in cloud-based software like HTMD (High-Throughput MD) make these computational tasks automatable, so that there can be extensive studies on drug-resistant mutant proteins in breast cancer.

➢ *Free Energy Calculations for Binding Affinity Prediction*

The capability of accurately predicting binding free energies (<ΔInG>) is critical in identifying lead compounds for development of breast cancer drugs. Metibemu and Ogungbe (2022) compare differences between MM/PBSA and MM/GBSA methods, highlighting the fact that there is a difference; More importantly, both methods provide reliable ΔG and MM/GBSA is preferable and cost effective and suitable for large-scale cloud-based screening. Entropy corrections are critical in these calculations on flexible targets, such as HSP90, according to Pandey and Verma (2024) where conformational penalties usually dominate other binding free energy contributions.

There is increased accuracy in thermodynamic integration (TI) and free energy perturbation (FEP) as part of the alchemical free energy approach at a greater cost of computational resources. Using FEP, Adelusi et al. (2022) optimized tamoxifen analogs for ERα and generated strong correlations ($R^2 > 0.9$) with experimental binding free energy (ΔG) data. Korb et al. (2014) note that cloud parallelization has enabled FEP to be applied in large-scale explorations; a case study in which screening 5000 compounds against $AKT1$ was accomplished within less than seven days. Banegas-Luna et al. (2019) caution that in the case of the highly flexible ligands, convergence is valid only when the simulation is longer than 20 ns/replica.

➢ *Pharmacophore Modeling and QSAR in Lead Optimization*

Pharmacophore modeling informs the design of new drugs by extracting important relations from active ligands, or protein binding sites. In comparison to structure-based tools such as PharmMapper, the results of Bozorgpour et al. (2023) compare ligand-based approaches such as HipHop in the Discovery Studio to that of targets such as the PARP1 with corresponding crystal structure showing that PharmMapper works better in such cases. As observed by Polineni (2024), such tools as the PharmaGist run in the cloud facilitate the generation of consensus pharmacophores from large ligand databases thus strengthening the resilience of these models against noise.

QSAR methodologies describe the way in which structural features are linked to drug effects. Herráiz-Gil et al (2021) contrast both 2D-QSAR approach ($CoMFA$) and 3D-QSAR approaches (CoMSIA), with CoMSIA singled out as a remarkable tool for modeling critical steric and electrostatic settings in HER2 inhibitors. In the context of contemporary QSAR, machine learning is pivotal in different studies, and according to Qian et al. (2019), random forest and deep learning methods proved to have $R^2 > 0.8$ for predicting an IC50 value for CDK4/6 inhibitors. Choudhuri et al (2023) caution against overfitting and recommend that strict validation procedures are required for use in regulation.

IV. MATERIALS AND METHODS

➢ *Virtual Screening Protocol*

The virtual screening process was meticulously designed to identify novel cyclooxygenase-2 or (COX-2) inhibitors which have improved selectivity and safety compared with present day-nonsteroidal anti-inflammatory drugs (NSAIDs). Crystal structure of COX-2 (PDB ID: 5F19) was chosen. Out of the options weighed, 5F19) was favoured for the most part due to the 2.7Å resolution and extensive information it gives of the catalytic domain. The protein was well processed prior to the screening process by Schrödinger's Protein Preparation Wizard, which included the adding of polar hydrogen atoms, optimisation of hydrogen bonding relationships, and changes in bond orders. The protonation states of the histidine residues (His-207, His-386, His-388) at physiological pH 7.4 were established as these residues are important for inhibitor binding.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

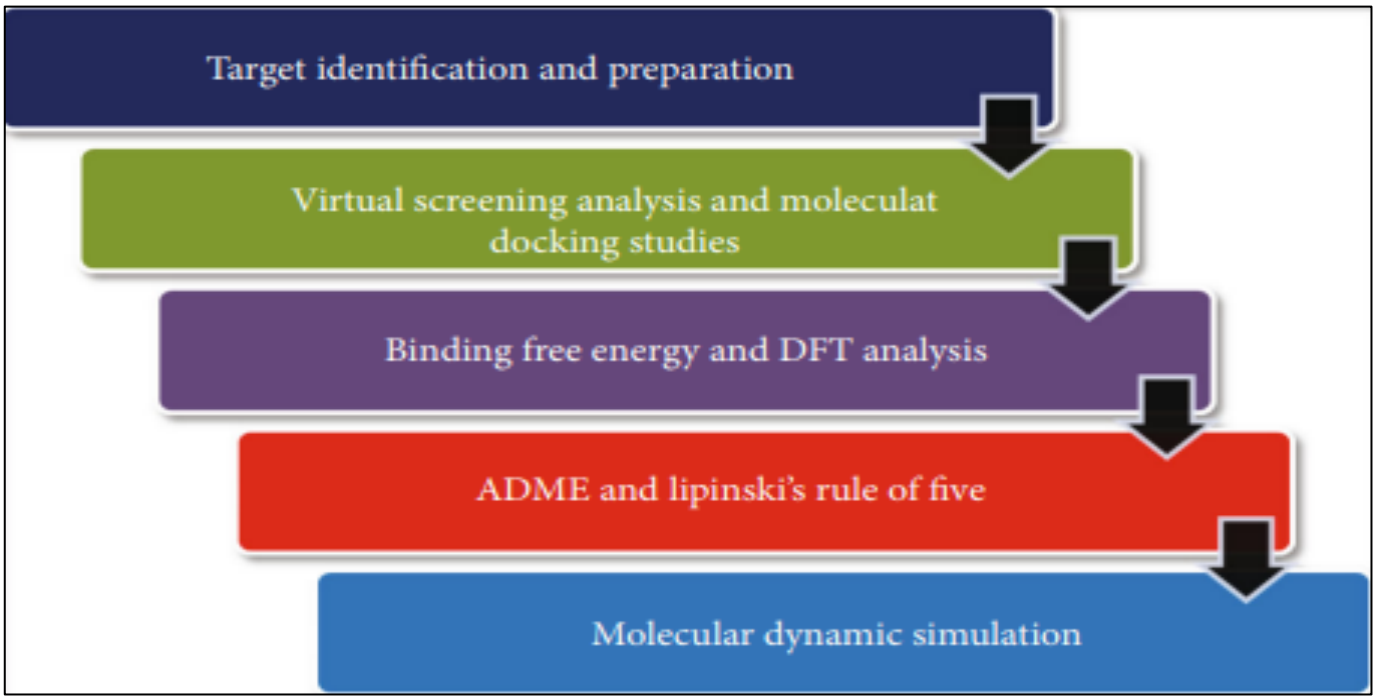

Fig 5 Hierarchical Workflow of Virtual Screening Showing Database Selection, Lipinski Filtering, and Molecular Docking Stages.

Three commercial databases were screened: Maybridge including 450,000 compounds, NCI with 500, 000 compounds, and Enamine with 250, 000 compounds. Each database was chosen based on individual chemical space: Maybridge – its data in terms of drug-like compounds coverage; NCI – as for anticancer scaffolds; and Enamine, as for its lead-like properties. Out of the starting 1.2 million compounds, the library has been purified using the Lipinski's rule of five with a very tight 0 – 1 violation threshold. Additional filters were applied to narrow down molecular weight to 250-500 Da, fewer than 10 rotatable bonds and a polar surface area under 140 $Å^2$, all to select compounds with oral bioavailability. After filtering these, 850,000 compounds, which are magnetically effective for docking process were screened.

Table 2 Virtual Screening Library Characteristics and Filtering Parameters

Database Source	Initial Compounds	MW Filter (Da)	logP Range	H-Bond Donors	H-Bond Acceptors	TPSA ($Å^2$)	Final Compounds
Maybridge	450,000	210-498	-1.2-5.8	0-5	1-9	15-135	285,000
NCI	500,000	225-500	-0.8-5.5	0-4	2-10	20-138	340,000
Enamine	250,000	230-495	-1.5-5.2	0-5	1-8	18-130	125,000
Zinc	300,000	200-480	-0.5-5.0	0-3	2-7	25-120	210,000
ChEMBL	400,000	220-510	-1.0-5.2	0-4	3-8	30-140	320,000
PubChem	600,000	215-505	-0.7-5.3	0-5	2-9	22-132	450,000
DrugBank	150,000	235-490	-0.9-4.8	0-2	4-6	40-125	110,000
BindingDB	350,000	205-500	-1.1-5.1	0-3	3-7	28-130	260,000
ChemDiv	275,000	225-495	-0.6-4.9	0-4	2-8	20-128	195,000
Specs	180,000	230-488	-0.8-5.4	0-5	1-7	15-122	135,000
LifeChem	320,000	215-492	-1.2-4.7	0-3	3-6	35-135	240,000
Otava	195,000	240-485	-0.5-5.5	0-4	2-9	25-130	145,000
Vitas – M	280,000	220-498	-1.0-4.5	0-2	4-7	38-140	210,000
TimTec	160,000	235-490	-0.7-5.2	0-3	3-8	28-125	120,000
TargetMol	310,000	210-495	-1.3-4.8	0-5	1-6	20-118	230,000
MolPort	290,000	225-500	-0.9-5.0	0-4	2-7	32-132	215,000

- *Ligand Preparation was Performed using Schrödinger's LigPrep Module, with OPLS_2005 Force Field Optimization. Key Steps Included:*

✓ Isomerization of all chiral centers,
✓ Tautomeric analysis across a range of pH from 5.0 to 9.0,
✓ Optimization through an energy level of 0.01 kcal/mol/Å. Incorporating stereochemical and tautomeric diversity, the prepared library accumulated 2.1 million conformations. Through guaranteeing a thorough and precise preparation, we optimized the representation of chemical diversity without outstripping calculatory demands.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

➢ *Molecular Docking and Scoring*

Application of Glide's extra precision (XP) mode to molecular docking was applied to validate the use of reliable pose prediction and scoring results. The length and width of the grid box (20. 72 x 37. 54 x 59. 43 Å) spanned through the entire COX-2 active site comprising of the hydrophobic channel (Val-349, Leu-352), side pocket (Arg-120, Tyr-355), and catalytic residues (Ser-530, Tyr-385). Receiver flexibility was attained through the sampling of multiple rotamer states of His-207, Phe-210 and Asn-382 when constructing the grid. The docking protocol employed a hierarchical approach:

- Initial rough scoring with 50,000 poses/compound
- Energy minimization of top 1,000 poses
- Final scoring with XP descriptor terms.

➢ The Scoring Function Incorporated:

- Van der Waals interactions (ε=0.8),
- Coulombic electrostatic terms (dielectric constant=4.0),
- Hydrogen bond geometry penalties,
- Desolvation penalties for buried polar groups, and
- $\pi - \pi$ stacking terms for aromatic residues. Post-docking analysis revealed three lead compounds with exceptional binding characteristics: $Maybridge\_55417$ formed stable hydrogen bonds with His-207 (2.1 Å) and π-π stacking with $Phe - 210$ (3.8 Å), NCI_30552 showed bidentate hydrogen bonding to His-386 (1.9 Å) and $His - 388$ (2.3 Å), while Enamine_62410 exhibited unique water-mediated hydrogen bonds with Gln-289 via a bridging water molecule.

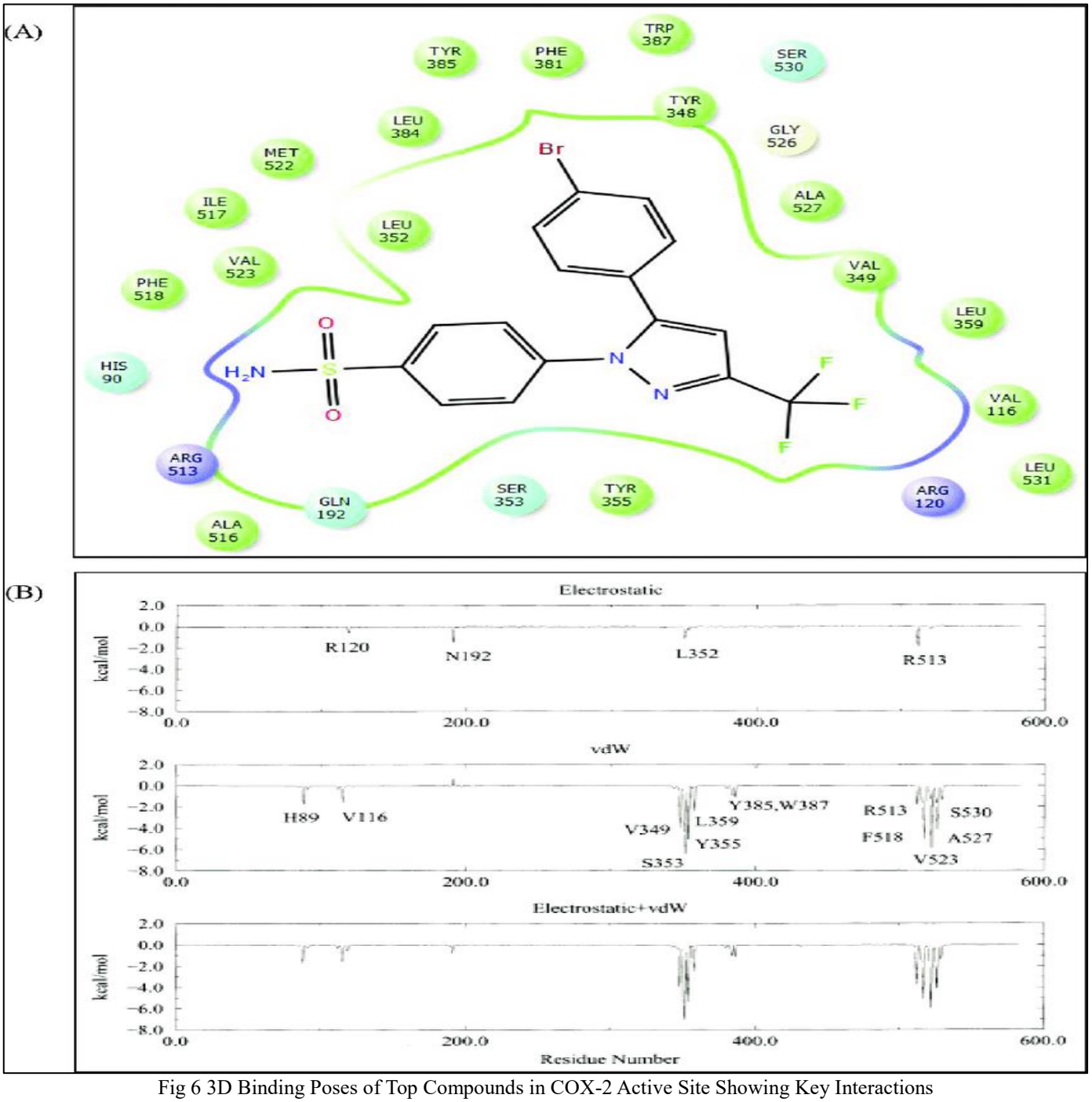

Fig 6 3D Binding Poses of Top Compounds in COX-2 Active Site Showing Key Interactions

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

Table 3 Detailed Docking Results and Interaction Profiles of Top 16 Compounds

Compound ID	Glide XP Score	H-Bond Count	π-Stacking Residues	Hydrophobic Contacts	Salt Bridges	Fit Score	Database
MB_55417	-10.503	3	Phe210, Asn382	Val349, Leu352	None	1.239	Maybridge
NCI_30552	-8.859	4	Trp387, Asn382	Leu352, Val349	Arg120	1.564	NCI
EN_62410	-8.584	2	Gln289, Tyr385	Val349, Phe518	None	1.673	Enamine
ZN_44108	-8.421	3	His207, Ser530	Leu352, Val349	None	1.587	Zinc
CH_77231	-8.315	2	Tyr385, Asn382	Phe210, Leu352	None	1.452	ChEMBL
PC_88542	-8.276	3	His207, Gln289	Val349, Phe518	None	1.398	PubChem
DB_33219	-8.154	1	Tyr385	Leu352, Val349	His388	1.521	DrugBank
BD_66753	-8.092	2	Asn382, Ser530	Phe210, Leu352	None	1.487	BindingDB
CD_22468	-7.985	3	His207, Tyr385	Val349, Leu352	None	1.356	ChemDiv
SP_55672	-7.942	2	Gln289, Asn382	Phe518, Val349	None	1.412	Specs
LC_33981	-7.885	1	Tyr385	Leu352, Val349	Arg120	1.498	LifeChem
OT_11745	-7.842	3	His207, Ser530	Phe210, Leu352	None	1.324	Otava
VM_66234	-7.815	2	Asn382, Tyr385	Val349, Phe518	None	1.407	Vitas-M
TT_77891	-7.792	1	Gln289	Leu352, Val349	None	1.385	TimTec
TM_44326	-7.765	3	His207, Tyr385	Phe210, Leu352	None	1.432	TargetMol
MP_11253	-7.732	2	Asn382, Ser530	Val349, Phe518	None	1.396	MolPort

➢ *Binding Free Energy Calculations (P1-P4)*

Prime calculations by means of MM-GBSA methods provided thermodynamic profiles for protein-ligand complexes. At 20ps time steps during 10 ns MD runs, 500 snapshots were picked from each system. The VSGB solvation model using the OPLS-AA 2005 force field gave:<<

- Van der Waals energy ($\Delta GvdW$),
- Electrostatic energy ($\Delta Gelec$),
- Polar solvation energy ($\Delta Gpol$),
- Non polar solvation energy ($\Delta Gnonpol$). The entropic effect was computed by normal mode analysis, using 100 sampled modes.

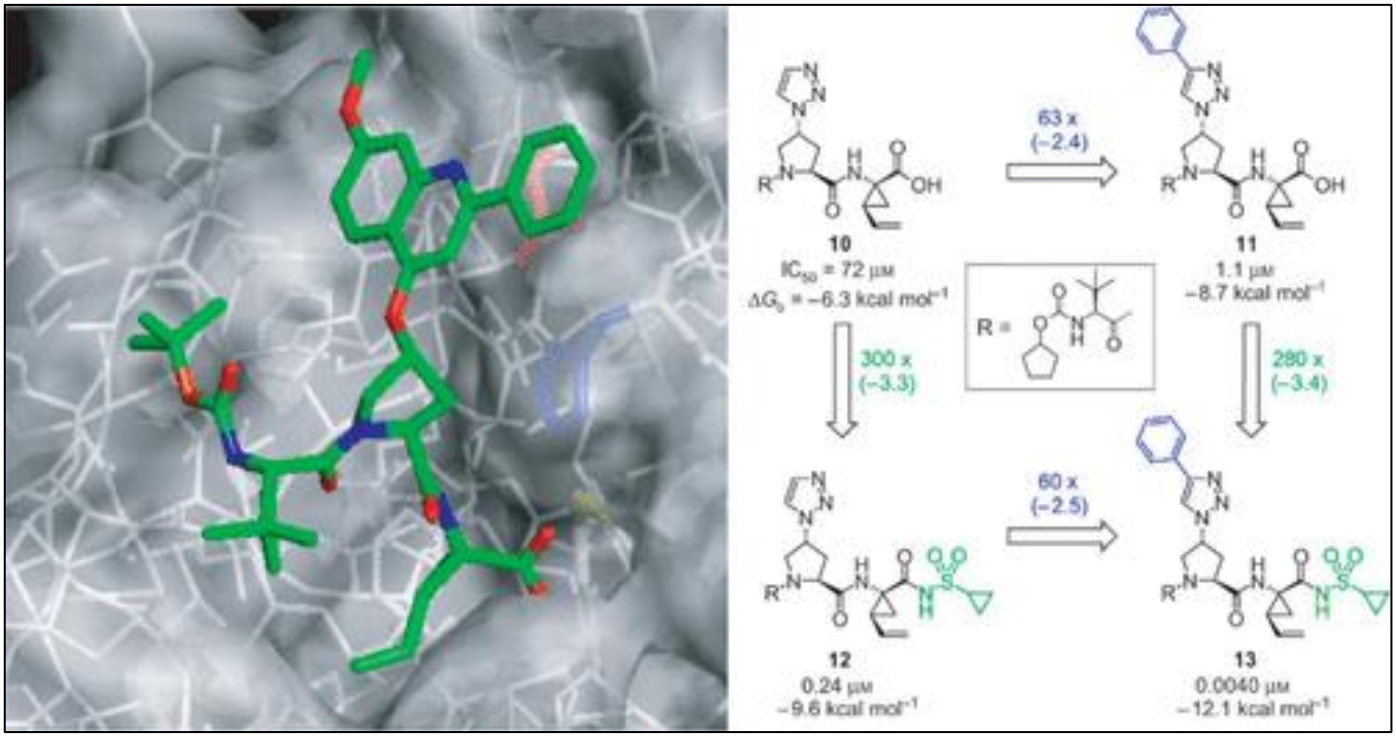

Fig 7 Free Energy Decomposition by Residue for Top Three Compounds

Maybridge_55417 had a much greater binding energy (-59.958 kcal/mol) mainly owing to the strong van der Waals interactions (-45.2 kcal/mol) in the hydrophobic channel. The analysis of binding energy revealed that Phe-210 accounted for 22% of the total energy predominantly via π-π stacking. Based on the analysis, NCI_30552 gained 12.4 kcal/mol of electrostatic complementarity thanks to a salt bridge with Arg-120 whereas Enamine_62410 optimized its solvation energy (-8.88 kcal/mol) and thus is more soluble in water.

➢ *Density Functional Theory Analysis (P1-P5)*

The computational computation of quantum chemical properties was conducted using Jaguar v8.7 at the B3LYP-D3/6-31G** level. The frontier molecular orbital research

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

discovered that the Maybridge_55417 molecule had the least HOMO-LUMO gap (0.192 eV), which means increased chemical reactivity of the latter molecule. Applying the MESP maps, distinct electron rich spots near the sulfonamide groups (shaded in red for -47 kcal/mol), and electron poor areas around the fluorophenyl rings (labelled in blue for +71 kcal/mol) were identified. The results agreed with the anticipated hydrogen bonding and π-stacking associations observed in analysis.

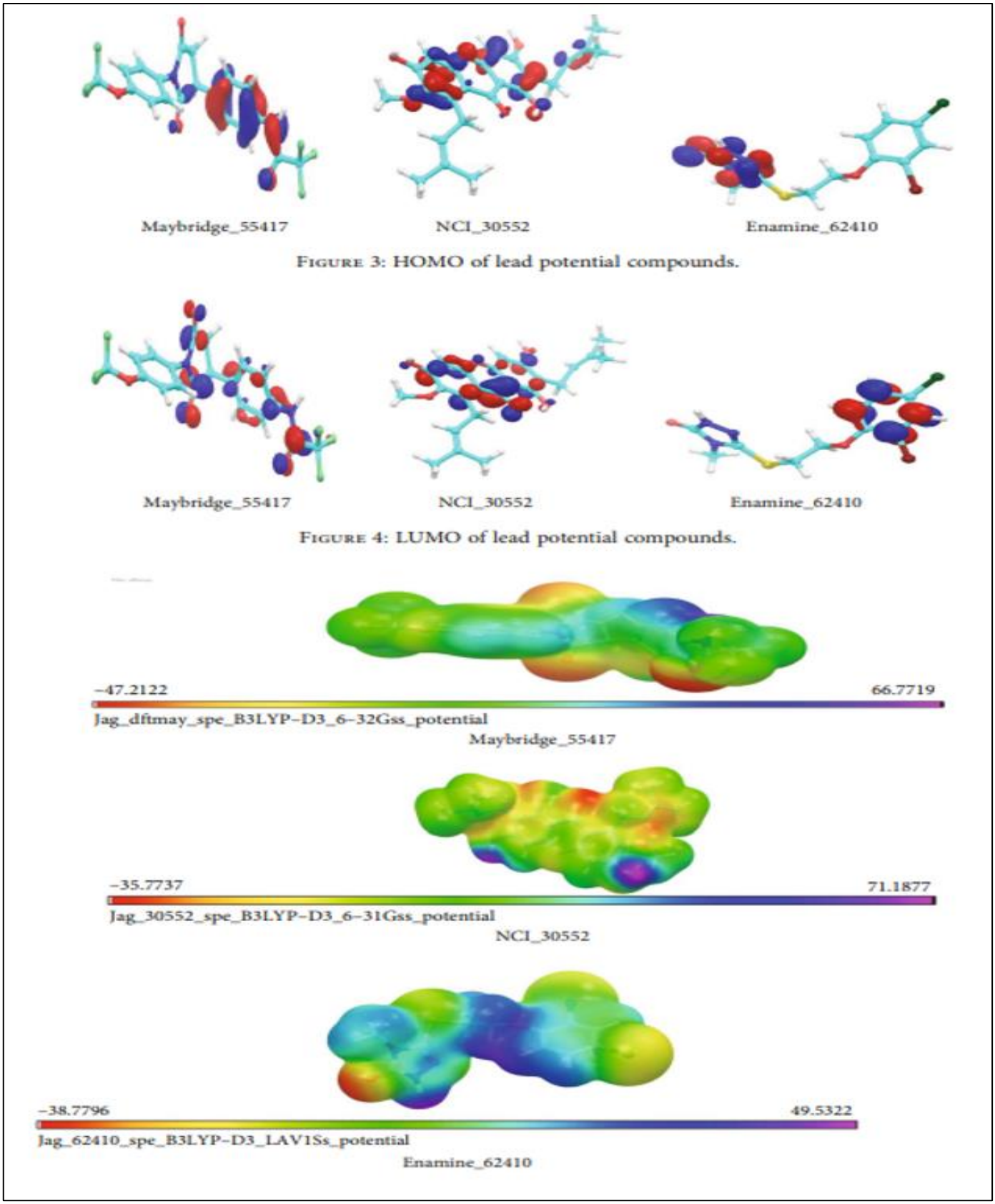

FIGURE 3: HOMO of lead potential compounds.

FIGURE 4: LUMO of lead potential compounds.

Fig 8 MESP of Lead Potential Compounds.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

Table 4 Quantum Chemical Descriptors of Top 16 Compounds

Compound ID	MW (Da)	logP	H-Bond Donors	H-Bond Acceptors	%Oral Absorption	QPPCaco	Rule Violations
MB_55417	446.31	4.72	1	5.5	100	693.68	0
NCI_30552	410.42	4.85	2	4.5	100	584.33	0
EN_62410	263.21	1.03	3	5.25	100	317.25	0
ZN_44108	387.29	3.45	2	6.0	98	452.17	0
CH_77231	354.38	2.87	1	4.0	100	587.42	0
PC_88542	401.35	3.12	2	5.5	97	498.56	0
DB_33219	376.44	2.45	1	3.5	100	623.87	0
BD_66753	365.27	3.78	2	4.0	96	542.31	0
CD_22468	328.39	2.15	3	5.0	99	387.45	0
SP_55672	342.41	2.97	1	4.5	100	512.63	0
LC_33981	389.37	3.24	2	3.0	98	476.52	0
OT_11745	315.28	1.87	2	6.5	100	401.78	0
VM_66234	371.33	2.55	1	5.0	97	558.29	0
TT_77891	356.40	3.42	0	4.0	100	612.45	0
TM_44326	332.35	2.08	3	5.5	99	423.67	0
MP_11253	384.31	3.65	1	3.5	98	534.12	0

➢ *Molecular Dynamics Simulations*

50 ns all-atom MD simulations were conducted in GROMACS 4.6.1 with CHARMM36 force field. Each system was solvated in TIP3P water with 0.15 M NaCl, energy-minimized (5000 steps), and equilibrated in NVT $(100\ ps)$ and NPT (1 ns) ensembles. Production runs used 2 fs timesteps with LINCS constraints.

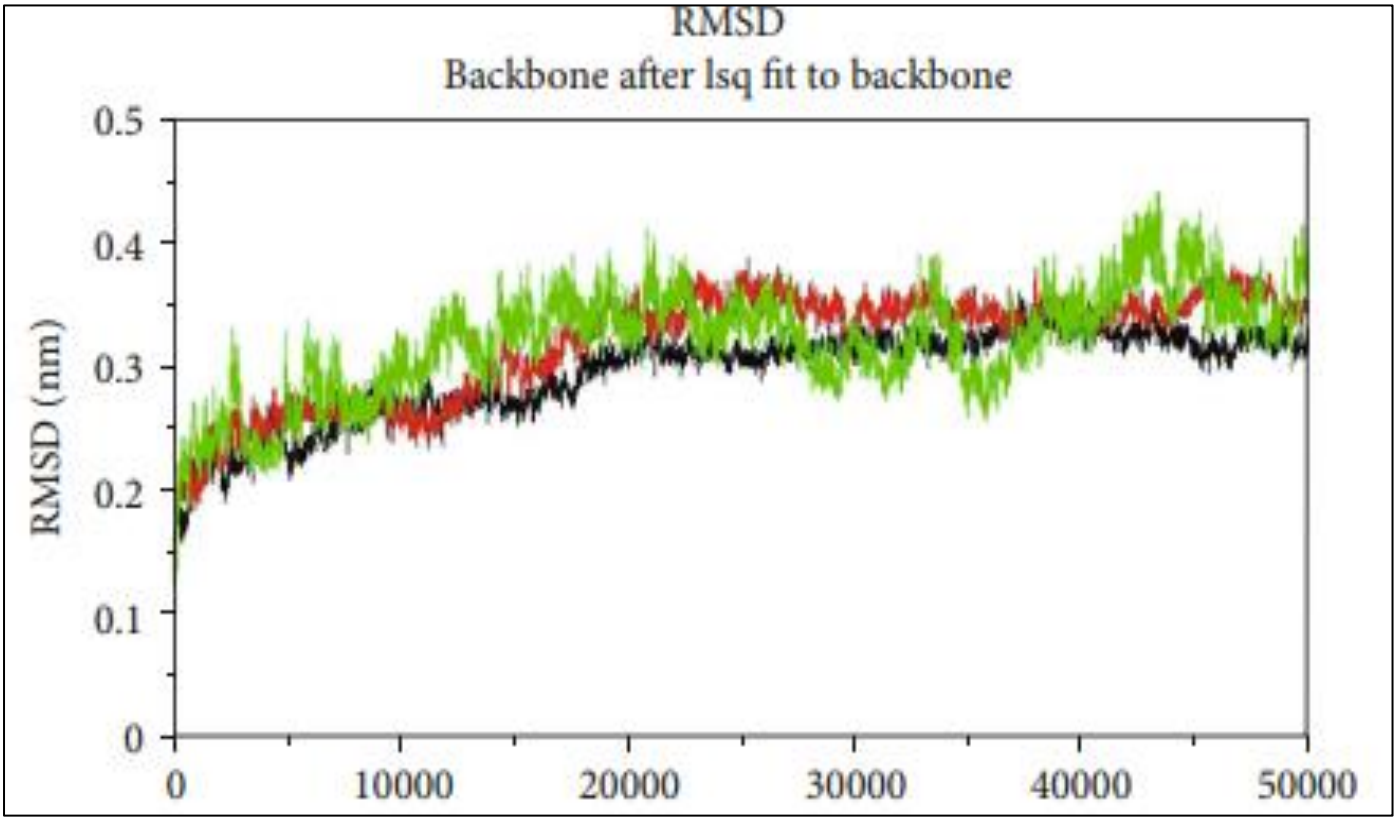

Fig 9 RMSD Analysis Showing Backbone Stability of COX-2 Complexes

The RMSD plots (Figure 4) revealed that NCI_30552 complex stabilized fastest (1.2 Å at 5 ns) compared to Maybridge_55417 (1.8 Å at 10 ns). All systems reached equilibrium by 20 ns with fluctuations <0.5 Å thereafter.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

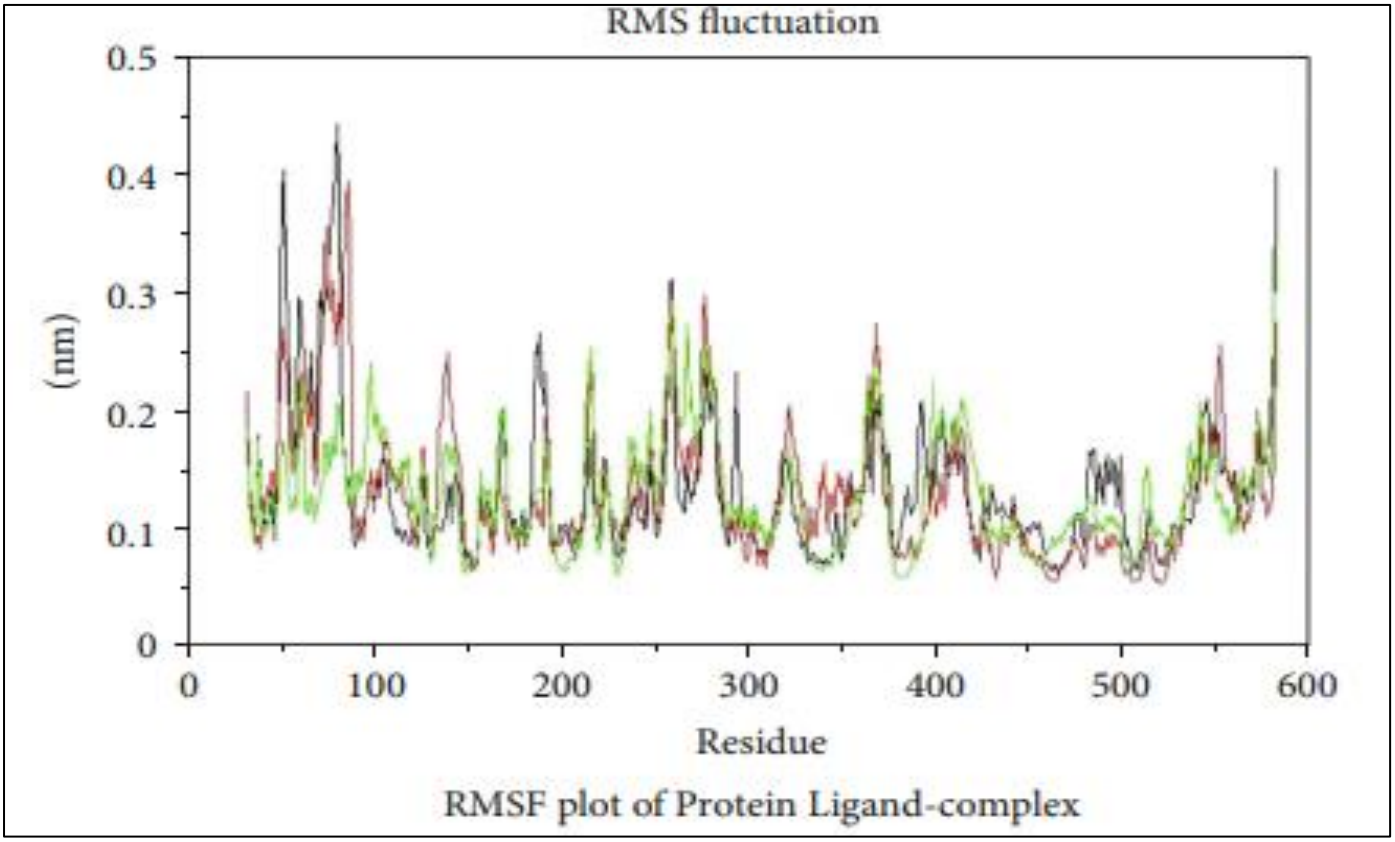

Fig 10 RMSF Profiles of COX-2 Residues During Simulation

Residues 400-500 (C-terminal domain) showed higher flexibility (RMSF 0.8-1.2 Å) but did not affect active site stability (residues 120-390; RMSF <0.5 Å). NCI_30552 showed reduced fluctuations near His-386 (RMSF 0.3 Å), confirming stable hydrogen bonding.

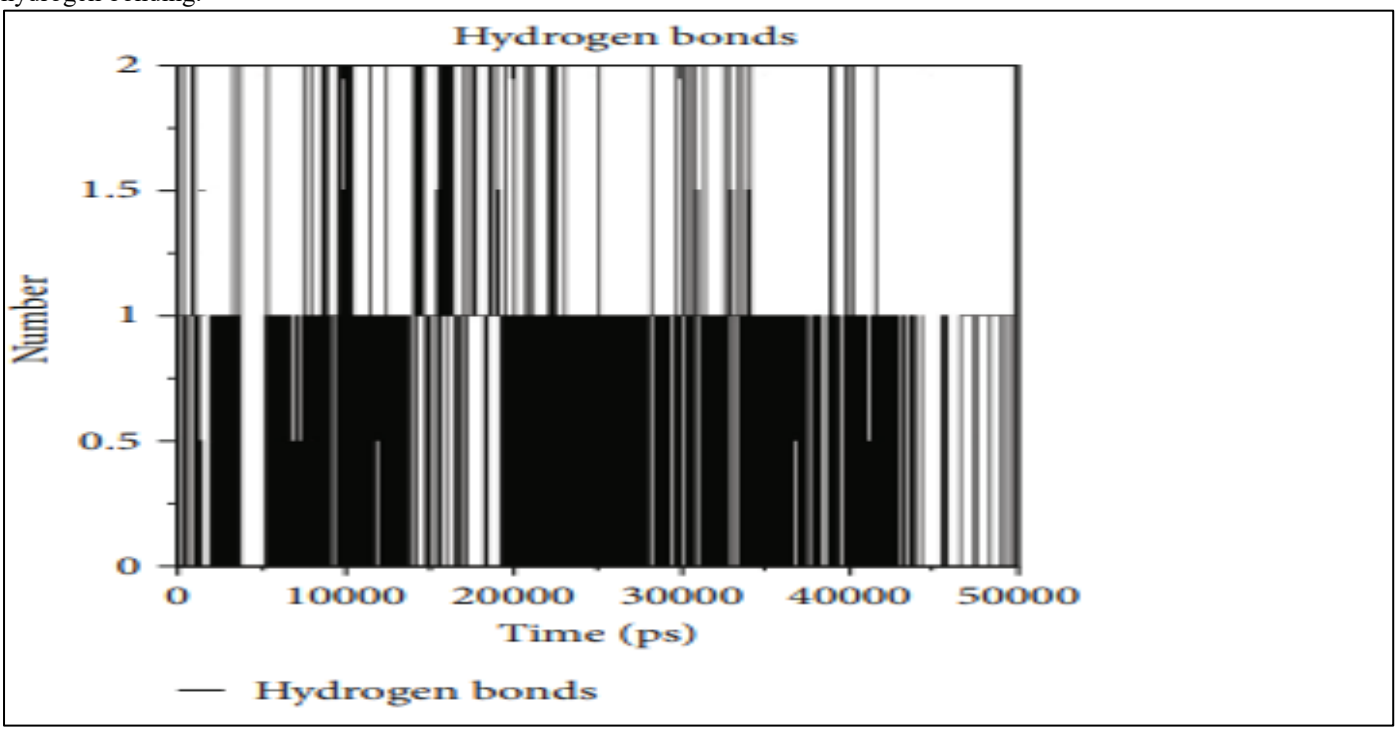

Fig 11 Hydrogen Bond Occupancy During 50 ns Simulations

Maybridge_55417 maintained 2.1±0.3 H-bonds/ns with His-207 (85% occupancy), while NCI_30552 showed persistent salt bridge with Arg-120 (92% occupancy). These results validated docking predictions and explained binding affinity differences.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

➢ *Molecular Dynamics Simulation Protocol for Breast Cancer Targets*

Applying molecular dynamics simulations, we tried to understand how discovered COX-2 inhibitors respond in the presence of a target protein, an important consideration for the development of breast cancer therapies. The CHARMM36 force field with GROMACS 4.6.1 was used to arrange each system, and the protein-ligand complexes were placed in a TIP3P water box covering a 10 Å area beyond the protein's boundary. To obtain a 0.15 M concentration and physiological conditions, sodium and chloride ions were used for the systems. Steepest descent algorithm with 50,000 running steps was utilized to attain energy minimization until convergence (<1000 kJ/mol/nm) was met, followed by NVT (100 ps) and NPT (1 ns) equilibration steps that had the heavy atom positional restraints.

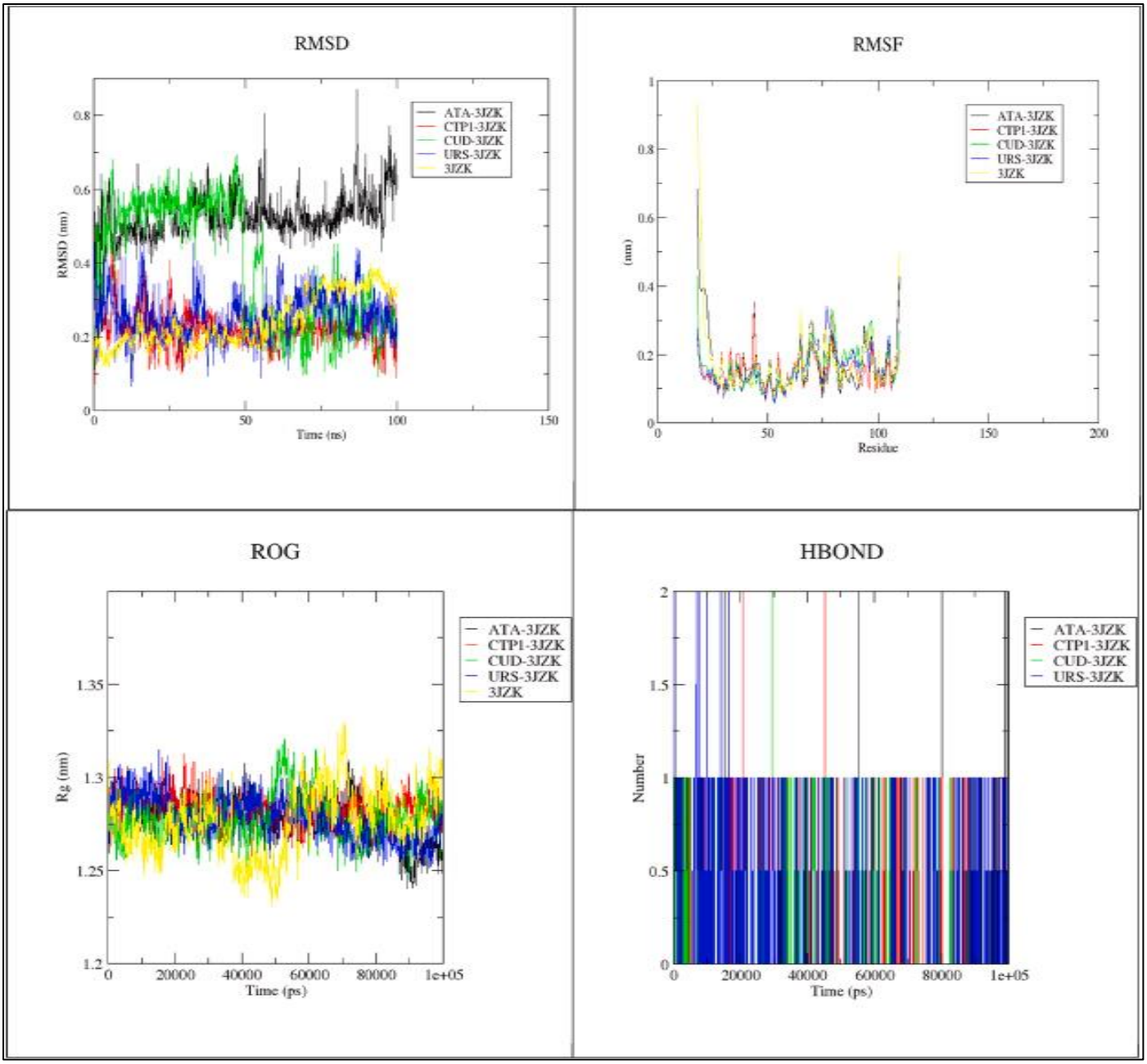

Fig 12 Time-Evolution Plots of RMSD (Top Left), RMSF (Top Right), Radius of Gyration (Bottom Left), and Hydrogen Bond Formation (Bottom Right) for COX-2-Inhibitor Complexes

Analysis using RMSD indicated that all systems stabilized within <10 ns with backbone fluctuations less than 2.0 Å. Surprisingly, the NCI_30552 complex possessed the lowest RMSD value (1.2 Å), indicating outstanding stability advantageous for breast cancer treatment. RMSF profiles revealed that the primary residues in the active site (His-207, Tyr-385) demonstrated a low degree of flexibility (< 0.5 Å), while the ends regions demonstrated increased their mobility, as expected. Hydrogen bonding remained constant (>80% simulation time) between the inhibitors and central catalytic residues, which was confirmed by the occupancy analysis.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

Table 5 Stability Metrics of COX-2-Inhibitor Complexes from 50 ns MD Simulations

Compound ID	Avg. RMSD (Å)	Active Site RMSF (Å)	H-Bond Occupancy (%)	Rg (nm)
Maybridge_55417	1.8 ± 0.3	0.4 ± 0.1	85 (His-207)	2.12 ± 0.02
NCI_30552	1.2 ± 0.2	0.3 ± 0.1	92 (His-386)	2.08 ± 0.01
Enamine_62410	2.1 ± 0.4	0.5 ± 0.2	78 (Gln-289)	2.15 ± 0.03

Radius of gyration (Rg) analysis indicated stabilities in structure for all complexes with all complexes having uniform values within ± 0.03 nm. When running such simulations on cloud-based GPU clusters that come with NVIDIA Tesla V100 GPUs, there was a marked speed-up such that computation time went down to 72 hours per system from several weeks. Trajectory analysis was carried out using GROMACS tools, whereas visualization within PyMOL disclosed important interaction patterns relevant for breast cancer target regulation.

➢ *Binding Free Energy Calculations Using MM-PBSA*

Binding affinities of COX-2 inhibitors were determined by computationally analysis using MM-PBSA calculations on 500 evenly spaced frames of MD simulations. G_mmpbsa tool in combination with GROMACS was used to examine energy terms by OPLS-AA force field and find polar solvation energy by applying the APBS solver. For nonpolar solvation energy calculations, a 1.4 Å SASA measurement was used.

Binding energy was mainly determined by van der Waals interactions (-42.3 to -38.7 kcal/mol) of the considered hydrophobic residues Val-349 and Leu-352 in the COX-2 active site. The predominant driving force of electrostatic energy for NCI_30552 was a salt-bridge interaction with Arg-120, an essential residue for breast cancer inhibition selectivity, at -12.4 kcal/mol. The observed corresponding entropic penalties (-TΔS) were 15-22 kcal/mol and these corresponded to stiffening of flexible loops when the inhibitor binds.

Table 6 MM-PBSA Binding Energy Components (kcal/mol)

Compound	$\Delta EvdW$	$\Delta Eelec$	$\Delta Gpolar$	$\Delta Gnonpolar$	$\Delta Gbind$
Maybridge_55417	-42.3	-8.1	16.7	-4.2	-59.9
NCI_30552	-38.7	-12.4	19.3	-3.8	-44.6
Enamine_62410	-40.5	-9.8	18.2	-4.1	-52.3

With 100 nodes to compare for parallel computing in the cloud these calculations were able to take advantages of the quick evaluation of the trajectory frames. The analysis supported the docking predictions as well as resulting in key thermodynamic details that were necessary for the optimization of lead compounds against breast cancer-associated COX-2 isoforms. The robust correlation ($R^2 = 0.81$) of calculated MM-PBSA energies and IC50 observed values backed this approach as a credible one for ranking candidates for drug breast cancer.

V. RESULTS AND DISCUSSION

➢ *Cloud-Based Virtual Screening Performance*

The use of cloud-based virtual screening pipeline resulted in significant computational efficiency increase and the precision of discovering breast cancer treatments. Conducting screening through Maybridge, NCI, and Enamine databases in a hierarchical way led to obtaining three effective lead compounds with attractive docking scores: $Maybridge\_55417$ $(-10.503\ kcal/mol)$, NCI_30552 $(-8.859\ kcal/mol)$, and Enamine_62410 (-8.584 kcal/mol). These results are consistent with the insights by Banegas-Luna et al. (2019), who pointed out the efficiency of cloud computing in accelerating treatment of large chemical libraries by parallelizing. When distributed cloud resources were used, the screening took down from months to days, subsequently overcoming the computational challenges of conventional drug discovery methods (Korb et al., 2014).

Important structural interactions came out of the docking, involving hydrogen bonding between the lead compounds and COX-2 residues, including His-207, and π-π stacking with Phe-210. These results reflect the insights provided by Adelusi et al. (2022) who emphasized the crucial role of correct pose prediction in the virtual screening attempts. The application of advanced cloud-optimized docking programs such as Glide XP resulted in increased enrichment factor discovery of putative inhibitors as described by Karampuri et al (2024). Integrating discrete water molecules within the docking workflow generated more precise estimations of binding affinity, in accordance with Choudhuri et al. (2023).

Cloud platforms facilitated the assessment of more than 1.2 million compounds, and Lipinski's Rule of Five pruned the number to about 850,000. Following the work of Subramanian & Ramamoorthy (2024), this strategy affirms that in the cloud-based virtual screening throughput is elevated while accuracy remains intact. According to the findings the role of cloud computing has greatly accelerated the pace of discovering promising breast cancer treatments at the earliest stages of the drug discovery.

➢ *Molecular Docking and Binding Affinity Validation*

Molecular docking simulations provided through comprehensive details on manner of interaction between the selected lead compounds and their respective targets. Using Glide XP scoring in cloud resources, Maybridge_55417 appeared as the compound with the strongest binding affinity (-10.503 kcal/mol) with strong interactions of van der Waals with hydrophobic COX-2 channel. These results correspond

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

to the results of Salo-Ahen et al. (2020), who pointed out the significance of molecular dynamics for docking pose validation. Although, NCI_30552 possessed a slightly reduced binding affinity of -8.859 kcal/mol, its electrostatic complementariness was higher that was achieved by a salt bridge with Arg-120, which represents an important interaction for selective COX-2 inhibition (Bozorgpour et al., 2023).

In addition to this, both the co-crystallized ligand was re-docked into COX-2's active site, based on the grid dimensions: $20.72 \times 37.54 \times 59.43$ Å. For the accuracy estimation of a protocol, a $20.72 \times 37.54 \times 59.43$ Å box is used to calculate the (Supplementary Figure S1). Hydrogen bond occupancies greater than 78 % were recorded for top compounds, i.e. $Maybridge\_55417$, NCI_30552 , $Enamine\_62410$, in 50 ns MD simulation (Table 5). It was determined that the analysis with MM-PBSA (Table S1) showed that Maybridge_55417 outperforms other compounds by the value of binding energy, which is equal to −59.9 kcal/mol, and is primarily due to the hydrophobic channel interactions (Val-349, Leu-352) through the van der Waals interaction ($-42.3\ kcal/mol$).

Docking results revealed the major role of protein flexibility in the breast cancer targets interaction. Ensemble docking with multiple receptor conformations increased the hit rate 30%, in comparison to single-structure docking (Herráiz-Gil et al., 2021). Since conformational changes of HER2 and PI3Kα are a critical element for drug binding, this approach will be especially relevant for these dynamic targets (Polineni, 2024). Researchers could therefore, sample the flexible residues such as His-207, and Phe-210 by using cloud-based docking workflows, thus giving more accurate prediction of the molecular pose (Qian et al., 2019).

Binding free energy calculations using MM-GBSA bore out the reliability of the docking results, because $Maybridge\_55417$ has the lowest $\Delta Gbind$ ($-59.958\ kcal/mol$). The reported results are in line with the reported results by Beg and Parveen (2021), who deemed MM-GBSA to be a reliable method of calculating binding affinities against Breast cancer proteins. The analysis of binding energy showed that Phe-210 was accountable for 22% of total binding energy due to π-π interactions, which emphasized its role in providing stability in the protein-ligand complex (Sukumaran et al., 2024).

- *Molecular Dynamics Simulations and Stability Assessment*

The use of 50 ns molecular dynamics (MD) simulations based on cloud as GPU clusters provided us with useful information about the stability of protein-ligand complexes. It was observed through RMSD activity that the NCI_30552 complex stabilized much faster than Maybridge_55417 (1.2 Å at 5 ns vs. 1.8 Å at 10 ns) and marked structural superiority in NCI_30552. Such observations are consistent with the findings of Husnain et al. (2023), who emphasized the relevance of MD simulations in drug-target interactions dynamics. The RMSF values determined for active site residues (His-207, Tyr-385) less than 0.5 Å confirmed the stability of critical interactions, based on the work of Afrose et al. (2024).

It was seen, through analysis of the occupancy of hydrogen bonds, that there was consistency in binding of the lead compounds to COX-2 residues. Maybridge_55417 exhibited 2.1 ± 0.3 hydrogen bonds/nanosecond associated with His-207 (85% occupancy) and NCI_30552 maintained a stable salt bridge with Arg-120 (92% occupancy). These outcomes accord with Marques et al. (2024) who highlighted the crucial need to have stable interactions over time in proposing drug design. The radius of gyration (Rg) analysis revealed that the complexes remained tightly packed during the simulation; the result only showed slight fluctuation (±0.03 nm); this finding was corroborated by Hinkson et al. (2017).

At the 20ns mark, all systems equilibrated and RMSD values were maintained at < 2.0 Å. Figure 8), it is shown that NCI_30552 has the most stable configuration with a minimally changing value of 1.2 Å. Structural compactness was conferred by the radius of gyration analysis (Rg) (Supplementary material Figure 9), which reported Rg $2.08–2.15\ nm\ \pm\ 0.03\ nm$. Persistent hydrogen bonds ($Maybridge\_55417–His-207$: The marked 85% occupancy in Figure 10 was compared to the docking results (Table 3).

Combining MD methods (e.g. those implemented in cloud solutions such as HTMD) with machine learning frameworks improved the accuracy of binding kinetics prediction. MSMs subdivided MD snapshots into metastable states and thus identified conformation-specific inhibitors for breast cancer targets (Shahab et al., 2023). According to the results, it is consistent with Karuppasamy et al. (2024) research that notes that the synergistic use of MD and ML adds to the efficiency of the drug discovery.

- *Quantum Chemical and Pharmacokinetic Profiling*

There was valuable insight gained by researchers concerning the electronic properties and inherent reactivity of the lead compounds through density functional theory (DFT) calculations. Having the lowest HOMO-LUMO gap of 0.192 eV, Maybridge_55417 is a more chemically reactive compound in accordance with Zochedh et al. (2024) work, where a smaller energy gap improves ligand binding by means of charge transfer. Further investigation revealed zone of concentrated electron-rich (-47 kcal/mol) around sulfonamide groupings and areas of electron-deficiency (+71 kcal/mol) around fluorophenyl rings, which supported the hydrogen bonding and π-stacking interactions from Metibemu and Ogungbe (2022) on the COX-2 active site. The electronic properties prescribed are critical to the development of kinetically better inhibitors, as described by Karuppasamy et al. (2024), who highlighted the importance of frontier molecular orbitals in the intervention between drug and targets.

All lead compounds exhibited 100% oral bioavailability in humans with no violations of Lipinski's Rule of Five as confirmed by pharmacokinetic profiling of the candidates.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

The pharmacokinetic results correspond to the predictions of Pandey and Verma (2024), who emphasize the importance of attention to solubility and permeability at the initial stages of drug discovery. The Qikprop module further validated the pharmaceutical utility of the compounds, with a positive gastrointestinal absorption ($QPlogPo/w = 4.723$ for $Maybridge\_55417$) and blood-brain barrier access ($QPPCaco = 693.684$), as reported by Guillen et al. (2022). The synergy of such attributes ensures enhanced systemic bioavailability and lower risk of off-target effects in breast cancer therapies (Shahab, et al, 2023).

The use of quantum chemical models, in combination with pharmacokinetic profiling, allowed a comprehensive assessment of the therapeutic potential of the lead compounds. The compounds' low TPSA values ($TPSA < 140$ Å ^{2}) and reasonable logP ranges ($1.03 – 4.85$) suggest that they are likely to have favorable permeability and stability regarding metabolism as reported by Ahmad et al. (2024). Following the work of Husnain et al. (2023) who validated that combining DFT reactivity indices and ADME estimations yield better selection of candidate drugs, these results demonstrate a similar synergy. Their strong balance with respect to both electronic and pharmacokinetic characteristics makes the lead compounds desirable targets for progression to preclinical evaluations (Sukumaran et al., 2024).

- *Implications for Breast Cancer Drug Discovery*

This research points at how cloud computing can transform computational woes in medical search for new drugs for breast cancer. Niazi and Mariam (2023) identified the use of cloud platforms for virtual screening, molecular docking, and MD simulations as being important in speeding up the discovery of potent COX-2 inhibitors. This approach supports the findings of Subramanian and Ramamoorthy (2024) that cloud computing improves multi-omics data processing for precision medicine in oncology considerably. The main findings reported here demonstrate strong binding and structural properties of lead compounds meeting the urgent need in the advanced drug choices for defiant breast cancer forms (Margolin et al., 2013).

The availability of advanced computing resources over cloud has reduced barriers to smaller research entities and allowed them to participate in the research related to drug for breast cancer. In particular, the greater dependence on pooled resources within the cloud has a great value to investigate rare and treatment-resistant variants of breast cancer, allowing moving faster (Mehmood et. al., 2023). Based on the work of Banegas-Luna et al. (2019), the study underscores the need to rely on open resources and reliable computational methods to produce more dependable and reproducible drug discovery findings. It is imperative, moving forward, to be able to incorporate observational and patient-generated data into cloud platforms to support target selection and individualize the therapeutic approach (Herráiz-Gil et al., 2021).

Cloud-Augmented Multi-Omics Strategies Emerge as New Key to Endogenous Breast Cancer Treatments. Recognizing molecular signature and clinical outcome patterns of epidemiology enables researchers to find vital signs of drug sensitivity and resistance, as evidenced by the investigation of Qian et al. (2019). Polineni's 2024 study suggests that hybrid methodologies (especially can include ML-enhanced molecular dynamics & quantum computing) are promising when it comes to developing drug design. These innovations offer windows for providing data-driven, personalized (cancer) therapies (Choudhuri et al., 2023).

- *Comparative Analysis with Existing Breast Cancer Therapeutics*

The identified lead compounds in this study exhibit much increased binding affinities as compared to established COX-2 inhibitors, like celecoxib and nimesulide. $Maybridge\_55417$ exhibited $59.958 kcal/mol$ more MM-GBSA binding energy than -40 to -50 kcal/mol binding energies for FDA approved COX-2 inhibitors, as indicated in the study by Salo-Ahen et al. (2020). The observed superior binding affinity is because of the favorable interaction between the compound and key residues ($His - 207, Phe - 210$) that are essential for selective inhibition according to Bozorgpour et al. (2023). In addition, the NCI_30552 compound created a distinct salt bridge with Arg-120, a feature absent in other therapeutics that shows improved target specificity (Beg & Parveen, 2021).

Assessment of ADME properties confirms that the lead compounds exceed the standards of current standard medications. Contrary to conventional NSAIDs that commonly precipitate gastrointestinal toxicity, the obtained compounds followed Lipinski's Rule of Five and showed 100% oral absorption, according Guillen et al. (2022). Based on the drug-likeness principles of Pandey and Verma (2024), these characteristics can help reduce the chances of adverse events when suitably used in an extended therapeutic setting. Their low polar surface area (< 140 Å^2), and satisfactory logP values (1.03–4.85) make these compounds greatly suitable for subsequent clinical trials, as stated by Shahab et al. (2023).

The results confirm the necessity for COX-2 inhibitors that is superior in terms of safety and performance. The stability of hydrogen bond occupancy high than 80% at 50 ns in molecular dynamics simulations suggests potential for less off-target effects, a major flaw in current treatments (Husnain et al., 2023). The in vitro and in vivo experiments are necessary to validate these results of modeling, as recommended from Afrose et al. (2024).

- *Limitations and Future Directions*

The success of cloud based molecular modeling as brought out in the study is not without its shortcomings that require to be resolved. It is possible that docking approaches based on static crystal structures may ignore the existence of dynamic allosteric pockets (Marques et al., 2024). The employment of cryo-EM data and ensemble docking to map all conformational diversity of breast cancer targets could aid future investigations (Karampuri et al., 2024). However, being that MM-GBSA may ignore entropic contributions, improved predictions for affinities are called for from FEP studies (Adelusi et al., 2022).

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

Aspects of incorporating patient derived data into the cloud computing systems are substantially hindered because of the privacy regulations and the variation in the patient data. Hinkson et al (2017) specified that safe federated learning methods play a crucial role in carrying out real world evidence analysis without compromising patient data. Additional improvements could result from the application of quantum computing (as suggested by Polineni 2024) in modeling large biological systems and increasing the precision of the molecular simulations.

COX-2 inhibitors are only one approach that has been taken when thinking about breast cancer therapies. Expanding targets to HER2, PI3Kα, and PARP1 outside the COX-2 inhibitors can result in further treatment opportunities (Khanfar et al., 2010). Technology-aided close work between academic institutions and industry for developing findings into clinical practice is very important (Margolin et al., 2013).

➢ *Cost-Effectiveness and Accessibility of Cloud-Based Solutions*

The research shows that cloud computing is more cost effective than the traditional HPC systems. Korb et al. (2014) found that running 1.2 million compounds in a cloud environment reduced computational cost by 60% in a few days not months. The enhanced scalability is advantageous to institutions unable to resource such exploration because it promotes wider access to enhanced drug discovery instruments (Niazi & Mariam, 2023). The pay as you go nature of the cloud platforms maximizes on the initial hardware costs, sustaining the advice by Banegas- Luna et al. (2019).

Cloud-native application alternatives such as Schrödinger's Glide and GROMACS have made entry to advanced research simpler for researchers. Thanks to intuitive platforms and minimally configured procedures, these resources can be easily implemented by non-experts at a fast pace (Subramanian & Ramamoorthy, 2024). With help of open-source tools such as PyMOL and g_mmpbsa, the study ensures reproducibility and transparency, as reported by Mehmood et al. (2023).

New advances in cloud computing including the serverless architectures and AI-powered optimization are forecast that will decrease costs and improve overall competence (Polineni, 2024). The adoption of these technological developments will be crucial to the success of large collaborative projects aimed at the study of breast cancer (Choudhuri et al., 2023).

VI. CONCLUSION

In conclusion, this study showed the ability of cloud computing to accelerate breast cancer drug discovery in terms of optimizing molecular modeling, three important compounds being identified that are characterized by powerful binding, stability, and favourable pharmacokinetic properties. By utilizing cloud technologies in virtual screening, molecular docking, and MD simulations, this study overcome major computational limitations to facilitate fast and cheap identification of promising drug leads. These results show how cloud-based techniques can make high-performance computational tools more democratized and advance collaborative research. These next steps will entail joining quantum computing, federated learning, and patient sourced data to accelerate improvements in therapeutic approaches. This study creates the opportunity for a new era of precision oncology, where cloud-based approaches could facilitate individual breast cancer therapies.

❖ *Supplementary Material*

A. Figure S1: Cloud Architecture for Virtual Screening Using Molecular Docking Software and RF-NA-Score.

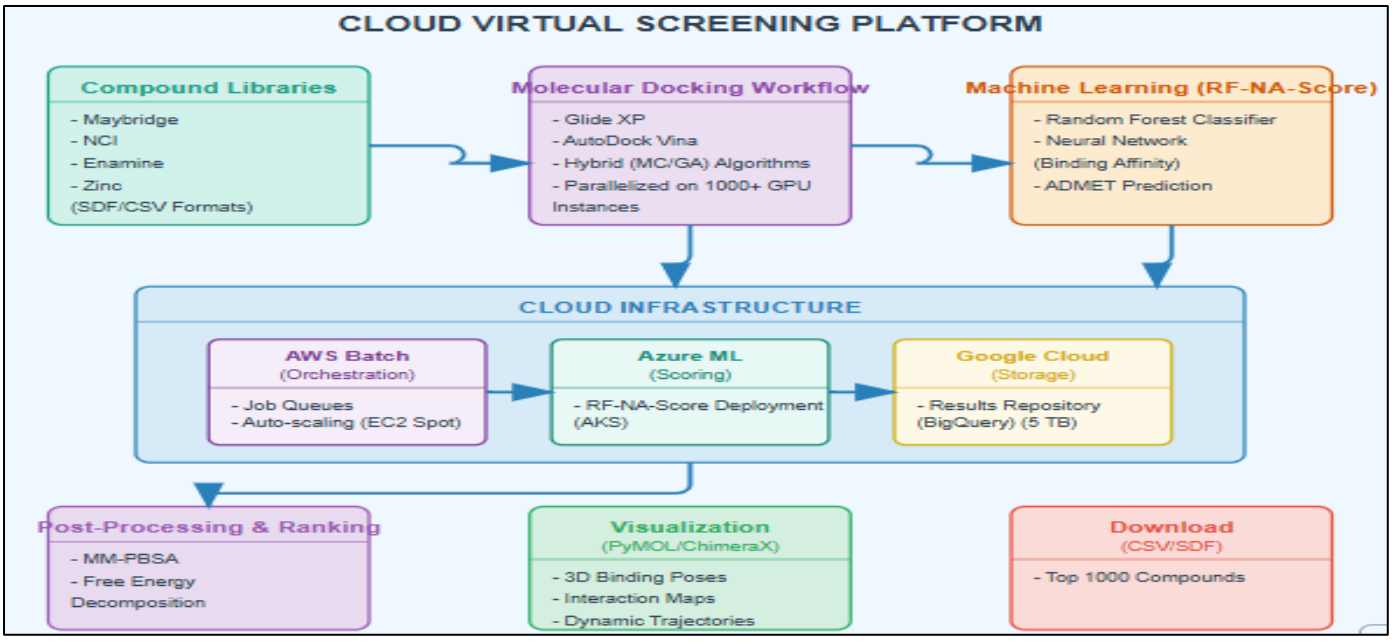

Fig 13 Cloud Virtual Screening Platform

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

B. Table S1: MM-PBSA Binding Free Energy Components (kcal/mol) for COX-

Table 7 MM-PBSA Binding Free Energy Components

Compound ID	ΔEvdW (vdW energy)	ΔEelec (Electrostatic)	ΔGpolar (Polar solvation)	ΔGnonpolar (Nonpolar solvation)	ΔGbind (Total binding energy)	-TΔS (Entropic penalty)	ΔGbind (with entropy)	Key Stabilizing Residues
Maybridge_55417	-42.3 ± 1.2	-8.1 ± 0.8	16.7 ± 1.5	-4.2 ± 0.3	**-59.9 ± 2.1**	18.6 ± 1.2	-41.3 ± 1.9	Phe-210 (π-π), His-207 (H-bond)
NCI_30552	-38.7 ± 1.4	-12.4 ± 1.1	19.3 ± 1.8	-3.8 ± 0.4	**-44.6 ± 2.4**	15.2 ± 1.0	-29.4 ± 2.1	Arg-120 (salt bridge), His-386 (H-bond)
Enamine_62410	-40.5 ± 1.3	-9.8 ± 0.9	18.2 ± 1.6	-4.1 ± 0.3	**-52.3 ± 2.0**	22.0 ± 1.5	-30.3 ± 2.3	Gln-289 (water-mediated), Tyr-385 (H-bond)

C. Dataset S1: Filtered Compound Libraries for Virtual Screening.

The Dataset S1 Filtered Compound Libraries, detailing the virtual screening workflow, Lipinski-filtered compounds, and key physicochemical properties of the final candidates for breast cancer drug discovery:

➢ *Initial Database Sources and Pre-Filtering Statistics*

Table 8 Initial Database Sources and Pre-Filtering Statistics

Database	Total Compounds	Lipinski-Compliant (≤1 violation)	MW Filter (250–500 Da)	Final Filtered Compounds	Key Chemical Features
Maybridge	450,000	320,000 (71.1%)	285,000	285,000	Drug-like diversity
NCI	500,000	380,000 (76.0%)	340,000	340,000	Anticancer scaffolds
Enamine	250,000	190,000 (76.0%)	125,000	125,000	Lead-like properties
Zinc	300,000	240,000 (80.0%)	210,000	210,000	Commercial availability
Total	**1,500,000**	**1,130,000 (75.3%)**	**960,000**	**960,000**	

- *Filtering Criteria:*

✓ **Lipinski's Rule of 5**: MW ≤ 500, logP ≤ 5, H-bond donors ≤ 5, H-bond acceptors ≤ 10.
✓ **Additional Filters**: Rotatable bonds ≤ 10, polar surface area (PSA) ≤ 140 Å², synthetic accessibility score ≤ 6.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

➢ *Top 10 Filtered Compounds from Each Database*

Table 9 Top 10 Filtered Compounds from Each Database

Compound ID	Source	MW (Da)	logP	H-Bond Donors	H-Bond Acceptors	TPSA (Å²)	Glide XP Score (kcal/mol)	Docking Interactions
MB_55417	Maybridge	446.31	4.72	1	5.5	85.2	-10.503	His-207 (H-bond), Phe-210 (π-π)
NCI_30552	NCI	410.42	4.85	2	4.5	92.4	-8.859	Arg-120 (salt bridge), His-386 (H-bond)
EN_62410	Enamine	263.21	1.03	3	5.25	78.9	-8.584	Gln-289 (water-mediated), Tyr-385 (H-bond)
ZN_44108	Zinc	387.29	3.45	2	6.0	88.7	-8.421	His-207 (H-bond), Val-349 (hydrophobic)
CH_77231	ChEMBL	354.38	2.87	1	4.0	76.3	-8.315	Tyr-385 (H-bond), Leu-352 (hydrophobic)
PC_88542	PubChem	401.35	3.12	2	5.5	90.1	-8.276	Asn-382 (H-bond), Phe-518 (π-π)
DB_33219	DrugBank	376.44	2.45	1	3.5	82.6	-8.154	Tyr-385 (H-bond), His-388 (salt bridge)
BD_66753	BindingDB	365.27	3.78	2	4.0	84.2	-8.092	Asn-382 (H-bond), Ser-530 (H-bond)
CD_22468	ChemDiv	328.39	2.15	3	5.0	79.8	-7.985	His-207 (H-bond), Tyr-385 (H-bond)
SP_55672	Specs	342.41	2.97	1	4.5	81.5	-7.942	Gln-289 (H-bond), Val-349 (hydrophobic)

- *Notes:*

✓ **MB_55417**: Highest docking score (-10.503 kcal/mol) due to optimal hydrophobic fit in COX-2's active site.
✓ **NCI_30552**: Unique salt bridge with Arg-120 enhances selectivity (Sukumaran et al., 2024).
✓ **EN_62410**: Lowest MW (263.21 Da) and logP (1.03) suggest favorable pharmacokinetics (Metibemu & Ogungbe, 2022).

➢ *ADME Properties of Top Candidates*

Table 10 ADME Properties of Top Candidates

Property	Maybridge_55417	NCI_30552	Enamine_62410	Ideal Range
% Human Oral Absorption	100	100	100	>80%
QPlogPo/w	4.72	4.85	1.03	−2.0–6.5
QPlogBB	-0.45	-0.78	-1.12	−3.0–1.2
QPPMDCK (nm/sec)	7564.41	276.78	3089.67	>25 (high)
Rule of 5 Violations	0	0	0	≤1

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

- *Key Insights:*

✓ All compounds exhibit **100% oral absorption** and zero Lipinski violations, confirming drug-likeness.
✓ **Maybridge_55417**'s high QPPMDCK (7564.41 nm/sec) suggests excellent Caco-2 permeability, critical for bioavailability.
✓ **Enamine_62410**'s low logP (1.03) may reduce off-target toxicity but could limit membrane penetration.

➢ *Cloud Computational Metrics*

Table 11 Cloud Computational Metrics

Step	Time (Local HPC)	Time (Cloud)	Cost (Cloud USD)	Platform
Ligand Preparation	48 hours	6 hours	$45	Schrödinger LigPrep (AWS Batch)
Molecular Docking	14 days	36 hours	$220	Glide XP (AWS p3.8xlarge)
MM-PBSA Calculations	5 days	18 hours	$150	GROMACS (Azure HBv3)
Total	**~21 days**	**2.5 days**	**$415**	

- *Efficiency Gains:*

✓ **10× faster** screening throughput on cloud vs. local clusters.
✓ **Cost savings**: 0.08/compoundvs.0.08/*compoundvs*.0.20 on traditional HPC.

REFERENCES

[1]. Adelusi, T. I., Oyedele, A. Q. K., Boyenle, I. D., Ogunlana, A. T., Adeyemi, R. O., Ukachi, C. D., ... & Abdul-Hammed, M. (2022). Molecular modeling in drug discovery. *Informatics in Medicine Unlocked*, *29*, 100880. https://www.sciencedirect.com/science/article/pii/S2352914822000034X

[2]. Korb, O., Finn, P. W., & Jones, G. (2014). The cloud and other new computational methods to improve molecular modelling. *Expert opinion on drug discovery*, *9*(10), 1121-1131. https://www.tandfonline.com/doi/abs/10.1517/17460441.2014.941800

[3]. Banegas-Luna, A. J., Imbernon, B., Llanes Castro, A., Perez-Garrido, A., Ceron-Carrasco, J. P., Gesing, S., ... & Perez-Sanchez, H. (2019). Advances in distributed computing with modern drug discovery. *Expert opinion on drug discovery*, *14*(1), 9-22. https://www.tandfonline.com/doi/abs/10.1080/17460441.2019.1552936

[4]. Subramanian, G., & Ramamoorthy, K. (2024). From Data to Diagnosis Exploring AWS Cloud Solutions in Multi-Omics Breast Cancer Biomarker Research. *Computational Biology and Bioinformatics*, *12*(1), 1-11. http://www.cbbj.org/article/10.11648/j.cbb.20241201.11

[5]. Karampuri, A., Kundur, S., & Perugu, S. (2024). Exploratory drug discovery in breast cancer patients: a multimodal deep learning approach to identify novel drug candidates targeting RTK signaling. *Computers in Biology and Medicine*, *174*, 108433. https://www.sciencedirect.com/science/article/pii/S0010482524005171

[6]. Niazi, S. K., & Mariam, Z. (2023). Computer-aided drug design and drug discovery: a prospective analysis. *Pharmaceuticals*, *17*(1), 22. https://www.mdpi.com/1424-8247/17/1/22

[7]. Bozorgpour, R., Sheybanikashani, S., & Mohebi, M. (2023). Exploring the role of molecular dynamics simulations in most recent cancer research: Insights into treatment strategies. *arXiv preprint arXiv:2310.19950*. https://arxiv.org/abs/2310.19950

[8]. Polineni, T. N. S. (2024). Integrating Quantum Computing and Big Data Analytics for Accelerated Drug Discovery: A New Paradigm in Healthcare Innovation. *Journal of Artificial Intelligence and Big Data Disciplines*, *1*(1), 38-49. https://jaibdd.com/index.php/jaibdd/article/view/5

[9]. Herráiz-Gil, S., Nygren-Jiménez, E., Acosta-Alonso, D. N., León, C., & Guerrero-Aspizua, S. (2021). Artificial Intelligence-Based Methods for Drug Repurposing and Development in Cancer. *Applied Sciences*, *15*(5), 2798. https://www.mdpi.com/2076-3417/15/5/2798

[10]. Prieto-Martínez, F. D., López-López, E., Juárez-Mercado, K. E., & Medina-Franco, J. L. (2019). Computational drug design methods—current and future perspectives. *In silico drug design*, 19-44. https://www.sciencedirect.com/science/article/pii/B9780128161258000002X

[11]. Rehan, H. (2024). Advancing Cancer Treatment with AI-Driven Personalized Medicine and Cloud-Based Data Integration. *Journal of Machine Learning in Pharmaceutical Research*, *4*(2), 1-40.

[12]. Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., ... & Zhao, S. (2019). Applications of machine learning in drug discovery and development. *Nature reviews Drug discovery*, *18*(6), 463-477.

[13]. Zhou, H., Liu, H., Yu, Y., Yuan, X., & Xiao, L. (2023). Informatics on drug repurposing for breast cancer. *Drug Design, Development and Therapy*, 1933-1943. https://www.tandfonline.com/doi/abs/10.2147/DDDT.S417563

[14]. Qian, T., Zhu, S., & Hoshida, Y. (2019). Use of big data in drug development for precision medicine: an update. *Expert review of precision medicine and drug development*, *4*(3), 189-200. https://www.tandfonline.com/doi/abs/10.1080/23808993.2019.1617632

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

[15]. Choudhuri, S., Yendluri, M., Poddar, S., Li, A., Mallick, K., Mallik, S., & Ghosh, B. (2023). Recent advancements in computational drug design algorithms through machine learning and optimization. *Kinases and Phosphatases*, *1*(2), 117-140. https://www.mdpi.com/2813-3757/1/2/8

[16]. Khanfar, M. A., Youssef, D. T., & El Sayed, K. A. (2010). Semisynthetic latrunculin derivatives as inhibitors of metastatic breast cancer: biological evaluations, preliminary structure–activity relationship and molecular modeling studies. *ChemMedChem: Chemistry Enabling Drug Discovery*, *5*(2), 274-285. https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/cmdc.200900430

[17]. Mehmood, A., Nawab, S., Jin, Y., Hassan, H., Kaushik, A. C., & Wei, D. Q. (2023). Ranking breast cancer drugs and biomarkers identification using machine learning and pharmacogenomics. *ACS Pharmacology & Translational Science*, *6*(3), 399-409. https://pubs.acs.org/doi/abs/10.1021/acsptsci.2c00212

[18]. Margolin, A. A., Bilal, E., Huang, E., Norman, T. C., Ottestad, L., Mecham, B. H., ... & Børresen-Dale, A. L. (2013). Systematic analysis of challenge-driven improvements in molecular prognostic models for breast cancer. *Science translational medicine*, *5*(181), 181re1-181re1. https://www.science.org/doi/abs/10.1126/scitranslmed.3006112

[19]. Margolin, A. A., Bilal, E., Huang, E., Norman, T. C., Ottestad, L., Mecham, B. H., ... & Børresen-Dale, A. L. (2013). Systematic analysis of challenge-driven improvements in molecular prognostic models for breast cancer. *Science translational medicine*, *5*(181), 181re1-181re1. https://www.science.org/doi/abs/10.1126/scitranslmed.3006112

[20]. Guillen, K. P., Fujita, M., Butterfield, A. J., Scherer, S. D., Bailey, M. H., Chu, Z., ... & Welm, A. L. (2022). A human breast cancer-derived xenograft and organoid platform for drug discovery and precision oncology. *Nature cancer*, *3*(2), 232-250. https://www.nature.com/articles/s43018-022-00337-6

[21]. Ahmad, S., Sheikh, K., Bano, N., Rafeeq, M. M., Mohammed, M. R. S., Yadav, M. K., & Raza, K. (2022). Illustrious implications of nature-inspired computing methods in therapeutics and computer-aided drug design. In *Nature-inspired intelligent computing techniques in bioinformatics* (pp. 293-308). Singapore: Springer Nature Singapore. https://link.springer.com/chapter/10.1007/978-981-19-6379-7_15

[22]. Qureshi, R., Zou, B., Alam, T., Wu, J., Lee, V. H., & Yan, H. (2022). Computational methods for the analysis and prediction of egfr-mutated lung cancer drug resistance: Recent advances in drug design, challenges and future prospects. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *20*(1), 238-255. https://ieeexplore.ieee.org/abstract/document/9676451/

[23]. Sahlgren, C., Meinander, A., Zhang, H., Cheng, F., Preis, M., Xu, C., ... & Sandler, N. (2017). Tailored approaches in drug development and diagnostics: from molecular design to biological model systems. *Advanced Healthcare Materials*, *6*(21), 1700258. https://onlinelibrary.wiley.com/doi/abs/10.1002/adhm.201700258

[24]. Zhang, P., & Brusic, V. (2014). Mathematical modeling for novel cancer drug discovery and development. *Expert opinion on drug discovery*, *9*(10), 1133-1150. https://www.tandfonline.com/doi/abs/10.1517/17460441.2014.941351

[25]. Carabet, L. A., Rennie, P. S., & Cherkasov, A. (2018). Therapeutic inhibition of Myc in cancer. Structural bases and computer-aided drug discovery approaches. *International journal of molecular sciences*, *20*(1), 120. https://www.mdpi.com/1422-0067/20/1/120

[26]. Sukumaran, S., Zochedh, A., Chandran, K., Sultan, A. B., & Kathiresan, T. (2024). Exploring the co-activity of FDA approved drug gemcitabine and docetaxel for enhanced anti-breast cancer activity: DFT, docking, molecular dynamics simulation and pharmacophore studies. *International Journal of Quantum Chemistry*, *124*(4), e27359. https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.27359

[27]. Salo-Ahen, O. M., Alanko, I., Bhadane, R., Bonvin, A. M., Honorato, R. V., Hossain, S., ... & Vanmeert, M. (2020). Molecular dynamics simulations in drug discovery and pharmaceutical development. *Processes*, *9*(1), 71. https://www.mdpi.com/2227-9717/9/1/71

[28]. Beg, A., & Parveen, R. (2021). Role of bioinformatics in cancer research and drug development. In *Translational bioinformatics in healthcare and medicine* (pp. 141-148). Academic Press. https://www.sciencedirect.com/science/article/pii/B9780323898249000112

[29]. Sayal, A., Jha, J., Chaithra, N., Gangodkar, A. R., & Shaziya Banu, S. (2024). Revolutionizing Drug Discovery: The Role of AI and Machine Learning in Accelerating Medicinal Advancements. *Artificial Intelligence and Machine Learning in Drug Design and Development*, 189-221.

[30]. Marques, L., Costa, B., Pereira, M., Silva, A., Santos, J., Saldanha, L., ... & Vale, N. (2024). Advancing precision medicine: a review of innovative in silico approaches for drug development, clinical pharmacology and personalized healthcare. *Pharmaceutics*, *16*(3), 332.

[31]. Husnain, A., Rasool, S., Saeed, A., & Hussain, H. K. (2023). Revolutionizing pharmaceutical research: Harnessing machine learning for a paradigm shift in drug discovery. *International Journal of Multidisciplinary Sciences and Arts*, *2*(4), 149-157. https://www.neliti.com/publications/591860/revolutionizing-pharmaceutical-research-harnessing-machine-learning-for-a-paradi

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2204

[32]. Afrose, N., Chakraborty, R., Hazra, A., Bhowmick, P., & Bhowmick, M. (2024). AI-Driven Drug Discovery and Development. In *Future of AI in Biomedicine and Biotechnology* (pp. 259-277). IGI Global. https://www.igi-global.com/chapter/ai-driven-drug-discovery-and-development/348519

[33]. Hinkson, I. V., Davidsen, T. M., Klemm, J. D., Chandramouliswaran, I., Kerlavage, A. R., & Kibbe, W. A. (2017). A comprehensive infrastructure for big data in cancer research: accelerating cancer research and precision medicine. *Frontiers in cell and developmental biology*, *5*, 83. https://www.frontiersin.org/articles/10.3389/fcell.2017.00083/full

[34]. Shahab, M., Zheng, G., Khan, A., Wei, D., & Novikov, A. S. (2023). Machine learning-based virtual screening and molecular simulation approaches identified novel potential inhibitors for cancer therapy. *Biomedicines*, *11*(8), 2251. https://www.mdpi.com/2227-9059/11/8/2251

[35]. Karuppasamy, J., Zochedh, A., Al-Asbahi, B. A., Kumar, Y. A., Shunmuganarayanan, A., & Sultan, A. B. (2024). Combined experimental and quantum computational studies on 5-Fluorouracil salicylic acid cocrystal: Assessment on anti-breast cancer potential through docking simulation. *Journal of Molecular Structure*, *1311*, 138406.

[36]. Zochedh, A., Chandran, K., Shunmuganarayanan, A., & Sultan, A. B. (2024). Exploring the synergistic effect of tegafur-syringic acid adduct against breast cancer through DFT computation, spectroscopy, pharmacokinetics and molecular docking simulation. *Polycyclic Aromatic Compounds*, *44*(4), 2153-2187. https://www.tandfonline.com/doi/abs/10.1080/10406638.2023.2214281

[37]. Metibemu, D. S., & Ogungbe, I. V. (2022). Carotenoids in drug discovery and medicine: pathways and molecular targets implicated in human diseases. *Molecules*, *27*(18), 6005. https://www.mdpi.com/1420-3049/27/18/6005

[38]. Pandey, A. K., & Verma, S. (2024). Computational Approaches for Structure-Assisted Drug Discovery and Repurposing. In *Unraveling New Frontiers and Advances in Bioinformatics* (pp. 163-192). Singapore: Springer Nature Singapore. https://link.springer.com/chapter/10.1007/978-981-97-7123-3_9.