STRUCTURAL DETERMINATION OF PLASMA KALLIKREIN PROTEIN AND IDENTIFICATION OF ITS INHIBITORS IN CANCER: AN *IN-SILICO* STUDY

Vani Kondaparthi[1], Madhavi Latha Bingi[1], Thirupathi Damera[1], Priyadarshini Gangidi[1], Mounika Badineni[1], Kiran Kumar Mustyala[3], Vasavi Malkhed[1,2*]

[1]Molecular Modelling Research Laboratory, Department of Chemistry, Osmania University, Hyderabad, Telangana, India – 500007;
[2]Department of Chemistry, University College of Science, Saifabad, Osmania University, Hyderabad, Telangana, India – 500004;
[3]Department of Chemistry, Nizam College, Osmania University, Hyderabad, Telangana, India – 500001.

*Corresponding Author: Vasavi Malkhed
*Department of Chemistry, University College of Science, Saifabad, Osmania University, P.O. Box: 500004, Hyderabad, India; E-mails: mvasavi@osmania.ac.in; vasavimalkhed@gmail.com; mvasavi2003@gmail.com.

ABSTRACT
Background: The serine protease known as KLK B1 is essential to the kallikrein-kinin system and has been linked to the development of cancer. Tumor growth, metastasis, and alterations in immunological responses may be caused by dysregulation of KLK B1 and its endogenous inhibitors.
Objective: Determine the expression levels of KLK B1 and its inhibitors in various cancer types, evaluate their function in tumor progression, and assess putative inhibitory mechanisms that contribute to the formation of cancer.
Method: Using homology modelling protocols, the theoretical structure of KLK B1 will be predicted and the resulting model will be validated by several server tools. To identify new scaffold compounds that are effective against KLK B1, the active site will be examined and the ligand database is used for virtual screening.
**Result:** Compared to normal tissues, KLK B1 was markedly elevated in pancreatic, prostate, and breast malignancies. HIS44, ASP93, and SER188 residues in the active site triad and protein residues from GLN25 to GLY211 will be chosen as a pocket for ligand molecule binding. With the results of virtual screening and ADME characteristics, the scaffolds containing the amide pharmacophore were recognized as potential hits against the plasma kallikrein protein. KLK B1 inhibitors decreased the migration and proliferation of cancer cells while encouraging apoptosis, indicating that they may have therapeutic value.
Conclusion: The development of cancer is significantly influenced by KLK B1. The research results determined that the chosen ligand molecules with ADME parameter values are more acceptable scaffolds, highlighting the ligand molecules' drug-like activity through inhibition of KLK B1 protein. Identification of novel therapeutic scaffolds for cancer is aided by structural data, active site details, and specific ligand molecules.

Keywords: Serine proteases, Cancer, Docking, Virtual screening, Kallikreins, Homology modelling, *In-Silico* studies.

INTRODUCTION

India has experienced a tremendous increase in cancer incidence. According to the 2020 Apollo Hospitals' Health of the Nation report, cancer is a major worry, with about 14 lakh cases reported in 2020 and expected to rise to 0.15 million by 2025. Breast, cervical, and ovarian cancers are the most frequent malignancies among Indian women, whereas males are more likely to get lung, oral, and prostate cancers. A key challenge in India is the low incidence of cancer screening, which is significantly below worldwide standards. The average age of cancer diagnosis is also lower than in other nations, with breast cancer being detected at 52, lung cancer at 54, and cervical cancer at 59. In terms of the global perspective, Cancer is a serious public health issue across the world. The report of WHO-IARC has been reported in Figure 1. The United States alone is expected to have nearly 2 million new instances of cancer and 600,000 deaths from the disease in 2020, according to the American Cancer Society's "Cancer Facts & Figures 2020". Many cancer types now have higher survival rates thanks to developments in early identification, therapy, and lifestyle modifications, but there are still major obstacles to overcome, especially in low and middle-income nations.
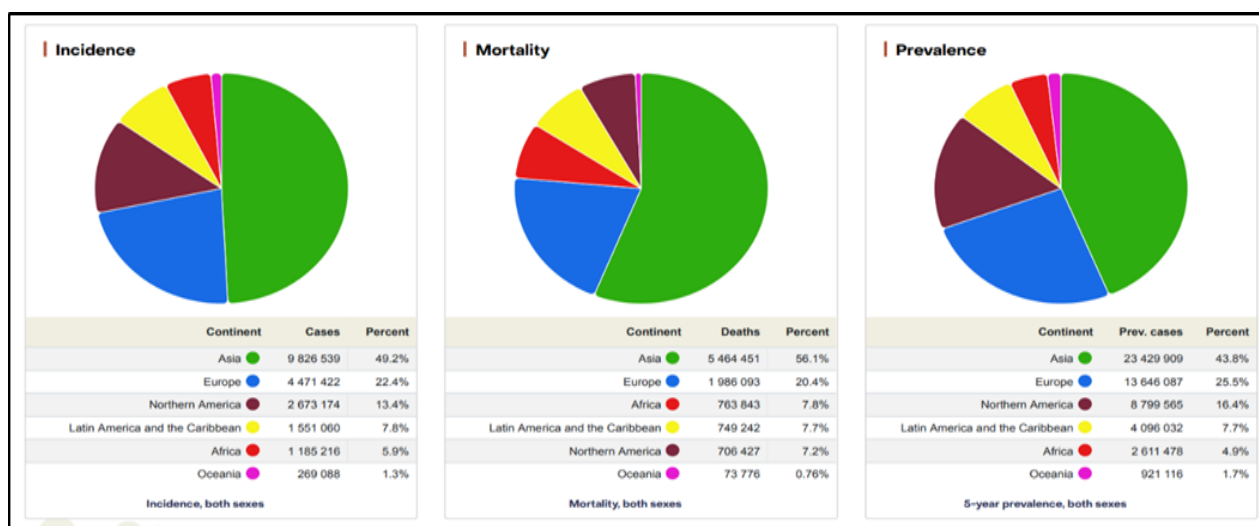
Fig.1 WHO-IARC Globocan report-2020.

A notable class of enzymes known as serine proteases are distinguished by their capacity to break peptide bonds in protein by use of serine residue located in the enzyme's active site. They are essential for several physiological functions, such as blood coagulation, immunological response, digestion, and cell signaling. Serine proteases, such as chymotrypsin, elastase, and trypsin, are often found in the digestive system and are necessary for the breakdown of proteins. The catalytic mechanism of these enzymes is extremely conserved and involves the serine residue, which temporarily establishes a covalent link with the substrate to facilitate its cleavage. Serine proteases are important in maintaining physiological homeostasis (1) and have the potential to be therapeutic targets due to their ability to cause mutations or dysregulation in several disorders, including emphysema, chronic pancreatitis, and some kinds of cancer. KLK B1, also called plasma kallikrein, belongs to the kallikrein family of serine proteases. This enzyme is mostly expressed in the pancreas, kidneys, and salivary glands, where critical roles in a variety of physiological functions are expressed. KLK B1 regulates blood pressure, inflammation, and tissue remodeling, principally through its capacity to cleave kininogen and produce bradykinin, a strong vasodilator (2, 3). The KLK B1 gene is found on chromosome 19q13.3-4, amid a cluster of 15 kallikrein genes. KLK B1 is synthesized as an inactive zymogen that needs proteolytic cleavage to activate (4). Once activated, KLK B1, a plasma kallikrein, has proteolytic activity (5) on a wide range of substrates, including peptide hormones, growth factors, and extracellular matrix proteins. In recent years, KLK B1 misimpression and dysregulation have

been linked to several clinical disorders, including hypertension, chronic inflammatory illnesses (6), and cancer(7). KLK B1 levels are elevated in a variety of cancers, indicating that it may have a role in tumor development, angiogenesis, and metastasis. KLK B1's role in cancer biology is noteworthy. The enzyme's proteolytic activity can alter the tumor microenvironment, impact cell signaling pathways, and promote cancer cell invasion and metastasis, as outlined in Figure 2. KLK B1's multiple activities make it an intriguing therapeutic target and a possible biomarker for cancer diagnosis and prognosis (8). The present study will offer a description of KLK B1, focusing on its structure and ligands acting against protein based on virtual screening and docking. Understanding the complexities of KLK B1 activity is intended to emphasize its potential as a target for innovative treatment as well as its value as a diagnostic biomarker in cancer (9, 10).
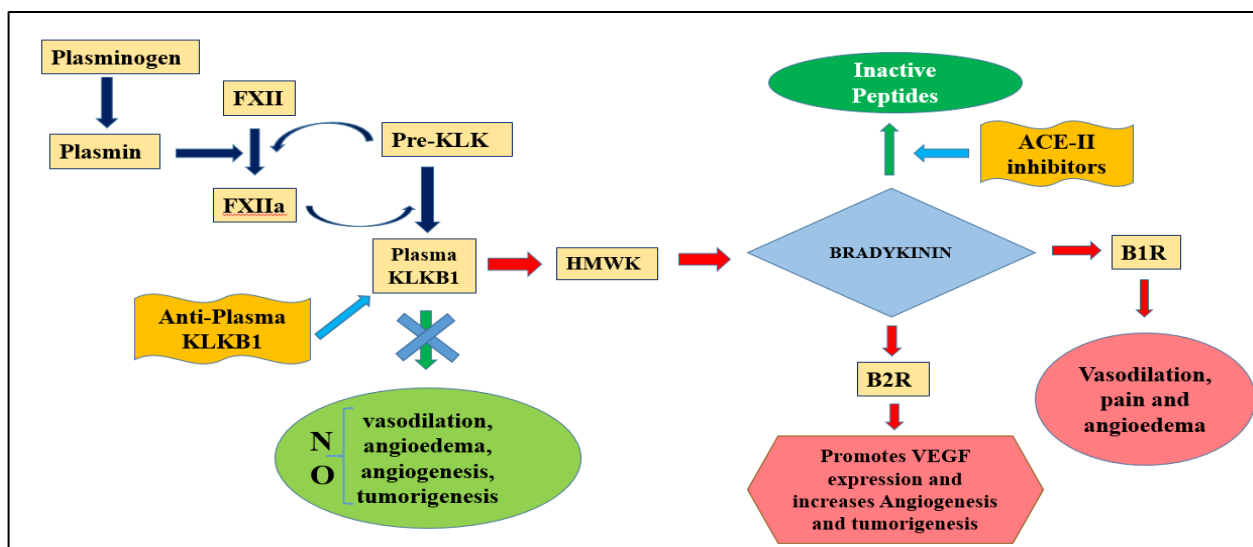
Fig.2 Biochemical pathway of KLKB1 protein.

MATERIALS AND METHODS

Homology Modelling: Homology modeling is a method to estimate the 3D framework of proteins. It entails identifying a model and aligning it in the order to the protein(11). The 3D framework of KLK B1 in the PDB was determined without using either X-ray crystallography or NMR. Modeller 9.20 (12, 13) is accustomed to construct a 3D model of the protein(14). The FASTA (15) of KLK B1 (Access. No. P03952) protein was chosen from UniProtKB records and the template (PDB ID: 2ANW) was discovered utilizing BlastP in BLAST(16) and JPRED (Jnet prediction)(17) servers by sending the FASTA sequence of KLK B1. Based on high query coverage and a statistically significant E-value, the optimal template was chosen. Modeller 9.20 was used to generate 25 models. The resulting 3D structure of KLK B1 selected for more research had the lowest modeller functional energy value.

Structure Validation: The lowest energy protein structure was verified utilizing a variety of validation techniques. ProSA (18) evaluates the overall quality of the 3D structure by comparing the protein's z-score to known protein structures. The z-score of the generated 3D framework of the protein reflects the model quality of typical native structures. Ramachandran plot analyses the psi (ψ) and phi (ϕ) angles of amino acids to assess stereochemical quality by checking if angles fall within allowed regions for proper folding(19). The Ramachandran plot of the predicted 3D structure from Modeller shows that KLK B1 has good stereochemical quality, with 91.5% of residues in the most favoured region, 8.5% in the additionally permitted region, 0.0% in the liberally allowed zone, and 0% in the unfavoured zone.

Virtual Screening:

Protein preparation: KLK B1 protein was optimized with the help of the protein preparation wizard in the Schrödinger suite's Maestro module(20). When the protein chains were reoriented, bond ordering was assigned, and missing sidechains were filled with Prime suite. Steric disagreements were resolved, hydrogens were included, and water molecules were eliminated. Using the Impref (21) module, energy minimization was executed with the default limitation of RMSD value of 0.30 Å using OPLS-2005 (20).

Ligand preparation: Schrödinger's LigPrep module was utilized for the ligand preparation. From the Asinex database, about 80000 ligands were used to construct high-quality conformations of molecules. Protonation states for every ligand component were obtained by Schrödinger's Epik tool with the force-field OPLS-2005 (22).

Active Site identification: The determination of the cavity that binds protein is critical for docking small compounds in drug discovery using structural pharmacology. The binding region of the KLK B1 protein was discovered by using literature, CASTp (23), and the SiteMap tool (24). The PDBsum website offers a ligand binding practically at the active site, i.e., Benzamidine binding itself to 2ANW. These residues were connected to the KLK B1 protein pattern to align several sequences in Clustal omega (25) and Swiss-model ExPASy server(26). CASTp provides an online resource for locating, delineating, and measuring concave surface regions as pockets and voids buried in the interiors of 3D protein structures. A SiteMap module aids in the identification of binding cavities and the prediction of site information, including quantifiable data and maps with graphical representations of hydrophobic and hydrophilic areas, bond donors, and bond acceptors. Glide, a tool in the Schrödinger suite, generates a grid taking into account the site residues of the KLK B1, with dimensions of 40x40x40 Å and 80x80x80 Å for receptor configuration and grid domain, respectively, using 1Å size. The grid box thus created has a binding cavity that is appropriate for virtual screening.

Screening: Virtual Screening is the simplest way to identify the optimally docked ligands as blockers of KLK B1. The KLK B1 protein was screened virtually at its operational site utilizing the GLIDE tool (27) of the Schrödinger suite. The screening procedure includes the LigPrep (28) results out file. Qikprop tool, used to predict ADME, was selected. Before docking, ligands were prefiltered using LROF and JROT. There are three hierarchical filtering stages in the virtual screening approach: the initial step filters ligands using HTVS docking, followed by SP docking(29), and finally XP docking(30). During each step, 10% of the molecules are filtered. The optimally docked structures were visualized with Biovia Discovery Studio 2017R2(31) for further study.

Prime MM-GBSA: Prime MM-GBSA (32) was utilized to calculate the free energies of ligands using XP docking results. The protein-ligand complex's binding free energy is estimated(33).

$$\Delta G\,Bind = \Delta E\,MM + \Delta G\,sol - T\,\Delta S$$

ΔG Bind represents the binding energy, whereas ΔE MM is the change in Molecular Mechanics energy, (ΔG sol means the solvation-free energy, and -TΔS indicates the conformational entropy upon binding, respectively.

Predicting ADME properties: The discovered hit compounds' pharmacological parameters were computed using a QikProp tool. The ADME properties of the above-hit compounds in drug discovery were determined to be satisfactory (34). Pharmacokinetic characteristics determine drug-likeness in ligand compounds. As a result, drug development is the most difficult step in the process of identifying lead molecules.

The ADME characteristics of organic compounds are critical for the creation of a successful pharmacological compound(35). The top hit compounds were identified through a sorting procedure by evaluating ADME parameters(36).

AutoDock: AutoDock 1.5.6 version(37) is a tool for determining the binding energies of KLK B1 and ligand molecules utilizing the AMBER force field. Eventually, 12 ligands were generated using XP docking as PDBQT files (38). The grid and docking parameter files were provided. The optimal configurations of ligands near the KLK B1 protein's active site were discovered using a Lamarckian genetic algorithm. Certain literature reviews used comparable procedures.

RESULTS AND DISCUSSION

Structure Analysis and Validation: The three-dimensional framework of the KLK B1 protein has not been disclosed in PDB. The target's primary KLK B1 protein (human) structure was obtained from UniProtKB using the access number P03952. KLK B1 protein has a light chain length of 240 amino acids. The template was identified by using the KLK B1 protein chain in the JPRED and BLAST servers. Template 2ANW was chosen for homology modelling of KLK B1 protein according to query coverage (99.59%), E-value (0.0), and sequence identity (97%), which were acquired as shown in Table 1. Low E-values indicate good protein uniqueness. The pairwise sequence alignment of KLK B1 and 2ANW protein sequences was performed in the Swiss Model EXPASY server tool and shown in Figure 3. Modeller 9.20 was predicated on homology protein modeling and was used to build 25 KLK B1 protein models.

Fig.3 Alignment of template and protein chain by the EXPASY server.

Table No.1 Comparison of the Template search (E-Scores and PDB codes) obtained from the various servers for KLK B1 protein.

S.NO	Name of the protein Database search server	Parameters considered for template selection	E-Score value	PDB ID	QUERY COVERAGE
1	BLAST	Sequence similarity	0.0	2ANW_A	99.59%
			1E-179	5F8T_A	99.58%
2	JPRED	Three-state (α-helix, β-strand, and coil) prediction of secondary structure	E-145	2ANW_A	
			E-143	5F8T_A	

The highlighted values in Table 1 represent the selection of 2ANW PDB ID containing chain as a template for determining the KLK B1 structure.

The structure with the minimum objective function (1322.72 KJ/mol) of the KLK B1 protein was chosen for additional investigation. The stabilized protein structure is tested for dependability using established validation techniques. The KLK B1 protein's 3D model predicted by the PDBsum server has 5α helices (Table 2), 14β strands (Table 3), 2β sheets (Table 4), and 4 disulfide bridges (Table 5) and is presented in Figure 4.

Table No.2 The secondary structure α-helices details of the KLK B1 protein using the PDBsum server.

S.No	Starting residues	Ending residues	No. of residues	Sequence
1	ALA 43	PHE 46	4	AHCF
2	LYS 117	SER 121	5	KGDTS
3	ASN 155	ARG 161	7	NEECQKR
4	VAL 224	TYR 227	4	VAEY
5	MET 228	SER 237	10	MDWILEKTQS

Table 2 presents the starting and ending residues in a α-helix and also the number of residues per sequence with one-letter codes of amino acid in a sequence.

Table No.3 The secondary structure β-strand details of the KLK B1 protein using the PDBsum server.

S.No.	Starting residues	Ending residues	No. of residues	Sequence
1	THR 5	ASN 6	2	TN
2	GLN 15	LYS 21	7	QVSLQVK
3	GLN 25	GLY 35	11	QRHLCGGSLIG
4	TRP 38	THR 41	4	WVLT
5	TRP 55	SER 59	5	WRIYS
6	ILE 76	ILE 81	6	IKEIII
7	ALA 95	LEU 99	5	ALIKL
8	CYS 127	GLY 131	5	CWVTG
9	GLN 146	ASN 149	4	QKVN
10	LEU 152	VAL 153	2	LV
11	MET 171	ALA 174	4	MVCA
12	PRO 191	HIS 196	6	PLVCKH
13	MET 199	TRP 208	10	MWRLVGITSW
14	GLY 219	LYS 223	5	GVYTK

Table 3 tabulates the details of β-strands in the protein chain with starting and ending residues along with single letter code of amino acids in a sequence.

**Table No.4** The β-sheets of KLK B1 protein results were predicted using the PDB sum server.

Sheet	No. of Strands	Type
A	8	Antiparallel
B	6	Antiparallel

The number of β-sheets predicted by the PDBsum server are two i.e., A and B sheets with 8 and 6 strands respectively in antiparallel type are presented in Table 4.

Table No.5 Disulphide bridges in the KLK B1 protein are predicted by the PDB sum server.

S.No.	1st Cysteine	2nd Cysteine
1	A 29	A 45
2	A 127	A 194
3	A 158	A 173
4	A 184	A 212

Table 5 tabulates the number of cysteine linkages present in the A chain of the protein along with starting and ending numbers of the amino acid.
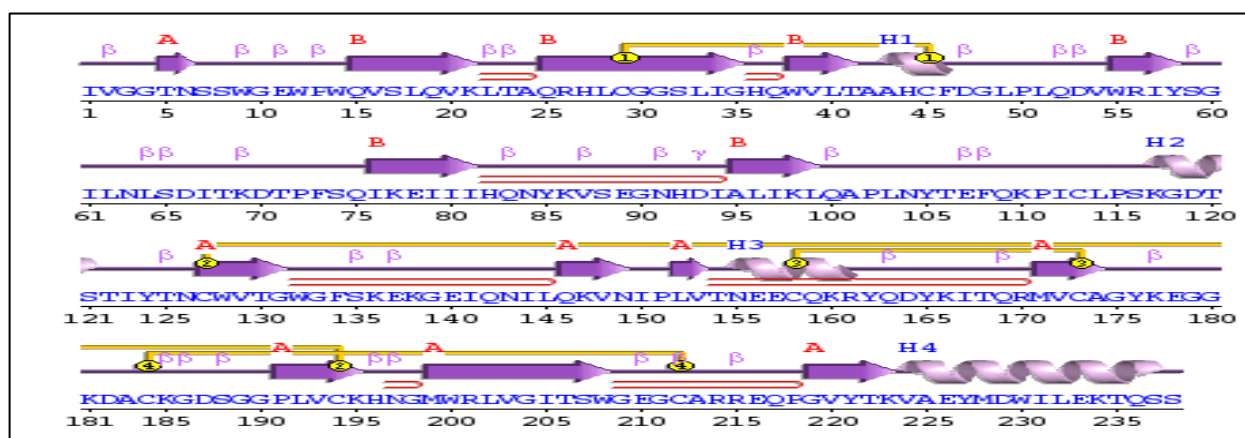

Fig.4 PDBsum server secondary structure determination.

The Ramachandran contour plot in Figure 5 of KLK B1 protein demonstrates that 91.5% of amino acids are in the most preferred position, indicating that the protein has high stereo-chemical model quality. ProSA is a tool for verifying the quality of the KLK B1 model. The black dot represents the Z-score of the modelled structure concerning various empirically determined structures via NMR and X-ray areas, which are displayed in navy blue and light blue, accordingly. Figure 6 displays a Z-score of -7.89, indicating a consistent overall structure quality of the KLK B1 protein in Figure 7. The 3D structure of KLK B1 predicted using PyMOL is depicted in Figure 8.
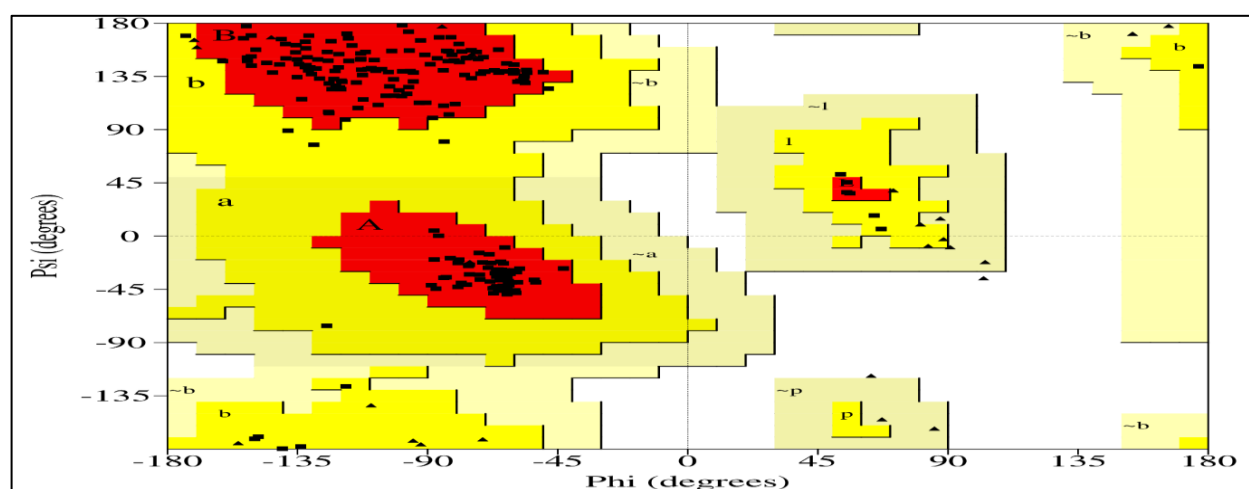
**Fig.5 The PROCHECK web server's Ramachandran contour plot.**

Residues in most favoured regions [A, B, L]	193	91.5%
Residues in additional allowed regions [a, b, l, p]	18	8.5%
Residues in generously allowed regions [~a, ~b, ~l, ~p]	0	0.0%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	211	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	25	
Number of proline residues	10	
Total number of residues	248	
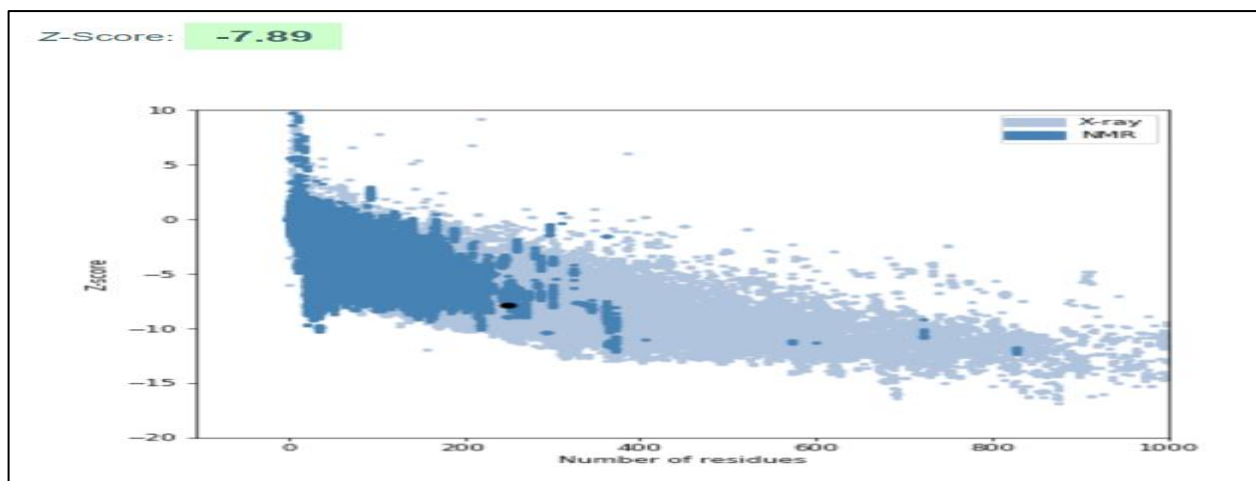
Fig.6 The Z-Score determination by ProSA.
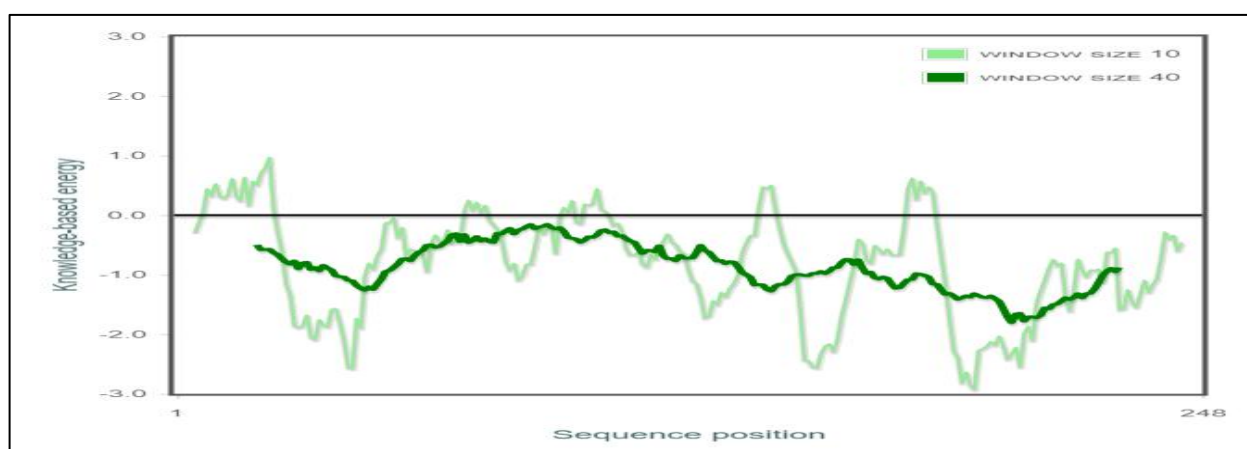
Fig.7 ProSA energy graph.
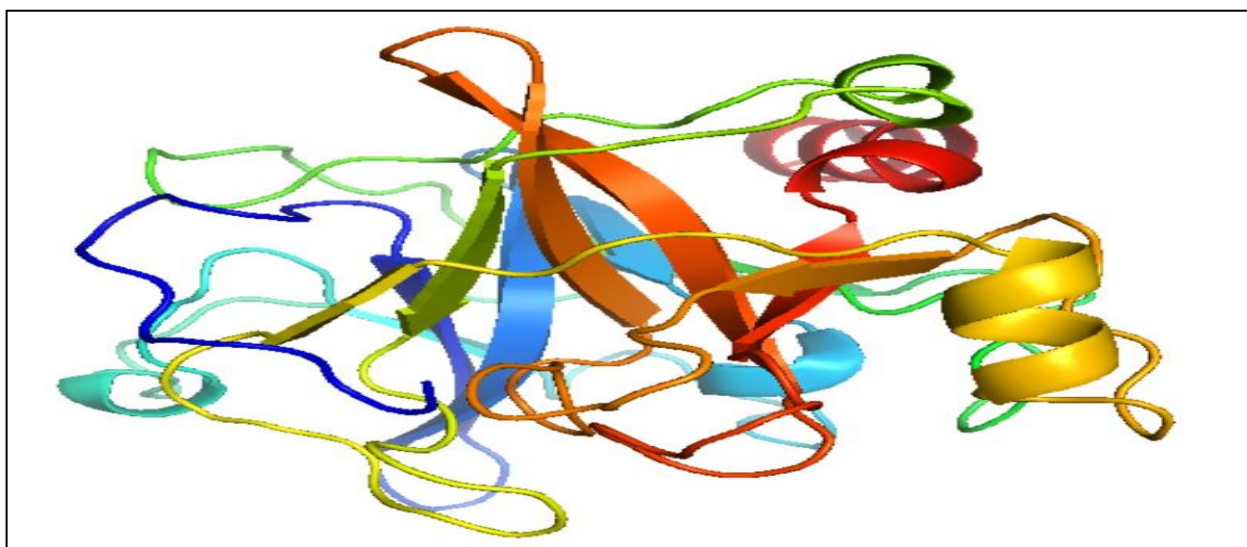

**Fig.8 KLK B1 3D secondary structure visualized using PYMOL software.**

Active Cavity Prediction of KLK B1:

The recognition of a possible active cavity for a target protein is critical in structure-based drug design(39). The 2ANW amino acids were connected to KLK B1's amino acid residue sequence in the COACH-D server, as illustrated in Figure 9 and tabulated in Table 6. The putative binding site in Figure 10 predicted by CASTp shows hydrophobic regions with volumes 65.754Å and 64.350Å, respectively, and presented in Table 7. The conserved domain of KLK B1 protein analyzed using

BLAST was suggested in Figure 11. The binding residues of the KLK B1 were extracted from the template 2ANW using LigPlot analysis and the PDBsum server as depicted in Figure 12. These residues found by KLK B1 protein correlation were then employed in docking experiments to find the lead inhibitors of KLK B1. The SiteMap module was additionally employed to locate a possible binding receptor for the KLK B1 protein using force-field in Figure 13. A SiteMap describes a Site score of 1.00 as having amino acids. The active site was generated with the center of X: 58.97, Y: 9.55, and Z: 41.50. The three-dimensional grid box was produced utilizing the Glide tool for the binding of ligands.
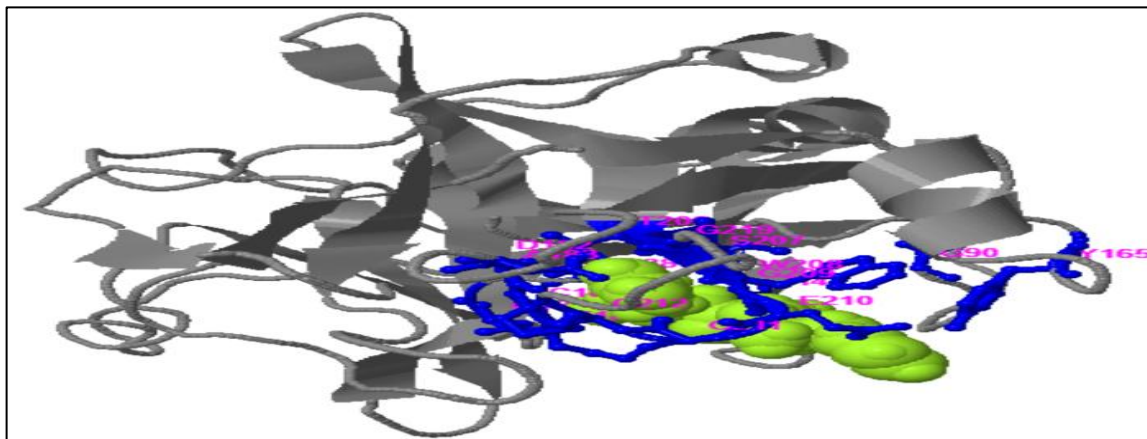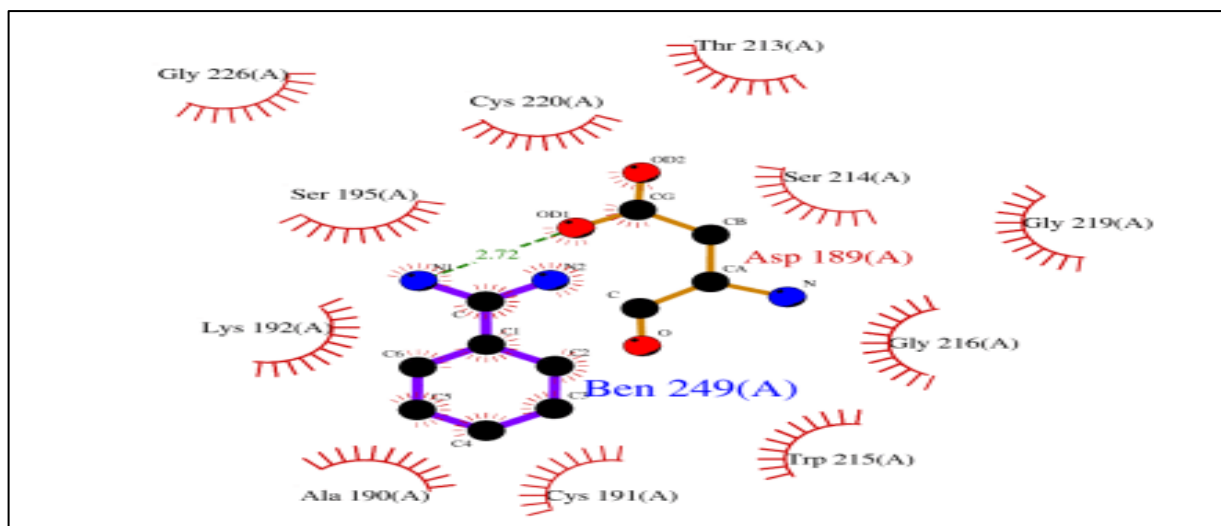

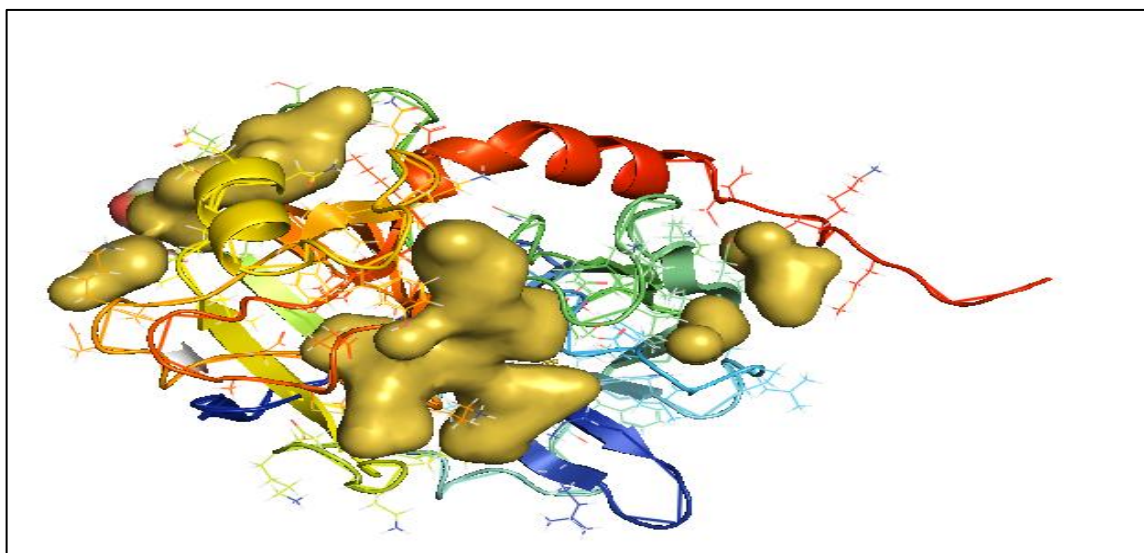**Fig.9 COACH-D server evaluation of binding pocket.**
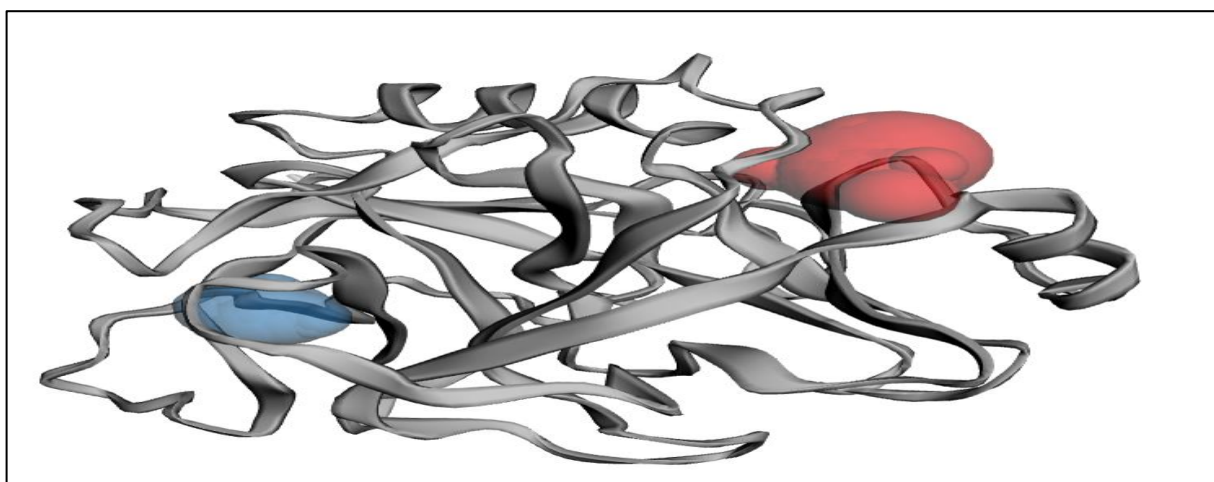
Fig.10 CASTp prediction of binding sites.
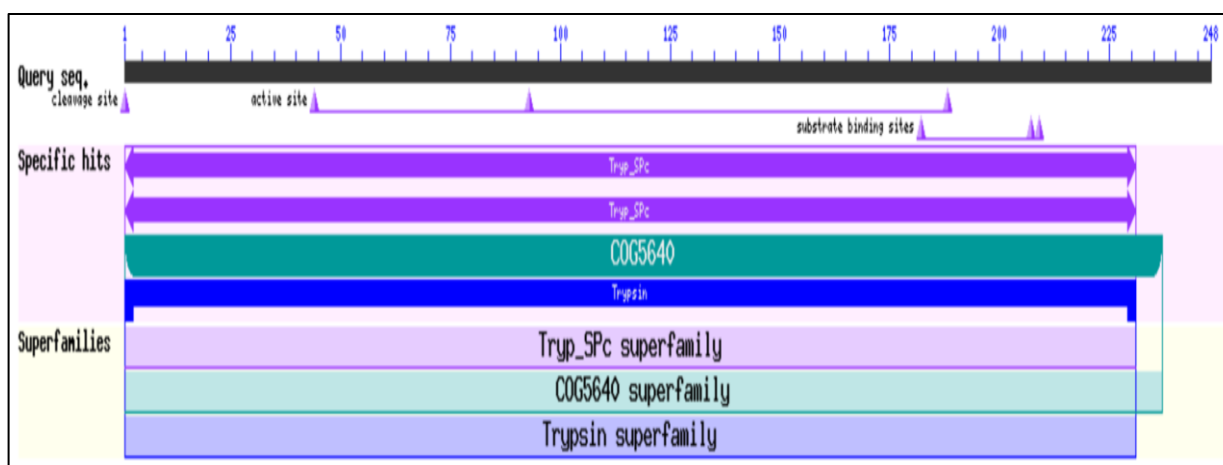
Fig.11 BLAST conserved domain of KLK B1.

Fig.12 PDBsum prediction of binding residues.

Fig.13 SiteMap module determination of active sites.

Table No.6 Binding residues based on COACH-D server.

S.No.	Rank	C-Score	Cluster size in Å	Energy	Predicted binding residues
1.	1	0.46	743	-4.7	77,76,93,132,133,136, 223,227,241,242,243,244, 245, 246,247, 257
2.	2	0.13	318	-4.1	76,77,78,183,223,226,227, 242,244,245, 246,248

Table 6 represents the COACH-D server-predicted binding residues. Active sites with rank 1, having the highest confidence score (C-score) and energy, are selected for further study.

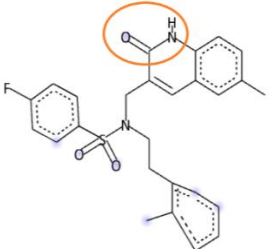**Table No.7** Binding pockets evaluation by CASTp server.

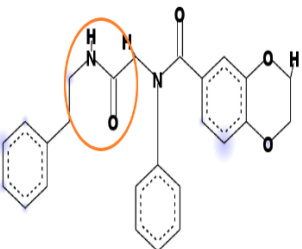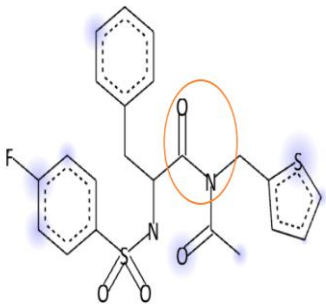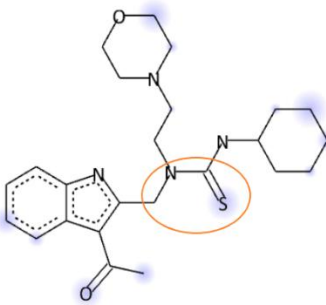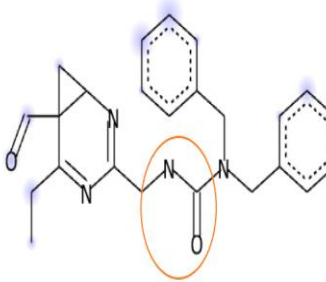| S. No	Server	POC ID	Volume (SA)Å^3	Amino acid residues
1.	CASTp	1	68.673	THR120,SER121,THR122,ILE123,TYR124, LEU152,VAL153,THR154,ASN155, MET171, VAL172, VAL207,LYS223
2.		3	24.559	SER7,SER8,TRP9,GLY10,GLU11,TRP12, PRO13,GLN15,GLY60,ILE61,LEU62, PHE108, TRP132,LEU145

The binding pockets of the KLK B1 protein cavity evaluated by the CASTp server tool are tabulated in table 7. Amino acid residues in pocket ID 1 with the highest volume are used as binding residues.
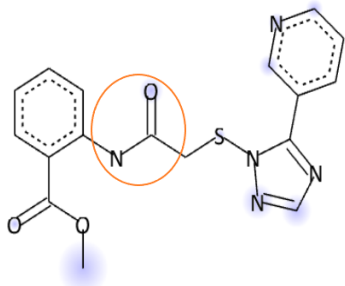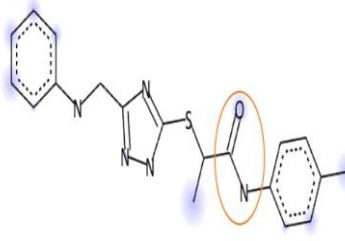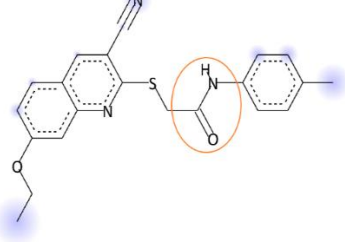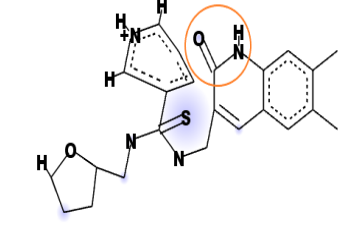
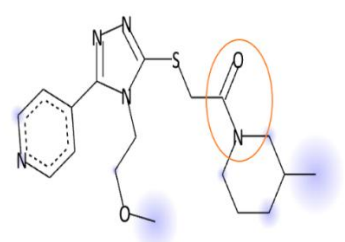**Docking Studies and Binding-Free Energies**

Docking utilizing Glide: Structure-based virtual screening was performed to identify new ligand compounds against KLK B1 protein. The Asinex information of 80000 ligands was produced with the LigPrep module, yielding 122734 lowest energy conformers as an output file. The LigPrep out file of 122734 ligand molecules underwent virtual screening in the Maestro module of the Schrödinger suite. Ten percent of compounds were docked accurately in HTVS mode after running the QikProp and Lipinski filter, yielding 5946 compounds. 10% of the above compounds were flexibly docked in SP mode, yielding 594 scaffolds. These 594 scaffolds were again filtered in XP mode, yielding 10% of the molecules and 60 ligands due to protein-ligand docking. The XP docked complexes' free energy associated with binding was computed using the Prime MM-GBSA module. The independent energies of protein-ligand structures (dG bind) evaluated the bonding intensity. The top twelve scaffolds with the lowest binding energy demonstrate the binding affinity of KLK B1 protein-ligand structures. The MM-GBSA results demonstrate that all twelve lead compounds discovered using molecular docking have favourable binding free energy ranging from -61.27 to -35.16 dG (kcal/mol). Prioritized lead scaffolds from the Schrödinger suite demonstrated consistent interactions, as given in Table 8.

Table No.8 The top-scoring ligands of KLK B1 and ligand docked complexes.

Ligand No.	Structure of the Molecule	Glide Score	Glide energy (K.Cal/Mol)	Hydrogen bond interactions	Distance (Å)
L1		-7.40	-60.86	L1:O:LYS185:HZ2 L1:H - GLY211:O L1:F - GLY90:HA2 L1:O LYS185:HE2 L1:F - GLY89:O L1 -TRP208 L1-CYS184:C,O; LYS185:N L1-CYS184:C,O; LYS185:N L1-TRP208:C.O; GLY209:N L1-TRP208:C.O; GLY209:N L1:C - ALA183 L1:C - TRP208 L1:C - TRP209 L1:C - TYR221 L1 - ALA183	2.35 1.92 2.80 2.56 2.96 5.42 4.00 4.96 4.10 3.25 3.76 4.20 4.37 3.70 4.34

L2		-6.97	-58.26	L2:H - GLY209:O L2:H - GLY211:O L2:O - LYS185:HE3 L2:H - GLY209:O L2:H - GLU89:O L2-CYS184:C,O; LYS185:N L2 -TRP208:C,O;GLY209:N L2 - ALA183 L2 - LYS138	2.02 2.96 2.54 2.83 2.36 4.31 3.54 4.73 5.26
L3		-7.24	-61.27	L3:N11 - GLY186:N L3 - HIS44 L3 - LEU28	3.21 4.84 4.45
L4		-7.17	-58.5	L4:N7 - GLY60:O L4:C5 - TRP9 L4 - TRP9 L4 - PHE108 L4 - PRO72 L4 - PRO72 L4 - LEU103	3.23 3.79 4.76 4.75 3.96 4.27 4.98
L5		-8.43	-58.45	L5:O10 - GLN19:NE2 L5:O10 - HIS27:N L5:N11 - GLN19:OE1 L5:C12 - LEU64:O L5 - ILE141:CG2 L5 - ILE67 L5 - ARG26	2.73 3.39 3.27 3.19 3.70 5.41 5.25

L6		-7.45	-35.16	L6:N19 - TYR165:OH L6:O13 - LYS185:NZ L5:O4 - HIS44:CD2 L5:O2 - SER88:CA L5 - HIS44 L5 - TYR165 L5 - TRP208 L5 - TRP208	3.07 2.82 3.65 3.49 3.80 4.80 5.10 4.75
L7		-7.22	-56.87	L7:N18 - ARG26:CA L7 - GLN25:NE2 L7:C12 - LYS69 L7 - LEU64 L7 - ILE141	3.74 3.40 4.63 4.50 4.16
L8		-7.09	-53.66	L8:O19 - GLN19:NE2 L8:H39 - GLN25:O L8 -ARG26:NH1 L8 - ILE141:CG2 L8:C15 - LEU28 L8 - ARG26 L8 - ARG26 L8 - ILE141 L8 - ILE67	3.24 2.79 4.83 3.80 5.41 5.43 4.47 3.97 5.17
L9		-7.00	-49.41	L9:H60 - ASP93:OD2 L9:H31 - SER188:OG L9:H43 - SER207:O L9:H50 - TYR85:OH L9:H50 - GLU89:O L9:H51 - SER88:O L9 - LYS185:NZ L9:C23 - ALA183 L9:C24 - CYS212 L9 - TYR165 L9 - TRP208 L9 - TRP208 L9 - LYS185	1.87 2.05 2.44 2.83 2.56 2.33 4.94 3.41 4.73 5.28 4.70 4.79 4.84
L10		-6.91	-46.45	L10:N7 - GLN19:NE2 L10:C12 - ASN143:OD1 L10 - ILE141:CG2 L10 - LEU64 L10 - ILE67	2.95 3.73 3.76 4.41 5.43

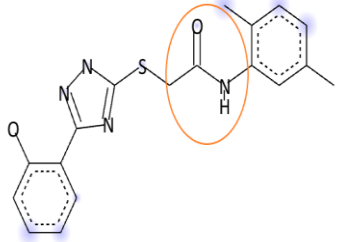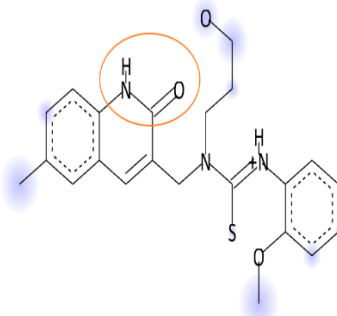|					
L11		-7.31	-44.01	L11:O24 - HIS27:N L11:H35 - GLN25:O L11 - ARG26:NH1 L11 - GLN25:CB L11 - ILE141:CG2 L11 - THR68:N L11:C8 - ILE67:C,O L11 - ARG26 L11 - ILE141	3.02 2.44 4.43 3.88 3.99 5.24 5.15 4.38 4.38
L12		-7.80	-59.5	L12:O29 - HIS27:N L12:S7 - LEU64:O L12:H37 - GLN25:O L12:H47 - HIS27:O L12 - ARG26:NH1 L12 - ILE141:CG2 L12 - ILE67 L12 - ARG26 L12 - ILE141	3.34 3.58 3.07 2.93 4.84 3.79 5.31 4.76 4.02

Table 8 presents the small molecule ligand structures prioritized based on Glide scores, and the circled part of the structures represents the amide pharmacophore.

Docking using AutoDock Vina: Docking was performed using Autodock Vina (40, 41) on 12 compounds acquired from XP virtual screening results and prioritized based on their binding energies (-5.9 to -7.5kcal/mol) and interactions with the KLK B1 protein. In comparison analysis, the findings of AutoDock docking were shown to be consistent with those obtained via XP docking of molecules in the Schrödinger suite. Based on the Glide score, Prime MM-GBSA, and AutoDock Vina docking, it was shown that ligand molecules had high binding free energy dG bind values in comparison to the protein-ligand complexes given in Table 9. The ligand molecules L2, L3, L4, L5, L6, L10, L11, and L12 were falling in the allowed range of QlogP of n-octanol/water value (i.e., <5)(42), JROT(43), excellent HOA(44), and the LROF(45), which are the critical components in determining the identification of new lead compounds against the KLK B1 protein for cancer treatment. Thus, all the ADME properties are presented in Table 10.

**Table No.9** The prime MM/GBSA calculations of ligand database.

Ligand No.	Prime MM/GBSA_dG_Bind (k.cal/mol)	MM/GBSA_dG_Bind_Coulomb (k.cal/mol)	MM/GBSA_dG_Bind_Covalent (k.cal/mol)	MM/GBSA_dG_Bind_vdW (k.cal/mol)	MM/GBSA_dG_Bind_Solv_GB (k.cal/mol)
L1	-63.42	-14.25	1.343	-52.19	31.75
L2	-63.12	-20.19	6.11	-58.19	32.65
L3	-63.01	-9.513	1.51	-58.43	24.76
L4	-59.41	-9.457	12.40	-57.73	19.46
L5	-59.08	-15.74	9.34	-57.85	29.79
L6	-58.98	-11.73	8.19	-42.30	9.44
L7	-58.98	-17.92	4.60	-54.75	33.30
L8	-58.73	-26.55	2.01	-48.37	35.18
L9	-58.10	-8.483	8.07	-44.12	13.22
L10	-57.38	-3.490	0.05	-46.19	7.96
L11	-56.31	-19.48	3.16	-40.34	25.38
L12	-55.44	-18.84	14.20	-61.49	34.49

The prime MM/GBSA calculated values are presented in Table 9.

Solvent Accessibility Surface Area: In structural biology, the term "Solvent accessible Surface Area" (SASA) refers to a compound's surface area that is accessible to a solvent. Measured in square angstroms ($Å^2$), it explains how the molecule interacts with its surroundings. SASA made it possible to understand protein folding, stability, and interactions with other molecules, such as ligands, substrates, and other proteins (46). By determining the molecule's hydrophobic and hydrophilic areas, this computation aids in predicting binding sites and the molecule's general behavior within a biological system. For instance, since they are more exposed to the solvent and possible interacting partners, areas with high SASA are frequently implicated in binding interactions. The graph in Figures 14 and 15 shows significant variation in the residues before and after docking. Comprehending SASA is essential for drug design as it facilitates the identification of possible binding sites and enhances how ligands interact with their targets. All the solubility properties of the ligand molecules are presented in Table 11.
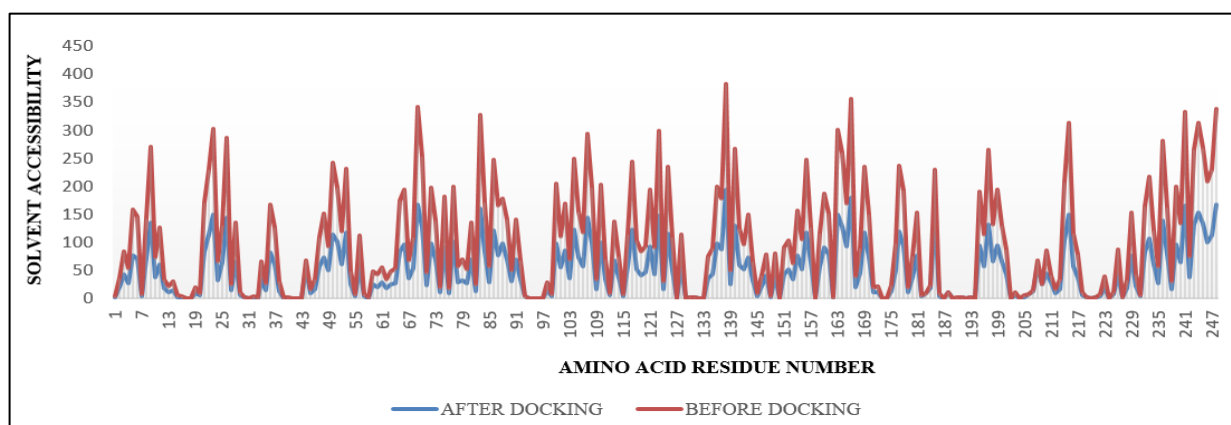
Fig.14 SASA values of KLK B1 protein.
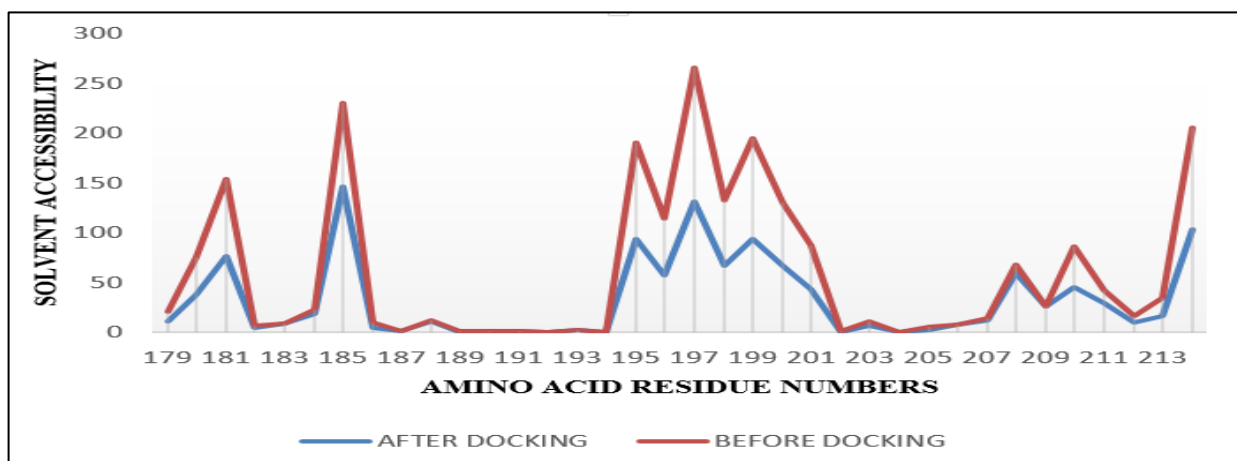
Fig.15 Change of SASA values of amino acid residues in complex formation.

Table No.10 The ADME properties of top-prioritized ligand molecules

Ligand No.	Docking score	M. Wt	Glide energy	QPlogHERG	QPlogS	CNS	QPlogPo/w	QPlogBB	ROT	ROF	HBD	HBA	%HOA
L5	-8.44	398.46	-58.45	-5.29	-5.03	-1	3.95	-0.79	0	0	2	5.5	100
L12	-7.81	411.52	-59.5	-6.22	-5.62	-1	3.87	-0.97	0	0	3	7.45	100
L6	-7.45	369.4	-35.16	-6.67	-5.12	-2	2.69	-1.55	0	0	1	7.5	84.68

L1	-7.4	464.55	-60.86	-6.28	-5.77	-1	4.83	-0.7	1	0	1	7	100
L11	-7.31	354.43	-44.01	-6.36	-5.52	-2	3.22	-1.17	0	0	3	5.75	93.27
L3	-7.24	489.58	-61.27	-4.77	-5.05	-2	3.21	-1.59	0	0	1.25	9.25	86.28
L7	-7.23	367.47	-56.87	-7.02	-5.95	-2	3.84	-1.11	1	0	3	6	100
L4	-7.18	442.62	-58.5	-5.54	-4.3	1	3.68	0.14	0	0	2	8.7	100
L8	-7.1	377.46	-53.66	-6.46	-6.95	-1	4.32	-0.85	1	0	1	5.75	100
L9	-7.01	436.57	-49.41	-5.89	-6.04	-1	4.21	-0.51	1	0	2	8.2	100
L2	-6.98	430.5	-58.26	-4.89	-4.62	0	4.11	-0.6	0	0	1	7	100
L10	-6.91	375.49	-46.45	-4.13	-2.98	-1	2.12	-0.95	0	0	0	8.2	86.68

Table 10 presents the ADME properties for L1 to L12 molecules predicted using the QikProp module of the Schrödinger suite are presented in table 10 with the highest priority to good docking score ligand. The ADME properties for the ligand molecules obtained from virtual screening are studied by using the QikProp module of the Schrödinger suite.

The permissible ranges are as follows: predicted Central nervous system (CNS): –2 (inactive) +2 (active); Molecular weight (Mol. wt.): (range 130–725); the number of hydrogen bond donors (HBD): (range 0.0–6.0); the number of hydrogen bond acceptors (HBA): (2.0–20.0); predicted octanol/water partition coefficient (QPlogPo/w), (range –2.0 to 6.5); QPlogHERG-Predicted IC50 value for blockage of HERG K+ channels (below -5.0); predicted brain–blood barrier partition coefficient (QPlogBB): (range –3.0 to 1.2); predicted human oral absorption (% HOA): >80% high, <25% low; the number of violations of Jorgensen's rule of three (ROT) (maximum 3); the number of violations of Lipinski's rule of five (ROF) (maximum 4).

Table No.11 The solubility properties of the ligand molecules.

Ligand No.	SASA	FISA	FOSA	PISA	PSA
L1	716.24	98.66	162.93	406.75	76.06
L2	706.81	70.53	221.86	414.42	78.43
L3	821.98	150.23	125.53	452.18	114.53
L4	724.57	59.94	498.31	118.91	70.04
L5	681.72	101.30	97.15	483.27	85.71
L6	679.59	173.85	131.66	342.59	122.49
L7	719.84	117.52	204.19	367.85	88.52
L8	702.98	98.89	266.23	305.34	79.99
L9	745.50	80.03	386.10	230.26	76.18
L10	691.78	113.83	407.94	156.83	81.00
L11	670.58	141.58	199.63	297.55	95.20
L12	729.24	111.43	281.32	286.02	85.71

The permissible range of the solubility properties presented in Table 11 are as follows:
Solvent Accessible Surface Area (SASA): 300 to 1000
FOSA Hydrophobic component of SASA: 0.0 to 750.0

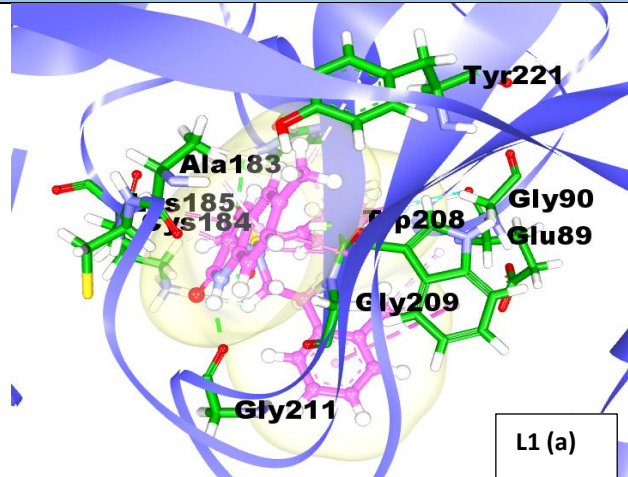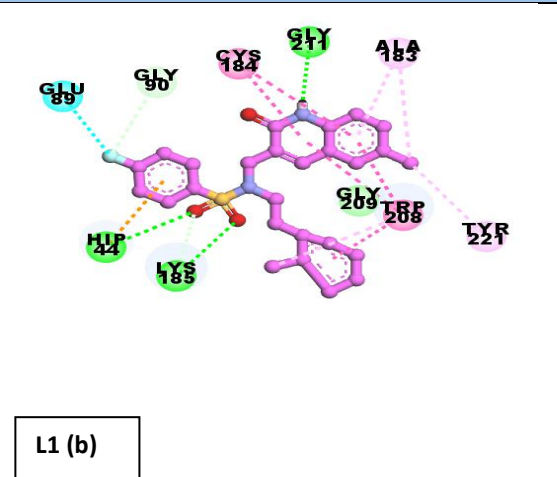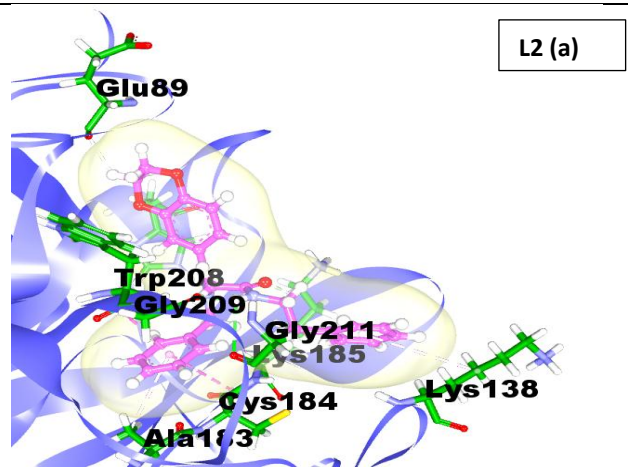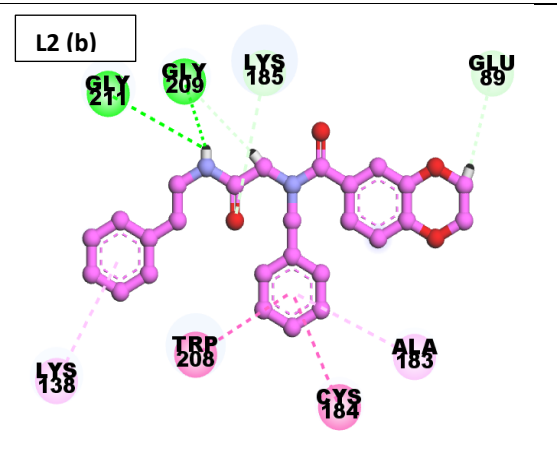FISA Hydrophilic component of SASA: 7.0 to 330.0

PISA π carbon and attached hydrogen component of SASA: 0.0 to 450.0

PSA Vander Waals surface area of polar nitrogen, oxygen, and carbonyl carbon atoms: 2 to 15

SIGNIFICANCE OF HIT MOLECULES

The final hit compounds with good binding energies from the MM-GBSA method had a high percentage of HOA, as shown in the table. L2, L4, L5, and L12 molecules had binding energies of -63.12, -59.41, -59.08, and -55.44 kcal/mol respectively, resulting in 100% oral absorption, respectively. All the docked complex structures of small molecule ligands with KLK B1 are depicted in Table 12. Among the above four lead compounds, L5 and L12 ligand molecules include an amide, a similar structural moiety.
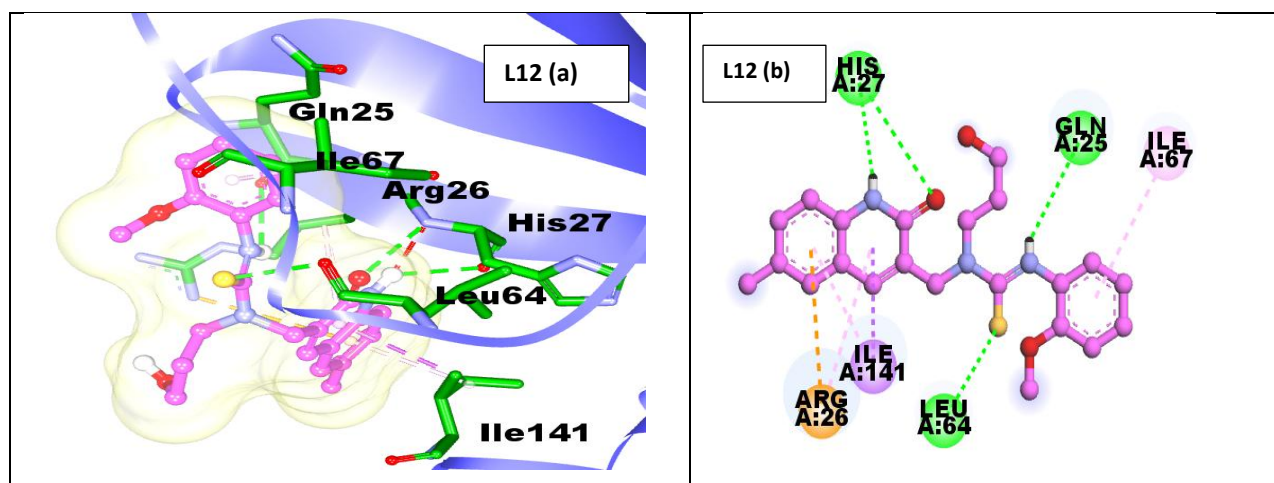
Table No.12 Docked complex structures of small molecule ligands (L1 to L12) with KLK B1 protein.

Binding interactions of KLK B1 with ligands	2D structure of ligand interactions with active site residues
L1 (a)	L1 (b)
L2 (a)	L2 (b)
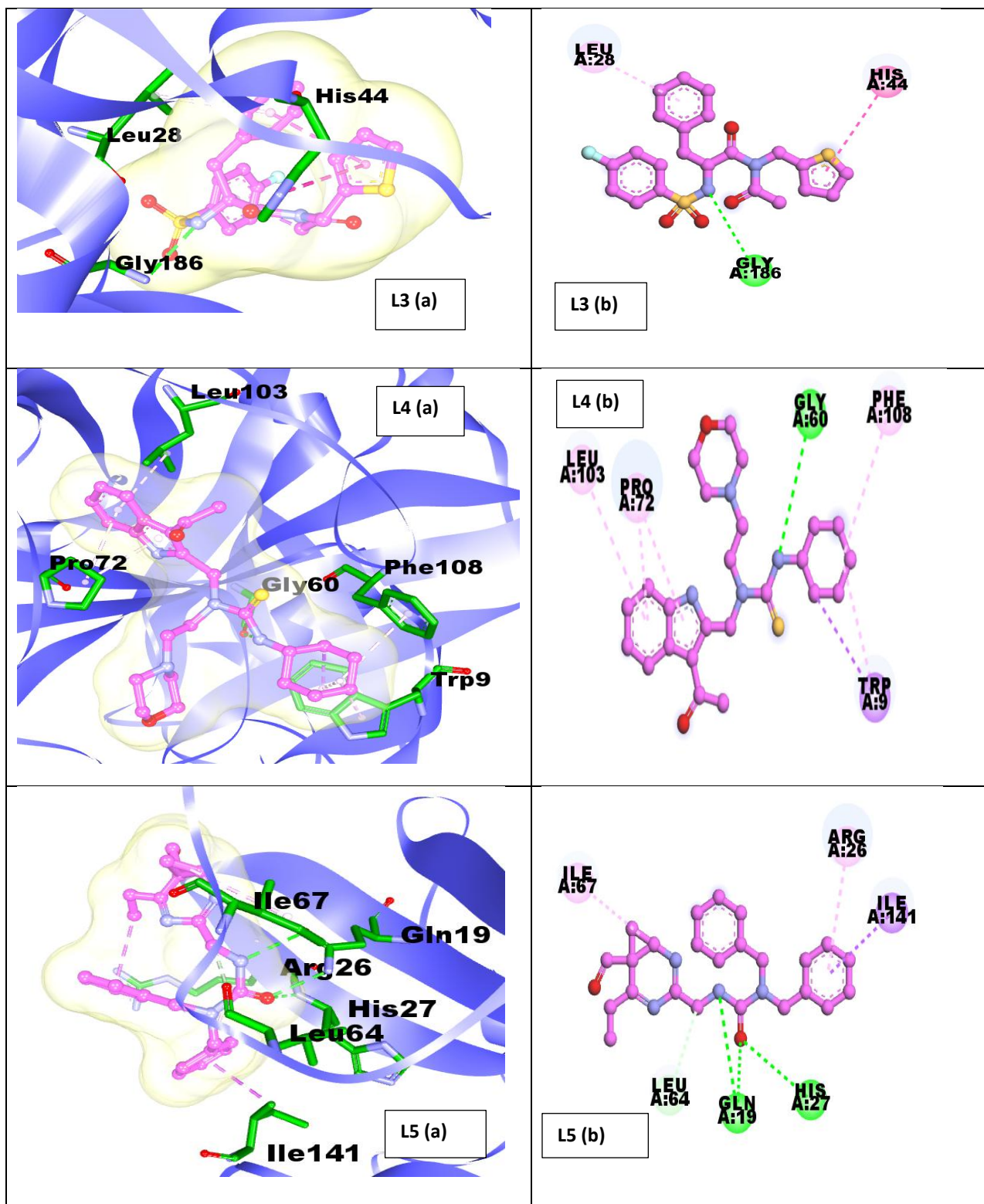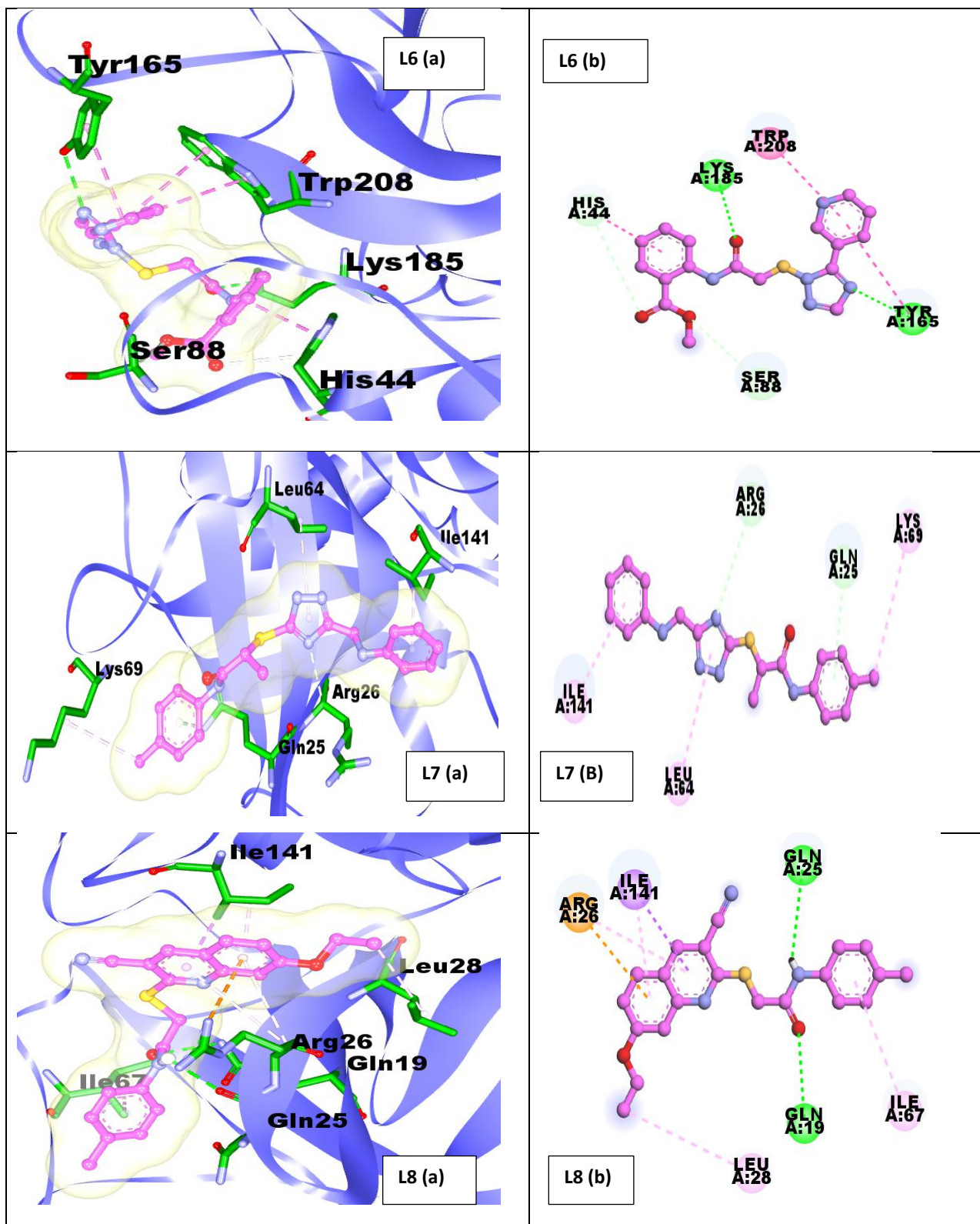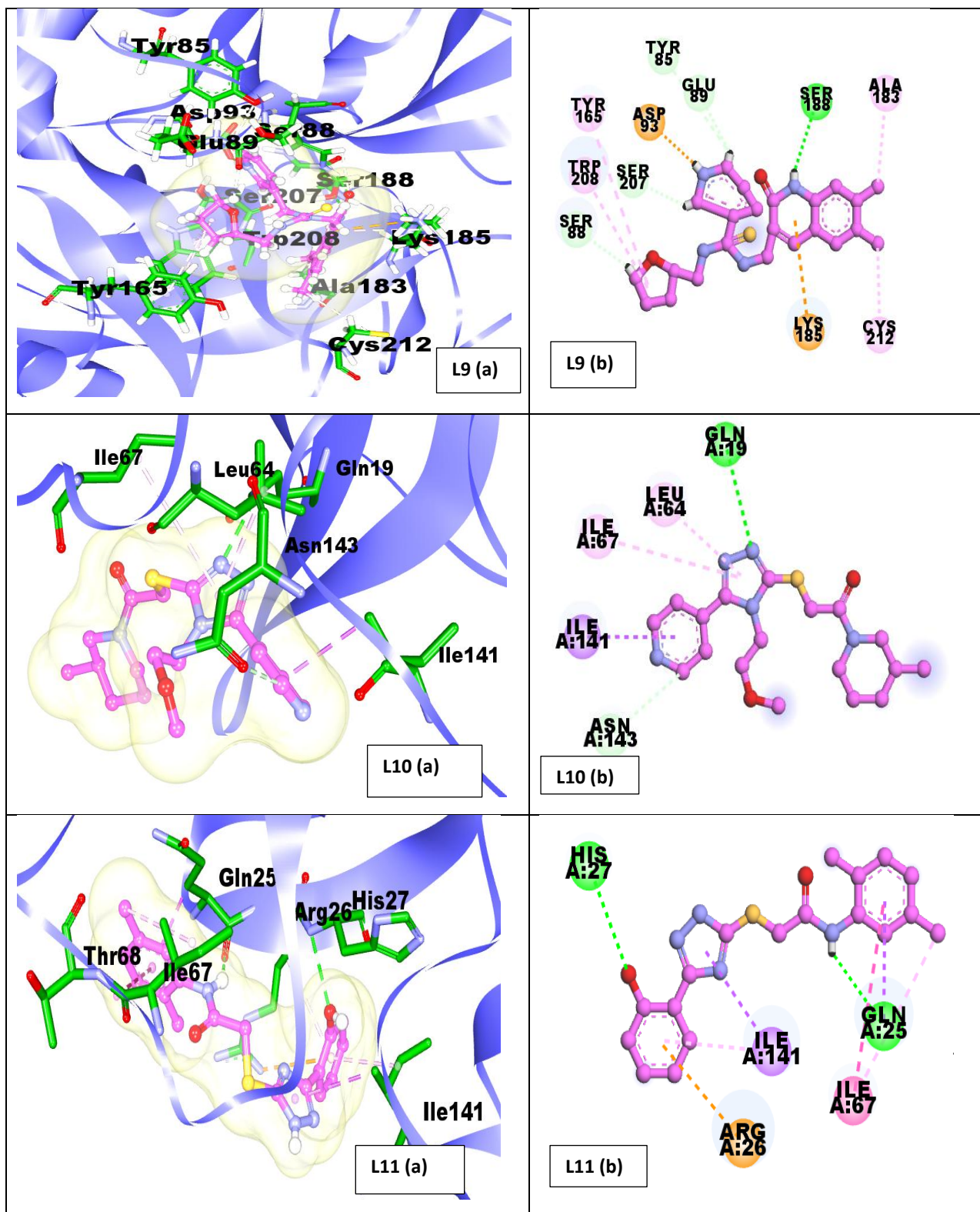
About 122734 small molecule ligands were screened based on physicochemical ADMET properties, QIKPROP, Lipinski filter, HTVS, SP, XP, and MMGBSA calculations out of which twelve ligands were shortlisted and presented in Table 12. The blue-colored flat ribbon represents the KLK B1 protein, the pink-colored ball, and stick-modeled structures are the small molecule ligand structures and the green-colored stick-modeled structures are the interacting amino acids of the protein with the ligand molecules. The cavity of the receptor is represented with a light yellow color.

Amides have emerged as potent inhibitors of the kallikrein family, a group of serine proteases linked to several physiological and pathological processes, including the development of cancer. By adhering themselves to the allosteric or active sites of kallikreins, these inhibitors prevent the proteolytic action of the protein. At a crucial stage in tumor invasion and metastasis, the breakdown of extracellular matrix components can be stopped by this inhibition. Studies have demonstrated the efficacious inhibition of kallikreins by certain amide-based drugs, especially in prostate cancer, where overexpression of kallikrein-related peptidases occurs.

Amphetamines inhibit the growth, invasion, and metastasis of cancer cells by going after these proteases. Apart from cancer, amide inhibitors of kallikreins may also be used to treat disorders involving aberrant blood coagulation and inflammatory illnesses.

CONCLUSION
According to the present study, KLK B1 protein represents a possible cancer treatment target. The 3D structure of KLK B1, which has 240 residues, was created by modelling. The quality of the modeled protein was confirmed utilizing ProSA and PROCHECK server tools. The Glide and AutoDock program demonstrated that L5 and L12 ligand molecules consistently bound to the amino acids of KLK B1 protein. These residues create interactions with ligand molecules, showing their importance in inhibiting the KLK B1 protein. The ligand molecules L5 and L12 exhibit 100% human oral absorption and efficient binding energies (-59.082 kcal/mol and -55.44 kcal/mol dG bind energy from Prime MMGBSA and -8.44 and -7.81 kcal/mol respectively from AutoDock Vina), resulting in unique protein-ligand combinations. Both of the docking investigations yielded comparable findings.

As a result, L5 and L12 might be regarded as the most notable hit compounds for cancer treatments that inhibit the KLK B1 protein. To maximize their clinical efficacy, reduce side effects, and improve selectivity and potency, amide-based kallikrein inhibitors must continue to be developed and optimized. Thus, this study offers a viable avenue for the creation of new cancer therapeutics linked to kallikreins.

ACKNOWLEDGMENTS

REFERENCES

[1]   Li S, Wang L, Sun S, Wu Q. Hepsin: a multifunctional transmembrane serine protease in pathobiology. Febs J. [Research Support, Non-U.S. Gov't Review]. 2021 Sep;288(18):5252-64.

[2]   Yarovaya GA, Neshkova EA. [Kallikrein-Kinin System. Long History and Present. (To 90th Anniversary of Discovery of the System)]. Bioorg Khim. [Review]. 2015 May-Jun;41(3):275-91.

[3]   Kashuba E, Bailey J, Allsup D, Cawkwell L. The kinin-kallikrein system: physiological roles, pathophysiology and its relationship to cancer biomarkers. Biomarkers. [Review]. 2013 Jun;18(4):279-96.

[4]   Bjorkqvist J, Jamsa A, Renne T. Plasma kallikrein: the bradykinin-producing enzyme. Thromb Haemost. [Research Support, Non-U.S. Gov't Review]. 2013 Sep;110(3):399-407.

[5]   Ivanov I, Verhamme IM, Sun MF, Mohammed B, Cheng Q, Matafonov A, et al. Protease activity in single-chain prekallikrein. Blood. [Research Support, N.I.H., Extramural]. 2020 Feb 20;135(8):558-67.

[6]   Di Paolo CT, Diamandis EP, Prassas I. The role of kallikreins in inflammatory skin disorders and their potential as therapeutic targets. Crit Rev Clin Lab Sci. [Review]. 2021 Jan;58(1):1-16.

[7]   Nauroy P, Nystrom A. Kallikreins: Essential epidermal messengers for regulation of the skin microenvironment during homeostasis, repair and disease. Matrix Biol Plus. 2020 May;6-7:100019.

[8]   Barco S, Sollfrank S, Trinchero A, Adenaeuer A, Abolghasemi H, Conti L, et al. Severe plasma prekallikrein deficiency: Clinical characteristics, novel KLKB1 mutations, and estimated prevalence. J Thromb Haemost. [Research Support, Non-U.S. Gov't]. 2020 Jul;18(7):1598-617.

[9]   Che YQ, Zhang Y, Li HB, Shen D, Cui W. Serum KLKB1 as a Potential Prognostic Biomarker for Hepatocellular Carcinoma Based on Data-Independent Acquisition and Parallel Reaction Monitoring. J Hepatocell Carcinoma. 2021;8:1241-52.

[10]   Park JE, Lim DS, Cho YH, Choi KY, Lee JJ, Kim BC, et al. Plasma contact factors as novel biomarkers for diagnosing Alzheimer's disease. Biomark Res. 2021 Jan 9;9(1):5.

[11]   Hameduh T, Haddad Y, Adam V, Heger Z. Homology modeling in the time of collective and artificial intelligence. Comput Struct Biotechnol J. [Review]. 2020;18:3494-506.

[12]   Dana H, Mahmoodi Chalbatani G, Gharagouzloo E, Miri SR, Memari F, Rasoolzadeh R, et al. In silico Analysis, Molecular Docking, Molecular Dynamic, Cloning, Expression and Purification of Chimeric Protein in Colorectal Cancer Treatment. Drug Des Devel Ther. 2020;14:309-29.

[13]   Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. Curr Protoc Bioinformatics. 2016 Jun 20;54:5 6 1-5 6 37.

[14]   Bitencourt-Ferreira G, de Azevedo WF, Jr. Homology Modeling of Protein Targets with MODELLER. Methods Mol Biol. [Research Support, Non-U.S. Gov't]. 2019;2053:231-49.

[15]   Pearson WR. Finding Protein and Nucleotide Similarities with FASTA. Curr Protoc Bioinformatics. [Research Support, N.I.H., Extramural]. 2016 Mar 24;53:3 9 1-3 9 25.

[16] Boratyn GM, Camacho C, Cooper PS, Coulouris G, Fong A, Ma N, et al. BLAST: a more efficient report with usability improvements. Nucleic Acids Res. [Research Support, N.I.H., Intramural]. 2013 Jul;41(Web Server issue):W29-33.

[17] Kim AK, Looger LL, Porter LL. A high-throughput predictive method for sequence-similar fold switchers. Biopolymers. 2021 Oct;112(10):e23416.

[18] Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. [Research Support, Non-U.S. Gov't]. 2007 Jul;35(Web Server issue):W407-10.

[19] Wlodawer A. Stereochemistry and Validation of Macromolecular Structures. Methods Mol Biol. [Review]. 2017;1607:595-610.

[20] Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des. 2013 Mar;27(3):221-34.

[21] Choudhury C. Fragment tailoring strategy to design novel chemical entities as potential binders of novel corona virus main protease. J Biomol Struct Dyn. [Research Support, Non-U.S. Gov't]. 2021 Jul;39(10):3733-46.

[22] Shelley JC, Cholleti A, Frye LL, Greenwood JR, Timlin MR, Uchimaya M. Epik: a software program for pK(a) prediction and protonation state generation for drug-like molecules. J Comput Aided Mol Des. 2007 Dec;21(12):681-91.

[23] Tian W, Chen C, Lei X, Zhao J, Liang J. CASTp 3.0: computed atlas of surface topography of proteins. Nucleic Acids Res. [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. 2018 Jul 2;46(W1):W363-W7.

[24] Halgren TA. Identifying and characterizing binding sites and assessing druggability. J Chem Inf Model. 2009 Feb;49(2):377-89.

[25] Sievers F, Higgins DG. The Clustal Omega Multiple Alignment Package. Methods Mol Biol. 2021;2231:3-16.

[26] Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. [Research Support, Non-U.S. Gov't]. 2018 Jul 2;46(W1):W296-W303.

[27] Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem. [Research Support, U.S. Gov't, P.H.S.]. 2004 Mar 25;47(7):1739-49.

[28] Sahayarayan JJ, Rajan KS, Vidhyavathi R, Nachiappan M, Prabhu D, Alfarraj S, et al. In-silico protein-ligand docking studies against the estrogen protein of breast cancer using pharmacophore based virtual screening approaches. Saudi J Biol Sci. 2021 Jan;28(1):400-7.

[29] Axenopoulos A, Daras P, Papadopoulos GE, Houstis EN. SP-dock: protein-protein docking using shape and physicochemical complementarity. IEEE/ACM Trans Comput Biol Bioinform. 2013 Jan-Feb;10(1):135-50.

[30] Elekofehinti OO, Iwaloye O, Josiah SS, Lawal AO, Akinjiyan MO, Ariyo EO. Molecular docking studies, molecular dynamics and ADME/tox reveal therapeutic potentials of STOCK1N-69160 against papain-like protease of SARS-CoV-2. Mol Divers. 2021 Aug;25(3):1761-73.

[31] Kemmish H, Fasnacht M, Yan L. Fully automated antibody structure prediction using BIOVIA tools: Validation study. PLoS One. [Validation Study]. 2017;12(5):e0177923.

[32] Wang E, Fu W, Jiang D, Sun H, Wang J, Zhang X, et al. VAD-MM/GBSA: A Variable Atomic Dielectric MM/GBSA Model for Improved Accuracy in Protein-Ligand Binding Free Energy Calculations. J Chem Inf Model. [Research Support, N.I.H., Extramural]. 2021 Jun 28;61(6):2844-56.

[33] Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov. [Research Support, Non-U.S. Gov't Review]. 2015 May;10(5):449-61.

[34] Lucas AJ, Sproston JL, Barton P, Riley RJ. Estimating human ADME properties, pharmacokinetic parameters and likely clinical dose in drug discovery. Expert Opin Drug Discov. [Research Support, Non-U.S. Gov't Review]. 2019 Dec;14(12):1313-27.

[35] Tibbitts J, Canter D, Graff R, Smith A, Khawli LA. Key factors influencing ADME properties of therapeutic proteins: A need for ADME characterization in drug discovery and development. MAbs. [Review]. 2016;8(2):229-45.

[36] Divyashri G, Krishna Murthy TP, Sundareshan S, Kamath P, Murahari M, Saraswathy GR, et al. In silico approach towards the identification of potential inhibitors from Curcuma amada Roxb against H. pylori: ADMET screening and molecular docking studies. Bioimpacts. 2021;11(2):119-27.

[37] Goodsell DS, Sanner MF, Olson AJ, Forli S. The AutoDock suite at 30. Protein Sci. [Research Support, N.I.H., Extramural]. 2021 Jan;30(1):31-43.

[38] Forli S, Huey R, Pique ME, Sanner MF, Goodsell DS, Olson AJ. Computational protein-ligand docking and virtual drug screening with the AutoDock suite. Nat Protoc. [Research Support, N.I.H., Extramural]. 2016 May;11(5):905-19.

[39] Zhao J, Cao Y, Zhang L. Exploring the computational methods for protein-ligand binding site prediction. Comput Struct Biotechnol J. [Review]. 2020;18:417-26.

[40] Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. J Chem Inf Model. [Research Support, N.I.H., Extramural]. 2021 Aug 23;61(8):3891-8.

[41] Nguyen NT, Nguyen TH, Pham TNH, Huy NT, Bay MV, Pham MQ, et al. Autodock Vina Adopts More Accurate Binding Poses but Autodock4 Forms Better Binding Affinity. J Chem Inf Model. [Research Support, Non-U.S. Gov't]. 2020 Jan 27;60(1):204-11.

[42] Kim T, Park H. Computational prediction of octanol-water partition coefficient based on the extended solvent-contact model. J Mol Graph Model. [Comparative Study Research Support, Non-U.S. Gov't]. 2015 Jul;60:108-17.

[43] Congreve M, Carr R, Murray C, Jhoti H. A 'rule of three' for fragment-based lead discovery? Drug Discov Today. [Comment]. 2003 Oct 1;8(19):876-7.

[44] Yim DS, Choi S, Bae SH. Predicting human pharmacokinetics from preclinical data: absorption. Transl Clin Pharmacol. 2020 Sep;28(3):126-35.

[45] Lipinski CA. Rule of five in 2015 and beyond: Target and ligand structural limitations, ligand chemistry structure and drug discovery project decisions. Adv Drug Deliv Rev. [Review]. 2016 Jun 1;101:34-41.

[46] Ali SA, Hassan MI, Islam A, Ahmad F. A review of methods available to estimate solvent-accessible surface areas of soluble proteins in the folded and unfolded states. Curr Protein Pept Sci. [Research Support, Non-U.S. Gov't Review]. 2014;15(5):456-76.