Structure-Based Drug Discovery for Breast Cancer: A Review of Modern Computational Approaches and Future Perspectives

Panshul Purusharthy[1], Shreosee Ghosh[2], Rupak Roy[3]
[1] *Corresponding Author Milton Academy, Massachusetts, 170 Centre Street, Milton, Massachusetts 02186.*
[2,3]*SHRM Biotechnologies Pvt. Ltd., Kolkata -700155, India.*

***Abstract*—Breast cancer is a heterogeneous disease in which differences in molecular characteristics, treatment response, drug resistance, and recurrence continue to complicate therapeutic development. Computational biology has become an important component of modern drug discovery by enabling the prediction and analysis of molecular interactions before extensive laboratory testing. This review summarizes the role of structure-based drug discovery (SBDD) and related computational approaches in breast cancer research, with emphasis on molecular docking, virtual screening, molecular dynamics simulations, quantitative structure–activity relationship analysis, fragment-based drug discovery, drug-likeness assessment, and ADMET profiling. The review also discusses the increasing integration of artificial intelligence and machine learning with structural biology and cheminformatics for candidate identification, molecular optimization, diagnosis, prognosis, and personalized treatment. It also highlights the relationship between computational prediction and experimental validation, as computational models still struggle to reproduce the dynamic, heterogeneous nature of biological systems. Overall, integrating computational methods with structural biology, artificial intelligence, and laboratory validation may improve the efficiency, selectivity, and personalization of future breast cancer drug development.**

***Index Terms*—breast cancer; structure-based drug discovery; molecular docking; molecular dynamics; artificial intelligence; ADMET; personalized medicine; computational drug discovery**

I. INTRODUCTION

Breast cancer is the most common cancer diagnosis in women and is also a leading cause of cancer-related mortality. Current treatment approaches include surgery, chemotherapy, radiotherapy, hormone therapy, targeted therapy, and immunotherapy. However, drug resistance, tumor recurrence, metastasis, treatment toxicity, and biological heterogeneity continue to limit therapeutic effectiveness [(1)]. Because tumors can differ substantially in their genetic and molecular characteristics, generalized treatment strategies may not adequately address the molecular drivers of an individual tumor.

These challenges have increased interest in bioinformatics and computational methods that can characterize molecular interactions and prioritize potential therapeutic candidates before laboratory testing. Structure-based drug discovery (SBDD) uses three-dimensional protein structures to investigate how small molecules interact with biological targets. Molecular docking can predict the preferred orientation of a ligand within a binding site, whereas molecular dynamics simulations can examine the stability and movement of protein–ligand complexes over time.

Docking and molecular dynamics therefore serve complementary purposes. Docking provides a comparatively static prediction of ligand binding pose and interaction strength, while molecular dynamics considers the dynamic behavior of the molecular system. These methods can be incorporated into virtual screening to prioritize promising compounds. Fragment-based approaches, cheminformatics filters, ADMET profiling, and Lipinski's Rule of Five can further support the evaluation of drug-like properties before experimental studies.

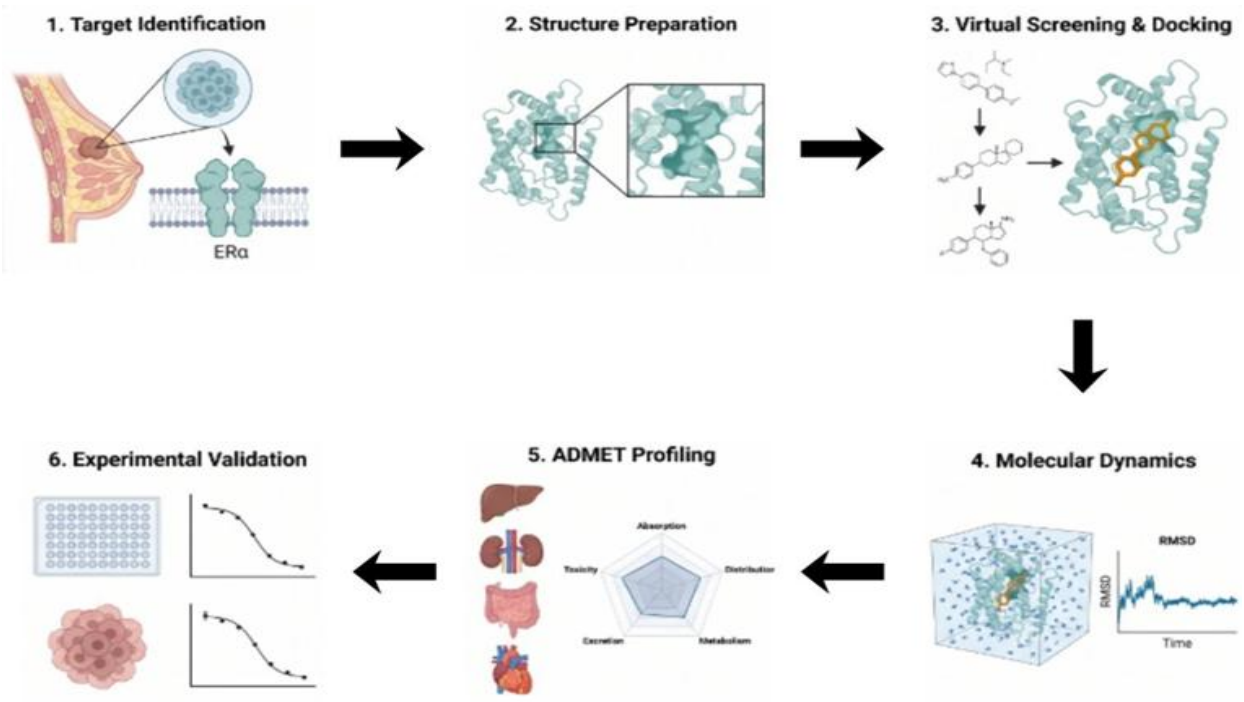


Figure 1: Overview of the computational drug-discovery workflow for breast cancer

Recent developments in artificial intelligence have further expanded computational drug discovery by enabling pattern recognition across large biological and chemical datasets. This review discusses the limitations of current breast cancer therapies, the molecular basis of personalized medicine, the role of SBDD and computational screening, emerging artificial intelligence applications, and the challenges that remain in applying computational biology to pharmaceutical development.

II. LIMITATIONS OF CURRENT BREAST CANCER TREATMENTS AND THE NEED FOR PERSONALIZED MEDICINE

Despite major advances in treatment, drug resistance, tumor recurrence, and metastasis remain important clinical challenges. Current treatment methods include surgery, radiotherapy, chemotherapy, hormone therapy, targeted therapy, and immunotherapy [(1)]. Surgery remains an important treatment when disease is localized and may involve breast-conserving surgery or mastectomy. Radiotherapy is commonly used after surgery to eliminate remaining cancer cells and reduce recurrence. Chemotherapy targets rapidly dividing cells, while hormone therapy can limit estrogen-mediated tumor growth in receptor-positive disease. Targeted therapies, including anti-HER2 agents, CDK4/6 inhibitors, and PARP inhibitors, focus on specific molecular pathways. Immunotherapy can enhance antitumor immune responses and has an important role in selected cases, including triple-negative breast cancer.

Breast cancer is highly heterogeneous, with tumors differing in genetic makeup, biological behavior, and response to treatment. Toss and Cristofanilli describe major molecular subtypes, including luminal A, luminal B, HER2-enriched, and triple-negative breast cancer, each with distinct molecular characteristics [(2)]. For example, luminal tumors may contain PIK3CA alterations, HER2-enriched tumors are characterized by HER2 overexpression, and triple-negative tumors may contain defects in DNA-repair pathways involving BRCA1 [(2)]. Identifying such molecular abnormalities supports the development of targeted therapies directed toward specific signaling pathways, including PI3K/AKT/mTOR [(2)].

The increasing use of molecular biomarkers and next-generation sequencing has therefore contributed to a shift from generalized treatment toward approaches tailored to individual tumor biology. This molecular heterogeneity also provides a rationale for computational strategies that can prioritize targets and drug candidates according to specific molecular

characteristics.

Table 1. Representative molecular characteristics of major breast cancer subtypes and their relevance to computational drug discovery.

Breast cancer subtype	Key molecular characteristic	Potential computational relevance
Luminal A	Hormone-receptor-associated ; may include PIK3CA alterations	Target identification and ligand screening against relevant signaling pathways
Luminal B	Hormone-receptor-associated with distinct proliferative features	Molecular profiling and target prioritization
HER2-enriched	HER2 overexpression	Structure-based screening of HER2-related therapeutic candidates
Triple-negative	Lacks ER, PR, and HER2; may involve DNA-repair alterations	Identification of alternative targets and computational screening for novel candidates

III. STRUCTURE-BASED DRUG DISCOVERY AND MOLECULAR MODELING

As knowledge of the molecular drivers of breast cancer increases, computational methods have become important for designing and prioritizing potential treatments. Batool, Ahmad, and Choi describe SBDD as an approach that uses three-dimensional protein structures to guide pharmaceutical development [3]. Traditional drug discovery may require physical screening of large compound collections, whereas SBDD can use molecular docking, virtual screening, and molecular dynamics to predict candidate interactions before laboratory testing.

By examining protein structures and binding sites, researchers can prioritize compounds with potentially favorable interactions while reducing the number of candidates requiring experimental evaluation. The growing use of artificial intelligence and machine learning may further improve screening by identifying patterns within large biological and chemical datasets [3].

Geromichalos further emphasizes the importance of molecular computer modeling in anticancer drug development [4]. Molecular docking can predict ligand–protein interactions and estimate binding strength, while virtual high-throughput screening can evaluate large numbers of compounds against a selected target. Quantitative structure–activity relationship (QSAR) analysis provides another computational strategy by relating molecular structure to biological activity [4]. Together, these methods can support the identification of compounds with desirable potency, selectivity, and toxicity profiles.

IV. PROTEIN–LIGAND INTERACTIONS AND COMPUTATIONAL DRUG SCREENING

A detailed understanding of protein–ligand interactions is essential for interpreting computational screening results. Du and colleagues discuss the molecular principles governing these interactions, including hydrogen bonding, electrostatic attractions, van der Waals interactions, and hydrophobic effects [5]. Gibbs free energy is also relevant to the stability of binding. The Lock-and-Key, Induced Fit, and Conformational Selection models provide conceptual frameworks for understanding how proteins and ligands achieve favorable interactions [5].

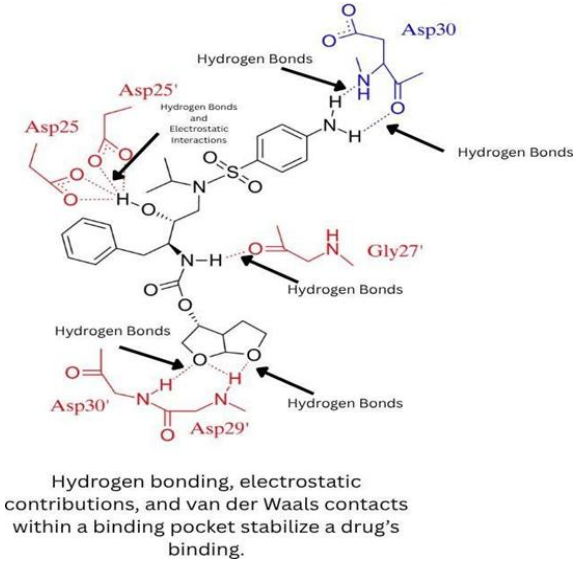


Figure 2: Schematic representation of protein–ligand recognition, including hydrogen bonding, electrostatic interactions, hydrophobic interactions, and van der Waals forces.

Arango and colleagues describe the use of molecular docking to identify favorable binding sites and calculate interaction scores between ligands and proteins [7]. Molecular dynamics simulations can subsequently be used to assess whether selected complexes remain stable under simulated biological conditions. ADMET profiling evaluates absorption, distribution, metabolism, excretion, and toxicity, helping to identify candidates with more suitable pharmacological properties [7].

Drug-likeness evaluation provides an additional filtering step. Walters and Murcko discuss Lipinski's Rule of Five, which considers molecular weight, lipophilicity, and hydrogen-bonding properties in relation to oral bioavailability [8]. Other computational filters, including REOS and RECAP, can support the removal of compounds with undesirable structural or chemical characteristics [8].

Table 2. Major computational approaches used in candidate screening and their principal applications and limitations

Method	Primary Purpose	Role in Drug Discovery	Key Limitations
Molecular docking	Predict ligand binding pose and score	Rapid prioritization of compounds	Often uses simplified/static representations
Virtual screening	Evaluate large compound libraries	Reduces number of candidates for testing	Prediction depends on library and scoring method
Molecular dynamics	Assess molecular movement	Refines docking predictions	Computationally intensive and model dependent
QSAR	Relate structure to biological activity	Predict activity of related compounds	Requires suitable training data
ADMET profiling	Predict pharmacokinetic/toxicity properties	Early drug-likeness filtering	In silico predictions require experimental confirmation

V. ADVANCED COMPUTATIONAL APPROACHES AND ARTIFICIAL INTELLIGENCE

Recent innovations have expanded computational drug discovery beyond conventional docking and screening. Chandraghatgi and colleagues describe a fragment-based drug discovery approach that uses computational optimization to generate candidate molecules [6]. Their FDSL-DD system identifies fragments that bind effectively to target proteins and uses genetic algorithms to combine and refine these fragments, with the aim of improving binding affinity and medicinal properties [6].

The integration of multiple computational approaches has further strengthened the drug-discovery workflow. Naithani and Guleria describe a framework combining structural biology, cheminformatics, artificial intelligence, and molecular simulations [9]. Databases such as the Protein Data Bank, PubChem, and ChEMBL can support target and compound identification, followed by structure-based and ligand-based drug design, molecular docking, molecular dynamics, and ADMET screening. Artificial intelligence and deep learning can then be used to identify relationships between molecular structure and biological activity.

Artificial intelligence also extends beyond drug discovery into breast cancer diagnosis, prognosis, and treatment planning. Feng, Yi, and Xu describe AI applications involving mammograms, MRI scans, pathology images, clinical records, and genomic data [11]. These applications illustrate how machine learning can contribute across multiple stages of breast cancer management, although clinical validation remains essential.

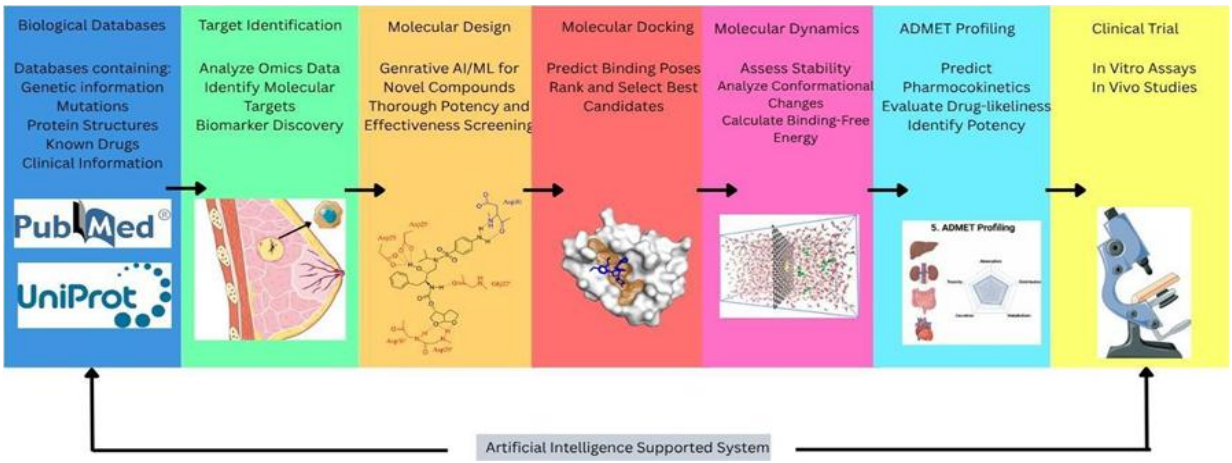


Figure 3: Integrated AI-assisted computational pipeline for breast cancer research, linking biological databases, target identification, molecular design, docking, molecular dynamics, ADMET prediction, and experimental/clinical validation.

VI. CHALLENGES AND FUTURE DIRECTIONS OF COMPUTATIONAL BIOLOGY

Despite substantial advances, computational biology continues to face important limitations. Dennis Bray argues that biological systems are highly dynamic and adaptive and therefore cannot be completely represented by current computational models [10]. Environmental changes, gene expression differences, and post-translational modifications can influence biological behavior in ways that remain difficult to reproduce computationally [10]. Similar difficulties occur in complex processes such as immune responses, tissue repair, and cancer progression.

These limitations are particularly relevant to cancer because tumor cells can adapt to changing environments and therapeutic pressure. Consequently, computational models may not fully capture the behavior of cancer cells under natural biological conditions. Artificial intelligence introduces additional challenges, including model transparency, reliability, validation, reproducibility, and regulation. Computational predictions should therefore be considered complementary to laboratory experimentation and clinical expertise rather than replacements for them.

Future progress will depend on integrating computational modeling with artificial intelligence, experimental validation, and patient-specific molecular information. Such integration may improve target selection, candidate prioritization, molecular optimization, and personalized therapeutic strategies.

Table 3. Major challenges in computational breast cancer research and potential directions for future development.

Challenge	Current limitation	Future direction
Biological complexity	Models may not reproduce dynamic cellular behavior	Multi-omics, structural and experimental data
Model reliability	Predictions depend on datasets and computational assumptions	Independent validation and standardized benchmarking
AI transparency	Some models provide limited interpretability	Develop interpretable and explainable models
Clinical translation	In silico findings may not directly translate to patients	Strengthen experimental and clinical validation
Tumor heterogeneity	A single model may not represent all tumor states	Use patient-specific and subtype-specific computational approaches

VII. CONCLUSION

Computational biology, structural modeling, and artificial intelligence are increasingly important in developing breast cancer therapies. Tumor heterogeneity, drug resistance, recurrence, metastasis, and treatment-related toxicity continue to limit conventional approaches and create a need for more

selective and personalized strategies. SBDD, molecular docking, molecular dynamics simulations, fragment-based drug discovery, QSAR, ADMET profiling, and drug-likeness assessment provide complementary tools for prioritizing candidate molecules and investigating their interactions with molecular targets.

At the same time, computational models remain constrained by the complexity and dynamic nature of biological systems. Artificial intelligence offers new opportunities for analyzing large datasets and supporting drug discovery, diagnosis, prognosis, and treatment planning, but its predictions require appropriate validation. Ultimately, the most promising approach is an integrated workflow in which computational predictions guide experimental investigation and patient-specific research. Such integration may contribute to the development of safer, more targeted, and more personalized therapies for breast cancer.

REFERENCES

[1]. Wang, J., & Wu, S. G. (2023). Breast Cancer: An Overview of Current Therapeutic Strategies, Challenges, and Perspectives. Dovepress, 721-730. https://doi.org/10.2147/BCTT.S432526

[2]. Toss, A., & Cristofanilli, M. (2015). Molecular characterization and targeted therapeutic approaches in breast cancer. Breast Cancer Research, 17(1). https://doi.org/10.1186/s13058-015-0560-9

[3]. Batool, M., Ahmad, B., & Choi, S. (2019). A Structure-Based Drug Discovery Paradigm. International Journal of Molecular Sciences, 20(11), 2783. https://doi.org/10.3390/ijms20112783

[4]. Geromichalos, G. (n.d.). Importance of molecular computer modeling in anticancer drug development.

[5]. Du, X., Li, Y., Xia, Y.-L., Ai, S.-M., Liang, J., Sang, P., Ji, X.-L., & Liu, S.-Q. (2016). Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods. International Journal of Molecular Sciences, 17(2), 144. https://doi.org/10.3390/ijms17020144

[6]. Chandraghatgi, R., Ji, H. F., Rosen, G. L., & Sokhansanj, B. A. (2024). Streamlining Computational Fragment-Based Drug Discovery through Evolutionary Optimization Informed by Ligand-Based Virtual Prescreening. Journal of chemical information and modeling, 64(9), 3826–3840. https://doi.org/10.1021/acs.jcim.4c00234

[7]. Arango, J. P. B., Rodriguez, D. Y. M., Cruz, S. L., & Ocampo, G. T. (2025). In silico evaluation of pharmacokinetic properties and molecular docking for the identification of potential anticancer compounds. Computational Biologyand Chemistry, 120, 108626. https://doi.org/10.1016/j.compbiolchem.2025.108626

[8]. Walters, W., & Murcko, M. (2002). Prediction of "drug-likeness." Advanced Drug Delivery Reviews, 54, 255–271.

[9]. Utkarsha Naithani, & Vandana Guleria. (2024). Integrative computational approaches for discovery and evaluation of lead compound for drug design. Frontiers in Drug Discovery, 4. https://doi.org/10.3389/fddsv.2024.1362456

[10]. Bray, D. (2015). Limits of computational biology. In Silico Biology, 12(1,2), 1–7. https://doi.org/10.3233/isb-140461

[11]. Feng, K., Yi, Z., & Xu, B. (2025). AI and breast cancer management: From data to the clinic. Cancer Innovation, 4(2). https://doi.org/10.1002/cai2.159