Using Immunoinformatics to Isolate a Pan-Hantavirus CD4+ Epitope for use in vaccine constructs

Sai Sahvir Bhaskaruni

Abstract

The project is aimed to determine if there is a possibility for the development of a pan-hantavirus vaccine by testing for the possibility of a creation of $CD4^+$ (Helper T-Cell) epitope that induces interferon gamma (IFN-γ). By using the available genomic sequences of all strains of hantavirus in GenBank, tests were conducted to test for conservancy. These results were then used to test for the existence of a $CD4^+$ epitope, from which tests were then conducted to determine if one of the epitopes were able to secrete IFN-γ. The top three isolated epitopes, ranked based on their MHC-II binding capacity, IFN-γ binding capacity, and antigenicity, were then synthesized to confirm their immunogenic potential *in-vitro*. After co-culture with a T-Cell and Macrophage, statistical significance of the immunogenetic potential of the epitopes will be analyzed. This data will then be used to determine the viability of the possible eventual creation of a pan-hantavirus vaccine.

Introduction

The recent pandemic has resulted in a wave of increased awareness about the threat that pathogens pose towards public health. In response to this threat a flurry of new research on Pan-Viral influenza or coronavirus vaccines is being conducted (Paules et al., 2017; Prakash et al., 2021). Though, the threat of another family of dangerous pathogens, hantaviruses, is being ignored. Hantaviruses are currently lacking a broadly approved treatment or vaccine to combat its infection. In fact it is even considered an emerging public health threat (Hangaragi, 2020).

Hantaviruses are a family of enveloped, negative-sense RNA viruses in the family Bunyaviridae. When in humans, hantaviruses can cause one of two diseases: Hantavirus Cardiopulmonary Syndrome (HCPS), and Hemorrhagic Fever with Renal Syndrome (HFRS). HCPS has a fatality rate of around 40%, and HFRS has a fatality rate of 15% (Liu et al., 2020). Combined with the lack of an efficacious treatment and a severely limited capability for prevention, Hantavirus poses an acute risk to public health.

This means that there is an immediate need for the development of a vaccine. Yet, in pursuit of a vaccine there are multiple factors that should be considered. Among them is the unique ability of hantaviruses to undergo whole genome reassortment, a feat that is a common occurrence in nature (Kim et al., 2016). Hantaviruses combine amongst themselves to transfer genetic material, creating new reassortant viruses with properties from both parent viruses. Any vaccine that is constructed has to take this into account, meaning that any vaccine constructed should be pan-viral .

The genome of hantavirus has three strands: the small (S) strand, the medium (M) strand, and the large (L) strand. The S strand encodes the nucleocapsid protein, the M strand encodes the external glycoproteins, and the L strand encodes viral polymerase (Muyangwa et al., 2015). The external glycoproteins are the main factor that enable virus-membrane fusion (Mittler et. al, 2017), and are the target of the immune response to hantaviruses. For this reason the majority of hantavirus vaccines in development target the M gene and the resulting surface glycoproteins (Liu et al., 2020).

This approach is no different. Aiming to determine if the construction of a pan-Hantavirus vaccine is viable, this study explores the possibility of pan-viral immunity through the $CD4^+$ pathway. In order to isolate the part of the M-gene that would be used in a vaccine, this study followed the precedent of a recent study that developed a pan-Coronavirus vaccine using *in-silico* methods (Prakash et al., 2021).

Activating the adaptive immune system through the release of interferon gamma (IFN-γ), the $CD4^+$ pathway is a vital part of the human immune response (Kak et al., 2018). A $CD4^+$ epitope that is part of a vaccine that aims to induce a strong immune response should result in the release of IFN-γ, and an *in-vitro* study should confirm the ability of isolated epitopes to do so.

Method

In the design of a multi-epitope pan-viral epitope, the first step is to collect and identify the target sequences. As this study is aimed at creating a pan-Hantavirus vaccine, this means collecting all of the sequences for the strains of hantavirus. GenBank was used to collect the recognized 47 strains of hantavirus (see figure 1) (Laenen et al., 2019). Whole M segment protein sequences were used, and as one could not be found for the Black Creek Canal orthohantavirus sequence it was excluded from this study.

After the remaining 46 strains were collected they were aligned using the default settings on the COBALT program through NCBI. This alignment can be seen through the NCBI Multiple Sequence Alignment viewer (see figure 2). From this alignment an initial judgment of homology can be conceived. The areas highlighted in red signify regions of strong conservancy, those in blue signify regions of low conservancy, while the rest in regions with no or limited conservancy.

The alignment was then used to construct a conserved protein sequence that represents conserved regions of all hantavirus strains. Using the default settings on the EMBOSS cons program a conserved sequence was constructed (see figure 3). The x's in the sequence signify places on the amino acid sequence where no conservancy could be found, while those with letters signify areas where a homological amino acid could be identified. From here the regions with sequences with that were 11 amino acids in length were identified and classified as candidate epitopes for further analysis (see figure 4).

The next step was to test the isolated epitopes for their major histocompatibility complex two (MHC-II) binding capacity. This is an important metric which factors into an epitope's ability to bind onto the complexes of an antigen presenting cell (APC), that helps to activate the adaptive immune system and the secretion of IFN-γ (Kak et al., 2018). The alleles that were used were based off an analysis of alleles with large population coverage: HLA-DBR1*01:01, HLA-DBR1*03:01, HLA-DBR1*04:01, HLA-DBR1*11:01, & HLA-DBR1*15:01 (Prakash et al., 2021). For each sequence analyzed, an average of the percentile rank was calculated and store for use in a final analysis. A low percentile rank indicates a high binding-affinity to the MHC-II, while a high binding-affinity indicates the opposite.

After the MHC-II binding analysis, the next step is to determine the ability of the epitopes to induce IFN-γ. The Support Vector Machine (SVM) was used and it was predicted by selecting IFN-γ versus non-IFN-γ. Furthermore, for each sequence analyzed the value associated with the capability to induce IFN-γ was recorded. A positive value indicates an ability to induce IFN-γ, while a negative value indicates the opposite. Those epitopes with a negative value were eliminated from further analysis, while the rest continued on to further analysis.

The next step is to determine the antigenicity of the isolated epitope. Using the VaxiJen 2.0 server, which is based on an auto and cross-covariance transformation of protein sequences into uniform vectors of major amino acid properties, the antigenicity of the epitopes was evaluated. Those sequences who failed to meet the 0.4 threshold value were eliminated from

further analysis as the sub-0.4 value indicates that the epitope would be a poor antigen. Those epitopes which met the threshold continued onto further analysis..

The penultimate step is to determine the allergenicity and toxicity of the final epitopes. Using the AllerTOP v.2.0 and ToxinPred servers respectively, the remaining epitopes were analyzed to determine whether they are allergenic or toxic. The default settings of both servers were used. Those epitopes that were determined to be allergenic or toxic were eliminated from further analysis, while the rest continued on to the final step.

The final step determined the top-3 epitopes that would be synthesized for *in-vitro* testing. From the epitopes that were not eliminated by aforementioned analysis, those that were 15-mer in length continued on to the final analysis as they are most likely to produce a strong immune response (Prakash et al., 2021). This elimination resulted in 9 final candidate epitopes. In order to determine the top-3 sequences an elimination sequence was conducted (see Figure 6).

For the quantifiable traits, MHC-II binding capacity, IFN-γ binding capacity, and antigenicity, each epitope was ranked with the most desirable being ranked one and the least desirable being ranked nine. If two values were the same, the average between the two next rankings were taken and given to both epitopes in contention. Finally, the average of the rankings were taken and the three epitopes with the lowest average were determined to be the top-3 epitopes. The epitopes highlighted in green are the top-3 epitopes, those highlighted in yellow

have an average ranking value close to the top-3, and those highlighted in red have an average ranking away from the top-3.

These top-3 epitopes were then synthesized, and are going to be tested *in-vitro*. A co-culture between a macrophage, the epitopes, and a helper T-cell will help to determine the ability of the epitopes to trigger an immune response. It is planned that an ELISpot assay will be run to determine if there is a statistically significant amount of IFN-γ released based on the interaction. This is an interaction that is aimed to mimic the human body's natural immune response.

Discussion

The results so far indicate that there is a possibility for the creation of a pan-hantavirus vaccine. The identification of significant regions of homology among the M gene indicates that there may be a significant chance that B-Cell and $CD8^+$ epitopes can also be isolated. Yet, it is not until the completion of *in-vitro* testing that we can get a clear picture of the scope of the data compiled until now.

If the *in-vitro* trials result in a statistically significant amount of IFN-γ secretion, then it is a good sign that pan-Hantavirus vaccination is possible. Otherwise, other forms of preparation must be taken in order to address the emerging public health threat in hantaviruses.

There is currently only a limited understanding of hantavirus virology and its interaction with human cells. Furthermore, the limited knowledge on the reassortant characteristics of hantavirus and our bodies' response to hantavirus infection means that it is essential to continue research into the functions of hantaviruses and the development of treatments and vaccines to help fight the virus.

As the ineffectiveness of the current approaches to hantavirus vaccination are well documented (Yi et al., 2018; Hooper et al., 2014), it is important to invest in research investigating other approaches to vaccination. The positive results of this study should result in further investment in research towards the development of an entire multi-epitope pan-Hantavirus vaccine.

Pursuing an effective solution to this threat is necessary considering the grave consequences that a hantavirus pandemic may have. With fatality rates of 40% and 15% HCPS and HFRS pose a public health threat (Liu et al., 2020), and any positive results towards resolving this threat should be pursued.

References

Arai, S., Ohdachi, S. D., Asakawa, M., Kang, H. J., Mocz, G., Arikawa, J., Okabe, N., & Yanagihara, R. (2008, October 21) Molecular phylogeny of a newfound hantavirus in the Japanese shrew mole (Urotrichus talpoides). *Proceedings of the National Academy of Sciences of the United States of America, 105*(42), 16296-16301. https://doi.org/10.1073/pnas.0808942105

Arai, S., Taniguchi, S., Aoki, K., Yoshikawa, Y., Kyuwa, S., Tanaka-Taya, K., Masangkay, J. S., Omatsu, T., Puentespina, R., Jr, Watanabe, S., Alviola, P., Alvarez, J., Eres, E., Cosico, E., Quibod, M., Morikawa S., Yanagihara, R., & Oishi, K. (2016, November) Molecular

phylogeny of a genetically divergent hantavirus harbored by the Geoffoy's rousette (Rousettus amplexicaudatus), a frugivorous bat species in the Philippines. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases, 45,* 26-32. https://doi.org/10.1016/j.meegid.2016.08.008

Bibi, S., Ullah, I., Zhu, B., Adnan, M., Liaqat, R., Kong, W. B., & Niu, S. (2021, January 13) In silico analysis of epitope-based vaccine candidate against tuberculosis using reverse vaccinology. *Scientific Reports, 11(1249)*. https://doi.org/10.1038/s41598-020-80899-6

de Bellocq, J. G., Těšíková, J., Meheretu, Y., Čížková, D., Bryjová, A., Leirs, H., & Bryja, J., (2016, November) Complete genome characterisation and phylogenetic position of Tigray hantavirus from the Ethiopian white-footed mouse, Stenocephalemys albipes. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases, 45*, 242-245.https://doi.org/10.1016/j.meegid.2016.09.009

Dhanda, S. K., Vir, P., & Raghava, G. P. S., (2013, December 5) Designing of interferon-gamma inducing MHC class-II binders. *Biology Direct, 8(30),* 1. https://doi.org/10.1186/1745-6150-8-30

Doytchinova, I. A., & Flower, D. R. (2007, January 5) VaxiJen: a server for prediction of protective antigens, tumor antigens and subunit vaccines. *BMC Bioinformatics, 8,* 4. https://doi.org/10.1186/1471-2105-8-4

Doytchinova, I. A., & Flower, D. R. (2007, January 15) Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. *Vaccine, 25(5),* 856-866. https://doi.org/10.1016/j.vaccine.2006.09.032

Doytchinova, I. A., & Flower, D. R. (2008, September) Bioinformatic Approach for Identifying Parasite and Fungal Candidate Subunit Vaccines. *The Open Vaccine Journal, 1(1),* 22-26. https://doi.org/10.2174/1875035400801010022

Dzagurova T. K., Witkowski, P. T., Tkachenko, E. A., Klempa, B., Morozov, V. G., Auste, B., Zavora, D. L., Lunicheva, L. V., Mutnih, E. S., & Kruger, D. H. (2012, January 1) Isolation of sochi virus from a fatal case of hantavirus disease with fulminant clinical course. *Clinical infectious diseases, 54(1),* e1–e4. https://doi.org/10.1093/cid/cir746

Fulhorst, C. F., Cajimat, M. N. B., Utrera, A., Milazzo, M. L., & Duno, G. M. (2004, September 1) Maporal virus, a hantavirus associated with the fulvous pygmy rice rat (Oligoryzomys

fulvescens) in western Venezuela. *Virus Research, 104(2),* 139-144. https://doi.org/10.1016/j.virusres.2004.03.009

Ge, X.-Y., Yang, W.-H., Pan, H., Zhou, J.-H., Han, X., Zhu, G.-J., Desmond, J. S., Daszak, P., Shi, Z.-L., Zhang, Y.-Z. (2016, February 16) Fugong virus, a novel hantavirus harbored by the small oriental vole (Eothenomys eleusis) in China, *Virology Journal, 13(27),* https://doi.org/10.1186/s12985-016-0483-9

Gu, S.H., Nicolas, V., Lalis, A., Sathirapongsasuti, N., and Yanagihara, R., (2013, August 27) Complete genome sequence and molecular phylogeny of a newfound hantavirus harbored by the Doucet's musk shrew (Crocidura douceti) in Guinea. *Infection, Genetics, and Evolution, 20*, 118-123. https://doi.org/10.1016/j.meegid.2013.08.016

Gu, S.H., Hejduk, J., Markowski, J., Markowski, M., Liberski, P. P. & Yanagihara, R. (2015, May 28) Whole-Genome Sequence of a Novel Hantavirus Isolated from the European Mole (Talpa europaea), *Genome announcements, 3(3),* e00508-15. https://doi.org/10.1128/genomeA.00508-15

Guo, W.-P., Lin, X.D., Wang,W., Tian, J.H., Cong, M.L., Zhange, H.L., Wang, M.R., Zhou, R.H., Wang, J.B., Li, M.H., Xu,J., Holmes, E.C., & Zhang, Y.Z. (2013, February 7) Phylogeny and origins of hantaviruses harbored by bats, insectivores, and rodents, *PLoS Pathogens, 9(2)*, e1003159. https://doi.org/10.1371/journal.ppat.1003159

Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., & Raghava, G. P. S., (2013, September 13) In Silico Approach for Predicting Toxicity of Peptides and Proteins. *PLoS One, 8(9)*, e73957. https://doi.org/10.1371/journal.pone.0073957

Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., & Raghava, G. P. S. (2014, December 11) Peptide Toxicity Prediction. *Computational Peptidology, 1268,* 143-157. https://doi.org/10.1007/978-1-4939-2285-7_7

Hangaragi, P. S. (2020, June 14) Hantavirus: An emerging global threat, *Asian Journal of Oral Health and Allied Sciences, 10(4),* 1-5. https://doi.org/10.25259/AJOHAS_6_2020

Hiltbrunner, M., & Heckel, G., (2020, July 11) Assessing Genome-Wide Diversity in European Hantaviruses through Sequence Capture from Natural Host Samples. *Viruses, 12(7),* 749. https://doi.org/10.3390/v12070749

Hooper, J. W., Moon, J. E., Paolino, K. M., Newcomer, R., McLain, D. E., Josleyn, M., Hannaman, D., & Schmaljohn, C. (2014, January 14) A Phase 1 clinical trial of Hantaan virus and Puumala virus M-segment DNA vaccines for haemorrhagic fever with renal syndrome delivered by intramuscular electroporation. *Clinical Microbiology and Infection, 20(5),* 110-117. https://doi.org/10.1111/1469-0691.12553

Johnson, A. M., Bowen, M. D., Ksiazek, T. G., Williams, R. J., Bryan, R. T., Mills, J. N., Peters, C. J., & Nichol, S. T., (1997, November 10) Laguna Negra virus associated with HPS in western Paraguay and Bolivia, *Virology, 238(1),* 115-127. https://doi.org/10.1006/viro.1997.8840

Kak, G., Raza, M., & Tiwari, B. K. (2018, May, 30) Interferon-gamma (IFN-γ): Exploring its implications in infectious diseases. *De Gruyter, 9(1),* 64-79. https://doi.org/10.1515/bmc-2018-0007

Kang, H. J., Bennett, S. N., Dizney, L., Sumibcay, L., Arai, S., Ruedas, L. A., Song, J.-W., & Yanagihara, R., (2009, May 25) Host switch during evolution of a genetically distinct hantavirus in the American shrew mole (Neurotrichus gibbsii), *Virology, 388(1)*, 8-14. https://doi.org/10.1016/j.virol.2009.03/019

Kang, H. J., Bennett, S. N., Hope, A. G., Cook, J. A., & Yanagihara, R., (2011, August) Shared ancestry between a newfound mole-borne hantavirus and hantavirus harbored by cricetid rodents, *Journal of Virology, 85(15)*, 7496-7503. https://doi.org/10.1128/JVI.02450-10

Kikuchi, F., Senoo, K., Arai, S., Tsuchiya, K., Són, N. T., Motokawa, M., Ranorosoa, M. C., Bawm, S., Lin, K. S., Suzuki, H., Unno, A., Nakata, K., Harada, M., Tanaka-Taya, K., Morikawa, S., Suzuki, M., Mizutani, T., & Yanagihara, R. (2021, July 12) Rodent-Borne Orthohantavirus in Vietnam, Madagascar and Japan, *Viruses, 13(7),* 1343. https://doi.org/10.3390/v13071343

Kim, J. -A., Kim, W. -K., No, J. S., Lee, S. -H., Lee, S. -K., Kim, J. H., Kho, J. H., Lee, D., Song, D. H., Gu, S. H., Jeong, S. T., Park, M. -S., Kim, H. -C., Klein, T. A., Song, J. -W. (2016, June 17) Genetic diversity and reassortment of Hantaan virus tripartite RNA genomes in nature, the Republic of Korea. *PLoS Neglected Tropical Diseases, 10(6),* 1-17. https://doi.org/10.1371/journal.pntd.0004650

Klempa, B., Witkowski, P. T., Popugaeva, E., Auste, B., Koivogui, L., Fichet-Calvet, E., Strecker, T., Meulen, J. T., & Krüger, D. H. (2012, April) Sangassou virus, the first

hantavirus isolate from Africa, displays genetic and functional properties distinct from those of other murinae-associated hantaviruses, *Journal of Virology, 86(7),* 3819-3827. https://doi.org/10.1128/JVI.05879-11

Laenen, L., Vergote, V., Kafetzopoulos, L. E., Wawina, T. B., Vassou, D., Cook, J.A., Hugo, J-P., Deboutte, W., Kang, H. J., Witkowski, P. T., Köppen-Rung, P., Krüger, D. H., Szemeš, T., Markowski, J., Hejduk, J., Kafetzopoulos, D., Van Ranst, M., Yanagihara, R., Klempa, B., & Maes, P. (2018, January 1) A Novel Hantavirus of the European Mole, Bruges Virus, Is Involved in Frequent Nova Virus Coinfections. *Genome Biology and Evolution, 10(1),* 45-55. https://doi.org/10.1093/gbe/evx268

Laenen, L., Vergote, V., Calsiher, C. H., Klempa, B., Klingström, J., Kuhm, J. H., & Maes, P. (2019, August 27). Hantaviridae: Current classification and future perspectives. *Viruses, 11(9),* 788. https://doi.org/10.3390/v11090788

Lee, S.-H., Kim, W.-K., Park, K., No, J. S., Lee, G. Y., Kim, H.-C., Klein, T.A., Min, M.-S., Lee, S.-J., Hwang, J., Park, M.-S., & Song, J.-W. (2020, April) Genetic diversity and phylogeography of *Jeju Orthohantavirus* (*Hantaviridae)* in the Republic of Korea. *Virology, 543,* 13-19. https://doi.org/10.1016/j.virol.2020.01.012

Lin, X., He, B., Xia, L. & Tu, C. (2014, July 4) Novel hantavirus identified in black-bearded tomb bats, China, *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases, 21,* 158-160. https://doi.org/10.1016/j.meegid.2015.01.018

Ling, S., Smura, T., Tamarit, D., Huitu, O., Voutilainen, L., Henttonen, H., Vaheri, A., Vapalahti, O., & Sironen, T., (2018, January) Evolution and postglacial colonization of Seewis hantavirus with Sorex araneus in Finland, *Intection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases, 57,* 88-97. https://doi.org/10.1016/j.meegid.2017.11.010

Liu, R., Ma, H., Shu, J., Zhang, Q., Han, M., Liu, Z., Jin, X., Zhang, F., & Wu, X. (2020, January 30) Vaccine and Therapeutics against Hantaviruses. *Frontiers in Microbiology, 10,* 2989. https://doi.org/10.3389/fmicb.2019.02989

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.-R. N., Potter, S. C., Finn, R. D., & Lopez, R. (2019, July 2) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research, 47(W1)*, W636-W641.

https://doi.org/10.1093/nar/gkz268

Mittler, E., Dieterle, M.E., Kleinfelter, L.M., Slough, M.M., Chandran, K., & Jangra, R.K. (2019, August 17) Hantavirus entry: Perspectives and recent advances. *Advances in Virus Research, 104(6),* 185-224. https://doi.org/10.1016/bs.aivir.2019.07.002

Montoya-Ruiz, C., Cajimat, M. N. B., Milazzo, M. L., Diaz, F. J., Rodas, J. D., Valbuena, G., & Fulhorts, C. F., (2015, July) Phylogenetic Relationship of Necoclí Virus to Other South American Hantaviruses (Bunyaviridae: Hantavirus), *Vector-Borne and Zoonotic Diseases, 15(7),* 438-445. https://doi.org/10.1089/vbz.2014.1739

Muyangwa, M., Martynova, E. V., Khaiboullina, S. F., Morzunov, S. P., & Rizvanov, A. A. (2015, November 27). Hantaviral Proteins: structure, functions, and role in Hantavirus Infection. *Frontiers in Microbiology, 6,* 1326. https://doi.org/10.3389/fmicb.2015.01326

Papadopoulos, J. S., & Agarwal, R., (2007, May 1) COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics, 23(9),* 1073-1079. https://doi.org/10.1093/bioinformatics/btm076

Park, K., Lee, S.-H., Kim, J., Lee, J., Budhathoka, S., Kim, Y.-J., Kim, Y.-S., Kim, H.-C., Klein, T. A., Kim, W.-K., & Song, J.-W. (2021, March 19) Multiplex PCR-Based Nanopore Sequencing and Epidemiological Surveillance of Hantaan orthohantavirus in Apodemus agrarius, Republic of Korea. *Viruses, 13,* 847. https://doi.org/10.3390/v13050847

Parrington, M. A., Lee, P.-W., & Kang, C. Y. (1991, August 1) Molecular Characterization of the Prospect Hill virus M RNA Segment: a Comparison with the M RNA Segments of Other Hantaviruses, *Journal of General Virology, 72(8)*, 1845-1854. https://doi.org/10.1099/0022-1317-72-8-1845

Paules, C. I., Marston, H. D., Eisinger, R. W., Baltimore, D., & Fauci, A. S. (2017, October 17) The pathway to a Universal Influenza Vaccine. *Immunity, 47(4),* 599-603. https://doi.org/10.1016/j.immuni.2017.09.007

Prakash, S., Srivastava, R., Coulon, P. G., Dhanushkodi, N. R., Chentoufi, A. A., Tifrea, D. F., Edwards, R. A., Figueroa, C. J., Schubl, S. D., Hsieh, L., Buchmeier, M. J., Bouzaine, M., Nesburn, A.B., Kuppermann, B.D., & BenMohamed, L. (2021, June 1) Genome-Wide B Cell, CD4+, and CD8+ T Cell Epitopes That Are Highly Conserved between Human and Animal Coronaviruses, Identified from SARS-CoV-2 as Targets for

Preemptive Pan-Coronavirus Vaccines. *The Journal of Immunology, 206(11),* 2566-2582. https://doi.org/10.4049/jimmunol.2001438

Protein [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information ; [1988] - [cited 2022 Feb 28]. Available from: https://www.ncbi.nlm.nih.gov/protein/ADI78324

Protein [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information ; [1988] - [cited 2022 Feb 28]. Available from: https://www.ncbi.nlm.nih.gov/protein/BAK08520

Protein [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information ; [1988] - [cited 2022 Feb 28]. Available from: https://www.ncbi.nlm.nih.gov/protein/YP_009362100

Protein [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information ; [1988] - [cited 2022 Feb 28]. Available from: https://www.ncbi.nlm.nih.gov/protein/YP_009506070

Protein [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information ; [1988] - [cited 2022 Feb 28]. Available from: https://www.ncbi.nlm.nih.gov/protein/YP_009506271

Protein [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information ; [1988] - [cited 2022 Feb 28]. Available from: https://www.ncbi.nlm.nih.gov/protein/QDI78324

Protein [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information ; [1988] - [cited 2022 Feb 28]. Available from: https://www.ncbi.nlm.nih.gov/protein/QXT50474

Protein [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information ; [1988] - [cited 2022 Feb 28]. Available from: https://www.ncbi.nlm.nih.gov/protein/QZR93723

Radosa, L., Schlegel, M., Gebauer, P., Ansorge, H., Heroldova, M., Janova, E., Stanko, M., Mosansky, L., Fricova, J., Pejcoch, M., Suchomel, J., Purchart, L., Groschup, M.H., Kruger, D.H., Ulrich, R.G., & Klempa, B. (2013, April 16) Detection of shrew-borne

hantavirus in Eurasian pygmy shrew (Sorex minutus) in Central Europe. I*nfection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases, 19*, 403-410. https:///doi.org/10.1016/j.meegid.2013.04.008

Roper, R.L. (2014, October, 14) Antigen Presentation Assays to Investigate Uncharacterized Immunoregulatory Genes, Methods in Molecular Biology, 890, 259-271. https://doi.org/10.1007/978-1-61779-876-4_15

Shi, M., Lin, X.-D., Chen, X., Tian, J.-H., Chen, L.-J., Li, K., Wang, W., Eden, J.-S., Shen, J.-J., Liu, L., Holmes, E. C., & Zhang, Y.-Z. (2018, April) The evolutionary history of vertebrate RNA viruses, Nature, 556(7700), 197-202. https://doi.org/10.1038/s41586-018-0012-7

Song, J.-W., Kang, H. J., Song, K.-J., Truong, T. T., Bennett, S. N., Arai, S., Truong, N. U., Yanagihara, R. (2007, November 13) Newfound hantavirus in Chinese mole shrew, Vietnam. Emerging infectious diseases, 13(11), 1784-1787. https://doi.org/10.3201/eid1311.070492

Torrez-Martinez, N., Song, W., & Hjelle, B. (1995, August 1) Nucleotide sequence analysis of the M genomic segment of El Moro Canyon hantavirus: antigenic distinction from four corners hantavirus. Virology 211(1), 336-338. https://doi.org/10.1006/viro.1995.1413

Venkatarajan, M. S., & Braun, W. (2001, December 8) New quantitative descriptors of amino acids based on multidimensional scaling of a large number of physical-chemical properties. Molecular modeling annual, 7, 445-453. https://doi.org/10.1007/s00894-001-0058-5

Vulin, J., Murri, S., Madriéres, S., Galan, M., Tatard, C., Piry, S., Vaccari, G., D'Agostino, C., Charbonnel, N., Castel, G., & Marianneau, P. (2021, March 16) Isolation and Genetic Characterization of Puumala Orthohantavirus Strains from France, Pathogens, 10(3), 349. https://doi.org/10.3390/pathogens10030349

Wang, P., Sidney, J., Dow, C., Mothé, B., Sette, A., & Peters, B., (2008, April 4) A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLos Computational Biology, 4(4), e10000048. https://doi.org/10.1371/journal.pcbi.10000048

Wang, P., Sidney, J., Kim, Y., Sette, A., Lund, O., Nielsen, M., & Peters, B., (2010, November

22) Peptide binding prediction for HLA DR, DP and DQ molecules. BMC Bioinformatics, 22(11), 568. https://doi.org/10.1186/1471-2015-11-568

Wang, C.-Q., Gao, J.-H., Li, M., Guo, W.-P., Lu, M.-Q., Wang, W., Hu, M.-X., Li, M.-H., Yang, J., Liang, H.-J., Tian, X.-F., Holmes, E. C., & Zhang, Y.-Z., (2014, October 13) Co-circulation of Hantaan, Kenkeme, and Khabarovsk Hantaviruses in Bolshoy Ussuriysky Island, China. Virus Research, 191, 51-58. https://doi.org/10.1016/j.virusres.2014.07.021

Wold, S., Jonsson, J., Sjörström, M., Sandberg, M., & Rännar, S., (1993, May 28) DNA and peptide sequences and chemical processes multivariate modeled by principal component analysis and partial least-squares projections to latent structures. Analytica Chimica Acta, 277(2), 239-253. https://doi.org/10.1016/0003-2670(93)80437-P

Yadav, P. D., Vincent, M. J., & Nichol, S. T., (2007, August 21) Thottapalayam virus is genetically distant to the rodent-borne hantaviruses, consistent with its isolation from the Asian house shrew (Suncus murinus), Virology Journal, 4, 80. https://doi.org/10.1186/1743-422X-4-80

Yi, Y., Park, H., & Jung, J. (2018, December 31) Effectiveness of inactivated hantavirus vaccine on the disease severity of hemorrhagic fever with renal syndrome, Kidney Research and Clinical Practice, 37(4), 366-372. https://doi.org/10.23876/j.krcp.18.0044

Zhang, Y., Yuan, J., Yang, X., Zhou, J., Yang, W., Peng, C., Zhang, H.-L., & Shi, Z., (2011, June 1) A novel hantavirus detected in Yunnan red-backed vole (Eothenomys miletus) in China. Journal of General Virology, 92(6), 1454-1457. https://doi.org/10.1099/vir.0.030122-0

Zou, Y., Wang, J.-B., Gaowa, H-S., Yao, L.-S., Hu, G.-W., Li, M.-H., Chen, H.-X., Plyusin, A., Shao, R., Zhang, Y-Z., (2008, February 22) Isolation and genetic characterization of hantaviruses carried by Microtus voles in China, Journal of Medical Virology, 80(4), 680-688. https://doi.org/10.1002/jmv.21119

Appendices

Figure 1: Compilation of GenBank Sequences Used

Protein Sequences				
Sub-Family	**Genus**	**Gene_ID**	**Protein_ID**	**Citation**
actinovirus	Batfish	MG599944	AVM87657	(Shi et. al, 2018)
actinovirus	Goosefish	MG599947	AVM87660	(Shi et. al, 2018)
actinovirus	Spikefish	MG599950	AVM87663	(Shi et. al, 2018)
agnathovirus	Hagfish	MG599953	AVM87666	(Shi et. al, 2018)
loanvirus	Longquan	NC_043127	YP_009664870	(Guo et. al, 2013)
mobatvirus	Laibin	KM102248	AJZ68871	(Lin et. al, 2014)
mobatvirus	Nova	NC_034470	YP_009362037	(Gu et. al, 2015)
mobatvirus	Quezon	NC_034393	YP_009361839	(Arai et. al, 2016)
orthohantavirus	Andes	MN095817.1	QDI78324	(Tan et. al, 2019)
orthohantavirus	Asama	NC_038274	ACI28500	(Arai et. al, 2008)
orthohantavirus	Asikkala	NC_043069	AGK36761	(Radosa et. al, 2013)
orthohantavirus	Bayou	GQ244521	ADE10201	(Richter et. al, 2009)
orthohantavirus	Black Creek Canal	NC_043073		
orthohantavirus	Bowe	NC_034406	YP_009361852	(Gu et. al, 2013)
orthohantavirus	Bruges	MF683845	AUI44484	(Laenen et. al, 2018)
orthohantavirus	Cano Delgadito	NC_034525	YP_009362100	(Milazzo et. al, 2007)
orthohantavirus	Cao Bang	NC_034474	YP_009362042	(Song et. al, 2007)
orthohantavirus	Choclo	NC_038374	YP_009506070	(Nelson et. al, 2005)
orthohantavirus	Dabieshan	NC_038383	YP_009506271	(Wang et. al, 2011)
orthohantavirus	Dobrava-Belgrade	JF920149	AES92928	(Dzagurova et. al, 2012)
orthohantavirus	El Moro Canyon	NC_038424	YP_009506355	(Torrez-Martinez et. al, 1995)
orthohantavirus	Fugong	NC_034466	YP_009362033	(Ge et. al, 2016)
orthohantavirus	Fusong	NC_038447	YP_009506412	(Zou et. al, 2008)
orthohantavirus	Hantaan	MW796169	QUW04913	(Park et. al, 2021)
orthohantavirus	Jeju	MN295686	QFX79483	(Lee et. al, 2020)
orthohantavirus	Kenkeme	NC_034565	YP_009362291	(Wang et. al, 2014)
orthohantavirus	Khabarovsk	NC_034518	YP_009362096	(Wang et. al, 2014)
orthohantavirus	Laguna Negra	NC_038506	YP_009506658	(Johnson et. al, 1997)
orthohantavirus	Luxi	NC_038528	YP_009507330	(Zhang et. al, 2011)
orthohantavirus	Maporal	NC_034552	YP_009362281	(Fulhorst et. al, 2004)
orthohantavirus	Montano	NC_034397	BAK08520	(Kariwa et. al, 2011)
orthohantavirus	Necocli	NC_043408	YP_009666010	(Montoya-Ruiz et. al, 2015)
orthohantavirus	Oxbow	NC_043176	YP_009665161	(Kang et. al, 2009)
orthohantavirus	Prospect Hill	NC_038940	YP_009508267	(Parrington et. al, 1991)
orthohantavirus	Puumala	MW148473	QTW01253	(Vulin et. al, 2021)
orthohantavirus	Rockport	NC_038694	YP_009507826	(Kang et. al, 2011)
orthohantavirus	Sangassour	NC_034516	YP_009362094	(Klempa et. al, 2012)
orthohantavirus	Seewis	KY651022	ATX68098	(Ling et. al, 2018)
orthohantavirus	Seoul	MZ343376	QXT50474	(Shepherd et. al, 2021)
orthohantavirus	Sin Nombre	MZ851474	QZR93723	(Goodfellow et. al, 2021)
orthohantavirus	Thailand	MZ343361	QZA58117	(Kikuchi et. al, 2021)
orthohantavirus	Tigray	KU934009	AOR07001	(de Bellocq et. al, 2016)
orthohantavirus	Tula	MT514290	QMU24953	(Hiltbrunner et. al, 2020)
orthohantavirus	Yakeshi	NC_038705	YP_009507845	(Guo et. al, 2013)
thottimvirus	Imjin	NC_034557	YP_009362285	(Lin et. al, 2014)
thottimvirus	Thottapalayam	NC_010708	YP_001911125	(Yadav et. al, 2007)
reptillovirus	Gecko	MG599941	AVM87654	(Shi et. al, 2018)

Figure 2: Multiple Sequence Alignment

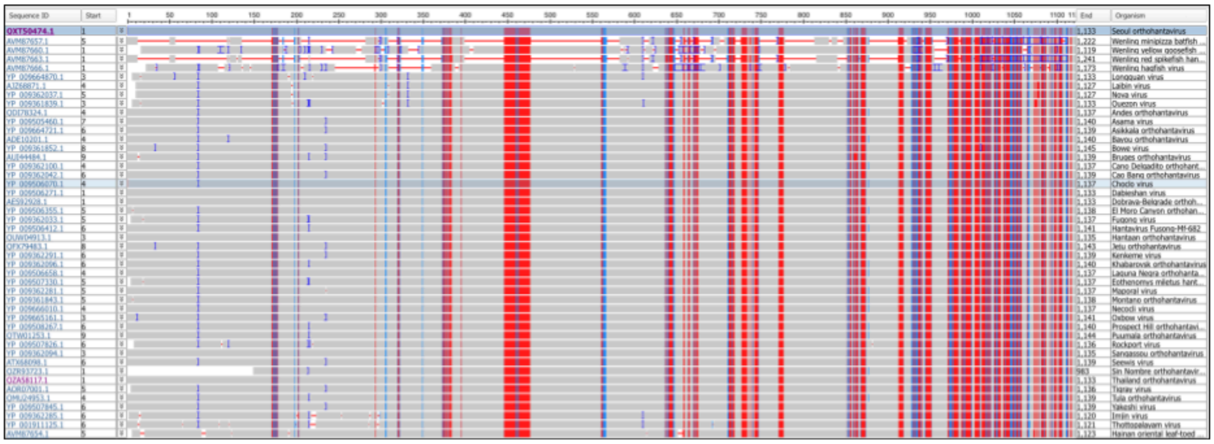

Figure 3: Conserved Sequence Constructed

```
xxxxxxxxxxxlxLLxVLxxvxxxxxxxxxRNVYELKLECPHTVxxxxGExxxxxxxxxVxGSVELPxIxLxEVxxxxxxSLK
xIESSCNFDIHxSxxxxQxFTQVtWxKKAdxxxTxNASSTTFExxSsEVNLKGxxxxxxxLxxxxxCVIxxxIIExxxKxx
xxRKTVICYDLSCNQTxxCKPTLHLIAPIxxxxxxxxxCxxMKSCLIxLGxxxxxxxxxxxxxxRIQVVYEKTYCVG
MLxVEGKxCFxPxxTLxxxxxxxxxxxxSxxxxxxxxxxxxxxxxxxYDvxxxxxxxxxxxxTLPVxCFLxxxIaKKxxxxxxx
xxxxxxxxxxxxxxxKIxExlEKIxxxxxkxxCTxxxxxENxxQGYYVCxIGxNSExIxVPSxDDxRSxExxIxxxxLSrMxxS
PHGEDHxxxxxxDxxxxxxxSLRIAGxxxxIExxxxxKVxPxTESSDxLxxxQGIAFSGxPMYSSLxxSVLxKxDPxxK
YVFSPGIIxxxxxPxxNxSxxxxxxxxxCDKKxLPLTWTGYxxIxIPGxxEKIxxxxxxxxxxxCTVFCTLSGPGASCEA
YSExxxxxxxGIFNISSPTCLVNKxxRFRbSEQQIxFVCQRxxxxxxxxVDxDIVVYCxxNGQKKVILTKTLVIGQC
IYTxTSLFSLLPxVAHSLAVELCVPGxHGWATIALLITFCFGWLLIPxITxIILKIxxxLKxIxxIxxxxxYNxESKFK
xILEKIKEEYQKTMGSMVCDVCxxxxxxxxxxxKHECETxKELKAHKKSCPQGQCPYCMxxxExTESALQAH
YKxVCKLTxxRFQEDLKKSIxxxxxxQxxGCxYRTLNIFRYKSRCYIxxVWIILLxIEsIIWAASAExxxxxxxxxxxx
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
xxxxxxxxxxxxxxxxxxxLEPxWNDxAxxxxxxxxxxxxxxxxxxxxxxxxHGVGxIPMxTDLELDFSLxxxxxxxxxxxxS
SSxYTYRRKLxNPxNEEExIpFxHIQIEKQVIxAEVQxLGHWMDAxxxxNIKTAFHCYGACxxKYSYPWQTAx
xxxxxxxCxFEKDYQYETxWGCNPxDxCPGVxxxxxxGTGCTACxGIYLDKLKxxxxxSVGxAYKIISLKYTRKV
CVQLGxExxCKxIDSNDxxxxxxxCLVTxNVKVCMIGTVSKFxxQxGDTLLFLGPLExGGLIxKQWCTTTCQF
GDPGDIMSxxNxGxxCPEYxGSFRKKCxFATTPVCEYxGNxVSGYKRMMATKDSFQSFNVTDxHxTxNKxxx
xxxxxxxxxxLEWxDxxxxxxxxxPDGxLRDHINIVVNxxxxxRDIxFxxxxxxxxxxxxEDLSENPCKVxLqTxSIEGAWG
SGVGFTLtCxVSLTECSxxTFLTSIKACDxxxAMCYGATSVTLVRGQNxTVxVxGKGGHSGSxxxxxkFKCCHD
xDCSxxGLLASAPxxxxxxHLDRVTGxNQIDNDKVYDDGAPeCGIxCWFxKSGEWLmGILxNGNWMVVVVL
IVILIISIILxSfLCPVRKxKKxxxxxxxxxxxxxxxxxxxx
```

Figure 4: 11+mer Sequence Classification

11+mer Sequence Classification	
Sequence ID	**Sequence**
A	CKPTLHLIAPI
B	CYRTLNLFRYK
C	FTQVtWKKKAd
D	GDTLLFLGPLE
E	GIYLDKLKSVG
F	DCPGVGTGCTAC
G	NIKTAFHCYGAC
H	QIDNDKVYDDGA
I	TESALQAHYKVC
J	VAHSLAVELCVPG
K	GIFNISSPTCLVNK
L	NVKVCMIGTVSKFQ
M	RKTVICYDLSCNQT
N	RNVYELKLECPHTV
O	SIEGAWGSGVGFTLtC
P	RIQVVYEKTYCVTGMLV
Q	VSLTECSTFLTSIKACD
R	WMVVVVLVVILIISIIL
S	AYKIISLKYTRKVCVQLG
T	KQWCTTTCQFGDPGDIMS
U	CTVFCTLSGPGASCEAYSE
V	KELKAHKKSCPQGQCPYCM
W	HGWATIALLITFCFGWLLIP
X	VSGYKRMMATKDSFQSFNVTD
Y	DIVVYCNGQKKVILTKTLVIGQCIYT
Z	ILEKIKEEYQKTMGSMVCDICKhECET

Figure 5: Elimination Sequence Table

Elimination Sequence								
	R23	WMVVVVLVVILIISI	15.36	Positive	0.5503569	0.7584	Non-Allergen	Non-Toxin
	R24	MVVVVLVVILIISII	14.44	Positive	0.70655736	0.5976	Non-Allergen	Non-Toxin
	R25	VVVVLVVILIISIIL	8.714	Positive	0.83196286	0.5386	Non-Allergen	Non-Toxin
	W35	HGWATIALLITFCFG	32.8	Positive	0.48341312	0.5922	Non-Allergen	Non-Toxin
	W36	GWATIALLITFCFGW	36.2	Positive	0.86239409	1.0371	Non-Allergen	Non-Toxin
	W37	WATIALLITFCFGWL	33.18	Positive	0.55717861	0.9842	Non-Allergen	Non-Toxin
	W38	ATIALLITFCFGWLL	33.18	Positive	0.55775663	0.7756	Non-Allergen	Non-Toxin
	W39	TIALLITFCFGWLLI	37.78	Positive	0.74836939	0.8847	Non-Allergen	Non-Toxin
	W40	IALLITFCFGWLLIP	39.78	Positive	0.66385574	0.8248	Non-Allergen	Non-Toxin
	R23	WMVVVVLVVILIISI	3	8	6	5.666666667	7	
	R24	MVVVVLVVILIISII	2	4	7	4.333333333	3	
	R25	VVVVLVVILIISIIL	1	2	9	4	2	
	W35	HGWATIALLITFCFG	4	9	8	7	9	
	W36	GWATIALLITFCFGW	7	1	1	3	1	
	W37	WATIALLITFCFGWL	5.5	7	2	4.833333333	5	
	W38	ATIALLITFCFGWLL	5.5	6	5	5.5	6	
	W39	TIALLITFCFGWLLI	8	3	3	4.666666667	4	
	W40	IALLITFCFGWLLIP	9	5	4	6	8	