Identifying Mutations in ACE2 That Influence Susceptibility to SARS-CoV-2

Zak White
zakerywhite@uvic.ca
Department of Computer Science
University of Victoria
Victoria, BC, Canada

Abstract

SARS-CoV-2 is an contagious virus established to affect not only humans, but other mammal species. Studies over the last two years have revealed certain species are distinctly immune to the virus, a reality this report aims to understand. By analyzing the differences between sequences of the virus' target protein, ACE2, in hosts with varying susceptibility to the virus, this study investigates mutations that may impact a host's vulnerability.

The analysis identifies eight mutations that may influence susceptibility (D31, A41, A66, F83, N113, H353, S426, V679), five of which are referenced in other studies.

CCS Concepts: • **Applied computing** → **Bioinformatics**; **Computational biology**; **Molecular sequence analysis**; **Molecular evolution**.

Keywords: sequence alignment, SARS-CoV-2, angiotensin-converting enzyme 2 (ACE2), protein mutations, viral susceptibility

ACM Reference Format:
Zak White. 2021. Identifying Mutations in ACE2 That Influence Susceptibility to SARS-CoV-2. In *UVic CSC482B: Computational Biology Algorithms, Fall 2021.* ACM, New York, NY, USA, 4 pages.

UVic CSC482B, November 2021, Victoria, British Columbia, Canada

1 Introduction

For over two years, the COVID-19 pandemic has been a hot-topic of research as it inescapably invades our world. COVID-19 is a disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a virus believed have zoonotic origins [1].

Recent animal studies have identified a range of mammal species – including some nonhuman primates, ferrets, hamsters, and bats [2] – that can be infected by and can transmit the virus. Conversely, certain experiments have supported the conclusion that some other species (e.g. the domesticated pig) are explicitly insusceptible to the virus [2][3]. This disjunction promotes the question, *what makes some species susceptible to the virus and others immune?*

The answer might correspond to how the the virus interacts with its target protein in different hosts. The SARS-COV-2 spike protein's target is the angiotensin-converting enzyme 2 (ACE2). Differences in the ACE2 sequence can affect the spike protein's binding affinity, thus affecting the host's susceptibility to the virus.

The objective of this research is to computationally identify mutations that may inhibit the spike protein's interaction with the ACE2 protein in a host.

1.1 Related Work

Li et al. (2005) studied the effects that changes in ACE2 protein had on the binding affinity of the SARS-CoV spike protein. By introducing residues of insusceptible ACE2 proteins into susceptible ones, they identified which mutations abolished interaction with the spike protein. Specifically, mutations D31, A41, 82-84NFS, A353, H353, and A357 restricted the interaction.

Liu et al. (2021) analyzed the structure of the human ACE2 protein to identify critical residues and binding sites; they concluded the amino acids at positions 31, 35, 38, 82, and 353 were crucial for interaction with the spike protein.

2 Methods

2.1 Data

Twenty-five angiotensin-converting enzyme 2 sequences with known susceptibility to the SARS-CoV-2 spike protein

were collected from the NCBI[1] and the UniProt[2] protein databases. This included the sequences for twelve mammal host species established as susceptible, and six mammal host species considered insusceptible.

Species	Suscept.
Human (*Homo sapiens*)	Yes
House cat (*Felis catus*)	Yes[2][6]
Ferret (*Mustela putorius furo*)	Yes[2]
European mink (*Mustela lutreola*)	Yes[2][6]
Cynomolgus macaque (*M. fascicularis*)	Yes[2][3]
Green monkey (*Chlorocebus sabaeus*)	Yes[2]
Common marmoset (*Callithrix jacchus*)	Yes[2][3]
European rabbit (*Oryctolagus cuniculus*)	Yes[2][3]
Big-eared horseshoe bat (*R. macrotis*)	Yes[2]
Short-nosed fruit bat (*C. sphinx*)	Yes[2][3]
White-tailed deer (*O. virginianus*)	Yes[6]
Siberian tiger (*Panthera tigris*)	Yes[6][7]
Pangolin (*Manis pentadactyla*)	No[6][8]
Raccoon (*Procyon lotor*)	No[3]
Greater horseshoe bat (*R. ferrumequinum*)	No[3]
Brown rat (*Rattus norvegicus*)	No[3]
House mouse (*Mus musculus*)	No[3]
Pig (*Sus domesticus*)	No[2][3]

Table 1. Host species from which ACE2 sequences were collected and their susceptibilities to SARS-CoV-2

Each of the mammal sequences was compared with the human sequence using the Needleman-Wunsch algorithm to produce an optimal global alignment. This processing step ensured that each enzyme sequence shared the same length – 805 amino acids – and that the indexing of the sequences correlated to that of the human ortholog.

Seven partial sequences of the ACE2 protein were also retrieved from the UniProt database; each was noted to be not susceptible to the SARS-CoV-2 spike protein. These partial sequences were aligned with the human sequence using the Smith-Waterman algorithm for local alignment.

2.2 Isolating Influential Mutations

The process used for identifying potentially influential mutations followed an iterative string comparison algorithm. For each acid in a negative (insusceptible) sequence, if the same acid did not exist in the same position of some positive (susceptible) sequence, the position and acid were recorded (see Algorithm 1). This process identified all mutations that appeared exclusively in negative sequences.

The algorithm assigns weights to each mutation based on the total number of potential influential indices for each sequence; i.e, each mutation is considered more influential in a sequence with fewer total mutations. The mutations with weights that summed above a specified threshold were considered most influential.

Algorithm 1: Finding potential influential mutations

```
for each negative sequence S⁻ do
    for each i < length of S⁻ do
        isInfluentialIndex ← True;
        for each positive sequence S⁺ do
            if i^th acid in S⁻ == i^th acid in S⁺ then
                isInfluentialIndex ← False;
            end
        end
        if isInfluentialIndex then
            // The mutation may be influential
            Record i, i^th acid in S⁻
        end
    end
end
```

3 Results

The analysis identified 181 total mutations in 149 different sites, and isolated the eight most influential (see Table 2). Each mutation appeared in at least two sequences, and all but one sequence included one of the mutations; the domestic pig sequence did not contain any of the resulting influential mutations. Table 3 shows the residues at the identified positions for each insusceptible sequence, emphasizing the influential mutations.

Pos	Mutation		
31	Lysine (K)	→	Aspartate (D)
41	Tyrosine (Y)	→	Alanine (A)
66	Glycine (G) Arginine (R)	⇒	Alanine (A)
83	Tyrosine (Y)	→	Phenylalanine (F)
113	Serine (S) Arginine (R)	⇒	Asparagine (N)
353	Lysine (K)	→	Histidine (H)
426	Proline (P)	→	Serine (S)
679	Isoleucine (I)	→	Valine (V)

Table 2. Summary of most influential mutations

Five of these selected mutations corresponded to findings in previous studies: Li et al. (2005) found that mutations D31, F83, and H353 abolished the interaction with the spike protein and mutation A41 inhibited it; Lee et al. (2020) found that the Lysine at 31, Tyrosine at 83, and Proline at 426 that appear in human sequences were each crucial residues for susceptibility.

[1] https://www.ncbi.nlm.nih.gov/protein
[2] https://www.uniprot.org/uniprot

There are no published studies suggesting the remaining three mutations (A66, N113, V679) have any effect on the protein's interaction with the spike protein. Similarly, the UniProt entry for the ACE2 protein does not include any reference to the acids at these positions impacting the degree of interaction [10].

4 Discussion

Of the eight influential mutations identified in this project, it is promising that five of them have been shown to influence susceptibility in previous studies. For the remaining three mutations, it is possible that the analysis incorrectly reported these positions as influential (i.e., a false positive). It could also be that these mutations do, in fact, abolish interaction between the SARS-CoV-2 spike protein and its target, and that they have yet to be identified in a published analysis of sequence or structure.

4.1 Considerations for Future Work

A primary restriction of this project was the limited amount of data available about species' susceptibilities to the virus; SARS-CoV-2 is ever-topical and more information about spike protein and ACE2 interaction is still surfacing. More samples of species known to be susceptible would increase the set of non-inhibiting mutations in the algorithm, possibly producing fewer false positives. A broader range of mammal species (more separated phylogenetically) could also help reduce the number of false positives.

Another consideration is the effect of mutations in consecutive sites. This analysis identified position 83 as crucial, but Li et al. (2005) and Li and Qin (2021) have shown that it is the mutation in sites 82-84 together holding an influence over the susceptibility. The analysis could involve a more complex algorithm to try to identify residues that are only influential when mutating together.

It's also important to consider that the issue of susceptibility is not binary: mutations in the ACE2 sequence can increase, inhibit, or abolish interaction with the spike protein [10] - not just abolish, as this study assumed. It would be challenging to perform an analysis that incorporated degree of interaction without more quantitative data about species' susceptibility.

5 Conclusion

There have been consistent results in identifying influential residues in ACE2 in former studies (Liu et al., 2021; Li and Qin, 2021): mutations positions 31, 41, 82-84, 353, 355, and 357 are known to abolish interaction with the SARS-CoV-2 spike protein. This computational analysis successfully identified five of these established positions by comparing amino acid sequences known to be susceptible to the virus with those that are not.

Acknowledgments

I'd like to thank Dr. Hosna Jabbari, who, with the help of TA Alison Ziesel, instructed a computation biology course that offered an interesting introduction to concepts in biology, and provided insight into the intersection between the fields of biology and computer science.

I'd also like to thank Samantha Addison Mitchell, who let me pick her brain as I worked through this project and learned new concepts.

References

[1] K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, and R. F. Garry, "The proximal origin of SARS-CoV-2," *Nature medicine*, vol. 26, no. 4, pp. 450–452, Apr 2020.

[2] B. B. Oude Munnink *et al.*, "Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans," *Science (New York, N.Y.)*, vol. 371, no. 6525, pp. 172–177, Jan 2021. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/33172935

[3] C. C. Sreenivasan, M. Thomas, D. Wang, and F. Li, "Susceptibility of livestock and companion animals to COVID-19," *Special Issue on New coronavirus (2019-nCoV or SARS-CoV-2) and the outbreak of the respiratory illness (COVID-19): Part-X*, vol. 93, pp. 1351–1360, Oct 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/jmv.26621

[4] W. Li *et al.*, "Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2," *The EMBO journal*, vol. 24, no. 8, pp. 1634–1643, Apr 2005. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/15791205

[5] Y. Liu *et al.*, "Functional and genetic analysis of viral receptor ACE2 orthologs reveals a broad potential host range of SARS-CoV-2," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 118, no. 12, p. e2025373118, Mar 2021. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/33658332

[6] M. V. Palmer *et al.*, "Susceptibility of white-tailed deer (Odocoileus virginianus) to SARS-CoV-2," *Journal of virology*, vol. 95, no. 11, pp. e00 083–21, Mar 2021. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/33692203

[7] E. M. Leroy, M. Ar Gouilh, and J. Brugère-Picoux, "The risk of SARS-CoV-2 transmission to pets and other wild and domestic animals strongly mandates a one-health strategy to control the COVID-19 pandemic," *One health (Amsterdam, Netherlands)*, vol. 10, pp. 100 133–100 133, Dec 2020. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/32363229

[8] R. Frutos, J. Serra-Cobo, T. Chen, and C. A. Devaux, "COVID-19: Time to exonerate the pangolin from the transmission of SARS-CoV-2 to humans," *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, vol. 84, pp. 104 493–104 493, Oct 2020. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/32768565

[9] I.-H. Lee, J.-W. Lee, and S. W. Kong, "A survey of genetic variants in SARS-CoV-2 interacting domains of ACE2, TMPRSS2 and TLR3/7/8 across populations," *Infection, Genetics and Evolution*, vol. 85, p. 104507, Nov 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1567134820303385

[10] UniProt. ACE2 - Angiotensin-converting enzyme 2 precursor - Homo sapiens (Human) - ACE2 gene & protein. Accessed November 21, 2021. [Online]. Available: https://www.uniprot.org/uniprot/Q9BYF1

[11] R. Li and C. Qin, "Expression pattern and function of SARS-CoV-2 receptor ACE2," *Biosafety and Health*, Aug 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2590053621000860

Position								
31	41	66	83	113	353	426	679	Sequence
N	Y	**A**	Y	R	K	P	I	Raccoon (*Procyon lotor*)
N	Y	**A**	**F**	**N**	**H**	**S**	**V**	House mouse (*Mus musculus*)
K	Y	G	Y	**N**	K	P	**V**	Pangolin (*Manis pentadactyla*)
K	Y	R	Y	S	K	P	I	Pig (*Sus domesticus*)
K	Y	**A**	**F**	**N**	**H**	**S**	**V**	Brown rat (*Rattus norvegicus*)
D	H	G	**F**	S	K	**S**	I	Greater horseshoe bat (*R. ferrumequinum*)
D	**A**	G	Y	-	-	-	-	Partial sequence 1
-	**A**	**A**	Y	R	-	-	-	Partial sequence 2

Table 3. Amino acid at each influential position of the insusceptible ACE2 proteins. Acids in bold indicate mutations that were computed to be influential.