

Identification of potential anticancer phytochemicals against TMPRSS4:

An application of structure-based docking studies

Poornima Eeram^{1,2*}, Pratima Srivastava^{1,*}

¹Aragen Lifesciences Private Ltd., Survey No.125(p) & 126, IDA Mallapur, Hyderabad-500 076, Telangana, India

²Department of Biotechnology, Jawaharlal Nehru Technological University, Hyderabad - 500 085, Telangana, India

*Corresponding Authors

Mrs. Poornima Eeram: Poornima.esram@aragen.com and +91 9848018878

&

Dr. Pratima Srivastava: Pratima.srivastava@aragen.com and +91 90002 32335

Abstract:

Proteolytic enzymes are directly or indirectly engaged in mostly all biological processes indicating their importance in maintaining health and causing diseases. It has been discovered that proteolysis of the cell surface is a crucial mechanism for producing physiologically active components that help mediate a large variety of functions of the cells. The dysregulation of proteolysis can cause several diseases including arthritis, cancer, cardiovascular diseases and even neurodegenerative diseases. TMPRSS4 is a unique transmembrane serine protease and found at the cell surface that is highly expressed in gastric, pancreatic, and colon cancer tissues. Molecular docking is a widely used tool in drug discovery to identify potential new therapeutic targets and plays a vital role in the visualization of ligand–protein interaction at an atomic level and enhancing our understanding of the ligand behavior thus aiding in the structure-based drug designing. This study aims to investigate the molecular complex of TMPRSS4 with the potent phytochemicals for drug discovery using an array of bioinformatics tools. The study would not only help us to find the structure and functional relationship of the TMPRSS4 but also help us in predicting site of molecular interaction between TMPRSS4 and the ligand, which could potentially help in going a step ahead in finding the role of TMPRSS4 as a druggable target to treat gastric, pancreatic, and colon cancer diseases.

1. Introduction

A regulatory process called proteolysis uses the targeted hydrolysis of peptide bonds to carry out its functions. It is due to this post-translational modification that the role played by various proteins is being controlled by proteases. Following that, these proteases control a number of biological mechanisms, for instance, protein secretion, signal transduction, apoptosis, cell growth etc. They also play an integral role in various physiological processes like reproduction, host defense and also in numerous pathologies, like inflammatory diseases, cancer etc. Proteolytic enzymes are directly or indirectly engaged in mostly all biological processes indicating their importance in maintaining health and causing diseases (de Aberasturi & Calvo, 2015; Ohler & Becker-Pauly, 2012; Puente et al, 2005). After the accomplishment of sequencing the human genome, it was found that around 600 proteases-encoding genes forming more than 2% of the overall human proteome is present in humans. The dysregulation of proteolysis can cause several diseases including arthritis, cancer, cardiovascular diseases and even neurodegenerative diseases (de Aberasturi & Calvo, 2015; Ohler & Becker-Pauly, 2012; Puente et al, 2005).

In accordance with the mechanisms of catalysis, proteases are categorized into serine, aspartyl, threonine, metallo and cysteine proteases (Puente et al, 2005). This large family of proteins could be found extracellularly or in the cytoplasm or in some organelles such as the lysosome. Only a very small number of serine proteases could be found that have a transmembrane domain that allows them to attach to the plasma membrane of the cells.

It has been discovered that proteolysis of the cell surface is a crucial mechanism for producing physiologically active components that help mediate a large variety of functions of the cells. Serine proteases could be divided into three groups based on the transmembrane domain structure: Type I, that has carboxy-terminal domain; TTSP or type II that has an amino-terminal domain; and GPI, which is bound to the membrane by glycosyl-phosphatidylinositol (Netzel-Arnett et al, 2003).

Formerly known as TMPRSS3 (Wallrapp et al, 2000), TMPRSS4 is a unique transmembrane serine protease (type II), found in the long arm of 11th chromosome (11q23.3). It contains 48,597bp with around 12 introns and 16 exons. Lot of studies showed that about 18 different transcripts of this serine protease could be generated, out of which 2 transcripts degrade with the reason being nonsense-mediated decay. About 8 transcripts don't give a protein as a product. From the remaining 8 transcripts 3 are found to have an incomplete coding sequence in 5' or 3', also it either doesn't have stop or start codon sequences. Therefore, TMPRSS4 codes for only five isoforms that have complete coding sequence. It has a predicted size of 48 kDa with two sites of glycosylation, one at 130 and the other at 178 amino acids, the canonic protein (TMPRSS4-1) has 437 amino acids, on the other hand isoforms second and third differ by only two and five amino acids, respectively (de Aberasturi & Calvo, 2015).

TMPRSS4 is found at the cell surface that is highly expressed in gastric, pancreatic, and colon cancer tissues. In spite of the above fact, the biological function of TMPRSS-4 in cancer is unknown (Jung et al, 2008). Few of the recent studies suggest that TMPRSS-4 activity is linked with the influenza virus spreading as well (Ohler & Becker-Pauly, 2012). Dysregulation of Proteases is also a key hallmark of cancer, by being a potential diagnostic marker in the disease and therefore proteases have been the subject of numerous cancer studies (Min et al, 2014).

The other members of the TTSP family and TMPRSS4 have the given domains in common: stem, transmembrane, cytoplasmic, and proteolytic domains (Hooper et al, 2001). The 'catalytic triad,' which consists of these amino acids- Histidine, Aspartic acid, and Serine, which are important for the action of the proteolytic domain, and are largely conserved among various TTSPs. A substrate binding pocket is also found that modulates the enzyme activity and determined the specificity of the enzyme (Antalis et al, 2010). A lot of study has been conducted on TTSP family that gives a lot of information about different domains and their roles including TMPRSS4. Additionally, it is discovered that the conditioned medium contains the active domain of the TMPRSS4 protease that was released from the cultured cells (Min et al, 2014). According to this knowledge, there is a chance that soluble fragments could be found in the serum of cancer patients, suggesting that TMPRSS4 could be used as a non-invasive diagnostic biomarker.

Since cancer genesis and metastasis research account for the majority of the information on TMPRSS4, it is reasonable to conclude that this membrane-anchored serine protease deserves additional study as a novel potential therapeutic target in solid tumors. Data on TMPRSS4 is still very less and several sources of evidence support that TMPRSS4, according to lot of research, promote metastasis in preclinical models, overexpression of TMPRSS4 in various cancer types in comparison with normal tissues, high levels of TMPRSS4 found to be linked with reduces disease-free survival (DFS) and overall survival (OS), TMPRSS4 is a membrane bound protein and so it can be employed for developing blocking antibodies or other bio-inhibitory tools, and also extracellular fragments of TMPRSS4 in conditioned media which could be detected in serum sample of various cancer patients (de Aberasturi & Calvo,2015). However, for utilizing the benefits of this serine protease TMPRSS4, more information and research is required.

Molecular docking has been proved to be an important in structural molecular biology and in computer aided drug designing. It is now routinely being used for virtual screening or lead optimization for drug screening and designing. It helps in exploring the behaviour of small molecules while binding to the target site. It is a tool that is increasingly being used for drug discovery. It is so efficient that it could be performed even for the structures that have been made by homology or comparative modelling such as in this study. The druggability of the compounds and their specificity against a particular target can also be found.

Molecular docking helps in predicting the predominant binding mode of the ligand with the protein. Successful docking methods usually try to search for high-dimensional spaces and use a scoring function that correctly ranks candidate dockings (Pagadala et al,2017; Morris & Lim-Wilby,2008). This study aims to investigate the molecular complex of Tmprss4 with a potent ligand for drug discovery using an array of bioinformatics tools. The study would not only help us to find the structure and functional relationship of the Tmprss4 but also help us in predicting site of molecular interaction between Tmprss4 and the ligand, which could potentially help in going a step ahead in finding the role of Tmprss4 as a diagnostic marker in various diseases.

2. Materials and Methods

Homology Modelling:

One of the computational methods for predicting the 3D structure of a protein from its amino acid sequence is homology modelling. It is regarded as the computational structure prediction method with the highest level of accuracy. It comprises of a number of simple and easy to follow steps (Muhammed & Aki-Yalcin,2019). For homology modelling, there are numerous different programmes and services that are used such as the Swiss model tool.

Uniprot:

There are more than 120 million protein sequences and annotations in this enormous database of protein information. Expert curators extract data from scientific papers, which is then put in the knowledge base's UniProtKB/Swiss-Prot section. It defines functional information using structured vocabularies like the Gene Ontology (Go) or ChEBI as well as human readable free text and regulated syntax summaries (The UniProt Consortium,2019).

Chimera:

The interactions were visualized and examined using UCSF Chimera. It is a flexible programme that is used to analyses molecular structures and other relevant data, such as density maps, sequence alignment findings, docking results, trajectory data, etc. In addition to the fundamental services, it also offers visualization and extensions with extremely high levels of capability. This structure ensures that the extension operation satisfies the developer's requirements for adding new functionality (Pettersen et al,2004).

Discovery Studio:

Along with the type of interaction and bond lengths, the Discovery Studio aids in the identification of interactions between the active sites in the target and ligand conformations. Discovery Studio is a single, unified, user-friendly graphical interface that is employed for effective drug design and protein modelling. It includes well-known and common applications like the Catalyst, MODELLER, CHARMm, etc. It has years of successful, validated, and published outcomes as well as cutting-edge science to address the difficulties in drug discovery that exist today. It was created using the open operating system server platform of SciTegic Pipeline PilotSciTegic Enterprise, which makes it simple to integrate protein modelling, structure-based design, and even outside design (Studio,2008).

SwissADME:

SwissADME was used for the initial analysis and screening (<https://www.swiss-adme.ch/>). It is a free online tool that rates the friendliness of medicinal chemistry of small compounds as well as their pharmacokinetics and drug-likeness (Daina et al,2017). In-house sophisticated approaches like the BOILED-Egg, iLOGP, and Bioavailability radar are among the prognostic models for physicochemical characteristics, pharmacokinetics, etc. that are made available through this platform for free. Using SwissADME, we eliminated a few compounds in accordance with Lipinski's rule of five parameters. The molecular weight of a certain chemical must be less than 500 Da, it must have high lipophilicity (Log P value must be less than 5), H-bond acceptors must be fewer than 10, and H-bond donors must be less than 5, all of which are requirements for a ligand. Any compound with two or more infractions was disqualified from further investigation.

Preparation of Target protein:

As the TMPRSS4 protein structure is not available in Protein data bank (PDB) database, the protein structure was built by homology modelling method by employing Swissmodel tool. The input data for the target protein was retrieved from UniProtKB, PubMed:24434139 (Min et al,2014). This input data serving as a query was used to search evolutionary related protein structures against the SWISS-MODEL template library (SMTL) (Biasini et al,2014). It executes this task using Basic local alignment search tool (BLAST) (Altschul et al,1997; Camacho et al,2009) and HMM-HMM-based lightning- fast iterative sequence search (HHblits) (Remmert et al, 2011). After this, the templates are ranked according to expected quality of the resulting model and are shown in tabular form with descriptive set of features. For the selected templates, 3D protein structure with the catalytic domain and the external domain are designed and after the estimation of the model quality the structure was used for docking using iGEMDOCK (Waterhouse et al, 2018).

Ligand Preparation:

Majority of the ligands were taken from PubChem database (<https://pubchem.ncbi.nlm.nih.gov/>) and Indian Medicinal Plants, Phytochemistry and Therapeutic (IMPPAT) database (<https://cb.imsc.res.in/imppat/>). The ligands were downloaded in .sdf format, and then were converted into .mol2 3D file using the chimera software. Ligand input files for docking was prepared using AutoDock Tools and saved as pdbqt files.

Docking Studies:

In order to accurately analyze the intermolecular interactions and the binding modes between the protein TMPRSS4 and the specific ligand, the molecular docking studies were applied to the selected ligands with the help of iGEMDOCK, A Graphical Environment for Recognizing Pharmacological Interactions and Virtual Screening. iGEMDOCK initially offers interactive interfaces for setting up the screening compound library as well as the target protein's binding site. The in-house docking tool GEMDOCK is then used to dock each compound in the library into the binding site (Yang & Chen,2004). Then, iGEMDOCK creates profiles of electrostatic (E), hydrogen-bonding (H), and van der Waals (V) interactions between proteins and compounds and then predicts the pharmacological interactions and groups the screening compounds for the post-screening analysis based on these profiles and compound structures. Finally, by merging the pharmacological interactions and energy-based scoring function of GEMDOCK, iGEMDOCK ranks and visualises the screening compounds (Hsu et al,2011)

3.5. Molecular Dynamics Simulation :

MD simulations were done by using Desmond (Desmond, Schrödinger, 2015) with Optimized Potentials for Liquid Simulations (OPLS) 2005 force field. Initially, the prepared structures were imported in Desmond system builder and solvated with the TIP3P water. Orthorhombic periodic boundary conditions were set up to specify the shape and size of the repeating unit buffered at 10 Å distances and to neutralize the system electrically, counter NA⁺/Cl⁻ ions were added based on the system charge and were placed randomly in the solvated system. The system was subjected to energy minimization with a maximum of 5000 steps or until a gradient threshold of 25 kcal/mol/Å was reached by applying a hybrid method of the steepest descent. This process was followed by Limited-memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) algorithms, till a convergence threshold of 1 kcal/mol/Å was touched. The minimized systems were equilibrated by applying the default protocol of NVT and NPT. Further, both systems were taken into MDS for 10 ns with default relaxation protocol followed by periodic boundary condition with a number of atoms, pressure, and temperature (NPT) ensemble and temperature, atmospheric pressure adjusted to 310 K, 1.013 atm respectively.

3. Results and Discussions

Homology Modelling

Human transmembrane serine protease (TMPRSS4) contains 437 amino acids and has a molecular mass of about 48,246 Da. The physicochemical characterization of this protein unveiled that it has a greater number of residues that are negatively charged compared to the positively charged.

Homology modelling was performed for the target sequence of external domain and catalytic domain using the Swiss-model server (<https://swissmodel.expasy.org/>) (Waterhouse et al.,2017; Bienert et al.,2018). It is a fully automated protein structure homology-modelling server which pioneered the field of automated modeling starting in 1993 and is one of the most widely-used free web-based automated modeling facility as of today (Schwede et al.,2003). High resolution 3D structures for TMPRSS4 were generated using Insilco prediction methods. After completing preliminary work for catalytic domain of TMPRSS4, it showed 47% homology concluding that it is sufficient to carry out homology modelling of the 3D structure of TMPRSS4 for extracellular and catalytic domain (Figure 1).

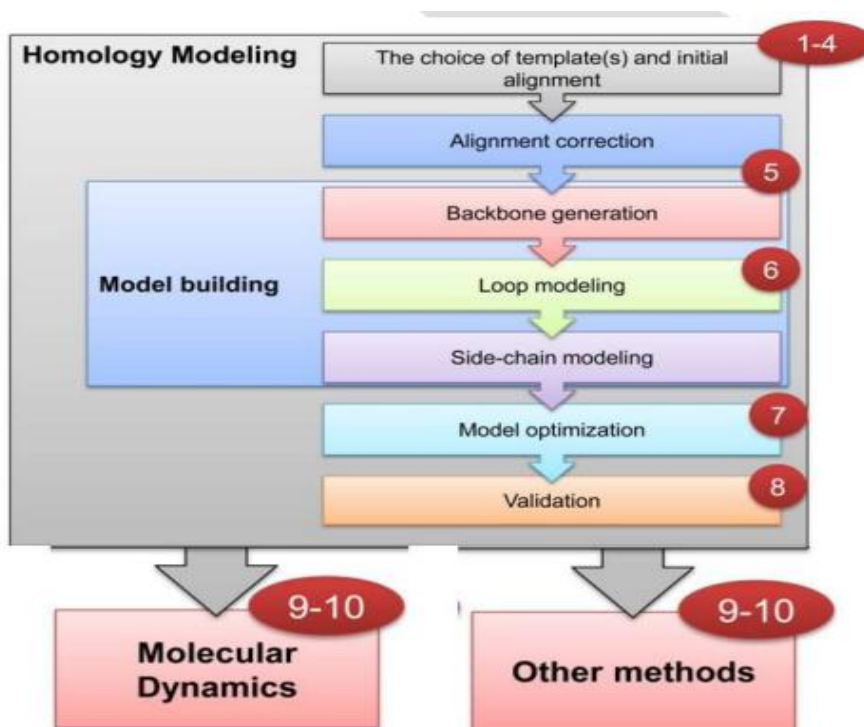
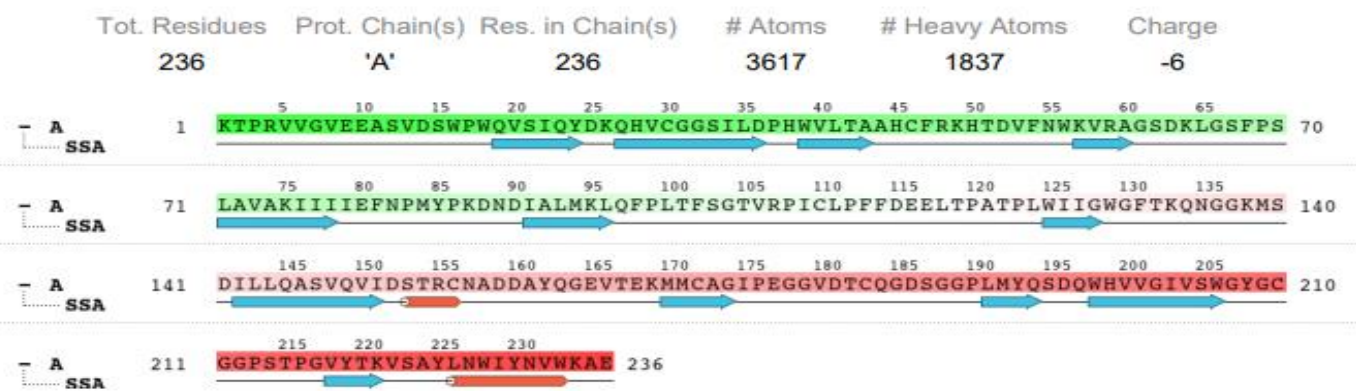


Figure 1: Schematic illustration of Homology modelling

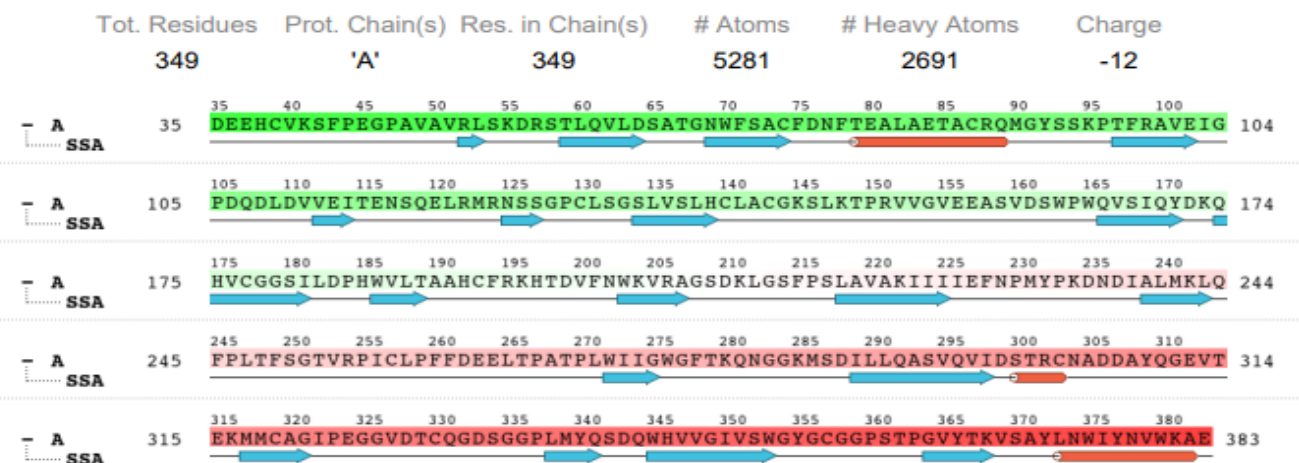
Protein-Protein BLAST was performed for the target sequence against PDB database to find suitable template for homology modelling. Then on the basis of the criteria of maximum identity with high score and lower e-value, PDB id: 7MEQ was selected as the template for modelling the catalytic domain and external domain of TMPRSS4. Sequence alignment between the target and template sequences based on structural information revealed the existence of 42% similarity for catalytic domain and 39% similarity for external domain, with GMQE score of 0.77 and QMEANDisCo Global score of 0.73 +/- 0.05 for homology modelled catalytic domain and GMQE score of 0.68 and QMEANDisCo Global score of 0.68 +/- 0.05 for homology modelled external domain.

Protein Information



(a)

Protein Information



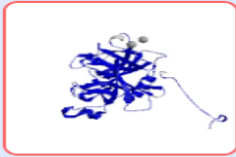
(b)

Figure 2: (a) Catalytic domain sequence; (b) External domain sequence of TMPRSS4

ITSort Coverage GMQE QSQE Identity Method Oligo State Ligands

7meq.1.A Transmembrane protease serine 2
Crystal structure of human TMPRSS2 in complex with Nafamostat

0.79 - 44.09 X-ray, 2.0Å monomer ✓ 1 x NAG-NAG, 1 x GBS



Method X-RAY DIFFRACTION 1.95 Å
Found By HHblits
Seq Similarity 0.42
Biounit Oligo State monomer
Target Prediction It is only possible to build a monomer.

Build Model

```

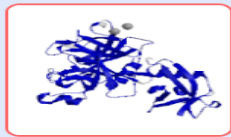
* Target GPCLSGSLVSLHCLACGKSL--KTPRVVGVVEEASVDSWFWQVSIQYDKQHVCGGSIIDPHWVLTAAHCFR 68
7meq.1.A ACSSKAVVSLRRCIACGVNLDNDKIVGGESALPGAWFWQVSLHVQNVHVCGGSIITPEWIVTAAHCFV 190
Target KHTD-VFNHKVVRAGSDKLG---SFPSLAVAKIIIEFNPMYPKNDIALMKLQFPLIFSGTIVRPICLPFF 134
7meq.1.A KPLNNPWHWTAFAGILRQSFMEYAGAGYQVEKVISHPNYDSKTKNNDIALMKLQKPLTFNDLVKPVCLPNE 260
Target DEELTPATFLWIIIGWFTKQNGGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMMCGAIPFEGGVDTCC 204
7meq.1.A GMMLOPEPQCWISGWGATEEKG-KTSEVLNAAKVVLIETQRCSNRYVYDNLITPAMICAGFLOGNVDSQC 229
Target GDSGGPLMYQS-DQWHVVGIVSWGYCGGPESTPGVYIKVSAAYLNWIYNVWKAEL 257
7meq.1.A GDSGGPLVTSKNNIWWLIGDTSWGGCAKAYRFEGVYGNVMVFTDWIYRQMRAD- 282
    
```

(a)

ITSort Coverage GMQE QSQE Identity Method Oligo State Ligands

7meq.1.A Transmembrane protease serine 2
Crystal structure of human TMPRSS2 in complex with Nafamostat

0.68 - 38.59 X-ray, 2.0Å monomer ✓ 1 x NAG-NAG, 1 x GBS



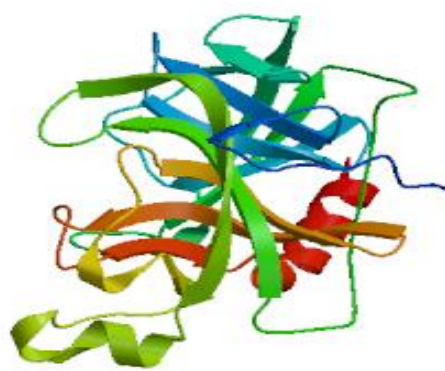
Method X-RAY DIFFRACTION 1.95 Å
Found By HHblits
Seq Similarity 0.39
Biounit Oligo State monomer
Target Prediction It is only possible to build a monomer.

Build Model

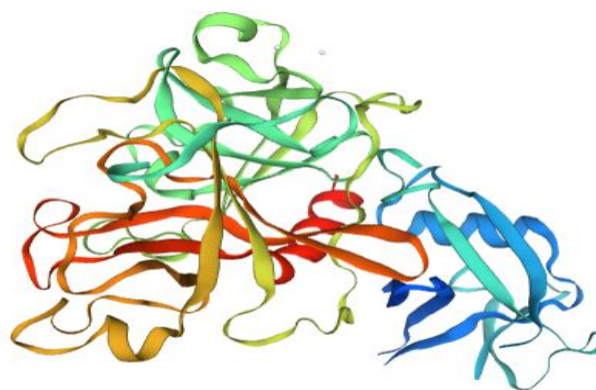
```

* Target KVILDKVYFLCGQPLHFIPRKQLCDGELDCPLGDEEHCVKRSFPEGPAVAVRLSKDRSTLQVLD SATGNW 70
7meq.1.A ---L---C---S---I---N---C---D---G---V---P---G---E---D---V---C---V---L---S---L---L---Q---V---S---S---K---K---W 60
Target FSACFDNFTFALAETACRQMGYSSKPTFRAVEIGFDQDLDVVEITE--NSQELMRNSSGGPCLSGSLVS 127
7meq.1.A LDCACGVNLDNDKIVGGESALPGAWFWQVSLHVQNVHVCGGSIITPEWIVTAAHCFV KPLNNPWHWTAFAGILRQSFMEYAGAGYQVEKVISHPNYDSKTKNNDIALMKLQKPLTFNDLVKPVCLPNE 200
Target LHCLACGKSL--KTPRVVGVVEEASVDSWFWQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHT-DVFNWK 204
7meq.1.A LDCACGVNLDNDKIVGGESALPGAWFWQVSLHVQNVHVCGGSIITPEWIVTAAHCFV KPLNNPWHWTAFAGILRQSFMEYAGAGYQVEKVISHPNYDSKTKNNDIALMKLQKPLTFNDLVKPVCLPNE 200
Target VRAGSDKLG---SFPSLAVAKIIIEFNPMYPKNDIALMKLQFPLIFSGTIVRPICLPFFDEELTPATFL 271
7meq.1.A LDCACGVNLDNDKIVGGESALPGAWFWQVSLHVQNVHVCGGSIITPEWIVTAAHCFV KPLNNPWHWTAFAGILRQSFMEYAGAGYQVEKVISHPNYDSKTKNNDIALMKLQKPLTFNDLVKPVCLPNE 200
Target WIIGWFTKQNGGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMMCGAIPFEGGVDTCCGGDSGGPLMYQ 341
7meq.1.A WISGWGATEEENG-KTSEVLNAAKVVLIETQRCSNRYVYDNLITPAMICAGFLOGNVDSQC GGDSGGPLVTSKNNIWWLIGDTSWGGCAKAYRFEGVYGNVMVFTDWIYRQMRAD- 339
Target S-DQWHVVGIVSWGYCGGPESTPGVYIKVSAAYLNWIYNVWKAEL 344
7meq.1.A S-DQWHVVGIVSWGYCGGPESTPGVYIKVSAAYLNWIYNVWKAEL 344
    
```

(b)



(c)

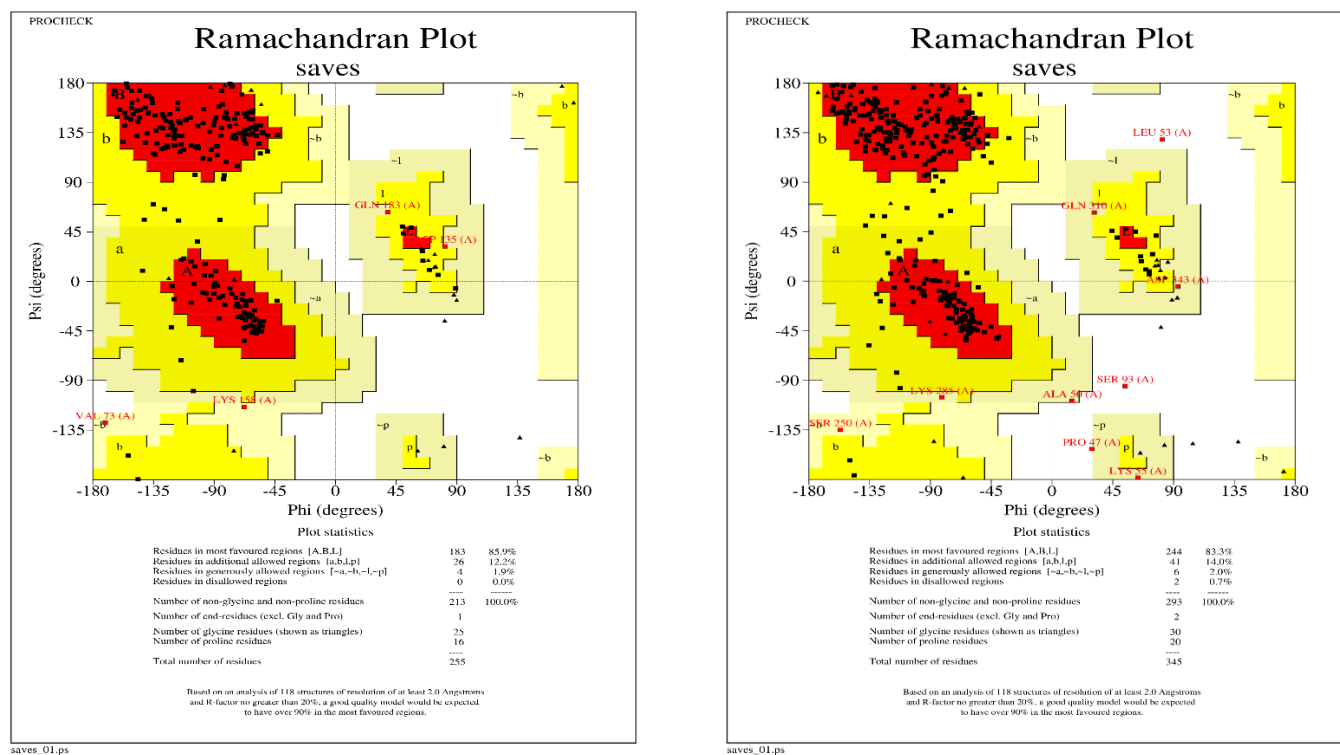


(d)

Figure 3: Alignment of target sequence and template sequence using structural information. The pink colored regions show the highly conserved regions (a) catalytic domain; (b) external domain; Homology modelled structures (c) catalytic domain; (d) external domain.

Validation of the 3D modeled structure

The Ramachandran contour map generated from PROCHECK analysis of the two-homology modelled structure revealed that homology modelled structure of the catalytic domain had 85.9% of residues in the favored region and 12.2% residues in the allowed region whereas for homology modelled structure of the external domain had 83.3% residues in the favored region and 14.0% residues in the allowed region suggesting its acceptable PHI-PSI statistics (Figure 4 a, b) and 3D structure reliability.



(a)

(b)

Figure 4: Ramachandran plot (a) Catalytic domain; (b) External Domain

Binding site Analysis

Computed Atlas of Surface Topography of proteins (CASTp) and binding site module of DS 4.0 was used for predicting the binding sites of homology modeled protein structure. Out of ten obtained possible results, the highest-ranked binding sites, with a surface area of 290.428 Å² and volume of 260.33 Å³ and with a surface area of 394.631 Å² and volume of 208.951 Å³, were selected for catalytic domain and external domain (Table 1) (Wei Tian et al, 2018; Binkowski, T, et al, 2003).

Table 1: Active site analysis results

	Domain	Area (SA) Å ²	Volume (SA) Å ³
1	Catalytic	290.428	260.33
2	External	394.631	208.951

Docking Based Virtual screening In this study compounds selected by two approaches one, using scaffold of 5 Hipsin inhibitors and one TMPSS4 inhibitor (Figure 5) filtered from 120 million compounds of PubChem, second, Phytochemicals were filtered by applying the scaffold filter from The Indian Medicinal Plants, Phytochemistry and Therapeutics 2.0 (Sunghyun Kan et al., 2013; Ilamathi et al., 2016).

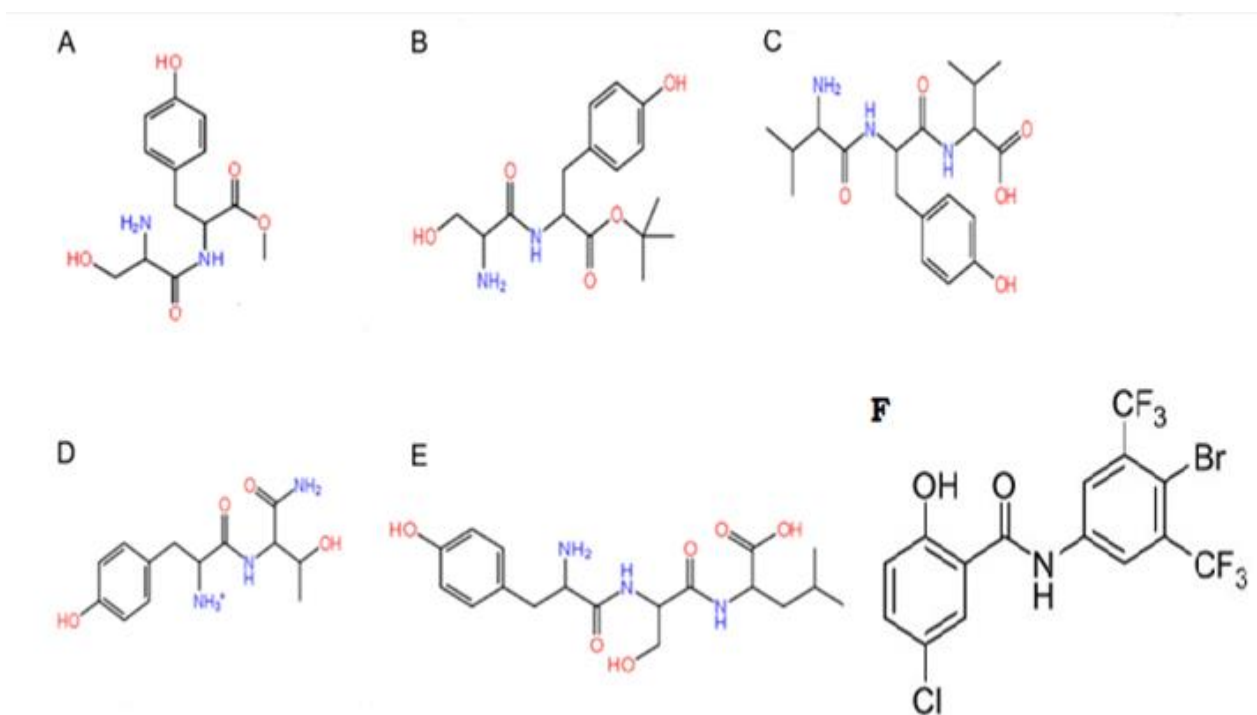


Figure 5: A-E compounds were Hipsin inhibitors, F compound was known TMPRSS4 inhibitors

Based on Hipsin inhibitors, a total of 2229 compounds were selected after which based on Lipinski rule 1640 compounds were filtered for docking (final.exe) against the homology modelled structure of both catalytic and external domains. Using the second approach a total of 550 compounds were taken for primary docking as well as virtual screening using iGEMDOCK, an automated graphical program which was used for the combined study of docking, screening, and post-screening investigation (Hsu et al, 2011). After primary docking and virtual screening, initially the top100 compounds were subjected for precision docking at both site using Autodock, a software designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure. The secondary docking top 10 compounds were selected from both of the databases for extracellular and catalytic domain site for biological studies (Table 2,3,4,5).

Interactions between TMPRSS4 and the top candidate ligands

The in-silico methods applied to the above experiment showed binding of the ligand to the important amino acid residues of the target domains and showed successful binding and fit within the external as well as internal domain. The top interactions have been discussed briefly and individually.

Interaction between external domain and top 10 phytochemicals from the IMPPAT database: **IMPHY001655** showed the highest binding affinity with the active sites of homology modelled external domain with a docking score of -7.4 K Cal/mol and formed two conventional Hydrogen bonds with LYSA:381, LEUA:121 and two Pi-Donor Hydrogen Bonds (Figure 6a).

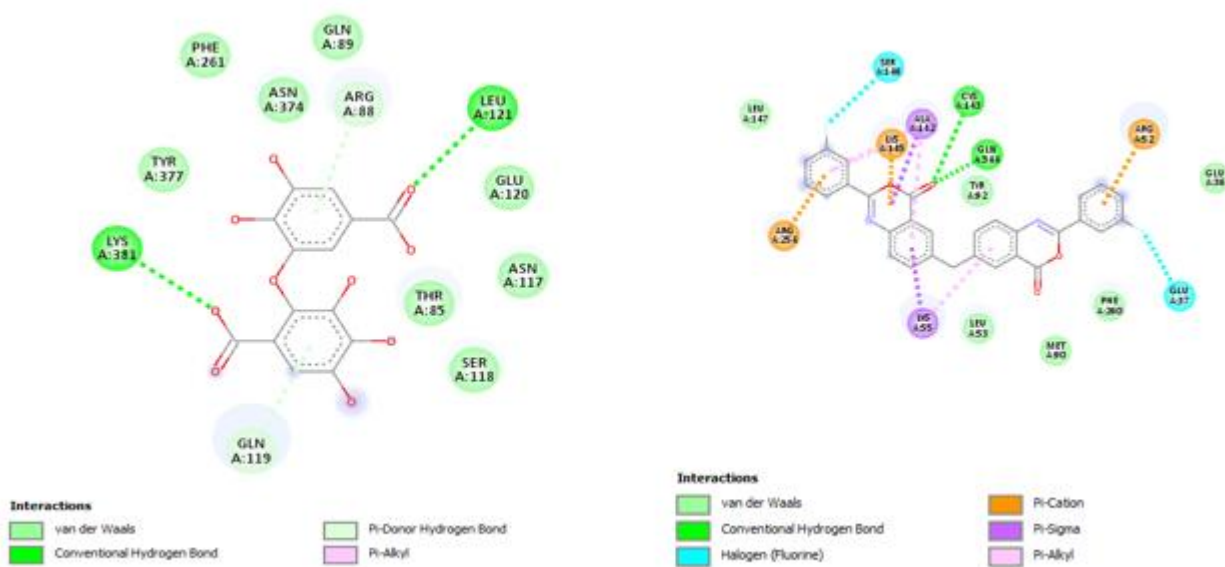


Figure 6: 2D Image of IMPHY001655 interaction with the external domain b) 2D Image of PubChem SID: 5770901 interactions with the external domain

IMPHY007825 also showed the second highest binding affinity towards the active site with a docking score of -7.4 K Cal/mol and formed six conventional Hydrogen bonds, two Carbon hydrogen bonds and one Pi-Donor Hydrogen Bond. Other interactions include one Pi-Alkyl interaction. **IMPHY001066** showed a docking score of -7.2 K Cal/mol and formed three conventional Hydrogen Bonds and. Other interactions include a Pi-Pi T shaped interaction.

Interaction between external domain and top 10 compounds from the PubChem database: PubChem SID: 5770901 showed the highest docking score of -9.0 K Cal/mol and formed 2 conventional hydrogen bonds with CYS A:143, GLNA:344. Additional interactions include two Halogen bonds, three Pi-Cation bonds and two Pi-Sigma bonds (Figure 6b).

PubChem SID: 101216213 showed the second highest binding affinity with the active sites of the external domain, showing a docking score of -7.8 K Cal/mol and formed two conventional Hydrogen Bonds and one Pi-Donor Hydrogen Bond. Other interactions include a Pi-Sigma interaction. **PubChem SID: 10875127** showed the third highest binding affinity with a docking score of -7.6 K Cal/mol forming 4 conventional Hydrogen Bonds and one carbon hydrogen Bond. Additional interactions include a Sulfur-X interaction and a Pi-Cation Interaction.

Interaction between catalytic domain and top 10 phytochemicals from the IMPPAT database:

IMPHY003078 showed the highest docking score of -9.3 K Cal/mol and formed one conventional Hydrogen Bond with ASPA:206. Additional interaction includes one Unfavorable Donor-Donor, one Amide-Pi Stacked and one Pi-Alkyl bond (Figure 7a).

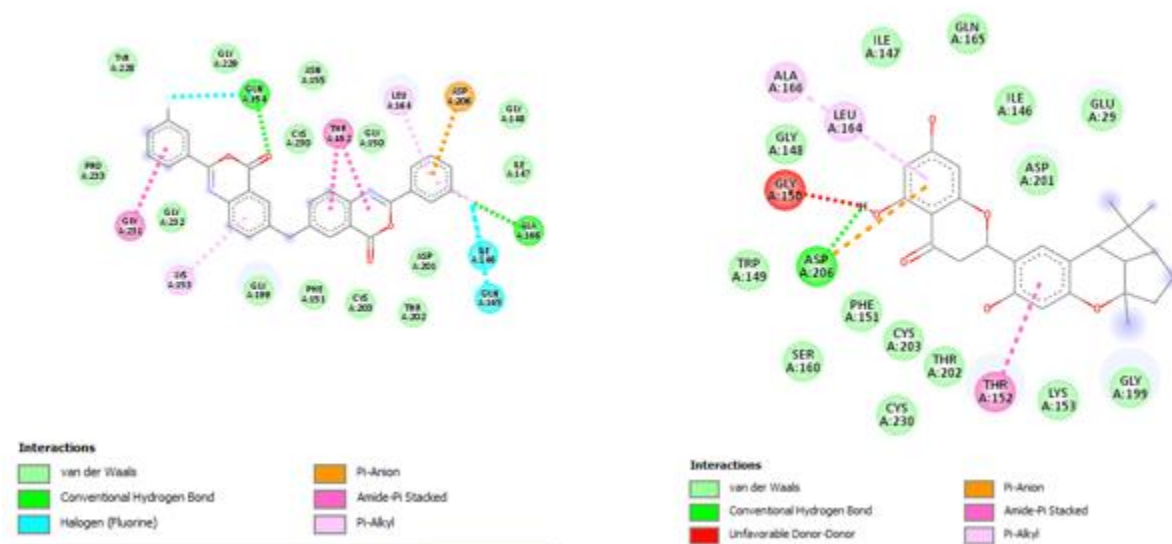


Figure 7a: 2D Image of IMPHY003078 interaction with the catalytic domain b) 2D Image of PubChem SID: 5770901 interactions with the catalytic domain

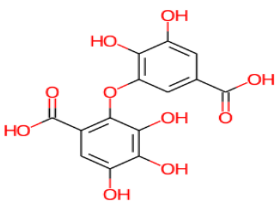
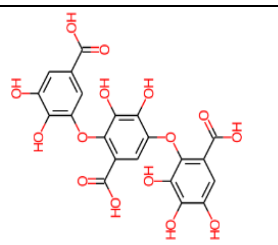
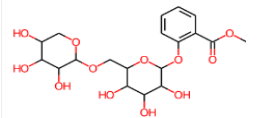
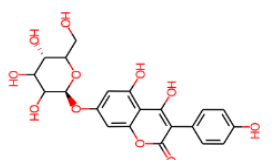
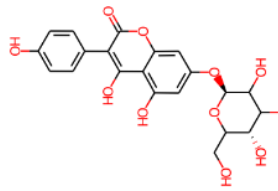
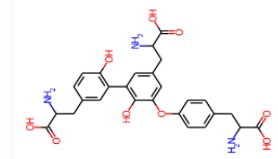
IMPHY007888 showed the second highest docking score of -9.1 K Cal/mol and formed one conventional hydrogen bond and one carbon hydrogen bond. Other interactions include one Pi-Anion, one Pi-Sigma and one Pi-Alkyl. **IMPHY013655** showed the third highest binding affinity towards the active site of the target with a docking score of -8.5 K Cal/mol and formed four conventional hydrogen bonds. Additional interactions include a Pi-Anion interaction.

Interaction between catalytic domain and top 10 compounds from the PubChem database: PubChem SID: 5770901 showed the highest binding affinity for the target with a docking score of -10.1 K Cal/mol and formed two conventional hydrogen bonds with GLNA:154, ALAA:166. Other interactions include a Halogen bond with fluorine, a Pi-Anion bond, one Amide-Pi Stacked and one Pi-Alkyl interaction (Figure 7b).

PubChem SID: 101216212 showed a binding score of -8.4 K Cal/mol and formed six conventional hydrogen bonds. Some of the additional interactions include two Pi-Anion bonds and one Pi-Sigma. **PubChem SID: 101216217** showed a docking score of -8.4 K Cal/mol and formed one conventional hydrogen bond and one carbon hydrogen bond. Other interaction includes one Pi-Anion, one Alkyl and one Pi-Alkyl bond.

IMPHY001655 from the IMPPAT database and 5770901 from the PubChem database showed the highest affinity for active sites of external domain whereas IMPHY003078 from the IMPPAT database and 5770901 from the PubChem database showed the highest affinity for active sites of catalytic domain.

Table 2: Docking and Amino acid interactions of the top 10 phytochemical compounds with the external domain.

S. No		Energy value (Kcal/mol)	Amino Acid Interaction
1	 IMPHY001655	-7.48	TYRA:377, PHEA:261, ASNA:374, GLNA:89, GLUA:120, ASNA:117, THRA:85, SERA:118 – Van der Waals; LYSA:381, LEUA:121 – Conventional Hydrogen Bond; GLNA:119, ARGA:88 – Pi-Donor Hydrogen Bond.
2	 IMPHY007825	-7.44	PHEA:78, ILEA:114, ASNA:378, TYRA:377, ASNA:374, GLNA:89, LEUA:121, GLUA:120 – Van der Waals; GLUA:84, LYSA:381, ASNA:117, SERA:118, GLNA:119, THRA:85 – Conventional Hydrogen Bond; GLUA:116, LEUA:82 – Carbon Hydrogen Bond; ARGA:88 – Pi-Donor Hydrogen Bond; ALAA:81 – Pi-Alkyl.
3	 IMPHY001066	-7.28	LEUA:373, TYRA:377, GLNA:89, GLUA:120, LEUA:121, SERA:118, THRA:85, ASNA:117, ALAA:81, LYSA:381 – Van der Waals; GLNA:119, ASNA:374, ARGA:88 – Conventional Hydrogen Bond; PHEA:261 – Pi-Pi T Shaped.
4	 IMPHY013673	-7.16	PHEA:261, GLNA:89, TYRA:377, ASNA:378, THRA:85, SERA:118, GLUA:84, LEUA:82, GLUA:116 – Van der Waals; GLNA:119, LYSA:381, ARGA:88 – Conventional Hydrogen Bond; ASNA:374, ASNA:117 – Carbon Hydrogen Bond; ALAA:81 – Pi-Alkyl.
5	 IMPHY 013655	-7.56	LEUA:141, HISA:139, ALAA:142, VALA:51, HISA:346, PHEA:260, SERA:342, LYSA:41, ASPA:35, GLUA:37, GLNA:344, LYSA:55, LEUA:53 – Van der Waals; CYSA:140, LYSA:145, GLUA:36, ASPA:343, TYRA:92 – Conventional Hydrogen Bond; ARGA:52 – Carbon Hydrogen Bond.
6	 IMPHY015916	-7.05	ASNA:378, TYRA:377, GLNA:89, PHEA:261, SERA:118, GLUA:116, LEUA:82, GLUA:84, PHEA:78 – Van der Waals; THRA:85, GLNA:119, ASNA:374, ASNA:117, ALAA:81, GLUA:80 – Conventional Hydrogen Bond; LYSA:381, ARGA:88 – Pi-Cation.

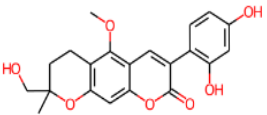
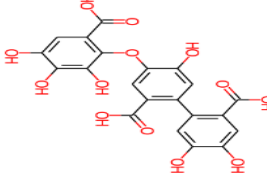
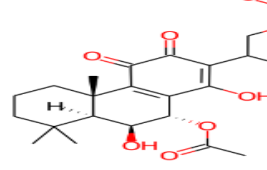
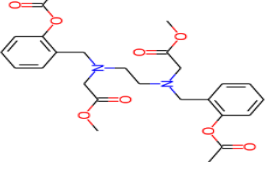
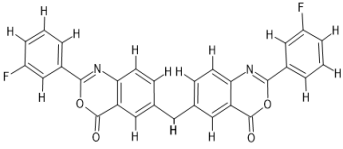
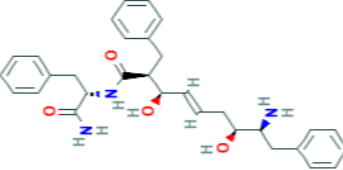
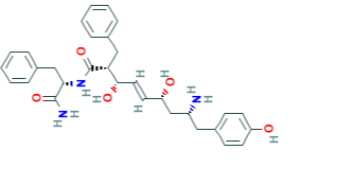
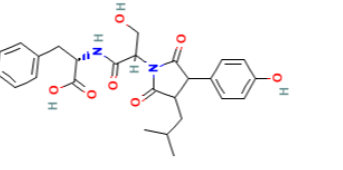
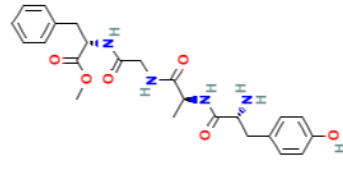
7	 <p>IMPHY006776</p>	-6.95	LYSA:55, ASPA:343, GLNA:344, TYRA:92, META:90, PHEA:260, GLUA:37, GLUA:36 – Van der Waals; SERA:342, ARGAs:52, LEUA:53 – Conventional Hydrogen Bond; ASPA:35 – Pi-Cation.
8	 <p>IMPHY014462</p>	-6.88	ASNA:378, TYRA:377, ASNA:374, GLNA:89, GLNA:119, GLUA:120, SERA:118, THRA:85, ASNA:117, GLUA:84 – Van der Waals; LYSA:381, ALAA:81 – Conventional Hydrogen Bond; ARGAs:88 – Pi-Cation.
9	 <p>IMPHY008511</p>	-6.55	ASPA:35, ASPA:343, GLUA:36, CYSA:39, HISA:38, GLUA:263, PHEA:260, HISA:346 – Van der Waals; ARGAs:52, SERA:342, LYSA:41 – Conventional Hydrogen Bond; GLUA:37 – Carbon Hydrogen Bond.
10	 <p>IMPHY013456</p>	-5.68	LYSA:145, GLNA:344, ASPA:343, ASPA:35, CYSA:39, LYSA:41, GLUA:263, GLUA:37, VALA:51 – Van der Waals; SERA:342, ARGAs:52 – Conventional Hydrogen Bond; HISA:346, GLUA:36, HISA:38 – Carbon Hydrogen Bond; META:90 – Pi-Sulfur; PHEA:260 – Pi-Pi T Shaped.

Table 3. Docking and Amino acid interactions of the 10 PubChem compounds with External domain

S. No	Compounds (PubChem id)	Energy value (Kcal/mol)	Amino Acid Interaction
1	5770901 	-9	LEUA:147, LEUA:53, META:90, PHEA:260, GLUA:36, TYRA:92 – Van der Waals; CYSA:143, GLNA:344 – Conventional Hydrogen Bond; SERA:146, GLUA:37 – Halogen (Fluorine); ARG:254, LYA:145, ARG:52 – Pi-Cation; ALAA:142, LYA:55 – Pi-Sigma
2	101216213 	-7.8	LEUA:373, GLNA:89, PHEA:261, ASNA:374, TYRA:377, GLNA:119, THRA:85, GLUA:116, LEUA:82, PHEA:78, GLUA:80, SERA:118, LYSA:381 – Van der Waals; ARG:88, GLUA:84 – Conventional Hydrogen Bond; ALAA:81 – Pi-Sigma; ASNA:117 – Pi-Donor Hydrogen Bond
3	10875127 	-7.6	GLUA:36, GLUA:37, HISA:38, CYSA:39, LYSA:41, VALA:51, ASPA:343, LYSA:145, PHEA:260, HISA:346, TYRA:92 – Van der Waals; GLUA:263, SERA:342, GLNA:344, LEUA:53 – Conventional Hydrogen Bond; ARG:52 – Carbon Hydrogen Bond; META:90 – Sulfur-x; ASPA:35, LYSA:55 – Pi-Cation
4	118986785 	-7.6	PHEA:261, GLNA:89, TYRA:377, THRA:85, ASNA:378, ASNA:117, GLUA:116 – Van der Waals; GLNA:119, LYSA:381, ASNA:374 – Conventional Hydrogen Bond; SERA:118 – Carbon Hydrogen Bond; ALAA:81, ARG:88 – Pi-Alkyl.
5	16396457 	-7.6	HISA:346, GLUA:263, LYSA:41, CYSA:39, HISA:38, GLUA:36, SERA:342, GLUA:37, ASPA:343, LYSA:145, TYRA:92, GLNA:61, SERA:54, META:90, ASPA:56 – Van der Waals; GLNA:344, LEUA:53, LYSA:55 – Conventional Hydrogen Bond; ARG:52 – Unfavorable Donor-Donor; TRPA:70 – Pi-Pi T Shaped; PHEA:260 – Pi-Pi Stacked

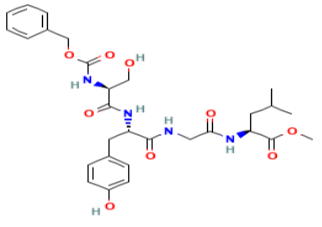
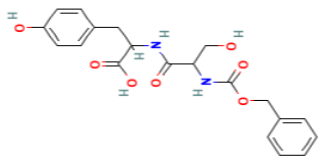
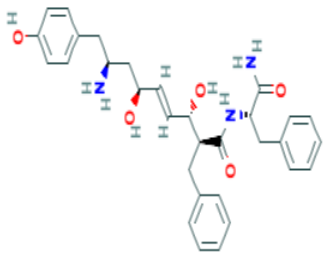
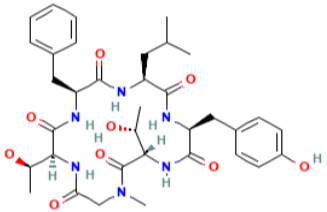
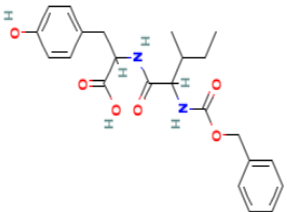
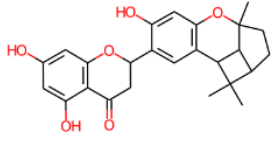
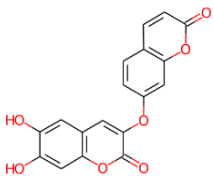
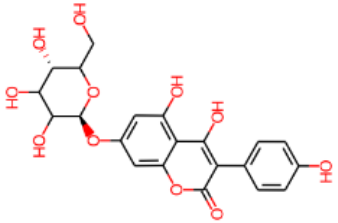
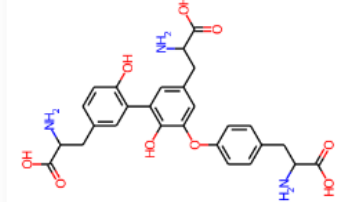
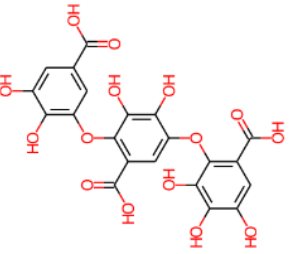
6	 <p>14802668</p>	-7.4	HISA:139, CYSA:140, TYRA:92, META:90, SERA:54, LYSA:41, CYSA:39, GLUA:263, HIA:346, ASPA:343, GLUA:37, GLUA:36, ASPA:35 – Van der Waals; LEUA:53, LYSA:145, ARGA:52, GLNA:344, SERA:342 – Conventional Hydrogen Bond; LYSA:55- Carbon Hydrogen Bond; Phea:260 – Pi-Pi Stacked; ALAA:142 - Pi- Alkyl
7	 <p>333329</p>	-7.4	GLYA:116, LEUA:82, THRA:85, SERA:118, ASNA:378, GLNA:89, PHEA:261, TYRA:377, LEUA:373 – Van der Waals; ALAA:81 – Pi-Sigma; PHEA:78 – Pi-Pi Stacked; ASNA:117 – Carbon Hydrogen Bond/Pi-Donor Hydrogen Bond
8	 <p>10995039</p>	-7.3	PHEA:261, GLNA:89, TYRA:377, ASNA:374, GLUA:84, THRA:85, GLUA:116, ILEA:114, LEUA:82 – Van der Waals; ASNA:378, GLNA:119 – Conventional Hydrogen Bond; SERA:118 -Carbon Hydrogen Bond; ASNA:117 – Pi Donor Hydrogen Bond; ALAA:81 – Pi-Sigma; PHEA:78 – Pi-Pi Stacked; ARGA:88, LYSA:381 – Pi-Cation.
9	 <p>118736526</p>	-7.3	TYRA:92, GLNA:344, LYSA:145, GLUA:37, PHEA:260, ARGA:52, HISA:38, ASPA:343, GLUA:36, LYSA:41 – Van der Waals; SERA:342, LYSA:55, ASPA:35, GLUA:263 – Conventional Hydrogen Bond; CYSA:39 – Carbon Hydrogen Bond.
10	 <p>333334</p>	-7.3	GLUA:116, LEUA:82, GLUA:84, SERA:118, ASNA:378, ASNA:374, PHEA:261, GLNA:89, TYRA:377 – Van der Waals; ALAA:81 – Pi-Sigma; PHEA: 78 - Pi-Pi Stacked; LYSA:381 – Alkyl; ASNA:117 – Pi-Donor Hydrogen Bond; GLNA:119, ARGA:88, THRA:85 – Conventional Hydrogen Bond.

Table 4: Docking and Amino acid interactions of the 10 Phytochemical compounds with Catalytic domain

S. No	Structure of the Compound IMPPACT database	Energy value (Kcal/mol)	Amino Acid Interaction
1	 IMPHY003078	-9.32	GLYA:199, LYSA:153, CYSA:230, THRA:202, CYSA:203, PHEA:151, SERA:160, TRPA:149, GLYA:148, ILEA:147, GLNA:165, ILEA:146, GLUA:29, ASPA:201 – Van der Waals; ASPA:206 – Conventional Hydrogen Bond; GLYA:150 – Unfavorable Donor-Donor; THRA:152 – Amide-Pi Stacked; ALAA:166, LEUA:164 – Pi-Alkyl.
2	 IMPHY00788	-9.15	VALA:28, GLNA:165, ASPA:201, THRA:152, CYSA:203, GLYA:150, PHEA:151, TRPA:149, GLYA:148, ILEA:147, GLLYA:199 – Van der Waals; ASPA:206 – Conventional Hydrogen Bond; VALA:200 – Carbon Hydrogen Bond; GLUA:29 – Pi-Anion; LEUA:164 – Pi-Sigma; ALAA:166 – Pi-Alkyl
3	 IMPHY013655	-8.52	GLNA:154, GLYA:232, CYSA:230, LYSA:153, CYSA:203, ASPA:201, PHEA:151, GLYA:150, GLYA:148, ILEA:147, GLNA:165, ALAA:166, THRA:152, PROA:233 – Van der Waals; GLYA:199, GLYA:231, LEUA:164, THRA:202 – Conventional Hydrogen Bond; ASPA:206 – Pi-Anion
4	 IMPHY015916	-8.35	GLYA:232, PROA:233, GLYA:199, LYSA:153, CYSA:230, GLNA:154, ASNA:155, CYSA:203, PHEA:151, SERA:160, GLYA:150, TRPA:149, LEUA:164, GLYA:148, THRA:152, TRPA:145, VALA:26, GLYA:27, SERA:167 – Van der Waals; ASPA:201, GLYA:231, ASPA:206, VALA:28, VALA:25, GLNA:165 – Conventional Hydrogen Bond; VALA:200 – Carbon Hydrogen Bond; GLUA:29 – Pi-Sigma; ALAA:166 – Pi-Alkyl.
5	 IMPHY007825	-8.05	ILEA:147, GLYA:148, TRPA:149, SERA:160, GLYA:150, THRA:152, PHEA:151, GLYA:231, THRA:202, ASNA:155, ALAA:166, GLUA:29, VALA:200, GLYA:199 – Van der Waals; LEUA:164, GLNA:165, CYSA:203, ASPA:201, CYSA:230 – Conventional Hydrogen Bond; GLNA:154 – Unfavorable Donor-Donor; ASPA:206 –

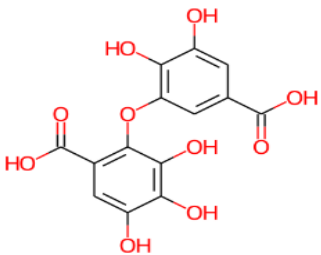
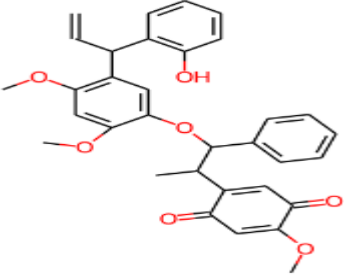
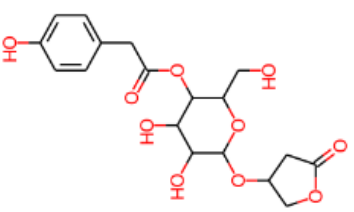
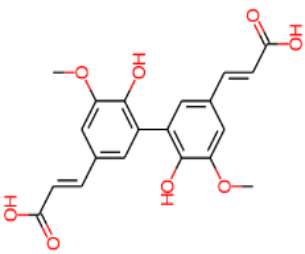
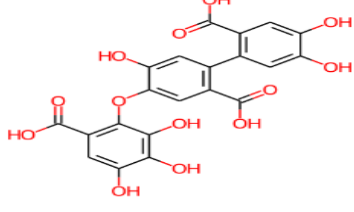
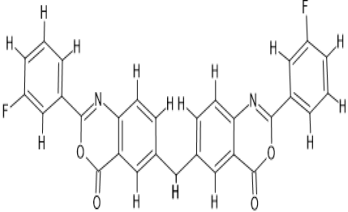
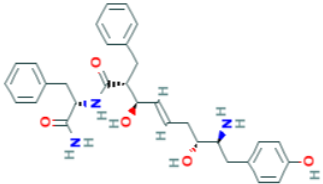
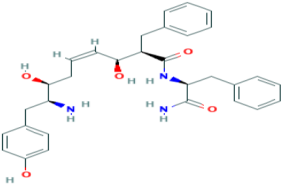
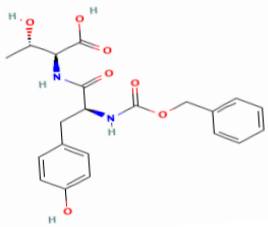
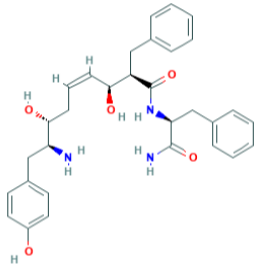
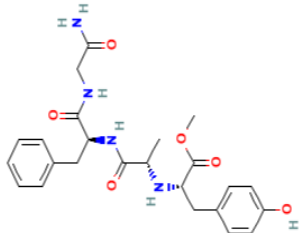
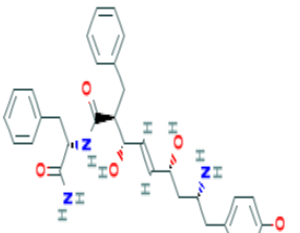
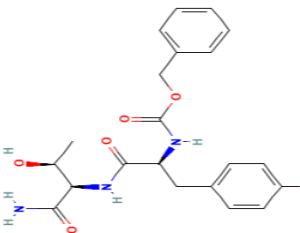
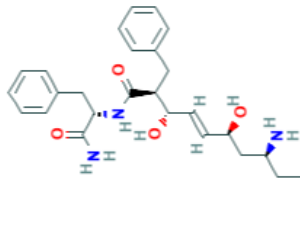
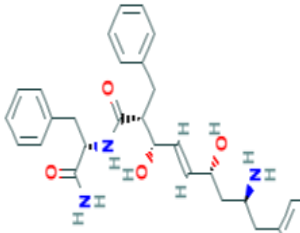
			Unfavorable Acceptor- Acceptor; LYSA:153 – Pi-Alkyl.
6	 <p>IMPHY001655</p>	-7.86	GLYA:199, VALA:200, GLYA:231, ASPA:201, THRA:202, CYSA:230, ALAA:166, CYSA:203, LYSA:153, ILEA:146, ILEA:147, GLNA:165, PHEA:151, TRPA:149, SERA:160, GLUA:29 – Van der Waals; LEUA:164, GLYA:148, GLYA:150 – Conventional Hydrogen Bond; THRA:152 – Carbon Hydrogen Bond; ASPA:206 – Pi-Anion
7	 <p>IMPHY007863</p>	-7.68	GLNA:165, ASPA:201, ILEA:146, ILEA:147, ASPA:206, GLYA:148, PHEA:151, GLUA:29, GLNA:154, GLYA:231 – Van der Waals; GLYA:199 – Conventional Hydrogen Bond; THRA:152, CYSA:230, THRA:202 – Carbon Hydrogen Bond; CYSA:203, LYSA:153 – Alkyl, LEUA:164, ALAA:166, VALA:200 – Pi-Alkyl
8	 <p>IMPHY013673</p>	-7.75	GLYA:231, LYSA:153, CYSA:203, PHEA:151, ASPA:206, ILEA:147, GLNA:165, ILEA:146, VALA:200 – Van der Waals; GLNA:154, CYSA:230, ASPA:201, GLYA:148 – Conventional Hydrogen Bond; THRA:152, GLYA:199 – Carbon Hydrogen Bond; LEUA:164, ALAA:166 – Pi-Alkyl
9	 <p>IMPHY005552</p>	-7.55	GLUA:29; ILEA:147, GLYA:148, SERA:160, TRPA:149, LEUA:164, PHEA:151, THRA:202, CYSA:203, GLYA:231, LYSA:153, GLYA:199, GLNA:165, TRPA:145 – Van der Waals; GLYA:150, ASPA:206, SERA:167 – Conventional Hydrogen Bond; ASPA:201 – Carbon Hydrogen Bond; THRA:152 – Amide-Pi Stacked; CYSA:230 – Alkyl; VALA:200, ALAA:166 – Pi-Alkyl.
10	 <p>IMPHY014462</p>	-7.65	GLUA:29, THRA:152, ALAA:166, GLNA:165, GLYA:148, ILEA:147, TRPA:149, GLYA:150, CYSA:203, VALA:200, PHEA:151, ASPA:201, LYSA:153, GLYA:199 – Van der Waals; ASPA:206 – Conventional Hydrogen Bond; LEUA:164 – Pi-Sigma.

Table 5: Docking and Amino acid interactions of the top 10 PubChem compounds with Catalytic domain

S. No	Compounds (PubChem id)	Energy value (Kcal/mol)	Amino Acid Interaction
1	 5770901	-10.1	TYRA:228, PROA:233, GLYA:232, GLYA:199, PHEA:15, CYSA:203, THRA:202, ASPA:201, ILEA:147, GLYA:148, GLYA:150, CYSA:230, ASNA:155, GLYA:229 – Van der Waals; GLNA:154, ALAA:166 – Conventional Hydrogen Bond; ILEA:146, GLNA:165 – Halogen (Fluorine), ASPA:206 – Pi-Anion; GLYA:231, THRA:152 – Amide-Pi Stacked; LYSA:153, LEUA:164 – Pi-Alkyl.
2	 101216212	-8.4	ILEA:195, GLYA:27, VALA:26, SERA:167, TRPA:145, GLNA:154, LYSA:153, CYSA:230, CYSA:203, GLYA:150, ILEA:147, LEUA:163, TRPA:149, ASPA:201, GLYA:199, ALAA:166, THRA:152 – Van der Waals; VALA:28, GLNA:165, VALA:25, PHEA:151, GLYA:148, LEUA:164 – Conventional Hydrogen Bond; GLUA:29; ASPA:206 – Pi-Anion; VALA:200 – Pi-Sigma.
3	 101216217	-8.4	GLYA:148, CYSA:203, ILEA:147, THRA:202, VALA:224, LEUA:164, ALAA:166, GLNA:165, GLUA:29, GLNA:154, GLYA:231, LYSA:153, GLYA:199, THRA:152, PHEA:151 – Van der Waals; ASPA:201 – Conventional Hydrogen Bond; VALA:203 – Carbon Hydrogen Bond; ASPA:206 – Pi-Anion; CYSA:230 – Alkyl; ILEA:146 – Pi-Alkyl.
4	 124437228	-8.4	SERA:160, TRPA:149, GLYA:148, GLYA:199, GLYA:231, CYSA:230, THRA:202, LYSA:153, CYSA:203, GLNA:154, PHEA:151, TRPA:145, VALA:25, GLNA:165, GLUA:29 – Van der Waals; GLYA:150, ASPA:201, THRA:152 – Conventional Hydrogen Bond; VALA:200 – Carbon Hydrogen Bond; ASPA:206 – Pi-Anion; SERA:167 – Pi-Donor Hydrogen Bond; LEUA:164 – Pi-Sigma; ALAA:166 – Pi-Alkyl.
5	 101216217	-8.3	GLYA:231, GLNA:154, CYSA:230, THRA:152, PHEA:151, GLYA:148, TRPA:149, GLYA:150, ALAA:166, GLUA:29, VALA:26, VALA:28. GLYA:27, VALA:25 – Van der Waals; CYSA:203, THRA:202, ASPA:201, GLYA:199, GLNA:165 – Conventional Hydrogen Bond; ASPA:206 – Pi-Anion; LEUA:164 – Pi-Sigma; LYSA:153 – Pi-Alkyl.

6	 <p>10412432</p>	-8.3	TRPA:145, VALA:26, VALA:28, GLUA:29, ILEA:146, ILEA:147, GLYA:148, THRA:202, PHEA:151, CYSA:230, GLYA:231, LYSA:153, THRA:152, GLYA:199, SERA:167 – Van der Waals; VALA:25, GLNA:165, ASPA:201, GLNA:154, CYSA:203 – Conventional Hydrogen Bond; VALA:200 – Carbon Hydrogen Bond; ASPA:206 – Pi-Anion; ALAA:166, LEUA:164 – Pi-Alkyl.
7	 <p>11082013</p>	-8.3	THRA:202, CYSA:203, GLYA:199, CYSA:230, GLYA:231, GLNA:154, ILEA:146, ILEA:147, GLYA:148, GLNA:165, GLUA:29, TRPA:145, VALA:26, SERA:167, THRA:152 – Van der Waals; PHEA:151, ASPA:201, VALA:25, VALA:28 – Conventional Hydrogen Bond; VALA:200 – Carbon Hydrogen Bond; ASPA:206 – Pi-Anion; LYSA:153, LEUA:164, ALAA:166 – Pi-Alkyl.
8	 <p>125180672</p>	-8.3	TRPA:145, VALA:25, GLNA:165, PHEA:151, CYSA:203, THRA:202, GLNA:154, CYSA:230, GLYA:231, LYSA:153, THRA:152, TRPA:149, SERA:160, GLYA:148, GLUA:29 – Van der Waals; GLYA:199, ASPA:201, GLYA:150 – Conventional Hydrogen Bond; ASPA:206 – Pi-Anion; VALA:200 – Carbon Hydrogen Bond; SERA:167 – Pi-Donor Hydrogen Bond; LEUA:164 – Pi-Sigma; ALAA:166 – Pi-Alkyl.
9	 <p>10995039</p>	-8.1	SERA:167, TRPA:145, GLYA:148, ILEA:147, ASPA:206, GLYA:150, SERA:160, PHEA:151, LYSA:153, GLYA:199, ILEA:146, THRA:152, LEUA:164, ALAA:166, ILEA:195 – Van der Waals; VALA:28, VALA:25, GLNA:165, ASPA:201 – Conventional hydrogen Bond; GLUA:29 – Pi-Anion; VALA:200 – Pi-Sigma; CYSA:203 – Pi-Alkyl.
10	 <p>10995040</p>	-8.1	THRA:202, CYSA:203, CYSA:230, LYSA:153, ASPA:206, GLYA:148, ILEA:147, ILEA:146, GLYA:231, THRA:152, TRPA:145, VALA:26, ILEA:195, GLYA:27, GLYA:199 – Van der Waals; PHEA:151, GLNA:165, VALA:25, VALA:28 – Conventional Hydrogen Bond; ASPA:201 – Unfavorable Acceptor-Acceptor; GLUA:29 – Pi-Anion; SERA:167 – Pi-Sigma; LEUA:164, ALAA:166, VALA:200 – Pi-Alkyl.

3.5. Molecular Dynamics Simulation:

MD simulations were executed to confirm the binding energy and molecular level interactions determine in the molecular docking. Based on docking results SID5770901 compound showing good activity with both domains, hence this compound selected for MD simulations to checks stability of complex for further. MD simulations were carried out for 10 ns for ECD and CD of TMPRSS4 complexed with SID5770901. The Root mean square deviation (RMSD) and Root mean square fluctuations (RMSF), were assessed the intermolecular distance between the protein C α backbone and the ligand along the trajectory were calculated. RMSD of ECD backbone was observed to be fluctuating between 2.5 and 3 Å (Figure 8) throughout the simulated timescale except 4 Å deviation at 8 ns. RMSD of CD backbone was observed to be fluctuating between 2.0 and 3 Å (Figure 9) throughout the simulated timescale. Clearly showing that the protein is much stabilized in its activity in the presence of the ligand. RMSF of the Protein individual residue were also found to be co-ordination. In graph peaks indicate areas of the protein that fluctuate the most during the simulation. For the ECD RMSF value was high for starting 30 residues of protein, and the values were subsequently stable with minimal fluctuations at the end of the simulation period, for CD RMSF stable with minimal fluctuations throughout the simulation (Figure 8 &9).

Protein–ligand interaction between the compound and two domains presented as histogram maps in Figures 6, 7. Amino acids ASP-35, ARG_52, LEU_53, LYS_55, MET_90, TYR 92, PHE260, interact with through hydrophobic interactions, GLU37, Ser342 interacting with water bridges, while cys140, asp343, gln344 forming hydrogen bond but it fluctuating with time frame in ECD complex with 5770901 ligand. In CD-5770901 complex Gln 154 showing Hydrophobic interactions, Gly134, Ala166 showing hydrogen bond interactions with ligand throughout the simulation. Analysis hydrogen bonding interactions of complex swere rationalized that the steady nature of the lead compound (Figure 8 &9).

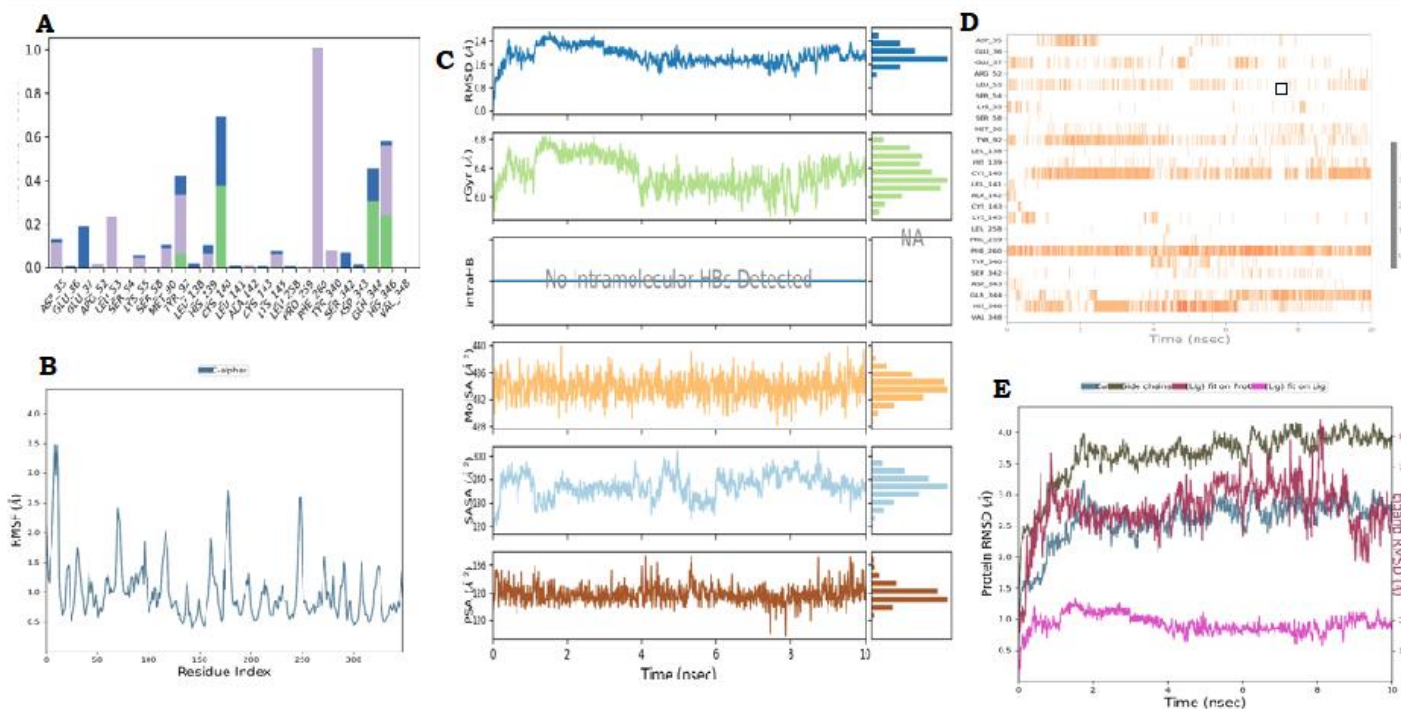


Figure 8 A) Molecular Interaction diagram of Extracellular domain complex with CID: 5770901. (B.) RMSF graph. (C) Ligand properties (D) Timeline representations of the Molecular Interaction (E) RMSD graph observed during the molecular dynamic simulation

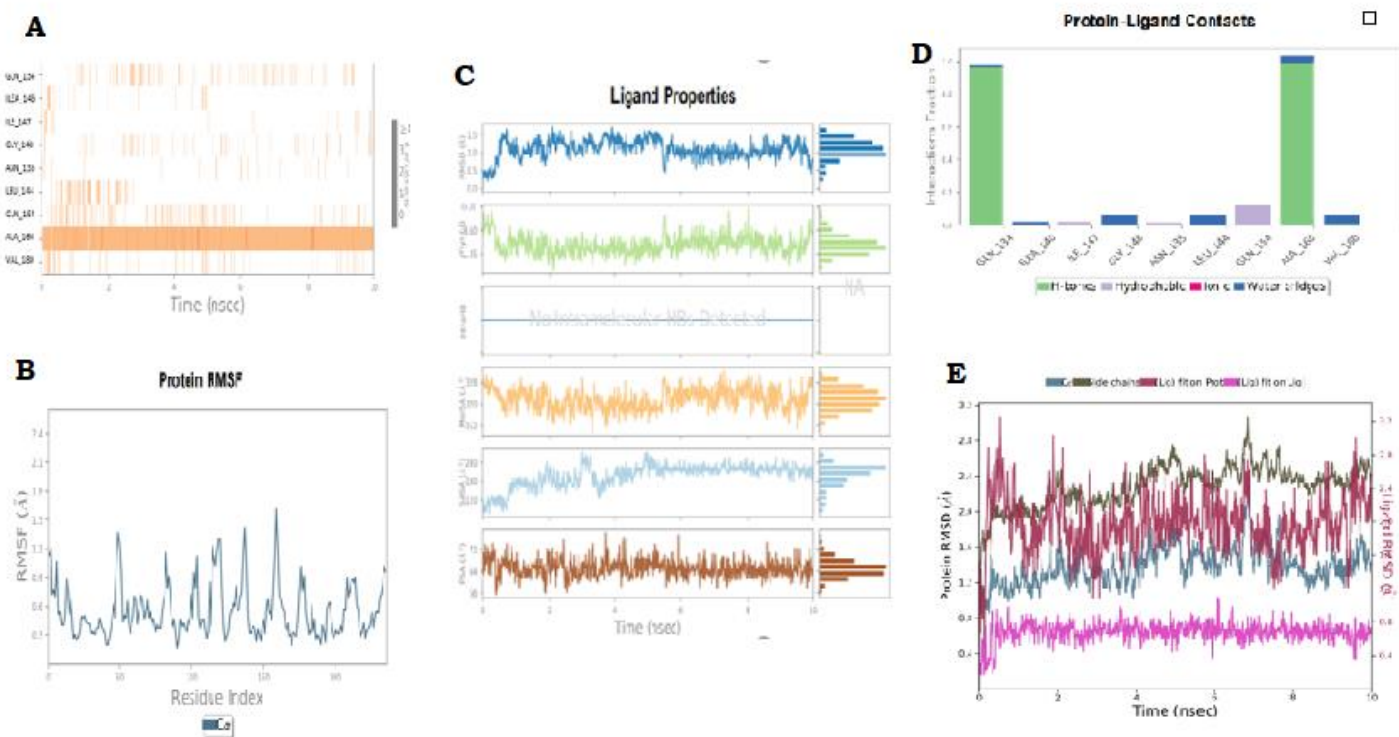


Figure 9 A) Molecular Interaction diagram of Catalytic domain complex with CID: 5770901. (B.) RMSF graph. (C) Ligand properties (D) Timeline representations of the Molecular Interaction (E) RMSD graph observed during the molecular dynamic simulation

Throughout the 10 ns trajectory, the ligand properties of the compound were measured using the following parameters: RMS, Polar Surface Area (PSA), Solvent Accessible Surface Area (SASA), Molecular Surface Area (MolSA) and Radius of Gyration (rGyr). The RMSD ranges from 0.8 Å to 2.5 Å after 2 ns it reaches 2.5 after that it was not much deviated throw-out simulation period in ECD complex. In CD complex mean RMSD value 1 Å it means ligand not deviated and showing stability between group of atoms (e.g. backbone atoms of a protein). The hardness of protein structure was estimated using the ROG (radius of gyration). Both complexes were found to be mildly stretched by maintaining a range of 6.0 to 7 Å throughout the simulation. For CD and ECD complexes values of SASA 120 to 240 Å², 120 to 300; PSA 60 to 74 Å² 128 to 136 Å² and MolSA of 312- 328 Å², 430-440 Å² respectively it helped to ensure its stability throughout MD simulation (Fig. 3† top). Analysis of hydrogen bonding, RMSF, RMSD, SASA, MolSA, PSA and RoG showed up steady nature of the 5770901 compound with both the domains and revealed that it was better potential inhibitor for TMPRSS4.

The docking and MD simulations studies were revealed that the binding affinity of 5770901 towards ECD and CD was highly stable with the proposed binding orientations.

Conclusion:

This is the first attempt to determine the 3D structure of the catalytic domain and extracellular domain of TMPRSS4 serine protease in humans. The 3D model generated using *Insilco* methods in this study provides necessary information for rational drug design. CID 5770901 compound was discovered through docking and MD simulations to have good activity against the catalytic and extracellular domains of TMPRSS4 serine protease. The development of selective TMPRSS4 inhibitors will be beneficial in the treatment of cancer. Nonetheless, more research and experiments are needed to confirm the above prediction.

Acknowledgment:

The authors would like to thank Aragen Lifesciences Private Ltd for allowing them to conduct their research. They also want to thank JNTUH for providing PhD admission and research guidance.

References:

Jung, H., Lee, K. P., Park, S. J., Park, J. H., Jang, Y-s., Choi, S-Y., Jung, J-G., Jo, K., Park, D. Y., Yoon, J. H., Park, J-H., Lim, D-S., Hong, G-R., Choi, C., Park, Y-K., Lee, J. W., Hong, H. J., Kim, S., Park, Y. W. (2008). TMPRSS4 promotes invasion, migration and metastasis of human tumor cells by facilitating an epithelial–mesenchymal transition. *Oncogene*, 27(18), 2635–2647. DOI: <https://doi.org/10.1038/sj.onc.1210914>.

de Aberasturi, A. L., Calvo, A. (2015). TMPRSS4: an emerging potential therapeutic target in cancer. *British Journal of Cancer*, 112(1), 4-8. DOI: <https://doi.org/10.1038/bjc.2014.403>.

Netzel-Arnett, S., Hooper, J. D., Szabo, R., Madison, E. L., Quigley, J. P., Bugge, T. H., Antalis, T. M. (2003). Membrane anchored serine proteases: A rapidly expanding group of cell surface proteolytic enzymes with potential roles in cancer. *Cancer and Metastasis Reviews*, 22(2), 237-258. DOI: <https://doi.org/10.1023/A:1023003616848>.

Ohler, A. & Becker-Pauly, C. (2012). TMPRSS4 is a type II transmembrane serine protease involved in cancer and viral infections. *Biological Chemistry*, 393(9), 907-914. DOI: <https://doi.org/10.1515/hsz-2012-0155>.

Min, H. J., Lee, M. K., Lee, J. W., Kim, S. (2014). TMPRSS4 induces cancer cell invasion through pro-uPA processing. *Biochemical and Biophysical Research Communications*, 446(1), 1-7. DOI: <https://doi.org/10.1016/j.bbrc.2014.01.013>.

Puente, X.S., Sánchez, L.M., Gutiérrez-Fernández, A., Velasco, G., López-Otín, C. (2005). A genomic view of the complexity of mammalian proteolytic systems. *Biochem Soc Trans*, 33(2), 331–334. DOI: <https://doi.org/10.1042/BST0330331>.

Wallrapp, C., Hähnel, S., Müller-Pillasch, F., Burghardt, B., Iwamura, T., Ruthenbürger, M., Lerch, M. M., Adler, G., & Gress, T. M. (2000). A novel transmembrane serine protease (TMPRSS3) overexpressed in pancreatic cancer. *Cancer research*, 60(10), 2602–2606. DOI: 15;60(10):2602-6.

Hooper, J. D., Clements, J. A., Quigley, J. P., & Antalis, T. M. (2001). Type II transmembrane serine proteases. Insights into an emerging class of cell surface proteolytic enzymes. *The Journal of biological chemistry*, 276(2), 857–860. DOI: <https://doi.org/10.1074/jbc.R000020200>.

Antalis, T. M., Buzza, M. S., Hodge, K. M., Hooper, J. D., & Netzel-Arnett, S. (2010). The cutting edge: membrane-anchored serine protease activities in the pericellular microenvironment. *The Biochemical journal*, 428(3), 325–346. DOI: <https://doi.org/10.1042/BJ20100046>.

Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2021. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2021

Pagadala, N.S., Syed, K. & Tuszynski, J. (2017). Software for molecular docking: a review. *Biophys Rev*, 9, 91–102. DOI: <https://doi.org/10.1007/s12551-016-0247-1>.

Morris, G.M., Lim-Wilby, M. (2008). Molecular Docking. In: Kukol, A. (eds) *Molecular Modeling of Proteins. Methods Molecular Biology™, Humana Press*, 443. DOI: https://doi.org/10.1007/978-1-59745-177-2_19

Sussman, J. L., Lin, D., Jiang, J., Manning, N. O., Prilusky, J., Ritter, O. & Abola, E. E. (1998). *Acta Cryst*, (D54), 1078-1084.

The UniProt Consortium. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*, 47 (D1), D506–D515. DOI: <https://doi.org/10.1093/nar/gky1049>.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13), 1605-1612.

Studio, D. (2008). Discovery studio. Accelrys [2.1].

Daina, A., Michielin, O. & Zoete, V. (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep*, 7, 42717. DOI: <https://doi.org/10.1038/srep42717>.

Trott, O., Olson, J., A. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. DOI: <https://doi.org/10.1002/jcc.21334>

Spoel, D. V. D., Lindahl, E., Hess, B., Groenhof, G., Mark, E. A., Berendsen, C. J. H. (2005). *Journal of Computational Chemistry* Volume, 26(16), 1701-1718. DOI: <https://doi.org/10.1002/jcc.20291>.

Muhammed, M. T., & Aki-Yalcin, E. (2019). Homology modeling in drug discovery: Overview, current applications, and future perspectives. *Chemical biology & drug design*, 93(1), 12–20. <https://doi.org/10.1111/cbdd.13388>

Yang, J. M., & Chen, C. C. (2004). GEMDOCK: a generic evolutionary method for molecular docking. *Proteins*, 55(2), 288–304. <https://doi.org/10.1002/prot.20035>

Min, H. J., Lee, M. K., Lee, J. W., & Kim, S. (2014). TMPRSS4 induces cancer cell invasion through pro-uPA processing. *Biochemical and biophysical research communications*, 446(1), 1–7. <https://doi.org/10.1016/j.bbrc.2014.01.013>

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L., & Schwede, T. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic acids research*, 42(Web Server issue), W252–W258. <https://doi.org/10.1093/nar/gku340>

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, 25(17), 3389–3402. <https://doi.org/10.1093/nar/25.17.3389>

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC bioinformatics*, 10, 421. <https://doi.org/10.1186/1471-2105-10-421>

Remmert, M., Biegert, A., Hauser, A., & Söding, J. (2011). HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature methods*, 9(2), 173–175. <https://doi.org/10.1038/nmeth.1818>

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. (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46(W1), W296–W303. DOI: <https://doi.org/10.1093/nar/gky427>

Hsu, K-C., Chen, Y-F., Lin, S-R., Yang, J-M. (2011). iGEMDOCK: a graphical environment of enhancing GEMDOCK using pharmacological interactions and post-screening analysis. *Journal of BMC Bioinformatics*, 12(1), S33. DOI: <https://doi.org/10.1186/1471-2105-12-S1-S33>.

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.* 46(W1), W296-W303 (2018).

Bienert, S., Waterhouse, A., de Beer, T.A.P., Tauriello, G., Studer, G., Bordoli, L., Schwede, T. The SWISS-MODEL Repository - new features and functionality. *Nucleic Acids Res.* 45, D313-D319 (2017).

Schwede, T., Kopp, J., Guex, N., & Peitsch, M. C. (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic acids research*, 31(13), 3381–3385. <https://doi.org/10.1093/nar/gkg520>

Wei Tian, Chang Chen, Xue Lei, Jiuling Zhao, Jie Liang, CASTp 3.0: computed atlas of surface topography of proteins, *Nucleic Acids Research*, Volume 46, Issue W1, 2 July 2018, Pages W363–W367, <https://doi.org/10.1093/nar/gky473>

Binkowski, T. A., Naghibzadeh, S., & Liang, J. (2003). CASTp: Computed Atlas of Surface Topography of proteins. *Nucleic acids research*, 31(13), 3352–3355. <https://doi.org/10.1093/nar/gkg512>, 2003)

Hsu, K. C., Chen, Y. F., Lin, S. R., & Yang, J. M. (2011). iGEMDOCK: a graphical environment of enhancing GEMDOCK using pharmacological interactions and post-screening analysis. *BMC bioinformatics*, 12 Suppl 1(Suppl 1), S33. <https://doi.org/10.1186/1471-2105-12-S1-S33>

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*, 389, 2016,253-262,

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*, 23, 6, 2013, 1748-1751,