**Abstract: Modeling the Cleavage Potential of Cross-Species TMPRSS2 Variants to Subunits of SARS-CoV-2**

**Tyshawn Ferrell1,2, Kelly Pierce1, Johnny Aldan1, Ethan Ho1, Dairian Balai 1**

1Texas Advance Computing Center**, University of Texas at Austin, Austin, TX, 78712-78758**

2Department of Biology, **Albany State University, Albany, GA 31705**

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a positive-stranded RNA virus, has caused >100,000 infections and >50,000 deaths during the 2020 pandemic. SARS-CoV-2 successful infection rate has been due to structural configurations of its envelope and specific cell receptor targeting. The spike (S) glycoprotein mediates cell entry by recognizing the angiotensin converting enzyme II (ACE2) receptor, which initializes cleavage of the protein to induce cell membrane fusion. However, in previous studies transmembrane protease/serine subfamily member 2 (TMPRSS2) have increased cleavage efficiency towards the S protein subunits. This study aims to characterize cross-species variants of TMPRSS2 and its potential to cleave deviates of S proteins in silico. We hypothesize that TMPRSS2-like endoprotease aids in the spillover of SARS-CoV-2 to alternate host by cleavage S protein subunits. We will investigate conservation of TMPRSS2 sequence and structure amongst species and assess their similarity to human isoforms. We will utilize protein and genomic databanks to characterize the amino acid changes that would alter TMPRSS2’s cleavage of SARS-CoV-2 S protein. Additionally, databanks on the variation of TMPRSS2 within human populations will be used to gain insight on genetic drift. This will be used to create a model that scores the susceptibility of various species to SARS-CoV-2 based off phylogenetic changes within TMPRSS2. This study will provide insight on possible spillover into alternate host with structurally similar TMPRSS2 strains compared to humans; while tracking the infectivity of SARS-CoV-2 by means of mutation in TMPRSS2 within human populations.

**Introduction**

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a novel betacoronavirus, that curtailed the world’s health system within a matter of months. As of July 20th, 2020, 14,348,858 laboratory-confirmed infections and 603,000 SARS-CoV-2 related deaths have been reported globally [12]. This is due to SARS-CoV-2 mode of transmission and high bonding affinity for the angiotensin converting enzyme II (ACE2) receptor in airway epithelial cells. More specifically the S1 subunit of the S protein binds to ACE2 via its receptor binding domain (RBD), while the S2 subunit is used for virus-cell membrane fusion [1,3,6]. In previous studies it has been shown that SARS-CoV have a preference for cells with an endoprotease, transmembrane protease/serine subfamily member 2 (TMPRSS2), which increases the ability for virus-host membrane fusion [6,7,9]. In silico studies have shown that SARS-CoV-2 ability to infect alternate host depends on ACE2 based on the conserveness of the enzyme compared to the human strain [8]. Though ACE2 is an essential component for attachment of the S protein, it is the host cell proteases that activates the S2 subunit that holds the transmembrane fusion machinery.

What is TMPRSS2?

TMPRSS2 is a member of the type II Transmembrane Serine Protease (TTSP), specifically belonging to the Hepsin/TMPRSS subfamily. TTSPs have four defining features: an N-terminal intracellular domain, transmembrane domain, “stem” domain, and a proteolytic domain. TMPRSS2 is synthesized as a zymogen that requires proteolytic processing to activate [2]. This activation results from the cleavage of extracellular substrates initialized by its Serine residue at the active site. In previous studies, TMPRSS2s proteolytic domain has shown to have increase pathogenesis by cleaving monobasic cleavage sites of virus transmembrane subunits within the coronavirus family [3-7]

How does TMPRSS2 affect the S protein

The S protein in SARS-Cov-2 has an multibasic S1/S2’ cleavage site containing several arginine residues that are cleave by host cell proteases, which increases the efficacy of cell-cell spread [3]. However, in order for the virus-cell transmembrane fusion to occur, the monobasic S2’ cleavage site must be cleaved by TMPRSS2 [3-4]. TMPRSS2 proteolytic activity results from its catalytic domain, which consist of Ser-His-Asp, but Ser initializes the peptide hydrolysis by attacking the acyl compound of a lys or Arg residue in the S protein [2,4]. During this reaction two tetrahedral intermediates and one acylenzyme, but finalizes with the protonation of an amine leaving group, carboxylic acid, water, and the catalytic traid reforms.

Why is TMPRSS 2 relevant to your project (what are you looking at specifically)

Even though TMPRSS2 is a crucial part for virus-host transmembrane fusion, much of its functionality is still unclear. In previous studies, ACE2 variants have been observed in multiple species, which may have lead SARS-CoV2 high infectivity in native and foreign host [1,8]. Since there are over 9000 variants of TMPRSS2, a comparative analysis of its amino acid sequence can predict its cleaving efficacy of SARS.CoV2 with similar ACE2 receptor sequences. This will yield valuable insight into SARS.CoV2 zoonotic transmission potential to spill over in specie populations.

**Work in progress below**

However, could TMPRSS2 variants in other species with similar ACE2 receptors (compared to human strains) lead to unfavorable bonding conditions for SARS-CoV-2. By comparing phylogenetic changes of TMPRSS2 amongst various species with humans, we’ll be able to score the susceptibility/severity of infection for each species, in theory. Also, we’ll look at the spike (S) glycoprotein variants to see their bonding fitness to ACE2/TMPRSS2. By determining the molecular configuration leading to cell entry, we’ll gain insight on spillover potential of SARS-CoV-2 to alternate host (such as domesticated animals). Which could arbor adaptive advantages for SARS-CoV-2 immune evasion and overall infectivity within broad host range.

**Methods**

Using protein blast to check for other organisms that have similar amino acid sequences of TMPRSS2 compared to human isoforms. Currently we are similar organisms are sheep (ovis aries), domestic yak (hypothetical protein- bos mutus), African grass rat (arvicanthis niloticus),Bison (predicted), Alpaca (vicugna pacos), Zebu (Bos indicus x Bos Taurus), African rodent (grammomys surdaster), dromedary (camelus dromedarius), sea otter (Enhydra lutris kenyoi), north American river otter (lontra canadensis), pale spear-nosed bat (phyllostomus discolor), house mouse (mus musculus),

**References**

1. Damas, J., Hughes, G. M., Keough, K. C., Painter, C. A., Persky, N. S., Corbo, M., Hiller, M., Koepfli, K. P., Pfenning, A. R., Zhao, H., Genereux, D. P., Swofford, R., Pollard, K. S., Ryder, O. A., Nweeia, M. T., Lindblad-Toh, K., Teeling, E. C., Karlsson, E. K., & Lewin, H. A. (2020). Broad Host Range of SARS-CoV-2 Predicted by Comparative and Structural Analysis of ACE2 in Vertebrates. bioRxiv : the preprint server for biology, 2020.04.16.045302. <https://doi.org/10.1101/2020.04.16.045302>
2. Hedstrom, L. (2002). Serine protease mechanism and specificity. *Chemical reviews*, *102*(12), 4501-4524.
3. Hoffmann, M., Kleine-Weber, H., & Pöhlmann, S. (2020). A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. *Molecular Cell*.
4. Hoffmann, M., Hofmann-Winkler, H., & Pöhlmann, S. (2018). Priming time: How cellular proteases arm coronavirus spike proteins. In *Activation of Viruses by Host Proteases* (pp. 71-98). Springer, Cham.
5. Hofmann, H., & Pöhlmann, S. (2004). Cellular entry of the SARS coronavirus. Trends in microbiology, 12(10), 466-472.
6. Matsuyama, S., Nao, N., Shirato, K., Kawase, M., Saito, S., Takayama, I., ... & Sakata, M. (2020). Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. Proceedings of the National Academy of Sciences, 117(13), 7001-7003.
7. Meng, T., Cao, H., Zhang, H., Kang, Z., Xu, D., Gong, H., ... & Wei, H. (2020). The insert sequence in SARS-CoV-2 enhances spike protein cleavage by TMPRSS. BioRxiv.
8. Paniri, A., Hosseini, M. M., & Akhavan-Niaki, H. (2020). First comprehensive computational analysis of functional consequences of TMPRSS2 SNPs in susceptibility to SARS-CoV-2 among different populations. Journal of Biomolecular Structure and Dynamics, (just-accepted), 1-18.
9. Shirato, K., Kawase, M., & Matsuyama, S. (2018*). Wild-type human coronaviruses prefer cell-surface TMPRSS2 to endosomal cathepsins for cell entry*. Virology, 517, 9-15.
10. Shulla, A., Heald-Sargent, T., Subramanya, G., Zhao, J., Perlman, S., & Gallagher, T. (2011). *A transmembrane serine protease is linked to the severe acute respiratory syndrome coronavirus receptor and activates virus entry*. Journal of virology, 85(2), 873-882.
11. Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veesler, D. (2020). *Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein*. Cell.
12. WHO, 2020 Who, *Novel Coronavirus(2019-nCoV) Situation Report 182*

https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200720-covid-19-sitrep-182.pdf?sfvrsn=60aabc5c\_2

Notes.

Partial cleavage can be accomplished by using enzymes, such as trypsin and chymotrypsin, that break bonds between specific amino acids.

* Trypsin: Cleaves the chain at the carboxyl groups of the basic amino acids lysine and arginine. The cleavage takes place in such a way that the amino acid with the charged side chain ends up at the C-terminal end of one of the peptides produced by the reaction. A peptide can be automatically identified as the C-terminal end of the original chain if its C-terminal amino acid is not a site of cleavage.
* Chymotryspin: Cleaves the chain at the carboxyl groups of the aromatic amino acids phenylalanine, tyrosine, and tryptophan.
* Protein domain- a conserved part of a given protein sequence and tertiary structure that can evolve, function, and exist independently of the rest of the protein chain.
* Quaternary structure is the final level of protein structure which is formed by two or more polypeptide chains. Each chain is termed a subunit of the structure.
  + Dimers, trimers, tetramers, consist of two, three, and four polypeptide chains, respectively.
  + Oligomer- an aggregate of several smaller units (monomers); bonding may be covalent or noncovalent.
    - Chains with noncovalent bond result in subtle changes that can alter the structure at one site of the protein molecule. This is known as allosteric hindess, which can cause conformational change in one subunit that induces a change in another subunit.
* Consensus Sequences- DNA sequences to which RNA polymerase binds; they are identical in many organisms
* Protein kinases- a class of enzymes that modify a protein by attaching a phosphate group to it
* I hypothesis that variation in the serine domain of TMPRSS2 will have an adverse effect on cross-species transmission of COVID-19. Specifically focusing on domesticated animal’s TMPRSS2 amino acid sequence. Protein folding and bonding affinity play a critical role in determining the cleavage potential for the proteolytic domain of TMPRSS2.

Code

* I want to compare two alignment sequences to return both similarities and differences with their positions labeled
* I want to pull a particular *species*  from the main list of species and place it as a variable
* Run a multisequence alignment on the species to find the consensus sequence in the Desired species list (DS)
* Take the position of the DS that matches with the positions of the trimmed Human sequence
* Convert the DS consensus sequence to a SeqRecord to compare to the human sequence
* Run a multiple sequence alignment of both DS and human
* Find the frequency of changes at specific sequence locations compare to the human sequence
  + Find out if these changes result in changes in amino acid properties (specifically hydrophobic to hydrophilic or basic to acid)
  + Also find other differences in the amino sequence such as changes in disulfide bonding
* Find the similarities in the two sequences and record the position of those similarities