Journal of Advanced Zoology

ISSN: 0253-7214

*Volume **44** Issue **03 Year 2023** Page **232:246***

Comprehensive Investigation of Recombinant Human TMPRSS4 Expression and Purification Across Diverse Expression Platforms: Bacterial, Insect (BVES), and Mammalian Systems

Poornima Esram [1], Premkumar Arumugam[2]

[1]*Aragen Lifesciences Private Ltd., Survey No. 125 & 126, IDA, Mallapur, Hyderabad-500076, Telangana, India. Email: Jessiesram@yahoo.com. ORCID: 0000-0003-0621-7174*

[2]*Institute of Science and Technology, Jawaharlal Nehru Technological University, Hyderabad-500085, Telangana, India. Email: premkumar1970@gmail.com. ORCID: 0009-0006-5490-3557*

**Corresponding author's E-mail: Poornima.esram@aragen.com*

Article History

Received: 06 June 2023
Revised: 25 August 2023
Accepted:11 Sept 2023

Abstract

TMPRSS4, an essential transmembrane protease serine 4, holds significant relevance in diverse biological contexts, making it a molecule of interest across various fields. This transmembrane serine protease, TMPRSS4, plays a pivotal role in multiple areas, including cancer research, disease processes, and potentially in the spread of viral infections. In light of its multifaceted significance, the research and production of active TMPRSS4 protein serve as a critical foundation for in-depth structural and functional studies, ultimately leading to the identification of potential inhibitors. Within this context, our study encompasses an extensive investigation of various expression systems, including microbial, insect cells, and mammalian cells, aimed at obtaining biologically active TMPRSS4 protein. This study underscores our efforts in optimizing solubility. We explored different E. coli expression strains, optimized expression conditions at various temperatures, employed both inducible and autoinduction-based approaches, fine-tuned the Multiplicity of Infection (MOI), and controlled cell density in Sf9 cells using BVES. Additionally, we delved into the exploration of alternative signal peptides for secretory expression in mammalian cells, followed by the characterization of purified protein variants. Through rigorous biochemical analyses and functional assays, we discovered the distinct advantages associated with mammalian cell expression, ultimately resulting in the production of biologically active TMPRSS4. Our findings provide critical insights into the optimization of expression systems, thereby enhancing the functionality of the protein.

Keywords: *Expi CHO S, Freestyle HEK F, Secretory expression, Alternative signal peptides, Functional characterization.*

1. Introduction

TMPRSS4, or Transmembrane Protease, Serine 4, is a member of the transmembrane serine protease family (TTSP) a class of cell surface proteolytic enzymes, which commonly have a proteolytic domain, transmembrane domain, a shot cytoplasmic domain and variable length stem region [Bugg et. Al., **1**] (Figure1). This family of proteases plays a crucial role in various physiological processes by cleaving and activating a wide range of substrates. TMPRSS4 is particularly interesting due to its diverse functions and its involvement in several pathological conditions [Wallrapp C et. Al., 2]. As a transmembrane serine protease, TMPRSS4 is embedded in cell membranes, where it exerts its effects on various cellular signalling pathways, protein activation, and cell-cell interactions. In recent years, research has revealed that TMPRSS4 is implicated in different cellular processes, including cell proliferation, invasion, and migration [Joyce et al., 3, Jung et al., 4]. Additionally, its role in proteolytic activation of viral glycoproteins highlights its significance in viral entry and infection. The expression of TMPRSS4 has been detected in multiple tissues and cell types, indicating its potential involvement in both normal physiological functions and disease states.

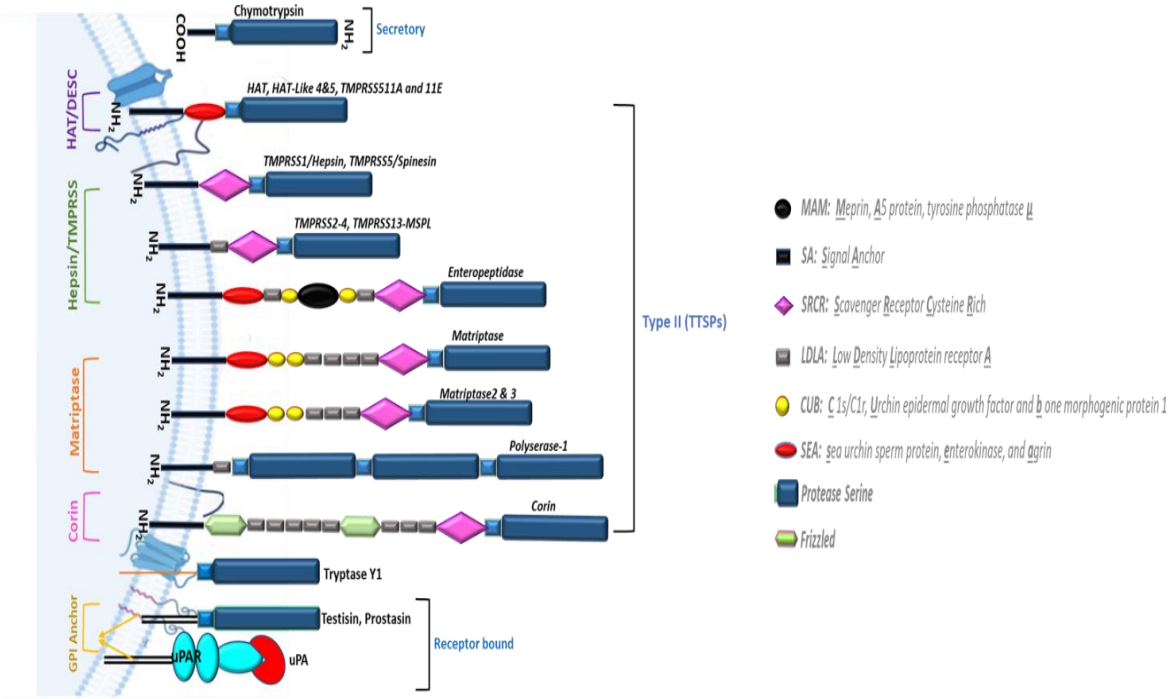

Figure 1: The figure provides a schematic representation of the structural organization of TTSP proteins and highlights the subfamilies within the TTSP family: HAT/DESC, hepsin/TMPRSS, matriptase, Corin and GPI anchor types. Typically, the TTSPs carry a signal anchor domain, followed by a serine protease domain which is involved in proteolytic activity, where the serine nucleophile attacks and cleaves Sissle bond of a specific amino acid.

Source: Made by Author, adapting from Bugg TH., Toni M. Antalis et al.

The protease's involvement in cancer, viral infections, and other diseases has raised interest in exploring its potential as a therapeutic target or biomarker. Recent research findings have shed light on the multifaceted roles of TMPRSS4 in tumorigenesis, emphasising its significance as a pivotal mediator in crucial processes such as invasion, metastasis, migration, adhesion, and the epithelial-to-mesenchymal transition (EMT) within human epithelial cancer cells. This recognition of TMPRSS4's influence positions it as a promising and novel target for cancer therapies [Min HJ et. Al., 5, Li T et. Al., 6].

TMPRSS4, comprised of 437 amino acids, presents a predictive molecular weight of 48 kDa. Notably, its structure unfolds into distinct domains, each contributing to its multifunctional nature. At its C-terminus (amino acids 234-437), TMPRSS4 hosts a serine protease domain. This protease domain is followed by a scavenger receptor cysteine-rich domain (SRDR) and a low-density lipoprotein receptor class A domain, as illustrated in Figure 2. Of particular significance is the catalytic triad within TMPRSS4—comprising His, Asp, and Ser—which underpins its enzymatic activity. Remarkably, this triad's conservation across a spectrum of type II transmembrane serine proteases (TTSPs) attests to its vital role in their collective functionality [Hooper et. Al.,7]. Moreover, TMPRSS4's enzymatic ability is further fine-tuned by a substrate-binding pocket, dictating the enzyme's specificity for substrates. This architectural intricacy forms the basis for TMPRSS4's dynamic engagement in various biological processes, ranging from cancer progression to viral entry.

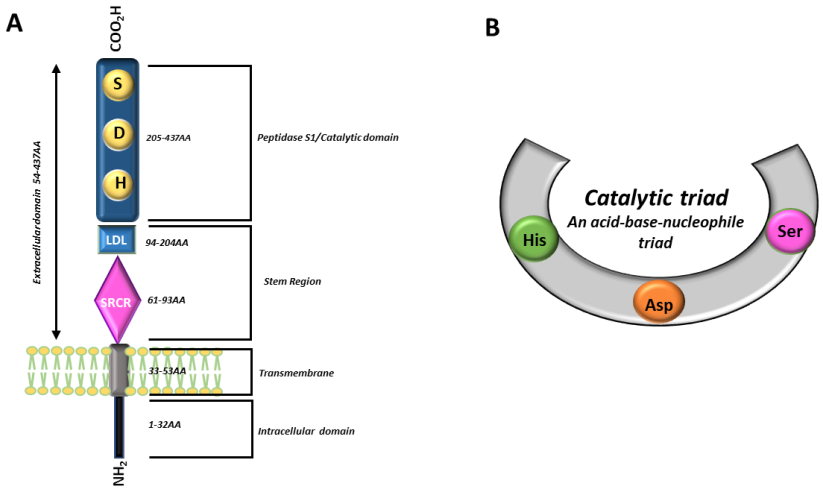

Figure 2: [A] Schematic diagram of the TMPRSS4 structure: TMPRSS4 is a single-pass type II membrane protein compassing a serine protease domain/catalytic domain (Peptidase S1), followed by a scavenger receptor cysteine-rich domain (SRCR) and a low-density lipoprotein domain. [B] Depiction of Catalytic triad or charge relay system of serine proteases, typically consisting of Asp, His and Her

Source: Made by Author adapting the concept from A L de Aberasturi et al

Recognizing its significance in cancer research, disease processes, and potentially in the spread of COVID-19, we undertook the task of generating active TMPRSS4 protein. We have aimed to express the truncated TMPRSS4 CD and ECD, skipping the cytoplasmic and transmembrane regions (1-53AA). Our initial efforts led us to explore a range of expression systems, including E. coli and insect cells utilizing the Baculovirus expression system. While these systems offered promising results, we encountered the challenge of protein aggregation, an obstacle known to compromise functionality and downstream analyses. Acknowledging the limitations of these platforms, we strategically transitioned to mammalian cell expression, with a particular focus on HEK and CHO cells. Renowned for their capacity to execute post-translational modifications and secrete complex proteins, HEK/CHO cells represented a promising avenue to generate soluble, functional TMPRSS4 protein. We have attempted to express ECD and CD domains of TMPRSS4 by fusing non-native signal sequences at their N terminus. As elucidated by Kober et. al. [Knappskog S. et. Al., 14, Zhang L. et. Al., 15, Kober L. et. Al., 16] it has been established that fusion of potent alternative signal peptides with the target of interest can aid or improve the secretion efficiency.

In this study, we present a comprehensive comparative analysis of three predominant expression systems, each harnessed in the pursuit of biologically active TMPRSS4 protein. Our objective was to acquire protein that would facilitate rigorous functional examinations also offer insights into the protein's structural attributes in future.

2. Materials And Methods

Expression and purification in E. *coli* cells

Expression construct generation

The E. coli strains employed in this study were BL21(DE3) (Novagen, Cat. No. 69450), Nico21(DE) (NEB, Cat. No. C252911), Tuner (DE3) (Novagen, Cat. No. 70623), Rosetta.2(DE3) (Novagen, Cat. No. 79397), and E. coli DH5α (Thermo Fisher Scientific, Cat. No.18265017) was utilized for the screening of ligated constructs. Cultures of these strains were prepared according to standard protocols. Bacterial strains were grown in Luria-Bertani (Sigma, LB Cat. No. L3522) medium supplemented with appropriate antibiotics as recommended by the respective suppliers.

The genes encoding the catalytic domain (205-437 amino acids) and extracellular domain (54 to 437 amino acids) of TMPRSS4 were codon optimized for expression in E. coli and synthesized (GeneArt). The synthetic genes were designed to contain an EcoRI restriction site at the 5' terminus and a XhoI

restriction site at the 3' terminus for subsequent cloning. Expression constructs were engineered using the pET-28a (+) vector backbone (Novagen, Cat. No. 69864) through a conventional restriction-based approach. The resulting constructs were designed to enable the expression of proteins with an N-terminal 6His tag, along with a Thrombin/enterokinase protease cleavage site for subsequent purification and downstream applications. For transformation and selection of transformants, Kanamycin (Sigma, Cat. No. K1377) and chloramphenicol (Sigma, Cat. No.C0378) were used depending on the strain characteristics to ensure proper maintenance of the expression constructs. The generation of these expression constructs was carried out in accordance with established molecular biology techniques, and the resulting plasmids were verified by Sangers sequencing to confirm the correct insertion of the target sequences.

The transformed expression strains were made by introducing the sequence verified expression constructs, following standard transformation protocols for E. coli. A single colony of each transformant was meticulously picked and inoculated into Luria-Bertani (LB) broth supplemented with the appropriate antibiotics as prescribed for the specific strain.

Inducible protein expression

Overnight cultures were initially developed at 37^0C, 220RPM, subsequently a subculture was initiated by inoculating ON culture into fresh LB broth containing the respective antibiotics. Expression was induced upon reaching an OD_{600} 0.4-0.5. Induction was achieved by adding isopropyl β-D-1-thiogalactopyranoside (IPTG, Sigma, Cat. No. I6758) at a concentration of 0.5mM. Protein expression was conducted in 25ml culture volume, at different temperatures: 16°C overnight (ON), 25°C overnight, and 37°C for 3 hours, all cultures were maintained at 150 during induction. After the designated incubation period, cultures were harvested by centrifugation at 5000 RPM for 20 minutes. The resulting cell pellets were collected and analysed for protein expression levels, potentially providing insights into expression strains, temperature, and time-dependent effects on protein production.

Protein expression analysis was performed using sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). Harvested cell pellets were suspended in phosphate-buffered saline (PBS; Sigma). Cell lysis was achieved by subjecting the cell suspension to brief sonication using a probe sonicator (Vibra cell VCX 130, Sonics). Sonication was carried out in pulses of 30 seconds ON followed by 30 seconds OFF, at 30% amplitude for 5-10cycles. Retaining an aliquot of whole cell lysate (WCL), clarified supernatant was collected by centrifuging the WCL at 12,000 RPM for 30 minutes for solubility confirmation.

For SDS-PAGE, 20 µL of each prepared fraction (WCL, soluble fraction, and insoluble fraction) were resolved using a precast polyacrylamide gel (Bio-Rad, 4-20% Mini protean TGX, Cat. No.4561094) under reducing conditions. The proteins were visualized by staining with Coomassie Brilliant Blue R-250 (CBBR250). The stained protein gels were imaged using a Chemidoc, Bio-Rad gel documentation system to capture and document the protein expression profiles.

Protein expression optimization using Autoinduction media (AI)

To optimize the expression of TMPRSS4 ECD and CD in autoinduction conditions transformants generated in, Nico21(DE) (NEB, Cat. No. C252911), and Rosetta.2(DE3) (Novagen, Cat. No. 79397), strains were employed. Single colonies from both strains were cultured in LB broth overnight at 37°C and then inoculated into AI media (0.6% of Na_2HPO_4, 0.3% of KH_2PO_4, 2%of Tryptone, 0.5% of YE, 5% of NaCl, 1% glycerol, 0.5% Lactose, 0.05% Glucose) at a 1:50 ratio (v/v), equivalent to a 2% inoculum. Cultures were allowed to grow until reaching an optical density at 600nm (OD_{600}) of 0.4-0.5 at 37^0C, then changed the temperature to 25°C, and 16°C while shaking at 150 RPM for ON. Cell pellets were obtained by centrifugation, suspended in phosphate-buffered saline (PBS), and subjected to cell lysis using a brief sonication step with a probe sonicator (Vibra cell VCX 130, Sonics). The whole cell lysate was then centrifuged at 12,000 RPM for 30 minutes to separate the soluble and insoluble fractions. Protein samples from both fractions were resolved by SDS-PAGE, and the proteins were visualized using Coomassie brilliant blue R250 staining.

Purification

A 2 Liter (L) batch of Rosetta.2 (DE3) E. coli cells was cultivated under optimized conditions with AI media. Following cell growth and protein expression, cell pellets were obtained and subsequently lysed using a PANDA plus homogenizer. The resulting lysate was clarified to remove cellular debris, and the cleared lysate was loaded onto a NiNTA column (GE Healthcare) previously equilibrated with Buffer A (50mM Sodium phosphate, pH 8.0, 150mM NaCl, 20mM Imidazole). After loading, a wash step was carried out with Buffer A to eliminate impurities, and the POI was eluted from the column using a linear gradient of imidazole concentration (20mM-500mM), with fractions collected throughout. Fractions containing the pure POI were identified through analysis, pooled, and subjected to buffer exchange into 50mM Tris, 100mM NaCl at pH 8.0 for storage in aliquots.

Expression in Insect cells, Sf9 (BVES)

Sf9 (Spodoptera frugiperda) cells adapted to suspension and serum free media (Thermo Fisher Scientific, Cat. No. 11496015) were utilized in these studies. Sf-900 II SFM (Thermo Fisher Scientific, Cat. No. 10902088) was selected as the primary culture medium and Sf9 cells were routinely propagated in suspension conditions using standard culturing techniques.

Expression construct generation

In this study, we engineered a baculovirus expression construct to produce the extracellular domain (ECD) and catalytic domain (CD) of TMPRSS4. The synthetic gene, optimized for expression in insect cells, was designed to incorporate NcoI and XhoI restriction sites at the 5' and 3' ends, respectively. The resulting expression construct enables the production of N-terminal 6xHis-tagged TMPRSS4 ECD (54-437 AA) and TMPRSS4 CD (207-437 AA) proteins, with a TEV protease cleavage site positioned between the target protein and the 6xHis tag, facilitating potential tag removal subsequently if needed.

Recombinant Bacmid, and recombinant virus generation

The sequence verified TMPRSS4 pFastBac HTA constructs were introduced into DH10Bac cells, a competent strain suitable for Bacmid generation. To identify colonies containing the recombinant Bacmid, we employed blue/white selection, which allowed us to distinguish between successful recombinants (white colonies) and non-recombinants. White colonies were selected and grown further to isolate the recombinant Bacmid for subsequent confirmation. Transposition of the gene of interest (GOI) onto the recombinant Bacmid was confirmed through PCR analysis using M3 forward and reverse primers. Confirmed recombinant Bacmids were used to generate the recombinant virus following standard transfection protocols. The harvested recombinant virus, referred to as P0, was further used to amplify thus generating first second generation of recombinant virus referred as P1 successive experiments. Viral particle titration was conducted using a plaque assay, providing quantification of virus particles.

Expression screening and optimization

The mortality of infection (MOI) was determined based on the results of the plaque assay, allowing for the optimization of expression conditions. We have used MOI of 1, 5 and 10 at the cell density $2X10^6$cells/ml. 1ml of cells were collected from 24-96 hours of post transfection with 24hours of interval to monitor the cell viability, and protein expression profile. The monoclonal anti-poly histidine antibody (Sigma, clone HIS-, Cat. No. SAB2105729) was employed to detect the presence of the N-terminal 6xHis-tagged TMPRSS4 proteins.

Expression in Mammalian cells

Culturing and maintenance of cells.

Expi CHO S cells (Life Technologies, A29127) and FreeStyle 293 F cells (Life Technologies, R79007) were cultured in a controlled suspension environment to support cell growth and protein expression. Expi CHO S cultures were maintained in an Eppendorf S41i CO2 Shaker incubator set at 120 RPM, 8% CO_2, and a temperature of 37°C, in Expi CHO expression media (Life Technologies, A29100) to provide the necessary nutrients and factors for cell growth and protein expression. FreeStyle 293 F cultures were maintained in an Eppendorf S41i CO_2 Shaker incubator, agitating at

120 RPM, temperature 37^0C, while maintaining the CO_2 concentration of 5%. Corning polycarbonate Erlenmeyer flasks equipped with vent caps were used as culture vessels.

Expression construct generation.

To express the TMPRSS4 catalytic domain (CD) and extracellular domain in mammalian cells, we used pcDNA3.4 expression vector (Thermo, A14308). The gene sequences for both the TMPRSS4 CD and ECD were codon-optimized to enhance expression in mammalian cells. Secretory signal peptides were ligated upstream of the TMPRSS4 CD and ECD gene sequences, both of which harboured a C-terminus 6xHis tag. Standard restriction enzyme (RE) cloning techniques were employed to insert the codon optimized TMPRSS4 CD and ECD gene sequences, along with the selected signal peptides, into the pcDNA3.4 vector. This step ensured that the gene constructs were in the appropriate expression context. Following gene insertion, the constructs were verified by restriction digestion, confirming the presence and proper orientation of the gene inserts within the pcDNA3.4 vector. Positive constructs, as confirmed by restriction digestion, followed by sequence verification, were further amplified to generate sufficient DNA for transient transfection experiments. Purification of the constructs was carried out using the PurLink Hi Pure plasmid midi prep kit (Life Technologies, K2100-14). This purification process ensured the isolation of high-quality plasmid DNA suitable for transfection studies.

Transfection and expression optimization.

In our study, we employed a well-defined protocol for transfecting Expi CHO cells and HEK F cells to express the TMPRSS4 domains. Expi CHO cells were transfected with Plasmid DNA constructs at a cell density of 6×10^6 cells/ml using Expifectamine CHO transfection reagent from the kit (Life Technologies, A29129). Standard protein expression procedures were followed, including the addition of Feed and Enhancers (Life Technologies, A29129) a day after transfection. Cultures were incubated at 120 RPM, 8% CO_2, and 37°C temperature, and expression was monitored for nine days. Cell viability was assessed by trypan blue staining, and media supernatant samples were collected for subsequent analysis. In the case of HEK F cells, transfections were performed at a cell density of 2×10^6 cells/ml using PEI max transfection grade linear polyethyleneimine Hydrochloride MW 40,000 (Poly biosciences 24765), and cultures were incubated at 120 RPM, 5% CO_2, and 37°C temperature for seven days of post-transfection. Cell viability was determined by microscopic observation. Media supernatant samples were collected, and protein expression was analysed using Western blotting on the Protein Simple WES platform. For large-scale protein production, 500 ml-sized cultures were harvested when cell viability dropped below 70-80%, typically occurring at 9 days for Expi CHO cells and 5-7 days for HEK F cells.

Purification.

Transiently transfected the cells with the sequence verified expression construct(s), and conditioned media were collected by centrifugation at 5000 RPM for 20 minutes. The media supernatant was then filtered using sterile 0.22µm PES rapid flow bottle filters. Clarified media supernatant was subjected to NiNTA column purification utilizing Pierce High-capacity Ni IMAC resin with EDTA compatible beads (Thermo, A50585). Beads were preconditioned with equilibration buffer (20mM Tris pH 8.0 and 150mM NaCl), and the media supernatant was allowed to bind to the resin overnight at 4°C. The next day, the beads were gently centrifuged at 3000 RPM for 30 minutes to collect the unbound fraction, while the beads retained approximately 50ml of the sample. The beads were then packed into an XK16/20 column (Cytiva) and the remaining 50ml was collected as flow-through. Beads were washed with 5 column volumes of Equilibration buffer (20mM Tris, pH 8.0, 150mM NaCl), and the wash fractions were collected in 5ml portions. Proteins were eluted in a linear gradient of 0-500mM Imidazole. Fractions were analyzed by SDS-PAGE, and proteins were visualized using Coomassie brilliant blue R250 staining. Fractions containing TMPRSS4 ECD (~50kDa) and CD (~26kDa) were pooled and concentrated while exchanging the buffer to 50mM Tris pH 8.0, 100mM NaCl. Glycerol was added to 20%, and protein aliquots were stored at -80°C for subsequent experiments.

Peptidase assay For Functional assessment.

The functional activity TMPRSS4 domains were evaluated using a fluorescence-based assay employing the fluorogenic peptide Boc.Gln-Ala-Arg-7amido4-methyl coumarin hydrochloride (Sigma, B4153) as a substrate, using 100mM Tris pH 8.0 with 10mM CaCl2 as assay buffer. Fluorescence was continuously monitored at 30°C using a Tcan Spark multimode reader with excitation and emission wavelengths set at λEx_{380} and λEm_{460}nm, respectively, and the rate of substrate hydrolysis was analyzed using the Spark Control Magellan software. Negative assay controls, comprising wells devoid of enzyme and wells containing substrate but lacking the enzyme in the assay buffer, were included to establish baseline measurements.

Statistical Analysis:

Statistical analysis was performed to ensure the accuracy of the enzymatic activity measurements. To achieve this, the background signal from the substrate was subtracted from the enzyme activity signal. The resulting data were then analyzed and plotted using GraphPad Prism software.

SignalP 4.0 Analysis:

For comprehensive proteomic analysis of cellular compartments and the determination of subcellular locations of specific proteins or protein complexes, SignalP 4.0 software was employed [Peterson et. Al., 17]. This bioinformatics tool played a pivotal role in elucidating the subcellular localization of the studied proteins, contributing valuable insights into their functional roles within the cellular context.

3. Results and Discussion

Expression and purification in E. coli

We have used BL21(DE3), Nico21(DE3), Tuner (DE3) and Roestta.2(DE3) expression hosts to express both the domains of TMPRSS4. Observed the expression of both the forms of TMPRSS4 in all the strains, yet the expression was insoluble, creating inclusion bodies (Figure3). Optimization parameters used to express these proteins as soluble fractions were unsuccessful while following IPTG based induction approach. We have selected Rosetta.2(DE3) and Nico21(DE3), strains for further optimizations using auto induction (AI) media. Both the strains observed to yield soluble expression of ECD at all temperatures tried, whereas expression levels in Rosetta.2(DE3) high in comparison with Nico21(DE3) at 25^0C and 16^0C. Nevertheless, these conditions did not work well with CD version. We have observed over 50% of soluble TMPRSS4 ECD expression with Rosetta.2(DE3) expression strain in AI media, at 16^0C temperature for overnight as confirmed based on the SDS PAGE data generated (Figure4). The clarified lysate after the protein expression with the optimized conditions is used to purify the protein by affinity NiNTA column chromatography on AKTA purification platform for the activity assessment yielded 5-10mg of the proteins with a purity of ~85%. Pooled the fractions having pure form of the proteins and concentrated to ~1mg/ml in 50mM Tris, pH8.0 having 100mM NaCl (Figure5).

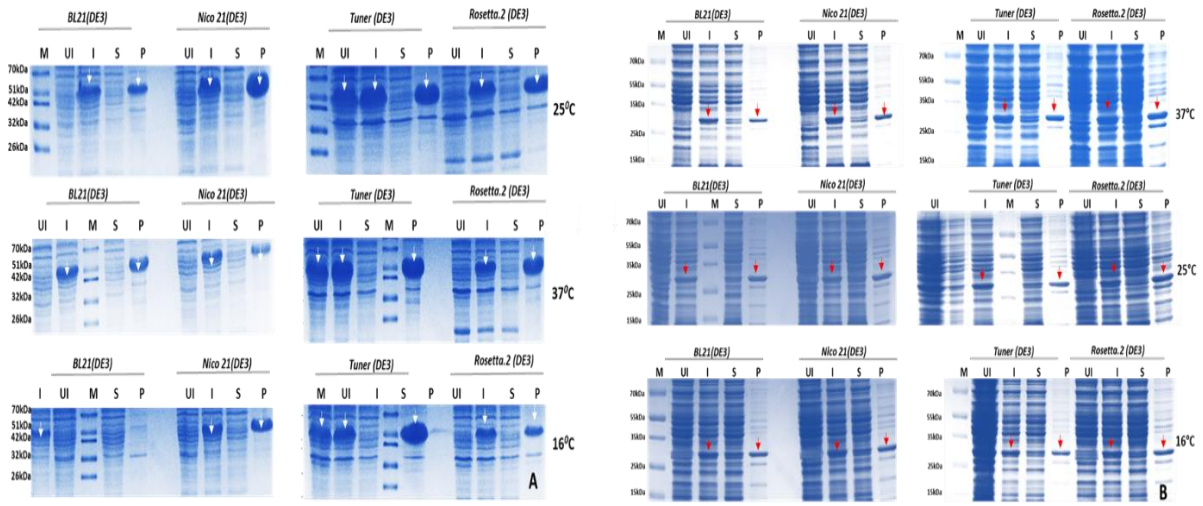

Figure 3: This figure provides visual evidence of the expression profiles of TMPRSS4 ECD and CD in four different expression systems. [A]The presence of a 46 kDa band in both the whole cell lysates and insoluble fractions for TMPRSS4 ECD suggests that the protein was expressed but remained insoluble under the tested conditions. [B] For TMPRSS4 CD, no soluble protein expression was detected in any of the expression systems, highlighting the need for further optimization strategies to achieve soluble expression.

Source: Author

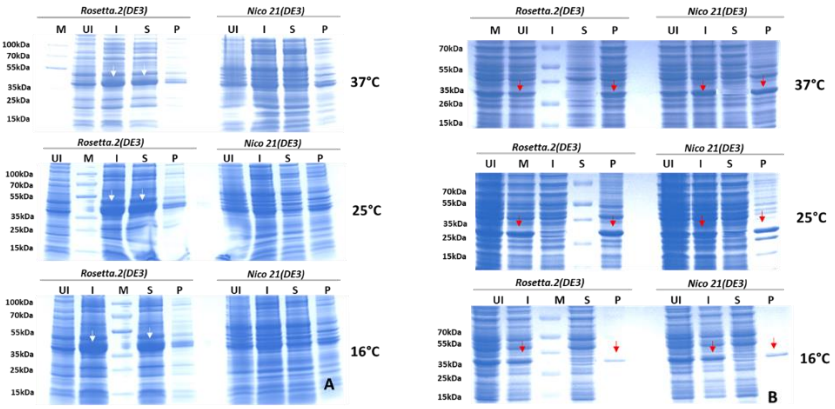

Figure 4: Soluble Expression Analysis of TMPRSS4 ECD and CD in Rosetta 2(DE3) and NiCo21(DE3) Under Auto-Inducible Conditions at Different Temperatures; This figure illustrates the soluble expression profiles of TMPRSS4 ECD and CD in Rosetta 2(DE3) and NiCo21(DE3) strains under different temperatures and auto-inducible conditions. While TMPRSS4 ECD exhibited approximately 50% soluble expression, with Rosetta 2(DE3) showing higher expression levels, TMPRSS4 CD remained insoluble under the experimental conditions.

Source: Author

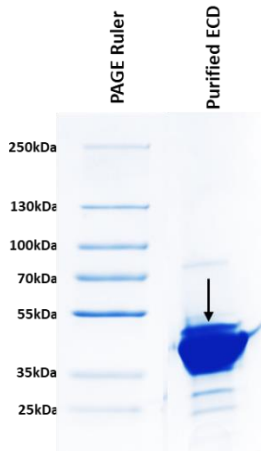

Figure 5: This figure showcases the successful expression and purification of His-tagged TMPRSS4 ECD in Rosetta 2(DE3) cells under optimized auto-inducible culture conditions. The use of the NiNTA column allowed for efficient purification, yielding highly pure TMPRSS4 ECD for downstream applications.

Source: Author

Expression and purification in Sf9 cells using Baculo Virus Expression System:

Expression construct to express TMPRSS4 ECD and CD in Sf9 cells were generated using pFastBac-HTA vector backbone. The pFastBac HTA TMPRSS4 ECD and CD expression constructs were confirmed by restriction digestion as well by Sangers sequencing confirming the GOI in frame. The sequence verified constructs were used to generate recombinant bacmids in DH10Bac cells. Positive

clones were selected by blue-white colony screening, and out of 10 clones screened 5-8 clones resulted to be positive indicating successful transposition of GOI as confirmed by PCR using M3 forward and reverse primers. Two of each recombinant bacmids were used to generate the recombinant virus, described as P0 in this article using Cellfectin II as transfecting agent. We have achieved over 60% of transfection efficiency, as confirmed based on the morphological changes from 24hours of transfection. P0 virus was generated in adherent conditions, 72hours of post transfection yielded viral particle formation. P0 virus was harvested between 7-9days of post transfection and stored as 1ml aliquots. P0 virus stocks were used to amplify the virus in large scale for MOI studies as well as protein expression scale up, and thus amplified virus is denoted as P1 virus here after. Recombinant TMPRSS4 ECD/CD baculovirus were titrated by plaque assay and obtained $2.2X10^8$ PFU/ml of the virus. Some of the P1 virus pellets used for expression verification by Western blot analysis using anti His antibody confirmed the expression of ECD and CD with corresponding molecular weights. Expression optimization was done with cell density of 2million cells, MOI of 1, 5 and 10. Checked the expression for post infection hours of 24, 48, 72 and 96 by Western blot analysis. Observed expression of ECD and CD with all the conditions tried, nevertheless all the protein observed in the pellets, confirming aggregation (Figure 6). We have tried different detergents in the cell lysis buffer to destruct the aggregations, but our attempts found to be ineffective.

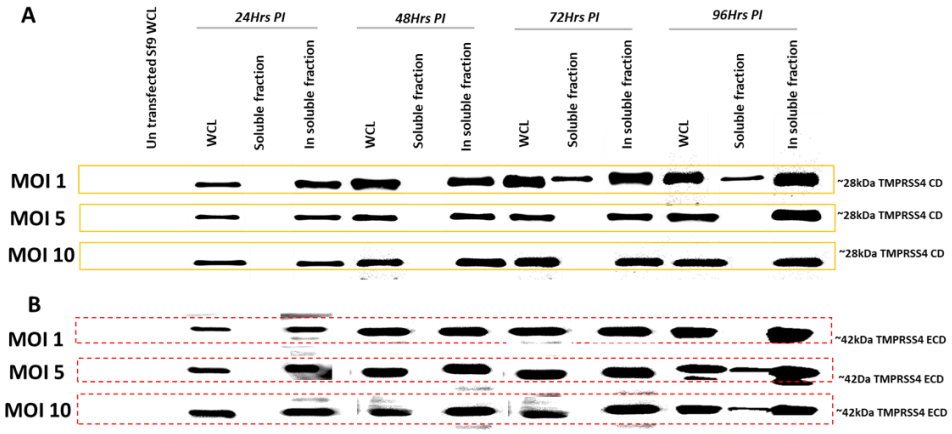

Figure 6: Western Blot Analysis of CD and ECD Expression in Insect (Sf9) Cells Using the Baculovirus Expression System (BVES); Expression of both CD and ECD was achieved through the Baculovirus expression system (BVES). Despite achieving high expression levels, neither the CD nor ECD versions were observed to be expressed in a soluble form under the conditions tested. However, the absence of soluble expression suggests that further optimization of the expression conditions or purification strategies may be required to obtain the desired soluble form of CD and ECD proteins.

Source: Author

Expression and purification in mammalian cells (Expi CHO and HEK-F)

TMPRSS4 (207-437AA) catalytic domain (Uniprot ID: Q9NRS4) expressed constructs were generated using the pcDNA3.4 expression vector to express in the ExpiCHO and HEK F Expression systems. We have achieved clones for all the five alternative signal peptides ligated upstream to the TMPRSS4 CD gene sequences which were having a C terminus 6 His Tag. The confirmed positive constructs using restriction digestion (HindIII, XhoI), and sangers sequencing were amplified further for expression screening experiments.

All the five SP(x)-TMPRSS4.CD-6His_pcDNA3.4 expression constructs were purified by midi prep for transfection. Our attempts of transfection ExpiCHO cells, grown in ExpiCHO culture media at 37 °C with 8% CO_2 with the TMPRSS4 CD expression constructs using Expi CHO Fectamine in suspension conditions were successful. Freestyle HEK F cells were grown in Freestyle culture media, at 37^0C, 5% CO_2 were transfected with the TMPRSS4 CD expression constructs using PEI 40Kda in suspensions conditions were also effective. The expression was monitored for nine days in Expi CHO Cells, and for seven days in HEK F cells by collecting 0.5ml culture volume samples. We have monitored the changes in cell morphology and viability through microscopic examination and protein

expression levels in clarified media by western blot. We have observed a drop in cell viability from 3-4th day of transfection in CHO S cells and reached to ~80% viability on 9th day of transfection. Whereas HEK F transfected cells reached to 70-80% of viability by 7th day of transfection. We have obtained expression of CD in HEK F as well in Expi CHO S from 24hours of transfection until we harvested in timely increasing manner as judged and confirmed by western blot analysis carried out using anti TMPRSS4 antibody (Sigma, catalogue#SAB2501048) (Figure 7). In our studies, we observed noteworthy expression levels when employing alternative signal peptides 1, 2, and 3 in both expression systems. SP1 is trypsin secretory signal peptide (SP1), shares a common serine protease classification with TMPRSS4, given the significant secretion levels and its relevance to TMPRSS4's protein class, we have chosen to employ SP1 in our subsequent experiments for ECD expression in both CHO and HEK cells.

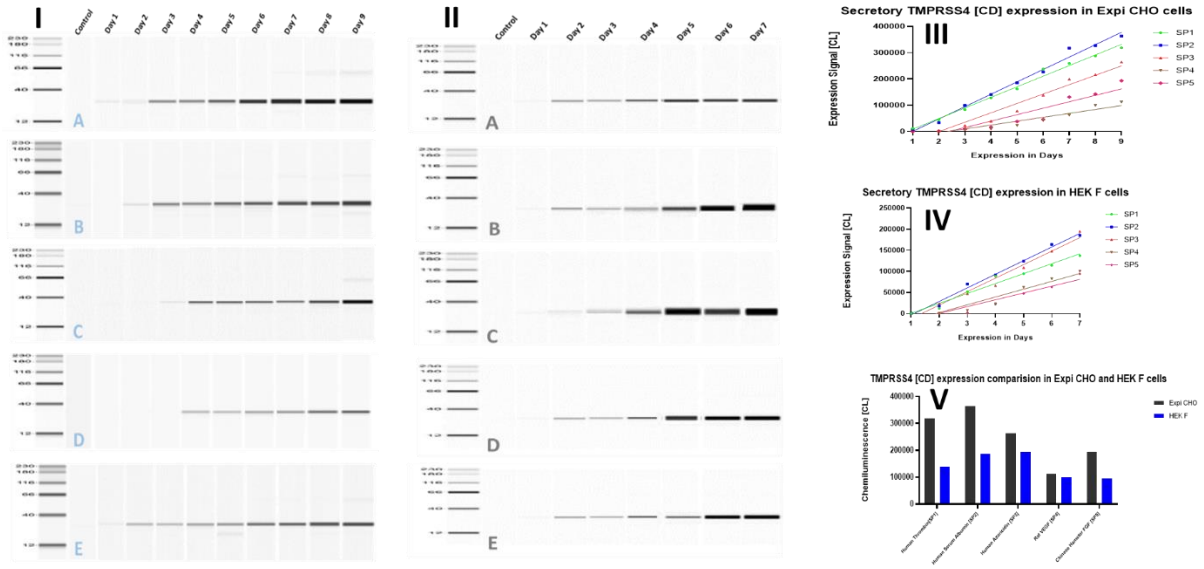

Figure 6: Confirmation of Soluble Secretory Expression of Catalytic Domain in Expi CHO S and Freestyle HEK F Cells; Western blot analysis was conducted to confirm the expression of the catalytic domain in Expi CHO S (I) and Freestyle HEK F (II) cells as soluble secretory proteins. Subfigures [I (A-E)] and [II (A-E)] represent the expression order of the alternative signal peptides fused with the GOI, and III and IV depict product yields plotted against the corresponding chemiluminescence signals, providing insights into the quantitative differences in expression levels between signal peptides variants. Figure V offers a graphical representation of the expression levels achieved in HEK and CHO cells using various signal peptides, aiding the selection of most suitable signal peptide for subsequent experiments.

Source: Author

We generated a single expression construct for the extracellular domain of TMPRSS4 by fusing it with SP1 at the N-terminus. This construct was successfully transfected into both ExpiCHO and HEK F cells, resulting in high expression levels in a time-dependent manner (Figure 8)

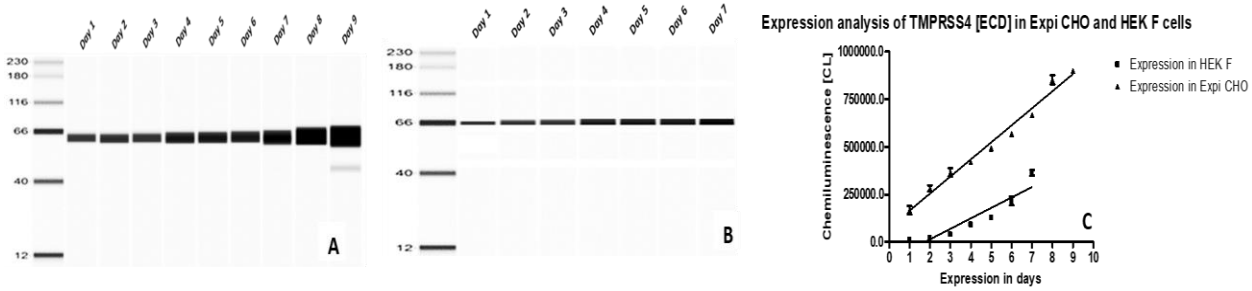

Figure 7: Expression Profile of TMPRSS4 ECD in Expi CHO [A] and HEK F Cells [B] Over Transfection Time Course; The figure displays the expression levels of TMPRSS4 ECD at various time points during the transfection period. The data is presented as product yields, which are plotted against the corresponding chemiluminescence signals. [C] (depicted in the graph) indicates that the expression levels exhibit a linear trend throughout the transfection duration, pattern suggests successful and reliable production of TMPRSS4 production, which is crucial for downstream application.

Source: Author

For large-scale production, we cultured a 500ml batch and harvested the cells. Conditioned media was collected when cell viability reached 70-80%, typically occurring on the 9th day for CHO cells and the 7th day for HEK F cells post-transfection.

We employed Pierce High-Capacity Ni IMAC resin for the purification of TMPRSS4 variants from clarified media. This resin is well-documented for its exceptional chelator stability and its compatibility with EDTA buffers and samples. Notably, ExpiCHO and FreeStyle media compositions include EDTA, making compatibility with EDTA moieties a critical factor during protein purification. Our efforts to bind the Protein of Interest (POI) and subsequent purification using these beads proved successful, yielding favourable recovery rates (Figure 9)

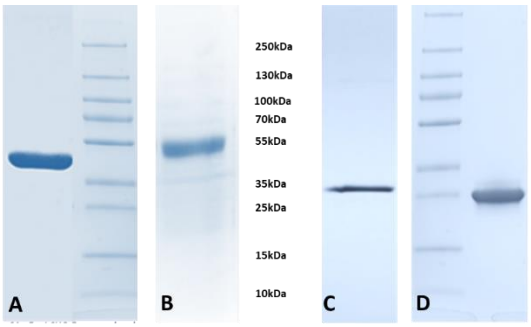

Figure 8: This figure provides a visual representation of the SDS-PAGE analysis, facilitating the comparison of protein purification between the Expi CHO S and HEK F expression systems and shedding light on the quality of the purified ECD and CD proteins.

Source: Author

Functional assay

To assess the proteins expressed and purified from E. coli, Expi CHO, and FreeStyle HEK F cells, we employed a fluorescent enzymatic assay that we developed, as previously described in the literature. In this assay, we utilized the fluorogenic substrate Boc.Gln-Ala-Arg-7-amido-4-methylcoumarin hydrochloride, a substrate known to be cleaved by several serine proteases. The enzymatic assay under optimal conditions, the 7-amino-4-methylcoumarin fluorophore is released and fluoresced at specific excitation and emission wavelengths confirming the proteolysis activity of the preparations.

All preparations exhibited good biological activity. However, it's worth noting that proteins expressed and purified from mammalian cells demonstrated a significantly higher level of activity. When

comparing the biological activity of the domains expressed in FreeStyle HEK F and CHO S cells, we observed superior activity with the proteins expressed in CHO cells (Figure 10)

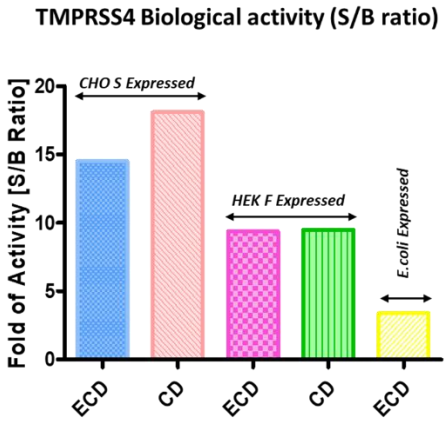

Figure 9: *In this figure, we present the enzymatic activity results obtained using the fluorogenic peptide Boc-Gln-Ala-Asp-AMC with various protein expression systems. The fold increase in enzymatic activity is indicative of the biological activity of the expressed proteins. The observed hierarchy in enzymatic activity suggests that Expi CHO and HEK expression systems are well-suited for achieving high activity in comparison with E. coli expressed protein these studies.*

Source: Author

Attaining fully functional recombinant proteins that support comprehensive biophysical and biochemical studies is of paramount importance in protein research. The successful production of recombinant proteins often entails a multifaceted optimization process, including the exploration of diverse host systems, fine-tuning of expression conditions, strategic gene sequence manipulations, and the incorporation of suitable purification tags.

In our study, we endeavoured to express human TMPRSS4 as truncated variants, focusing on the catalytic domain spanning amino acids 207-437 and the extracellular domain encompassing amino acids 54-437, derived from the full-length 437 amino acid sequence. To ensure a comprehensive assessment, we conducted our expression screening using three well-established expression systems: E. coli, insect (Baculovirus Expression Vector System - BVES), and mammalian cell-based systems. Each of these systems is recognized for its efficacy in heterologous recombinant protein production.

We conducted a thorough screening of expression in four E. coli strains: BL21(DE3), Nico21(DE3), Tuner (DE3), and Rosetta.2(DE3). This screening aimed to evaluate the unique advantages that each of these strains offers when dealing with challenging or difficult-to-express proteins, with a focus on achieving soluble fractions. All the strains observed to yield good expression levels with the inducible as well auto induced approaches yet had challenges with solubility majorly with CD version. Majority of the protein observed to be in insoluble fractions forming inclusions bodies in induced based experiments. While switching to autoinduction approach we have gained substantial soluble expression for ECD. It is well known that auto-inducible systems provide a gradual increase in expression levels, mimicking natural regulation of genes, allowing cells to adapt to the increased protein production slowly. This gradual increase can be less stressful for the cells, reducing the risk of misfolding and aggregation leading to soluble expression. And also, it is reported that auto induction often leads to a balanced expression profile, where the protein of interest is produced alongside other cellular components in a synchronized manner, thus helping to maintain less competitive cellular environment resulting improved protein solubility. This approach did not work well for CD version in our hands. Nevertheless, employing a combination of strategies such as growth conditions optimization, co-expression of chaperons, altering expression levels, using solubility enhancing fusion tags like GST, MBP may improve the soluble expression of TMPRSS4 CD versions as well ECD using inducible and auto inducible approaches.

Our attempts to express TMPRSS4 variants in Sf9 cells were met with significant challenges. Although we achieved high expression levels under the tried conditions, a mere 5% of the protein was observed in the soluble fraction. Furthermore, we encountered substantial protein aggregation in our preparations. In an effort to enhance solubility without compromising the tertiary structure of our target proteins, we explored the incorporation of various non-ionic detergents in extraction buffers. Regrettably, none of these conditions proved effective in achieving the desired solubility.

Given that TMPRSS4 is a protease, it is plausible to hypothesize that its overexpression may have led to the degradation of cellular proteins due to its catalytic activity. Additionally, the less favourable or resource-limited cellular environment may have contributed to the formation of aggregates.

To address these challenges, we recognize the need for further research in a controlled environment. Exploring alternative expression modes, such as secretion using an appropriate signal peptide to release the protein into the extracellular matrix, such as the culture media, holds promise for enhancing the solubility of TMPRSS4 in Sf9 cells while minimizing aggregation. These strategies warrant further investigation to optimize the expression of TMPRSS4 variants in this system

We selected mammalian cells for expressing TMPRSS4 proteins due to their ability to closely mimic human protein properties. This study represents the first successful expression and purification of TMPRSS4 proteins, including its catalytic domain (CD) and extracellular domain (ECD), in a soluble secretory form. To achieve this, we explored the potential of expressing TMPRSS4 as a soluble protein in mammalian cells, with a focus on secretion into the culture media using alternative signal peptides.

After an extensive literature search, we identified five promising signal sequences to fuse with TMPRSS4 CD. These included a signal peptide from the serine protease family (Thrombin), two well-documented signal peptides for heterologous protein secretion (Human Serum Albumin and Azuricidin), and two signal peptides from CHO endogenous proteins (Vascular Endothelial Growth Factor C - VEGF C and Fibroblast Growth Factor - FGF). The amino acid sequences of these signal peptides are listed in Table 1.

S.No	Signal Peptide	Secretory Signal	Secretory Signal Amino Acid Sequence	Discrimination Score
1	Sp1	Human Trypsin	Mnplliltfvaaala	0.758
2	Sp2	Human Serum Albumin	Kwvtflsllflfssays	0.892
3	Sp3	Human Azuricidin	Mtrltvlallagllassra	0.876
4	Sp4	Rat Vegf C	Mhllcflslacsllaaalipgpreapatvaa	0.825
5	Sp5	Chinese Hamster Fgf	Mlgtclrllvgvlcsacslgtvra	0.848

Table 1: List of the alternative signal peptides (SP) used in the study to express TMPRSS4 CD as secretory in HEK F and Expi CHO Cells.

Source: Author

All five signal sequences were fused to the N-terminus of TMPRSS4 CD and ECD. We assessed their potential by predicting discrimination scores (D-scores) using the Signal P 4.1 secretome computation tool, confirming their suitability as potent secretory signal peptides. We generated expression constructs by fusing each signal peptide-coding gene with the TMPRSS4 CD gene in the pcDNA3.4 vector backbone. These constructs were transiently transfected into Expi CHO and HEK F cells, and protein expression was monitored for 7-9 days. Expression was confirmed by Western blot analysis using a TMPRSS4 antibody in culture media samples collected from Day 1 of transfection.

Although the prediction algorithm scores correlated with the observed secretory protein expression profiles, there were significant differences in protein secretion levels among the signal peptides. Human Serum Albumin, Thrombin, and Azuricidin signal peptides demonstrated the highest performance as heterologous secretory signal peptides, aligning with previous reports. In contrast,

VEGF and FGF signal peptides resulted in weaker protein secretion, despite their significant secretome prediction scores.

Understanding the role of signal peptides in protein secretion, significance of the signal recognition particle (SRP) in recognizing signal peptides during protein synthesis is crucial. The interaction between the hydrophobic groove of the M-domain in SRP and the hydrophobic H-domain of the signal peptide is fundamental. Furthermore, specific amino acid positions in the signal peptide, such as positions -1 and -3, are essential for efficient secretion, and careful selection of nonnative signal peptides helps in attaining secretory expression.

4. Conclusion

Our Western blot analyses confirmed the expression of TMPRSS4 CD and ECD in the culture media, helping identify suitable alternative signal peptides for secretory protein expression. In summary, this study reports the successful expression and purification of biologically active TMPRSS4 domains in both E. coli and mammalian cells, with the latter being achieved in a secretory manner. This approach offers an effective means of obtaining active TMPRSS4 variants for structural and functional investigations. We anticipate that the strategies employed here could be applied to express challenging proteins with significant biological importance. Cautious experimental design, including the use of fusion tags and specific signal sequences, can enhance expression levels and address solubility challenges.

Acknowledgments

We extend our heartfelt appreciation to the management of Aragen Life Sciences for their generous support in facilitating the execution of these experiments and essential analyses for our research. Our sincere thanks also go to Dr. Lakshman Rajagopalan for granting us the opportunity to conduct these studies. We would like to convey our gratitude to Raghu Bheemanathi, Aparna Bhagavathula, Alderson Joshi, Dr. Parag Khursade, and Ramesh Ramagoni for their invaluable assistance during the execution of our experimental work. The success of this endeavours was made possible through the collective support and collaborative efforts of all those mentioned above.

References:

1. Bugge TH, Antalis TM, Wu Q, 2009. Type II transmembrane serine proteases. J Biol Chem. 2009 Aug 28;284(35):23177-81. https://doi.org/10.1074/jbc.R109.021006. Epub, PMID: 19487698; PMCID: PMC2749090.
2. Wallrapp C, Hähnel S, Müller-Pillasch F, Burghardt B, Iwamura T, Ruthenbürger M, Lerch MM, Adler G, Gress TM, 2000. A novel transmembrane serine protease (TMPRSS3) overexpressed in pancreatic cancer. Cancer Res. 2000 May 15;60(10):2602-6. PMID: 10825129.
3. Joyce JA, Pollard JW, 2009. Microenvironmental regulation of metastasis. Nat Rev Cancer. Apr;9(4):239-52. https://doi.org/10.1038/nrc2618 Epub 2008 Mar 12. PMID: 19279573; PMCID: PMC3251309.
4. Jung H, Lee KP, Park SJ, Park JH, Jang YS, Choi SY, Jung JG, Jo K, Park DY, Yoon JH, Park JH, Lim DS, Hong GR, Choi C, Park YK, Lee JW, Hong HJ, Kim S, Park YW, 2008. TMPRSS4 promotes invasion, migration and metastasis of human tumor cells by facilitating an epithelial-mesenchymal transition. Oncogene, Apr 17;27(18):2635-47. https://doi.org/10.1038/sj.onc.1210914. Epub 2007 Oct 29. PMID: 17968309.
5. Min HJ, Lee Y, Zhao XF, Park YK, Lee MK, Lee JW, Kim S., 2014. TMPRSS4 upregulates uPA gene expression through JNK signaling activation to induce cancer cell invasion. Cell Signal. Feb;26(2):398-408. https://doi.org/10.1016/j.cellsig.2013.08.002 Epub 2013 Aug 24. PMID: 23978400.
6. Li, T., Zeng, ZC., Wang, L. *et al., 2011.* Radiation enhances long-term metastasis potential of residual hepatocellular carcinoma in nude mice through TMPRSS4-induced epithelial–mesenchymal transition. *Cancer Gene Ther* **18**, 617–626 https://doi.org/10.1038/cgt.2011.29
7. Hooper JD, Clements JA, Quigley JP, Antalis TM., 2001. Type II transmembrane serine proteases. Insights into an emerging class of cell surface proteolytic enzymes. J Biol Chem. Jan 12;276(2):857-60. https://doi.org/10.1074/jbc.R000020200. PMID: 11060317.
8. Page MJ, Di Cera E., 2008. Evolution of peptidase diversity. J Biol Chem. Oct 31;283(44):30010-4. https://doi.org/10.1074/jbc.M804650200. Epub 2008 Sep 3. PMID: 18768474; PMCID: PMC2573091.
9. Page MJ, Di Cera E., 2008. Serine peptidases: classification, structure and function. Cell Mol Life Sci. Apr;65(7-8):1220-36. https://doi.org/10.1007/s00018-008-7565-9 PMID: 18259688.

10. Murray AS, Varela FA, List K., 2016. Type II transmembrane serine proteases as potential targets for cancer therapy. Biol Chem. Sep 1;397(9):815-26. https://doi.org/10.1515/hsz-2016-0131. PMID: 27078673; PMCID: PMC4981541.
11. Hames, B. D., & Hoopepr, N. M. (2005). *Biochemistry* (3rd ed). Taylor & Francis.
12. Netzel-Arnett S, Hooper JD, Szabo R, Madison EL, Quigley JP, Bugge TH, Antalis TM, 2003. Membrane anchored serine proteases: a rapidly expanding group of cell surface proteolytic enzymes with potential roles in cancer. Cancer Metastasis Rev. Jun-Sep;22(2-3):237-58. https://doi: 10.1023/a:1023003616848. PMID: 12784999.
13. Choi SY, Shin HC, Kim SY, Park YW., 2008. Role of TMPRSS4 during cancer progression. Drug News Perspect. Oct;21(8):417-23. https://doi.org/10.1358/dnp.2008.21.8.1272135. PMID: 19034347.
14. Knappskog S, Ravneberg H, Gjerdrum C, Trösse C, Stern B, Pryme IF., 2007. The level of synthesis and secretion of Gaussia princeps luciferase in transfected CHO cells is heavily dependent on the choice of signal peptide. J Biotechnol. Mar 10;128(4):705-15. https://doi.org/10.1016/j.jbiotec.2006.11.026 Epub 2006 Dec 12. PMID: 17316861.
15. Zhang L, Leng Q, Mixson AJ., 2005. Alteration in the IL-2 signal peptide affects secretion of proteins in vitro and in vivo. J Gene Med. Mar;7(3):354-65. https://doi.org/10.1002/jgm.677 PMID: 15619290.
16. Kober L, Zehe C, Bode J., 2013. Optimized signal peptides for the development of high expressing CHO cell lines. Biotechnol Bioeng. Apr;110(4):1164-73. https://doi.org/10.1002/bit.24776 Epub 2013 Jan 17. PMID: 23124363.
17. Petersen TN, Brunak S, von Heijne G, Nielsen H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. Sep 29;8(10):785-6. https://doi.org/10.1038/nmeth.1701. PMID: 21959131.
18. de Aberasturi AL, Calvo A., 2015. TMPRSS4: an emerging potential therapeutic target in cancer. Br J Cancer. 2015 Jan 6;112(1):4-8. https://doi.org/10.1038/bjc.2014.403 Epub 2014 Sep 9. PMID: 25203520; PMCID: PMC4453593.