

Host factors associated with RNA virus pathogenesis: investigating the role of Tetherin in viral replication

Ph.D. Thesis

By

RITUDHWAJ TIWARI



**DISCIPLINE OF BIOSCIENCE AND BIOMEDICAL
ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY INDORE
MARCH, 2021**

Host factors associated with RNA virus pathogenesis: investigating the role of Tetherin in viral replication.

*Submitted in partial fulfillment of the
requirements for the award of the degree*

of
DOCTOR OF PHILOSOPHY

By

RITUDHWAJ TIWARI



**DISCIPLINE OF BIOSCIENCE AND BIOMEDICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY INDORE
MARCH, 2021**



INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled Host factors associated with RNA virus pathogenesis: investigating the role of Tetherin in viral replication in the partial fulfillment of the requirements for the award of the degree of DOCTOR OF PHILOSOPHY and submitted in the DEPARTMENT/SCHOOL OF Biosciences and Biomedical Engineering, Indian Institute of Technology Indore, is an authentic record of my own work carried out during the time period from July 2015 to march 2021 under the supervision of Dr. Debasis Nayak, Associate Professor, Indian Institute of Technology, Indore.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

Ritudwhaj Tiwari
06/01/2022

Signature of the student with date

(RITUDWHAJ TIWARI)

This is to certify that the above statement made by the candidate is correct to the best of my/our knowledge.

Debasis Nayak

(January 05, 2022)

Signature of Thesis Supervisor #1 with date

Signature of

Thesis Supervisor #2 with date

(Dr. DEBAIS NAYAK)
THESIS SUPERVISOR)

(NAME OF

Ritudhwaj Tiwari has successfully given his/her Ph.D. Oral Examination held on 5
January 2022

Debasu Nayan

(January 05, 2022)

Signature of Thesis Supervisor #1 with date
of Thesis Supervisor #2 with date

Signature

(Dr. DEBAIS NAYAK)
(NAME OF THESIS SUPERVISOR)

ACKNOWLEDGMENTS

“Fall seven times and stand up eight”

This is something that my Ph.D. journey has taught me. I am grateful to the God for the good health and strength that were necessary to complete this thesis work.

My special and hearty thanks to my supervisor, Dr. Debasis Nayak, who encouraged and directed me during my Ph.D. work. Without his leadership and motivation, this research work could not have been completed. On the academic level, Dr. Nayak encouraged me to learn the fundamentals of conducting scientific research. Under his supervision, I got the opportunity to learn how to define a research problem, organize the literature search and data, planning of the experiments and time management to find a solution, and finally, translating the benchwork into the publication.

On a personal level, I feel Sir, as a parent. As I remembered my first meeting with Sir, he told, Ritu come with me I want to talk to you, and he said, "Science is all about passion; if you want to earn money, you can go and do MBA, but research is just like a marriage. You need to be faithful with your work, and you will enjoy the journey". I have been trying to follow those words throughout these years. In the last five years, Sir always stands with me even in the harsh condition of life. There are many memories where we have done the party, at his home, the delicious food cooked by Sir and many things. The most significant learning from my association with him was simplicity, dedication, and happiness that will be the key mantra for the rest of my life. I was fortunate enough to work under the guidance of such a person, who not only makes you a good researcher but also helps you to be a good human being. I will always be grateful to him and his family for constant support and encouragement during my stay in IIT Indore.

I also place on record, my sense of gratitude to the members of the PSPC committee, Dr. Amit Kumar and Dr. S. Vasudevan for their insightful comments and for sharing their tremendous experience with me. I am thankful to Dr. Debasis Nayak (Head, Discipline of Biosciences & Biomedical Engineering) for his

continuous support in various aspects. I wish to express my sincere thanks to all faculty members of BSBE discipline for all their suggestions and support. I would like to acknowledge all the teachers I learned from since my childhood; I would not have been here without their guidance, blessings, and support. I am also thankful to all Non-teaching staff of the BSBE family, Mr. Arif Patel, Mr. Murphy Bhaskar Ganveer, Mr. Gaurav Singh. I am grateful to the collaborators for lending me their expertise to my scientific and technical problems: Dr. Ujjwal Neogi, Karolinska Institutet, Stockholm, Sweden, Dr. Anirban Basu, NBRC, Delhi, Dr. Siddappa N. Byrareddy, , University of Nebraska Medical Center, Omaha, NE, USA, Prof. Asit K. Pattnaik University of Nebraska-Lincoln, Dr. Sandeep Danda St. Jude Children Research Hospital, Memphis, Tennessee, USA.

I am very grateful to Prof. Neelesh Kumar Jain, Director, Indian Institute of Technology, Indore, for his continuous encouragement, support, and the research facilities within an amiable working environment. I also wish to express a very warm thanks to Prof. Neelesh Kumar Jain, Dean of Academic Affairs, Indian Institute of Technology, Indore, for his excellent support, motivation, and encouragement throughout my Ph.D. tenure.

I thank my fellow labmates, Anurag Mishra, Basant Pravas Sahu, Prativa Majee, Shruti Pyasi, Suman Bishnoi, Venkat Are Narayana, Saumya Jaiswal, Sanchit Neema, Sheeba Rahman, for their constant support and motivation. My immense love and care to all my junior labmates for always being there and bearing with me in the good and bad during my wonderful days of Ph.D. I would also like to thank all my seniors and juniors from Discipline of Biosciences and Biomedical Engineering, IIT Indore, for all the love that they have showered upon me. I could not have asked for more than what I got from them.

I would also like to thank my friends Dr. Subodh Mishra, Dr. Eshan Khan, Dr. Anshu Mishra, Uma Shankar, Neha Jain, Dr. Omprakash, Dr. Neha Bahrill, Dr. Amit Singh and Mr. Jai Prakash who were always there and never let me down during my Ph.D. research work.

Lastly, I would like to thank my family for all their care, love, affection and encouragement. I sincerely express my love and gratitude to my father Mr. Shrinivas Tiwari and my mother Mrs. Shakuntala Tiwari for their unconditional love, trust, timely encouragement, and endless patience which always gives me the motivation to stand up again when I got weary. I would also like to thanks my sister Mrs. Arti Dubey and her husband Mr. Deepak Dubey.

I would like to express my sincere thanks to IIT Indore for infrastructure and University Grant Commission (UGC), India, for my fellowship.

RITUDHWAJ TIWARI

***DEDICATED
TO
MY PARENTS***

SYNOPSIS

A virus is a small biological substance, regarded as an obligate intracellular parasite. The virus replicates only within the living cells of plants, animals, and bacteria. In contrast, a virus particle composed of only two macromolecules, nucleic acids (DNA or RNA) as genetic material, and a protein coat that encapsulates this genetic material.[1] [2] Viral infections cause millions of deaths each year and are a primary health concern globally. Many researchers delve into understanding how the virus interacts with its hosts and strategizing appropriate prevention and mitigation measures to address the problems. While many successful antiviral drugs targeting viral proteins are often used, their administration is usually limited to specific viral species or strains. Baltimore classified the virus into seven branches based on their genetic makeup.[2] Compared to DNA viruses, the RNA viruses have low-fidelity polymerase enzymes and therefore prone to rapid mutations, leading to drug-resistant strains. Additionally, viruses encode very few proteins, thus limiting the number of available targets for drug discovery [3].

Hence to discover new antivirals, it's essential to know about host-virus pathogenesis, including how the virus enters the host cells, which proteins are involved in virus entry, replication, translation, assembly, virus antagonizes the host immune responses, etc. In recent years, genome-wide high-throughput RNA interference (RNAi) screening and CRISPR-Cas ribonuclease-based technologies have accelerated the search for such factors involved in virus-host interactions.[4-7] As a result, we now have detailed knowledge of host factors involved in virus life cycles and their involvement mechanisms.

When a virus infects its host, two kinds of host factors get activated. One category associated with viral lifecycle is called essential host factors, and the second types put a restriction on a viral infection called restriction host factors. Both kinds of host factors information are necessary to understand viral pathogenesis. Moreover, this virus and host interactions have resulted in various evolutionary

outcomes. During this deep-rooted common history, viruses and hosts have always put stress on each other for survival. Over time, hosts have developed the first defense mechanisms against viruses, called restriction factors. Restriction factors are host proteins that are part of the innate immune system and are potentially inhibit pathogen infection.[8] These restriction proteins target the various stages of the viral life cycle. Some factors directly interact with the virus or viral protein, while others indirectly inhibit virus infection by stopping macromolecule synthesis, apoptosis, etc. But with the evolution, viruses have also evolved antagonism mechanisms against these restriction factors. These antagonistic connections between the host factors and the virus proteins induced an evolutionary genetic conflict between the two entities.[9] This genetic contest is also called the "virus-host arms race," based on the Red Queen hypothesis, where organisms continuously develop and adapt to survive with changing environments.[10] Moreover, host factors information is essential to understanding viral pathogenesis and discovering new antiviral drugs.

The current thesis work highlights the host factors' information associated with the severe acute respiratory syndrome coronavirus-2 (SARS-COV-2) and Chandipura virus (CHPV). Our in silico and in vitro study shows how the complement system regulates the thrombosis process in COVID19 patients. In the same line of research, we have found that SARS-CoV-2 induces degradation of CD4 cells.

In respect of the CHPV, we have come with a novel viral entry mechanism. Interestingly, we discovered that a host restriction factor, Bone marrow stromal cell antigen 2 (BST-2), helps in virus entry. Further, we learned that the N terminal of BST-2 is essential for virus entry. Additionally, we have developed a virus-host factor interaction database called VHFIDB (WWW.VHFIDB.COM). It contains detailed information on 72 viral species and their respective host factors.

The thesis work includes seven chapters that are discussed below in detail. We aimed to define the importance of the host factors in the contest of the RNA viral infection.

Chapter 1: Review of literature for a few important viral diseases, associated host factors, virus-host evolution, and BST-2.

This chapter describes a detailed review regarding the viruses, their associated disease, mode of transmission, associated host factors, and their possible roles in virus restriction and virus pathology. This portion also elaborates on the part of host factors in virus evolution and vice versa. It further explains the specific host factor BST-2, its structure, topology, and role in viral pathogenesis.

Chapter 2: Materials, methodology, and techniques involved in the research study.

This chapter describes the materials, methodology, and techniques involved in conducting the research studies. Additionally, this chapter briefs about the basic principle of various techniques and tools used to predict viral protein structure, virus-similar host protein identification, identification of weighted proteins in associated pathways, etc. It also describes the standard assays, such as Western blot, RT-PCR, plaque assay, TCID-50, and cutting-edge gene-editing methods like CRISPER/cas9. Chapter 2 also describes the pseudovirus generation method for studied viruses such as CHPV, vesicular stomatitis virus (VSV), and murine leukemia virus (MuLV).

Chapter 3: VHFIDB: The Virus-Host Factor Interaction Database

This chapter describes the VHFIDB database. Here we have curated host factor information from peer-reviewed research articles and made an open online resource for the VHF database. The Virus-Host Factor Interaction Database (VHFIDB, <https://vhfidb.com/index.php>) is a freely available online platform, where each entry is stemmed from the published work of peer researchers. With broader coverage of virus families and incorporation of analytical tools, the VHFIDB provides a unique user-friendly enhanced learning experience compared to the other published databases, such as EHFPI and vhfRNAi. For example, the EHFPI and vhfRNAi databases comprise information only from RNAi screen-

based research articles. Simultaneously, the VHFIDB collects data from RNAi screen, CRISPR Cas9, or supported by other essential molecular techniques.

Additionally, the VHFIDB contains information on 72 virus species and 9,921 host genes. It covers some highly contagious viruses like Enterovirus A 71 (EVA-71), severe acute respiratory syndrome coronavirus (SARS-CoV), SARS-CoV-2, Zaire ebolavirus (ZEBOV), Marburg virus (MARV), etc. We have also compiled the emerging and re-emerging viruses in the current context. Information on contagious animal viruses like porcine reproductive and respiratory syndrome virus (PRRSV), equine arteritis virus, etc. is also available. Most importantly, the VHFIDB is also associated with three powerful analytical tools: VHF pathogen network, VHF overlap analysis, and gene enrichment analysis, which increases the usability of VHFIDB. We believe this database is a valuable resource tool for investigators in virology, biomedical sciences, and those involved in discovering and developing antiviral therapeutics.

Chapter 4: Bone Marrow Stromal Cell Antigen 2 (BST-2) enhances CHPV entry in BHK cells.

This chapter describes the role of BST-2 in CHPV pathogenesis and associated novel entry mechanism. BST-2 or Tetherin is a cell surface protein associated with lipid rafts whose expression is induced by type I Interferon (IFN-I) signaling. Induced expression of BST-2 blocks the release of a diverse group of enveloped viruses from infected cells' surface. These include many human pathogens such as HIV-I, JEV, Ebola virus, CHIKV, Dengue virus, etc. However, the action of BST-2 against the Chandipura virus is not yet studied. In our study, unexpectedly, we noticed that human and murine BST-2 overexpression significantly increases the infection of CHPV in a dose-dependent manner.

Further investigation shows that BST-2 enhances the virus entry mechanism. In - Silico study reveals that Tyrosine residues at 6 and 8 positions in the cytoplasmic domain of BST-2 interact with glu-405, glu-38, and thr-188 of CHPV glycoprotein via electrostatic interaction. Similarly, the threonine at position 4 of BST-2 cytoplasmic domain interacts with glu-399 and glu -402 of CHPV

glycoprotein. We have got similar results in in vitro studies. Moreover, we have also found that the virus used clathrin-mediated endocytosis for its entry.

Chapter 5: Structural similarity-based prediction of host factors associated with SARS-CoV-2 infection and pathogenesis

This chapter describes the host protein associated with the virus's life cycle and pathogenesis. This study involves implementing a computational method for predicting the interactions between SARS-CoV-2 and host proteins. This approach is based on protein structural similarity. At first, we determined the structural similarities between SARS-CoV-2 and human proteins using an established method by analyzing protein crystal structures. Further, we identified known interactions for these SARS-CoV-2 -similar human proteins. We assumed that these interacting proteins of SARS-CoV-2 - similar proteins would also interact with the SARS-CoV-2 proteins. We predicted and shortlisted an interaction map for SARS-CoV-2 and host proteins using cellular co-localization information. We then validated this interaction by using previously published data of host factors of coronaviruses and other RNA viruses.

Further analysis of these interacting proteins revealed that SARS-CoV-2 might use the clathrin-mediated endocytosis pathway for its entry. We also delineated the interplay of host proteins associated with the SARS-CoV-2 life cycle. We primarily focused on viral genome replication, translation, assembly, and predicted the pathways and the host factor requirements. Gene enrichment analysis of these interacting proteins reveals that apoptosis, IFN-gamma signaling, and CD4T cells proteasomal degradation pathways are positively associated with SARS-CoV-2 pathogenesis.

Chapter 6: In Silico and In vitro Studies Reveal Complement System Drives Coagulation Cascade in SARS-CoV-2 Pathogenesis

This chapter described the mechanism of SARS-CoV-2 associated pathogenesis. This study utilized *bioinformatics and computational biology* methods to predict the interactions between SARS-CoV-2 and the host proteins.

Further, we analyzed transcriptome and proteomics results of the SARS-CoV-2 infected Huh7 cell-line. Gene enrichment analysis of these interacting host proteins advocates that cytokine storm and neutrophil degranulation drive acute respiratory disease syndrome (ARDS) in SARS CoV-2 patients. The degree centrality analysis shows cytokine-associated pathways are regulated by key players such as TP53, TNF, MAPK3, and MAPK1 proteins. Similarly, degree centrality analysis of neutrophils degranulation genes highlighted VAMP8, ITGM, and STOM to be the highly weighted proteins in this pathway. We also discovered that to counter IFN-I, a few SARS-CoV-2 proteins may interact with signaling molecules like MAVS, IRF3, TRIM21, TRAF6, and IRK1 and thus could inhibit their functions. On further study, we have found that the Complement and Coagulation cascade is most strikingly interconnected and potentially a key driver of the innate immune response against SARS-CoV-2. This prediction was further validated by the KEGG pathway in gene enrichment analysis of combined transcriptomics and proteomics data sets resulted from *in vitro* SARS-CoV-2 infections in cell lines. Thus, the current study helps us understand the SARS-CoV-2 infection's molecular mechanism and its role in pathogenesis driven by the combined activation of cytokine storm, neutrophil degranulation, and the complement system.

Chapter 7: Scope of the thesis work and the future perspective:

Viruses cause significant acute and chronic infectious diseases. They are major patron to the global burden of disease—however, very few vaccines and antiviral drugs are available for these emerging viruses. Recent studies denote how viruses evolve to hijack the host cellular pathway and evade innate immune responses by modulating essential host protein and signaling pathways. Using *in silico* and in-vitro methods, we have described the host factor information and critical pathways associated with the SARS-CoV-2 and CHPV virus pathogenesis. In the SARS-CoV-2 study case, we explain how a host complement system drives the thrombosis process in COVID19; we also explain and give host factors information associated with the virus entry, replication, assembly, release, and another associated pathway that regulates virus pathogenesis. The current thesis

work also describes the CHPV virus use the clathrin-mediated endocytosis pathway for entry, and at the same time, cellular restriction factor BST-2 helps in virus entry.

Further *in silico* and in vitro study reveals that the cytoplasmic region of BST-2 is essential for virus entry. Additionally, we have also developed a comprehensive database called VHFIDB., which contains information about virus classification, virus pathogenesis, associated virus host factors. Moreover, VHFIDB is also hitched with three robust analysis tools VHF pathogen network, VHF overlap analysis, and gene enrichment analysis. Conclusively, the study opens up a new dimension of host-pathogen interactions and provides new therapeutic perspectives for treating viral diseases.

Reference:

1. Abrescia, N. G., D. H. Bamford, J. M. Grimes and D. I. Stuart. (2012), Structure unifies the viral universe, *Annu Rev Biochem*, 81, 795-822 (10.1146/annurev-biochem-060910-095130)
2. Mahmoudabadi, G. and R. Phillips. (2018), A comprehensive and quantitative exploration of thousands of viral genomes, *eLife*, 7, e31955 (10.7554/eLife.31955)
3. Sanjuán, R. and P. Domingo-Calap. (2016), Mechanisms of viral mutation, *Cell Mol Life Sci*, 73, 4433-4448 (10.1007/s00018-016-2299-6)
4. Elbashir, S. M., J. Harborth, W. Lendeckel, A. Yalcin, K. Weber and T. Tuschl. (2001), Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells, *Nature*, 411, 494-498 (10.1038/35078107)
5. Panda, D., A. Das, P. X. Dinh, S. Subramaniam, D. Nayak, N. J. Barrows, J. L. Pearson, J. Thompson, D. L. Kelly, I. Ladunga *et al.* (2011), RNAi screening reveals requirement for host cell secretory pathway in infection by diverse families of negative-strand RNA viruses, *Proc Natl Acad Sci U S A*, 108, 19036-19041 (10.1073/pnas.1113643108)
6. Wiedenheft, B., S. H. Sternberg and J. A. Doudna. (2012), RNA-guided genetic silencing systems in bacteria and archaea, *Nature*, 482, 331-338 (10.1038/nature10886)
7. Bhaya, D., M. Davison and R. Barrangou. (2011), CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation, *Annu Rev Genet*, 45, 273-297 (10.1146/annurev-genet-110410-132430)
8. Duggal, N. K. and M. Emerman. (2012), Evolutionary conflicts between viruses and restriction factors shape immunity, *Nat Rev Immunol*, 12, 687-695 (10.1038/nri3295)
9. Daugherty, M. D. and H. S. Malik. (2012), Rules of engagement: molecular insights from host-virus arms races, *Annu Rev Genet*, 46, 677-700 (10.1146/annurev-genet-110711-155522)
10. Van Valen, L. (1973). Theory.

LIST OF PUBLICATIONS

(A) Publication from the Thesis

(a) Published

1. **Tiwari, R.**; de la Torre, J. C.; McGavern, D. B.; Nayak, D. (2019), Beyond Tethering the Viral Particles: Immunomodulatory Functions of Tetherin (BST-2). *DNA, Cell Biol*, 38, (11), 1170-1177. DOI: 10.1089/dna.2019.4777 [**Impact Factor: 3.3**]
2. **Tiwari, R.**, Mishra, A., Mikaeloff, F., Gupta, S., Mirazimi, A., Byrareddy, S., Ujjwal, N., and Nayak, D (2020). In silico and In vitro Studies Reveal Complement System Drives Coagulation Cascade in SARS-CoV-2 Pathogenesis. *Comp Str Bio J*. DOI: 10.1016/j.csbj.2020.11.005 (**Impact Factor: 6.018**).
3. **Tiwari R.**, Mishra A. R., Gupta A., and Nayak D. (2021) Structural similarity-based prediction of host factors associated with SARS-CoV-2 infection and pathogenesis, *J Biomol Struct Dyn*, Jan 28;1-12. doi: 10.1080/07391102.2021.1874532 (**Impact Factor: 3.22**).

(b) Under peer review

4. **Tiwari R.**, Shankar U., Mishra K. S., Majee P., Mishra A., Pyasi S., Kumar A., Pattnaik A., Nayak D. VHFIDB: The Virus-Host Factor Interaction Database, *Cell host and microbes*. (**Impact Factor: 15.923**)
5. **Tiwari R.**, Mishra A. R., Are N. V., and Nayak D. Interferon-induced host factor bone marrow stromal antigen-2 (BST-2) enhances Chandipura virus (CHPV) entry. (Under prepration)

(B) Publication apart from the Thesis

(a) Published

6. Bishnoi S, **Tiwari R**, Gupta S, Byrareddy S, Nayak D. (2018). Oncotargeting by Vesicular Stomatitis Virus (VSV): Advances in Cancer Therapy. *Viruses* 10:90. DOI: 10.3390/v10020090 (**Impact Factor: 3.816**)
7. Das M, Kumar Kundu B, **Tiwari R**, Mandal P, Nayak D, Ganguly R, Mukhopadhyay S. 2018. Investigation on chemical protease, nuclease and catecholase activity of two copper complexes with flexidentate Schiff base ligands, *Inorganica Chim Acta* 469:111–122. DOI: 10.1016/j.ica.2017.09.013 (**Impact Factor: 2.046**)
8. Kundu, B.K., Mandal, P., Mukhopadhyay, B.G., **Tiwari, R.**, Nayak, D., Ganguly, R., Mukhopadhyay, S. (2019), Substituent dependent sensing behavior of Schiff base chemosensors in detecting Zn²⁺ and Al³⁺ ions: Drug sample analysis and living cell imaging, *Sensors and Actuators B: Chemical*, 282, 347-358.(DOI: 10.1016/j.snb.2018.11.076) (**Impact Factor: 7.1**)
9. Kundu, B.K., Singh, R., **Tiwari, R.**, Nayak, D., Mukhopadhyay, S. (2019), An amide probe as a selective Al³⁺ and Fe³⁺ sensor inside the HeLa and a549 cell lines: Pictet–Spengler reaction for the rapid detection of tryptophan amino acid, *New Journal of Chemistry*, 43, 4867-4877.(DOI: 10.1039/C9NJ00138G) (**Impact Factor: 3.288**)

(b) Under peer review

10. Sahu B.P., George B, Majee P., Singh . R.R., Mishra A., **Tiwari R**, Nayak D. comprehensive analysis of Simple Sequence Repeats in Picornaviruses, *Computational biology and chemistry* (Manuscript ID: CBAC_2020_1340). (**Impact Factor: 1.850**)
11. Venkat N. Are, Sandeep Kumar Dhanda, Sanchit Neema, **Ritudhwaj Tiwari**, and Debasis Nayak, Sequence homology and in-silico approach predicting candidate targets for immune responses to Kyasanur forest

disease virus (KFDV), Alkhumra hemorrhagic fever virus (ALKV), and Tick-borne encephalitis virus (TBEV), ***CSBJ* (Impact Factor: 6.08)**

TABLE OF CONTENTS

1. LIST OF FIGURES	xxv - xxxiii
2. LIST OF TABLES	xxxv
3. NOMENCLATURE	xxxvii
4. ACRONYMS	xxxix - xli
Chapter 1 – Introduction.....	1-45
1.1 Virus overview	1
1.1.1 History.....	1
1.1.2 Classification of viruses	2
1.1.3 Host-virus pathogenesis	6
1.1.4 Host virus interaction	8
1.1.5 Role of host factors in zoonosis	8
1.1.6 Coevolution of virus and host	12
1.2 Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2)	14
1.2.1 Classification lineage and history.....	14
1.2.2 Clinical impact	14
1.2.3 Genomic Structure and Replication cycle	15
1.2.4 Treatment and prevention.....	20
1.3 Chandipura virus (CHPV).....	20
1.3.1 Classification lineage and history.....	20
1.3.2 Clinical impact	21
1.3.3 Genomic Structure and Replication cycle	21
1.3.4 Treatment and prevention.....	22
1.4 Bone marrow stromal antigen 2 (BST-2)	24
1.4.1 History and molecular characterization.....	24
1.4.2 Antiviral activity of BST-2.....	26
1.4.3 BST-2 and cell signaling pathways	26
1.4.3 Immunomodulatory role of BST-2 following infection	27

1.4.3.1 Interferons & innate immunity	27
1.4.3.2 Adaptive Immunity	31
1.4.4 BST-2 and cancer	34
1.5 AIM and Scope.....	36
1.6. Reference:	37
Chapter 2 Material, methods, and instrumentation.....	49-77
2.1 Materials	49
2.1.1 Chemicals	49
2.1.2 Cloning vectors and competent cells	50
2.1.3 Cell lines.....	50
2.1.4 Enzymes, Cell culture media, Antibodies, and Kits	50
2.2 Methods and Instrumentation.....	51
2.2.1 E. coli competent cell preparation	51
2.2.2 Transformation of E. coli competent cells.....	51
2.2.3 Plasmid DNA isolation	53
2.2.4 Polymerase chain reaction	53
2.2.5 PCR purification.....	53
2.2.6 Restriction digestion	53
2.2.7 Agarose gel electrophoresis	54
2.2.8 Gel extraction.....	54
2.2.9 Ligation	54
2.2.10 cell culture	54
2.2.11 Virus propagation	54
2.2.12 CHPV Pseudovirus generation	55
2.2.13 Plaque assay	57
2.2.14 Construction of BST2 knock-in cells.....	59
2.2.14.1 Generation of MULV based retrovirus	59
2.2.14.2 Retroviral Transduction	59
2.2.15 Construction of BST2 knock-out cells	61
2.2.16 RNA isolation and cDNA synthesis	63
2.2.17 Quantitative real-time polymerase chain reaction assay	63

2.2.18 Protein sample preparation	66
2.2.19 SDS polyacrylamide gel electrophoresis (SDS-PAGE) and transfer to PVDF membrane	66
2.2.20 Western blot assay	67
2.2.21 Immunofluorescence microscopy	69
2.2.22 LC-MS analysis	71
2.2.23 TEM analysis	73
2.3 Bio-informatic tools and server used in the studies	75
2.3.1 Molecular Docking	75
2.3.2 Determination of structural similarities between virus and human proteins	75
2.3.2 Gene enrichment analysis	75
2.4 Reference	76
Chapter 3 VHFIDB: The Virus-Host Factor Interaction Database.....	81-113
3.1 Introduction	81
3.2 Results and discussion	85
3.2.1 Database statistics	85
3.2.2 Data retrieval	88
3.2.3 Search	88
3.2.4 Browse	90
3.2.5 Advanced search	92
3.2.6 Analysis	92
3.2.6.1 Overlap analysis	94
3.2.6.2 VHF pathogen network analysis	97
3.2.6.3 Gene enrichment analysis	99
3.2.7 VHFs as a potential source of antiviral drug information	101
3.3 Conclusions	110
3.4 Materials and Methods	110
3.4.1 Data source	110
3.4.2 Data organization	111

3.4.3 Data update	111
3.4.4 Data analysis	111
3.4.4.1 VHF Pathogen Network Analysis:.....	112
3.4.4.2 VHF Overlap Analysis	112
3.4.4.3 Gene Enrichment Analysis.....	112
3.5 References	113

Chapter 4 CHPV Pseudovirus generation and identification of virus entry mechanism.....117-145

4.1 Introduction	119
4.2. Results	121
4.2.1. Generation and characterization of CHPV Pseudovirus	121
4.2.2. Clathrin is essential for CHPV entry:	126
4.2.3. CHPV entry is pH-dependent	129
4.2.4. Lipid raft is essential for CHPV entry	132
4.2.5. BST-2 enhance CHPV entry	135
4.2.6. In-Silico study reveals that the cytoplasmic domain of BST-2 interacts with CHPV Glycoprotein.....	137
4.3. Discussion	139
4.4. Material method.....	142
4.4.1 Chemicals	142
4.4.2 Antibody	142
4.4.2 Cell lines.....	142
4.4.3 Viruses:	142
4.4.4 Pseudovirus production:	142
4.4.5 LC-MS analysis of Pseudovirus	143
4.4.6 Transmission electron microscopy Imaging	143
4.4.7 RNA extraction and RT PCR:.....	143
4.4.8 Drug inhibition of pseudovirus entry:.....	144
4.4.9 Statistical Analysis	144
4.5. Reference:	145

Chapter 5 Structural similarity-based prediction of host factors associated with SARS-CoV-2 infection and pathogenesis.....149-178

5.1 Introduction	151
5.2 Results	152
5.2.1 Identification of SARS-COV-2 host interactor proteins and gene enrichment analysis	152
5.2.2 SARS-CoV-2 use clathrin-mediated endocytosis for its cellular entry	155
5.2.3 Intracellular replication and assembly of SARS-CoV-2	156
5.2.3.1 Host proteins involved in SARS-CoV-2 translation	156
5.2.3.2 Host proteins involved in a double-membrane vesicle (DMV) formation during SARS-CoV-2 infection	159
5.2.3.3 Host proteins involved in SARS-CoV-2 replication	160
5.2.3.4 Host proteins involved in SARS-CoV-2 assembly and release	162
5.2.4 IRF1/9/7 suggested to have a regulatory role in IFN- γ mediated signaling pathway during SARS-CoV-2 infection	162
5.2.5 TP53 and CASP3 are the critical players of SARS-CoV-2 mediated apoptosis	166
5.2.6 SARS-CoV-2 protein-induced proteasomal degradation of CD4 T cell	169
5.3 Discussion	172
5.4 Material and method	173
5.4.1 Data sources	173
5.4.2 Determination of structural similarity between SARS-CoV-2 and human proteins	174
5.4.3 Interaction prediction	174
5.4.4 Cellular compartmentalization (CC) and GO enrichment analysis	174
5.4.5 Validation of Predictions	175
5.5 REFERENCE	176

Chapter 6 In silico and In vitro Studies Reveal Complement System Drives Coagulation Cascade in SARS-CoV-2 Pathogenesis.....181-215

6.1. Introduction	183
6.2. Results	184

6.2.1. Identification of SARS-CoV-2-similar human proteins (hSARS-CoV-2) and hSARS-CoV-2 host interaction	184
6.2.2. Gene enrichment analysis of SARS-CoV-2 interactor proteins	187
6.2.3. MAPK is predicted to be the key player in the innate immune response.....	190
6.2.4. TP53 and VAMP8 respectively are the key proteins of Cytokines storm and neutrophil degranulation process associated with acute respiratory distress syndrome (ARDS).....	193
6.2.5. The complement system-induced thrombosis in SARS CoV-2 patients.....	198
6.2.6. Genes associated with complement and coagulation pathways are upregulated in the SARS-CoV-2 infected Huh7 cells.....	203
6.3. Discussion	206
6.4. Materials and Methods	209
6.4.1 Data sources.....	209
6.4.2. Determination of structural similarities between SARS-CoV-2 and human proteins	209
6.4.3. Interaction prediction	210
6.4.4. Cellular compartmentalization (CC) and Gene enrichment analysis.....	210
6.4.5. Validation of Predictions.....	210
6.4.6. Complement and coagulation cascades pathway in SARS-CoV-2 infection model	211
6.5. References	213
Chapter 7 Conclusion and future prospective.....	219-222
7.1 Conclusion of the thesis.....	221
7.2 Future prospective.....	224
APPENDIX-A.....	227-231
APPENDIX-B	235-234
APPENDIX-C.....	237-251

List of Figures

Figure 1.1. Schematic diagram of Baltimore classification. Baltimore divided viruses into seven groups based on virus genomes and their replication mode. The intermediate steps involved in mRNA synthesis are also described in the diagram.....	5
Figure 1.2. Schematic diagram of virus infection and pathogenesis	7
Figure 1.3. Viral adaptation to the human and viral disease emergence. On the left side of the figure shows that most of the zoonosis occur from birds and animal to human, the middle part of the model shows a zoonosis pyramid which carries a concept that that animal viruses become frequently adapted to human through a series of evolutionary steps represented from bottom to top. The right part of the figure represents the status of immune system functionality against zoonosis.....	10
Figure 1.4. Role of Host protein in virus zoonosis process. Host carries two kinds of protein, proviral and antiviral. Whenever a virus infects the host, the innate immune system will be activated, and simultaneously proviral gens also activated. For proper virus adaptation and zoonosis, the virus is required to use most of the proviral host factors for their infection and replication, and at the same time, it is also needed to protect itself from restriction factors.....	11
Figure 1.5. Schematic representation of virus-host co-evolution. In this figure, we describe a host-virus arm-race, an antiviral protein antagonized by the virus, but human protein evolution allows the host protein to win this battle against the virus. but with time virus again evolved and overcome by the host protein effect. Moreover, most of the time, the virus wins the 'virus-host arms race.' because the virus has a small life span and can evolve faster than its hosts.....	13
Figure 1.6. Genomic organization and molecular structure of SARS-CoV-2.	17
Figure 1.7. A schematic diagram displaying all the significant steps of the SARS-CoV-2 life cycle	18
Figure 1.8. Genomic organization and molecular structure of CHPV	23

Figure 1.9. A schematic diagram displaying all the significant steps of the CHPV life cycle	23
Figure 1.10. Molecular characterization of BST-2. BST-2 is a type II transmembrane protein and contains ~180 amino acids (aa). Its cytoplasmic domain contains The mature protein comprising of a short N-terminal cytoplasmic domain (1-20aa) followed by an α -helical transmembrane domain (21-48aa), an ectodomain (49-161aa), and a C-terminal glycosylphosphatidylinositol (GPI) domain (162-180aa). The cytoplasmic tail of BST-2 has a highly conserved YXY tyrosine motif which plays a crucial role in NF-kB-mediated signaling and clathrin-mediated endocytosis.	25
Figure 1.11. BST-2 regulation of the IFN-I response.	29
Figure 1.12. Antiviral and immunomodulatory functions of BST-2.	32
Figure 1.13. The influence of BST-2 on tumor cell survival, invasion, and migration.	35
Figure 2.1. Schematic representation of transformation process.	52
Figure 2.2. Schematic representation of CHPV Pseudovirus generation	56
Figure 2.3. Schematic representation of Plaque assay	58
Figure 2.4. Schematic representation of Retrovirus generation and transduction	60
Figure 2.5. Schematic representation of CRISPER-Cas9.	62
Figure 2.6. Schematic representation of Real time PCR	65
Figure 2.7. Schematic representation of western blot.....	68
Figure 2.8. Schematic representation of Fluorescence Microscopy	70
Figure 2.9. Schematic representation of TMT-LC-MS	72
Figure 2.10. Schematic representation of negative stain Transmission Electron Microscopy.....	74
Figure 3.1. Overview of VHFIDB	84
Figure 3.2. Screenshot of the home page of VHFIDB.....	86
Figure 3.3. Bar graph representing the virus and host factors distribution in 25 different families—example of the number of host factors associated	

with the virus family. The total number of factors related to a virus family is presented in a bar graph. 87

Figure 3.4. Screenshot of options available on the home page (A) Provides search options and (B) the output results, including VHFIDB ID, host factor symbol, protein name, HF function, UniProt ID, Interacting Proteins, Virus name, Virus family, and reference PubMed ID..... 89

Figure 3.5. Screenshot of browse options. The figure shows host factors of different viruses based on the Baltimore system of virus classification. 91

Figure 3.6. Schematic representation of analysis tools. Three analysis tools of VHFIDB to facilitate the in-depth understanding of host factors for pathogenic infection. 93

Figure 3.7. Screenshot of the VHF overlap analysis tool and its corresponding results. (A) Users can select virus class, family, and species to do overlap analysis. Heatmap is exhibiting overlapping host factors in VHFIDB based on (B) Family, (C) Species, and (D) Class..... 96

Figure 3.8. Screenshot of VHF Pathogen Network Analysis and its corresponding result. (A) Users can select a virus name or put the host factor name to get overlap analysis; (B) selection of SARS-CoV-2 and network analysis of it as an example..... 98

Figure 3.9. Screenshot of gene enrichment analysis tool. (A) Users can select virus names or put the host factor name to get gene enrichment analysis, and (B, C) shows the analysis results in the form of an annotation summary and function annotation table. 100

*Figure 4.1. Schematic diagram of CHPV Pseudovirus production. At first, to generate VSV pseudovirus, we have replaced VSV Glycoprotein from TFP and use pCAG VSV-G for packing (G*ΔG-VSV). 70% confluent 100 mm dish of BHK-21 cells were transfected with N, P, L, G, *ΔG-VSV-TFP and generate G*ΔG-VSV-TFP, and used for further experiment. Next, for CHPV pseudoviruses construction, envelope genes of CHPV from Nagpur strain (strain No. 1653514) were cloned into the eukaryotic expression plasmid pCAG and transfected in*

*BHK21 cells, which is transcribed and expresses CHPV glycoprotein that is transferred to and attached to the cell membrane. Twenty-four hours later, transfected cells infected with G*ΔG-VSV-TFP with 1 MOI. one hour after infection, cells were washed with PBS twice, and then 2% DMEM was added. Twenty-four hours post-infection, CHPV pseudovirus carrying culture supernatants were collected, filtered, tested, and stored in –80°C..... 123*

Figure 4.2. LC-MS data of CHPV Pseudovirus.70% confluent 100 mm dish of BHK-21 cells were transfected with 5 µg of pCAG-CHPV-G by using Lipofectamine 2000. Thirty hours later, transfected cells infected with G*ΔG-VSV-TFP with 1 MOI. one hour after infection, cells were washed with PBS twice, and then 2% DMEM was added. Twenty-four hours post-infection, CHPV pseudovirus carrying culture supernatants were collected, filtered (0.45-µm pore size, Millipore, SLHP033RB). after filtration centrifuge it at one lakh RPM for 2 hours, and send it to V-proteomics a Delhi (India) based company for LC-MS. LC-MS data confirms CHPV glycoprotein. 124

Figure 4.3.The morphological characterization of CHPV Wt (left) and CHPV pseudovirus (right) CHPV pseudovirus generated as mentioned above. For the TEM imaging, samples were placed onto 400-mesh copper grids (SIGMA aldrich chemicals) and incubated for 5 minutes. The grids were then rinsed twice with DI water, wicked dry, and stained with 2% uranyl acetate for 5 min. CHPV virus looked bullet shape (left) while CHPV Pseudovirus looked circular, and its size was around 70-100 nm..... 125

Figure 4.4. Clathrin is essential for CHPV entry. Vero and Vero E6 cells were treated with 40 µM CPZ and .45 M sucrose. After 45 min, cells were infected with CHPV pseudovirus in the presence of a drug. Microscopic imaging was performed after 24 hours (A, D). Fluorescence cells counted by Image j software (B, E), CHPV pseudovirus TFP

infectivity was measured using a spectrophotometer at 24 h of infection(C, F). Error Bar represents the SD of three independent experiments. 128

Figure 4.5. CHPV entry is pH dependent.Vero and Veo E6 cells were treated with 25 mM NH₄Cl and 50 mM NH₄Cl. After 45 min, cells were infected with CHPV pseudovirus in the presence of a drug. Microscopic imaging was performed after 24 hours (A, D). Fluorescence cells counted by Image j software (B, E), CHPV pseudovirus TFP infectivity was measured using a spectrophotometer at 24 h of infection(C, F). Error Bar represents the SD of three independent experiments. 131

Figure 4.6. Lipid Raft is important for CHPV entry.Vero and Veo E6 cells were treated with 1 mM M β CD and 5 mM M β CD. After 45 min, cells were infected with CHPV pseudovirus in the presence of a drug. Microscopic imaging was performed after 24 hours (A, D). Fluorescence cells counted by Image j software (B, E), CHPV pseudovirus TFP infectivity was measured using a spectrophotometer at 24 h of infection(C, F). Error Bar represents the SD of three independent experiments..... 134

Figure 4.7. BST-2 enhance CHPV entry. BHK-21 cells were transfected with Murine and Human BST-2. After 32 hours of transfection, cells were infected with 10 MOI of CHPV (A, B, E, F) and UV inactivated CHPV (C, D, G, H) for 2h at 4 C (p > 0.05). then, cells were washed with citric acid buffer (pH 3), following PBS wash, viral RNA was extracted from BHK-21 cells and quantified by qPCR..... 136

Figure 4.8. Cytoplasmic domain of BST-2 interact with CHPV-G.We have performed molecular docking by using Autodock vina. The cytoplasmic chain of BST-2 was extracted from PDB: 4P6Z and performed docking with CHPV-G, PDB: 4DEW (A, B). The cytoplasmic domain of BST-2 directly interacts with CHPV-G and, Tyr6 and Tyr8 of BST-2 cytoplasmic domain interact with Glu-405,

Glu-38, Thr-188 of CHPV glycoprotein via electrostatic interaction, and Thr4 of BST-2 cytoplasmic domain interact with Glu-399 and Glu-402 of CHPV-G (C, D).	138
Figure 4.9. Proposed schematic representation of CHPV entry. CHPV binds with cells and enters into the cells by clathrin-mediated endocytosis. Lipid raft and BST-2 also play a crucial role in virus entry. After entry, first, it will move to early endosome than late endosome. After that, viral RNA is released in the cytoplasm at low pH.	141
Figure 5.1. Gene enrichment analysis of SARS-CoV-2 interactor human proteins. (A). Enriched biological pathways obtained from Reactome database (B). Cellular expression data of interacting protein obtained from HPA. Bonferroni corrected p-values were transformed by $-\log_{10}$. (The bar graphs were created by using GraphPad-Prism software)	154
Figure 5.2. Schematic representation of the SARS-CoV-2 life cycle and its associated host proteins. This cartoon depicts the endocytotic pathway for viral entry, followed by uncoating and releasing the viral genome to the host cell cytoplasm. Subsequently, viral RNA is associated with the replication and transcription complex (RTC) membrane and double-membrane vesicles (DMVs) during the viral genome replication and transcriptions.....	161
Figure 5.3. A predicted interaction map of the top 20 IFN- γ associated human proteins and their interacting SARS-CoV-2 protein. Blue color represents the virus proteins, and brown color represents the human interactor proteins. Twelve SARS-CoV-2 proteins interact with human proteins related to IFN- γ signaling pathway. IRF1, IRF7, and IRF9 (illustrated in green color) are highly weighted proteins and may regulate IFN- γ signaling cascade. (Interaction map was created using Cytoscape.)	164
Figure 5.4. A predicted interaction map of the top 20 apoptosis-associated human proteins and their interacting SARS-CoV-2 protein. Blue color	

represents the virus proteins, and brown color represents the human interactor proteins. Fourteen SARS-CoV-2 proteins interact with human proteins connected with apoptosis. Degree centrality analysis reveals that TP53, CASP3, and CTNNB1 (illustrated in green color) are the principal players in the SARS-CoV-2 mediated apoptosis. (Interaction map was created using Cytoscape.) 167

Figure 5.5. A predicted interaction map of the top 20 proteasomal degradations of CD4-associated human proteins and their interacting SARS-CoV-2 protein. Blue color represents the viral proteins, and brown color represents the human interactor proteins. Thirteen SARS-CoV-2 proteins interact with human proteins connected with proteasomal degradation of CD4 T cells. Degree centrality analysis reveals that UBA52, UBC, and PSMC5 (represented in green color) are the key player in the process of proteasomal degradation of CD4 T cells. (Interaction map was created using Cytoscape.) 170

Figure 6.1. Overlap with previous studies: (A) Figure represents the overlap proteins between primary interactors list and that of list published by Gordon et al. Note ~91 % of the list provided in Gordon et al. study overlap with the candidates predicted in our study. (B, C) Diagram represents the overlap proteins between CC and literature filtered protein with Gordon et al. proteins..... 186

Figure 6.2. GOST multi-query Manhattan plot and Degree centrality analysis of overlapping proteins in the profoundly enriched pathways. (a) g: GOST multi-query Manhattan plot shows significantly enriched GO: BP, KEGG terms, and Reactome enhanced data for SARS CoV-2 interacting proteins. (b) Network analysis based on the input pathways, showing overlapping proteins between 6 highly enriched pathways. Of these, the subnetwork shows MAPK1, MAPK3, AKT1, and SRC proteins are connected maximally. Abbreviations: M1: MAPK signaling pathway, M2: C-type lectin receptor signaling pathway, M3: Platelet activation, M4: PD-L1 expression, and PD-1

checkpoint pathway in cancer, M5: Innate Immune System, M6: Cytokine Signaling in Immune system. 189

Figure 6.3. Type I interferon induction and signaling during SARS-CoV-2 infection and virus-mediated inhibition of IFN-I and ISGs. The schematic diagram represents key players of INF-I pathways associated with SARS-CoV-2 infection. After entering the cells, the innate sensors such as MDA5, RIG1, and PAMS recognize the viral proteins and nucleic acids, which then activate IFN-I and pro-inflammatory cytokine production. Subsequently, INF-I enhances the production of ISGs and sets the stage of the potent antiviral immune response. At the same time, the SARS CoV-2 proteins interact with the MAVS, TRAF6, IRAK1 and negatively regulate IFN-I signaling and dampen host immune response. 191

Figure 6.4. Predicted interaction map of SARS-CoV-2 proteins and top 20 host proteins associated with cytokines signaling pathway and Neutrophils degranulation. Blue color represents the virus proteins, and brown color represents the human interactor proteins. (a) Sixteen SARS-CoV-2 proteins interact with human proteins associated with the cytokines pathway. TP53, TNF, and MAPK3 (illustrated in green color) are highly weighted proteins and may regulate cytokine production. (b) Thirteen SARS-CoV-2 proteins interact with human proteins associated with Neutrophils degranulation. Degree centrality analysis reveals that VAMP8, ITGAM, and STOM (represented in green color) are the key players in the neutrophils degranulation process. 197

Figure 6.5. Predicted interaction map of SARS-CoV-2 proteins and the top 20 host proteins associated with complement and coagulation cascade. Blue color represents the virus proteins, and brown color represents the human interactor proteins. Four SARS-CoV-2 proteins interact with human proteins associated with complementing and coagulation cascade. Most of the proteins interact with spike protein. Degree

centrality analysis reveals that KNG1, C3, and FGG (represented in green color) are the key player in the complement and coagulation cascade..... 200

Figure 6.6. Schematic representation of the mechanism of complement system-mediated thrombosis. SARS-CoV-2- Spike (S) protein interacts with C3 and activates proteolytic processing of C3, C3 hydrolyzes in C3a and C3b, further C3b cleaves C5 into C5a and C5b. C3a and C5a induce inflammatory cytokines production. Further, C3a, C5a, and C5b9 activate the coagulation pathway. C5a also activates IL6 and TNF alpha production, which also magnifies the coagulation pathway. 202

Figure 6.7. In vitro validation of the complement and coagulation cascade in Huh7cells. a) Heatmap of 59 genes associated with KEGG pathway “Complement and coagulation cascades” that were detected in the transcriptomics over the indicated time of infection. (b) Heatmap of 36 proteins associated with KEGG pathway “Complement and coagulation cascades” that were detected in the proteomics data over the indicated time of infection. Data were quantile normalized and Z-score transformed. Lower values are represented in yellow and higher values in red. (c) Cytoscape network of KEGG pathway “Complement and coagulation cascades” labeled with transcriptomics results. 59 transcripts were detected, and 44 have a differential expression over time (d) Cytoscape network of KEGG pathway “Complement and coagulation cascades” labeled with proteomics results..... 204

List of Tables

Table 1.1. List of non-structural proteins of SARS-CoV-2 and their function.....	19
Table 3.1. List of virus-host factors associated with the SARS-CoV-2 and their associated function. The DrugBank IDs and ChEMBL IDs represent the drugs against the host factor that can be used to treat the Coronaviruses.	102
Table 5.1. List of human protein kinase and their interactor SARS-CoV-2 proteins.	158
Table 5.2. Degree centrality analysis of proteins associated with IFN- γ signaling pathway.	165
Table 5.3. Degree centrality analysis of proteins associated with apoptosis pathway.	168
Table 5.4. Degree centrality analysis of proteins associated with proteasomal degradation of CD4 T cells	171
Table 6.1. List of significantly enriched KEGG terms and Reactome enhanced data. Compilation of 2,647 unique proteins entries resulting in highly enriched biomedical relevant KEGG term and the Reactome pathways listed above having p-adjusted p-values.....	188
Table 6.2. Degree centrality analysis of proteins associated with Cytokines storm.	194
Table 6.3. Degree centrality analysis of proteins associated with neutrophil degranulation.	196
Table 6.4. Degree centrality analysis of proteins associated with complement and coagulation cascade.	199

Nomenclature

+ve	Positive-sense
-ve	Negative-sense
cM	Centimorgan
MB	Megabases
GB	Gigabases
mV	Millivolt
pA	Picoampere
μl	Microliter
pmol	Picomole
°	Degree
C	Centigrade
bp	Basepair
s	Second

ACRONYMS

AP2B1	AP-2 complex subunit beta
AP2M1	AP-2 complex subunit alpha-1
COPG1	COPI Coat Complex Subunit Gamma 1
IRF 3/9	Interferon regulatory factor 3/9
TMED1	Transmembrane P24 Trafficking Protein 1
ABIN2	TNFAIP3 interacting protein 3
ACE2	Angiotensin-converting enzyme 2
ACTA1	Actin, alpha skeletal muscle
ACTB	Actin, cytoplasmic 1
ACTC1	Actin, alpha cardiac muscle 1
ANPEP	Alanyl Aminopeptidase, Membrane
API	Activator protein 1
AP2A1	AP-2 complex subunit alpha-1
AP2A2	AP-2 complex subunit alpha-2
ARCN1	Archain 1
ATG12	Ubiquitin-like protein ATG12
ATG16L1	Autophagy Related 16 Like 1
ATP13A2	ATPase Cation Transporting 13A2
ATP2C1	Calcium-transporting ATPase type 2C member 1
BECN1	Beclin-1
C9ORF72	C9orf72-SMCR8 Complex Subunit
CDC	Centers for Disease Control
CLTC	Clathrin heavy chain 1
COPA	Coatomer subunit alpha
COPB1	Coatomer subunit beta
COPG2	COPI Coat Complex Subunit Gamma 2
DAB2	Disabled homolog 2-interacting protein
DDX3Y	DEAD-Box Helicase 3 Y-Linked
DMEM	Dulbecco's Modified Eagle's Medium
DNM1	Dynamin-related protein DNMI
DNM2	Dynamin-2
DNM3	Dynamin-3
dNTPs	Deoxynucleotides
Ds	Double-stranded
DYNC1H1	Cytoplasmic dynein 1 heavy chain 1
EDEM1	ER Degradation Enhancing Alpha-Mannosidase Like Protein 1

EDTA	Ethylene diaminetetraacetic acid
ERGIC1	Endoplasmic Reticulum-Golgi Intermediate Compartment 1
ERK1/2	extracellular signal-regulated kinases
FBS	Fetal bovine serum
GBF1	Golgi-specific brefeldin A-resistance guanine nucleotide exchange factor 1
GCN1L1	GCN1 Activator Of EIF2AK4
HAdV	Human Adenoviruses
HBV	Hepatitis B virus
HCMV	Human cytomegalovirus
HCV	Hepatitis C virus
HPV	Human papillomavirus
HSV	Herpes simplex virus
ICTV	International Committee on Taxonomy of Virus
IFN-1	Type I Interferon, RIP1
IGF2BP1	Insulin Like Growth Factor 2 mRNA Binding Protein 1
IKK α	I κ B Kinase α
IKK β	I κ B Kinase β
IRAK1/2/4	Interleukin-1 receptor-associated kinase
ISG	Interferon-stimulated gene
ISRE	Interferon-Stimulated Response Element
JAK1	Janus kinase 1
JNK1/2	c-Jun N-terminal kinases
LARP	La Ribonucleoprotein 1, Translational Regulator
MAP1LC3A	Microtubule Associated Protein 1 Light Chain 3 Alpha
MAP1LC3B	Microtubule Associated Protein 1 Light Chain 3 Beta
MAVS	Mitochondrial antiviral-signaling protein
MDA5	melanoma differentiation-associated protein 5
MK2	MAPK-activated protein kinase 2
MKK1/2/3/4/6/7	Mitogen-activated protein kinase
MYD88	Myeloid differentiation primary response protein MyD88
NCCS	National Centre for Cell Sciences
NCK1	NCK Adaptor Protein 1
NEMO	NF-Kappa-B essential modulator
ORFs	Open reading frames
OS-9	OS9 Endoplasmic Reticulum Lectin
PAMP	Pathogen-associated molecular pattern
PBS	Phosphate buffer saline

RAB11B	Ras-related protein Rab-11B
RAB13	Ras-related protein Rab-13
RAB5	Ras-related protein Rab-5A
RAB5B	Ras-related protein Rab-5B
RAB5C	Ras-related protein Rab-5C
RAB7A	Ras-related protein Rab-7a
RAB7B	Ras-related protein Rab-7b
RAB7L1	Ras-related protein Rab-7L1
RIGI	retinoic acid-inducible gene I
RV	Rabies virus
SARS-CoV	Severe acute respiratory syndrome-coronavirus
SARS-CoV-2	Severe acute respiratory syndrome-coronavirus-2
SEC61	Protein transport protein Sec61 subunit alpha isoform 1
ss	Single-stranded
STAT1/2	Signal Transducer and Activator Of Transcription
STX17	Syntaxin-17
TAB2/3	TGF-Beta Activated Kinase Binding Protein
TAK1	TGF-beta-activated kinase
TBK1	TANK Binding Kinase 1
TLP2	Thioredoxin
TLR	Toll-like receptors
TMPRSS2	Transmembrane protease serine 2
TMV	Tobacco mosaic virus
TRAF3	TNF Receptor Associated Factor 3
TRAF6	TNF Receptor Associated Factor 6
TRIF	TIR domain-containing adapter molecule 1
TyK2	Tyrosine kinase 2
VAPA	VAMP Associated Protein A
VAPB	Vesicle-associated membrane protein-associated protein B/C
VPS11	Vacuolar protein sorting-associated protein 11 homolog
VPS33A	Vacuolar protein sorting-associated protein 33A
VSV	Vesicular stomatitis virus

CHAPTER 1

Chapter 1

Introduction

1.1 Virus overview

Viruses are the smallest and the most abundant biological substance on the earth. According to one of the estimates, around 10^{31} viruses are present in the biosphere, so practically, viruses are everywhere. They can infect almost all kinds of life, including plant, animal, bacteria, fungi, amoeba, and archaea bacteria [1,2]. In contrast, a virus particle comprises only two macromolecules, nucleic acids (DNA or RNA) as genetic material, and a protein coat that encapsulates this genetic material[3]. It is well known that viruses do not carry metabolic components. Most notably, viruses cannot generate ATP by themselves and do not bear the required machinery essential for translation. Specifically, they don't have ribosomes and cannot autonomously make proteins from messenger RNA. That's why they are considered obligate intracellular parasites[4].

1.1.1 History

When we talk about the virus, the first question that appears in our mind is how long ago the human viruses first appear? Human life originated about 34 million years ago, and the oldest document of the virus in history was found 4000 years back in ancient Egypt. The oldest virus evidence was observed in a stele from 13th-century BC in Egypt, where a man is standing with a stick is believed to be a victim of poliovirus. Except for poliovirus, Smallpox was also explained in ancient China's literature (700 BC), and the first physical evidence was found in the Egyptian mummies[5,6].

Even after a long history of viral infection, the first virus was discovered in 1892 by a Russian biologist named Dmitry Losifovich Ivanovsky. He observed an exciting and unexpected finding during his study on the Tobacco mosaic disease of

a plant. He observed that even after filtration, individual infectious particles remained in the solution to infect fresh tobacco plants. After six-years, Martinus Beijerinck, a Dutch scientist, separately made similar observations in his research on a plant's tobacco mosaic disease and speculated these infectious particles exist only in living cells. He gave the term 'virus' for these infectious particles[7]. Consequently, at the same time, two German scientists, Loeffler and Frosch, made a similar observation for an animal virus that causes foot and mouth disease in cows[8]. The first human virus was discovered in the early 20th century when Walter Reed demonstrated that a filter inoculum from an infected person could infect healthy volunteers. The moment of discovery of the yellow fever virus (YFV) in 1902 was the first human virus ever isolated. From the discovery of YFV to now, more than 219 species are known to infect humans[9,10]. However, 3-4 new viral species are still discovered every year and cause millions of deaths globally.

1.1.2 Classification of viruses

The scientific community proposed in 1953 that viruses should be separated from non-viruses. To justify their demand, they have pointed out a few discriminative characters of viruses: the first one is that viruses only carry one type of nucleic acid, either DNA or RNA; another type contains both DNA and RNA.

The second one is that Virions are reproduced from their nucleic acid, whereas other agents are reproduced from their constituents' integrated sum. The third one is that Virions are unable to expand and undergo binary fission. After this, many classification systems came for viruses, i.e., Cooper's system, Hamparian, Hillman, Ketlers's system, The Loff Horne Tournier system, and the Baltimore classification system[11].

Out of those, Baltimore is the most conventional and widely accepted classification system of viruses. In 1971, David Baltimore proposed animal viruses classification, and he divided them into seven groups: DNA viruses (Group I and II), RNA viruses (Group III, IV, and V), and RT viruses (Group VI and VII)[12].

Group I: Double-stranded (ds) DNA virus

Group II: Single-stranded (ss) DNA virus

Group III: Double-stranded (ds) RNA virus

Group IV: Single-stranded (ss) positive-sense (+ve) RNA virus

Group V: Single-stranded (ss) negative-sense (-ve) RNA virus

Group VI: Single-stranded RNA (ss) with DNA intermediate with the help of reverse transcriptase enzyme

Group VII: Double-stranded DNA (ds) with RNA intermediate with the help of reverse transcriptase enzyme

Whether some viral genomic strand directly replicates and makes mRNA, which translates and makes viral structural proteins. Different mRNA transcription approaches represent the trademark of each virus group. Viruses belonging to Group I have ds DNA genome structure and directly synthesize mRNA from the DNA genome template. GROUP II viruses have ssDNA genome, and their genome first converts into dsDNA, which is then used as a template for mRNA transcription. Group III viruses have dsRNA genome, and they directly manufacture mRNA by transcription from their double-stranded RNA template. Group IV viruses have +ssRNA genome. They use genomic RNA now as mRNA.

Viruses including a negative-stranded RNA genome represent group V. Group V viruses made mRNA by transcription from their RNA genome template. Group VI and VII viruses are called reverse transcriptase (RT) viruses. These viruses have either RNA or ds DNA genome. RT viruses shared an important feature that the viral DNAs are synthesized via a reverse transcription mechanism. Even Group VI viruses contain an RNA genome, but the genomic RNA does not work as mRNA, unlike Group IV (Figure 1.1).

It is noteworthy to mention here that Viral species are officially classified and named by an international committee called the International Committee on Taxonomy of Virus (ICTV) and was established in 1966. ICTV can be organized and categorizes recorded viral species into different hierarchical ranks, starting with orders, then families, then genera, and then species. Species can be further subdivided into genotypes[13].

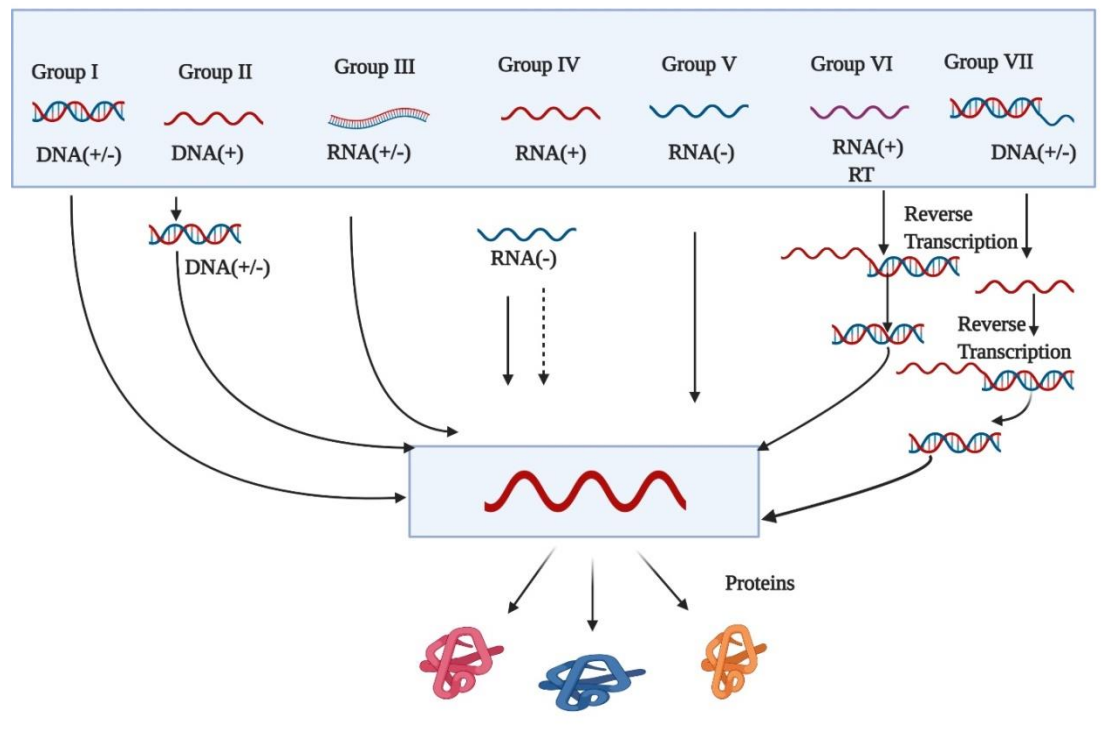


Figure 1.1. Schematic diagram of Baltimore classification. Baltimore divided viruses into seven groups based on virus genomes and their replication mode. The intermediate steps involved in mRNA synthesis are also described in the diagram.

1.1.3 Host-virus pathogenesis

Virus infections are not analogs of disease or death. Many microbial infections are subclinical or asymptomatic. Typically, three kinds of host-pathogen interaction exist in nature; mutualism, commensalism, and parasitism. The virus comes under parasites; as mentioned above, it doesn't carry essential substrates for its replication and other processes. It depends on the host for its replication and progeny process. Although all host has a common defense mechanism which may include:

- Skin and mucosal secretions
- Non-specific local responses (e.g., pH)
- Non-specific inflammatory responses (IL2, IL10)
- Specific immune reactions (B cell, T cells mediated response)

Some viruses invade these host defense mechanisms and cause pathogenesis. Pathogenesis is the process by which a virus develops the disease (Figure 1.2). It can occur when a virus enters the cells, colonizes, invades the host defense mechanism, increases its numbers, damages host tissue and exit from the host body, and infects other individuals[14,15].

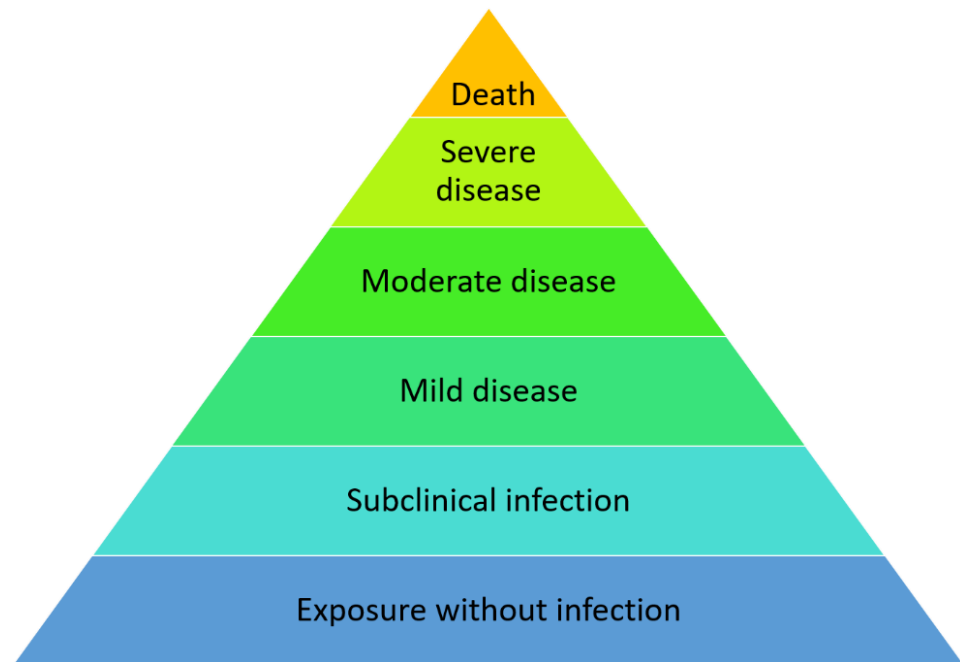


Figure 1.2. Schematic diagram of virus infection and pathogenesis

1.1.4 Host virus interaction

As we discussed earlier, a virus is not able to propagate alone. For propagation and pathogenesis, it interacts with the host factors. Host factors are host proteins that activate after viral infection. It can be divided into two groups. The first is essential host- factors (proviral), which is critical for viral infection, and the second is restriction host factor, which inhibits viral infection. Identification of viral restriction host factors can lead to new advances in antiviral therapies[16].

Recent advancements in functional genomics and proteomics have provided an unbiased platform for identifying cellular factors involved in viral infection. Several screening methods have been used in this course, including loss of genomic function and gain of function screens[17]. Genome-wide techniques, RNAi-mediated transient silencing of mRNAs in mammalian cells (siRNA) and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas ribonuclease system are widely used to identify cellular requirements for viral infection[18-21]. A combination of other omics technologies, including proteomics and transcriptomics, can give a more comprehensive picture of the complex interactions between viruses and their hosts. Commonly genome-wide screening occurs in appropriate cells; most probably, cells have been chosen based on the virus, primarily infection site is the natural host. With the time, with extensive scale screening of host factors in different cell lines, researchers will get information about common host factors and pathways between cells or viruses.[22-24] This information may allow us to design broadly acting antiviral drugs.

1.1.5 Role of host factors in zoonosis

When animal viruses infect humans, this process is called zoonosis, which can be a reason for the epidemic and pandemic conditions. In contrast, humans are regularly exposed to animal viruses through food, pets, and interactions with nature. Most of the viruses that enter our bodies pass through our gastrointestinal tracts without any effect or, if they enter the cells, they can be neutralized by our immune systems[25]. However, on the minimal condition, an animal virus infects humans and starts replication itself[26]. Replication of an animal virus within the first

human host is crucial in the zoonotic process because it possibly renders two things. The first thing is that the virus does mutations and adapts and improves itself for virus replication in the new host. Secondly, the replication produces a high titer, and the virus is ready to infect the second human. With time, only highly replicating and spreading viruses can go for positive selection. Now we know that host genetics play a crucial role in determining which animal virus can cross the species barrier and replicate in a first human host (Figure 1.3).[27]

Animal viruses replicate in human cells, where they interact with most of the essential host factors. Simultaneously, they need to avoid interaction with all immune and restriction proteins that would inhibit them. For most animal viruses, these conditions are very stringent[28]. It is just a chance when a virus matches these conditions. Moreover, a virus with a few or no genetic barriers to replicate in human cells poses the greatest risk to humans (Figure 1.4).

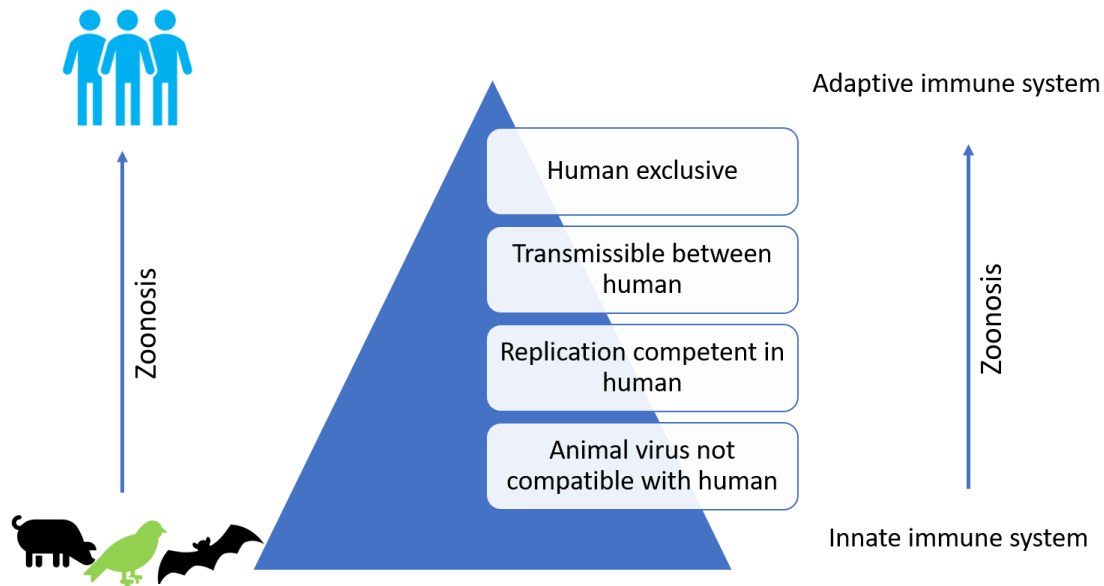


Figure 1.3. Viral adaptation to the human and viral disease emergence. On the left side of the figure shows that most of the zoonosis occur from birds and animal to human, the middle part of the model shows a zoonosis pyramid which carries a concept that that animal viruses become frequently adapted to human through a series of evolutionary steps represented from bottom to top. The right part of the figure represents the status of immune system functionality against zoonosis. In the early phase of zoonosis, the innate immune plays a crucial role in inhibiting virus infection, but when the virus is adopted in the new host, the adaptive immune system plays a key role.

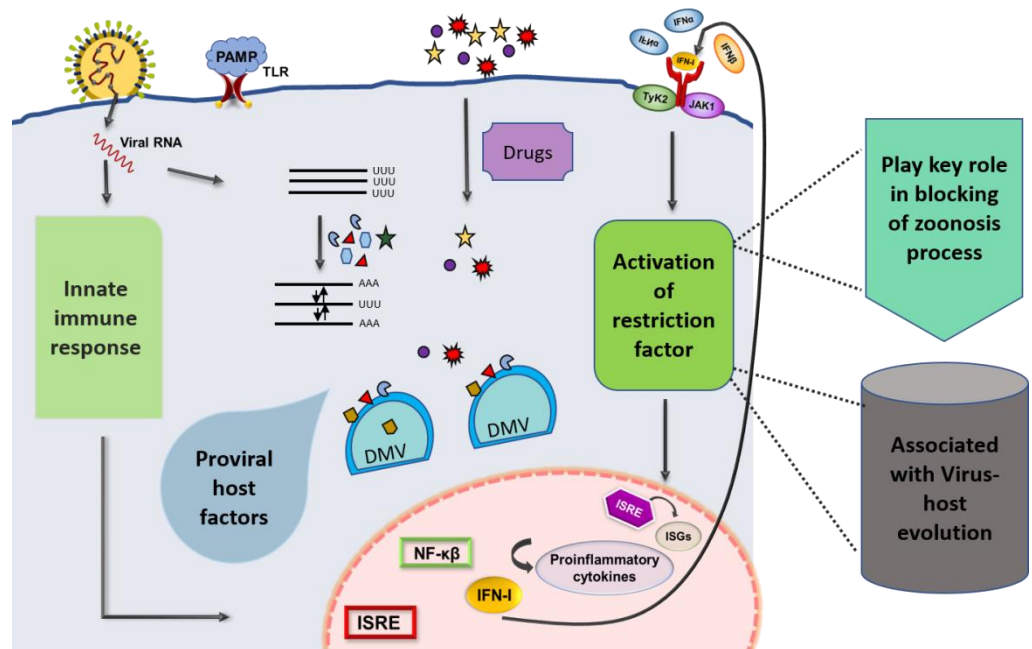


Figure 1.4. Role of Host protein in virus zoonosis process. Host carries two kinds of protein, proviral and antiviral. Whenever a virus infects the host, the innate immune system will be activated, and simultaneously proviral gens also activated. For proper virus adaptation and zoonosis, the virus is required to use most of the proviral host factors for their infection and replication, and at the same time, it is also needed to protect itself from restriction factors.

1.1.6 Coevolution of virus and host

As discussed earlier, viruses interact with the host, and host proteins play a crucial role in virus infection. Moreover, this virus and host interactions have resulted in various evolutionary outcomes. During this deep-rooted shared history, viruses and hosts have always stressed each other for survival. Over time, hosts have acquired the first defense mechanisms against viruses, called restriction factors. Restriction factors are host proteins that are part of the innate immune system, and they potently inhibit pathogen infection[29,30]. These restriction proteins target the various stages of the viral life cycle. Some factors directly interact with the virus or viral protein, while others indirectly inhibit virus infection by stopping macromolecule synthesis, apoptosis, etc. But with the evolution, viruses have also evolved antagonistic mechanisms against restriction factors[31]. These antagonistic connections between the host factors and the virus proteins induced an evolutionary genetic conflict between the two entities. This genetic contest is also called the "virus-host arms race," based on the Red Queen hypothesis, where organisms continuously develop and adapt to survive with changing environments[32]. But most of the time, the virus wins the 'virus-host arms race.' because the virus has a small life span and can evolve faster than its hosts. That's why the current-day host's innate immune system works against the ancient virus rather than more recent viral infection[33]. Therefore, the evolutionary history of restriction factors may explain why a host is responsive or resistant to a newly emerging virus (Figure 1.5). A newly emerging coronavirus (SARS-CoV-2) is an excellent example of virus-host arms race evolution. Evolutionary arms race dynamics develop the heterogeneity of viruses and their receptors. In a recent study, it is reported that Chinese horseshoe bat populations show high polymorphism in ACE2. These ACE2 modifications support SARS-CoV and SARSr-CoV infection, but it shows different binding affinities to different-different spike proteins. The authors found that the SARSr-CoV spike shows a higher binding affinity to human ACE2[34,35]. This study suggests that these viruses can cross the species barrier and infect humans.

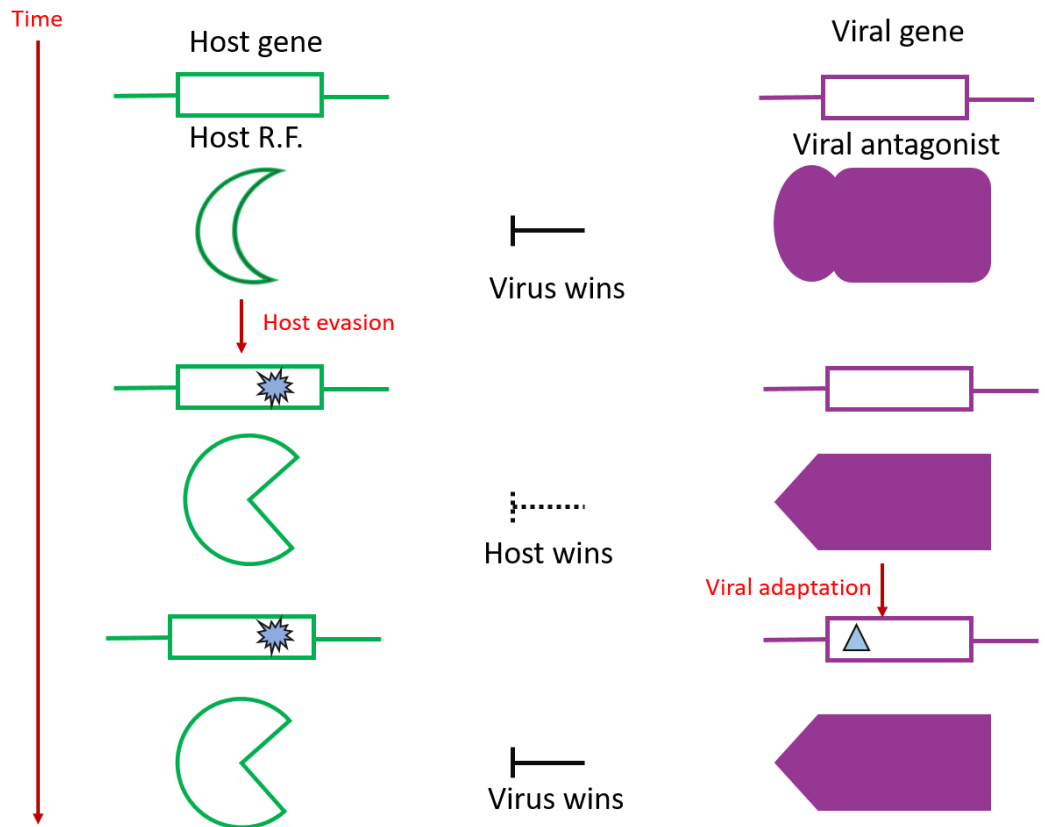


Figure 1.5. Schematic representation of virus-host co-evolution. In this figure, we describe a host-virus arm-race, an antiviral protein antagonized by the virus, but human protein evolution allows the host protein to win this battle against the virus. but with time virus again evolved and overcome by the host protein effect. Moreover, most of the time, the virus wins the 'virus-host arms race.' because the virus has a small life span and can evolve faster than its hosts.

1.2 Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2)

1.2.1 Classification lineage and history

SARS-CoV-2 is an enveloped positive-sense, single-stranded RNA betacoronavirus, which belongs to the Coronaviridae family. A Coronaviridae family has a diverse group of viruses that infect many different animals. Six human coronaviruses (HCoVs) were previously identified; HCoV-229E and HCoV-NL63 belong to the alpha coronavirus group. The rest members, HCoV-HKU1, HCoV-OC43, severe acute respiratory syndrome coronavirus (SARS-CoV), and Middle East respiratory syndrome coronavirus (MERS-CoV) belong to the betacoronavirus group[36,37]. However, over the past few years, highly pathogenic human coronaviruses have emerged. The appearance of SARS-CoV in 2002 exhibited ~8,000 cases worldwide with a mortality rate of ~10%. Similarly, MERS-CoV in 2012 marked with 2,500 patients with a higher mortality rate of 36%[36]. In late December 2019, a novel coronavirus named SARS-CoV-2 emerged in Wuhan, China, which causes unusual viral pneumonia[38]. This novel coronavirus disease, also called coronavirus disease 2019 (COVID-19), is highly transmissible and spreads fast. The COVID-19 patients showed similar symptoms like SARS and MARS. Most of the first 27 known hospitalized patients are epidemiologically linked to Huanan Seafood Wholesale Market. According to a retrospective study, patient zero of SARS-CoV-2 was observed on 8 December 2019[39]. On 31 December china government notified and informed the world health organization about a pneumonia outbreak from an unknown cause; after that virus spread very fast from China to the whole world, and the rest is history[40].

1.2.2 Clinical impact

SARS-CoV-2 infects almost all ages of the people, and the median age of infection is around 50 years. However, the clinical conditions differ with age; people age above 60 develop a severe respiratory disease that may require hospitalization or even die. Most young people and children have only mild disorders or are asymptomatic. Whenever a virus infects most common symptoms are fever, dry cough, fatigue, olfactory and taste disorders, and less common

symptoms are sputum production, headache, haemoptysis, diarrhea, anorexia, sore throat, chest pain, chills, and nausea and vomiting[41,42].

1.2.3 Genomic Structure and Replication cycle

SARS-CoV-2 shares genomic similarity with other betacoronavirus. It shows 79% genome sequence similarity with SARS-CoV and 50% with MERS-CoV. It is a single-stranded positive-sense RNA virus with a genome size of 29.9 kB. The viral genome carries 14 open reading frames (ORFs), which make 27 different proteins. First, six ORFs are organized in order from 5' to 3' direction: spike (S), replicase (ORF1a/ORF1b) envelope (E), membrane (M), and nucleocapsid (N). Besides, seven putative ORFs encoding accessory proteins are distributed between the structural genes (Figure 1.6). The replicase gene of SARS-CoV-2 covers two-thirds of the 5' genome. It encodes a large polyprotein (pp1ab), further cleaved into 16 non-structural proteins (NSPs) that participate in transcription and virus replication (Table 1.1). The 3'UTR comprises four structural genes and eight accessory genes[43,44]. The accessory genes are located between the structural genes, and their function is not known till now.

The SARS-CoV-2 genome work as a template for both replication and translation. The virus enters the cell and releases its genome in the cytoplasm. The virus controls its proteins' relative expression through a conserved molecular mechanism, identified as -1 programmed ribosomal frameshifting (-1 PRF). SARS also observed a similar mechanism. The two ORFs, ORF1a and ORF1b, translate into non-structural proteins (NSP1-NAP16), which participate in virus replication, immune suppression, and other viral beneficiary works.[45] The viral genome's replication and transcription process is mediated by RNA-dependent RNA polymerase activity (RdRP/nsp12). The NSP12 catalyzes the synthesis of viral RNA with the help of nsp7 and nsp8 as cofactors. Usually, RNA virus lacking proofreading activity, but coronavirus has a specific protein named NSP14, which provides proofreading activity. Virus replication occurs in the replication complex; viral and host both proteins make this complex and produce full-length negative-sense RNA intermediates, which serve as a template for the synthesis of positive-

sense RNA. The nucleocapsid protein encapsulates the genomic RNA(positive-sense RNA), and The virus assembly and maturation process occur inside the Golgi. after the maturation virus is released from the cells (Figure 1.7)[46-48].

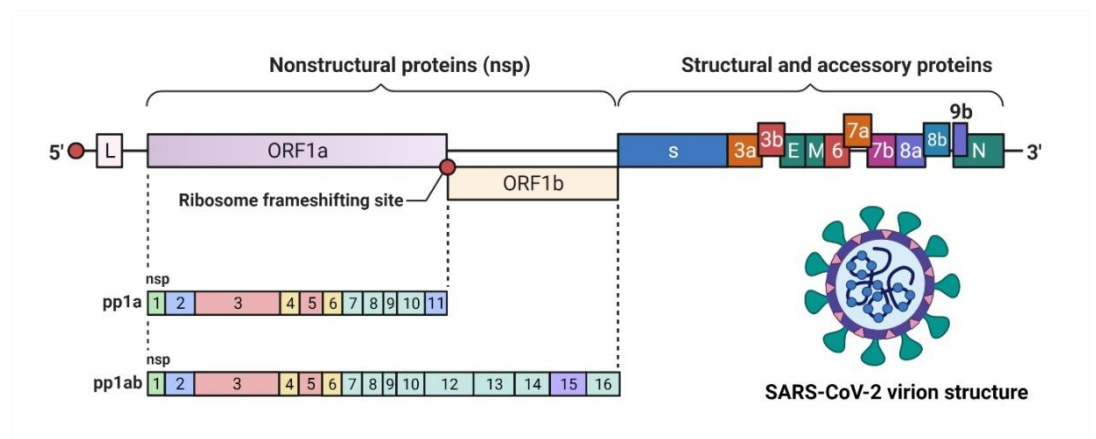


Figure 1.6. Genomic organization and molecular structure of SARS-CoV-2.

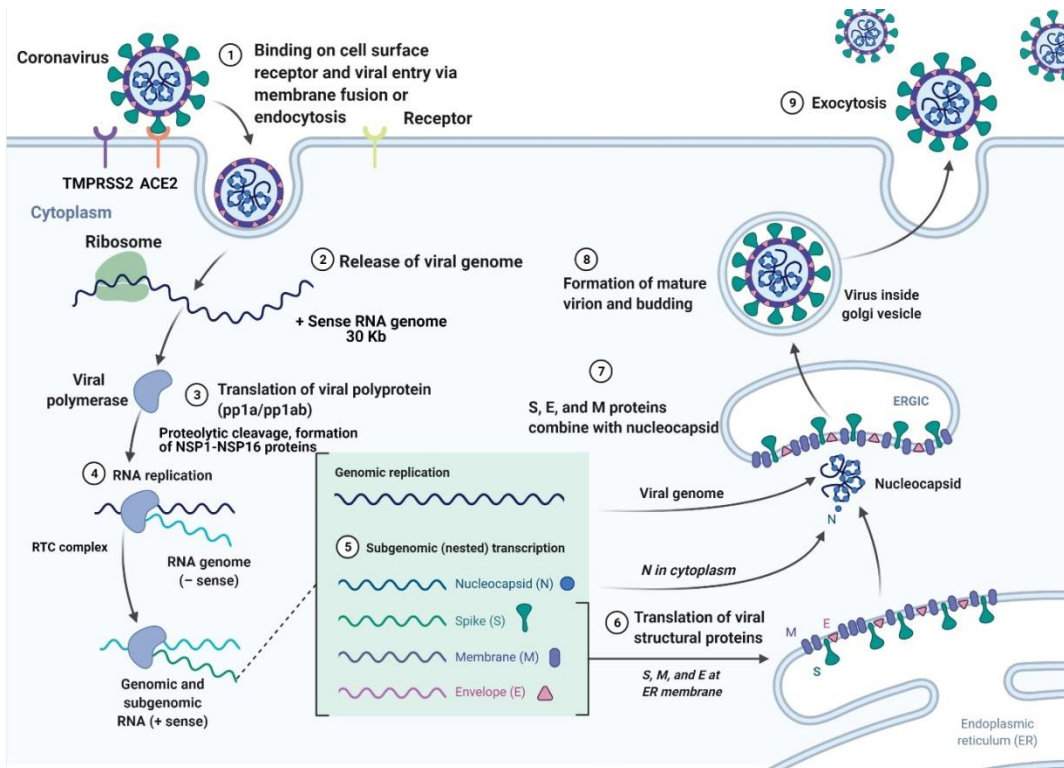


Figure 1.7. A schematic diagram displaying all the significant steps of the SARS-CoV-2 life cycle

Table 1.1. List of non-structural proteins of SARS-CoV-2 and their function

S.No.	Protein name	Function
1.	nsp1	Interferes with the mRNA binding and suppresses hosts' immune functions, Anchors the replication complex to cellular membranes Degrades host's mRNA by interacting with the human 40S ribosomal subunit
2.	nsp2	Harbours mutations that make it more contagious Might play a role in the modulation of host cell survival; also known as p65 homolog
3.	nsp3	Papain-like protease 2 (PL2 ^{pro}) involved in proteolytic cleavage
4.	nsp4	Responsible for the formation of the double-membrane vesicle during replication, Anchors the viral replication-transcription complex to the membranes of the endoplasmic reticulum
5.	nsp5	Proteases (3CL ^{pro} , M ^{pro}) involved in polypeptide cleaving
6.	nsp6	Prevents the expansion of autophagosome, Help information of double-membrane vesicle; suppresses IFN-I signaling
7.	nsp7	Forms a hexadecamer with nsp8 and acts as a primase in viral replication
8.	nsp8	Acts as a primase with nsp7
9.	nsp9	Acts as ssRNA binding protein
10.	nsp10	Plays role in the methylation of viral mRNA cap. Stimulates the nsp14 3'-5' exoribonuclease and 2'-O-methyltransferase (NSP16) activities
11.	nsp11	Unknown
12.	nsp12	Catalytic subunit of the RNA-dependent RNA polymerase; Catalyses the synthesis of viral RNA, using nsp7 and nsp8 as cofactors
13.	nsp13	Helicase and NTPase activity: hydrolyze the NTPs and unwind the duplex RNA and DNA with a 5' single-stranded tail in a 5' to 3' direction A potent interferon antagonist
14.	nsp14	Guanine-N7 methyltransferase, a multienzyme complex Acts on both sides ssRNA and dsRNA in a 3'-> 5' direction A potent interferon antagonist It plays a role in genome replication, sub-genomic RNA synthesis, and recombination
15.	nsp15	It is a nidoviral RNA uridylylate-specific endoribonuclease (NendoU); plays a role in viral replication and transcription A potent interferon antagonist
16.	nsp16	Acts as 2'-O-methyltransferase that mediates mRNA cap 2'-O-ribose methylation to the 5'-cap structure of viral mRNAs

1.2.4 Treatment and prevention

Limited initial drugs are suggested against the SARS-CoV-2, but their efficacy remains poor. Many drugs suggest and are used to treat COVID19 patients; these drugs belong to different classes include nucleotide analogue, monoclonal antibody, protease inhibitor, RNA polymerase inhibitor, endonuclease inhibitor, immune modulator, interferon-alpha, and immune suppressor. Scientists across the world are trying to make a vaccine against SARS CoV-2. According to the WHO report, there are 48 vaccine candidates under the clinical trials' advanced stages. Out of them, 11 are currently in the phase III trial[49-52].

1.3 Chandipura virus (CHPV)

1.3.1 Classification lineage and history

Chandipura virus belongs to the vesiculovirus genus and Rhabdoviridae family, which comes under the mononegaviridae. Viruses belong to mononegaviridae are characterized by a non-segmented, single-stranded RNA genome. Among all mononegaviridae Rhabdoviridae family got significant attention because of its wide host range[53]. Viruses from the Rhabdoviridae family can infect humans, other vertebrates, arthropods, fishes, and even infect members of the plant kingdom. They have typical bullet shape morphology[54]. Similarities between CHPV and VSV in genetic structure, polypeptide composition, and life cycle include CHPV within the vesiculovirus genus. Comparative sequence analysis of CHPV and VSV shows that CHPV is evolutionarily equidistant from new world vesiculoviruses VSV Indiana (VSVind) and VSV New Jersey (VSVnj), and closely related to its Asian kin Isfahan[55].

Chandipura virus was first discovered and isolated accidentally from the two adults' blood with a febrile illness in a village in Nagpur district, Maharashtra, India, in 1965[56]. The only other instance when it was isolated from a human in 1980, in Madhya Pradesh, India, from an acute encephalitis patient[57]. However, the retrospective serological studies indicate that CHPV history is more senior and its infected human population as early as 1957-58[58].

However, due to the low case reports, it has not got much attention until the 2003 outbreak in Andhra Pradesh, India. The epidemic started in June 2003 and continued next 3-4 months. Around 329 children (age between 9 months -14 years) were infected with the virus, showed encephalitis, and 183 children died[59]. Simultaneously CHPV encephalitis outbreak was also reported in Fifteen districts of Maharashtra during the same time. In the subsequent year 2005, another outbreak was reported in the Baroda district of Gujrat, India, with 70 % cases of mortality in the pediatric population. However, many small outbreaks are reported throughout the country with different periods[60-62]. Moreover, the CHPV virus outbreak is not only limited to India. It is also reported in other countries of Asia (Bhutan, Nepal, Sri Lanka) and Africa (Nigeria, Senegal) subcontinent[63,64].

1.3.2 Clinical impact

CHPV is an emerging tropical virus with a high mortality rate in children below 15. Usually, CHPV causes cerebral edema and cerebral encephalitis. These patients also show high-grade fever, vomiting, loose motion, hypertonia of the limbs, hyperreflexia, bilateral extensor plantar response, with a sudden drop in blood pressure.[53,54]

1.3.3 Genomic Structure and Replication cycle

The CHPV is an enveloped, non-segmented RNA virus with a genome size of 11 kb. CHPV genomic RNA contains a 49 nucleotide leader gene (I), followed by five transcriptional units separated by intragenic spacer regions and a short non-transcribed 46 nucleotide trailer sequence (t) arranged in the order 3' I-N-P-M-G-L-t 5'. It has a lipoprotein envelope containing a helical ribonucleoprotein (RNP), which further enwraps an RNA genome. CHPV genome codes five polypeptides; namely, glycoprotein (G), large protein (L), nucleocapsid protein (N), matrix protein (M), and phosphoprotein (P) (Figure 1.8) The G protein spikes present on the outer membrane. They play a role in virus entry, receptor recognition, the fusion of viral and cellular membrane, and a major antigenic determinant. The M protein acts as an adhesive layer between the lipid bilayers' inner membrane and the core nucleocapsid. The core N protein encapsulates viral genomic RNA and protects it

from cellular RNase. It also enwraps the viral genome and forms the template for viral transcription. L and P protein form viral RNA-dependent RNA polymerase, where L performs the catalytic activity for RNA polymerization, and P acts as a transcriptional activator[65,66].

CHPV virus life cycle can be divided into many steps. At first, the viral glycoprotein interacts with the host cell plasma membrane and is internalized by endocytic vesical; uncoating occurs in the late endosome in the low Ph condition and virus genome release in the cytoplasm. Released viral RNA starts the transcription process, and viral polymerase synthesizes five discrete mRNAs and obeys to stop signals present at the genomic RNA or transcribed in decreasing order of molar ratio as N>P>M>G>L and enter the translation phase. After accumulating the subsequent amount of viral proteins in the host cell, L protein acts as a replicase. It ignores the gene junctions to generate a polycistronic antigenomic analog that works as a template for further replication to create many more copies of the genome RNA. The newly developed genomic RNA goes for viral protein packaging and is released as a mature particle from the cells (Figure 1.9)[67,68].

1.3.4 Treatment and prevention

Till now, there is no such vaccine or medicine available for the Chandipura virus. Most of the time, doctors gave symptoms-based treatment to the patients. However, some recombinant and killed virus vaccine approaches are in a clinical trial[69].

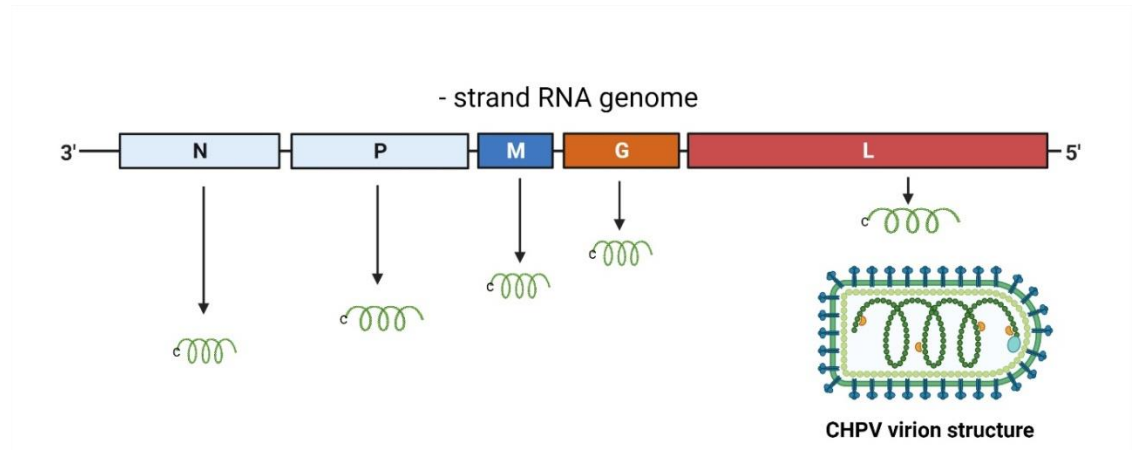


Figure 1.8. Genomic organization and molecular structure of CHPV

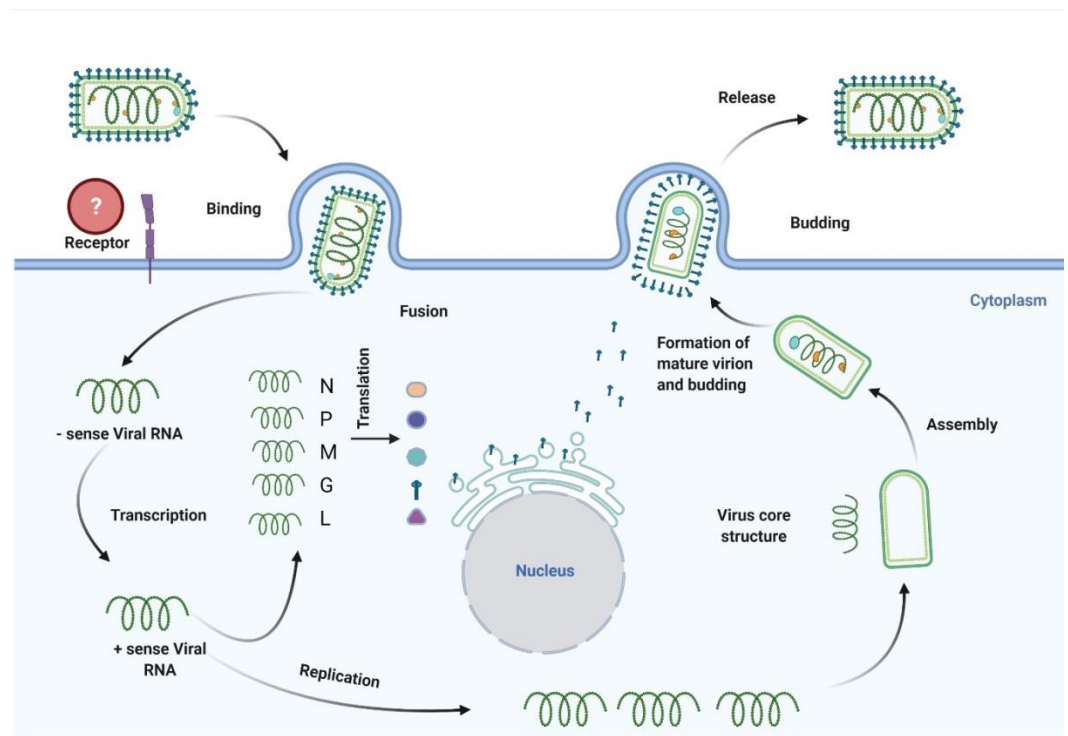


Figure 1.9. A schematic diagram displaying all the significant steps of the CHPV life cycle

1.4 Bone marrow stromal antigen 2 (BST-2)

1.4.1 History and molecular characterization

Initially discovered as a surface marker for terminally differentiated and neoplastic B cells, bone marrow stromal antigen-2 (BST-2) was later reported to have diverse cellular functions[70,71]. The protein is expressed in almost all cell types; however, the degree of expression varies from cell to cell[72,73]. BST-2 is a type II transmembrane protein and contains ~180 amino acids (aa). The mature protein adopts a unique topology comprising a short N-terminal cytoplasmic tail (1-20aa) followed by an α -helical transmembrane domain (21-48aa), an ectodomain (49-161aa), and a C-terminal glycosylphosphatidylinositol (GPI) domain (162-180aa)[74,75]. The cytoplasmic tail of BST-2 has a highly conserved YXY tyrosine motif known to play a role in NF- κ B-mediated signaling and clathrin-mediated endocytosis (Figure 1.10)[76-78]. Simultaneously, its ectodomain contains two evolutionarily conserved motifs - three cysteine residues and two glycosylation sites[79]. These conserved cysteine residues covalently link monomers to form dimeric or tetrameric forms of BST-2[80]. Two N-linked glycosylation sites (at N65 and N92) are required for the proper folding of BST-2, and the GPI domain anchors BST-2 to cell surface lipid rafts. In addition to the plasma membrane, BST-2 expression is also detected in the trans-Golgi network and within the recycling endosomes in the cytoplasm[81].

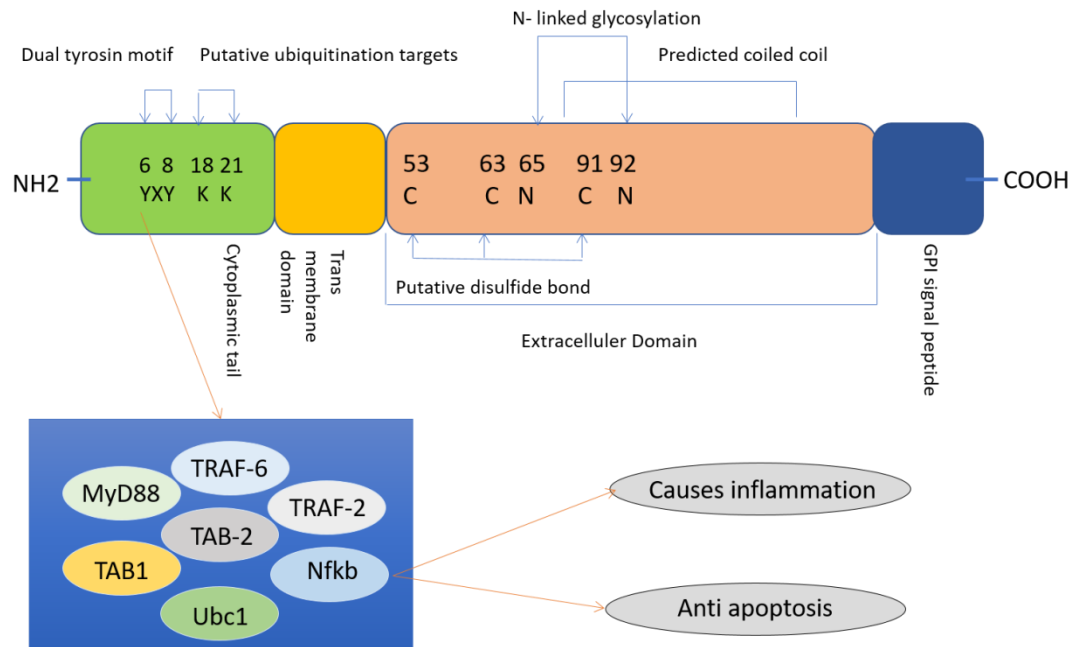


Figure 1.10. Molecular characterization of BST-2. BST-2 is a type II transmembrane protein and contains ~180 amino acids (aa). Its cytoplasmic domain contains The mature protein comprising of a short N-terminal cytoplasmic domain (1-20aa) followed by an α -helical transmembrane domain (21-48aa), an ectodomain (49-161aa), and a C-terminal glycosylphosphatidylinositol (GPI) domain (162-180aa). The cytoplasmic tail of BST-2 has a highly conserved YXY tyrosine motif which plays a crucial role in NF- κ B-mediated signaling and clathrin-mediated endocytosis.

1.4.2 Antiviral activity of BST-2

Bieniasz and colleagues reported in 2008 that BST-2 restricted the release of human immunodeficiency virus-1 (HIV-1). BST-2 inhibited virion release of a recombinant HIV-1 lacking the Vpu gene by tethering the nascent virion to the host cell plasma membrane [82]. This prompted the name “tetherin” for BST-2. Subsequently, BST-2 was shown to tether a broad range of enveloped viruses, but its tethering activity against enveloped viruses is not a universal phenomenon. The antiviral activity of BST-2 is related to its membrane anchoring topology [79]. The two anchoring domains of BST-2 form a bridge between the budding virion and the host plasma membrane, thereby physically restricting virion release [82,83]. Perez Caballero et al. demonstrated that the membrane anchoring domains are necessary and sufficient for viral tethering.

Interestingly, the action of tetherin occurred independently of other host cofactors. Engineered forms of BST-2 lacking any significant sequence homology but retain the transmembrane domain, dimeric ectodomain, and GPI anchor domains restricted HIV-1 release from the cell surface, demonstrating the importance of these domains for BST-2 antiviral activities [84]. Tethered virions are either retained at the cell surface or mobilized for endocytic internalization and subsequent ubiquitin-based degradation [85,86]. BST-2 antiviral activities against a variety of enveloped viruses, including retroviruses, alphaviruses, rhabdoviruses, and mammarenaviruses, have been discussed in several excellent review articles [78,87-91]. Here, we highlight other aspects of BST-2 biology, such as cell signaling, immunomodulatory functions, and immunity.

1.4.3 BST-2 and cell signaling pathways

BST-2 expression is induced by type 1 and type 2 interferons in response to viral infection. The role of BST-2 in cell biology was later reconsidered after it was established as a potent inducer of NF- κ B [92]. Still, this finding was initially overshadowed by studies focusing on its antiviral functions. More recent studies confirmed a role for BST-2 in regulating NF- κ B signaling [77,93,94]. BST-2 induction of NF- κ B depends on its multimerization or viral sensing actions [93],

but no experimental data are linking its tethering action with that of NF- κ B induction. Early studies identified TGF β -activated kinase 1 (TAK1) as being critical for BST-2-induced NF- κ B mediated signal transduction[77,93]. However, subsequent studies identified additional intermediate signaling molecules have a role in the activation of the NF- κ B signaling pathway by BST-2. For example, knockdown of TNF receptor-associated factor 6 (TRAF6) or TNF receptor-associated factor 2 (TRAF2)/ Ubiquitin-Conjugating Enzyme E2N (Ubc13) gene expression blocked BST-2 mediated NF- κ B activation[77]. Yet, Myeloid differentiation primary response protein MyD88, TRAF2, TGF β activated kinase 1 (TAB1), and TGF β activated kinase 2 (TAB2) were found dispensable for BST-2-induced activation of the NF- κ B signaling pathway[93]. These findings revealed that BST-2, via induction of NF- κ B signaling, can influence the host inflammatory response to a virus [95,96], but the underlying mechanisms remain to be elucidated.

1.4.3 Immunomodulatory role of BST-2 following infection

1.4.3.1 Interferons & innate immunity

Host-pathogen recognition and responses are multifaceted. Upon pathogen encounter, pattern recognition receptors (PRR), including Toll-like receptors (TLRs), Nod-like receptors (NLRs), and RIG-1 like receptors (RLRs)[97-100] can initiate antiviral responses, including IFN-I and pro-inflammatory cytokine production[101]. IFN-I signaling via the IFN receptor and JAK/STAT pathway induces the expression of hundreds of interferon-stimulated genes (ISGs) that contribute to establishing an antiviral state that limits virus propagation within the infected host[102]. IFN-I responses are highly controlled and short-lived, but if unchecked, excessive expression of IFNs may harm the host [103] and negatively affect hematopoiesis[104].

Plasmacytoid dendritic cells (pDCs) are among the highest producers of IFNs and inflammatory cytokines upon sensing bacterial or viral nucleic acids through TLR7 and TLR9 receptors [105,106]. In this context, Cao et al. demonstrated that BST-2 negatively regulates the IFN-I response in pDCs. BST-2

is a biological ligand for the human pDC-specific receptor immunoglobulin-like transcript 7 (ILT7), and binding of BST-2 to ILT7 can initiate signaling via the ILT7-FcεRIγ (a high-affinity IgE receptor) complex. FcεRIγ contains an immunoreceptor tyrosine-based activation motif (ITAM) in its cytoplasmic tail that mediates a calcium-dependent signaling cascade that inhibits the production of IFNs and inflammatory cytokines by pDCs[107,108]. However, the detailed mechanisms underlying this negative feedback loop are largely unknown. Co-culture of pDCs with BST-2-expressing cells reduced IFN-I production by pDCs following stimulation by the oligodeoxynucleotide, CpG-A[109]. Moreover, treatment with a MEK1/2 inhibitor significantly increased interferon production [109], suggesting that the BST-2-ILT7 mediated downregulation of IFN-I is at least partially associated with MEK1/2 signaling. This pathway is highly specific for human pDCs, as ILT7 is only present in human and primate pDCs[110].

BST-2 mediated down-regulation of the IFN-I response was also linked to this protein's ability to counteract the RLR-mediated IFN-I signaling pathway[111]. Specifically, BST-2 recruits the E3 ubiquitin ligase, MARCH 8, which catalyzes the lysine (K27) linked polyubiquitin chains on the mitochondrial antiviral-signaling protein (MAVS)[111]. This activity subsequently targets MAVS for autophagic degradation via nuclear domain 10 protein 52 (NDP52) receptor. As MAVS is an essential host signaling adaptor protein for IFN-I production, its degradation negatively affected the interferon response[111](figure 1.11). Interestingly, unlike humans, murine NDP52 lacks the ubiquitin-binding domain LIM-L[112-114], which should prevent BST-2-mediated degradation of MAVS. In fact, pDCs from BST-2 KO mice showed reduced IFN-I secretion in response to viral challenges [115]. These results suggest that the evolutionary selection of the LIM-L domain in NDP-52 and ILT7 expression by human pDCs cells influences how BST-2 regulates IFN-I production in these cells relative to mice pDCs. Additional research is required to determine how BST-2 affects the innate antiviral immune response in different species.

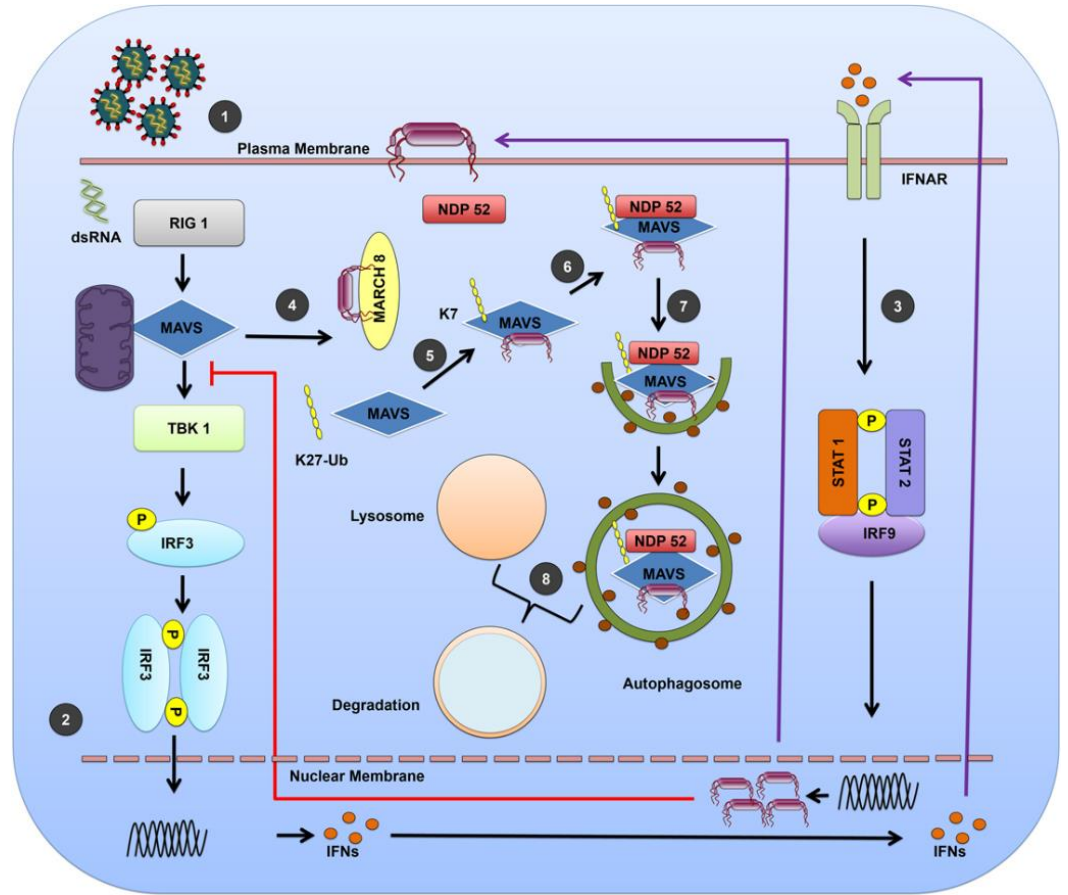


Figure 1.11. BST-2 regulation of the IFN-I response.

(1) Virion interacts with cell surface receptors to enter the cell. (2) Viral genome (RNA) is recognized by the RLRs. RIG-I signals are transduced to the transcription factors through stimulation of MAVS at the mitochondrion-associated membrane. Activation of MAVS leads to phosphorylation of IRF3. Phosphorylated dimers of IRF3 then translocate to the nucleus where they bind and activate specific promoters triggering expression of IFNs. (3) Type-I IFNs interact with IFNAR, recruit, and phosphorylate the STAT1 and STAT2. STAT1 and STAT2 form a heterodimer that, in turn, recruits the IRF9 to make a complex. This complex translocates to the nucleus and induces expression of genes (e.g., BST-2) regulated by IFN-stimulated response elements. (4) BST-2 recruits the E3 ubiquitin ligase MARCH 8. (5) MARCH 8 then catalyzes the K27-linked polyubiquitin chains on MAVS at K7 position. (6) Cargo receptor NDP52 recognizes ubiquitinated MAVS. (7) NDP52 delivers MAVS to autophagosome

for degradation. (8) BST-2-mediated autosomal degradation of MAVS and terminal of RIG-I, MAVS-mediated IFN α production via a negative feedback manner. BST-2, bone marrow stromal antigen 2; IFN, interferon; IRF3, IFN response factor 3; IFNAR, IFN- α/β receptor; MAVS, mitochondrial antiviral-signaling protein; RLR, RIG-I-like receptor; STAT, signal transducers and activators of transcription.

1.4.3.2 Adaptive Immunity

Although BST-2 has conserved coding regions, polymorphic forms do exist across species. BST-2 alleles lacking the endosomal-sorting motif (YxY) are found in NZW/LacJ (NZW) mice and show higher cell surface expression than C57BL/6 mice whose BST-2 contains the YxY motif. BST-2 from NZW mice restricted Friend murine leukemia retrovirus (F-MuLV) more potently than BST-2 from C57BL/6 mice [116]. Because BST-2 is expressed at higher levels on the cell surface in NZW mice (due to defective endocytosis), these mice should have an increased potential to restrict Friend retrovirus distribution in vivo. However, the endocytosis-competent version of BST-2 in C57BL/6 mice actually showed a greater ability to control viremia, suggesting a role for immune-modulatory functions linked to BST-2 [117]. For example, enhanced restriction of F-MuLV was associated with a stronger IFN γ response in NK cells, CD4 $^{+}$ T cells, and CD8 $^{+}$ T cells [117]. This study further proposed that increased endocytosis of virions by pDCs could trigger TLR3-mediated IFN-I production, leading to augmented NK function, as these cells are highly responsive to TLR3- and TLR7-dependent cytokine stimuli [115, 117, 118] (figure 1.12). This might account for lower IFN-I production observed in BST-2 deficient pDCs [115]. Because BST-2 can influence IFN-I levels, it has the potential to modulate host defense during both the early and late phases of viral infection. Using the chronic lymphocytic choriomeningitis virus (LCMV) infection model in mice, we showed that the early confinement and sequestration of virions in the splenic marginal zone was compromised in the absence of BST-2 [119]. This resulted in alterations in antiviral T cell priming, leading to reduced T cell proliferation and effector functions (e.g. IFN γ and TNF α). Peripheral control of a chronic LCMV infection was also compromised in BST-2 deficient mice, and the virus established long-term persistence in the brain [119]. Collectively, these studies demonstrate how BST-2 can influence both innate and adaptive immune responses to viral infections. Further studies are required to understand the direct versus indirect contributions of BST-2 to antiviral immunity and how this protein's functionality can be enhanced to help fight infections.

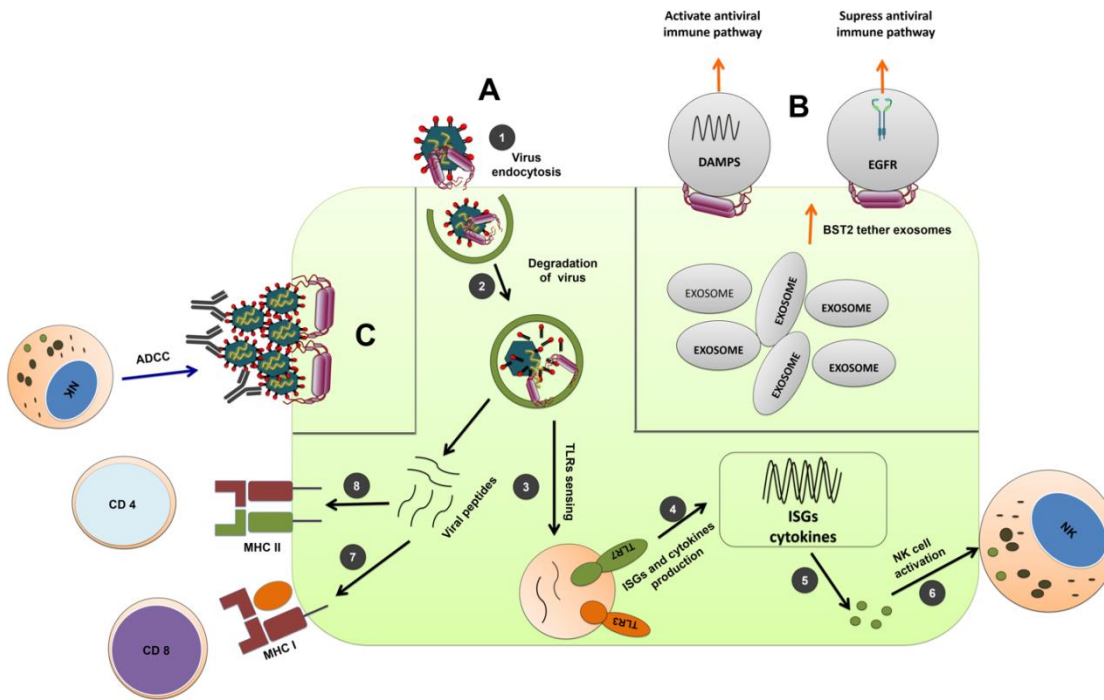


Figure 1.12. Antiviral and immunomodulatory functions of BST-2.

(A) (1) BST-2 interacts with the viral envelope and restricts cellular egress of nascent virion that, in turn, internalizes the virion through endocytosis. (2) Also, endosomally expressed BST-2 halts virion trafficking and likely allows more time for endosomal proteases to act upon and degrade the virions. (3) Endosomal degradation of viral envelope facilitates the release of genomic RNA that activates TLR3 and TLR7-mediated innate immune pathways. (4, 5) Activation of TLR3 and TLR7, along with other costimulatory molecules, can further enhance the expression of ISGs and cytokines in antigen-presenting cells. (6) Cytokines such as IL-15 can promote NK cell activation and function. (7) Proteolytically degraded viral proteins generate a plethora of viral peptides that

are often cross-presented by MHC I and can stimulate CD8 T cells. (8) Similarly, peptides loaded into MHC II can promote CD4 T cell activation. (B) In HIV-1–infected cells, the interaction between BST-2 and viral Env protein can increase the accumulation of Env at the surface of the cell. This can facilitate interactions with circulating antibodies against HIV-1 and stimulate ADCC-mediated elimination of the infected cell. (C) BST-2 can tether exosomes like viral envelopes restricting their movement. Exosomes often carry signaling molecules such as DAMPs and activated EGFR. DAMP-carrying exosomes can activate antiviral immunity, whereas EGFR-carrying exosomes can suppress it. ADCC, antibody-dependent cellular cytotoxicity; DAMPs, damage-associated molecular patterns; EGFR, epidermal growth factor receptor; HIV-1, human immunodeficiency virus-1; IL, interleukin; ISGs, IFN-stimulated genes; TLR, Toll-like receptor.

1.4.4 BST-2 and cancer

Increased BST-2 expression is shown in different cancer tissues, including ovarian, lung, head and neck, cervical, thyroid, breast, endometrial, pancreatic glioblastoma, and myeloma[120-127]. BST-2 overexpression at the early stage of multiple myeloma suggested that BST-2 might be a suitable target for cancer immune therapy. However, not all cancer types exhibit increased BST-2 expression. BST-2 expression remains unchanged in thyroid and lung adenocarcinomas and is downregulated in the liver, kidney, lung squamous, and prostate cancer relative to normal cells[123]. Almost all breast tumors express BST-2 to a certain level, and a higher expression level of BST-2 is associated with aggressive and progressive malignancy [124]. The functional significance of BST-2 expression in malignancy remains to be elucidated, but BST-2 homodimers appear critical in certain instances for the promotion of cancer cell adhesion[128]. In addition, BST-2 enhances cancer cell survival and growth by promoting proteosomal degradation of pro-apoptotic proteins, such as BIM - a member of the Bcl-2 protein family [128](figure 1.13A). Evidence indicates that the cytoplasmic tail of BST-2 is responsible for cell migration and invasion, but the detailed mechanisms underlying such functions are largely unknown[129] (figure 1.13B). Studying the role of BST-2 in different cancers is a very active area of research. In the future, we expect that novel insights will emerge regarding the relationship between BST-2 and malignancy.

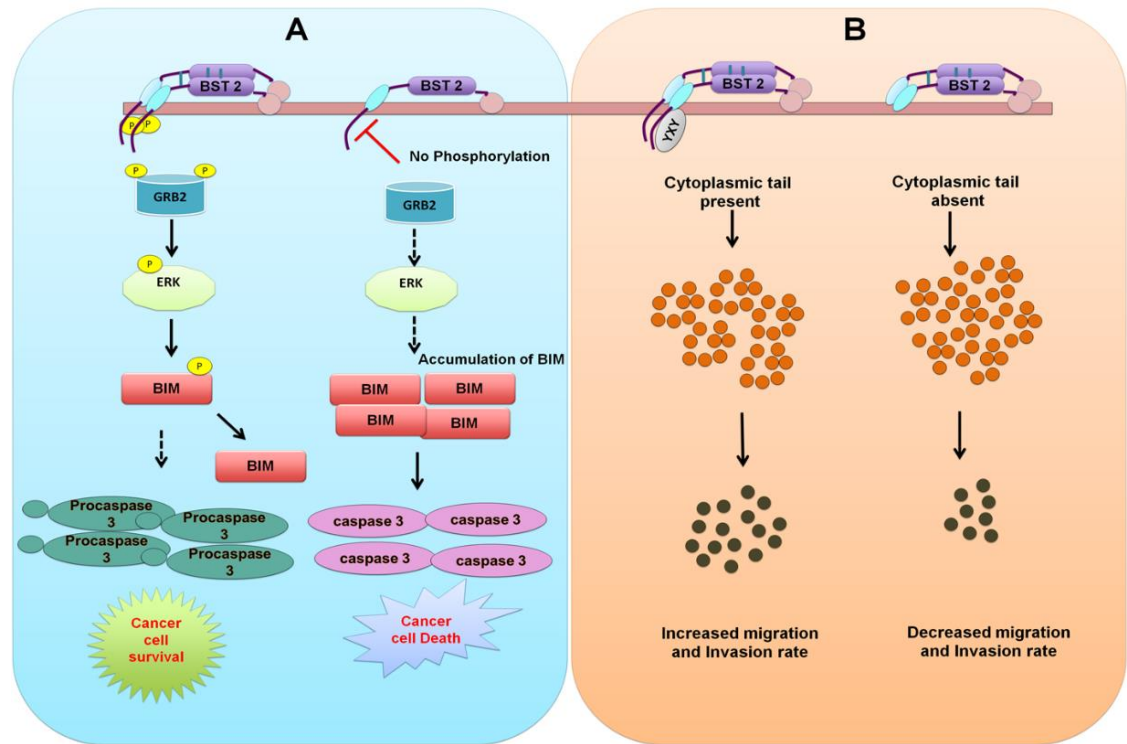


Figure 1.13. The influence of BST-2 on tumor cell survival, invasion, and migration.

(A) The dimeric form of BST-2 can facilitate cell-to-cell interactions or extracellular matrix interactions. BST-2 activation leads to phosphorylation of its cytoplasmic tail (most likely in the tyrosine-6 and tyrosine-8 positions). Phosphorylated BST-2 recruits GRB-2 and activates a kinase (unknown) that phosphorylates ERK (pERK), which, in turn, phosphorylates BIM, resulting in subsequent proteasomal degradation of BIM. In the absence of BIM, procaspase-3 is neither cleaved nor activated. This results in cancer cell survival. In a monomeric form, the cytoplasmic domain of BST-2 is not phosphorylated, which can promote apoptosis of cancer cells. (B) The YXY motif of BST-2 is responsible for cancer cell migration and invasion. In the absence of the YXY motif, cancer cells exhibit a reduced migration rate.

1.5 AIM and Scope

The discipline of virology is the most senior subject of infectious biology, and for a long time, it was focused on the pathogen itself. It is established that the host response is uniformly or more important in establishing the pathological outcome of virus infection. Based on the recent outbreak of highly infectious and deadly viruses like SARS-CoV-2, Influenza viruses, etc., vaccine failures against continuously mutating viruses, and insufficient availability of antiviral therapeutics. To overcome contagious virus pathogenesis requires that the virology and viral immunology community should focus on host-pathogen interaction and identify new antiviral targets. Our study has tried to fill this gap; this thesis contains a database VHFIDB (<https://vhfidb.com>). VHFIDB includes information on 72 viral species and their associated host factors. Moreover, it hitched with 18 other databases. VHFIDB has three comprehensive tools that are very useful for the scientific community.

After that, we have focused our study on two highly pathogenic viruses CHPV and SARS-CoV-2. CHPV is an emerging tropical virus with a high mortality rate in children below the age of 15 years. We have generated and characterized the VSV-based CHPV fluorescence pseudovirus and also described the CHPV entry mechanism. Our study found that CHPV virus entry required clathrin, lipid raft, and low pH. more interestingly, we have found that an antiviral protein BST-2 enhance CHPV virus entry, In-silico docking of the cytoplasmic domain of BST-2, and glycoprotein of CHPV reveals that YXY region of BST-2 directly binds with virus and increases virus entry. By studying the SARS-CoV-2, we have described the host factor information and critical pathways associated with the SARS-CoV-2 and CHPV virus pathogenesis. In the SARS-CoV-2 study case, we explain how a host complement system drives the thrombosis process in COVID19; we also explain and give host factors information associated with the virus entry, replication, assembly, release, and another associated pathway that regulates virus pathogenesis.

1.6. Reference:

1. Zablocki, O., E. M. Adriaenssens and D. Cowan. (2016), Diversity and Ecology of Viruses in Hyperarid Desert Soils, *Appl Environ Microbiol*, 82, 770-777 (10.1128/aem.02651-15)
2. Mushegian, A. R. (2020), Are There 10³¹ Virus Particles on Earth, or More, or Fewer?, *J Bacteriol*, 202 (10.1128/jb.00052-20)
3. Gajdusek, D. C. (1977), Unconventional viruses and the origin and disappearance of kuru, *Science*, 197, 943-960 (10.1126/science.142303)
4. van Helvoort, T. (1994), History of virus research in the twentieth century: the problem of conceptual continuity, *Hist Sci*, 32, 185-235 (10.1177/007327539403200204)
5. Blomqvist, S., L. El Bassioni, E. M. El Maamoon Nasr, A. Paananen, S. Kaijalainen, H. Asghar, E. de Gourville and M. Roivainen. (2012), Detection of imported wild polioviruses and of vaccine-derived polioviruses by environmental surveillance in Egypt, *Appl Environ Microbiol*, 78, 5406-5409 (10.1128/aem.00491-12)
6. Thèves, C., E. Crubézy and P. Biagini. (2016), History of Smallpox and Its Spread in Human Populations, *Microbiol Spectr*, 4 (10.1128/microbiolspec.PoH-0004-2014)
7. Oldstone, M. B. A. (2014) In Schmidt, T. M. (ed.), *Encyclopedia of Microbiology (Fourth Edition)*. Academic Press, Oxford, pp. 608-612.
8. Brown, F. (2003), The history of research in foot-and-mouth disease, *Virus Res*, 91, 3-7 (10.1016/s0168-1702(02)00268-x)
9. Norrby, E. (2007), Yellow fever and Max Theiler: the only Nobel Prize for a virus vaccine, *J Exp Med*, 204, 2779-2784 (10.1084/jem.20072290)
10. Ryu, W. (2016), Discovery and Classification, *Molecular Virology of Human Pathogenic Viruses*, 2017:3-20, Epub 2016 May 2016
11. Lwoff, A. and P. Tournier. (1966), The classification of viruses, *Annu Rev Microbiol*, 20, 45-74 (10.1146/annurev.mi.20.100166.000401)
12. Baltimore, D. (1971), Expression of animal virus genomes, *Bacteriol Rev*, 35, 235-241
13. Lefkowitz, E. J., D. M. Dempsey, R. C. Hendrickson, R. J. Orton, S. G. Siddell and D. B. Smith. (2018), Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV), *Nucleic Acids Res*, 46, D708-d717 (10.1093/nar/gkx932)
14. Casadevall, A. and L. A. Pirofski. (1999), Host-pathogen interactions: redefining the basic concepts of virulence and pathogenicity, *Infect Immun*, 67, 3703-3713 (10.1128/iai.67.8.3703-3713.1999)
15. Casadevall, A. and L. A. Pirofski. (2000), Host-pathogen interactions: basic concepts of microbial commensalism, colonization, infection, and disease, *Infect Immun*, 68, 6511-6518 (10.1128/iai.68.12.6511-6518.2000)
16. Enard, D., L. Cai, C. Gwennap and D. A. Petrov. (2016), Viruses are a dominant driver of protein adaptation in mammals, *Elife*, 5 (10.7554/eLife.12469)
17. Phizicky, E. M. and S. Fields. (1995), Protein-protein interactions: methods for detection and analysis, *Microbiol Rev*, 59, 94-123
18. Yasunaga, A., S. L. Hanna, J. Li, H. Cho, P. P. Rose, A. Spiridigliozzi, B. Gold, M. S. Diamond and S. Cherry. (2014), Genome-wide RNAi screen identifies broadly-

- acting host factors that inhibit arbovirus infection, *PLoS Pathog*, 10, e1003914 (10.1371/journal.ppat.1003914)
19. Panda, D., P. P. Rose, S. L. Hanna, B. Gold, K. C. Hopkins, R. B. Lyde, M. S. Marks and S. Cherry. (2013), Genome-wide RNAi screen identifies SEC61A and VCP as conserved regulators of Sindbis virus entry, *Cell Rep*, 5, 1737-1748 (10.1016/j.celrep.2013.11.028)
 20. Moser, T. S., R. G. Jones, C. B. Thompson, C. B. Coyne and S. Cherry. (2010), A kinome RNAi screen identified AMPK as promoting poxvirus entry through the control of actin dynamics, *PLoS Pathog*, 6, e1000954 (10.1371/journal.ppat.1000954)
 21. Strich, J. R. and D. S. Chertow. (2019), CRISPR-Cas Biology and Its Application to Infectious Diseases, *J Clin Microbiol*, 57 (10.1128/jcm.01307-18)
 22. Ramage, H. and S. Cherry. (2015), Virus-Host Interactions: From Unbiased Genetic Screens to Function, *Annu Rev Virol*, 2, 497-524 (10.1146/annurev-virology-100114-055238)
 23. Gillen, J. and A. Nita-Lazar. (2019), Experimental Analysis of Viral-Host Interactions, *Front Physiol*, 10, 425 (10.3389/fphys.2019.00425)
 24. Peng, X., E. Y. Chan, Y. Li, D. L. Diamond, M. J. Korth and M. G. Katze. (2009), Virus-host interactions: from systems biology to translational research, *Curr Opin Microbiol*, 12, 432-438 (10.1016/j.mib.2009.06.003)
 25. Heesterbeek, H., R. M. Anderson, V. Andreasen, S. Bansal, D. De Angelis, C. Dye, K. T. Eames, W. J. Edmunds, S. D. Frost, S. Funk *et al.* (2015), Modeling infectious disease dynamics in the complex landscape of global health, *Science*, 347, aaa4339 (10.1126/science.aaa4339)
 26. Geoghegan, J. L. and E. C. Holmes. (2017), Predicting virus emergence amid evolutionary noise, *Open Biol*, 7 (10.1098/rsob.170189)
 27. Parrish, C. R., E. C. Holmes, D. M. Morens, E. C. Park, D. S. Burke, C. H. Calisher, C. A. Laughlin, L. J. Saif and P. Daszak. (2008), Cross-species virus transmission and the emergence of new epidemic diseases, *Microbiol Mol Biol Rev*, 72, 457-470 (10.1128/mmbr.00004-08)
 28. Warren, C. J. and S. L. Sawyer. (2019), How host genetics dictates successful viral zoonosis, *PLoS Biol*, 17, e3000217 (10.1371/journal.pbio.3000217)
 29. Pincus, T., W. P. Rowe and F. Lilly. (1971), A major genetic locus affecting resistance to infection with murine leukemia viruses. II. Apparent identity to a major locus described for resistance to friend murine leukemia virus, *J Exp Med*, 133, 1234-1241 (10.1084/jem.133.6.1234)
 30. Takaoka, A. and H. Yanai. (2006), Interferon signalling network in innate defence, *Cell Microbiol*, 8, 907-922 (10.1111/j.1462-5822.2006.00716.x)
 31. Duggal, N. K. and M. Emerman. (2012), Evolutionary conflicts between viruses and restriction factors shape immunity, *Nat Rev Immunol*, 12, 687-695 (10.1038/nri3295)
 32. Van Valen, L. (1973). Theory.
 33. Daugherty, M. D. and H. S. Malik. (2012), Rules of engagement: molecular insights from host-virus arms races, *Annu Rev Genet*, 46, 677-700 (10.1146/annurev-genet-110711-155522)
 34. Damas, J., G. M. Hughes, K. C. Keough, C. A. Painter, N. S. Persky, M. Corbo, M. Hiller, K. P. Koepfli, A. R. Pfenning, H. Zhao *et al.* (2020), Broad host range of SARS-

- CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates, *Proc Natl Acad Sci U S A*, 117, 22311-22322 (10.1073/pnas.2010146117)
35. Devaux, C. A., L. Pinault, I. O. Osman and D. Raoult. (2020), Can ACE2 Receptor Polymorphism Predict Species Susceptibility to SARS-CoV-2?, *Front Public Health*, 8, 608765 (10.3389/fpubh.2020.608765)
 36. Ksiazek, T. G., D. Erdman, C. S. Goldsmith, S. R. Zaki, T. Peret, S. Emery, S. Tong, C. Urbani, J. A. Comer, W. Lim *et al.* (2003), A novel coronavirus associated with severe acute respiratory syndrome, *N Engl J Med*, 348, 1953-1966 (10.1056/NEJMoa030781)
 37. van der Hoek, L. (2007), Human coronaviruses: what do they cause?, *Antivir Ther*, 12, 651-658
 38. Zhu, N., D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu *et al.* (2020), A Novel Coronavirus from Patients with Pneumonia in China, 2019, *N Engl J Med*, 382, 727-733 (10.1056/NEJMoa2001017)
 39. Tay, M. Z., C. M. Poh, L. Rénia, P. A. MacAry and L. F. P. Ng. (2020), The trinity of COVID-19: immunity, inflammation and intervention, *Nat Rev Immunol*, 20, 363-374 (10.1038/s41577-020-0311-8)
 40. Hu, B., H. Guo, P. Zhou and Z. L. Shi. (2021), Characteristics of SARS-CoV-2 and COVID-19, *Nat Rev Microbiol*, 19, 141-154 (10.1038/s41579-020-00459-7)
 41. Tiwari, R., A. R. Mishra, F. Mikaeloff, S. Gupta, A. Mirazimi, S. N. Byraredy, U. Neogi and D. Nayak. (2020), In silico and in vitro studies reveal complement system drives coagulation cascade in SARS-CoV-2 pathogenesis, *Computational and Structural Biotechnology Journal*, 18, 3734-3744 (<https://doi.org/10.1016/j.csbj.2020.11.005>)
 42. Wang, C., P. W. Horby, F. G. Hayden and G. F. Gao. (2020), A novel coronavirus outbreak of global health concern, *Lancet*, 395, 470-473 (10.1016/s0140-6736(20)30185-9)
 43. Wang, H., X. Li, T. Li, S. Zhang, L. Wang, X. Wu and J. Liu. (2020), The genetic sequence, origin, and diagnosis of SARS-CoV-2, *Eur J Clin Microbiol Infect Dis*, 39, 1629-1635 (10.1007/s10096-020-03899-4)
 44. Lu, R., X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu *et al.* (2020), Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, *The Lancet*, 395, 565-574 ([https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8))
 45. V'kovski, P., A. Kratzel, S. Steiner, H. Stalder and V. Thiel. (2021), Coronavirus biology and replication: implications for SARS-CoV-2, *Nature Reviews Microbiology*, 19, 155-170 (10.1038/s41579-020-00468-6)
 46. Romano, M., A. Ruggiero, F. Squeglia, G. Maga and R. Berisio. (2020), A Structural View of SARS-CoV-2 RNA Replication Machinery: RNA Synthesis, Proofreading and Final Capping, *Cells*, 9 (10.3390/cells9051267)
 47. Fehr, A. R. and S. Perlman. (2015), Coronaviruses: an overview of their replication and pathogenesis, *Methods Mol Biol*, 1282, 1-23 (10.1007/978-1-4939-2438-7_1)
 48. Tiwari, R., A. R. Mishra, A. Gupta and D. Nayak. (2021), Structural similarity-based prediction of host factors associated with SARS-CoV-2 infection and pathogenesis, *J Biomol Struct Dyn*, 1-12 (10.1080/07391102.2021.1874532)
 49. Wu, R., L. Wang, H. D. Kuo, A. Shannar, R. Peter, P. J. Chou, S. Li, R. Hudlikar, X. Liu, Z. Liu *et al.* (2020), An Update on Current Therapeutic Drugs Treating COVID-19, *Curr Pharmacol Rep*, 1-15 (10.1007/s40495-020-00216-7)

50. Weston, S., C. M. Coleman, R. Haupt, J. Logue, K. Matthews, Y. Li, H. M. Reyes, S. R. Weiss and M. B. Frieman. (2020), Broad Anti-coronavirus Activity of Food and Drug Administration-Approved Drugs against SARS-CoV-2 In Vitro and SARS-CoV In Vivo, *J Virol*, 94 (10.1128/jvi.01218-20)
51. McKee, D. L., A. Sternberg, U. Stange, S. Laufer and C. Naujokat. (2020), Candidate drugs against SARS-CoV-2 and COVID-19, *Pharmacol Res*, 157, 104859 (10.1016/j.phrs.2020.104859)
52. Chien, M., T. K. Anderson, S. Jockusch, C. Tao, X. Li, S. Kumar, J. J. Russo, R. N. Kirchdoerfer and J. Ju. (2020), Nucleotide Analogues as Inhibitors of SARS-CoV-2 Polymerase, a Key Drug Target for COVID-19, *J Proteome Res*, 19, 4690-4697 (10.1021/acs.jproteome.0c00392)
53. Ghosh, S., K. Dutta and A. Basu. (2013), Chandipura virus induces neuronal death through Fas-mediated extrinsic apoptotic pathway, *J Virol*, 87, 12398-12406 (10.1128/jvi.01864-13)
54. Sapkal, G. N., P. M. Sawant and D. T. Mourya. (2018), Chandipura Viral Encephalitis: A Brief Review, *Open Virol J*, 12, 44-51 (10.2174/1874357901812010044)
55. Marriott, A. C. (2005), Complete genome sequences of Chandipura and Isfahan vesiculoviruses, *Arch Virol*, 150, 671-680 (10.1007/s00705-004-0452-2)
56. Bhatt, P. N. and F. M. Rodrigues. (1967), Chandipura: a new Arbovirus isolated in India from patients with febrile illness, *Indian J Med Res*, 55, 1295-1305
57. Rodrigues, J. J., P. B. Singh, D. S. Dave, R. Prasan, V. Ayachit, B. H. Shaikh and K. M. Pavri. (1983), Isolation of Chandipura virus from the blood in acute encephalopathy syndrome, *Indian J Med Res*, 77, 303-307
58. Sudeep, A. B., Y. K. Gurav and V. P. Bondre. (2016), Changing clinical scenario in Chandipura virus infection, *Indian J Med Res*, 143, 712-721 (10.4103/0971-5916.191929)
59. Rao, B. L., A. Basu, N. S. Wairagkar, M. M. Gore, V. A. Arankalle, J. P. Thakare, R. S. Jadi, K. A. Rao and A. C. Mishra. (2004), A large outbreak of acute encephalitis with high fatality rate in children in Andhra Pradesh, India, in 2003, associated with Chandipura virus, *Lancet*, 364, 869-874 (10.1016/s0140-6736(04)16982-1)
60. Chadha, M. S., V. A. Arankalle, R. S. Jadi, M. V. Joshi, J. P. Thakare, P. V. Mahadev and A. C. Mishra. (2005), An outbreak of Chandipura virus encephalitis in the eastern districts of Gujarat state, India, *Am J Trop Med Hyg*, 73, 566-570
61. Gurav, Y. K., B. V. Tandale, R. S. Jadi, R. S. Gunjekar, S. S. Tikute, A. V. Jamgaonkar, R. K. Khadse, S. V. Jalgaonkar, V. A. Arankalle and A. C. Mishra. (2010), Chandipura virus encephalitis outbreak among children in Nagpur division, Maharashtra, 2007, *Indian J Med Res*, 132, 395-399
62. Dwibedi, B., J. Sabat, R. K. Hazra, A. Kumar, D. S. Dinesh and S. K. Kar. (2015), Chandipura virus infection causing encephalitis in a tribal population of Odisha in eastern India, *Natl Med J India*, 28, 185-187
63. Peiris, J. S., W. P. Dittus and C. B. Ratnayake. (1993), Seroepidemiology of dengue and other arboviruses in a natural population of toque macaques (*Macaca sinica*) at Polonnaruwa, Sri Lanka, *J Med Primatol*, 22, 240-245
64. Kemp, G. E. (1975), Viruses other than arenaviruses from West African wild mammals. Factors affecting transmission to man and domestic animals, *Bull World Health Organ*, 52, 615-620

65. Marriott, A. C. and C. A. Hornsey. (2011), Reverse genetics system for Chandipura virus: tagging the viral matrix protein with green fluorescent protein, *Virus Res*, 160, 166-172 (10.1016/j.virusres.2011.06.007)
66. Cherian, S. S., R. S. Gunjekar, A. Banerjee, S. Kumar and V. A. Arankalle. (2012), Whole genomes of Chandipura virus isolates and comparative analysis with other rhabdoviruses, *PLoS One*, 7, e30315 (10.1371/journal.pone.0030315)
67. Ghosh, S. and A. Basu. (2017), Neuropathogenesis by Chandipura virus: An acute encephalitis syndrome in India, *Natl Med J India*, 30, 21-25
68. Roy, S., D. Pavitrakar, R. Gunjekar, V. M. Ayachit, V. P. Bondre and G. N. Sapkal. (2016), Monocytes and B cells support active replication of Chandipura virus, *BMC Infectious Diseases*, 16, 487 (10.1186/s12879-016-1794-6)
69. Jadi, R. S., A. B. Sudeep, P. V. Barde, V. A. Arankalle and A. C. Mishra. (2011), Development of an inactivated candidate vaccine against Chandipura virus (Rhabdoviridae: Vesiculovirus), *Vaccine*, 29, 4613-4617 (10.1016/j.vaccine.2011.04.063)
70. Goto, T., S. J. Kennel, M. Abe, M. Takishita, M. Kosaka, A. Solomon and S. Saito. (1994), A novel membrane antigen selectively expressed on terminally differentiated human B cells, *Blood*, 84, 1922-1930
71. Ishikawa, J., T. Kaisho, H. Tomizawa, B. O. Lee, Y. Kobune, J. Inazawa, K. Oritani, M. Itoh, T. Ochi, K. Ishihara *et al.* (1995), Molecular cloning and chromosomal mapping of a bone marrow stromal cell surface gene, BST2, that may be involved in pre-B-cell growth, *Genomics*, 26, 527-534
72. Jones, P. H., H. V. Mehta, M. Maric, R. J. Roller and C. M. Okeoma. (2012), Bone marrow stromal cell antigen 2 (BST-2) restricts mouse mammary tumor virus (MMTV) replication in vivo, *Retrovirology*, 9, 10 (10.1186/1742-4690-9-10)
73. Erikson, E., T. Adam, S. Schmidt, J. Lehmann-Koch, B. Over, C. Goffinet, C. Harter, I. Bekeredjian-Ding, S. Sertel, F. Lasitschka *et al.* (2011), In vivo expression profile of the antiviral restriction factor and tumor-targeting antigen CD317/BST-2/HM1.24/tetherin in humans, *Proc Natl Acad Sci U S A*, 108, 13688-13693 (10.1073/pnas.1101684108)
74. Kupzig, S., V. Korolchuk, R. Rollason, A. Sugden, A. Wilde and G. Banting. (2003), Bst-2/HM1.24 is a raft-associated apical membrane protein with an unusual topology, *Traffic*, 4, 694-709
75. Hinz, A., N. Miguët, G. Natrajan, Y. Usami, H. Yamanaka, P. Renesto, B. Hartlieb, A. A. McCarthy, J. P. Simorre, H. Gottlinger *et al.* (2010), Structural basis of HIV-1 tethering to membranes by the BST-2/tetherin ectodomain, *Cell Host Microbe*, 7, 314-323 (10.1016/j.chom.2010.03.005)
76. Masuyama, N., T. Kuronita, R. Tanaka, T. Muto, Y. Hirota, A. Takigawa, H. Fujita, Y. Aso, J. Amano and Y. Tanaka. (2009), HM1.24 is internalized from lipid rafts by clathrin-mediated endocytosis through interaction with alpha-adaptin, *J Biol Chem*, 284, 15927-15941 (10.1074/jbc.M109.005124)
77. Galao, R. P., A. Le Tortorec, S. Pickering, T. Kueck and S. J. Neil. (2012), Innate sensing of HIV-1 assembly by Tetherin induces NFkappaB-dependent proinflammatory responses, *Cell Host Microbe*, 12, 633-644 (10.1016/j.chom.2012.10.007)
78. Tokarev, A., M. Skasko, K. Fitzpatrick and J. Guatelli. (2009), Antiviral activity of the interferon-induced cellular protein BST-2/tetherin, *AIDS Res Hum Retroviruses*, 25, 1197-1210 (10.1089/aid.2009.0253)

10.1089/aid.2009.9991)

79. Andrew, A. J., E. Miyagi, S. Kao and K. Strebel. (2009), The formation of cysteine-linked dimers of BST-2/tetherin is important for inhibition of HIV-1 virus release but not for sensitivity to Vpu, *Retrovirology*, 6, 80 (10.1186/1742-4690-6-80)
80. Swiecki, M., S. M. Scheaffer, M. Allaire, D. H. Fremont, M. Colonna and T. J. Brett. (2011), Structural and biophysical analysis of BST-2/tetherin ectodomains reveals an evolutionary conserved design to inhibit virus release, *J Biol Chem*, 286, 2987-2997 (10.1074/jbc.M110.190538)
81. Hammonds, J., J. J. Wang, H. Yi and P. Spearman. (2010), Immunoelectron microscopic evidence for Tetherin/BST2 as the physical bridge between HIV-1 virions and the plasma membrane, *PLoS Pathog*, 6, e1000749 (10.1371/journal.ppat.1000749)
82. Neil, S. J., T. Zang and P. D. Bieniasz. (2008), Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu, *Nature*, 451, 425-430 (10.1038/nature06553)
83. Van Damme, N., D. Goff, C. Katsura, R. L. Jorgenson, R. Mitchell, M. C. Johnson, E. B. Stephens and J. Guatelli. (2008), The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein, *Cell Host Microbe*, 3, 245-252 (10.1016/j.chom.2008.03.001)
84. Perez-Caballero, D., T. Zang, A. Ebrahimi, M. W. McNatt, D. A. Gregory, M. C. Johnson and P. D. Bieniasz. (2009), Tetherin inhibits HIV-1 release by directly tethering virions to cells, *Cell*, 139, 499-511 (10.1016/j.cell.2009.08.039)
85. Neil, S. J., S. W. Eastman, N. Jouvenet and P. D. Bieniasz. (2006), HIV-1 Vpu promotes release and prevents endocytosis of nascent retrovirus particles from the plasma membrane, *PLoS Pathog*, 2, e39 (10.1371/journal.ppat.0020039)
86. Miyakawa, K., A. Ryo, T. Murakami, K. Ohba, S. Yamaoka, M. Fukuda, J. Guatelli and N. Yamamoto. (2009), BCA2/Rabring7 promotes tetherin-dependent HIV-1 restriction, *PLoS Pathog*, 5, e1000700 (10.1371/journal.ppat.1000700)
87. Evans, D. T., R. Serra-Moreno, R. K. Singh and J. C. Guatelli. (2010), BST-2/tetherin: a new component of the innate immune response to enveloped viruses, *Trends Microbiol*, 18, 388-396 (10.1016/j.tim.2010.06.010)
88. Arias, J. F., Y. Iwabu and K. Tokunaga. (2011), Structural Basis for the Antiviral Activity of BST-2/Tetherin and Its Viral Antagonism, *Front Microbiol*, 2, 250 (10.3389/fmicb.2011.00250)
89. Swiecki, M., N. S. Omattage and T. J. Brett. (2013), BST-2/tetherin: structural biology, viral antagonism, and immunobiology of a potent host antiviral factor, *Mol Immunol*, 54, 132-139 (10.1016/j.molimm.2012.11.008)
90. Mahauad-Fernandez, W. D. and C. M. Okeoma. (2016), The role of BST-2/Tetherin in host protection and disease manifestation, *Immun Inflamm Dis*, 4, 4-23 (10.1002/iid3.92)
91. Sarojini, S., T. Theofanis and C. S. Reiss. (2011), Interferon-induced tetherin restricts vesicular stomatitis virus release in neurons, *DNA Cell Biol*, 30, 965-974 (10.1089/dna.2011.1384)
92. Matsuda, A., Y. Suzuki, G. Honda, S. Muramatsu, O. Matsuzaki, Y. Nagano, T. Doi, K. Shimotohno, T. Harada, E. Nishida *et al.* (2003), Large-scale identification and characterization of human genes that activate NF-kappaB and MAPK signaling pathways, *Oncogene*, 22, 3307-3318 (10.1038/sj.onc.1206406)

93. Tokarev, A., M. Suarez, W. Kwan, K. Fitzpatrick, R. Singh and J. Guatelli. (2013), Stimulation of NF-kappaB activity by the HIV restriction factor BST2, *J Virol*, 87, 2046-2057 (10.1128/jvi.02272-12)
94. Cocka, L. J. and P. Bates. (2012), Identification of alternatively translated Tetherin isoforms with differing antiviral and signaling activities, *PLoS Pathog*, 8, e1002931 (10.1371/journal.ppat.1002931)
95. Liu, T., L. Zhang, D. Joo and S.-C. Sun. (2017), NF-κB signaling in inflammation, *Signal Transduction And Targeted Therapy*, 2, 17023 (10.1038/sigtrans.2017.23)
96. Moynagh, P. N. (2005), The NF-kappaB pathway, *J Cell Sci*, 118, 4589-4592 (10.1242/jcs.02579)
97. Chan, Y. K. and M. U. Gack. (2016), Viral evasion of intracellular DNA and RNA sensing, *Nat Rev Microbiol*, 14, 360-373 (10.1038/nrmicro.2016.45)
98. Liu, B., M. Zhang, H. Chu, H. Zhang, H. Wu, G. Song, P. Wang, K. Zhao, J. Hou, X. Wang *et al.* (2017), The ubiquitin E3 ligase TRIM31 promotes aggregation and activation of the signaling adaptor MAVS through Lys63-linked polyubiquitination, *Nat Immunol*, 18, 214-224 (10.1038/ni.3641)
99. Cao, X. (2015), Self-regulation and cross-regulation of pattern-recognition receptor signalling in health and disease, *Nature Reviews Immunology*, 16, 35 (10.1038/nri.2015.8)
100. Goubau, D., S. Deddouch and C. Reis e Sousa. (2013), Cytosolic sensing of viruses, *Immunity*, 38, 855-869 (10.1016/j.immuni.2013.05.007)
101. Takeuchi, O. and S. Akira. (2010), Pattern recognition receptors and inflammation, *Cell*, 140, 805-820 (10.1016/j.cell.2010.01.022)
102. Schneider, W. M., M. D. Chevillotte and C. M. Rice. (2014), Interferon-stimulated genes: a complex web of host defenses, *Annu Rev Immunol*, 32, 513-545 (10.1146/annurev-immunol-032713-120231)
103. Gota, C. and L. Calabrese. (2003), Induction of Clinical Autoimmune Disease by Therapeutic Interferon-α, *Autoimmunity*, 36, 511-518 (10.1080/08916930310001605873)
104. Lin, Q., C. Dong and M. D. Cooper. (1998), Impairment of T and B cell development by treatment with a type I interferon, *J Exp Med*, 187, 79-87
105. Colonna, M., G. Trinchieri and Y. J. Liu. (2004), Plasmacytoid dendritic cells in immunity, *Nat Immunol*, 5, 1219-1226 (10.1038/ni1141)
106. Honda, K. and T. Taniguchi. (2006), IRFs: master regulators of signalling by Toll-like receptors and cytosolic pattern-recognition receptors, *Nat Rev Immunol*, 6, 644-658 (10.1038/nri1900)
107. Cao, W., L. Bover, M. Cho, X. Wen, S. Hanabuchi, M. Bao, D. B. Rosen, Y. H. Wang, J. L. Shaw, Q. Du *et al.* (2009), Regulation of TLR7/9 responses in plasmacytoid dendritic cells by BST2 and ILT7 receptor interaction, *J Exp Med*, 206, 1603-1614 (10.1084/jem.20090547)
108. Cho, M., K. Ishida, J. Chen, J. Ohkawa, W. Chen, S. Namiki, A. Kotaki, N. Arai, K. Arai and Y. Kamogawa-Schifter. (2008), SAGE library screening reveals ILT7 as a specific plasmacytoid dendritic cell marker that regulates type I IFN production, *Int Immunol*, 20, 155-164 (10.1093/intimm/dxm127)
109. Janovec, V., B. Aouar, A. Font-Haro, T. Hofman, K. Trejbalova, J. Weber, L. Chaperot, J. Plumas, D. Olive, P. Dubreuil *et al.* (2018), The MEK1/2-ERK Pathway Inhibits Type I IFN Production in Plasmacytoid Dendritic Cells, *Front Immunol*, 9, 364 (10.3389/fimmu.2018.00364)

110. Brown, D., J. Trowsdale and R. Allen. (2004), The LILR family: modulators of innate and adaptive immune pathways in health and disease, *Tissue Antigens*, 64, 215-225 (10.1111/j.0001-2815.2004.00290.x)
111. Jin, S., S. Tian, M. Luo, W. Xie, T. Liu, T. Duan, Y. Wu and J. Cui. (2017), Tetherin Suppresses Type I Interferon Signaling by Targeting MAVS for NDP52-Mediated Selective Autophagic Degradation in Human Cells, *Mol Cell*, 68, 308-322.e304 (10.1016/j.molcel.2017.09.005)
112. Deretic, V., T. Saitoh and S. Akira. (2013), Autophagy in infection, inflammation and immunity, *Nat Rev Immunol*, 13, 722-737 (10.1038/nri3532)
113. Thurston, T. L., G. Ryzhakov, S. Bloor, N. von Muhlinen and F. Randow. (2009), The TBK1 adaptor and autophagy receptor NDP52 restricts the proliferation of ubiquitin-coated bacteria, *Nat Immunol*, 10, 1215-1221 (10.1038/ni.1800)
114. Inomata, M., S. Niida, K. Shibata and T. Into. (2012), Regulation of Toll-like receptor signaling by NDP52-mediated selective autophagy is normally inactivated by A20, *Cell Mol Life Sci*, 69, 963-979 (10.1007/s00018-011-0819-y)
115. Swiecki, M., Y. Wang, S. Gilfillan, D. J. Lenschow and M. Colonna. (2012), Cutting edge: paradoxical roles of BST2/tetherin in promoting type I IFN response and viral infection, *J Immunol*, 188, 2488-2492 (10.4049/jimmunol.1103145)
116. Barrett, B. S., D. S. Smith, S. X. Li, K. Guo, K. J. Hasenkrug and M. L. Santiago. (2012), A single nucleotide polymorphism in tetherin promotes retrovirus restriction in vivo, *PLoS Pathog*, 8, e1002596 (10.1371/journal.ppat.1002596)
117. Li, S. X., B. S. Barrett, K. J. Heilman, R. J. Messer, R. A. Liberatore, P. D. Bieniasz, G. Kassiotis, K. J. Hasenkrug and M. L. Santiago. (2014), Tetherin promotes the innate and adaptive cell-mediated immune response against retrovirus infection in vivo, *J Immunol*, 193, 306-316 (10.4049/jimmunol.1400490)
118. Gibbert, K., S. Francois, A. M. Sigmund, M. S. Harper, B. S. Barrett, C. J. Kirchning, M. Lu, M. L. Santiago and U. Dittmer. (2014), Friend retrovirus drives cytotoxic effectors through Toll-like receptor 3, *Retrovirology*, 11, 126 (10.1186/s12977-014-0126-4)
119. Urata, S. and E. Kenyon. (2018), BST-2 controls T cell proliferation and exhaustion by shaping the early distribution of a persistent viral infection, 14, e1007172 (10.1371/journal.ppat.1007172)
120. Milutin Gasperov, N., S. A. Farkas, T. K. Nilsson and M. Grce. (2014), Epigenetic activation of immune genes in cervical cancer, *Immunol Lett*, 162, 256-257 (10.1016/j.imlet.2014.09.019)
121. Wang, W., Y. Nishioka, S. Ozaki, A. Jalili, S. Abe, S. Kakiuchi, M. Kishuku, K. Minakuchi, T. Matsumoto and S. Sone. (2009), HM1.24 (CD317) is a novel target against lung cancer for immunotherapy using anti-HM1.24 antibody, *Cancer Immunol Immunother*, 58, 967-976 (10.1007/s00262-008-0612-4)
122. Fang, K. H., H. K. Kao, L. M. Chi, Y. Liang, S. C. Liu, C. Hseuh, C. T. Liao, T. C. Yen, J. S. Yu and K. P. Chang. (2014), Overexpression of BST2 is associated with nodal metastasis and poorer prognosis in oral cavity cancer, *Laryngoscope*, 124, E354-360 (10.1002/lary.24700)
123. Mahauad-Fernandez, W. D., N. C. Borchering, W. Zhang and C. M. Okeoma. (2015), Bone marrow stromal antigen 2 (BST-2) DNA is demethylated in breast tumors and breast cancer cells, *PLoS One*, 10, e0123931 (10.1371/journal.pone.0123931)

124. Mahauad-Fernandez, W. D., K. A. DeMali, A. K. Olivier and C. M. Okeoma. (2014), Bone marrow stromal antigen 2 expressed in cancer cells promotes mammary tumor growth and metastasis, *Breast Cancer Res*, 16, 493 (10.1186/s13058-014-0493-8)
125. Yokoyama, T., T. Enomoto, S. Serada, A. Morimoto, S. Matsuzaki, Y. Ueda, K. Yoshino, M. Fujita, S. Kyo, K. Iwahori *et al.* (2013), Plasma membrane proteomics identifies bone marrow stromal antigen 2 as a potential therapeutic target in endometrial cancer, *Int J Cancer*, 132, 472-484 (10.1002/ijc.27679)
126. Wainwright, D. A., I. V. Balyasnikova, Y. Han and M. S. Lesniak. (2011), The expression of BST2 in human and experimental mouse brain tumors, *Exp Mol Pathol*, 91, 440-446 (10.1016/j.yexmp.2011.04.012)
127. Tai, Y. T., H. M. Horton, S. Y. Kong, E. Pong, H. Chen, S. Cemerski, M. J. Bennett, D. H. Nguyen, S. Karki, S. Y. Chu *et al.* (2012), Potent in vitro and in vivo activity of an Fc-engineered humanized anti-HM1.24 antibody against multiple myeloma via augmented effector function, *Blood*, 119, 2074-2082 (10.1182/blood-2011-06-364521)
128. Mahauad-Fernandez, W. D. and C. M. Okeoma. (2017), Cysteine-linked dimerization of BST-2 confers anoikis resistance to breast cancer cells by negating proapoptotic activities to promote tumor cell survival and growth, *Cell Death Dis*, 8, e2687 (10.1038/cddis.2017.68)
129. Naushad, W., W. D. Mahauad-Fernandez and C. M. Okeoma. (2017), Structural determinant of BST-2-mediated regulation of breast cancer cell motility: a role for cytoplasmic tail tyrosine residues, *Oncotarget*, 8, 110221-110233 (10.18632/oncotarget.22753)

CHAPTER 2

Chapter 2

Material, methods, and instrumentation

2.1 Materials

2.1.1 Chemicals

All oligos used in this study were procured from Sigma-Aldrich Chemicals Ltd. (St. Louis, MO, USA) and Integrated DNA Technologies (Iowa, United States), respectively (Appendix B). The chemical agent used in this study are as follows: Trizma base, Glycine ($C_2H_5NO_2$), Ethylene diamine tetra acetic acid (EDTA), Potassium hydrogen phosphate (K_2HPO_4), Potassium dihydrogen phosphate (KH_2PO_4), Sodium hydrogen phosphate (Na_2HPO_4), Sodium dihydrogen phosphate (NaH_2PO_4), Ethidium bromide (EtBr), Potassium chloride (KCl), Sodium chloride (NaCl), Magnesium chloride ($MgCl_2$), Calcium chloride ($CaCl_2$), Sodium hydroxide (NaOH), Sodium thiosulphate ($Na_2S_2O_3$), Sodium hydrogen carbonate ($NaHCO_3$), Silver nitrate ($AgNO_3$), Sodium carbonate (Na_2CO_3), Dimethyl formamide (C_3H_7NO), Sodium dodecyl sulphate or Sodium lauryl sulphate (SDS), Commassie brilliant blue G-250, β -mercaptoethanol ($HOCH_2CH_2SH$), Agarose, Luria-Bertani agar (LB-agar), Luria-Bertani broth (LB-broth), Ampicillin, Kanamycin, Acrylamide (C_3H_5NO), N, N'-Methylene bisacrylamide (MBAA), Triethylenetetramine or Tetramethylethylene diamine (TEMED), Ammonium per sulphate ($(NH_4)_2S_2O_8$), Polysorbate 20 or Tween 20 ($C_{58}H_{114}O_{26}$), non-fat dry skimmed milk powder, Glycerol ($C_3H_8O_3$), bromophenol blue or 3',3'',5',5''-tetrabromophenolsulfonphthalein ($C_{19}H_{10}Br_4O_5S$), Bovine serum albumin (BSA), 4',6-Diamidino-2-Phenylindole, Dihydrochloride ($C_{16}H_{15}N_5$), Paraformaldehyde ($OH(CH_2O)_nH$ ($n=8-100$)), Isopropyl- β -D-thiogalactopyranoside ($C_9H_{18}O_5S$), Imidazole ($C_3H_4N_2$), 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside ($C_{14}H_{15}BrClNO_6$), Nonidet P-40 (tergitol), Triton X-100 $C_{14}H_{22}O(C_2H_4O)_n$, ethylene glycol-bis (β -aminoethyl ether)-N,N,N',N'-tetraacetic acid ($C_{14}H_{24}N_2O_{10}$), Glycerol ($C_3H_8O_3$), Xylene cyanol ($C_{25}H_{27}N_2NaO_6S_2$), RNase A, Methanol (CH_3OH), Glacial acetic acid (CH_3COOH), Formalin or formaldehyde (CH_2O ($H-CHO$)), Ethanol (C_2H_5OH), Fetal bovine serum

(FBS), Dimethyl sulphoxide (C_2H_6OS), 2- mercaptoethanol ($HOCH_2CH_2SH$), Glutamine ($C_5H_{10}N_2O_3$), Pyruvic acid ($C_3H_4O_3$), Penicillin G streptomycin, Phenol (C_6H_5OH), Chloroform ($CHCl_3$), Isoamyl alcohol ($C_5H_{12}O$), N-lauroylsarcosine ($C_{15}H_{28}NNaO_3$), phenylmethylsulfonylfluoride ($C_7H_7FO_2S$), Protease inhibitor cocktail, HEPES buffer ($C_8H_{18}N_2O_4S$), Tamoxifen (Nolvadex), Lipofectamine 2000, X-tremeGENE HP, Chlorpromazine, Ammonium chloride, Methyl- β -cyclodextrin, Sucrose. All chemicals were from molecular biology grade and procured from different makes like Sigma Aldrich Chemical Pvt. Ltd., Invitrogen Pvt. Ltd., MP Biomedical USA, Himedia Pvt. Ltd. India, Alpha aesar Pvt. Ltd. and, Sisco Research Lab Pvt. Ltd. India,

2.1.2 Cloning vectors and competent cells

The mammalian expression vectors used are pCAG and pcDNA 3.1, kindly gifted by Prof. Asit K. Pattnaik (University of Nebraska-Lincoln, USA). Plasmid related to the retrovirus system and BST-2 k/o, PQCXIP-BST2-HA, pUMVC, and pSpCas9(B.B.)-2A-Puro (PX459) V2.0 were purchased from Addgene. BST-2 shRNA plasmid was procured from Sigma-Aldrich Chemicals Ltd. Bacterial competent cells used, *E. coli* DH5 α were obtained from Himedia Pvt. Ltd. India.

2.1.3 Cell lines

The human embryonic kidney cells 293T (HEK 293T), Human embryonic kidney 293 (HEK 293), Vero, and Vero E6 cells were obtained from NCCS Pune. Baby hamster kidney cells (BHK-21) were obtained from Sigma-Aldrich Chemicals Ltd.

2.1.4 Enzymes, Cell culture media, Antibodies, and Kits

The PCR purification kit, Plasmid isolation kit, and Gel extraction kit were purchased from QIAGEN and Favorgen Biotech Corporation. T4 DNA ligase, 10X Buffer with Mg, dNTPs, Taq DNA Polymerase, Q5 DNA polymerase, 100bp DNA ladder, 1kb DNA ladder, unstained protein marker, prestained protein marker, and all restriction enzymes were purchased from New England Biolabs, USA. TRIzol reagents, Superscript IV cDNA synthesis kit, Anti Flag antibody, Cell culture media (DMEM, MEM OptiMEM, etc.), and SYBR green master mix were purchased from Invitrogen Pvt. Mouse monoclonal antibody against BST-2 was purchased from Santa Cruz Biotechnology, CHPV antibody was

obtained from Abegenex, Rabbit polyclonal anti-actin antibody, anti-rabbit secondary antibody, anti-mouse secondary antibody were obtained from Sigma Aldrich Chemical Pvt. Ltd.

2.2 Methods and Instrumentation

2.2.1 E. coli competent cell preparation

Competent cells were prepared by using a 50 mM CaCl_2 solution. *E. coli* was grown on Luria Broth (L.B.) agar plates. Single colonies from overnight grown cultures plate were transferred to L.B. broth medium. Cells were incubated in a bacterial incubator shaker at 250 rpm at 37 °C. After the optical density (O.D.) reached 0.4 nm, the flask was placed on ice for 20 min, and then transferred to sterile centrifuge tubes and spinned at 6000 rpm for 5 min at 4 °C. Next, the pellets were gently resuspended in ice-cold sterile CaCl_2 , kept on ice for 20 min, and spinned at 6000 rpm for 5 min at 4 °C. The pellet was resuspended in 2 ml of CaCl_2 (100 mM), containing 15% glycerol, kept on ice for 20 min, and at the end, 50 μl each was aliquoted into Eppendorf tubes and put in liquid nitrogen until they were frozen and stored them in -80 °C.

2.2.2 Transformation of E. coli competent cells

We have used *E. coli* strains DH5 α for the transformation of a specific plasmid. For each transformation, 1-100 ng of plasmid DNA were added to 50 μl of competent cells. And incubated on ice for 20 min, followed by heat shock at 42 °C for 90 seconds in a water bath and set it on ice for the next 5 min. The cells were allowed to recover in 1 ml Luria-broth (L.B. broth: 1 % Bacto-Tryptone, 1 % NaCl, and 0.5 % Bacto-Yeast extract) and then incubated for 90 min at 37 °C and shaken at 220 rpm in a bacterial shaker incubator. Transformed *E. coli* cells were plated on the appropriate antibiotic-containing agar plates and incubated at 37 °C overnight in a bacterial incubator to select the transformants (Figure 2.1).

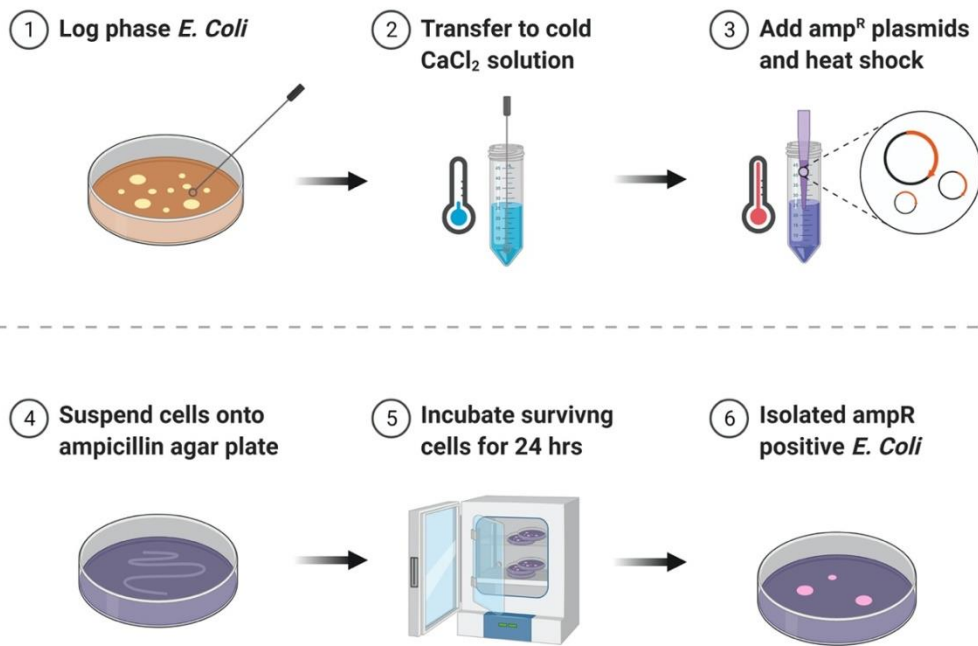


Figure 2.1. Schematic representation of transformation process.

2.2.3 Plasmid DNA isolation

Plasmid DNA was isolated from transformed cells using QIAGEN and Favorgen kits. The single isolated colony from transformed cells was inoculated in 50 ml L.B. broth containing appropriate antibiotics and incubated at 250 rpm at 37 °C for overnight. It was then centrifuged, and following the kit manufactures protocol, the plasmid DNA was isolated. The recovered plasmid DNA was then stored at -20 °C.

2.2.4 Polymerase chain reaction

Gene sequences were amplified from genomic DNA or cDNA by using PCR. Plasmid DNA was used as a positive control. A PCR reaction mixture of 20-50 µl containing 10X PCR Buffer with Mg⁺⁺(NEB), 2 mM dNTPs (Sigma), Primers (Forward and Reverse), Taq DNA Polymerase or Q5 DNA Polymerase (NEB). The thermal cycle was programmed for 5 min at 95 °C as initial denaturation, followed by 30 cycles of 30 sec at 95 °C for denaturation, annealing temperature based on primer, extension based on template size, and a final extension at 72 °C for 10 min.

2.2.5 PCR purification

The PCR amplified product was purified by a PCR purification kit (Thermo scientific) and a PCI manual method. Further, PCR purified DNA concentration and band size were checked on agarose gel electrophoresis. PCR purified DNA was stored at -20 °C and used for further cloning.

2.2.6 Restriction digestion

To perform restriction digestion, we have taken appropriate amounts of plasmid DNA as a vector and PCR product as an insert. DNA was mixed with sterile ddH₂O to a final volume of 50 µl. Further, 10X cut smart buffer was added in a 1X dilution. Finally, the mixture was supplemented with 1U of restriction enzyme per µg of DNA. Digestion was carried out for 3 hours for plasmid DNA and PCR products at an appropriate temperature as recommended by the manufacturers in a water bath. The samples were gel electrophoresed and gel purified as described below.

2.2.7 Agarose gel electrophoresis

Presence of amplified PCR product or digested plasmid vector/ insert size were confirmed by using 0.8 % Agarose in 0.5X TAE Buffer. Ethidium bromide was added to the gel of 100 µg/ml from the stock concentration. Samples were loaded with 6X Loading Dye. 1 Kb DNA ladder and 100 bp DNA Ladder were used during the process. Bands were visualized using a U.V. transilluminator, and images were taken by Image Quant (G.E. Healthcare).

2.2.8 Gel extraction

PCR amplified product of respective DNA fragments were gel eluted by Gel extraction kit (Fibrinogen). The eluted DNA fragments concentration was further checked on agarose gel electrophoresis. Eluted DNA was stored at -20 °C and used for cloning into corresponding vectors.

2.2.9 Ligation

For a ligation reaction, the vector and insert following solution was mixed as 1:3 molar ratio (linearized purified vector DNA: a purified insert DNA), 1 µl of 10X T4 DNA ligase buffer, 1 µl of T4 DNA ligase (NEB) to the total volume of 10 µl. Ligation control was without the insert DNA. After that, the ligation mixture was incubated at 16 °C for 16 hours. 2-5 µl of ligation mixture and ligation control were used for transformation.

2.2.10 cell culture

293T human embryonic kidney cells (HEK 293T), HeLa, Vero, and Vero E6 cells were obtained from NCCS Pune. Baby hamster kidney cells (BHK-21) were obtained from Sigma-Aldrich Chemicals Ltd. BST K/O and BST-2 expressing cells generated in the lab. All cells were grown in Dulbecco's modified essential medium (DMEM: Invitrogen), supplemented with glutamate (Invitrogen) and 10% fetal bovine serum (Invitrogen). All cells were maintained in a humidified 37 °C incubator with 5 % CO₂.

2.2.11 Virus propagation

To propagate the virus (CHPV, VSV, etc.), 90% confluent Vero cells in T-75 flask were infected at .01 MOI in serum-free DMEM. Cells were incubated for 12 to 24 hours, and

cells were checked at different time intervals when all cells looked rounded or loosely attached. We have collected the cell supernatants, spun at 12000 rpm for 30 min at 4 °C. The supernatant was filtered through a 0.22-micron filter and aliquoted in small amounts (50-100 ul), and stored at -80 °C.

2.2.12 CHPV Pseudovirus generation

For the generation of VSVΔG TFP stock, a single plaque was picked and subsequently resuspended in 500 µl of DMEM containing 1X PS. This diluted virus inoculum was used to infect VSV G expressing BHK -21 cell plated in 100mm tissue culture dish. The following day, the supernatant was collected, centrifuged at 350 g to clarify for 20 minutes at 4 °C, and filtered through a 0.22-µm filter. This VSVΔG TFP stock was later used to generate CHPV G pseudotyped VSV ΔG TFP particles. 293T cells were plated 3.5×10^6 per 10-cm dish in 10 ml in DMEM containing 5 % FBS. The following day, cells were transfected with the five µg of pCAG-CHPV-G using the lipofectamine 2000 in a 1:3 ratio and according to the procedure recommended by the manufacturer. Incubate the cells in a CO2 incubator for 48 hr. During this incubation period, the 293T cells grow to reach confluence. After 48 hr, the transfected cells were infected with the purified virus stock. The infection was performed for 90 minutes adsorption with rocking at each 10 min interval. The following day, the supernatant was collected, centrifuged at 350 g to clarify for 20 minutes at 4 °C, and filtered through a 0.22-µm filter (Figure 2.2).^[1] Aliquots were prepared and stored at -70°C for further experiments. The stock titer was determined as the pseudotype's infectious unit (IU) using the plaque assay.

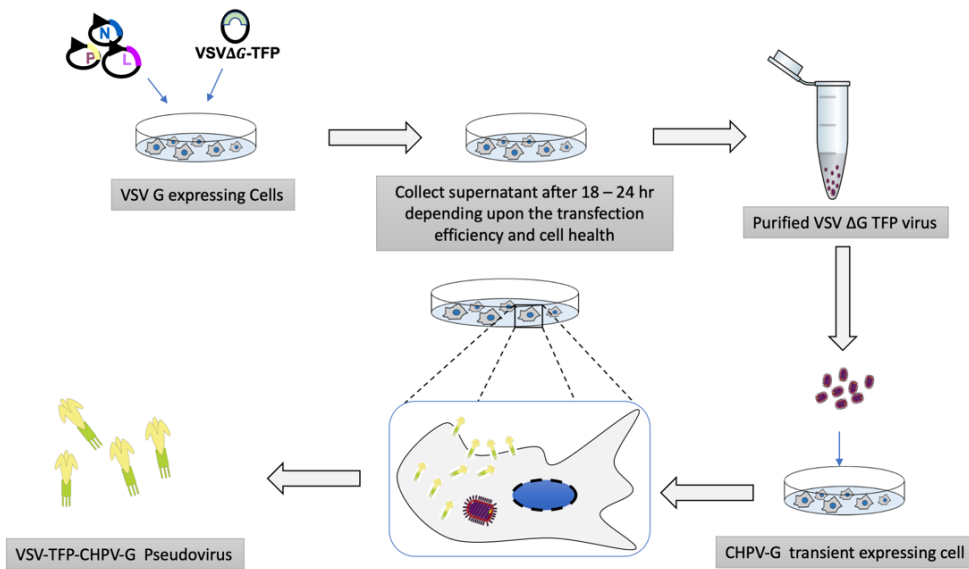


Figure 2.2. Schematic representation of CHPV Pseudovirus generation

2.2.13 Plaque assay

One day before the assay (in the evening), Vero cells were plated in 6-well plates at 80-90 % confluent density. The next day, virus stock was thawed, and serial dilutions of the virus are prepared in plain DMEM. Make a 1:10 dilution by adding 100 µl viral stock to 900 µl medium (Opti – MEM) and starting with your 1:10 dilution up to 10⁻⁵ to 10⁻¹⁰ depending on your viral stock's expected concentration. Infect cells with 0.5 ml of each dilution, add it to the evenly distribute cell monolayer. Incubate the cells for 60 mins at 37°C for adsorption and shake the plate gently occasionally. The cell monolayer was then washed twice with PBS and subsequently overlaid with 3 ml of DMEM containing 2% FBS and 0.5 % of Agarose. Let the agarose layer solidify at RT and incubated at 37°C until the development of plaques (18h for VSV and 48h for vTF7-3). Following the development of plaques, the cell was fixed and stained using the Fixing and staining solution (FS) (0.1% solution of Crystal violet and 2 % of glutaraldehyde) 12 h at room temperature, and then the plaques were removed carefully. The following day, the cell was stained with a 0.1% crystal violet solution for 10 minutes, followed by gentle washing with water (Figure 2.3). The number of plaques was counted, and the initial viral suspension concentration in PFU/ml was determined using the following formula.[2]

Viral titer (PFU/ML) = No. of Plaques

$$(D \times V)$$

D = Dilution factor

V = Volume of diluted virus/well

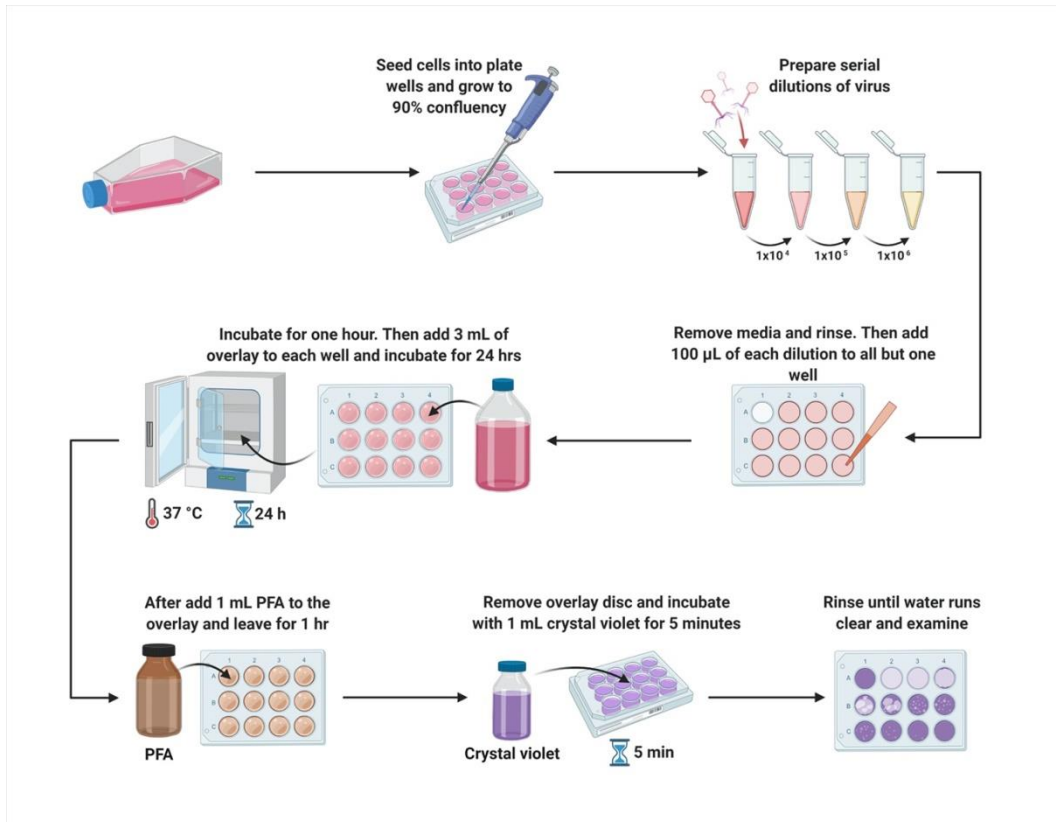


Figure 2.3. Schematic representation of Plaque assay

2.2.14 Construction of BST2 knock-in cells

The establishment of BST-2 expressing cell lines by retrovirus system is a two-step process. The first step is the retrovirus generation, and the second step is the transduction of the retrovirus in appropriate cells.

2.2.14.1 Generation of MULV based retrovirus

To make stable cell lines first, we need to generate retrovirus; for that, we have seeded 293T cells with 60 % confluency in a 100 mm dish. After 24 hours, cells were transfected with three different plasmids, desiring protein plasmid (PQCXIP-HA-BST2), packaging plasmid (pMULV), and envelope plasmid (pCAG-VSV-G), using lipofectamine 2000, after 48 h of transfection supernatant were collected, centrifuged at 12000 rpm for 30 min at 4 °C. Then centrifuged supernatant was filtered by a 0.45-micron filter and stored at -80 °C.[3] (Note- fresh retrovirus show more transduction efficiency)

2.2.14.2 Retroviral Transduction

To introduce the desired protein permanently in a cell line, we need to perform retroviral transduction. We have seeded 80% 293T, BHK-21, and HeLa cells in a 100 mm dish and, after that, added 4 ml of retrovirus containing media supplemented with 8 µg/mL Polybrene. Polybrene can increase transduction by 10-fold or more, but it's toxic for cells. To make the 4 µg/mL final concentration of Polybrene, we added 4 ml extra media after 6 hours of transduction in cells.[4] After 72 hours, cells were split and poured into selection media (Puromycin, 3µg/ml). Cells that carry the gene of interest only survived (Figure 2.4). After selection, the desired gene expressing cells was confirmed by western blot.

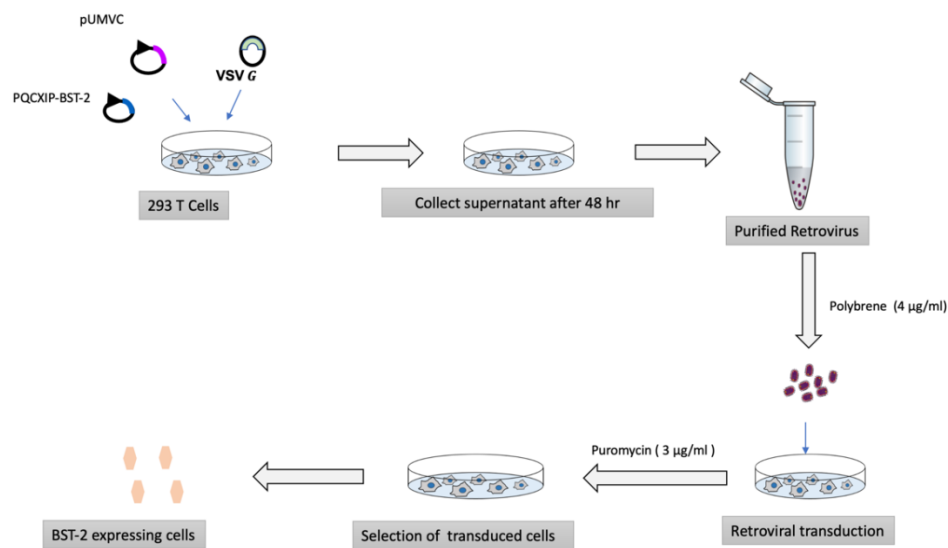


Figure 2.4. Schematic representation of Retrovirus generation and transduction

2.2.15 Construction of BST2 knock-out cells

To generate BST K/O cell lines, we have used the CRISPER-Cas9 system. CRISPER-Cas9 is the only known adaptive immune system in bacteria that memorizes previous infections by integrating short sequences of invading genomes called spacers into the CRISPR locus. The spacers are interspaced with repeats expressed as small guide CRISPR RNAs (crRNAs) employed by Cas proteins to target invaders sequence-specifically upon a reoccurring infection.[5]

We have designed gRNA from the Zhan lab and cloned it in pSpCas9(B.B.)-2A-Puro (PX459) V2.0 vector. Cells were seeded on 100 mm dish with 70% confluency and transfected with BST K/O plasmid DNA using lipofectamine. After 36 h, the old media was removed and filled with the selection media (Puromycin, 3µg/ml) (Figure 2.5). After selection, confirmed K/O cells were analyzed by western blot.

CRISPR/Cas9 Gene Editing

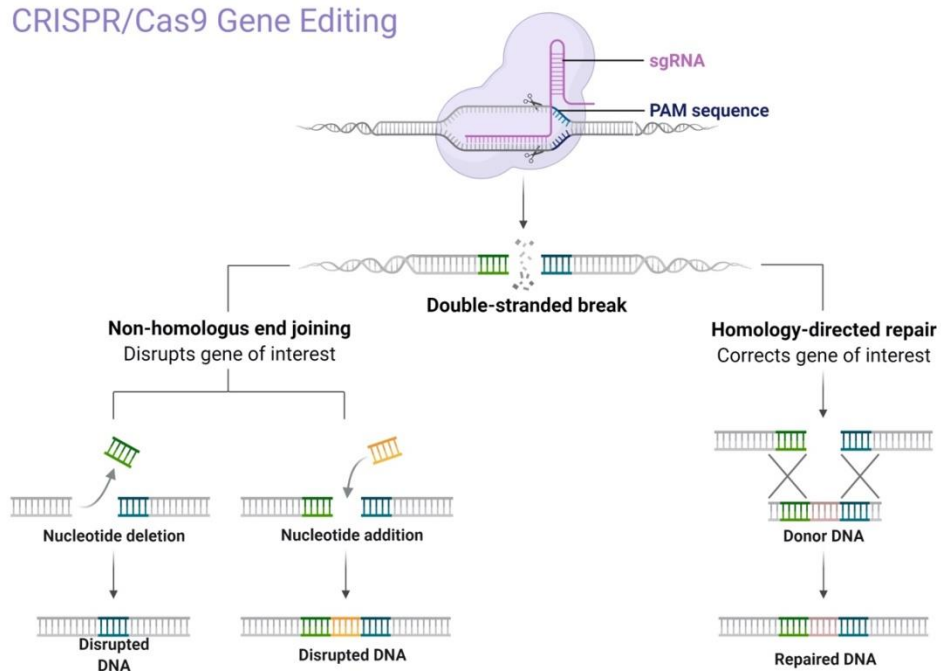


Figure 2.5. Schematic representation of CRISPER-Cas9.

2.2.16 RNA isolation and cDNA synthesis

Cells seeded on 12 well culture plates were transiently transfected with plasmid or treated with the drug in a separate set of experiments. After the appropriate time in each experiment, total RNA was isolated from virus-infected and control cells by using Trizol reagent, according to the manufacturer's protocol (Invitrogen). RNA concentration and purity were quantified using the nanodrop spectrophotometer (Thermo Scientific). 1 µg of RNA was used to synthesize cDNA using the iScript™ cDNA Synthesis Kit (Bio-Rad).

2.2.17 Quantitative real-time polymerase chain reaction assay

The qRT-PCR technique is an advanced method of PCR that is mostly used in molecular biology studies. It allows the detection and quantification of the amplified products formed during each PCR cycle in real-time. It gives the advantage of using the RNA as the primary template material, which is first transcribed into the complementary DNA (cDNA) with reverse transcriptase enzyme that is subsequently used as PCR template for DNA amplification.[6] The primary PCR protocol includes three necessary steps: denaturation, annealing, and elongation. Furthermore, the RT-PCR has an additional step of converting RNA to cDNA before PCR and uses fluorescent labels to obtain the real-time data profile. Two well-known RT-PCR methods are widely used; the probe-based qPCR and the other process take advantage of a dye that can specifically bind to double-stranded(ds) DNA like SYBR green. With each PCR cycle, the copy number of ds DNA increases, and the dye binds with these newly formed amplified products and increases the fluorescence intensity. So, in this case, also the fluorescence intensity released by the dye is directly proportional to the number of amplified products. The SYBR green-based RT-PCR method is cost-effective but not sequence-specific as that of the TaqMan process and, therefore, very commonly used for RNA quantification (Figure 2.6).

We have run and analyzed RT-PCR on a StepOnePlus system with SYBR green master mix during our study. We have analyzed CHPV-G, pCAG-CHPV-G, and GAPDH transcription levels using the SYBR Green PCR Master Mix (Invitrogen or Bio-Rad). PCR conditions used for a run as 92 °C for 30 sec, 52 °C for 45 sec, and 72 °C for 60 sec for 40 cycles, and a melt curve was also added in the reaction. The comparative method threshold cycle was used, and data were normalized with a pCAG-CHPV-G plasmid. Results were analyzed by StepOnePlus software (Applied Biosystems). Fold change values for change

in gene expression were calculated concerning the control. At least three independent experiments were performed. Error bars represent the standard deviation of performed technical replicates. Standard deviations of ΔCT and $\Delta\Delta CT$ were calculated as per the manufacturer's guidelines.

Fluorescent Dye-Based Real Time PCR (qPCR)

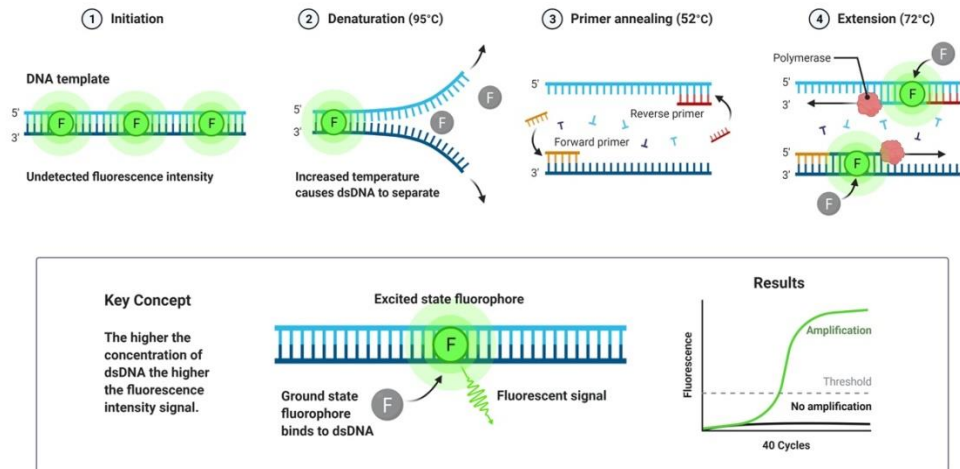


Figure 2.6. Schematic representation of Real time PCR

2.2.18 Protein sample preparation

Cells seeded on 12 well culture plates were transiently transfected in a separate set of experiments with plasmids of interest. Then each set of transfected cells was treated either with the virus for 2 hours each. Similarly, for drug inhibition assay, cells were treated with different drugs and infected by CHPV pseudovirus after 45 minutes. Simultaneously, in these two different experimental conditions, after 2 and 24 hours, the medium was aspirated, and cells were washed with ice-cold PBS. Cells were lysed using RIPA lysis buffer (Thermo scientific-89900) and centrifuged at 4 °C at 14000 g for 15 min. Protein concentration was defined by using the Bradford method (Bio-Rad)[7]. BSA was used as a standard, and samples were analyzed in duplicates.

2.2.19 SDS polyacrylamide gel electrophoresis (SDS-PAGE) and transfer to PVDF membrane

After that, proteins were resolved by SDS-PAGE according to the standard methodology using 10 % percent resolving and 5 % stacking gel. Running buffer (0.3% Tris base, 1.44% glycine, and 0.1% SDS) was added to the Mini PROTEAN III Cell electrophoresis unit (Bio-Rad). Protein samples were prepared by adding 6x loading dye (Laemmli buffer) and boiling at 5 min. A prestained protein marker (Himedia-MBT092) was used, and electrophoresis was performed at 70 V until the bromophenol blue dye front had reached the bottom of the gel.

After completing electrophoresis, the SDS-PAGE gels were removed from the gel apparatus and transferred to the PVDF membrane (Millipore). The transfer was performed using the Mini Trans-Blot®Electrophoretic Transfer Cell unit. The tank was filled with transfer buffer (25 mM Tris; 193 mM glycine; 20 % methanol), and transfer of blot was performed at 250 V for 70-80 min, and proper cooling was maintained during transfer. Following the transfer, the membrane was stained with Ponceau S (0.1% (w/v) in 5% acetic acid) to verify the effective transfer of proteins. After that, the membrane was destained with dH₂O and subjected to western blot analysis.

2.2.20 Western blot assay

Western blotting PVDF membranes were blocked in 5 % blocking buffer (5% milk in TBST) at room temperature for 1 hour on a rotating chamber. Antibodies were diluted in 5% milk TBST, with the preceding block being 5% milk TBST. The membrane was washed with TBST 3-5 times at room temperature (R.T.) and incubated with desired primary antibody (Dilution as per manufacturer's instructions). The membrane was incubated with the relevant HRP-conjugated secondary antibody (Diluted as per manufacturer's instructions) for 2 hours at R.T. Bands were detected using chemiluminescence reagents (Invitrogen), and the image was acquired and analyzed using ImageQuant LAS 4000 (G.E. Healthcare, Biosciences Ltd., Sweden) (Figure 2.7).[8]

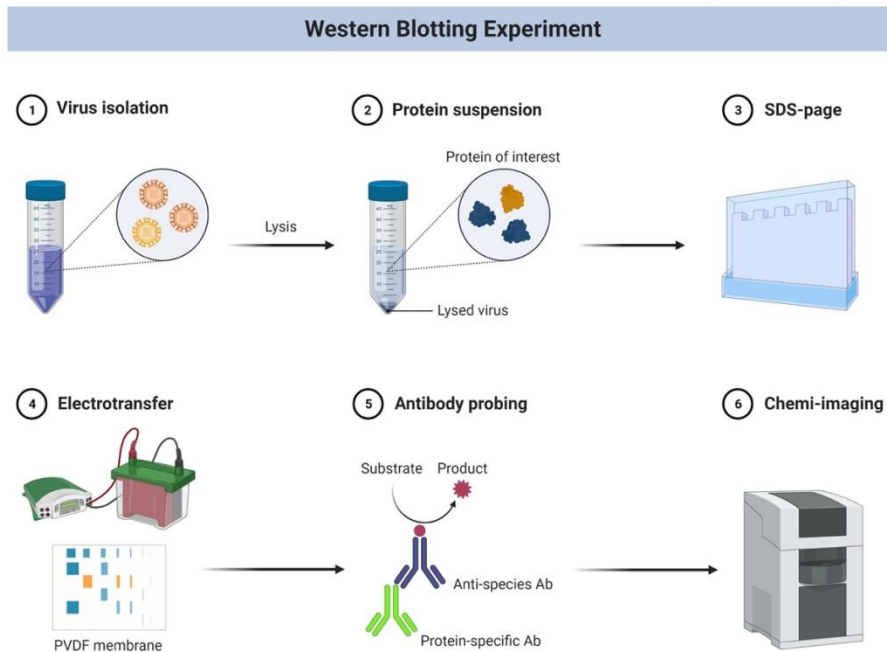


Figure 2.7. Schematic representation of western blot

2.2.21 Immunofluorescence microscopy

To identify the CHPV pseudovirus entry mechanism, we have plated Vero and Vero E6 cells in 12 well plates, after 45 min virus were infected for 1 h. After 24 hours of infection, the fluorescence was observed using the Olympus Multi-Photon Laser Scanning Microscope (FV1200MPE, IX83 Model) (Figure 2.8).

Fluorescence Microscopy

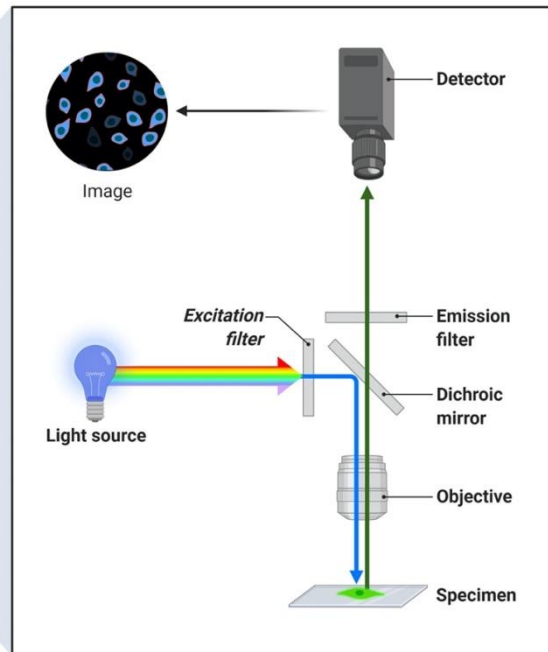


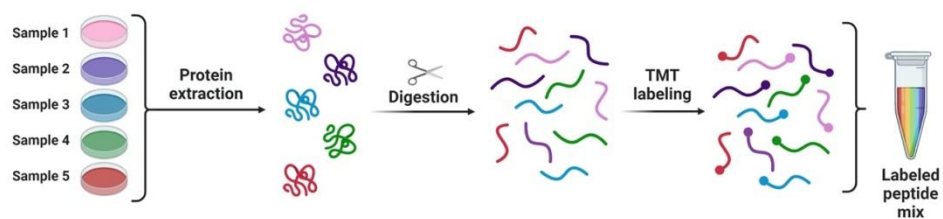
Figure 2.8. Schematic representation of Fluorescence Microscopy

2.2.22 LC-MS analysis

The Huh7 cells were infected with SARS-CoV-2 wherever needed (uninfected, 24hpi, 48hpi, and 72hpi) and were lysed using lysis buffer (5% glycerol, 150 mM NaCl, 10 mM Tris, 10% SDS, and protease inhibitor) after the required time. NuPAGE™ LDS sample buffer (Thermo Fisher Scientific, U.S.) was further added, and the samples were boiled at 99°C for 10 min using a water bath.

Cell lysates were aliquoted and transferred to sample tubes, incubated at 37°C for 5 min at 550 rpm using a block heater, and further sonicated in the water bath for 5 min. Each of the samples prepared was reduced by adding 7 µL of 0.5 M dithiothreitol (DTT) and incubating at 37°C for 30 minutes. Further, alkylation was done by adding 14 µL of 0.5 M iodoacetamide and incubating in the dark for 30 min at room temperature (R.T.). Then, 2 µL of concentrated phosphoric acid and 1211 µL of binding buffer were added, following which the protein capturing was performed using S-Trap™ Micro spin columns (Protifi, Huntington, NY) according to the manufacturer's protocol. The samples were washed with 150 µL of binding buffer four times and then performed proteolytic digestion by using 1.2 µg trypsin (sequencing grade, Promega) for 2 h at 47°C. After that 40 µL of 50 mM TEAB was added, followed by acidification using 40 µL of 0.2% formic acid (F.A.) and elution with 40 µL of 50% acetonitrile (AcN)/0.2% F.A. The eluents were then dried with the help of a Vacufuge vacuum concentrator (Eppendorf, U.S.). The produced peptides were cleaned up in a HyperSep filter plate with a bed volume of 40 µL (Thermo Fisher Scientific, Rockford, IL). Briefly, the plate containing the peptides was washed with 80% AcN/0.1% F.A. and equilibrated with 0.1% F.A. Samples were then filtered in the plate and washed with 0.1% F.A. Further, the peptides were eluted with 30% AcN/0.1% F.A. and 80% AcN/0.1% F.A. and dried using a vacuum concentrator prior to tandem mass tag (TMT) labeling. Further TMT-Pro labeling and RPLC-MS/MS analysis performed by following the previously published paper. Further to identify the relative gene changes, temporal differential abundance analysis was performed using a univariate time series model from the R package LIMMA. In the LIMMA design matrix, separated coefficients were associated with time and replicate to extract the difference as a contrast. Moderated paired t-test using LIMMA with adjustment for replicates was used. Benjamini-Hochberg (BH) adjustment was applied, and only proteins and transcripts with adjusted p values <0.05 were selected (Figure 2.9).[9]

① TMT labeling protocol



② Data collection and analysis

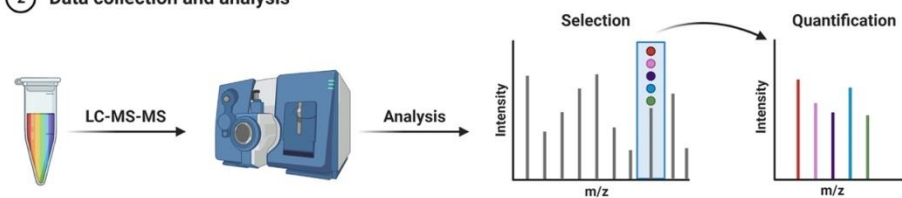


Figure 2.9. Schematic representation of TMT-LC-MS

2.2.23 TEM analysis

The morphological assessment of the CHPV Pseudovirus was done using a transmission electron microscope (TEM, JEOL 2100F). Transmission electron microscopes (TEM) use the electron beam for imaging specimens and generate a highly-magnified image. TEMs can increase objects up to 2 million times. TEMs apply a high voltage electron beam to create an image. It has an electron gun at the top and emits electrons that move through the microscope's vacuum tube. Rather than having a glass lens focusing the light, TEM applies an electromagnetic lens that focuses the electrons into a magnificent beam. This beam then passes through the sample, which is very thin, and the electrons either scatter or hit a fluorescent screen at the bottom of the microscope. The specimen's image with its assorted parts shown in different shades according to its density appears on the screen. To generate the virus's morphological structure, we placed samples on 400-mesh copper grids (SIGMA ALDRICH CHEMICALS PVT LT) and incubated them for 5 minutes. The grids were then rinsed twice with DI water, wicked dry, and stained with 2% uranyl acetate for 5 min and used for TEM imaging (Figure 2.10).[10]

Negative-Stain Transmission Electron Microscopy (TEM)

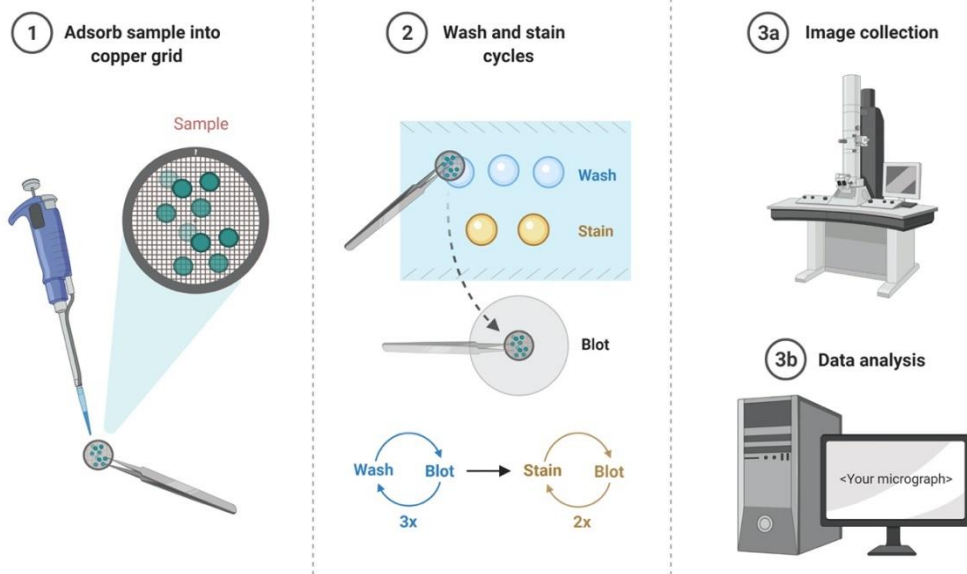


Figure 2.10. Schematic representation of negative stain Transmission Electron Microscopy

2.3 Bio-informatic tools and server used in the studies

2.3.1 Molecular Docking

Molecular docking is a valuable method that involves predicting ligand orientation and conformation within a target site. Molecular docking is a multi-step complex process with two essential aims: 1) Correct prediction of activity and 2) Accurate structural modeling.[11] A computational tool such as Autodock, Discovery studio, and Schrodinger has emerged as valuable tools for drug discovery purposes.[12] Out of these tools, Autodock is a non-commercial docking tool, which uses a stochastic Lamarckian genetic algorithm to deduce the ligand conformations and different scoring functions. These two values help to elucidate the thermodynamic stability of the ligand-macromolecule complex.[13,14] Rizvi *et al.* have been discussed the detailed description of the protocol to perform the docking of ligand and macromolecules.

2.3.2 Determination of structural similarities between virus and human proteins

We discovered the structural similarities among SARS-CoV-2 and human proteins from DaliLite v. 5 web servers.[15] The Dali server compares 3D structural coordinates of two PDB entries by an alignment of alpha carbon distance matrices, allowing for differences in domain order, and produces a structural similarity score. It provides all similar proteins available in the PDB database. From these results, we then filtered only to include those structures pertaining to the human host.

2.3.2 Gene enrichment analysis

To perform gene enrichment analysis, many databases like DAVID, PANTHER, Reactome, etc., and servers available. For our study, we have used g: Profiler g: GOST tools and manually selected G.O.: biological process (G.O.: B.P.), KEGG pathways, and Reactome database in the data source option and set the threshold value at 0.05. We collected the result data and generated their interaction networks and degree centrality calculations using the STRING and CytoNCA tool in Cytoscape.[16-20]

2.4 Reference

1. Whitt, M. A. (2010), Generation of VSV pseudotypes using recombinant Δ G-VSV for studies on virus entry, identification of entry inhibitors, and immune responses to vaccines, *J Virol Methods*, 169, 365-374 (10.1016/j.jviromet.2010.08.006)
2. Baer, A. and K. Kehn-Hall. (2014), Viral concentration determination through plaque assays: using traditional and novel overlay systems, *J Vis Exp*, e52065 (10.3791/52065)
3. Jang, J., J. T. Lee, K. Lee, S. Kim, J. Y. Kim, K. Yoon and S. Kim. (2011), Development of murine leukemia virus-based retroviral vectors with a minimum possibility of cis-activation, *Gene Therapy*, 18, 240-249 (10.1038/gt.2010.135)
4. Cepko, C. and W. Pear. (2001), Overview of the retrovirus transduction system, *Curr Protoc Mol Biol*, Chapter 9, Unit9.9 (10.1002/0471142727.mb0909s36)
5. Adli, M. (2018), The CRISPR tool kit for genome editing and beyond, *Nature Communications*, 9, 1911 (10.1038/s41467-018-04252-2)
6. Heid, C. A., J. Stevens, K. J. Livak and P. M. Williams. (1996), Real time quantitative PCR, *Genome Res*, 6, 986-994 (10.1101/gr.6.10.986)
7. Carlsson, N., A. Borde, S. Wölfel, B. Kerman and A. Larsson. (2011), Quantification of protein concentration by the Bradford method in the presence of pharmaceutical polymers, *Anal Biochem*, 411, 116-121 (10.1016/j.ab.2010.12.026)
8. Thorpe, G. H. and L. J. Kricka. (1986), Enhanced chemiluminescent reactions catalyzed by horseradish peroxidase, *Methods Enzymol*, 133, 331-353 (10.1016/0076-6879(86)33078-7)
9. Appelberg, S., S. Gupta, S. Svensson Akusjärvi, A. T. Ambikan, F. Mikaeloff, E. Saccon, Á. Végvári, R. Benfeitas, M. Sperk, M. Ståhlberg *et al.* (2020), Dysregulation in Akt/mTOR/HIF-1 signaling identified by proteo-transcriptomics of SARS-CoV-2 infected cells, *Emerg Microbes Infect*, 9, 1748-1760 (10.1080/22221751.2020.1799723)
10. Prasad, S., V. Potdar, S. Cherian, P. Abraham and A. Basu. (2020), Transmission electron microscopy imaging of SARS-CoV-2, *Indian J Med Res*, 151, 241-243 (10.4103/ijmr.IJMR_577_20)
11. Brooijmans, N. and I. D. Kuntz. (2003), Molecular recognition and docking algorithms, *Annu Rev Biophys Biomol Struct*, 32, 335-373 (10.1146/annurev.biophys.32.110601.142532)
12. Gilbert, D. (2004), Bioinformatics software resources, *Brief Bioinform*, 5, 300-304 (10.1093/bib/5.3.300)
13. Forli, S., R. Huey, M. E. Pique, M. F. Sanner, D. S. Goodsell and A. J. Olson. (2016), Computational protein-ligand docking and virtual drug screening with the AutoDock suite, *Nat Protoc*, 11, 905-919 (10.1038/nprot.2016.051)
14. Morris, G. M., R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson. (2009), AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, *J Comput Chem*, 30, 2785-2791 (10.1002/jcc.21256)
15. Holm, L. (2019), Benchmarking fold detection by Dalilite v.5, *Bioinformatics*, 35, 5326-5327 (10.1093/bioinformatics/btz536)

16. Dennis, G., Jr., B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane and R. A. Lempicki. (2003), DAVID: Database for Annotation, Visualization, and Integrated Discovery, *Genome Biol*, 4, P3
17. Fabregat, A., K. Sidiropoulos, G. Viteri, O. Forner, P. Marin-Garcia, V. Arnau, P. D'Eustachio, L. Stein and H. Hermjakob. (2017), Reactome pathway analysis: a high-performance in-memory approach, *BMC Bioinformatics*, 18, 142 (10.1186/s12859-017-1559-2)
18. Otasek, D., J. H. Morris, J. Bouças, A. R. Pico and B. Demchak. (2019), Cytoscape Automation: empowering workflow-based network analysis, *Genome Biol*, 20, 185 (10.1186/s13059-019-1758-4)
19. Reimand, J., T. Arak, P. Adler, L. Kolberg, S. Reisberg, H. Peterson and J. Vilo. (2016), g:Profiler-a web server for functional interpretation of gene lists (2016 update), *Nucleic Acids Res*, 44, W83-89 (10.1093/nar/gkw199)
20. Szklarczyk, D., J. H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N. T. Doncheva, A. Roth, P. Bork *et al.* (2017), The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible, *Nucleic Acids Res*, 45, D362-d368 (10.1093/nar/gkw937)

CHAPTER 3

Chapter 3

VHFIDB: The Virus-Host Factor Interaction Database

3.1 Introduction

Infectious viral diseases resulted in millions of deaths each year and are regarded as a primary health concern globally [1]. Many researchers delve into understanding how the virus interacts with its hosts and strategizing appropriate prevention and mitigation measures to address the problems. While many successful antiviral drugs targeting viral proteins are often used, their administration is limited to specific virus species or strains. Some RNA viruses have low fidelity polymerase enzymes such as influenza virus, human immunodeficiency virus 1 (HIV-1), enteroviruses (e.g. EV71), etc.[2,3] and are prone to rapid mutations, leading to drug-resistant strains. Additionally, viruses encode very few proteins, thus limiting the number of available targets for drug discovery.

If affordable to the host, host protein targeting could emerge as an alternative and attractive tool to counter the viral life cycle. Viruses depend on host factors (proteins, nucleotides, etc.) at various stages of their life cycle. Traditionally, multiple techniques such as co-immunoprecipitation, chromatography, yeast two-hybrid method, phage display, fluorescence system, etc., are employed to identify interacting factors [4]. In recent years, genome-wide high-throughput RNA interference (RNAi) screening and CRISPR-Cas ribonuclease-based technologies have accelerated the search for factors involved in virus-host interactions [5-7]. As a result, we now have detailed knowledge of host factors involved in virus life cycles and their involvement mechanisms. The genome-wide loss-of-function analysis method is a valuable tool to identify host genes that enhance or inhibit virus infections without overtly impacting the host; thus, they are assigned as Virus-Host Factors (VHFs). Sayda M. Elbashir et al. firstly demonstrated the RNAi-

mediated transient silencing of mRNAs in mammalian cells to discover essential host factors, which others later adopted to investigate the basis of host-pathogen interactions [5].

Furthermore, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas ribonuclease system is widely used in studying host-virus interactions. The CRISPR-Cas is an RNA and protein-based adaptive immune system that protect bacteria and archaea against invading viruses and foreign nucleic acid [8,9]. Genome-wide CRISPR-Cas screens are being used to identify host factors required for virus replication, entry, and assembly. It has begun by transducing *S. pyogenes* Cas9+ target cells with a lentiviral sgRNA library where host dependency factors, particularly in the context of Epstein-Barr virus-transformed B cells. Multiple genome-wide sgRNA libraries are now commercially available for conducting such studies [10]. With the availability of sgRNA designing tools [11] and progress in reducing unwanted Cas9 off-target effects, the technology is being widely used in eukaryotic genome engineering and aided the researchers in screening host factors in the context of virus infection [12,13].

This study has curated host factor information from peer-reviewed articles and made an open online resource for the VHF database for human and animal viruses. The Virus-Host Factor Interaction Database (VHFIDB, <https://www.vhfidb.com/index.php>) is a freely available online platform, where each entry is stemmed from the published work of peer researchers. With broader coverage of virus families and incorporation of analytical tools, the VHFIDB provides a unique user-friendly enhanced learning experience compared to the other published databases, such as EHFPI and vhfRNAi [14,15]. For example, the EHFPI and vhfRNAi databases comprise information only from RNAi screen-based research articles, while the VHFIDB collects data from RNAi screen, CRISPR Cas9, or supported by other essential molecular techniques. Additionally, the VHFIDB contains information on 72 viruses and 9,921 host genes. It covers some highly contagious viruses like enterovirus, coronaviruses (Human coronavirus 229E, Human SARS coronavirus, SARS-CoV-2), Zaire ebolavirus, Marburg virus, etc., taking into view the emergence and re-emergence of these

viruses in the current context. We have also incorporated some relevant information on contagious animal viruses like porcine reproductive and respiratory syndrome virus (PRRSV), equine arteritis virus, etc. Most importantly, the VHFIDB is also associated with three powerful analytical tools, namely, VHF pathogen network, VHF overlap analysis and, gene enrichment analysis, which increases the usability of VHFIDB (Figure 3.1). We believe this database is a valuable resource tool for investigators in virology, biomedical sciences, and those involved in discovering and developing antiviral therapeutics.



Figure 3.1. Overview of VHFIDB

Virus host factors related to research articles were retrieved from peer-viewed papers through a search on PubMed and Google Scholar. The current VHFIDB database is hitched with the external databases, e.g., UniProt, KEGG ID, GO ID, MINT ID, STRING, OMIM ID, PANTHER, PDB ID, Pfam ID, DrugBank ID, ChEMBL ID, etc.

3.2 Results and discussion

3.2.1 Database overview

The VHFIDB is a collection of experimentally validated viral host factors curated manually from publications. This online resource is primarily obtained from results or work using high throughput CRISPR Cas9, RNAi, and other molecular virology technology. This online resource also comprises analytical bioinformatics tools to interpret the data sets. The VHFIDB can be accessed at the web URL <https://www.vhfidb.com/index.php>. Figure 3.2 represents the VHFIDB home page's screenshot showing search, browse, advanced search, tools, help, download, and contact us options. Currently, the VHFIDB gives information on 9,921 host genes affecting 25 virus families and 72 virus species (Figure 3.3). These host factors exhibit both functions, either they are essential for the virus infection or showing any antiviral activities. We also provide specific information on host factors' roles concerning viral pathogenesis. Our study shows that the maximum number of host factors are attributed to the Parvoviridae family with 1,494 entries followed by the Flaviviridae (1,467 entries), the Orthomyxoviridae (1409), and so on, as displayed in Figure 3. These host factors have been validated on different cell lines like U2OS, MRC5, Vero, HeLa, HEK293, etc. Some information is also supported by the murine model, which is published by peer researchers. With this, we have also provided brief details on viral pathogenesis, like virus-associated diseases, cell tropism, and mode of transmission, except for their classification.

<https://www.vhfdb.com/index.php>



Welcome to Virus Host Factor Interaction Database

[Host Factors against the corona virus including SARS-CoV-2](#)

[Recent Publications on COVID19](#)



Updates on COVID19

Epidemiology

[WHO](#)

[JOHN HOPKINS](#)

[India](#)

[Advise on COVID19](#)

[Genomic Epidemiology of COVID19](#)

Virus-Host Factor Interaction Database (VHFIDB) is a comprehensive and beneficial database which includes a wide array of host factors (essential and antiviral) associated with virus infections. This database encompasses detailed information regarding different host factors and their functions in respect of virus biology. The data has been keenly curated from the available published articles which focus on host factors identified through siRNA based gene knockdown approaches, CRISPR-cas based gene knockout approaches, etc. or by any other techniques.

Currently, VHFIDB contains thorough information about published articles, gene information and virus information for 9921 host factors of 72 virus species from 25 different virus families. Beside this most extensive coverage, VHFIDB is also hitched with five powerful analysis tool, namely, VHF pathogen network, VHF overlap analysis, VHF pathogen protein interacting analysis, gene enrichment analysis, and GWAS candidate gene analysis which increases the domain of VHFIDB.

Database Currently Contains

9921

Host factors

72

Viral Species

25

Viral Family

Figure 3.2. Screenshot of the home page of VHFIDB

The database's home page depicts the search, advanced search, browse, tools, help, and contact us options.

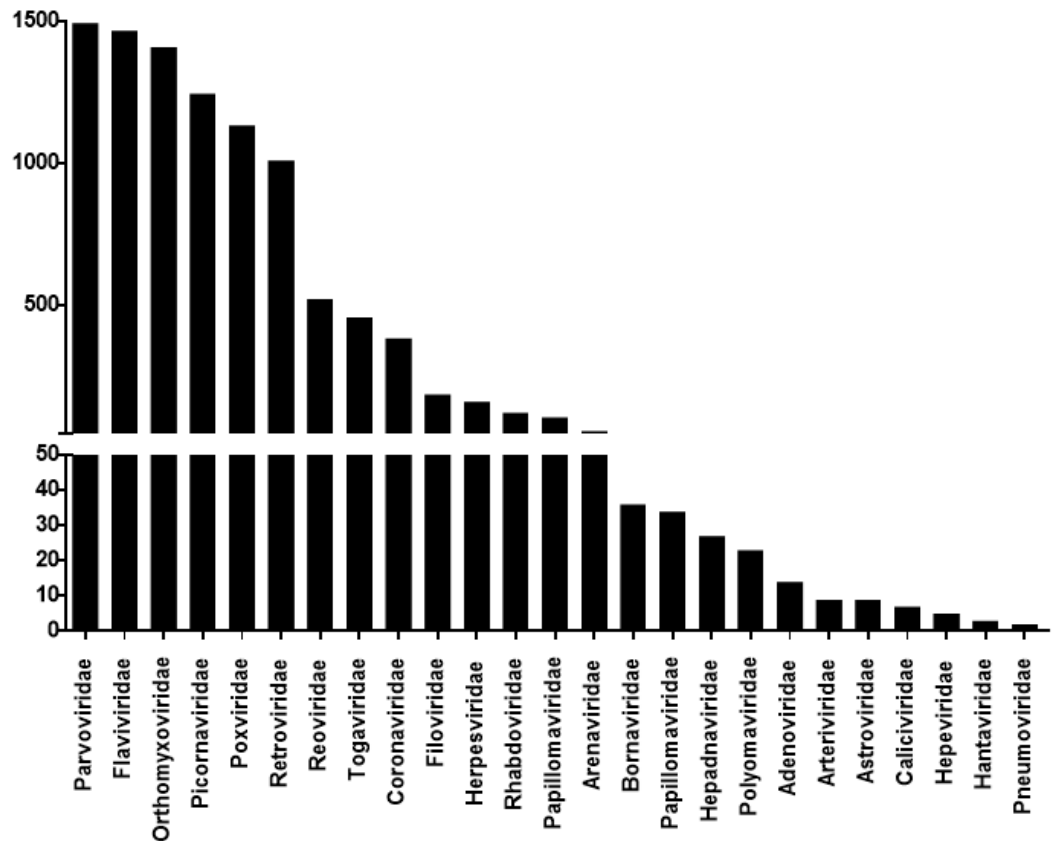



Figure 3.3. Bar graph representing the virus and host factors distribution in 25 different families—example of the number of host factors associated with the virus family. The total number of factors related to a virus family is presented in a bar graph.

3.2.2 Data retrieval

The VHFIDB provides many web-based search tools to reduce the time cost of database users and get the desired information in a minimum number of clicks.

3.2.3 Search

To explore the database content fast, we have provided a search option on the home page. Users can enter the query like host factor name, VHFID ID, PubMed ID, UniProt ID, virus name, virus family name in the box and select the given option in the dashboard. The product includes various components like VHFIDB unique ID, gene name, protein name, host factor function, UniProt ID, virus name, virus family, and references PubMed ID. If a viewer wants more information about the host factor, he needs to click on VHFIDB ID. Moreover, we are also giving a clickable option below the virus name to view the interactive map for a specific virus and its host factors. (Figure 3.4).


Virus Host factor Interaction Database
Help Download

Search Result

[Home](#)
[Advanced Search](#)
[Browse](#)
[Tools](#)
[Contact Us](#)

Search Result for SARS-CoV-2 in virus

Show 10 entries
 Search:

ID	Gene Name	Protein Name	HF Function	UniProt ID	Virus Name	Virus Family	Reference-PMID
VHFI09594	PRIM1	DNA primase small subunit	Associated with the viral infection	P49642	Severe acute respiratory syndrome coronavirus-2 View interaction map	Coronaviridae	32353859
VHFI09595	PRIM2	DNA primase large subunit	Associated with the viral infection	P49643	Severe acute respiratory syndrome coronavirus-2 View interaction map	Coronaviridae	32353859
VHFI09596	PKP2	Plakophilin-2	Associated with the viral infection	Q99959	Severe acute respiratory syndrome coronavirus-2 View interaction map	Coronaviridae	32353859
VHFI09597	COLGALT1	Procollagen galactosyltransferase 1	Associated with the viral infection	Q8N8J5	Severe acute respiratory syndrome coronavirus-2 View interaction map	Coronaviridae	32353859
VHFI09598	EIF4E2	Eukaryotic translation initiation factor 4E type 2	Associated with the viral infection	Q60573	Severe acute respiratory syndrome coronavirus-2 View interaction map	Coronaviridae	32353859
VHFI09599	SLC27A2	Very long-chain acyl-CoA synthetase	Associated with the viral infection	O14975	Severe acute respiratory syndrome coronavirus-2 View interaction map	Coronaviridae	32353859
VHFI09600	RAP1GDS1	Rap1 GTPase-GDP dissociation stimulator 1	Associated with the viral infection	P52306	Severe acute respiratory syndrome coronavirus-2 View interaction map	Coronaviridae	32353859
VHFI09601	FKBP15	FK506-binding protein 15	Associated with the viral infection	Q5T1M5	Severe acute respiratory syndrome coronavirus-2 View interaction map	Coronaviridae	32353859
VHFI09602	WASHC4	WASH complex subunit 4	Associated with the viral infection	Q2M389	Severe acute respiratory syndrome coronavirus-2 View interaction map	Coronaviridae	32353859
VHFI09603	POR	NADPH-cytochrome P450 reductase	Associated with the viral infection	P16435	Severe acute respiratory syndrome coronavirus-2 View interaction map	Coronaviridae	32353859

Showing 51 to 60 of 250 entries
 1 ... 5 **6** 7 ... 25

Figure 3.4. Screenshot of options available on the home page (A) Provides search options and (B) the output results, including VHFIDB ID, host factor symbol, protein name, HF function, UniProt ID, Interacting Proteins, Virus name, Virus family, and reference PubMed ID.

3.2.4 Browse

The browse option contains the list of virus families described by the Baltimore classification system. For example, single strand (+) RNA virus group include eight virus families, the user can directly click on a family to go to its host factors or can click on "+" sign before the family name to go on further classification, including genus, and species (Figure 3.5).

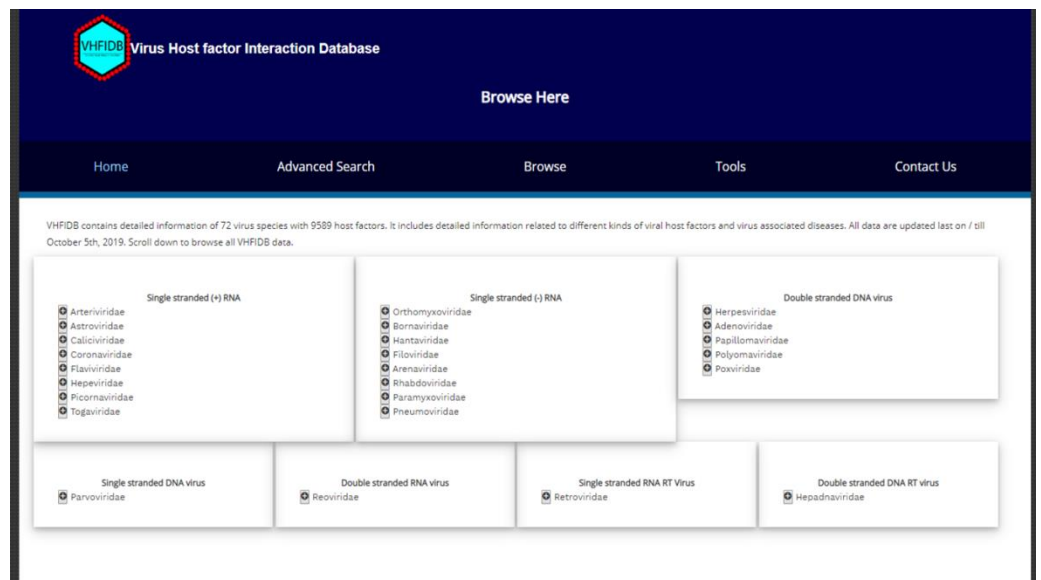


Figure 3.5. Screenshot of browse options. The figure shows host factors of different viruses based on the Baltimore system of virus classification.

3.2.5 Advanced search

We have incorporated an advanced search tool to search the database by virus classification. According to the Baltimore classification system, viruses can be classified into seven classes, single-strand (+) RNA virus, single-strand (-) RNA virus, double-strand DNA virus, single-strand DNA virus, double-strand RNA virus, single-strand RNA reverse transcribing (RT) virus, and double-strand DNA RT virus. For further specifications, we have provided three more drop-down menus. The user can select specifically virus family, virus genus, and virus species to get information on their host factors.

3.2.6 Analysis

VHFIDB also provides useful analytical tools such as overlap analysis, network analysis, and gene enrichment analysis to help understand host-pathogen biology (Figure 3.6).

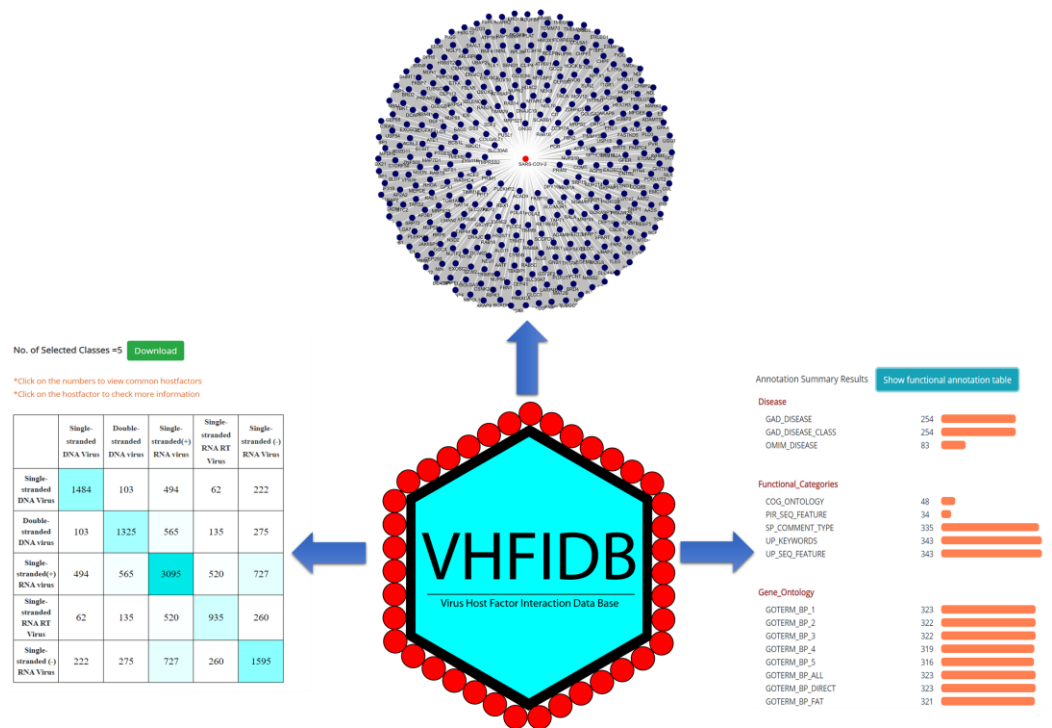


Figure 3.6. Schematic representation of analysis tools. Three analysis tools of VHFIDB to facilitate the in-depth understanding of host factors for pathogenic infection.

3.2.6.1 Overlap analysis

Overlap analysis identifies host factors linked to various viral infection pathways. The VHFIDB allows multicentric comparisons of VHF data and turns results into a heatmap representation. The maximum number of overlapping host factors was found between the vaccinia virus (VACV) and the Influenza-A virus (IAV), with 118 common host factors. The IAV has the maximum number of overlapping host genes with VACV, followed by HIV-1, Hepatitis C virus (HCV), and Adeno associated virus-2 (AAV2), having 118, 117, 98, and 91 entries, respectively. While HIV-1 shared the highest number of overlapping host genes with IAV, VACV, AAV2, HCV, and EV71 that have 117, 78, 49, 45, and 31 entries, correspondingly. Similarly, Betacoronavirus shared a maximum number of overlapping host genes with IAV, HIV-1, HCV, and AAV-2 with 90, 49, 31, 20 matching entries. These overlapping host factors can also be observed in a family-wise and class-wise manner. Coronaviridae shared the highest number of overlapping host genes with Flaviviridae, Picornaviridae, Retroviridae, Orthomyxoviridae, Poxviridae with 105, 93, 44, 41, and 36 entries, respectively. In the case of group-wise overlapping, single-strand (+) RNA virus ((+) ssRNA) has the highest frequency of overlapping host genes with single strand (-) RNA virus ((-) ssRNA), double-strand DNA virus (dsDNA), single-strand RNA (RT) virus (ssRNA-RT) and single-strand DNA virus (ssDNA) with 727, 565, 520 and 494 entries, respectively (Figure 3.7) moreover that a user can directly download the results and use it for their publication and other purposes.

Interactive heatmap displays the common host genes shared by the viruses. The heatmap can be analyzed between different virus classes, virus families or between species of viruses.

Select any

Species

Analyze

B

No. of Selected Families =6

Download

*Click on the numbers to view common hostfactors

*Click on the hostfactor to check more information

	Coronaviridae	Retroviridae	Flaviviridae	Picornaviridae	Orthomyxoviridae	Poxviridae	Common Hostfactors :
Coronaviridae	375	44	105	93	41	36	<div> TMPRSS2 RTN4 NUP88 MARK3 CUL2 NEU1 ATP13A3 POFUT1 PPIA TRIM56 IFITM3 IFITM2 RB1 DDX5 ACE2 PI4KB RYBP PPIA NOMO3 FKBP1A PPIG PPIH RCAN3 </div>
Retroviridae	44	935	162	129	131	97	
Flaviviridae	105	162	1360	206	214	152	
Picornaviridae	93	129	206	1160	156	146	
Orthomyxoviridae	41	131	214	156	1257	123	
Poxviridae	36	97	152	146	123	1112	

C

No. of Selected Families =6

Download

*Click on the numbers to view common hostfactors

*Click on the hostfactor to check more information

	Coronaviridae	Retroviridae	Flaviviridae	Picornaviridae	Orthomyxoviridae	Poxviridae	Common Hostfactors :
Coronaviridae	375	44	105	93	41	36	<div> TMPRSS2 RTN4 NUP88 MARK3 CUL2 NEU1 ATP13A3 POFUT1 PPIA TRIM56 IFITM3 IFITM2 RB1 DDX5 ACE2 PI4KB RYBP PPIA NOMO3 FKBP1A PPIG PPIH RCAN3 </div>
Retroviridae	44	935	162	129	131	97	
Flaviviridae	105	162	1360	206	214	152	
Picornaviridae	93	129	206	1160	156	146	
Orthomyxoviridae	41	131	214	156	1257	123	
Poxviridae	36	97	152	146	123	1112	

D No. of Selected Classes =5 [Download](#)

*Click on the numbers to view common hostfactors

*Click on the hostfactor to check more information

	Single-stranded DNA Virus	Double-stranded DNA virus	Single-stranded(+) RNA virus	Single-stranded RNA RT Virus	Single-stranded (-) RNA Virus
Single-stranded DNA Virus	1484	103	494	62	222
Double-stranded DNA virus	103	1325	565	135	275
Single-stranded(+) RNA virus	494	565	3095	520	727
Single-stranded RNA RT Virus	62	135	520	935	260
Single-stranded (-) RNA Virus	222	275	727	260	1595

Common Hostfactors :

KMT5A
WWC2
KRBA1
DKFZP434P162
CETN2
LMNA
CSGALNACT1
SPAG6
APPL2
TAP1
SYT1
ZYG11B
PHYHD1
IL27RA
BACE1
C15orf56
UTS2B
ACTRT2
ZNF599
TOR4A
COA5
EMC10
SPHAR

Figure 3.7. Screenshot of the VHF overlap analysis tool and its corresponding results. (A) Users can select virus class, family, and species to do overlap analysis. Heatmap is exhibiting overlapping host factors in VHFIDB based on (B) Family, (C) Species, and (D) Class.

3.2.6.2 VHF pathogen network analysis

The VHF-pathogen network displays the global connection of VHFs for selected viruses in an imaged version. The connectivity of nodes (i.e., pathogens or VHF genes) exhibits different viruses' associations based on shared VHFs. The graphical network shows that human SARS-CoV-2 shares a different number of host factors with 30 other viruses. Especially, SARS-CoV-2 shares the most host factors with Influenza A virus (IAV) and Human immunodeficiency virus type 1 (HIV-1), followed by Hepatitis C virus (HCV), Vaccinia virus (VACV), Dengue (DENV-2), and Enterovirus 71 (EV71), etc. (Figure 3.8). the viewer can also precisely select interacting virus or host factor whatever he wants for his study. Moreover, by clicking on the virus name or host factor name, viewer can redirect to search results for the specific virus or host factor.

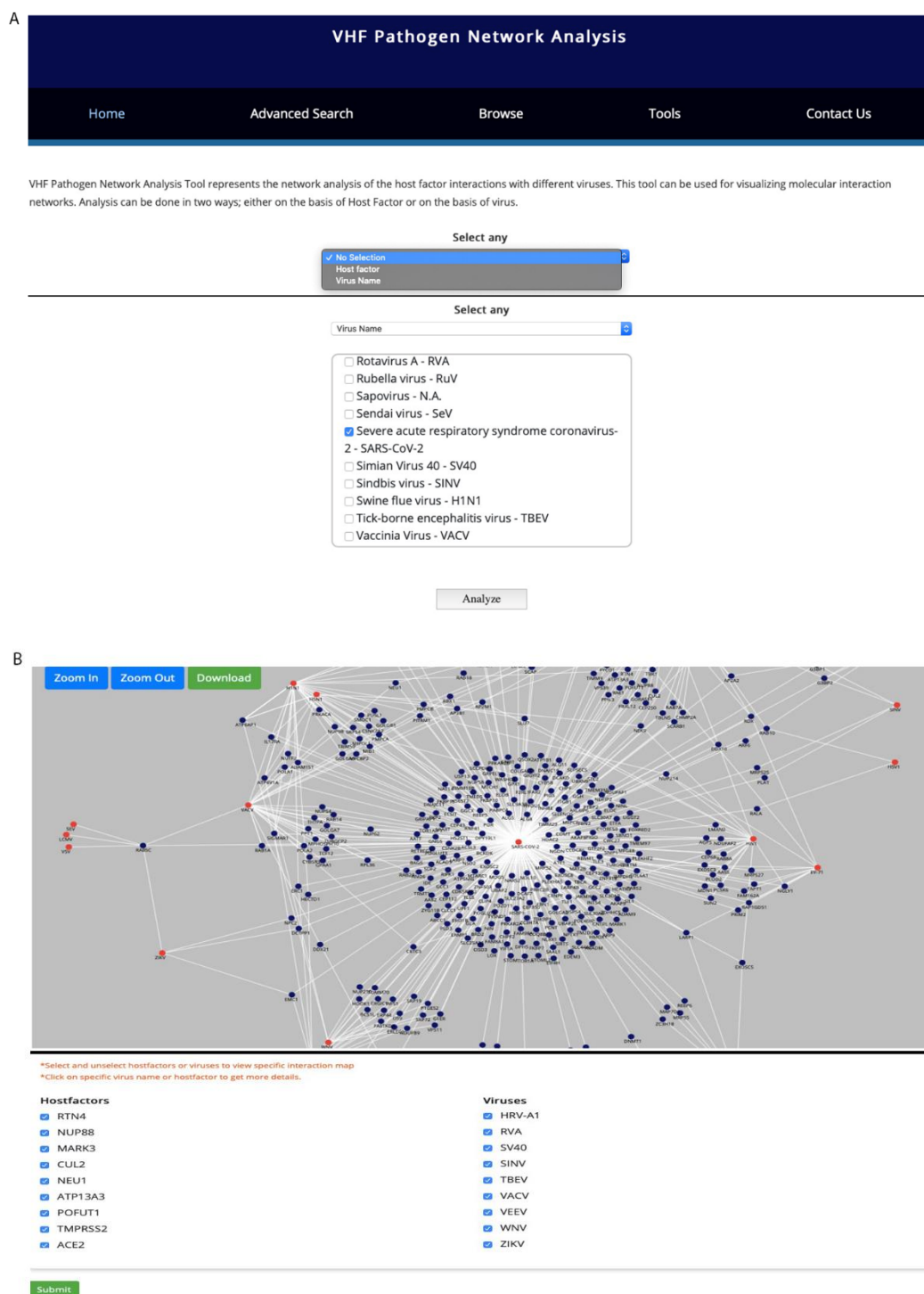


Figure 3.8. Screenshot of VHF Pathogen Network Analysis and its corresponding result. (A) Users can select a virus name or put the host factor name to get overlap analysis; (B) selection of SARS-CoV-2 and network analysis of it as an example.

3.2.6.3 Gene enrichment analysis

In gene enrichment analysis, users can analyze host factors/proteins based on evolutionary and functional classification. In this tool, we have linked the host factor from the DAVID database [16]. Gene enrichment analysis results come in two forms, annotation summary results and function annotation table. In annotation summary results, results appear as a bar graph of genes associated with different terms.

If users want to get more results, they need to click on the function annotation table, and a new window with a table will appear, and the user can get information according to his/her necessity (Figure 3.9).



Figure 3.9. Screenshot of gene enrichment analysis tool. (A) Users can select virus names or put the host factor name to get gene enrichment analysis, and (B, C) shows the analysis results in the form of an annotation summary and function annotation table.

3.2.7 VHF as a potential source of antiviral drug information

It is always challenging to develop a specific antiviral drug against the virus because of the small number of molecular targets involved in pathogenesis and the rapid mutation rate of the viral genome. Viruses require host factors for their attachment, entry, genome transcription, replication, translation, and assembly, thus targeting the host factor(s) offers a unique opportunity to develop novel antiviral drugs. Genome-wide RNAi, CRISPR Cas9, and some other molecular techniques screening for viral infections have offered virus host factors a promising class of potential drug targets for viruses. The VHFIDB also provides information on FDA-approved drugs and other compounds from DrugBank and ChEMBL's IDs that target host factors [17,18].

For example, with respect to the SARS-CoV-2, 48 out of 333 VHF genes were identified as targets of 196 drugs from the DrugBank database, and 70 out of 333 VHF genes were identified as drug targets of 70 drugs from the ChEMBL database, respectively (Table 3.1). It was also observed that many drugs targeted VHFs have commonalities with other pathogens too. Therefore, the potential of VHFs as drug targets for developing broad-spectrum antivirals can be discovered. For example, ATP6V1A, a human gene, is essential for replicating four viruses (Human SARS coronavirus, Swine flu virus, West Nile virus, Avian influenza virus) that can be targeted by four different drugs. Similarly, SARS-CoV2 (COVID-19) and Human SARS coronavirus use ACE2 as an entry receptor, and its also crucial for Swine flu virus infection that can be targeted by four drugs Moexipril (DB00691), Lisinopril (DB00722), SPP1148 (DB05203), Azilsartan medoxomil (DB05358) [19-21]. Thus, the VHFIDB provides a multifaceted platform for answering queries of the wider virology field.

Table 3.1. List of virus-host factors associated with the SARS-CoV-2 and their associated function. The DrugBank IDs and ChEMBL IDs represent the drugs against the host factor that can be used to treat the Coronaviruses.

VHFID	Gene Name	Function	Viral protein	DrugBank	ChEMBL
VHFID9590	ACE2	Cell entry receptor for the SARS-CoV-2	Spike	DB00722;DB00691;DB05203;DB05358	CHEMBL3736
VHFID9591	TMPRSS2	Cell entry receptor for the SARS-CoV-2	Spike	N.A.	CHEMBL1795140
VHFID9592	POLA1	Associated with the viral infection	Nsp1	DB00242;DB00631;DB01073;DB01280	CHEMBL1828
VHFID9593	POLA2	Associated with the viral infection	Nsp1	DB00851	N.A.
VHFID9599	SLC27A2	Associated with the viral infection	Nsp2	N.A.	CHEMBL4326
VHFID9603	POR	Associated with the viral infection	Nsp2	DB03461;DB00865;DB00694;DB00997;DB01466;DB03147;DB00166;DB00305;DB00665;DB00698;DB03247;	CHEMBL2169731
VHFID9604	GIGYF2	Associated with the viral infection	Nsp2	N.A.	CHEMBL2331055
VHFID9610	IDE	Associated with the viral infection	Nsp4	DB00626;DB00030;DB00071;	CHEMBL1293287
VHFID9613	HDAC2	Associated with the viral infection	Nsp5	DB01223;DB05015;DB00227;DB05651;DB01303;DB06603;DB06176;DB05223;DB00277;DB00313;DB02546;	CHEMBL1937
VHFID9615	GPX1	Associated with the viral infection	Nsp5	DB00143;	CHEMBL2163186
VHFID9618	ATP6AP1	Associated with the viral infection	Nsp6	N.A.	CHEMBL4790

VHFID96 19	SIGM AR1	Associated with the viral infection	Nsp6	DB00321;DB09014;DB0051 4;DB01488;DB00540;DB006 52;DB03575;DB01708;DB00 409;	CHEMBL 287
VHFID96 21	SCA RB1	Associated with the viral infection	Nsp7	DB00144;	CHEMBL 1914272
VHFID96 23	PTG ES2	Associated with the viral infection	Nsp7	N.A.	CHEMBL 4411
VHFID96 24	CYB5 R3	Associated with the viral infection	Nsp7	DB03147;DB00157;	CHEMBL 2146
VHFID96 28	RAB7 A	Associated with the viral infection	Nsp7	DB04315;	N.A.
VHFID96 36	NDU FAF2	Associated with the viral infection	Nsp7	N.A.	CHEMBL 2363065
VHFID96 37	COM T	Associated with the viral infection	Nsp7	DB02342;DB03336;DB0028 6;DB00255;DB00841;DB009 88;DB00494;DB00968;DB01 141;DB04820;DB00118;DB0 1420;DB00323;	CHEMBL 2023
VHFID96 40	RALA	Associated with the viral infection	Nsp7	DB04315;	N.A.
VHFID96 41	RHO A	Associated with the viral infection	Nsp7	DB04315;	CHEMBL 6052
VHFID96 42	MTA RC1	Associated with the viral infection	Nsp7	N.A.	CHEMBL 3706559
VHFID96 48	MOG S	Associated with the viral infection	Nsp7	N.A.	CHEMBL 4684
VHFID96 49	ACSL 3	Associated with the viral infection	Nsp7	DB00159;	N.A.
VHFID96 56	SEP SEC S	Associated with the viral infection	Nsp8	DB00114;	N.A.

VHFID96 59	NSD 2	Associated with the viral infection	Nsp8	N.A.	CHEMBL 3108645
VHFID96 69	NAR S2	Associated with the viral infection	Nsp8	DB00174;	N.A.
VHFID96 76	EIF4 H	Associated with the viral infection	Nsp9	N.A.	CHEMBL 1293274
VHFID96 82	MAT 2B	Associated with the viral infection	Nsp9	DB00134;	N.A.
VHFID96 89	NEK9	Associated with the viral infection	Nsp9	N.A.	CHEMBL 5257
VHFID96 92	EIF4 H	Associated with the viral infection	Nsp9	N.A.	CHEMBL 1293274
VHFID96 93	GFE R	Associated with the viral infection	Nsp1 0	DB03147;	CHEMBL 1741189
VHFID96 99	RIPK 1	Associated with the viral infection	Nsp1 2	N.A.	CHEMBL 5464
VHFID97 12	BCK DK	Associated with the viral infection	Nsp1 2	DB01660;	N.A.
VHFID97 27	USP1 3	Associated with the viral infection	Nsp1 3	N.A.	CHEMBL 3407324
VHFID97 32	CIT	Associated with the viral infection	Nsp1 3	N.A.	CHEMBL 5579
VHFID97 34	TBK1	Associated with the viral infection	Nsp1 3	N.A.	CHEMBL 5408
VHFID97 54	PRK ACA	Associated with the viral infection	Nsp1 3	DB07107;DB06959;DB0785 7;DB07860;DB08073;DB069 77;DB08568;DB07858;DB08 756;DB07855;DB07876;DB0 8070;DB08113;DB07859;DB 07996;DB07856;DB04098;D	CHEMBL 4101

				B02611;DB01940;DB02155; DB08846;DB04707;DB0794 7;DB08231;DB07995;DB079 97;DB07854;DB01919;DB04 522;DB02482;	
VHFID97 55	PRK AR2A	Associated with the viral infection	Nsp1 3	DB05798;	N.A.
VHFID97 56	PRK AR2B	Associated with the viral infection	Nsp1 3	DB02527;	N.A.
VHFID97 60	IMPD H2	Associated with the viral infection	Nsp1 4	DB04566;DB01033;DB0068 8;DB01024;DB00157;DB008 11;DB03070;DB06103;	CHEMBL 2002
VHFID97 61	SIRT 5	Associated with the viral infection	Nsp1 4	DB03478;DB02059;DB0270 1;DB04786;	CHEMBL 2163183
VHFID97 67	BRD 4	Associated with the viral infection	E	N.A.	CHEMBL 1163125
VHFID97 68	BRD 2	Associated with the viral infection	E	N.A.	CHEMBL 1293289
VHFID97 72	SLC4 4A2	Associated with the viral infection	E	DB00122;	N.A.
VHFID97 75	CSN K2A2	Associated with the viral infection	N	DB07546;	CHEMBL 4070
VHFID97 76	CSN K2B	Associated with the viral infection	N	N.A.	CHEMBL 2358
VHFID97 80	PAB PC1	Associated with the viral infection	N	N.A.	CHEMBL 1293286
VHFID97 82	PAB PC4	Associated with the viral infection	N	N.A.	CHEMBL 5333
VHFID97 87	PITR M1	Associated with the viral infection	M	N.A.	CHEMBL 3124731

VHFID97 90	COQ 8B	Associated with the viral infection	M	N.A.	CHEMBL 5753
VHFID97 91	GGC X	Associated with the viral infection	M	DB01125;DB00100;DB0003 6;DB00055;DB00142;DB001 70;DB01022;	CHEMBL 2012
VHFID97 92	ATP6 V1A	Associated with the viral infection	M	DB00630;DB06733;DB0673 4;DB01077;DB01133;	N.A.
VHFID97 97	ATP1 B1	Associated with the viral infection	M	DB09479;	CHEMBL 2095186
VHFID97 98	AAS S	Associated with the viral infection	M	DB00142;DB04207;DB0015 7;DB02338;	N.A.
VHFID97 99	ACA DM	Associated with the viral infection	M	DB03415;DB03147;DB0291 0;	N.A.
VHFID98 03	RTN4	Associated with the viral infection	M	N.A.	CHEMBL 3712895
VHFID98 12	TAR S2	Associated with the viral infection	M	DB00156;	CHEMBL 3351186
VHFID98 14	PSM D8	Associated with the viral infection	M	N.A.	CHEMBL 2364701
VHFID98 18	HMO X1	Associated with the viral infection	Orf3a	DB07342;DB02468;DB0390 6;DB02073;DB01942;DB001 57;DB04912;DB04803;DB00 163;	CHEMBL 2823
VHFID98 30	NEU 1	Associated with the viral infection	Orf8	DB00198;	CHEMBL 2726
VHFID98 31	PLAT	Associated with the viral infection	Orf8	DB07684;DB00513;DB0640 4;DB09228;DB01050;DB010 88;DB00013;	CHEMBL 1873
VHFID98 41	DNM T1	Associated with the viral infection	Orf8	DB00928;DB01262;DB0109 9;DB01035;	CHEMBL 1993
VHFID98 42	ITGB 1	Associated with the	Orf8	DB00098;	CHEMBL 1905

		viral infection			
VHFID98 43	ADA M9	Associated with the viral infection	Orf8	DB05033;	CHEMBL 5982
VHFID98 44	PCS K6	Associated with the viral infection	Orf8	N.A.	CHEMBL 2951
VHFID98 47	IL17 RA	Associated with the viral infection	Orf8	N.A.	CHEMBL 3580485
VHFID98 48	GGH	Associated with the viral infection	Orf8	DB00158;DB00563;	CHEMBL 2223
VHFID98 51	PLO D2	Associated with the viral infection	Orf8	DB00126;	N.A.
VHFID98 52	COL6 A1	Associated with the viral infection	Orf8	N.A.	CHEMBL 2364188
VHFID98 53	PVR	Associated with the viral infection	Orf8	DB08231;DB03203;	N.A.
VHFID98 54	LOX	Associated with the viral infection	Orf8	N.A.	CHEMBL 2249
VHFID98 58	ADA MTS 1	Associated with the viral infection	Orf8	N.A.	CHEMBL 5133
VHFID98 59	SDF2	Associated with the viral infection	Orf8	DB02201;	N.A.
VHFID98 63	HYO U1	Associated with the viral infection	Orf8	N.A.	CHEMBL 2216741
VHFID98 64	SIL1	Associated with the viral infection	Orf8	N.A.	N.A.
VHFID98 65	ERO 1B	Associated with the viral infection	Orf8	DB03147;	N.A.

VHFID98 75	GDF 15	Associated with the viral infection	Orf8	N.A.	CHEMBL 3120039
VHFID98 77	MAR K1	Associated with the viral infection	Orf9b	N.A.	CHEMBL 5940
VHFID98 78	MAR K2	Associated with the viral infection	Orf9b	N.A.	CHEMBL 3831
VHFID98 79	MAR K3	Associated with the viral infection	Orf9b	N.A.	CHEMBL 5600
VHFID98 83	DPH 5	Associated with the viral infection	Orf9b	DB01752;	N.A.
VHFID98 87	DCT PP1	Associated with the viral infection	Orf9b	N.A.	CHEMBL 3769292
VHFID98 99	F2RL 1	Associated with the viral infection	Orf9c	N.A.	CHEMBL 5963
VHFID99 01	ABC C1	Associated with the viral infection	Orf9c	DB05812;DB00345;DB0070 1;DB01072;DB01076;DB048 51;DB02659;DB01394;DB00 286;DB00091;DB00970;DB0 0694;DB00586;DB01248;DB 00997;DB00445;DB00773;D B00693;DB01645;DB00143; DB01016;DB00365;DB0105 0;DB01177;DB00224;DB003 28;DB00762;DB00602;DB00 709;DB00563;DB00834;DB0 1204;DB02375;DB03467;DB 01165;DB01229;DB01174;D B01032;DB00396;DB04216; DB03825;DB01045;DB0050 3;DB01098;DB01232;DB063 35;DB01138;DB04348;DB09 161;DB05294;DB08881;DB0 0661;DB00570;DB00541;DB 00399;	CHEMBL 3004
VHFID99 05	NDU FAF1	Associated with the viral infection	Orf9c	N.A.	CHEMBL 2363065
VHFID99 07	NDU FB9	Associated with the	Orf9c	DB00157;	CHEMBL 2363065

		viral infection			
VHFID99 14	PPT1	Associated with the viral infection	Orf10	DB02035;DB03796;	CHEMBL 2331051
VHFID99 19	RBX1	Associated with the viral infection	Orf10	N.A.	CHEMBL 3833061
VHFID99 20	ELO B	Associated with the viral infection	Orf10	N.A.	CHEMBL 3301400
VHFID99 21	ELO C	Associated with the viral infection	Orf10	N.A.	CHEMBL 3301400

3.3 Conclusions

We have developed a comprehensive database comprising host factors involved in viral pathogenesis identified by experimentally validated techniques with supporting literature. Currently, the VHFIDB contains 9,921 entries related to 25 virus families and 72 virus species. We have also incorporated five powerful tools for VHFIDB analysis and visualization with exhaustive information about host factors, viruses, and publications. We anticipate that this database will be of significant help to the researchers engaged in biomedicine, microbiology, pharmaceutical industries, etc. As an open resource, it will not only help in understanding the host-virus interactions and their associated diseases but also facilitate the research and developments of translational medicine and host factor-directed antiviral drug development. The VHFIDB database will continue to include newly discovered host factors on a half-yearly basis to keep abreast with this rapidly developing field. Therefore, we are committed to making this database more feature-rich and valuable to researchers worldwide.

3.4 Materials and Methods

3.4.1 Data source

The VHFIDB database includes VHFIs backed by strong experimental evidence from the published scientific literature. The primary source of references was obtained from the advanced search on PubMed, and Google Scholar using the keywords such as genome-wide RNAi or RNAi screen of host factors in virus pathogenesis, genome-wide CRISPR/Cas9 screen identifies host factors for virus pathogenesis and essential antiviral host factors. Literature reporting the screening and identification of host genes necessary or inhibitory effect on viral infections was accurately extracted by reviewing the articles. The initial search retrieved many published papers. However, only a limited number of Research articles had relevant, usable information for the database. Finally, 557 peer-reviewed articles became the foundation for the data collection.

3.4.2 Data organization

The VHFIDB provides comprehensive information about essential as well as antiviral host factors and viruses. We have divided the host factor information page into three parts: host factor information, pathogen information, and publication information.

Host factor information is hitched with: (1) host factor (HF) name, (2) HF protein name, (3) HF function, (4) UniProt ID [22], (5) protein sequence, (6) NCBI gene ID, (7) gene name, (8) gene synonyms, (9) Ensemble gene ID [23], (10) Ensemble transcript, (11) KEGG ID [24], (12) GO ID [25], (13) MINT ID [26], (14) STRING ID [27], (15) OMIM ID [28], (16) PANTHER ID [16], (17) PDB ID [29], (18) Pfam ID [30], (19) DrugBank ID [17], (20) ChEMBL ID [18], (21) Organism.

The virus information is associated with: (1) virus name, (2) virus short name, (3) order, (4) virus family, (5) virus subfamily, (6) genus, (7) species, (8) host, (9) cell tropism, (10) associated disease, (11) mode of transmission, (12) VIPR DB link [31], (13) ICTV DB link [32], (14) virus-host DB link [33]. Similarly, the publication information is hitched with: (1) paper title, (2) author's name, (3) journal name, (4) Pubmed ID, (5) abstract, (6) used model, (7) DOI.

3.4.3 Data update

The VHFIDB database will combine newly published research papers and related VHF data on a half-yearly basis.

3.4.4 Data analysis

The VHFIDB is a flexible web interface that allows both basic and advanced search and browsing options. Besides this, the database also provides three different analysis tools to help in-depth understand host factors involved in viral pathogenesis (Figure 3.2). They are serially enlisted below:

3.4.4.1 VHF Pathogen Network Analysis:

The VHF Pathogen Network Analysis Tool represents the network analysis of the host factor interactions with various viruses. This tool provides the user with a broader view regarding molecular interaction networks. One can either select the host factor's name or the virus's name to visualize the molecular interaction network between them.

3.4.4.2 VHF Overlap Analysis

The interactive heatmap displays the common host genes shared by different viruses. The heatmap generated by this tool may be based on different classes of viruses, families of viruses, or between different viruses.

3.4.4.3 Gene Enrichment Analysis

It is an essential and general analytical tool for functional annotation of VHF genes by using DAVID. Users can investigate the VHF's genes as involved in different metabolic or signaling pathways [16].

3.5 References

1. Nichol, S. T., J. Arikawa and Y. Kawaoka. (2000), Emerging viral diseases, *Proc Natl Acad Sci U S A*, 97, 12411-12412 (10.1073/pnas.210382297)
2. Sanjuan, R., M. R. Nebot, N. Chirico, L. M. Mansky and R. Belshaw. (2010), Viral mutation rates, *J Virol*, 84, 9733-9748 (10.1128/JVI.00694-10)
3. Kautz, T. F. and N. L. Forrester. (2018), RNA Virus Fidelity Mutants: A Useful Tool for Evolutionary Biology or a Complex Challenge?, *Viruses*, 10 (10.3390/v10110600)
4. Phizicky, E. M. and S. Fields. (1995), Protein-protein interactions: methods for detection and analysis, *Microbiol Rev*, 59, 94-123
5. Elbashir, S. M., J. Harborth, W. Lendeckel, A. Yalcin, K. Weber and T. Tuschl. (2001), Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells, *Nature*, 411, 494-498 (10.1038/35078107)
6. Panda, D., A. Das, P. X. Dinh, S. Subramaniam, D. Nayak, N. J. Barrows, J. L. Pearson, J. Thompson, D. L. Kelly, I. Ladunga *et al.* (2011), RNAi screening reveals requirement for host cell secretory pathway in infection by diverse families of negative-strand RNA viruses, *Proc Natl Acad Sci U S A*, 108, 19036-19041 (10.1073/pnas.1113643108)
7. Shalem, O., N. E. Sanjana, E. Hartenian, X. Shi, D. A. Scott, T. Mikkelsen, D. Heckl, B. L. Ebert, D. E. Root, J. G. Doench *et al.* (2014), Genome-scale CRISPR-Cas9 knockout screening in human cells, *Science*, 343, 84-87 (10.1126/science.1247005)
8. Wiedenheft, B., S. H. Sternberg and J. A. Doudna. (2012), RNA-guided genetic silencing systems in bacteria and archaea, *Nature*, 482, 331-338 (10.1038/nature10886)
9. Bhaya, D., M. Davison and R. Barrangou. (2011), CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation, *Annu Rev Genet*, 45, 273-297 (10.1146/annurev-genet-110410-132430)
10. Ma, H., Y. Dang, Y. Wu, G. Jia, E. Anaya, J. Zhang, S. Abraham, J. G. Choi, G. Shi, L. Qi *et al.* (2015), A CRISPR-Based Screen Identifies Genes Essential for West-Nile-Virus-Induced Cell Death, *Cell Rep*, 12, 673-683 (10.1016/j.celrep.2015.06.049)
11. Doench, J. G., N. Fusi, M. Sullender, M. Hegde, E. W. Vaimberg, K. F. Donovan, I. Smith, Z. Tothova, C. Wilen, R. Orchard *et al.* (2016), Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9, *Nat Biotechnol*, 34, 184-191 (10.1038/nbt.3437)
12. Smith, I., P. G. Greenside, T. Natoli, D. L. Lahr, D. Wadden, I. Tirosh, R. Narayan, D. E. Root, T. R. Golub, A. Subramanian *et al.* (2017), Evaluation of RNAi and CRISPR technologies by large-scale gene expression profiling in the Connectivity Map, *PLoS Biol*, 15, e2003213 (10.1371/journal.pbio.2003213)
13. Kleinstiver, B. P., V. Pattanayak, M. S. Prew, S. Q. Tsai, N. T. Nguyen, Z. Zheng and J. K. Joung. (2016), High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects, *Nature*, 529, 490-495 (10.1038/nature16526)
14. Liu, Y., D. Xie, L. Han, H. Bai, F. Li, S. Wang and X. Bo. (2015), EHFPI: a database and analysis resource of essential host factors for pathogenic infection, *Nucleic Acids Res*, 43, D946-955 (10.1093/nar/gku1086)

15. Thakur, A., A. Qureshi and M. Kumar. (2017), vhfRNAi: a web-platform for analysis of host genes involved in viral infections discovered by genome wide RNAi screens, *Mol Biosyst*, 13, 1377-1387 (10.1039/c6mb00841k)
16. Mi, H., A. Muruganujan, D. Ebert, X. Huang and P. D. Thomas. (2019), PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools, *Nucleic Acids Res*, 47, D419-D426 (10.1093/nar/gky1038)
17. Wishart, D. S., Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda *et al.* (2018), DrugBank 5.0: a major update to the DrugBank database for 2018, *Nucleic Acids Res*, 46, D1074-D1082 (10.1093/nar/gkx1037)
18. Gaulton, A., A. Hersey, M. Nowotka, A. P. Bento, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L. J. Bellis, E. Cibrian-Uhalte *et al.* (2017), The ChEMBL database in 2017, *Nucleic Acids Res*, 45, D945-D954 (10.1093/nar/gkw1074)
19. Shapira, S. D., I. Gat-Viks, B. O. Shum, A. Dricot, M. M. de Grace, L. Wu, P. B. Gupta, T. Hao, S. J. Silver, D. E. Root *et al.* (2009), A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection, *Cell*, 139, 1255-1267 (10.1016/j.cell.2009.12.018)
20. Hoffmann, M., H. Kleine-Weber, S. Schroeder, N. Kruger, T. Herrler, S. Erichsen, T. S. Schiergens, G. Herrler, N. H. Wu, A. Nitsche *et al.* (2020), SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor, *Cell* (10.1016/j.cell.2020.02.052)
21. Li, W., M. J. Moore, N. Vasilieva, J. Sui, S. K. Wong, M. A. Berne, M. Somasundaran, J. L. Sullivan, K. Luzuriaga, T. C. Greenough *et al.* (2003), Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus, *Nature*, 426, 450-454 (10.1038/nature02145)
22. Apweiler, R., A. Bairoch, C. H. Wu, W. C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane *et al.* (2004), UniProt: the Universal Protein knowledgebase, *Nucleic Acids Res*, 32, D115-119 (10.1093/nar/gkh131)
23. Zerbino, D. R., P. Achuthan, W. Akanni, M. R. Amode, D. Barrell, J. Bhai, K. Billis, C. Cummins, A. Gall, C. G. Giron *et al.* (2018), Ensembl 2018, *Nucleic Acids Res*, 46, D754-D761 (10.1093/nar/gkx1098)
24. Kanehisa, M., M. Furumichi, M. Tanabe, Y. Sato and K. Morishima. (2017), KEGG: new perspectives on genomes, pathways, diseases and drugs, *Nucleic Acids Res*, 45, D353-D361 (10.1093/nar/gkw1092)
25. Harris, M. A., J. Clark, A. Ireland, J. Lomax, M. Ashburner, R. Foulger, K. Eilbeck, S. Lewis, B. Marshall, C. Mungall *et al.* (2004), The Gene Ontology (GO) database and informatics resource, *Nucleic Acids Res*, 32, D258-261 (10.1093/nar/gkh036)
26. Licata, L., L. Briganti, D. Peluso, L. Perfetto, M. Iannuccelli, E. Galeota, F. Sacco, A. Palma, A. P. Nardozza, E. Santonico *et al.* (2012), MINT, the molecular interaction database: 2012 update, *Nucleic Acids Res*, 40, D857-861 (10.1093/nar/gkr930)
27. Szklarczyk, D., J. H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N. T. Doncheva, A. Roth, P. Bork *et al.* (2017), The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible, *Nucleic Acids Res*, 45, D362-D368 (10.1093/nar/gkw937)

28. Hamosh, A., A. F. Scott, J. S. Amberger, C. A. Bocchini and V. A. McKusick. (2005), Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders, *Nucleic Acids Res*, 33, D514-517 (10.1093/nar/gki033)
29. Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne. (2000), The Protein Data Bank, *Nucleic Acids Res*, 28, 235-242 (10.1093/nar/28.1.235)
30. Finn, R. D., A. Bateman, J. Clements, P. Coghill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry *et al.* (2014), Pfam: the protein families database, *Nucleic Acids Res*, 42, D222-230 (10.1093/nar/gkt1223)
31. Pickett, B. E., E. L. Sadat, Y. Zhang, J. M. Noronha, R. B. Squires, V. Hunt, M. Liu, S. Kumar, S. Zaremba, Z. Gu *et al.* (2012), ViPR: an open bioinformatics database and analysis resource for virology research, *Nucleic Acids Res*, 40, D593-598 (10.1093/nar/gkr859)
32. Lefkowitz, E. J., D. M. Dempsey, R. C. Hendrickson, R. J. Orton, S. G. Siddell and D. B. Smith. (2018), Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV), *Nucleic Acids Res*, 46, D708-D717 (10.1093/nar/gkx932)
33. Mihara, T., Y. Nishimura, Y. Shimizu, H. Nishiyama, G. Yoshikawa, H. Uehara, P. Hingamp, S. Goto and H. Ogata. (2016), Linking Virus Genomes with Host Taxonomy, *Viruses*, 8, 66 (10.3390/v8030066)

CHAPTER 4

Chapter 4

CHPV Pseudovirus generation and identification of virus entry mechanism

4.1 Introduction

Chandipura virus (CHPV) is an emerging human pathogen, mainly reported from many areas of the Indian subcontinent, with a fatality of around 55 to 77%. CHPV causes brain encephalitis in children below the age of 15[1]. Its symptoms are similar to other encephalitides and are characterized by acute fever, diarrhea, altered sensorium, seizures, and vomiting[2].

CHPV is an enveloped (–) single-stranded RNA virus, belongs to the Rhabdoviridae family. It's a vector-borne virus transmitted by sandflies. CHPV genome size is 11 kb; it encodes five genes, namely, nucleoprotein (N), matrix protein (M), phosphoprotein (P), glycoprotein (G), large polymerase protein (L), and glycoprotein (G). Chandipura virus was first discovered and isolated accidentally from the two adults' blood with a febrile illness in a village in Nagpur district, Maharashtra, India, in 1965.[3] The only other instance when it was isolated from a human in 1980 was in Madhya Pradesh, India, from an acute encephalitis patient.[4] Although the retrospective serological studies indicate the enduring history of CHPV and that it infected the human population as early as 1957-58.[5] However, due to the low case reports, it has not got much attention until the 2003 outbreak in Andhra Pradesh, India. The epidemic started in June 2003 and continued for the next 3-4 months. Around 329 children (age between 9 months -14 years) infected with the virus showed encephalitis, and 183 died.[6] Simultaneously, a CHPV encephalitis outbreak was also reported in fifteen districts of Maharashtra during the same time. In the subsequent year 2005, another outbreak was reported in the Baroda district of Gujrat, India, with a 70 % mortality rate in the pediatric population.[7] However, many small outbreaks are reported

eventually throughout the country during different periods. [8,9] Although CHPV survives endemically in limited parts of the Indian subcontinent, it has a high potential of epidemic or pandemic because its vector can cross the current borders. Due to limited research facilities, the detailed investigation on CHPV is slow yet necessary.

All viruses are a non-cellular life form that must find some way to enter the host cell to complete their life cycle.[10] Most of the enveloped viruses enter the cells by the two primary pathways. Some viruses bind to the cell surface receptor and deliver their genome to the cytoplasm, while others use cellular endocytosis machinery.[11,12] Different viruses use different paths, including clathrin-mediated endocytosis (CME), caveolae-mediated endocytosis, lipid raft-mediated endocytosis, macropinocytosis, etc.[13,14]

The most common and well-studied cellular endocytosis pathway is the clathrin-mediated pathway, which many of the viruses use. After endocytosis, some viruses like vesicular stomatitis virus (VSV), Crimean-Congo hemorrhagic fever virus (CCHFV), and influenza viruses require acidified conditions to activate their fusion proteins, and they require internalization by endocytic vesicles to reach the cytosol.[15-17]

Bone Marrow Stromal Antigen 2 (BST-2) was discovered as a surface marker of neoplastic B cells. The protein present in almost all cell types; however, the degree of expression varies from cell to cell. BST-2 is a type II transmembrane protein and contains ~180 amino acids (aa). BST-2 inhibits most of the envelope viruses released from the cell surface.[18]

It was also reported once that it enhances human cytomegalovirus (HCMV) entry[19]. The receptor and entry mechanism for CHPV is unknown. A member of rhabdovirus VSV uses a Low-density lipoprotein receptor (LDL-R) as a primary receptor. VSV internalization occurs by clathrin-mediated endocytosis while acidic pH and, lipid-raft is also required.[20-22]

To identify the molecular mechanism of CHPV entry. We have generated a VSV backbone-based fluorescence CHPV pseudovirus, and we have also characterized it. We have demonstrated that CHPV pseudovirus shows clathrin-mediated entry with that lipid-raft and low pH also required for CHPV entry. Moreover, we have reported that the mouse and human tetherin host factor enhance CHPV virus entry.

4.2. Results

4.2.1. Generation and characterization of CHPV Pseudovirus

The CHPV G pseudotyped virus is produced using a VSV pseudotyped virus packaging system. This system was previously used to generate various pseudotyped viruses, such as SARS-CoV-2, Ebola virus, Nipah virus, Hantaan virus, Hepatitis c virus, and Rift valley fever virus.[23-28] The pseudotyped virus's backbone came from the VSV virus. The G gene of VSV was replaced with the Teal fluorescence protein (TFP) gene, and the G protein from CHPV was provided as the membrane protein on the surface of the VSV pseudotyped virus. The schematic representation of the CHPV pseudotyped virus generation is shown in figure 4.1. During the preparation of CHPV pseudovirus, BHK-21 cells were transfected with the CHPV G expressing plasmid. After 30 hours of transfection, the cells were infected with G*ΔG-VSV. After infection, the uncoated ΔG-VSV genome expressed the structural proteins and the enzymes from the VSV genome and completes genome replication.

The expressed incomplete genome assembled into virus particles without an envelope. The viruses were discharged from the cells by budding using the CHPV envelope protein present on the cell membrane. So, the new virus particles were termed CHPV pseudovirus. After 24 hours of infection, the supernatant was collected, filtered with a 0.45 μm filter, and kept at -80°C. CHPV pseudovirus presence was further confirmed by western blot analysis and LC-MS (Figure 4.2). Most interestingly, we have found some human proteins attached to the CHPV pseudovirus (Appendix C). Because we have used VSV Pseudovirus as a control

for LC-MS and respective human proteins were not observed with its data, Which confirmed that these proteins specifically bind with CHPV glycoprotein. Further, we have examined the CHPV Pseudovirus morphology by using Transmission Electron Microscope (TEM). It looked circular, and its size was around 70-100 nm (Figure 4.3).

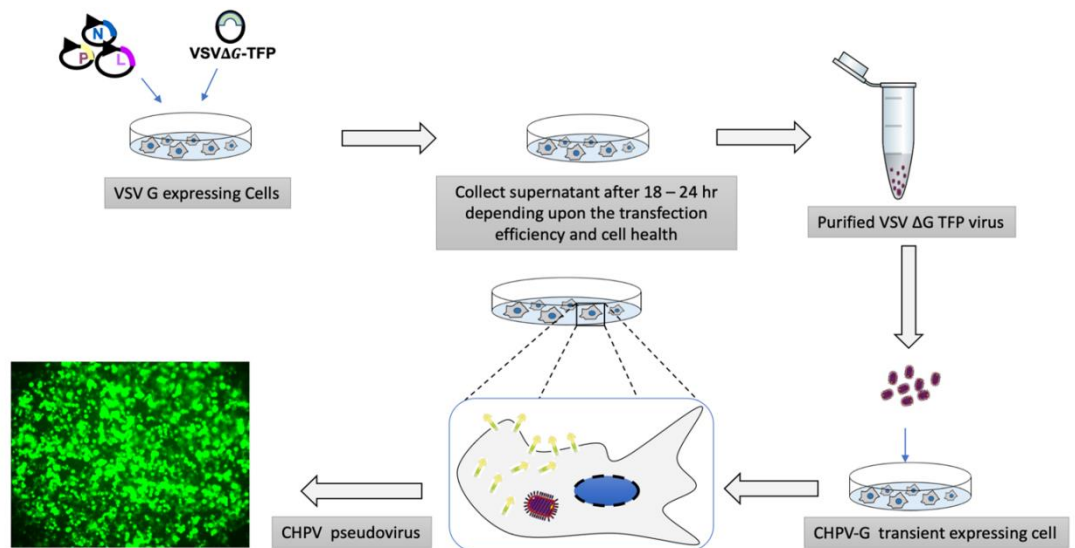


Figure 4.1. Schematic diagram of CHPV Pseudovirus production. At first, to generate VSV pseudovirus, we have replaced VSV Glycoprotein from TFP and use pCAG VSV-G for packing (G*ΔG-VSV). 70% confluent 100 mm dish of BHK-21 cells were transfected with N, P, L, G, *ΔG-VSV-TFP and generate G*ΔG-VSV-TFP, and used for further experiment. Next, for CHPV pseudoviruses construction, envelope genes of CHPV from Nagpur strain (strain No. 1653514) were cloned into the eukaryotic expression plasmid pCAG and transfected in BHK21 cells, which is transcribed and expresses CHPV glycoprotein that is transferred to and attached to the cell membrane. Twenty-four hours later, transfected cells infected with G*ΔG-VSV-TFP with 1 MOI. one hour after infection, cells were washed with PBS twice, and then 2% DMEM was added. Twenty-four hours post-infection, CHPV pseudovirus carrying culture supernatants were collected, filtered, tested, and stored in -80°C .

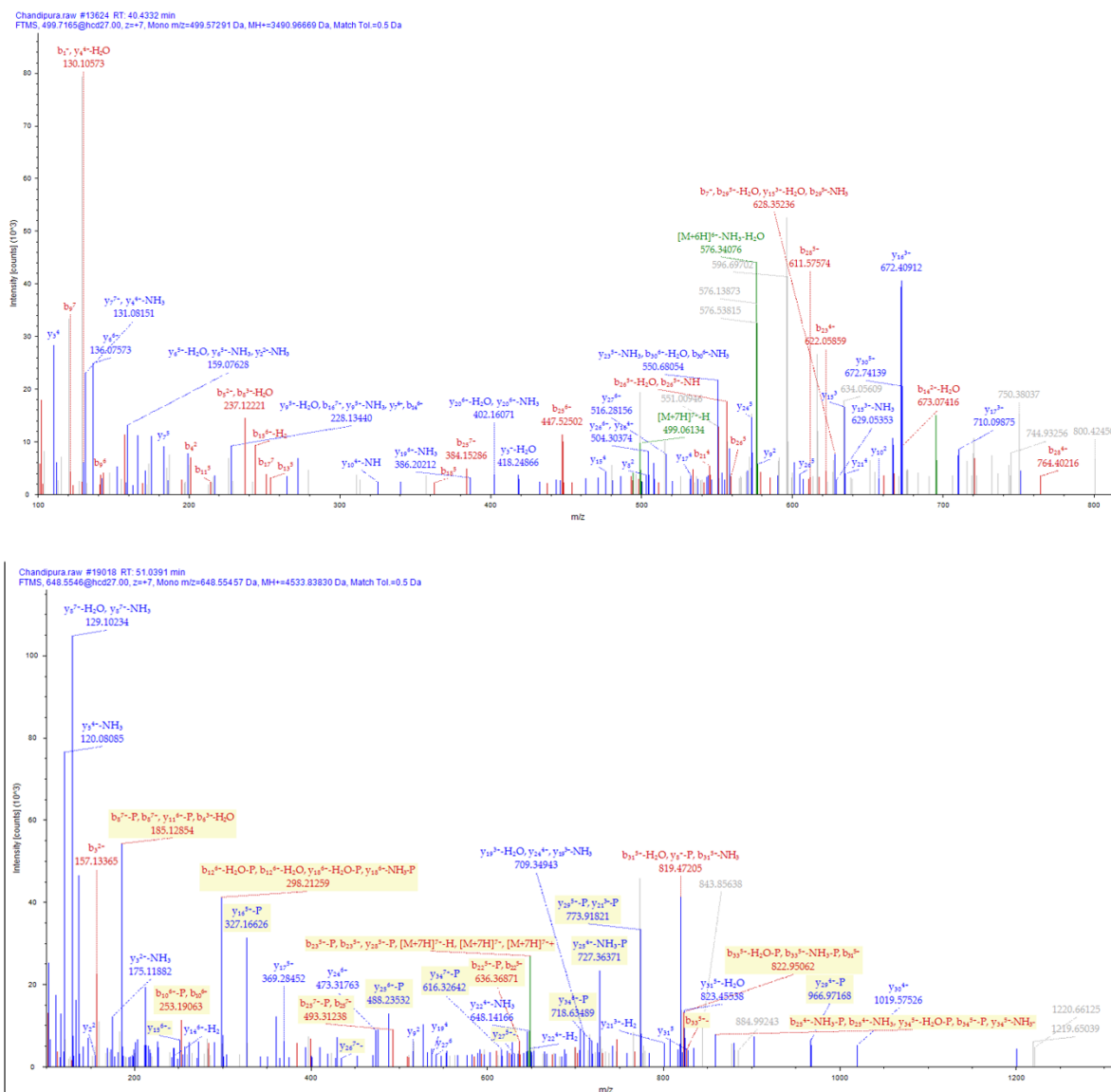


Figure 4.2. LC-MS data of CHPV Pseudovirus. 70% confluent 100 mm dish of BHK-21 cells were transfected with 5 μ g of pCAG-CHPV-G by using Lipofectamine 2000. Thirty hours later, transfected cells infected with G* Δ G-VSV-TFP with 1 MOI. one hour after infection, cells were washed with PBS twice, and then 2% DMEM was added. Twenty-four hours post-infection, CHPV pseudovirus carrying culture supernatants were collected, filtered (0.45- μ m pore size, Millipore, SLHP033RB). After filtration centrifuge it at 1×10^6 RPM for 2 hours, and send it to V-proteomics a Delhi (India) based company for LC-MS. LC-MS data confirms CHPV glycoprotein.

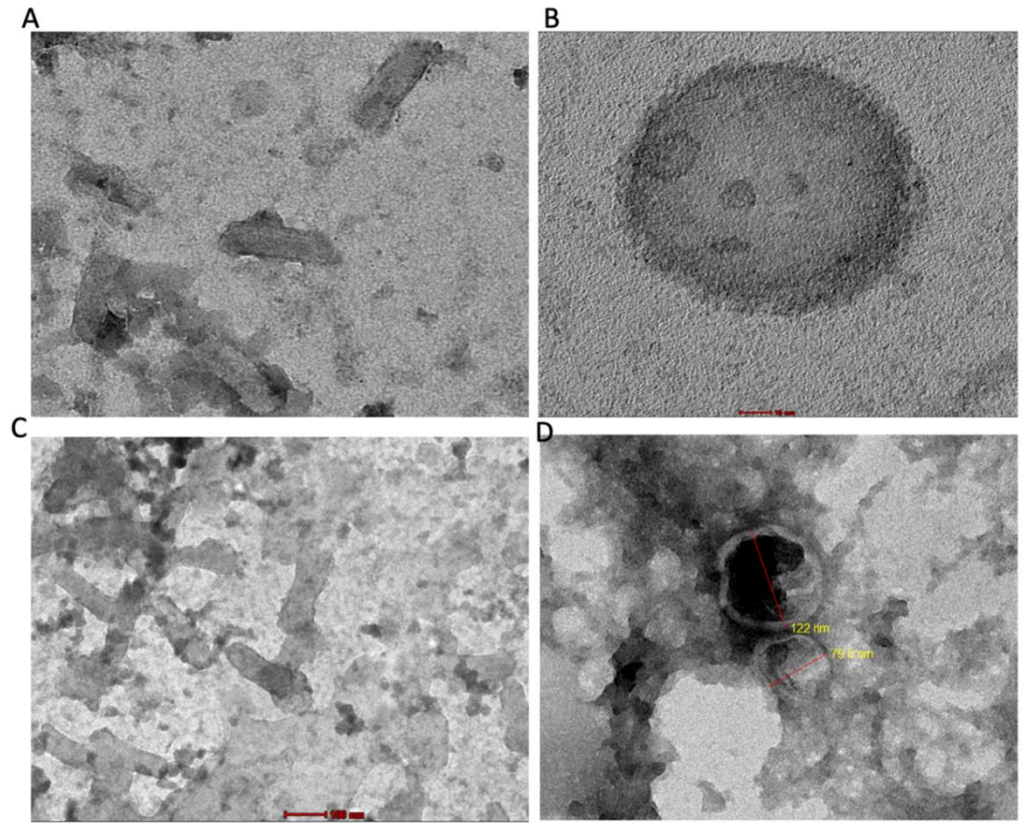
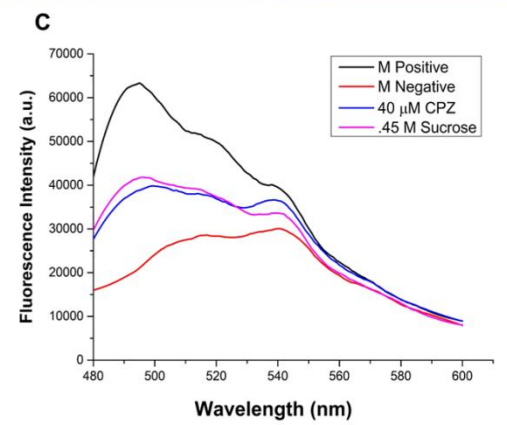
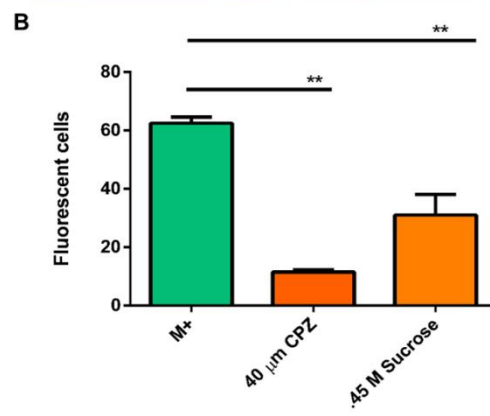
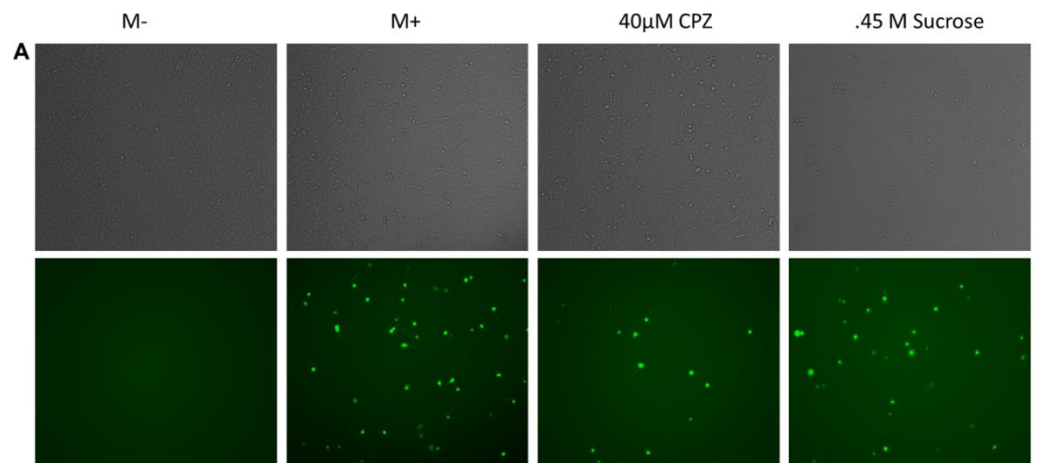


Figure 4.3. *The morphological characterization of CHPV Wt (left) and CHPV pseudovirus (right) CHPV pseudovirus generated as mentioned above. For the TEM imaging, samples were placed onto 400-mesh copper grids (SIGMA aldrich chemicals) and incubated for 5 minutes. The grids were then rinsed twice with DI water, wicked dry, and stained with 2% uranyl acetate for 5 min. CHPV virus looked bullet shape (left) while CHPV Pseudovirus looked circular, and its size was around 70-100 nm.*

4.2.2. Clathrin is essential for CHPV entry:

After generating the CHPV pseudovirus, we attempted to identify the specific endocytosis pathway used by the virus.[14] The clathrin-dependent pathway is the most common pathway used by animal viruses. To obtain the role of clathrin-dependent endocytosis in CHPV entry, we have investigated the inhibitory role of chlorpromazine (CPZ) and sucrose. CPZ and sucrose cause clathrin lattice to gather on endosomal membranes and, simultaneously, prevent coated pits' assembly at the cell surface. We have examined the effect of CPZ and sucrose on CHPV pseudovirus using two techniques: microscopy and spectrofluorimetry. At first, we have treated Vero and Vero E6 cells with 40 μ m of CPZ. CHPV pseudoviruses were infected after 45 min of drug treatment in the presence of drug. After 1.5 h, cells were washed, and infection was maintained in fresh 2% culture media till 24 hours post-infection. After 24 hpi, TFP fluorescence was observed with microscopy (figure 4.4). Relative infectivity of the viruses was determined by measuring TFP expression level using the spectrofluorometer (figure 4.4). Our results demonstrated that CPZ and sucrose significantly inhibit virus entry, and CHPV enters into Vero and Vero E6 cells by using clathrin-mediated endocytosis.



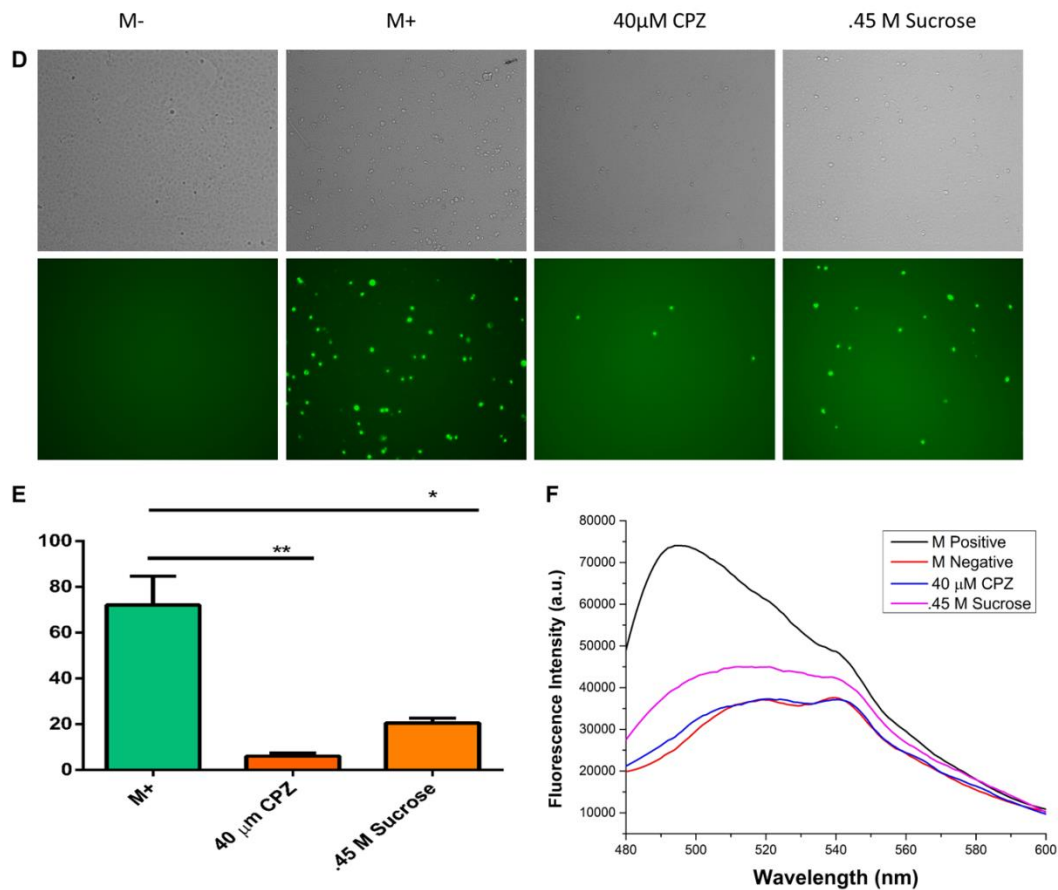
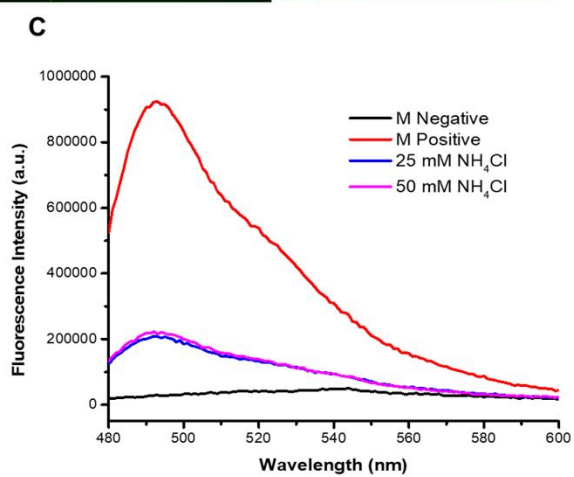
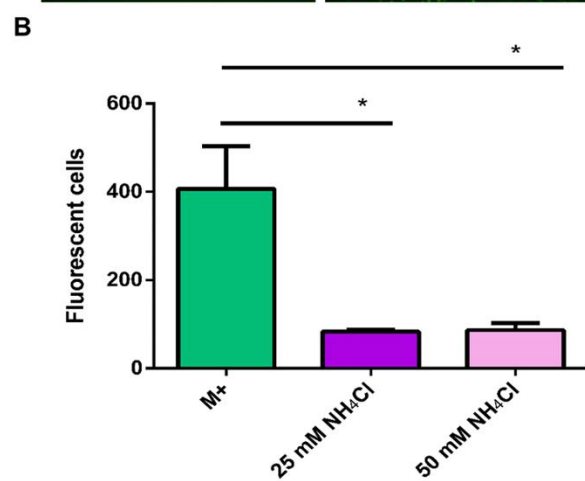
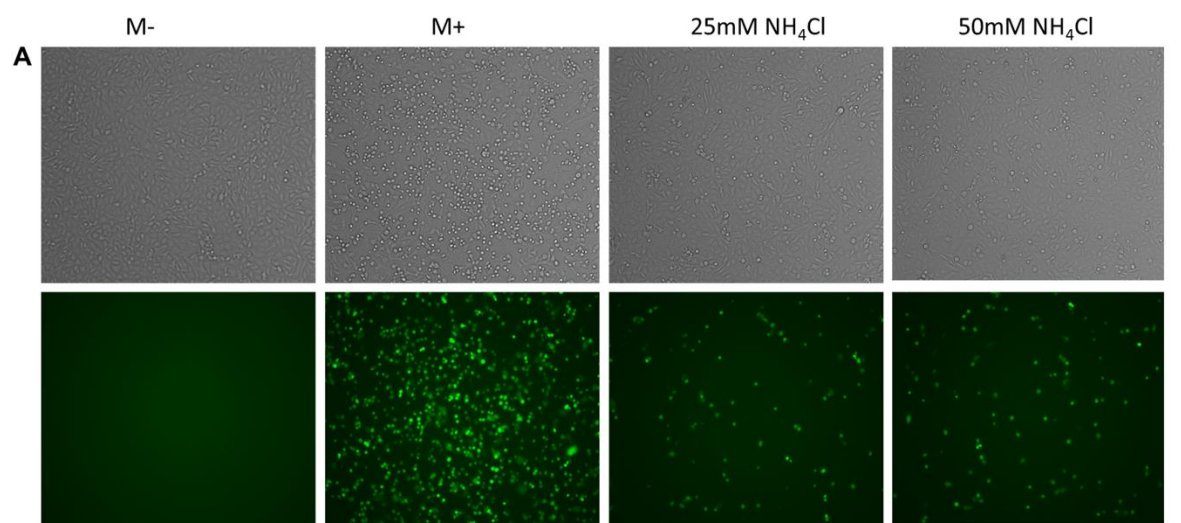


Figure 4.4. Clathrin is essential for CHPV entry. Vero and Veo E6 cells were treated with 40 μ m CPZ and .45 M sucrose. After 45 min, cells were infected with CHPV pseudovirus in the presence of a drug. Microscopic imaging was performed after 24 hours (A, D). Fluorescence cells counted by Image j software (B, E), CHPV pseudovirus TFP infectivity was measured using a spectrophotometer at 24 h of infection(C, F). Error Bar represents the SD of three independent experiments.

4.2.3. CHPV entry is pH-dependent

Viruses that enter the host cell via endocytosis are normally trafficked to early endosomes for sorting and then transported to late the late endosome.[29] Since we have observed that CHPV pseudovirus entry depends upon clathrin-mediated endocytosis. We have also investigated the potential role of pH in CHPV entry. For that, we have used NH_4Cl , an inhibitor of endosome acidification.[30] Vero and Vero E6 cells were pretreated with two different concentrations of NH_4Cl (25mM & 50 mM) for 45 min and infected with CHPV Pseudovirus in the presence of NH_4Cl . Cells were washed, and 2% fresh media was added after 1.5 h. After 24 hours, fluorescence was determined by microscopy and spectrofluorometer (Figure 4.5). We have observed that interference with endosome acidification significantly -reduced CHPV fluorescence level, and CHPV also displayed a similar mechanism as VSV, HCV, SACRS-COV do.[22,31,32]



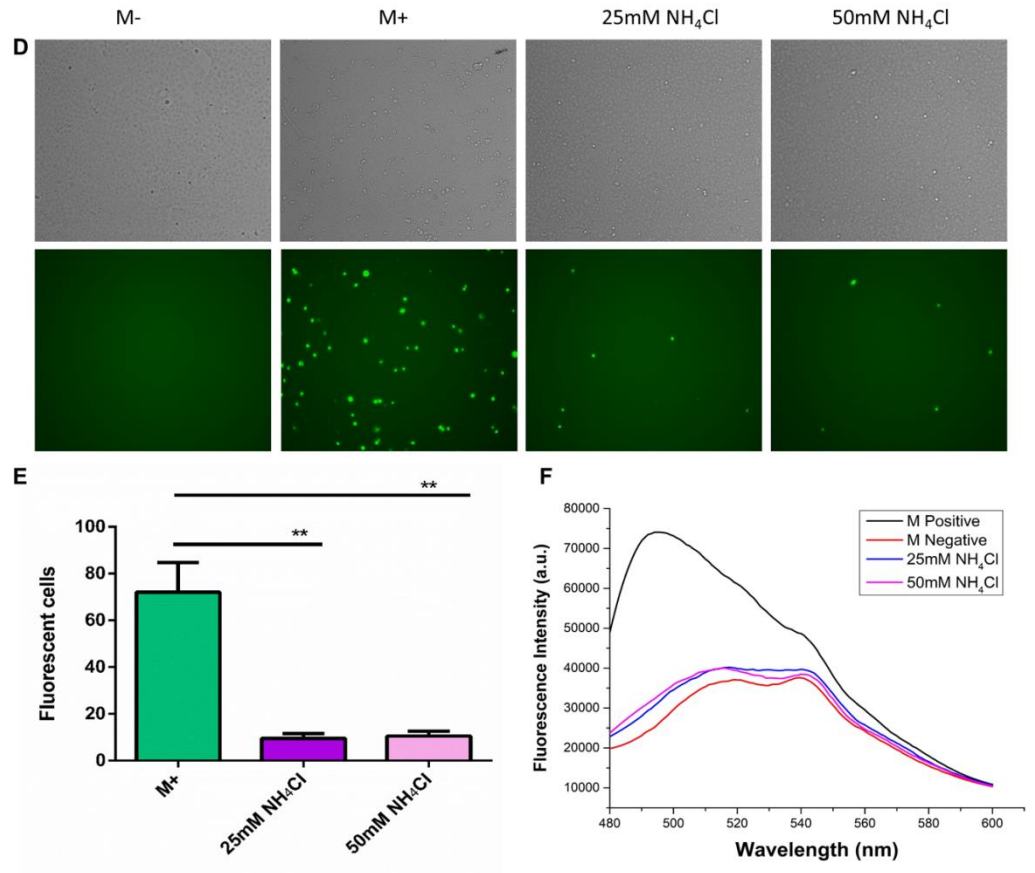
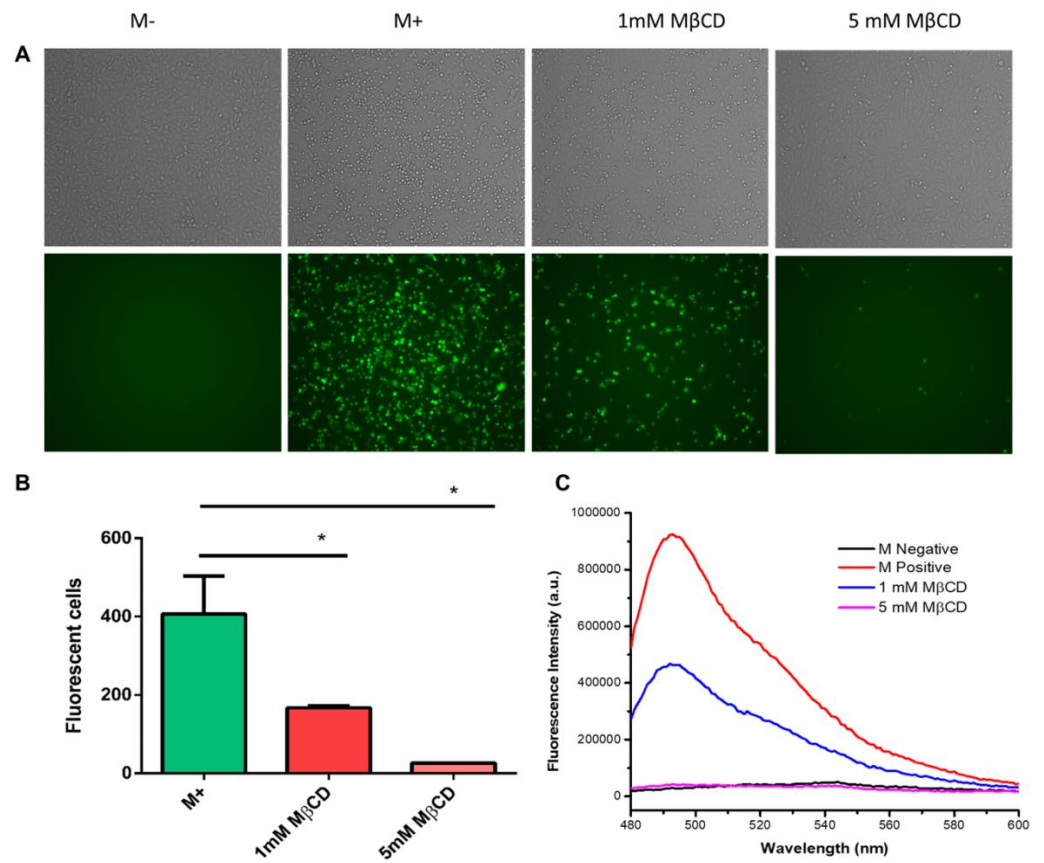


Figure 4.5. CHPV entry is pH dependent. Vero and Veo E6 cells were treated with 25 mM NH₄Cl and 50 mM NH₄Cl. After 45 min, cells were infected with CHPV pseudovirus in the presence of a drug. Microscopic imaging was performed after 24 hours (A, D). Fluorescence cells counted by Image j software (B, E), CHPV pseudovirus TFP infectivity was measured using a spectrophotometer at 24 h of infection (C, F). Error Bar represents the SD of three independent experiments.

4.2.4. Lipid raft is essential for CHPV entry

After discerning the importance of clathrin and pH for CHPV entry, we have also investigated the role of lipid raft in virus entry. Lipid rafts are membrane microdomains loaded with sphingolipids and cholesterol and contain various proteins associated with signalling.[33] In respect of viral access, lipid raft is usually used as entry gateways.[34-36] To investigate lipid raft's role, we have used a drug named methyl- β -cyclodextrin (M β CD). Vero and Vero E6 cells pretreated with 1 & 5 mM of M β CD for 45 min, then CHPV pseudovirus were infected for 1.5 h. Cells were washed, and fresh 2% media was added. After 24 hours, fluorescence was determined by microscopy and spectrofluorometer. In our results, we have found that CHPV virus entry was reduced in an M β CD concentration-dependent manner (Figure 4.6). In 5 mM of M β CD, very few viruses reached inside the cells compared to that control and in the presence of 1mM M β CD. These results suggest that lipid raft is essential for CHPV pseudovirus entry.



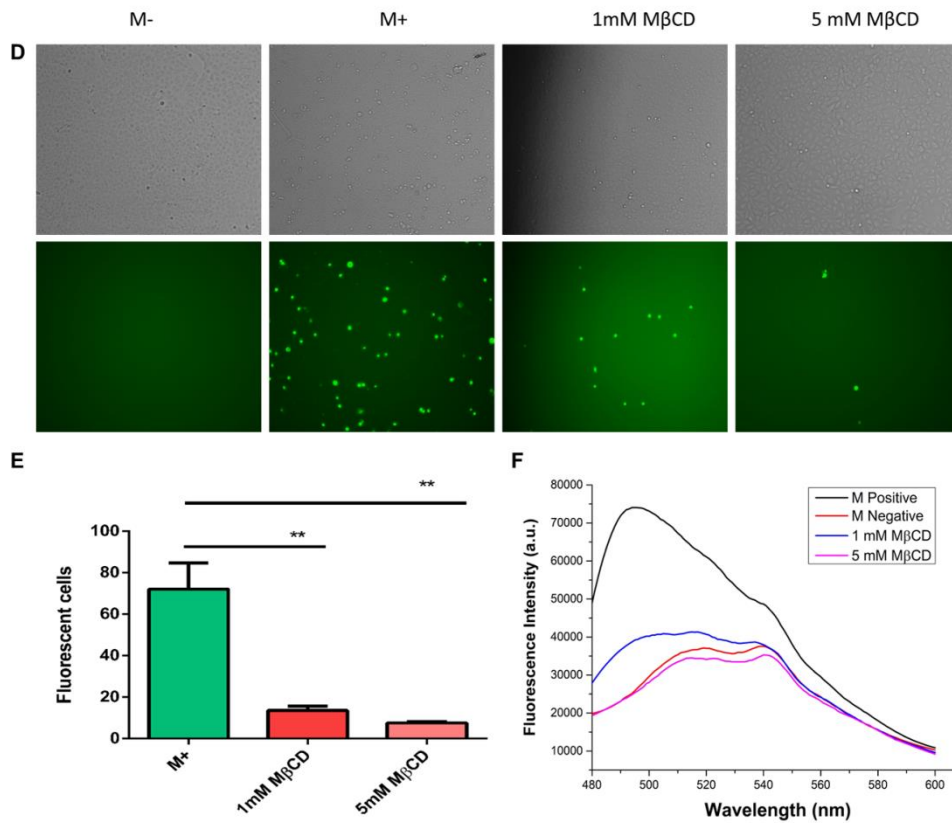


Figure 4.6. Lipid Raft is important for CHPV entry. Vero and Vero E6 cells were treated with 1 mM MβCD and 5 mM MβCD. After 45 min, cells were infected with CHPV pseudovirus in the presence of a drug. Microscopic imaging was performed after 24 hours (A, D). Fluorescence cells counted by Image j software (B, E), CHPV pseudovirus TFP infectivity was measured using a spectrophotometer at 24 h of infection (C, F). Error Bar represents the SD of three independent experiments.

4.2.5. BST-2 enhance CHPV entry

Further, we have investigated the role of BST-2 in CHPV infection. Normally BST-2 inhibits the release of enveloped viruses.[37,38] Our primary investigation found that virus titer will be high in BST-2 expressing cells. To check the role of BST-2 in CHPV entry, we determined the amount of CHPV glycoprotein by qPCR from BST2-transient expressing or control BHK-21 cells immediately after virus infection. We have transfected BHK-21 cells with 2ug of pCAG-M-BST-2 or pCAG-H-BST-2 (M: Mouse, H: Human). After 32 hours of transfection, cells were infected with 10 MOI of CHPV for 2 h at 4°C. The cells were then washed with citric acid buffer (pH = 3), following PBS wash, to remove adhered virions from the cell surface. Cells were then trypsinized, Viral RNA was extracted from BHK-21 cells, and qPCR determined viral genome copies. At 1hpi, BST2 expressing cells carried an increased amount of CHPV viral genome. This result concludes that BST-2 helps in viral entry (Figure 4.7 A, B, E, F). To rule out that increased viral copy numbers were due to increased viral genome replication. We infected U.V. inactivated CHPV, which can enter the cells but unable to replicate. We have taken all experiment conditions similar as we have taken for the untreated virus. An Increased amount of viral RNA was recovered from BST-2 expressing cells (Figure 4.7 C, D, G, H). This result confirmed that BST-2 increases viral infection by enhancing the virus entry.

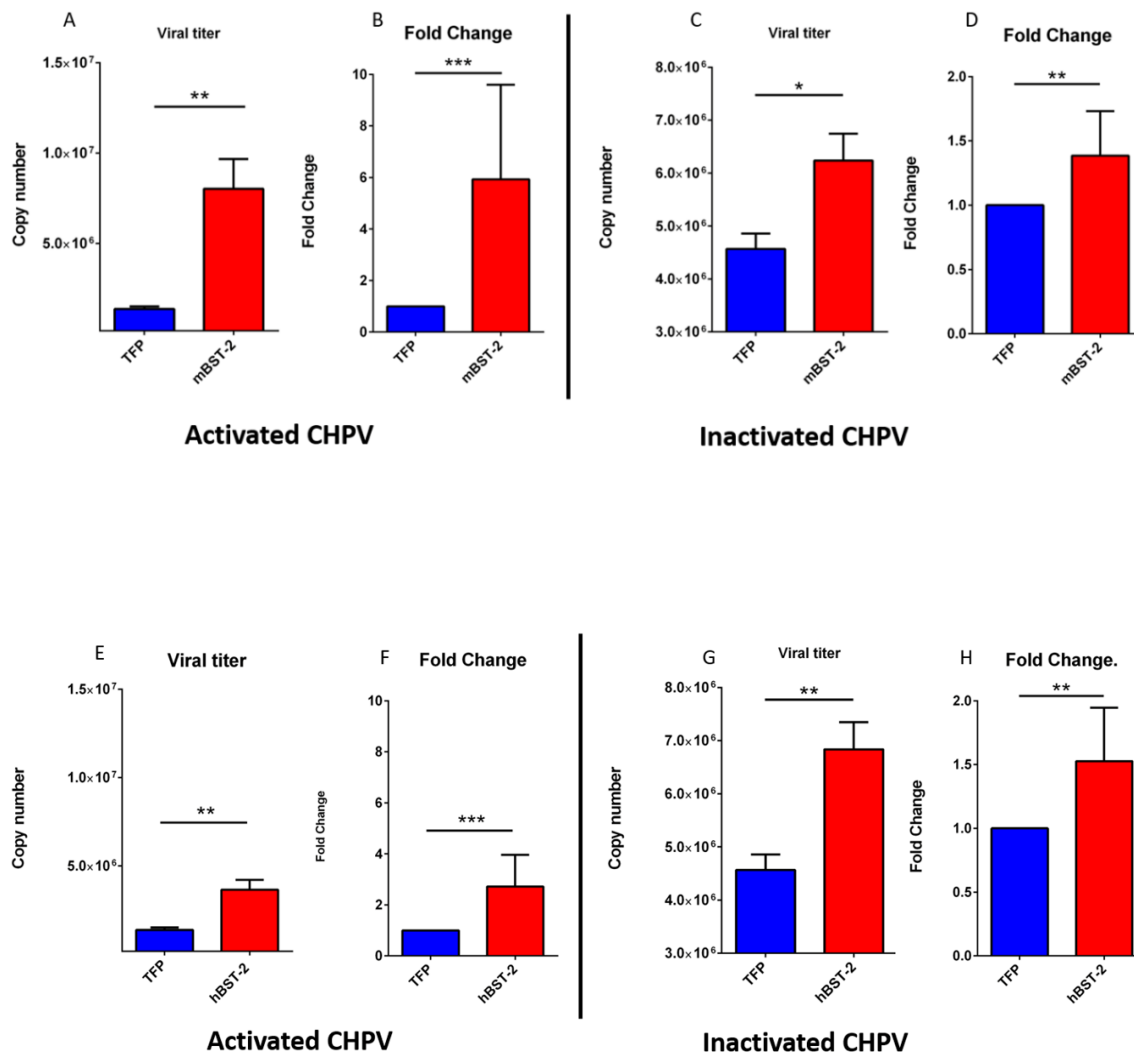


Figure 4.7. *BST-2 enhance CHPV entry. BHK-21 cells were transfected with Murine and Human BST-2. After 32 hours of transfection, cells were infected with 10 MOI of CHPV (A, B, E, F) and UV inactivated CHPV (C, D, G, H) for 2h at 4 C (p > 0.05). then, cells were washed with citric acid buffer (pH 3), following PBS wash, viral RNA was extracted from BHK-21 cells and quantified by qPCR.*

4.2.6. In-Silico study reveals that the cytoplasmic domain of BST-2 interacts with CHPV Glycoprotein

In the previous experiment, we have confirmed that BST-2 enhances CHPV entry. Further, we want to know how BST-2 increases viral infection. Either it directly interacts with BST-2, or it increases infection by other means. To investigate this, we have performed a docking experiment between the cytoplasmic domain of BST2 (PDB ID: 4P6Z: T) and CHPV glycoprotein (PDB ID: 4DEW). In our result, we have found that the cytoplasmic domain of BST-2 directly interacts with CHPV-G and, Tyr6 and Tyr8 of BST-2 cytoplasmic domain interact with Glu-405, Glu-38, Thr-188 of CHPV glycoprotein via electrostatic interaction and Thr4 of BST-2 cytoplasmic domain interact with Glu-399 and Glu-402 of CHPV-G.

This result is also supported by the previous publication, which shows that the YXY region of the cytoplasmic domain of BST-2 is essential for clathrin-mediated endocytosis, and we have also demonstrated that CHPV uses clathrin-mediated endocytosis for entry (Figure 4.8).[39]

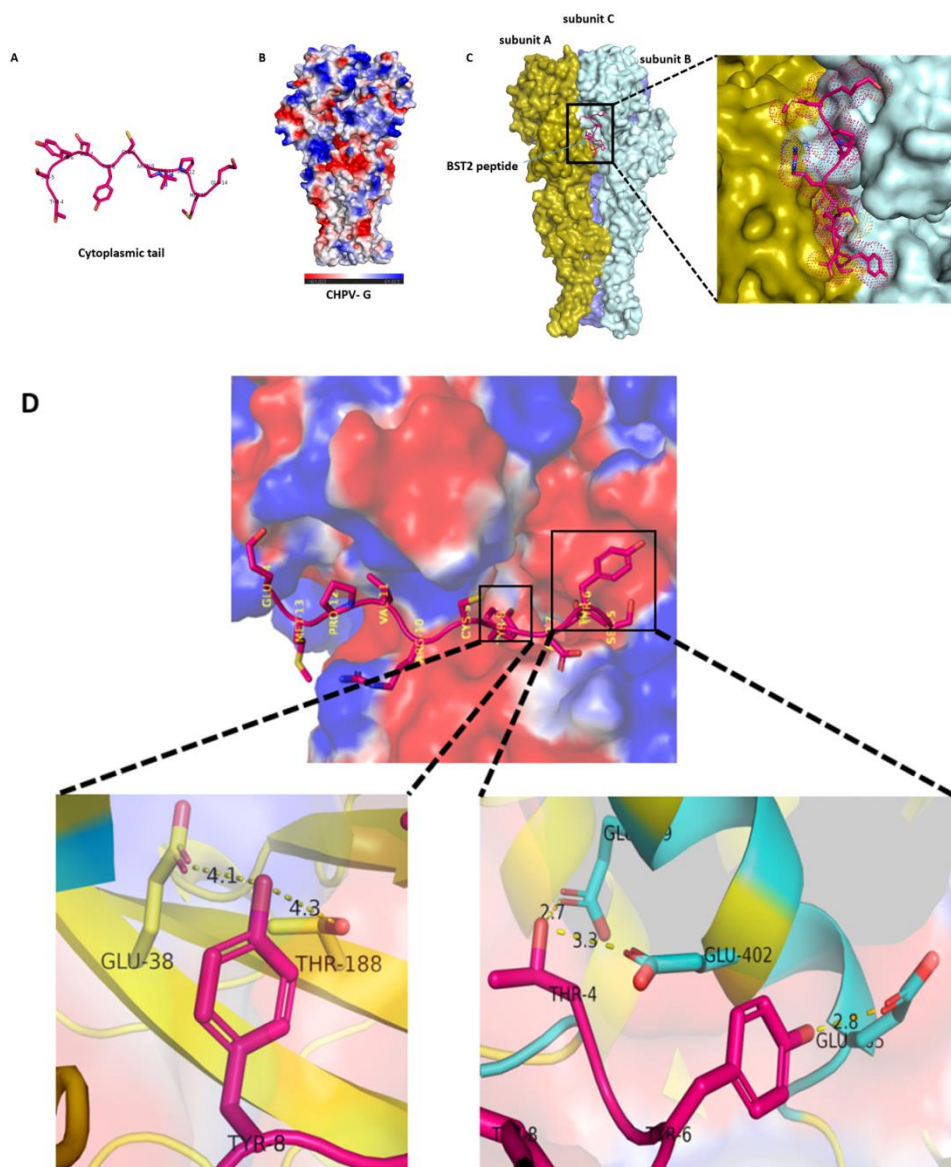


Figure 4.8. Cytoplasmic domain of BST-2 interact with CHPV-G. We have performed molecular docking by using Autodock vina. The cytoplasmic chain of BST-2 was extracted from PDB: 4P6Z and performed docking with CHPV-G, PDB: 4DEW (A, B). The cytoplasmic domain of BST-2 directly interacts with CHPV-G and, Tyr6 and Tyr8 of BST-2 cytoplasmic domain interact with Glu-405, Glu-38, Thr-188 of CHPV glycoprotein via electrostatic interaction, and Thr4 of BST-2 cytoplasmic domain interact with Glu-399 and Glu-402 of CHPV-G (C, D).

4.3. Discussion

In this study, we have generated and characterized the VSV-based CHPV fluorescence pseudovirus and also described the CHPV entry mechanism. Viruses are obligate parasites.[31] They must enter the host cell before they can start their life cycle. Some viruses interact with their receptor and directly discharge their genome to the cytoplasm. While other use cells endocytic machinery. Usually, direct interaction with the plasma membrane does not require low pH, while the endocytic pathway's entry depends upon the low pH of late endocytic vesical.[12,40,41] It was previously thought that a single virus could use either direct membrane fusion or endocytosis method for their entry, not both. For example, before 2000, HIV and Newcastle disease virus was only known for direct fusion entry, but later, it was reported that they also used endocytic pathways for access in the cells.[42,43] Our study observed that CHPV pseudovirus shows more negligible fluorescence in CPZ and sucrose-treated cells. CPZ and sucrose are known for clathrin-mediated endocytosis inhibitors. This confirms CHPV uses clathrin for their entry. Other rhabdoviruses also used a similar mechanism(ref). Further, we have investigated the role of pH in CHPV entry, and we have observed that low pH facilitates CHPV entry. VSV, CCHFV, and HCV also required low pH in the late endosome for productive infection.[16,22] In our next step, we have examined the role of lipid raft in CHPV entry. In respect of viral entry, lipid raft is commonly used as an entry portal.[44,45] In our results, we have found that virus fluorescence is decreased in a dose-dependent manner on MBCD treatment, suggesting that lipid raft also essential for CHPV entry. Moreover, we have also checked the effect of a host factor, BST-2, on virus infection. BST-2 is highly famous for its antiviral property; it inhibits enveloped viruses released from the cell membrane.[18] More interestingly, we have got different results. BST -2 increased CHPV infection. Although a similar observation was reported for HCMV.[19] In our further investigation, we have found that BST-2 enhances CHPV entry. in-silico study reveals that BST-2 cytoplasmic domain directly interacts with CHPV-G.

However, CHPV is a highly contagious virus reported in the Indian subcontinent, with a case fatality of around 55 to 77%. But there are significantly fewer studies occur related to virus pathogenesis, specifically regarding the virus entry domain. We are probably the first who made the VSV-based fluorescence pseudovirus and identified the molecular entry mechanism of CHPV to the best of our knowledge (Figure 4.9).

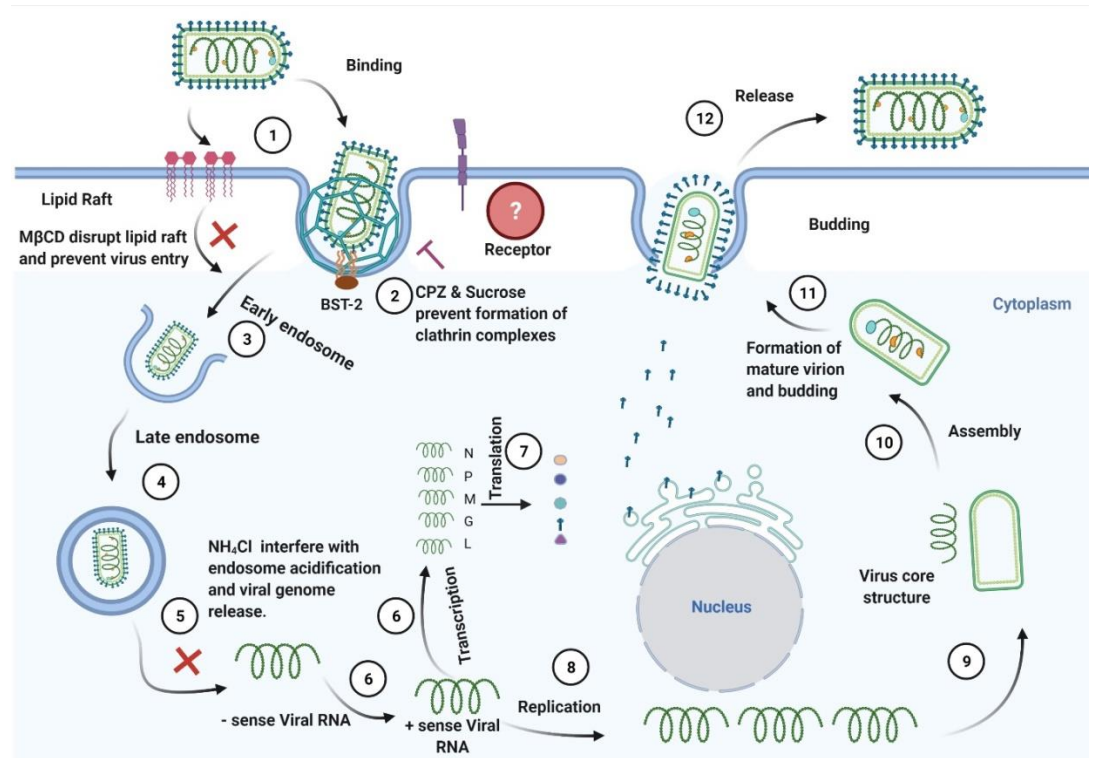


Figure 4.9. Proposed schematic representation of CHPV entry. CHPV binds with cells and enters into the cells by clathrin-mediated endocytosis. Lipid raft and BST-2 also play a crucial role in virus entry. After entry, first, it will move to early endosome than late endosome. After that, viral RNA is released in the cytoplasm at low pH.

4.4. Material method

4.4.1 Chemicals

Dulbecco's modified Eagle's medium (DMEM), OptiMEM, L-glutamine, Sodium pyruvate and goat, fetal bovine serum (FBS), Lipofectamine 2000 were purchased from Invitrogen. X-tremeGENE HP, Chlorpromazine, Ammonium chloride, methyl- β -cyclodextrin, sucrose, and all other chemicals were from Sigma.

4.4.2 Antibody

Mouse monoclonal antibody against BST-2 was purchased from SANTA CRUZ ((E-4): sc-390719), CHPV antibody was obtained from Abegenex, Rabbit polyclonal anti-actin antibody (A2066), anti-rabbit secondary antibody (A9169), anti-mouse secondary antibody (A4416) were obtained from Sigma.

4.4.2 Cell lines

293T human embryonic kidney cells (HEK 293T), Vero, and Vero E6 cells were obtained from NCCS Pune. Baby hamster kidney cells (BHK-21) were obtained from SIGMA. All cells were grown in Dulbecco's modified essential medium (DMEM: Invitrogen), supplemented with glutamate (Invitrogen) and 10% fetal bovine serum (Invitrogen).

4.4.3 Viruses:

Chandipura virus (CHPV) was kindly gifted by Dr. Anirban Basu, National Brain Research Center (NBRC). VSVwt and VSV-eGFP were generated by reverse genetics in our lab.

4.4.4 Pseudovirus production:

For pseudoviruses construction, envelope genes of CHPV from Nagpur strain (strain No. 1653514) were cloned into the eukaryotic expression plasmid pCAG. To generate VSV pseudovirus, we have replaced VSV Glycoprotein from TFP and use pCAG VSV-G for packing (G* Δ G-VSV). 70% confluent 100 mm dish of BHK-21 cells were transfected with 5 μ g of pCAG-CHPV-G using Lipofectamine 2000 (Invitrogen, 11668019) following the manufacturer's direction. Twenty-four

hours later, transfected cells were infected with G*ΔG-VSV-TFP with 1 MOI. One hour after infection, cells were washed with PBS twice, and then 2% DMEM was added. Twenty-four hours post-infection, CHPV pseudovirus carrying culture supernatants were collected, filtered (0.45-μm pore size, Millipore, SLHP033RB), aliquoted, and stored in -80°C.

4.4.5 LC-MS analysis of Pseudovirus

70% confluent 100 mm dish of BHK-21 cells were transfected with 5 μg of pCAG-CHPV-G using Lipofectamine 2000 (Invitrogen, 11668019) following the manufacturer's direction. Twenty-four hours later, transfected cells infected with G*ΔG-VSV-TFP with 1 MOI. One hour after infection, cells were washed with PBS twice, and then 2% DMEM was added. Twenty-four hours post-infection, CHPV pseudovirus carrying culture supernatants were collected, filtered (0.45-μm pore size, Millipore, SLHP033RB). After filtration, it was centrifuged at one lakh *ref* for 2 hours. The obtained pellet was dissolved in 1 ml PBS and sent V-proteomics a Delhi (India) based company for LC-MS.

4.4.6 Transmission electron microscopy Imaging

The morphological assessment of the CHPV Pseudovirus was done using a transmission electron microscope (TEM, JEOL 2100F), operating at an accelerating voltage of 200 kV. For the TEM imaging, samples were placed onto 400-mesh copper grids (SIGMA ALDRICH CHEMICALS PVT LT) and incubated for 5 minutes. The grids were then rinsed twice with DI water, wicked dry, and stained with 2% uranyl acetate for 5 min.

4.4.7 RNA extraction and RT PCR:

Total RNA from BHK-21, Vero, and Vero E6 cells infected with CHPV Wt, CHPV U.V. inactivated, and CHPV pseudovirus was extracted with TRIzol reagent (Invitrogen) by following the standard protocol. cDNA was synthesized by using iScript (BIORAD) by following the manufacture's protocol. The primer sequence for CHPV-G was designed by a real-time PCR tool of IDT. The sense sequence used was 5'-GCCCCGGTGTGAAAGAAATG-3', and the antisense sequence was

5'-GTTTGGGCCTATCTCCATATCC-3'. The pCAG-CHPV-G plasmid was taken as a control.

4.4.8 Drug inhibition of pseudovirus entry:

The disruption of lipid raft from the plasma membrane was performed using methyl - β -cyclodextrin (MBCD). Vero and VeroE6 cells were seeded in 12 well plates one day before the experiment and at 90% confluency were incubated with one and five mM MBCD for 40 min then infect with 1 MOI of Pseudovirus after 1.5 hours, cells were washed three times and 2% culture media was added for 24 hours. For the inhibition of clathrin-mediated endocytosis, cells were incubated with 40 μ M of chlorpromazine and .45 M sucrose for 1 hour and then infected with the Pseudovirus for 1.5 hours. After extensive washes with PBS, cells were further incubated for 24 hours. To determine the pH dependency of the virus, cells were incubated with 25mM and 50 mM NH_4Cl for 1 hour then infected with Pseudovirus for 1.5 hours. The TFP fluorescence was determined by microscopy and spectrofluorimetry.

4.4.9 Statistical Analysis

Data are represented as means \pm SD, and the significance of the difference between groups was evaluated by using ANOVA Test. The data were analyzed and plotted using the Graphpad Prism (v6.01) (Graphpad Prism, San Diego, CA, USA).

4.5. Reference:

1. Ghosh, S., K. Dutta and A. Basu. (2013), Chandipura virus induces neuronal death through Fas-mediated extrinsic apoptotic pathway, *J Virol*, 87, 12398-12406 (10.1128/jvi.01864-13)
2. Sapkal, G. N., P. M. Sawant and D. T. Mourya. (2018), Chandipura Viral Encephalitis: A Brief Review, *Open Virol J*, 12, 44-51 (10.2174/1874357901812010044)
3. Bhatt, P. N. and F. M. Rodrigues. (1967), Chandipura: a new Arbovirus isolated in India from patients with febrile illness, *Indian J Med Res*, 55, 1295-1305
4. Rodrigues, J. J., P. B. Singh, D. S. Dave, R. Prasan, V. Ayachit, B. H. Shaikh and K. M. Pavri. (1983), Isolation of Chandipura virus from the blood in acute encephalopathy syndrome, *Indian J Med Res*, 77, 303-307
5. Sudeep, A. B., Y. K. Gurav and V. P. Bondre. (2016), Changing clinical scenario in Chandipura virus infection, *Indian J Med Res*, 143, 712-721 (10.4103/0971-5916.191929)
6. Rao, B. L., A. Basu, N. S. Wairagkar, M. M. Gore, V. A. Arankalle, J. P. Thakare, R. S. Jadi, K. A. Rao and A. C. Mishra. (2004), A large outbreak of acute encephalitis with high fatality rate in children in Andhra Pradesh, India, in 2003, associated with Chandipura virus, *Lancet*, 364, 869-874 (10.1016/s0140-6736(04)16982-1)
7. Chadha, M. S., V. A. Arankalle, R. S. Jadi, M. V. Joshi, J. P. Thakare, P. V. Mahadev and A. C. Mishra. (2005), An outbreak of Chandipura virus encephalitis in the eastern districts of Gujarat state, India, *Am J Trop Med Hyg*, 73, 566-570
8. Gurav, Y. K., B. V. Tandale, R. S. Jadi, R. S. Gunjekar, S. S. Tikute, A. V. Jamgaonkar, R. K. Khadse, S. V. Jalgaonkar, V. A. Arankalle and A. C. Mishra. (2010), Chandipura virus encephalitis outbreak among children in Nagpur division, Maharashtra, 2007, *Indian J Med Res*, 132, 395-399
9. Dwibedi, B., J. Sabat, R. K. Hazra, A. Kumar, D. S. Dinesh and S. K. Kar. (2015), Chandipura virus infection causing encephalitis in a tribal population of Odisha in eastern India, *Natl Med J India*, 28, 185-187
10. Marsh, M. and A. Helenius. (2006), Virus entry: open sesame, *Cell*, 124, 729-740 (10.1016/j.cell.2006.02.007)
11. Sieczkarski, S. B. and G. R. Whittaker. (2002), Dissecting virus entry via endocytosis, *J Gen Virol*, 83, 1535-1545 (10.1099/0022-1317-83-7-1535)
12. Yamauchi, Y. and A. Helenius. (2013), Virus entry at a glance, *J Cell Sci*, 126, 1289-1295 (10.1242/jcs.119685)
13. Kaksonen, M. and A. Roux. (2018), Mechanisms of clathrin-mediated endocytosis, *Nat Rev Mol Cell Biol*, 19, 313-326 (10.1038/nrm.2017.132)
14. Doherty, G. J. and H. T. McMahon. (2009), Mechanisms of endocytosis, *Annu Rev Biochem*, 78, 857-902 (10.1146/annurev.biochem.78.081307.110540)
15. Blanchard, E., S. Belouzard, L. Goueslain, T. Wakita, J. Dubuisson, C. Wychowski and Y. Rouillé. (2006), Hepatitis C virus entry depends on clathrin-mediated endocytosis, *J Virol*, 80, 6964-6972 (10.1128/jvi.00024-06)
16. Simon, M., C. Johansson and A. Mirazimi. (2009), Crimean-Congo hemorrhagic fever virus entry and replication is clathrin-, pH- and cholesterol-dependent, *J Gen Virol*, 90, 210-215 (10.1099/vir.0.006387-0)

17. Sieczkarski, S. B. and G. R. Whittaker. (2002), Influenza virus can enter and infect cells in the absence of clathrin-mediated endocytosis, *J Virol*, 76, 10455-10464 (10.1128/jvi.76.20.10455-10464.2002)
18. Tiwari, R., J. C. de la Torre, D. B. McGavern and D. Nayak. (2019), Beyond Tethering the Viral Particles: Immunomodulatory Functions of Tetherin (BST-2), *DNA Cell Biol*, 38, 1170-1177 (10.1089/dna.2019.4777)
19. Viswanathan, K., M. S. Smith, D. Malouli, M. Mansouri, J. A. Nelson and K. Fröh. (2011), BST2/Tetherin enhances entry of human cytomegalovirus, *PLoS Pathog*, 7, e1002332 (10.1371/journal.ppat.1002332)
20. Sun, X., V. K. Yau, B. J. Briggs and G. R. Whittaker. (2005), Role of clathrin-mediated endocytosis during vesicular stomatitis virus entry into host cells, *Virology*, 338, 53-60 (10.1016/j.virol.2005.05.006)
21. Johannsdottir, H. K., R. Mancini, J. Kartenbeck, L. Amato and A. Helenius. (2009), Host cell factors and functions involved in vesicular stomatitis virus entry, *J Virol*, 83, 440-453 (10.1128/jvi.01864-08)
22. Roberts, P. C., T. Kipperman and R. W. Compans. (1999), Vesicular stomatitis virus G protein acquires pH-independent fusion activity during transport in a polarized endometrial cell line, *J Virol*, 73, 10447-10457 (10.1128/jvi.73.12.10447-10457.1999)
23. Whitt, M. A. (2010), Generation of VSV pseudotypes using recombinant Δ G-VSV for studies on virus entry, identification of entry inhibitors, and immune responses to vaccines, *J Virol Methods*, 169, 365-374 (10.1016/j.jviromet.2010.08.006)
24. Quinn, K., M. A. Brindley, M. L. Weller, N. Kaludov, A. Kondratowicz, C. L. Hunt, P. L. Sinn, P. B. McCray, Jr., C. S. Stein, B. L. Davidson *et al.* (2009), Rho GTPases modulate entry of Ebola virus and vesicular stomatitis virus pseudotyped vectors, *J Virol*, 83, 10176-10186 (10.1128/jvi.00422-09)
25. Negrete, O. A., E. L. Levroney, H. C. Aguilar, A. Bertolotti-Ciarlet, R. Nazarian, S. Tajyar and B. Lee. (2005), EphrinB2 is the entry receptor for Nipah virus, an emergent deadly paramyxovirus, *Nature*, 436, 401-405 (10.1038/nature03838)
26. Lee, B. H., K. Yoshimatsu, K. Araki, M. Okumura, I. Nakamura and J. Arikawa. (2006), A pseudotype vesicular stomatitis virus containing Hantaan virus envelope glycoproteins G1 and G2 as an alternative to hantavirus vaccine in mice, *Vaccine*, 24, 2928-2934 (10.1016/j.vaccine.2005.12.040)
27. Moeschler, S., S. Locher, K. K. Conzelmann, B. Krämer and G. Zimmer. (2016), Quantification of Lyssavirus-Neutralizing Antibodies Using Vesicular Stomatitis Virus Pseudotype Particles, *Viruses*, 8 (10.3390/v8090254)
28. Nie, J., Q. Li, J. Wu, C. Zhao, H. Hao, H. Liu, L. Zhang, L. Nie, H. Qin, M. Wang *et al.* (2020), Quantification of SARS-CoV-2 neutralizing antibody by a pseudotyped virus-based assay, *Nature Protocols*, 15, 3699-3715 (10.1038/s41596-020-0394-5)
29. Elkin, S. R., A. M. Lakoduk and S. L. Schmid. (2016), Endocytic pathways and endosomal trafficking: a primer, *Wien Med Wochenschr*, 166, 196-204 (10.1007/s10354-016-0432-7)
30. Misinzo, G., P. L. Delputte and H. J. Nauwynck. (2008), Inhibition of endosome-lysosome system acidification enhances porcine circovirus 2 infection of porcine epithelial cells, *J Virol*, 82, 1128-1135 (10.1128/jvi.01229-07)

31. Codran, A., C. Royer, D. Jaeck, M. Bastien-Valle, T. F. Baumert, M. P. Kieny, C. A. Pereira and J. P. Martin. (2006), Entry of hepatitis C virus pseudotypes into primary human hepatocytes by clathrin-dependent endocytosis, *J Gen Virol*, 87, 2583-2593 (10.1099/vir.0.81710-0)
32. Shang, J., Y. Wan, C. Luo, G. Ye, Q. Geng, A. Auerbach and F. Li. (2020), Cell entry mechanisms of SARS-CoV-2, *Proc Natl Acad Sci U S A*, 117, 11727-11734 (10.1073/pnas.2003138117)
33. Simons, K. and D. Toomre. (2000), Lipid rafts and signal transduction, *Nat Rev Mol Cell Biol*, 1, 31-39 (10.1038/35036052)
34. Campbell, S. M., S. M. Crowe and J. Mak. (2001), Lipid rafts and HIV-1: from viral entry to assembly of progeny virions, *J Clin Virol*, 22, 217-227 (10.1016/s1386-6532(01)00193-7)
35. Chung, C. S., C. Y. Huang and W. Chang. (2005), Vaccinia virus penetration requires cholesterol and results in specific viral envelope proteins associated with lipid rafts, *J Virol*, 79, 1623-1634 (10.1128/jvi.79.3.1623-1634.2005)
36. Lee, C. J., H. R. Lin, C. L. Liao and Y. L. Lin. (2008), Cholesterol effectively blocks entry of flavivirus, *J Virol*, 82, 6470-6480 (10.1128/jvi.00117-08)
37. Homann, S., D. Smith, S. Little, D. Richman and J. Guatelli. (2011), Upregulation of BST-2/Tetherin by HIV infection in vivo, *J Virol*, 85, 10659-10668 (10.1128/jvi.05524-11)
38. Ooi, Y. S., M. Dubé and M. Kielian. (2015), BST2/tetherin inhibition of alphavirus exit, *Viruses*, 7, 2147-2167 (10.3390/v7042147)
39. Evans, D. T., R. Serra-Moreno, R. K. Singh and J. C. Guatelli. (2010), BST-2/tetherin: a new component of the innate immune response to enveloped viruses, *Trends Microbiol*, 18, 388-396 (10.1016/j.tim.2010.06.010)
40. Cullen, P. J. and F. Steinberg. (2018), To degrade or not to degrade: mechanisms and significance of endocytic recycling, *Nat Rev Mol Cell Biol*, 19, 679-696 (10.1038/s41580-018-0053-7)
41. Lanzrein, M., A. Schlegel and C. Kempf. (1994), Entry and uncoating of enveloped viruses, *Biochem J*, 302 (Pt 2), 313-320 (10.1042/bj3020313)
42. Fackler, O. T. and B. M. Peterlin. (2000), Endocytic entry of HIV-1, *Curr Biol*, 10, 1005-1008 (10.1016/s0960-9822(00)00654-0)
43. Cantín, C., J. Holguera, L. Ferreira, E. Villar and I. Muñoz-Barroso. (2007), Newcastle disease virus may enter cells by caveolae-mediated endocytosis, *J Gen Virol*, 88, 559-569 (10.1099/vir.0.82150-0)
44. Rojek, J. M., M. Perez and S. Kunz. (2008), Cellular entry of lymphocytic choriomeningitis virus, *J Virol*, 82, 1505-1517 (10.1128/jvi.01331-07)
45. Wang, H., P. Yang, K. Liu, F. Guo, Y. Zhang, G. Zhang and C. Jiang. (2008), SARS coronavirus entry into host cells through a novel clathrin- and caveolae-independent endocytic pathway, *Cell Res*, 18, 290-301 (10.1038/cr.2008.15)

CHAPTER 5

Chapter 5

Structural similarity-based prediction of host factors associated with SARS-CoV-2 infection and pathogenesis

5.1 Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causes a rainbow of diseases, ranging from flu-like symptoms to pneumonia, acute respiratory distress syndrome (ARDS), thrombosis, and fatal consequences.[1] SARS-CoV-2 is a Coronaviridae family member that encapsulates a positive-sense, single-stranded RNA genome. Six human coronaviruses (HCoVs) are identified earlier; these are HCoV-229E and HCoV-NL63, belonging to alpha coronavirus group. The rest members, HCoV-HKU1, HCoV-OC43, severe acute respiratory syndrome coronavirus (SARS-CoV), and Middle East respiratory syndrome coronavirus (MERS-CoV), belong to the betacoronavirus group.[2] However, over the past few years, highly pathogenic human coronaviruses have emerged. The appearance of SARS-CoV in 2002 exhibited 8,000 cases worldwide with mortality of ~10% and MERS-CoV in 2012, marked with 2,500 cases with a higher mortality rate of 36%.[3,4] Although the mortality rate of current SARS-CoV-2 is comparatively low, it is exceptionally contagious in nature. The SARS-CoV-2 mainly spread through aerosolized droplets and can be transmitted by direct contact and oral-fecal route.[5] Lack of efficacious vaccine or antivirals is a measure concern for SARS-CoV-2 infection. The use of some antiviral drugs like remdesivir, lopinavir plus, and symptom-based management are currently available options for COVID patients.[6] Therefore to develop new therapeutic interventions, a better understanding of the virus biology and host-pathogen interactions is necessary.

Our study involves implementing a computational method for predicting the interactions between SARS-CoV-2 and host proteins. The approach is based on

protein structural similarity. At first, we determined the structural similarities between SARS-CoV-2 and human proteins using an established method by analyzing protein crystal structures. Further, we identified known interactions for these SARS-CoV-2 similar human proteins. We assumed that these interacting proteins of SARS-CoV-2 similar proteins would also interact with the SARS-CoV-2 proteins. This approach is also reported earlier for the prediction of HIV, Chandipura virus (CHPV), chikungunya virus (CHIKV), and dengue virus (DENV)- human interactions.[7-10] We predicted and shortlisted an interaction map for SARS-CoV-2 and host proteins using cellular co-localization information. We then validated these interactions using previously published host factors datasets associated with coronaviruses and other RNA viruses.[11-13] The predicted interactions are highlighted based on their functional importance during SARS-CoV-2 infection and prioritized based on other related RNA viruses' available information. In a similar line, we also predicted the host complement system driving the coagulation process in SARS-CoV-2 infection.[14] While, in this study, these interacting proteins revealed that SARS-CoV-2 might use the clathrin-mediated endocytosis pathway for its entry. We also delineated the interplay of host proteins associated with the SARS-CoV-2 life cycle. We primarily focused on viral genome replication, translation, assembly and predicted the pathways and the host factor requirements. Gene enrichment analysis of these interacting proteins reveals that apoptosis, IFN- γ signaling, and proteasomal degradation of CD4 T cells are positively associated with SARS-CoV-2 pathogenesis.

5.2 Results

5.2.1 Identification of SARS-COV-2 host interactor proteins and gene enrichment analysis

The initial study identified 3,735 human proteins showing similarity to 16 SARS-CoV-2 proteins. These SARS-CoV-2 similar proteins were predicted to interact with 12,872 (endogenous interaction) human proteins through 57,359 unique

interactions. To filter out the noise and curate the data, we performed CC analysis. This resulted in finding the involvement of 6,877 proteins through 19,047 unique interactions. Interestingly about 35% of resultant proteins are already reported to have some role in the pathogenesis of other coronavirus and RNA viruses. Next, we performed gene enrichment analysis by using g: GOST tool of g: Profiler. For this, we uploaded the SARS-CoV-2 interactors protein lists (CC filtered) in g: GOST tool and manually selected "Reactome database" and "Human protein atlas database" as data source option and set the threshold value at 0.05. The g: Profiler gave us a list of associated pathways and their adjusted_p_value for the input gene list. Subsequent Reactome dataset analysis revealed that SARS-CoV-2 interactors proteins enriched with the terms of Axon guidance, Membrane trafficking, Vesicle-mediated transport, Programmed cell death, Apoptosis, etc. (Figure 5.1A). Similarly, the HPA database showed a higher expression of human cells for SARS-CoV-2 interacting human protein (Figure 5.1B).

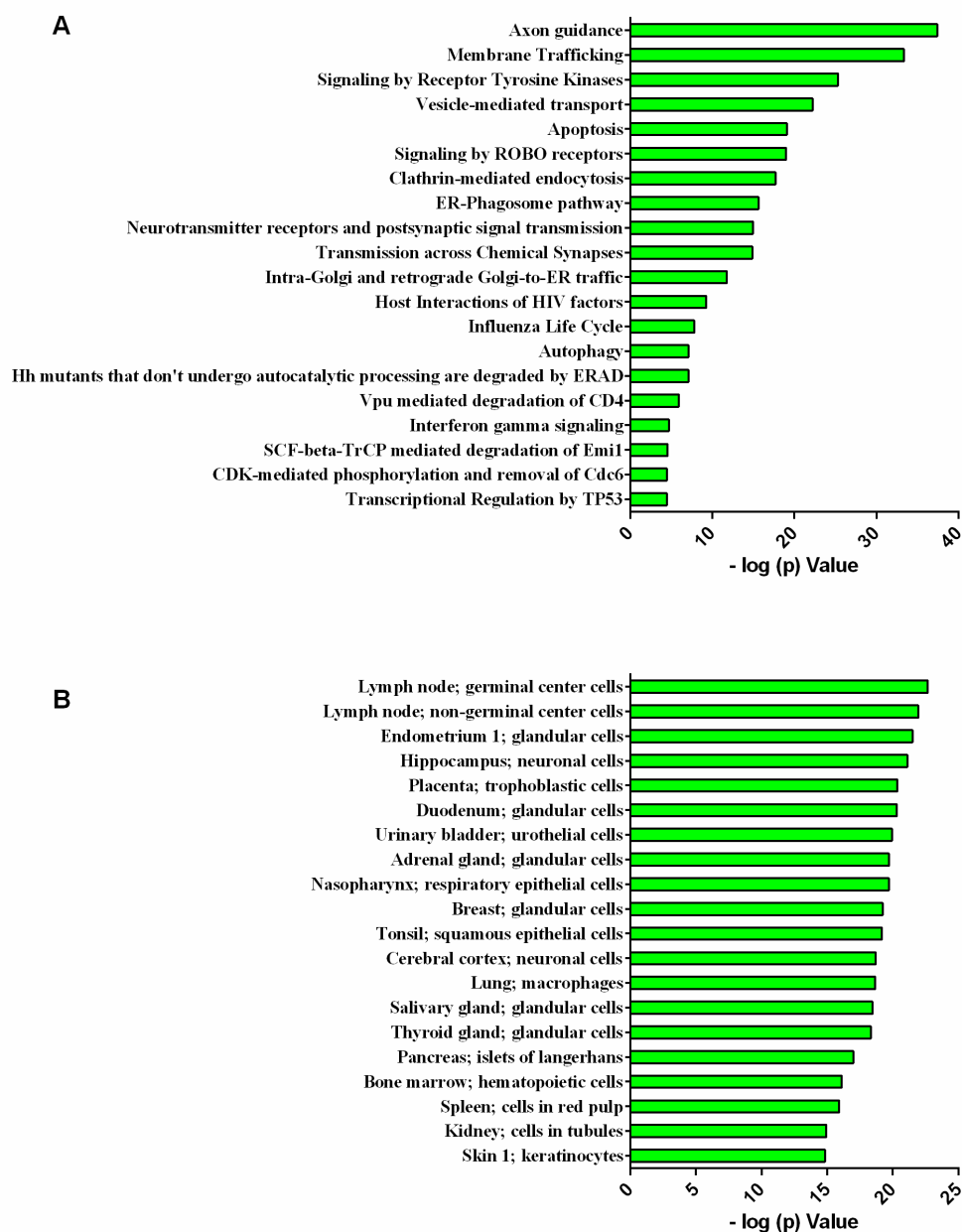


Figure 5.1. Gene enrichment analysis of SARS-CoV-2 interactor human proteins. (A). Enriched biological pathways obtained from Reactome database (B). Cellular expression data of interacting protein obtained from HPA. Bonferroni corrected p-values were transformed by $-\log_{10}$. (The bar graphs were created by using GraphPad-Prism software)

5.2.2 SARS-CoV-2 use clathrin-mediated endocytosis for its cellular entry

The productive viral infection begins after virions get access to a highly specific entry pathway. This allows the viral components to enter the host cytoplasm for the following processes, virus uncoating, gene expression, genome replication, virion assembly, and nascent virion release. Most of the viruses accomplish this task via endocytosis. Recently it was reported that host proteins ACE2 and TMPRSS2 are essential for SARS-CoV-2 entry. In the present study, we found that SARS-CoV-2 spike (S) interacts with several other host proteins for its entry through clathrin-coated pits. These include CLTC, DNM2, AP2A1, AP2A2, AP2B1, AP2M1, and DAB2. This result suggested that the SARS-CoV-2 internalization is dependent on clathrin-mediated endocytosis. Overall the SARS-CoV-2 entry process is very similar to that of SARS CoV, HCV, and VSV.[15-17] The host proteins such as AP2A1, AP2A2, AP2B1, and AP2M1 are the subunits of the heterotetrameric adaptor protein complex AP2, while DAB2 is present on the clathrin-coated vesicles. These host proteins continuously construct clathrin-coated pits and have a crucial role in the entry of SARS-CoV and MHV viruses.[15,18] The cytoskeletal machinery containing actin and microtubule is essential for the entry process of the mouse hepatitis virus (MHV).[19] This machinery is recruited through endocytosis after the coat formation, which triggers cell membrane deformation and virion internalization. The clathrin components, such as actin, and microtubule complex, when associated with dynamin proteins, result in the activation of binding sites for several other proteins. For example, cortactin (CTTN) and intersectin 1 (ITSN1) can activate the Arp2/3 complex of the actin cytoskeleton.[20] This leads to enhanced localized polymerization of actin and microtubules, favoring virus internalization. Here, SARS-CoV-2 glycoprotein is predicted to interact with CTTN and ITSN1 along with various types of actin (ACTA1, ATP2C1, ACTB, and ACTC1), dynamin (DNM1, DNM2, and DNM3), and microtubules (DYNC1H1) molecules. Together, these components facilitate clathrin-mediated endocytosis of virion and crucial for the intracellular trafficking of viral components during the virion assembly process.[21] Our data suggest that the interaction of SARS-CoV-2 spike (S) protein with Rab5b and Rab5c. It is established that upon clathrin-

mediated uptake, SARS-CoV-2 is trafficked to the early endosomes in a RAB5-dependent manner. [22] The S protein also interacts with Rab11b and Rab13, associated with the early endosome, and regulates the recycling of receptors and other ligands to the membrane. [23] After endocytosis, the lysosomal protease cathepsin L cleaves the S protein in the early endosome. Thus initiating fusion of the viral envelope with the endosome membrane.[24] The S protein interacts with the late endosomal proteins (RAB7A, RAB7B, RAB7L1) for lysosome maturation (VPS11, VPS33A) proteins. The HOPS complex regulates the late endosomes to the lysosomal maturation process (Figure 5.2).[25] Although the HOPS complex does not affect the infection process of other viruses such as VSV and IAV, viral entry is more dependent upon low pH conditions. While in the case of SARS-CoV-2, low pH is not sufficient to trigger virus entry. Hence, it appears that the HOPS complex could play a significant role in comparison to the pH factors during the SARS-CoV-2 entry process.

5.2.3 Intracellular replication and assembly of SARS-CoV-2

5.2.3.1 Host proteins involved in SARS-CoV-2 translation

Upon cellular entry and uncoating, the viral RNA serves as a transcript for the cap-dependent translation of ORF1a, ORF1b and produces polyproteins (pp1a and pp1ab). The autoproteolytic cleavage of both polyproteins generates 15–16 non-structural proteins (NSP's) associated with various functions. Simultaneously, the host mounts innate immune response by augmenting antiviral protein synthesis primarily orchestrated by type-I interferon (IFN-I) signaling mechanism. To counter this, viruses (including +ve sense RNA viruses) modulate host protein synthesis by restricting the host mRNAs translation and favoring viral protein synthesis.

Host mRNA translation is initiated by forming the heterotrimeric eIF2 complex comprising eIF2 α , eIF2 β , and eIF2 γ . The complex recruits Met-tRNA_i to start ribosomal translation in a GTP-dependent manner. Once the initiation process is completed, eIF2-GDP gets discharged from the ribosome, and GTP substitutes GDP to form an active eIF2-GTP complex to participate in another round of

translation initiation.[26] The eIF2 complex can be inactivated by phosphorylation at a single serine (Ser51) of its alpha subunit (eIF2a) by one of the four mammalian kinases in response to various stimuli. In this study, we found that three mammalian kinases interact with SARS-CoV-2 proteins, protein kinase R (PKR), PKR-like endoplasmic reticulum kinase (PERK), and heme-regulated inhibitor (HRI) can phosphorylate eIF2a and shut off the host cell's translation machinery (Table 5.1).

Table 5.1. List of human protein kinase and their interactor SARS-CoV-2 proteins.

S.No	Host Protein	Protein Name	Interacting Viral Protein
1	EIF2AK1/HRI	Eukaryotic Translation Initiation Factor 2 Alpha Kinase 1 / Heme-Regulated Inhibitor	NSP4, NSP7, NSP8, NSP14
2	EIF2AK2/PKR	Eukaryotic Translation Initiation Factor 2 Alpha Kinase 2/ Protein kinase R	NSP3, NSP4, NSP7, NSP8
3	EIF2AK3/PERK	Eukaryotic Translation Initiation Factor 2 Alpha Kinase 3 / PKR-like endoplasmic reticulum kinase	NSP4, NSP7, NSP8

5.2.3.2 Host proteins involved in a double-membrane vesicle (DMV) formation during SARS-CoV-2 infection

Positive-sense RNA viruses perform RNA replication in the cytoplasm. Genome replication is associated with virus-induced structures derived from cellular endomembranes.[27] In the case of coronaviruses, replication takes place in double-membrane vesicles (DMVs). The mechanism by which these DMVs are formed is not fully understood. In our study, we predict that the viral NSP4 interacts with the host proteins, namely, SEC61, VAPB, COPA, ARCN1, COPB1, GBF1, VAPA, VAPB, STX17, and TMED1, while NSP15 interacts with protein disulfide isomerase (PDI) associated with the ER. Earlier studies have reported that SEC 61 A and PDI are present on the inner surface of DMV.[28] Vesicle-associated membrane protein-associated protein A (VAP-A) and VAP-B are crucial for viral RNA replication and present in the DMV. Electron microscope tomography studies earlier confirmed that DMVs are part of a reticular network of modified endoplasmic reticulum (ER) membranes. The inner surface of DMVs also contains double-stranded RNA (dsRNA) derived from viral replication intermediate.[29] The outer layer of SARS-CoV-2 induced DMVs can be continuous with ER cisternae suggesting that the secretory pathway is also essential for virus replication. The depletion of COPB1 and GBF1, associated with the same path, affected SARS-CoV replication, intensifying this phenomenon.[30]

Previous studies suggested that coronaviruses take advantage of the cellular autophagy system for DMV biogenesis.[30] We observed that NSP4 interacts with several factors associated with autophagy, namely, ATG12, ATG16L1, ATP13A2, BECN1, C9ORF72, MAP1LC3A, MAP1LC3B, etc. suggesting that autophagy could play a crucial role in SARS-CoV-2 mediated DMV formation. Interestingly, NSP13 shows interaction with ERDA regulators EDEM1 and OS-9. Both EDEM1 and OS-9 are associated with the ERAD tuning pathway. It is observed that the prototypic coronavirus, MHV interferes with ERAD. This directs to intracellular accumulation of EDEM1 and OS-9, leading to relocalization and limitation of these chaperones, and enhance MHV-induced DMV formation (Figure 5.2).[31] Our

study suggests that SARS-CoV-2 mediated DMV formation is associated with the endoplasmic reticulum, autophagosome, and ERAD machinery.

5.2.3.3 Host proteins involved in SARS-CoV-2 replication

The +ve sense RNA genome of coronavirus serves as a template for the replicase that synthesizes a full-length negative-sense RNA, which serves as a template for the further synthesis of nascent genomic RNA. This process occurs in virus replication/transcription complex (RTC), a complex of viral and host proteins. The viral proteins (NSP3-NSP16) having various enzymatic activities are considered to be part of the RTC. These enzymes' actions include deubiquitination, protease, helicase, polymerase, Exo and endonuclease, and N7- and 2'O-methyltransferases. Interestingly, we observed that many host interactor of RTC complex are associated with RNA processing, [ELAVL like protein 1 (ELAVL1), ribosomal proteins (40S ribosomal protein S6 (RPS6), 40S ribosomal protein S4, X isoform (RPS4X), 40S ribosomal protein S3a (RPS3A), polyadenylate-binding protein 1 (PABPC1)]. Protein involved in translation initiation, especially multiple subunits of eukaryotic translation initiation factors 2, 3, 4, 5(eIF2, eIF3, eIF4, eIF5) and DDX3Y helicase. The viral proteins also showed interactions with the 60S ribosomal protein L13A (RPL13A) and its regulatory elements, such as IGF2BP1, GCN1L1, LARP, and NCK1 (Figure 5.2). These findings indicate that the host cell translation machinery is present near the RTC complex.

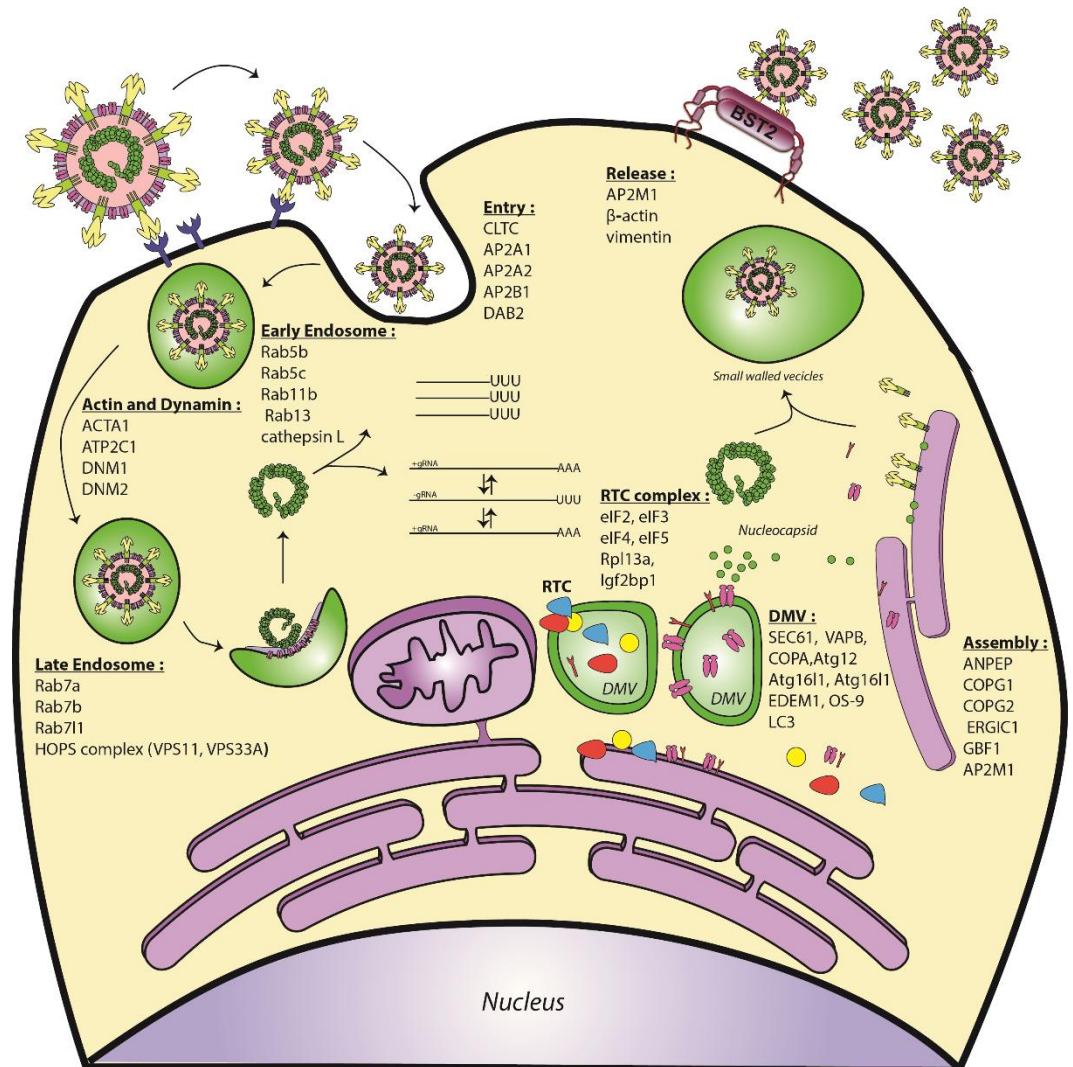


Figure 5.2. Schematic representation of the SARS-CoV-2 life cycle and its associated host proteins. This cartoon depicts the endocytotic pathway for viral entry, followed by uncoating and releasing the viral genome to the host cell cytoplasm. Subsequently, viral RNA is associated with the replication and transcription complex (RTC) membrane and double-membrane vesicles (DMVs) during the viral genome replication and transcriptions.

5.2.3.4 Host proteins involved in SARS-CoV-2 assembly and release

Coronavirus assembly occurs in the ER-Golgi intermediate compartment (ERGIC) and organized by the viral matrix (M) protein.[32,33] The M-S and M-nucleoprotein (N) interactions facilitate the recruitment of structural components to the assembly site. Contrastingly our analysis did not pick any hSARS-CoV-2 for N protein interactions. SARS-CoV-2 S and M proteins interact with 43 host proteins (i.e., ANPEP, COG1, COG2, ERGIC1, and GBF1) associated with the ERGIC inter-organelle compartment. This result shows SARS-CoV-2 also uses a similar mechanism as other coronaviruses. This mechanism is also conserved with other viruses such as HIV and VSV and can be inhibited by Brefeldin A (BFA).[34] BFA is an antiviral antibiotic drug containing a 13-member macrocyclic lactone ring.[35] It causes disassembly of the Golgi complex and inhibits lipid vesicle secretion.[36] Besides ERGIC associated proteins, SARS-CoV-2 also interacts with β -actin, vimentin (an intermediate filament protein), and AP2M1, which are essential for virus assembly and release.[37] The viral S protein interacts with a host antiviral protein, bone marrow stromal antigen 2 (BST-2), which is known to inhibit the release of the nascent virion from the cell. The BST-2 is a well-defined antiviral protein that tethers many enveloped viruses release like HIV1/2, Dengue, HCV, VSV, etc.[38]

5.2.4 IRF1/9/7 suggested to have a regulatory role in IFN- γ mediated signaling pathway during SARS-CoV-2 infection

The innate immune response is a conserved defense strategy of the host; this is critical for the first detection and localized restriction of pathogens. Various case reports indicate that SARS and SARS-CoV-2 downregulate IFN-I, but at the same time, type II interferon (IFN- γ) present in higher concentration in SARS and SARS-CoV-2 patients.[39,40] Hence, it is highly possible that IFN- γ to be a critical player in SARS-CoV-2 mediated lung injury and could be a potential biomarker for disease severity. Our results also suggest a significant role of IFN- γ pathways associated with SARS-CoV-2 infection. We show that 62 candidates

(associated with IFN- γ) interact with SARS-CoV-2 proteins. The degree centrality analysis suggests that IRF1, IRF9, IRF7, and HLA-DRB1 protein (HLA-DRB1) have crucial and central roles in cytokines IFN- γ signaling pathway. (Figure 5.3, Table 5.2).

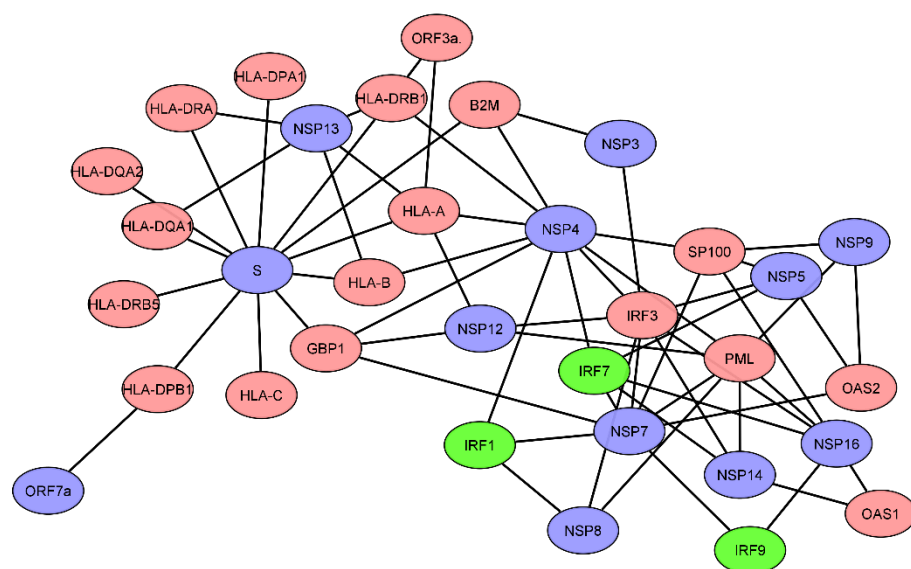


Figure 5.3. *A predicted interaction map of the top 20 IFN- γ associated human proteins and their interacting SARS-CoV-2 protein. Blue color represents the virus proteins, and brown color represents the human interactor proteins. Twelve SARS-CoV-2 proteins interact with human proteins related to IFN- γ signaling pathway. IRF1, IRF7, and IRF9 (illustrated in green color) are highly weighted proteins and may regulate IFN- γ signaling cascade. (Interaction map was created using Cytoscape.)*

Table 5.2. Degree centrality analysis of proteins associated with IFN- γ signaling pathway.

Serial Number	Host Protein	Ensemble Gene ID	Degree Centrality (Weight)
1	IRF1	ENSP00000245414	46.187
2	IRF9	ENSP00000380073	45.519
3	IRF7	ENSP00000380697	44.485
4	HLA-DRB1	ENSP00000353099	44.311
5	HLA-DRA	ENSP00000378786	44.263
6	HLA-DQA1	ENSP00000339398	44.217
7	HLA-DPA1	ENSP00000393566	44.048
8	IRF3	ENSP00000471896	44.038
9	HLA-DQA2	ENSP00000364076	44.028
10	PML	ENSP00000268058	43.965
11	HLA-DPB1	ENSP00000408146	43.961
12	HLA-DRB5	ENSP00000364114	43.878
13	B2M	ENSP00000452780	43.652
14	HLA-A	ENSP00000379873	43.531
15	OAS1	ENSP00000388001	43.358
16	SP100	ENSP00000343023	43.092
17	HLA-C	ENSP00000365402	42.578
18	HLA-B	ENSP00000399168	42.534
19	GBP1	ENSP00000359504	42.492
20	OAS2	ENSP00000342278	42.453

5.2.5 TP53 and CASP3 are the critical players of SARS-CoV-2 mediated apoptosis

Apoptosis is a sort of programmed cell death characterized by the highly controlled disassembling of cellular structures released in membrane-bound vesicles engulfed by neighboring cells or phagocytes. [41] Apoptosis occurs by two pathways, extrinsic and intrinsic, regulated by a preformed cascade of proteases called caspases. Individually, these pathways stimulate downstream caspases that initiate morphological and biochemical changes in the cell that eventually begin apoptosis.[42] Many viruses encode specific proteins that modulate the apoptosis pathway. In coronavirus-infected patients, one of the common abnormalities manifested is lymphopenia resulting from the depletion of the T cell population.[43] Autopsy findings of SARS-CoV-2 and SARS cases show T cell apoptosis in various infected tissues, such as lungs, liver, and thyroid.[44] In line with this, we observed apoptosis pathways to be highly enriched in gene enrichment analysis. Here 137 host proteins (associated with apoptosis) showed interaction with SARS-CoV-2 proteins. The degree centrality analysis of these genes suggests that cellular tumor antigen p53 (TP53), followed by caspase-3 (CASP3), catenin beta-1 (CTNNB1), and ubiquitin-60S ribosomal protein L40 (UBA52) are having crucial roles in the apoptosis pathway (Figure 5.4) (Table 5.3).

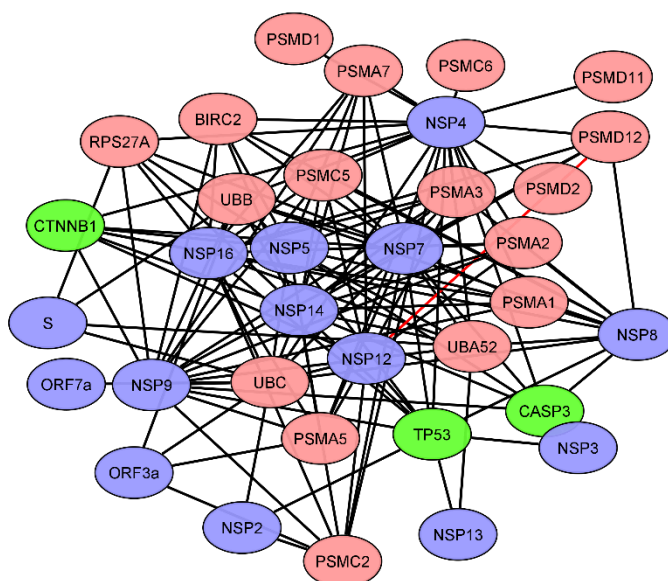


Figure 5.4. A predicted interaction map of the top 20 apoptosis-associated human proteins and their interacting SARS-CoV-2 protein. Blue color represents the virus proteins, and brown color represents the human interactor proteins. Fourteen SARS-CoV-2 proteins interact with human proteins connected with apoptosis. Degree centrality analysis reveals that TP53, CASP3, and CTNNB1 (illustrated in green color) are the principal players in the SARS-CoV-2 mediated apoptosis. (Interaction map was created using Cytoscape.)

Table 5.3. Degree centrality analysis of proteins associated with apoptosis pathway.

Serial Number	Host Protein	Ensemble Gene ID	Degree Centrality (Weight)
1	TP53	ENSP00000269305	73.562
2	CASP3	ENSP00000311032	57.312
3	CTNNB1	ENSP00000344456	53.542
4	UBA52	ENSP00000388107	52.511
5	UBB	ENSP00000304697	52.417
6	BIRC2	ENSP00000477613	52.298
7	UBC	ENSP00000441543	51.775
8	RPS27A	ENSP00000272317	51.714
9	PSMD1	ENSP00000309474	47.781
10	PSMC6	ENSP00000401802	46.648
11	PSMA5	ENSP00000271308	46.036
12	PSMC5	ENSP00000310572	46.01
13	PSMA3	ENSP00000216455	45.966
14	PSMD11	ENSP00000261712	45.945
15	PSMD12	ENSP00000348442	45.92
16	PSMD2	ENSP00000310129	45.898
17	PSMA1	ENSP00000414359	45.89
18	PSMA2	ENSP00000223321	45.848
19	PSMA7	ENSP00000359910	45.761
20	PSMC2	ENSP00000391211	45.214

5.2.6 SARS-CoV-2 protein-induced proteasomal degradation of CD4 T cell

In eukaryotes, most intracellular proteins are diminished by the ubiquitin (Ub)–proteasome pathway (UPP). [45] Ubiquitin is a 76-amino acid polypeptide present and conserved in every eukaryotic cell.[46] The covalent modification of proteins with ubiquitin chains forms a strong targeting signal driving recognition and destruction by the 26S proteasomes. The covalent interaction of ubiquitin to lysine residues requires the action of at least three ubiquitin enzymes: the ubiquitin-activating enzyme E1, one of several ubiquitin-conjugating enzymes, E2, and one of the multiple ubiquitin ligases, E3.[47]

Some viruses use ubiquitin (Ub)–proteasome pathway for cellular protein degradation, and the best example for this is HIV-I. Vpu of HIV-I targets CD4 protein for degradation. CD4 is a class I integral membrane glycoprotein expressed on the surface of a subset of T lymphocytes that recognize MHC-II associated peptides. This process is vital for construction and maintenance of the immune system. The HIV-Vpu interacts with the CD4 in the endoplasmic reticulum and triggers its proteolytic degradation. At first, Vpu interacts with the cytoplasmic domain of CD4 and connects it to β TrCP (a member of the F-box protein family first characterized as components of ubiquitin-ligase complexes). Following this the, N terminal of β TrCP interact with SKP1, a targeting factor for ubiquitin-mediated proteolysis.[48] Interestingly, our data highlights 43 candidates out of 52 known proteins associated with Vpu mediated degradation of CD4 T cells. Degree centrality analysis reveals that ubiquitin-60S ribosomal protein L40 (UBA52), followed by ubiquitin C (UBC), and 26S proteasome regulatory subunit 8 (PSMC5) having crucial roles in the proteasomal degradation of CD4 expressing T cells (Figure 5.5 and Table 5.4). We found that core elements of the 26S and 20S proteasome complex (Psm1, Psmc2, and Psm3) interact with SARS-CoV-2 proteins. These findings suggest that SARS-CoV-2 could also use a similar pathway to degrade CD4 expressing cells. A recent case study on COVID19 patients having lower CD4 and CD8 T cell counts corroborates this notion. We hypothesize that SARS-CoV-2 proteins directly bind to CD4 and direct its degradation using ubiquitin (Ub)–proteasome pathway.

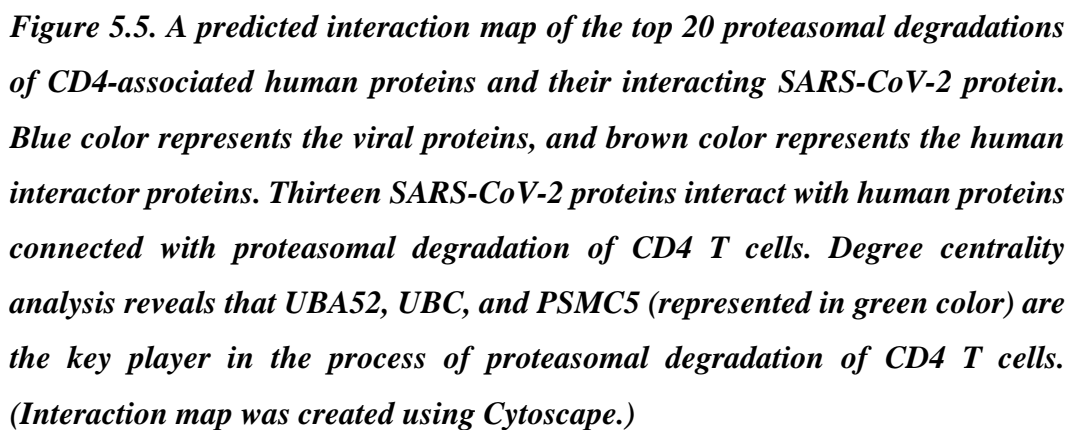


Table 5.4. Degree centrality analysis of proteins associated with proteasomal degradation of CD4 T cells

Serial Number	Host Protein	Ensemble Gene ID	Degree Centrality (Weight)
1	UBA52	ENSP00000388107	40.771
2	UBC	ENSP00000441543	40.754
3	PSMC5	ENSP00000310572	40.751
4	UBB	ENSP00000304697	40.748
5	PSMC2	ENSP00000391211	40.719
6	PSMB2	ENSP00000362334	40.717
7	PSMA3	ENSP00000216455	40.716
8	PSMD1	ENSP00000309474	40.703
9	PSMA5	ENSP00000271308	40.686
10	PSMA6	ENSP00000261479	40.684
11	PSMA4	ENSP00000044462	40.684
12	PSMB1	ENSP00000262193	40.68
13	PSMB4	ENSP00000290541	40.675
14	PSMC4	ENSP00000157812	40.673
15	PSMA1	ENSP00000414359	40.662
16	PSMD12	ENSP00000348442	40.653
17	PSMB3	ENSP00000483688	40.643
18	PSMC1	ENSP00000261303	40.641
19	PSMD7	ENSP00000219313	40.64
20	PSMC6	ENSP00000401802	40.63

5.3 Discussion

A successful viral infection depends on the virus's ability to manipulate host biological pathways to evade the host immune system during the pathogenesis. In the present study, we constructed a network of the protein-protein interactome of the host and SARS-CoV-2 proteins by applying a structure-based computational approach. This computational method's principle is based on the assumption that proteins with similar structures would share related interaction partners.

Based on this analysis, we craft possible chronological events associated with SARS-CoV-2 pathogenesis and brought light into the molecular partners and their potential role in host-pathogen interactions. The GO analysis of interacting proteins reveals that SARS-CoV-2 may use clathrin-mediated endocytosis for its entry. Among the viral proteins, the spike protein was found to interact with a series of host proteins. These include actin and microtubule cytoskeletal proteins (CTTN and ITSN1), early endosome (RAB5, RAB11B, and RAB13), late endosome (RAB7A, RAB7B, and RAB7L1), and HOPS complex (VPS11, VPS33A) proteins. These factors are known for their association with clathrin-mediated virus entry.[49] We also predict the role of several host factors essential for SARS-CoV2 replication and assembly. Hence, therapeutic modalities could be targeted to restrict these factors (if affordable) to mitigate viral infection. Most of the positive-sense RNA viruses replicate at virus-induced endomembranes. [50] While in the case of coronaviruses, genome replication occurs at the DMV sites.[51,52] Various cellular organelles reorganized and form DMV after virus infection. Here we report that the endoplasmic reticulum, the autophagosome, and ERAD machinery are associated with DMV formation. After the DMV formation, SARS-CoV-2 non-structural proteins (NSP3-NSP16) interact with the host proteins and make an RTC complex. Among these host proteins, ELAVL1, RPS6, RPS4X, PABPC1, eIF2/3/4/5, and 60S ribosomal protein contribute to RTC complex formation, which is essential for virus replication. Similarly, these viral protein interacts with candidates associated with ERGIC, where virus assembly takes place.[53]

Further, we elaborate significantly enriched pathways associated with SARS-CoV-2 pathogenesis. These include SARS-CoV-2 mediated apoptosis, IFN- γ

signaling pathway, and proteasomal degradation of CD4 pathway playing an essential role in viral pathogenesis. Various clinical reports show elevated blood levels of IFN- γ , associated with a higher degree of SARS-CoV-2 pathogenesis.[54] The Degree centrality analysis of IFN- γ signaling suggests that IRF1/9 and IRF7 are the key players of this pathway. Previously it was observed that coronavirus induced apoptosis in host cells. However, this is not required for virus replication and infection.[55] It might be a possible mechanism by which viruses degrade immune cells and cause leukocytopenia.[43] In this case, the degree centrality analysis reveals TP53 and CASP3 are the most weighted protein in SARS-CoV-2 mediated apoptosis process. Apart from core apoptosis, we uncovered one more interesting pathway, "Vpu mediated degradation of CD4 cells." Similar to HIV-I Vpu causing proteasomal degradation of CD4 T cells, [48,56] we learn that core elements of the 26S and 20S proteasome complex (Psmc1, Psmc2, and Psmb3) interact with SARS-Cov-2 proteins. Considering these pieces of evidence, we postulate that SARS-CoV-2 induces ubiquitination-mediated proteasomal degradation of CD4 T cells and other immune cells in the affected host. Lower CD4 T cell counts in severe SARS-CoV-2 patients support this hypothesis. Overall, we bring crucial mechanistic insight to SARS-CoV-2 pathogenesis in this study. These observations could be the foundation for future clinical and in vivo studies.

5.4 Material and method

5.4.1 Data sources

The PDB files for the crystal structures of SARS-COV-2 were obtained from the two primary sources: RCSB PDB [main protease (PDB ID:5R7Y), spike glycoproteins (PDB ID: 6VSB, 6VXX), HR2 domain (PDB ID:6LVN), NSP15 (PDB ID:6VWW), NSP3 (PDB ID:6W02) and NSP9 (PDB ID: 6W4B)] and from Zhang Lab (modeled these structure by using I-TASSER).[57] The PDB files were uploaded to the DaliLite v.5 webserver for structural similarities identification.[58] The Dali server provides the PDB codes as an output file mapped to their corresponding Uniprot ID and gene name by DAVID Gene ID

conversion or Uniprot ID mapping.[59,60] Reactome and g: Profiler database was used to perform the gene enrichment analysis, whereas the network interactions were analyzed in Cytoscape software. [61,62]

5.4.2 Determination of structural similarity between SARS-CoV-2 and human proteins

The Dali server uses the alpha carbon distance matrix, allowing them for differences in domain order, and produces a structural similarity score (z) to identify the structurally similar human protein.[58] For the current study, we have used the cutoff z score of 2.0. The human protein showing the result above the cutoff score was selected and referred to as " SARS-CoV-2 similar" proteins.

5.4.3 Interaction prediction

The SARS-CoV-2 proteins' interacting partners were determined by identifying the interaction partner for " SARS-CoV-2 similar" proteins. We mined BIOGRID, HPRD, and MINT database to find the endogenous target proteins.[63-65] These databases are sources of literature-curated interaction partners of human proteins. The assumption is that the human protein that interacts with "SARS-CoV-2 similar" proteins will also interact with the SARS-CoV-2 viral proteins.

5.4.4 Cellular compartmentalization (CC) and GO enrichment analysis

The identified host protein interactors of the SARS-CoV-2 proteome were shortlisted based on their localization and associated functions. Therefore, the interactor host proteins must share at least one cellular compartment with the viral protein for direct communication. The information regarding cellular compartmentalization protein-protein interactions is available in the UniProt database, curated by published literature. The host and viral protein sharing at least one cellular compartment were identified, and later, these proteins were submitted to the g: profiler for gene enrichment analysis.[61] The Reactome functional annotation charts and human protein atlas organized as a tree structure. As the distance from root increased, the terms were supposed to be more specific. The Bonferroni procedure was used to obtain the corrected p-values, and these values

were then plotted in $-\log_{10}$ terms for graphical representation of data. The identified proteins were imported to Cytoscape software for further analysis.[62] STRING and CytoNCA tool in Cytoscape was used for network generation and degree centrality calculation.[66]

5.4.5 Validation of Predictions

The Dali server provides the PDB ID that was converted to the Uniport ID. We observed many duplicates in the datasets during the analysis, as multiple PDB structures are available for the same protein, hence leading to repetition in the interaction predictions. Therefore, we removed the duplicate copies by analyzing the prediction based on a single pair of individual Uniprot accessions and SARS CoV-2 protein names. For better understanding, a SARS-CoV-2 protein is labeled with the protein name, while Entrez Gene ID was assigned for the host protein. For further validation of the output data, we also compared the predicted datasets with published experimental results carried out for SARS-CoV-2 and other related RNA viruses.

5.5 REFERENCE

1. Tay, M. Z., C. M. Poh, L. Rénia, P. A. MacAry and L. F. P. Ng. (2020), The trinity of COVID-19: immunity, inflammation and intervention, *Nat Rev Immunol*, 20, 363-374 (10.1038/s41577-020-0311-8)
2. van der Hoek, L. (2007), Human coronaviruses: what do they cause?, *Antivir Ther*, 12, 651-658
3. Ksiazek, T. G., D. Erdman, C. S. Goldsmith, S. R. Zaki, T. Peret, S. Emery, S. Tong, C. Urbani, J. A. Comer, W. Lim *et al.* (2003), A novel coronavirus associated with severe acute respiratory syndrome, *N Engl J Med*, 348, 1953-1966 (10.1056/NEJMoa030781)
4. Graham, R. L., E. F. Donaldson and R. S. Baric. (2013), A decade after SARS: strategies for controlling emerging coronaviruses, *Nat Rev Microbiol*, 11, 836-848 (10.1038/nrmicro3143)
5. Liu, J., X. Liao, S. Qian, J. Yuan, F. Wang, Y. Liu, Z. Wang, F. S. Wang, L. Liu and Z. Zhang. (2020), Community Transmission of Severe Acute Respiratory Syndrome Coronavirus 2, Shenzhen, China, 2020, *Emerg Infect Dis*, 26, 1320-1323 (10.3201/eid2606.200239)
6. Wang, Y., D. Zhang, G. Du, R. Du, J. Zhao, Y. Jin, S. Fu, L. Gao, Z. Cheng, Q. Lu *et al.* (2020), Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial, *Lancet*, 395, 1569-1578 (10.1016/s0140-6736(20)31022-9)
7. Rajasekharan, S., J. Rana, S. Gulati, S. K. Sharma, V. Gupta and S. Gupta. (2013), Predicting the host protein interactors of Chandipura virus using a structural similarity-based approach, *Pathog Dis*, 69, 29-35 (10.1111/2049-632x.12064)
8. Doolittle, J. M. and S. M. Gomez. (2010), Structural similarity-based predictions of protein interactions between HIV-1 and Homo sapiens, *Viol J*, 7, 82 (10.1186/1743-422x-7-82)
9. Doolittle, J. M. and S. M. Gomez. (2011), Mapping protein interactions between Dengue virus and its human and insect hosts, *PLoS Negl Trop Dis*, 5, e954 (10.1371/journal.pntd.0000954)
10. Rana, J., R. Sreejith, S. Gulati, I. Bharti, S. Jain and S. Gupta. (2013), Deciphering the host-pathogen protein interface in chikungunya virus-mediated sickness, *Arch Virol*, 158, 1159-1172 (10.1007/s00705-013-1602-1)
11. Gordon, D. E., G. M. Jang, M. Bouhaddou, J. Xu, K. Obernier, K. M. White, M. J. O'Meara, V. V. Rezelj, J. Z. Guo, D. L. Swaney *et al.* (2020), A SARS-CoV-2 protein interaction map reveals targets for drug repurposing, *Nature* (10.1038/s41586-020-2286-9)
12. Pfefferle, S., J. Schopf, M. Kogl, C. C. Friedel, M. A. Muller, J. Carbajo-Lozoya, T. Stellberger, E. von Dall'Armi, P. Herzog, S. Kallies *et al.* (2011), The SARS-coronavirus-host interactome: identification of cyclophilins as target for pan-coronavirus inhibitors, *PLoS Pathog*, 7, e1002331 (10.1371/journal.ppat.1002331)
13. Xiong, Y., Y. Liu, L. Cao, D. Wang, M. Guo, A. Jiang, D. Guo, W. Hu, J. Yang, Z. Tang *et al.* (2020), Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients, *Emerg Microbes Infect*, 9, 761-770 (10.1080/22221751.2020.1747363)

14. Tiwari, R., A. R. Mishra, F. Mikaeloff, S. Gupta, A. Mirazimi, S. N. Byraredddy, U. Neogi and D. Nayak. (2020), In silico and in vitro studies reveal complement system drives coagulation cascade in SARS-CoV-2 pathogenesis, *Comput Struct Biotechnol J*, 18, 3734-3744 (10.1016/j.csbj.2020.11.005)
15. Inoue, Y., N. Tanaka, Y. Tanaka, S. Inoue, K. Morita, M. Zhuang, T. Hattori and K. Sugamura. (2007), Clathrin-dependent entry of severe acute respiratory syndrome coronavirus into target cells expressing ACE2 with the cytoplasmic tail deleted, *J Virol*, 81, 8722-8729 (10.1128/jvi.00253-07)
16. Sun, X., V. K. Yau, B. J. Briggs and G. R. Whittaker. (2005), Role of clathrin-mediated endocytosis during vesicular stomatitis virus entry into host cells, *Virology*, 338, 53-60 (10.1016/j.virol.2005.05.006)
17. Blanchard, E., S. Belouzard, L. Goueslain, T. Wakita, J. Dubuisson, C. Wychowski and Y. Rouille. (2006), Hepatitis C virus entry depends on clathrin-mediated endocytosis, *J Virol*, 80, 6964-6972 (10.1128/jvi.00024-06)
18. Eifart, P., K. Ludwig, C. Bottcher, C. A. de Haan, P. J. Rottier, T. Korte and A. Herrmann. (2007), Role of endocytosis and low pH in murine hepatitis virus strain A59 cell entry, *J Virol*, 81, 10758-10768 (10.1128/jvi.00725-07)
19. Burkard, C., M. H. Verheije, O. Wicht, S. I. van Kasteren, F. J. van Kuppeveld, B. L. Haagmans, L. Pelkmans, P. J. Rottier, B. J. Bosch and C. A. de Haan. (2014), Coronavirus cell entry occurs through the endo-/lysosomal pathway in a proteolysis-dependent manner, *PLoS Pathog*, 10, e1004502 (10.1371/journal.ppat.1004502)
20. Uruno, T., J. Liu, P. Zhang, Y. Fan, C. Egile, R. Li, S. C. Mueller and X. Zhan. (2001), Activation of Arp2/3 complex-mediated actin polymerization by cortactin, *Nat Cell Biol*, 3, 259-266 (10.1038/35060051)
21. Lamason, R. L. and M. D. Welch. (2017), Actin-based motility and cell-to-cell spread of bacterial pathogens, *Curr Opin Microbiol*, 35, 48-57 (10.1016/j.mib.2016.11.007)
22. Mire, C. E., J. M. White and M. A. Whitt. (2010), A spatio-temporal analysis of matrix protein and nucleocapsid trafficking during vesicular stomatitis virus uncoating, *PLoS Pathog*, 6, e1000994 (10.1371/journal.ppat.1000994)
23. Urbe, S., L. A. Huber, M. Zerial, S. A. Tooze and R. G. Parton. (1993), Rab11, a small GTPase associated with both constitutive and regulated secretory pathways in PC12 cells, *FEBS Lett*, 334, 175-182 (10.1016/0014-5793(93)81707-7)
24. Huang, I. C., B. J. Bosch, F. Li, W. Li, K. H. Lee, S. Ghiran, N. Vasilieva, T. S. Dermody, S. C. Harrison, P. R. Dormitzer *et al.* (2006), SARS coronavirus, but not human coronavirus NL63, utilizes cathepsin L to infect ACE2-expressing cells, *J Biol Chem*, 281, 3198-3203 (10.1074/jbc.M508381200)
25. Balderhaar, H. J. and C. Ungermann. (2013), CORVET and HOPS tethering complexes - coordinators of endosome and lysosome fusion, *J Cell Sci*, 126, 1307-1316 (10.1242/jcs.107805)
26. Jackson, R. J., C. U. Hellen and T. V. Pestova. (2010), The mechanism of eukaryotic translation initiation and principles of its regulation, *Nat Rev Mol Cell Biol*, 11, 113-127 (10.1038/nrm2838)
27. Romero-Brey, I. and R. Bartenschlager. (2016), Endoplasmic Reticulum: The Favorite Intracellular Niche for Viral Replication and Assembly, *Viruses*, 8 (10.3390/v8060160)

28. Hagemeyer, M. C., I. Monastyrska, J. Griffith, P. van der Sluijs, J. Voortman, P. M. van Bergen en Henegouwen, A. M. Vonk, P. J. Rottier, F. Reggiori and C. A. de Haan. (2014), Membrane rearrangements mediated by coronavirus nonstructural proteins 3 and 4, *Virology*, 458-459, 125-135 (10.1016/j.virol.2014.04.027)
29. Knoop, K., M. Kikkert, S. H. Worm, J. C. Zevenhoven-Dobbe, Y. van der Meer, A. J. Koster, A. M. Mommaas and E. J. Snijder. (2008), SARS-coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum, *PLoS Biol*, 6, e226 (10.1371/journal.pbio.0060226)
30. de Wilde, A. H., K. F. Wannee, F. E. Scholte, J. J. Goeman, P. Ten Dijke, E. J. Snijder, M. Kikkert and M. J. van Hemert. (2015), A Kinome-Wide Small Interfering RNA Screen Identifies Proviral and Antiviral Host Factors in Severe Acute Respiratory Syndrome Coronavirus Replication, Including Double-Stranded RNA-Activated Protein Kinase and Early Secretory Pathway Proteins, *J Virol*, 89, 8318-8333 (10.1128/jvi.01029-15)
31. Cali, T., C. Galli, S. Olivari and M. Molinari. (2008), Segregation and rapid turnover of EDEM1 by an autophagy-like mechanism modulates standard ERAD and folding activities, *Biochem Biophys Res Commun*, 371, 405-410 (10.1016/j.bbrc.2008.04.098)
32. Klumperman, J., J. K. Locker, A. Meijer, M. C. Horzinek, H. J. Geuze and P. J. Rottier. (1994), Coronavirus M proteins accumulate in the Golgi complex beyond the site of virion budding, *J Virol*, 68, 6523-6534
33. Masters, P. S. (2006), The molecular biology of coronaviruses, *Adv Virus Res*, 66, 193-292 (10.1016/s0065-3527(06)66005-3)
34. Pal, R., S. Mumbauer, G. M. Hoke, A. Takatsuki and M. G. Sarngadharan. (1991), Brefeldin A inhibits the processing and secretion of envelope glycoproteins of human immunodeficiency virus type 1, *AIDS Res Hum Retroviruses*, 7, 707-712 (10.1089/aid.1991.7.707)
35. Misumi, Y., Y. Misumi, K. Miki, A. Takatsuki, G. Tamura and Y. Ikehara. (1986), Novel blockade by brefeldin A of intracellular transport of secretory proteins in cultured rat hepatocytes, *J Biol Chem*, 261, 11398-11403
36. Fujiwara, T., K. Oda, S. Yokota, A. Takatsuki and Y. Ikehara. (1988), Brefeldin A causes disassembly of the Golgi complex and accumulation of secretory proteins in the endoplasmic reticulum, *J Biol Chem*, 263, 18545-18552
37. Neveu, G., R. Barouch-Bentov, A. Ziv-Av, D. Gerber, Y. Jacob and S. Einav. (2012), Identification and targeting of an interaction between a tyrosine motif within hepatitis C virus core protein and AP2M1 essential for viral assembly, *PLoS Pathog*, 8, e1002845 (10.1371/journal.ppat.1002845)
38. Tiwari, R., J. C. de la Torre, D. B. McGavern and D. Nayak. (2019), Beyond Tethering the Viral Particles: Immunomodulatory Functions of Tetherin (BST-2), *DNA Cell Biol*, 38, 1170-1177 (10.1089/dna.2019.4777)
39. Wong, C. K., C. W. Lam, A. K. Wu, W. K. Ip, N. L. Lee, I. H. Chan, L. C. Lit, D. S. Hui, M. H. Chan, S. S. Chung *et al.* (2004), Plasma inflammatory cytokines and chemokines in severe acute respiratory syndrome, *Clin Exp Immunol*, 136, 95-103 (10.1111/j.1365-2249.2004.02415.x)
40. Yao, Z., Z. Zheng, K. Wu and Z. Junhua. (2020), Immune environment modulation in pneumonia patients caused by coronavirus: SARS-CoV, MERS-CoV and SARS-CoV-2, *Aging (Albany NY)*, 12, 7639-7651 (10.18632/aging.103101)

41. Elmore, S. (2007), Apoptosis: a review of programmed cell death, *Toxicol Pathol*, 35, 495-516 (10.1080/01926230701320337)
42. Galluzzi, L., I. Vitale, S. A. Aaronson, J. M. Abrams, D. Adam, P. Agostinis, E. S. Alnemri, L. Altucci, I. Amelio, D. W. Andrews *et al.* (2018), Molecular mechanisms of cell death: recommendations of the Nomenclature Committee on Cell Death 2018, *Cell Death Differ*, 25, 486-541 (10.1038/s41418-017-0012-4)
43. Tavakolpour, S., T. Rakhshandehroo, E. X. Wei and M. Rashidian. (2020), Lymphopenia during the COVID-19 infection: What it shows and what can be learned, *Immunol Lett*, 225, 31-32 (10.1016/j.imlet.2020.06.013)
44. Zhao, Q., M. Meng, R. Kumar, Y. Wu, J. Huang, Y. Deng, Z. Weng and L. Yang. (2020), Lymphopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: A systemic review and meta-analysis, *Int J Infect Dis*, 96, 131-135 (10.1016/j.ijid.2020.04.086)
45. Huang, Q. and M. E. Figueiredo-Pereira. (2010), Ubiquitin/proteasome pathway impairment in neurodegeneration: therapeutic implications, *Apoptosis*, 15, 1292-1311 (10.1007/s10495-010-0466-z)
46. Li, W. and Y. Ye. (2008), Polyubiquitin chains: functions, structures, and mechanisms, *Cell Mol Life Sci*, 65, 2397-2406 (10.1007/s00018-008-8090-6)
47. Wilkinson, K. D. (1995), Roles of ubiquitinylation in proteolysis and cellular regulation, *Annu Rev Nutr*, 15, 161-189 (10.1146/annurev.nu.15.070195.001113)
48. Schubert, U., L. C. Antón, I. Bacík, J. H. Cox, S. Bour, J. R. Bennink, M. Orlowski, K. Strebel and J. W. Yewdell. (1998), CD4 glycoprotein degradation induced by human immunodeficiency virus type 1 Vpu protein requires the function of proteasomes and the ubiquitin-conjugating pathway, *J Virol*, 72, 2280-2288
49. Inoue, Y., N. Tanaka, Y. Tanaka, S. Inoue, K. Morita, M. Zhuang, T. Hattori and K. Sugamura. (2007), Clathrin-Dependent Entry of Severe Acute Respiratory Syndrome Coronavirus into Target Cells Expressing ACE2 with the Cytoplasmic Tail Deleted, *Journal of Virology*, 81, 8722-8729 (10.1128/jvi.00253-07)
50. Wolff, G., C. E. Melia, E. J. Snijder and M. Bárcena. (2020), Double-Membrane Vesicles as Platforms for Viral Replication, *Trends Microbiol* (10.1016/j.tim.2020.05.009)
51. V'kovski, P., A. Kratzel, S. Steiner, H. Stalder and V. Thiel. (2020), Coronavirus biology and replication: implications for SARS-CoV-2, *Nature Reviews Microbiology* (10.1038/s41579-020-00468-6)
52. Wolff, G., C. E. Melia, E. J. Snijder and M. Bárcena. (2020), Double-Membrane Vesicles as Platforms for Viral Replication, *Trends Microbiol*, 28, 1022-1033 (10.1016/j.tim.2020.05.009)
53. V'kovski, P., M. Gerber, J. Kelly, S. Pfaender, N. Ebert, S. Braga Lagache, C. Simillion, J. Portmann, H. Stalder, V. Gaschen *et al.* (2019), Determination of host proteins composing the microenvironment of coronavirus replicase complexes by proximity-labeling, *Elife*, 8 (10.7554/eLife.42037)
54. Gadotti, A. C., M. de Castro Deus, J. P. Telles, R. Wind, M. Goes, R. Garcia Charello Ossoski, A. M. de Padua, L. de Noronha, A. Moreno-Amaral, C. P. Baena *et al.* (2020), IFN- γ is an independent risk factor associated with mortality in patients with moderate and severe COVID-19 infection, *Virus Res*, 289, 198171 (10.1016/j.virusres.2020.198171)
55. Bordi, L., C. Castilletti, L. Falasca, F. Ciccosanti, S. Calcaterra, G. Rozera, A. Di Caro, S. Zaniratti, A. Rinaldi, G. Ippolito *et al.* (2006), Bcl-2 inhibits the caspase-

- dependent apoptosis induced by SARS-CoV without affecting virus replication kinetics, *Arch Virol*, 151, 369-377 (10.1007/s00705-005-0632-8)
56. Binette, J., M. Dubé, J. Mercier, D. Halawani, M. Latterich and E. A. Cohen. (2007), Requirements for the selective degradation of CD4 receptor molecules by the human immunodeficiency virus type 1 Vpu protein in the endoplasmic reticulum, *Retrovirology*, 4, 75 (10.1186/1742-4690-4-75)
 57. Zhang, C., W. Zheng, X. Huang, E. W. Bell, X. Zhou and Y. Zhang. (2020), Protein Structure and Sequence Reanalysis of 2019-nCoV Genome Refutes Snakes as Its Intermediate Host and the Unique Similarity between Its Spike Protein Insertions and HIV-1, *J Proteome Res*, 19, 1351-1360 (10.1021/acs.jproteome.0c00129)
 58. Holm, L. (2019), Benchmarking fold detection by DaliLite v.5, *Bioinformatics*, 35, 5326-5327 (10.1093/bioinformatics/btz536)
 59. (2019), UniProt: a worldwide hub of protein knowledge, *Nucleic Acids Res*, 47, D506-d515 (10.1093/nar/gky1049)
 60. Dennis, G., Jr., B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane and R. A. Lempicki. (2003), DAVID: Database for Annotation, Visualization, and Integrated Discovery, *Genome Biol*, 4, P3
 61. Fabregat, A., K. Sidiropoulos, G. Viteri, O. Forner, P. Marin-Garcia, V. Arnau, P. D'Eustachio, L. Stein and H. Hermjakob. (2017), Reactome pathway analysis: a high-performance in-memory approach, *BMC Bioinformatics*, 18, 142 (10.1186/s12859-017-1559-2)
 62. Otasek, D., J. H. Morris, J. Boucas, A. R. Pico and B. Demchak. (2019), Cytoscape Automation: empowering workflow-based network analysis, *Genome Biol*, 20, 185 (10.1186/s13059-019-1758-4)
 63. Oughtred, R., C. Stark, B. J. Breitkreutz, J. Rust, L. Boucher, C. Chang, N. Kolas, L. O'Donnell, G. Leung, R. McAdam *et al.* (2019), The BioGRID interaction database: 2019 update, *Nucleic Acids Res*, 47, D529-d541 (10.1093/nar/gky1079)
 64. Goel, R., H. C. Harsha, A. Pandey and T. S. Prasad. (2012), Human Protein Reference Database and Human Proteinpedia as resources for phosphoproteome analysis, *Mol Biosyst*, 8, 453-463 (10.1039/c1mb05340j)
 65. Chatr-aryamontri, A., A. Ceol, L. M. Palazzi, G. Nardelli, M. V. Schneider, L. Castagnoli and G. Cesareni. (2007), MINT: the Molecular INTERaction database, *Nucleic Acids Res*, 35, D572-574 (10.1093/nar/gkl950)
 66. von Mering, C., M. Huynen, D. Jaeggi, S. Schmidt, P. Bork and B. Snel. (2003), STRING: a database of predicted functional associations between proteins, *Nucleic Acids Res*, 31, 258-261 (10.1093/nar/gkg034)

CHAPTER 6

Chapter 6

In silico and In vitro Studies Reveal Complement System Drives Coagulation Cascade in SARS-CoV-2 Pathogenesis

6.1. Introduction

Coronavirus disease 2019 (COVID-19) is a respiratory disease that primarily manifests with pneumonia-like symptoms, gastrointestinal symptoms and occasionally associated with multiorgan failure.[1,2] The causative agent SARS-CoV-2 is an enveloped positive-sense, single-stranded RNA betacoronavirus, which belongs to the Coronaviridae family. Six human coronaviruses (HCoVs) have been previously identified. However, over the past two decades, highly pathogenic HCoVs have emerged. These include SARS-CoV in 2002 with a mortality rate of 10%, and MERS-CoV emerged in 2012 with a case fatality rate of 36%.[3,4] Although the overall mortality rate due to the SARS-CoV-2 relatively lower, it seemingly spread more efficiently, making it far more challenging to contain and increases its pandemic potential. Despite continuous research activities by the scientific community, efficacious vaccines or antivirals may be months or even years away. In order to understand the pathophysiology of the virus and to develop new therapeutics, it is necessary to develop a broad understanding of host-pathogen interactions, like how SARS-CoV-2 uses the host machinery during infection, and apply this information towards developing both new drugs and repurposing of existing drugs.

In this study, we utilized *in silico* methods for predicting the interactions between SARS-CoV-2 and the host proteins, followed by transcriptomics and proteomics analysis of the SARS-CoV-2 infected Huh7 cell-line. The approach is based on protein structural similarity. We predicted an interaction map for SARS-CoV-2 cellular co-localization information by using previously published data sets of coronaviruses and other respiratory disease-causing RNA viruses.[5-7]

Collectively, with the knowledge of protein-protein interactions of SARS-CoV-2, these predictions provide an additional platform for a better understanding of viral pathogenesis and identification of potential clinical targets. Gene enrichment analysis of these interacting host proteins advocates that cytokine storm and neutrophil degranulation drive acute respiratory disease syndrome (ARDS) in SARS CoV-2 patients. Most strikingly, the Complement and Coagulation cascade are interconnected and potentially key driver of the innate immune response against SARS-CoV-2. This prediction was further validated by the KEGG pathway in gene set enrichment analysis of combined transcriptomics and proteomics data sets resulted from *in vitro* SARS-CoV-2 infections in cell lines as reported earlier.[8] Thus the current study helps us understand the molecular mechanism of the SARS-CoV-2 infection and its role in pathogenesis through the combined activation of cytokine storm, neutrophil degranulation, and the complement system.

6.2. Results

6.2.1. Identification of SARS-CoV-2-similar human proteins (hSARS-CoV-2) and hSARS-CoV-2 host interaction

To identify the list of human host proteins having structural similarity to SARS-CoV-2 proteins, we employed a previously established protocol [9] by collecting the available PDB structures of viral proteins and predicted structures from the Zhang lab.[10] The viral protein structure was then submitted to DaliLite v. five web servers for predicting structure similarity to human proteins. We refer these identified human proteins bearing a domain of high structural similarity to the SARS-CoV-2 protein as "SARS-CoV-2 -similar". Next, we determined the known interactions for these SARS-CoV-2 -similar human (hSARS-CoV-2) proteins. The structural similarity analysis identified 3,735 human proteins (hSARS-CoV-2) similar to the 16 SARS-CoV-2 proteins. Next, we identified all possible interaction partners for these proteins and identified the interacting partners of the 16 proteins of SARS-CoV-2 from all interacting partners of the hSARS-CoV-2 similar proteins downloaded from three different databases (HPRD, BIOGRID, and MINT). A total

of 57,359 unique interaction partners were identified for these 16 SARS-CoV-2 proteins, involving 12,872 unique human interacting proteins. To further enrich the protein pool, we used two different kinds of filters (Cellular compartment analysis (CC), and Literature analysis). First, we used GO cellular compartment (CC) annotation to refine and filter the candidates using the cellular localization pattern. This exercise reduced the number of unique interaction partners to 19,047, with 6,876 unique proteins. Later we curated the data by analyzing the RNA virus dataset, filtered, and obtained 5,903 unique interaction partners with 2,647 unique proteins. This information was subsequently used for the functional pathway analysis and for determining the role of genes in associated pathways.

We compared our predictions with a previously published study by Gordon et al. Experimental data sets published in this study showed SARS-CoV-2 and human proteins resulted in 332 unique interactions. When we aligned our interacting protein list with Gordon et al., we noticed around 91% overlap with our primary interacting list and about 54% genes with our CC and literature filtered list. (Figure 6.1)

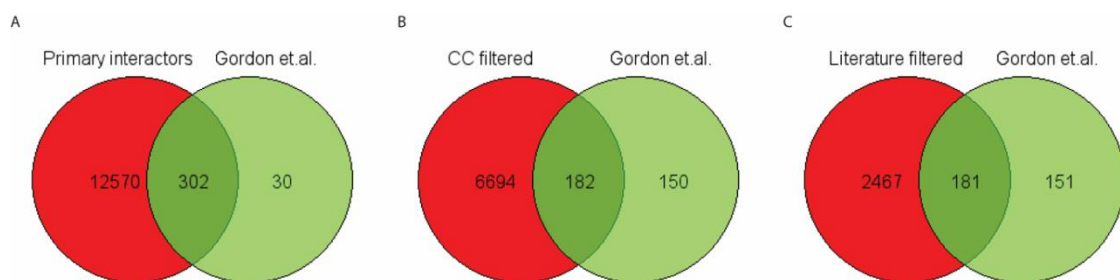


Figure 6.1. Overlap with previous studies: (A) Figure represents the overlap proteins between primary interactors list and that of list published by Gordon et al. Note ~91 % of the list provided in Gordon et al. study overlap with the candidates predicted in our study. (B, C) Diagram represents the overlap proteins between CC and literature filtered protein with Gordon et al. proteins.

6.2.2. Gene enrichment analysis of SARS-CoV-2 interactor proteins

Next, we performed gene enrichment analysis by using g: GOST tool of g: Profiler. g: GOST is the core tool for performing functional enrichment analysis on the input gene list.[11] It mapped a handler-provided list of genes to find relevant information sources and detect statistically significant enriched biological processes and pathways. At first, we uploaded the SARS-CoV-2 interactors protein lists in. g: GOST tools and manually selected GO: biological process (GO: BP), KEGG pathways, and Reactome database in the data source option and set threshold value at .05. Finally, we got g: GOST multi-query Manhattan plot, which shows significantly enriched GO: BP, KEGG terms, and Reactome enhanced data (Table 6.1, Figure 6.2A). Subsequent Reactome dataset analysis revealed that SARS-CoV-2 interactors proteins are enriched with the terms of biomedical pathways. Similarly, the KEGG pathway showed most of the genes associated with salmonella infection, MAPK signaling pathway, complement, and coagulation cascades, endocytosis, PD-L1 expression, and PD-1 checkpoint pathway in cancer and C-type lectin receptor signaling pathway. To check the essential proteins of these signaling pathways and their overlapping genes, we performed the degree centrality analysis of genes associated with highly enriched pathways by using the CytoNCA tool of Cytoscape. The degree centrality analysis revealed that the MAPK1, MAPK3, AKT1, and SRC proteins play a crucial role in six highly relevant biomedical pathways. These include cytokine signaling in the immune system, MAPK signaling pathway, PD-L1 expression, and PD-1 checkpoint pathway in cancer, platelet activation, Innate immune system, and C-type lectin receptor signaling pathways. (Figure 6.2B).

ID	Source	Term ID	Term Name	Padj (query_1)
1	KEGG	KEGG:05132	Salmonella infection	3.974×10^{-17}
2	KEGG	KEGG:04010	MAPK signaling pathway	8.579×10^{-15}
3	KEGG	KEGG:04610	Complement and coagulation cascades	9.573×10^{-14}
4	KEGG	KEGG:04144	Endocytosis	4.423×10^{-13}
5	KEGG	KEGG:04722	Neurotrophin signaling pathway	6.123×10^{-12}
6	KEGG	KEGG:04145	Phagosome	4.708×10^{-11}
7	KEGG	KEGG:04611	Platelet activation	2.669×10^{-10}
8	KEGG	KEGG:04068	FoxO signaling pathway	1.514×10^{-9}
9	KEGG	KEGG:05235	PD-L1 expression and PD-1 checkpoint pathway in...	3.956×10^{-9}
10	KEGG	KEGG:04625	C-type lectin receptor signaling pathway	5.300×10^{-9}
11	KEGG	KEGG:04066	HIF-1 signaling pathway	1.328×10^{-8}
12	KEGG	KEGG:03050	Proteasome	1.720×10^{-6}
13	KEGG	KEGG:04012	ErbB signaling pathway	5.535×10^{-6}
14	KEGG	KEGG:04151	PI3K-Akt signaling pathway	7.975×10^{-6}
15	KEGG	KEGG:04915	Estrogen signaling pathway	2.851×10^{-5}
16	KEGG	KEGG:04062	Chemokine signaling pathway	4.189×10^{-5}
17	KEGG	KEGG:04014	Ras signaling pathway	1.038×10^{-4}
18	KEGG	KEGG:04660	T cell receptor signaling pathway	4.742×10^{-4}
19	KEGG	KEGG:04668	TNF signaling pathway	8.981×10^{-4}
20	KEGG	KEGG:04662	B cell receptor signaling pathway	9.830×10^{-4}
21	REAC	REAC:R-HSA-42...	Axon guidance	4.425×10^{-42}
22	REAC	REAC:R-HSA-16...	Innate Immune System	7.175×10^{-36}
23	REAC	REAC:R-HSA-12...	Cytokine Signaling in Immune system	4.669×10^{-32}
24	REAC	REAC:R-HSA-37...	Signaling by ROBO receptors	7.459×10^{-28}
25	REAC	REAC:R-HSA-67...	Neutrophil degranulation	1.957×10^{-27}
26	REAC	REAC:R-HSA-90...	Signaling by Receptor Tyrosine Kinases	7.213×10^{-26}
27	REAC	REAC:R-HSA-16...	Influenza Life Cycle	1.548×10^{-23}
28	REAC	REAC:R-HSA-72...	Eukaryotic Translation Initiation	7.206×10^{-23}
29	REAC	REAC:R-HSA-56...	Vesicle-mediated transport	9.396×10^{-17}
30	REAC	REAC:R-HSA-10...	Apoptosis	2.639×10^{-16}
31	REAC	REAC:R-HSA-56...	MAPK family signaling cascades	1.206×10^{-14}
32	REAC	REAC:R-HSA-10...	Hemostasis	4.371×10^{-12}
33	REAC	REAC:R-HSA-11...	Downstream signaling events of B Cell Receptor (B...	4.880×10^{-12}
34	REAC	REAC:R-HSA-16...	Toll Like Receptor 3 (TLR3) Cascade	7.405×10^{-12}
35	REAC	REAC:R-HSA-88...	Clathrin-mediated endocytosis	3.820×10^{-10}
36	REAC	REAC:R-HSA-91...	Interferon Signaling	5.476×10^{-10}

version e99_eg46_p14_f929183
date 5/25/2020, 11:07:11 AM
organism hsapiens

Table 6.1. List of significantly enriched KEGG terms and Reactome enhanced data. Compilation of 2,647 unique proteins entries resulting in highly enriched biomedical relevant KEGG term and the Reactome pathways listed above having *p*-adjusted *p*-values.

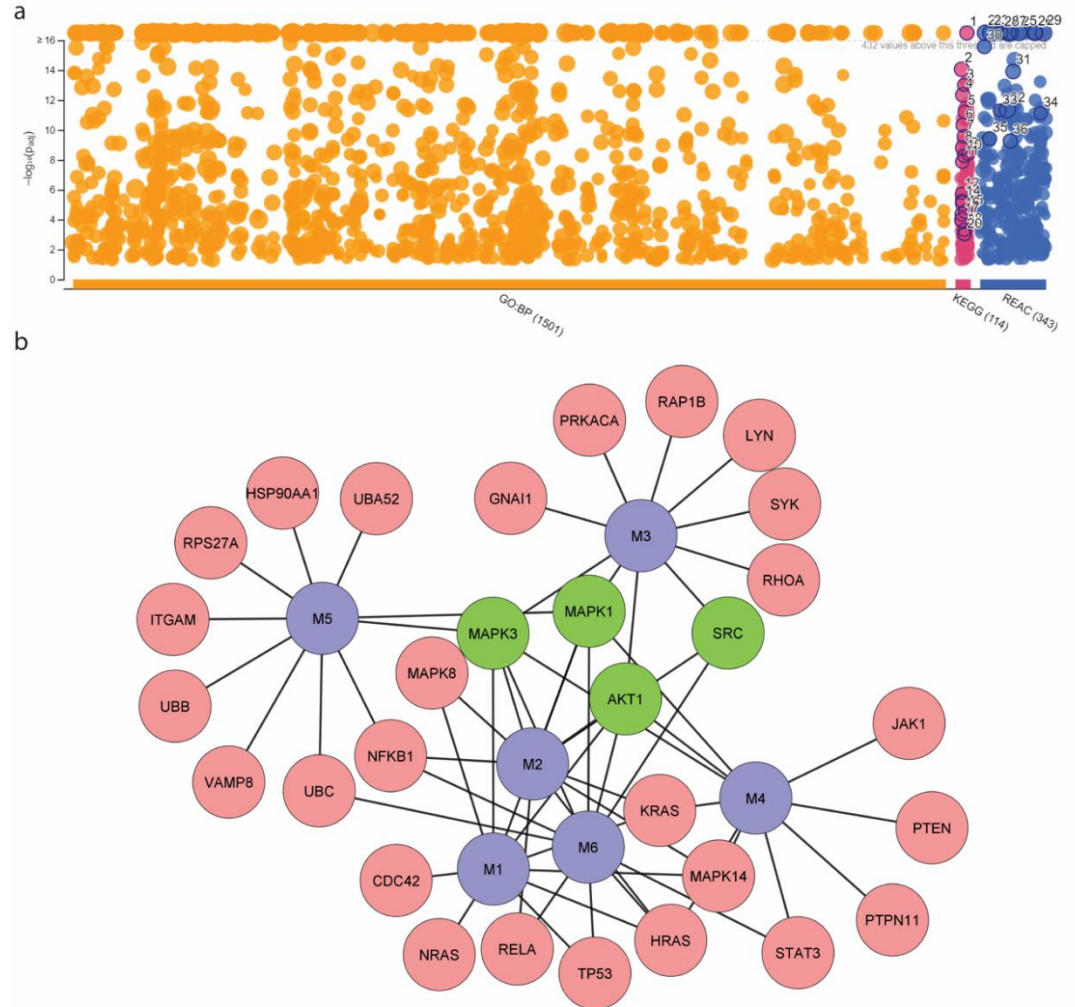


Figure 6.2. GOST multi-query Manhattan plot and Degree centrality analysis of overlapping proteins in the profoundly enriched pathways. (a) g: GOST multi-query Manhattan plot shows significantly enriched GO: BP, KEGG terms, and Reactome enhanced data for SARS CoV-2 interacting proteins. (b) Network analysis based on the input pathways, showing overlapping proteins between 6 highly enriched pathways. Of these, the subnetwork shows MAPK1, MAPK3, AKT1, and SRC proteins are connected maximally. Abbreviations: M1: MAPK signaling pathway, M2: C-type lectin receptor signaling pathway, M3: Platelet activation, M4: PD-L1 expression, and PD-1 checkpoint pathway in cancer, M5: Innate Immune System, M6: Cytokine Signaling in Immune system.

6.2.3. MAPK is predicted to be the key player in the innate immune response

Coronavirus can evade the host innate immune response by interacting with key components of signaling pathways. GO enrichment analysis of predicted candidates showed 369 host genes associated with the innate immune system to interact with SARS-CoV-2 proteins. *In silico* analysis revealed that viral proteins like S, NSP13, NSP15, NSP7, NSP8, NSP 5, and NSP4 interact with mitochondrial antiviral-signaling protein (MAVS), an essential intermediate of RIG-I signaling-mediated IFN-I production. This interaction of the viral proteins with MAVS can inhibit IFN-I production. For example, by blocking CARD-CARD domain interaction between RIG-I and MAVS, promoting proteasomal degradation of MAVS or interfering with the recruitment of tumor necrosis factor receptor-associated factor (TRAF) family proteins.[12] Significantly, the same viral proteins also can interact with TRIM-25 E3 ligase, which is essential for the ubiquitination of RIG-I and activation of the signaling cascade. The SARS-CoV-2 proteins also can interact with effector targets of the RIG-I signaling like IRF3 and IRF7 and inhibit its translocation to the nucleus and downregulate IFN-I production.[13] We also observed that viral protein might interact with components of TLR-mediated IFN response like TRAF6 and IRAK1.[14]

The IFN-I signaling can also activate mitogen-activated protein kinase (MAPK) signaling that regulates several cellular processes ranging from inflammation, cell differentiation, stress to apoptosis and metabolism. On the other hand, the MAPK signaling pathway is activated by several viruses. Interestingly, MAPK is essential for coronavirus replication and, also associated with the production of inflammatory cytokines (Figure 6.3).[15,16] In this context, the degree centrality analysis of the predicted 369 host genes revealed that the MAPK signaling pathway could play a central role in the innate immune signaling in response to SARS-CoV-2 infection. In the downstream, the MAPK-signaling could result in a cascade of other potent pathways that potentially influence the outcome of viral pathogenesis.[17]

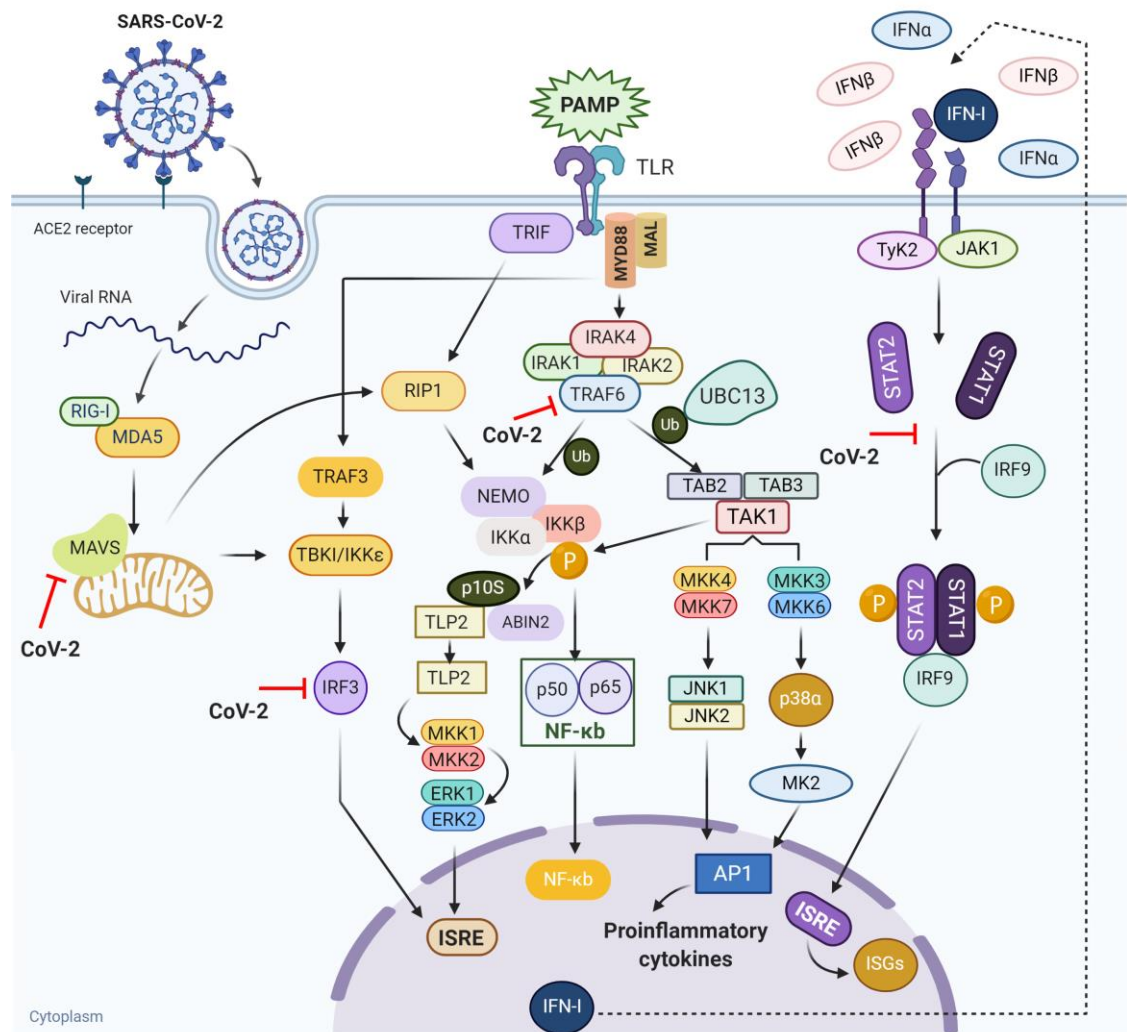


Figure 6.3. Type I interferon induction and signaling during SARS-CoV-2 infection and virus-mediated inhibition of IFN-I and ISGs. The schematic diagram represents key players of IFN-I pathways associated with SARS-CoV-2 infection. After entering the cells, the innate sensors such as MDA5, RIG1, and PAMS recognize the viral proteins and nucleic acids, which then activate IFN-I and pro-inflammatory cytokine production. Subsequently, IFN-I enhances the production of ISGs and sets the stage of the potent antiviral immune response. At the same time, the SARS CoV-2 proteins interact with the MAVS, TRAF6,

IRAK1 and negatively regulate IFN-I signaling and dampen host immune response.

Abbreviations: TyK2: Tyrosine kinase 2, JAK1: Janus kinase 1, MK2: MAPK-activated protein kinase 2, MKK1/2/3/4/6/7 : Mitogen-activated protein kinase, ERK1/2: extracellular signal-regulated kinases, JNK1/2: c-Jun N-terminal kinases, TAB2/3: TGF-Beta Activated Kinase Binding Protein, TAK1: TGF-beta-activated kinase, ABIN2: TNFAIP3 interacting protein 3, IRF 3/9: Interferon regulatory factor 3/9, STAT1/2: Signal Transducer and Activator Of Transcription, IKK α : I κ B Kinase α , IKK β : I κ B Kinase β , NEMO: NF-Kappa-B essential modulator, RIGI: retinoic acid-inducible gene I, MDA5: melanoma differentiation-associated protein 5, MAVS: Mitochondrial antiviral-signaling protein, TBK1: TANK Binding Kinase 1, TRAF3: TNF Receptor Associated Factor 3, TRAF6: TNF Receptor Associated Factor 6, TRIF: TIR domain-containing adapter molecule 1, MYD88: Myeloid differentiation primary response protein MyD88, IRAK1/2/4: Interleukin-1 receptor-associated kinase, PAMP: Pathogen-associated molecular pattern, TLR: Toll-like receptors, TLP2: Thioredoxin, ISG: Interferon-stimulated gene, AP1: Activator protein 1, ISRE: Interferon-Stimulated Response Element, IFN-I: Type I Interferon, RIP1: Receptor Interacting Serine/Threonine Kinase 1.

6.2.4. TP53 and VAMP8 respectively are the key proteins of Cytokines storm and neutrophil degranulation process associated with acute respiratory distress syndrome (ARDS)

The SARS-CoV-2 causes acute lung injury and acute respiratory distress syndrome (ARDS), resulting in severe lethal conditions.[18] Our results suggest that cytokine signaling and neutrophils degranulation pathways are the most significant pathways associated with SARS-CoV-2 infection. We show that 299 candidates (associated with cytokine signaling) interact with SARS-CoV-2 proteins. The degree centrality analysis of these genes suggests that cellular tumor antigen p53 (TP53), followed by tumor necrosis factor (TNF), mitogen-activated protein kinase 3 (MAPK3), and mitogen-activated protein kinase 1 (MAPK1) have crucial and central roles in cytokines signaling (Figure 6.4A) (Table 6.2). $\text{TNF}\alpha$ is a potent pro-inflammatory factor and key regulator of immune cell functions. In viral infections, $\text{TNF}\alpha$ can induce pro-inflammatory cytokine production and activates TNF-dependent pathways. This primarily directs NF- κ B-mediated cytokines production, i.e., IL6, IL4.[19] Also, SARS-CoV-2 viral proteins can stimulate the MAPK and NF- κ B pathways (Figure 6.4A), leading to the induction of IL6 production.

Table 6.2. Degree centrality analysis of proteins associated with Cytokines storm.

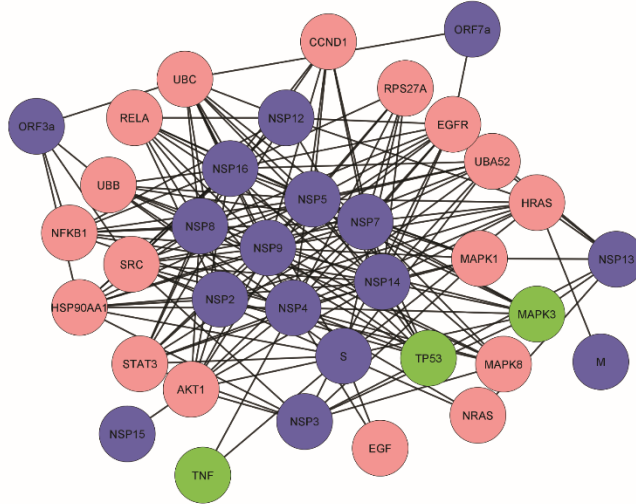
Serial Number	Host Protein	Ensemble Gene ID	Degree Centrality (Weight)
1	TP53	ENSP00000269305	114.272
2	TNF	ENSP00000398698	113.552
3	MAPK3	ENSP00000263025	110.762
4	MAPK1	ENSP00000215832	106.453
5	AKT1	ENSP00000451828	103.622
6	HRAS	ENSP00000407586	98.912
7	STAT3	ENSP00000264657	98.127
8	SRC	ENSP00000362680	98.033
9	RELA	ENSP00000384273	87.107
10	UBC	ENSP00000441543	84.869
11	NFKB1	ENSP00000226574	84.853
12	NRAS	ENSP00000358548	83.778
13	UBA52	ENSP00000388107	83.556
14	RPS27A	ENSP00000272317	82.657
15	UBB	ENSP00000304697	82.254
16	EGFR	ENSP00000275493	82.104
17	MAPK8	ENSP00000378974	79.934
18	HSP90AA1	ENSP00000335153	79.78
19	EGF	ENSP00000265171	79.206
20	CCND1	ENSP00000227507	78.367

Neutrophils are present during many lung diseases associated with acute respiratory distress syndrome (ARDS) and may be involved in severe lung injury [20] in COVID-19 patients. Neutrophils form extracellular web-like structures of DNA and proteins called neutrophil extracellular trap (NET). These NETs induce mucus accumulation in patients' airways and facilitate ARDS during viral infection.[21] Presently, we noticed that 193 proteins are associated with neutrophils degranulation to interact with SARS-CoV-2 proteins. Degree centrality analysis suggests that vesicle-associated membrane protein 8 (VAMP8), integrin alpha-M (ITGAM), and erythrocyte band 7 integral membrane protein (STOM) are the key regulators of neutrophil degranulation (Figure 6.4B) (Table 6.3). VAMP8 is a member of the soluble *N*-ethylmaleimide-sensitive fusion protein attachment protein receptor (SNARE) family of fusion proteins. It confines on secretory granules, and degranulation is inhibited in VAMP8-deficient macrophages and mast cells.[22] VAMP8 is also expressed in human neutrophils and may follow a similar mechanism of degranulation.[23] The complement system also activates neutrophils degranulation. The C5a triggers the phagocytic NADPH-oxidative burst and enhances phagocytosis and the release of granule molecules from neutrophils and macrophages. [24,25]

Table 6.3. Degree centrality analysis of proteins associated with neutrophil degranulation.

Serial Number	Host Protein	Ensemble Gene ID	Degree Centrality (Weight)
1	VAMP8	ENSP00000263864	72.583
2	ITGAM	ENSP00000441691	64.175
3	STOM	ENSP00000286713	62.88
4	C3AR1	ENSP00000302079	60.08
5	ITGB2	ENSP00000380948	58.371
6	DNAJC5	ENSP00000354111	57.133
7	TOM1	ENSP00000413697	56.9
8	CD59	ENSP00000379191	56.825
9	MAPK1	ENSP00000215832	56.629
10	CKAP4	ENSP00000367265	56.388
11	SNAP23	ENSP00000249647	54.835
12	CYBB	ENSP00000367851	54.389
13	TMEM30A	ENSP00000230461	54.252
14	CMTM6	ENSP00000205636	54.165
15	GAA	ENSP00000305692	52.909
16	ATP6V0C	ENSP00000329757	52.449
17	ADAM10	ENSP00000260408	52.051
18	PLD1	ENSP00000342793	51.552
19	UBR4	ENSP00000364403	50.564
20	CYBA	ENSP00000261623	50.427

a



b

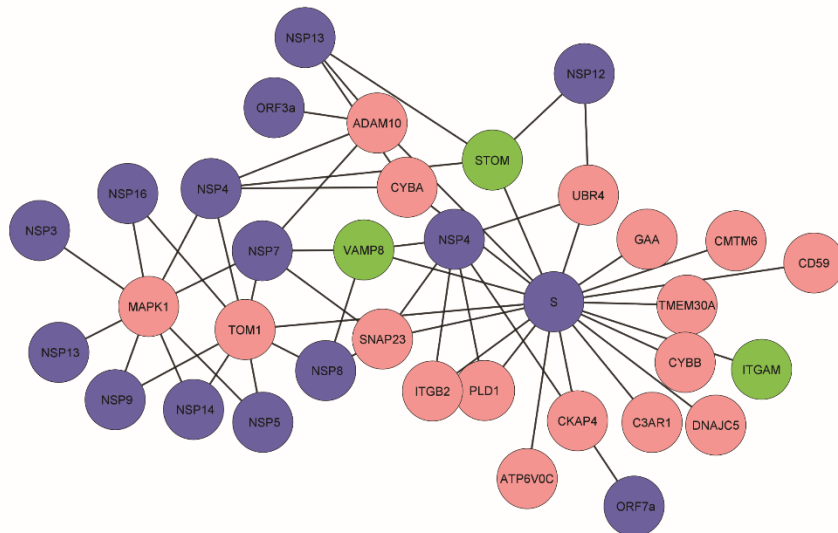


Figure 6.4. Predicted interaction map of SARS-CoV-2 proteins and top 20 host proteins associated with cytokines signaling pathway and Neutrophils degranulation. Blue color represents the virus proteins, and brown color represents the human interactor proteins. (a) Sixteen SARS-CoV-2 proteins interact with human proteins associated with the cytokines pathway. TP53, TNF, and MAPK3 (illustrated in green color) are highly weighted proteins and may regulate cytokine production. (b) Thirteen SARS-CoV-2 proteins interact with human proteins associated with Neutrophils degranulation. Degree centrality analysis reveals that VAMP8, ITGAM, and STOM (represented in green color) are the key players in the neutrophils degranulation process.

6.2.5. The complement system-induced thrombosis in SARS CoV-2 patients

Extra-pulmonary microvascular injury associated with blood thickening and clotting is being reported in different organs and blood vessels of SARS-CoV-2 infected patients. These virus-induced thrombotic changes manifest in endotheliitis, and intravascular coagulation is presumed to be the primary cause of stroke and other complications in patients. Earlier, Margo et al. published a case study on five seriously ill SARS CoV-2 infected patients and reported that complement components C3, C4, the terminal complex C5b-9 [also known as the membrane attack complex (MAC)] were deposited on the coagulation area in the lungs alveoli.[26] The COVID-19 patients showed thrombocytopenia and possibly elevated D-dimer level, a fibrinolysis-specific degradation product. [27],[28] The thrombocytopenia and elevated D-dimer can be described by the extreme activation of the coagulation cascade and platelets.[29]

In our KEGG pathway and gene enrichment analysis, we discovered two highly enriched, prothrombotic, i.e., Complement and coagulation cascade, and platelet activation pathways, presumably directing thrombotic activities in severe COVID-19 patients. Our analysis shows that 52 proteins related to Complement and coagulation cascades interact with SARS-CoV-2 proteins. Degree centrality analysis of these genes suggests that Kininogen-1 (KNG1), followed by C3, and Fibrinogen gamma chain (FGG), play a crucial role in activating Complement and coagulation cascades (Figure 6.5) (Table 6.4). Our study suggests that among all SARS-CoV-2 proteins, the spike protein showed high weighted interaction with proteins associated with complement and coagulation cascade and potentially could activate proteolytic processing of C3. The C3 hydrolysis leads to C3a and C3b, subsequently further cleavage of C5 into C5a and C5b. Activation and generation of C3a and C5a are well-known inducers for the production of inflammatory cytokines that leads to tissue damage in the lungs.

Table 6.4. Degree centrality analysis of proteins associated with complement and coagulation cascade.

S.No.	Host Protein	Ensemble Gene ID	Degree Centrality (Weight)
1	KNG1	ENSP00000265023	30.174
2	C3	ENSP00000245907	29.864
3	FGG	ENSP00000336829	29.062
4	FGA	ENSP00000306361	28.534
5	PLG	ENSP00000308938	28.522
6	SERPINC1	ENSP00000356671	28.022
7	F2	ENSP00000308541	27.394
8	FGB	ENSP00000306099	25.675
9	PROC	ENSP00000234071	23.989
10	F5	ENSP00000356771	23.368
11	C4A	ENSP00000396688	22.361
12	F3	ENSP00000334145	22.34
13	SERPINE1	ENSP00000223095	21.536
14	F8	ENSP00000353393	21.432
15	F10	ENSP00000364709	20.901
16	SERPINA1	ENSP00000416066	20.644
17	THBD	ENSP00000366307	20.425
18	CD59	ENSP00000379191	18.566
19	ITGAM	ENSP00000441691	18.562
20	VTN	ENSP00000226218	17.758

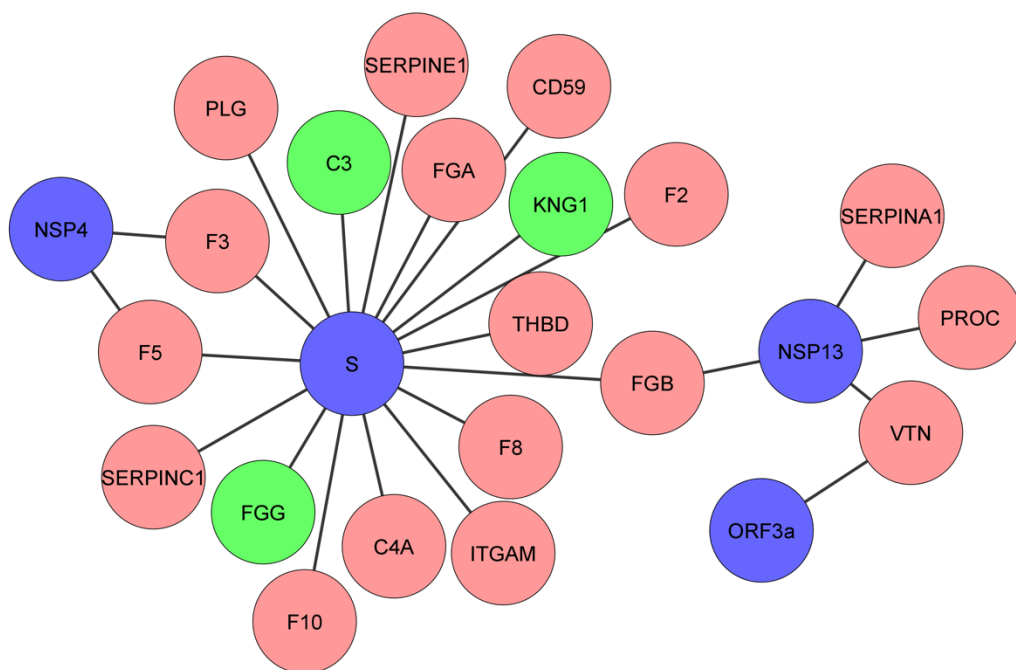


Figure 6.5. Predicted interaction map of SARS-CoV-2 proteins and the top 20 host proteins associated with complement and coagulation cascade. Blue color represents the virus proteins, and brown color represents the human interactor proteins. Four SARS-CoV-2 proteins interact with human proteins associated with complementing and coagulation cascade. Most of the proteins interact with spike protein. Degree centrality analysis reveals that KNG1, C3, and FGG (represented in green color) are the key player in the complement and coagulation cascade.

The alteration of cellular membranes is essential in platelet activation, which is required for the primary clot formation. Association of the complement C5b-9 complex into the cell membrane activates platelets and results in the appearance of procoagulant lipids.[30] Modification of the phospholipid is also required for the tissue factor (TF)-based coagulation pathway. TF starts the extrinsic pathway of coagulation and is highly expressed in various tissues, including the brain, lung, kidney, and placenta.[31]. Our analysis shows that the activated complement substrate, specifically C5, can cause an increase in the expression of functionally active TF in leukocytes and endothelial cells (Figure 6.6).[32] Although neutrophils carry TF on their membranes, it is unclear whether they can produce TF alone. It is tempting for us to speculate that the enhanced coagulation process observed in SARS-CoV-2 infection is due to higher neutrophils degranulation.

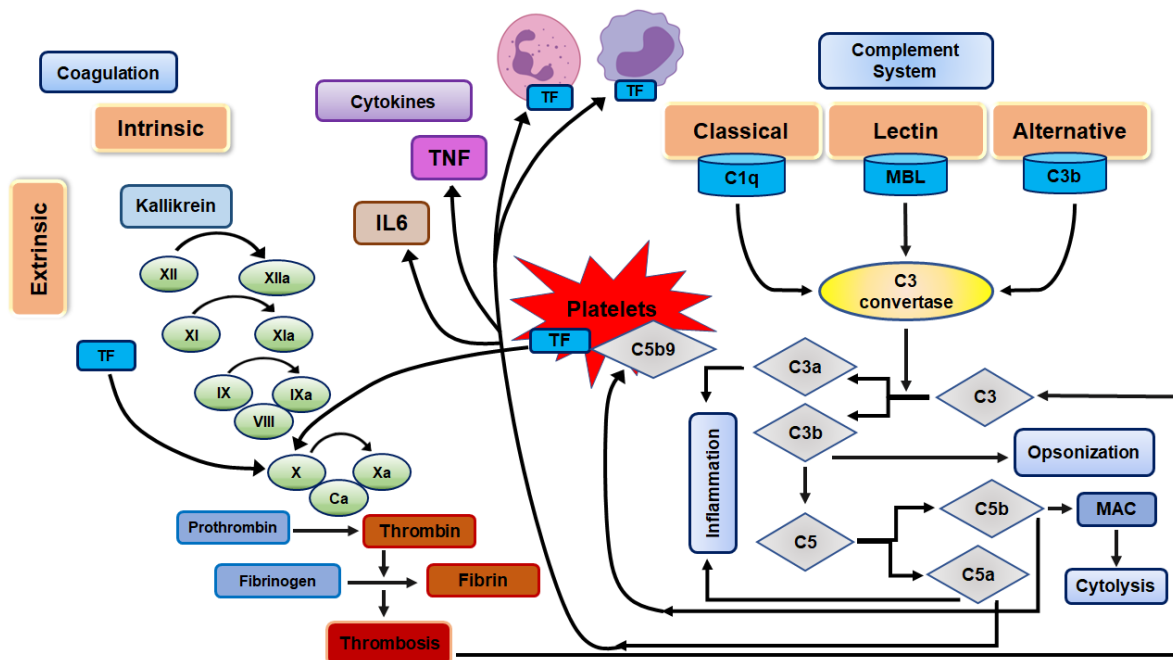


Figure 6.6. Schematic representation of the mechanism of complement system-mediated thrombosis. SARS-CoV-2- Spike (S) protein interacts with C3 and activates proteolytic processing of C3, C3 hydrolyzes in C3a and C3b, further C3b cleaves C5 into C5a and C5b. C3a and C5a induce inflammatory cytokines production. Further, C3a, C5a, and C5b9 activate the coagulation pathway. C5a also activates IL6 and TNF alpha production, which also magnifies the coagulation pathway.

6.2.6. Genes associated with complement and coagulation pathways are upregulated in the SARS-CoV-2 infected Huh7 cells

Our *in silico* analysis showed strong evidence of complement and coagulation cascade in SARS-CoV-2 infection. We then validated our findings in *vitro* experiments by re-analyzing the proteomics and transcriptomics data from the study by Appelberg et al., where Huh7 cells were infected with SARS-CoV-2 for 72 h. Proteomics and transcriptomics data sets were analyzed against 85 genes associated with the ‘Complement and coagulation pathways’ obtained from the KEGG database. In globality, there was a constant increase in transcript expression (Figure 6.7A) and protein abundance (Figure 6.7B) of coagulation associated genes from the uninfected stage to 72 hours post-infection. Fold change of hits was relatively low because proteins or transcript abundance was not increasing constantly. In the main study by Appelberg *et al.*, the authors showed that the most significant increase in transcriptomics was between 48 and 72 h, while for proteomics, it was between 24 and 48 h. The network of experimentally observed genes associated with complement and coagulation cascade shows a highly interconnected network (Figure 6.7 C, D). The network was annotated with transcriptomics (Figure 6.7C) and proteomics data (Figure 6.7D). Among 85 genes associated with complement and coagulation, 59 transcripts were detected, of which 44 showed significant changes in the transcriptomics data set. While in the proteomics data set, 36 proteins were detected, and 28 showed significant changes. There was a considerable overlap of 24 genes/proteins. Thus, by analyzing both (*in silico* and *in vitro*) data sets, we suggest that most proteins and transcripts are upregulated during the course of infection and strongly activate the complement and coagulation pathways. These findings can be extrapolated to existing pathophysiological observations of thrombotic events in COVID-19 disease, which could be directly influenced by activation of complement and coagulation pathways.

*transcripts were detected, and 44 have a differential expression over time (d)
Cytoscape network of KEGG pathway “Complement and coagulation cascades”
labeled with proteomics results.*

*36 proteins were detected, and 28 have differential abundance over time.
Genes or proteins are represented as circles. Gradient color was applied to
proteins depending on temporal fold change calculated by LIMMA (low =
green to high = red). Nonsignificant proteins or transcripts are represented
with transparency. Non-detected proteins are represented grey. The size of the
significant protein is proportional to the fold change.*

6.3. Discussion

In the study, we established a network of protein-protein interactions between the host and SARS-CoV-2 proteins by using a protein structure-based computational approach. The GO analysis of these proteins identified their involvement majorly in the innate immune system, cytokine signaling, MAPK signaling pathway, neutrophils degranulation, complement, and coagulation cascades. On further analysis, we discovered that all these pathways overlap considerably with a key set of highly connected genes. The degree centrality analysis revealed that MAPK1, MAPK3, AKT1, and SRC genes are crucial for these pathways and overlapped maximally. This information suggests the existence of a nexus of robust inflammatory events upon SARS-CoV-2 infection. As MAPKs are heterogeneous kinases capable of phosphorylation of serine and threonine residues present in many proteins, the emergence of MAPK-associated pathways is seen as the likely reason for a profound pro-inflammatory cytokines production: stimulation and induction of strong innate response in the host. We also discovered that to counter IFN-I, a few SARS-CoV-2 proteins may interact with IFN-I signaling pathways associated proteins like MAVS, IRF3, TRIM21, TRAF6, and IRK1 and could inhibit their functions.[14,33] By combining these observations, we suggest that the MAPK signaling pathway is central to the innate immune response to SARS-CoV-2 infection and associated pathophysiology. We have also found some other enriched pathways, which may be crucial for SARS-CoV-2 mediated pathogenesis, such as PD-L1 expression and PD-1 checkpoint pathway in cancer and C-type lectin receptor signaling pathways. A recent clinical study shows that PD-L1 expression is associated with COVID-19 patient's severity. It is highly activated in COVID-19 severe patients.[34] as it was earlier reported that PD-1 and PD-L1 are vital mediators in T cell depletion in sepsis and cancer patients.[35,36] Our observations are in line with the clinically approved drug Camrelizumab (human anti-programmed cell death 1 (PD-1) monoclonal antibody) with thymosin is already in use against SARS-Cov-2.[37] Similarly, C-type lectin receptor signaling pathways are also enriched in our study and might be associated with induction of the pro-inflammatory response after virus infection.[38]

Clinical studies and case reports suggest the COVID-19 disease severity is related to the degree of ARDS, lung injury, hypoxia, thrombosis, ischemia, and reperfusion injury. These pathological manifestations are also observed along with the increased cytokine storm, neutrophils degranulation, and activation of components of the complement and coagulation pathway.[39,40] We show highly enriched pathways exactly supporting these clinical observations and further unravel the plausible mechanism behind these phenomena. The degree centrality analysis shows cytokine-associated pathways are regulated by key players such as TP53, TNF, MAPK3, and MAPK1 proteins. Similarly, degree centrality analysis of neutrophils degranulation genes highlighted VAMP8, ITGM, and STOM to be the highly weighted proteins in this pathway. Thus, we propose that these molecules could be the target to mitigate viral infection.

Most strikingly, our observations were narrowed down to the Complement and Coagulation cascade's activation and showed their interconnectivity in SARS-CoV-2 pathogenesis. These findings could be a step in solving the thrombotic events observed in severe COVID19 patients.[41] our *In vitro* experimental results identified a vast majority of known players associated with complement and coagulation during viral infection. On the other hand, the complement system also inhibits an anticoagulation process. The C4b-binding protein (C4BP) is an essential cofactor for the enzymatic degradation of C4b. It can make a complex with protein S (PS), a vitamin-K-dependent anticoagulant glycoprotein. The PS is the cofactor of activated protein C (APC), which degrades coagulation factors Va and VIIIa. Formation of the PS–C4BP complex results in a loss of PS cofactor activity, thereby reducing its anticoagulant effects.[42] After proteolytic cleavage of complement cascade results in the formation of C3a and C5a, which regulates cytokines response and enhance the production of TNF and IL6.[43] Cytokines like TNF and IL6 also are strongly associated with the coagulation process. TNF is a potent enhancer of TF expression on monocytes, and IL-6 enhances the production and thrombogenicity of platelets, [44]. Therefore, they can also decrease the anticoagulant factors like; thrombomodulin, the endothelial cell protein C receptor, and PS.[44] Activation of the complement system, coagulation cascade, and

neutrophils degranulation are interconnected and could result in the pleiotropic effect in response to the SARS-CoV-2 infection. This could exacerbate pathogenesis of lung parenchymal cells, diminished oxygen uptake, and cause rapid thrombotic events and intravascular coagulation, which could result in multiorgan failure, including central nervous system complications and death. [45]

Among the viral proteins, spike protein interacts with complement components such as with C3, C4A, C5, and ORF 3a product interact with mannose-binding lectin2 (MBL2). This observation is in line with earlier studies conducted on SARS-CoV-1, which revealed direct interaction with MBL2 and activation of the complement system.[46,47] The C3 is a critical regulatory protein of the complement system. This notion is supported by a recent study on SARS-CoV-2, which revealed that activation of complement component C3 exacerbates disease pathology in SARS-CoV-associated ARDS. Mice deficient with C3 ($C3^{-/-}$) exhibited significantly lesser respiratory dysfunction despite having equivalent viral loads in the lungs. This was associated with considerably fewer neutrophils and inflammatory monocytes present in the lungs of $C3^{-/-}$ compared to the control mice. Subsequent investigations exhibited reduced lung pathology and lower cytokine and chemokine levels in the tissue and the sera of $C3^{-/-}$ mice.[48] Thus, the Complement system chemoattractant property may cause neutrophilia in damaged tissue. Based on previous studies and our experimental results, we propose the complement system, along with neutrophils, significantly modulate the thrombogenesis pathway during SARS-CoV-2 infection.

Since the complement system plays a crucial role during the entire course of pathological events during COVID-19 disease progression, a recent study shows that the anti-complement C3 inhibitor AMY-101 successfully treats COVID-19 patients. [49] This supports our findings that the complement system is a key pathway of the innate immune system associated with SARS-CoV-2 mediated pathogenesis.[50] We propose active research on antiviral drug discovery front targeting the key molecules on the complement and coagulation cascade pathway. as our study is limited to in silico and in vitro experiments and warrants for detailed clinical studies of the samples obtained from critically ill COVID-19 patients.

Future studies with animal models with genetically modified hosts could also help unravel the exact mechanism of this phenomenon and warrants future studies.

6.4. Materials and Methods

6.4.1 Data sources

The crystal and cryo-EM structures of SARS-CoV-2 [main protease (PDB ID:5R7Y), Spike glycoproteins (PDB ID: 6VSB, 6VXX), HR2 Domain (PDB ID:6LVN), NSP15 (PDB ID:6VWW), NSP3 (PDB ID:6W02) and NSP9 (PDB ID: 6W4B)] were obtained from RCSB Protein Data Bank (PDB). The rest of SARS-CoV-2 protein structures (NSP1, NSP2, NSP4, NSP6, NSP7, NSP8, NSP10, NSP12, NSP13, NSP14, NSP16, ORF3a, E, M, ORF6, ORF7a, ORF8, N, ORF10) were curated from Zhang Lab, which modeled these structures by using I-TASSER.[10] Each of the structures for SARS-CoV-2 proteins was compared with proteins of known structure for structural similarities using the DaliLite v.5 webserver.[51] The PDB codes obtained from Dali were mapped to their corresponding Uniprot ID and gene Name by DAVID Gene ID Conversion or Uniprot ID mapping.[52,53] Gene Ontology (GO) analysis of these proteins was performed using the g: Profiler tool and Reactome database. Their interaction networks and degree centrality calculation were annotated using the STRING and CytoNCA tools in Cytoscape. [11,54-56]

6.4.2. Determination of structural similarities between SARS-CoV-2 and human proteins

We determined the structural similarities among SARS-CoV-2 and human proteins from DaliLite v. 5 webserver.[51] The Dali server compares 3D structural coordinates of two PDB entries by aligning alpha carbon distance matrices, allowing them for differences in domain order, and producing a structural similarity score. In the present study, we submitted each SARS-CoV-2 protein (PDB and Zhang Lab) into the Dali web server that searched against the entire PDB dataset for structurally similar proteins, with a z-score above 2.0. As a result, we retrieved all the proteins having structures similar to SARS-CoV-2 proteins available in the

PDB database. From these results, we then collected only those structures that were from the human host. These human SARS-CoV-2 similar proteins are hence referred to as "hSARS-CoV-2 similar" proteins.

6.4.3. Interaction prediction

To predict the human endogenous proteins interacting with the SARS-CoV-2 proteins, we investigated the target proteins that interact with the hSARS-CoV-2 similar proteins during cellular processes. We found known interactions between hSARS-CoV-2 similar proteins and target human endogenous proteins using data from the BIOGRID, HPRD, and MINT database.[57-59] These datasets are sourced from literature-curated interactions among human proteins. It is presumed that these cellular proteins, which are known to interact with human protein (similar to SARS-CoV-2 structures), might interact with SARS-CoV-2 proteins due to their structural similarity.

6.4.4. Cellular compartmentalization (CC) and Gene enrichment analysis

The assumed host protein interactors of the SARS-CoV-2 proteome were shortlisted based on their functionality and cellular localization. The interactor host proteins were primarily annotated based on protein localization, as theoretically, the two proteins must share at least one cellular compartment for direct communication among them. In our study, cellular compartments were chosen based on published literature and as available in the UniProt database, suggesting the localization of viral proteins in the host cell. Classified proteins from CC study were then submitted to the g: Profiler and Reactome database to retrieve the list of terms enriched with these proteins and the biological pathways, including them.[11,54]

6.4.5. Validation of Predictions

As there may be multiple PDB structures representing the same protein in Dali, there is some repetition in the interaction predictions. In some cases, SARS-CoV-2 proteins were found to have numerous PDB structures pointing to the same interaction predictions. Hence, the predictions counted as single based on a pair of

human Uniprot accessions and SARS-CoV-2 protein names. The predicted interactions supported by the data obtained from experimental studies carried out for SARS-CoV-2 and datasets from related RNA viruses obtained from our VHFIDB database (www.vhfidb.com) is very comprehensive and manually curated database, which contains 9921 host factors information for 72 viral species.

6.4.6. Complement and coagulation cascades pathway in SARS-CoV-2 infection model

We re-analyzed the proteo-transcriptomics data of SARS-CoV-2 infected (1 MOI) Huh7 cells to identify the temporal changes resulted from infection over time.[8] Briefly, Huh7 cells infected with SARS-CoV-2 were collected at 24, 48, and 72 h in triplicates. Transcript and protein abundance was quantified using RNA-seq and LC-MS/MS methods. To identify the relative genes changes, temporal differential abundance analysis was performed using a univariate time series model from the R package LIMMA. In the LIMMA design matrix, separated coefficients were associated with time and replicated in order to extract the difference as a contrast. Moderated paired t-test using LIMMA with adjustment for replicates was used. Benjamini-Hochberg (BH) adjustment was applied, and only proteins and transcripts with adjusted p values < 0.05 were selected. Eighty-five genes associated with the KEGG pathway "Complement and coagulation cascades - Homo sapiens" were retrieved from the KEGG database (<https://www.genome.jp/kegg-bin>). Proteins and transcripts related to complement and coagulation networks were created separately with Cytoscape ver 3.6.1. For each node, fold change and q-value from LIMMA were added to the network template file. Edges between nodes were taken from the string database (<https://string-db.org/>) with high evidence (interaction score > 700), and interactions retrieve from experiments and databases only. Nodes refer to connected objects in the network and edges to the connections between nodes. Heatmap was created using R package heatmap 2.

6.5. References

1. Wang, C., P. W. Horby, F. G. Hayden and G. F. Gao. (2020), A novel coronavirus outbreak of global health concern, *Lancet*, 395, 470-473 (10.1016/s0140-6736(20)30185-9)
2. Zhu, N., D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu *et al.* (2020), A Novel Coronavirus from Patients with Pneumonia in China, 2019, *N Engl J Med*, 382, 727-733 (10.1056/NEJMoa2001017)
3. Ksiazek, T. G., D. Erdman, C. S. Goldsmith, S. R. Zaki, T. Peret, S. Emery, S. Tong, C. Urbani, J. A. Comer, W. Lim *et al.* (2003), A novel coronavirus associated with severe acute respiratory syndrome, *N Engl J Med*, 348, 1953-1966 (10.1056/NEJMoa030781)
4. Graham, R. L., E. F. Donaldson and R. S. Baric. (2013), A decade after SARS: strategies for controlling emerging coronaviruses, *Nat Rev Microbiol*, 11, 836-848 (10.1038/nrmicro3143)
5. Gordon, D. E., G. M. Jang, M. Bouhaddou, J. Xu, K. Obernier, K. M. White, M. J. O'Meara, V. V. Rezeli, J. Z. Guo, D. L. Swaney *et al.* (2020), A SARS-CoV-2 protein interaction map reveals targets for drug repurposing, *Nature* (10.1038/s41586-020-2286-9)
6. Pfefferle, S., J. Schopf, M. Kogl, C. C. Friedel, M. A. Muller, J. Carbajo-Lozoya, T. Stellberger, E. von Dall'Armi, P. Herzog, S. Kallies *et al.* (2011), The SARS-coronavirus-host interactome: identification of cyclophilins as target for pan-coronavirus inhibitors, *PLoS Pathog*, 7, e1002331 (10.1371/journal.ppat.1002331)
7. Xiong, Y., Y. Liu, L. Cao, D. Wang, M. Guo, A. Jiang, D. Guo, W. Hu, J. Yang, Z. Tang *et al.* (2020), Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients, *Emerg Microbes Infect*, 9, 761-770 (10.1080/22221751.2020.1747363)
8. Appelberg, S., S. Gupta, S. Svensson Akusjärvi, A. T. Ambikan, F. Mikaeloff, E. Saccon, Á. Végvári, R. Benfeitas, M. Sperk, M. Ståhlberg *et al.* (2020), Dysregulation in Akt/mTOR/HIF-1 signaling identified by proteo-transcriptomics of SARS-CoV-2 infected cells, *Emerg Microbes Infect*, 9, 1748-1760 (10.1080/22221751.2020.1799723)
9. Doolittle, J. M. and S. M. Gomez. (2011), Mapping protein interactions between Dengue virus and its human and insect hosts, *PLoS Negl Trop Dis*, 5, e954 (10.1371/journal.pntd.0000954)
10. Zhang, C., W. Zheng, X. Huang, E. W. Bell, X. Zhou and Y. Zhang. (2020), Protein Structure and Sequence Reanalysis of 2019-nCoV Genome Refutes Snakes as Its Intermediate Host and the Unique Similarity between Its Spike Protein Insertions and HIV-1, *J Proteome Res*, 19, 1351-1360 (10.1021/acs.jproteome.0c00129)
11. Reimand, J., T. Arak, P. Adler, L. Kolberg, S. Reisberg, H. Peterson and J. Vilo. (2016), g:Profiler-a web server for functional interpretation of gene lists (2016 update), *Nucleic Acids Res*, 44, W83-89 (10.1093/nar/gkw199)
12. Ren, Z., T. Ding, Z. Zuo, Z. Xu, J. Deng and Z. Wei. (2020), Regulation of MAVS Expression and Signaling Function in the Antiviral Innate Immune Response, *Frontiers in immunology*, 11, 1030-1030 (10.3389/fimmu.2020.01030)

13. Matthews, K., A. Schäfer, A. Pham and M. Frieman. (2014), The SARS coronavirus papain like protease can inhibit IRF3 at a post activation step that requires deubiquitination activity, *Virology*, 11, 209 (10.1186/s12985-014-0209-9)
14. Shi, C. S., H. Y. Qi, C. Boularan, N. N. Huang, M. Abu-Asab, J. H. Shelhamer and J. H. Kehrl. (2014), SARS-coronavirus open reading frame-9b suppresses innate immunity by targeting mitochondria and the MAVS/TRAF3/TRAF6 signalosome, *J Immunol*, 193, 3080-3089 (10.4049/jimmunol.1303196)
15. Kumar, R., N. Khandelwal, R. Thachamvally, B. N. Tripathi, S. Barua, S. K. Kashyap, S. Maherchandani and N. Kumar. (2018), Role of MAPK/MNK1 signaling in virus replication, *Virus Res*, 253, 48-61 (10.1016/j.virusres.2018.05.028)
16. Arthur, J. S. and S. C. Ley. (2013), Mitogen-activated protein kinases in innate immunity, *Nat Rev Immunol*, 13, 679-692 (10.1038/nri3495)
17. Bouhaddou, M., D. Memon, B. Meyer, K. M. White, V. V. Rezeli, M. Correa Marrero, B. J. Polacco, J. E. Melnyk, S. Ulferts, R. M. Kaake *et al.* (2020), The Global Phosphorylation Landscape of SARS-CoV-2 Infection, *Cell*, 182, 685-712.e619 (10.1016/j.cell.2020.06.034)
18. Chen, L., H. G. Liu, W. Liu, J. Liu, K. Liu, J. Shang, Y. Deng and S. Wei. (2020), [Analysis of clinical features of 29 patients with 2019 novel coronavirus pneumonia], *Zhonghua Jie He He Hu Xi Za Zhi*, 43, E005 (10.3760/cma.j.issn.1001-0939.2020.0005)
19. Jia, Q., W. Cheng, Y. Yue, Y. Hu, J. Zhang, X. Pan, Z. Xu and P. Zhang. (2015), Cucurbitacin E inhibits TNF- α -induced inflammatory cytokine production in human synovial cells via suppression of PI3K/Akt/NF- κ B pathways, *Int Immunopharmacol*, 29, 884-890 (10.1016/j.intimp.2015.08.026)
20. Camp, J. V. and C. B. Jonsson. (2017), A Role for Neutrophils in Viral Respiratory Disease, *Front Immunol*, 8, 550 (10.3389/fimmu.2017.00550)
21. Barnes, B. J., J. M. Adrover, A. Baxter-Stoltzfus, A. Borczuk, J. Cools-Lartigue, J. M. Crawford, J. Daßler-Plenker, P. Guerci, C. Huynh, J. S. Knight *et al.* (2020), Targeting potential drivers of COVID-19: Neutrophil extracellular traps, *J Exp Med*, 217 (10.1084/jem.20200652)
22. Tiwari, N., C. C. Wang, C. Brochetta, G. Ke, F. Vita, Z. Qi, J. Rivera, M. R. Soranzo, G. Zabucchi, W. Hong *et al.* (2008), VAMP-8 segregates mast cell-preformed mediator exocytosis from cytokine trafficking pathways, *Blood*, 111, 3665-3674 (10.1182/blood-2007-07-103309)
23. Mollinedo, F., J. Calafat, H. Janssen, B. Martín-Martín, J. Canchado, S. M. Nabokina and C. Gajate. (2006), Combinatorial SNARE complexes modulate the secretion of cytoplasmic granules in human neutrophils, *J Immunol*, 177, 2831-2841 (10.4049/jimmunol.177.5.2831)
24. Melendez, A. J. and F. B. M. Ibrahim. (2004), Antisense Knockdown of Sphingosine Kinase 1 in Human Macrophages Inhibits C5a Receptor-Dependent Signal Transduction, Ca^{2+} Signals, Enzyme Release, Cytokine Production, and Chemotaxis, *The Journal of Immunology*, 173, 1596-1603 (10.4049/jimmunol.173.3.1596)
25. Ibrahim, F. B., S. J. Pang and A. J. Melendez. (2004), Anaphylatoxin signaling in human neutrophils. A key role for sphingosine kinase, *J Biol Chem*, 279, 44802-44811 (10.1074/jbc.M403977200)
26. Magro, C., J. J. Mulvey, D. Berlin, G. Nuovo, S. Salvatore, J. Harp, A. Baxter-Stoltzfus and J. Laurence. (2020), Complement associated microvascular injury

and thrombosis in the pathogenesis of severe COVID-19 infection: a report of five cases, *Transl Res* (10.1016/j.trsl.2020.04.007)

27. Guan, W. J., Z. Y. Ni, Y. Hu, W. H. Liang, C. Q. Ou, J. X. He, L. Liu, H. Shan, C. L. Lei, D. S. C. Hui *et al.* (2020), Clinical Characteristics of Coronavirus Disease 2019 in China, *N Engl J Med*, 382, 1708-1720 (10.1056/NEJMoa2002032)
28. Demelo-Rodríguez, P., E. Cervilla-Muñoz, L. Ordieres-Ortega, A. Parra-Virto, M. Toledano-Macías, N. Toledo-Samaniego, A. García-García, I. García-Fernández-Bravo, Z. Ji, J. de-Miguel-Diez *et al.* (2020), Incidence of asymptomatic deep vein thrombosis in patients with COVID-19 pneumonia and elevated D-dimer levels, *Thromb Res*, 192, 23-26 (10.1016/j.thromres.2020.05.018)
29. Subramaniam, S. and I. Scharrer. (2018), Procoagulant activity during viral infections, *Front Biosci (Landmark Ed)*, 23, 1060-1081 (10.2741/4633)
30. Eriksson, O., C. Mohlin, B. Nilsson and K. N. Ekdahl. (2019), The Human Platelet as an Innate Immune Cell: Interactions Between Activated Platelets and the Complement System, *Front Immunol*, 10, 1590 (10.3389/fimmu.2019.01590)
31. Drake, T. A., W. Ruf, J. H. Morrissey and T. S. Edgington. (1989), Functional tissue factor is entirely cell surface expressed on lipopolysaccharide-stimulated human blood monocytes and a constitutively tissue factor-producing neoplastic cell line, *J Cell Biol*, 109, 389-395 (10.1083/jcb.109.1.389)
32. Muhlfelder, T. W., J. Niemetz, D. Kreutzer, D. Beebe, P. A. Ward and S. I. Rosenfeld. (1979), C5 chemotactic fragment induces leukocyte production of tissue factor activity: a link between complement and coagulation, *J Clin Invest*, 63, 147-150 (10.1172/jci109269)
33. Sa Ribero, M., N. Jouvenet, M. Dreux and S. Nisole. (2020), Interplay between SARS-CoV-2 and the type I interferon response, *PLoS Pathog*, 16, e1008737 (10.1371/journal.ppat.1008737)
34. Patel, H., N. J. Ashton, R. J. Dobson, L.-m. Anderson, A. Yilmaz, K. Blennow, M. Gisslen and H. Zetterberg. (2020), Proteomic blood profiling in mild, severe and critical COVID-19 patients, *medRxiv*, 2020.2006.2022.20137216 (10.1101/2020.06.22.20137216)
35. Liechtenstein, T., I. Dufait, C. Bricogne, A. Lanna, J. Pen, K. Breckpot and D. Escors. (2012), PD-L1/PD-1 Co-Stimulation, a Brake for T cell Activation and a T cell Differentiation Signal, *Journal of clinical & cellular immunology*, S12, 006 (10.4172/2155-9899.S12-006)
36. Diskin, B., S. Adam, M. F. Cassini, G. Sanchez, M. Liria, B. Aykut, C. Buttar, E. Li, B. Sundberg, R. D. Salas *et al.* (2020), PD-L1 engagement on T cells promotes self-tolerance and suppression of neighboring macrophages and effector T cells in cancer, *Nature Immunology*, 21, 442-454 (10.1038/s41590-020-0620-x)
37. Lythgoe, M. P. and P. Middleton. (2020), Ongoing Clinical Trials for the Management of the COVID-19 Pandemic, *Trends Pharmacol Sci*, 41, 363-382 (10.1016/j.tips.2020.03.006)
38. Zhao, X., H. Chu, B. H.-Y. Wong, M. C. Chiu, D. Wang, C. Li, X. Liu, D. Yang, V. K.-M. Poon, J. Cai *et al.* (2020), Activation of C-Type Lectin Receptor and (RIG)-I-Like Receptors Contributes to Proinflammatory Response in Middle East Respiratory Syndrome Coronavirus-Infected Macrophages, *The Journal of infectious diseases*, 221, 647-659 (10.1093/infdis/jiz483)

39. Jose, R. J. and A. Manuel. (2020), COVID-19 cytokine storm: the interplay between inflammation and coagulation, *Lancet Respir Med* (10.1016/s2213-2600(20)30216-2)
40. Risitano, A. M., D. C. Mastellos, M. Huber-Lang, D. Yancopoulou, C. Garlanda, F. Ciceri and J. D. Lambris. (2020), Complement as a target in COVID-19?, *Nature Reviews Immunology*, 20, 343-344 (10.1038/s41577-020-0320-7)
41. Connors, J. M. and J. H. Levy. (2020), COVID-19 and its implications for thrombosis and anticoagulation, *Blood*, 135, 2033-2040 (10.1182/blood.2020006000)
42. Rezende, S. M., R. E. Simmonds and D. A. Lane. (2004), Coagulation, inflammation, and apoptosis: different roles for protein S and the protein S-C4b binding protein complex, *Blood*, 103, 1192-1201 (10.1182/blood-2003-05-1551)
43. Markiewski, M. M., R. A. DeAngelis and J. D. Lambris. (2006), Liver inflammation and regeneration: two distinct biological phenomena or parallel pathophysiologic processes?, *Mol Immunol*, 43, 45-56 (10.1016/j.molimm.2005.06.019)
44. Harada, M., D. R. Van Wagoner and S. Nattel. (2015), Role of inflammation in atrial fibrillation pathophysiology and management, *Circ J*, 79, 495-502 (10.1253/circj.CJ-15-0138)
45. Mishra, A. R., S. N. Byraredddy and D. Nayak. (2020), IFN-I Independent Antiviral Immune Response to Vesicular Stomatitis Virus Challenge in Mouse Brain, *Vaccines (Basel)*, 8 (10.3390/vaccines8020326)
46. Zhou, Y., K. Lu, S. Pfefferle, S. Bertram, I. Glowacka, C. Drosten, S. Pöhlmann and G. Simmons. (2010), A Single Asparagine-Linked Glycosylation Site of the Severe Acute Respiratory Syndrome Coronavirus Spike Glycoprotein Facilitates Inhibition by Mannose-Binding Lectin through Multiple Mechanisms, *Journal of Virology*, 84, 8753-8764 (10.1128/jvi.00554-10)
47. Noris, M., A. Benigni and G. Remuzzi. The case of Complement activation in COVID-19 multiorgan impact, *Kidney International* (10.1016/j.kint.2020.05.013)
48. Gralinski, L. E., T. P. Sheahan, T. E. Morrison, V. D. Menachery, K. Jensen, S. R. Leist, A. Whitmore, M. T. Heise and R. S. Baric. (2018), Complement Activation Contributes to Severe Acute Respiratory Syndrome Coronavirus Pathogenesis, *mBio*, 9 (10.1128/mBio.01753-18)
49. Mastaglio, S., A. Ruggeri, A. M. Risitano, P. Angelillo, D. Yancopoulou, D. C. Mastellos, M. Huber-Lang, S. Piemontese, A. Assanelli, C. Garlanda *et al.* (2020), The first case of COVID-19 treated with the complement C3 inhibitor AMY-101, *Clin Immunol*, 215, 108450 (10.1016/j.clim.2020.108450)
50. Magro, C., J. J. Mulvey, D. Berlin, G. Nuovo, S. Salvatore, J. Harp, A. Baxter-Stoltzfus and J. Laurence. (2020), Complement associated microvascular injury and thrombosis in the pathogenesis of severe COVID-19 infection: a report of five cases, *Transl Res*, 220, 1-13 (10.1016/j.trsl.2020.04.007)
51. Holm, L. (2019), Benchmarking fold detection by DaliLite v.5, *Bioinformatics*, 35, 5326-5327 (10.1093/bioinformatics/btz536)
52. (2019), UniProt: a worldwide hub of protein knowledge, *Nucleic Acids Res*, 47, D506-d515 (10.1093/nar/gky1049)
53. Dennis, G., Jr., B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane and R. A. Lempicki. (2003), DAVID: Database for Annotation, Visualization, and Integrated Discovery, *Genome Biol*, 4, P3
54. Fabregat, A., K. Sidiropoulos, G. Viteri, O. Forner, P. Marin-Garcia, V. Arnau, P. D'Eustachio, L. Stein and H. Hermjakob. (2017), Reactome pathway analysis: a

- high-performance in-memory approach, *BMC Bioinformatics*, 18, 142 (10.1186/s12859-017-1559-2)
55. Otasek, D., J. H. Morris, J. Boucas, A. R. Pico and B. Demchak. (2019), Cytoscape Automation: empowering workflow-based network analysis, *Genome Biol*, 20, 185 (10.1186/s13059-019-1758-4)
 56. Szklarczyk, D., J. H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N. T. Doncheva, A. Roth, P. Bork *et al.* (2017), The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible, *Nucleic Acids Res*, 45, D362-d368 (10.1093/nar/gkw937)
 57. Oughtred, R., C. Stark, B. J. Breitkreutz, J. Rust, L. Boucher, C. Chang, N. Kolas, L. O'Donnell, G. Leung, R. McAdam *et al.* (2019), The BioGRID interaction database: 2019 update, *Nucleic Acids Res*, 47, D529-d541 (10.1093/nar/gky1079)
 58. Goel, R., H. C. Harsha, A. Pandey and T. S. Prasad. (2012), Human Protein Reference Database and Human Proteinpedia as resources for phosphoproteome analysis, *Mol Biosyst*, 8, 453-463 (10.1039/c1mb05340j)
 59. Chatr-aryamontri, A., A. Ceol, L. M. Palazzi, G. Nardelli, M. V. Schneider, L. Castagnoli and G. Cesareni. (2007), MINT: the Molecular INTeraction database, *Nucleic Acids Res*, 35, D572-574 (10.1093/nar/gkl950)

CHAPTER 7

Chapter 7

Conclusion and Future prospective

7.1 Conclusion of the thesis

The discipline of virology is the most senior subject of infectious biology. And for a long time, the research was focused on the pathogen itself. Viruses cause significant acute and chronic infectious diseases. They are major patron to the global burden of disease—however, very few vaccines and antiviral drugs are available for these emerging viruses. While many successful antiviral drugs targeting viral proteins are often used, their administration is limited to specific virus species or strains. Some RNA viruses have low fidelity polymerase enzymes such as influenza virus, human immunodeficiency virus 1 (HIV-1), enteroviruses (e.g., EV71), etc. and are prone to rapid mutations, leading to drug-resistant strains.

Additionally, viruses encode very few proteins, thus limiting the number of available targets for drug discovery.

It is established that the host response is uniformly or more important in establishing the pathological outcome of virus infection. Based on the recent outbreak of highly infectious and deadly viruses like SARS-CoV-2, Influenza viruses, etc., we saw vaccine failures against continuously mutating viruses, and insufficient availability of antiviral therapeutics. To overcome contagious virus pathogenesis, it requires that the virology and viral immunology community should focus on host-pathogen interaction and identify new antiviral targets. In this thesis work, first we created a database called VHFIDB (<https://vhfidb.com>). VHFIDB contains information about host factors associated with 72 viral species. Moreover, it hitched with 18 other databases. We have also incorporated some relevant information on contagious animal viruses like porcine reproductive and respiratory syndrome virus (PRRSV), equine arteritis virus, etc. Most importantly, the VHFIDB is also associated with three powerful analytical tools, namely, VHF pathogen network, VHF overlap analysis and, gene enrichment analysis, which

increases the usability of VHFIDB. We believe this database is a valuable resource tool for investigators in virology, biomedical sciences, and those involved in discovering and developing antiviral therapeutics. Further we have studied 2 highly pathogenic human viruses, CHPV and SARS-COV-2.

CHPV is an emerging tropical virus with a high mortality rate in children below the age of 15 years. We have generated and characterized the VSV-based CHPV fluorescence pseudovirus and also described the CHPV entry mechanism. We have found that CHPV enters into the cells by using clathrin and lipid raft. We have also observed that low pH is required for virus uncoating and genome release in the cytoplasm. More interestingly, during our investigation, the antiviral role of BST-2 against CHPV. We have found that BST-2 enhances CHPV virus entry, *in silico* docking of the cytoplasmic domain of BST-2, and glycoprotein of CHPV reveals that the YXY region of BST-2 directly binds with the virus and increases virus entry. Further we have discussed SARS-CoV-2 infection, access, replication, and pathogenesis.

SARS-CoV-2 is a newly emerged virus. It was observed first in late December 2019 in Wuhan, China which causes unusual viral pneumonia. The virus is highly transmissible and has spread fast worldwide. Till today, there have been 111,762,965 COVID-19 cases reported, including 2,479,678 deaths. We have performed a detailed *in silico* and in-vitro study to understand host-pathogen interaction against SARS-CoV-2. Our study was based on structural similarity. We have found 3,735 human proteins (hSARS-CoV-2) similar to the SARS-CoV-2 proteins during our primary investigation by using Dali V5 webserver. We identified all possible interaction partners for these proteins and identified the interacting partners of SARS-CoV-2 from all interacting partners of the hSARS-CoV-2 similar proteins downloaded from three different databases (HPRD, BIOGRID, and MINT). Total of 12,872 unique human interacting proteins, have been identified. Further, we have filtered the list by Cellular compartment analysis (CC) and Literature analysis. Then we have performed the gene enrichment analysis of these viral interacting host proteins.

The GO analysis of interacting proteins reveals that SARS-CoV-2 may use clathrin-mediated endocytosis for its entry. Among the viral proteins, the spike protein was found to interact with a series of host proteins. These include actin and microtubule cytoskeletal proteins, early endosome, late endosome, and HOPS complex proteins. We also predict the role of several host factors essential for SARS-CoV-2 replication and assembly. In our study, we have found that viral genome replication occurs at the DMV sites. Various cellular organelles reorganized and form DMV after virus infection. Here we have reported that the endoplasmic reticulum, the autophagosome, and ERAD machinery are associated with DMV formation. After the DMV formation, SARS-CoV-2 non-structural proteins (NSP3-NSP16) interact with the host proteins and make an RTC complex essential for virus replication. Similarly, these viral protein interacts with candidates associated with ERGIC, where virus assembly takes place.

Further, we elaborated significantly enriched pathways associated with SARS-CoV-2 pathogenesis. These include SARS-CoV-2 mediated apoptosis, IFN- γ signaling pathway, and proteasomal degradation of CD4 pathway, MAPK signaling pathway, complement, and coagulation cascades, cytokines storm, neutrophils degranulation, endocytosis, PD-L1 expression, and PD-1 checkpoint pathway in cancer and C-type lectin receptor signaling pathway. Further, we have performed the degree centrality analysis to know critical proteins in these pathways. Our most exciting finding was that the complement system regulates the coagulation process in SARS-CoV-2 patients. Our in vitro data also support our conclusions; moreover, we came with all possible SARS-CoV-2 pathogenesis modes and predicted the essential proteins associated with that.

7.2 Future prospective

Several outbreaks of old viruses occur each year. At the same time, new viruses also emerge, which increases the burden on governments and scientists to generate vaccines and other viral protein targeting drugs. Creating a vaccine and cure, from lab to production, will take huge time. At that time, our study opens a new path in virus pathogenesis. We have created a database with vast information on the virus and its associated host factors. The research related to virus pathogenesis and drug discovery can choose a host proviral protein essential for many viruses and develop or repurposing drugs against them.

At the same time, we have also come with novel findings related to CHPV and SARS-CoV-2 viruses. We have generated a VSV backbone-based TFP expressing CHPV pseudovirus. A similar platform was used to develop a vaccine against Ebola, Influenza, and many more other viruses. So we can also use this for further investigation as a vaccine candidate against CHPV. CHPV shows an outbreak in India almost every year and causes encephalitis in infants with a higher mortality rate because our pseudovirus is fluorescent, so it can be easy to use for drug testing.

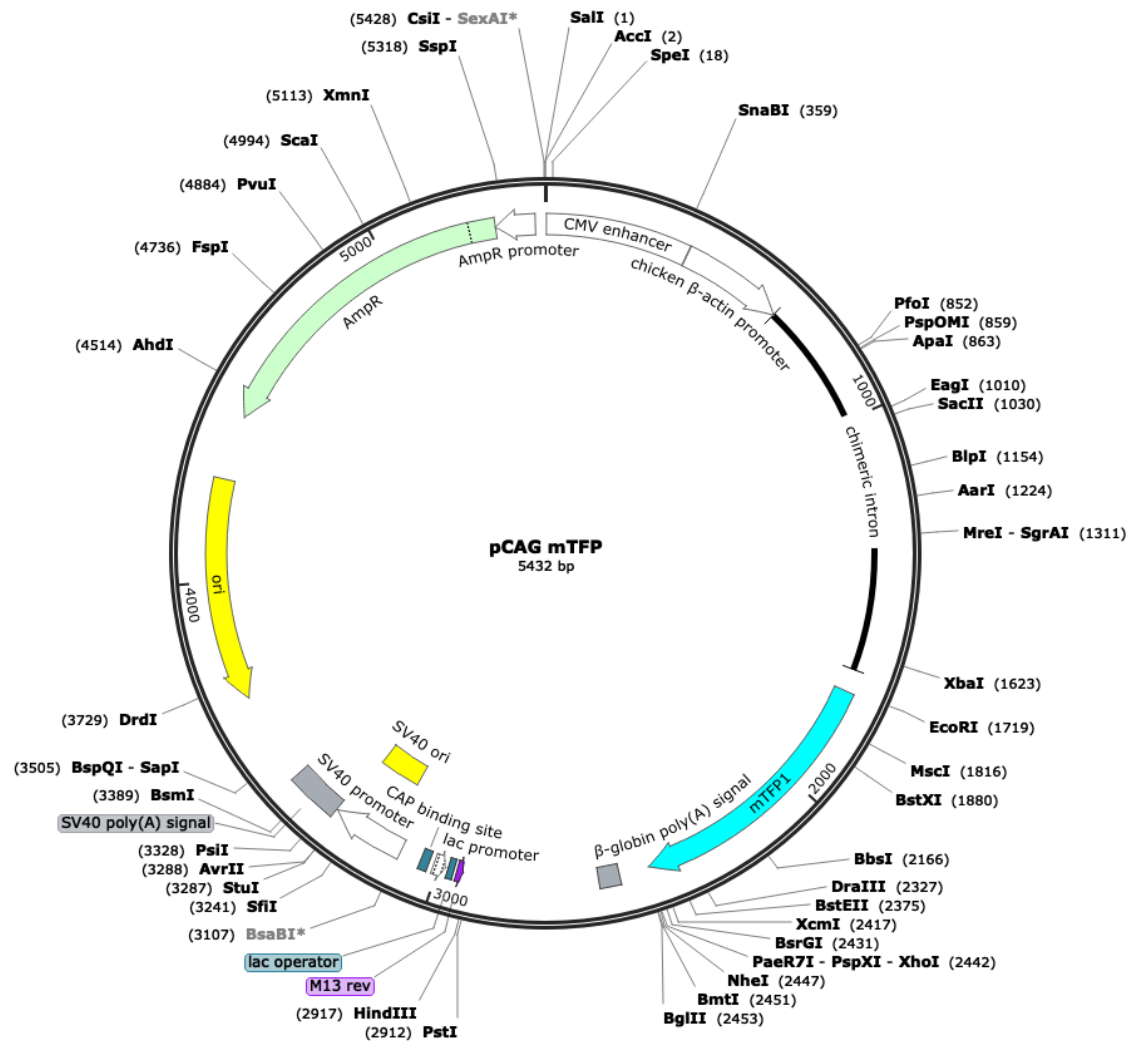
Moreover that we also come with a novel entry mechanism for CHPV. This information can be helpful for researchers associated with CHPV pathogenesis. Similarly, we have also identified many cellular pathways and associated proteins with respect to SARS-CoV-2 pathogenesis, which might help SARS-CoV-2 pathogenesis research. We have given a mechanism of how a complement system regulates the coagulation process during SARS-CoV-2 infection. It was a mystery that SARS-CoV-2 patients show coagulation and heart attack even after a long time of infection. Recently it was also reported that sometimes the vaccine also causes thrombosis. We believe our finding helps the researchers and drug-making companies to target virus pathogenesis by using host factors information for respective viruses.

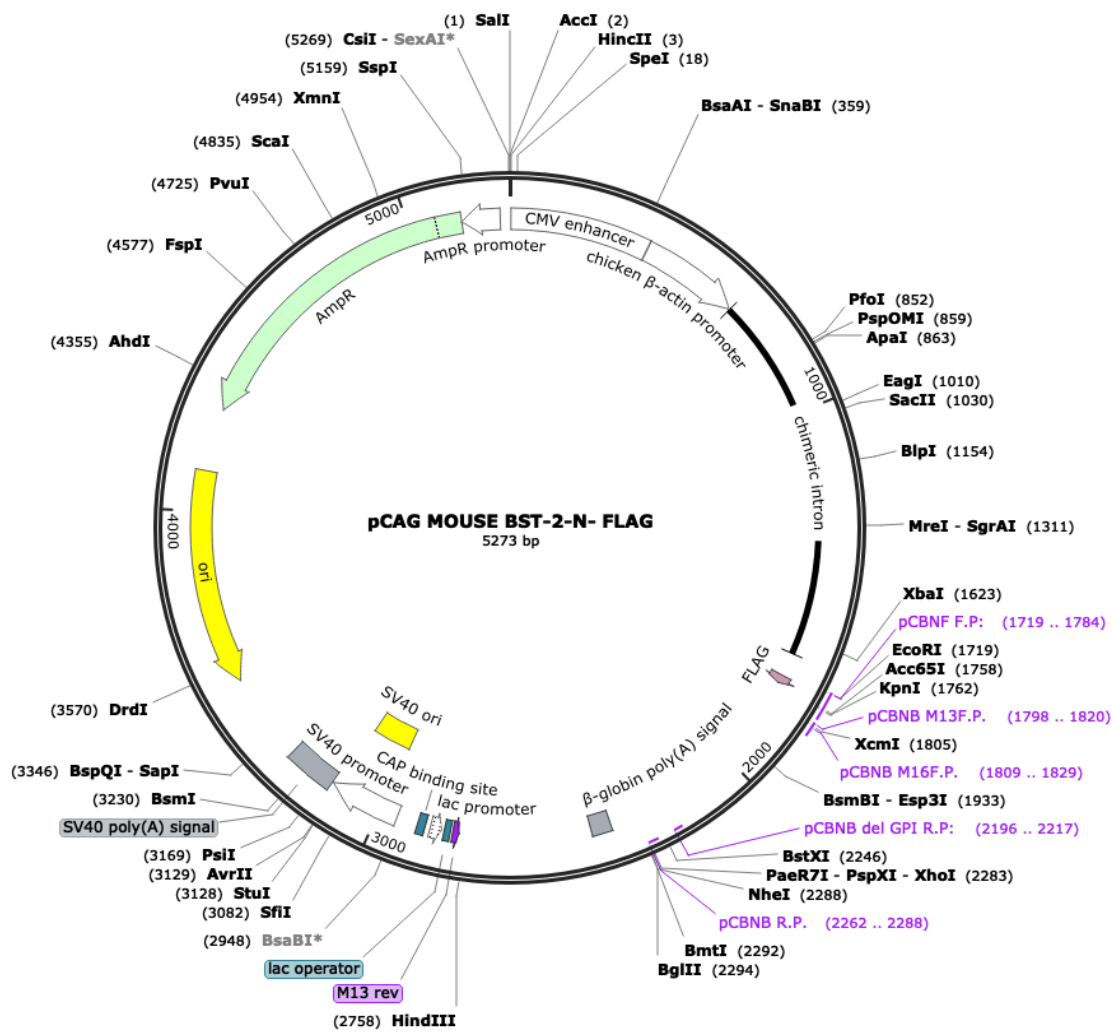
APPENDIX

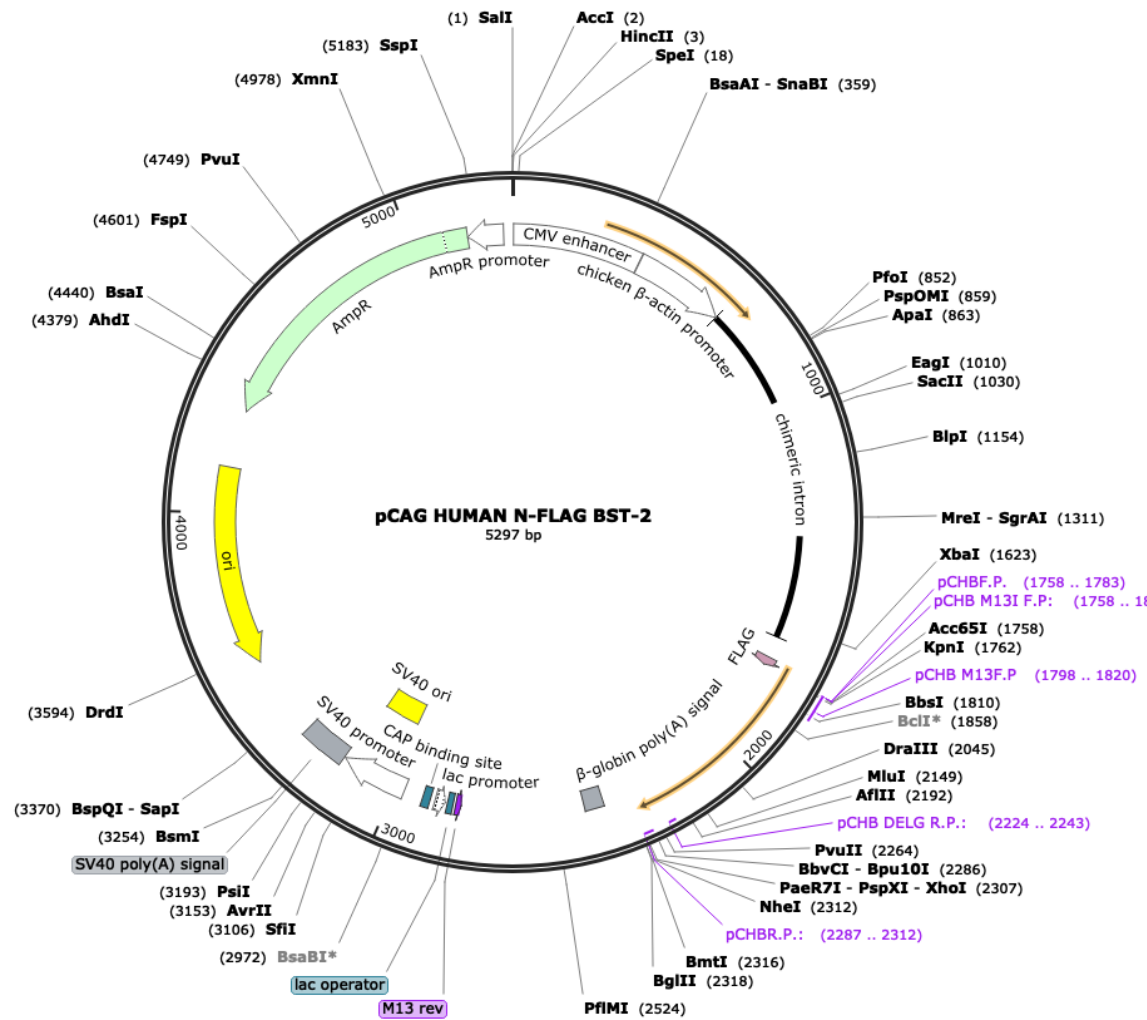
APPENDIX-A

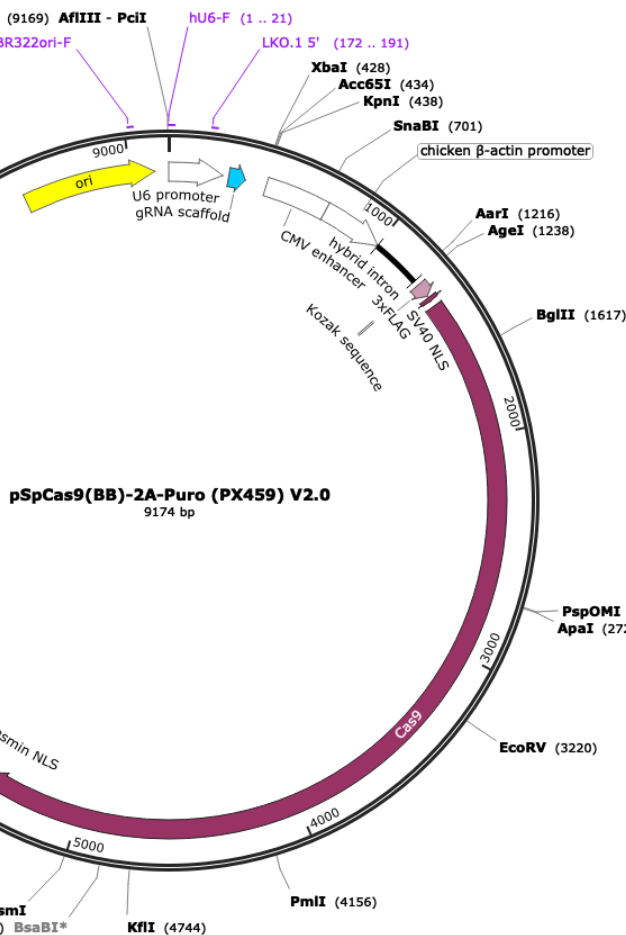
Vector Maps

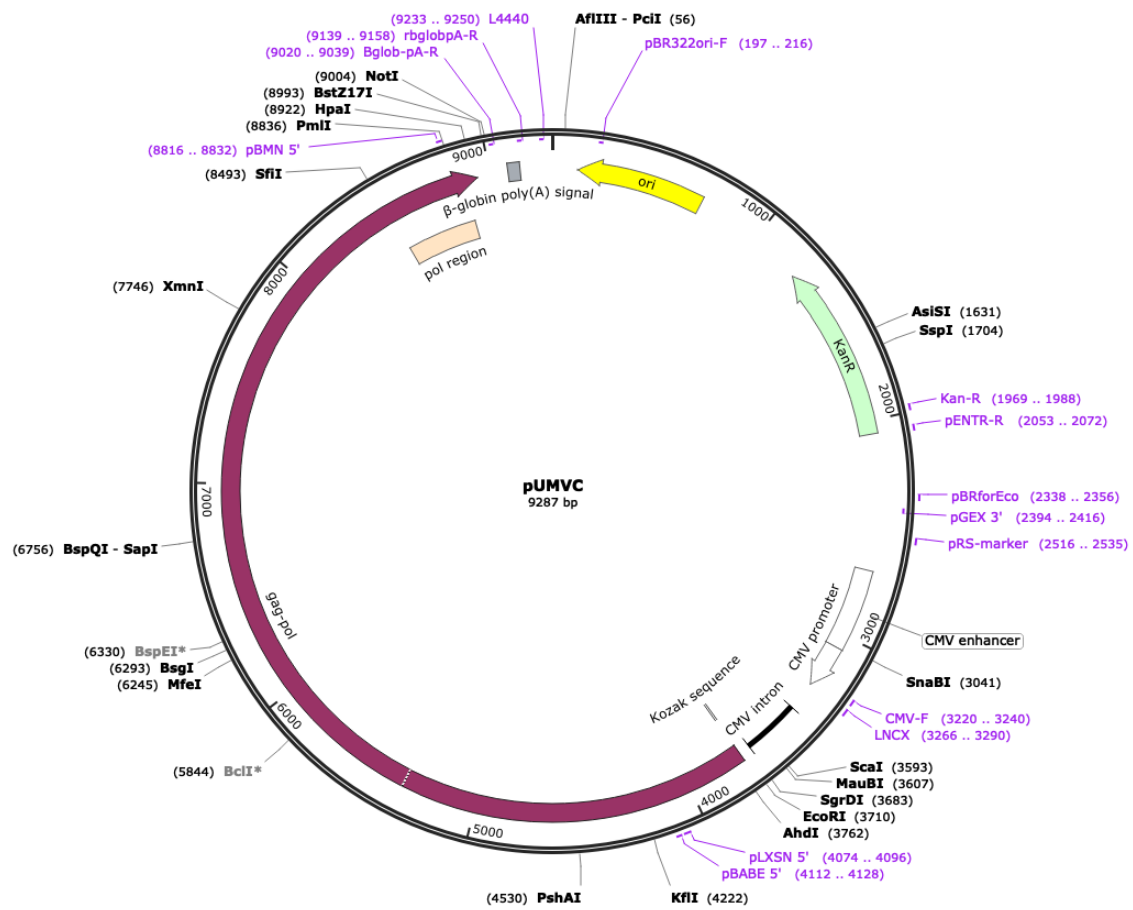
Created with SnapGene®

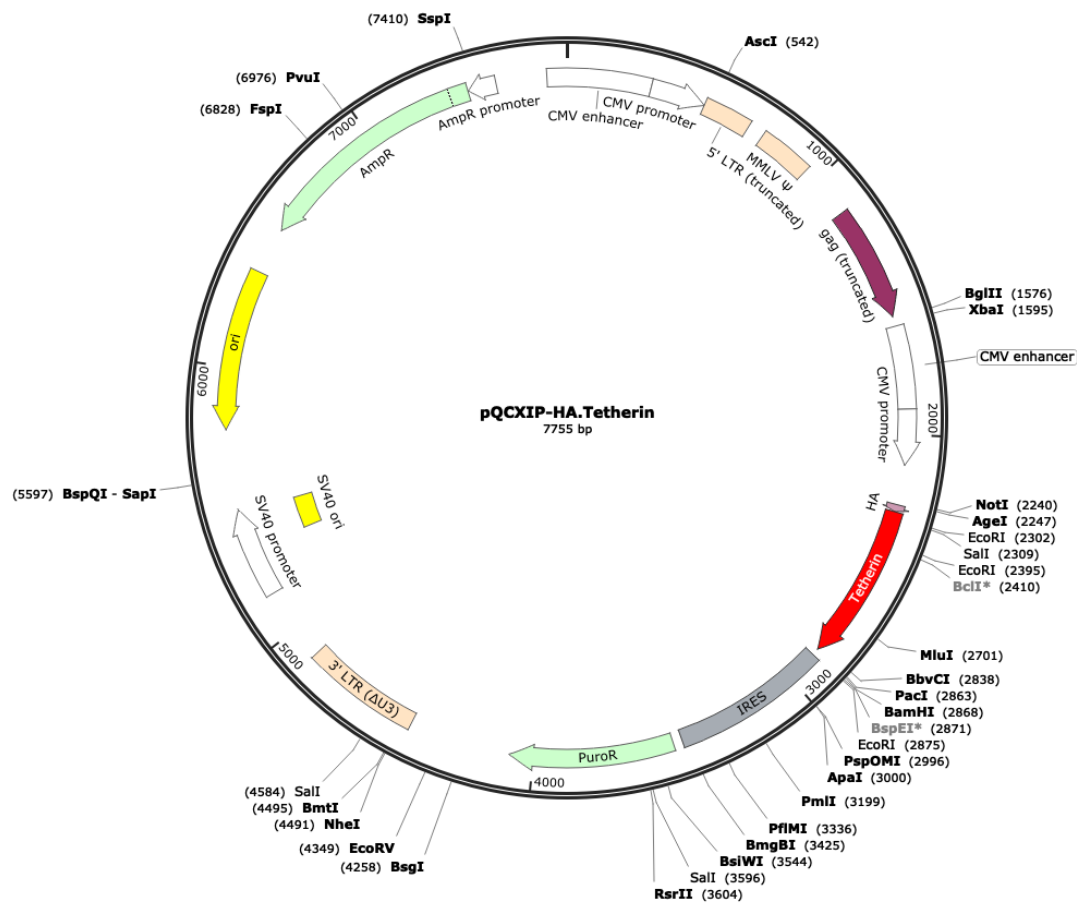


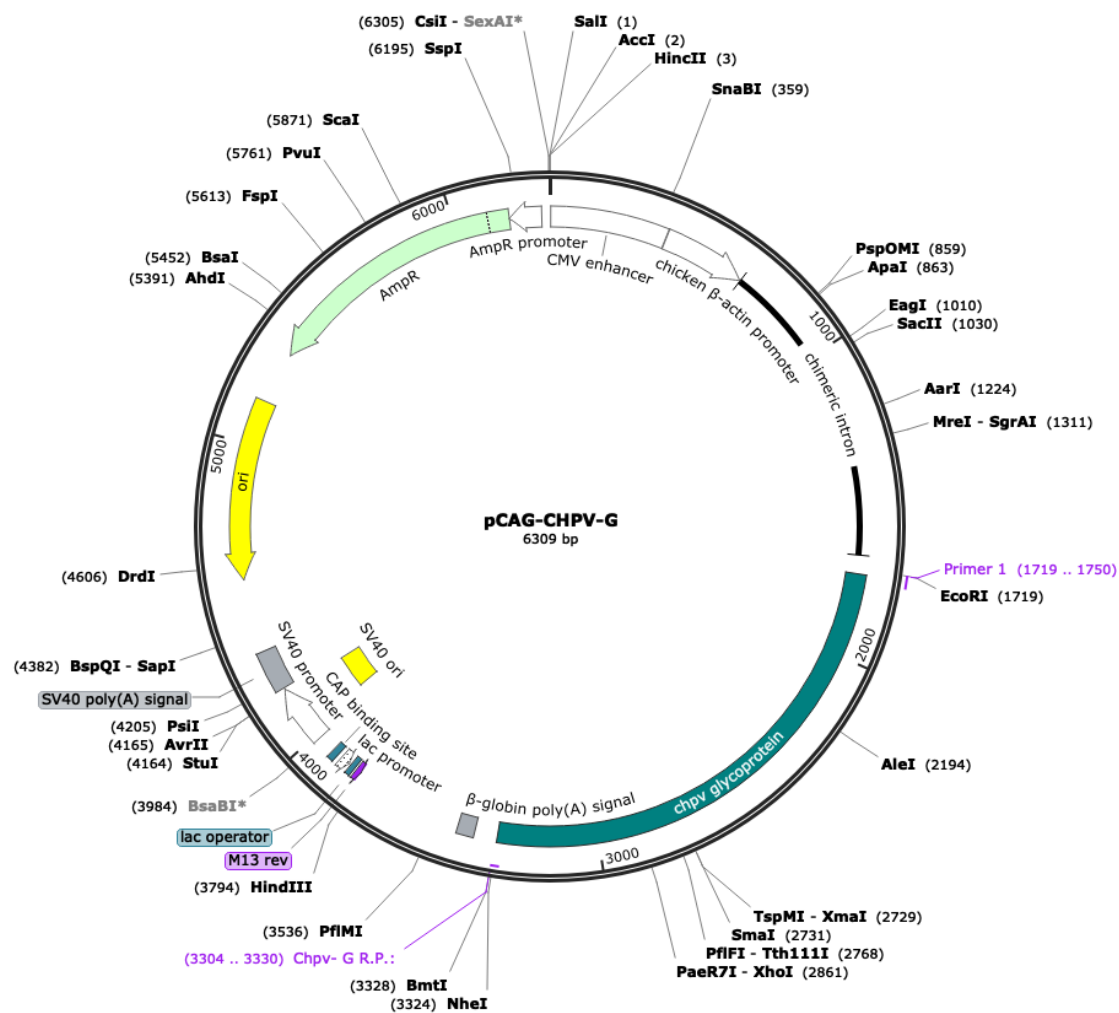












APPENDIX-B

Primers used for PCR and RT-qPCR

Primer	Sequence 5'to 3'
RDT1	ATATATGAATTCaccgccATGGACTACAAAGACGATGACGACAAGGGTAC CATGGCGCCCTCTTTCTATCAC
RDT2	ATATATCTCGAGTCAAAAGAGCAGGAACAGTGA
RDT3	ATATATGGTACCATGGATGAGATGGGGGGGAAG
RDT4	ATATATGGTACCATGGGGGGGAAGCAAGGATGG
RDT5	ATATATCTCGAGTCAGTTCACCTGCACTGTGCTAGA
RDT6	ATATATGGTACCATGGCATCTACTTCGTATGA
RDT7	ATATATGGTACCATGGAAGACGGGGATAAGCGC
RDT8	ATATATGGTACCATGGCATCTACTTCGTATGACTATTGCAGAGTGCCCAT CGAAG
RDT9	ATATATCTCGAGTCAGGAGTCCTGGGAGCTGGGGT
RDT10	ACACTGTGATGGCCCTAATG
RDT11	CGTCCTGAAGCTTATGGTTTAATG
RDT12	CAGAAGGGCTTTCAGGATGT
RDT13	TTTGTCTTGGGCCTTCTC
RDT14	ATATATGGTACCATGGCACCTATTTTGATGA
RDT15	ATATATCTCGAGTCACAGCAGCAGAGCGCTCAAG
RDT16	TGTCGCAATGTCACCTATCTC
RDT17	GCATCCAGGGAAGCCATTA
RDT18	CTCCTGGTCATAGTGCTTCTG
RDT19	TAGGTGACATTGCGCACTC
RDT20	GCCCGGTGTTGAAAGAAATG
RDT21	GTTTGGGCCTATCTCCATATCC
RDT22	ACAGCAGGAGGGTACAAATTC
RDT23	GGTAGCTGTGGATGGTCTAAAG

RDT24	GTGACGTGGACATCCGTAAA
RDT25	CAGGGCAGTAATCTCCTTCTG
RDT26	GCTCTCTTCCAACCTTCCTTC
RDT27	CGTACAGGTCTTTACGGATGTC
RDT28	ATATATGAATTCaccgccATGCACCACCACCACCACGGTACCATGGCG CCCTCTTTCTATCAC

APPENDIX-C

LC-MS data of CHPV pseudovirus interacting proteins

Accession	Description	Sum PEP Score
P58876	Histone H2B type 1-D OS=Homo sapiens GN=HIST1H2BD PE=1 SV=2	95.729
P60709	Actin, cytoplasmic 1 OS=Homo sapiens GN=ACTB PE=1 SV=1	94.784
P62805	Histone H4 OS=Homo sapiens GN=HIST1H4A PE=1 SV=2	92.576
P63261	Actin, cytoplasmic 2 OS=Homo sapiens GN=ACTG1 PE=1 SV=1	86.575
P57053	Histone H2B type F-S OS=Homo sapiens GN=H2BFS PE=1 SV=2	85.808
Q71DI3	Histone H3.2 OS=Homo sapiens GN=HIST2H3A PE=1 SV=3	84.294
P68431	Histone H3.1 OS=Homo sapiens GN=HIST1H3A PE=1 SV=2	74.204
P84243	Histone H3.3 OS=Homo sapiens GN=H3F3A PE=1 SV=2	59.571
P08238	Heat shock protein HSP 90-beta OS=Homo sapiens GN=HSP90AB1 PE=1 SV=4	48.041
P11142	Heat shock cognate 71 kDa protein OS=Homo sapiens GN=HSPA8 PE=1 SV=1	43.248
P62736	Actin, aortic smooth muscle OS=Homo sapiens GN=ACTA2 PE=1 SV=1	43.163
P06748	Nucleophosmin OS=Homo sapiens GN=NPM1 PE=1 SV=2	42.311
P62917	60S ribosomal protein L8 OS=Homo sapiens GN=RPL8 PE=1 SV=2	41.449
P39019	40S ribosomal protein S19 OS=Homo sapiens GN=RPS19 PE=1 SV=2	41.371
P04406	Glyceraldehyde-3-phosphate dehydrogenase OS=Homo sapiens GN=GAPDH PE=1 SV=3	39.245
P10412	Histone H1.4 OS=Homo sapiens GN=HIST1H1E PE=1 SV=2	37.143
P07437	Tubulin beta chain OS=Homo sapiens GN=TUBB PE=1 SV=2	36.124
P07900	Heat shock protein HSP 90-alpha OS=Homo sapiens GN=HSP90AA1 PE=1 SV=5	35.544
P62081	40S ribosomal protein S7 OS=Homo sapiens GN=RPS7 PE=1 SV=1	35.364

P83731	60S ribosomal protein L24 OS=Homo sapiens GN=RPL24 PE=1 SV=1	35.339
P68363	Tubulin alpha-1B chain OS=Homo sapiens GN=TUBA1B PE=1 SV=1	34.587
P04908	Histone H2A type 1-B/E OS=Homo sapiens GN=HIST1H2AB PE=1 SV=2	33.2
P26373	60S ribosomal protein L13 OS=Homo sapiens GN=RPL13 PE=1 SV=4	32.868
P68371	Tubulin beta-4B chain OS=Homo sapiens GN=TUBB4B PE=1 SV=1	31.666
P16403	Histone H1.2 OS=Homo sapiens GN=HIST1H1C PE=1 SV=2	31.514
Q7L7L0	Histone H2A type 3 OS=Homo sapiens GN=HIST3H2A PE=1 SV=3	31.455
P0C0S8	Histone H2A type 1 OS=Homo sapiens GN=HIST1H2AG PE=1 SV=2	30.441
P04264	Keratin, type II cytoskeletal 1 OS=Homo sapiens GN=KRT1 PE=1 SV=6	29.746
P25398	40S ribosomal protein S12 OS=Homo sapiens GN=RPS12 PE=1 SV=3	29.638
P23528	Cofilin-1 OS=Homo sapiens GN=CFL1 PE=1 SV=3	29.077
P62249	40S ribosomal protein S16 OS=Homo sapiens GN=RPS16 PE=1 SV=2	28.455
P08708	40S ribosomal protein S17 OS=Homo sapiens GN=RPS17 PE=1 SV=2	27.982
Q00839	Heterogeneous nuclear ribonucleoprotein U OS=Homo sapiens GN=HNRNPU PE=1 SV=6	27.377
P62269	40S ribosomal protein S18 OS=Homo sapiens GN=RPS18 PE=1 SV=3	26.916
P62280	40S ribosomal protein S11 OS=Homo sapiens GN=RPS11 PE=1 SV=3	26.369
P62913	60S ribosomal protein L11 OS=Homo sapiens GN=RPL11 PE=1 SV=2	25.792
P36578	60S ribosomal protein L4 OS=Homo sapiens GN=RPL4 PE=1 SV=5	24.262
P52272	Heterogeneous nuclear ribonucleoprotein M OS=Homo sapiens GN=HNRNPM PE=1 SV=3	23.607

P62851	40S ribosomal protein S25 OS=Homo sapiens GN=RPS25 PE=1 SV=1	23.26
P60866	40S ribosomal protein S20 OS=Homo sapiens GN=RPS20 PE=1 SV=1	22.99
P14618	Pyruvate kinase PKM OS=Homo sapiens GN=PKM PE=1 SV=4	22.241
P16104	Histone H2AX OS=Homo sapiens GN=H2AFX PE=1 SV=2	21.792
Q9NZI8	Insulin-like growth factor 2 mRNA-binding protein 1 OS=Homo sapiens GN=IGF2BP1 PE=1 SV=2	19.081
P62263	40S ribosomal protein S14 OS=Homo sapiens GN=RPS14 PE=1 SV=3	19.001
P62854	40S ribosomal protein S26 OS=Homo sapiens GN=RPS26 PE=1 SV=3	18.538
P62277	40S ribosomal protein S13 OS=Homo sapiens GN=RPS13 PE=1 SV=2	18.537
P16401	Histone H1.5 OS=Homo sapiens GN=HIST1H1B PE=1 SV=3	17.734
P05386	60S acidic ribosomal protein P1 OS=Homo sapiens GN=RPLP1 PE=1 SV=1	17.21
P62241	40S ribosomal protein S8 OS=Homo sapiens GN=RPS8 PE=1 SV=2	17.095
P61247	40S ribosomal protein S3a OS=Homo sapiens GN=RPS3A PE=1 SV=2	16.737
P62753	40S ribosomal protein S6 OS=Homo sapiens GN=RPS6 PE=1 SV=1	16.489
P19338	Nucleolin OS=Homo sapiens GN=NCL PE=1 SV=3	15.398
P68104	Elongation factor 1-alpha 1 OS=Homo sapiens GN=EEF1A1 PE=1 SV=1	15.2
P02768	Serum albumin OS=Homo sapiens GN=ALB PE=1 SV=2	15.188
P30050	60S ribosomal protein L12 OS=Homo sapiens GN=RPL12 PE=1 SV=1	15.115
Q02878	60S ribosomal protein L6 OS=Homo sapiens GN=RPL6 PE=1 SV=3	14.759
P55072	Transitional endoplasmic reticulum ATPase OS=Homo sapiens GN=VCP PE=1 SV=4	14.729
P62701	40S ribosomal protein S4, X isoform OS=Homo sapiens GN=RPS4X PE=1 SV=2	13.986

P62424	60S ribosomal protein L7a OS=Homo sapiens GN=RPL7A PE=1 SV=2	13.738
P62826	GTP-binding nuclear protein Ran OS=Homo sapiens GN=RAN PE=1 SV=3	13.577
P62829	60S ribosomal protein L23 OS=Homo sapiens GN=RPL23 PE=1 SV=1	13.493
P46782	40S ribosomal protein S5 OS=Homo sapiens GN=RPS5 PE=1 SV=4	13.312
Q86V81	THO complex subunit 4 OS=Homo sapiens GN=ALYREF PE=1 SV=3	13.261
P18621	60S ribosomal protein L17 OS=Homo sapiens GN=RPL17 PE=1 SV=3	13.207
O00571	ATP-dependent RNA helicase DDX3X OS=Homo sapiens GN=DDX3X PE=1 SV=3	12.883
P62857	40S ribosomal protein S28 OS=Homo sapiens GN=RPS28 PE=1 SV=1	12.578
P62316	Small nuclear ribonucleoprotein Sm D2 OS=Homo sapiens GN=SNRPD2 PE=1 SV=1	12.469
P62318	Small nuclear ribonucleoprotein Sm D3 OS=Homo sapiens GN=SNRPD3 PE=1 SV=1	12.176
P69905	Hemoglobin subunit alpha OS=Homo sapiens GN=HBA1 PE=1 SV=2	12.118
P32969	60S ribosomal protein L9 OS=Homo sapiens GN=RPL9 PE=1 SV=1	12.102
P08670	Vimentin OS=Homo sapiens GN=VIM PE=1 SV=4	12.1
P62899	60S ribosomal protein L31 OS=Homo sapiens GN=RPL31 PE=1 SV=1	11.634
P26641	Elongation factor 1-gamma OS=Homo sapiens GN=EEF1G PE=1 SV=3	10.753
P61254	60S ribosomal protein L26 OS=Homo sapiens GN=RPL26 PE=1 SV=1	10.681
P52926	High mobility group protein HMGI-C OS=Homo sapiens GN=HMGA2 PE=1 SV=1	10.613
P35527	Keratin, type I cytoskeletal 9 OS=Homo sapiens GN=KRT9 PE=1 SV=3	10.46

P46779	60S ribosomal protein L28 OS=Homo sapiens GN=RPL28 PE=1 SV=3	10.315
P63220	40S ribosomal protein S21 OS=Homo sapiens GN=RPS21 PE=1 SV=1	10.246
P09651	Heterogeneous nuclear ribonucleoprotein A1 OS=Homo sapiens GN=HNRNPA1 PE=1 SV=5	10.198
P06733	Alpha-enolase OS=Homo sapiens GN=ENO1 PE=1 SV=2	10.015
P37108	Signal recognition particle 14 kDa protein OS=Homo sapiens GN=SRP14 PE=1 SV=2	9.52
P16989	Y-box-binding protein 3 OS=Homo sapiens GN=YBX3 PE=1 SV=4	9.376
P84098	60S ribosomal protein L19 OS=Homo sapiens GN=RPL19 PE=1 SV=1	9.365
P13645	Keratin, type I cytoskeletal 10 OS=Homo sapiens GN=KRT10 PE=1 SV=6	9.337
P51991	Heterogeneous nuclear ribonucleoprotein A3 OS=Homo sapiens GN=HNRNPA3 PE=1 SV=2	8.954
P53999	Activated RNA polymerase II transcriptional coactivator p15 OS=Homo sapiens GN=SUB1 PE=1 SV=3	8.948
P49207	60S ribosomal protein L34 OS=Homo sapiens GN=RPL34 PE=1 SV=3	8.867
Q9H7E9	UPF0488 protein C8orf33 OS=Homo sapiens GN=C8orf33 PE=1 SV=1	8.727
P06454	Prothymosin alpha OS=Homo sapiens GN=PTMA PE=1 SV=2	8.556
Q9Y3U8	60S ribosomal protein L36 OS=Homo sapiens GN=RPL36 PE=1 SV=3	8.517
P13639	Elongation factor 2 OS=Homo sapiens GN=EEF2 PE=1 SV=4	8.456
P23526	Adenosylhomocysteinase OS=Homo sapiens GN=AHCY PE=1 SV=4	8.399
P46776	60S ribosomal protein L27a OS=Homo sapiens GN=RPL27A PE=1 SV=2	8.352
Q9UQ35	Serine/arginine repetitive matrix protein 2 OS=Homo sapiens GN=SRRM2 PE=1 SV=2	8.077
P62266	40S ribosomal protein S23 OS=Homo sapiens GN=RPS23 PE=1 SV=3	7.884

P62861	40S ribosomal protein S30 OS=Homo sapiens GN=FAU PE=1 SV=1	7.817
Q9Y281	Cofilin-2 OS=Homo sapiens GN=CFL2 PE=1 SV=1	7.629
P62750	60S ribosomal protein L23a OS=Homo sapiens GN=RPL23A PE=1 SV=1	7.579
P46778	60S ribosomal protein L21 OS=Homo sapiens GN=RPL21 PE=1 SV=2	7.29
P62314	Small nuclear ribonucleoprotein Sm D1 OS=Homo sapiens GN=SNRPD1 PE=1 SV=1	7.259
P05387	60S acidic ribosomal protein P2 OS=Homo sapiens GN=RPLP2 PE=1 SV=1	7.013
P67809	Nuclease-sensitive element-binding protein 1 OS=Homo sapiens GN=YBX1 PE=1 SV=3	6.874
P62987	Ubiquitin-60S ribosomal protein L40 OS=Homo sapiens GN=UBA52 PE=1 SV=2	6.496
Q9BWJ5	Splicing factor 3B subunit 5 OS=Homo sapiens GN=SF3B5 PE=1 SV=1	6.478
Chandipura_G_protein	Chandipura_G_protein	6.474
P17096	High mobility group protein HMG-I/HMG-Y OS=Homo sapiens GN=HMGA1 PE=1 SV=3	6.355
P62937	Peptidyl-prolyl cis-trans isomerase A OS=Homo sapiens GN=PPIA PE=1 SV=2	6.298
P07910	Heterogeneous nuclear ribonucleoproteins C1/C2 OS=Homo sapiens GN=HNRNPC PE=1 SV=4	6.261
Q9Y230	RuvB-like 2 OS=Homo sapiens GN=RUVBL2 PE=1 SV=3	6.237
P62906	60S ribosomal protein L10a OS=Homo sapiens GN=RPL10A PE=1 SV=2	6.116
P42766	60S ribosomal protein L35 OS=Homo sapiens GN=RPL35 PE=1 SV=2	6.102
P07355	Annexin A2 OS=Homo sapiens GN=ANXA2 PE=1 SV=2	6.028
P67936	Tropomyosin alpha-4 chain OS=Homo sapiens GN=TPM4 PE=1 SV=3	5.968
P09496	Clathrin light chain A OS=Homo sapiens GN=CLTA PE=1 SV=1	5.783

Q92598	Heat shock protein 105 kDa OS=Homo sapiens GN=HSPH1 PE=1 SV=1	5.783
P18669	Phosphoglycerate mutase 1 OS=Homo sapiens GN=PGAM1 PE=1 SV=2	5.77
P62910	60S ribosomal protein L32 OS=Homo sapiens GN=RPL32 PE=1 SV=2	5.576
P0C0S5	Histone H2A.Z OS=Homo sapiens GN=H2AFZ PE=1 SV=2	5.562
P40227	T-complex protein 1 subunit zeta OS=Homo sapiens GN=CCT6A PE=1 SV=3	5.481
P52701	DNA mismatch repair protein Msh6 OS=Homo sapiens GN=MSH6 PE=1 SV=2	5.293
P81605	Dermcidin OS=Homo sapiens GN=DCD PE=1 SV=2	5.29
P10809	60 kDa heat shock protein, mitochondrial OS=Homo sapiens GN=HSPD1 PE=1 SV=2	5.22
Q9BSD7	Cancer-related nucleoside-triphosphatase OS=Homo sapiens GN=NTPCR PE=1 SV=1	5.107
P14625	Endoplasmin OS=Homo sapiens GN=HSP90B1 PE=1 SV=1	5.105
P61353	60S ribosomal protein L27 OS=Homo sapiens GN=RPL27 PE=1 SV=2	5.045
P31943	Heterogeneous nuclear ribonucleoprotein H OS=Homo sapiens GN=HNRNPH1 PE=1 SV=4	4.899
Q8TF09	Dynein light chain roadblock-type 2 OS=Homo sapiens GN=DYNLRB2 PE=1 SV=1	4.797
P39023	60S ribosomal protein L3 OS=Homo sapiens GN=RPL3 PE=1 SV=2	4.734
P50914	60S ribosomal protein L14 OS=Homo sapiens GN=RPL14 PE=1 SV=4	4.728
P83881	60S ribosomal protein L36a OS=Homo sapiens GN=RPL36A PE=1 SV=2	4.703
Q9Y3I0	tRNA-splicing ligase RtcB homolog OS=Homo sapiens GN=RTCB PE=1 SV=1	4.649
O75934	Pre-mRNA-splicing factor SPF27 OS=Homo sapiens GN=BCAS2 PE=1 SV=1	4.623
P01023	Alpha-2-macroglobulin OS=Homo sapiens GN=A2M PE=1 SV=3	4.542
P09429	High mobility group protein B1 OS=Homo sapiens GN=HMGB1 PE=1 SV=3	4.462

O75531	Barrier-to-autointegration factor OS=Homo sapiens GN=BANF1 PE=1 SV=1	4.448
P62841	40S ribosomal protein S15 OS=Homo sapiens GN=RPS15 PE=1 SV=2	4.414
Q13263	Transcription intermediary factor 1-beta OS=Homo sapiens GN=TRIM28 PE=1 SV=5	4.388
P78371	T-complex protein 1 subunit beta OS=Homo sapiens GN=CCT2 PE=1 SV=4	4.353
P07477	Trypsin-1 OS=Homo sapiens GN=PRSS1 PE=1 SV=1	4.315
P18124	60S ribosomal protein L7 OS=Homo sapiens GN=RPL7 PE=1 SV=1	4.188
Q9Y3Y2	Chromatin target of PRMT1 protein OS=Homo sapiens GN=CHTOP PE=1 SV=2	4.136
Q9UKV3	Apoptotic chromatin condensation inducer in the nucleus OS=Homo sapiens GN=ACIN1 PE=1 SV=2	4.118
P23246	Splicing factor, proline- and glutamine-rich OS=Homo sapiens GN=SFPQ PE=1 SV=2	4.064
P14678	Small nuclear ribonucleoprotein-associated proteins B and B' OS=Homo sapiens GN=SNRPB PE=1 SV=2	4.007
O14818	Proteasome subunit alpha type-7 OS=Homo sapiens GN=PSMA7 PE=1 SV=1	3.733
P55209	Nucleosome assembly protein 1-like 1 OS=Homo sapiens GN=NAP1L1 PE=1 SV=1	3.718
Q92785	Zinc finger protein ubi-d4 OS=Homo sapiens GN=DPF2 PE=1 SV=2	3.425
P09234	U1 small nuclear ribonucleoprotein C OS=Homo sapiens GN=SNRPC PE=1 SV=1	3.423
Q01105	Protein SET OS=Homo sapiens GN=SET PE=1 SV=3	3.204
Q9NPA8	Transcription and mRNA export factor ENY2 OS=Homo sapiens GN=ENY2 PE=1 SV=1	3.176
Q9UMS4	Pre-mRNA-processing factor 19 OS=Homo sapiens GN=PRPF19 PE=1 SV=1	3.14
Q14683	Structural maintenance of chromosomes protein 1A OS=Homo sapiens GN=SMC1A PE=1 SV=2	3.124
Q9H444	Charged multivesicular body protein 4b OS=Homo sapiens GN=CHMP4B PE=1 SV=1	3.106

P60900	Proteasome subunit alpha type-6 OS=Homo sapiens GN=PSMA6 PE=1 SV=1	3.061
Q12904	Aminoacyl tRNA synthase complex-interacting multifunctional protein 1 OS=Homo sapiens GN=AIMP1 PE=1 SV=2	3.041
Q01658	Protein Dr1 OS=Homo sapiens GN=DR1 PE=1 SV=1	3.01
Q15019	Septin-2 OS=Homo sapiens GN=SEPT2 PE=1 SV=1	2.954
P63173	60S ribosomal protein L38 OS=Homo sapiens GN=RPL38 PE=1 SV=2	2.935
Q86X55	Histone-arginine methyltransferase CARM1 OS=Homo sapiens GN=CARM1 PE=1 SV=3	2.927
P04075	Fructose-bisphosphate aldolase A OS=Homo sapiens GN=ALDOA PE=1 SV=2	2.865
P62304	Small nuclear ribonucleoprotein E OS=Homo sapiens GN=SNRPE PE=1 SV=1	2.841
P35268	60S ribosomal protein L22 OS=Homo sapiens GN=RPL22 PE=1 SV=2	2.745
P60981	Dextrin OS=Homo sapiens GN=DSTN PE=1 SV=3	2.732
Q6NZI2	Polymerase I and transcript release factor OS=Homo sapiens GN=PTRF PE=1 SV=1	2.72
P20290	Transcription factor BTF3 OS=Homo sapiens GN=BTF3 PE=1 SV=1	2.713
P18077	60S ribosomal protein L35a OS=Homo sapiens GN=RPL35A PE=1 SV=2	2.696
P41091	Eukaryotic translation initiation factor 2 subunit 3 OS=Homo sapiens GN=EIF2S3 PE=1 SV=3	2.673
P02042	Hemoglobin subunit delta OS=Homo sapiens GN=HBD PE=1 SV=2	2.656
P84090	Enhancer of rudimentary homolog OS=Homo sapiens GN=ERH PE=1 SV=1	2.637
E9PAV3	Nascent polypeptide-associated complex subunit alpha, muscle-specific form OS=Homo sapiens GN=NACA PE=1 SV=1	2.615
P63313	Thymosin beta-10 OS=Homo sapiens GN=TMSB10 PE=1 SV=2	2.554
P52655	Transcription initiation factor IIA subunit 1 OS=Homo sapiens GN=GTF2A1 PE=1 SV=1	2.544
P61513	60S ribosomal protein L37a OS=Homo sapiens GN=RPL37A PE=1 SV=2	2.521

Q06830	Peroxiredoxin-1 OS=Homo sapiens GN=PRDX1 PE=1 SV=1	2.438
Q09028	Histone-binding protein RBBP4 OS=Homo sapiens GN=RBBP4 PE=1 SV=3	2.426
Q9UKM9	RNA-binding protein Raly OS=Homo sapiens GN=RALY PE=1 SV=1	2.381
Q96K17	Transcription factor BTF3 homolog 4 OS=Homo sapiens GN=BTF3L4 PE=1 SV=1	2.345
O15126	Secretory carrier-associated membrane protein 1 OS=Homo sapiens GN=SCAMP1 PE=1 SV=2	2.342
Q9H299	SH3 domain-binding glutamic acid-rich-like protein 3 OS=Homo sapiens GN=SH3BGRL3 PE=1 SV=1	2.331
Q15008	26S proteasome non-ATPase regulatory subunit 6 OS=Homo sapiens GN=PSMD6 PE=1 SV=1	2.304
P27635	60S ribosomal protein L10 OS=Homo sapiens GN=RPL10 PE=1 SV=4	2.296
Q99733	Nucleosome assembly protein 1-like 4 OS=Homo sapiens GN=NAP1L4 PE=1 SV=1	2.266
Q92804	TATA-binding protein-associated factor 2N OS=Homo sapiens GN=TAF15 PE=1 SV=1	2.098
O43390	Heterogeneous nuclear ribonucleoprotein R OS=Homo sapiens GN=HNRNPR PE=1 SV=1	1.972
P62888	60S ribosomal protein L30 OS=Homo sapiens GN=RPL30 PE=1 SV=2	1.954
Q13409	Cytoplasmic dynein 1 intermediate chain 2 OS=Homo sapiens GN=DYNC112 PE=1 SV=3	1.951
P07737	Profilin-1 OS=Homo sapiens GN=PFN1 PE=1 SV=2	1.931
P15880	40S ribosomal protein S2 OS=Homo sapiens GN=RPS2 PE=1 SV=2	1.927
P47914	60S ribosomal protein L29 OS=Homo sapiens GN=RPL29 PE=1 SV=2	1.923
P22626	Heterogeneous nuclear ribonucleoproteins A2/B1 OS=Homo sapiens GN=HNRNPA2B1 PE=1 SV=2	1.917
Q96I24	Far upstream element-binding protein 3 OS=Homo sapiens GN=FUBP3 PE=1 SV=2	1.878
P62847	40S ribosomal protein S24 OS=Homo sapiens GN=RPS24 PE=1 SV=1	1.854

Q5SSJ5	Heterochromatin protein 1-binding protein 3 OS=Homo sapiens GN=HP1BP3 PE=1 SV=1	1.852
P38919	Eukaryotic initiation factor 4A-III OS=Homo sapiens GN=EIF4A3 PE=1 SV=4	1.839
P11387	DNA topoisomerase 1 OS=Homo sapiens GN=TOP1 PE=1 SV=2	1.812
Q9NP79	Vacuolar protein sorting-associated protein VTA1 homolog OS=Homo sapiens GN=VTA1 PE=1 SV=1	1.788
P01008	Antithrombin-III OS=Homo sapiens GN=SERPINC1 PE=1 SV=1	1.786
P46777	60S ribosomal protein L5 OS=Homo sapiens GN=RPL5 PE=1 SV=3	1.785
P52597	Heterogeneous nuclear ribonucleoprotein F OS=Homo sapiens GN=HNRNPF PE=1 SV=3	1.785
P23396	40S ribosomal protein S3 OS=Homo sapiens GN=RPS3 PE=1 SV=2	1.743
P50991	T-complex protein 1 subunit delta OS=Homo sapiens GN=CCT4 PE=1 SV=4	1.724
P59768	Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-2 OS=Homo sapiens GN=GNG2 PE=1 SV=2	1.711
Q99832	T-complex protein 1 subunit eta OS=Homo sapiens GN=CCT7 PE=1 SV=2	1.708
P62244	40S ribosomal protein S15a OS=Homo sapiens GN=RPS15A PE=1 SV=2	1.706
Q7KZF4	Staphylococcal nuclease domain-containing protein 1 OS=Homo sapiens GN=SND1 PE=1 SV=1	1.658
Q15046	Lysine--tRNA ligase OS=Homo sapiens GN=KARS PE=1 SV=3	1.643
P09874	Poly [ADP-ribose] polymerase 1 OS=Homo sapiens GN=PARP1 PE=1 SV=4	1.629
Q969G5	Protein kinase C delta-binding protein OS=Homo sapiens GN=PRKCDBP PE=1 SV=3	1.535
P22087	rRNA 2'-O-methyltransferase fibrillarin OS=Homo sapiens GN=FBL PE=1 SV=2	1.518
Q15637	Splicing factor 1 OS=Homo sapiens GN=SF1 PE=1 SV=4	1.508
P62328	Thymosin beta-4 OS=Homo sapiens GN=TMSB4X PE=1 SV=2	1.462
O75533	Splicing factor 3B subunit 1 OS=Homo sapiens GN=SF3B1 PE=1 SV=3	1.431

Q04917	14-3-3 protein eta OS=Homo sapiens GN=YWHAH PE=1 SV=4	1.428
Q9Y5B9	FACT complex subunit SPT16 OS=Homo sapiens GN=SUPT16H PE=1 SV=1	1.41
Q92945	Far upstream element-binding protein 2 OS=Homo sapiens GN=KHSRP PE=1 SV=4	1.391
Q9UQN3	Charged multivesicular body protein 2b OS=Homo sapiens GN=CHMP2B PE=1 SV=1	1.385
Q09666	Neuroblast differentiation-associated protein AHNAK OS=Homo sapiens GN=AHNAK PE=1 SV=2	1.38
P08686	Steroid 21-hydroxylase OS=Homo sapiens GN=CYP21A2 PE=1 SV=1	1.376
P21333	Filamin-A OS=Homo sapiens GN=FLNA PE=1 SV=4	1.351
Q13242	Serine/arginine-rich splicing factor 9 OS=Homo sapiens GN=SRSF9 PE=1 SV=1	1.349
Q9BRL6	Serine/arginine-rich splicing factor 8 OS=Homo sapiens GN=SRSF8 PE=1 SV=1	1.333
Q15149	Plectin OS=Homo sapiens GN=PLEC PE=1 SV=3	1.33
P14649	Myosin light chain 6B OS=Homo sapiens GN=MYL6B PE=1 SV=1	1.322
O43633	Charged multivesicular body protein 2a OS=Homo sapiens GN=CHMP2A PE=1 SV=1	1.285
P02788	Lactotransferrin OS=Homo sapiens GN=LTF PE=1 SV=6	1.285
P00966	Argininosuccinate synthase OS=Homo sapiens GN=ASS1 PE=1 SV=2	1.231
Q14152	Eukaryotic translation initiation factor 3 subunit A OS=Homo sapiens GN=EIF3A PE=1 SV=1	1.214
P02810	Salivary acidic proline-rich phosphoprotein 1/2 OS=Homo sapiens GN=PRH1 PE=1 SV=2	1.194
Q15052	Rho guanine nucleotide exchange factor 6 OS=Homo sapiens GN=ARHGEF6 PE=1 SV=2	1.192
Q3KQU3	MAP7 domain-containing protein 1 OS=Homo sapiens GN=MAP7D1 PE=1 SV=1	1.173
P43243	Matrin-3 OS=Homo sapiens GN=MATR3 PE=1 SV=2	1.167
P19823	Inter-alpha-trypsin inhibitor heavy chain H2 OS=Homo sapiens GN=ITI2 PE=1 SV=2	1.163

Q14240	Eukaryotic initiation factor 4A-II OS=Homo sapiens GN=EIF4A2 PE=1 SV=2	1.149
P50502	Hsc70-interacting protein OS=Homo sapiens GN=ST13 PE=1 SV=2	1.148
P02545	Prelamin-A/C OS=Homo sapiens GN=LMNA PE=1 SV=1	1.133
P02452	Collagen alpha-1(I) chain OS=Homo sapiens GN=COL1A1 PE=1 SV=5	1.112
Q9H3N1	Thioredoxin-related transmembrane protein 1 OS=Homo sapiens GN=TMX1 PE=1 SV=1	1.092
P42677	40S ribosomal protein S27 OS=Homo sapiens GN=RPS27 PE=1 SV=3	1.083
P13667	Protein disulfide-isomerase A4 OS=Homo sapiens GN=PDIA4 PE=1 SV=2	1.075
P13647	Keratin, type II cytoskeletal 5 OS=Homo sapiens GN=KRT5 PE=1 SV=3	1.073
P08865	40S ribosomal protein SA OS=Homo sapiens GN=RPSA PE=1 SV=4	1.052
Q5T7W0	Zinc finger protein 618 OS=Homo sapiens GN=ZNF618 PE=1 SV=1	1.036
Q9NQT5	Exosome complex component RRP40 OS=Homo sapiens GN=EXOSC3 PE=1 SV=3	1.035
P61981	14-3-3 protein gamma OS=Homo sapiens GN=YWHAG PE=1 SV=2	1.033
P35579	Myosin-9 OS=Homo sapiens GN=MYH9 PE=1 SV=4	1
O60244	Mediator of RNA polymerase II transcription subunit 14 OS=Homo sapiens GN=MED14 PE=1 SV=2	0.996
Q6I9Y2	THO complex subunit 7 homolog OS=Homo sapiens GN=THOC7 PE=1 SV=3	0.965
P0CG43	Putative protein FAM157C OS=Homo sapiens GN=FAM157C PE=3 SV=1	0.963
P34932	Heat shock 70 kDa protein 4 OS=Homo sapiens GN=HSPA4 PE=1 SV=4	0.946
P63104	14-3-3 protein zeta/delta OS=Homo sapiens GN=YWHAZ PE=1 SV=1	0.946
O43823	A-kinase anchor protein 8 OS=Homo sapiens GN=AKAP8 PE=1 SV=1	0.944

P40429	60S ribosomal protein L13a OS=Homo sapiens GN=RPL13A PE=1 SV=2	0.943
Q08170	Serine/arginine-rich splicing factor 4 OS=Homo sapiens GN=SRSF4 PE=1 SV=2	0.943
Q02543	60S ribosomal protein L18a OS=Homo sapiens GN=RPL18A PE=1 SV=2	0.936
O95661	GTP-binding protein Di-Ras3 OS=Homo sapiens GN=DIRAS3 PE=1 SV=1	0.925
Q7LBR1	Charged multivesicular body protein 1b OS=Homo sapiens GN=CHMP1B PE=1 SV=1	0.916
Q9UDR5	Alpha-aminoadipic semialdehyde synthase, mitochondrial OS=Homo sapiens GN=AASS PE=1 SV=1	0.902
Q13162	Peroxiredoxin-4 OS=Homo sapiens GN=PRDX4 PE=1 SV=1	0.891
P15924	Desmoplakin OS=Homo sapiens GN=DSP PE=1 SV=3	0.848
O00763	Acetyl-CoA carboxylase 2 OS=Homo sapiens GN=ACACB PE=1 SV=3	0.845
P02458	Collagen alpha-1(II) chain OS=Homo sapiens GN=COL2A1 PE=1 SV=3	0.844
P04083	Annexin A1 OS=Homo sapiens GN=ANXA1 PE=1 SV=2	0.837
Q13395	Probable methyltransferase TARBP1 OS=Homo sapiens GN=TARBP1 PE=1 SV=1	0.762
Q7Z408	CUB and sushi domain-containing protein 2 OS=Homo sapiens GN=CSMD2 PE=1 SV=2	0.756
Q9NVA2	Septin-11 OS=Homo sapiens GN=SEPT11 PE=1 SV=3	0.748
Q99961	Endophilin-A2 OS=Homo sapiens GN=SH3GL1 PE=1 SV=1	0.744
Q5JSP0	FYVE, RhoGEF and PH domain-containing protein 3 OS=Homo sapiens GN=FGD3 PE=1 SV=1	0.742
Q9BXP5	Serrate RNA effector molecule homolog OS=Homo sapiens GN=SRRT PE=1 SV=1	0.733
Q9Y496	Kinesin-like protein KIF3A OS=Homo sapiens GN=KIF3A PE=1 SV=4	0.733
Q12851	Mitogen-activated protein kinase kinase kinase 2 OS=Homo sapiens GN=MAP4K2 PE=1 SV=2	0.713
Q8TE73	Dynein heavy chain 5, axonemal OS=Homo sapiens GN=DNAH5 PE=1 SV=3	0.71

Q16186	Proteasomal ubiquitin receptor ADRM1 OS=Homo sapiens GN=ADRM1 PE=1 SV=2	0.706
P01024	Complement C3 OS=Homo sapiens GN=C3 PE=1 SV=2	0.704
Q9NQX4	Unconventional myosin-Vc OS=Homo sapiens GN=MYO5C PE=1 SV=2	0.703
P35443	Thrombospondin-4 OS=Homo sapiens GN=THBS4 PE=1 SV=2	0.691
Q07666	KH domain-containing, RNA-binding, signal transduction- associated protein 1 OS=Homo sapiens GN=KHDRBS1 PE=1 SV=1	0.677
Q5VU43	Myomegalin OS=Homo sapiens GN=PDE4DIP PE=1 SV=1	0.675
O00628	Peroxisomal targeting signal 2 receptor OS=Homo sapiens GN=PEX7 PE=1 SV=1	0.673
O75569	Interferon-inducible double-stranded RNA-dependent protein kinase activator A OS=Homo sapiens GN=PRKRA PE=1 SV=1	0.671
P20929	Nebulin OS=Homo sapiens GN=NEB PE=1 SV=5	0.669
Q9Y265	RuvB-like 1 OS=Homo sapiens GN=RUVBL1 PE=1 SV=1	0.665
Q9UJ83	2-hydroxyacyl-CoA lyase 1 OS=Homo sapiens GN=HACL1 PE=1 SV=2	0.653
O75439	Mitochondrial-processing peptidase subunit beta OS=Homo sapiens GN=PMPCB PE=1 SV=2	0.647
P28066	Proteasome subunit alpha type-5 OS=Homo sapiens GN=PSMA5 PE=1 SV=3	0.646
Q9Y2Z9	Ubiquinone biosynthesis monooxygenase COQ6, mitochondrial OS=Homo sapiens GN=COQ6 PE=1 SV=2	0.641
P62191	26S protease regulatory subunit 4 OS=Homo sapiens GN=PSMC1 PE=1 SV=1	0.627
P31948	Stress-induced-phosphoprotein 1 OS=Homo sapiens GN=STIP1 PE=1 SV=1	0.626
Q99447	Ethanolamine-phosphate cytidyltransferase OS=Homo sapiens GN=PCYT2 PE=1 SV=1	0.612
Q15005	Signal peptidase complex subunit 2 OS=Homo sapiens GN=SPCS2 PE=1 SV=3	0.612
Q8WWQ2	Inactive heparanase-2 OS=Homo sapiens GN=HPSE2 PE=1 SV=3	0.611
O76011	Keratin, type I cuticular Ha4 OS=Homo sapiens GN=KRT34 PE=2 SV=2	0.608

Q9H0L4	Cleavage stimulation factor subunit 2 tau variant OS=Homo sapiens GN=CSTF2T PE=1 SV=1	0.608
Q96EL3	39S ribosomal protein L53, mitochondrial OS=Homo sapiens GN=MRPL53 PE=1 SV=1	0.603
Q9Y2D8	Afadin- and alpha-actinin-binding protein OS=Homo sapiens GN=SSX2IP PE=1 SV=3	0.602
O75369	Filamin-B OS=Homo sapiens GN=FLNB PE=1 SV=2	0.6
Q03701	CCAAT/enhancer-binding protein zeta OS=Homo sapiens GN=CEBPZ PE=1 SV=3	0.598
P46059	Solute carrier family 15 member 1 OS=Homo sapiens GN=SLC15A1 PE=2 SV=1	0.588
P46781	40S ribosomal protein S9 OS=Homo sapiens GN=RPS9 PE=1 SV=3	0.588
P57721	Poly(rC)-binding protein 3 OS=Homo sapiens GN=PCBP3 PE=2 SV=2	0.585
Q9NR09	Baculoviral IAP repeat-containing protein 6 OS=Homo sapiens GN=BIRC6 PE=1 SV=2	0.585
Q92526	T-complex protein 1 subunit zeta-2 OS=Homo sapiens GN=CCT6B PE=1 SV=5	0.579
P84103	Serine/arginine-rich splicing factor 3 OS=Homo sapiens GN=SRSF3 PE=1 SV=1	0.563
Q9Y2X9	Zinc finger protein 281 OS=Homo sapiens GN=ZNF281 PE=1 SV=1	0.559
Q9Y620	DNA repair and recombination protein RAD54B OS=Homo sapiens GN=RAD54B PE=1 SV=1	0.55
Q969F1	General transcription factor 3C polypeptide 6 OS=Homo sapiens GN=GTF3C6 PE=1 SV=1	0.544
O15230	Laminin subunit alpha-5 OS=Homo sapiens GN=LAMA5 PE=1 SV=8	0.541
Q9Y3R5	Protein dopey-2 OS=Homo sapiens GN=DOPEY2 PE=1 SV=5	0.541
P08069	Insulin-like growth factor 1 receptor OS=Homo sapiens GN=IGF1R PE=1 SV=1	0.536
Q13200	26S proteasome non-ATPase regulatory subunit 2 OS=Homo sapiens GN=PSMD2 PE=1 SV=3	0.528
Q9HD42	Charged multivesicular body protein 1a OS=Homo sapiens GN=CHMP1A PE=1 SV=1	0.526

O00471	Exocyst complex component 5 OS=Homo sapiens GN=EXOC5 PE=1 SV=1	0.517
Q13155	Aminoacyl tRNA synthase complex-interacting multifunctional protein 2 OS=Homo sapiens GN=AIMP2 PE=1 SV=2	0.499
Q00536	Cyclin-dependent kinase 16 OS=Homo sapiens GN=CDK16 PE=1 SV=1	0.492
Q9Y5K6	CD2-associated protein OS=Homo sapiens GN=CD2AP PE=1 SV=1	0.492
Q9UII4	E3 ISG15--protein ligase HERC5 OS=Homo sapiens GN=HERC5 PE=1 SV=2	0.485
Q15904	V-type proton ATPase subunit S1 OS=Homo sapiens GN=ATP6AP1 PE=1 SV=2	0.481
Q9Y2R5	28S ribosomal protein S17, mitochondrial OS=Homo sapiens GN=MRPS17 PE=1 SV=1	0.469
P52630	Signal transducer and activator of transcription 2 OS=Homo sapiens GN=STAT2 PE=1 SV=1	0.46
B2RTY4	Unconventional myosin-IXa OS=Homo sapiens GN=MYO9A PE=1 SV=2	0.455
A4D1F6	Leucine-rich repeat and death domain-containing protein 1 OS=Homo sapiens GN=LRRD1 PE=2 SV=2	0.448
O43150	Arf-GAP with SH3 domain, ANK repeat and PH domain- containing protein 2 OS=Homo sapiens GN=ASAP2 PE=1 SV=3	0.446
O95396	Adenylyltransferase and sulfurtransferase MOCS3 OS=Homo sapiens GN=MOCS3 PE=1 SV=1	0.437
Q8N9W8	Protein FAM71D OS=Homo sapiens GN=FAM71D PE=2 SV=2	0.43