

Structural, Sequence, and Germline Analysis of SARS-CoV-2 Antibodies Across Humans and Mice

Saanvi Dogra

Del Norte High School, San Diego, California, USA, 92127

KEYWORDS. SARS-CoV-2, Antibodies, Spike Protein, Virus, Mouse

BRIEF. This research compared the structure, sequence, and germline of SARS-CoV-2 antibodies that bind to the RBD between humans and mice, making important contributions to the information known about this family of viruses and potentially helping develop possible treatment methods like synthetic antibodies.

ABSTRACT. Antibodies are proteins that neutralize foreign substances in our body, such as viruses. SARS-CoV-2 is a virus that has caused a respiratory illness, called COVID-19. This project compares the structure, sequence, and germline of SARS-CoV-2 antibodies across humans and mice. It was hypothesized that these characteristics would be similar in both human and mouse antibodies as they target the spike protein of the virus. The antibodies were categorized based on epitope, neutralizing properties, heavy and light V gene germlines, and number of residues in their CDRH3. Three human antibodies and three mouse antibodies with unknown structures were modeled. Their structures were superimposed with PyMOL, and their sequences were aligned with Clustal Omega. The results partially confirmed the hypothesis. The prominent heavy V gene germline for antibodies binding to RBD was IGHV3 for humans and IGHV1 for mice. These antibody germlines were also vastly different for the light V gene (IGKV1 prevalent in humans and IGKV3 prevalent in mice). While the sequences of the antibodies were very different (less than 50% similarity), the structures were similar (RMSD = 0.7Å). This study makes important contributions to the information known about SARS-CoV-2 and may help develop possible treatment methods like synthetic antibodies.

INTRODUCTION.

Antibodies are Y-shaped proteins used to recognize and neutralize foreign substances in our body. Each antibody is specific to a unique pathogen, an infectious agent, called an antigen [1]. Antibodies consist of 4 polypeptide chains—2 identical heavy chains and 2 identical light chains. The heavy chains are more often studied as they are more variable, therefore accounting for more differences between antibodies. At the tips of the chains there are 6 hyper-variable loops, three of which come from the VL (variable light) chain and three of which come from VH (variable heavy) chain. These groups are called CDRs (complementary-determining regions). CDRH3, specifically, is the most variable part of the antibody and plays a crucial role in antigen recognition [2].

Additionally, the germline theory proposes that each antibody was encoded in a different germline gene and antibodies were collectively inherited, explaining antibody diversity. From cloning the immunoglobulin genes, we can see that the antibody collection was generated by DNA rearrangements during B-cell development. Each antibody has a germline, and multiple antibodies can have the same germline yet still be fundamentally different [3]. The V (variable) region of an antibody is at the tip and gives the antibody specificity to bind to an antigen. The V domain is made of two different DNA segments. The first DNA segment contains 95-101 amino acids and is called the V gene. The second segment encodes the remaining segment, up to 13 amino acids, and is called the J (joining) gene. The germlines being investigated in this study are for the light and heavy V genes, names as follows: IGHV (Immunoglobulin Heavy Variable) for the heavy V

gene and IGKV (Immunoglobulin Kappa Variable) or IGLV (Immunoglobulin Lambda Variable) for the light V gene. The V gene is the most important for identification since it has the longest sequence and the most variants in the sequence [4].

In December 2019, SARS-CoV-2, part of the beta coronavirus 2B lineage of the coronavirus family, was identified and has since caused the COVID-19 illness, resulting in more than 6 million deaths. At the time, quarantining for weeks had become normal, stories of our friends and family falling sick were common, and schools and workplaces are slowly dealt with shutting down. The impact SARS-CoV-2 had on our lives inspired me to research this virus, potentially limiting its spread and aiding treatment of other related viruses in the future.

The spike protein on SARS-CoV-2 is divided into S1 and S2. S1 contains the receptor-binding domain (RBD). The RBD is essential for the viral entry into the host cell and the most common binding site, so only antibodies binding to RBD will be studied in this research [5]. ACE2 (Angiotensin-Converting Enzyme 2) is an enzyme that allows SARS-CoV-2 to infect and destroy our cells. Since the mouse ACE2 is so different from the human one, mice only get mild symptoms when infected with this virus. However, scientists have developed new strains of SARS-CoV-2 that recognize the mouse ACE2 and cause much more severe symptoms in mice [6]. Through this approach, many more mouse antibodies targeting SARS-CoV-2 have been produced and sequenced, causing the research in this area to rapidly progress. Examining the difference in antibodies between both species could provide more information on how humans can more effectively deal with the virus, just as mice do, and reveal the viability of a mouse model for vaccine testing [6].

To investigate the differences in infections between humans and mice, this study compared the structures, sequences, and germlines of natural human and natural mouse antibodies targeting the SARS-CoV-2 virus. It was hypothesized that the mouse and human antibodies will have similar structures, germlines, and sequences since they all target the SARS-CoV-2 envelope glycoprotein.

MATERIALS AND METHODS.

Collecting Data.

Three groups of antibodies were downloaded from the Oxford Protein Informatics Group Coronavirus Antibody Database (87 antibodies binding to SARS-CoV-2 on the S1, 233 antibodies binding to SARS-CoV-2 on the S2 protein, 1584 antibodies binding to SARS-CoV-2 on the RBD). Synthetic antibodies and antibodies with undetermined sequences were removed.

Categorizing Human and Mouse Data.

The antibody groups were categorized into human and mouse antibodies. The human antibodies binding to RBD were further categorized into neutralizing or non-neutralizing antibodies, then based on their heavy V gene and light V gene. Those with the IGHV3 germline (most common heavy V gene germline) were categorized based on their heavy V gene subclassification and based on the number of residues

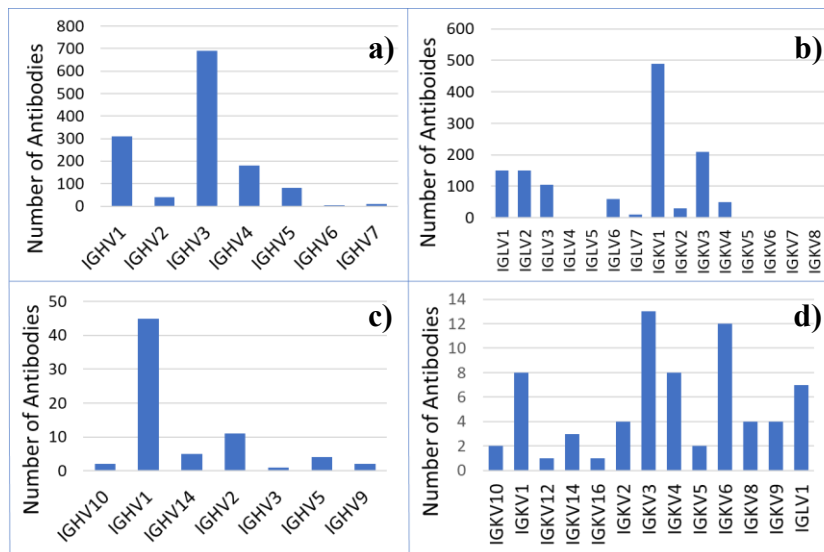


Figure 1. Germline classification of the **a)** heavy V gene and **b)** light V gene for natural human antibodies binding to RBD, as well as the germline classification of the **c)** heavy V gene and **d)** light V gene for natural mouse antibodies binding to RBD

in the CDRH3. Those with the IGKV1 germline (most common light V gene germline) were then further categorized based on their light V gene subclassification and also based on the number of residues in their CDRH3. The process was repeated for mouse antibodies binding to RBD, with IGHV1 as the most common heavy V gene germline and IGKV3 as the most common light V gene germline.

Comparing Antibody Structure.

Three human antibodies that bind to RBD (4A3, Ab 510H2, and C112) were chosen and BlastP was conducted on PDB to determine that they had no known structure. The antibody sequence was inputted into the abYsis modeling tool and modeled the heavy chain structure of the antibodies. The color of CDRH3 of the antibody was changed on PyMOL to make it visible. The basic structure, the surface structure, and the electrostatic potential structure were modeled with PyMOL. The three antibodies were superimposed on Pymol to find the RMSD value. The CDRH3 of the antibodies were extracted and superimposed on Pymol to find their RMSD value. The heavy chain of 4A3, Ab 510H2, and C112 was aligned with Clustal Omega, and the identity matrix was produced. The CDRH3 of 4A3, Ab 510H2, and C112 was aligned with Clustal Omega, and their identity matrix was also produced. The process was repeated with three mouse antibodies that bind to RBD (1E02, 1H06, 2C02). One of the three human antibodies (Ab_510H2) and one of the three mouse antibodies (C202) was chosen and superimposed on PyMOL to determine the RMSD value. The CDRH3 of Ab_510H2 and C202 was also extracted and superimposed on Pymol to find their RMSD value. The heavy chain sequence of Ab_510H2 and C202 was aligned on Clustal Omega, and the identity matrix was produced. The CDRH3 of Ab_510H2 and C202 was aligned on Clustal Omega, and their identity matrix was produced.

RESULTS.

Figure 1 shows the number of antibodies for each respective germline category in the heavy V or light V gene of natural human or mouse antibodies taken from the database. For human antibodies, IGHV3 is the most prevalent for the heavy V gene and IGKV1 for the light V gene. For mouse antibodies, IGHV1 is the most common for the heavy V gene and IGKV3 for the most common light V gene.

In the heavy chain sequence alignment of Ab_510H2 and C202 we can see a relatively low identity percentage of 48.72%. In the CDRH3 sequence alignment there is a lower percentage of 11.11%.

	Ab_510H2	2C02
a) Ab_510H2	100	48.72
2C02	48.72	100

	Ab_510H2	2C02
b) Ab_510H2	100	11.11
2C02	11.11	100

Figure 2. Matrix comparing the percentage similarity between the **a)** heavy chain sequence alignment and **b)** CDRH3 sequence alignment of the Ab_510H2 human antibody and C202 mouse antibody.

The heavy chain superimposition of the human antibody Ab_510H2 and the mouse antibody C202 had a low RMSD value of 0.7 Å and they had an RMSD value of 0.01 Å, indicating high structural similarity. CDRH3s

DISCUSSION.

The prevalent germlines of the natural human antibodies and the prevalent germlines of the natural mouse antibodies binding to RBD on SARS-CoV-2 are different (Figure 1, S1). The similarity between the heavy chain sequences and between the CDRH3 sequences of the human antibody and the mouse antibody was also very low (Figure 2). However, the superimposition of the human and mouse antibody had a low RMSD values, indicating high structural similarity in the heavy chains and CDRH3s (Figure 3). Therefore, the structures of the human and mouse antibodies were similar, but the germlines and sequences differed. The results can be explained by the fundamental physical and chemical difference between human and mice, causing each organism to develop antibodies specifically suiting their needs. On the other hand, the structures of the antibodies were similar because similar structures are needed to target the RBD on the SARS-CoV-2 virus.

The CDRH3 length of human and mouse antibodies were also compared. The most common number of residues in the CDRH3 of human antibodies and in the CDRH3 of mouse antibodies binding to RBD with their respective prevalent heavy V gene germlines are the same but with their respective prevalent light V gene germlines are different (Fig-

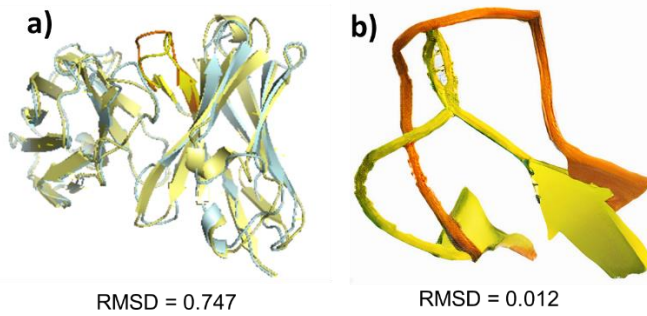


Figure 3. RMSD value and superimposition of **a)** the heavy chains and of **b)** the CDRH3s of the human Ab_510H2 and mouse C202 antibodies

ure S1). Again, these results were partially similar due to the same epitope being targeted.

When comparing antibodies of the same species against one another, more results were concluded. For the three human antibodies modeled, their superimpositions showed the heavy chain structures were very similar and the CDRH3 structure similarity widely varied (Figure S2). Also, the heavy chain sequences had high similarity and the CDRH3 sequences had low similarity percentage (Figure S3). Comparison with the three modeled mouse antibodies followed the same pattern (Figure S4, S5). These results can be explained by the fact that different antibodies are very specific to different antigens, and they evolve to target these evolving antigens better. The heavy chain sequence is more consistent among antibodies, and the antibodies all target the same virus, so there is a higher similarity percentage. The CDRH3 is highly variable since it interacts with SARS-CoV-2, causing the CDRH3 across different antibodies to have a low similarity percentage.

Since the CDRH3 of the three modeled human antibodies and the three modeled mouse antibodies is negatively charged, we can predict that the corresponding binding sites on SARS-CoV-2 are positively charged, since the CDRH3 interacts with the virus (Figure S6, S7). Opposite charges attract, so this interaction would only be possible if the epitope had the opposite, positive charge. To dive deeper into the impact a shorter or longer CDRH3 has on the antibody's ability to neutralize the virus. This project can also be used to develop synthetic antibodies since the epitope of the antibodies and other various characteristics are contributed to knowing more about the virus and help develop possible treatment methods.

Overall, the goal to analyze SARS-CoV-2 antibodies and compare the structure, sequence, and germline in natural human and mice antibodies targeting SARS-CoV-2 was achieved. In the future, studies comparing human antibodies targeting SARS-CoV-2 and antibodies of other animals could also be conducted. Experimentation could be used to dive deeper into the impact a shorter or longer CDRH3 has on the

antibody's ability to neutralize the virus. This project can also be used to develop synthetic antibodies since the epitope of the antibodies and other various characteristics are categorized. This research could also be used to determine if the corresponding binding sites on SARS-CoV-2 of the antibodies with negatively charged CDRH3s have a positive charge, as predicted. Learning more information about the antibodies targeting SARS-CoV-2 can contribute to knowing more about the virus and help develop possible treatment methods.

SUPPORTING INFORMATION.

Supporting information includes graphs comparing the CDRH3 residue length in human and mouse antibodies; structural and sequence similarity between all the human antibodies; structural and sequence similarity between all the mouse antibodies; and the basic structure, surface structure, and electrostatic potential models portraying both the analyzed human and mouse antibodies.

ACKNOWLEDGMENTS.

I would like to thank my mentor Sonu Kumar, the Senior Scientist in Discovery Biotherapeutics at Bristol Myers Squibb, for providing me with the tools necessary for this research.

REFERENCES

1. A. Reinholm, S. Maljanen, P. Jalkanen, Neutralizing antibodies after the third COVID-19 vaccination in healthcare workers with or without breakthrough infection. *Commun Med* **4**, 28 (2024)
2. K. D. Elgert, Ed., *Immunology: Understanding the Immune System* (Wiley, ed. 2, 2009)
3. C. Gaebler, Z. Wang, J. C. C. Lorenzi, Evolution of Antibody Immunity to SARS-CoV-2. *Nature* **591**, 639-644 (2021)
4. C. A. Janeway, P. Travers, M. Walport, M. Shlomchik, *Immunobiology: The Immune System in Health and Disease* (Garland Publishing, ed. 5, 2001)
5. S. Ludwig, A. Zarbock, Coronaviruses and SARS-CoV-2: A Brief Overview. *Anesthesia and Analgesia* **131**, 93-96 (2020)
6. S. Clever, A. Volz, Mouse models in COVID-19 research: analyzing the adaptive immune response. *Med Microbiol Immunol* **212**, 165-183 (2023)



Saanvi Dogra is a student at Del Norte High School in San Diego, CA.