



## EXPLORING INNOVATIVE INHIBITOR CANDIDATES TARGETING THE RHOF PROTEIN: AN *IN-SILICO* INVESTIGATION.

Madhavi Latha Bingi<sup>1</sup>, Vani Kondaparthi<sup>1</sup>, Thirupathi Damera<sup>1</sup>, Priyadarshini Gangidi<sup>1</sup>,  
Hareesh Reddy Badepally<sup>1</sup>, Mounika Badineni<sup>1</sup>, Kiran Kumar Mustyala<sup>3</sup>, Vasavi Malkhed<sup>1,2\*</sup>

<sup>1</sup>Molecular Modeling Research Laboratory, Department of Chemistry, Osmania University,  
Hyderabad – 500007, Telangana, India.

<sup>2</sup>Department of Chemistry, University College of Science, Saifabad, Osmania University,  
Hyderabad – 500004, Telangana, India.

<sup>3</sup>Department of Chemistry, Nizam College, Osmania University, Hyderabad – 500001, Telangana,  
India.

**\*Corresponding Author:** Vasavi Malkhed

\*Department of Chemistry, University College of Science, Saifabad, Osmania University,  
Hyderabad – 500004, Telangana, India.

### Abstract

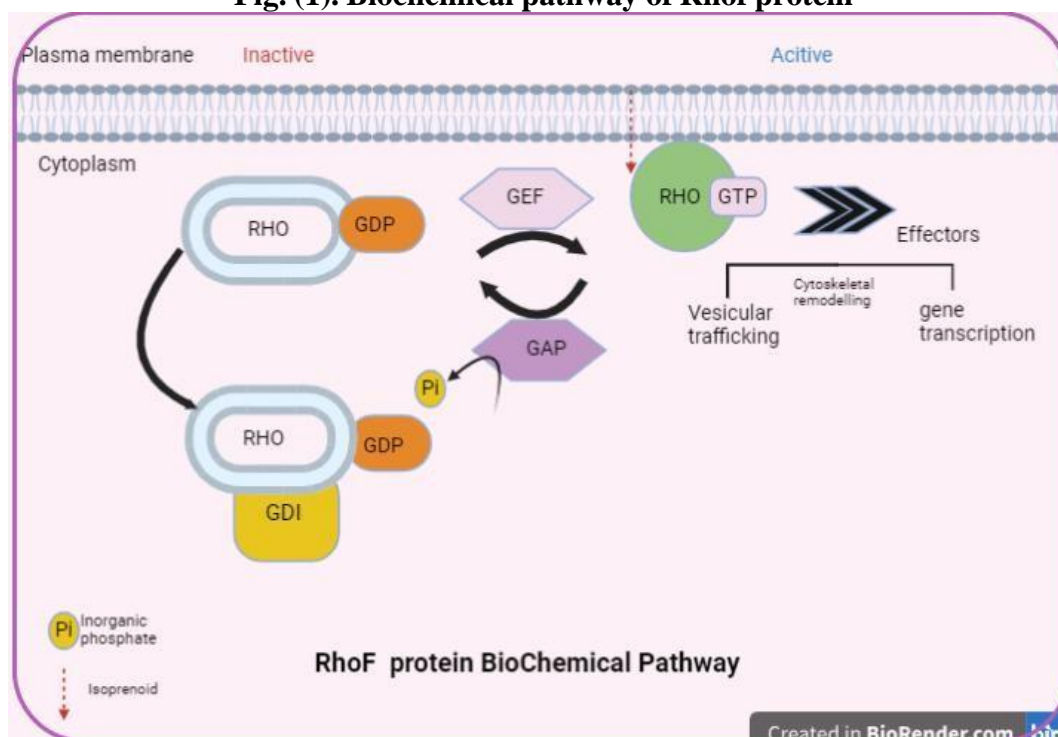
Cancer is a broad category of diseases characterized by untamed cell division and proliferation that damages nearby tissues and progresses to other parts of the body. The Rhof family, which is a subgroup of the broader Rhof family of tiny signaling G proteins, controls several physiological functions such as actin cytoskeleton, cell motility, and gene expression. Dysregulation or mutations in Rho GTPases can result in disorders such as cancer, as they are essential for cellular motility and proliferation. The present investigation assesses and validates the three-dimensional arrangement of the Rhof protein utilizing In-silico approaches. Virtual screening is underway to create potent inhibitors that selectively target the GTP interacting domain of the Rhof protein. The results of the research indicate that the amino acid residues ARG122, GLY123, ILE124 and PRO125, ARG166, ALA195, LEU196, LYS198, GLU200 ARG201, and LYS204 have strong interactions with ligands and significant influence on complex formation. Computational approaches reveal that the ADME features of screened ligand molecules suggest the optimum level of permissibility of drug-like properties. The utilization of Rhof structural information, active site details, and selected ligand molecules greatly aid in identifying new therapeutic structures for Rhof protein.

### 1. INTRODUCTION

Cancer refers to the uncontrolled proliferation of aberrant cells. Certain malignancies have the potential to metastasize towards distant tissues. Cancer develops when genetic alterations induce excessive growth and proliferation in a solitary cell or a limited number of cells, which can result in the formation of a malignant growth called a tumor <sup>[1]</sup>. According to the Global Cancer Statistics report by the American Cancer Society in 2020, there were an estimated 19.3 million new cases of cancer diagnoses in 2020 and 10 million mortality worldwide. Cancer affects one in five people, with lung cancer being the most frequently diagnosed and Primary cause of mortality. In 2020, lung cancer was responsible for 11.7 million new cases and 18% of all cancer deaths, followed by breast, colorectal, prostate, and stomach cancers <sup>[2]</sup>. Rhof /Rif protein, also known as Rhof protein,

is a small GTPase belonging to the Rho family. The Rho family is categorized into subfamilies such as Rho, Rac, Cdc42, Rnd, and RhoBTB. RhoF proteins are in the Rho family and with a structure that has not changed much over time with a GTP-binding domain and a lipid modification site at the very end<sup>[3]</sup>. GEF, GAP, and GRI regulate the activities of the Rho family of proteins, which play a role in signaling pathways, and their degradation can result in diseases such as cancer and neurological disorders. Aberrant overexpression or activity of RhoF has been associated with tumor progression, spread, and infiltration in cancer<sup>[4]</sup>. In addition, the involvement of RhoF in neuronal growth and synaptic plasticity within the brain has been suggested to regulate actin dynamics and cell responses<sup>[5]</sup>. RhoF facilitates the regulation of cell migration through coordination and fibro podia and recognition of actin cytoskeleton. GTPase-Activating Proteins (GAPs) and Guanine-nucleotide Dissociation Inhibitors (GDIs), render RhoF inactive by cleaving bound GTP to GDP and storing it in the cytoplasm depicted in Figure 1<sup>[6]</sup>. The involvement of RhoF in signaling networks facilitated by integrin's and growth factor receptors has been observed to influence many cellular responses, including migration, proliferation, and survival.<sup>[7]</sup> Thus targeting the pathway could potentially serve as a viable technique for cancer treatment. Although no inhibitors have been identified for the RhoF protein in current research, It has been considered a promising target.<sup>[8]</sup> Advances in new cancer therapies can be accelerated by utilizing methodologies such as three-dimensional structure prediction and the creation of innovative scaffold molecules. The three-dimensional conformation of the RhoF protein is evaluated and validated using *In-silico* techniques in this study<sup>[9]</sup>. Virtual screening studies are being performed to develop competitive inhibitors that specifically target the GTP binding region of the RhoF protein. The Primary objective of this investigation was to identify possible competitive inhibitors targeting the GTP binding domain of the RhoF protein<sup>[10]</sup>.

**Fig. (1). Biochemical pathway of RhoF protein**



RhoF protein activation involves Rho GTPases acting as switches between an inactive GDP-bound state and an active GTP-bound state. Guanine nucleotide exchange factors (GEFs) and GTPase-activating proteins (GAPs) control these transitions. When activated, Rho GTPases engage effector molecules to elicit responses. Post-translational modifications, such as prenylation, allow isoprenoid group attachment via Geranyl Geranyl Transferases (GGTases), which target proteins to the plasma membrane.

## 2. METHODOLOGY

Current drug discovery techniques have encountered significant limitations that need to be circumvented, making advances in computational chemistry indispensable. The primary purpose of this investigation is to identify Rhof as a novel target for cancer therapy<sup>[11]</sup>. At present, neither empirical nor theoretical approaches can extract the intricate, three-dimensional Rhof Structure. To evaluate and validate the Rhof three-dimensional model, computational methods, specifically *in silico* approaches were employed<sup>[12]</sup>. Rhof FASTA sequence is obtained in the first phase from sources such as Universal Protein Resource (UniProt)<sup>[13]</sup>. Proteins exhibiting comparable secondary structures, domains, and folds are subsequently identified as potential templates from sources such as BLAST and JPred4 servers.<sup>[14]</sup> Metrics such as the E-value are used to statistically quantify the degree of similarity between a target protein and its templates, in addition to assessing evolutionary conservation from sequence to structure. This approach facilitates the assessment of the reliability and accuracy of the developed three-dimensional Rhof protein, thereby creating opportunities for potential advancement in Cancer clinical trials<sup>[15]</sup>.

### 2.1. Alignment and model generation

The amino acid sequence of the template is aligned with the target protein sequence using the CLUSTALW tool, which facilitates the comparison of their structural similarities. After this alignment, 3D Rhof models are constructed using Modeller 9.12 software<sup>[16]</sup>. This software uses the CHARMM22 force field, an integral part of the accurate simulation of molecular interactions within protein structure. Several criteria, including the model's objective function, must be evaluated before the desired protein model is chosen for additional studies. This function determines the best-fit model for additional optimization and refinement by analyzing parameters including collision and energy scores, as well as stereo-chemical qualities<sup>[17]</sup>.

### 2.2. Energy minimization and validation

Energy optimization is a crucial phase to obtain the stability of the newly constructed 3D model. Enhancing the quality of a 3D model requires several important actions. Swiss PDB viewer 4.1<sup>[18]</sup> is utilized for energy reduction. Using the OPLS 2005 force field and an RMSD threshold of 0.3 Å, energy minimization is executed using the Impref component of the Schrodinger suite to preserve the structural integrity of the protein, this limiting range is necessary to maintain the carbon chain in its original configuration. Without affecting the carbon chain coordinates, the Impref module creates a near-minimum energy state conformation by rigidly maintaining the amino acid placement in the backbone while allowing flexibility in the side chains.

Molecular dynamics experiments using the online tool locPREFMD<sup>[19]</sup> allow further refinement of the model quality. The three-dimensional structure is optimized utilizing the Protein preparation Wizard feature of the Schrödinger suite using the OPLS force field. The server tools PROCHECK, ProSA<sup>[20]</sup>, and VERIFY 3D<sup>[21]</sup> are used to verify the model. Investigation of secondary structural features of the Rhof protein and putative active site regions is part of the analysis of its most stable conformation. The thorough approach improves the protein's 3D structure and aids in finding new inhibitors for drug development. The force field used for energy optimization in OPLS 2005 consists of selective equations and parameters designed to simulate the dynamics and interactions of molecules<sup>[22]</sup>. After the energy optimization procedure, further validation studies are performed on the resulting structure to verify its stability and suitability for further investigation<sup>[23],[24]</sup>.

### 2.3 Active site identification by *In-silico* methods

Drug designing research demands the determination of the active site of a protein which is crucial to understanding its specific biological activity<sup>[25]</sup>. The use of computational approaches is of fundamental importance in the prediction of probable binding sites in protein hydrophobic zones, which are predicted by CASTp and the SiteMap module of Schrödinger's suite<sup>[26],[27],[28]</sup>. The aforementioned analysis offers significant insight into the functional aspects of the protein and identifies possible lead molecules.

## 2.4. Docking of Rhof with GTP and Virtual Screening

AutoDock Vina is a widely used open-source software utilized in the field of computational biology and bioinformatics for molecular docking, an essential approach. The technique predicts the optional alignment of a tiny molecule when it is attached to a protein, essential for comprehending molecular interactions and developing novel medications<sup>[29]</sup>. Natural substrate binding was studied using AutoDock Vina.

The first step is to create a grid within the Rhof binding domain using the Glide module from Schrödinger's suite<sup>[30]</sup>. The grid is then used to perform virtual screening and docking studies. One part of the Schrodinger suite called LigPrep matches the stereochemistry, Ionization states, and ring conformations ligands from structural databases so that they can be virtually screened more easily. Glide software has different modes for virtual screening, namely High-throughput Virtual evaluation, Standard Precision (SP), and extra precision (XP), which are used to perform docking simulations<sup>[32]</sup>. After the docking process, the ligands go through a classification process that is determined based on their Glide score, providing valuable insights into their potential binding affinities. Comprehensive computational analysis facilitates an understanding of the interaction between novel scaffolds and Rhof, which is crucial for future drug discovery efforts<sup>[33]</sup>.

## 2.5. MMGBSA Calculations

The calculation of the free energy of the receptor-ligand complex is achieved by the MM-GBSA approach which was developed by Gleneden and Ryde in 2015. The Prime module in the Schrodinger framework, which is specifically tailored for MMGBSA calculations, is used to perform this estimation<sup>[34]</sup>.

The equation employed to calculate the binding free energy for each ligand works as follows:

The formula for  $G_{\text{bind}} = E_{\text{MM}} + E_{\text{sol}} + G_{\text{SA}}$

In this context, the term ' $E_{\text{MM}}$ ' refers to the disparity in minimized energy within the receptor-ligand combination and the aggregate energies of unbound receptor and ligands. The term ' $E_{\text{sol}}$ ' refers to the disparity in the solvation energy of the receptor-ligand compared to the combined energies of the receptor and the ligands. Finally,  $G_{\text{SA}}$  serves as a measure of the disparity in the surface energy between the receptor and ligands complex and the aggregated energies of unbound receptor and unbound ligands.

## 2.4. ADME Calculations

Assessment of the Absorption, Distribution, Metabolism, Elimination, and Toxicity (ADMET) properties is crucial to deciding the efficacy of lead compounds in both clinical and commercial contexts<sup>[35]</sup>. ADMET characteristics of ligand molecules extracted from virtual screening studies through the QikProp module of the Schrödinger suit<sup>[36]</sup>. The SwissADME web application is used to assess the toxicity and feasibility of synthesizing lead molecules. Promising lead candidates for cancer are selected based on their ADMET parameter values<sup>[37]</sup>.

Solubility characteristics of identified ligands such as FISA (hydrophobic part of SASA on N, O, H on heteroatoms) PISA (Carbon and associated hydrogen), and PSA values (Vander Waals surface of polar nitrogen and oxygen atoms) are determined using the QikProp module in the Schrödinger's suite<sup>[40]</sup>.

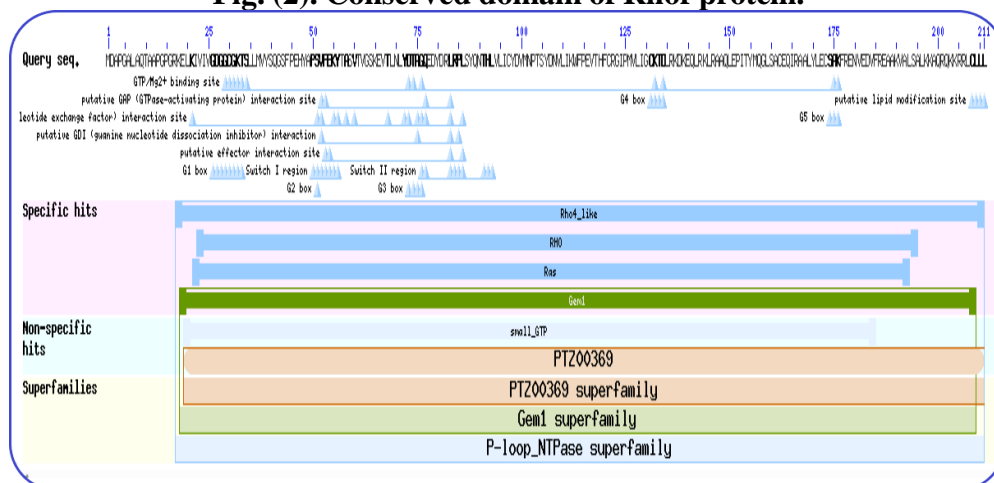
## 3. RESULTS AND ANALYSIS

### 3.1 Evaluation and verification of structure

This study retrieved the amino acid sequence of the Rhof protein from UniProtKB. The PSI-BLAST server submitted the amino acid sequence (FASTA format) of the Rhof protein in order to identify the conserved domain and potential templates.<sup>[39]</sup> Figure 2 depicts the conserved parts of the Rhof protein, emphasizing the importance of amino acid regions 26-32 considered  $\text{Mg}^{+2}$  binding regions, 26-33, 73-77, and 131-134 are considered GTP binding regions. The residues above are important in driving the protein's functioning and facilitating its interaction with GTP and magnesium ions. Utilizing bioinformatics techniques such as PSI-BLAST facilitates the

comprehension of the structural as well as functional characteristics of proteins, namely in the identification of conserved domains and crucial residues necessary for their biological activity.

**Fig. (2). Conserved domain of Rhof protein.**



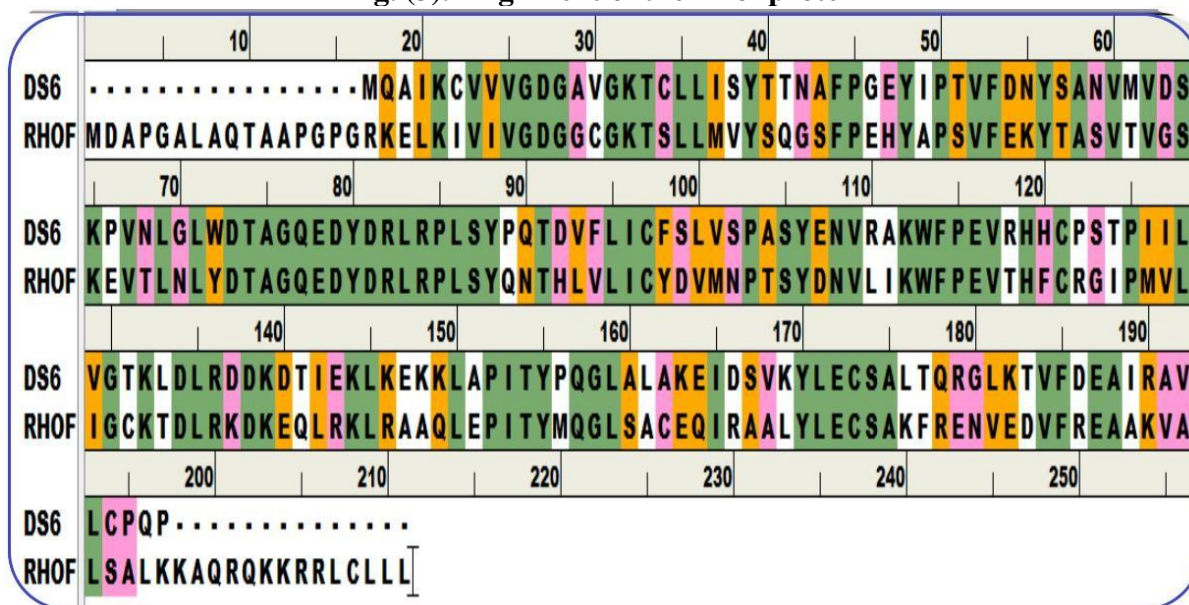
The GLY13, VAL14, GLY15, LYS16, SER17, PHE28, ILE29, and GLU30 residues are said to be involved in the protein's catalytic region, according to the graphical representation of the local similarity search results generated by BLAST.

Identification of the conserved Rhof protein's domain was performed using BLAST, Phyre2, and JPred3 servers. Based on parameters such as sequence similarity (54%), and statistical E- Value. Additionally query coverage as a parameter, the protein designated PDB ID: 1DS6 was chosen as the template. Predictions of the secondary structure and coiled-coil region of Rhof were predicted using the JPred3 server. Table 1 lists the E-values obtained from the different servers employed for template selection. Sequence alignments between Rhof and the template (1DS6) were performed using ClustalW as shown in Figure 3. Green represents identical residues, pink indicates highly similar, orange indicates similar, and pink indicates non- identical residues. 3D models of Rhof were created based on the PDB structure of the template protein (1DS6) employing Modeller 9.11 server, generating twenty-five models. The model exhibiting the lowest Modeller objective value was chosen for subsequent validation.

**Table 1. RhoF protein template identification**

S. No.	Template search tool	E- value	% of identity	PDB code of protein
1	BLAST	1e-65	50.26	1DS6-A
2	JPred3	3e-50	-	1DS6-A
3	Phyre2	100	51	1DS6-A



**Fig. (3). Alignment of the Rhof protein**

Pair-wise alignment of the Rhof protein and its template was performed using the Clustal W server program. The residues that are identical, strongly similar, and weakly similar are visually depicted using the colors green, orange, and pink respectively.

### 3.2 Energy minimization and structure validation

The Molecular dynamics simulations within PREFMD software allow researchers to accurately replicate the movements and interactions of atoms and molecules during the evolution of a protein's structure. To enhance the precision of the protein structure, simulations include several factors such as atomic forces, energy reduction, and sharp impacts significant progress in the stereochemistry of the Rhof protein, as shown in Table 2, which has been achieved through advancements in energetic and structural modifications. The model verification process involved using software tools Verify 3D, and PROCHECK. While SwissPDB viewer 4.0 was used to evaluate secondary structural features including  $\alpha$ -helices and  $\beta$ -strands. Figure 4 Tables 3a, 3b, and 3c provide comprehensive information regarding the 3D structure of the Rhof protein, revealing the existence of 9 $\alpha$ -helices, 5 helix-helix interactions and one  $\beta$ -sheet

**Table 2. List of the parameters for the Rhof protein that were molecular dynamics stabilized**

S.No.	Parameter checked	Initial value	Final value	Goal
1	Phi-Psi backbone favored region	93.5	95.2	>90%
2	Phi-psi backbone allowed region	5.4	3.8	-
3	Phi-Psi backbone general region	1.1	1.1	<1%
4	Phi-psi backbone disfavored region	0.0	0.0	< 0.2%
5	Phi-Psi backbone unallowed region	0.024	0.019	<0.2%
6	G-factor covalent bonds	0.19	0.2	>-0.5
7	G-factor overall interactions	-0.050	-0.060	>-0.5
8	Favorable main chain bond lengths	100	100.0	100%
9	Favorable main chain bond angles	95.5	95.5	100%
10	Side chain ring planarity	97.2	95.8	100%

**Table 3. Secondary structural information of Rhof protein.****a. Helices**

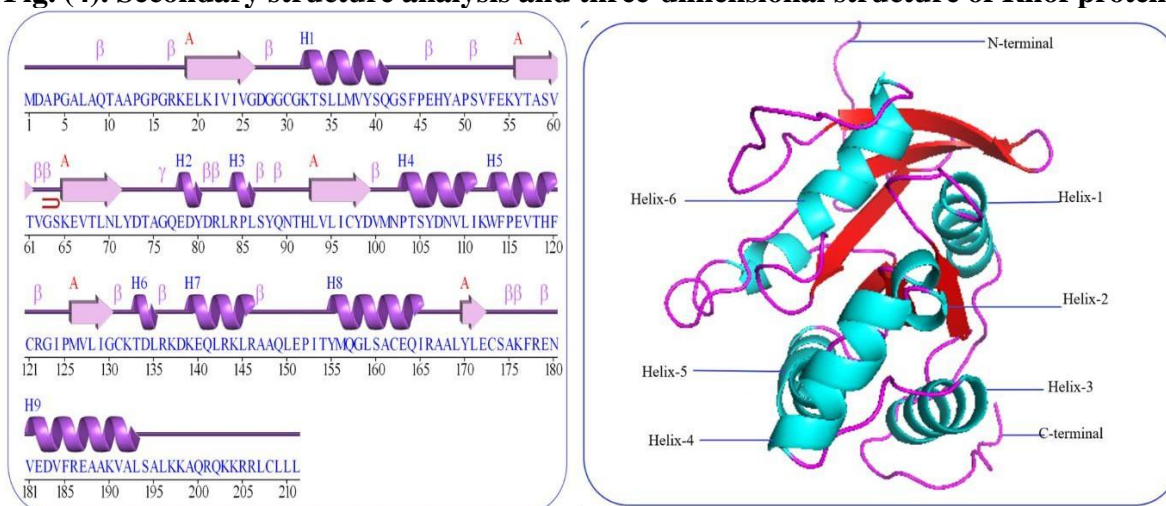
S.No	Start	End	Sheet	No. of residues	Sequence
1.	LYS32	GLN41	H	10	KTSLLMVYSQ
2.	GLU78	TYR80	G	3	EDY
3.	ARG84	LEU	G	3	RPL
4.	PRO103	ILE111	H	9	PTSYDNVLI
5.	TRP113	PHE120	G	8	WFPEVTHF
6.	THR133	LEU135	A	3	TDL
7.	LYS139	ARG146	H	8	KEQLRKLR
8.	TYR155	ILE165	H	11	YMQGLSACEQQI
9.	VAL181	LEU193	H	13	VEDVFREAAKVAL

**b. Beta sheets**

Sheet	No. of strands	Type	Barrel
A	6	Mixed	No

**C. Helix-helix interactions**

S.No.	Interactions between helices	Distance	Number of interacting residues
1	A1-A9	9.1	5
2	A4-A5	2.5	3
3	A4-A7	10.8	1
4	A4-A8	10.3	6

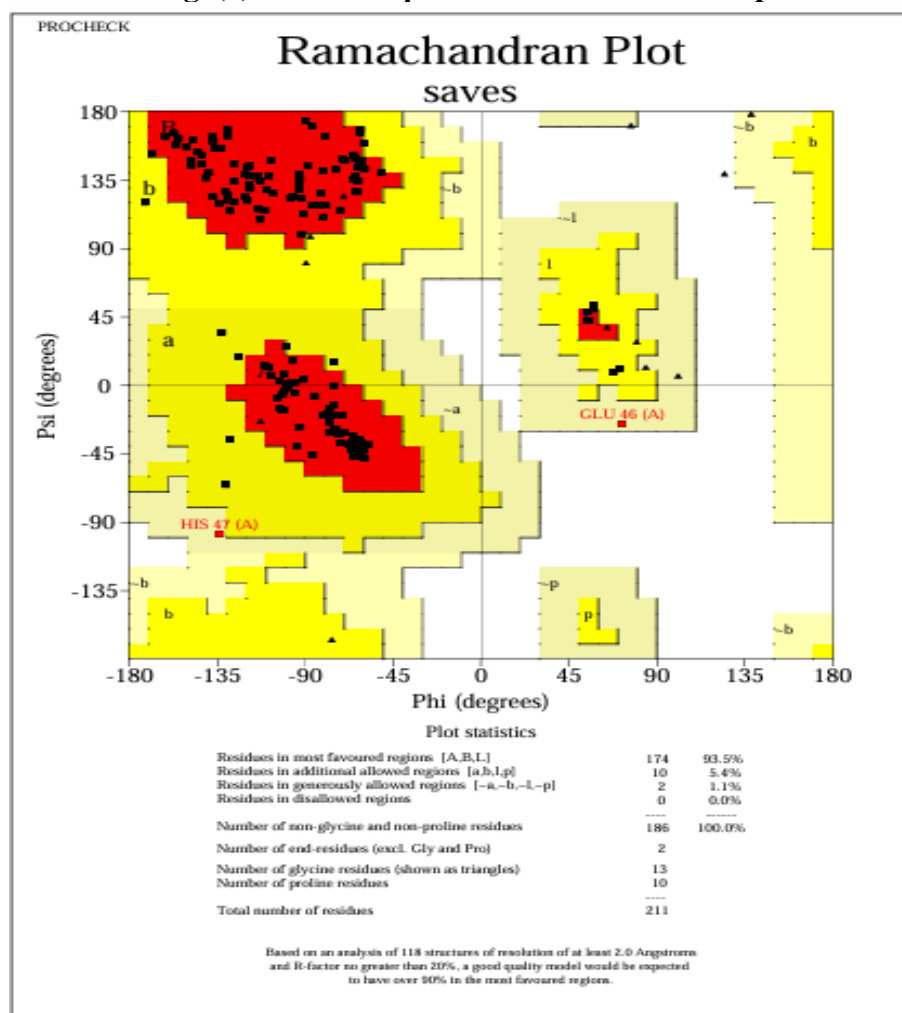
**Fig. (4). Secondary structure analysis and three-dimensional structure of Rhof protein.**

Two-dimensional configuration of the Rhof protein. The 2D structure of Rhof is generated using the PDBsum online server. It has nine beta turns, six beta strands, nine alpha helices, and one sheet.

The statistics of the Ramachandran plot for the Rhof protein are depicted in Figure 5, indicating that 94.6% of the residues are in the most favorable region and additionally permitted locations showing good stereochemical quality. The validity of the three-dimensional representation of the Rhof protein was confirmed through ProSA validation as evidenced by a Z-score of -5.44 (Figure 6a). A substantial portion of residues in the negative region suggests a well-designed model with favorable local energy. A thorough examination and validation of the three-dimensional model of the Rhof

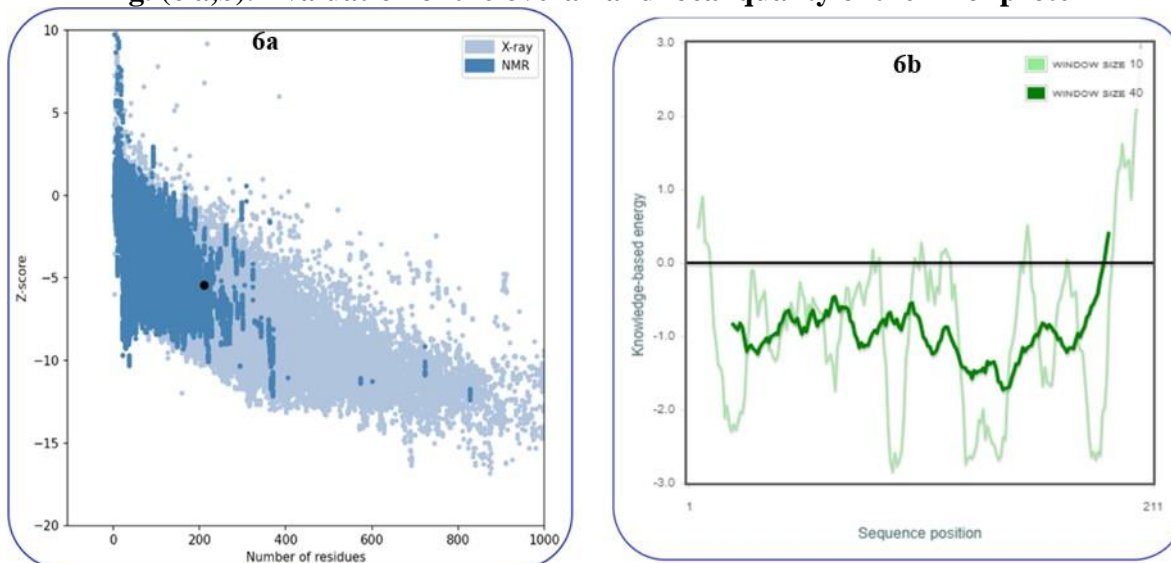
protein demonstrates its exceptional quality and reliability and creates a solid foundation for further investigation of its structure and function [40]. Proteins usually have properties that make a lot of their residues in the c-terminal domain consistently show strong folding energies that are lower than the baseline. The ProSA image (Figure 6b) shows a strong association between the characterized Rhof and native structure in most of the sequence.

**Fig. (5). The Rhof protein's Ramachandran plot**



The Ramachandran plot visually represents the plot of dihedral angles (Phi and Psi angles) of amino acids in a protein. The permitted area is in a vibrant yellow-blue, while the restricted area is in white, and the central area is in red.

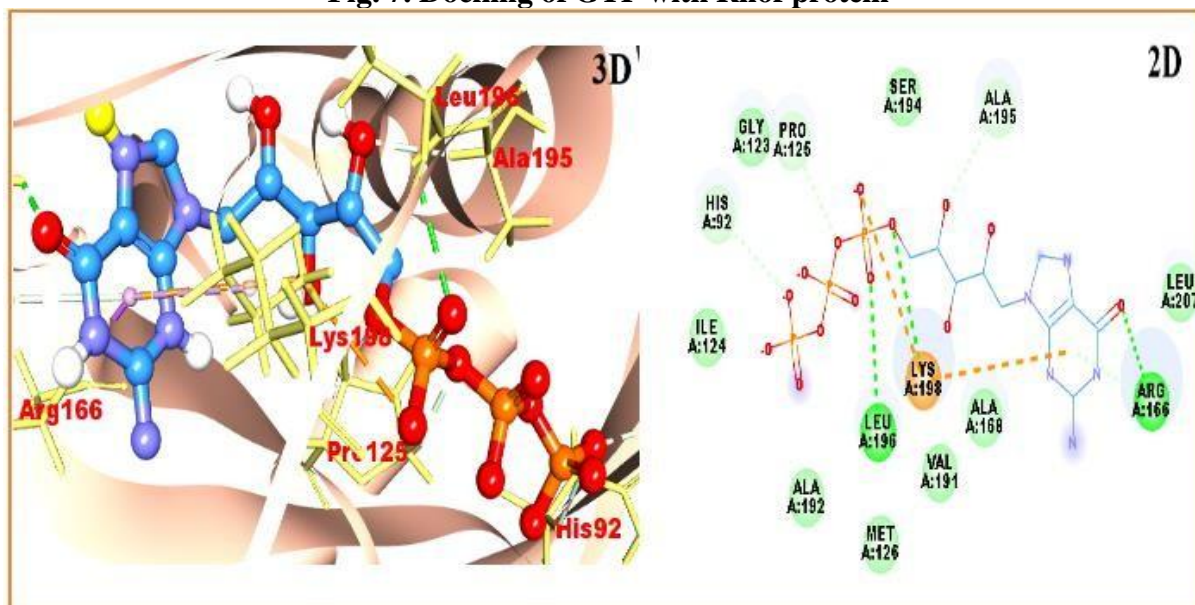


**Fig. (6 a,b). Evaluation of the overall and local quality of the Rhof protein**

ProSA (a) The total quality of the protein model is shown on the graph. The resulting protein structure is compared to the NMR and X-ray structures of the known protein. The X-ray and NMR calculated structural Z-scores are represented as light blue and dark blue patches, respectively. The total quality of the Rhof protein is shown by a black area in the NMR region with a Z-score of -5.44. In the first plot, a negative Z-score indicates that the protein's energy is lower than would be expected for a native-like fold, and ProSA (b) shows that a dark green line suggests a region that is favorable or stable.

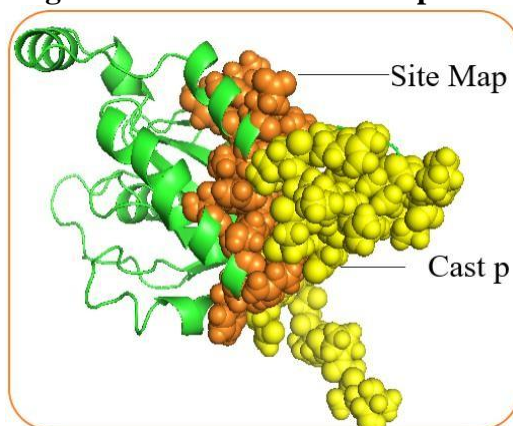
### 3.3 Active site identification Docking of Rhof and GTP

The Rhof protein was docked with its endogenous substrate, GTP, using the AutoDock Vina. The docking interaction between Rhof and GTP is shown in Figure 7. The Rhof protein and its natural substrate GTP were uploaded to the Auto Dock Vina where docking was performed. This process resulted in the production of eight docked complexes. Complexes that were docked were subsequently scored using their complementary geometry scores. Figure 7 shows the docked complex that emerged with Discovery Studio 3.5. It shows how hydrogen bonds are formed between the Rhof protein residues ASP27, GLY28, GLY29, LYS32, GLY76, GLN77, and GLU78.

**Fig. 7. Docking of GTP with Rhof protein**

Identification of the potential active site required using the binding region of the Rhof protein to its receptor as a reference point. Identification of active site residues and pocket volumes was performed using CASTp and Site Map discovery servers. The binding pockets in the Rhof proteins, as well as the respective cavity dimensions, are listed in Figure 8, Table 4. These cavities were identified by employing the CASTp and Sitemap servers. CASTp analysis identified two binding cavities, and the Site Map revealed the presence of two cavities shown in Table 4. The binding site of the Rhof protein consists of a sequence of amino acid residues, namely ASP27-LYS32, GLY76-GLU78, PHE114-MET126, LEU159- ALA168, and ARG201- LEU211. The binding site was identified as being located in a predicted region of the conserved domain.

**Fig. 8. Active site of the Rhof protein**



The hydrophobic area of the protein produced by the CASTp server is shown by the significant amino acid residues depicted in yellow spheres. The SiteMap provides data on binding pockets, with orange spheres representing the hydrophobic area of amino acid residues.

**Table 4: Active site prediction of Rhof protein**

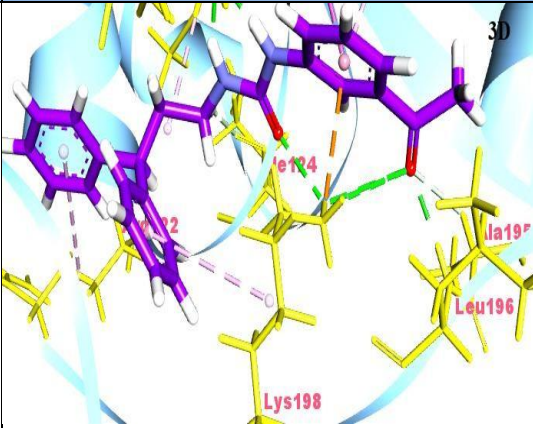
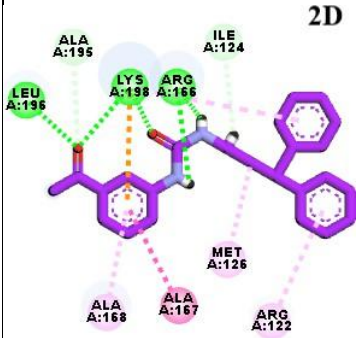
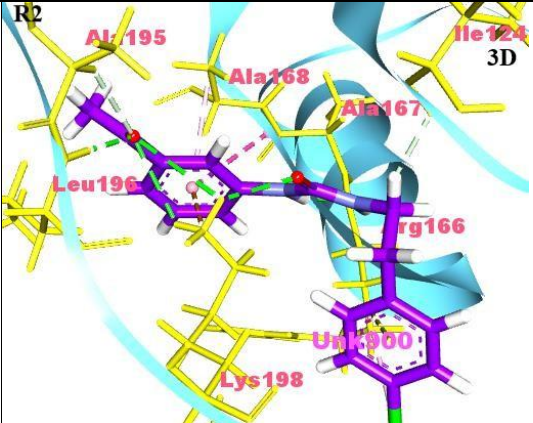
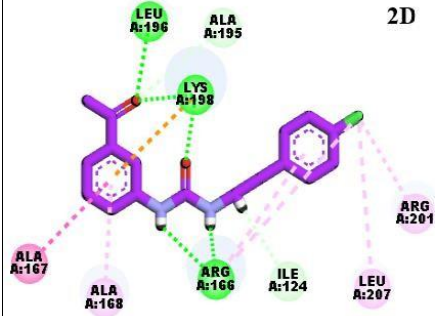
S.No	Active site prediction tool	Site No	Volume Å <sup>3</sup>	Amino acid residues
1	CASTp	1	1584.5	THR10, ALA11, ALA12, PRO13, GLY14, PRO15, GLY16, LYS18, GLU19, ASN90, THR91, HIS92, PHE114, THR118, PHR120, MET126, ILE165, ARG166, ALA167, ALA168, VAL191, ALA192, SER194, LEU196, LYS197, LYS198, ALA199, GLN200, ARG201, GLN202, LYS203, LYS204, ARG205, LEU207, LEU209, LEU211
		2	166.4	MET1, ASP2, PRO4, ALA12, GLY14, PRO15
		3	33.9	ASP27, GLY28, GLY29, GLY31, LYS32, THR33, SER34, ALA49, SER51, VAL52, ASP73, THR74, ALA75, GLY76, GLN77, TYR80
2	SiteMap	1	504.1	PHE114, THR118, HIS119, ARG122, GLY123, ILE124, PRO125, MET126, LEU159, CYS162, GLU163, ILE165, ARG166, ALA167, ALA168, LEU169, TYR170, GLU187, LYS190, VAL191, SER194, ALA195
		2	112.8	GLY28, GLY29, CYS30, GLY31, LYS32, THR33, SER34, ALA49, PRO50, SER51, VAL52, HRr74, ALA75, GLY76, GLN77, LYS132, VAL152, PHE53, GLU54, LYS55, TYR72, ASP73, THR74, ALA75, TYR80, LEU83, ARG84, SER87

The table lists the binding cavities found by applying different active site prediction techniques. Site map has identified two major hydrophobic zones, and CASTp has identified three hydrophobic sites

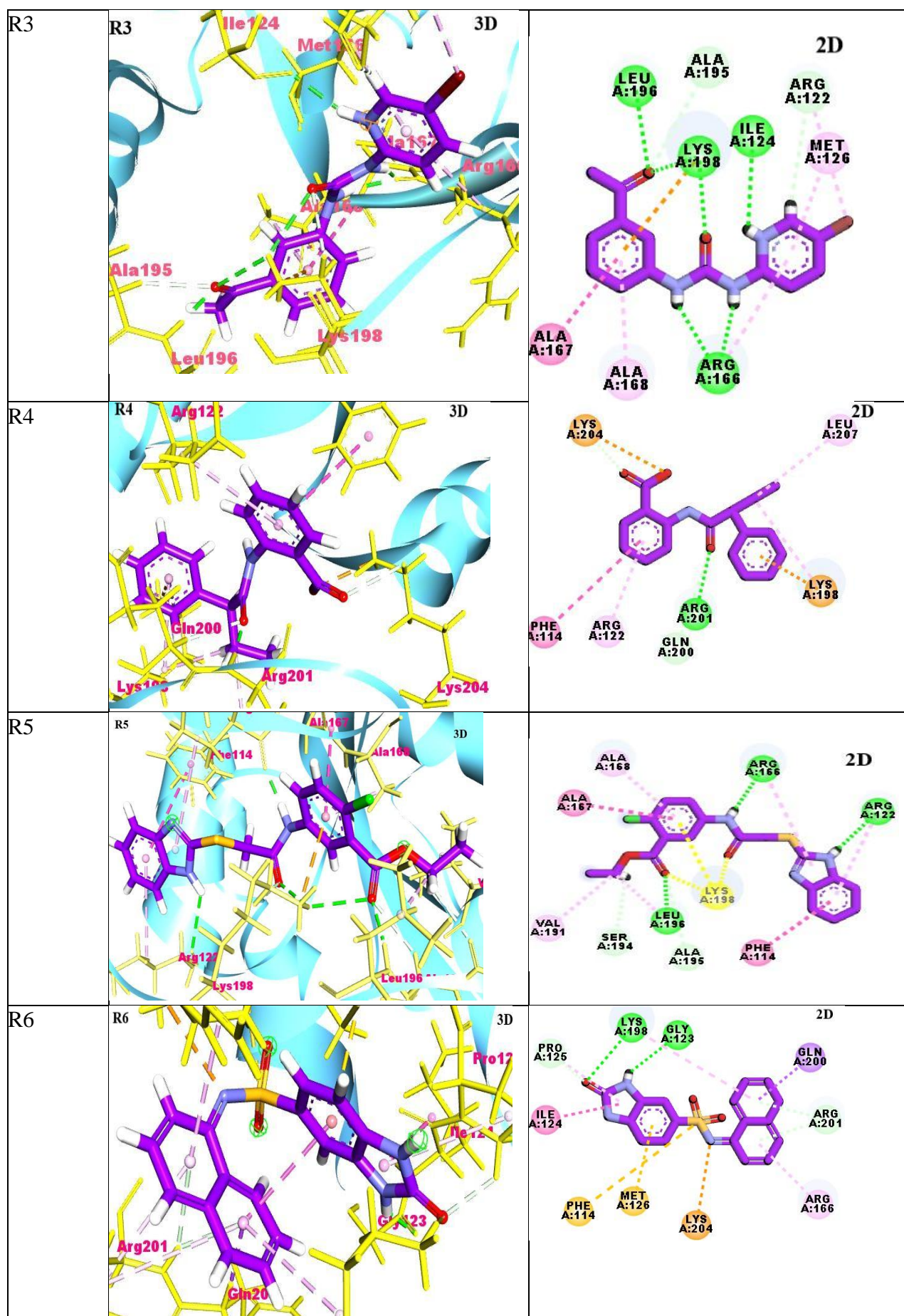
### 3.4 Virtual screening

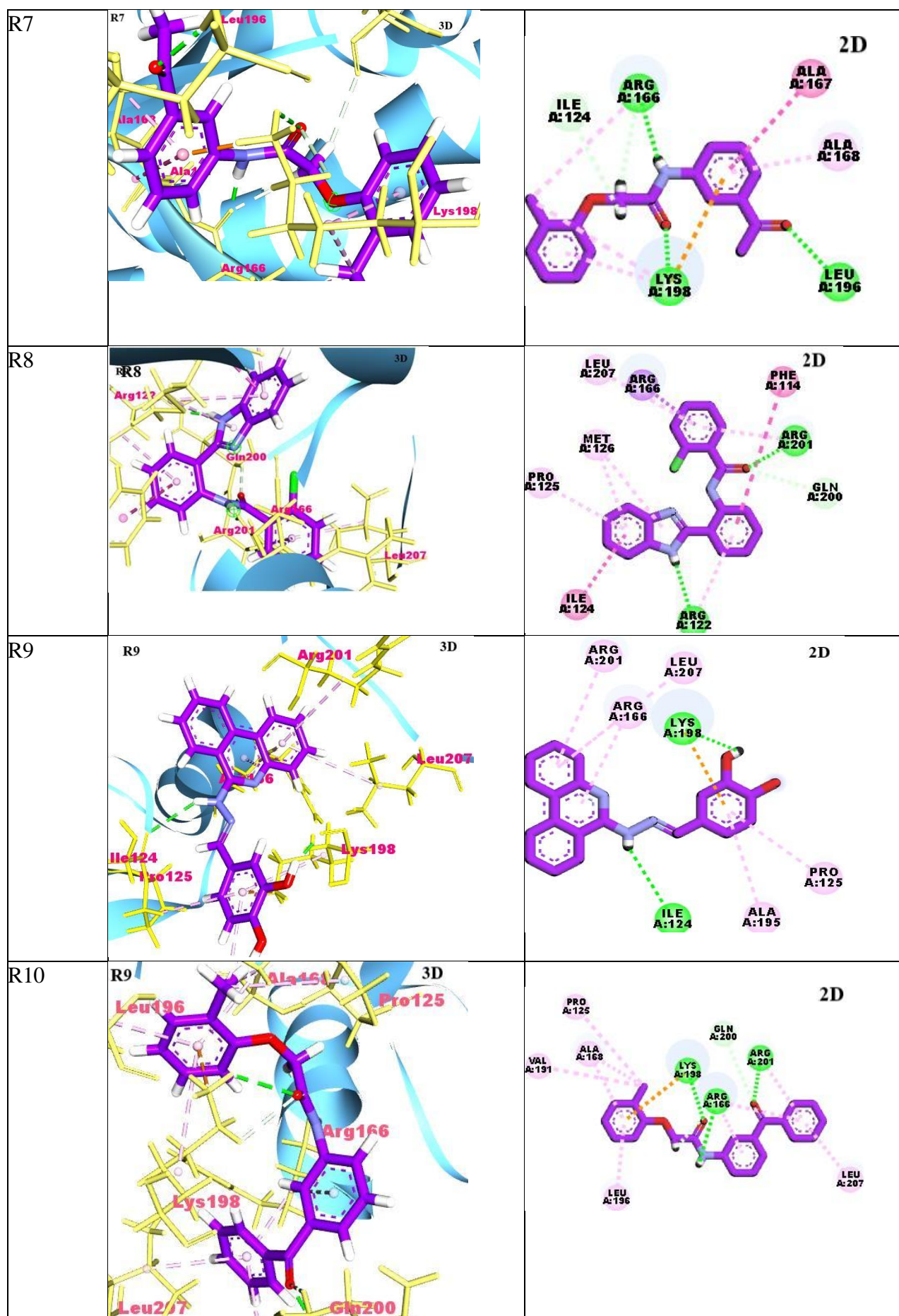
Virtual screening is an efficient methodology for identifying inhibitors that have innovative chemical scaffolds.<sup>[41]</sup> Approximately 35,000 compounds were screened and all were then retained for subsequent molecular docking. The conformation that recorded the lowest binding energy during docking was selected as the most likely binding conformation. The Rhof active site was docked with a total of 35,000 molecules that were tested. The docking energy of all compounds was expressed in kilocalories for each mole (k.cal/mol). Table 5 visually depicts the Rhof protein, small molecules, and their interactions with the active site region. The protein-ligand combination is stabilised by hydrogen bonding. The study shows that all small molecules contained in the Rhof protein's active site form hydrogen bond at the same site. The most efficient docking small molecules, namely R1, R2, R3, R5, R6, R7, and R8 have docking score values of -8.8, -7.69, -7.66, -7.38, -6.96, -6.89 and -6.86 k.cal/mol, respectively. The active area of the target protein exhibited strong binding affinity with the ligands and facilitated the formation of hydrogen bonds through ILE124, ARG166, ALA195, LEU196, and LYS198 amino acid residues respectively as depicted in Figure 9. Hydrogen bonding interactions are key to maintaining the stability of the complex<sup>[42]</sup>. Table 6 shows top 16 docked small molecules were visualized in 2D representation using the Discovery Studio visualizer.

**Table 5. Results of virtual screening**

Ligand No.	Interactions between the protein and ligand molecules	2D Structure of the ligand with the protein amino acids
R1		
R2		

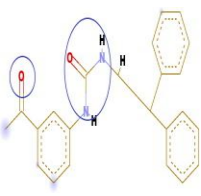
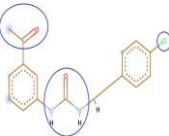
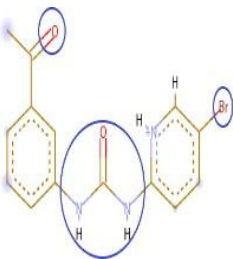
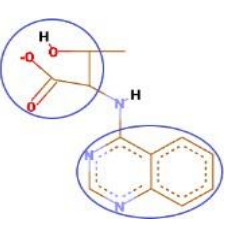
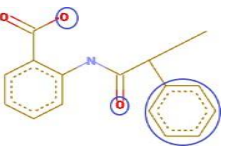
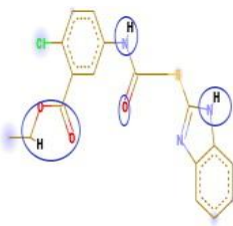




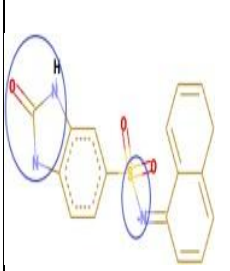
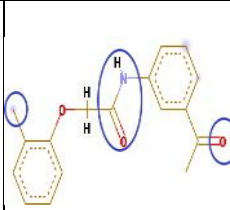
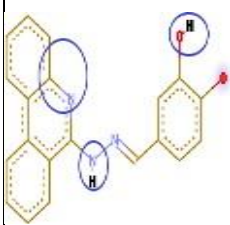
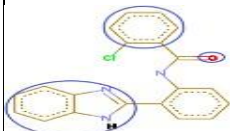




**Table 6. List of ligands selected in virtual screening.**

S. No.	Structure of the molecule	Molecular weight	H-bonding interactions	Glide Score	Glide energy	Prime MMG BSA $\Delta G$ Bind K.cal/mol	Free energy of ligand
R1		372.4	R1:O-LEU196:H- R1:O-LYS198:HZ2 R1:O-LYS198:HZ2 R1:O-LYS198:HZ3 R1:HARG166:OARG 166 R1:O-ALA195:HA	-8.81	-52.60	-59.90	-18.24
R2		316.78	R2: O-LEU196: R2:O-LYS198:HZ2 R2:O-LYS198:HZ2 R2:O-LYS198:HZ3 R2:OH-ARG166:O- R2:H-ARG166:HA R2:O-ILE124:O- R2:H	-7.698	-43.06	-48.36	-36.84
R3		334.17	R3:O10-LEU196 R3:O7-LYS198:HZ2 R3:O10-LYS198:HZ2 R3:H2O-A OARG166 R3:H21-A:ARG166 O R3:H33-A:O-LEU124 R3:O10-LA195:HA R3:H23-AO:ARG122	-7.66	-44.52	-35.68	-49.17
R4		247.2	R4:O30- LYS204:HZ1- R4:O13- GLN202:HE22 R4:H18-ARG201:O R4:H23-ARG201:O R4:H23-R4O30	-7.57	-31.59	-12.99	-31.59
R5		283.3	R5:O37-LYS204:NZ R5: O-ARG201: H R5: OGLN200: HA R5: O-LYS204:HE2	-7.45	-32.60	-25.05	8.48
R6		389.8	LEU196: H-R6: O LYS198:HZ2-R6: O LYS198:HZ2-R6:O ARG122:O- R6:H-A ARG166:OR6:H ALA195: HA-R6: O LYS198:HE2-R6: O SER194:O-R6:H	-7.38	-53.54	-55.74	-24.71



R7		339.3	LYS204:NZ-R7:N36 LYS198:HZ2-R7:O13 GLY123:O-R7:H24 PRO125:HD3-R7O13 ARG201:H-R7 ARG201:H-R7	-6.96	-24.49	-14.07	-51.07
R8		283.2	LEU196: H-R8: O LYS198:HZ2-R8: O ARG166:O-R8:H LYS198:HE2-R8: O ARG166:O-R8H ILE124:O-R8:H	-6.86	-40.56	-10.79	-5.72
R9		329.3	GLN202:HE22-R9: O ARG122:O-R9:H22 ARG122:O-R9H23 LYS198:O-R9:H24	-6.72	-43.16	-36.17	-34.88
R10		347.8	ARG201:H-R10:O ARG122:O-R10:H GLN200: HA-R10:O	-6.66	-38.86	-41.82	-8.97

The table displays the outcomes of docking and virtual screening conducted on the active site of the Rhof protein. By their glide score, ligands are organized in a descending order. The observation and analysis of the ligands interaction with the target protein Rhof were conducted.

### 3.5 Free energy calculation

The study analyzed protein-ligand complexes utilizing the Prime MM-GBSA component within Schrödinger software. <sup>[43]</sup> The binding free energy investigation revealed that the affinities of ten ligand molecules formed complexes with the Rhof protein were stable. The Rhof protein's active site domain had a 2.5-minute hydrogen bond, indicating structural stability <sup>[44]</sup>. High scores for ligands R1, R2, R3, R5, R6, R7 and R8 indicated thermodynamic feasibility and compatibility. Our research findings show that there was a substantial correlation and similarity between the binding free energy and the results achieved with Glide XP.

### 3.6 ADME properties

Evaluation of these compounds for further research on their ADME properties is worthwhile to aid in the development of novel oncological pharmaceuticals. Using the QikProp component of the Schrodinger suite, the ADME properties of the compounds were determined. Table 7 lists ADME characteristics for ligands (R1-R10). The visual representation of ADME characteristic values is grey for values that are within the acceptable range and pink for values that are outside the acceptable range with the exception of QPlog HERG values of R5, and R10, which are considered undesirable, all properties are shown in Table 7.

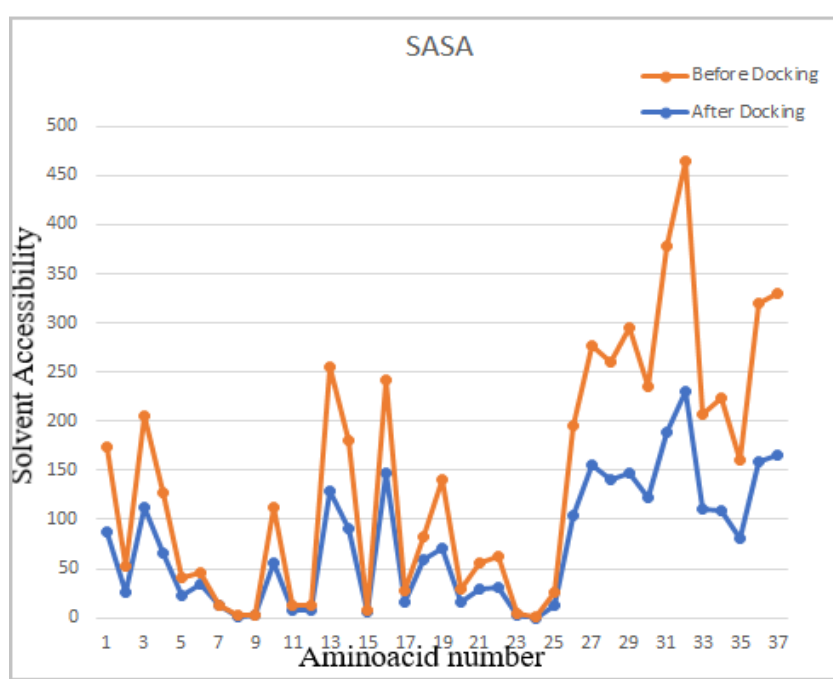
**Table 7. ADME parameters of the top-ranked ligands acquired from virtual screening**

Ligand No.	Molecular Weight	CNS	Donor HB	Accept HB	QPlog Po/w	QPlog HERG	QPlog BB	%Human Oral Absorption	Rule of five	Rule of Three
R1	372.4	-2	2.0	4.0	4.60	-6.15	-1.18	100	0	1
R2	316.7	-1	2.0	4.0	3.22	-4.18	-0.79	94.75	0	0
R3	334.1	-1	2.0	5.0	1.99	-4.54	-0.74	84.96	0	0
R4	283.3	-1	1.0	3.5	3.82	-3.62	-0.70	92.24	0	0
R5	389.8	-2	2.0	6.0	3.84	-6.74	-1.14	100	0	1
R6	339.3	-2	3.0	6.5	1.36	-5.23	-1.43	71.37	0	1
R7	282.3	-2	3.0	5.5	2.32	-1.66	-1.76	53.90	0	1
R8	283.3	-1	1.0	5.2	2.98	-6.03	-0.76	100	0	0
R9	329.3	-2	3.0	5.0	2.84	-6.57	-1.30	89.32	0	0
<b>R10</b>	349.8	0	2.0	4.0	4.38	-6.48	-0.04	100	0	0

The permissible ranges are as follows: Predicted central nervous system(CNS): 2(active)+-2(inactive), Predicted molecular weight:(130-725); Estimated Donor Hydrogen Bond(DHB) range: (0.0-0.6); Estimated water /octanol partition coefficient(QPlogPo/w): range (-2.0-6.5); Estimated IC50 value for blocking; QPlogHERG-Estimated IC50 for HERG channel blockage: (below6.18 to -5.35): Estimated Human Oral Absorption (%HOA): >80%high,<25%low; Estimated brain-blood barrier partition coefficient(QPlogBB): range-3.0to1.2; The number of times Lipinski's rule of five (ROF); the number of times Jorgensen's rule of three (ROT): (maximum 3)

### 3.7 SASA properties

The assessment of surface area accessible by solvent (SASA) values necessitated a direct evaluation of the surface accessibility of amino acids in the RHOF protein before and after the subsequent docking procedure. The study identified a reduction in SASA values for certain amino acid residues, namely ARG122, GLY123, ILE 124, PRO125, ARG166, ALA195, LEU 196, LYS198, GLN200, ARG201, and LYS204. The study findings demonstrate that the amino acid residues ARG122, GLY123, ILE 124, PRO125, ARG166, ALA195, LEU196, LYS198, GLN200, ARG201, and LYS204 played a substantial role in the docking process with the Rhof protein.both before and after their participation in ligand binding.



## Conclusion

This study created a three-dimensional representation of Rhof protein using modeler 9.11 software. The stable structure is then used for the docking method of the screened compounds. Subsequently; conventional approaches were used to enable the identification and characterization of the active part of the protein. Findings derived from virtual screening performed on an active site area of the Rhof protein indicate the existence of distinct amino acids known as ARG122, GLY123, ILE124, PRO125, ARG166, ALA195, IEU196, LYS198, GLN200, ARG201, LYS204. These amino acids serve a crucial function in the binding mechanism. The data presented in the study suggest that the ligands are more likely to interact with specific amino acid residues through hydrogen bonding. The ADME characteristic values are represented for values that fall within the acceptable range.

## Conflict of interest

The authors declare no conflict of interest.

## Acknowledgments

The authors, MLB, VK, TD, PG, HRB, and MB extend their heartfelt appreciation to the Principal and Head, Department of Chemistry at the University College of Science, Osmania University, for generously providing the necessary resources. Additionally, the authors would like to thank the Principal of the University College of Science, Saifabad, Osmania University, and the DST-FIST-sponsored Department of Chemistry, UCS, Saifabad, OU, for their facilities. The authors also acknowledge the resources established by the DST-FIST and UGC-UPE programs.

## References

- [1]. Ferlay J, Colombet M, Soerjomataram I, et.al. Global and Regional Estimates of the Incidence and Mortality for 38 Cancers: *GLOBOCAN 2018*. International Agency for Research on Cancer/World Health Organization; 2018.
- [2]. World Health Organization (WHO). *Global Health Estimates 2020: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2019*. WHO; 2020.
- [3]. Etienne-Manneville S, Hall A. Rho GTPases in cell biology. *Nature*. 2002 Dec 12;420(6916):629-35.
- [4]. Jung, H., Yoon, S. R., Lim, J., Cho, H. J., & Lee, H. G. (2020). Dysregulation of Rho GTPases in Human Cancers. *Cancers*.
- [5]. Senior, A.W., Evans, R., Jumper, J. *et al*. Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).
- [6]. Ellenbroek SI, Collard JG. Rho GTPases: functions and association with cancer. *Clin. Exp Metastasis*. 2007;24(8):657-72.
- [7]. Xu,H., Niu,M., Yuan,X. et.al. CD44 as a tumor biomarker and therapeutic target. *Exp. Hematol Oncol* 9,36 (2020)
- [8]. Hamidi, H., Ivaska, J. every step of the way: integrins in cancer progression and metastasis. *Nat Rev Cancer* 18, 533-548 (2018).
- [9]. Koledova Z. 3D Cell Culture: An Introduction. *Methods Mol Biol*. 2017;1612:1-11.
- [10]. Li Q, Shah S. Structure-Based Virtual Screening. *Methods Mol Biol*. 2017;1558:111-124.
- [11]. Lin, X., Li, X., & Lin, X. (2020). *A Review on Applications of Computational Methods in Drug Screening and Design*.
- [12]. Davidavičienė, V. (2018). Research Methodology: An Introduction. In: Marx Gómez, J., Mouselli, S. (eds) *Modernizing the Academic Teaching and Research Environment*. Progress in IS. Springer, Cham.
- [13]. The UniProt Consortium, UniProt: a worldwide hub of protein knowledge, *Nucleic Acids Research*, Volume 47, Issue D1, 08 January 2019, Pages D506–D515.
- [14]. Drozdetskiy, A., Cole, C., Procter, J., & Barton, G. J. (2015). JPred4: a protein secondary structure prediction server. *Nucleic Acids Research*.
- [15] Anderson, Nicole M.et.al, The tumor microenvironment, *Current Biology*, Volume 30,

Issue 16, R921 – R925

- [16]. Webb, B. & S. (2016). Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics*, . . . *Current Protocols in Bioinformatics*, 54(1).
- [17]. Andrew Waterhouse, Martino Bertoni, Stefan Bienert, Gabriel Studer, Gerardo Tauriello, Rafal Gumienny, Florian T Heer, Tjaart A P de Beer, Christine Rempfer, Lorenza Bordoli, Rosalba Lepore, Torsten Schwede, SWISS-MODEL: homology modelling of protein structures and complexes, *Nucleic Acids Research*, Volume 46, Issue W1, 2 July 2018, Pages W296–W303.
- [18]. Byrne, B. M. (2005). Factor Analytic Models: Viewing the Structure of an Assessment Instrument From Three Perspectives. *Journal of Personality Assessment*, 85(1), 17–32. [https://doi.org/10.1207/s15327752jpa8501\\_02](https://doi.org/10.1207/s15327752jpa8501_02)
- [19]. Brocke, S. A., Degen, A., MacKerell, A. D., Jr, Dutagaci, B., & Feig, M.. (2018). Prediction of Membrane Permeation of Drug Molecules by Combining an Implicit Membrane Model with Machine Learning. . *Journal of Chemical Information and Modeling*, 59(3), 1147–1162.
- [20]. Valckenaers, P. (2018). RTI Reference Architecture – PROSA Revisited. *SpringerLink*. [21]. Al-Refaei, M. A., Makki, R. M., & Ali, H. M. (2020). *Structure prediction of transferrin receptor protein 1 (TfR1) by homology modelling, docking, and molecular dynamics simulation studies*. 6(1(e03221).
- [22]. Banach M, Konieczny L, Roterman I. Secondary and Supersecondary Structure of Proteins in Light of the Structure of Hydrophobic Cores. *Methods Mol Biol*. 2019;1958:347-378.
- [23]. Thomas Simonson, Georgios Archontis, and Martin Karplus. Free Energy simulations Come of Age: Protein-Ligand Recognition. *Acc. Chem. Res*. 2002, 35, 6, 430–437.
- [24]. Monticelli, L., Tieleman, D.P. (2013). Force Fields for Classical Molecular Dynamics. In: Monticelli, L., Salonen, E. (eds) *Biomolecular Simulations. Methods in Molecular Biology*, vol 924. Humana Press, Totowa, NJ.
- [25]. Simone Brogi, Teodorico Castro Ramalho, Kamil Kuca, José L. Medina-Franco, Marian Valko. In silico Methods for Drug Design and Discovery. *Frontiers in Chemistry*, August 2020
- [26]. 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.
- [27]. Pavlo.Polishchuk. Interpretation of Quantitative structure- Activity Relationship Models: Past, Present, and future. *J. Chem. Inf. Model*. 2017, 57, 11, 2618–2639
- [28]. Lu, H., Zhou, Q., He, J. *et al*. Recent advances in the development of protein–protein interactions modulators: mechanisms and clinical trials. *Sig Transduct Target Ther* **5**, 213 (2020).
- [29]. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 2010 Jan 30;31(2):455-61.
- [30]. Pattar, S.V., Adhoni, S.A., Kamanavalli, C.M. *et al*. In silico molecular docking studies and MM/GBSA analysis of coumarin-carbonodithioate hybrid derivatives divulge the anticancer potential against breast cancer. *Beni-Suef Univ J Basic Appl Sci* **9**, 36 (2020).
- [31]. Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., & Sali, A. (2000). *Comparative Protein Structure Modeling of Genes and Genomes. Annual Review of Biophysics and Biomolecular Structure*,. 29(1), 295–325.
- [32]. Prajapat, M., Shekhar, N., Sarma, P., Avti, P., Singh, S., Kaur, H., Bhattacharyya, A., Kumar, S., Sharma, S., Prakash, A., & Medhi, B. (2020). Virtual screening and molecular dynamics study of approved drugs as inhibitors of spike protein S1 domain and ACE2 interaction in SARS-CoV-2. *Journal of Molecular Graphics and Modelling*,. *Journal of Molecular Graphics and Modelling*, 101(107716).
- [33]. Forli, S., Huey, R., Pique, M. *et al*. Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nat Protoc* **11**, 905–919 (2016).
- [34] Wang E, Liu H, Wang J, Weng G, Sun H, Wang Z, Kang Y, Hou T. Development and

- Evaluation of MM/GBSA Based on a Variable Dielectric GB Model for Predicting Protein-Ligand Binding Affinities. *J Chem Inf Model*. 2020 Nov 23; 60(11):5353-5365.
- [35]. Yim DS, Choi S, Bae SH. Predicting human pharmacokinetics from preclinical data: absorption. *Transl Clin Pharmacol*. 2020 Sep; 28(3):126-135.
- [36]. 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*. 2019 Dec; 14(12):1313-1327.
- [37]. Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep*. 2017 Mar 3;7: 42717.
- [38]. Gimeno A, Ojeda-Montes MJ, Tomás-Hernández S, Cereto-Massagué A, Beltrán-Debón R, Mulero M, Pujadas G, Garcia-Vallvé S. The Light and Dark Sides of Virtual Screening: What Is There to Know? *Int J Mol Sci*. 2019 Mar 19; 20(6):1375.
- [39]. Pundir S, Martin MJ, O'Donovan C; UniProt Consortium. UniProt Tools. *Curr Protoc Bioinformatics*. 2016 Mar 24; 53: 1.29.1-1.29.15.
- [40]. Wright GJ. Signal initiation in biological systems: the properties and detection of transient extracellular protein interactions. *Mol Biosyst*. 2009 Dec; 5(12):1405-12.
- [41]. Vadija, R., Mustyala, K. K., Nambigari, N., Dulapalli, R., Dumpati, R. K., Ramatenki, V., Vellanki, S. P., & Vuruputuri, U. (2016). *Homology modeling and virtual screening studies of FGF-7 protein—a structure-based approach to design new molecules against tumor angiogenesis*. . 9(3), 69–78. T.
- [42] Chen, D., Oezguen, N., Urvil, P., Ferguson, C., Dann, S.M., & Savidge, T.C. (2016). Regulation of protein-ligand binding affinity by hydrogen bond pairing. *Science Advances*, 2. (2016). Regulation of protein-ligand binding affinity by hydrogen bond pairing. *Science Advances*, 2(3).
- [43]. Perumalsamy, H., Sankarapandian, K., Veerappan, K., Natarajan, S., Kandaswamy, N., Thangavelu, L., & Balusamy, S. R. (2020). In silico molecular docking studies and MM/GBSA analysis of coumarin-carbonodithioate hybrid derivatives divulge the anticancer potential against breast cancer. *Beni-Suef University Journal of Basic and Applied Sciences*.
- [44]. Zoe Cournia, Bryce Allen, and Woody Sherman. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *Journal of Chemical Information and Modeling* **2017** 57 (12), 2911-2937.