

Homologous Recombination Repair In Breast Cancer: Computational Strategies To Inhibit RAD51C Function

U S H Raghavendra Prasad¹, N Sravanthi, N Ravi², M Kavitha³

^{1,2,3}Department of Chemistry, University College of Science, Osmania University, Hyderabad-500007, Telangana, India.

Abstract

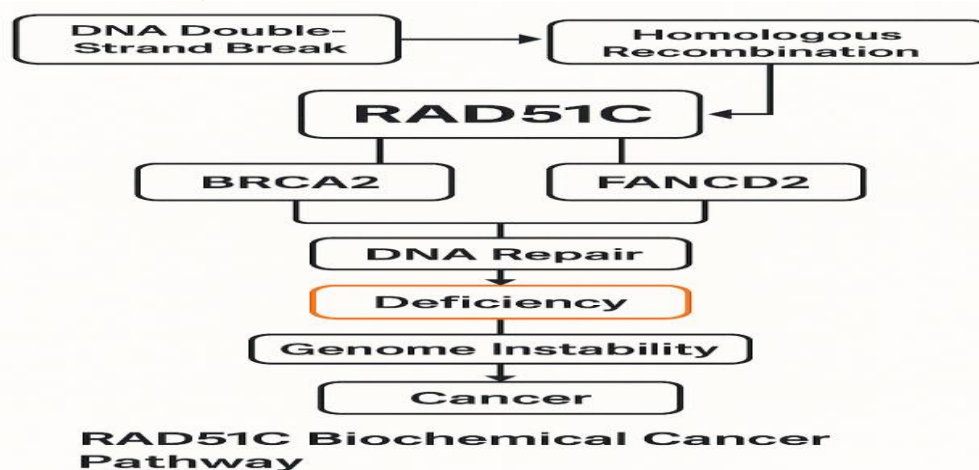
RAD51C is essential in the homologous recombination (HR) ^[1] pathway for repairing DNA double-strand breaks (DSBs) ^[2]. It coordinates with BRAC2 ^[3] and FANCD2 ^[3] to facilitate RAD51 loading and strand invasion during repair. Mutations in RAD51C ^[4] impair this process, leading to genomic instability and promoting tumorigenesis in tissues like breast and ovary. To identify potential Rad51C inhibitors, computational drug discovery approaches such as molecular docking, and virtual screening are employed. These techniques verifies 3D structure of RAD51C's to predict binding affinities of small molecules, helping to screen thousands of compounds efficiently. Coupled with in silico ADMET profiling, this pipeline accelerates the discovery of selective inhibitors that may enhance sensitivity to DNA-damaging therapies or PARP inhibitors in RAD51C deficient cancers.

Keywords: RAD51C, BRACA2, FANCD2, ADMET, PARP inhibitors.

INTRODUCTION

Breast Cancer is a malignant disease originating in breast tissue, most commonly in the ducts or lobules. It arises from uncontrolled cell division due to genetic mutations, hormonal imbalances, or environmental triggers. One major cause is impaired DNA repair mechanisms, especially homologous recombination (HR). The RAD protein family, particularly RAD51 and its paralogs ^{[5] [6] [7]} (including RAD51C, RAD51D, XRCC2, and XRCC3), plays a central role in HR by repairing DNA double-strand breaks. Defects or mutations in these genes compromise genomic stability, increasing susceptibility to breast cancer. RAD proteins are also potential therapeutic targets, especially in HR-deficient tumors responsive to PARP inhibitors. RAD51C one of the paralog of RAD51 family ^[8], interacts closely with other DNA Repair proteins, notably BRCA2, and FANCD2. BRCA2 helps load RAD51 onto the single stranded DNA, a critical step for Strand invasion during HR, FANCD2, part of the Fanconi anemia pathway, coordinates repair mechanisms at stalled replication forks and DNA interstrand crosslinks, often acting in concert with RAD51C. Deficiencies or mutations in RAD51C disrupts this repair cascade, leading to incomplete or erroneous repair of DSBs. Such defects can cause genomic instability-a hallmark of cancer-by allowing accumulation of mutations, chromosomal aberrations, and aneuploidy. Over time, this instability may lead to tumorigenesis, particularly in tissues with high rates of cell division, such as breast cancer ^{[9] [10]}.

Figure 1. Cancerous pathway of RAD51 C



2. METHODOLOGY

a) Retrieval of Protein biological information and 3D structure

The UniProt database is a vital tool in bioinformatics, offering an extensive and carefully curated collection of protein sequence data and related annotations. Widely used across diverse areas of biological research, UniProt^[11] compiles information from reputable sources such as Swiss-Prot, TrEMBL, and PIR. It provides insights into protein function, sequence characteristics, structural features, and taxonomic information. The Protein Data Bank (PDB)^[12] is an open-access online repository that stores three-dimensional structural data of biological macromolecules such as proteins, nucleic acids, and complexes. It provides experimentally determined structures submitted by researchers worldwide, primarily obtained through X-ray crystallography, NMR spectroscopy and cryo- electron microscopy. The PDB serves as a key resource for structural biology, drug design and computational modeling.

b) Validation

The 3D structure of RAD51C was then validated for its stereochemical properties through the PROCHECK^{[13][14]} program, accessible via the SAVES^{[15][16]} (Structural Analysis and Verification Server) platform. A Ramachandran plot^[17] was used to further assess the structural model by analyzing the distribution of backbone dihedral angles in relation to amino acid residues. Additional evaluations included verifying sequence-to-structure alignment and stability using energy-based scoring systems such as the Z-score from the ProSA tool^[18], which compares the model's quality to that of experimentally resolved protein structures.

c) Putative sites Prediction

Precisely identifying the active site of a protein is essential for understanding its function and is a key component in structure-based drug design. Computational methods are frequently used to forecast likely ligand-binding sites by analyzing the protein's 3D structure. Popular tools like CASTp^[19] and SiteMap^[20] included in the Schrödinger suite, are commonly applied to detect hydrophobic pockets and regions with geometries conducive to ligand binding.

d) Structure-based virtual screening using molecular docking

Molecular docking, a computational approach used to model interactions between small molecules and protein targets, is instrumental in the search for new drug candidates and the refinement of lead compounds. This technique is designed to identify novel chemical entities capable of binding effectively to specific proteins, thereby eliciting the intended therapeutic response. The success of virtual screening largely depends on a comprehensive understanding of the receptor's structural features and energetic landscape. Docking is a fundamental technique used to explore various ligand conformations and anticipate their interactions within protein binding pockets. It plays a pivotal role in structure-based drug design. Among the commonly employed tools for this purpose is GLIDE (Grid-Based Ligand Docking with Energetics)^[19], which efficiently forecasts ligand binding poses and approximates their binding strengths. This is achieved through a series of hierarchical filtering steps that examine the active site of the target protein. In this research, a structurally optimized form of the RAD51C protein underwent structure-based virtual screening utilizing GLIDE. The Vander Waals parameters were set with a scaling factor of 1.0 and a partial charge cutoff of 0.25 Å. A docking grid measuring X Å x Y Å x Z Å was created to cover the target binding site. The ligands selected from the Comprehensive Marine Natural Products Database (CMNPD)^{[21][23]} libraries-for the docking analysis. Ligand structures were prepared in three dimensions at a physiological pH of 7.0±2.0 using the LigPrep module within Maestro (Schrodinger, LLC, New York), employing the OPLS_2004 force field. Tautomer's, ionization variants, and stereoisomers were generated using default settings, ensuring low-energy conformations. After ligand preparation, compounds exhibiting the most favorable energy characteristics were subjected to flexible docking against the predicted active site of the RAD51C protein. This was carried out using a multi-stage process in the GLIDE module. The initial phase employed High Throughput Virtual Screening (HTVS) mode to filter candidates, followed by further evaluation of the top 10% of hits using Standard Precision (SP) mode. The best-performing ligands were then refined using Extra Precision (XP) mode to obtain the most reliable binding conformations. Post-docking, these top poses were adjusted to optimize bond geometries and were rescored using the Glide Score function. The most promising candidates were

subsequently assessed for their pharmacokinetic properties, focusing on ADME^{[24] [25]} (Absorption, Distribution, metabolism, and Excretion) parameters.

e) ADME

Analyzing the computational ADME (Absorption, Distribution, Metabolism, and Excretion) properties of ligand molecules provides essential information about their suitability as potential drug candidates. This evaluation is a key step in the drug discovery process, helping to improve the chances of success in clinical trials by eliminating compounds with unfavorable pharmacokinetic profiles at an early stage.

Ligands showing strong Glide scores and energy values were further analyzed for their pharmacokinetic and physicochemical profiles using QikProp^[26] Schrödinger suite module and toxic profiling using the Pro Tox 3.0^{[27] [28]} online platform. By combining these ADMET properties predictions with molecular docking results, and detailed structural examination, a selection of top candidates emerged as potential lead compounds for targeting the RAD51C protein.

3. RESULTS AND DISCUSSION

Protein structure retrieval, Structure analysis and validation of protein

a) Acquisition of protein structure

The three-dimensional crystal structure of RAD51C was obtained from the Protein Data Bank (PDB) using the PDB ID: 8FAZ_C. Selection criteria included high - resolution quality of the protein 2.30Å (em), structural completeness, and relevance for molecular docking applications. Before proceeding with docking simulations, the structure was refined by removing water molecules, non-essential chains, and heteroatoms using Schrodinger software. Polar hydrogens were added and Kollman charges were assigned as part of the preparation process.

b) Model verification of RAD51C

The Ramachandran plot, generated using the PROCHECK^{[29] [30]} server, evaluates the distribution of ϕ (phi) and ψ (psi) backbone torsion angles within the protein model. As shown in Figure 314 residues (91.7%) are located within the most favored regions, 23 residues (8.3%) fall into additionally allowed regions. Notably no residues were found in the disallowed regions. For a high-quality protein model, it is generally expected that more than 90% of residues should reside within the most favored region.

Figure 3.1. The Ramachandran Plot

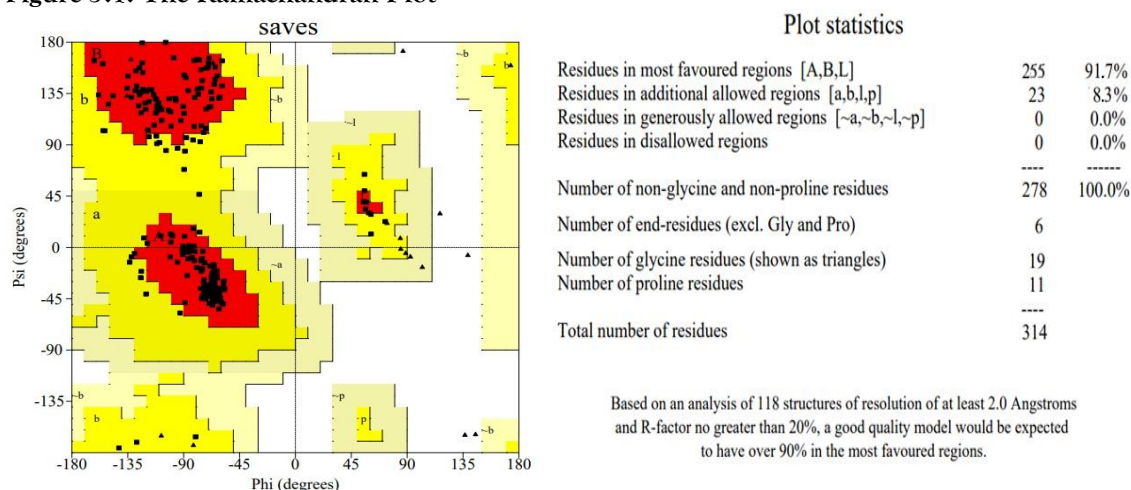


Figure presents the Ramachandran plot for the RAD51C protein, illustrating the distribution of backbone dihedral angles for its residues. The plotted residues (black dots) are mapped across regions of varying energetic favorability, which are color-coded for clarity: red indicated the most favored regions, yellow denotes additionally allowed regions, and light yellow represents generously allowed conformational spaces.

Figure 3.2. ProSA plot of the RAD51C protein

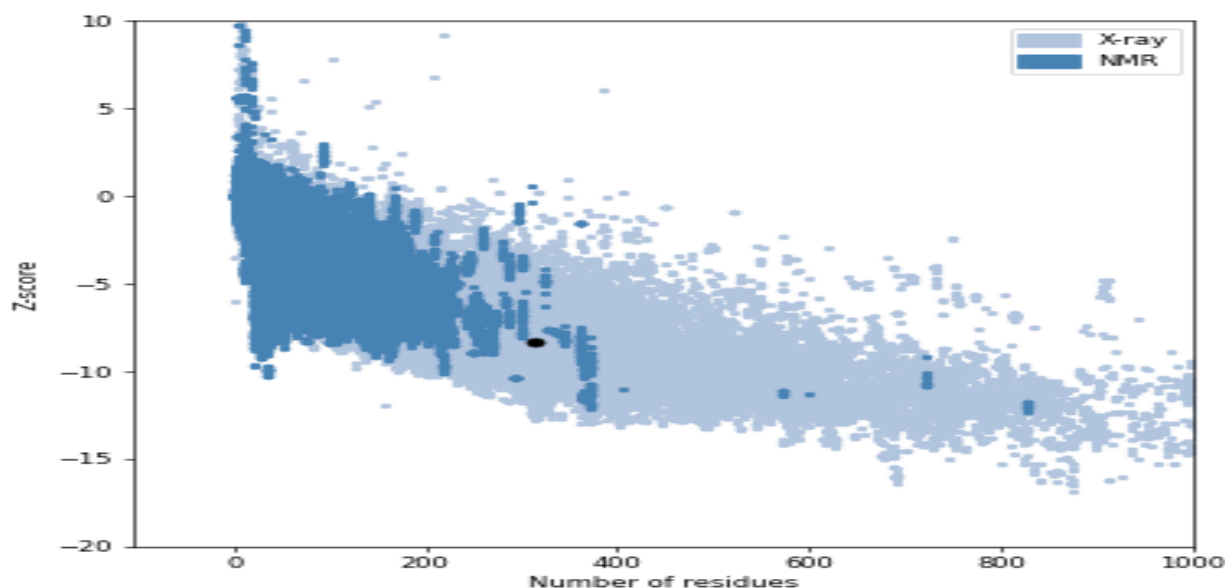


Figure illustrates the distribution of Z-scores for all protein structures available in the Protein Data Bank (PDB) as a function of their amino acid chain lengths. The Z-score^[31] of the RAD51C protein model, represented by a black dot, is compared against the Z-scores of experimentally resolved structures obtained via X-ray crystallography and NMR spectroscopy (depicted in light and dark blue regions, respectively). The RAD51C model exhibits a Z-score of -8.36 (Figure 3.2), which falls within the acceptable range, indicating a high-quality and reliable overall three-dimensional structure. Additionally, the ProSA energy profile^[32] (Figure3.3) presents the local energy distribution across the protein sequence using two sliding window sizes -10 and 40 amino acids – providing insight into the region-specific quality of the model.

Figure 3.3. Local model quality of the RAD51C protein

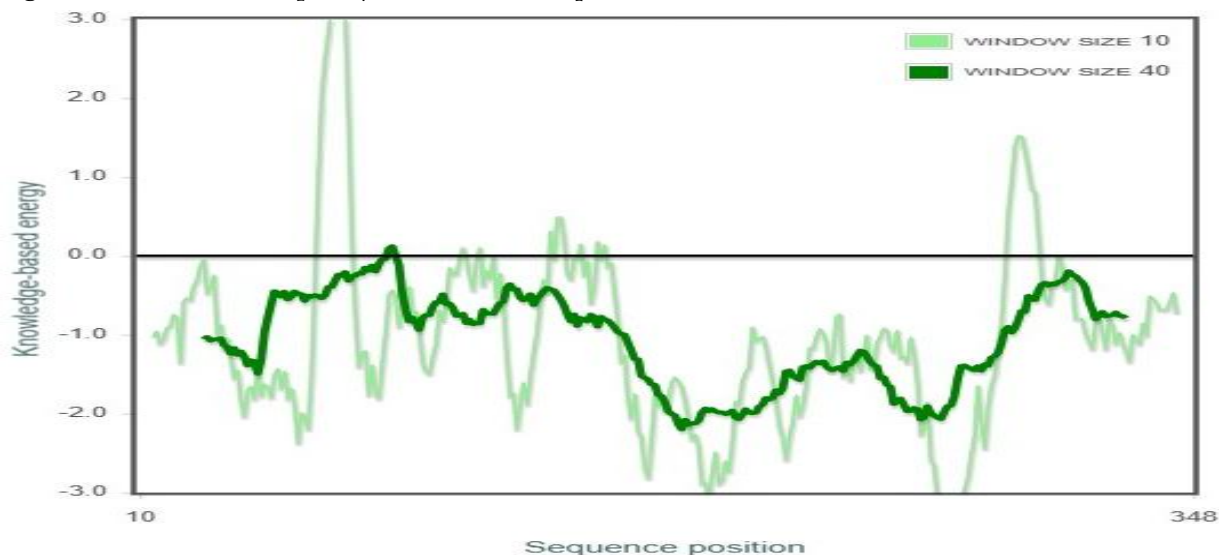


Figure presents the knowledge-based energy profile for the amino acid residues of the RAD51C protein, evaluated using sliding window sizes of 10 residues (light green) and 40 residues (dark green). The majority of the energy values fall below the baseline, suggesting favorable local structural quality and stability throughout most regions of the protein model.

4. ANALYSIS OF STRUCTURE OF RAD51C PROTEIN

Figure 4.1 showing three dimensional structure of RAD51C protein and visualized using Accelrys discovery studio.

Figure 4.1. The 3D- structure of the RAD51C protein

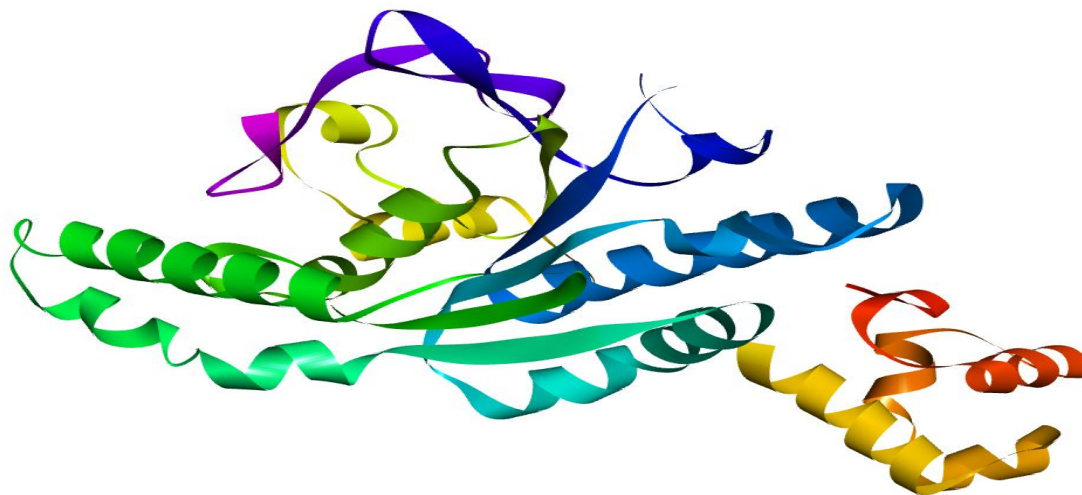


Figure 4.2. The PDBsum Server Predicted Secondary Structure

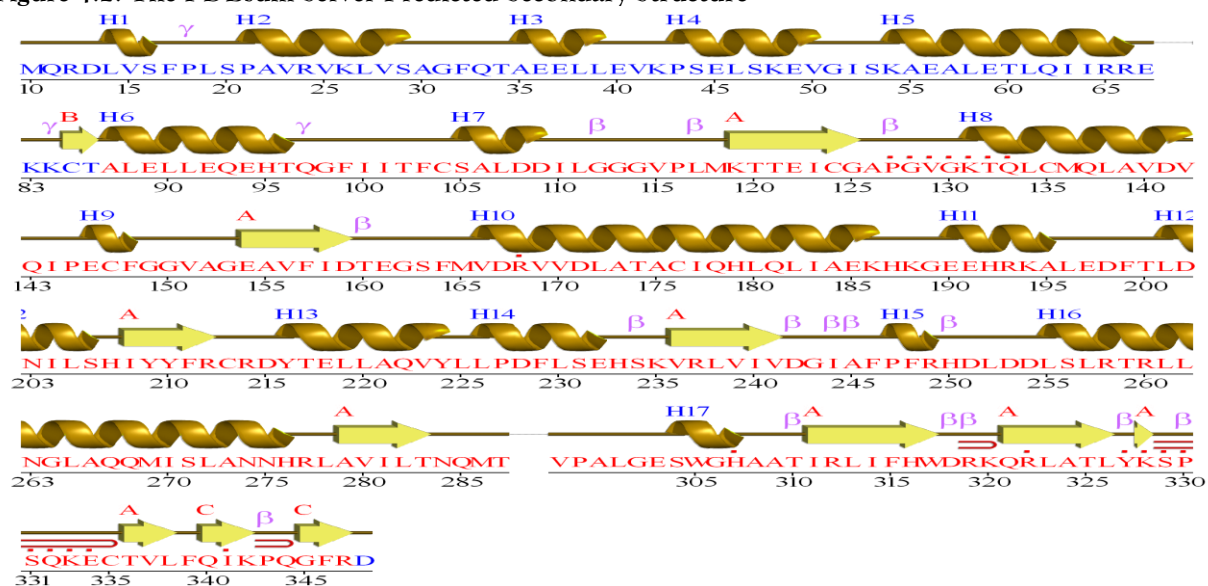


Figure 4.2 showing the secondary structure architecture of the RAD51C protein, as generated using the PDB-sum server^[33]. In the illustration, α -helices are represented in dark yellow, β -sheets are indicated by yellow arrows and details are shown in table 4.1.

4.1 Table Structural analysis of RAD51C Protein

S NO	Type of Secondary structure	Amino acids From to To
1	α -helices	13 to 16 21 to 29 35 to 39 43 to 50 54 to 66 87 to 96 105 to 109

		131 to 142 146 to 148 166 to 186 190 to 195 201 to 206 216 to 224 226 to 232 247 to 249 255 to 276 304 to 307
2	β -sheets β – hair pin (one)	85 to 86 119 to 125 154 to 159 208 to 212 236 to 241 279 to 283 311 to 317 321 to 326 336 to 338

5. PREDICTION OF ACTIVE SITE

Active site identification involves pinpointing the specific regions within a protein-dimensional structure that are responsible for its biological activity, typically where substrate binding and catalysis occur. Computational tools, such as CASTp, SiteMap, are often employed to detect surface cavities, conserved residues, and geometric features indicative of functional sites. Accurate identification of these regions is essential for structure-based drug design, ligand docking, and functional annotation of proteins.

Table 5.1. Active site identification

S. NO.	Active site identification server / tool	Amino acids From to To	Volume (Å)
1	CASTp	15,17,18,19,20,21 24,59,62,63,66, 67,83,84,223, 224,226,227, 228,231,260, 261,263,264, 267,268,269, 271,272,275, 276	2614.844
2	SiteMap	97,143,145,146, 147,151,152, 153,154,195, 198,199,203, 207,235,237	130.340

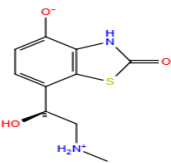
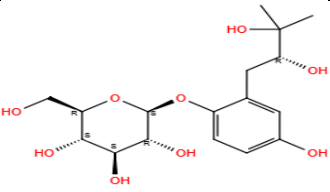
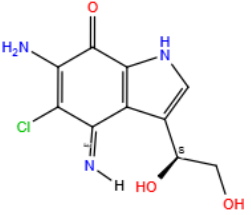
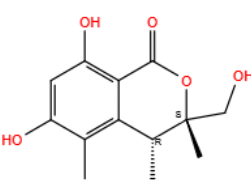
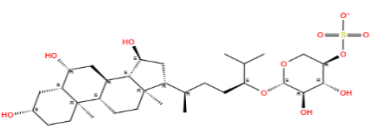
Table 5.1 presents identified active site regions from CASTp and SiteMap which are same as identified from experimentally and from literature.

6. Structure-Based Virtual screening of ligands targeting the RAD51C protein

A structure-based virtual screening (SBVS) strategy was employed to identify ligand candidates the RAD51C protein. A receptor grid was generated at the predicted active site of the RAD51C structure,

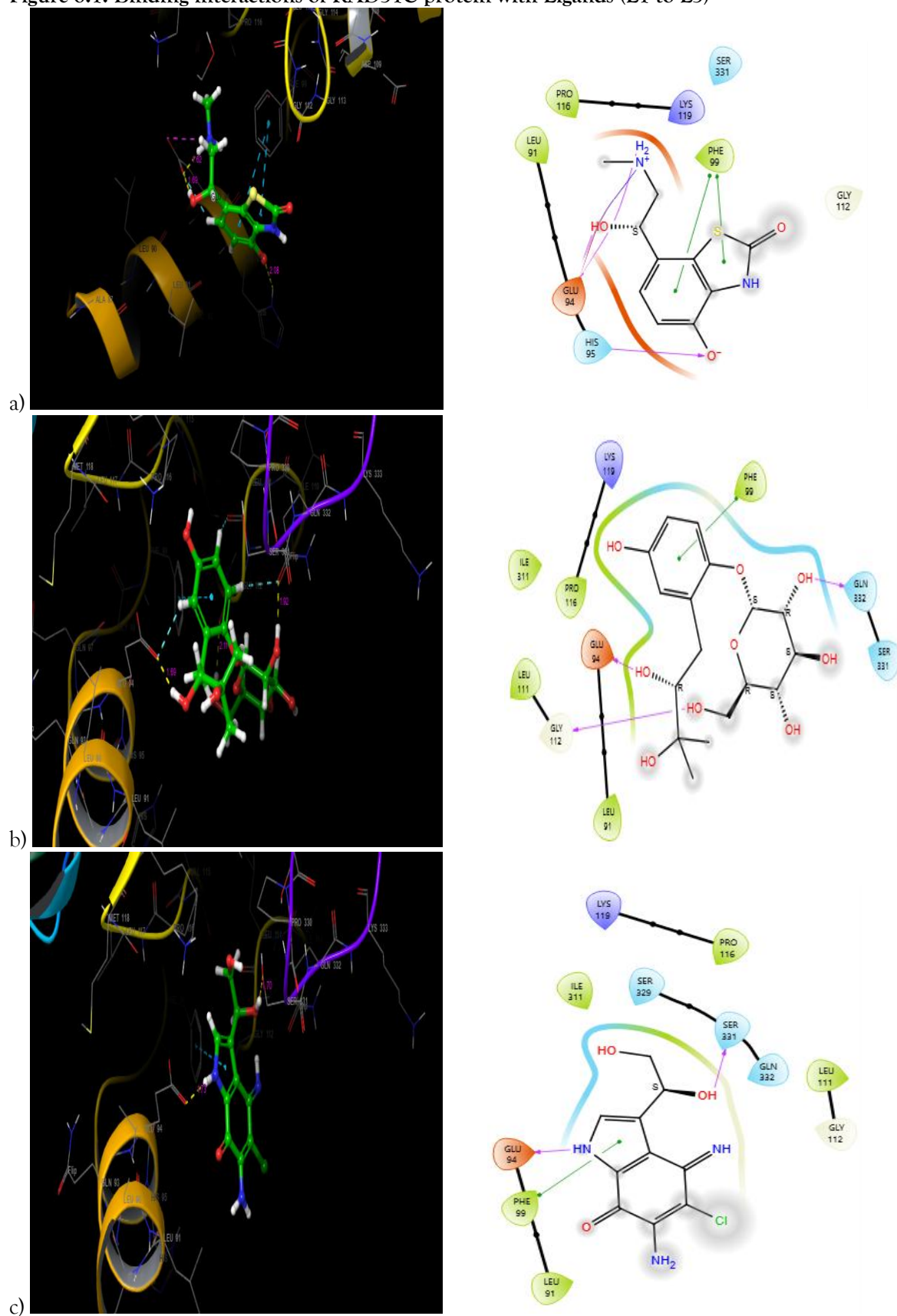
with dimensions set at $60 \text{ \AA} \times 60 \text{ \AA} \times 60 \text{ \AA}$ to define the docking space. Ligands were prepared using the LigPrep module of the Schrodinger suite, generating multiple ionization states per molecule to ensure biological relevance. A total of 30,000 compounds from the CMNPD (Comprehensive Marine Natural Products Database) were processed, yielding 47,000 structurally distinct ligand conformers. The prepared ligands underwent hierarchical virtual screening using the Glide module, which incorporates three filtering stages: High throughput Virtual Screening (HTVS), Standard Precision (SP), and Extra Precision (XP). At each level, the top 10% of ligands, based on Glide Score ranking, were selected for further screening. This process resulted in 45 final ligand-protein complexes.

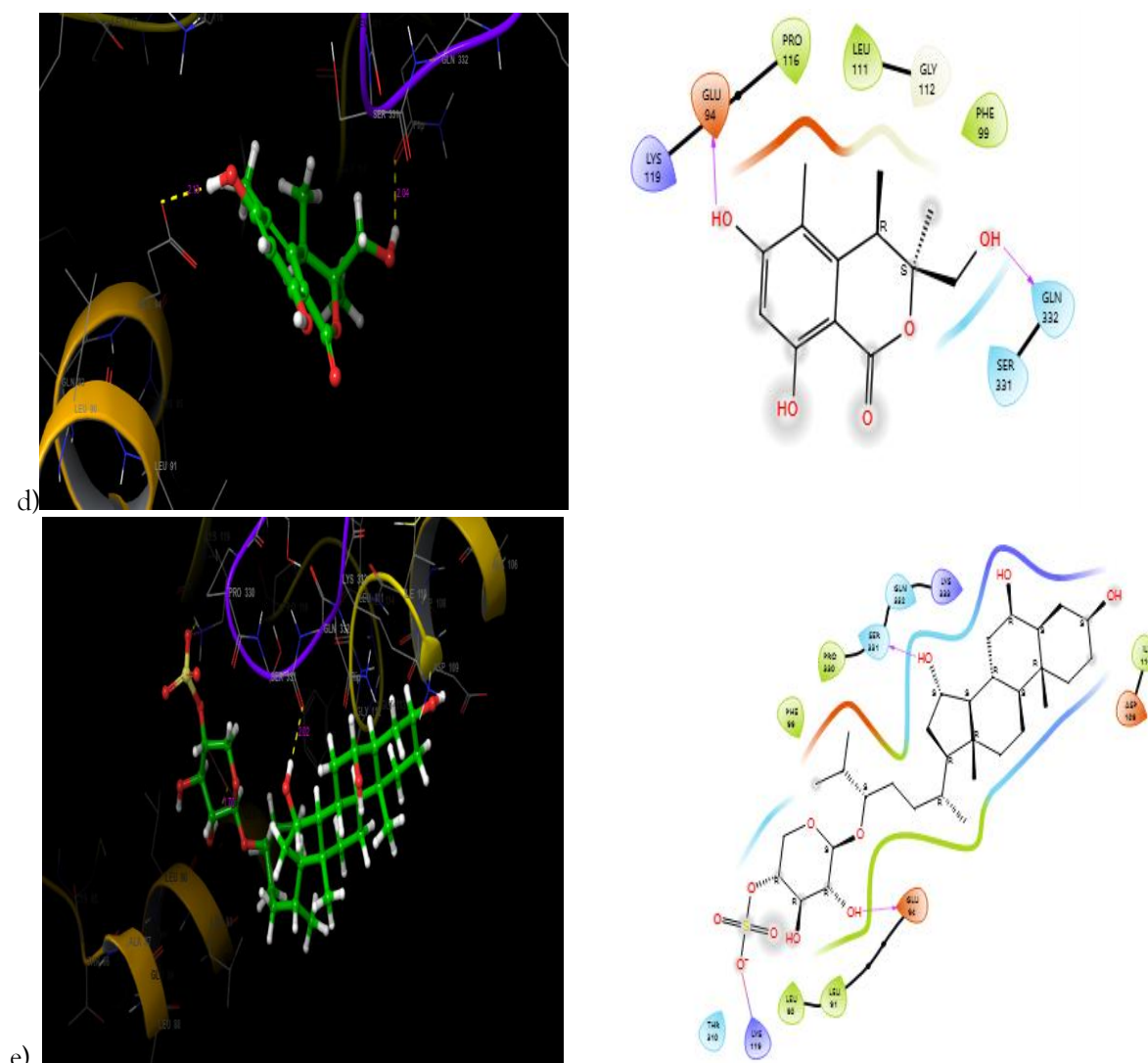
Table 6.1. Binding interactions

Ligand	Ligand structure	Glide Score	Glide Energy Kcal/mol	Hydrogen Bond	Hydrogen bond distance
L1		-7.099	-54.995	L1-HIS95 L1-GLU94 L1-GLU94	2.08 1.96 1.62
L2		-6.995	-50.258	L2-GLU94 L2-GLY112 L2- GLY332	1.69 2.11 1.92
L3		-6.830	-43.229	L3-GLU94 L3- SER331	1.73 1.70
L4		-6.710	-39.420	L4-GLU94 L4-GLN332	2.13 2.04
L5		-6.707	-38.313	L5- SER331 L5-GLU94	2.02 1.70

A representative selection of 5 high-scoring docked ligands is presented in Table 6.1. The binding interactions were analyzed, and the results indicate strong affinity of these ligands for the RAD51C protein's active site (Figure 6.1). Hydrogen bonding patterns were visualized using Discovery Studio Visualizer^{[34][35]}. All observed hydrogen bonds exhibited interatomic distance under $2.5 \text{ \AA}$, suggesting favorable and specific interactions.

Figure 6.1. Binding interactions of RAD51C protein with Ligands (L1 to L5)





7. ADME OUTLINE

a) Physicochemical Characteristics

The drug-likeness and pharmacokinetic profiles of the highest-ranking compounds were analyzed using QikProp (Schrödinger Suite) ^[36] ^[37]. Each molecule exhibited acceptable physicochemical parameters, including molecular weights not exceeding 374.387 (Table 7.1), hydrogen bond donor's ≤ 5.5 , and acceptors ≤ 12.45 , indicating favorable structural and chemical characteristics (refer to Table 7.2)

b) Pharmacokinetic Assessment

Predicted human oral absorption (HOA) values ranges from 28.138 to 77.567, highlighting strong oral bioavailability. Water solubility values (QPlogS) were maintained between -0.766 to -3.431, falling within the desired threshold. The compounds also demonstrated efficient intestinal permeability, with Caco-2 cell permeability values (QPPCaco) between 14.532 to 186.196. Protein binding predictions (QPlogKhsa) varied from -0.356 to -0.999, while blood penetration (QPlogBB) values ranged from -0.957 to -3.052, suggesting limited central nervous system (CNS) exposure and reduced neurotoxicity risk. CNS activity scores were consistently negative, reinforcing the low potential for neurological side effects. Additionally, predicted hERG inhibition scores ranged from -3.31 to -5.874, supporting a favorable cardiac safety profile (Table 7.2).

c) Evaluation of Drug-likeness

All molecules satisfied both Lipinski's Rule of Five ^[38] and Jorgensen's Rule of Three ^[39], which are established guidelines for assessing oral drug-likeness. The calculated lipophilicity values (QPlogPo/w) fell within the acceptable range of -1.136 to 1.725, (Table 7.2) further supporting the potential of these compounds as viable drug candidates.

Table 7.1. Physicochemical, Pharmacokinetic, Drug Likeness Properties

Ligand Number	Physicochemical Properties				Pharmacokinetic Properties							Drug Likeness Property		
	mol_MW	donorHB	accptHB	QPlogS	HDA%	QPPCaco	QPlogKhsa	QPlogPw	QPlogBB	CNS	QPlogHERG	RuleOfFive	RuleOfThree	QPlogPo/w
L1	240.276	4	6.45	-0.766	51.019	34.273	-0.659	13.856	-0.957	-1	-4.651	0	0	-0.58
L2	374.387	7	12.45	-1.876	28.138	14.532	-0.999	23.53	-3.052	-2	-4.582	1	2	-1.136
L3	255.66	5.5	7.4	-1.632	40.619	57.021	-0.808	16.853	-1.609	-2	-3.522	1	0	-0.819
L4	292.293	2	6	-3.431	77.567	183.666	-0.185	12.525	-1.463	-2	-5.874	0	0	1.725
L5	252.266	2	5.2	-2.191	71.942	186.196	-0.356	9.602	-1.152	-2	-3.31	0	0	0.746

Table 7.2. ADME considerations

S. No.	Descriptor	ADME property	permissible ranges or recommended value
1	CNS	Predicted central nervous system activity on -2 to +2 scale	-2 (inactive) to +2 (active)
2	mol_MW	Molecular weight of the molecule	130 to 725
3	DHB	Estimated number of hydrogen bonds that would be donated by the solute to water molecules in an aqueous solution	0 to 6
4	AHB	Estimated number of hydrogen bonds that would be accepted by the solute from water molecules in an aqueous solution	2 to 20
5	QPPCaco	Predicted apparent Caco-2 cell permeability in nm/sec	<25 poor, >500 great
6	QPlogPw	Predicted water/gas partition coefficient.	4.0 – 45.0
7	QPlogPo/w	Predicted octanol/water partition coefficient	-2.0 – 6.5
8	QPlogS	Predicted aqueous solubility, log S, S in mol dm ⁻³	-6.5 – 0.5
9	QPlogKhsa	Prediction of binding to human serum albumin	-1.5 – 1.5
10	QPlogHERG	Predicted IC ₅₀ value for blockage of HERG K ⁺ channels	Below +5.0
11	QPlogBB	Predicted blood / partition coefficient	-3.0 – 1.2
12	% human oral absorption	Predicted human oral absorption on 0 to 100% scale	>80% is high <25% is poor
13	Rule Of Five	Number of violations of Lipinski's rule of five	maximum is 4
14	Rule Of Three	Number of violations of Jorgensen's rule of three	maximum is 3
15	Synthetic feasibility	Predicted synthetic feasibility: on scale of 1 to 10;	0 = high feasibility 10 = least feasible
16	Lipophilicity	Predicted lyophilic nature of the ligand calculated from pIC ₅₀ -LogP	min -6; max +3

d) Toxicity

To evaluate potential metabolic liabilities, the selected compounds were assessed for interactions with Cytochrome P450 enzymes using The ProTox 3.0^[40] platform (Table 7.3). The analysis revealed that several ligands exhibited either inhibitory or non-inhibitory effects on major CYP450 isoforms—an important factor in predicting drug-drug interactions and metabolic stability. Additionally, the compounds demonstrated no significant activity related to hepatotoxicity or cardiotoxicity. Collectively, these findings indicate that the identified compounds possess favorable drug-like characteristics with minimal toxicity concerns, reinforcing their potential as therapeutic candidates for the treatment of Breast cancer.

Figure 7.3 Hepatotoxicity, Cardiotoxicity and Cytotoxicity profile

S.NO.	LIGANDS	HEPATOTOXICITY	CARDIOTOXICITY	CYP-1A2	CYP-2C19	CYP-2C9	CYP-2D6	CYP-3A4
1	L1	INACTIVE	INACTIVE	INACTIVE	INACTIVE	INACTIVE	INACTIVE	INACTIVE
2	L2	INACTIVE	INACTIVE	ACTIVE	INACTIVE	INACTIVE	INACTIVE	INACTIVE
3	L3	INACTIVE	INACTIVE	INACTIVE	INACTIVE	INACTIVE	INACTIVE	ACTIVE
4	L4	INACTIVE	INACTIVE	INACTIVE	INACTIVE	INACTIVE	INACTIVE	INACTIVE
5	L5	ACTIVE	INACTIVE	INACTIVE	INACTIVE	INACTIVE	INACTIVE	INACTIVE

CONCLUSION

Targeting the RAD51C mutation represents a promising avenue for developing precision therapies, particularly in cancers associated with homologous recombination deficiency. In this study, computational approaches—including structure-based drug design, virtual screening, molecular docking, and binding site prediction—proved effective in identifying potential inhibitors of mutant RAD51C. Tools such as SiteMap and CASTp enabled accurate localization of ligand-binding pockets, while docking simulations revealed several lead compounds (D1 to D5) with strong binding affinities and favorable interactions. These findings lay the groundwork for future in vitro and in vivo validation, offering a strategic point for the development of targeted RAD51C inhibitors that may enhance therapeutic outcomes in cancers with defective DNA repair mechanisms.

ACKNOWLEDGEMENT

Authors are thankful to Osmania University for providing computational software's.

REFERENCES

1. Somyajit, K., Subramanya, S., & Nagaraju, G. (2012). RAD51C: A novel cancer susceptibility gene is the new addition to the HR family. *DNA Repair*, 11(12), 1023–1028. <https://doi.org/10.1016/j.dnarep.2012.09.006>
2. Kuznetsov et al., *JCB* (2008) Showed RAD51C accumulation at DSB sites alongside RAD51, retention post-RAD51 disassembly, and direct involvement in checkpoint signaling post-DSB
3. Zhang et al. (2018) – review in MDPI *Genes* Discusses direct interaction of RAD51C with PALB2, a scaffold that simultaneously binds RAD51, RAD51C, and BRCA2—suggesting a coordinated complex that couples BRCA2 removal to RAD51 filament nucleation
4. Janavicius et al. (2012) – *Breast Cancer Research and Treatment*
In a European cohort, detected RAD51C pathogenic variants in ~2.6% of BRCA-negative breast/ovarian families
5. Schild, D. et al. (2000) “Evidence for a new RAD51-like gene, RAD51C, important for homologous recombination in human cells.” *EMBO Journal*, 19(3): 556–566.
<https://doi.org/10.1093/emboj/19.3.556>
6. Liu, N. et al. (1998) “XRCC2 and XRCC3: New human RAD51-family members that promote chromosome stability.” *Molecular and Cellular Biology*, 18(10): 6146–6153.
<https://doi.org/10.1128/MCB.18.10.6146>
7. Park, J. Y. et al. (2022) “Homologous recombination-deficient mutation cluster in tumor suppressor RAD51C.” *PNAS*, 119(37): e2203109119. <https://doi.org/10.1073/pnas.2203109119>
8. Suwaki, N. et al. (2011) “DNA damage response mediated by RAD51 paralogs.” *Journal of Nucleic Acids*, Article ID 368623. <https://doi.org/10.4061/2011/368623>
9. Meindl, A., et al. (2010). Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nature Genetics*, 42(5), 410–414. <https://doi.org/10.1038/ng.569>

10. Loveday, C., et al. (2012). Germline mutations in RAD51D confer susceptibility to ovarian cancer. **Nature Genetics**, 43(9), 879–882. (Includes analysis of RAD51C alongside RAD51D) <https://doi.org/10.1038/ng.893>
11. The UniProt Consortium. (2023). **UniProt: the Universal Protein knowledgebase in 2023**. *Nucleic Acids Research*, 51(D1), D523–D531. <https://doi.org/10.1093/nar/gkac1052>
12. Berman, H. M., Westbrook, J., Feng, Z., et al. (2000). **The Protein Data Bank**. *Nucleic Acids Research*, 28(1), 235–242. <https://doi.org/10.1093/nar/28.1.235>
13. Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). **PROCHECK: a program to check the stereochemical quality of protein structures**. *Journal of Applied Crystallography*, 26(2), 283–291. <https://doi.org/10.1107/S0021889892009944>
14. Morris et al. (1992) – Validation of Protein Models Using PROCHECK Morris, A. L., MacArthur, M. W., Hutchinson, E. G., & Thornton, J. M. (1992). Stereochemical quality of protein structure coordinates. *Proteins: Structure, Function, and Genetics*, 12(4), 345–364. <https://doi.org/10.1002/prot.340120407>
15. Colovos, C., & Yeates, T. O. (1993). **Verification of protein structures: patterns of nonbonded atomic interactions**. *Protein Science*, 2(9), 1511–1519
16. Bowie, J. U., Lüthy, R., & Eisenberg, D. (1991). **A method to identify protein sequences that fold into a known three-dimensional structure**. *Science*, 253(5016), 164–170.
17. Lüthy, R., Bowie, J. U., & Eisenberg, D. (1992). **Assessment of protein models with three-dimensional profiles**. *Nature*, 356(6364), 83–85.
18. Wiederstein, M., & Sippl, M. J. (2007). **ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins**. *Nucleic Acids Research*, 35(Web Server issue), W407–W410. <https://doi.org/10.1093/nar/gkm290>
19. Tian, W., Chen, C., Lei, X., Zhao, J., & Liang, J. (2018). **CASTp 3.0: Computed atlas of surface topography of proteins**. *Nucleic Acids Research*, 46(W1), W363–W367. <https://doi.org/10.1093/nar/gky473>
20. Halgren, T. A. (2009). **Identifying and characterizing binding sites and assessing druggability**. *Journal of Chemical Information and Modeling*, 49(2), 377–389. <https://doi.org/10.1021/ci800324m>
21. Lyu, C., Shen, J., Qiang, B., Liu, N., Wang, H., Zhang, L., & Liu, Z. (2021). Chemical Space, Scaffolds, and Halogenated Compounds of CMNPD: A Comprehensive Chemoinformatic Analysis. *Nucleic Acids Research*, 49(D1), D509–D515 (supplemental analysis).
22. Wang, J., Khan, Y. D., Khan, M., Zhang, Q., Liu, B., & Liu, X. (2023). **Pharmacophore-based virtual screening of CMNPD database for identification of novel Hsp90 inhibitors: molecular docking, ADMET prediction, and MD simulation**. *Molecules*, 28(24), 8074. <https://doi.org/10.3390/molecules28248074>
23. Li, Y., Khan, Y. D., Khan, M., Wang, J., Zhang, Q., & Liu, X. (2024). **Structure-based pharmacophore modeling, virtual screening, docking, ADME and MD simulation for identification of potent aromatase inhibitors from marine natural product database**. *BMC Chemistry*, 18, Article 46. <https://doi.org/10.1186/s13065-024-01350-9>
24. Pires, D. E. V., Blundell, T. L., & Ascher, D. B. (2015). **pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures**. *Journal of Medicinal Chemistry*, 58(9), 4066–4072. <https://doi.org/10.1021/acs.jmedchem.5b00104>
25. Jorgensen, W. L., & Duffy, E. M. (2002). **Prediction of drug solubility from structure**. *Advanced Drug Delivery Reviews*, 54(3), 355–366. [https://doi.org/10.1016/S0169-409X\(02\)00008-X](https://doi.org/10.1016/S0169-409X(02)00008-X)
26. Khan, Y. D., Zhang, Q., Wang, J., Liu, B., & Liu, X. (2023). **Structure-based virtual screening of CMNPD for identification of Hsp90 inhibitors using docking, QikProp ADMET, and MD simulation**. *Molecules*, 28(24), 8074. <https://doi.org/10.3390/molecules28248074>
27. Banerjee, P., Eckert, A. O., Schrey, A. K., & Preissner, R. (2018). **ProTox-II: A webserver for the prediction of toxicity of chemicals**. *Nucleic Acids Research*, 46(W1), W257–W263. <https://doi.org/10.1093/nar/gky318>
28. Drwal, M. N., Banerjee, P., Dunkel, M., Wettig, M. R., & Preissner, R. (2014). **ProTox: a web server for the in silico prediction of rodent oral toxicity**. *Nucleic Acids Research*, 42(Web Server issue), W53–W58. <https://doi.org/10.1093/nar/gku401>
29. Ramachandran, G. N., Ramakrishnan, C., & Sasisekharan, V. (1963). **Stereochemistry of polypeptide chain configurations**. *Journal of Molecular Biology*, 7(1), 95–99. [https://doi.org/10.1016/S0022-2836\(63\)80023-6](https://doi.org/10.1016/S0022-2836(63)80023-6)
30. Morris, A. L., MacArthur, M. W., Hutchinson, E. G., & Thornton, J. M. (1992). Stereochemical quality of protein structure coordinates. *Proteins: Structure, Function, and Genetics*, 12(4), 345–364. <https://doi.org/10.1002/prot.340120407>
31. Wiederstein, M., & Sippl, M. J. (2007). **ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins**. *Nucleic Acids Research*, 35(Web Server issue), W407–W410. <https://doi.org/10.1093/nar/gkm290>
32. Meslamani et al. (2007) In Evaluation of the structural quality of modeled proteins, ProSA II Z-scores were plotted alongside other validation metrics (e.g., GDT_TS, ModCheck) to evaluate models from CASP targets

33. Laskowski, R. A. (2009). PDBsum new things. *Nucleic Acids Research*, 37(Web Server issue), W349–W353. <https://doi.org/10.1093/nar/gkp313>
34. BIOVIA Discovery Studio, Release 2021 (or your version), Dassault Systèmes, San Diego, CA, USA.
35. Gurung, A. B., & Bhattacharjee, A. (2020). Structure-based virtual screening and molecular docking of natural compounds for potential inhibition of the SARS-CoV-2 main protease. *Journal of Biomolecular Structure and Dynamics*, 39(16), 6286–6295. <https://doi.org/10.1080/07391102.2020.1772885>
36. Gupta, A., Sharma, N., & Dube, T. (2021). Identification of potential inhibitors of SARS-CoV-2 main protease via QikProp-based virtual screening. *Journal of Biomolecular Structure and Dynamics*, 39(18), 6625–6637. <https://doi.org/10.1080/07391102.2020.1843777>
37. Schrödinger Release 2024.1, QikProp, Schrödinger, LLC, New York, NY, 2024.
38. Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 23(1–3), 3–25. [https://doi.org/10.1016/S0169-409X\(96\)00423-1](https://doi.org/10.1016/S0169-409X(96)00423-1)
39. Zhao, Y., & Abraham, M. H. (2002). Lipophilicity, hydrogen bonding and molecular size as major determinants of blood–brain barrier penetration. *Journal of Pharmacy and Pharmacology*, 54(8), 1171–1182. <https://doi.org/10.1211/002235702760345414>
40. Banerjee, P., Eckert, A. O., Schrey, A. K., & Preissner, R. (2018). ProTox-II: A webserver for the prediction of toxicity of chemicals. *Nucleic Acids Research*, 46(W1), W257–W263. <https://doi.org/10.1093/nar/gky318>