



## Identification of Potential Inhibitors against TMPRSS4 through Molecular Docking and Dynamics Simulations

Poornima Eeram<sup>1</sup>, Dr. Premkumar Arumugam

<sup>1</sup>Reagent Generation, Aragen Lifesciences Private Ltd., Survey No.125(p) & 126, IDA Mallapur, Hyderabad-500 076, India. [Poornima.esram@aragen.com](mailto:Poornima.esram@aragen.com)

<sup>2</sup>4, ARANI Main Road, OTTERI, Vellore-2 Tamil Nādu- 632107, India. [aprem70@yahoo.com](mailto:aprem70@yahoo.com)

**Corresponding Author:** Poornima Eeram  
[Poornima.esram@aragen.com](mailto:Poornima.esram@aragen.com)

---

### Abstract:

TMPRSS4, a member of the transmembrane serine protease (TTSP) family, is notably overexpressed in various cancer types, including pancreatic, ovarian, thyroid, colorectal, lung, breast, cervical, gall bladder, gastric, and liver cancer. This suggests its potential as an innovative therapeutic target for solid tumors. However, despite its relevance in cancer research, limited structural information exists for TMPRSS4. In this pioneering study, we present the first 3-dimensional models of the extracellular domain (ECD) and catalytic domain (CD) of human TMPRSS4, crucial for rational drug design. Employing in-silico methods, we constructed these models and evaluated the compound CID 5770901 through molecular docking and molecular dynamics (MD) simulations. Our findings demonstrate the compound's remarkable efficacy in inhibiting both TMPRSS4 domains, with a calculated docking score of -9 kcal/mole. MD simulations revealed stable complex formation, with an RMSD value of approximately 3.8 Å for ECD-5770901 and minimal fluctuations in RMSF values for CD. These results highlight the potential of tailored inhibitors for TMPRSS4 as a promising avenue in cancer therapy, though further research is necessary.

**Keywords:** TMPRSS4; Molecular docking; MD simulations; TMPRSS4 inhibitors; Cancer therapeutic targets.

---

**DOI: 10.31838/ecb/2023.12.7.353**

### INTRODUCTION

Proteolysis, a pivotal regulatory mechanism, hinges on the precise cleavage of peptide bonds by proteases, enabling its multifaceted involvement in diverse biological activities. Proteases or proteolytic enzymes are of paramount importance in various array of biological processes, exerting significant influence through both direct and indirect mechanisms. This highlights their significance in maintaining overall health as well as their involvement in the development of many disorders [1-3].

The dysregulation of proteolysis has been identified as a contributing factor in the development of various diseases, such as arthritis, cancer, cardiovascular disorders, and neurological disorders [1-3]. Proteases are classified into different categories based on the processes of catalysis, including serine, aspartyl, threonine, metallo, and cysteine proteases [3]. This extensive protein family exhibits a distribution pattern that encompasses extracellular, cytoplasmic, and organelle-localized locations, including the lysosome. A select subgroup of serine proteases exhibits a distinctive characteristic: the presence of a transmembrane domain, which enables their anchoring to the cell's plasma membrane. This intriguing feature underpins the significance of proteolysis at the cell surface, where it plays an essential role in the generation of biologically active molecules that orchestrate a wide spectrum of essential

cellular functions. Serine proteases can be categorized into three distinct groups based on their transmembrane domain structure. The first group, termed Type I, is characterized by a carboxy-terminal domain. The second group, known as TTSP or Type II, exhibits an amino-terminal domain. The third group, referred to as GPI, anchors to the membrane via glycosylphosphatidylinositol [4].

TMPRSS4, previously identified as TMPRSS3 according to Wallrapp et al., (2000), is a distinctive type II transmembrane serine protease located on the long arm of chromosome 11 (11q23.3).

"TMPRSS4, transmembrane in nature is localized on the cell membrane and displayed significant expression in tissues affected by gastric, pancreatic, and colon cancers. While current data on TMPRSS4 remain limited, multiple lines of evidence strongly indicate its involvement in promoting metastasis, as demonstrated in preclinical models. Notably, numerous studies have highlighted TMPRSS4's overexpression in various cancer types relative to normal tissues, alongside a correlation between elevated TMPRSS4 levels and reduced disease-free survival (DFS) and overall survival (OS). Given that the majority of available research on TMPRSS4 primarily revolves around cancer genesis and metastasis, it is compelling to propose that this membrane-tethered serine protease holds promise as an innovative therapeutic target for solid tumors. Furthermore, the presence of TMPRSS4 fragments in conditioned media and serum samples from cancer patients underscores its potential clinical relevance as well [1]. Nonetheless, to fully unlock the therapeutic potential offered by TMPRSS4, further investigation and scholarly inquiry are imperative.

Molecular docking has emerged as a pivotal tool in the realm of structural molecular biology and computer-aided drug design. Within this domain, virtual screening and lead optimization have become widely adopted techniques, propelling advancements in drug discovery and design. These methods facilitate the exploration of interactions exhibited by small molecules upon binding to specific molecular sites, holding substantial promise for novel drug identification. The increasing adoption of molecular docking in drug discovery reflects its growing prominence. The methodology delineated in this paper underscores its remarkable efficiency, extending its applicability to structures constructed via homology or comparative modeling. Furthermore, it enables the assessment of a compound's druggability and specificity toward a designated target, enhancing our capabilities in rational drug design.

The application of molecular docking serves as a valuable tool for predicting the binding conformation between ligands and proteins, offering insights into critical molecular interactions. Within the realm of docking, effective strategies often entail the exploration of high-dimensional spaces and the utilization of precise scoring functions to prioritize potential docking configurations [6-7]. In this study, our primary objective is to investigate the molecular complex formed between TMPRSS4 and a potent ligand, with the overarching goal of advancing the drug discovery process. To accomplish this objective, a diverse array of bioinformatics methods will be harnessed. This research aims to unravel both the structural and functional attributes of TMPRSS4 while facilitating the identification of potential molecular interaction sites between TMPRSS4 and its ligand. Such insights have the potential to significantly contribute to our comprehension of the diagnostic role of TMPRSS4 in various diseases.

## MATERIALS AND METHODS

### **Homology Modelling:**

Homology modeling stands as a computational technique employed to predict the three-dimensional structure of a protein, relying on its amino acid sequence as a blueprint. As attested by Muhammed and Aki-Yalcin in 2019, this computational method for structure prediction is widely acclaimed for its exceptional accuracy [8-9]. Within the domain of homology modeling, an array of applications and services are readily accessible. For the purposes of this study, the Swiss Model tool was selected as the modeling platform. To validate the resulting protein models, the freely available Procheck web program was employed.

### **Molecular Docking Studies**

In order to identify a potential candidate for inhibiting TMPRSS4, molecular docking studies were conducted on the homology model of TMPRSS4's ECD and CD domains. The docking process operated the PubChem database (<https://pubchem.ncbi.nlm.nih.gov/>) and the Indian Medicinal Plants, Phytochemistry and Therapeutic (IMPPAT) database (<https://cb.imsc.res.in/imppat/>). The ligands were obtained in sdf format and subsequently converted into a mol2 3D file format using the chimera software. AutoDock Tools was employed to prepare the ligand input files for docking, which were then saved as pdbqt files. The homology modelled structure was prepared. The structures were modified by introducing hydrogen atoms in their standard geometry, followed by the removal of all solvent molecules. Subsequently, energy minimization was performed on the modified structures. The ultimate optimised structures were stored in the designated working directory. To facilitate a comprehensive investigation of the intermolecular interactions and binding modes between the protein TMPRSS4 and the specific ligand, molecular docking studies were conducted using iGEMDOCK, a graphical environment intended for the recognition of pharmacological interactions and virtual screening. According to Yang and Chen (2004), iGEMDOCK provides interactive interfaces for the configuration of screening compounds and the identification of the binding site of the target protein [10]. The energy-based scoring mechanism employed by GEMDOCK is opted to rank and visually represent the screening chemicals [11]. The secondary docking was conducted using the AutoDock Vina tool, employing a rigid receptor as a refinement methodology. The objective of this approach was to determine the best 20 poses from a pool of 100 poses generated for each drug under investigation. The scoring algorithms underwent modifications to restore them to their original settings [12]. After the completion of the docking operations, the obtained poses were analysed. The poses that displayed good characteristics, including acceptable rmsd\_refine values and preservation of the native ligand's binding mode, were selected.

### **Molecular Dynamics Simulation:**

The molecular dynamics (MD) simulations were conducted via the Desmond software package (Desmond, Schrödinger, 2015) in conjunction with the Optimised Potentials for Liquid Simulations (OPLS) 2005 force field. The prepared structures were initially loaded into the Desmond system builder and then solvated using TIP3P water. Orthorhombic periodic boundary conditions were established in order to define the dimensions and geometry of the repeating unit, which was buffered at intervals of 10 angstrom (Å). To maintain electrical neutrality, NA<sup>+</sup>/Cl<sup>-</sup> ions were introduced into the solvated system dependent on the overall charge of the system. The ions were positioned in a random manner throughout the system. The energy reduction of the system was achieved through the use of a hybrid methodology that combined the steepest descent algorithm. The algorithm was executed for a maximum of 5000

iterations or until a gradient threshold of 25 kcal/mol/Å was reached. The experiment was conducted employing the Limited-memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) algorithms until a convergence threshold of 1 kcal/mol/Å was attained.

The minimized systems underwent equilibration using the standard NVT and NPT protocols. Additionally, both systems underwent molecular dynamics simulations (MDS) for a duration of 10 nanoseconds (ns) using the default relaxation technique. The simulations were conducted under periodic boundary conditions, employing the NPT ensemble is established by assigning suitable values to the number of atoms, pressure, and temperature. The temperature was adjusted to 310 Kelvin, while the atmospheric pressure was set to 1.013 atmospheres.

## RESULTS

### Homology Modelling

The transmembrane serine protease TMPRSS4, consisting of a 437-amino acid sequence, exhibits a molecular mass of approximately 48,246 Da. To model the external domain and catalytic domain, homology modeling was conducted utilizing the Swiss-model server [13-14].

To facilitate the homology modeling process, the target sequences underwent Protein-Protein BLAST analysis against the PDB database, with the primary goal of identifying suitable templates. The selection of PDB id: 7MEQ as the template for modeling both the catalytic and external domains of TMPRSS4 was guided by criteria emphasizing maximal identity, a high score, and a lower e-value. Comparing the target and template sequences using structural data revealed a 42% similarity in the catalytic domain and a 39% similarity in the external domain. The resulting homology models exhibited promising quality metrics, with the catalytic domain model achieving a GMQE score of 0.77 and a QMEANDisCo Global score of 0.73 +/- 0.05. Similarly, the external domain model demonstrated a GMQE score of 0.68 and a QMEANDisCo Global score of 0.68 +/- 0.05."

### MD stimulation for the Catalytic and External Domain:

MD simulation played a pivotal role in refining the homology models of both the catalytic and extracellular domains of human TMPRSS4 [15]. The stability assessment of these domains during the MD simulation was performed by monitoring their deviation from the initial structure, quantified as the root mean square deviation (RMSD). The RMSD values for the backbone atoms of the human TMPRSS4 protein and all atoms across the entire MD simulation trajectory are illustrated in Figure 1.

Notably, the three-dimensional conformations of the extracellular domain of human TMPRSS4 attained a stable state after a 5-nanosecond simulation, with the RMSD of protein backbone atoms converging to 2.2 Å. Similarly, the catalytic domain of human TMPRSS4 achieved structural stability in its 3D conformation after a 6-nanosecond simulation, with the RMSD of all atoms in the protein backbone converging to 1.6 Å.

To further investigate the mobility of individual residues, root-mean-square fluctuations (RMSF) were computed for all structures of human TMPRSS4. The RMSF profiles in Figure 2 reveal notable peaks in side chain residue fluctuations within both the Catalytic and External domains, indicative of substantial variations among residues residing in helical or sheet structures (Figure 4, 5).

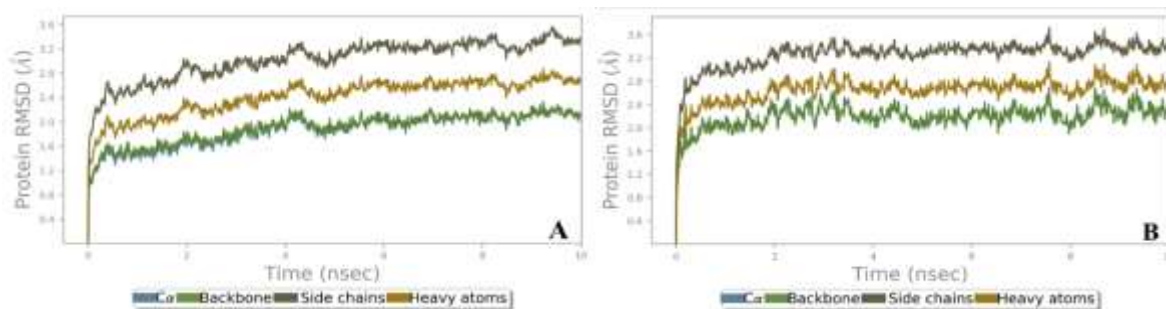


Figure 1: Protein-Ligand RMSD for [A]. Catalytic domain (CD) of TMPRSS4 and B. Extracellular domain (ECD) of TMPRSS4

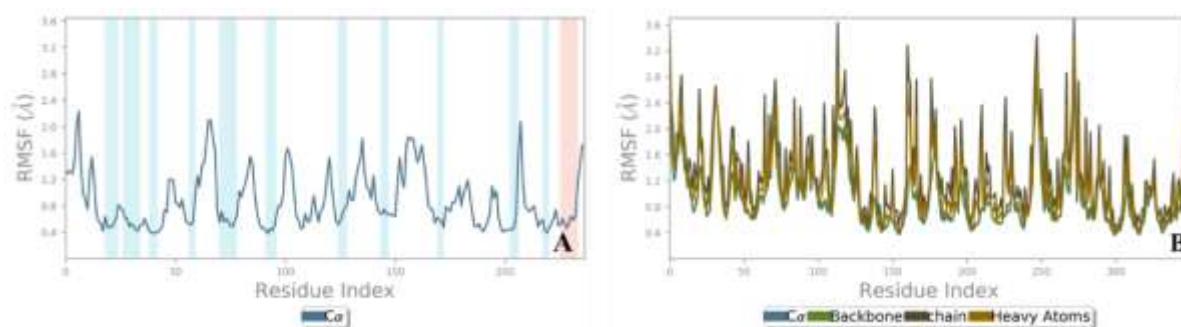


Figure 2: Protein-Ligand RMSF for [A] Catalytic domain (CD) of TMPRSS4 and [B] Extracellular domain (ECD) of TMPRSS4

The Ramachandran plot, depicted in Supplementary Figure S1 in additional file 1, reveals that the optimized structure of ECD TMPRSS4 exhibits a notably higher percentage of residues (84.8%) positioned within the allowed regions. Similarly, the CD TMPRSS4 structure displays a higher percentage of residues (80.6%) within the allowed regions, in contrast to the initial 3D structures obtained through homology modeling, which exhibited percentages of 70.4% and 75.0% for allowed regions, respectively. This enhancement in structural quality suggests that the optimized extracellular domain (ECD) and catalytic domain (CD) of TMPRSS4 are characterized by enhanced stability and reliability, rendering them suitable candidates for investigating the binding mechanisms of ECD and CD TMPRSS4 with ligands through molecular docking and subsequent molecular dynamics (MD) simulations.

### Molecular Docking

The identification of binding sites in the protein structures, modeled through homology, was accomplished through two distinct methods: the Computed Atlas of Surface Topography of proteins (CASTp) and the binding site module of DS 4.0. For the catalytic domain and external domain, we selected the highest-ranked binding sites based on the criteria established by Wei Tian et al. in 2018 and Binkowski, T. et al. in 2003, respectively [16-17].

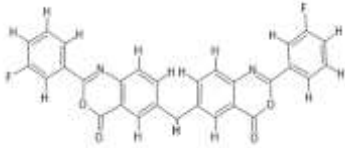
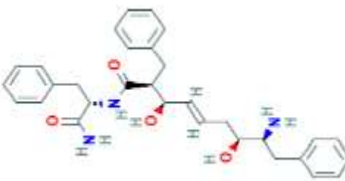

The binding sites chosen for the catalytic domain possessed a surface area of 290.428 Å<sup>2</sup> and a volume of 260.33 Å<sup>3</sup>, whereas those identified for the external domain exhibited a surface area of 394.631 Å<sup>2</sup> and a volume of 208.951 Å<sup>3</sup>.

In this study, compounds were identified through two distinct approaches. Firstly, a scaffold filter was applied to screen compounds from five Hepsin inhibitors (PubMed id: 15613534, 10892684, 71371232, 7020194, and 10045387), along with one TMPSS4 inhibitor (5-chloro-

N-[4-fluoro-3,5-bis(trifluoromethyl)phenyl]-2-hydroxybenzamide), sourced from a vast pool of 120 million compounds within PubChem. Secondly, phytochemicals were subjected to the same scaffold filter, originating from The Indian Medicinal Plants, Phytochemistry, and Therapeutics 2.0 [18-19]. The initial screening resulted in the selection of 2,229 compounds, primarily based on their similarity to Hepsin inhibitors. Subsequently, 1,640 compounds were subjected to Lipinski rule screening, specifically for docking against the homology-modeled structures of both the catalytic and exterior domains.

In parallel, a second methodology involved the identification of 550 compounds for primary docking and virtual screening. This process harnessed iGEMDOCK, a comprehensive software program encompassing docking, screening, and post-screening analysis [20]. Following the primary docking and virtual screening, the top 100 compounds were selected for precision docking at both target sites, utilizing Autodock—a specialized software tool designed for predicting the binding of small molecules, including drug candidates or substrates, to receptors with known three-dimensional structures.

Ultimately, the top 10 compounds were chosen from the databases of both the extracellular and catalytic domains, serving as key candidates for subsequent biological research (Table 1, 2).

Sl. No.	Structure of the compound	Energy value (Kcal/mol)	Amino Acid Interactions
1		-9	LEU147, LEU53, MET90, PHE260, GLU36, TYR92 – Van der Waals; CYS143, GLN344 – Conventional Hydrogen Bond; SER146, GLU37 – Halogen (Fluorine); ARG254, LYS145, ARG52 – Pi-Cation; ALA142, LYS55 – Pi-Sigma
2		-7.8	LEU373, GLN89, PHE261, ASN374, TYR377, GLN119, THR85, GLU116, LEU82, PHE78, GLU80, SER118, LYS381 – Van der Waals; ARG88, GLU84 – Conventional Hydrogen Bond; ALA81 – Pi-Sigma; ASN117 – Pi-Donor Hydrogen Bond
3		-7.6	GLU36, GLU37, HIS38, CYS39, LYS41, VAL51, ASP343, LYS145, PHE260, HIS346, TYR92 – Van der Waals; GLU263, SER342, GLN344, LEU53 – Conventional Hydrogen Bond; ARG52 – Carbon Hydrogen Bond;

	10875127		MET90 – Sulfur-x; ASP35, LYS55 – Pi-Cation
4	118986785	-7.6	PHE261, GLN89, TYR377, THR85, ASN378, ASN117, GLU116 – Van der Waals; GLN119, LYS381, ASN374 – Conventional Hydrogen Bond; SER118 – Carbon Hydrogen Bond; ALA81, ARG88 – Pi-Alkyl.
5	16396457	-7.6	HIS346, GLU263, LYS41, CYS39, HIS38, GLU36, SER342, GLU37, ASP343, LYS145, TYR92, GLN61, SER54, MET90, ASP56 – Van der Waals; GLN344, LEU53, LYS55 – Conventional Hydrogen Bond; ARG52 – Unfavourable Donor-Donor; TRP70 – Pi-Pi T Shaped; PHE260 – Pi-Pi Stacked
6	IMPHY001655	-7.48	TYR377, PHE261, ASN374, GLN89, GLU120, ASN117, THR85, SER118 – Van der Waals; LYS381, LEU121 – Conventional Hydrogen Bond; GLN119, ARG88 – Pi-Donor Hydrogen Bond.
7	IMPHY007825	-7.44	PHE78, ILE114, ASN378, TYR377, ASN374, GLN89, LEU121, GLU120 – Van der Waals; GLU84, LYS381, ASN117, SER118, GLN119, THR85 – Conventional Hydrogen Bond; GLU116, LEU82 – Carbon Hydrogen Bond; ARG88 – Pi-Donor Hydrogen Bond; ALA81 – Pi-Alkyl.
8	IMPHY001066	-7.28	LEU373, TYR377, GLN89, GLU120, LEU121, SER118, THR85, ASN117, ALA81, LYS381 – Van der Waals; GLN119, ASN374, ARG88 – Conventional Hydrogen Bond; PHE261 – Pi-Pi T Shaped.
9	IMPHY013673	-7.16	PHE261, GLN89, TYR377, ASN378, THR85, SER118, GLU84, LEU82, GLU116 – Van der Waals; GLN119, LYS381, ARG88 – Conventional Hydrogen Bond; ASN374, ASN117 – Carbon Hydrogen Bond; ALA81 – Pi-Alkyl.

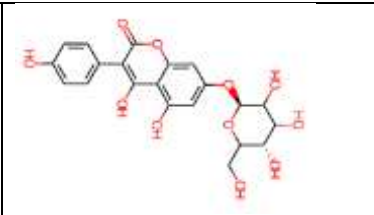
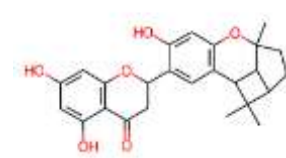
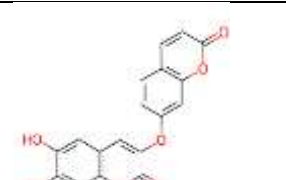
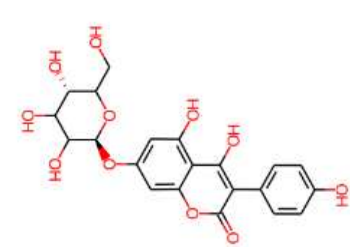
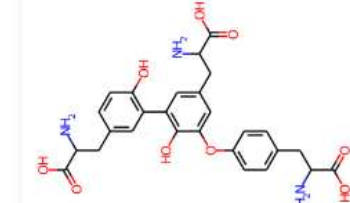

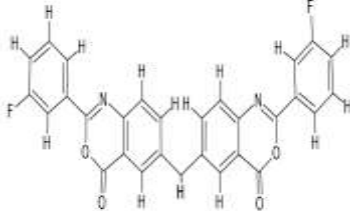
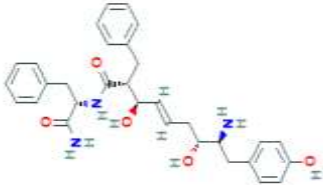

10	 IMPHY 013655	-7.56	LEU141, HIS139, ALA142, VAL51, HIS346, PHE260, SER342, LYS41, ASP35, GLU37, GLN344, LYS55, LEU53 – Van der Waals; CYS140, LYS145, GLU36, ASP343, TYR92 – Conventional Hydrogen Bond; ARG52 – Carbon Hydrogen Bond.
----	---	-------	--

Table 1: Docking and Amino acid interactions of ten most prominent compounds with the Extracellular domain (ECD) of TMPRSS4

S. No	Structure of the Compound	Energy value (Kcal/mol)	Amino Acid Interaction
1	 IMPHY003078	-9.32	GLY199, LYS153, CYS230, THR202, CYS203, PHE151, SER160, TRP149, GLY148, ILE147, GLN165, ILE146, GLU29, ASP201 – Van der Waals; ASP206 – Conventional Hydrogen Bond; GLY150 – Unfavourable Donor-Donor; THR152 – Amide-Pi Stacked; ALA166, LEU164 – Pi-Alkyl.
2	 IMPHY00788	-9.15	VAL28, GLN165, ASP201, THR152, CYS203, GLY150, PHE151, TRP149, GLY148, ILE147, GLLY199 – Van der Waals; ASP206 – Conventional Hydrogen Bond; VAL200 – Carbon Hydrogen Bond; GLU29 – Pi-Anion; LEU164 – Pi-Sigma; ALA166 – Pi-Alkyl
3	 IMPHY013655	-8.52	GLN154, GLY232, CYS230, LYS153, CYS203, ASP201, PHE151, GLY150, GLY148, ILE147, GLN165, ALA166, THR152, PRO233 – Van der Waals; GLY199, GLY231, LEU164, THR202 – Conventional Hydrogen Bond; ASP206 – Pi-Anion
4		-8.35	GLY232, PRO233, GLY199, LYS153, CYS230, GLN154, ASN155, CYS203, PHE151, SER160, GLY150, TRP149, LEU164, GLY148, THR152, TRP145, VAL26, GLY27, SER167 – Van der Waals; ASP201,



	IMPHY015916		GLY231, ASP206, VAL28, VAL25, GLN165 – Conventional Hydrogen Bond; VAL200 – Carbon Hydrogen Bond; GLU29 – Pi-Sigma; ALA166 – Pi-Alkyl.
5	 IMPHY007825	-8.05	ILE147, GLY148, TRP149, SER160, GLY150, THR152, PHE151, GLY231, THR202, ASN155, ALA166, GLU29, VAL200, GLY199 – Van der Waals; LEU164, GLN165, CYS203, ASP201, CYS230 – Conventional Hydrogen Bond; GLN154 – Unfavourable Donor-Donor; ASP206 – Unfavourable Acceptor-Acceptor; LYS153 – Pi-Alkyl.
6.	 5770901	-10.1	TYR228, PRO233, GLY232, GLY199, PHE15, CYS203, THR202, ASP201, ILE147, GLY148, GLY150, CYS230, ASN155, GLY229 – Van der Waals; GLN154, ALA166 – Conventional Hydrogen Bond; ILE146, GLN165 – Halogen (Fluorine), ASP206 – Pi-Anion; GLY231, THR152 – Amide-Pi Stacked; LYS153, LEU164 – Pi-Alkyl.
7.	 101216212	-8.4	ILE195, GLY27, VAL26, SER167, TRP145, GLN154, LYS153, CYS230, CYS203, GLY150, ILE147, LEU163, TRP149, ASP201, GLY199, ALA166, THR152 – Van der Waals; VAL28, GLN165, VAL25, PHE151, GLY148, LEU164 – Conventional Hydrogen Bond; GLU29; ASP206 – Pi-Anion; VAL200 – Pi-Sigma.
8.	 101216217	-8.4	GLYA48, CYS203, ILE147, THR202, VAL224, LEU164, ALA166, GLN165, GLU29, GLN154, GLY231, LYS153, GLY199, THR152, PHE151 – Van der Waals; ASP201 – Conventional Hydrogen Bond; VAL203 – Carbon Hydrogen Bond; ASP206 – Pi-Anion; CYS230 – Alkyl; ILE146 – Pi-Alkyl.


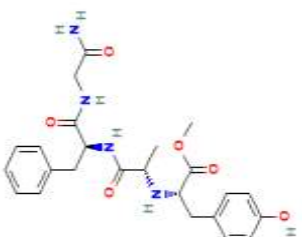
9.	 <p>124437228</p>	-8.4	SER160, TRP149, GLY148, GLY199, GLY231, CYS230, THR202, LYS153, CYS203, GLN154, PHE151, TRP145, VAL25, GLN165, GLU29 – Van der Waals; GLY150, ASP201, THR152 – Conventional Hydrogen Bond; VAL200 – Carbon Hydrogen Bond; ASP206 – Pi-Anion; SER167 – Pi-Donor Hydrogen Bond; LEU164 – Pi-Sigma; ALA166 – Pi-Alkyl.
10.	 <p>10412432</p>	-8.3	TRP145, VAL26, VAL28, GLU29, ILE146, ILE147, GLY148, THR202, PHE151, CYS230, GLY231, LYS153, THR152, GLY199, SER167 – Van der Waals; VAL25, GLN165, ASP201, GLN154, CYS203 – Conventional Hydrogen Bond; VAL200 – Carbon Hydrogen Bond; ASP206 – Pi-Anion; ALA166, LEU164 – Pi-Alkyl.

Table 2: Docking and Amino acid interactions of the ten most prominent compounds with the catalytic domain (CD) of TMPRSS4

The computational techniques employed in this experiment elucidated the interactions between the ligand and critical amino acid residues within the target domains, affirming the successful binding of the ligand and its compatibility with both the extracellular and catalytic domains. These interactions have been comprehensively assessed on an individual basis.

Specifically, the docking and amino acid interactions of the 10 compounds sourced from PubChem with the External domain were thoroughly investigated. The outcomes revealed that compounds 5770901, 101216213, and 10875127 displayed the most notable docking energy values, with scores of -9 (Kcal/mol), -7.8 (Kcal/mol), and -7.6 (Kcal/mol), respectively. Among the tested compounds, IMPHY001655 from the IMPPAT database demonstrated the highest docking energy, registering at -7.48 (Kcal/mol) when interacting with the active site of the external domain. The docking and amino acid interactions of ten phytochemical compounds with the catalytic domain were systematically analyzed. Notably, IMPHY003078, IMPHY00788, and IMPHY013655 exhibited considerable docking energies, measuring -9.32 (Kcal/mol), -9.15 (Kcal/mol), and -8.52 (Kcal/mol), respectively.

Furthermore, the docking and amino acid interactions of the top 10 compounds from PubChem with the Catalytic domain were scrutinized, revealing that compound 5770901, 101216212, and 101216217 achieved docking scores of -10.1 (Kcal/mol), -8.4 (Kcal/mol), and -8.4 (Kcal/mol), respectively. IMPHY001655 from the IMPPAT database and compound 5770901 from the PubChem database displayed the highest affinity for the active sites within the external domain. Conversely, IMPHY003078 from the IMPPAT database and compound

5770901 from the PubChem database exhibited the strongest binding affinity for the active sites situated within the catalytic domain.

The protein-ligand interactions between the compounds and the two domains are depicted in histogram maps in Figure 3. Key amino acids, including ASP35, ARG52, LEU53, LYS55, MET90, TYR92, and PHE260, engage in hydrophobic contacts, while GLU37 and Ser342 establish water bridge interactions in the ECD complex with the 5770901 ligands. Additionally, CYS140, ASP343, and GLN344 form hydrogen bonds, although these interactions exhibit temporal fluctuations. In the CD-5770901 complex, GLN154 was observed to participate in hydrophobic contacts, while GLY134 and ALA166 engaged in hydrogen bond interactions with the ligand. The hydrogen bonding interactions within the complex were meticulously analyzed and elucidated to provide a comprehensive understanding of the consistent behavior observed in the lead molecule (Figure 4 and 5)

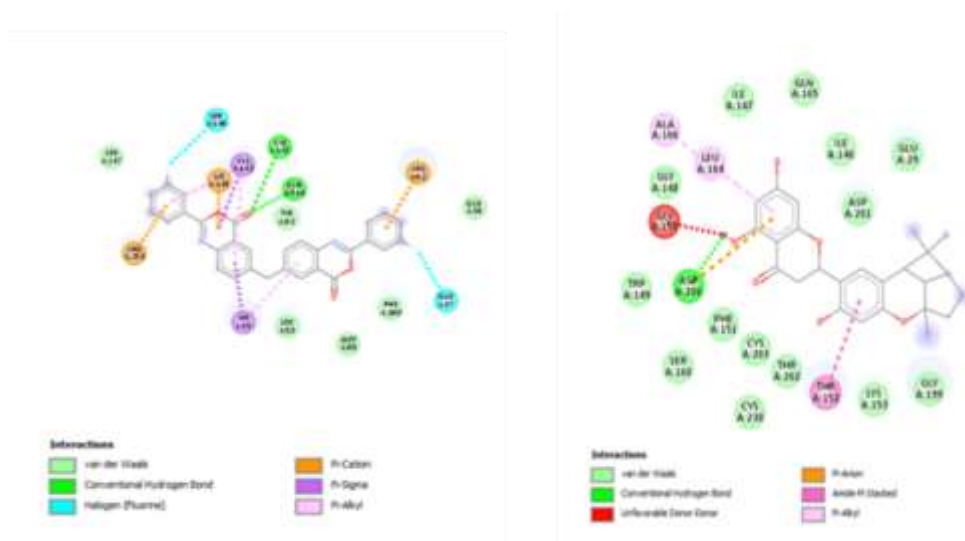


Figure 3:

2D image of PubChem SID: 5770901 interactions with the [A] Extracellular domain and [B] Catalytic domain.

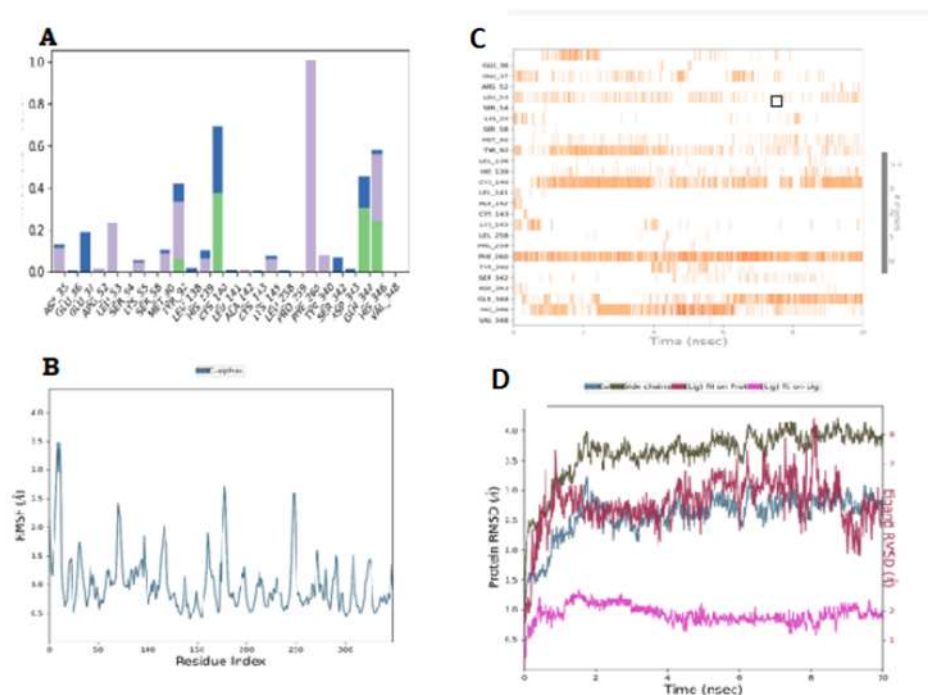


Figure 4: A) Molecular Interaction diagram of Extracellular domain complex with CID: 5770901. (B.) RMSF graph. (C) Timeline representations of the Molecular Interaction (D) RMSD graph observed during the molecular dynamic simulation

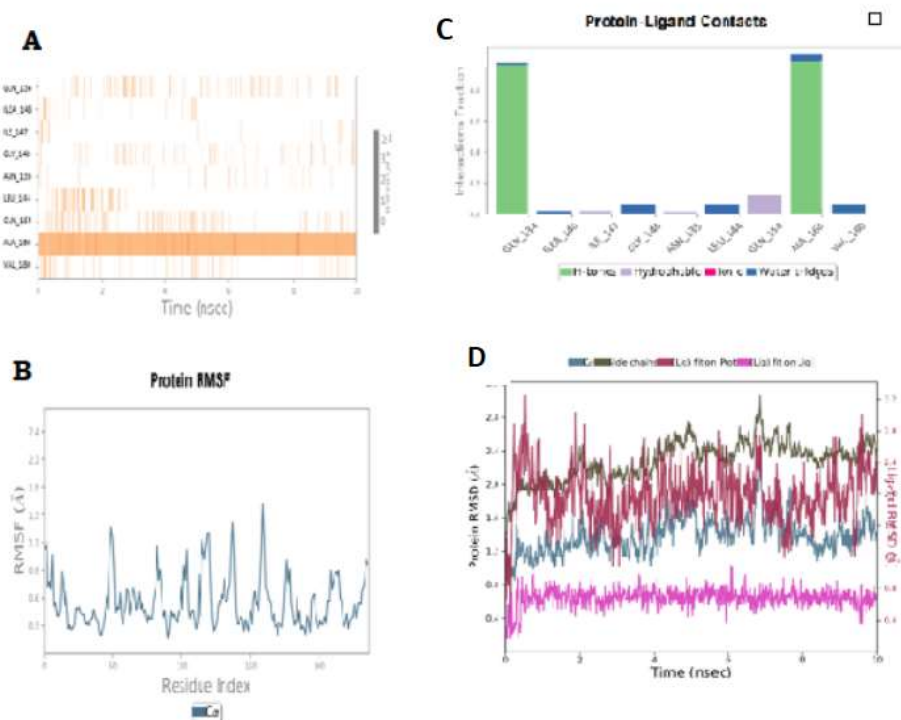


Figure 5: A) Molecular Interaction diagram of Catalytic domain complex with CID: 5770901. (B.) RMSF graph. (C) Timeline representations of the Molecular Interaction (D) RMSD graph observed during the molecular dynamic simulation

### Molecular dynamics analysis of receptor-ligand complex:

The MD simulation approach stands as a widely utilized tool for assessing the stability of docking conformations and quantifying the contribution of active site residues within a protein. To evaluate the top 10 compounds, chosen based on their docking scores against two specific domains, MD simulations were conducted. From this selection, three compounds, identified by their PubMed identifiers as SID5770901, CID101216213, and CID10995039, along with the top two compounds from the IMPACT database bearing identifiers IMPHY001655 and IMPHY007825, were earmarked for MD simulations. The optimal docking conformations were selected to perform 10 ns MD simulations, enabling a comprehensive exploration of the molecular orientations and offering insights into the pivotal residues responsible for their stability.

#### **Molecular dynamics simulation of top five compounds with ECD:**

The RMSD analysis of the ECD-5770901 complex revealed an initial increase, reaching 4.0 Å within the first 4 ns of simulation, followed by a stable RMSD value of approximately 3.8 Å for the remainder of the simulation. Similarly, the ECD-10995039 complex achieved stability around 4.0 Å after 4 ns of simulation. In contrast, the ECD-101216212 complex displayed fluctuations spanning from 2.5 Å to 4.5 Å throughout the simulation. The ECD-IMPHY001655 complex reached a stable state at around 3.5 Å after 3 ns of simulation, maintaining this conformation for the duration of the experiment. On the other hand, the ECD-IMPHY007825 complex exhibited significant variations, with RMSD values fluctuating between 0.1 Å and 4.8 Å.

Additionally, the inhibitor complex of TMPRSS4 initially displayed an increase in distance, reaching 3.5 Å during the first nanosecond, followed by fluctuations that led to a maximum distance of 5.4 Å by the end of the simulation. When comparing the known TMPRSS4 inhibitor to all the tested compounds, most of the compounds exhibited moderate RMSD values ranging from 2.5 Å to 4.5 Å, with the exception of IMPHY007825.

The RMSF plots were employed to analyze the fluctuations exhibited by individual residues in the presence of the ligand, which correlated with secondary structure elements (SSE) observed during the simulation. In the ECD-5770901 complex, the initial 30 residues of the protein exhibited high variability in RMSF values. However, as the simulation progressed, RMSF values stabilized with minimal fluctuations. Similarly, the RMSF values for CD remained steady with few variations throughout the entire simulation period, as depicted in Figure 4 and 5.

On average, all complexes exhibited an RMSF value ranging from 2.5 to 3.5 Å, except for the IMPHY007825-ECD complex, which reached 5.6 Å in the 55-60 amino acid region. Notably, secondary structures such as  $\alpha$  helices and  $\beta$  strands displayed higher rigidity compared to loop regions within the protein, with loop regions exhibiting an average RMSF fluctuation ranging from 3.5 to 4.5 Å, while secondary structural elements exhibited fluctuations ranging from 2.0 to 3.0 Å.

Furthermore, the RMSF analysis provided insights into the dynamic fluctuations of the ligand molecule, particularly in relation to its two-dimensional structure and interactions with protein fragments. The complexes displayed an average Ligand RMSF value ranging from 0.5 to 1 Å, except for the IMPHY007825-ECD complex, which exhibited a notable 30 Å fluctuation, suggesting inadequate binding to the ECD active site.

To elucidate the interactions, a histogram map was generated, encompassing hydrogen bonds, hydrophobic interactions, water bridges, and ionic interactions. The histogram depicted residues involved in interactions, with the y-axis representing the fraction of interactions, expressed as a percentage.

In the ECD-5770901 complex, hydrogen bonds were observed between the SER residue and a nitrogen atom, contributing 32%, and between the GLN residue at position 344 and the hydroxyl group, contributing 49% to the overall complex. Hydrophobic interactions occurred with amino acids PHE260 and HIS346, contributing 33% and 39%, respectively.

In the ECD-10995039 complex, hydrogen bonding interactions were observed with specific amino acid residues, including THR85, ARG88, GLU116, ASN117, SER118, and ASN374, contributing 38%, 35%, 39%, 73%, 42%, and 38% to the overall binding affinity, respectively. Except for ASN374, which interacted with the nitrogen group, all other amino acids formed interactions with various hydroxyl groups on the ligand.

The ECD-101216212 complex displayed interactions with amino acid residues Glutamic acid 84, Glutamine 89, Asparagine 117, Glutamine 119, Asparagine 374, Tyrosine 377, and Lysine 381, contributing 24%, 36%, 42%, 62%, 11%, 15%, and 59% to the respective activities. The ligand also exhibited water bridging interactions with ARG88 for approximately 46% of the total interaction duration.

In the ECD-IMPHY007825 complex, interactions involved amino acid residues CYS140, ALA142, PHE249, SER250, and ARG254, forming hydrogen bonds at rates of 18%, 33%, 30%, 15%, and 42%, respectively. Additionally, LEU247 formed a water bridge for approximately 30% of the simulation time.

The ECD-IMPHY001655 complex displayed consistent hydrogen bond formation throughout the simulation duration with ARG88, ASN117, and GLN119, while GLN89 only formed hydrogen bonds for 44% of the contact time, exclusively between the ligand's hydroxyl groups.

The inhibitor of TMPRSS4 established three hydrogen bonds with amino acids GLU36, GLU263, and SER342 located in the ECD, with contact durations of 41%, 84%, and 44%, respectively. Notably, all comparable inhibitors, except for IMPHY007825, demonstrated favorable interactions with the extracellular domain of TMPRSS4.

Molecular dynamics analysis of CD and top compounds:

"The RMSD values for the CD-5770901 complex displayed an initial increase to approximately 3.5 Å within the first 3 ns of the simulation, followed by stabilization at around 3.8 Å for the remainder of the study. The CD-10995039 complex reached stability at a distance of approximately 3.2 Å after 3 ns of simulation. In contrast, the CD-101216213 complex exhibited an initial peak at approximately 3.8 Å at 4 nanoseconds, followed by fluctuations in the range of 3 to 4 Å. Both the CD-IMPHY001655 and CD-IMPHY007825 complexes achieved stability at approximately 3.8 Å at 4 ns, with subsequent fluctuations between 3 to 4 Å. The inhibitor complex of TMPRSS4 displayed an initial increase to a distance of 3.6 Å during the first nanosecond, followed by fluctuations ranging up to 5.4 Å until the conclusion of the simulation. When comparing these compounds with the known inhibitor of TMPRSS4, it was observed that the compounds exhibited low RMSD values ranging from 2.5 Å to 4.5 Å.

RMSF analysis of the CD-5770901 complex revealed large values for the initial 30 residues of the protein, which stabilized with minimal variations toward the end of the simulation period, as depicted in Figure 4 and 5. On average, all complexes exhibited an RMSF value ranging from 3.0 to 4.0 Å, except for one complex. The loop region displayed higher average RMSF fluctuation (ranging from 3.5 to 4.5 Å) compared to secondary structural elements (ranging from 2.0 to 3.0 Å) in all complexes.

The ligand RMSF values for all complexes averaged between 1 to 2 Å, except for the CD-10995039 complex, which exhibited a significant fluctuation of 6 Å in residues 13-20 of the ligand. Importantly, all complexes demonstrated favorable interactions with CD TMPRSS4 compared to a known inhibitor of TMPRSS4.

The histogram analysis highlighted specific interactions within the CD-5770901 complex, including hydrogen bonds formed with amino acid residues PHE131, GLN134, ASP181, and CYS210. PHE131 and ASP181 displayed interactions throughout the entire simulation time, while GLN134 and CYS210 exhibited interaction times of 40% and 24%, respectively. Additionally, hydrophobic interactions occurred with the amino acids PHE260 and HIS346, contributing 33% and 39%, respectively. Hydrogen bonds formed between PHE131 and GLN134 amino acids and the hydroxyl groups of the ligand, as well as a hydrogen bond between ASP181 and the amino group of the ligand.

In the CD-10995039 complex, residues GLY27, VAL28, PHE151, THR152, GLN165, and ASP206 were involved in hydrogen bond formation with the ligand. Interactions between VAL28, GLN165, and ASP206 occurred throughout the simulation duration. GLY27, PHE151, and THR152 exhibited interaction times of 24%, 70%, and 24%, respectively. Water bridges formed by VAL25, GLU29, and GLY199 persisted for the majority of the simulation time.

The CD-101216213 complex exhibited hydrogen bond interactions with GLY27, GLU29, PHE151, THR152, and ASP206. ASP206 interactions occurred throughout the simulation duration. GLY27, GLU29, PHE151, and THR152 displayed interaction times of 22%, 64%, 62%, and 21%, respectively. Water bridges formed between aspartic acid residue at position 201 (ASP201) and the ligand's hydroxyl group. A hydrogen bond also formed between PHE151 and the amino group of the ligand.

In the CD-IMPHY007825 complex, hydrogen bonding interactions involved residues GLY150, THR152, GLN154, GLY199, ASP201, ASP206, and CYS230. ASP201 and ASP206 exhibited interactions throughout the entire duration of the simulation. GLY150, THR152, GLN154, GLY199, and CYS230 had interaction times of 22%, 40%, 24%, 75%, and 36%, respectively. Additionally, LEU247 formed a water bridge for approximately 30% of the simulation time.

The CD-IMPHY001655 complex exhibited consistent hydrogen bond formation throughout the simulation duration with residues PHE151, ASP201, ASP206, and CYS230. These interactions contributed 84%, 26%, 79%, and 64% of the total interaction time, respectively, with hydrogen bonds exclusively forming between the ligand's hydroxyl groups and the amino acids.

The inhibitor of TMPRSS4 established hydrogen bonds with three specific amino acids: GLU36, GLU263, and SER342 located in the ECD, with contact durations of 41%, 84%, and

44%, respectively. Residues GLN29, GLN165, and GLY199 were associated with the formation of water bridges.

Comparatively, all compounds, except for IMPHY007825, exhibited favorable interactions with CD TMPRSS4. The docking and MD simulation studies revealed that the binding affinity of 5770901 toward CD TMPRSS4 remained highly stable with the proposed orientations.

## **DISCUSSION**

The transmembrane protease, serine 4 (TMPRSS4), has emerged as a significant player in the context of cancer, with close associations to other proteases like hepsin, TMPRSS2, and matriptase. It is noteworthy that TMPRSS4 has been consistently found to be upregulated in various solid tumors, often correlating with unfavorable prognoses. In a groundbreaking study [21], it was demonstrated that the methylation status of TMPRSS4 could serve as a valuable diagnostic tool for the early detection of cancer and the post-surgical monitoring of recurrence risk. These findings underscore the potential of TMPRSS4 as a biomarker for identifying individuals at heightened risk of disease recurrence.

An important study by Kang et al. in 2013 [22] explored a unique set of 2-hydroxydiarylamide derivatives. These compounds were meticulously synthesized and evaluated for their capacity to inhibit TMPRSS4 activity and mitigate cancer cell invasion. The study yielded remarkable results, revealing substantial inhibitory effects on TMPRSS4 activity. Importantly, these inhibitory effects translated into reduced invasiveness in colon cancer cells characterized by TMPRSS4 overexpression. This underscores the therapeutic promise of targeting TMPRSS4 in the battle against cancer.

In a more recent investigation by Kim et al. in 2019 [23], novel 2-hydroxydiarylamide derivatives, namely IMD-0354 and KRT1853, were scrutinized for their potential as TMPRSS4 inhibitors. IMD-0354, a known inhibitor of I $\kappa$ B kinase (IKK)- $\beta$ , displayed considerable efficacy in both acute and subacute inflammatory conditions, exhibiting an inhibitory concentration (IC<sub>50</sub>) of 11  $\mu$ M. Notably, KRT1853, a derivative of IMD-0354, exhibited a twofold increase in inhibitory potency. These compounds not only inhibited TMPRSS4 but also impacted TMPRSS4-mediated cellular signaling pathways, including the activation of transcription factors SP1, AP-1, and NF- $\kappa$ B, as well as the induction of anti-apoptotic proteins bcl-2 and surviving. Furthermore, both compounds demonstrated substantial efficacy in reducing cancer cell invasion and proliferation while inducing apoptosis.

## **CONCLUSION:**

The present study represents the initial endeavor to ascertain the three-dimensional structure of the ECD and CD of TMPRSS4 serine protease in the human species. The exploitation of *in silico* methodologies to construct a three-dimensional model yields essential data for the purpose of rational drug design. The compound CID 5770901 was subjected to molecular docking and MD simulations, which indicated its significant activity against both the catalytic and extracellular domains of the TMPRSS4 serine protease. The docking score of CID 5770901 was determined to be -9 kcal/mole. The RMSD value of the complex formed by ECD-5770901



displayed an early rise to 4.0 Å over the course of the first 4 ns, after which it maintained a steady value of about 3.8 Å until the completion of the simulation. In a similar vein, the RMSF readings for CD exhibited a consistent pattern with little fluctuations across the whole duration of the simulation. TMPRSS4, a member of the transmembrane serine protease (TTSP) family, demonstrates elevated expression in several cancers. The advancement of targeted inhibitors for TMPRSS4 holds promise for the therapeutic management of cancer. However, further investigation and empirical studies are required in order to validate the aforementioned hypothesis.

## REFERENCES

1. de Aberasturi, A. L., Calvo, A., TMPRSS4: an emerging potential therapeutic target in cancer. *British Journal of Cancer*, 2015; pp.112(1), 4-8.
2. Ohler, A. & Becker-Pauly, C., TMPRSS4 is a type II transmembrane serine protease involved in cancer and viral infections. *Biological Chemistry*, 2012; 393(9), pp. 907-914.
3. Puente, X.S., Sánchez, L.M., Gutiérrez-Fernández, A., Velasco, G., López-Otín, C., A genomic view of the complexity of mammalian proteolytic systems. *Biochem Soc Trans*, 2005; 33(2), pp. 331–334.
4. Netzel-Arnett, S., Hooper, J. D., Szabo, R., Madison, E. L., Quigley, J. P., Bugge, T. H., Antalis, T. M., Membrane anchored serine proteases: A rapidly expanding group of cell surface proteolytic enzymes with potential roles in cancer. *Cancer and Metastasis Reviews*, 2003; 22(2), pp. 237-258.
5. Wallrapp, C., Hähnel, S., Müller-Pillasch, F., Burghardt, B., Iwamura, T., Ruthenbürger, M., Lerch, M. M., Adler, G., & Gress, T. M., A novel transmembrane serine protease (TMPRSS3) overexpressed in pancreatic cancer. *Cancer research*, 2000; 60(10), pp. 2602–2606.
6. Pagadala, N.S., Syed, K. & Tuszynski, J., Software for molecular docking: a review. *Biophys Rev*, 2017; 9, pp. 91– 102.
7. Morris, G.M., Lim-Wilby, M., Molecular Docking. In: Kukol, A. (eds) *Molecular Modeling of Proteins. Methods Molecular Biology™*, Humana Press, 2008; 443.
8. Muhammed, M. T., & Aki-Yalcin, E., Homology modeling in drug discovery: Overview, current applications, and future perspectives. *Chemical biology & drug design*, 2019; 93(1), pp. 12–20.
9. Kaczanowski S, Zielenkiewicz P, Why similar protein sequences encode similar three-dimensional structures? *Theoretical Chemistry Accounts*, 2009; 125: pp. 643–50.
10. Yang, J. M., & Chen, C. C., GEMDOCK: a generic evolutionary method for molecular docking. *Proteins*, 2004; 55(2), pp. 288–304.
11. Hsu, K-C., Chen, Y-F., Lin, S-R., Yang, J-M., iGEMDOCK: a graphical environment of enhancing GEMDOCK using pharmacological interactions and post-screening analysis. *Journal of BMC Bioinformatics*, 2011; 12(1), pp. S33.
12. Samra R. M., Soliman A. F., Zaki A. A., Ashour A., Al-Karmalawy A. A., Hassan M. A., & Zaghoul A. M., Bioassay-guided isolation of a new cytotoxic ceramide from *Cyperus rotundus* L. *South African Journal of Botany*, 2021; 139, pp. 210–216.
13. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T., SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*, 2018, 46(W1), pp. W296-W303.
14. Bienert, S., Waterhouse, A., de Beer, T.A.P., Tauriello, G., Studer, G., Bordoli, L., Schwede, T., The SWISSMODEL Repository - new features and functionality. *Nucleic Acids Res*, 2017; 45, pp. D313-D319.

15. Raval A, Piana S, Eastwood M.P., Dror R.O., Shaw D.E., Refinement of protein structure homology models via long, all-atom molecular dynamics simulations. *Proteins*, 2012; 80(8), pp. 2071-2079.
16. Wei Tian, Chang Chen, Xue Lei, Jieliang Zhao, Jie Liang, CASTp 3.0: computed atlas of surface topography of proteins, *Nucleic Acids Research*, 2018; 46 (W1), pp. W363–W367.
17. Binkowski, T. A., Naghibzadeh, S., & Liang, J., CASTp: Computed Atlas of Surface Topography of proteins. *Nucleic acids research*, 2003; 31(13), pp. 3352–3355.
18. M. Ilamathi, R. Hemanth, S. Nishanth, V. Sivaramakrishnan, Identification of potential transmembrane protease serine 4 inhibitors as anti-cancer agents by integrated computational approach, *Journal of Theoretical Biology*, 2016; 389, pp. 253-262.
19. Sunghyun Kang, Hye-Jin Min, Min-Seo Kang, Myung-Geun Jung, Semi Kim, Discovery of novel 2- hydroxydiarylamide derivatives as TMPRSS4 inhibitors. *Bioorganic & Medicinal Chemistry Letters*, 2013; 23(6), pp. 1748-1751.
20. Hsu, K. C., Chen, Y. F., Lin, S. R., & Yang, J. M., iGEMDOCK: a graphical environment of enhancing GEMDOCK using pharmacological interactions and post-screening analysis. *BMC bioinformatics*, 2011; 12(Suppl 1), pp. S33.
21. Villalba, M., Exposito, F., Pajares, M. J., Sainz, C., Redrado, M., Ramirez, A., Wistuba, I., Behrens, C., Jantus-Lewintre, E., Camps, C., Montuenga, L. M., Pio, R., Lozano, M. D., de Andrea, C., & Calvo, A., TMPRSS4: A Novel Tumor Prognostic Indicator for the Stratification of Stage IA Tumors and a Liquid Biopsy Biomarker for NSCLC Patients. *Journal of Clinical Medicine*, 2019; 8(12), pp. 2134.
22. Kang, S., Min, H.-J., Kang, M.-S., Jung, M.-G., & Kim, S., Discovery of novel 2-hydroxydiarylamide derivatives as TMPRSS4 inhibitors. *Bioorganic & Medicinal Chemistry Letters*, 2013; 23(6), pp. 1748–1751.
23. Kim, S., Ko, D., Lee, Y., Jang, S., Lee, Y., Lee, I. Y., & Kim, S., Anti-cancer activity of the novel 2-hydroxydiarylamide derivatives IMD-0354 and KRT1853 through suppression of cancer cell invasion, proliferation, and survival mediated by TMPRSS4. *Scientific Reports*, 2019; 9.