INFLUENCE OF PROTEIN-PROTEIN INTERACTIONS (PPIs) ON THE OUTCOME OF VIRAL INFECTIONS

EDITED BY: Gorka Lasso Cabrera, Pablo Guardado-Calvo, Rohit K. Jangra, Eva Mittler and Mercè Llabrés

PUBLISHED IN: Frontiers in Microbiology and Frontiers in Virology

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88976-614-7
DOI 10.3389/978-2-88976-614-7

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

INFLUENCE OF PROTEIN-PROTEIN INTERACTIONS (PPIs) ON THE OUTCOME OF VIRAL INFECTIONS

Topic Editors:
Gorka Lasso Cabrera, Albert Einstein College of Medicine, United States
Pablo Guardado-Calvo, Institut Pasteur, France
Rohit K. Jangra, LSU Health Sciences Center-Shreveport, LA, United States
Eva Mittler, Albert Einstein College of Medicine, United States
Mercè Llabrés, University of the Balearic Islands, Spain

Citation: Cabrera, G. L., Guardado-Calvo, P., Jangra, R. K., Mittler, E., Llabrés, M., eds. (2022). Influence of Protein-Protein Interactions (PPIs) on the Outcome of Viral Infections. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-614-7

Table of Contents

frontiers | Frontiers in Microbiology

EDITORIAL
published: 27 June 2022
doi: 10.3389/fmicb.2022.943379

Editorial: Influence of Protein-Protein Interactions (PPIs) on the Outcome of Viral Infections

*Rohit K. Jangra[1], Mercè Llabrés[2], Pablo Guardado-Calvo[3], Eva Mittler[4] and Gorka Lasso[4]**

[1] Department of Microbiology and Immunology, Louisiana State University Health Sciences Center-Shreveport, Shreveport, LA, United States, [2] Department of Mathematics and Computer Science, University of the Balearic Islands, Palma, Spain, [3] Structural Virology Unit, Institut Pasteur, Université de Paris Cité, CNRS UMR 3569, Paris, France, [4] Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, NY, United States

Keywords: virus-host protein-protein interactions, PPI, amino acid variation, protein binding affinity, antibody recognition, virus-host PPI network, PPI database, viral infection

Editorial on the Research Topic

Influence of Protein-Protein Interactions (PPIs) on the Outcome of Viral Infections

In the last two decades, the (re)emergence of zoonotic viruses [e.g., Severe acute respiratory syndrome coronavirus (SARS-CoV), Middle-Eastern respiratory syndrome coronavirus (MERS-CoV), SARS-CoV-2, H1N1 and H5N1 influenza viruses, Ebola virus, and Zika virus] has resulted in devastating consequences from a health, economic, and social perspective. Changes in ecological and environmental factors, demographics, and socio-economic behavior have increased the risk of spillover events and of (re)emergence of zoonotic viruses (Ahmed et al., 2019; Gibb et al., 2020; Johnson et al., 2020; Carlson et al., 2022). It is therefore imperative to establish preventive and therapeutic measures, as well as epidemiological surveillance to mitigate the effect of future outbreaks (Abubakar et al., 2012; Watsa, 2020).

Viruses are genetic parasites that exploit the host's molecular machinery by employing specific virus-host protein-protein interactions (PPIs) that mediate critical steps in virus replication and immune evasion. Thus, PPIs are prime targets for the development of therapeutics and vaccines. However, their characterization is an urgent albeit challenging task that benefits drastically from the integration of computational and experimental approaches. This Research Topic brings together nine articles (including original research and review articles) that collectively leverage experimentally and computationally derived information to describe important biological processes mediated by virus-host PPIs.

Opening this Research Topic, four review articles present an overview of virus-host PPIs and their biological role by focusing on specific host proteins, virus family or providing a more holistic view. Chung and Song summarize interactions of proteins expressed by oncogenic gammaherpesviruses [e.g., Epstein-Barr virus (EBV), Kaposi's sarcoma-associated herpesvirus (KSHV), and murine gammaherpesvirus 68 (MHV-68)] with host poly (ADP-Ribose) polymerase 1 (PARP1), a nuclear enzyme that regulates diverse cellular pathways. PARP1's interaction with several viral proteins supports establishment of viral latency by down-regulating viral DNA replication, and reducing virus production to prevent reactivation. Simultaneously, these viruses also employ multiple mechanisms to down-regulate PARP1 expression to then promote their own replication. Fishburn et al. systematically review virus-host PPIs that mediate virus entry and replication of various flaviviruses, including dengue virus (DENV), Zika virus (ZIKV), West Nile virus (WNV), yellow fever virus (YFV), and Japanese encephalitis virus (JEV). The

OPEN ACCESS

Edited and reviewed by:
Anna Kramvis,
University of the Witwatersrand, South Africa

***Correspondence:**
Gorka Lasso
gorka.lasso@gmail.com

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *13 May 2022*
Accepted: *23 May 2022*
Published: *27 June 2022*

Citation:
Jangra RK, Llabrés M, Guardado-Calvo P, Mittler E and Lasso G (2022) Editorial: Influence of Protein-Protein Interactions (PPIs) on the Outcome of Viral Infections. Front. Microbiol. 13:943379. doi: 10.3389/fmicb.2022.943379

authors also summarize the role of virus-host and intra-host PPIs mediated by cellular proteins involved in autophagy, mitochondrial, and innate immune responses including the antagonism of host immunity. Experimentally determined virus-host PPIs are compiled in PPI databases. Saha et al. give a detailed summary of virus-host PPI repositories and illustrate how publicly available data can be leveraged to identify shared and unique strategies employed by four emerging viruses to co-opt cellular processes. Notably, Saha et al. and Gabriel Valiente emphasize the poor overlap between the different repositories, highlighting the need to use meta-databases that combine different primary resources and inspect annotated PPIs during dataset assembly.

Experimental characterization of the virus-host protein interactome is far from complete and only few viruses have been extensively studied. In order to narrow this knowledge gap, computational tools can provide a catalog of high-confidence PPI predictions to be tested experimentally (Lasso et al., 2019). Recently, structural bioinformatics has experienced a major breakthrough by the introduction of Deep Learning (DL) methods to predict protein structure and PPIs from sequence (Yang et al., 2020; Baek et al., 2021; Jumper et al., 2021; Kryshtafovych et al., 2021; Bryant et al., 2022; Evans et al., 2022; Gao et al., 2022). Yang et al. summarize technical details of DL in the context of viral-host PPI prediction, including the different types of architecture, dataset preparation, feature engineering and performance assessment. While we expect DL-based methods to play a major role in inter-species PPI prediction in the near future, Yang et al. highlight important aspects of the technique that require further improvements and careful examination.

The following research articles illustrate important aspects of virus-host PPIs, including amino acid variations at protein interfaces, through a wide range of approaches such as X-ray crystallography, cryo-electron microscopy (Cryo-EM), molecular dynamics (MD), protein structure modeling, and binding affinity assays. Ford et al. combined DL-based protein structure modeling and protein docking to evaluate the potential binding between the spike (S) protein of the Omicron variant of SARS-CoV-2 and four neutralizing monoclonal antibodies (mAbs) targeting S with known structure. This study highlights amino acid variations that are predicted to decrease mAb-binding affinities without completely abrogating interactions and has important implications in the rapid assessment of neutralization escape potential of emerging viral strains. Zhu et al. experimentally studied the interaction between a PDZ-domain binding motif (PBM) found in SARS-CoV-2 envelope (E) protein and PDZ-containing cellular proteins, which are commonly targeted by other viruses (Javier and Rice, 2011). The authors identified an interaction of E protein with several PDZ domains of host proteins involved in cellular junctions and cell polarity, resulting in the sequestration of these host proteins in the Golgi compartment. Structural studies on PDZ:PBM complexes highlighted structure and sequence preferences at the interface. Ongoing studies focus on a point mutation in E protein localized in proximity to its PBM in the SARS-CoV-2 variant of concern beta that was shown to influence the binding affinity of E protein for PDZ domains. Glycans also play a key role in modulating the interaction with host proteins (Thompson et al., 2019; Watanabe et al., 2019). However, their intrinsic flexibility and cell-type specific composition makes them difficult to study experimentally. Stagnoli et al. combine Cryo-EM and MD to investigate the composition and dynamics of the glycan shield in the SARS-CoV-2 S protein. The authors show that the conformation of the glycans that best fit the Cryo-EM density map are those in which the movement of the most external carbohydrates are more geometrically restricted, providing an understanding of why these glycans are visible by Cryo-EM. Finally, Sabariegos et al. describe how the interaction between the cellular kinase Akt and the Hepatitis C virus (HCV) RNA-dependent RNA polymerase NS5B modulates this viral protein via phosphorylation of conserved residues. Site-directed mutagenesis of key NS5B residues to mimic phosphorylation significantly reduced RNA polymerase activity and prevented rescue of HCV from infectious clones, thus, describing a mechanism of viral polymerase inactivation whose biological role remains to be determined.

In conclusion, this Research Topic provides an overview of computational and experimental approaches that, when combined, can significantly accelerate our understanding of virus-host PPIs and their biological role in viral infectious diseases. We are grateful for the valuable contributions of authors, reviewers, and members of the Editorial team at Frontiers.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

RKJ would like to acknowledge funding from Louisiana State University Health Sciences Center-Shreveport and National Institute of Allergy and Infectious Diseases of the National Institutes of Health (R21 AI156482).

REFERENCES

Abubakar, I., Gautret, P., Brunette, G. W., Blumberg, L., Johnson, D., Poumerol, G., et al. (2012). Global perspectives for prevention of infectious diseases associated with mass gatherings. *Lancet Infect. Dis.* 12, 66–74. doi: 10.1016/S1473-3099(11)70246-8

Ahmed, S., Dávila, J. D., Allen, A., Haklay, M. M., Tacoli, C., and Fèvre, E. M. (2019). Does urbanization make emergence of zoonosis more likely? Evidence, myths and gaps. *Environ. Urban* 31, 443–460. doi: 10.1177/0956247819866124

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373, 871–876. doi: 10.1126/science.abj8754

Bryant, P., Pozzati, G., and Elofsson, A. (2022). Improved prediction of protein-protein interactions using AlphaFold2. *Nat. Commun*. 13, 1265. doi: 10.1038/s41467-022-28865-w

Carlson, C. J., Albery, G. F., Merow, C., Trisos, C. H., Zipfel, C. M., Eskew, E. A., et al. (2022). Climate change increases cross-species viral transmission risk. *Nature* doi: 10.1038/s41586-022-04788-w. [Epub ahead of print].

Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., et al. (2022). Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 2021, 10.04.463034. doi: 10.1101/2021.10.04.463034

Gao, M., An, D. N., Parks, J. M., and Skolnick, J. (2022). Predicting direct physical interactions in multimeric proteins with deep learning. *bioRxiv* 2021, 11.09.467949. doi: 10.1101/2021.11.09.467949

Gibb, R., Franklinos, L. H. V., Redding, D. W., and Jones, K. E. (2020). Ecosystem perspectives are needed to manage zoonotic risks in a changing climate. *BMJ* 371, m3389–m3389. doi: 10.1136/bmj.m3389

Javier, R. T., and Rice, A. P. (2011). Emerging theme: cellular PDZ proteins as common targets of pathogenic viruses. *J. Virol*. 85, 11544–11556. doi: 10.1128/JVI.05410-11

Johnson, C. K., Hitchens, P. L., Pandit, P. S., Rushmore, J., Evans, T. S., Young, C. C. W., et al. (2020). Global shifts in mammalian population trends reveal key predictors of virus spillover risk. *Proc. Biol. Sci*. 287, 20192736. doi: 10.1098/rspb.2019.2736

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi: 10.1038/s41586-021-03819-2

Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K., and Moult, J. (2021). Critical assessment of methods of protein structure prediction (CASP)-Round XIV. *Proteins* 89, 1607–1617. doi: 10.1002/prot.26237

Lasso, G., Mayer, S. V., Winkelmann, E. R., Chu, T., Elliot, O., Patino-Galindo, J. A., et al. (2019). A structure-informed atlas of human-virus interactions. *Cell* 178, 1526 e16–1541 e16. doi: 10.1016/j.cell.2019.08.005

Thompson, A. J., Vries, D. E., and Paulson, R. P. (2019). Virus recognition of glycan receptors. *Curr. Opin. Virol*. 34, 117–129. doi: 10.1016/j.coviro.2019.01.004

Watanabe, Y., Bowden, T. A., Wilson, I. A., and Crispin, M. (2019). Exploitation of glycosylation in enveloped virus pathobiology. *Biochim. Biophys. Acta Gen. Subj*. 1863, 1480–1497. doi: 10.1016/j.bbagen.2019.05.012

Watsa, M. (2020). Rigorous wildlife disease surveillance. *Science* 369, 145–147. doi: 10.1126/science.abc0017

Yang, J., Anishchenko, V., Park, H., Peng, Z., Ovchinnikov, S., and Baker, D. (2020). Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci. U.S.A*. 117, 1496–1503. doi: 10.1073/pnas.1914677117

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 22 October 2021
doi: 10.3389/fmicb.2021.754664

Akt Phosphorylation of Hepatitis C Virus NS5B Regulates Polymerase Activity and Hepatitis C Virus Infection

*Rosario Sabariegos[1,2,3], Laura Albentosa-González[1], Blanca Palmero[1], Pilar Clemente-Casares[1,3,4], Eugenio Ramírez[1], Carlos García-Crespo[5,6], Isabel Gallego[5,6], Ana Isabel de Ávila[5,6], Celia Perales[5,6,7], Esteban Domingo[3,5,6] and Antonio Mas[1,3,4]**

[1] Centro Regional de Investigaciones Biomédicas, Universidad de Castilla-La Mancha, Albacete, Spain, [2] Facultad de Medicina, Universidad de Castilla-La Mancha, Albacete, Spain, [3] Unidad de Biomedicina UCLM-CSIC, Madrid, Spain, [4] Facultad de Farmacia, Universidad de Castilla-La Mancha, Albacete, Spain, [5] Centro de Biología Molecular "Severo Ochoa" (CBMSO) (CSIC-UAM), Consejo Superior de Investigaciones Científicas (CSIC), Madrid, Spain, [6] Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd) del Instituto de Salud Carlos III, Madrid, Spain, [7] Department of Clinical Microbiology, Instituto de Investigación Sanitaria-Fundación Jiménez Díaz University Hospital, Universidad Autónoma de Madrid (IIS-FJD, UAM), Madrid, Spain

OPEN ACCESS

Edited by:
Eva Mittler, Albert Einstein College of Medicine, United States

Reviewed by:
Leiliang Zhang, Shandong First Medical University, China
David McGivern, United States Food and Drug Administration, United States

***Correspondence:**
Antonio Mas
Antonio.Mas@uclm.es

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *06 August 2021*
Accepted: *20 September 2021*
Published: *22 October 2021*

Citation:
Sabariegos R, Albentosa-González L, Palmero B, Clemente-Casares P, Ramírez E, García-Crespo C, Gallego I, de Ávila AI, Perales C, Domingo E and Mas A (2021) Akt Phosphorylation of Hepatitis C Virus NS5B Regulates Polymerase Activity and Hepatitis C Virus Infection. Front. Microbiol. 12:754664. doi: 10.3389/fmicb.2021.754664

Hepatitis C virus (HCV) is a single-stranded RNA virus of positive polarity [ssRNA(+)] that replicates its genome through the activity of one of its proteins, called NS5B. This viral protein is responsible for copying the positive-polarity RNA genome into a negative-polarity RNA strand, which will be the template for new positive-polarity RNA genomes. The NS5B protein is phosphorylated by cellular kinases, including Akt. In this work, we have identified several amino acids of NS5B that are phosphorylated by Akt, with positions S27, T53, T267, and S282 giving the most robust results. Site-directed mutagenesis of these residues to mimic (Glu mutants) or prevent (Ala mutants) their phosphorylation resulted in a reduced NS5B *in vitro* RNA polymerase activity, except for the T267E mutant, the only non-conserved position of all those that are phosphorylated. In addition, *in vitro* transcribed RNAs derived from HCV complete infectious clones carrying mutations T53E/A and S282E/A were transfected in Huh-7.5 permissive cells, and supernatant viral titers were measured at 6 and 15 days post-transfection. No virus was rescued from the mutants except for T53A at 15 days post-transfection whose viral titer was statistically lower as compared to the wild type. Therefore, phosphorylation of NS5B by cellular kinases is a mechanism of viral polymerase inactivation. Whether this inactivation is a consequence of interaction with cellular kinases or a way to generate inactive NS5B that may have other functions are questions that need further experimental work.

Keywords: HCV (hepatitis C), NS5B (non-structural protein) polymerase, Akt, virus replication, phosphorylation

INTRODUCTION

Hepatitis C virus (HCV) is a positive single-stranded RNA virus [ssRNA(+)] belonging to Flaviviridae family. HCV genome replication takes place in replication complexes where the non-structural NS5B protein produces positive single-stranded genome copies [RNA(+)] through an intermediary of negative polarity [RNA(−)] (Neufeldt et al., 2018; Tabata et al., 2020). The HCV

protein NS5B is an RNA-dependent RNA polymerase (RDRP) which, like other polymerases in its class, shows a structure that has been compared to the shape of a right hand with three subdomains called fingers, palm, and thumb (Sesmero and Thorpe, 2015). The subdomain of the palm comprises three well-preserved motifs A (D^{220}-X(4)-D^{225}), B (S^{282}-X(8)-N^{291}), and C ($G^{317}D^{318}D^{319}$), which define the catalytic center. The aspartic acid residues D^{220} in motif A and D^{318} and D^{319} in motif C are involved in the coordination of the divalent cations (Mg^{2+} and/or Mn^{2+}) essential for the formation of the phosphodiester bond. Residues D^{225} of motif A, and S^{282} and N^{291} residues of motif B are involved in selection of ribonucleoside triphosphates over dNTPs and, thus, determine whether RNA rather than DNA is synthesized (Sesmero and Thorpe, 2015).

Viruses are obligate intracellular parasites that usurp the cellular machinery to complete their replicative cycle. This means that viral components need to interact with cellular components to direct cellular activity according to viral interests. HCV is no exception, and an enormous amount of interactions between viral components and cellular proteins have already been described (de Chassey et al., 2008; Dolan et al., 2013; Hagen et al., 2014). HCV proteins NS3, NS5A, and core have shown the most extensive network of interactions with host factors (de Chassey et al., 2008). NS5B also interacts with cellular proteins (Hamamoto et al., 2005; Munakata et al., 2005; Watashi et al., 2005; Kusakawa et al., 2007; Inoue et al., 2011; Hillung et al., 2012), and some of its interactions require partial denaturation of NS5B and subsequent loss of RDRP activity. For example, the retinoblastoma tumor suppressor is down regulated by interacting with NS5B via amino acids located in the catalytic center (C motif) (Munakata et al., 2005). For this interaction to occur, the NS5B protein must expose amino acids located in the active center. These data indicate that an inactive NS5B may be modulating an important surveillance pathway.

We have previously described the interaction of NS5B with the cellular kinase Akt and the changes in subcellular localization related with this protein:protein interaction (Valero et al., 2016). Furthermore, we documented that Akt phosphorylates NS5B and that inhibitors of Akt affect HCV replication in cell culture (Valero et al., 2016). Viral RNA polymerase phosphorylation has also been described in other viral systems previously (Barik et al., 1995; Jakubiec et al., 2006; Schmid et al., 2007), and norovirus RNA polymerase is phosphorylated by Akt as well (Eden et al., 2011). The PI3K-Akt pathway, transiently activated during HCV entry (Liu et al., 2012; Qian et al., 2020), has been linked to HCV infection and related metabolic disorders (Qadri et al., 2012; Zhang et al., 2018). Therefore, modulation of Akt and Akt-related proteins could be of great importance for HCV replication.

In this study, we describe the positions of the HCV NS5B protein that are phosphorylated by the cellular kinase Akt. Furthermore, we show that mutants mimicking phosphorylation at these positions lead to proteins with very low RDRP activities. In addition, viruses carrying mutations mimicking Akt phosphorylation of T53 or S282 residues were unable to replicate in cultured cells. Akt:NS5B interaction produces an inactive viral polymerase whose role needs to be elucidated.

MATERIALS AND METHODS

Reagents, Expression Plasmids, and Inhibitors

Plasmid pET_NS5BΔ21 encoding HCV NS5B from strain HC-J4 with a 21 amino acid deletion at the C-terminal end has been described previously (Lopez-Jimenez et al., 2014). Recombinant Akt/PKB was purchased from Biaffin (PK-PKBA-020, Biaffin

TABLE 1 | Oligonucleotides used in this study.

Name	Sequence (5′–>3′)
pJ4 S27A s	gcccatcaacccgttggccaactctttgctgcgt
pJ4 S27A as	acgcagcaaagagttggccaacgggttgatgggc
pJ4 S29A s	aacccgttgagcaacgctttgctgcgtcacc
pJ4 S29A as	ggtgacgcagcaaagcgttgctcaacgggtt
pJ4 S27A/S29A s	cccatcaacccgttggccaacgctttgctgcgtcacc
pJ4 S27A/S29A as	ggtgacgcagcaaagcgttggccaacgggttgatggg
pJ4 T53A s	ccggcagaagaaggtcgcctttgacagattgca
pJ4 T53A as	tgcaatctgtcaaaggcgaccttcttctgccgg
pJ4 T267A s	cgggggtcccctggctaactcaaaaggg
pJ4 T267A as	ccctttgagttagccaggggacccccg
pJ4 T269A s	gggtcccctgactaacgcaaaagggcagaactg
pJ4 T269A as	cagttctgccctttgcgttagtcaggggaccc
pJ4 T267A/T269A s	cgggggtcccctggctaacgcaaaagggcagaa
pJ4 T267A/T269A as	ttctgccctttgcgttagccaggggacccccg
pJ4 T282A s	cgccggtgccgcgcagctggcgtgc
pJ4 T282A as	gcacgccagctgcgcggcaccggcg
pJ4 S27E s	ctgcccatcaacccgttggagaactctttgctgcgtcac
pJ4 S27E as	gtgacgcagcaaagagttctccaacgggttgatgggcag
pJ4 S29E s	ccatcaacccgttgagcaacgagttgctgcgtcaccacaacat
pJ4 S29E as	atgttgtggtgacgcagcaactcgttgctcaacgggttgatgg
pJ4 S27E/S29E s	agtaagctgcccatcaacccgttggagaa cgagttgctgcgtcaccacaacatggtc
pJ4 S27E/S29E as	gaccatgttgtggtgacgcagcaa ctcgttctccaacgggttgatgggcagcttact
pJ4 T282E s	tatcgccggtgccgcgcagagggcgtgctgacg
pJ4 T282E as	cgtcagcacgccctctgcgcggcaccggcgata
pJ4 T53E s	ggacttgcaatctgtcaaactcgaccttcttctgccggagg
pJ4 T53E as	cctccggcagaagaaggtcgagtttgacagattgcaagtcc
pJ4 T267E/T269E s	ggctttacatcgggggtcccctggaga acgagaaagggcagaactgcggttatcg
pJ4 T267E/T269E as	cgataaccgcagttctgccctttctcgttctccaggggacccccgatg-taaagcc
pJ4 T267E s	cagttctgccctttgagttctccaggggacccccgatgtaaa
pJ4 T267E as	tttacatcgggggtcccctggagaactcaaaagggcagaactg
pJ4 T269E s	gcagttctgccctttctcgttagtcaggggacccccga
pJ4 T269E as	tcgggggtcccctgactaacgagaaagggcagaactgc
NS5B T53E s	tcacagagggctaaaaaggtagagtttgacaggacgcaagtgctc
NS5B T53E as	gagcacttgcgtcctgtcaaactctacctttttagccctctgtga
NS5B T282E s	agacgttgccgcgccgaggggtgctaaccact
NS5B T282E as	agtggttagcacccccctcggcgcggcaacgtct
NS5B T53A s	cagagggctaaaaaggtagcttttgacaggacgcaag
NS5B T53A as	cttgcgtcctgtcaaaagctacctttttagccctctg
NS5B T282A s	acgttgccgcgccgccggggtgctaacc
NS5B T282A as	ggttagcaccccggcggcgcggcaacgt

GmbH&Co), and lipofectamine 2000 from Invitrogen. Huh7.5 cells were kindly provided by Dr. R. Bartenschlager (University of Heidelberg, Germany).

Hepatitis C Virus NS5B WT and Mutants Purification

Point mutants in NS5B were generated by site-directed mutagenesis following the manufacturer's instructions (QuikChange Site-Directed Mutagenesis, Agilent Technologies). Synthetic oligonucleotides used for point mutant generation are described in **Table 1**. NS5B wild type and mutants were over-expressed and purified as described previously (Lopez-Jimenez et al., 2014; Valero et al., 2016).

In vitro Kinase Assay

Kinase assays were performed as previously described (Albentosa-Gonzalez et al., 2021). Briefly, HCV NS5B (3 μg) was incubated in 20 mM Hepes pH 7.4, 10 mM $MgCl_2$, 10 mM $MnCl_2$, 1 μCi of $\gamma[^{32}P]$-ATP, 1 mM DTT, in the presence of 0.5 μg of recombinant Akt/PKB (PK-PKBA-A020, Biaffin GmbH&Co). Following SDS-PAGE electrophoresis, the gel was dried and exposed to phosphorimager screens, and scanned with Typhoon 9600 (Molecular Dynamics) to detect radiolabeled products.

In-Gel Digestion and Reverse Phase-Liquid Chromatography RP-LC-MS/MS Analysis

The identification of phosphorylated NS5B residues was performed as previously described (Albentosa-Gonzalez et al., 2021). Briefly, phosphorylated protein was digested *in situ* with sequencing grade trypsin (Promega, Madison, WI, United States) and analyzed by RP-LC-MS/MS in an Easy-nLC II system coupled to an ion trap LTQ-Orbitrap-Velos-Pro mass spectrometer (Thermo Scientific). The MS/MS spectra from the peptides were analyzed by assigning the fragment ions to the candidate sequence after calculating the series of theoretical fragmentations.

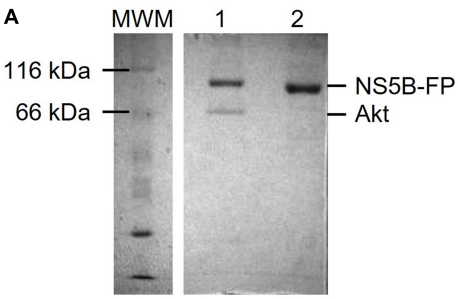
In vitro RNA-Dependent RNA Polymerase Replication Assays

RNA polymerase assays were performed using the symmetric substrate LE-19, which is capable of *de novo* initiation (DN) and primer-extension (PE), as previously described (Clemente-Casares et al., 2011; Lopez-Jimenez et al., 2014). 200 nM NS5B was pre-incubated for 30 min in a reaction mixture containing 20 mM MOPS, pH 7.3, and 5 mM $MnCl_2$. Reactions were started by adding 500 μM GTP, 100 μM ATP, and UTP, and 1 μCi $\alpha[^{32}P]$CTP (3000 Ci mmol, PerkinElmer Life Sciences). Reactions were stopped by adding EDTA/formamide loading buffer at different time points as indicated. Products were separated using denaturing polyacrylamide (23% PAA, 7 M urea) gel electrophoresis. Gels were exposed to phosphorimager screens and scanned with Typhoon (Molecular Dynamics). Quantification was achieved by running samples on parallel gels and determining band volumes using ImageQuant software (GE Healthcare).

Cell Culture Virus Infection

The origin of the Huh-7.5 cell line, procedures for cell growth in Dulbecco's modified Eagle's medium (DMEM), the virus used in the experiments rescued from plasmid Jc1FLAG2(p7-nsGluc2A) (a chimera of J6 and JFH-1 from genotype 2a), and the procedures used to prepare the initial virus stock HCVp0, to titrate viral infectious particles, and to quantify viral RNA have been described previously (Perales et al., 2013; Sheldon et al., 2014). To perform infections for immunofluorescence and RNA quantification assays, 1×10^5 Huh-7.5 cells were infected with HCVp0 at a multiplicity of infection (MOI) of 0.5 Tissue Culture Infectious Dose (50%) TCID50/cell. The infected cells were further incubated at 37°C for 6 and 15 days. Absence of contamination was checked by maintaining and titrating mock-infected cells and their supernatants in parallel with the infected cultures. No infectivity in the mock-infected cultures was detected in any of the experiments.

Plasmid Jc1FLAG2(p7-nsGluc2A) was used as a template for constructing T53A, T53E, S282A, and S282E mutants by site-directed mutagenesis using the oligonucleotides described in **Table 1**. Plasmids carrying the selected mutations were transcribed *in vitro* and the genomic RNA from WT or mutant virus was used to transfect cells as described above to produce HCVp0 virus stock. The supernatant from these transfections



B

```
          10         20         30         40         50         60         70
SMSYTWTGAL ITPCAAEESK LPINPLSNSL LRHHNMVYAT TSRSASLRQK KVTFDRLQVL DDHYRDVLKE
          80         90        100        110        120        130        140
MKAKASTVKA KLLSIEEACK LTPPHSAKSK FGYGAKDVRN LSSRAVNHIR SVWEDLLEDT ETPIDTTIMA
         150        160        170        180        190        200        210
KSEVFCVQPE KGGRKPARLI VFPDLGVRVC EKMALYDVVS TLPQAVMGSS YGFQYSPKQR VEFLVNTWKS
         220        230        240        250        260        270        280
KKCPMGFSYD TRCFDSTVTE SDIRVEESIY QCCDLAPEAR QAIRSLTERL YIGGPLTNSK GQNCGYRRCR
         290        300        310        320        330        340        350
ASGVLTTSCG NTLTCYLKAT AACRAAKLQD CTMLVNGDDL VVICESAGTQ EDAAALRAFT EAMTRYSAPP
         360        370        380        390        400        410        420
GDPPQPEYDL ELITSCSSNV SVAHDASGKR VYYLTRDPTT PLARAAWETA RHTPINSWLG NIIMYAPTLW
         430        440        450        460        470        480        490
ARMILMTHFF SILLAQEQLE KALDCQIYGA CYSIEPLDLP QIIERLHGLS AFTLHSYSPG EINRVASCLR
         500        510        520        530        540        550        560
KLGVPPLRTW RHRARSVRAK LLSQGGRAAT CGRYLFNWAV RTKLKLTPIP AASQLDLSGW FVAGYSGGDI
         570
YHSLSRARPR
```

FIGURE 1 | (A) SDS-PAGE gel showing products obtained after Akt phosphorylation of NS5B-FP (lane 1). Unphosphorylated NS5B-FP is shown in lane 2. Molecular weight marker (MWM) is on the left of the gel and sizes for NS5B-FP and Akt are indicated on the right. **(B)** Amino acid sequence of HCV NS5B from strain HC-J4. Phosphorylated sites identified in phosphoproteomics are in red (with increased font size). Sequence uncovered by the phosphoproteomic analysis is in gray color.

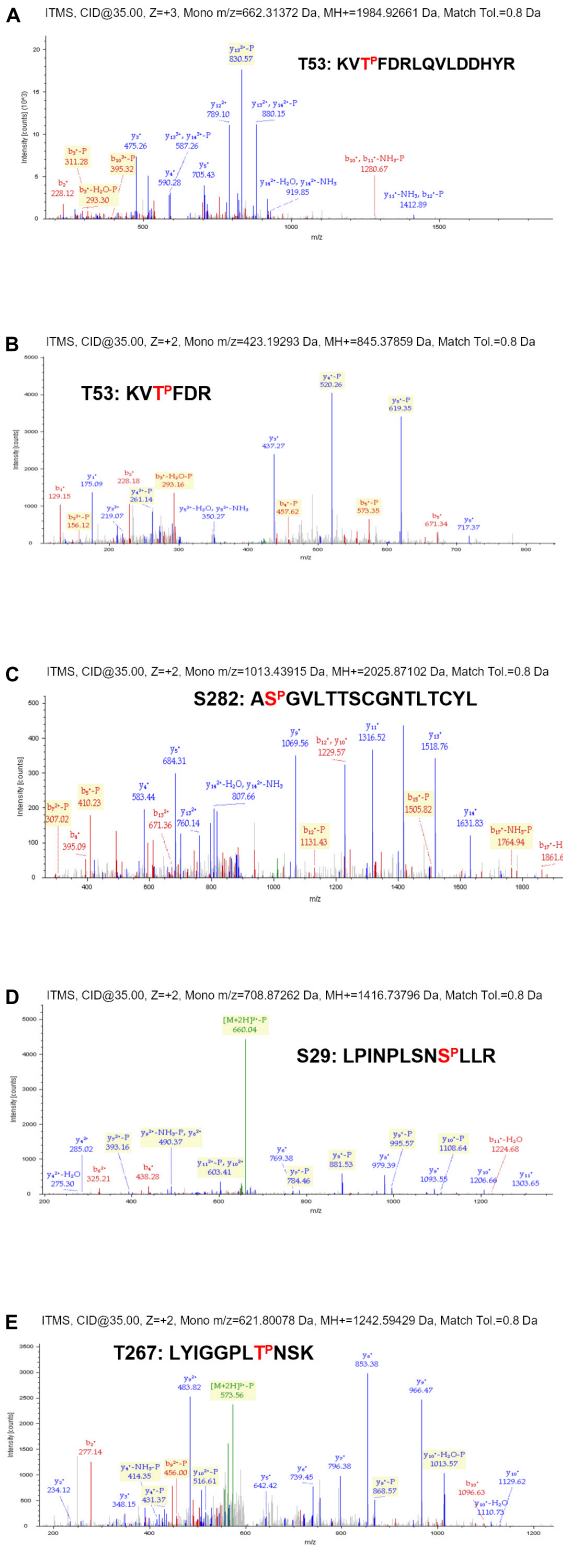


FIGURE 2 | Phosphoproteomic details from the phosphorylation experiment described in **Figure 1**. MS/MS profiles for NS5B-FP tryptic peptides encompassing residues T53 **(A,B)**, S282 **(C)**, S29 **(D)**, and T267 **(E)**. Peptide sequences corresponding to each spectrum are indicated with the identified phosphorylated residue marked in red.

was used to titer the virus obtained. Supernatant from infections with mutant virus in which viral titer was obtained was used to purify viral RNA that was retrotranscribed, PCR amplified, and sequenced to test for mutation reversal.

RESULTS

Phosphoproteomic Studies

Previous results from our laboratory have shown that the Akt protein phosphorylates the NS5B polymerase of HCV (Valero et al., 2016). To analyze the effect that this phosphorylation causes in the biology of the virus, we first wanted to identify the residues phosphorylated by this Ser/Thr kinase. To this end, we performed an *in vitro* Akt phosphorylation assay using recombinant protein NS5B (66 kDa) fused to the fluorescent protein EGFP (NS5B-FP) and $\gamma[^{32}P]$-ATP as substrates. The fused protein shows a molecular mass of approximately 95 kDa that can easily be distinguished from Akt (68 kDa). The products of the phosphorylation reaction were resolved by SDS-PAGE and two bands were detected, one at about 100 kDa corresponding to NS5B-FP, and the other at about 65 kDa corresponding to the autophosphorylated Akt (**Figure 1A**). The phosphorylation experiment was repeated with cold ATP for proteomic purposes. The NS5B-FP band was extracted and digested with trypsin, yielding a sequence coverage of 70% in the proteomic analysis (**Figure 1B**).

Proteomic analysis of the trypsin digestion products allowed the identification of the phosphorylated residues S29, T53, T267, and S282 (**Figure 2**). Ser29 is located in the subdomain called fingertips, and is involved in the interaction with the thumb (**Figure 3A**). Residues T53 and T267 are located in the fingers subdomain, and are involved in the helix-helix interactions that stabilize this subdomain (**Figure 3A**). Finally, the S282 residue is part of the motif B site and interacts with the D225 residue of motif A for the correct positioning of the NTP during the formation of the phosphodiester bond. Two other residues (S27 and S269) could be phosphorylated and were not well identified because of their proximity to the most likely ones (S29 and T267). In any case, it is much less likely that they are phosphorylated and this was the reason they were not included in some experiments.

NS5B Activity Studies

Next, we sought to determine the effect of the phosphorylation of the residues described above on the RDRP activity of the NS5B protein. To this end, mutations that mimic phosphorylated Ser or Thr residues at these positions were introduced by site-directed mutagenesis, yielding proteins carrying the mutations S27E, S29E, T53E, T267E, S269E, and S282E. We also generated mutants carrying Ala in these positions. All these proteins were overexpressed in *E. coli* and purified to homogeneity by affinity chromatography, as judged by SDS-PAGE (**Figure 3B**).

RDRP activity assays were carried out using oligonucleotide LE19, which allows *de novo* initiation and primer extension to be analyzed at the same time. Results with Ala mutants showed activity levels below 50% compared to WT in all cases and for both types of activity. Of all the Ala mutants tested, T267A

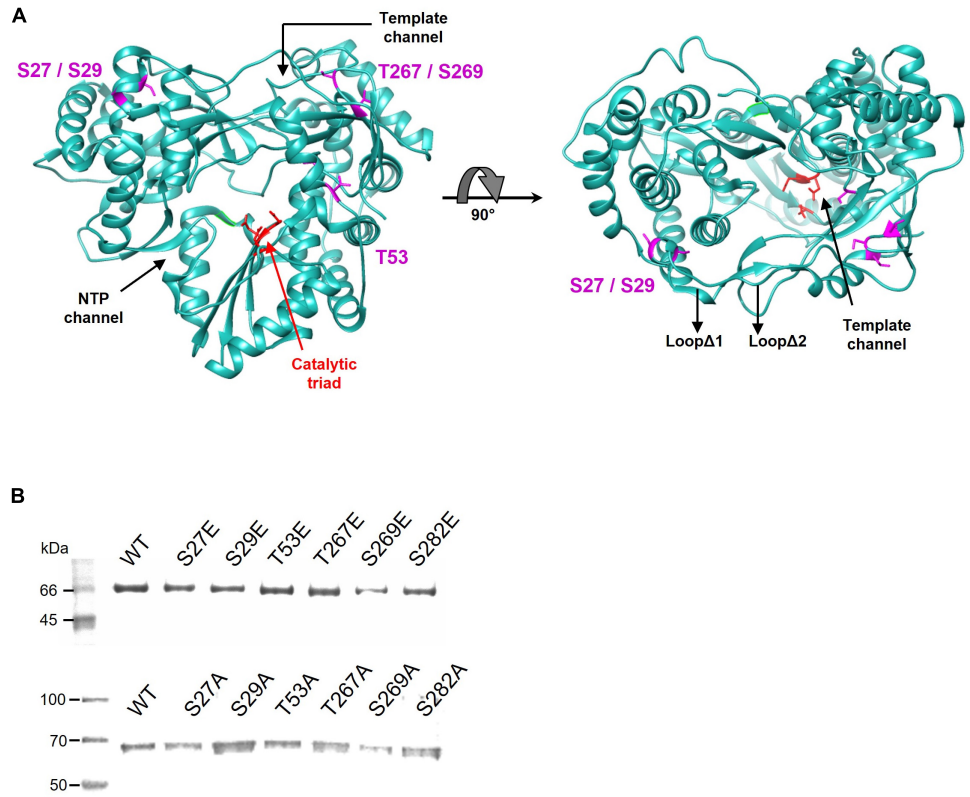


FIGURE 3 | (A) Localization of the residues phosphorylated by Akt. The positions of the amino acids identified by proteomics are shown in magenta on the structure of the NS5B HC-J4 (PDB: 1NB4) protein. The amino acids of the catalytic triad are shown in red. The most important structural elements (template channel, nucleotide entry channel and loops Δ1 and Δ2) are also indicated. **(B)** SDS-PAGE gel showing purified WT and Glu (upper panel) or Ala (lower panel) mutant proteins for each position. The molecular weight marker is shown on the left. The identity of the protein in each lane is indicated above the protein bands.

showed the highest activity levels (**Figure 4**). Assays with the Glu mutants showed imperceptible activity levels for all mutants except T267E, which showed RNA polymerase activity similar to or even higher than the WT protein (**Figure 5**). Therefore, changes that mimic or abrogate phosphorylation at the NS5B protein positions identified to be phosphorylated by Akt yield proteins with much lower levels of RDRP activity than the WT protein for both *de novo* and primer-dependent synthesis, with the sole exception of the T267E mutant.

Phosphorylation of Ser and Thr residues leads to the introduction of negative net charges in these positions. Ser282, one of the NS5B residues that becomes phosphorylated by Akt, is part of the active center of the polymerase. We therefore wanted to analyze whether the mutant that mimics phosphorylation in this position maintains the ability to bind RNA. For this purpose, we carried out electromobility shift assay (EMSA) experiments comparing WT and S282E proteins. First, we determined the concentration of WT protein that was able to bind and delay a fixed RNA concentration (corresponding to 10,000 cpm) and at a 100 mM NaCl concentration. The results show that at 100 nM protein all RNA is forming RNA:protein complexes (**Figure 6A**). Then, with those conditions (10,000 cpm RNA and 100 nM protein) as starting point, we performed electromobility shift assays varying the NaCl concentration to compare WT and S282E proteins. Since RNA:protein interactions are predominantly electrostatic, an increase in ionic strength will result in the loss of interaction. When the experiment was carried out at the lowest NaCl concentration (30 mM), all of the RNA with WT protein is forming RNA:protein complexes (**Figure 6B**). However, with the protein carrying the S282E mutation under these conditions, a free-form RNA band is observed, indicating less interaction between RNA and mutant protein compared to the WT protein (**Figure 6B**). In addition, whereas the WT protein even shows an intense band corresponding to RNA: protein complexes at the highest NaCl concentration tested (530 mM), the mutant protein shows very faint complex bands at concentrations higher than 330 mM NaCl (**Figure 6B**). These results indicate that, under the same conditions, the interaction of RNA with the S282E mutant protein is weaker than with the WT protein.

Effect of Mutations in T53 and S282 Residues on Virus Replication in Cell Culture

Next, the effect of NS5B phosphorylation by Akt or, alternatively, its inability to completely phosphorylate the polymerase, on HCV's replicative capacity was analyzed. To do this, we constructed T53A, T53E, S282A, and S282E mutants in the

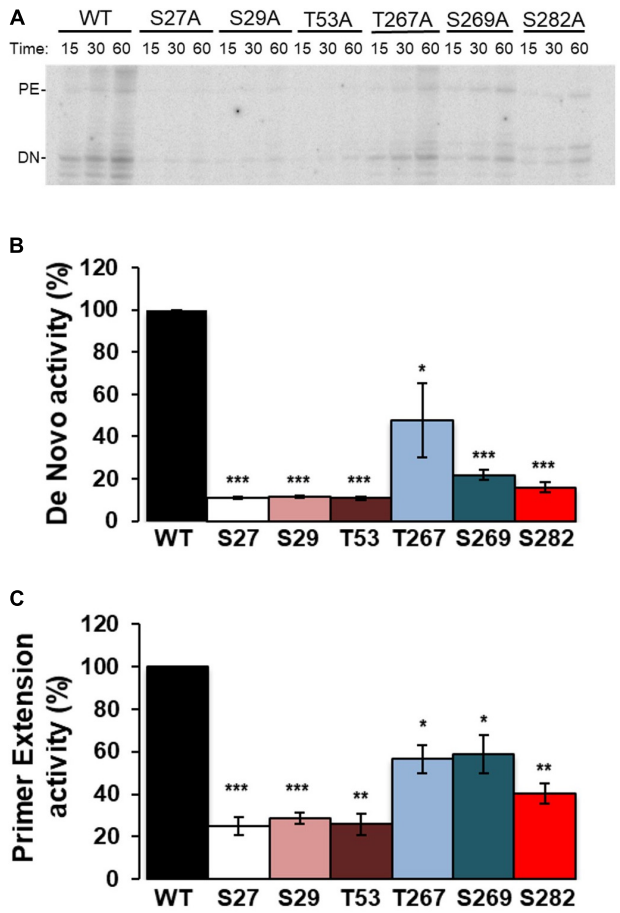


FIGURE 4 | Elongation of LE19 oligonucleotide by *de novo* initiation and primer extension RDRP activity of NS5B Ala mutants. NS5B WT and Ala mutants of residues S27, S29, T53, T267, S269, and S282 were assayed as described in "Materials and Methods" section. Aliquots were stopped at 15, 30, and 60 min, and resolved in denaturing polyacrylamide gels (PAA 23%). **(A)** Representative experiment showing products obtained by primer extension (PE) and *de novo* initiation (DN) using LE19 RNA as a template. Recombinant proteins used are indicated at the top. **(B)** RNA-dependent RNA polymerase *de novo* activity products at 60 min were quantified and represented in arbitrary units. **(C)** RNA-dependent RNA polymerase primer extension activity products at 60 min were quantified and represented in arbitrary units. The mutated residue (Ala mutant) corresponding to each bar is indicated at the bottom. Values in **(B,C)** are the averages (normalized to WT) and corresponding Standard Errors of the Mean (SEM) from at least four independent experiments. Statistically significant differences (Student's *t*-test) are represented as follows: $*p < 0.05$; $**p < 0.005$; $***p < 0.0005$. Asterisks over the bar indicate the *p*-value comparing to WT.

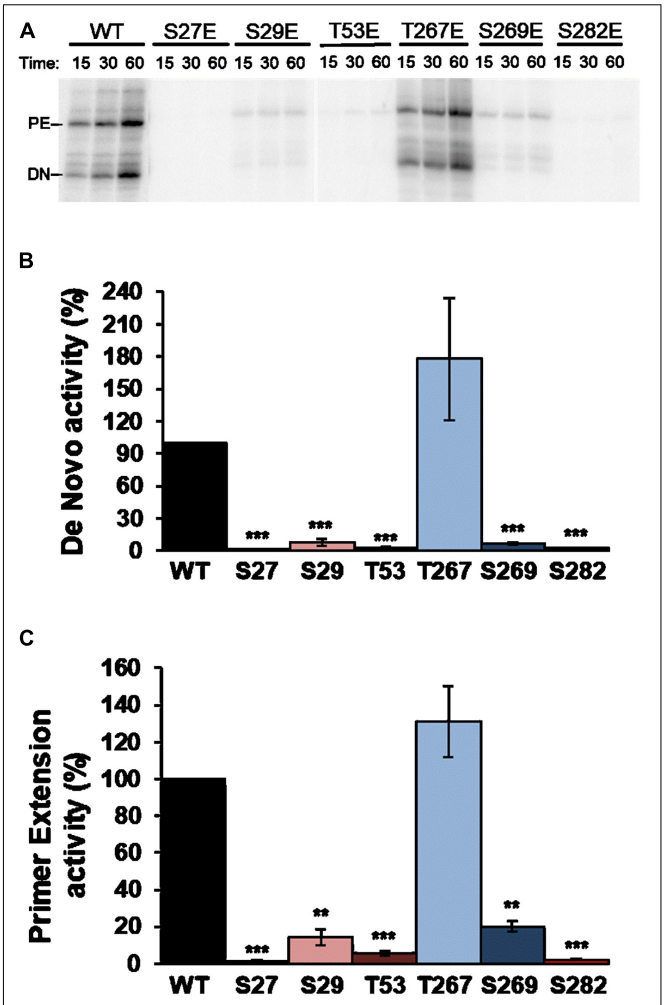


FIGURE 5 | Elongation of LE19 oligonucleotide by *de novo* initiation and primer extension of NS5B Glu mutants. **(A)** Representative experiment showing products obtained by primer extension (PE) and *de novo* initiation (DN) by NS5B WT and mutants imitating phosphorylation (Glu) of residues S27, S29, T53, T267, S269, and S282, using LE19 RNA as a template. Recombinant proteins used are indicated at the top. Quantitative analysis of RNA-dependent RNA polymerase *de novo* **(B)** and primer extension **(C)** activity products at 60 min quantified as described in **Figure 4**. The mutated residue (Glu mutant) corresponding to each bar is shown at the bottom. Values in **(B,C)** are the averages (normalized to WT) and corresponding Standard Errors of the Mean (SEM) from at least four independent experiments. Statistically significant differences (Student's *t*-test) are represented as follows: $**p < 0.005$; $***p < 0.0005$. Asterisks over the bar indicate the *p*-value comparing to WT.

Jc1FLAG2(p7-nsGluc2A) plasmid, the resulting plasmids were transcribed *in vitro,* and the RNA product of the transcription was used to transfect Huh-7.5 cells. The supernatant from these transfected cells was used to titrate for the presence of HCV following a described procedure (Perales et al., 2013; Sheldon et al., 2014). WT virus could be recovered from the supernatant of all three replicates of transfected cells with the corresponding RNA (**Figure 7A**). The average viral titer was 4.40×10^3 TCID50/ml at 6 days post-transfection, and 1.96×10^3 TCID50/ml at 15 days post-transfection. On the contrary, only two of the mutant virus replicates, both corresponding to the T53A mutant 15 days post-transfection, gave a viral titer value above the cut-off value (**Figure 7A**). T53A TCID50/ml values were 4.64×10^1 and 2.43×10^1.

To rule out that the virus collected from the supernatant of the cells transfected with the mutant T53A was a product of mutation reversal, we purified the viral RNA, subjected it to an RT-PCR reaction and sequenced the DNA product. The result allowed

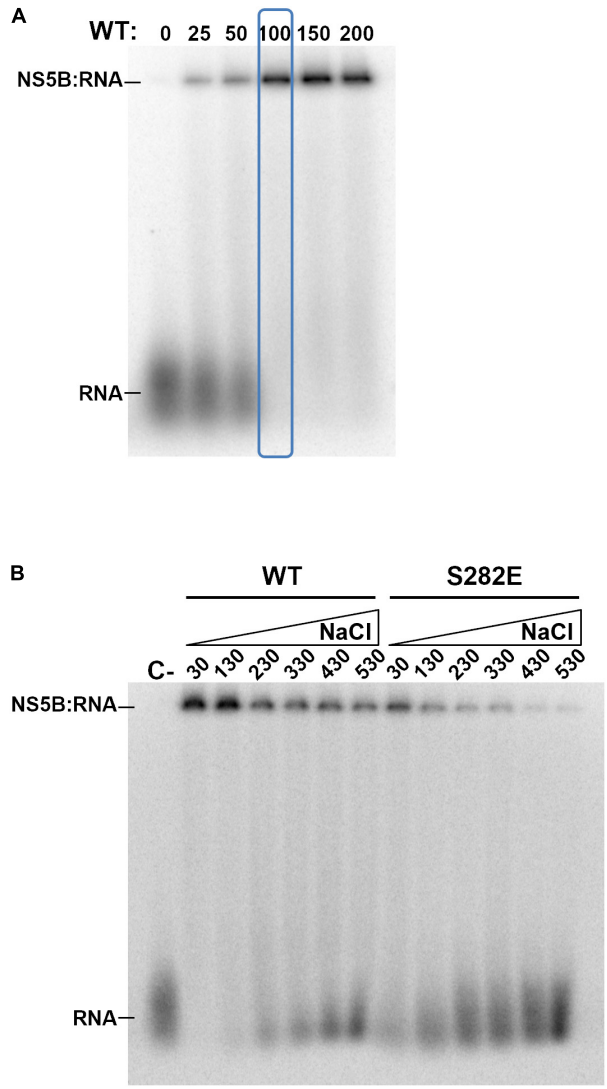


FIGURE 6 | Electromobility shift assays. **(A)** Gel showing retardation of a labeled RNA (10,000 cpm each lane) by increasing concentrations of WT NS5B. Concentration of WT NS5B is indicated at the top (nM). The protein concentration chosen for the following experiments is indicated by a blue box. **(B)** Gel showing retardation of a labeled RNA by WT and S282E NS5B proteins at increasing concentrations (30, 130, 230, 330, 430, and 530 mM) of NaCl. Free labeled probe (RNA) and retarded products (NS5B:RNA) are indicated.

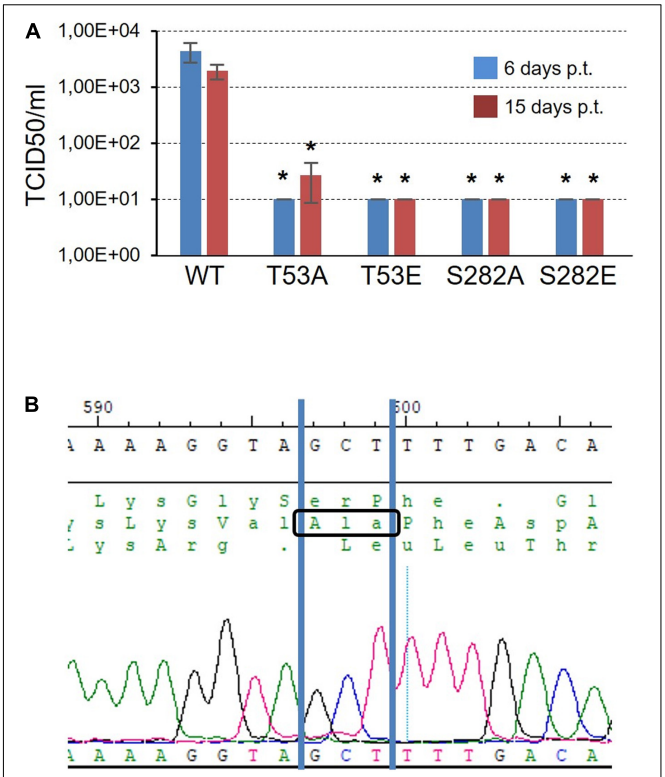


FIGURE 7 | Infectivity of T53 and S282 mutants. **(A)** TCID50/ml values obtained 6 and 15 days post-transfection (blue and red bars, respectively) after transfecting Huh7.5 cells with RNA corresponding to WT virus and the T53A, T53E, S282A, and S282E mutants. The threshold value was established at 10 TCID50/ml. Statistically significant differences (Student's *t*-test) are represented as follows: $*p < 0.05$. Asterisk over the bar indicate the *p*-value comparing to WT. **(B)** Chromatogram showing the sequence obtained from the RNA of the T53A mutant rescued 15 days post-transfection. The GCT codon corresponding to the amino acid Ala is identified.

us to confirm that the T53A change was present in the rescued virus (**Figure 7B**).

DISCUSSION

Previous work in our laboratory has shown that the cellular kinase Akt interacts with and phosphorylates the HCV polymerase NS5B (Valero et al., 2016). Now, we have identified phosphorylated positions, and analyzed the effect these modifications have on RDRP activity and virus replication. We identified positions S29 (or much less likely S27), T53, S269 (or much less probable T267) and S282 as substrates of the cellular Akt kinase (**Figure 1**). Coverage of the NS5B proteomic analysis was 70% of the sequence, leaving the possibility that some other phosphorylable positions have not been identified. Previous work has shown that Akt is important for virus replication in HCV-infected cells in culture (13 and references therein), suggesting that the HCV-Akt relationship is important *in vivo* and supporting the *in vitro* results described in the present work. Residue T267 is not conserved among genotypes (**Figure 8A**), and mutations at this position rendered the highest RDRP activity values among all mutants analyzed in this study, and even better than WT in the case of the T267E mutant. Residue T53 has not been described so far in relation to NS5B activity nor HCV replication. S29 and S282 residues have been previously described as important for both RDRP activity and viral replication. S29 has already been described previously as substrate of other important kinases in the HCV replicative cycle (Han et al., 2014; Hernandez et al., 2015). Furthermore, S29 is the homologous position of T33 of norovirus, which is also phosphorylated by Akt (Eden et al., 2011). These antecedents indicate us that phosphorylation of S29 seems to be very

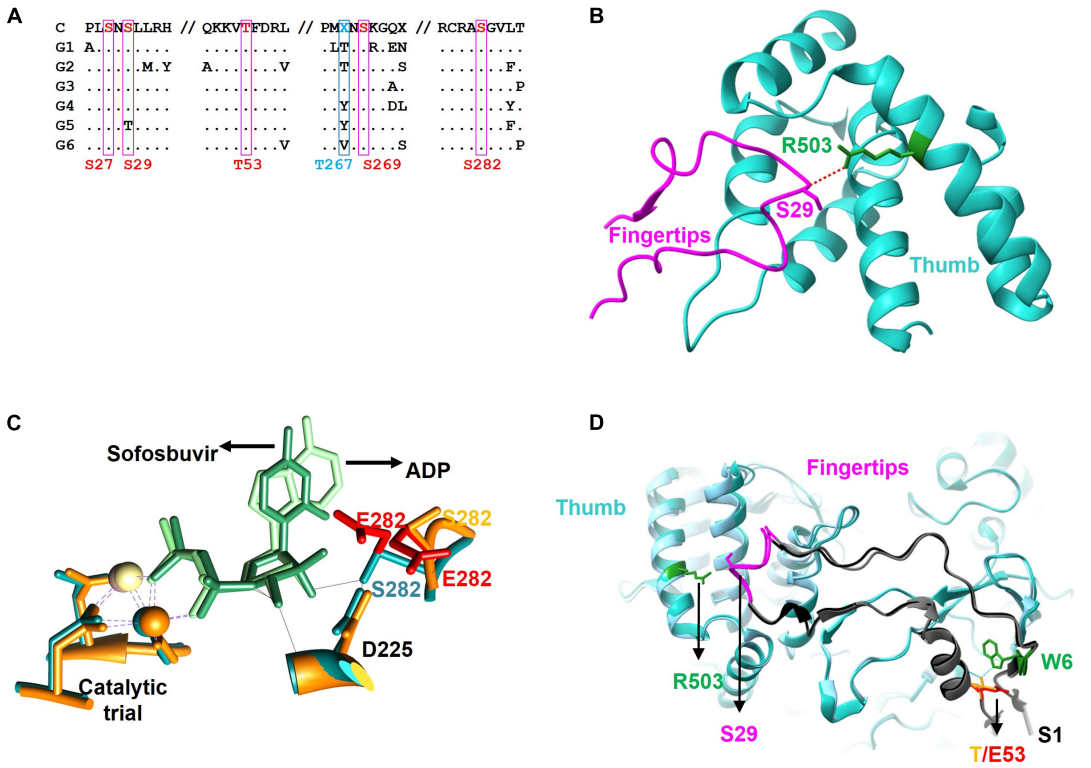


FIGURE 8 | Sequence and structure of the NS5B protein. **(A)** Partial amino acid sequence alignment of the consensus sequences of the major HCV genotypes (G1, G2, G3, G4, G5, and G6) showing the positions described in this work (S27, S29, T53, S269, and S282 in red and T267 in blue) as well as the surrounding residues. The consensus sequence is shown at the top **(C)**. Conserved residues are represented as a dot. **(B)** Zoom of the fingertip region of the NS5B protein showing residues S29 (magenta, fingertip domain) and R503 (blue, thumb domain) and the hydrogen bond connecting them (red dashed line). **(C)** Zoom of the catalytic center of NS5B obtained by aligning the protein structures in the presence of ADP (blue backbone) and sofosbuvir (orange backbone). The catalytic triad coordinated with beta and gamma phosphates of the incoming nucleotide (ADP in light green and sofosbuvir in dark green), and positions D225 and S282 are shown. The most likely rotamers for the S282E mutant are shown in red. **(D)** Zoom showing the Δ1 loop in dark gray. Fingertips with the amino acid S29 and its partner R503 are shown on the left side of the image. On the right are residues S1 (gray), W6 (green) and T53 (orange), as well as the most likely rotamer for the T53E mutant (in red). Hydrogen-bridging bonds between T53 and W6 are shown with dotted blue lines.

important for the replicative cycle of HCV (Han et al., 2014). Finally, S282 is a residue involved in substrate recognition by interacting with other amino acids (D225) and ribonucleotide substrates (**Figure 8B**; Appleby et al., 2015). Residue S282 has also been related to sofosbuvir and ribavirin resistance (Aloia et al., 2012; Ji et al., 2015; Kulkarni et al., 2016; Fourati et al., 2019). Therefore, with the exception of position T267, the positions identified in this work that are modified by Akt are critical for the RDRP activity of the NS5B protein as any change greatly affects RDRP activity.

S29 is located in the fingertips subdomain and is part of the residues involved in fingers-thumb interaction (Cai et al., 2005; **Figures 3**, **8B**). Previous *in vitro* work focused on this NS5B region showed that fingers:thumb interactions were key for RDRP activity, binding of RNA and for the transition from *de novo* initiation to primer extension (Chinnaswamy et al., 2008, 2010). In addition, viruses carrying the S29A mutation have serious replication defects in cell culture (Han et al., 2014), which is consistent with the low levels of RDRP activity we have found (**Figures 4**, **5**). The S29E change decreases the distance with R503, a residue that is at binding distance and is also part of the fingers:thumb interaction region (**Figure 8B**). The distance between S29 and R503 is between 3.3 and 3.8 Å, while the distance between the most stable rotamer of mutant S29E and R503 is between 3.4 and 2.6 Å. In addition, the phosphorylation of S29 introduces a negative charge in that location, and, considering the proximity of the R503 residue (≈3 Å), the interactions of this area could be distorted. This NS5B region must undergo large conformational changes for allowing the transition from *de novo* synthesis to primer extension (Chinnaswamy et al., 2008, 2010; Appleby et al., 2015). Mutations in these positions (e.g., phosphorylation of S29) could prevent these conformational changes or even impair a proper folding of the NS5B protein to carry out the replication of the genome.

Position S282 is part of the NS5B catalytic pocket (**Figure 8C**). This residue is part of the motif B next to D225 and is essential for the accurate positioning of NTP and the subsequent formation of the phosphodiester bond. The crystalline structure of NS5B in the presence of ADP shows that S282 has a hydrogen bond with the ribose of the nucleotide (**Figure 8C**; Appleby et al., 2015). By introducing the mutation S282E to mimic the phosphorylation in that position, all the rotamers including the more stable

ones lost that hydrogen bond (**Figure 8A**). Previously, it has been described that changes S282T/G/C/R are associated with resistance to sofosbuvir (Lam et al., 2012; Perales et al., 2015; Chen et al., 2016; Gane et al., 2017). Changes at this position, even the most subtle (S282T) that could be also phosphorylated by Akt, give rise to viruses with very low replication efficiency (Suda et al., 2019; **Figure 7**). The crystalline structure of the NS5B protein with sofosbuvir located in its active center has been determined (Appleby et al., 2015). This structure showed that S282 is displaced to avoid steric hindrance (compare the protein structures with ADP in cyan and with sofosbuvir in orange in **Figure 8C**). In addition, the S282E mutation does not allow the hydrogen bond with the ribose ring of the incoming nucleotide in any of the structures, neither in the presence of ADP nor sofosbuvir, as the WT protein does. The lack of an interaction network together with the displacement of protein motifs could be related to the results shown in **Figures 4**–**7**.

Position T53 is the least studied of the above-mentioned phosphorylable residues. It is fully conserved in all HCV genotypes. The amino acid T53 is probably involved in maintaining the structure of the Δ1 loop and fingertips through interactions with other amino acids such as W6, with which it establishes hydrogen bonds (**Figure 8D**). Other hydrogen bonds involved would be those formed by residues S1:R56, M2:R56, S3:V52, S3:F54, and E17:S42. Notably, the Δ1 loop S42 residue is a substrate of the serine kinase PRK2, and mutations at this position also affect RDRP activity and HCV replication (Han et al., 2014). Therefore, phosphorylation of S42 and T53, together with S29, could be destabilizing the interactions that allow the maintenance of an active NS5B structure. The large number of interactions between positions 1–6 and 52–56 could be responsible for maintaining the structure of this region of the NS5B protein and thus for the replication of the T53A mutant (**Figure 7**), albeit at much lower levels than the parental virus.

In summary, the positions described in this work are phosphorylated *in vitro* by Akt, a Ser/Thr kinase involved in many cellular processes. The positions identified are critical for NS5B RDRP activity, and Ser/Thr to Glu changes at these positions (imitating phosphorylation) as well as changes to Ala result in loss of HCV polymerase activity. Furthermore, mutants that mimic phosphorylation are unable to replicate in cell culture. The reason why HCV polymerase is a substrate for cellular kinases is currently unclear. The inactive NS5B protein product of phosphorylation might be able to interact with other viral or cellular proteins, thus modulating the viral and cellular replicative cycle. These lines of research are currently being explored in our laboratory.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

RS and AM: conceptualization, formal analysis, and writing—original draft preparation. RS, LA-G, BP, PC-C, ER, CG-C, IG, AÁ, and AM: investigation. RS, CP, ED, and AM: methodology. RS, LA-G, PC-C, and AM: writing—review and editing. CP, ED, and AM: funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

The proteomic analysis: Protein Identification by LC-MS/MS was carried out in the "CBMSO PROTEIN CHEMISTRY FACILITY" that belongs to ProteoRed, PRB3-ISCIII, supported by Grant PT17/0019. The work at UCLM was funded by MINISTERIO DE CIENCIA E INNOVACIÓN, Grant Nos. SAF2016-80451-P and PID2019-106068GB-I00 and UNIVERSIDAD DE CASTILLA-LA MANCHA Grant Nos. GI20163501, GI20174100, and 2019-GRIN-27080. BP is supported by predoctoral contract 2020-PREDUCLM-16195 from UCLM. The work at CBMSO and IIS-FJD was supported by grants SAF2017-87846-R and BFU2017-91384-EXP from Ministerio de Ciencia, Innovación y Universidades (MCIU), PI18/00210 from Instituto de Salud Carlos III, and S2018/BAA-4370 (PLATESA2 from Comunidad de Madrid/FEDER). CP was supported by the Miguel Servet program of the Instituto de Salud Carlos III (CPII19/00001), cofinanced by the European Regional Development Fund (ERDF). CG-C was supported by the predoctoral contract PRE2018-083422 from MCIU. CIBERehd (Centro de Investigación en Red de Enfermedades Hepáticas y Digestivas) was funded by the Instituto de Salud Carlos III. Institutional grants from the Fundación Ramón Areces and Banco Santander to the CBMSO are also acknowledged. The team at CBMSO belongs to the Global Virus Network (GVN).

ACKNOWLEDGMENTS

The Huh7.5 cell line was kindly supplied by Bartenschlager (University of Heidelberg). Dr. Sandra Franco is acknowledged for valuable discussions about generation of HCVp0 wt and mutant viruses. Dr. Piet de Groot is acknowledged for critical reading the manuscript. José Joaquín Lorenzo was acknowledged for technical assistance.

REFERENCES

Albentosa-Gonzalez, L., Sabariegos, R., Arias, A., Clemente-Casares, P., and Mas, A. (2021). Akt interacts with usutu virus polymerase, and its activity modulates viral replication. *Pathogens* 10:244. doi: 10.3390/pathogens10020244

Aloia, A. L., Locarnini, S., and Beard, M. R. (2012). Antiviral resistance and direct-acting antiviral agents for HCV. *Antivir Ther.* 17(6 Pt B), 1147–1162. doi: 10.3851/imp2426

Appleby, T. C., Perry, J. K., Murakami, E., Barauskas, O., Feng, J., Cho, A., et al. (2015). Viral replication. Structural basis for RNA replication by the hepatitis C virus polymerase. *Science* 347, 771–775.

Barik, S., McLean, T., and Dupuy, L. C. (1995). Phosphorylation of Ser232 directly regulates the transcriptional activity of the P protein of human respiratory syncytial virus: phosphorylation of Ser237 may play an accessory role. *Virology* 213, 405–412. doi: 10.1006/viro.1995.0013

Cai, Z., Yi, M., Zhang, C., and Luo, G. (2005). Mutagenesis analysis of the rGTP-specific binding site of hepatitis C virus RNA-dependent RNA polymerase. *J. Virol.* 79, 11607–11617. doi: 10.1128/jvi.79.18.11607-11617.2005

Chen, Z. W., Li, H., Ren, H., and Hu, P. (2016). Global prevalence of pre-existing HCV variants resistant to direct-acting antiviral agents (DAAs): mining the GenBank HCV genome data. *Sci. Rep.* 6:20310.

Chinnaswamy, S., Murali, A., Li, P., Fujisaki, K., and Kao, C. C. (2010). Regulation of de novo-initiated RNA synthesis in hepatitis C virus RNA-dependent RNA polymerase by intermolecular interactions. *J. Virol.* 84, 5923–5935. doi: 10.1128/jvi.02446-09

Chinnaswamy, S., Yarbrough, I., Palaninathan, S., Kumar, C. T., Vijayaraghavan, V., Demeler, B., et al. (2008). A locking mechanism regulates RNA synthesis and host protein interaction by the hepatitis C virus polymerase. *J. Biol. Chem.* 283, 20535–20546. doi: 10.1074/jbc.m801490200

Clemente-Casares, P., Lopez-Jimenez, A. J., Bellon-Echeverria, I., Encinar, J. A., Martinez-Alfaro, E., Perez-Flores, R., et al. (2011). De novo polymerase activity and oligomerization of hepatitis C virus RNA-dependent RNA-polymerases from genotypes 1 to 5. *PLoS One* 6:e18515. doi: 10.1371/journal.pone.0018515

de Chassey, B., Navratil, V., Tafforeau, L., Hiet, M. S., Aublin-Gex, A., Agaugue, S., et al. (2008). Hepatitis C virus infection protein network. *Mol. Syst. Biol.* 4:230.

Dolan, P. T., Zhang, C., Khadka, S., Arumugaswami, V., Vangeloff, A. D., Heaton, N. S., et al. (2013). Identification and comparative analysis of hepatitis C virus-host cell protein interactions. *Mol. Biosyst.* 9, 3199–3209. doi: 10.1039/c3mb70343f

Eden, J. S., Sharpe, L. J., White, P. A., and Brown, A. J. (2011). Norovirus RNA-dependent RNA polymerase is phosphorylated by an important survival kinase. *Akt. J. Virol.* 85, 10894–10898. doi: 10.1128/jvi.05562-11

Fourati, S., Rodriguez, C., Hezode, C., Soulier, A., Ruiz, I., Poiteau, L., et al. (2019). Frequent antiviral treatment failures in patients infected with hepatitis C virus genotype 4. Subtype 4r. *Hepatology* 69, 513–523. doi: 10.1002/hep.30225

Gane, E. J., Metivier, S., Nahass, R., Ryan, M., Stedman, C. A., Svarovskaia, E. S., et al. (2017). The emergence of NS5B resistance associated substitution S282T after sofosbuvir-based treatment. *Hepatol. Commun.* 1, 538–549. doi: 10.1002/hep4.1060

Hagen, N., Bayer, K., Rosch, K., and Schindler, M. (2014). The intraviral protein interaction network of hepatitis C virus. *Mol. Cell Proteomics* 13, 1676–1689. doi: 10.1074/mcp.m113.036301

Hamamoto, I., Nishimura, Y., Okamoto, T., Aizaki, H., Liu, M., Mori, Y., et al. (2005). Human VAP-B is involved in hepatitis C virus replication through interaction with NS5A and NS5B. *J. Virol.* 79, 13473–13482. doi: 10.1128/jvi.79.21.13473-13482.2005

Han, S. H., Kim, S. J., Kim, E. J., Kim, T. E., Moon, J. S., Kim, G. W., et al. (2014). Phosphorylation of hepatitis C virus RNA polymerases ser29 and ser42 by protein kinase C-related kinase 2 regulates viral RNA replication. *J. Virol.* 88, 11240–11252. doi: 10.1128/jvi.01826-14

Hernandez, S., Figueroa, D., Correa, S., Diaz, A., Aguayo, D., and Villanueva, R. A. (2015). Phosphorylation at the N-terminal finger subdomain of a viral RNA-dependent RNA polymerase. *Biochem. Biophys. Res. Commun.* 466, 21–27. doi: 10.1016/j.bbrc.2015.08.082

Hillung, J., Ruiz-Lopez, E., Bellon-Echeverria, I., Clemente-Casares, P., and Mas, A. (2012). Characterization of the interaction between hepatitis C virus NS5B and the human oestrogen receptor alpha. *J. Gen. Virol.* 93(Pt 4), 780–785. doi: 10.1099/vir.0.039396-0

Inoue, Y., Aizaki, H., Hara, H., Matsuda, M., Ando, T., Shimoji, T., et al. (2011). Chaperonin TRiC/CCT participates in replication of hepatitis C virus genome via interaction with the viral NS5B protein. *Virology* 410, 38–47. doi: 10.1016/j.virol.2010.10.026

Jakubiec, A., Tournier, V., Drugeon, G., Pflieger, S., Camborde, L., Vinh, J., et al. (2006). Phosphorylation of viral RNA-dependent RNA polymerase and its role in replication of a plus-strand RNA virus. *J. Biol. Chem.* 281, 21236–21249. doi: 10.1074/jbc.m600052200

Ji, H., Kozak, R. A., Biondi, M. J., Pilon, R., Vallee, D., Liang, B. B., et al. (2015). Next generation sequencing of the hepatitis C virus NS5B gene reveals potential novel S282 drug resistance mutations. *Virology* 477, 1–9. doi: 10.1016/j.virol.2014.12.037

Kulkarni, A. S., Damha, M. J., Schinazi, R. F., Mo, H., Doehle, B., Sagan, S. M., et al. (2016). A complex network of interactions between S282 and G283 of hepatitis C virus nonstructural protein 5b and the template strand affects susceptibility to sofosbuvir and ribavirin. *Antimicrob Agents Chemother.* 60, 2018–2027. doi: 10.1128/aac.02436-15

Kusakawa, T., Shimakami, T., Kaneko, S., Yoshioka, K., and Murakami, S. (2007). Functional interaction of hepatitis C Virus NS5B with Nucleolin GAR domain. *J. Biochem.* 141, 917–927. doi: 10.1093/jb/mvm102

Lam, A. M., Espiritu, C., Bansal, S., Micolochick Steuer, H. M., Niu, C., Zennou, V., et al. (2012). Genotype and subtype profiling of PSI-7977 as a nucleotide inhibitor of hepatitis C virus. *Antimicrob Agents Chemother.* 56, 3359–3368. doi: 10.1128/aac.00054-12

Liu, Z., Tian, Y., Machida, K., Lai, M. M., Luo, G., Foung, S. K., et al. (2012). Transient activation of the PI3K-AKT pathway by hepatitis C virus to enhance viral entry. *J. Biol. Chem.* 287, 41922–41930. doi: 10.1074/jbc.m112.414789

Lopez-Jimenez, A. J., Clemente-Casares, P., Sabariegos, R., Llanos-Valero, M., Bellon-Echeverria, I., Encinar, J. A., et al. (2014). Hepatitis C virus polymerase-polymerase contact interface: significance for virus replication and antiviral design. *Antiviral Res.* 108, 14–24. doi: 10.1016/j.antiviral.2014.04.009

Munakata, T., Nakamura, M., Liang, Y., Li, K., and Lemon, S. M. (2005). Down-regulation of the retinoblastoma tumor suppressor by the hepatitis C virus NS5B RNA-dependent RNA polymerase. *Proc. Natl. Acad. Sci. U S A.* 102, 18159–18164. doi: 10.1073/pnas.0505605102

Neufeldt, C. J., Cortese, M., Acosta, E. G., and Bartenschlager, R. (2018). Rewiring cellular networks by members of the flaviviridae family. *Nat. Rev. Microbiol.* 16, 125–142. doi: 10.1038/nrmicro.2017.170

Perales, C., Beach, N. M., Gallego, I., Soria, M. E., Quer, J., Esteban, J. I., et al. (2013). Response of hepatitis C virus to long-term passage in the presence of alpha interferon: multiple mutations and a common phenotype. *J. Virol.* 87, 7593–7607. doi: 10.1128/jvi.02824-12

Perales, C., Quer, J., Gregori, J., Esteban, J. I., and Domingo, E. (2015). Resistance of hepatitis c virus to inhibitors: complexity and clinical implications. *Viruses* 7, 5746–5766. doi: 10.3390/v7112902

Qadri, I., Choudhury, M., Rahman, S. M., Knotts, T. A., Janssen, R. C., Schaack, J., et al. (2012). Increased phosphoenolpyruvate carboxykinase gene expression and steatosis during hepatitis C virus subgenome replication: role of nonstructural component 5A and CCAAT/enhancer-binding protein beta. *J. Biol. Chem.* 287, 37340–37351. doi: 10.1074/jbc.m112.384743

Qian, X., Xu, C., Wu, B., Tang, H., Zhao, P., and Qi, Z. (2020). SNORD126 promotes hepatitis C virus infection by upregulating Claudin-1 via activation of PI3K-AKT signaling pathway. *Front. Microbiol.* 11:565590. doi: 10.3389/fmicb.2020.565590

Schmid, S., Mayer, D., Schneider, U., and Schwemmle, M. (2007). Functional characterization of the major and minor phosphorylation sites of the P protein of Borna disease virus. *J. Virol.* 81, 5497–5507. doi: 10.1128/jvi.02233-06

Sesmero, E., and Thorpe, I. F. (2015). Using the hepatitis C virus RNA-Dependent RNA polymerase as a model to understand viral polymerase structure. *Funct. Dynam. Viruses* 7, 3974–3994. doi: 10.3390/v7072808

Sheldon, J., Beach, N. M., Moreno, E., Gallego, I., Pineiro, D., Martinez-Salas, E., et al. (2014). Increased replicative fitness can lead to decreased drug sensitivity of hepatitis C virus. *J. Virol.* 88, 12098–12111. doi: 10.1128/jvi.01860-14

Suda, G., Kimura, M., Shigesawa, T., Suzuki, K., Nakamura, A., Ohara, M., et al. (2019). Effects of resistance-associated variants in genotype 2 hepatitis C virus on viral replication and susceptibility to antihepatitis C virus drugs. *Hepatol. Res.* 49, 1275–1285. doi: 10.1111/hepr.13401

Tabata, K., Neufeldt, C. J., and Bartenschlager, R. (2020). Hepatitis C virus replication. *Cold Spring Harb. Perspect. Med.* 10:a037093.

Valero, M. L., Sabariegos, R., Cimas, F. J., Perales, C., Domingo, E., Sanchez-Prieto, R., et al. (2016). Hepatitis C virus RNA-Dependent RNA polymerase interacts with the Akt/PKB kinase and induces its subcellular relocalization. *Antimicrob Agents Chemother* 60, 3540–3550. doi: 10.1128/aac.03019-15

Watashi, K., Ishii, N., Hijikata, M., Inoue, D., Murata, T., Miyanari, Y., et al. (2005). Cyclophilin B is a functional regulator of hepatitis C virus RNA polymerase. *Mol. Cell* 19, 111–122. doi: 10.1016/j.molcel.2005.05.014

Zhang, H., Zhang, C., Tang, H., Gao, S., Sun, F., Yang, Y., et al. (2018). CD2-Associated protein contributes to hepatitis C, virus propagation and steatosis by disrupting insulin signaling. *Hepatology* 68, 1710–1725. doi: 10.1002/hep.30073

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

REVIEW
published: 14 January 2022
doi: 10.3389/fmicb.2021.811671

Virus–Host Interplay Between Poly (ADP-Ribose) Polymerase 1 and Oncogenic Gammaherpesviruses

*Woo-Chang Chung and Moon Jung Song**

Virus-Host Interactions Laboratory, Department of Biotechnology, College of Life Sciences and Biotechnology, Korea University, Seoul, South Korea

The gammaherpesviruses, include the Epstein–Barr virus, Kaposi's sarcoma-associated herpesvirus, and murine gammaherpesvirus 68. They establish latent infection in the B lymphocytes and are associated with various lymphoproliferative diseases and tumors. The poly (ADP-ribose) polymerase-1 (PARP1), also called ADP-ribosyltransferase diphtheria-toxin-like 1 (ARTD1) is a nuclear enzyme that catalyzes the transfer of the ADP-ribose moiety to its target proteins and participates in important cellular activities, such as the DNA-damage response, cell death, transcription, chromatin remodeling, and inflammation. In gammaherpesvirus infection, PARP1 acts as a key regulator of the virus life cycle: lytic replication and latency. These viruses also develop various strategies to regulate PARP1, facilitating their replication. This review summarizes the roles of PARP1 in the viral life cycle as well as the viral modulation of host PARP1 activity and discusses the implications. Understanding the interactions between the PARP1 and oncogenic gammaherpesviruses may lead to the identification of effective therapeutic targets for the associated diseases.

Keywords: poly (ADP-ribose) polymerase 1 (PARP1), gammaherpesvirus replication, Epstein-Barr virus (EBV), Kaposi's sarcoma-associated herpesvirus (KSHV), murine gammaherpesvirus 68 (MHV-68), virus-host interaction, ADP-ribosyltransferase diphtheria-toxin-like 1 (ARTD1)

OPEN ACCESS

Edited by:
Rohit K. Jangra, LSU Health Sciences Center-Shreveport, United States

Reviewed by:
Hongyan Guo, Louisiana State University Health Science Center Shreveport, United States
Qiliang Cai, Fudan University, China

***Correspondence:**
Moon Jung Song
moonsong@korea.ac.kr

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *09 November 2021*
Accepted: *23 December 2021*
Published: *14 January 2022*

Citation:
Chung W-C and Song MJ (2022) Virus–Host Interplay Between Poly (ADP-Ribose) Polymerase 1 and Oncogenic Gammaherpesviruses. Front. Microbiol. 12:811671. doi: 10.3389/fmicb.2021.811671

INTRODUCTION

Post-translational modifications are important molecular mechanisms through which a cell regulates the diverse functions of proteins in various biological processes. Among the various cellular post-translational modifications, ADP-ribosylation constitutes a major modification. It involves the covalent attachment of the ADP-ribose unit from nicotinamide adenine dinucleotide (NAD +) onto a target protein. Depending on the acceptor proteins or catalyzing enzymes, ADP-ribosylation may involve the addition of a single unit or a polymer of ADP ribose units. The main proteins catalyzing ADP-ribosylation are PAR polymerases (PARPs) (Gupte et al., 2017). To date, 18 PARPs have been identified in the human genome, based on the conserved PARP motif (Amé et al., 2004; Jubin et al., 2016). Among these, PARP1 is the most abundant protein and has been widely studied as a major PARylating enzyme (Bai, 2015). PARP1 is involved in several cellular events, such as DNA repair, transcription, DNA replication, chromatin remodeling, energy metabolism, and cell death via PARylating target proteins (Kim et al., 2005; Ko and Ren, 2011; Gibson and Kraus, 2012; Luo and Kraus, 2012). Moreover, PARP1 is critical in viral infections, including the retroviruses, herpesviruses, influenza virus, hepatitis B virus, and chikungunya virus (Kameoka et al., 1999, 2005;

Dandri et al., 2002; Zhang et al., 2002; Ko and Ren, 2011; Grady et al., 2012; Bueno et al., 2013; Rom et al., 2015; Na et al., 2016; Xia et al., 2020).

The human gammaherpesviruses such as the Epstein–Barr virus (EBV) and Kaposi's sarcoma-associated herpesvirus (KSHV), are important pathogens associated with various tumors and proliferative diseases (Arvin et al., 2007; Jha et al., 2016). The murine gammaherpesvirus 68 (MHV-68) is genetically and biologically related to the human gammaherpesviruses. MHV-68 is extensively studied for elucidating the virus-host interactions and pathogenesis of gammaherpesviruses (Speck and Virgin, 1999; Virgin and Speck, 1999; Nash et al., 2001). These oncogenic herpesviruses display host cell tropisms, infecting epithelial, endothelial, fibroblastic, and lymphoblastic cells for their replication and establishing latency in lymphoid tissues, mainly T or B cells, through which they establish a persistent life-long infection in the hosts (Roizman et al., 1981). Interestingly, various host proteins have been identified to regulate gammaherpesvirus infections (Ye et al., 2011; Lee et al., 2012; Murata, 2014; Niller et al., 2014). Among these, PARP1 plays a key role in regulating the life cycle of the gammaherpesvirus. In this review, we have summarized the roles of PARP1 in the latent and lytic infection phases of the gammaherpesviruses, such as KSHV and EBV, and discussed how these oncogenic herpesviruses modulate PARP1 to promote their replication in the host. Furthermore, this review discusses the implications and the future perspectives of the activities of PARP1 on modulating these viruses.

PARP1: Biological Functions

PARPs or ADP-ribosyltransferase diphtheria-toxin-like (ARTDs) are enzymes that catalyze the transfer of the ADP-ribose unit from NAD^+ to specific residues in the target proteins, resulting in the addition of ADP-ribose polymers, a process termed as PARylation (Kawaichi et al., 1980; Ogata et al., 1980a,b). Therefore, the availability of the NAD^+ pool in the cells is a key factor for the activity of PARPs. PARP1 is the founding member among the 18 human PARPs. This highly conserved eukaryotic nuclear protein is known to play a role in DNA damage response, chromatin modification, transcriptional regulation, inflammation, and cell death (Gibson and Kraus, 2012). PARP1 mediates approximately 90% of the cellular PARylation in response to DNA damage (Bai, 2015). PARP1 is responsible for the majority of cellular PARP activity, followed by PARP2, while the activities of other PARPs seem to be negligible in comparison (Bai, 2015). The PARylated proteins can be reversely hydrolyzed by PAR glycohydrolase (PARG), ADP-ribosylhydrolase 3 (ARH3), or O-acyl-ADP-ribose deacylase 1 (OARD1) to remove the PAR chain from the target protein and catabolize ADP-ribose (Oka et al., 1984, 2006; Davidovic et al., 2001; Rosenthal et al., 2013; Sharifi et al., 2013). A schematic diagram for addition or removal of PAR to an acceptor protein is shown in **Figure 1A**.

PARP1 contains three domains, an N-terminal DNA-binding domain (DBD), a central auto-modification domain (AD), and a C-terminal catalytic domain (CAT) (Alvarez-Gonzalez et al., 1999; Affar et al., 2001; Kim et al., 2005; **Figure 1B**). The DBD contains three zinc-finger motifs (ZnFI, ZnFII, and ZnFIII) and a nuclear localization sequence and are responsible for recognizing the DNA breaks such as a single-strand break (SSB) or double-strand break (DSB) (Caldecott et al., 1996; Langelier et al., 2008, 2011, 2012). The AD includes the BRCA1 C-terminus (BRCT) motif, flanked by lysine and glutamate residues as auto-PARylation sites (Altmeyer et al., 2009; Tao et al., 2009). The BRCT motif mediates interactions with other proteins involved in various cellular pathways (Masson et al., 1998; Beernink et al., 2005; Cuneo et al., 2011). The CAT domain contains a helical WGR domain and an ADP-ribosyltransferase (ART) domain that is conserved in the PARP family proteins (Hottiger et al., 2010). It is responsible for catalyzing the addition of ADP-ribose units onto target proteins and their subsequent elongation and branching (Mendoza-Alvarez and Alvarez-Gonzalez, 1993). PARP1 has been reported to interact with and/or PARylate various cellular factors including PARP1 itself to modulate their functions (Gibson and Kraus, 2012; Ciccarone et al., 2017; Kamaletdinova et al., 2019). Although domain-mapping studies for protein-protein interactions using domain mutants of PARP1 have not been done for all the interacting proteins, representative binding partners for its individual domains are listed and marked with the associated cellular activities in **Figure 1B** (Kraus and Lis, 2003; Glover et al., 2004; Strosznajder et al., 2005; Li et al., 2006; Kanno et al., 2007; Hassa and Hottiger, 2008; Kotoglou et al., 2009; Krishnakumar and Kraus, 2010; Chang et al., 2011; Rodríguez et al., 2011; Gibson and Kraus, 2012; Ko and Ren, 2012; Mangerich and Bürkle, 2012; Xie et al., 2015; Bonfiglio et al., 2017; Ray Chaudhuri and Nussenzweig, 2017; Alemasova and Lavrik, 2019; Kamaletdinova et al., 2019; Ke et al., 2019; Fehr et al., 2020; Demény and Virág, 2021).

PARP1 and Viral Infections

PARP1 plays important roles in infections of viruses with DNA genome. For example, PARP1 binds to the hepatitis B virus (HBV) core promoter and enhances viral gene transcription and HBV replication (Ko and Ren, 2011). A study on PARP1 inhibitors suggested that PARP1 activity may limit the integration of the HBV DNA (Dandri et al., 2002). As an essential factor for HBV replication, HBx interacts with PARP1 and inhibits the recruitment of the DNA repair machinery to the damaged DNA sites, thereby promoting hepatocarcinogenesis (Na et al., 2016). In vaccina virus infection, PARP1 plays a role in NK cell migration to the site of infection and promotes CCL2 production (Shou et al., 2019). Herpes simplex virus 1 degrades the poly (ADP-ribose) glycohydrolase (PARG) using the viral E3-ligase, ICP0 and activates PARP1 to facilitate virus replication (Grady et al., 2012).

PARP1 activity also regulates the replication of many RNA viruses. In human immunodeficiency virus 1 (HIV-1) infection, PARP1 activation increases the integration of the HIV-1 genome into the host chromosome (Ha et al., 2001; Kameoka et al., 2005). PARP-1 activity also increased long terminal repeat (LTR)-mediated viral gene transcription (Kameoka et al., 1999; Rom et al., 2015). In contrast, there are reports showing a negative role of PARP-1 in viral gene transcription (Parent et al., 2005; Bueno et al., 2013). For human T-lymphotropic virus 1 (HTLV-1), PARP1 activates transcription of Tax binding elements

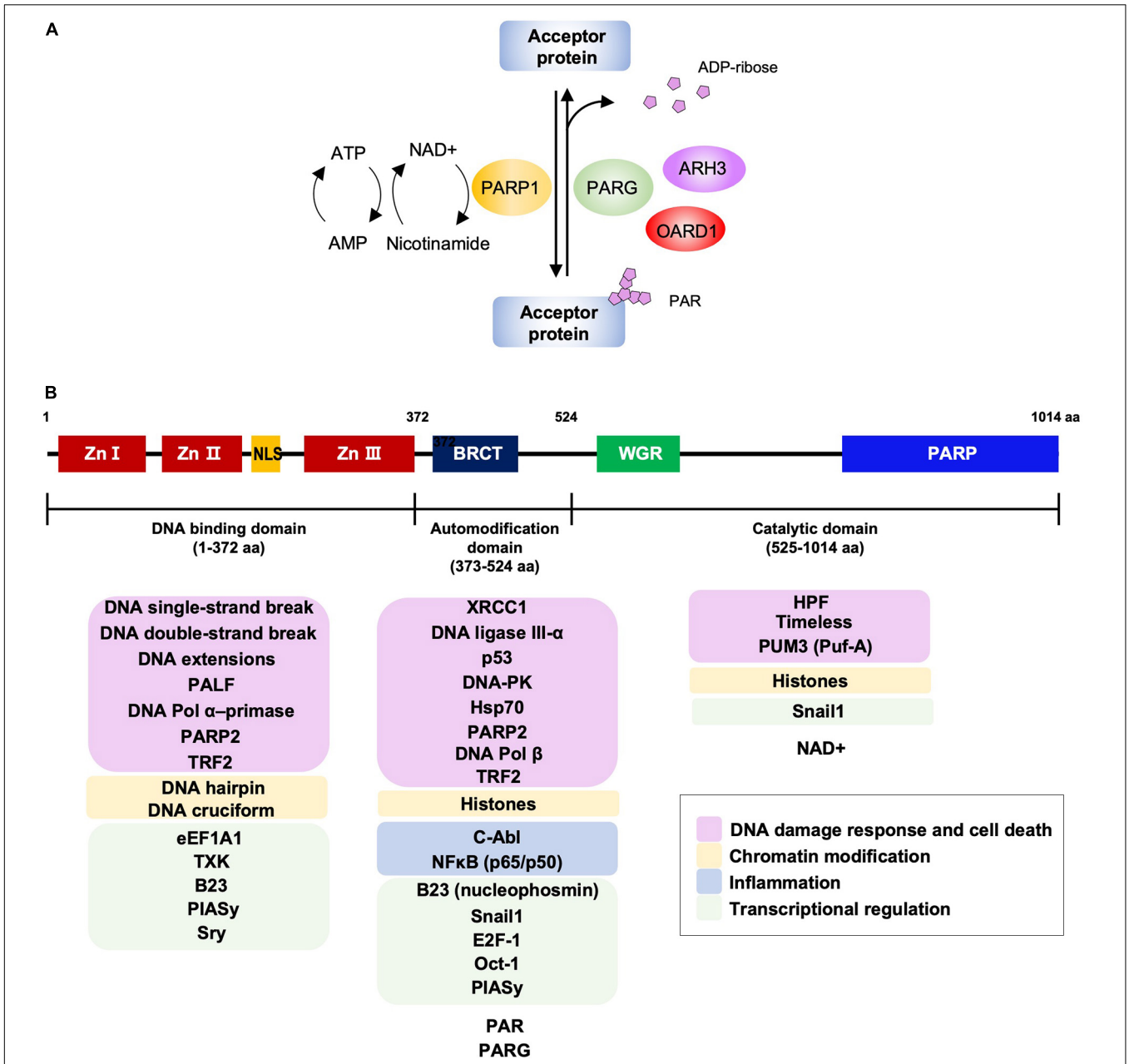


FIGURE 1 | PARP1-mediated PARylation and PARP1 structural domains with interacting partners. **(A)** The PARylation reaction by PARP1. The activated PARP1 hydrolyses NAD^+ and catalyzes the transfer of ADP-ribose moiety onto target residues of the acceptor protein (PARylation). The PARylated proteins are involved in various biological processes, including DNA repair, cell death, chromatin remodeling, inflammation, and transcription. PAR chains from the PARylated acceptor proteins are rapidly catabolized by PAR glycohydrolase (PARG), ADP-ribosylhydrolase 3 (ARH3), and O-acyl-ADP-ribose deacylase 1 (OARD1). NAD^+, nicotinamide adenine dinucleotide. **(B)** Schematic representation of PARP1 domains and interaction partners. PARP1 contains three domains; the N-terminal DNA binding domain that has three zinc finger motifs with DNA binding activity and an NLS that is responsible for the localization of PARP1 into the nucleus; the auto modification domain that has a BRCT domain and is responsible for its auto-PARylation; the catalytic domain that contains a homologous PARP family motif and catalyzes PARylation of the target proteins. According to the studies using PARP1 domain mutants, PARP1 interacting partners for each domain are listed and marked for the associated cellular activities. Zn I, Zn II, and Zn III: zinc finger motifs; NLS, nuclear localization signal; BRCT, BRCA1 C-terminus motif; WGR, tryptophan-glycine-arginine-rich domain; PARP, PARP activity domain.

(Zhang et al., 2002). In Influenza A virus, hemagglutinin (HA) interacts with PARP1 and degrades type I interferon receptor (IFNAR), thereby enhancing virus replication (Xia et al., 2020). PARP1 interacts with the nucleocapsid (N) protein of porcine reproductive and respiratory syndrome virus (PRRSV) and its activity promotes the virus replication (Liu et al., 2015). Sindbis virus (SINV), belonging to the *Alphavirus,* activates PARP1 during infection (Ubol et al., 1996; Nargi-Aizenman et al., 2002).

PARP1 is found in the SINV replication complex and interacts with nsP3, suggesting its positive role in the replication of SINV (Park and Griffin, 2009). Overall effects of PARP1 on regulation of viral replication can be either proviral or antiviral, while the effects appear to be both in HIV-1 infection.

PARP1 IN EBV INFECTION

EBV, a member of *Lymphocryptovirus* genus, is the first discovered human oncovirus that is associated with Burkitt's lymphoma, infectious mononucleosis, Hodgkin's disease, nasopharyngeal carcinoma, gastric carcinoma, and various other lymphomas (Rickinson, 2014). It has a worldwide prevalence with over 95% of the adults being seropositive for EBV (Luzuriaga and Sullivan, 2010). As it spreads primarily through saliva, most of the individuals are infected during infancy and early childhood asymptomatically or with non-specific symptoms.

Life Cycle of EBV

EBV is known to spread mainly through body fluids, such as saliva, blood, and semen, while organ transplantation is also a potential transmission route (Chang et al., 2009). After primary infection, the virus remains latent in the individuals for the rest of their lives. EBV establishes latency mainly in the B lymphocytes and is reactivated intermittently by the induction of lytic replication (Dugan et al., 2019). A subset of viral genes is expressed to maintain latency that causes proliferation of the infected B lymphocytes. EBV-infected cells exhibit three latency programs, latency I, II, or III. Latency I restricts the expression to that of the most latent genes, while latency III involves the expression of a full repertoire of the known latent genes. During latency III, the EBV nuclear antigens (EBNAs 1, 2, 3A, 3B, 3C, and LP), latent membrane proteins (LMPs 1, 2A, and 2B), and viral non-coding RNAs (EBERs, miRNAs, and BARTs) are expressed, establishing the lymphoblastoid cell lines (Rickinson, 2002). The selective pressure from the EBV-specific cytotoxic lymphocytes leads to the transition from latency III to more restricted forms of latency. Latency II is characterized by the lack of *EBNA2* and *EBNA3* expression but exhibits the expression of other latent genes (Price and Luftig, 2015). *EBNA1* is the only viral gene expressed in latency I. In latency 0, the viral genome persists in the host cells without viral gene expression, which is associated with infection in the non-dividing memory B-cells (Dugan et al., 2019). When EBV reactivation is induced, the immediate-early (IE) genes, *BZLF1* and *BRLF1* are transcribed to produce Zta (also called ZEBRA or EB1) and Rta (also called R), respectively. After the synthesis of these transcription factors, the downstream lytic genes, including their promoters, are activated, resulting in a full cascade of lytic gene expression. The early (E) genes encode proteins required for DNA replication and metabolism, such as the viral DNA polymerase, viral DNA primase, processivity factor, and thymidine kinase. Following viral DNA replication, the late (L) genes are transcribed for the synthesis of structural proteins to assemble the virion particles (Young et al., 2007).

Roles of PARP1 in EBV Infection

PARP1 is involved in the regulation of EBV latent infection (**Figure 2A**). During latency, the replication of viral DNA depends on the interaction between EBNA1 and origin of plasmid replication (*OriP*) in the virus (Kennedy and Sugden, 2003; Lindner and Sugden, 2007). PARP1 was identified as one of the cellular factors that bind to the dyad symmetry (DS) elements of the *OriP*, the site for EBNA1 binding (Deng et al., 2002, 2005; Tempera et al., 2010). Treatment of D98 cells with the PARP inhibitors, such as niacinamide (NA) and 3-aminobenzamide (3-ABA), resulted in a significant increase in the *OriP* plasmid maintenance after 3 weeks of treatment (Deng et al., 2002, 2005; Tempera et al., 2010). In contrast, a ribonucleotide reductase inhibitor, hydroxyurea (HU), that increases PARP1 activity, caused a substantial loss of the *OriP* plasmids in these cells, which was consistent with a previous report that HU accelerated the loss of EBV genomes from Akata cells (Srinivas et al., 1998; Deng et al., 2002, 2005; Tempera et al., 2010). Mechanistically, PARP1 was shown to induce PARylation of EBNA1, which reduced the binding affinity of EBNA1 to the DS of *OriP* (Tempera et al., 2010). Studies on PARP1 knockdown and treatment with PARP1 inhibitors further suggest that PARP1 suppresses EBNA1 binding and the recruitment of origin recognition complex 2 (ORC2) onto *OriP*, leading to a reduction in the replication of viral DNA during latency (Tempera et al., 2010). In addition, PARP1 is colocalized with the CCCTC-binding factor (CTCF), a host insulator protein that binds to specific sites throughout the EBV genome including the latency promoter Cp (Lupey-Green et al., 2018). PARP1 PARylates CTCF and studies involving PARP1 inhibitors showed that PARP1 stabilizes CTCF binding and maintains the open chromatin structure of the EBV genome at the active *Cp* promoter during type III latency (Lupey-Green et al., 2018). Taken together, these results suggest that PARP1 and PARylation of EBNA1 and CTCF are important mechanisms regulating the viral episome during EBV latency.

A recent study indicated the involvement of PARP1 in the lytic cycle of the EBV replication. PARP1 binds to the *BZLF1* promoter in the type I and type III latently infected B-cells and prevents binding of Zta to its promoter for autoregulation, thereby limiting the EBV reactivation (Lupey-Green et al., 2017). In another study, treatment with the PARP1 inhibitor, 3-ABA, during EBV reactivation increased early antigen expression (Mattiussi et al., 2007). However, PARP1 inhibition increased LMP1 and EBNA2 expression, but decreased the expression of BFRF1, a nuclear egress protein, reducing the overall virion production in the culture media (Mattiussi et al., 2007). These results suggest that PARP1 enzymatic activity may play a role in the progression of the EBV lytic cycle.

Viral Modulation of PARP1 in EBV Infection

In EBV-associated tumorigenicity, LMP1 employs PARP1 for epigenetic control. LMP1 expressed in type II and III latency programs is the major transforming protein that is critical for EBV-induced B-cell transformation and modulate several

processes, such as differentiation, cell migration and survival, and tumorigenicity (Li and Chang, 2003; Shair et al., 2012; Liu et al., 2018). Martin et al. (2016) reported that LMP1 hijacks PARP1 to enhance the cellular PARylation level and thus enhances cellular transformation. They showed that PARP1 suppressed the expression of enhancer of zeste 2 polycomb repressive complex 2 subunits (EZH2), a histone methyltransferase, and a catalytic component of the initiation complex, polycomb repressive complex 2 (PRC2); this, in turn, reduced the level of the repressive histone marker, trimethylation of lysine 27 on histone H3 (H3K27me3), and induced the expression of LMP1-related protooncogenes *c-Fos* and *EGR1* (Martin et al., 2016). Hulse et al. (2018) showed that LMP1 activated PARP1 to increase the hypoxia-inducible factor 1-alpha (HIF-1α)-dependent gene expression. In latently infected cells, LMP1-activated PARP1 acts as a coactivator of HIF-1α, forming the PARP1–PARylated–HIF-1α complex, as depicted in **Figure 2A**. This complex bound to the promoters of genes downstream of *HIF-1α* and increased their expression, resulting in altered cell metabolism and a switch from mitochondrial respiration to a glycolytic "Warburg" metabolism (Hulse et al., 2018).

EBV suppresses the inhibitory effect of PARP1 to achieve successful lytic replication (**Figure 3A**). Upon reactivation, PARP1 was downregulated and *BZLF1*-encoded Zta overexpression reduced PARP1 levels to some extent, which contributes to the easy accessibility of *BZLF1* promoter. This makes Zta bind to *BZLF1* promoter and facilitates lytic replication (Lupey-Green et al., 2017). In KSHV and MHV-68, viral processivity factors were shown to downregulate PARP1 in a proteosome-dependent manner. In EBV infection, BMRF1 gene-encoded EA-D primarily acts as a DNA polymerase processivity

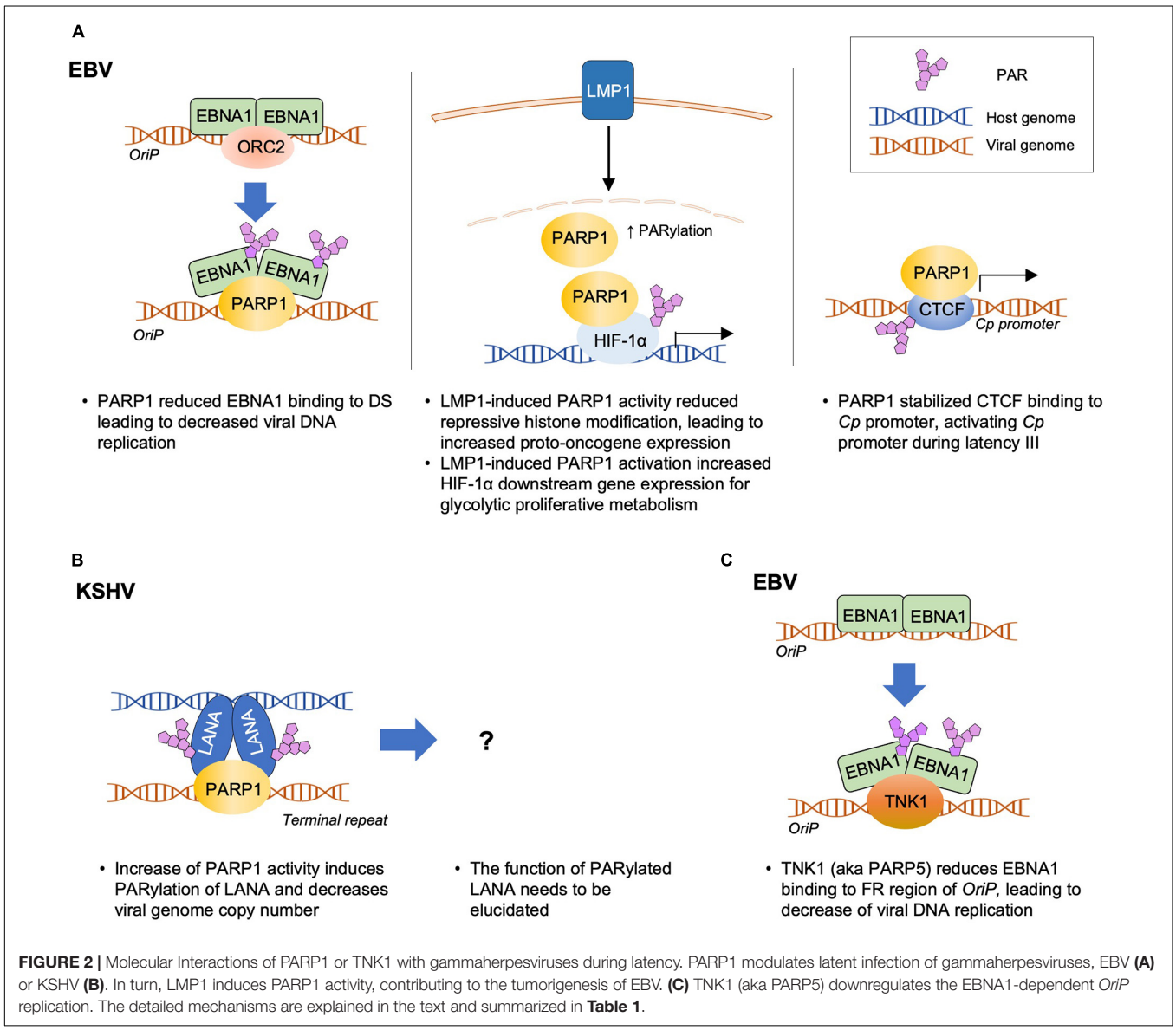


FIGURE 2 | Molecular Interactions of PARP1 or TNK1 with gammaherpesviruses during latency. PARP1 modulates latent infection of gammaherpesviruses, EBV **(A)** or KSHV **(B)**. In turn, LMP1 induces PARP1 activity, contributing to the tumorigenesis of EBV. **(C)** TNK1 (aka PARP5) downregulates the EBNA1-dependent *OriP* replication. The detailed mechanisms are explained in the text and summarized in **Table 1**.

factor as a component of the viral DNA replication complex to facilitate DNA synthesis (Kiehl and Dorsky, 1991; Tsurumi, 1993; Neuhierl and Delecluse, 2006). It will be intriguing to find out whether EBV EA-D may downregulate PARP1 in a similar way, suggesting a conserved viral mechanism for PARP1 downregulation among gammaherpesviruses.

PARP1 IN KSHV AND MHV-68 INFECTIONS

After its identification as an infectious agent of Kaposi's sarcoma (KS), KSHV was confirmed as the eighth member of the human herpesviruses (HHV-8) and classified as a member of the

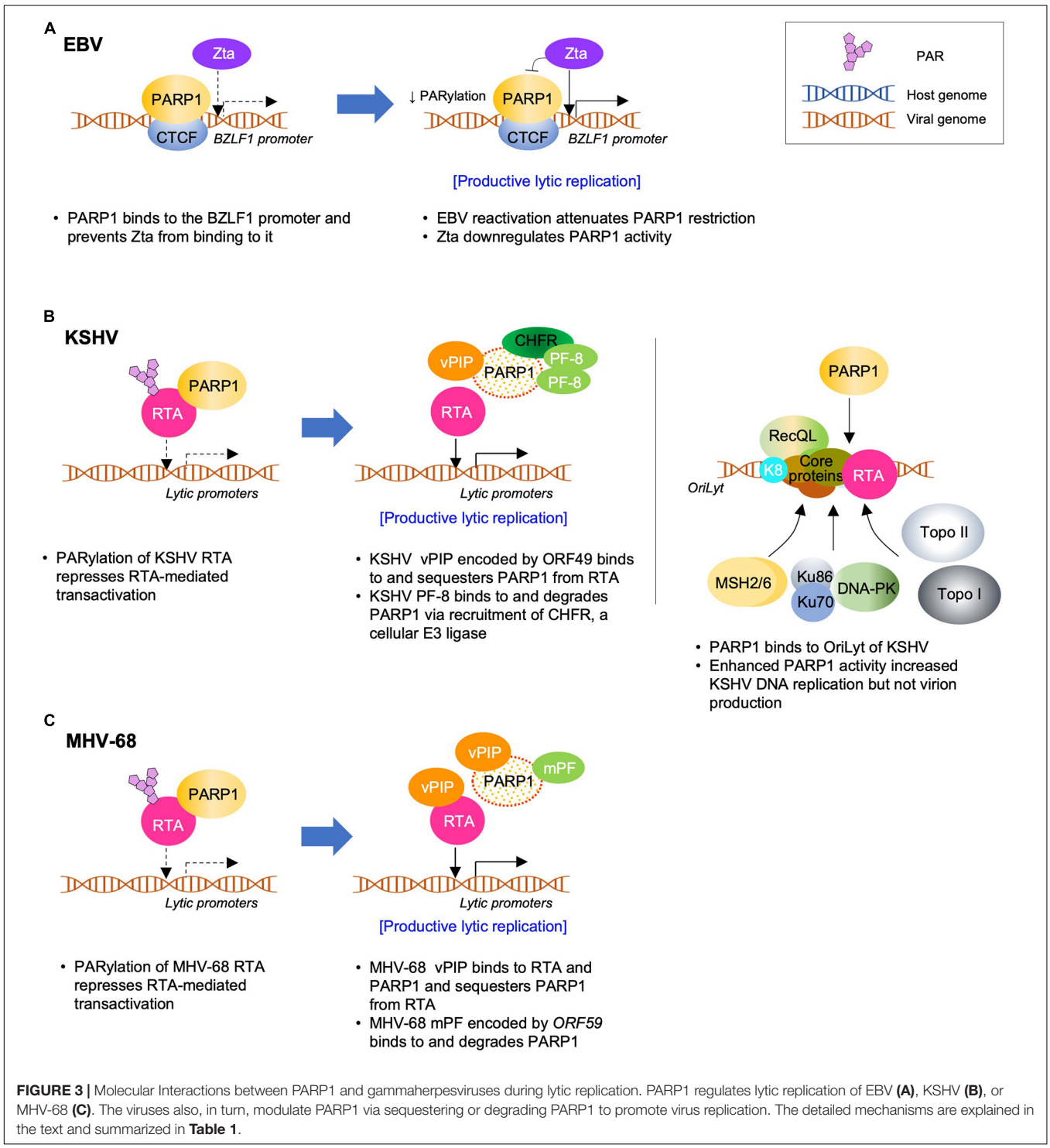


FIGURE 3 | Molecular Interactions between PARP1 and gammaherpesviruses during lytic replication. PARP1 regulates lytic replication of EBV **(A)**, KSHV **(B)**, or MHV-68 **(C)**. The viruses also, in turn, modulate PARP1 via sequestering or degrading PARP1 to promote virus replication. The detailed mechanisms are explained in the text and summarized in **Table 1**.

Rhadinovirus genus of the gammaherpesvirus subfamily, based on its genome sequence and homology analysis (Chang et al., 1994; McGeoch and Davison, 1999). In addition to KS, KSHV is associated with primary effusion lymphoma (PEL), multicentric Castleman's disease (MCD), body-cavity-based lymphoma, and other lymphoproliferative diseases (Di Alberti et al., 1997; Rettig et al., 1997; Ganem, 2007). Epidemiology showed that 50% of the adults in sub-Saharan Africa and less than 5% of the adults in northern Europe, North America, and Asia are seropositive for KSHV (Simpson et al., 1996; Calabro et al., 1997; Whitby and Boshoff, 1998). Similar to EBV, saliva is the main transmission route for KSHV, but sexual contact, blood transfusion, and organ transplantation are also considered important routes (Mayama et al., 1998; Corey et al., 2002; Henke-Gendo and Schulz, 2004; Martro et al., 2004). MHV-68 (also known as γHV-68), another member of the *Rhadinovirus* genus, is homologous to EBV and KSHV in terms of genome organization and pathogenesis and is therefore considered an excellent small animal model to study virus-host interactions and viral pathogenesis of human gammaherpesviruses (Simas and Efstathiou, 1998). MHV-68 also provides an amenable genetic system to investigate viral gene functions *in vitro* and *in vivo*.

Life Cycle of KSHV and MHV-68

After primary infection of the permissive cells, KSHV enters the latent phase and only expresses the latency-associated genes without productive lytic replication. Unlike EBV, KSHV does not induce immortalization of the B-cells nor direct transformation of any infected cells (Myoung and Ganem, 2011). Moreover, there are no diverse latency programs in KSHV, albeit a subset of latent genes may be differentially expressed depending on the cell type. During latency, only a few genes are expressed; these include the *open reading frame (ORF)73* encoding the latency-associated nuclear antigen (LANA or LNA), *ORF72* encoding a viral cyclin D homolog (vCyclin), and *ORF71* encoding a viral Fas-associated protein with death domain-like interleukin-1β-converting enzyme/caspase-8-inhibitory protein (vFLIP), as well as those encoding K12, viral interferon regulatory factors (vIRFs), and viral miRNAs (Cai et al., 2010). LANA is a major latent protein expressed in KSHV-positive malignancies serving as a hallmark for KSHV latency (Rainbow et al., 1997; Gao et al., 1999). It is necessary for DNA replication, maintenance, and segregation of the KSHV episomes during host cell mitosis (Hu et al., 2002; Purushothaman et al., 2016). LANA also acts as a multifunctional nuclear protein, interacting with various cellular proteins involved in tumorigenesis, cellular transcription, and chromatin remodeling (Purushothaman et al., 2016; Wei et al., 2016). When reactivation is induced in the KSHV latently-infected cells, transcription cascade of viral lytic genes activates transcription of the *IE* genes, followed by *E* and *L* genes, leading to the production of infectious virions (Sun et al., 1998; Gradoville et al., 2000; Lukac and Yuan, 2007; Schulz and Chang, 2007; Cai et al., 2010). Notably, the replication and transcription activator (RTA), mainly encoded by *ORF50*, acts as a key switch molecule to induce the expression of lytic genes (Lukac et al., 1998, 1999; Sun et al., 1998; Gradoville et al., 2000; Xu et al., 2005). Several cellular factors including PARP1 regulate the RTA activity either negatively or positively, thereby regulating the virus life cycle (Li et al., 2014; Aneja and Yuan, 2017; Yan et al., 2019).

The life cycle of MHV-68 is similar to that of KSHV in many aspects with some discrepancies. *De novo* infection of MHV-68 results in robust lytic replication in the fibroblasts and epithelial cells (Wu et al., 2000; Rochford et al., 2001). MHV-68 RTA, mainly encoded by *ORF50*, is also essential for activating the cascade of downstream lytic gene expression and inducing lytic replication (Liu et al., 2000; Wu et al., 2000). Moreover, MHV-68 LANA, a homolog of the KSHV LANA, plays an important role in establishing and maintaining viral latency (Fowler et al., 2003; Moorman et al., 2003; Habison et al., 2012).

Roles of PARP1 in KSHV and MHV-68 Infections

PARP1 affects the latent infection of KSHV (**Figure 2B**). Together with other known replication factors, such as ORC2, CDC6, and Mcm7, PARP1 was identified as a cellular factor that binds to the terminal repeat (TR) of the KSHV genome (Ohsaki et al., 2004). Although PARP1 was colocalized with LANA and induced its PARylation, PARP1 binding to a specific region within the TR was independent of LANA, which leaves the function of PARylated LANA to be elucidated. Nevertheless, NA (a chemical inhibitor of PARP1) treatment increased the viral genome copy number in BC-3 cells, whereas HU (a PARP1 activator) treatment decreased it, suggesting that PARP1 may play a role in maintaining KSHV latency. PARP1 and LANA interaction was also confirmed using the affinity purification method using the N-terminus of LANA (Barbera et al., 2006a). Although Ku70, Ku80, and PARP1 were found to be interacting partners in this study, neither of these factors have been shown to mediate LANA chromosome association (Barbera et al., 2006a). However, the binding of PARP1 to the TR might play a role in genome circularization and latent genome maintenance associated with LANA. Further studies are needed to identify this role (Barbera et al., 2006b).

PARP1 is also an important cellular factor that negatively regulates lytic replication of KSHV (**Figure 3B**). PARP1 suppresses the KSHV RTA activity by directly binding to and PARylating RTA (Gwack et al., 2003). RTA PARylation decreases the RTA-mediated transcriptional activity by inhibiting the recruitment of RTA to the promoters of lytic genes. Gwack et al. (2003) showed that PARP1 and Ste20-like kinase hKFC interacted with the serine/threonine-rich region of RTA. These two factors synergistically enhanced their interactions with RTA as well as post-translational modifications of RTA, such as PARylation and phosphorylation, thereby acting as strong repressors of the RTA activity on lytic gene expression (Gwack et al., 2003). Using mutant RTA constructs, Ko et al. (2012) showed that the Thr-366 and Thr-367 residues of RTA formed the primary motif for O-GlcNAcylation *in vivo*, which plays a role in the recruitment of PARP1 to RTA. In contrast, PARP1 reportedly plays a positive role in regulating DNA replication during the lytic phase. Using DNA affinity purification, Wang et al. (2008) identified topoisomerases (Topo) I and II, MSH2/6, RecQL, DNA-PK, Ku86/70, scaffold attachment factor A (SAF-A), and PARP1 as KSHV *oriLyt*-bound proteins. Inhibition of

PARP1 using chemical inhibitors (3-ABA and NA) resulted in decreased *oriLyt*-dependent DNA replication, whereas HU increased PARP1 activity and DNA replication, suggesting a positive role for PARP1 during the lytic replication of KSHV (Wang et al., 2008). Although the roles of PARP1 in KSHV lytic replication were seemingly inconsistent, Wang et al. (2008) also showed that the overall effect of PARP1 lies in reducing KSHV virion production; the result of lytic replication, suggesting that viral DNA replication may not be the rate-limiting step in virion production and that PARP1-mediated inhibition of RTA activity may be the key step that regulates the KSHV life cycle. Interestingly, the inhibitory effect of PARP1 on virus lytic replication is also conserved in MHV-68 (Gwack et al., 2003). PARP1 reduces the transcriptional activity of MHV-68 RTA and virus lytic replication during *de novo* infection (**Figure 3C**).

Viral Modulation of PARP1 in KSHV and MHV-68 Infections

To counteract the inhibitory effect of PARP1 on virus lytic replication, KSHV and/or MHV-68 have been shown to employ two viral proteins during different stages of the life cycle (**Figures 3B,C**). The MHV-68 *ORF49*-encoded protein, a tegument protein, interacts with and sequesters PARP1 from RTA binding (Noh et al., 2012; Chung et al., 2018). The purified recombinant ORF49 protein from MHV-68 and KSHV interacts directly with PARP1 without any additional cellular factors and is therefore termed viral PARP1 interacting protein (vPIP) (Noh et al., 2012; Chung et al., 2018). vPIP enhances RTA-mediated transactivation by reducing the level of PARylated RTA. Owing to its association with the virion, vPIP is likely to modulate PARP1 function during the early phase of *de novo* infection. Consistent with this hypothesis, the viral growth of a recombinant MHV-68 with transposon or triple stop codons inserted at the ORF49 locus is significantly attenuated *in vitro* and *in vivo*. Furthermore, vPIP residues that are critical for PARP1 interaction were identified based on the vPIP crystal structure (Chung et al., 2018). A recombinant MHV-68 harboring mutations at the three critical residues of vPIP exhibited defective PARP1 interaction and was highly attenuated in viral growth both *in vitro* and *in vivo*, suggesting the significant role of vPIP and PARP1 interaction in derepressing RTA for viral replication.

Other studies have reported that KSHV and MHV-68 downregulate PARP1 in a proteosome-dependent manner during lytic replication (Chung et al., 2015, 2021). KSHV *ORF59* encodes the viral processivity factor, PF-8, which interacts with and degrades PARP1, thereby enhancing the RTA-mediated transactivation, especially on the *RTA* promoter (Chung et al., 2015, 2021). PARP1 degradation is dependent on PF-8 interaction and recruitment of a cellular ubiquitin E3-ligase, checkpoint with FHA and RING finger domains (CHFR) (Chung et al., 2015, 2021). As the viral processivity factor is expressed as an early lytic gene following RTA expression, the PF-8-induced degradation of PARP1 may be important in reducing the inhibitory effect of PARP1 on RTA and further reinforcing the RTA positive feedback to promote lytic replication.

OTHER PARPs THAT REGULATE GAMMAHERPESVIRUSES

In addition to PARP1, other PARPs are reportedly involved in gammaherpesvirus replication. Tankyrase 1 (TNK1; PARP5) uses NAD^+ to PARylate the acceptors by recognizing the RxxPDG motif (TNKS-binding motif) (Smith et al., 1998; Sbodio and Chi, 2002). TNK1, TNK2 (PARP6), and telomeric repeat binding factor 2 (TRF2) interact with *OriP* in an EBNA1-dependent manner, as determined by the DNA affinity purification assay (Deng et al., 2002). Like PARP1, TNK1 PARylates EBNA1 and downregulates the EBNA1-dependent *OriP* replication (Deng et al., 2005; Tempera et al., 2010), as shown in **Figure 2C**. Interactome studies have identified interactions between PARPs and viral proteins. A yeast two-hybrid screening for EBV viral proteins with a human cDNA library showed that EBV BRRF1, a homolog of KSHV and MHV-68 vPIP interacts with PARP4 (Calderwood et al., 2007). Furthermore, an interactome study with KSHV ORF libraries using affinity purification/LC-MS has identified PARP2 as an interaction partner of LANA (Davis et al., 2015). However, further studies should be warranted to confirm their genuine interactions and to elucidate the significance of these interactions in the context of virus life cycle.

DISCUSSION

PARP1 plays a role in various cellular mechanisms, such as DNA damage repair, cell death, proliferation, differentiation, gene transcription, and inflammation (Kim et al., 2005; Ko and Ren, 2011; Gibson and Kraus, 2012; Luo and Kraus, 2012). Additionally, it plays a role in oncogenic gammaherpesvirus infection and modulates the viral life cycle in either a positive or negative manner. **Table 1** summarizes the effects of PARP1 on the life cycle of gammaherpesviruses, while **Figures 2**, **3** show. Although the overall effects of PARP1 on the virus life cycle either promote or inhibit the lytic replication or latency, there are a few discrepancies regarding the roles of PARP1, with reports of both positive and negative effects on the distinct stages of the life cycle of the virus. For example, several studies reported that PARP1 exerts a negative effect on the lytic replication of KSHV (Gwack et al., 2003; Wang et al., 2008; Lupey-Green et al., 2017), while a few reported its positive effect on the lytic replication of KSHV or downregulation of latent genome (Ohsaki et al., 2004; Wang et al., 2008). These discrepancies may be attributed to the differences in the experimental methods employed in individual studies. Moreover, the PARP1 inhibitor, 3-ABA is also a potent inhibitor of apoptosis, which makes it difficult to interpret the contradicting results. In addition, the virus may potentially utilize or overcome the effects of PARP1 depending on the stage of its life cycle. Furthermore, the mode of action of PARP1 can be either dependent or independent of its enzymatic activity, depending on the target proteins or DNA. Therefore, these factors should be considered when attempting to elucidate the roles

TABLE 1 | Summary of PARP1 actions and virus interactions on regulation of gammaherpesvirus life cycle.

Overall outcome	Life cycle	Virus	Effects of PARP1 and virus interactions	References
Positive	Latency	EBV	• LMP1 interacts with PARP1 and enhances the cellular PARylation levels and cellular transformation. • PARP1 suppresses the EZH2 expression, reducing repressive histone modification, and induces the LMP1-related protooncogene expression. • The LMP1-induced PARP1 activation increases the HIF-1α-dependent gene expression. • The PARylated HIF-1α–PARP1 complex binds to promoters of the *HIF-1α* downstream genes, activating their expression and altering the cellular metabolism.	Martin et al., 2016; Hulse et al., 2018
			• PARP1 is colocalized with CTCF, stabilizing its binding to latency *Cp* promoter, thereby activating *Cp* promoter during latency III.	Lupey-Green et al., 2018
	Lytic infection	EBV	• Treatment with a PARP1 inhibitor increases EA expression, but decreases expression of BFRF1, a nuclear egress protein. • Inhibition of PARP1 activity reduces the overall virion production. • Inhibition of PARP1 activity increases LMP1 and EBNA2 expression.	Mattiussi et al., 2007
		KSHV	• PARP1 binds to KSHV *ori-Lyt* DNA. • Increased PARP1 activity enhances the *ori-Lyt* DNA replication.	Wang et al., 2008
Negative	Latency	EBV	• PARP1 binds to the dyad symmetry (DS) element of *OriP*, the site for EBNA1 binding. • PARP1 PARylates EBNA1 and reduces EBNA1 binding to DS, leading to decreased viral DNA replication.	Deng et al., 2002, 2005; Tempera et al., 2010
		KSHV	• PARP1 binds to LANA. • PARP1 is recruited to the terminal repeat in the KSHV genome and colocalized with LANA. • Increased PARP activity induces PARylation of LANA and decreases viral genome copy number.	Ohsaki et al., 2004; Barbera et al., 2006a,b
	Lytic infection	EBV	• PARP1 binds to the *BZLF1* promoter and prevents Zta from binding to the same promoter. • Zta attenuates the PARylation activity of PARP1.	Lupey-Green et al., 2017
		KSHV	• PARP1 directly binds to and PARylates KSHV RTA, repressing the RTA-mediated transcriptional activity by inhibiting the recruitment of RTA to the lytic gene promoters. • KSHV vPIP directly binds to and sequesters PARP1 from RTA. • KSHV lytic replication induces PARP1 degradation. • KSHV PF-8 binds to and degrades PARP1 via recruitment of a cellular ubiquitin E3 ligase, CHFR, which enhances RTA transactivation.	Gwack et al., 2003; Ko et al., 2012; Noh et al., 2012; Chung et al., 2015, 2018, 2021
		MHV-68	• PARP1 directly binds to MHV-68 RTA and inhibits its transactivation. • MHV-68 RTA transactivation and lytic replication are enhanced in the PARP1 knockout cells. • MHV-68 vPIP binds and sequesters PARP1 from RTA, thereby promoting RTA-mediated transactivation. • MHV-68 vPIP mutant viruses with transposon insertion or triple stop codons were highly attenuated in viral growth *in vitro* and *in vivo*. • MHV-68 recombinant virus harboring a vPIP mutation defective in PARP1 interaction exhibits severely attenuated viral growth both *in vitro* and *in vivo*. • MHV-68 lytic replication and mPF induces PARP1 degradation.	Gwack et al., 2003; Noh et al., 2012; Chung et al., 2015, 2018

of PARP1 by thoroughly examining the effects of *PARP1* knockdown or overexpression on virus DNA replication, lytic gene transcription, viral protein expression, and ultimately, virion production.

As PARP1 activity requires and thereby lowers the cellular NAD^+ levels, the activity of other NAD^+-dependent deacetylases, sirtuins (SIRTs) may be affected (Houtkooper et al., 2012). PARP1 activation downregulates SIRT activity accompanied by lowered NAD^+ level, while SIRT1 activation reduces PARP activity, suggesting that SIRTs and PARPs compete for the cofactor NAD^+ (Zhang, 2003; Pillai et al., 2005; Kolthur-Seetharam et al., 2006; Bai et al., 2011). SIRT1 inhibition or knockdown reportedly induced KSHV lytic replication, increased active histone H3K4me3 mark, and decreased histone H3K27me3 mark in the RTA promoter. SIRT1 also interacted with RTA, inhibiting its transactivation activity, thereby regulating KSHV lytic replication (Li et al., 2014). Similarly, SIRT6 reportedly bound to KSHV genome and suppressed lytic gene expression (Hu et al., 2019). It will be interesting to study how these two competing factors, SIRTs and PARPs, regulate and balance the cellular NAD^+ level during virus life cycle and how the gammaherpesviruses modulate these factors for their own benefit.

In conclusion, PARP1 modulates gammaherpesvirus life cycle at lytic replication or latency stage. Interestingly, gammaherpesviruses surmount the inhibitory effects of PARP1 using viral proteins that interact with PARP1 or exploit PARP1 activity for efficient replication. However, the detailed modes of action of PARP1 need to be elucidated under various conditions during virus lytic replication or latency that result in different consequences. Further investigations regarding these aspects will expand our understanding of the interactions between PARP1 and oncogenic gammaherpesviruses, which in turn may lead to identification of effective therapeutic targets for the associated diseases.

AUTHOR CONTRIBUTIONS

W-CC and MJS: conceptualization, writing—review and editing, and visualization. MJS: supervision, project administration, and

funding acquisition. Both authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by the National Research Foundation of Korea (NRF) grants, the Bio and Medical Technology Development Program, and the Basic Research Laboratory Program of NRF funded by the Korea government (MSIT) (Nos. 2020R1A2C2013827, 2021M3A9I2080487, and 2020R1A4A1018019). W-CC was supported, in part, by a postdoctoral research grant funded by MSIT (No. NRF2019R1A6A3A01093571).

ACKNOWLEDGMENTS

We apologize the colleagues whose work has not been cited in this manuscript due to the text limit.

REFERENCES

Affar, E. B., Germain, M., Winstall, E., Vodenicharov, M., Shah, R. G., Salvesen, G. S., et al. (2001). Caspase-3-mediated processing of poly (ADP-ribose) glycohydrolase during apoptosis. *J. Biol. Chem.* 276, 2935–2942. doi: 10.1074/jbc.M007269200

Alemasova, E. E., and Lavrik, O. I. (2019). Poly(ADP-ribosyl)ation by PARP1: reaction mechanism and regulatory proteins. *Nucleic Acids Res.* 47, 3811–3827. doi: 10.1093/nar/gkz120

Altmeyer, M., Messner, S., Hassa, P. O., Fey, M., and Hottiger, M. O. (2009). Molecular mechanism of poly (ADP-ribosyl) ation by PARP1 and identification of lysine residues as ADP-ribose acceptor sites. *Nucleic Acids Res.* 37, 3723–3738. doi: 10.1093/nar/gkp229

Alvarez-Gonzalez, R., Watkins, T. A., Gill, P. K., Reed, J. L., and Mendoza-Alvarez, H. (1999). Regulatory mechanisms of poly (ADP-ribose) polymerase. *Mol. Cell. Biochem.* 193, 19–22.

Amé, J. C., Spenlehauer, C., and De Murcia, G. (2004). The PARP superfamily. *Bioessays* 26, 882–893. doi: 10.1002/bies.20085

Aneja, K. K., and Yuan, Y. (2017). Reactivation and lytic replication of kaposi's sarcoma-associated herpesvirus: an update. *Front. Microbiol.* 8:613. doi: 10.3389/fmicb.2017.00613

Arvin, A., Campadelli-Fiume, G., Mocarski, E., Moore, P. S., Roizman, B., Whitley, R., et al. (2007). *Human Herpesviruses: Biology, Therapy, And Immunoprophylaxis*. Cambridge: Cambridge University Press.

Bai, P. (2015). Biology of Poly(ADP-Ribose) polymerases: the factotums of cell maintenance. *Mol. Cell* 58, 947–958. doi: 10.1016/j.molcel.2015.01.034

Bai, P., Cantó, C., Oudart, H., Brunyánszki, A., Cen, Y., Thomas, C., et al. (2011). PARP-1 inhibition increases mitochondrial metabolism through SIRT1 activation. *Cell Metab.* 13, 461–468. doi: 10.1016/j.cmet.2011.03.004

Barbera, A. J., Chodaparambil, J. V., Kelley-Clarke, B., Joukov, V., Walter, J. C., Luger, K., et al. (2006a). The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA. *Science* 311, 856–861. doi: 10.1126/science.1120541

Barbera, A. J., Chodaparambil, J. V., Kelley-Clarke, B., Luger, K., and Kaye, K. M. (2006b). Kaposi's sarcoma-associated herpesvirus LANA hitches a ride on the chromosome. *Cell Cycle* 5, 1048–1052. doi: 10.4161/cc.5.10.2768

Beernink, P. T., Hwang, M., Ramirez, M., Murphy, M. B., Doyle, S. A., and Thelen, M. P. (2005). Specificity of protein interactions mediated by BRCT domains of the XRCC1 DNA repair protein. *J. Biol. Chem.* 280, 30206–30213. doi: 10.1074/jbc.M502155200

Bonfiglio, J. J., Fontana, P., Zhang, Q., Colby, T., Gibbs-Seymour, I., Atanassov, I., et al. (2017). Serine ADP-ribosylation depends on HPF1. *Mol. Cell* 65, 932.e–940.e. doi: 10.1016/j.molcel.2017.01.003

Bueno, M. T., Reyes, D., Valdes, L., Saheba, A., Urias, E., Mendoza, C., et al. (2013). Poly(ADP-ribose) polymerase 1 promotes transcriptional repression of integrated retroviruses. *J. Virol.* 87, 2496–2507. doi: 10.1128/JVI.01668-12

Cai, Q., Verma, S. C., Lu, J., and Robertson, E. S. (2010). Molecular biology of Kaposi's sarcoma-associated herpesvirus and related oncogenesis. *Adv. Virus Res.* 78, 87–142. doi: 10.1016/B978-0-12-385032-4.00003-3

Calabro, M., Sheldon, J., Favero, A., Simpson, G., Fiore, J., Gomes, E., et al. (1997). Seroprevalence of Kaposi's sarcoma-associated herpesvirus/human herpesvirus 8 in several regions of Italy. *J. Hum. Virol.* 1, 207–213 .

Caldecott, K. W., Aoufouchi, S., Johnson, P., and Shall, S. (1996). XRCC1 polypeptide interacts with DNA polymerase β and possibly poly (ADP-ribose) polymerase, and DNA ligase III is a novel molecular 'nick-sensor'in vitro. *Nucleic Acids Res.* 24, 4387–4394. doi: 10.1093/nar/24.22.4387

Calderwood, M. A., Venkatesan, K., Xing, L., Chase, M. R., Vazquez, A., Holthaus, A. M., et al. (2007). Epstein-Barr virus and virus human protein interaction maps. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7606–7611. doi: 10.1073/pnas.0702332104

Chang, C. M., Yu, K. J., Mbulaiteye, S. M., Hildesheim, A., and Bhatia, K. (2009). The extent of genetic diversity of Epstein-Barr virus and its geographic and disease patterns: a need for reappraisal. *Virus Res.* 143, 209–221. doi: 10.1016/j.virusres.2009.07.005

Chang, H. Y., Fan, C. C., Chu, P. C., Hong, B. E., Lee, H. J., and Chang, M. S. (2011). HPuf-A/KIAA0020 modulates PARP-1 cleavage upon genotoxic stress. *Cancer Res.* 71, 1126–1134. doi: 10.1158/0008-5472.CAN-10-1831

Chang, Y., Cesarman, E., Pessin, M. S., Lee, F., Culpepper, J., Knowles, D. M., et al. (1994). Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* 266, 1865–1869. doi: 10.1126/science.7997879

Chung, W. C., Lee, S., Kim, Y., Seo, J. B., and Song, M. J. (2021). Kaposi's sarcoma-associated herpesvirus processivity factor (PF-8) recruits cellular E3 ubiquitin ligase CHFR to promote PARP1 degradation and lytic replication. *PLoS Pathog.* 17:e1009261. doi: 10.1371/journal.ppat.1009261

Chung, W. C., Park, J. H., Kang, H. R., and Song, M. J. (2015). Downregulation of Poly(ADP-Ribose) polymerase 1 by a viral processivity factor facilitates lytic replication of *Gammaherpesvirus*. *J. Virol.* 89, 9676–9682. doi: 10.1128/JVI.00559-15

Chung, W.-C., Kim, J., Kim, B. C., Kang, H.-R., Son, J., Ki, H., et al. (2018). Structure-based mechanism of action of a viral poly (ADP-ribose) polymerase 1-interacting protein facilitating virus replication. *IUCrJ* 5, 866–879. doi: 10.1107/S2052252518013854

Ciccarone, F., Zampieri, M., and Caiafa, P. (2017). PARP1 orchestrates epigenetic events setting up chromatin domains. *Semin. Cell Dev. Biol.* 63, 123–134. doi: 10.1016/j.semcdb.2016.11.010

Corey, L., Brodie, S., Huang, M. L., Koelle, D. M., and Wald, A. (2002). HHV-8 infection: a model for reactivation and transmission. *Rev. Med. Virol.* 12, 47–63. doi: 10.1002/rmv.341

Cuneo, M. J., Gabel, S. A., Krahn, J. M., Ricker, M. A., and London, R. E. (2011). The structural basis for partitioning of the XRCC1/DNA ligase III-α BRCT-mediated dimer complexes. *Nucleic Acids Res.* 39, 7816–7827. doi: 10.1093/nar/gkr419

Dandri, M., Burda, M. R., Burkle, A., Zuckerman, D. M., Will, H., Rogler, C. E., et al. (2002). Increase in de novo HBV DNA integrations in response to oxidative DNA damage or inhibition of poly(ADP-ribosyl)ation. *Hepatology* 35, 217–223. doi: 10.1053/jhep.2002.30203

Davidovic, L., Vodenicharov, M., Affar, E. B., and Poirier, G. G. (2001). Importance of poly (ADP-ribose) glycohydrolase in the control of poly (ADP-ribose) metabolism. *Exp. Cell Res.* 268, 7–13. doi: 10.1006/excr.2001.5263

Davis, Z. H., Verschueren, E., Jang, G. M., Kleffman, K., Johnson, J. R., Park, J., et al. (2015). Global mapping of *herpesvirus*-host protein complexes reveals a transcription strategy for late genes. *Mol. Cell* 57, 349–360. doi: 10.1016/j.molcel.2014.11.026

Demény, M. A., and Virág, L. (2021). The PARP enzyme family and the hallmarks of cancer part 1. cell intrinsic hallmarks. *Cancers (Basel)* 13:2042. doi: 10.3390/cancers13092042

Deng, Z., Atanasiu, C., Zhao, K., Marmorstein, R., Sbodio, J. I., Chi, N. W., et al. (2005). Inhibition of Epstein-Barr virus OriP function by tankyrase, a telomere-associated poly-ADP ribose polymerase that binds and modifies EBNA1. *J. Virol.* 79, 4640–4650. doi: 10.1128/JVI.79.8.4640-4650.2005

Deng, Z., Lezina, L., Chen, C. J., Shtivelband, S., So, W., and Lieberman, P. M. (2002). Telomeric proteins regulate episomal maintenance of Epstein-Barr virus origin of plasmid replication. *Mol. Cell* 9, 493–503. doi: 10.1016/s1097-2765(02)00476-8

Di Alberti, L., Piattelli, A., Artese, L., Favia, G., Patel, S., Saunders, N., et al. (1997). Human herpesvirus 8 variants in sarcoid tissues. *Lancet* 350, 1655–1661. doi: 10.1016/s0140-6736(97)10102-7

Dugan, J. P., Coleman, C. B., and Haverkos, B. (2019). Opportunities to target the life cycle of epstein-barr virus (EBV) in EBV-associated lymphoproliferative disorders. *Front. Oncol.* 9:127. doi: 10.3389/fonc.2019.00127

Fehr, A. R., Singh, S. A., Kerr, C. M., Mukai, S., Higashi, H., and Aikawa, M. (2020). The impact of PARPs and ADP-ribosylation on inflammation and host-pathogen interactions. *Genes Dev.* 34, 341–359. doi: 10.1101/gad.334425.119

Fowler, P., Marques, S., Simas, J. P., and Efstathiou, S. (2003). ORF73 of murine herpesvirus-68 is critical for the establishment and maintenance of latency. *J. Gen. Virol.* 84, 3405–3416. doi: 10.1099/vir.0.19594-0

Ganem, D. (2007). *KSHV-Induced Oncogenesis*. Cambridge: Cambridge University Press.

Gao, S. J., Zhang, Y. J., Deng, J. H., Rabkin, C. S., Flore, O., and Jenson, H. B. (1999). Molecular polymorphism of Kaposi's sarcoma-associated herpesvirus (Human herpesvirus 8) latent nuclear antigen: evidence for a large repertoire of viral genotypes and dual infection with different viral genotypes. *J. Infect. Dis.* 180, 1466–1476. doi: 10.1086/315098

Gibson, B. A., and Kraus, W. L. (2012). New insights into the molecular and cellular functions of poly(ADP-ribose) and PARPs. *Nat. Rev. Mol. Cell Biol.* 13, 411–424. doi: 10.1038/nrm3376

Glover, J. N., Williams, R. S., and Lee, M. S. (2004). Interactions between BRCT repeats and phosphoproteins: tangled up in two. *Trends Biochem. Sci.* 29, 579–585. doi: 10.1016/j.tibs.2004.09.010

Gradoville, L., Gerlach, J., Grogan, E., Shedd, D., Nikiforow, S., Metroka, C., et al. (2000). Kaposi's sarcoma-associated *herpesvirus* open reading frame 50/Rta protein activates the entire viral lytic cycle in the HH-B2 primary effusion lymphoma cell line. *J. Virol.* 74, 6207–6212. doi: 10.1128/jvi.74.13.6207-6212.2000

Grady, S. L., Hwang, J., Vastag, L., Rabinowitz, J. D., and Shenk, T. (2012). Herpes simplex virus 1 infection activates poly(ADP-ribose) polymerase and triggers the degradation of poly(ADP-ribose) glycohydrolase. *J. Virol.* 86, 8259–8268. doi: 10.1128/JVI.00495-12

Gupte, R., Liu, Z., and Kraus, W. L. (2017). PARPs and ADP-ribosylation: recent advances linking molecular functions to biological outcomes. *Genes Dev.* 31, 101–126. doi: 10.1101/gad.291518.116

Gwack, Y., Nakamura, H., Lee, S. H., Souvlis, J., Yustein, J. T., Gygi, S., et al. (2003). Poly(ADP-ribose) polymerase 1 and Ste20-like kinase hKFC act as transcriptional repressors for gamma-2 herpesvirus lytic replication. *Mol. Cell Biol.* 23, 8282–8294. doi: 10.1128/MCB.23.22.8282-8294.2003

Ha, H. C., Juluri, K., Zhou, Y., Leung, S., Hermankova, M., and Snyder, S. H. (2001). Poly(ADP-ribose) polymerase-1 is required for efficient HIV-1 integration. *Proc. Natl. Acad. Sci. U.S.A.* 98, 3364–3368. doi: 10.1073/pnas.051633498

Habison, A. C., Beauchemin, C., Simas, J. P., Usherwood, E. J., and Kaye, K. M. (2012). Murine Gammaherpesvirus 68 LANA acts on terminal repeat DNA to mediate episome persistence. *J. Virol.* 86, 11863–11876. doi: 10.1128/JVI.01656-12

Hassa, P. O., and Hottiger, M. O. (2008). The diverse biological roles of mammalian PARPS, a small but powerful family of poly-ADP-ribose polymerases. *Front. Biosci.* 13, 3046–3082. doi: 10.2741/2909

Henke-Gendo, C., and Schulz, T. F. (2004). Transmission and disease association of Kaposi's sarcoma-associated herpesvirus: recent developments. *Curr. Opin. Infect. Dis.* 17, 53–57. doi: 10.1097/00001432-200402000-00011

Hottiger, M. O., Hassa, P. O., Lüscher, B., Schüler, H., and Koch-Nolte, F. (2010). Toward a unified nomenclature for mammalian ADP-ribosyltransferases. *Trends Biochem. Sci.* 35, 208–219. doi: 10.1016/j.tibs.2009.12.003

Houtkooper, R. H., Pirinen, E., and Auwerx, J. (2012). Sirtuins as regulators of metabolism and healthspan. *Nat. Rev. Mol. Cell Biol.* 13, 225–238. doi: 10.1038/nrm3293

Hu, J., Garber, A. C., and Renne, R. (2002). The latency-associated nuclear antigen of Kaposi's sarcoma-associated herpesvirus supports latent DNA replication in dividing cells. *J. Virol.* 76, 11677–11687. doi: 10.1128/jvi.76.22.11677-11687.2002

Hu, M., Armstrong, N., Seto, E., Li, W., Zhu, F., Wang, P. C., et al. (2019). Sirtuin 6 Attenuates Kaposi's Sarcoma-Associated herpesvirus reactivation by suppressing ori-lyt activity and expression of RTA. *J. Virol.* 93, e02200–e2218. doi: 10.1128/JVI.02200-18

Hulse, M., Caruso, L. B., Madzo, J., Tan, Y., Johnson, S., and Tempera, I. (2018). Poly(ADP-ribose) polymerase 1 is necessary for coactivating hypoxia-inducible factor-1-dependent gene expression by Epstein-Barr virus latent membrane protein 1. *PLoS Pathog* 14:e1007394. doi: 10.1371/journal.ppat.1007394

Jha, H. C., Banerjee, S., and Robertson, E. S. (2016). The role of Gammaherpesviruses in cancer pathogenesis. *Pathogens* 5:18. doi: 10.3390/pathogens5010018

Jubin, T., Kadam, A., Jariwala, M., Bhatt, S., Sutariya, S., Gani, A., et al. (2016). The PARP family: insights into functional aspects of poly (ADP-ribose) polymerase-1 in cell growth and survival. *Cell Prolif.* 49, 421–437. doi: 10.1111/cpr.12268

Kamaletdinova, T., Fanaei-Kahrani, Z., and Wang, Z. Q. (2019). The enigmatic function of PARP1: from PARylation activity to PAR readers. *Cells* 8:1625. doi: 10.3390/cells8121625

Kameoka, M., Nukuzuma, S., Itaya, A., Tanaka, Y., Ota, K., Inada, Y., et al. (2005). Poly(ADP-ribose)polymerase-1 is required for integration of the human immunodeficiency virus type 1 genome near centromeric alphoid DNA in human and murine cells. *Biochem. Biophys. Res. Commun.* 334, 412–417. doi: 10.1016/j.bbrc.2005.06.104

Kameoka, M., Tanaka, Y., Ota, K., Itaya, A., and Yoshihara, K. (1999). Poly (ADP-ribose) polymerase is involved in PMA-induced activation of HIV-1 in U1 cells by modulating the LTR function. *Biochem. Biophys. Res. Commun.* 262, 285–289. doi: 10.1006/bbrc.1999.1146

Kanno, S., Kuzuoka, H., Sasao, S., Hong, Z., Lan, L., Nakajima, S., et al. (2007). A novel human AP endonuclease with conserved zinc-finger-like motifs involved in DNA strand break responses. *EMBO J.* 26, 2094–2103. doi: 10.1038/sj.emboj.7601663

Kawaichi, M., Ueda, K., and Hayaishi, O. (1980). Initiation of poly(ADP-ribosyl) histone synthesis by poly(ADP-ribose) synthetase. *J. Biol. Chem.* 255, 816–819. doi: 10.1016/s0021-9258(19)86100-8

Ke, Y., Wang, C., Zhang, J., Zhong, X., Wang, R., Zeng, X., et al. (2019). The Role of PARPs in inflammation-and metabolic-related diseases: molecular mechanisms and beyond. *Cells* 8:1047. doi: 10.3390/cells8091047

Kennedy, G., and Sugden, B. (2003). EBNA-1, a bifunctional transcriptional activator. *Mole. Cell. Biol.* 23, 6901–6908. doi: 10.1128/MCB.23.19.6901-6908.2003

Kiehl, A., and Dorsky, D. I. (1991). Cooperation of EBV DNA polymerase and EA-D(BMRF1) in vitro and colocalization in nuclei of infected cells. *Virology* 184, 330–340. doi: 10.1016/0042-6822(91)90849-7

Kim, M. Y., Zhang, T., and Kraus, W. L. (2005). Poly(ADP-ribosyl)ation by PARP-1: 'PAR-laying' NAD+ into a nuclear signal. *Genes Dev.* 19, 1951–1967. doi: 10.1101/gad.1331805

Ko, H. L., and Ren, E. C. (2011). Novel poly (ADP-ribose) polymerase 1 binding motif in hepatitis B virus core promoter impairs DNA damage repair. *Hepatology* 54, 1190–1198. doi: 10.1002/hep.24502

Ko, H. L., and Ren, E. C. (2012). Functional aspects of PARP1 in DNA repair and transcription. *Biomolecules* 2, 524–548. doi: 10.3390/biom2040524

Ko, Y.-C., Tsai, W.-H., Wang, P.-W., Wu, I.-L., Lin, S.-Y., Chen, Y.-L., et al. (2012). Suppressive regulation of KSHV RTA with O-GlcNAcylation. *J. Biomed. Sci.* 19:12. doi: 10.1186/1423-0127-19-12

Kolthur-Seetharam, U., Dantzer, F., Mcburney, M. W., De Murcia, G., and Sassone-Corsi, P. (2006). Control of AIF-mediated cell death by the functional interplay

of SIRT1 and PARP-1 in response to DNA damage. *Cell Cycle* 5, 873–877. doi: 10.4161/cc.5.8.2690

Kotoglou, P., Kalaitzakis, A., Vezyraki, P., Tzavaras, T., Michalis, L. K., Dantzer, F., et al. (2009). Hsp70 translocates to the nuclei and nucleoli, binds to XRCC1 and PARP-1, and protects HeLa cells from single-strand DNA breaks. *Cell Stress Chaperones* 14, 391–406. doi: 10.1007/s12192-008-0093-6

Kraus, W. L., and Lis, J. T. (2003). PARP goes transcription. *Cell* 113, 677–683.

Krishnakumar, R., and Kraus, W. L. (2010). The PARP side of the nucleus: molecular actions, physiological outcomes, and clinical targets. *Mol. Cell* 39, 8–24. doi: 10.1016/j.molcel.2010.06.017

Langelier, M. F., Planck, J. L., Roy, S., and Pascal, J. M. (2012). Structural basis for DNA damage-dependent poly(ADP-ribosyl)ation by human PARP-1. *Science* 336, 728–732. doi: 10.1126/science.1216338

Langelier, M. F., Servent, K. M., Rogers, E. E., and Pascal, J. M. (2008). A third zinc-binding domain of human poly(ADP-ribose) polymerase-1 coordinates DNA-dependent enzyme activation. *J. Biol. Chem.* 283, 4105–4114. doi: 10.1074/jbc.m708558200

Langelier, M.-F., Planck, J. L., Roy, S., and Pascal, J. M. (2011). Crystal structures of poly (ADP-ribose) polymerase-1 (PARP-1) zinc fingers bound to DNA: structural and functional insights into DNA-dependent PARP-1 activity. *J. Biol. Chem.* 286, 10690–10701. doi: 10.1074/jbc.M110.202507

Lee, H.-R., Brulois, K., Wong, L., and Jung, J. U. (2012). Modulation of immune system by Kaposi's sarcoma-associated herpesvirus: lessons from viral evasion strategies. *Front. Microbiol.* 3:44. doi: 10.3389/fmicb.2012.00044

Li, H. P., and Chang, Y. S. (2003). Epstein-Barr virus latent membrane protein 1: structure and functions. *J. Biomed. Sci.* 10, 490–504. doi: 10.1007/BF02256110

Li, Q., He, M., Zhou, F., Ye, F., and Gao, S.-J. (2014). Activation of Kaposi's sarcoma-associated herpesvirus by inhibitors of class III histone deacetylases: identification of SIRT1 as a regulator of KSHV life cycle. *J. Virol.* 88, 6355–6367. doi: 10.1128/JVI.00219-14

Li, Y., Oh, H. J., and Lau, Y. F. (2006). The poly(ADP-ribose) polymerase 1 interacts with Sry and modulates its biological functions. *Mol. Cell Endocrinol.* 25, 35–46. doi: 10.1016/j.mce.2006.06.008

Lindner, S. E., and Sugden, B. (2007). The plasmid replicon of Epstein-Barr virus: mechanistic insights into efficient, licensed, extrachromosomal replication in human cells. *Plasmid* 58, 1–12. doi: 10.1016/j.plasmid.2007.01.003

Liu, L., Lear, Z., Hughes, D. J., Wu, W., Zhou, E. M., Whitehouse, A., et al. (2015). Resolution of the cellular proteome of the nucleocapsid protein from a highly pathogenic isolate of porcine reproductive and respiratory syndrome virus identifies PARP-1 as a cellular target whose interaction is critical for virus biology. *Vet. Microbiol.* 176, 109–119. doi: 10.1016/j.vetmic.2014.11.023

Liu, S., Pavlova, I. V., Virgin, H. W. T., and Speck, S. H. (2000). Characterization of Gammaherpesvirus 68 gene 50 transcription. *J. Virol.* 74, 2029–2037. doi: 10.1128/jvi.74.4.2029-2037.2000

Liu, X., Li, Y., Peng, S., Yu, X., Li, W., Shi, F., et al. (2018). Epstein-Barr virus encoded latent membrane protein 1 suppresses necroptosis through targeting RIPK1/3 ubiquitination. *Cell Death Dis.* 9:53. doi: 10.1038/s41419-017-0081-9

Lukac, D. M., and Yuan, Y. (2007). "Reactivation and lytic replication of KSHV," in *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis*, eds A. Arvin, G. Campadelli-Fiume, E. Mocarski, P. S. Moore, B. Roizman, R. Whitley, et al. (Cambridge: Cambridge University Press).

Lukac, D. M., Kirshner, J. R., and Ganem, D. (1999). Transcriptional activation by the product of open reading frame 50 of Kaposi's sarcoma-associated herpesvirus is required for lytic viral reactivation in B cells. *J. Virol.* 73, 9348–9361. doi: 10.1128/JVI.73.11.9348-9361.1999

Lukac, D. M., Renne, R., Kirshner, J. R., and Ganem, D. (1998). Reactivation of Kaposi's sarcoma-associated herpesvirus infection from latency by expression of the ORF 50 transactivator, a homolog of the EBV R protein. *Virology* 252, 304–312. doi: 10.1006/viro.1998.9486

Luo, X., and Kraus, W. L. (2012). On PAR with PARP: cellular stress signaling through poly(ADP-ribose) and PARP-1. *Genes Dev.* 26, 417–432. doi: 10.1101/gad.183509.111

Lupey-Green, L. N., Caruso, L. B., Madzo, J., Martin, K. A., Tan, Y., Hulse, M., et al. (2018). PARP1 stabilizes CTCF binding and chromatin structure to maintain Epstein-Barr virus latency type. *J. Virol.* 92, e755–e818. doi: 10.1128/JVI.00755-18

Lupey-Green, L. N., Moquin, S. A., Martin, K. A., Mcdevitt, S. M., Hulse, M., Caruso, L. B., et al. (2017). PARP1 restricts Epstein Barr virus lytic reactivation by binding the BZLF1 promoter. *Virology* 507, 220–230. doi: 10.1016/j.virol.2017.04.006

Luzuriaga, K., and Sullivan, J. L. (2010). Infectious mononucleosis. *N. Engl. J. Med.* 362, 1993–2000. doi: 10.1056/NEJMcp1001116

Mangerich, A., and Bürkle, A. (2012). Pleiotropic cellular functions of PARP1 in longevity and aging: genome maintenance meets inflammation. *Oxid. Med. Cell Longev.* 2012:321653. doi: 10.1155/2012/321653

Martin, K. A., Lupey, L. N., and Tempera, I. (2016). Epstein-Barr Virus Oncoprotein LMP1 mediates epigenetic changes in host gene expression through PARP1. *J. Virol.* 90, 8520–8530. doi: 10.1128/JVI.01180-16

Martro, E., Bulterys, M., Stewart, J., Spira, T., Cannon, M., Thacher, T., et al. (2004). Comparison of human herpesvirus 8 and Epstein–Barr virus seropositivity among children in areas endemic and non-endemic for Kaposi's sarcoma. *J. Med. Virol.* 72, 126–131. doi: 10.1002/jmv.10548

Masson, M., Niedergang, C., Schreiber, V., Muller, S., Menissier-De Murcia, J., and De Murcia, G. (1998). XRCC1 is specifically associated with poly (ADP-ribose) polymerase and negatively regulates its activity following DNA damage. *Mol. Cell Biol.* 18, 3563–3571. doi: 10.1128/MCB.18.6.3563

Mattiussi, S., Tempera, I., Matusali, G., Mearini, G., Lenti, L., Fratarcangeli, S., et al. (2007). Inhibition of Poly(ADP-ribose)polymerase impairs Epstein Barr Virus lytic cycle progression. *Infect. Agent Cancer* 2:18. doi: 10.1186/1750-9378-2-18

Mayama, S., Cuevas, L. E., Sheldon, J., Omar, O. H., Smith, D. H., Okong, P., et al. (1998). Prevalence and transmission of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in Ugandan children and adolescents. *Int. J. Cancer* 77, 817–820. doi: 10.1002/(sici)1097-0215(19980911)77:6<817::aid-ijc2>3.0.co;2-x

McGeoch, D. J., and Davison, A. J. (1999). The descent of human herpesvirus 8. *Semin. Cancer Biol.* 9, 201–209. doi: 10.1006/scbi.1999.0093

Mendoza-Alvarez, H., and Alvarez-Gonzalez, R. (1993). Poly (ADP-ribose) polymerase is a catalytic dimer and the automodification reaction is intermolecular. *J. Biol. Chem.* 268, 22575–22580. doi: 10.1016/s0021-9258(18)41568-2

Moorman, N. J., Willer, D. O., and Speck, S. H. (2003). The gammaherpesvirus 68 latency-associated nuclear antigen homolog is critical for the establishment of splenic latency. *J. Virol.* 77, 10295–10303. doi: 10.1128/jvi.77.19.10295-10303.2003

Murata, T. (2014). Regulation of Epstein-Barr virus reactivation from latency. *Microbiol. Immunol.* 58, 307–317. doi: 10.1111/1348-0421.12155

Myoung, J., and Ganem, D. (2011). Infection of primary human tonsillar lymphoid cells by KSHV reveals frequent but abortive infection of T cells. *Virology* 413, 1–11. doi: 10.1016/j.virol.2010.12.036

Na, T. Y., Ka, N. L., Rhee, H., Kyeong, D., Kim, M. H., Seong, J. K., et al. (2016). Interaction of hepatitis B virus X protein with PARP1 results in inhibition of DNA repair in hepatocellular carcinoma. *Oncogene* 35, 5435–5445. doi: 10.1038/onc.2016.82

Nargi-Aizenman, J. L., Simbulan-Rosenthal, C. M., Kelly, T. A., Smulson, M. E., and Griffin, D. E. (2002). Rapid activation of poly(ADP-ribose) polymerase contributes to *Sindbis virus* and staurosporine-induced apoptotic cell death. *Virology* 293, 164–171. doi: 10.1006/viro.2001.1253

Nash, A. A., Dutia, B. M., Stewart, J. P., and Davison, A. J. (2001). Natural history of murine gamma-herpesvirus infection. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356, 569–579. doi: 10.1098/rstb.2000.0779

Neuhierl, B., and Delecluse, H. J. (2006). The Epstein-Barr virus BMRF1 gene is essential for lytic virus replication. *J. Virol.* 80, 5078–5081. doi: 10.1128/JVI.80.10.5078-5081.2006

Niller, H. H., Szenthe, K., and Minarovits, J. (2014). Epstein-Barr virus-host cell interactions: an epigenetic dialog? *Front. Genet.* 5:367. doi: 10.3389/fgene.2014.00367

Noh, C.-W., Cho, H.-J., Kang, H.-R., Jin, H. Y., Lee, S., Deng, H., et al. (2012). The virion-associated open reading frame 49 of murine gammaherpesvirus 68 promotes viral replication both in vitro and in vivo as a derepressor of RTA. *J. Virol.* 86, 1109–1118. doi: 10.1128/JVI.05785-11

Ogata, N., Ueda, K., and Hayaishi, O. (1980a). ADP-ribosylation of histone H2B. Identification of glutamic acid residue 2 as the modification site. *J. Biol. Chem.* 255, 7610–7615.

Ogata, N., Ueda, K., Kagamiyama, H., and Hayaishi, O. (1980b). ADP-ribosylation of histone H1. Identification of glutamic acid residues 2, 14, and the COOH-terminal lysine residue as modification sites. *J. Biol. Chem.* 255, 7616–7620.

Ohsaki, E., Ueda, K., Sakakibara, S., Do, E., Yada, K., and Yamanishi, K. (2004). Poly(ADP-ribose) polymerase 1 binds to Kaposi's sarcoma-associated herpesvirus (KSHV) terminal repeat sequence and modulates KSHV replication in latency. *J. Virol.* 78, 9936–9946. doi: 10.1128/JVI.78.18.9936-9946.2004

Oka, J., Ueda, K., Hayaishi, O., Komura, H., and Nakanishi, K. (1984). ADP-ribosyl protein lyase. Purification, properties, and identification of the product. *J. Biol. Chem.* 259, 986–995.

Oka, S., Kato, J., and Moss, J. (2006). Identification and characterization of a mammalian 39-kDa poly(ADP-ribose) glycohydrolase. *J. Biol. Chem.* 281, 705–713. doi: 10.1074/jbc.M510290200

Parent, M., Yung, T. M., Rancourt, A., Ho, E. L., Vispé, S., Suzuki-Matsuda, F., et al. (2005). Poly (ADP-ribose) polymerase-1 is a negative regulator of HIV-1 transcription through competitive binding to TAR RNA with tat· positive transcription elongation factor b (p-TEFb) complex. *J. Biol. Chem.* 280, 448–457. doi: 10.1074/jbc.M408435200

Park, E., and Griffin, D. E. (2009). Interaction of *Sindbis* virus non-structural protein 3 with poly(ADP-ribose) polymerase 1 in neuronal cells. *J. Gen. Virol.* 90, 2073–2080. doi: 10.1099/vir.0.012682-0

Pillai, J. B., Isbatan, A., Imai, S., and Gupta, M. P. (2005). Poly(ADP-ribose) polymerase-1-dependent cardiac myocyte cell death during heart failure is mediated by NAD+ depletion and reduced Sir2alpha deacetylase activity. *J. Biol. Chem.* 280, 43121–43130. doi: 10.1074/jbc.M506162200

Price, A. M., and Luftig, M. A. (2015). To be or not IIb: a multi-step process for Epstein-Barr virus latency establishment and consequences for B cell tumorigenesis. *PLoS Pathog.* 11:e1004656. doi: 10.1371/journal.ppat.1004656

Purushothaman, P., Dabral, P., Gupta, N., Sarkar, R., and Verma, S. C. (2016). KSHV genome replication and maintenance. *Front. Microbiol.* 7:54. doi: 10.3389/fmicb.2016.00054

Rainbow, L., Platt, G. M., Simpson, G. R., Sarid, R., Gao, S. J., Stoiber, H., et al. (1997). The 222- to 234-kilodalton latent nuclear protein (LNA) of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) is encoded by orf73 and is a component of the latency-associated nuclear antigen. *J. Virol.* 71, 5915–5921. doi: 10.1128/JVI.71.8.5915-5921.1997

Ray Chaudhuri, A., and Nussenzweig, A. (2017). The multifaceted roles of PARP1 in DNA repair and chromatin remodelling. *Nat. Rev. Mol. Cell Biol.* 18, 610–621. doi: 10.1038/nrm.2017.53

Rettig, M. B., Ma, H. J., Vescio, R. A., Pold, M., Schiller, G., Belson, D., et al. (1997). Kaposi's sarcoma-associated herpesvirus infection of bone marrow dendritic cells from multiple myeloma patients. *Science* 276, 1851–1854. doi: 10.1126/science.276.5320.1851

Rickinson, A. (2002). Epstein-Barr virus. *Virus Res.* 82, 109–113. doi: 10.1016/s0168-1702(01)00436-1

Rickinson, A. B. (2014). Co-infections, inflammation and oncogenesis: future directions for EBV research. *Semin. Cancer Biol.* 26, 99–115. doi: 10.1016/j.semcancer.2014.04.004

Rochford, R., Lutzke, M. L., Alfinito, R. S., Clavo, A., and Cardin, R. D. (2001). Kinetics of murine gammaherpesvirus 68 gene expression following infection of murine cells in culture and in mice. *J. Virol.* 75, 4955–4963. doi: 10.1128/JVI.75.11.4955-4963.2001

Rodríguez, M. I., González-Flores, A., Dantzer, F., Collard, J., De Herreros, A. G., and Oliver, F. J. (2011). Poly(ADP-ribose)-dependent regulation of Snail1 protein stability. *Oncogene* 30, 4365–4372. doi: 10.1038/onc.2011.153

Roizman, B., Carmichael, L., Deinhardt, F., Nahmias, A., Plowright, W., Rapp, F., et al. (1981). Herpesviridae. *Intervirology* 16, 201–217. doi: 10.1159/000149269

Rom, S., Reichenbach, N. L., Dykstra, H., and Persidsky, Y. (2015). The dual action of poly(ADP-ribose) polymerase -1 (PARP-1) inhibition in HIV-1 infection: HIV-1 LTR inhibition and diminution in Rho GTPase activity. *Front. Microbiol.* 6:878. doi: 10.3389/fmicb.2015.00878

Rosenthal, F., Feijs, K. L., Frugier, E., Bonalli, M., Forst, A. H., Imhof, R., et al. (2013). Macrodomain-containing proteins are new mono-ADP-ribosylhydrolases. *Nat. Struct. Mol. Biol.* 20, 502–507. doi: 10.1038/nsmb.2521

Sbodio, J. I., and Chi, N. W. (2002). Identification of a tankyrase-binding motif shared by IRAP. TAB182, and human TRF1 but not mouse TRF1. NuMA contains this RXXPDG motif and is a novel tankyrase partner. *J. Biol. Chem.* 277, 31887–31892. doi: 10.1074/jbc.M203916200

Schulz, T. F., and Chang, Y. (2007). "KSHV gene expression and regulation," in *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis*, eds A. Arvin, G. Campadelli-Fiume, E. Mocarski, P. S. Moore, B. Roizman, R. Whitley, et al. (Cambridge: Cambridge University Press).

Shair, K. H., Bendt, K. M., Edwards, R. H., Nielsen, J. N., Moore, D. T., and Raab-Traub, N. (2012). Epstein-Barr virus-encoded latent membrane protein 1 (LMP1) and LMP2A function cooperatively to promote carcinoma development in a mouse carcinogenesis model. *J. Virol.* 86, 5352–5365. doi: 10.1128/JVI.07035-11

Sharifi, R., Morra, R., Appel, C. D., Tallis, M., Chioza, B., Jankevicius, G., et al. (2013). Deficiency of terminal ADP-ribose protein glycohydrolase TARG1/C6orf130 in neurodegenerative disease. *EMBO J.* 32, 1225–1237. doi: 10.1038/emboj.2013.51

Shou, Q., Fu, H., Huang, X., and Yang, Y. (2019). PARP-1 controls NK cell recruitment to the site of viral infection. *JCI Insight* 4:e121291 doi: 10.1172/jci.insight.121291

Simas, J. P., and Efstathiou, S. (1998). Murine gammaherpesvirus 68: a model for the study of gammaherpesvirus pathogenesis. *Trends Microbiol.* 6, 276–282. doi: 10.1016/s0966-842x(98)01306-7

Simpson, G. R., Schulz, T. F., Whitby, D., Cook, P. M., Boshoff, C., Rainbow, L., et al. (1996). Prevalence of Kaposi's sarcoma associated herpesvirus infection measured by antibodies to recombinant capsid protein and latent immunofluorescence antigen. *Lancet* 348, 1133–1138. doi: 10.1016/S0140-6736(96)07560-5

Smith, S., Giriat, I., Schmitt, A., and De Lange, T. (1998). Tankyrase, a poly(ADP-ribose) polymerase at human telomeres. *Science* 282, 1484–1487. doi: 10.1126/science.282.5393.1484

Speck, S. H., and Virgin, H. W. (1999). Host and viral genetics of chronic infection: a mouse model of gamma-herpesvirus pathogenesis. *Curr. Opin. Microbiol.* 2, 403–409. doi: 10.1016/s1369-5274(99)80071-x

Srinivas, S. K., Sample, J. T., and Sixbey, J. W. (1998). Spontaneous loss of viral episomes accompanying Epstein-Barr virus reactivation in a Burkitt's lymphoma cell line. *J. Infect. Dis.* 177, 1705–1709. doi: 10.1086/517427

Strosznajder, R. P., Jesko, H., and Zambrzycka, A. (2005). Poly(ADP-ribose) polymerase: the nuclear target in signal transduction and its role in brain ischemia-reperfusion injury. *Mol. Neurobiol.* 31, 149–167. doi: 10.1385/MN:31:1-3:149

Sun, R., Lin, S. F., Gradoville, L., Yuan, Y., Zhu, F., and Miller, G. (1998). A viral gene that activates lytic cycle expression of Kaposi's sarcoma-associated herpesvirus. *Proc. Natl. Acad. Sci. U.S.A.* 95, 10866–10871. doi: 10.1073/pnas.95.18.10866

Tao, Z., Gao, P., and Liu, H.-W. (2009). Identification of the ADP-ribosylation sites in the PARP-1 automodification domain: analysis and implications. *J. Am. Chem. Soc.* 131, 14258–14260. doi: 10.1021/ja906135d

Tempera, I., Deng, Z., Atanasiu, C., Chen, C. J., D'erme, M., and Lieberman, P. M. (2010). Regulation of Epstein-Barr virus OriP replication by poly(ADP-ribose) polymerase 1. *J. Virol.* 84, 4988–4997. doi: 10.1128/JVI.02333-09

Tsurumi, T. (1993). Purification and characterization of the DNA-binding activity of the Epstein-Barr virus DNA polymerase accessory protein BMRF1 gene products, as expressed in insect cells by using the baculovirus system. *J. Virol.* 67, 1681–1687. doi: 10.1128/JVI.67.3.1681-1687.1993

Ubol, S., Park, S., Budihardjo, I., Desnoyers, S., Montrose, M. H., Poirier, G. G., et al. (1996). Temporal changes in chromatin, intracellular calcium, and poly(ADP-ribose) polymerase during *Sindbis* virus-induced apoptosis of neuroblastoma cells. *J. Virol.* 70, 2215–2220. doi: 10.1128/JVI.70.4.2215-2220.1996

Virgin, H. W., and Speck, S. H. (1999). Unraveling immunity to gamma-herpesviruses: a new model for understanding the role of immunity in chronic virus infection. *Curr. Opin. Immunol.* 11, 371–379. doi: 10.1016/s0952-7915(99)80063-6

Wang, Y., Li, H., Tang, Q., Maul, G. G., and Yuan, Y. (2008). Kaposi's sarcoma-associated herpesvirus ori-Lyt-dependent DNA replication: involvement of host cellular factors. *J. Virol.* 82, 2867–2882. doi: 10.1128/JVI.01319-07

Wei, F., Gan, J., Wang, C., Zhu, C., and Cai, Q. (2016). Cell cycle regulatory functions of the KSHV oncoprotein LANA. *Front. Microbiol.* 7:334. doi: 10.3389/fmicb.2016.00334

Whitby, D., and Boshoff, C. (1998). Kaposi's sarcoma herpesvirus as a new paradigm for virus-induced oncogenesis. *Curr. Opin. Oncol.* 10, 405–412. doi: 10.1097/00001622-199809000-00007

Wu, T. T., Usherwood, E. J., Stewart, J. P., Nash, A. A., and Sun, R. (2000). Rta of murine gammaherpesvirus 68 reactivates the complete lytic cycle from latency. *J. Virol.* 74, 3659–3667. doi: 10.1128/jvi.74.8.3659-3667.2000

Xia, C., Wolf, J. J., Sun, C., Xu, M., Studstill, C. J., Chen, J., et al. (2020). PARP1 enhances influenza a virus propagation by facilitating degradation of host type i interferon receptor. *J. Virol.* 94, e01572–e01619. doi: 10.1128/JVI.01572-19

Xie, S., Mortusewicz, O., Ma, H. T., Herr, P., Poon, R. Y., Helleday, T., et al. (2015). Timeless interacts with PARP-1 to promote homologous recombination repair. *Mol. Cell* 60, 163–176. doi: 10.1016/j.molcel.2015.07.031

Xu, Y., Aucoin, D. P., Huete, A. R., Cei, S. A., Hanson, L. J., and Pari, G. S. (2005). A Kaposi's sarcoma-associated herpesvirus/human herpesvirus 8 ORF50 deletion mutant is defective for reactivation of latent virus and DNA replication. *J. Virol.* 79, 3479–3487. doi: 10.1128/JVI.79.6.3479-3487.2005

Yan, L., Majerciak, V., Zheng, Z. M., and Lan, K. (2019). Towards better understanding of KSHV life cycle: from transcription and posttranscriptional regulations to pathogenesis. *Virol. Sin.* 34, 135–161. doi: 10.1007/s12250-019-00114-3

Ye, F., Lei, X., and Gao, S.-J. (2011). Mechanisms of Kaposi's sarcoma-associated herpesvirus latency and reactivation. *Adv. Virol.* 2011:193860. doi: 10.1155/2011/193860

Young, L. S., Arrand, J. R., and Murray, P. G. (2007). "EBV gene expression and regulation," in *Human Herpesviruses: Biology, therapy, and immunoprophylaxis*, eds A. Arvin, G. Campadelli-Fiume, E. Mocarski, P. S. Moore, B. Roizman, R. Whitley, et al. (Cambridge: Cambridge University Press).

Zhang, J. (2003). Are poly(ADP-ribosyl)ation by PARP-1 and deacetylation by Sir2 linked? *Bioessays* 25, 808–814. doi: 10.1002/bies.10317

Zhang, Z., Hildebrandt, E. F., Simbulan-Rosenthal, C. M., and Anderson, M. G. (2002). Sequence-specific binding of poly(ADP-ribose) polymerase-1 to the human T cell leukemia virus type-I tax responsive element. *Virology* 296, 107–116. doi: 10.1006/viro.2002.1385

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Virology

BRIEF RESEARCH REPORT
published: 07 February 2022
doi: 10.3389/fviro.2022.830202

Predictions of the SARS-CoV-2 Omicron Variant (B.1.1.529) Spike Protein Receptor-Binding Domain Structure and Neutralizing Antibody Interactions

Colby T. Ford[1,2]*, *Denis Jacob Machado*[1] and *Daniel A. Janies*[1]

[1] Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC, United States,
[2] School of Data Science, University of North Carolina at Charlotte, Charlotte, NC, United States

OPEN ACCESS

Edited by:
Gorka Lasso Cabrera,
Albert Einstein College of Medicine,
United States

Reviewed by:
Michael Gromiha,
Indian Institute of Technology Madras,
India
Matthew Raybould,
University of Oxford, United Kingdom
Reda Rawi,
National Institutes of Health (NIH),
United States

***Correspondence:**
Colby T. Ford
colby.ford@uncc.edu

Specialty section:
This article was submitted to Bioinformatic and Predictive Virology, a section of the journal Frontiers in Virology

Received: 06 December 2021
Accepted: 04 January 2022
Published: 07 February 2022

Citation:
Ford CT, Jacob Machado D and Janies DA (2022) Predictions of the SARS-CoV-2 Omicron Variant (B.1.1.529) Spike Protein Receptor-Binding Domain Structure and Neutralizing Antibody Interactions. Front. Virol. 2:830202. doi: 10.3389/fviro.2022.830202

The genome of the SARS-CoV-2 Omicron variant (B.1.1.529) was released on November 22, 2021, which has caused a flurry of media attention due the large number of mutations it contains. These raw data have spurred questions around vaccine efficacy. Given that neither the structural information nor the experimentally-derived antibody interaction of this variant are available, we have turned to predictive computational methods to model the mutated structure of the spike protein's receptor binding domain and posit potential changes to vaccine efficacy. In this study, we predict some structural changes in the receptor-binding domain that may reduce antibody interaction without completely evading existing neutralizing antibodies (and therefore current vaccines).

Keywords: SARS-CoV-2, COVID-19, bioinformatics, computational biology, protein-protein interaction (PPI), machine learning, protein structure, antibodies

INTRODUCTION

A team of researchers from the Botswana-Harvard HIV Reference Laboratory submitted a new SARS-CoV-2 genome sequence to GISAID on November 22, 2021 (GISAID accession. EPI_ISL_6752027). The specimen was taken from a living 59-year-old male from Gaborone, Botswana using a nasopharyngeal swab and was sequenced using a Nanopore MinION device.

This sample's genome contains 60 mutations from the Wuhan-derived reference genome (GenBank accession no. NC_045512.2) (1), 37 of which are in the Spike (S) protein. This variant was given the identifier B.1.1.529 by PANGO lineages (2). On November 26, 2021, the WHO has designated B.1.1.529 as a Variant of Concern (VOC), named Omicron (3).

The emergence of new SARS-CoV-2 variants is expected. Therefore, scientists have advocated for close international monitoring to determine the need for vaccination boosters and/or redesign (4). Hence, the identification of the omicron variant is not surprising. What is surprising is the number of mutations that the omicron variant accumulated compared to the first sequenced genome of SARS-CoV-2.

Different authors have warned that limited SARS-CoV-2 sampling and sequencing from positive cases, especially from asymptomatic and symptomatic cases that did not require hospitalization, would make it challenging to identify new mutations in the virus. For example, Brito et al. (5)

analyzed the spatiotemporal heterogeneity in each country's SARS-CoV-2 genomic surveillance efforts using metadata submitted to GISAID until May 30, 2021. These authors calculate that sequencing capacity should be at least 0.5% of cases per week when incidence is more than 100 positive cases every 100,000 people. Unfortunately, most countries are not reaching this sequence threshold.

While sampling bias can explain why we may miss new mutations and fail to identify new variants of low prevalence, the emergence of new variants is due to factors that favor the transmission of SARS-CoV-2, including low vaccination rates in some regions, especially in low and middle-income countries (LMICs). Therefore, disparities in vaccination rates combined with sampling bias explain why scientists may continue to be surprised by the mutations in new SARS-CoV-2 variants.

There are many questions regarding genomic epidemiology and the lessons we can learn from the COVID-19 pandemic. Those questions are beyond the scope of this manuscript, and we addressed them elsewhere (6). While the origin and evolution of the Omicron variant are still open questions, here we focus on the potential implications of the mutations observed in this variant.

This seemingly hyper-mutated variant is of public health concern with unanswered questions surrounding vaccine protection (from available vaccines), the possibility of reinfection, transmissibility, and pathogenicity.

Regarding vaccine efficacy, we must look at the receptor-binding domain (RBD), part of the S1 subunit, of the spike protein as this is the binding site for neutralizing antibodies. This domain exists between positions 319 and 541 of the spike protein. Omicron contains 15 mutations in the RBD, none of which are deletions or insertions. In contrast, the Delta variant contains 7 mutations across the entire spike protein, only 2 of which are in the RBD.

Given that an experimentally-derived structure of the Omicron spike protein is not yet available, we must derive a predicted structure from its sequence *in silico*. Then, we can use available neutralizing antibody structures to computationally model the interaction between Omicron and the paratopes of the antibodies, thus allowing us to compare potential affinity changes due to the mutations and posit their effects to vaccine efficacy.

METHODS

Sequence Comparison Among VBMs and VOCs

We downloaded the reference genome of SARS-CoV-2 (Wuhan-Hu-1, NCBI's RefSeq accession no. NC_045512.2) as well as the first 100 complete genome sequences (≥29,000 bp) of each Variant of Concern (VOC) and Variant Being Monitored (VBM). The total number of input sequences was 1,301. We aligned all of these complete genomes using `MAFFT` version 7.475 (7) with the "auto" option and trimmed the alignment to remove the 5′-UTR and 3′-UTR regions. We also removed duplicated sequences or sequences with more than 5% of missing data, leaving us with 1,026 sequences.

We annotated each of the 1,026 remaining sequences using the strategy described in Machado et al. (8). Once we had all the predicted spike proteins for each of the 1,026 genomes, we aligned those sequences based on their translation with the help of `MAFFT` using the `TranslatorX` pipeline (9). We removed duplicated sequences and sequences with more than 5% of missing data. Finally, we identified the receptor binding motif on that alignment based on sequence similarity with the reference.

We then calculated the pairwise p-distances between each pair of sequences were calculated using `MEGA` version 11.0.10 (10). This distance is the proportion (p) of nucleotide sites at which two sequences being compared are different. The p-distances were calculated for the whole spike alignments (nucleotides) but also for the alignment of its receptor binding motif (RBM; position 430–522 of the spike amino acid sequence, a subset of the positions in the RBD).

This variant nucleotide sequence for the spike protein was then translated into amino acids using the standard translation table. This sequence was then trimmed to only contain the RBD of the spike protein (positions 319 to 541).

Receptor-Binding Domain Structural Prediction

Using the derived RBD amino acid sequence for Omicron, we used `AlphaFold2` and `RoseTTAFold` to create a predicted 3D protein structures. `AlphaFold2` is a neural network-based deep learning model created by Google DeepMind (11). The algorithm first searches for homologous sequences with existing structures to use as a scaffold on which to place the new sequence. `RoseTTAFold` is a similar neural network-based system by the Institute for Protein Design at the University of Washington (12).

The `AlphaFold2`-based prediction was run with the "single sequence" mode using the predicted TM-score (PTM) method. We also specified that the algorithm should run an Amber relaxation procedure to repair any structural violations in the predicted model (13). The `RoseTTAFold`-based prediction was run with the "mmseqs2" mode [by ColabFold (14)].

Both systems each resulted in a .PDB file of the predicted RBD structure for Omicron along with metrics surrounding the multiple sequence alignment coverage, predicted aligned error (PAE), and predicted confidence (pLDDT) by position (Available online at: https://github.com/colbyford/SARS-CoV-2_B.1.1.529_Spike-RBD_Predictions).

Given that this study focuses on antibodies that bind to the top of the RBD of the spike protein. Since that `AlphaFold2` and `RoseTTAFold` are template-based models that generate the predicted RBD structures using homologous sequences for which we have actual structures, we can avoid modeling the entire spike protein. See **Figure 1**.

Neutralizing Antibody Interaction Simulation

Using the predicted structures of the Omicron RBD, we simulated the interaction with four available neutralizing antibody structures: C105, CC12.1, CC12.3, and CV30 (PDBs: 6XCM, 6XC2, 6XC7, and 6XE1, respectively) (15–17). We used

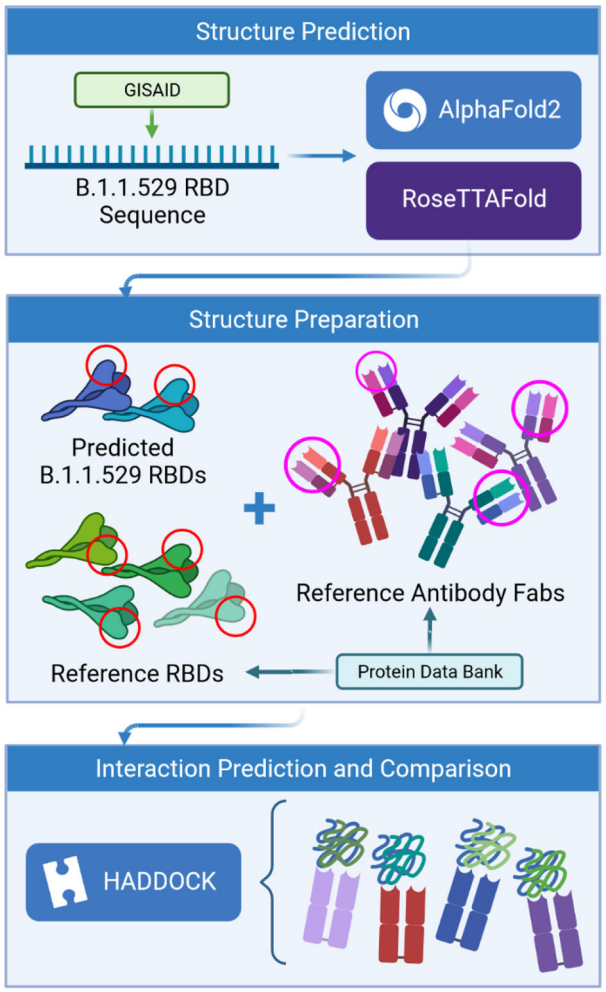


FIGURE 1 | Process flow of the prediction analysis steps.

TABLE 1 | List of analyses performed, comparing reference and predicted RBD structures in complex with reference Fab structures.

Antibody Fab	Analysis type	RBD source
C105 (6XCM, chains N and S)	Reference	6XCM (chain B)
	Prediction	B.1.1.529 (from `AlphaFold2`)
	Prediction	B.1.1.529 (from `RoseTTAFold`)
CC12.1 (6XC2, chains H and L)	Reference	6XC2 (chain A)
	Prediction	B.1.1.529 (from `AlphaFold2`)
	Prediction	B.1.1.529 (from `RoseTTAFold`)
CC12.3 (6XC7, chains C and D)	Reference	6XC7 (chain A)
	Prediction	B.1.1.529 (from `AlphaFold2`)
	Prediction	B.1.1.529 (from `RoseTTAFold`)
CV30 (6XE1, chains H and L)	Reference	6XE1 (chain B)
	Prediction	B.1.1.529 (from `AlphaFold2`)
	Prediction	B.1.1.529 (from `RoseTTAFold`)

only a single fragment antigen-binding (Fab) region of the antibody structures as the paratope location against which to dock. Each of the RBD structures from these reference files have identical sequences to the Wuhan-Hu-1 spike RBD.

Each of these neutralizing antibody structures were collected from patients who had been infected with SARS-CoV-2. Thus, these structures are serologically-derived antibodies rather than structures of therapeutic antibodies. All of them bind to the same "up" location of the S1 subunit of the spike protein (class I binders). This is a similar location to the interaction site between the human ACE2 receptor epitope. Thus, the neutralizing mechanism of these antibodies is in the prevention of SARS-CoV-2 binding to ACE2 on human cells.

We used `HADDOCK` version 2.4, a biomolecular modeling software that provides docking predictions for provided structures, to predict the binding affinity between the epitope of the RBD with the paratope of the neutralizing antibody structures (18). This takes in two or more .PDB files as inputs and outputs multiple predicted protein complexes in .PDB format along with docking metrics.

We first renumbered the residues according to `HADDOCK`'s requirements and then specified the interacting residues between the RBD structure and the Fab. Specifically, we ensure there are not overlapping residue IDs between the chains of a .PDB file and then specify the residues that are assumed to interact between the structures. This analysis was performed on the antibody-RBD structure pairs shown in **Table 1**.

The assessment of these interactions was measured by multiple biophysical factors including van der Waals energy, electrostatic energy, desolvation energy, and restraints violation energy, which were collectively used to derive a HADDOCK score to quantify changes in protein-protein interaction resulting from mutations in the RBD. Further, interfacing residues between the RBD and Fab structures were determined by identifying residues above the 1.0 Å^2 difference in surface area cutoff between the chains of the RBD and the Fab using the *InterafaceResidues*[1] functionality in `PyMol` version 2.4.1 (19).

We then compared the metrics of the actual complexes (i.e., the real RBD structure and the Fab) vs. the predicted RBD structures of Omicron (with the same Fab). This comparison provides a baseline interaction that was then measured against the mutated interactions with each respective Fab.

Differences between HADDOCK results were assessed using Kruskal-Wallis tests and *ad hoc* Wilcoxon pairwise comparisons. All statistical analyses were performed using `R` version 4.0.4 (20).

[1] *InterafaceResidues* Function Documentation: https://pymolwiki.org/index.php/InterfaceResidues.

Variant	1	2	3	4	5	6	7	8	9	10	11
1. Reference: Wuhan-Hu-1	0.00%										
2. VBM: B.1.617.3	0.19%	0.00%									
3. VBM: Beta	0.20%	0.32%	0.00%								
4. VBM: Epsilon	0.15%	0.23%	0.31%	0.00%							
5. VBM: Eta	0.23%	0.33%	0.33%	0.32%	0.00%						
6. VBM: Gamma	0.33%	0.43%	0.38%	0.41%	0.46%	0.00%					
7. VBM: Iota	0.17%	0.29%	0.27%	0.26%	0.33%	0.43%	0.00%				
8. VBM: Kappa	0.24%	0.20%	0.37%	0.28%	0.39%	0.49%	0.32%	0.00%			
9. VBM: Mu	0.22%	0.26%	0.27%	0.32%	0.33%	0.40%	0.27%	0.33%	0.00%		
10. VBM: Zeta	0.11%	0.22%	0.21%	0.21%	0.23%	0.29%	0.20%	0.27%	0.23%	0.00%	
11. VOC: Delta	0.23%	0.14%	0.39%	0.26%	0.40%	0.50%	0.34%	0.29%	0.33%	0.29%	0.00%
12. VOC: Omicron	0.81%	0.88%	0.90%	0.90%	0.88%	0.98%	0.84%	0.90%	0.79%	0.86%	0.88%

Proportion (p) 0.00% – 0.98%

FIGURE 2 | Genetic distance matrix of the spike gene (using nucleotides) for 9 Variants Being Monitored (VBM) and 2 Variants of Concern (VOC). The distance is the average proportion (*p*) of nucleotide sites at which two sequences being compared are different.

Variant	1	2	3	4	5	6	7	8	9	10	11
1. Reference: Wuhan-Hu-1	0.00%										
2. VBM: B.1.617.3	2.20%	0.00%									
3. VBM: Beta	2.20%	3.33%	0.00%								
4. VBM: Epsilon	1.09%	1.09%	3.33%	0.00%							
5. VBM: Eta	1.25%	2.36%	1.25%	2.36%	0.00%						
6. VBM: Gamma	2.20%	3.33%	0.00%	3.33%	1.25%	0.00%					
7. VBM: Iota	0.82%	2.67%	2.30%	1.92%	1.34%	2.30%	0.00%				
8. VBM: Kappa	2.20%	0.00%	3.33%	1.09%	2.36%	3.33%	2.67%	0.00%			
9. VBM: Mu	2.20%	3.33%	0.00%	3.33%	1.25%	0.00%	2.30%	3.33%	0.00%		
10. VBM: Zeta	1.13%	2.28%	1.21%	2.24%	0.28%	1.21%	1.22%	2.28%	1.21%	0.00%	
11. VOC: Delta	2.12%	2.12%	4.42%	1.02%	3.42%	4.42%	2.97%	2.12%	4.42%	3.30%	0.00%
12. VOC: Omicron	11.09%	12.43%	9.78%	12.43%	11.28%	9.78%	10.76%	12.43%	9.78%	11.09%	11.27%

Proportion (p) 0.00% – 12.43%

FIGURE 3 | Genetic distance matrix of the receptor binding motif (RBM) of spike gene (using amino acids) for 9 Variants Being Monitored (VBM) and 2 Variants of Concern (VOC). The distance is the average proportion (*p*) of nucleotide sites at which two sequences being compared are different.

RESULTS

Variant Sequence Comparison

Although the Omicron RBM (spike amino acid sequence, positions 430–522) can be efficiently categorized by nine characteristic mutations (S:N440K, S:G446S, S:S447N, S:T478K, S:E484A, S:Q493R, S:G496S, S:Q298R, S:N501Y), at least two of them (S:N440K and S:G446S) may be missing from some samples classified as Omicron. Also, some Omicron RBMs contains an additional mutation at S:Y505H.

The Omicron variant is the variant more distantly related to the reference genome (SARS-CoV-2 Wuhan-Hu-1; NCBI's RefSeq accession no. NC_045512.2) in the proportion of shared nucleotides. Also, Omicron is the variant that is more distantly related to Gamma. See **Figures 2**, **3**.

Mutational Analysis

Comparing the RBD of Omicron to the reference genome, there are 15 mutations, all of which are single amino acid substitutions. Most of the substitutions result in a change in the residue type (see **Table 2**).

The resulting RBD structure from `AlphaFold2` predicts that there is little conformational change from the reference structure. Conversely, there is a significant conformational change in the predicted RBD structure from `RoseTTAFold`. See **Figure 4**.

There are multiple mutated residues (shown in red in **Figures 5**, **6**) in positions that may affect the ability of a

TABLE 2 | Mutations in the receptor-binding domain (RBD) of the spike protein in the Omicron variant (B.1.1.529).

Position	Ref	Alt	Ref type	Alt type	Type difference
339	G	D	Non-polar	Negative	Yes
371	S	L	Polar	Non-polar	Yes
373	S	P	Polar	Non-polar	Yes
375	S	F	Polar	Non-polar	Yes
417	K	N	Positive	Polar	Yes
440	N	K	Polar	Positive	Yes
446	G	S	Non-polar	Polar	Yes
477	S	N	Polar	Polar	No
478	T	K	Polar	Positive	Yes
484	E	A	Negative	Non-polar	Yes
493	Q	R	Polar	Positive	Yes
496	G	S	Non-polar	Polar	Yes
498	Q	R	Polar	Positive	Yes
501	N	Y	Polar	Polar	No
505	Y	H	Polar	Positive	Yes

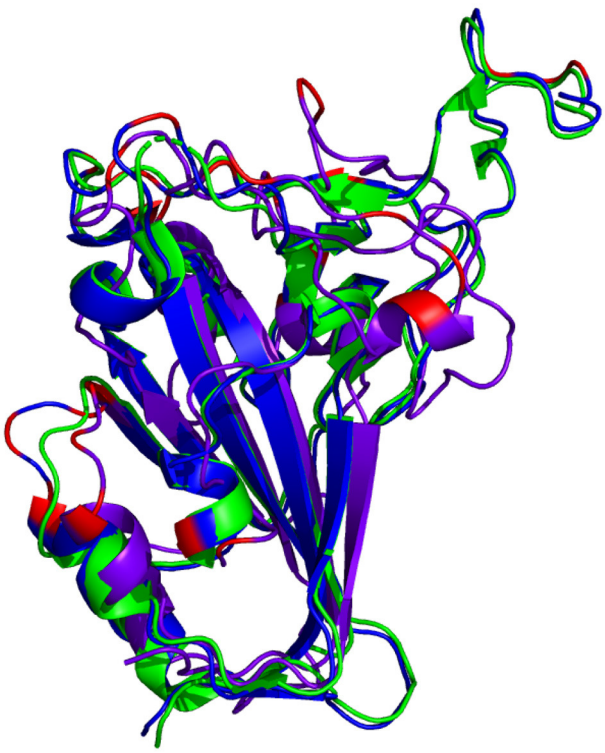


FIGURE 4 | Comparison of reference RBD structure (PDB: 6XC2, shown in green) and the predicted Omicron (B.1.1.529) RBD structures (`AlphaFold2` shown in blue, `RoseTTAFold` shown in purple). Mutated residues are highlighted in red.

FIGURE 5 | Possible inhibitory mutated RBD residues in the AlphaFold2 Omicron (B.1.1.529) (structure shown in blue with mutated residues of interest shown in red) superimposed on a reference RBD structure. Note: the reference RBD structure (PDB: 6XC2) is shown in green with equivalent position residues highlighted in yellow. CC12.1 antibody Fab (from PDB 6XC2) is shown in magenta/pink.

FIGURE 6 | Possible inhibitory mutated RBD residues in the RoseTTAFold Omicron (B.1.1.529) (structure shown in purple with mutated residues of interest shown in red) superimposed on a reference RBD structure. Note: the reference RBD structure (PDB: 6XC2) is shown in green with equivalent position residues highlighted in yellow. CC12.1 antibody Fab (from PDB 6XC2) is shown in magenta/pink.

neutralizing antibody to sufficiently bind. Some of these mutated residues change to much longer side-chained or differently-charged amino acids. For example, there are two "to lysine" mutations: N440K and T478K (i.e., from polar, smaller side chain residues to a positive-charged, longer side chain residue). These types of changes may have an effect on the binding affinity between the RBD and an antibody, either by changing the surface charge on the protein or by inhibiting a tighter antibody interaction.

Superimposing the predicted RBDs on the reference RBD with the CC12.1 antibody in place, shown in **Figures 5**, **6**, shows that mutations Q493R, Q498R, and N501Y in the AlphaFold2 structure and the E484A mutation in the RoseTTAFold structure may affect the binding position of the antibody. These side chains clash with particular residues of the Fab, which may cause a less effective and more distant binding. Longer/larger side chains may increase the distance between the Fab paratope of the antibody and the epitope of the RBD.

Antibody Binding Analyses

The results of all four antibody docking exercises with the predicted AlphaFold2 RBD structure show that the Fab of the respective neutralizing antibodies continue to bind to the RBD of Omicron, though not as well as the reference interaction. Note that there is a consistent decrease (increase in value) in the electrostatic energy and an increase in restraints violation energy between the binding from the reference RBDs and the predicted RBD of Omicron. The HADDOCK score is worse (higher) across the board and it appears that the interaction of the Omicron RBD with the antibodies are more distant, as shown by the buried surface area changes below.

TABLE 3 | HADDOCK metrics for the C105 docking prediction, comparing the 6XCM RBD vs. the Omicron (B.1.1.529) predicted RBD structures.

Metric	6XCM RBD w/C105	Predicted B.1.1.529 RBD w/C105	
		AlphaFold	RoseTTAFold
van der Waals energy	–85	–71.8 (16%)	–44.2 (48%)
Electrostatic energy	–280.9	–268.9 (4%)	–324.5 (–16%)
Desolvation energy	-18.6	–17.1 (8%)	–3.2 (83%)
Restraints violation energy	154.1	158.8 (3%)	338.6 (120%)
HADDOCK score	–144.4	–126.8 (12%)	–78.4 (46%)
Buried surface area	2417.8	2299.6 (–5%)	1555.1 (–36%)

When looking at the antibody docking exercises with the predicted RoseTTAFold RBD structure, there is agreement with the AlphaFold2 results in that there still seems to be interaction with the neutralizing antibodies. However, the reduction in binding affinity is much more severe given the conformational changes seen only in the RoseTTAFold structure.

Note: All values in **Tables 3–6** below represent the best docking predictions from HADDOCK. Also, values in parentheses

TABLE 4 | HADDOCK metrics for the CC12.1 docking prediction, comparing the 6XC2 RBD vs. the Omicron (B.1.1.529) predicted RBD structures.

Metric	6XC2 RBD w/CC12.1	Predicted B.1.1.529 RBD w/CC12.1	
		AlphaFold	RoseTTAFold
van der Waals energy	–106.9	–90.7 (15%)	–27.5 (74%)
Electrostatic energy	–342.6	–284.9 (17%)	–259.2 (24%)
Desolvation energy	–37.5	–28.8 (23%)	–12.4 (67%)
Restraints violation energy	143.2	152.6 (7%)	201.4 (41%)
HADDOCK score	–198.6	–163.3 (18%)	–71.6 (64%)
Buried Surface Area	2778.5	2584.3 (–7%)	1578.4 (–43%)

TABLE 5 | HADDOCK metrics for the CC12.3 docking prediction, comparing the 6XC7 RBD vs. the Omicron (B.1.1.529) predicted RBD structures.

Metric	6XC7 RBD w/CC12.3	Predicted B.1.1.529 RBD w/CC12.3	
		AlphaFold	RoseTTAFold
van der Waals energy	–86.1	–93.1 (–8%)	–48.1 (44%)
Electrostatic energy	–248.1	–194.1 (22%)	–189.2 (24%)
Desolvation energy	–41.7	–37.9 (9%)	–23.5 (44%)
Restraints violation energy	159.1	104.7 (–34%)	221.4 (39%)
HADDOCK score	–161.5	–159.4 (1%)	–87.4 (46%)
Buried Surface Area	2489.8	2416.7 (–3%)	1420.6 (–43%)

TABLE 6 | HADDOCK metrics for the CV30 docking prediction, comparing the 6XE1 RBD vs. the Omicron (B.1.1.529) predicted RBD structures.

Metric	6XE1 RBD w/CV30	Predicted B.1.1.529 RBD w/CV30	
		AlphaFold	RoseTTAFold
van der Waals energy	–86.1	–61.7 (28%)	–41.1 (52%)
Electrostatic energy	–354.4	–171.8 (52%)	–207.9 (41%)
Desolvation energy	–13.8	–19.5 (–41%)	–10.2 (26%)
Restraints violation energy	147.6	159.4 (8%)	286.0 (93%)
HADDOCK score	–156.0	–99.6 (36%)	–43.9 (71%)
Buried Surface Area	2479.0	1992.6 (–20%)	1600.2 (–35%)

represent the percentage difference between the given metric for the predicted structure and the reference structure.

C105 Antibody Binding

Resulting binding metrics from the C105 HADDOCK docking analysis are shown in **Table 3**. This interaction shows that there is a ∼ 4% reduction in the electrostatic energy and an ∼ 5% decrease in buried surface area comparing between the 6XCM RBD and the predicted RBD of Omicron from AlphaFold2. From RoseTTAFold, the interaction shows that there is ∼ 16% increase in the electrostatic energy but with a ∼ 36% decrease in buried surface area. See **Figure 7**.

FIGURE 7 | HADDOCK docking prediction using C105 (shown in magenta/pink), comparing the 6XCM RBD (shown in shown in green) vs. the Omicron (B.1.1.529) predicted RBD structures (AlphaFold2 shown in blue, RoseTTAFold shown in purple).

CC12.1 Antibody Binding

Resulting binding metrics from the CC12.1 HADDOCK docking analysis are shown in **Table 4**. This interaction shows that there is a ∼ 17% reduction in the electrostatic energy and an ∼ 7% decrease in buried surface area comparing between the 6XC2 RBD and predicted RBD of Omicron from AlphaFold2. From RoseTTAFold, the interaction shows that there is ∼ 24% reduction in the electrostatic energy but and a ∼ 43% decrease in buried surface area. See **Figure 8**.

CC12.3 Antibody Binding

Resulting binding metrics from the CC12.3 HADDOCK docking analysis are shown in **Table 4**. This interaction shows that there is a ∼ 22% reduction in the electrostatic energy and an ∼ 3% decrease in buried surface area comparing between the 6XC7 RBD and predicted RBD of Omicron from AlphaFold2. From RoseTTAFold, the interaction shows that there is ∼ 24% reduction in the electrostatic energy and a ∼ 43% decrease in buried surface area. See **Figure 9**.

CV30 Antibody Binding

Resulting binding metrics from the CV30 HADDOCK docking analysis are shown in **Table 3**. This interaction shows that there is a ∼ 52% reduction in the electrostatic energy and an ∼ 20%

FIGURE 8 | HADDOCK docking prediction using CC12.1 (shown in magenta/pink), comparing the 6XC2 RBD (shown in shown in green) vs. the Omicron (B.1.1.529) predicted RBD structures (`AlphaFold2` shown in blue, `RoseTTAFold` shown in purple).

FIGURE 9 | HADDOCK docking prediction using CC12.3 (shown in magenta/pink), comparing the 6XC7 RBD (shown in shown in green) vs. the Omicron (B.1.1.529) predicted RBD structures (`AlphaFold2` shown in blue, `RoseTTAFold` shown in purple).

decrease in buried surface area comparing between the 6XE1 RBD and predicted RBD of Omicron from `AlphaFold2`. From `RoseTTAFold`, the interaction shows that there is ∼ 41% reduction in the electrostatic energy and a ∼ 35% decrease in buried surface area. See **Figure 10**.

Antibody Interaction Comparison

All of the interaction predictions among the four antibodies tested in this study (C105, CC12.1, CC12.3, and CV30) agree that there is a decrease in binding affinity when comparing the respective RBD interactions with the Omicron RBD interactions. Across all of the docking predictions using the `AlphaFold2` RBD structure, we see a drop in electrostatic interaction (increase in the electrostatic energy value) ranging from ∼ 4% to ∼ 52% and a consistent decrease in buried surface area (increase in distance) of the RBD and the antibody Fab. In addition, we see a variable increase (worsening) in the HADDOCK score, indicating that all of the Omicron RBD structures have a lower binding affinity when compared to their respective reference RBD structures as a benchmark. See **Figure 11**.

Similarly, we see a more extreme reduction in the binding affinity of the Fab structures and the RBD structure predicted by `RoseTTAFold`. We see a drop in electrostatic interaction (increase in the electrostatic energy value) ranging from ∼ 24% to ∼ 41% in three of the four interactions (with an odd ∼ 16% increase in the C105 interaction) and a consistent ∼ 35% to ∼ 43% decrease in buried surface area (increase in distance) of the RBD and the antibody Fab.

We performed Kruskal-Wallis tests with each of the HADDOCK metrics, which show that there is a statistically-significant difference between the three sets of binding experiments (Alpha and the two predicted Omicron sets). See

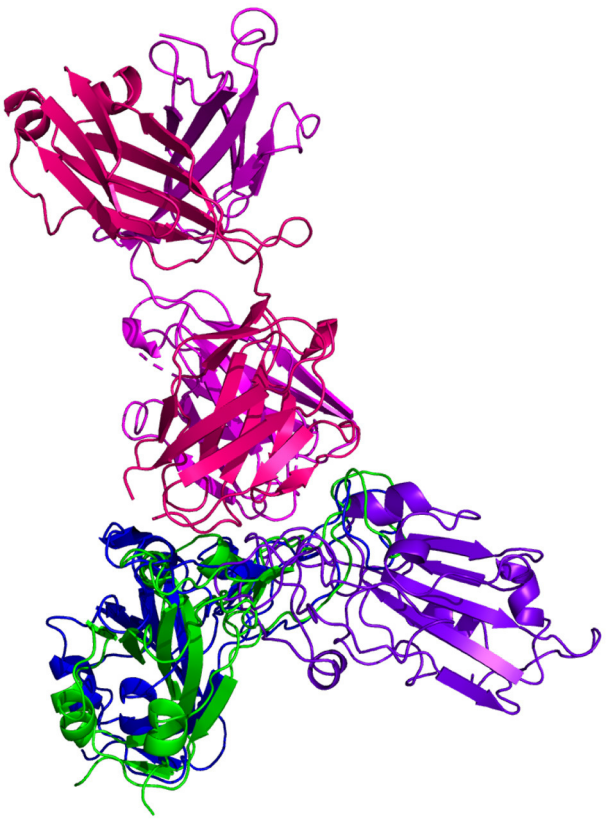


FIGURE 10 | HADDOCK docking prediction using CV30 (shown in magenta/pink), comparing the 6XE1 RBD (shown in shown in green) vs. the Omicron (B.1.1.529) predicted RBD structures (`AlphaFold2` shown in blue, `RoseTTAFold` shown in purple).

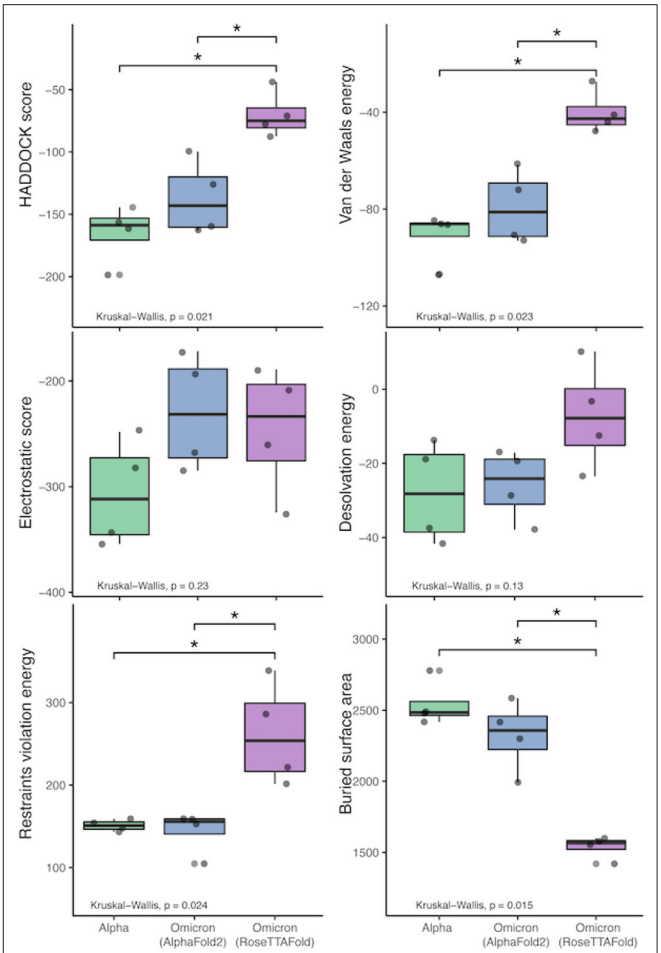


FIGURE 11 | HADDOCK results comparison between the reference RBD structures and the predicted Omicron (B.1.1.529) RBD structures. The bars indicate results that are significantly different and the "*" indicates Wilcoxon Rank-Sum's p-value ≤ 0.005).

Figure 11. Interestingly, performing the Wilcoxon Rank-Sum tests on these metrics to compare the differences between the predictions and reference results shows that there is no statistically-significant difference at the $\alpha = 0.05$ level between the `AlphaFold2` RBD structure binding and the reference structures. However, when comparing the `RoseTTAFold` RBD structure to the reference results, differences in the restraints violation energy, van der Waals energy, HADDOCK score, and buried surface area are statistically significant at the $\alpha = 0.05$ level and the differences in desolvation energy is statistically significant at the $\alpha = 0.10$ level. Also, the same metrics are significantly different between the `AlphaFold2` and `RoseTTAFold`, further showing that these predicted structures are quite different from one another.

Fab-RBD Interfacing Residues

Furthermore, when comparing residues that are interfaced between the Fab and RBD, there is agreement in that particular residues in Omicron are no longer interfacing with the antibodies analyzed in this study. In particular, residues 448N, 484A, and 494S may not interface with the Fab structure as they are in the reference RBD-Fab complexes. However, the aforementioned N501Y and S477N mutations (along with a variety of other mutations) do not appear to affect the interfacing of the residues at these positions.

This implies that there are certain positions that are more sensitive to mutations in that substitutions at these loci are more likely to affect the interface of the RBD with the antibody's Fab (denoted by a △ symbol in **Table 7**).

In contrast, there are other positions that have been substituted between the reference RBDs and the predicted Omicron RBDs that continue to interface with most of the Fab structures (denoted by a ♦ symbol in **Table 7**).

Finally, there are some residues that remain unchanged in the Omicron variant RBD amino acid sequence relative to the Wuhan reference isolate, yet we see changes in the interfacing at these loci (denoted by a ○ symbol in **Table 7**). This suggests that there are other mutated residues around these stable positions that may be affecting their ability to interface.

Many of these findings concur with the findings in Sharma et al. (21). Specifically, we also continue to see a reliance on residues at positions K417, S477, Q493, Q498, N501, and

TABLE 7 | Interfacing residue changes of interest between the Fab paratope and the RBD structures.

Interface position		417 (K▶N)	448 (N)	477 (S▶N)	484 (E▶A)	493 (Q▶R)	494 (S)	498 (Q▶R)	501 (N▶Y)	505 (Y▶H)
6XC2 vs. B.1.1.529	Reference	K	N	S	E	Q	S	Q	N	Y
	AlphaFold2 Prediction	N	-	N	-	R	-	R	Y	H
	RoseTTAFold Prediction	-	N	N	A	-	-	-	-	H
6XC7 vs. B.1.1.529	Reference	K	N	S	E	Q	S	Q	N	Y
	AlphaFold2 Prediction	N	-	N	-	R	-	R	Y	H
	RoseTTAFold Prediction	-	-	N	-	R	-	R	Y	-
6XCM vs. B.1.1.529	Reference	K	-	S	E	Q	S	Q	N	Y
	AlphaFold2 Prediction	N	-	N	-	R	S	R	Y	H
	RoseTTAFold Prediction	-	N	N	A	-	-	-	-	H
6XE1 vs. B.1.1.529	Reference	K	-	S	-	Q	S	Q	N	Y
	AlphaFold2 Prediction	N	-	-	A	R	S	R	Y	H
	RoseTTAFold Prediction	-	N	N	A	-	-	-	-	H
Legend: ♦ - Mutated positions with few interface changes, ○ - Non-mutated positions with interface changes, △ - Mutated positions with interface changes		△	○	♦	△	△	○	△	△	♦

(Note that a '-' means that the residue at this position no longer interfaces with the Fab structure).

Y505, which are stated to increase binding affinity. Interestingly, the interfacing residues that are non-mutated N448 and S494 (denoted with a ○ symbol) are rarely, if ever, listed as interfacing residues in **Table 3** in Sharma et al. (21). This further supports the notion that there are residue positions that are important for interfacing with neutralizing antibodies, but that the mutations seen here in the Omicron RBD may not seriously affect this RBD-antibody interface.

CONCLUSION AND DISCUSSION

While *in vitro* experiments are needed to validate these predictions, the predicted results here suggest that existing neutralizing antibodies may still bind to the mutated spike protein of the Omicron variant. However, it appears that the affinity of Omicron's RBD for neutralizing antibodies is reduced compared to the reference RBD structures. The results of both AlphaFold2 and RoseTTAFold suggest that antibodies elicited from previous infection will provide at least some protection against Omicron. Additionally, these results indicate that the SARS-CoV-2 Omicron variant will note completely evade vaccines based on the spike protein.

Though there are many mutations in the RBD of Omicron, the predicted structure from AlphaFold2 suggests these mutations do not appear to be causing any large conformational change that would totally evade antibody interaction. However, we do see some amino acid substitutions to different, longer side chained residues at the binding site. This result may be due to the slightly more distant interaction with the antibody and therefore may reduce the binding affinity.

The results of RoseTTAFold suggest that the mutations have a different effect on the overall 3D structure of the Omicron RBD. The conformational change seen in this structure may contribute to antibody evasion or more severely reduce antibody binding affinity.

AlphaFold2 has not been validated for predicting the effect of mutations and is not expected to produce a completely unfolded structure if the Omicron sequence contains any destabilizing point mutations. This may explain the extremely similar backbone structure to the reference RBD structures.

Given that the antibody docking results between AlphaFold2 and RoseTTAFold are statistically different, this further posits that this potential conformational change seen in the RoseTTAFold RBD may directly affect the binding of neutralizing antibodies.

Though the predicted RBD structures between AlphaFold2 and RoseTTAFold are quite different, the HADDOCK antibody docking results seem to converge on specific positions with which the Fab will interface. S477N and Y505H may be useful positions on which to focus for future vaccine design or mutational surveillance in new variants.

Further analyses are needed using a broader range of different classes of antibodies, including therapeutic antibodies and neutralizing antibodies that bind to other locations on the spike protein. Recent preprint articles show more drastic reductions in the binding affinity of some other antibodies like CB6, a neutralizing antibody similar to ones used in this study, as well as a variety of therapeutic antibodies (22, 23). Thus, while we fail to see complete evasion of the antibodies used in this study, there are far more Omicron-antibody interactions to be understood.

Once a true structure of the Omicron RBD is determined, it will be of interest to compare the true structure to both of the predicted RBD structures from AlphaFold2 and RoseTTAFold, mainly to see if a large conformational change occurs (as is seen in the RoseTTAFold) or if the overall backbone/3D structure is very similar to the reference structures used in this study (as is seen in the AlphaFold2). In addition, it will be necessary to validate the antibody interactions predicted using HADDOCK with true, experimentally-derived binding measurements.

Determining the actual structure of a protein is a time-consuming process. Further, quantifying protein-protein interactions (like spike-to-antibody interactions) are also experimentally difficult to perform *in vitro*. Given the public health urgency in understanding the impacts of new SARS-CoV-2 variants quickly requires that we act quicker than is possible in a lab. Thus, *in silico* predictive tools like AlphaFold2, RoseTTAFold, and HADDOCK are important for quickly understanding the biochemistry of variants and can help us to infer the epidemiological implications of the variant.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://github.com/colbyford/SARS-CoV-2_B.1.1.529_Spike-RBD_Predictions,GitHub (Zenodo: https://doi.org/10.5281/zenodo.5733161).

AUTHOR CONTRIBUTIONS

CF created the predicted RBD structures and performed the antibody docking analyses. DJ and DAJ performed the sequence analysis and interpretation for the comparison of the variants. All authors wrote and reviewed the manuscript.

ACKNOWLEDGMENTS

We acknowledge these entities at UNC Charlotte: The Department of Bioinformatics and Genomics, the Bioinformatics Research Center, the College of Computing and Informatics, the College of Liberal Arts and Sciences, Research and Economic Development, University Research Computing, and the Graduate School. We also thank the Belk family for support. Also, we appreciate the initial feedback from members of the Google DeepMind team on the usage of AlphaFold2 for point mutations. This prompted our further investigation and comparison with RoseTTAFold and the differing results we now show in this study.

REFERENCES

1. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. Nature. (2020) 579:265–9. doi: 10.1038/s41586-020-2008-3
2. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol. (2020) 5:1403–7. doi: 10.1038/s41564-020-0770-5
3. World Health Organization. Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern (2021). Available online at: https://www.who.int/news/item/26-11-2021-classification-of-omicron-%28b.1.1.529%29-sars-cov-2-variant-of-concern (accessed December 1, 2021).
4. Farooqi T, Malik JA, Mulla AH, Al Hagbani T, Almansour K, Ubaid MA, et al. An overview of SARS-CoV-2 epidemiology, mutant variants, vaccines, and management strategies. J Infect Public Health. (2021) 14:1299–312. doi: 10.1016/j.jiph.2021.08.014
5. Brito AF, Semenova E, Dudas G, Hassler GW, Kalinich CC, Kraemer MUG, et al. Global disparities in SARS-CoV-2 genomic surveillance. medRxiv. (2021) doi: 10.1101/2021.08.21.21262393
6. Machado DJ, White III RA, Kofsky J, Janies DA. Fundamentals of genomic epidemiology, lessons learned from the coronavirus disease 2019 (COVID-19) pandemic, and new directions. Antimicrob Stewardship Healthcare Epidemiol. (2021). 1:E6. doi: 10.1017/ash.2021.222
7. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evolut. (2013) 30:772–80. doi: 10.1093/molbev/mst010
8. Jacob Machado D, Scott R, Guirales S, Janies DA. Fundamental evolution of all Orthocoronavirinae including three deadly lineages descendent from Chiroptera-hosted coronaviruses: SARS-CoV, MERS-CoV and SARS-CoV-2. Cladistics. (2021) 37:461–88. doi: 10.1111/cla.12454
9. Abascal F, Zardoya R, Telford MJ. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. (2010) 38:W7–W13. doi: 10.1093/nar/gkq291
10. Stecher G, Tamura K, Kumar S. Molecular evolutionary genetics analysis (MEGA) for macOS. Mol Biol Evolut. (2020) 37:1237–9. doi: 10.1093/molbev/msz312
11. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. (2021) 596:583–9. doi: 10.1038/s41586-021-03819-2
12. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. Science. (2021) 373:871–6. doi: 10.1126/science.abj8754
13. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins. (2006) 65:712–25. doi: 10.1002/prot.21123
14. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold-Making protein folding accessible to all. bioRxiv. (2021) doi: 10.1101/2021.08.15.456425
15. Barnes CO, West AP Jr, Huey-Tubman KE, Hoffmann MAG, Sharaf NG, Hoffman PR, et al. Structures of human antibodies bound to SARS-CoV-2 spike reveal common epitopes and recurrent features of antibodies. Cell. (2020) 182:828–842.e16. doi: 10.1016/j.cell.2020.06.025
16. Yuan M, Liu H, Wu NC, Lee CCD, Zhu X, Zhao F, et al. Structural basis of a shared antibody response to SARS-CoV-2. Science. (2020) 369:1119–23. doi: 10.1126/science.abd2321
17. Hurlburt NK, Seydoux E, Wan YH, Edara VV, Stuart AB, Feng J, et al. Structural basis for potent neutralization of SARS-CoV-2 and role of antibody affinity maturation. Nat Commun. (2020) 11:5413. doi: 10.1038/s41467-020-19231-9
18. van Zundert GCP, Rodrigues JPGLM, Trellet M, Schmitz C, Kastritis PL, Karaca E, et al. The HADDOCK2.2 web server: user-friendly integrative

modeling of biomolecular complexes. J Mol Biol. (2016) 428:720–5. doi: 10.1016/j.jmb.2015.09.014
19. Schrödinger LLC. The PyMOL Molecular Graphics System, version 2.4.1 (2015). Available online at: http://www.pymol.org/pymol (accessed November 29,2021).
20. R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Core Team (2021).
21. Sharma D, Rawat P, Janakiraman V, Gromiha MM. Elucidating important structural features for the binding affinity of spike - SARS-CoV-2 neutralizing antibody complexes. Proteins. (2021) 1–11. doi: 10.1002/prot.26277
22. Liu L, Iketani S, Guo Y, Chan JFW, Wang M, Liu L, et al. Striking antibody evasion manifested by the omicron variant of SARS-CoV-2. bioRxiv. (2021) doi: 10.1101/2021.12.14.472719
23. VanBlargan LA, Errico JM, Halfmann PJ, Zost SJ, Crowe JE, Purcell LA, et al. An infectious SARS-CoV-2 B.1.1.529 Omicron virus escapes neutralization by several therapeutic monoclonal antibodies. bioRxiv. (2021) doi: 10.21203/rs.3.rs-1175516/v1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 23 February 2022
doi: 10.3389/fmicb.2022.829094

Interactions of Severe Acute Respiratory Syndrome Coronavirus 2 Protein E With Cell Junctions and Polarity PSD-95/Dlg/ZO-1-Containing Proteins

Yanlei Zhu[1†], Flavio Alvarez[1†], Nicolas Wolff[1], Ariel Mechaly[2], Sébastien Brûlé[3], Benoit Neitthoffer[4], Sandrine Etienne-Manneville[4], Ahmed Haouz[2], Batiste Boëda[4‡] and Célia Caillet-Saguy[1‡]*

[1] Channel Receptors Unit, CNRS, UMR 3571, Institut Pasteur, Université de Paris, Paris, France, [2] Crystallography Platform-C2RT, CNRS, UMR 3528, Institut Pasteur, Université de Paris, Paris, France, [3] Molecular Biophysics Platform-C2RT, CNRS, UMR 3528, Institut Pasteur, Université de Paris, Paris, France, [4] Cell Polarity, Migration and Cancer Unit, Institut Pasteur, UMR 3691 CNRS, Université de Paris, Equipe Labellisée Ligue Contre le Cancer, Paris, France

The C-terminus of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) protein E contains a PBM (PDZ-binding motif) targeting PDZ (PSD-95/Dlg/ZO-1) domains, which is identical to the PBM of SARS-CoV. The latter is involved in the pathogenicity of the virus. Recently, we identified 10 human PDZ-containing proteins showing significant interactions with SARS-CoV-2 protein E PBM. We selected several of them involved in cellular junctions and cell polarity (TJP1, PARD3, MLLT4, and LNX2) and MPP5/PALS1 previously shown to interact with SARS-CoV E PBM. Targeting cellular junctions and polarity components is a common strategy by viruses to hijack cell machinery to their advantage. In this study, we showed that these host PDZ domains TJP1, PARD3, MLLT4, LNX2, and MPP5/PALS1 interact in a PBM-dependent manner *in vitro* and colocalize with the full-length E protein *in cellulo*, sequestrating the PDZ domains to the Golgi compartment. We solved three crystal structures of complexes between human LNX2, MLLT4, and MPP5 PDZs and SARS-CoV-2 E PBM highlighting its binding preferences for several cellular targets. Finally, we showed different affinities for the PDZ domains with the original SARS-CoV-2 C-terminal sequence containing the PBM and the one of the beta variant that contains a mutation close to the PBM. The acquired mutations in the E protein localized near the PBM might have important effects both on the structure and the ion-channel activity of the E protein and on the host machinery targeted by the variants during the infection.

Keywords: SARS-CoV-2, envelope protein, host–pathogen interactions, PDZ-binding motif, PDZ-containing protein, protein–protein interaction, cell junctions and polarity

OPEN ACCESS

Edited by:
Eva Mittler,
Albert Einstein College of Medicine, United States

Reviewed by:
Jérôme Bouchet,
Université Paris Descartes, France
Marc Kvansakul,
La Trobe Institute for Molecular Science, La Trobe University, Australia

****Correspondence:***
Célia Caillet-Saguy
celia.caillet-saguy@pasteur.fr

† These authors have contributed equally to this work and share first authorship

‡ These authors have contributed equally to this work and share last authorship

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *04 December 2021*
Accepted: *06 January 2022*
Published: *23 February 2022*

Citation:
Zhu Y, Alvarez F, Wolff N, Mechaly A, Brûlé S, Neitthoffer B, Etienne-Manneville S, Haouz A, Boëda B and Caillet-Saguy C (2022) Interactions of Severe Acute Respiratory Syndrome Coronavirus 2 Protein E With Cell Junctions and Polarity PSD-95/Dlg/ZO-1-Containing Proteins. Front. Microbiol. 13:829094. doi: 10.3389/fmicb.2022.829094

INTRODUCTION

Currently, there have been more than 263 million individuals infected by SARS-CoV-2, including more than 5.2 million deaths.[1] While significant constant advances are made in understanding this virus, the knowledge of the SARS-CoV-2 protein interactions with host cell proteins is still limited.

Targeting multiple cellular PDZ (PSD-95/Dlg/ZO-1)-containing proteins through short linear PDZ-binding motifs (PBMs) is a common strategy used by viruses to facilitate their viral replication and dissemination to new hosts (Javier and Rice, 2011). Indeed, PDZ proteins are involved in processes of particular interest in viral infection: cell junction formation, cell polarity establishment, and immune system signaling (Javier and Rice, 2011; James and Roberts, 2016; Gutiérrez-González and Santos-Mendoza, 2019). The two SARS-CoV-2 viroporins, proteins E and 3a, contain a C-terminal PBM targeting specific PDZ domain-containing proteins (Castaño-Rodriguez et al., 2018; Caillet-Saguy et al., 2021).

PDZ domains are a large family of protein–protein interaction domains widespread in the human proteome (Luck et al., 2012). PBMs are mainly located at the C-terminus of target proteins and interact directly with PDZ domains. They are classified into three types: type I PBM (-X-S/T-X-ϕ_{COOH}), type II PBM (-X-ϕ-X-ϕ_{COOH}), and type III PBM (-X-D/E-X-ϕ_{COOH}), with ϕ signifying a hydrophobic residue. The C-terminal PBM sequence of SARS-CoV E protein is of type II (-DLLV$_{COOH}$) and has been identified as a virulence factor (Jimenez-Guardeño et al., 2014). It is likely that the abilities to target PDZ proteins also make significant contributions to the pathogenesis of the E protein of SARS-CoV-2. Indeed, the E protein is highly conserved with 94.7% identity between SARS-CoV and SARS-CoV-2, and their PBMs are strictly conserved.

Protein E is a small transmembrane protein of 75 residues involved in several phases of the virus life cycle, such as assembly, budding, envelope formation, and pathogenesis (Schoeman and Fielding, 2019). Protein E has a hydrophobic helical transmembrane domain of 30 residues flanked by an N-terminal domain of 8 residues and a C-terminal domain of 37 residues (**Figure 1A**; Mandala et al., 2020). It is present in virions in small quantities, whereas it is very abundant in infected cells at the level of the intermediate compartment between the Golgi and the endoplasmic reticulum (ERGIC) where its expression increases the pH inside the organelle (Cabrera-Garcia et al., 2021). Protein E actively participates in the budding, morphogenesis, and trafficking of the virus. A cytoplasmic orientation of the C-terminal domain and a luminal orientation for the N-terminal domain has been reported (Nieto-Torres et al., 2011; Duart et al., 2020). The E protein can oligomerize *via* its transmembrane domain into a homopentameric ion channel called viroporin that inserts into the host cell endomembrane system (Li et al., 2014) possibly modulated by changes in pH (Cabrera-Garcia et al., 2021). Recently, the pentameric structure of the transmembrane domain within a lipid bilayer reconstituting the ERGIC membrane has been reported by NMR in the absence of the cytoplasmic part (Mandala et al., 2020). The C-terminal cytoplasmic part of protein E is important for interactions with different partners such as its PBM sequence to interact with PDZ-containing proteins.

Previously, we used a high-throughput quantitative approach using a library covering all the human PDZ domains (Vincentelli et al., 2015) to establish the list of PDZ-containing proteins potentially targeted by SARS-CoV-2 E protein through its PBM. Among the 10 PDZ-containing proteins identified, four are involved in cellular junction and polarity: ZO-1 (also called TJP1), LNX2, PARD3, and MLLT4 (also called Afadin) (Caillet-Saguy et al., 2021; **Table 1**).

Indeed, ZO-1 is one of the essential proteins that connect transmembrane tight-junction proteins to the actin cytoskeleton. It consists of three PDZ domains, the second being the one that has a high affinity for the PBM motif of the E protein of SARS-CoV-2 (Caillet-Saguy et al., 2021; Shepley-McTaggart et al., 2021). Robinot et al. (2021) reported a disruption of the epithelial barrier integrity and an alteration of the ZO-1 distribution at the tight junctions (TJs) during infection with SARS-CoV-2. LNX2 acts as a molecular scaffold for Numb family proteins, essential players in the regulation of cell adhesion and polarity (Wang et al., 2009). PARD3 is an essential protein in asymmetric cell division and in polarized growth. It plays a central role in the establishment of TJs (Chen et al., 2017). MLLT4 is an adapter protein linking nectin to the actin cytoskeleton, essential for the formation of adherent junctions and the regulation of cell adhesion (Ikeda et al., 1999). In addition, the E protein of SARS-CoV was previously reported to interact with the PDZ-containing protein MPP5 (also called PALS1) through PBM–PDZ interaction altering TJ formation and the mammalian epithelium structure (Teoh et al., 2010). A recent study reported that the E protein PBM recognizes a pocket formed by residues from the PDZ and SH3 domains of MPP5 (Chai et al., 2021). As a key component of the Crumbs complex that controls the apical–basal polarity and TJ formation (Roh et al., 2002), MPP5 may contribute to the lung epithelium breakdown observed in patients infected by SARS-CoV. The relevance of these five PBM/PDZ interactions is not yet fully understood. Understanding the pathogenicity of this virus is currently a global health issue and targeting these virus–host interactions could reduce damage to the respiratory tract barrier and moderate the virus spread.

In this study, we showed that these five PDZ domains interact *in vitro* with the full-length E protein in a PBM-dependent manner and colocalize with the full-length E protein *in cellulo*, sequestrating the PDZ domains to the Golgi compartment. We further solved the three crystal structures of human PDZ/SARS-CoV-2 E PBM complexes for LNX2, MLLT4, and MPP5, highlighting their specific binding modes (**Table 1**). A mutation in the E protein localized near the PBM (P71L) was reported in the variant of concern (VOC) beta (B.1.351) (**Figure 1B**). We showed here different affinities between the PDZ domains and the SARS-CoV-2 WT and the beta variant E C-terminals encompassing the PBM (**Figure 1C**). Thus, the acquired mutations might have important consequences on

[1] https://covid19.who.int/

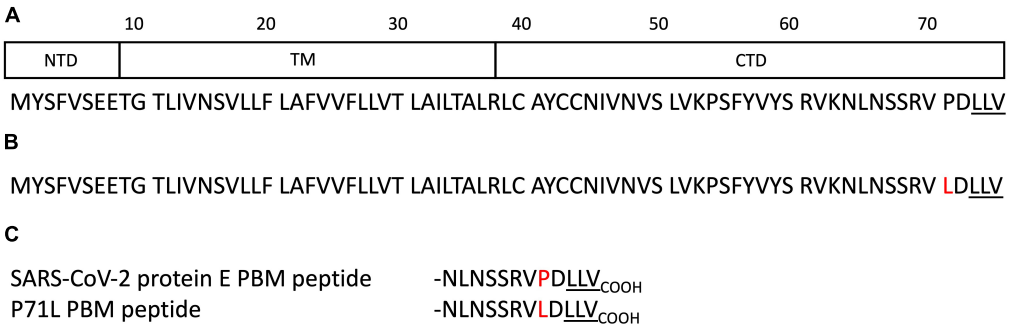


FIGURE 1 | Schematic domain representation and sequences of the protein E constructs. **(A)** Domain delimitations and sequence of the full-length protein E of SARS-CoV-2. NTD, TM, and CTD correspond to the N-terminal domain, the transmembrane domain, and the C-terminal domain respectively. **(B)** Sequence of the full-length protein E of SARS-CoV-2 beta variant (B.1.351). **(C)** Sequences of the SARS-CoV-2 WT and the beta variant E C-terminal encompassing the PSD-95/Dlg/ZO-1-binding motif (PBM). The PBMs are underlined. The mutation is highlighted in red.

TABLE 1 | Selected human PSD-95/Dlg/ZO-1 (PDZ) domains targeted by the SARS-CoV-2 E protein PSD-95/Dlg/ZO-1-binding motif (PBM).

PDZ domain	Uniprot	Delimitations	PDB structure/alphafold model	α 2 residue facing p-2
ZO-1 PDZ2	Q07157	184-264	2JWE	Leu
LNX2 PDZ2	Q8N448	334-426	7QCT[a]	Pro
PARD3 PDZ3	Q8TEW0	582-685	Alphafold prediction (Jumper et al., 2021)	Asn
MLLT4	P55196	1002-1095	7QCR[a]	Gln
MPP5	Q8N3R9	238-336	7QCS[a]	Val

PDZ name, Uniprot code, delimitations of the constructs used in this work (residue numbers correspond to the numbering in the full-length human protein), PDB of the structure model used, and the residue of PDZ helix α2 facing the PBM residue at position -2.
[a]*Structure solved in this work.*

host machinery targeted during the infection in addition to a potential effect both on the structure and the ion-channel activity of the E protein.

RESULTS

PDZ Domains of the Proteins Involved in Cell Junctions and Cell Polarity Bind to the Full-Length E Protein in a PBM-Dependent Manner

We selected five PDZ-containing proteins involved in cellular junction and polarity from our previous high-throughput study on the specificity profile of the C-terminal SARS-CoV-2 E PBM sequence (**Figure 1C**) against our library of all human PDZ domains (PDZome) using the automated holdup assay (Vincentelli et al., 2015; Duhoo et al., 2019; Caillet-Saguy et al., 2021). These are ZO-1, LNX2, PARD3, MLLT4, and MPP5 previously reported to interact with SARS-CoV-2 E PBM through PBM–PDZ interactions (**Table 1**).

To investigate whether the five PDZ domains can interact with the full-length E protein and in a PBM-dependent manner, glutathione S-transferase (GST) pull-down assays were performed with the purified GST-tagged PDZ domains of ZO-1, LNX2, PARD3, MLLT4, and MPP5 and the lysates of HEK293 cells overexpressing constructs of GFP alone, GFP-tagged wild-type full-length E protein (GFP-E-WT), GFP-tagged full-length E protein with the PBM mutated with glycines (GFP-E-GGGG), and GFP-tagged cytoplasmic tail of protein E (last 12 residues; GFP-E last 12 aa) designated 1, 2, 3, and 4 in **Figure 2** (**Figure 2A**). The GFP-tagged construct expression was assessed by examining the fluorescence emitted by the GFP by microscopy and by Western blot on the cell lysates using anti-GFP antibody (**Figure 2B**). GST alone was used as a negative control. We confirmed by using Ponceau S staining that equal amounts of GST-tagged protein constructs were bound to the GST resin (**Figure 2C**).

Glutathione S-transferase alone and its fusion with PDZ ZO1, PDZ LNX2, PDZ MLLT4, PDZ MPP5, and PDZ PARD3 are used as baits and were immobilized on glutathione beads and tested for their ability to pull down GFP alone, GFP-E-WT, GFP-E-GGGG, and GFP-E last 12 aa by Western blot using the anti-GFP antibody. After washing, identical amounts of beads were analyzed for the presence of GFP-tagged proteins.

The GFP-E last 12 aa was detected in all interactions with GST-PDZ domains but not detected with GST alone confirming the interactions identified in our high-throughput holdup assay (Caillet-Saguy et al., 2021), validating the interactions within the context of lysates of HEK293 cells overexpressing SARS-CoV-2 protein E constructs (**Figure 2C**).

GFP-tagged wild-type full-length E protein was also detected in all interactions with GST-PDZ domains but with GST-PDZ PARD3 that showed no clear band (**Figure 2C**). In all cases, the bands have a weaker intensity than GFP-E last 12 aa in

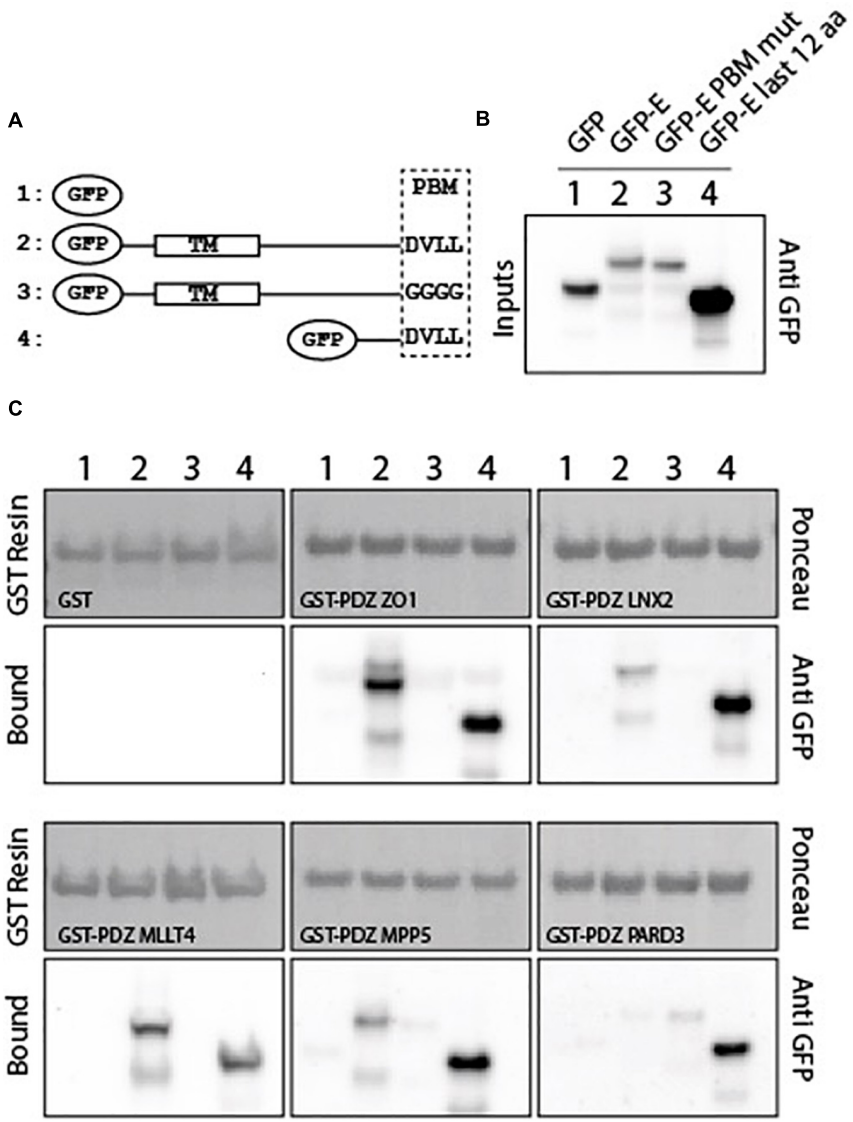


FIGURE 2 | PDZ domains interact with the SARS-CoV-2 full-length E protein in a PBM-dependent manner. **(A)** Schematic representation of the GFP-E constructs used in this study. Construct 1 corresponds to the GFP alone, construct 2 and 3 to the full-length WT and mutated E protein respectively, and construct 4 to the last 12 C-terminal residues of E protein. TM and PBM correspond to the transmembrane domain and to the PDZ-binding motif, respectively. **(B)** Input fraction of GFP-tagged viral E gene construct expressed in HEK293 cells and used for pull-down assay. Samples were analyzed by immunoblotting using anti-GFP antibody. **(C)** Gluthatione S-transferase (GST) pull-down binding results with immobilized GST or GST-PDZ domains (GST-PDZ ZO1, GST-PDZ LNX2, GST-PDZ MLLT4, GST-PDZ MPP5, and GST-PDZ PARD3) used as affinity resin and incubated with the HEK293 cell lysates. GST-tagged proteins were stained by Ponceau S (top panels). The bound fraction was analyzed by immunoblotting using an anti-GFP antibody (bottom panels). Note that blots represent discontinuous panels from the gels when black line delimitation is present. The images have been cropped to frame the relevant region.

agreement with a significant lower expression in cells, as shown in the inputs for GFP-E-WT compared with GFP-E last 12 aa (**Figure 2B**). We failed to obtain a clear detection of the band for GST-PDZ PARD3 (**Figure 2C**). Conversely, GFP-E-GGGG was not detected in all interactions with GST-PDZ domains except with GST-PDZ PARD3 that showed a weak band (**Figure 2C**).

Altogether, these results indicate a specific interaction between the SARS-CoV-2 WT protein E and the PDZ domains of ZO-1, LNX2, MLLT4, and MPP5 with variations in binding intensities except for PARD3. These interactions are PBM dependent in the context of the full-length E protein.

Structures of Complexes Between PDZ Domains of MLLT4, MPP5, and LNX2 and the C-Terminal of Protein E

We deciphered the molecular basis of recognition of the C-terminal sequence of the protein E of SARS-CoV-2 encompassing the PBM by the PDZ domains of the proteins

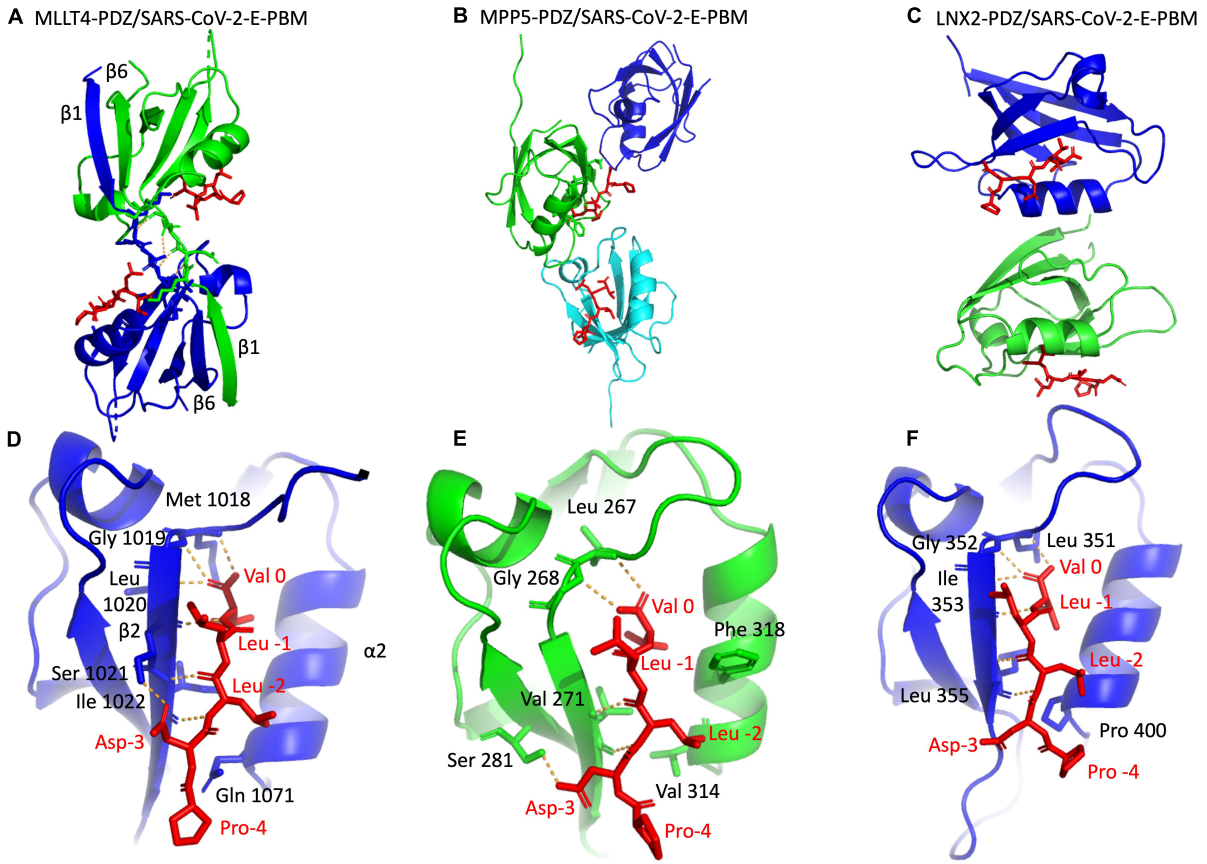


FIGURE 3 | X-ray structures of the PDZ domains of MLLT4, MPP5, and LNX2 bound to the SARS-CoV-2 protein E PBM. **(A–C)** The asymmetric unit of the PDZ domains of MLLT4, MPP5, and LNX2, respectively, bound to the SARS-CoV-2 protein E PBM shown as red sticks. **(A)** Selected interchain polar contacts related to the swapped dimer between the fragments Lys 1014–Gly 1017 of each chain are shown in orange, and the associated residues are shown as sticks. **(D–F)** Detailed views of the PDZ domains bound to SARS-CoV-2 protein E PBM. Important residues are labeled and shown as sticks. Intermolecular H-bonds and polar contacts are reported as orange dashed lines.

engaged in cellular junction and polarity. To this aim, we solved the crystal structures of the complex formed by the C-terminal peptide of protein E encompassing the PBM (**Figure 1C**) and the PDZ domains of MLLT4, MPP5, and LNX2 (**Table 1**) by molecular replacement (**Figure 3** and **Table 2**). We have described the set of intermolecular bonds to gain structural insights into the binding mode of the C-terminal sequence of the protein E of SARS-CoV-2 with PDZ domains.

Structure of the PDZ Domain of MLLT4 in Complex With the PBM of Protein E

The crystal structure of MLLT4-PDZ in complex with SARS-CoV-2-E-PBM peptide was solved at a resolution of 2.28 Å. The final refined model contains two PDZ domains per asymmetric unit that adopt a swapped dimer conformation with the PDZ folding comprising five β strands and two α helices. Two peptides are bound to each PDZ dimer (**Figure 3A**). An electron density is observed for the last five and last three residues of the peptide containing the PBM indicating a well-defined conformation of these last C-terminal residues. The last three residues of the SARS-CoV-2-E-PBM peptide bind to the α2/β2-groove in each PDZ unit (**Figure 3D**).

The swapped dimer comprises an intermolecular interaction between the two PDZ domains through the β1/β6 pair (**Figure 3A**), and several pairs of inter-domain H-bonds between the fragments Lys 1014–Gly 1017 of each chain (**Figure 3A**).

The E PBM binds to the PDZ in a conventional manner as an antiparallel extension of the β2 strand by inserting into a binding groove formed by the β2 strand, the α2 helix, and the "GLGF" motif. The last four residues of the SARS-CoV-2-E-PBM peptide were in contact with the PDZ domain, whereas only the last three residues were in contact and the upstream residues were distant from the PDZ domain surface in the previously reported complex between MLLT4-PDZ and the nectin-3 or the Bcr PBM peptides (Chen et al., 2007; Fujiwara et al., 2015).

The C-terminal carboxylate of Val at position 0 of the PBM forms three hydrogen bonds with the amide protons of Met 1018, Gly 1019, and Leu 1020 of the "GLGF" loop of MLLT4-PDZ (**Figure 3D**). A hydrogen bond is also formed between

TABLE 2 | X-ray data collection and refinement statistics.

	LNX2-PDZ/SARS-CoV-2-E-PBM	MLLT4-PDZ/SARS-CoV-2-E-PBM	MPP5-PDZ/SARS-CoV-2-E-PBM
Crystallization conditions	2.5 M sodium chloride 0.1 M sodium acetate pH 4.5 0.2 M lithium sulfate	1 M tri-sodium citrate, 0.1 M TRIS-HCl pH 7.0 0.2 M NaCl	0.5 M lithium chloride 1.6 M ammonium sulfate
Beamline	Proxima-1	Proxima-1	Proxima-2A
Resolution range	48.97–3.197 (3.312–3.197)	37.81–2.281 (2.362–2.281)	44.38–2.804 (2.904–2.804)
Space group	I 4 3 2	C 1 2 1	C 2 2 21
Unit cell	183.23 183.23 183.23 90 90 90	76.16 43.49 56.07 90 96.88 90	88.75 131.55 70.93 90 90 90
Total reflections	708,906 (68,100)	57,314 (4,867)	44,758 (4,462)
Unique reflections	9,009 (861)	8,373 (746)	8,040 (799)
Multiplicity	78.7 (77.6)	6.8 (6.5)	5.6 (5.6)
Completeness (%)	99.69 (97.95)	98.90 (89.23)	75.85 (78.61)
Mean I/sigma(I)	20.81 (1.39)	6.59 (1.16)	9.33 (1.31)
Wilson B-factor	123.76	44.58	67.85
R-merge	0.2679 (3.061)	0.2185 (1.454)	0.1238 (1.321)
R-meas	0.2697 (3.082)	0.2367 (1.577)	0.1365 (1.455)
R-pim	0.03029 (0.3451)	0.08997 (0.6014)	0.05584 (0.5944)
CC1/2	0.999 (0.459)	0.988 (0.506)	0.999 (0.689)
CC*	1 (0.793)	0.997 (0.82)	1 (0.903)
Reflections used in refinement	8,992 (860)	8,373 (746)	8,036 (794)
Reflections used for R-free	451 (44)	419 (37)	399 (39)
R-work	0.2281 (0.3597)	0.2208 (0.3409)	0.2633 (0.4147)
R-free	0.2395 (0.3810)	0.2523 (0.4239)	0.3049 (0.4974)
C(work)	0.908 (0.646)	0.916 (0.670)	0.936 (0.687)
CC(free)	0.905 (0.593)	0.911 (0.399)	0.891 (0.744)
Number of non-hydrogen atoms	1,483	1,410	2,204
Macromolecules	1,460	1,331	2,204
Ligands	0	5	0
Solvent	23	74	0
Protein residues	190	183	287
RMS(bonds)	0.012	0.012	0.013
RMS(angles)	1.53	1.56	1.66
Ramachandran favored (%)	96.15	97.66	96.39
Ramachandran allowed (%)	3.30	1.75	3.25
Ramachandran outliers (%)	0.55	0.58	0.36
Rotamer outliers (%)	11.69	7.64	12.24
Clashscore	10.15	6.51	5.61
Average B-factor	130.45	51.33	89.21
Macromolecules	130.97	51.32	89.21
Ligands		128.47	
Solvent	97.36	46.31	
PDB entry ID	7QCT	7QCR	7QCS

Statistics for the highest-resolution shell are shown in parentheses.

the proton amide of this Val 0 and the carbonyl of Leu 1020 of the β2 strand. In addition, the carbonyl and the amide group of Leu at position -2 interacts through a hydrogen bond with the proton amide and the carbonyl group of Ile 1022, respectively. The Ile -2 side chain establish hydrophobic contacts with the side chains of Gln 1071 and Ile 1022. A hydrogen bond is also formed between the carboxyl at the end of the side chain of Asp at position -3 of the PBM and the hydroxyl of Ser 1021 of the β2 strand (**Figure 3D**). Thus, the formed β sheet involved the last three residues of the SARS-CoV-2 E protein.

Structure of the PDZ Domain of MPP5 in Complex With the PBM of Protein E

The crystal structure of the MPP5-PDZ in complex with SARS-CoV-2-E-PBM was solved at a resolution of 2.80 Å. There are three MPP5-PDZs per asymmetric unit (**Figure 3B**). Each domain adopts a compact globular PDZ fold, and an electron density was observed for the last five residues of the peptide in two MPP5-PDZs present in the asymmetric unit.

As for MLLT4, the E PBM conventionally binds to the MPP5-PDZ. The C-terminal carboxylate of the Val at position 0 of the PBM forms two hydrogen bonds: one with the amide proton of

Leu 267 and one with Gly 268 of the "GLGF" loop of MPP5-PDZ (**Figure 3E**). In addition, the amide and carbonyl groups of Leu at position -2 form hydrogen bonds with the carbonyl and amide proton of valine 271, respectively. Interestingly, as for MLLT4, a hydrogen bond is also formed between the carboxyl at the end of the side chain of Asp at position -3 of the PBM and the hydroxyl group of Ser 281 of the β3 strand (**Figure 3E**). The key hydrogen bonds of Val 0 and Leu -2 are similar with the structure recently reported (Javorsky et al., 2021). However, the side chains of Asp at position -3 and Arg 272 form an ionic bond within this structure. The alternative hydrogen bond with Ser 281 identified in our structure was previously reported in the complex formed by MPP5-PDZ and the PBM peptide of Crumbs (Javorsky et al., 2021). Furthermore, the side chain of Phe 318 in α2 helix, which prevents access to the binding groove in the unbound form (Ivanova et al., 2015), is located out of the binding groove in our structure allowing SARS-CoV-2-E-PBM peptide binding (**Figure 3E**) as previously stated for MPP5-PDZ in complex with SARS-CoV-1 E and SARS-CoV-2 E PBMs, as well as for Crumbs peptides (Javorsky et al., 2021).

Structure of the PDZ Domain of LNX2 in Complex With the PBM of Protein E

The crystal structure of LNX2-PDZ2 in complex with SARS-CoV-2-E-PBM was solved at a resolution of 3.3 Å. Each of the two LNX2-PDZs present in the asymmetric unit are bound to the peptide (**Figure 3C**). The PDZ fold agrees with the unbound form of LNX2-PDZ (PDB 5e1y) with an overall RMSD of the backbone atoms of the two PDZs of 0.44 Å. Electron density was observed for the last five residues of the two peptides (**Figure 3F**). In both cases, the last three residues of the SARS-CoV-2-E-PBM peptide bind to the canonical α2 helix/β2 strand groove in each PDZ unit (**Figure 3C**).

As for MLLT4 and MPP5, the PBM of protein E conventionally binds to the PDZ of LNX2. The C-terminal carboxylate of Val 0 of the PBM forms three hydrogen bonds: two with the amide protons of Leu 351 and Gly 352 of the "GLGF" loop of LNX2-PDZ2 and one with the Ile 353 amide proton (**Figure 3F**). The proton amide of Val 0 also forms an H-bond with the carbonyl of Ile 353. In addition, the amide and carbonyl groups of Leu -2 form hydrogen bonds with the carbonyl and amide proton of Leu 355, respectively (**Figure 3F**). In one of the two complexes, an ionic bond is formed between the carboxylate of Val at position -3 of the PBM and the amine group of Arg 357 at the end of the β 2 strand.

In summary, we have shown that the SARS-CoV-2-E PBM peptide interacts with the PDZ domains of MLLT4, MPP5, and LNX2 with similar binding modes at positions 0 and -2. The SARS-CoV-2-E PBM is a class II PBM with a hydrophobic residue at position -2 that should contact a hydrophobic residue or the aliphatic part of a lysine at the N-terminal of the α2 helix of the PDZ domain (Songyang et al., 1997; Harris and Lim, 2001). Indeed, the N-terminal of the α2 helix is occupied by a valine (Val 314), a proline (Pro 400), and a glutamine (Gln 1071) in MPP5-PDZ, LNX2-PDZ, and MLLT4-PDZ, respectively. A glutamine at this position is different from a canonical class II PDZ but is still classified as a class II PDZ (**Figure 3** and **Table 1**). The side chain of aspartic acid at position -3 is involved in an H-bond with MLLT4-PDZ and MPP5-PDZ, whereas the proline at position -4 does not interact with any of the PDZ domains. Thus, we established the structural basis of SARS-CoV-2-E PBM binding to the PDZ domains of MLLT4, MPP5, and LNX2.

The Viral E Protein Sequestrates PDZ Domains to the Golgi Compartment

Then we explore the PDZ–PBM interactions between the full-length E protein and the selected PDZ domains in cells.

Severe acute respiratory syndrome coronavirus 1 or SARS-CoV-2 E proteins expressed from cDNA were reported to localize to the endoplasmic reticulum (ER), ERGIC, or the Golgi complex depending on the nature and localization of the tag (N or C) used in the study (Lopez et al., 2005; Cohen et al., 2011; Pearson et al., 2021). Here, we used the recently developed ALFA tag, that is, small (14 residues) and electroneutral (Götzke et al., 2019) to tag the SARS-CoV-2 E protein at the N-terminal. HeLa cells transiently transfected with SARS-CoV-2 ALFA-E encoding plasmids displayed strong Golgi expression of the viral protein as revealed by the costaining with Golgi marker GM130 (**Supplementary Figure 1A**).

We then investigated the ability of the viral E protein to recruit GFP-tagged PDZ domains to the Golgi compartment. When transfected alone, GFP-tagged PDZ domains from ZO1, MLLT4, MPP5/PALS1, LNX2, and PARD3 do not accumulate in the Golgi apparatus (**Supplementary Figure 1B**). Interestingly, all these GFP-tagged PDZ domains relocalize to the Golgi compartment when they are cotransfected with the ALFA-E construct as shown by the GM130 costaining for ZO1 (**Figure 4A**) and the other PDZ domains (**Supplementary Figure 1C**). ZO-1 PDZ2/E protein colocalization is strictly PBM dependent as no colocalization is observed between ZO1-PDZ2, and the ALFA E PBM mutant for which the PBM sequence DVLL was substituted by four glycines (**Figure 4B**). Likewise, this interaction was specific as the SCRIB-PDZ1 domain, an unrelated PDZ domain that was not identified in our screen, showed no tropism for the Golgi compartment when it was cotransfected with the ALFA E construct (**Figure 4C**). These results indicate that the viral E protein can bind to ZO1, MLLT4, MPP5, LNX2, and PARD3 PDZ domains *in cellulo* in a PBM-specific manner and recruit them to the Golgi apparatus.

The PDZ Domains Bind With Different Affinities With the Original C-Terminal SARS-CoV-2-E PBM and the C-Terminal SARS-CoV-2-E PBM P71L Mutant of the Variant of Concern Beta

Acquired mutations in SARS-CoV-2 protein E localized close to the PBM were reported, and the most common non-synonymous mutations were S68F and P71L (Hassan et al., 2020). We focused on the P71L mutation found in the beta variant. The VOC beta was first identified in South Africa in September 2020 and has been reported in more than 130 countries. The mutations on the beta variant make it more transmissible (Tegally et al., 2021), with greater antibody resistance compared with earlier variants of SARS-CoV-2. The

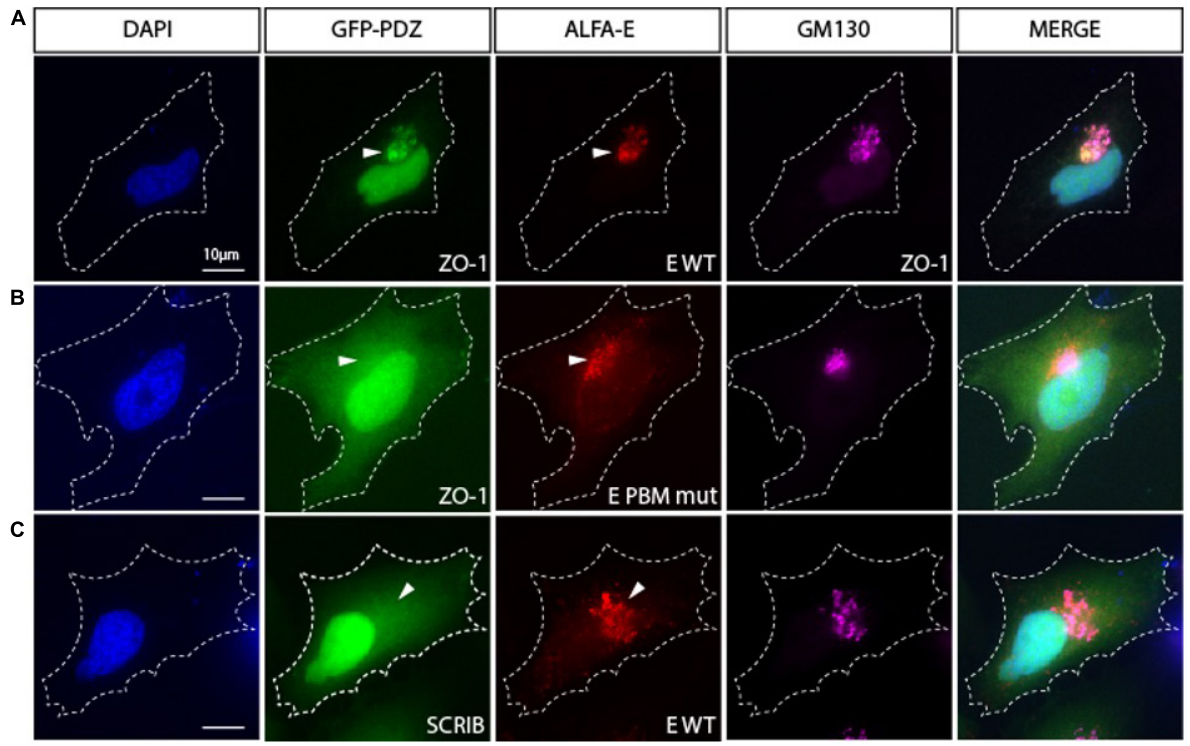


FIGURE 4 | The viral E protein recruits the PDZ domain of ZO1 to the Golgi apparatus. **(A)** Hela cell cotransfected with encoding constructs GFP-PDZ2 of ZO1 and ALFA-E (red) displays Golgi recruitment of both proteins as indicated by GM130 staining (Magenta). **(B)** Cotransfection of GFP-ZO1 with ALFA-E PBM mutant (red) does not lead to the recruitment of GFP-ZO1 to the Golgi apparatus. **(C)** Hela cell cotransfected with GFP-PDZ1 of SCRIB and ALFA-E (red) does not show recruitment of the proteins to the Golgi apparatus (Magenta). White arrowheads indicate Golgi apparatus position. Nuclei are stained with DAPI. Bars correspond to 10 μm.

TABLE 3 | Kd values between the E protein PBM of SARS-CoV-2 WT and of the P71L mutant for the selected PDZ domains.

	Kd (μ M)	
	SARS-CoV-2 WT	**SARS-CoV-2 P71L**
ZO1 PDZ2	15 (±11)	133 (±28)
MPP5 PDZ	30 (±24)	22 (±2)
LNX2 PDZ2	289 (±116)	47 (±20)
PARD3 PDZ3	341 (±57)	No interaction detected
MLLT4 PDZ	569 (±129)	No interaction detected

The data are representative of two independent experiments and error bars correspond to the standard deviation.

beta variant is estimated to be 50% more infectious than the original coronavirus strain (Tegally et al., 2021) and could be of a higher risk of hospitalization, admission to intensive care, and death (Funk et al., 2021).

The P71L mutation in the E protein of the beta variant is in the vicinity of the PBM at position -4. Upstream residues can also affect the PDZ-binding affinity and specificity in addition to the C-terminal positions 0 and -2 of the PBM, which are important for canonical PDZ domain binding (Luck et al., 2012).

We examined the binding affinities of the C-terminal SARS-CoV-2-E PBM peptide (12-mers peptide NLNSSRVPDLLV$_{COOH}$) from the original strain and the P71L PBM peptide (12-mers peptide sequence NLNSSRVLDLLV$_{COOH}$) against the PDZ domains of ZO1, MLLT4, MPP5, LNX2, and PARD3 using microscale thermophoresis (MST) (**Table 3** and **Supplementary Figure 2**). The averaged affinity constants (Kds) of the two PBM peptides for the different PDZ partners are summarized in **Table 3**. The best affinities for the original SARS-CoV-2 peptide E are obtained with the PDZ2 of ZO-1 and the PDZ of MPP5 with Kd values of 15 and 30 μM, respectively. The affinities of SARS-CoV-2 peptide E for PARD3-PDZ3 and MLLT4-PDZ are lower with Kd values of 341 and 569 μM, respectively. The beta variant P71L mutation does not significantly impact the affinity for the PDZ of MPP5 but considerably reduces the one for the ZO-1-PDZ2 with a Kd of 133 μM, and the PDZ domains of MLLT4 and PARD3 with affinities not detectable in the tested concentration range. Thus, either the P71L mutant does not interact with the PDZs of MLLT4 and PARD3, or the affinities are outside the tested concentration range (Kd > 1 mM).

Conversely, the affinity of LNX2-PDZ2 is improved from 289 to 47 μM with the mutant P71L. Thus, this mutation leads to a significant increase in the affinity for LNX2-PDZ2 with a Kd almost six times lower compared with the one of SARS-CoV-2 WT.

To resume, significant interactions are detected between the original C-terminal SARS-CoV-2-E PBM peptide and all

the PDZ domains we tested by MST, in agreement with our previous results using the high-throughput screening holdup assay (Caillet-Saguy et al., 2021). Interestingly, we observed different binding affinities for four over five PDZ domains with the P71L mutant. This strongly suggests a differential specificity profile against the PDZome between the original C-terminal SARS-CoV-2-E PBM and the C-terminal SARS-CoV-2-E PBM P71L mutant.

DISCUSSION

Deletion of protein E strongly reduces the replication of SARS-CoV-1 in cells and suppresses virus-induced mortality in mice (DeDiego et al., 2007). The protein E of SARS-CoV-1 contains a C-terminal type II PBM involved in viral pathogenesis (Jimenez-Guardeño et al., 2014) that binds to PDZ domains of cellular proteins. Interestingly, mice infected with an E PBM mutant virus showed 100% survival rate, mimicking the phenotype seen with the deletion of the full-length E protein (Jimenez-Guardeño et al., 2014). The protein E from SARS-CoV-2 presents also a PBM like that of SARS-CoV-1, and we recently identified the PDZ proteins targeted by the PBM of SARS-CoV-1, SARS-CoV-2, and MERS by screening *in vitro* the full human PDZ library (Caillet-Saguy et al., 2021). From this previous study and others (Teoh et al., 2010; Toto et al., 2020), we selected five human PDZ-containing proteins expressed in SARS-CoV-2 target cells showing significant interactions with the SARS-CoV-2 protein E PBM and involved in cellular junctions and cell polarity: ZO-1/TJP1, PARD3, MLLT4, LNX2, and MPP5/PALS1. Indeed, targeting cellular junctions and polarity machineries is a shared strategy by viruses to facilitate the infectious cycle improving either viral entry, replication, dissemination, or egress. Notably, ZO-1 and PALS1 are targeted by other viruses, such as influenza and SARS-CoV, to disrupt and open TJs to efficiently exit the airway epithelia to spread and disseminate (Torres-Flores and Arias, 2015). SARS-CoV-2 was shown to transiently impair bronchial epithelium altering the distribution of ZO-1 (Robinot et al., 2021), and PALS1 was shown to translocate from TJ to ERGIC when it is targeted by protein E through its PBM resulting in the dissociation of TJ in the epithelia of various organs (Chai et al., 2021). In this study, we showed that the five PDZ domains TJP1, PARD3, MLLT4, LNX2, and MPP5/PALS1 interact in a PBM-dependent manner *in vitro*, and these five PDZ domains colocalize with the full-length E protein *in cellulo* sequestrating the PDZ domains into the Golgi compartment. This is consistent with the exposition of the C-terminus of the protein E to the cytoplasmic side as previously reported (Duart et al., 2020), allowing interactions with viral and host proteins. The mutation P71L of protein E most likely affects the interaction with PDZ domains in cells and the targeting of the Golgi, knowing the affinity changes measured by the MST experiments, in particular, the PDZ of MLLT4 and PARD3 should not be targeted anymore since their affinities are greater than 1 mM.

Interestingly, our pull-down experiments showed that the binding partners identified in the holdup screening can interact not only with the last 12 amino acids, as previously shown in the holdup screening, but also with the full-length E protein. We noted a weaker detection of the interaction with the full-length WT protein E compared with the C-terminal peptide because the protein is less expressed, formed oligomers, and the C-terminal peptide is less likely accessible in the context of the entire protein inserted in the membrane (Park et al., 2021). Moreover, we observed by microscopy that the overexpressed protein E induced a cellular mortality (data not shown). It was previously proposed that the cation channel formed by the overexpressed protein E could disrupt the host cell membrane (Cao et al., 2021). Less cells are recovered after transfection with WT protein E than with other constructs. To address this issue, we transfected more cells. Interestingly, no cellular toxicity was observed with the GFP-E-GGGG construct. This is in agreement with a previous work on SARS-CoV-1 that showed that the mutant E-GGGG is no longer inducing lethality in mice (Jimenez-Guardeño et al., 2014), illustrating the impact of the E PBM on the homeostasis of infected cells.

Altogether, our results indicate significant interactions between the SARS-CoV-2 WT protein E and the PDZ domains of ZO-1, LNX2, MLLT4, and MPP5. These interactions are dependent on the presence of PBM in the full-length protein E since the mutation of the PBM led to a loss of interaction. PARD3 binds the SARS-CoV-2 E PBM *in vitro* and colocalizes with the full-length protein E, and not with the mutated PBM E protein that also strongly suggest a specific interaction. We solved the X-ray crystal structures of the complexes between the SARS-CoV-2 E C-terminal encompassing the PBM, and the human LNX2, MLLT4, and MPP5/PALS1 PDZs highlighting the binding modes for three of the potential cellular targets of protein E. We found a swapped dimer in the structure of MLLT4-PDZ with the β1 and β6 strands swapped (PDB 7QCR). Previous NMR structures of MLLT4-PDZ complexed or not with other compounds (peptides or small molecules) only reported monomers (PDB 1XZ9, 2EXG, 1T2M, and 2AIN). One X-ray structure reported a dimer due to a fusion of a PBM at the C-terminus of MLLT4-PDZ that created a new dimer interface forming an antiparallel β sheet between β2 strands (PDB 3AXA). To investigate the presence of a dimer in solution, we performed analytical ultracentrifugation (AUC) experiments on MLLT4-PDZ (**Table 4** and **Supplementary Figure 3**). The MLLT4-PDZ is found mainly monomeric in solution with a sedimentation coefficient of 1.4 S, suggesting that the swapped dimer is probably an artifact of crystallization. Interestingly, MLLT4-PDZ is able to recognize type II PBM and also type I PBM due to an unexpected glutamine in the α2 helix as previously reported (Zhou et al., 2005).

Remarkably, ZO-1-PDZ2 forms a very stable swapped dimer with an extended antiparallel inter-domain β sheet. The β1 and β2 strands of one domain is swapped with those from the second domain allowing the formation of the binding groove (Fanning et al., 2007; Chen et al., 2008). We also performed AUC experiments on ZO-1-PDZ2 to verify its oligomeric state in solution (**Table 4** and **Supplementary Figure 3**). We found that ZO-1-PDZ2 adopts two oligomeric

TABLE 4 | Hydrodynamic parameters of ZO-1 PDZ2 and MLLT4 PDZ derived from the analysis of analytical ultracentrifugation.

	ZO-1 1 mg/mL			MLLT4-PBM 1 mg/mL		
Sedimentation coefficient (S)	2.1	2.8	3.7	0.5	1.4	2.3
Frictional ratio f/f0 (average)		1.4			1.2	
Molecular weight estimated by analytical ultracentrifugation (AUC) (kDa)	26	40	60	2.8	12	26
Peak proportion%	49	49	2	25	65	10

states in solution: one dimeric form and one higher oligomeric state possibly compatible with a tetramer in solution that might be consistent with a swapped dimer in solution. Conclusively, the X-ray structures have provided the structural basis for SARS-CoV-2-E PBM binding to the PDZ domains of MLLT4, MPP5, and LNX2 and these results offer a mechanistic beginning for SARS-CoV-2 perturbation of PDZ-containing proteins.

Finally, we showed different affinities for the cellular PDZ domains of ZO-1, LNX2, PARD3, MLLT4, and MPP5, with the original SARS-CoV-2 E C-terminal peptide (-NLNSSRVPDLLV$_{COOH}$) and the beta variant E C-terminal peptide (-NLNSSRVLDLLV$_{COOH}$) containing the P71L mutation close to the PBM (**Figure 1B**). This position -4 is not well defined within an electronic density in the crystal structures, and when a density is observed, it is the last upstream residue modeled for the C-terminal of the protein E. In the three complexes we studied, the residue in position -4 is not involved in the interaction network with the PDZ domains. Surprisingly, we determined that P71L mutation markedly altered affinities with host PDZ domains. While it noticeably reduces the affinities for the PDZ domains of ZO-1, MLLT4, and PARD3, it enhances its affinity for LNX2 compared with SARS-CoV-2 WT. Influences of PBM upstream residues on the affinity for PDZ domains have already been reported previously (Lee and Zheng, 2010; Terrien et al., 2012). These findings show that the acquired P71L mutation might have important effects on the human PDZome targeted by the variant during the infection.

Less prevalent than the alpha and delta variants, the beta variant accounted for approximately 10% of virus samples in France in June 2021.[2] Mutations in the beta variant make it more transmissible, with greater antibody resistance, higher risk of hospitalization, ICU admission, and death compared with earlier SARS-CoV-2 E variants (Veneti et al., 2021; Wang et al., 2021). In addition to the P71L mutation, the beta variant has other mutations, including three on its spike protein, which may help the virus to escape antibodies and to bind more tightly to human cells (Han et al., 2021). The importance of the P71L mutation in the greater pathogenicity of the beta variant remains to be demonstrated. Nevertheless, because viroporins like E protein are linked to inflammasome activation during viral infection (Schoeman and Fielding, 2019), mutations that can alter the specificity of E for its protein partners could potentially impact the viral E proinflammatory behavior and, thus, be associated with a change in viral pathogenicity.

[2]https://www.santepubliquefrance.fr/

MATERIALS AND METHODS

Construct Cloning

DNA sequences encoding the PDZ domains of ZO-1-2 (PDZ2; residues 184–264), LNX2-2 (PDZ2; residues 334–426), PARDD3-3 (PDZ3; residues 582–685), MLLT4 (residues 961–1056), and MPP5 (residues 238–336) come from the PDZome library with PDZ domains cloned into a Gateway system (Vincentelli et al., 2015; Duhoo et al., 2019). The PDZ domains were cloned into pETG-41A plasmid vectors as an N-terminal fusion to a histidine tag and a maltose-binding protein (His-MBP-tag). Optimization of expression and purification conditions led us to test other tags for most constructs using the Gateway vector recombination system. Following the protocol of the manufacturer of the Gateway^{TM} LR Clonase^{TM} II Enzyme mix (Invitrogen), PDZ domains of ZO-1, LNX2, and PARD3 were subcloned into pDEST^{TM}17 vector and MPP5-PDZ into pDEST^{TM}15 vector, allowing the production of recombinant proteins as a fusion to an N-terminal histidine and GST tag, respectively. A TEV cleavage site was introduced between the N-terminal tags and the PDZ sequences.

The pCMV ALFA E vector was constructed as follows: DNA sequence encoding the ALFA tag (MSRLEEELRRRLTE) followed by a linker (GGGGS) fused to the sequence corresponding to the alpha variant of SARS-CoV-2 E cDNA (GenBank: BCM16077.1) synthetized by Eurofins. The ALFA-linker-E sequence was subsequently cloned into a pCMV backbone vector using *Cla*I and *Xho*I. The E PBM C-terminal mutation (DVLL > GGGG) was performed using the Q5 site-directed mutagenesis kit (NEB). The sequences corresponding to the PDZ domains were cloned into the pCMV GFP vector using *Eco*RI and *Xho*I sites.

Protein Expression and Purification

The vectors were used to transform *E. coli* BL21 Star (DE3) (Invitrogen) strain. Bacteria were grown in LB medium supplemented with ampicillin (100 mg/L) at 37°C. Protein expression was induced at OD600 nm 0.8–1 with 0.2 mM IPTG at 18°C overnight. Bacteria were then harvested, and PDZ domains of MLLT4, LNX2, PARD3, and MPP5 were resuspended in lysis buffer [Tris 50 mM (pH 7.5), NaCl 250 mM, β-mercaptoethanol 2 mM, protease inhibitors, one tablet per 50 mL of buffer (EDTA-free, Roche Diagnostics), and benzonase (E1014-25KU > 250 U/L of culture)]. Then the cells were broken under pressure using a cell disruptor CellD (ROQUEMAURE, France) (1.3 kBar at 4°C). The debris were pelleted by centrifugation (30,000 × *g*, 1 h, 4°C). The GST tag fused proteins were purified by affinity

chromatography column with GSTrap HP (GE Healthcare) and the His or His-MBP tag fused proteins by nickel-chelated HiTrap Chelating HP (GE Healthcare), followed by a TEV protease cleavage at 16°C overnight. A final step of size-exclusion chromatography was achieved using a Sephacryl S-100 HR 16/600 (GE Healthcare) with a buffer of Tris 50 mM (pH 7.5), NaCl 150 mM, TCEP 0.5 mM, protease inhibitors, and one tablet per 100 mL of buffer (Complete, EDTA 2%, Roche Diagnostics). Regarding the purification of ZO-1-PDZ2, a denaturation protocol was applied. The bacterial pellet was resuspended in a denaturing solution containing guanidine hydrochloride (6 M) and imidazole 20 mM, and cells were lysed by sonication. The debris were pelleted by centrifugation (30,000 × *g*, 1 h, 4°C). The His-tagged ZO-1-PDZ2 was purified using a HiTrap Chelating HP (GE Healthcare), where a renaturation gradient by exchange of denaturation buffer to a gel filtration buffer [Tris 50 mM (pH 7.5), NaCl 150 mM, TCEP 0.5 mM, protease inhibitors, one tablet per 100 mL of buffer (Complete, EDTA 2%, Roche Diagnostics)] was performed. Protein elution was then carried out by a gradient of imidazole from 0 to 500 mM in gel filtration buffer. Finally, a size-exclusion chromatography step was achieved using a Sephacryl® S-200 HR (GE Healthcare).

Peptide Synthesis

The acetylated peptides containing the C-terminal PBM sequence of E protein (12 residues long) of SARS-CoV-2 ($_{ac}$NLNSSRVPDLLV$_{COOH}$) or of its South African mutant P71L ($_{ac}$NLNSSRVLDLLV$_{COOH}$) were synthesized in a solid phase using the Fmoc strategy (Proteogenix, Schiltigheim, France). Peptides were resuspended in water to prepare stock solutions.

Crystallization, Data Collection, and Structure Determination

The PDZ–PBM complexes for crystallization were generated by mixing LNX2, MLLT4, or MPP5 PDZ domains with SARS-CoV-2 E protein PBM peptide at a ratio of 1:2, ensuring that at least 92% of complexes were formed. Initial screening of crystallization conditions was carried out by the vapor diffusion method using a MosquitoTM nanoliter-dispensing system (TTP Labtech, Melbourn, United Kingdom) following the established protocols (Weber et al., 2019). Sitting drops were set up using 400 nl of a 1:1 mixture of each sample protein and crystallization solutions (672 different commercially available conditions) equilibrated against a 150-µl reservoir in multiwell plates (Greiner Bio-one, GmbH, Frichenhausen, Germany). The crystallization plates were stored at 4°C in a RockImager1000® (Formulatrix, Bedford, MA, United States) automated imaging system to monitor crystal growth. The best crystals were obtained in crystallization conditions described in **Table 2**. Crystals were then flash cooled in liquid nitrogen using the condition of crystallization supplemented with 30% (V/V) of glycerol as cryoprotectant.

X-ray diffraction data were collected on the beamlines Proxima-1 and Proxima-2A at Synchrotron SOLEIL (St. Aubin, France). The data were processed by XDS (Kabsch, 2010), and the structures were solved by molecular replacement with PHASER (McCoy, 2007) using the search atomic models PDB id 3AXA, 4UU5, and 5E1Y for MLLT4-PDZ, MPP5-PDZ, and LNX2-PDZ2, respectively. The positions of the bound peptides were determined from an Fo–Fc difference electron density maps. Models were rebuilt using COOT (Emsley et al., 2010), and refinement was done with phenix.refine of the PHENIX suite (Adams et al., 2010). The crystal parameters, data collection statistics, and final refinement statistics are shown in **Table 2**. The multiple Ramachandran outliers correspond to residues from either the N- or C-terminal tails of the PDZ domains. More specifically, the outlier residues are Gly 422 (LNX2), Asp 3 (MPP5), and Gly 89 (MLLT4). With respect to the MPP5/E complex, the R/R-free values reported are relatively high due to the presence of a highly mobile/dynamic PDZ domain (chain E). The density for this molecule is very poor, most likely as a consequence of the few contacts established with the other molecules present in the asymmetric unit. The structure factors and coordinates have been deposited in the Protein Data Bank under accession codes 7QCR, 7QCS, and 7QCT for MLLT4-PDZ, MPP5-PDZ, and LNX2-PDZ, respectively. All structural figures were generated with the PyMOL Molecular Graphics System, Version (Schrödinger).

Glutathione S-Transferase Pulldown Assay

N-terminal GST fusion constructs containing the PDZ domains of ZO-1, MPP5, LNX2, PARD3, and MLLT4 were expressed and purified without the cleavage step by TEV protease. Purified GST constructs were individually incubated with glutathione–agarose beads for 1 h at 4°C with mild shaking. The beads were pelleted by centrifugation and washed four times with binding buffer (Tris 50 mM pH 7.5, NaCl 150 mM, TCEP 0.5 mM, antiprotease with EDTA 2% 1 tablet/100 mL). HEK 293 cells were transiently transfected with the GFP-tagged constructs using the phosphate calcium method. Cell lysates were prepared by scraping cells in lysis buffer Tris 50 mM pH7.5, triton 2%, NP40 1%, NaCl 200 mM with complete protease inhibitor tablet (Roche, Indianapolis, IN, United States), and centrifuged for 10 min at 13,000 rpm 4°C to pellet cell debris. Soluble detergent extracts were incubated with glutathione resins for 2 h at 4°C prior to washing three times with PBS supplemented with NaCl 200 mM and 0.1% Triton and processed for Western blot analysis with GFP antibody (Novus NB600-313).

Immunofluorescence

HeLa cells were transfected using the Genejuice transfection reagent (Novagen) according to the protocol of the manufacturer. At 24 h after transfection, cells were fixed with PBS PFA 4% for 10 min and permeabilized in PBS Triton 0.1% for 5 min before being processed for immunofluorescence using DAPI, anti-ALFA tag (NanoTag Biotechnologies; cat#N1502-SC3), and anti-GM130 (BD Transduction Lab; cat#610823). Images were acquired on a Leica DM6B microscope with a 63X objective.

Microscale Thermophoresis

The binding affinities between the C-terminal SARS-CoV-2-E PBM peptide (sequence $NLNSSRVPDLLV_{COOH}$) or the P71L PBM peptide (sequence $NLNSSRVLDLLV_{COOH}$) and the PDZ domains of ZO1, MLLT4, MPP5, LNX2, and PARD3 were measured using a Monolith NT.115 instrument (NanoTemper, Gmbh). The purified PDZ domains were covalently labeled using a fluorescent dye reactive on amine following the protocol of the manufacturer (Protein Labelling Kit RED-NHS, Nanotemper). A serial dilution of the WT and the P71L mutant of the E protein PBM peptide was prepared in the buffer containing PBS tween 20 0.05%. A volume of 10 μl of peptide was serially diluted 1:1 in the buffer and mixed with an equal volume of labeled PDZ and loaded on capillaries. The concentration of the various labeled PDZ domains was kept constant, and the peptide concentration varied.

We used 80% LED and 20% microscale thermophoresis (MST) power at room temperature and an MST time of 30 s for all MST measurements. The data analysis and curve fitting with a Kd model were performed with NanoTemper programs MO.Control 2 and MO.Affinity 1 Analysis. The change in the thermophoretic mobility upon titration is measured as a delta of normalized fluorescence. All experiments were made in duplicate.

Analytical Ultracentrifugation

ZO-1-PDZ2 and MLLT4-PDZ were prepared at 1 mg/mL. MLLT4-PDZ was prepared with an excess of protein E PBM peptide. Samples were prepared in Tris 50 mM pH 7.5, NaCl 150 mM, and TCEP 0.5 mM. A sample of 400 μl was loaded into 1.2-cm double-sector cells between two sapphire windows. Cells were incubated for 2 h at 20°C in an AN60-Ti rotor in the Optima-AUC analytical ultracentrifuge (Beckman Coulter) before data acquisition for 15 h at 42,000 rpm. Sedimentation profiles were monitored over time by absorbance measurement at 280 nm and by interferometry. The data were analyzed using the continuous size distribution model c(s) of the Sedfit 16.36 software. All distributions were calculated with a floating frictional ratio f/f0 and a maximum entropy regularization procedure with a confidence level of 0.68. The buffer viscosity (η = 0.01031 Poise), the density (ρ = 1.0059), and the partial specific volume of 0.746 for MLLT4 and 0.736 for Z0-1 were estimated at 20°C from the amino acid sequences at 20°C using the software SEDTERP 3.0.3.[3]

[3] http://www.jphilo.mailway.com/sednterp.htm

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: http://www.wwpdb.org/, 7QCT; http://www.wwpdb.org/, 7QCR; http://www.wwpdb.org/, 7QCS.

AUTHOR CONTRIBUTIONS

CC-S designed and directed the project. CC-S and BB conceived and designed the experiments. YZ, FA, AH, SB, BN, BB, and CC-S collected the data. YZ, FA, AM, SB, BB, and CC-S analyzed the data. NW and SE-M provided the equipment and funding. CC-S and BB wrote the manuscript with input from the other authors. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the URGENCE COVID-19 fundraising campaign of Institut Pasteur, the ANR Recherche Action Covid19–FRM PDZCov2 Program, and the DON MICHELIN COVID PFR-5 Cov-2-Cvnet. BB was supported by the Institut National de la Santé et de la Recherche Médicale (INSERM).

ACKNOWLEDGMENTS

We thank the staff of the Crystallography Core Facility at the Institut Pasteur for carrying out robot-driven crystallization screenings, and the staff at the beamlines of Proxima 1 and Proxima 2 at the French National Synchrotron Facility (SOLEIL, St Aubin, France) for help with data collection. We thank the DIM 1HEALTH region Ile-de-France for funding the Centrifection project that has allowed the Optima ultracentrifuge investment.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.829094/full#supplementary-material

REFERENCES

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66, 213–221. doi: 10.1107/S0907444909052925

Cabrera-Garcia, D., Bekdash, R., Abbott, G. W., Yazawa, M., and Harrison, N. L. (2021). The envelope protein of SARS-CoV-2 increases intra-Golgi pH and forms a cation channel that is regulated by pH. *J. Physiol.* 599, 2851–2868. doi: 10.1113/JP281037

Caillet-Saguy, C., Durbesson, F., Rezelj, V. V., Gogl, G., Tran, Q. D., Twizere, J.-C., et al. (2021). Host PDZ-containing proteins targeted by SARS-CoV-2. *FEBS J.* 2021:15881. doi: 10.1111/febs.15881

Cao, Y., Yang, R., Lee, I., Zhang, W., Sun, J., Wang, W., et al. (2021). Characterization of the SARS-CoV -2 E Protein: Sequence, Structure, Viroporin, and Inhibitors. *Prot. Sci.* 30, 1114–1130. doi: 10.1002/pro.4075

Castaño-Rodriguez, C., Honrubia, J. M., Gutiérrez-Álvarez, J., DeDiego, M. L., Nieto-Torres, J. L., Jimenez-Guardeño, J. M., et al. (2018). Role of Severe Acute Respiratory Syndrome Coronavirus Viroporins E, 3a, and 8a in Replication and Pathogenesis:*mBio* 9, 17. doi: 10.1128/mBio.02325-17

Chai, J., Cai, Y., Pang, C., Wang, L., McSweeney, S., Shanklin, J., et al. (2021). Structural basis for SARS-CoV-2 envelope protein recognition of human cell junction protein PALS1. *Nat. Commun.* 12:3433. doi: 10.1038/s41467-021-23533-x

Chen, J., Pan, L., Wei, Z., Zhao, Y., and Zhang, M. (2008). Domain-swapped dimerization of ZO-1 PDZ2 generates specific and regulatory connexin43-binding sites. *EMBO J.* 27, 2113–2123. doi: 10.1038/emboj.2008.138

Chen, Q., Niu, X., Xu, Y., Wu, J., and Shi, Y. (2007). Solution structure and backbone dynamics of the AF-6 PDZ domain/Bcr peptide complex. *Protein Sci.* 16, 1053–1062. doi: 10.1110/ps.062440607

Chen, X., An, Y., Gao, Y., Guo, L., Rui, L., Xie, H., et al. (2017). Rare Deleterious PARD3 Variants in the aPKC-Binding Region are Implicated in the Pathogenesis of Human Cranial Neural Tube Defects *Via* Disrupting Apical Tight Junction Formation. *Hum. Mutat.* 38, 378–389. doi: 10.1002/humu.23153

Cohen, J. R., Lin, L. D., and Machamer, C. E. (2011). Identification of a Golgi complex-targeting signal in the cytoplasmic tail of the severe acute respiratory syndrome coronavirus envelope protein. *J Virol* 85, 5794–5803. doi: 10.1128/JVI.00060-11

DeDiego, M. L., Alvarez, E., Almazán, F., Rejas, M. T., Lamirande, E., Roberts, A., et al. (2007). A severe acute respiratory syndrome coronavirus that lacks the E gene is attenuated in vitro and in vivo. *J. Virol.* 81, 1701–1713. doi: 10.1128/JVI.01467-06

Duart, G., García-Murria, M. J., Grau, B., Acosta-Cáceres, J. M., Martínez-Gil, L., and Mingarro, I. (2020). SARS-CoV-2 envelope protein topology in eukaryotic membranes. *Open Biol.* 10:200209. doi: 10.1098/rsob.200209

Duhoo, Y., Girault, V., Turchetto, J., Ramond, L., Durbesson, F., Fourquet, P., et al. (2019). High-Throughput Production of a New Library of Human Single and Tandem PDZ Domains Allows Quantitative PDZ-Peptide Interaction Screening Through High-Throughput Holdup Assay. *Methods Mol. Biol.* 2025, 439–476. doi: 10.1007/978-1-4939-9624-7_21

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501. doi: 10.1107/S0907444910007493

Fanning, A. S., Lye, M. F., Anderson, J. M., and Lavie, A. (2007). Domain swapping within PDZ2 is responsible for dimerization of ZO proteins. *J. Biol. Chem.* 282, 37710–37716. doi: 10.1074/jbc.M707255200

Fujiwara, Y., Goda, N., Tamashiro, T., Narita, H., Satomura, K., Tenno, T., et al. (2015). Crystal structure of afadin PDZ domain-nectin-3 complex shows the structural plasticity of the ligand-binding site. *Protein Sci.* 24, 376–385. doi: 10.1002/pro.2628

Funk, T., Pharris, A., Spiteri, G., Bundle, N., Melidou, A., Carr, M., et al. (2021). Characteristics of SARS-CoV-2 variants of concern B.1.1.7, B.1.351 or P.1: data from seven EU/EEA countries, weeks 38/2020 to 10/2021. *Euro Surveill.* 26:2100348. doi: 10.2807/1560-7917.ES.2021.26.16.2100348

Götzke, H., Kilisch, M., Martínez-Carranza, M., Sograte-Idrissi, S., Rajavel, A., Schlichthaerle, T., et al. (2019). The ALFA-tag is a highly versatile tool for nanobody-based bioscience applications. *Nat. Commun.* 10:4403. doi: 10.1038/s41467-019-12301-7

Gutiérrez-González, L. H., and Santos-Mendoza, T. (2019). Viral targeting of PDZ polarity proteins in the immune system as a potential evasion mechanism. *FASEB J.* 33, 10607–10617. doi: 10.1096/fj.201900518R

Han, P., Su, C., Zhang, Y., Bai, C., Zheng, A., Qiao, C., et al. (2021). Molecular insights into receptor binding of recent emerging SARS-CoV-2 variants. *Nat. Commun.* 12:6103. doi: 10.1038/s41467-021-26401-w

Harris, B. Z., and Lim, W. A. (2001). Mechanism and role of PDZ domains in signaling complex assembly. *J. Cell Sci.* 114, 3219–3231.

Hassan, S. S., Choudhury, P. P., and Roy, B. (2020). SARS-CoV2 envelope protein: non-synonymous mutations and its consequences. *Genomics* 112, 3890–3892. doi: 10.1016/j.ygeno.2020.07.001

Ikeda, W., Nakanishi, H., Miyoshi, J., Mandai, K., Ishizaki, H., Tanaka, M., et al. (1999). Afadin: a key molecule essential for structural organization of cell-cell junctions of polarized epithelia during embryogenesis. *J. Cell Biol.* 146, 1117–1132. doi: 10.1083/jcb.146.5.1117

Ivanova, M. E., Fletcher, G. C., O'Reilly, N., Purkiss, A. G., Thompson, B. J., and McDonald, N. Q. (2015). Structures of the human Pals1 PDZ domain with and without ligand suggest gated access of Crb to the PDZ peptide-binding groove. *Acta Crystallogr. D Biol. Crystallogr.* 71, 555–564. doi: 10.1107/S139900471402776X

James, C. D., and Roberts, S. (2016). Viral Interactions with PDZ Domain-Containing Proteins—An Oncogenic Trait? *Pathogens* 5:5010008. doi: 10.3390/pathogens5010008

Javier, R. T., and Rice, A. P. (2011). Emerging theme: cellular PDZ proteins as common targets of pathogenic viruses. *J. Virol.* 85, 11544–11556. doi: 10.1128/JVI.05410-11

Javorsky, A., Humbert, P. O., and Kvansakul, M. (2021). Structural basis of coronavirus E protein interactions with human PALS1 PDZ domain. *Commun. Biol* 4:724. doi: 10.1038/s42003-021-02250-7

Jimenez-Guardeño, J. M., Nieto-Torres, J. L., DeDiego, M. L., Regla-Nava, J. A., Fernandez-Delgado, R., Castaño-Rodriguez, C., et al. (2014). The PDZ-binding motif of severe acute respiratory syndrome coronavirus envelope protein is a determinant of viral pathogenesis. *PLoS Pathog.* 10:e1004320. doi: 10.1371/journal.ppat.1004320

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi: 10.1038/s41586-021-03819-2

Kabsch, W. (2010). XDS. *Acta Crystallogr. D Biol. Crystallogr.* 66, 125–132. doi: 10.1107/S0907444909047337

Lee, H.-J., and Zheng, J. J. (2010). PDZ domains and their binding partners: structure, specificity, and modification. *Cell Comm. Sig.* 8:8.

Li, Y., Surya, W., Claudine, S., and Torres, J. (2014). Structure of a conserved Golgi complex-targeting signal in coronavirus envelope proteins. *J. Biol. Chem.* 289, 12535–12549. doi: 10.1074/jbc.M114.560094

Lopez, L. A., Jones, A., Arndt, W. D., Hogue, B. G. (2005). *Subcellular localization of sars-cov structural proteins*. New York, NY: Springer.

Luck, K., Charbonnier, S., and Travé, G. (2012). The emerging contribution of sequence context to the specificity of protein interactions mediated by PDZ domains. *FEBS Lett.* 586, 2648–2661. doi: 10.1016/j.febslet.2012.03.056

Mandala, V. S., McKay, M. J., Shcherbakov, A. A., Dregni, A. J., Kolocouris, A., and Hong, M. (2020). Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers. *Nat. Struct. Mol. Biol.* 27, 1202–1208. doi: 10.1038/s41594-020-00536-8

McCoy, A. J. (2007). Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D Biol. Crystallogr.* 63, 32–41. doi: 10.1107/S0907444906045975

Nieto-Torres, J. L., Dediego, M. L., Alvarez, E., Jiménez-Guardeño, J. M., Regla-Nava, J. A., Llorente, M., et al. (2011). Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein. *Virology* 415, 69–82. doi: 10.1016/j.virol.2011.03.029

Park, S. H., Siddiqi, H., Castro, D. V., De Angelis, A. A., Oom, A. L., Stoneham, C. A., et al. (2021). Interactions of SARS-CoV-2 envelope protein with amilorides correlate with antiviral activity. *PLoS Pathog.* 17:e1009519. doi: 10.1371/journal.ppat.1009519

Pearson, G. J., Broncel, M., Snijders, A. P., and Carlton, J. G. (2021). Exploitation of the Secretory Pathway by SARS-CoV-2 Envelope. *bioRxiv* doi: 10.1101/2021.06.30.450614

Robinot, R., Hubert, M., de Melo, G. D., Lazarini, F., Bruel, T., Smith, N., et al. (2021). SARS-CoV-2 infection induces the dedifferentiation of multiciliated cells and impairs mucociliary clearance. *Nat. Commun.* 12:4354. doi: 10.1038/s41467-021-24521-x

Roh, M. H., Makarova, O., Liu, C.-J., Shin, K., Lee, S., Laurinec, S., et al. (2002). The Maguk protein, Pals1, functions as an adapter, linking mammalian homologues of Crumbs and Discs Lost. *J. Cell Biol.* 157, 161–172. doi: 10.1083/jcb.200109010

Schoeman, D., and Fielding, B. C. (2019). Coronavirus envelope protein: current knowledge. *Virol. J.* 16:69. doi: 10.1186/s12985-019-1182-0

Shepley-McTaggart, A., Sagum, C. A., Oliva, I., Rybakovsky, E., DiGuilio, K., Liang, J., et al. (2021). SARS-CoV-2 Envelope (E) protein interacts with PDZ-domain-2 of host tight junction protein ZO1. *PLoS One* 16:e0251955.

Songyang, Z., Fanning, A. S., Fu, C., Xu, J., Marfatia, S. M., Chishti, A. H., et al. (1997). Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science* 275, 73–77. doi: 10.1126/science.275.5296.73

Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., et al. (2021). Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* 592, 438–443. doi: 10.1038/s41586-021-03402-9

Teoh, K.-T., Siu, Y.-L., Chan, W.-L., Schlüter, M. A., Liu, C.-J., Peiris, J. S. M., et al. (2010). The SARS coronavirus E protein interacts with PALS1 and alters tight junction formation and epithelial morphogenesis. *Mol. Biol. Cell* 21, 3838–3852. doi: 10.1091/mbc.E10-04-0338

Terrien, E., Chaffotte, A., Lafage, M., Khan, Z., Préhaud, C., Cordier, F., et al. (2012). Interference with the PTEN-MAST2 interaction by a viral protein leads to cellular relocalization of PTEN. *Sci. Signal* 5:ra58.

Torres-Flores, J. M., and Arias, C. F. (2015). Tight Junctions Go Viral! *Viruses* 7, 5145–5154. doi: 10.3390/v7092865

Toto, A., Ma, S., Malagrinò, F., Visconti, L., Pagano, L., Stromgaard, K., et al. (2020). Comparing the binding properties of peptides mimicking the Envelope protein of SARS-CoV and SARS-CoV-2 to the PDZ domain of the tight junction-associated PALS1 protein. *Protein Sci.* 29, 2038–2042. doi: 10.1002/pro.3936

Veneti, L., Seppälä, E., Larsdatter Storm, M., Valcarcel Salamanca, B., Alnes Buanes, E., Aasand, N., et al. (2021). Increased risk of hospitalisation and intensive care admission associated with reported cases of SARS-CoV-2 variants B.1.1.7 and B.1.351 in Norway, December 2020 –May 2021. *PLoS One* 16:e0258513. doi: 10.1371/journal.pone.0258513

Vincentelli, R., Luck, K., Poirson, J., Polanowska, J., Abdat, J., Blémont, M., et al. (2015). Quantifying domain-ligand affinities and specificities by high-throughput holdup assay. *Nat. Methods* 12, 787–793. doi: 10.1038/nmeth.3438

Wang, P., Nair, M. S., Liu, L., Iketani, S., Luo, Y., Guo, Y., et al. (2021). Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* 593, 130–135. doi: 10.1038/s41586-021-03398-2

Wang, Z., Sandiford, S., Wu, C., and Li, S. S.-C. (2009). Numb regulates cell-cell adhesion and polarity in response to tyrosine kinase signalling. *EMBO J.* 28, 2360–2373. doi: 10.1038/emboj.2009.190

Weber, P., Pissis, C., Navaza, R., Mechaly, A. E., Saul, F., Alzari, P. M., et al. (2019). High-Throughput Crystallization Pipeline at the Crystallography Core Facility of the Institut Pasteur. *Molecules* 24:E4451. doi: 10.3390/molecules24244451

Zhou, H., Xu, Y., Yang, Y., Huang, A., Wu, J., and Shi, Y. (2005). Solution Structure of AF-6 PDZ Domain and Its Interaction with the C-terminal Peptides from Neurexin and Bcr. *J. Biol. Chem.* 280, 13841–13847. doi: 10.1074/jbc.M411065200

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

REVIEW
published: 03 March 2022
doi: 10.3389/fmicb.2022.847588

Let's Get Physical: Flavivirus-Host Protein–Protein Interactions in Replication and Pathogenesis

*Adam T. Fishburn[1], Oanh H. Pham[1], Matthew W. Kenaston[1], Nitin S. Beesabathuni[1,2] and Priya S. Shah[1,2]**

[1]Department of Microbiology and Molecular Genetics, University of California, Davis, Davis, CA, United States, [2]Department of Chemical Engineering, University of California, Davis, Davis, CA, United States

Flaviviruses comprise a genus of viruses that pose a significant burden on human health worldwide. Transmission by both mosquito and tick vectors, and broad host tropism contribute to the presence of flaviviruses globally. Like all viruses, they require utilization of host molecular machinery to facilitate their replication through physical interactions. Their RNA genomes are translated using host ribosomes, synthesizing viral proteins that cooperate with each other and host proteins to reshape the host cell into a factory for virus replication. Thus, dissecting the physical interactions between viral proteins and their host protein targets is essential in our comprehension of how flaviviruses replicate and how they alter host cell behavior. Beyond replication, even single interactions can contribute to immune evasion and pathogenesis, providing potential avenues for therapeutic intervention. Here, we review protein interactions between flavivirus and host proteins that contribute to virus replication, immune evasion, and disease.

Keywords: flavivirus, protein–protein interactions, virus replication, virus pathogenesis, virus host interactions, autophagy

OPEN ACCESS

Edited by:
Rohit K. Jangra,
LSU Health Sciences Center-Shreveport, United States

Reviewed by:
Denise Haslwanter,
Albert Einstein College of Medicine, United States
Yaw Shin Ooi,
Duke-NUS Medical School, Singapore

***Correspondence:**
Priya S. Shah
prsshah@ucdavis.edu

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *03 January 2022*
Accepted: *31 January 2022*
Published: *03 March 2022*

Citation:
Fishburn AT, Pham OH, Kenaston MW, Beesabathuni NS and Shah PS (2022) Let's Get Physical: Flavivirus-Host Protein–Protein Interactions in Replication and Pathogenesis. Front. Microbiol. 13:847588. doi: 10.3389/fmicb.2022.847588

INTRODUCTION

Flavivirus is a genus of positive-sense, single-stranded RNA (ssRNA+), arthropod-transmitted viruses within the family *Flaviviridae*. The ssRNA genome contains a single open-reading frame, which is translated by host ribosomes into a large viral polyprotein. This polyprotein is co-translationally processed by viral and host proteases into 10 individual viral proteins. Three of these proteins are referred to as structural proteins which include Capsid (C), pre-Membrane (prM), and Envelope (Env) proteins, which form the physical virion. The remaining seven proteins are referred to as non-structural (NS) proteins, which include NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5. These proteins are not components of infectious virions but rather play broad roles within infected cells in generating virus progeny. Four distinct enzymatic activities are encoded within two NS proteins. NS3 serves as the helicase. It also interacts with NS2B as a cofactor (NS2B3) to form the viral protease (Falgout et al., 1991). NS5 is both the RNA-dependent RNA-polymerase and methyltransferase, which synthesizes and caps new RNA genomes (Tan et al., 1996; Egloff et al., 2002; Ray et al., 2006). The RNA genome also contains 3' and 5' untranslated regions (UTRs) with loop-like structures that play roles in genome stability and translation (Alvarez et al., 2005; Lodeiro et al., 2009). Genome replication occurs within the remodeled ER in involuted structures

referred to as virus replication organelles or replication compartments (Gillespie et al., 2010). These substructures serve to concentrate replication substrates and hide viral nucleic acids from detection by the host immune response. Here, the viral NS proteins assemble into the replication complex, which performs the enzymatic steps of RNA synthesis (Welsch et al., 2009; Yi et al., 2012). Viral ssRNA+ is initially used as a template for the synthesis of negative-sense ssRNA (ssRNA-), which in turn is used as a template to synthesize more ssRNA+. As replication progresses these genomes are either further amplified or packaged into progeny virions. In addition to genome replication by the replication complex, viral NS proteins mediate different aspects of virus replication, such as ER remodeling and modulating the host immune response.

The most well-studied flaviviruses are those that cause significant disease in humans. For mosquito-transmitted viruses this includes dengue virus (DENV), Zika virus (ZIKV), West Nile virus (WNV), yellow fever virus (YFV), and Japanese encephalitis virus (JEV). These flaviviruses are all transmitted by mosquitoes of either *Aedes* or *Culex* spp. (Huang et al., 2014). DENV is the most widespread and threatening flavivirus. Currently, there are four well-described serotypes of DENV, referred to as DENV1–DENV4, that each have distinct molecular and physiological characteristics (Chandramouli et al., 2010; Yung et al., 2015). World-wide there are an estimated 390 million cases of DENV infection per year, occurring across 128 countries, although most infections occur in Asia (Brady et al., 2012; Bhatt et al., 2013). Recently, the emergence of a fifth DENV serotype (DENV5) with a sylvatic replication cycle has been reported (da Silva Voorham, 2014; Mustafa et al., 2015). However, DENV5 remains a controversial topic, as the evidence to support the existence of this serotype is limited and mathematical modeling suggests a low probability for the emergence of new DENV serotypes (Sánchez-González et al., 2021). ZIKV recently received major research due to the 2015–2016 epidemic and the revelation that congenital ZIKV infection causes birth defects, collectively referred to as congenital Zika syndrome (CZS; de Araújo et al., 2016; Moore et al., 2017). ZIKV infection in adults is usually limited to mild flu-like illness but can be rarely associated with Guillain-Barré Syndrome, a condition where nerves are damaged, usually in the extremities (Cao-Lormeau et al., 2016). While less common, WNV and JEV can also cause encephalitis (Solomon et al., 2000; DeBiasi and Tyler, 2006). Tick-borne flaviviruses are transmitted by many different ticks, including *Haemaphysalis*, *Ixodes*, *Dermacentor*, and *Ornithidoros* spp. (de la Fuente et al., 2017). These account for much fewer total human infections, many of which are in vastly different geographical settings compared to the tropical climates which host mosquitoes. The most notable of these are tick-borne encephalitis virus (TBEV) and Powassan virus (POWV). While the number of human infections arising from tick-borne viruses is relatively limited the resulting disease can be very severe. Encephalitis resulting from TBEV infection can appear in several forms, with an overall mortality rate of around 2% (Ruzek et al., 2019). Given the severity of disease caused by flaviviruses, it is critical to understand mechanisms of replication and pathogenesis.

In general, flaviviruses have a conserved replication cycle, which includes viral entry, virion fusion with the endosome and release of viral RNA, genome replication and protein production in the ER, virion packaging and processing through the secretory pathway, and viral release *via* exocytosis (**Figure 1**). At each of these stages, flaviviruses are dependent on host machinery to perform necessary functions. The limited flavivirus genome size requires them to maximize the functions of each protein they encode. Flavivirus replication is therefore largely dependent on the interactions between viral proteins and host proteins to manipulate their biology through direct and indirect mechanisms. These protein interactions can be identified using targeted and comprehensive screening approaches (Coyaud et al., 2018; Scaturro et al., 2018; Shah et al., 2018; Cao et al., 2019; Li et al., 2019; Wang et al., 2019; Golubeva et al., 2020; Tan et al., 2020; Zeng et al., 2020; Lemasson et al., 2021). This review will focus on virus-host protein–protein interactions (PPIs) emerging from both targeted and comprehensive studies that directly facilitate flavivirus replication, dampen host immune response, or disrupt cellular processes to cause disease. While not covered here, it is worth noting that additional virus-host interactions, such as RNA-protein, and RNA–RNA interactions also play important roles in flavivirus replication and disease (Funk et al., 2010; Chavali et al., 2017; Damas et al., 2019; Ooi et al., 2019).

FLAVIVIRUS-HOST PPIS FACILITATE FUNDAMENTAL ASPECTS OF FLAVIVIRUS REPLICATION

In this section, we review the data emerging from both comprehensive and targeted studies of flavivirus-host PPIs as they relate to various stages of flavivirus replication.

Virus Attachment Factors

The first step in any virus replication cycle is entry into the host cell and involves the classic virus-host protein interaction between a virion structural protein and a host attachment factor. In the case of flaviviruses, Env proteins on the virion exterior interact and attach to host factors on the plasma membrane surface. Flavivirus Env proteins are quite promiscuous and can bind many different host factors. While each flavivirus appears to bind multiple host factors, not all flaviviruses use the same set of host factors for entry. Generally, flaviviruses use TAM (e.g., Tyro3, Axl, and Mer) family receptor tyrosine kinases (Meertens et al., 2012; Richard et al., 2017), phosphatidyl serine receptor T-cell immunoglobulin (TIM; Dejarnac et al., 2018; Niu et al., 2018; Zhang et al., 2022), C-type lectin receptors (e.g., DC-SIGN; Miller et al., 2008; Pereira et al., 2019; Routhu et al., 2019), integrins (Chu and Ng, 2004; Schmidt et al., 2013), heat-shock proteins 70/90 (Reyes-Del Valle et al., 2005; Das et al., 2009; Pujhari et al., 2019), laminin receptor (LAMR1; Tio et al., 2005; Thongtan et al., 2012), and heparan sulfate (Chen et al., 1997; Germi et al., 2002) as means of attachment. Subtle differences in Env protein sequence likely contribute to differences in host

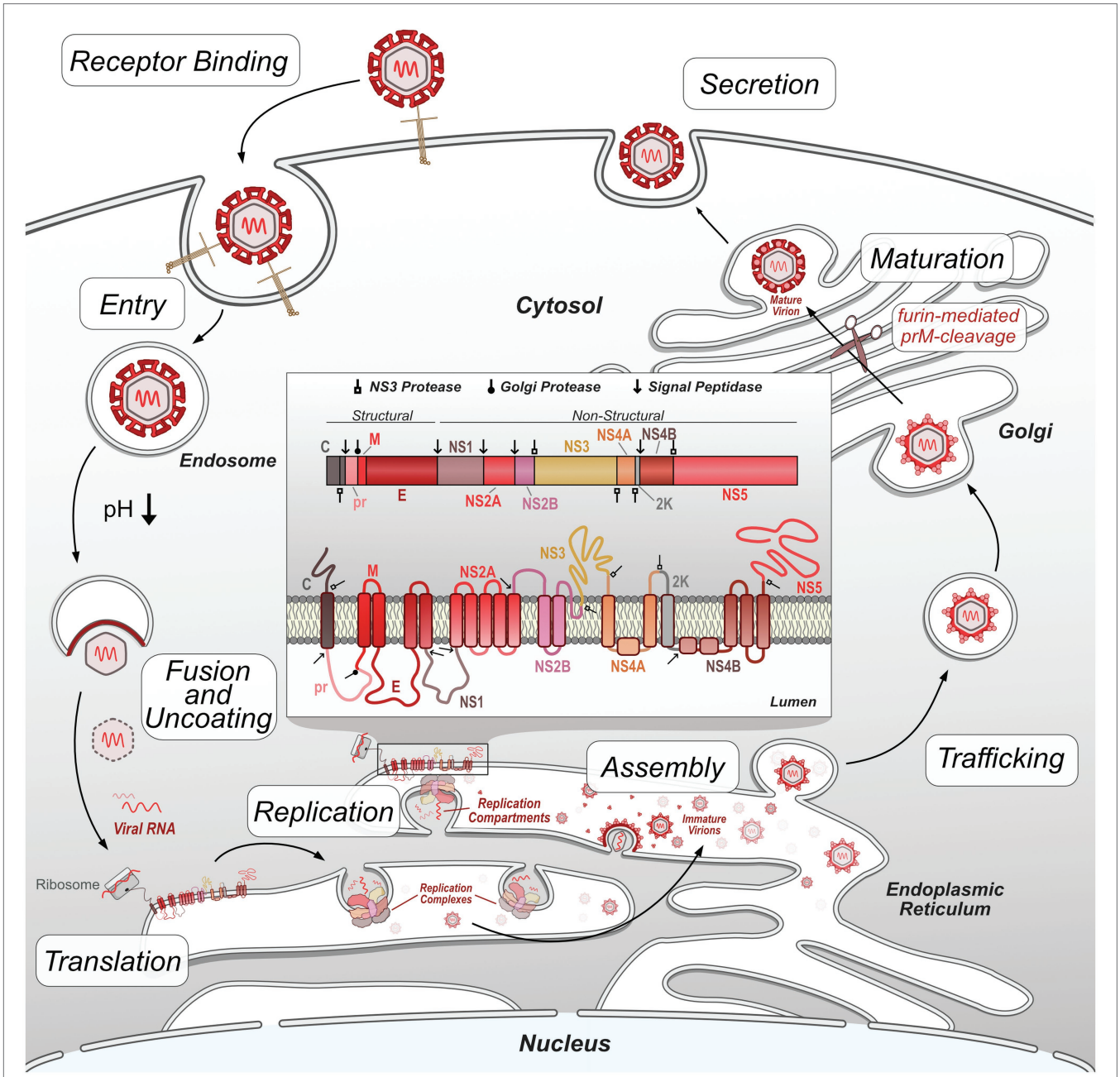


FIGURE 1 | Flavivirus replication cycle. Flavivirus infection begins by receptor-mediated binding to the host cell and entry *via* clathrin-mediated endocytosis. Decreases in endosome pH trigger virion envelope fusion with the endosome membrane, releasing the genome into the host cytosol. After uncoating, the viral RNA genome is translated by host ribosomes into the viral polypeptide, which is co-translationally processed, including insertion of transmembrane proteins into the ER and cleavage of the polypeptide by host and viral proteases into individual proteins. Non-structural viral proteins form replication complexes, which replicate viral RNA genomes within invaginated ER compartments. Structural viral proteins are assembled and loaded with viral genetic material in the ER prior to entering the trans-Golgi network. In the Golgi, immature virions are processed by furin protease cleavage of prM, resulting in mature, infectious virions. These virions exit the cell by exocytosis and continue the replication cycle by initiating infection of other host cells.

factor usage, and distinct tissue tropisms between flaviviruses. The well-established flavivirus-host attachment factors are described in **Figure 2**; however, it is essential to note that there are likely others that each virus uses that have not been identified.

Interestingly, some entry determinants are dependent on specific intracellular virus-host interactions that provide newly generated progeny virions with additional receptor targets. A recent study elegantly showed how the interaction between TRIM7, an E3-ubiquitin ligase, and Env resulted in specific polyubiquitination in the infected cell that allowed progeny virion binding to Tim-1 of the new target cell (Giraldo et al., 2020). Here, ZIKV Env was ubiquitinated on three lysine

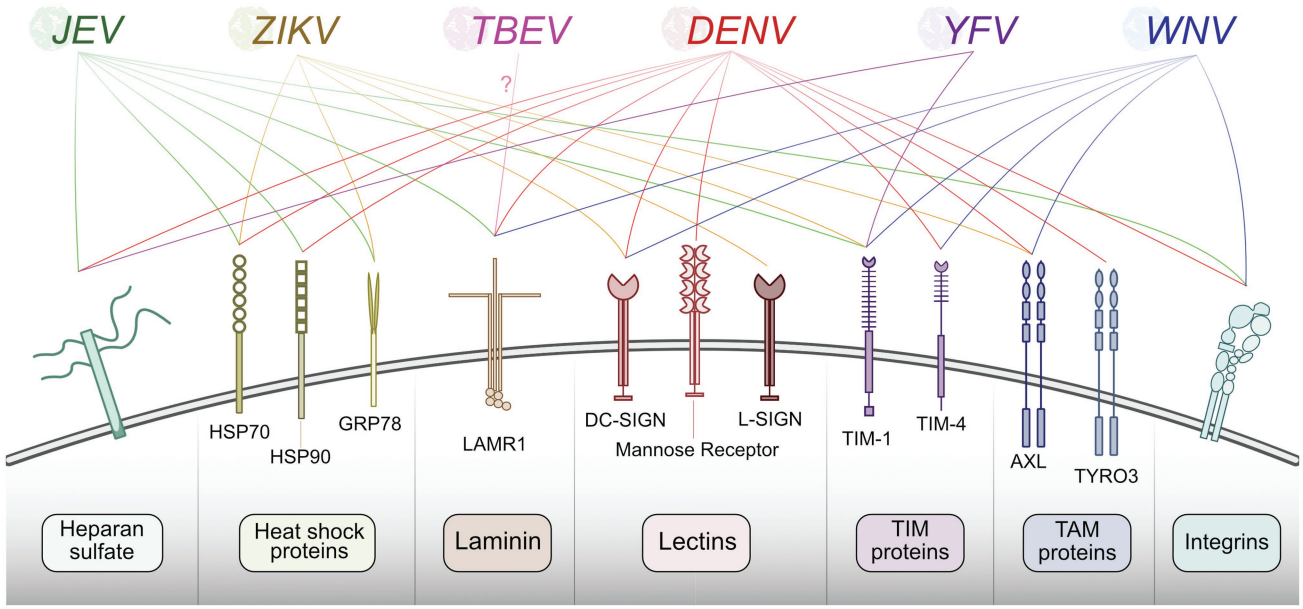


FIGURE 2 | Summary of host proteins used by flaviviruses for entry. Flaviviruses recognize and bind plasma membrane host factors to initiate entry into the host cell. Different flaviviruses utilize a similar pool of host proteins for entry.

residues: K38, K63, and K281. A recombinant virus in which one of these lysines was swapped with arginine (E-K38R) was significantly attenuated in JEG-3 placental trophoblast cells and *in vivo* in mice, but not in mosquitoes. Intriguingly, the ZIKV E-K38R titers *in vivo* varied significantly by tissue, suggesting that Env ubiquitination may drive tissue specific tropism. To further explore this, they generated TRIM7 knockout cells which attenuated ZIKV replication in JEG-3 cells but did not affect DENV replication in A549 lung epithelial cells. Similar results were observed in *Trim7*$^{-/-}$ knockout mice. In this model, ZIKV replicated similarly to WT in the heart, liver, lung, and muscle, whereas ZIKV replication in the brain, eyes, and reproductive tissues was significantly reduced. Finally, they identified that Tim-1 interacted with wild-type ZIKV but very minimally with K38R viral particles, suggesting that ubiquitination at this site is critical for the interaction and virus entry through Tim-1. This was supported both by reduced attachment of ZIKV to Tim-1 knockout JEG-3 cells and reduced replication of ZIKV in the brains of *Tim1*$^{-/-}$ knockout mice (Giraldo et al., 2020). All together, these results indicate an important interaction between ZIKV Env and TRIM7, providing ubiquitination that mediates entry into the brain and other tissues that are major contributors to ZIKV pathogenesis. Currently, it is unknown if other flaviviruses use TRIM7 ubiquitination of Env to mediate entry *via* TIM-1, as many other flaviviruses use TIM-1 as a host receptor (Meertens et al., 2012; Dejarnac et al., 2018; Niu et al., 2018; Zhang et al., 2022).

Intriguingly, a recent study identified this same ZIKV Env ubiquitination is also targeted by host factors and restricts virus infection (Hu et al., 2021). The laminin receptor LAMR1 consists of an intracellular domain, a transmembrane domain, and a larger extracellular domain, which is known to be utilized as an attachment factor by several flaviviruses that are not ZIKV (Tio et al., 2005; Sakoonwatanyoo et al., 2006; Bogachek et al., 2008; Thongtan et al., 2012). Unsurprisingly, ZIKV Env was also found to interact with LAMR1. However, it only interacts with the intracellular region, not the extracellular region that would mediate extracellular virion attachment. Overexpression of LAMR1 reduced virus replication and repression of LAMR1 by shRNA resulted in significant increases in viral titer. The interaction between ZIKV Env and LAMR1 is mediated by a single amino acid in Env, G282. Interestingly, G282 is very highly conserved among ZIKV strains, but is not conserved at all with other flaviviruses. Further, the authors found that LAMR1 recruits eukaryotic translation initiation factor 3 subunit 5 (EIF3S5), a member of the ubiquitin proteasome system (UPS). Knockdown of EIF3S5 reduced Env deubiquitination and increased the levels of NS5, Env, and viral RNA in infected HeLa cells (Hu et al., 2021). Thus, Env ubiquitination can have opposing effects mediated by different Env-host protein interactions.

It is worth noting that there is plentiful information on the host entry factors of mosquito-borne flaviviruses. However, knowledge on the attachment factors utilized by tick-borne flaviviruses is extremely limited. One recent study attempted to identify attachment factors for Langat virus (LGTV). They found that LGTV did not utilize heparin sulfate, *O*- or *N*-linked glycans, or glycolipids for entry, suggesting that the host receptor is protein in nature. However, they were unable to definitively identify such a protein (Rodrigues et al., 2019). A pair of studies suggests one such attachment factor might be LAMR1, although additional studies are required (Protopopova et al.,

1997; Malygin et al., 2009). Another recently published study used multiple methods to identify TIM-1 as an entry factor for TBEV (Zhang et al., 2022).

Fusion and Uncoating

After a flavivirus binds to an extracellular host entry factor, it enters the intracellular space by clathrin-mediated endocytosis (Acosta et al., 2008; van der Schaar et al., 2008). There is several cargo internalization factors involved in this process that are necessary for flavivirus infection. Specifically, LY6E has been shown to reorganize itself into tubule-like structures to support entry of WNV, ZIKV, and DENV (Hackett and Cherry, 2018), though direct virus-host interactions have yet to be identified in this case. Following endocytosis, the membrane of the virus envelope must fuse with the endosome membrane. This process happens through a mechanism not requiring any physical virus-host protein interactions. V-ATPase pumps protons from the cytoplasm into the lumen of various organelles, including endosomes, decreasing the intra-endosomal pH (Lafourcade et al., 2008; Kozik et al., 2013). This triggers conformational changes in viral Env proteins that ultimately lead to their insertion into the endosome membrane and formation of the fusion pore, releasing the nucleocapsid to the cytosol (Allison et al., 1995; Chao et al., 2014). Once released the viral genome is not immediately capable of being translated. It first must be stripped of the capsid proteins that otherwise stabilize and protect viral RNA. This occurs through the ubiquitination of Capsid proteins by host UBA1 (Byk et al., 2016), with subsequent nucleocapsid disassembly shown to be mediated by VCP (Ramanathan et al., 2020). Once uncoated, the flavivirus RNA genome may be translated into the viral polyprotein.

INTERACTIONS INVOLVED IN VIRAL PROTEIN TRANSLATION AND STABILITY

Translation of flavivirus genomes into functional viral proteins is dependent on the activity of several host pathways. As is true of all viruses, host ribosomes are required to initially translate the genome into the viral polyprotein. Several host proteases are necessary for polyprotein cleavage into singular proteins. However, many of these NS proteins contain multiple transmembrane domains, specifically NS2A, NS2B, NS4A, and NS4B, which all must be correctly inserted into the ER membrane in the correct orientation in order to be functional (Miller et al., 2007; Xie et al., 2013; Li et al., 2015, 2016). The stability and insertion of these proteins are performed by the signal-recognition particle (SRP), host SEC61 translocon, and ER membrane complex (EMC). These complexes have been shown to be critical host factors for many viruses, including flaviviruses (Sessions et al., 2009; Heaton et al., 2016; Marceau et al., 2016; Zhang et al., 2016). In this section, we will review recent work that has advanced our understanding of how flaviviruses co-opt these complexes during infection (**Figure 3**).

SRP-Translocon Pathway

Flaviviruses replicate within the ER and utilize ER-associated ribosomes for translation. The translated polyprotein contains many transmembrane domains that must be properly integrated into the ER membrane. This function is performed by the SRP-translocon pathway, in which the SRP ribonucleotide complex binds and identifies a hydrophobic transmembrane region of the nascent polypeptide, arrests translation, and brings the ribosome to a translocon where translation continues (Keenan et al., 2001). In eukaryotes, the Sec translocon is made up of the SEC61 complex (SEC61α/β/γ), SEC62/63, and a number of other proteins that can vary depending on substrate (Denks et al., 2014). Flavivirus polypeptide insertion into the ER membrane is thus at least partially reliant on the interaction with SRP and SEC proteins. Unsurprisingly, these proteins have been found in several flavivirus-host protein interaction studies, including interactions between ZIKV/DENV NS4A with SEC62, SEC61γ, and SRPR, NS4A/2B with SEC61β, and NS4B with SEC61α (Scaturro et al., 2018; Shah et al., 2018). Along with these interaction-based screens, these proteins have been identified in genetic screens as host factors supporting flavivirus replication (Krishnan et al., 2008; Sessions et al., 2009; Marceau et al., 2016; Zhang et al., 2016; Hoffmann et al., 2021). Interestingly, while the SEC61 translocon is essential for protein biogenesis, pharmacological modulation of this complex inhibits DENV and ZIKV replication (Heaton et al., 2016; Monel et al., 2017; Shah et al., 2018). Together, these results highlight the importance of the SRP/SEC61 translocon on flavivirus replication and the potential for a pan-flaviviral drug target.

ER Membrane Complex

The EMC co-translationally interacts with nascent proteins and prevents their degradation by associating with chaperones. The EMC preferentially stabilizes multipass membrane proteins that may otherwise have difficulty being inserted into the ER membrane, thereby avoiding misfolding and degradation (Jonikas et al., 2009; Shurtleff et al., 2018). Similar to the SEC61 complex, EMC proteins have been identified in a number of flavivirus protein interaction (Coyaud et al., 2018; Shah et al., 2018) and genetic screens (Ma et al., 2015; Marceau et al., 2016; Savidis et al., 2016; Hoffmann et al., 2021), underlining their importance. Three recent papers dissect their role in flavivirus replication. Through a combination of biochemical assays and pulse-chase experiments with gene knockdown, Lin et al. (2019) demonstrated that EMC1 promoted NS4B biogenesis, but not its post-translational stability. Interestingly, NS4B's dependence on the EMC arises from its two N-terminal transmembrane regions, which are marginally hydrophobic, as altering the nature of these regions in either direction, more or less hydrophobic, rescued expression in EMC knockout cells. Together, this suggests that the generation and co-translational stability of flavivirus multi-pass proteins, including NS4B, depends on the interaction with EMC for protection from degradation and integration into the membrane (Lin et al., 2019). Similar inhibition of virus replication and decreases in viral protein production were also shown by (Ngo et al., 2019), using a

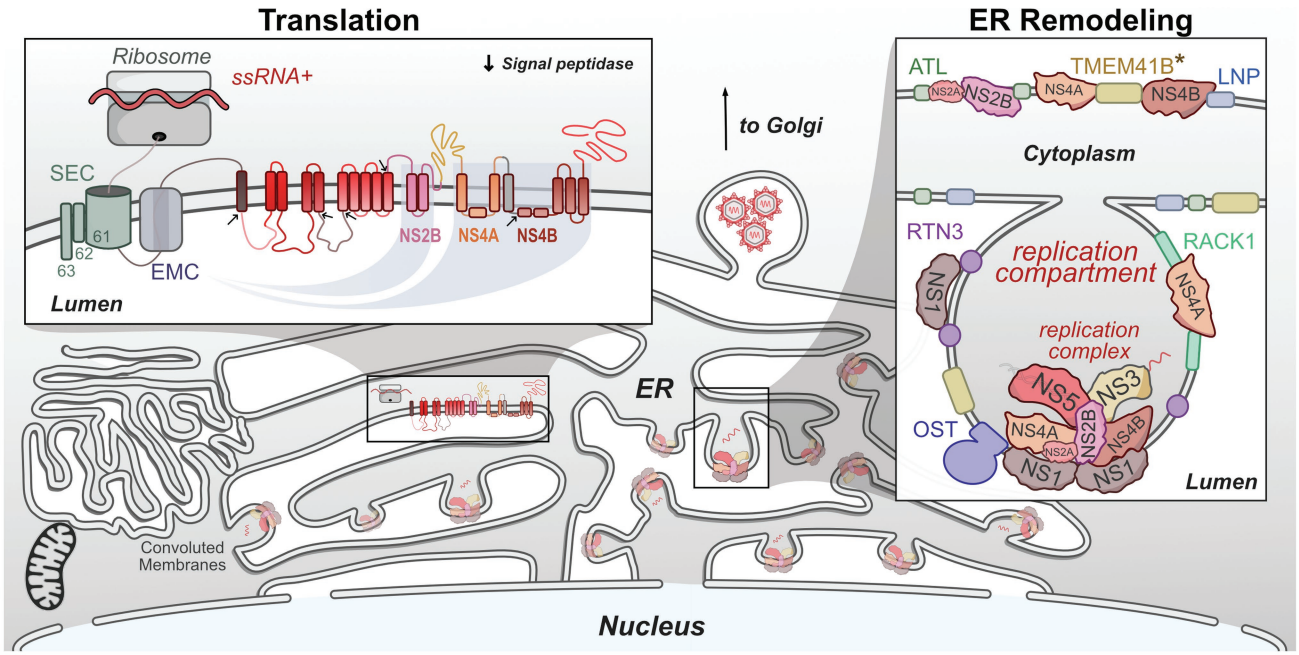


FIGURE 3 | Flaviviruses co-opt host proteins to remodel the ER. Flaviviruses dramatically alter the morphology of the host ER to create a niche that maximizes the efficiency of genome replication and virion packaging. Replication compartments are formed by the involution of the ER membrane by both viral and host proteins. Viral replication complexes reside within these compartments and carry out RNA replication. These complexes also physically associate with host proteins. Viral single-stranded RNA (ssRNA)+ genomes are translated on the ER by host ribosomes. The resulting viral polyprotein is co-translationally processed to ensure its stability, insertion into the ER membrane, and proper cleavage into individual viral proteins. *TMEM41B is known to interact with either ZIKV NS4A or YFV NS4B and may facilitate ER remodeling.

similar CRISPR knockout setup. Using a dual-fluorescence reporter system they were able to identify that NS4B's underlying reliance on the EMC stems from its link to NS4A through the 2K peptide, a transmembrane region which serves as a signal sequence for the translocation of NS4B (Ngo et al., 2019). Barrows et al. (2019) additionally found that knockout of EMC4 led to near complete loss in replication of DENV, ZIKV, and YFV, but did not affect WNV replication at all. They speculated this difference may arise from the WNV being transmitted by *Culex* mosquitoes, rather than *Aedes*. This vector specific hypothesis was supported by their finding that DENV titer in *Aedes* mosquito midguts was reduced after siRNA targeting of EMC2/3/4, although the decreases in replication were not nearly as severe as what was observed in human cells (Barrows et al., 2019). Altogether, the EMC is a vital host factor utilized by *Aedes*-transmitted flaviviruses for correct viral protein insertion into the ER membrane.

Signal Peptidase and Oligosaccharyltransferase Complexes

Gene perturbation screens have long been used to identify essential flavivirus host factors. With the advances of CRISPR, these screens have become even more powerful. A pair of such screens published together in 2016 mapped host factors of multiple flaviviruses (Marceau et al., 2016; Zhang et al., 2016). Unsurprisingly, proteins involved in ER translocation of and polypeptide stability, including SEC61B and EMC proteins, were among the hits in these screens. Additionally, signal peptidase complex (SPCS) and oligosaccharyltransferase (OST) complex were found to be essential for replication of many flaviviruses. Knockout of SPCS1, a major component of the SPCS, in HEK293T embryonic kidney cells completely ablated replication of all tested flaviviruses, but had little effect on other unrelated RNA viruses, suggesting its role in virus replication was specific to *Flaviviridae*. Further experiments revealed that the SPCS1 is responsible for several polyprotein cleavage events, specifically C-prM, prM-E, E-NS1, and 2K-NS4B (Zhang et al., 2016). The cleavage between NS1 and NS2A occurs through an unknown signal peptidase pathway mechanism. Interestingly, the OST complex plays a role in replication separate from its enzymatic activity. Normally, the OST complex is responsible for the *N*-linked glycosylation of host proteins. Knockout of major OST complex component STT3A had major effects on the replication of DENV, YFV, WNV, JEV, and ZIKV (Marceau et al., 2016; Zhang et al., 2016). However, these replication defects could be rescued by the expression of catalytically dead STT3A mutants, suggesting that the OST complex serves virus replication through function outside its ability to glycosylate proteins (Marceau et al., 2016). Physical interactions with flavivirus replication complex members NS1, NS2B, NS3, and NS4B along with its close association with sites of virus replication in the ER suggest that the OST complex may

serve a structural role in genome replication (Marceau et al., 2016; Hafirassou et al., 2017).

ER REMODELING AND VIRUS REPLICATION COMPARTMENT FORMATION

The majority of flavivirus replication occurs within the ER membrane. Flaviviruses employ a variety of mechanisms to remodel the host ER into a niche, which maximizes the efficiency of genome replication and viral packaging. The task of remodeling the ER is primarily performed by the viral NS proteins through a combination of direct remodeling and specific virus-host interactions. NS4A and NS4B contain transmembrane domains, which pass through the ER and helices that lie in the plane of the ER lumen to induce positive curvature of the membrane (Roosendaal et al., 2006; Miller et al., 2007; Kaufusi et al., 2014). Membrane alteration is further driven by the oligomerization of NS4A (Lee et al., 2015). In addition to the action of these viral proteins on their own, they recruit and hijack the function of a number of other host proteins involved in ER morphology. The highly curved and tubular host ER membrane system is stabilized and maintained by several protein families including the reticulon (RTN) family, the atlastin (ATL) family, and the Lunapark (LNP) protein (Goyal and Blackstone, 2013; Wang et al., 2016). Even beyond these canonically ER proteins, other host factors have been shown to be involved in flavivirus-mediated ER modifications. Recently several groups have evaluated the roles of these protein families in flavivirus replication and ER remodeling during infection.

Reticulon

Proteins in the RTN family all have a domain including two transmembrane regions separated by a single hydrophilic loop which, similarly to NS4A, induce membrane curvature (Zurek et al., 2011). Reticulon 3A (RTN3A) is known to be involved in the replication of other viruses, including the *Flaviviridae* family member Hepatitis C virus (HCV; Tang et al., 2007; Diaz et al., 2010; Barajas et al., 2014). Aktepe et al. (2017) found that the broadly expressed RTN3.1A plays a role in the replication of several flaviviruses, including WNV, DENV, and ZIKV. RTN3.1A colocalized with sites of virus replication and siRNA silencing resulted in significant decreases in viral titer after infection. ZIKV infection in RTN3.1A-silenced cells displayed dramatically less membrane curvature with fewer replication complexes. Using a combination of immunofluorescence microscopy and fluorescence resonance energy transfer (FRET), the authors determined that RTN3.1A specifically interacted with WNV NS4A, whereas ZIKV and DENV NS4A did not (Aktepe et al., 2017). However, later proteomics studies did identify an interaction between ZIKV NS4A and RTN3 (Shah et al., 2018). A yeast-two hybrid screen also showed that ZIKV NS4A and NS2B interact with RTN1, suggesting that flaviviruses may utilize RTN family members differentially for roles in ER remodeling (Golubeva et al., 2020).

Atlastins

The ATL family of proteins is composed of three (ATL1/2/3) membrane-bound, dynamin-related GTPases that function in maintaining Golgi (ATL1) and ER (ATL2/3) morphogenesis through the formation of three-way junctions (Rismanchi et al., 2008; Hu et al., 2009). ZIKV is known to actively remodel the ER and induce the formation of large ER-derived cytoplasmic vacuoles. Ultimately this results in cell death through paraptosis, a caspase-independent, non-apoptotic form of cell death (Monel et al., 2017). The formation of these vacuoles in HeLa cells is dependent on the activity of ATLs. Knockout of ATL2 and ATL3 led to nearly a complete loss in the formation of these vacuoles during ZIKV infection and significant reduction in ZIKV replication. These phenotypes could be rescued by expression of wild-type ATL3, but not a GTPase-deficient mutant (Monel et al., 2019). Another group similarly found that knockdown of ATL2/3 reduced the replication of both ZIKV and DENV and that ATL3 played an important role in DENV maturation (Neufeldt et al., 2019). There appears to be multiple methods by which flaviviruses physically interact with ATL proteins. Using co-immunoprecipitation (Co-IP) and immunofluorescence analysis, one group determined that ATL3 strongly interacted with both ZIKV NS2A and NS2B3, although they did identify partial interaction with both NS4A and NS4B as well (Monel et al., 2019). ATL2/3 was found to interact with DENV NS2B, NS3, and NS5. Interestingly, ATL3 was also found to further interact with DENV NS1, envelope, and capsid proteins (Neufeldt et al., 2019). Flavivirus-ATL interactions have also been identified in a number of proteomic screens including WNV NS4B with ATL2 (Li et al., 2019) and ZIKV NS4A and NS2A with both ATL1 and ATL2 (Coyaud et al., 2018). Thus, while multiple flaviviruses hijack atlastin proteins, the molecular mechanisms appear to be unique.

Lunapark

While RTN family proteins induce curvature within the ER membrane and ATL form three-way tubular junction, the LNP protein stabilizes these junctions and is required for their mobility, a necessary feature of the dynamic ER (Chen et al., 2015). Similar to RTN and ATL, siRNA silencing of LNP results in significant reduction in flavivirus induced replication compartments and corresponding decreases in genome replication. Using Co-IP, Tran et al. (2021) identified that TBEV NS4B interacted with LNP through its C-terminal region (Tran et al., 2021). Additionally, ZIKV NS4A has been identified to interact with LNP, and may constitute virus-specific mechanisms of ER-remodeling (Shah et al., 2018). All together, these findings show that flaviviruses physically hijack a number of host pathways to remodel the ER membrane system to create a space conducive to virus replication.

TMEM41B and VMP1

A recent CRISPR genetic screen assessed host factors involved in flavivirus infection (Hoffmann et al., 2021). In addition to identifying many ER proteins discussed previously, the authors also identified two transmembrane ER proteins, TMEM41B

and VMP1. These proteins function as phospholipid scramblases and have similar roles in lipid mobilization, lipoprotein biogenesis, autophagy, and the induction of membrane curvature (Zhao et al., 2017; Moretti et al., 2018; Morita et al., 2019; Huang et al., 2021). Knockout of either gene dramatically inhibited the replication of a wide range of mosquito- and tick-borne flaviviruses. TMEM41B was also shown to be critical for infection across multiple cell types, including mosquito C6/36 cells. The authors found TMEM41B interacts and colocalizes with ZIKV NS4A and YFV NS4B during infection, which is supported by the previous identification of ZIKV NS4B's interaction with TMEM41B (Scaturro et al., 2018). Given its role in inducing membrane curvature, this suggested that TMEM41B may be involved in the formation of viral replication compartments in the ER. Intriguingly, TMEM41B-deficient cells were observed to have heightened innate immune responses after infection. Using YFV replicons, this was elegantly shown to be due to increased sensing of viral dsRNA in TMEM41B knockout cells (Hoffmann et al., 2021). Together, these results show that TMEM41B is a pan-flavivirus host factor that is likely involved in the formation of replication compartments in the ER, and loss of this protein results in the inability to retain viral dsRNA in the ER, leading to detection by host immune response sensors.

Vimentin

The RTN, ATL, and LNP family proteins are logical targets for virus-mediated ER remodeling based on their canonical roles in regulating ER morphology. However, flaviviruses are also capable of hijacking host proteins with more divergent functions to establish replication compartments within the ER. A 2014 study by Teo and Chu (2014) established that DENV NS4A interacted with host vimentin, a major component of cytoskeletal intermediate filaments, to anchor replication compartments in the ER. They found that the N-terminal cytoplasmic region of NS4A mediated this interaction and that DENV infection increasing the phosphorylation of vimentin, promoting depolymerization and reassembly to the perinuclear region where it was utilized for virus replication. Phosphorylation of vimentin was shown to be crucial for replication as siRNA silencing of the vimentin-targeting kinase CaMKIIγ led to significant decreases in DENV replication (Teo and Chu, 2014).

Receptor for Activated C Kinase 1

The Receptor for Activated C Kinase 1 (RACK1) protein is a known scaffolding protein with roles in protein shuttling, anchoring, and stabilization, as well as mediating cellular pathways through protein interactions (Adams et al., 2011). The interaction between DENV NS1 and host RACK1 was first identified in a DENV NS1 specific proteomics screen (Hafirassou et al., 2017) but the role of this interaction was not fully explored until recently by Shue et al. (2021). They performed a genome-wide CRISPR knockout screen in Huh7 cells to identify host genes involved in ZIKV replication. This identified several potential host genes including members of the EMC (discussed earlier in this review), as well as RACK1. Additionally, they found that silencing of RACK1 impacted the replication of several flaviviruses, including ZIKV, DENV, WNV, POWV, and LGTV, and even SARS-CoV-2. However, they found that YFV, herpes simplex virus (a DNA virus), and vesicular stomatitis virus (ssRNA- virus) were not affected by RACK1 silencing. Using a *Renilla luciferase* DENV replicon they determined that RACK1 specifically played a role in viral genome replication, rather than viral entry or translation. Using replication-independent expression system that induces the formation of replication compartments in the ER without virus infection they found that RACK1 silencing led to reduced formation of these compartments in the ER (Shue et al., 2021). These studies are a great example of the power of integrating proteomic and genetic screens to identify mechanisms of virus replication. In the future utilization of existing screens will advance our understanding of these mechanisms and identify new interactions that are necessary for flavivirus replication.

One interesting feature of flavivirus infection worth noting is the induction of convoluted membranes (**Figure 3**). These peculiar membranous structures contain vast arrangements of smooth ER, however, they appear to form only under certain conditions, as their presence can vary with virus or cell type (Junjhon et al., 2014; Hanners et al., 2016; Cortese et al., 2017; Offerdahl et al., 2017). Convoluted membranes contain viral proteins but lack viral RNA, suggesting these are not sites of genome replication (Welsch et al., 2009). The virus-host PPIs that contribute to the formation of these membrane structures are still under investigation. It has been shown that NS4B associated with mitochondria physically contact these structures, potentially to tether them near sites of virus replication or assembly or to dampen innate immune response signaling (Chatel-Chaix et al., 2016).

FLAVIVIRUS INTERACTIONS FOR HOST PROCESSES OUTSIDE THE ER

While the ER is a major site of flavivirus replication, virus-host protein interactions in other organelles are critical for replication. Soluble viral proteins such as NS3 and NS5 are known to have dispersed localizations during infection, thus it is unsurprising that identified interacting host proteins also have a wide range of localizations. Here, we will review important and recently identified virus-host PPIs outside context of ER replication that promotes virus replication.

Trafficking

After virions are packaged and assembled in the ER lumen they must be processed prior to release. Specifically, the prM protein on the outermost part of the virion must be cleaved by furin, a host protease within the Golgi apparatus. Cleavage sites on prM are only made accessible by the relatively acidic environment of the Golgi and secretory vesicles (Stadler et al., 1997; Yu et al., 2008; Zheng et al., 2014). This prM maturation is required to allow future viral entry into host cells after release. Vesicles containing immature virions reach the Golgi

through the host's secretory pathway or trans-Golgi network (TGbN). Golgi proteins and others involved in TGN trafficking have been identified in proteomic screens (Carpp et al., 2014), but in-depth studies on the role of these interactions in flavivirus replication are very limited. Recently several specific virus-host PPIs have been identified here with roles in virus maturation and replication.

As trafficking through the TGN is essential for flaviviruses, one anti-viral host mechanism is to limit this processing by halting virion progression at the Golgi, preventing release. One well-studied protein with this function is bone marrow stromal cell antigen 2 (BST2), also known as tetherin. BST2 is known to restrict the replication of many viruses, including filoviruses, retroviruses, and alphaviruses, by tethering virions to the cell surface or by interrupting virion release from the TGN prior to exit from the cell (Jouvenet et al., 2009; Sakuma et al., 2009; Liu et al., 2015; Ooi et al., 2015). Accordingly, several viruses have evolved measures to counteract this inhibition. For example, the Vpu protein of HIV-1 inhibits the anti-viral tethering effects of BST2, allowing release of infectious virions from the cell (Neil et al., 2008). However, there are some conflicting reports about the effects of BST2 on flaviviruses. A 2012 study described significant BST2-mediated inhibition of DENV release from Huh7 cells (Pan et al., 2012). Conversely, another study found only modest effects on non-infectious, "virus-like particle" release from TRex HEK293 cells expressing BST2 and transfected to express DENV Env (Ooi et al., 2015). Whether these discrepancies are methodological or cell-type derived is unclear. More recently, Li et al. (2017) investigated the potential mechanisms by which JEV escapes BST2 restriction. Endogenous BST2 proteins levels were actively decreased during JEV infection and expression of JEV Env alone was sufficient to reduce BST2 expression. JEV Env physically interacted with BST2 at its transmembrane and cytoplasmic loop domains, and targeted it for lysosomal degradation (Li et al., 2017). Thus, the interaction between Env and BST2 promotes virus replication by eliminating the anti-viral activity of BST2. Whether other flaviviruses interact with and inhibit BST2 using similar mechanisms requires further study. Previously we also discussed the interaction between ATL3 and multiple viral proteins, and the role of this interaction in ER remodeling. Interestingly, this study also identified a role of ATL3 in flavivirus maturation and furin recycling. Knockdown of ATL3 increased levels of extracellular un-cleaved prM and altered furin localization away from the Golgi. The relocalization of furin was specifically observed after knockdown of ATL2 and ATL3, whereas knockdown of other ER remodeling proteins RTN3 and LNP had no effect (Neufeldt et al., 2019).

Autophagy

Autophagy is an essential intracellular degradative process that recycles cytoplasmic components (Bento et al., 2016). Autophagy involves three major steps, the formation of autophagosomes and simultaneous capture of cytoplasmic material, the fusion of autophagosomes with lysosomes to form autolysosomes, and the turnover of autolysosomes. The cytoplasmic components often referred to as cargo, can either be selectively or non-selectively degraded (Gatica et al., 2018). For selective autophagy, cargo such as mitochondria is tagged by cargo receptors which are then encapsulated and degraded by autophagy. Autophagy is involved in the replication of various flaviviruses. The overall role of autophagy as proviral or antiviral in flavivirus replication is complex and has no clear consensus (Choi et al., 2018; Ke, 2018; Echavarria-Consuegra et al., 2019). Here, we discuss the studies that have implicated the role of autophagy-related proteins in virus replication through physical interactions with viral proteins.

DENV and ZIKV hijack various aspects of selective autophagy for efficient virus replication. Regulation of lipid metabolism during DENV infection has been reported by multiple groups (Heaton and Randall, 2010; Perera et al., 2012; Jordan and Randall, 2017; Chotiwan et al., 2018; Koh et al., 2020). DENV NS4A physically interacts with unubiquitinated AUP1 to translocate lipid droplets to autophagosomes to induce lipophagy (Zhang et al., 2018). Interestingly, DENV NS4B or DENV infection is essential for this interaction. Ubiquitination of AUP1 impeded its interaction with NS4A, which led to defective lipophagy and reduced viral titers. This study highlights the importance of lipophagy during virus infection that is regulated by virus-host protein interactions. Regulation of apoptosis through autophagy is another strategy utilized by DENV for prolonged virus replication (McLean et al., 2011), and DENV NS1 interacts with Beclin-1 to activate autophagy and prevent apoptosis at early stages of infection (Lu et al., 2020). FAM134B, an ER phagy (reticulophagy) selective cargo receptor was identified to interact with DENV and ZIKV NS2B3 (Lennemann and Coyne, 2017). The researchers demonstrated that DENV and ZIKV NS2B3 cleave FAM134B to inhibit the degradation of viral proteins through reticulophagy. Additionally, overexpression of FAM134B leads to decreased virus replication. These results indicate selective degradation of ER is subverted by viruses even though overall autophagy could be upregulated during infection. In a recent study, Ponia and colleagues observed inhibition of mitophagy through the interaction of ZIKV NS5 with the host protein Ajuba (Ponia et al., 2021). Ajuba is a key regulator of mitophagy and is translocated to depolarized mitochondria to initiate PINK1-Parkin mediated mitophagy. NS5 interaction with Ajuba impeded its translocation to depolarize mitochondria, thus inhibiting mitophagy. The authors further use *in vivo* ZIKV infection studies in mice to demonstrate increased early pro-inflammatory chemokines and viral load in tissue due to inhibition of mitophagy, further underlining the importance of the NS5-Ajuba interaction. These studies point towards the regulation of selective autophagy by DENV and ZIKV. Systematic measurements of degraded cargo during virus infection can provide key insights. In the future, it will also be valuable to explore if modulation of selective autophagy is a common theme for other flaviviruses and other types of selective autophagy (e.g., pexophagy and xenophagy).

Interactions of general autophagy-related proteins with WNV and JEV proteins have also been identified. WNV Capsid protein interacts with AMPK, an autophagy inducer (Kobayashi et al., 2020). This interaction mediated the degradation of

AMPK through the proteasome pathway and led to the accumulation of ubiquitinated protein aggregates. A mutant Capsid protein reduced the interaction with AMPK and its degradation. Even though disrupting this interaction did not affect virus replication, it led to lower protein aggregates in mouse brain and reduced neurological symptoms. For JEV, Sharma et al. (2021) have shown that autophagy acts as an antiviral response during JEV infection in neuronal cells. They also observed NS1 colocalization with LC3-I, an important autophagy protein whose depletion caused decreased viral titers. In a more recent study, the same group also demonstrated Capsid protein interaction with LC3-I, using immunoprecipitation (Sarkar et al., 2021). The functional role of these interactions in virus replication and autophagy is uncharacterized and could be a potential study.

Future efforts can be focused on investigating known uncharacterized physical interactions. The mTOR pathway is an important autophagy pathway that is differentially regulated during flavivirus infection (Shives et al., 2014; Liang et al., 2016; Jordan and Randall, 2017; Sahoo et al., 2020; Lahon et al., 2021). Moreover, viral protein interactions with mTOR were also found using proteomic approaches (Shah et al., 2018). However, characterizing the role of virus-mTOR PPIs in the context of virus infection is largely unexplored and could be a potential future direction. Selective autophagy during virus infection is another interesting attribute for potential study. Viruses appear to exploit the selective nature of autophagy by variably regulating specific cargo degradation. For example, DENV upregulates lipophagy while downregulating reticulophagy for effective replication (Lennemann and Coyne, 2017; Zhang et al., 2018). Thus, further investigating the interactions found between the viral proteins and cargo receptors could be a promising direction. We explored the literature to generate a list of virus-autophagy PPIs that have been identified in various proteomic screens, which could further explain the role of autophagy during virus infection (**Table 1**). Interestingly, while many autophagy proteins were identified in these screens, none were pursued mechanistically in those published studies, leaving the door open to many systematic studies of these PPIs. Finally, capturing the temporal change in cargo degraded during virus infection may also provide novel insights into the dynamic replication cycle.

TABLE 1 | Protein–protein interaction (PPI) found between autophagy proteins and viral proteins from seven data sets (Coyaud et al., 2018; Scaturro et al., 2018; Shah et al., 2018; Li et al., 2019; Golubeva et al., 2020; Zeng et al., 2020).

Autophagy protein	Autophagy related role	Viral proteins
ACBD5	Pexophagy receptor	NS4A[1]
AMBRA1	Key regulator of autophagy by modulating the BECN1-PIK3C3 complex	NS1[7], NS2B[7]
ATG9A	Supplies membrane for the growing autophagosome	Env[7]
BNIP3 (NIP3)	Mitophagy receptor	NS5[1,2]
EI24 (EPG4)	Regulates formation of degradative autolysosomes	NS1[2], NS4B[3]
LGALS8	Restricts infection by initiating autophagy *via* interaction with CALCOCO2/NDP52	NS3[6]
MTOR	Key regulator of autophagy through phosphorylation of ULK1, DAP, AMBRA1, and RUBCNL	NS4A[1,2]
PHB2	Mitophagy receptor	NS2B3[3], NS4B[3]
SQSTM1 (p62)	Multiple cargo receptor	NS4B[2]
STX17	Regulates autophagosome fusion with lysosomes	NS2A[7]
VCP	Essential for the maturation of ubiquitin-containing autophagosomes and the clearance of ubiquitinated protein by autophagy	NS2B3[5]
WAC	Regulator of autophagy	NS2B[6]
AUP1	Lipophagy regulator	NS2A[4], NS4B[3,4]
FAM134C	Reticulophagy receptor	NS4A[1], NS4B[3]
RTN3	Reticulophagy receptor	NS4A[1]
SEC62	Reticulophagy receptor	NS4A[2]
CALCOCO1	Reticulophagy receptor	NS5[5]
NBR1	Aggrephagy, pexophagy, and xenophagy receptor	NS2A[4]
VMP1	Required for autophagosome biogenesis	NS4A[4]
TMEM41B	Required for autophagosome biogenesis	NS4B[3]

[1]*Shah et al. (2018) (ZIKV).*
[2]*Shah et al. (2018) (DENV).*
[3]*Scaturro et al. (2018) (ZIKV),*
[4]*Coyaud et al. (2018) (ZIKV).*
[5]*Li et al. (2019) (WNV).*
[6]*Golubeva et al. (2020) (ZIKV).*
[7]*Zeng et al. (2020) (ZIKV).*
PPI that were found significant by the authors were considered for the search. Approximately, 100 autophagy proteins were probed for interactions based on a list of proteins mentioned in these studies (Galluzzi et al., 2017; Gatica et al., 2018; Gubas and Dikic, 2021).

Mitochondrial Dynamics and Morphology

Mitochondria are dynamic organelles with widespread functions in cellular homeostasis, including ATP production, immune response signaling, and apoptosis activation. Unsurprisingly, many viruses interact with and perturb these functions to benefit their own replication (Anand and Tikoo, 2013). Recently the mechanisms and protein interactions that flaviviruses use to modulate these mitochondrial functions have revealed dynamic alterations in mitochondrial morphology that impact virus replication.

The morphology of host mitochondria is constantly changing. The constant fusion and fission of mitochondria is critical for cellular homeostasis. The fusion of mitochondria together is mediated by mitofusin 1 (MFN1) and MFN2 in the outer mitochondrial membrane and optic atrophy protein 1 (OPA1) in the inner membrane. Fission is mediated by dynamin-related protein 1 (DRP1), which is soluble and recruited to mitochondria by mitochondrial fission protein 1 (FIS1; Westermann, 2010). ZIKV and DENV both impact mitochondrial morphology, albeit in cell-type and virus specific manners (Chatel-Chaix et al., 2016; Barbier et al., 2017; García et al., 2020; Yang et al., 2020). DENV and ZIKV infection in Huh7 hepatocytes induces dramatic mitochondrial elongation. This is associated with significant decreases in DRP1 fission activity, specifically

through decreased phosphorylation at S616, a site which induces fission by DRP1. NS4B interacts with many mitochondrial proteins (Scaturro et al., 2018; Shah et al., 2018) and its expression alone is sufficient to alter mitochondrial morphology (Chatel-Chaix et al., 2016). NS4B expression is linked with decreased expression of CDK1, the kinase which phosphorylates DRP1 at S616. Knockdown of DRP1 further increases DENV and ZIKV replication, while increasing fusion through knockdown of MFN2 decreases replication (Chatel-Chaix et al., 2016; Barbier et al., 2017). Interestingly, knockdown of DRP1 did not impact the replication of fellow flavivirus WNV or the closely related HCV (Chatel-Chaix et al., 2016). Together, this suggests that DENV and ZIKV specifically induce the elongation of mitochondria in these cells. It appears this elongation may serve two functions for virus replication. Firstly, elongated mitochondria have increased respiratory function, resulting in greater energy production which may be utilized directly for virus replication or promote host cell survival (Barbier et al., 2017). Secondly, this elongation impedes mitochondrial innate immune response signaling by preventing the translocation of RIG-I to mitochondrial-associated membranes and decreasing mitochondrial antiviral signaling protein (MAVS)-associated interferon (IFN) production (Yu et al., 2015; Chatel-Chaix et al., 2016). The interplay between flaviviruses and innate immune signaling, including through MAVS and RIG-I, will be discussed in more detail later in this review.

Intriguingly, mechanisms that promote mitochondrial fission, rather than fusion, have been observed during DENV and ZIKV infection in other cell types (Yu et al., 2015; Yang et al., 2020). In A549 cells, DENV infection also leads to abnormal mitochondrial dynamics, however, independent of DRP1. Rather, MFN1 and MFN2 are cleaved by the DENV NS2B3 protease, resulting in decreased fusion and more mitochondrial fragmentation. Specifically, cleavage of MFN1 results in decreased MAVS-mediated IFN production, while cleavage of MFN2 decreased the activation of cell-death associated caspases. Again, this activity does not appear to be conserved across all flaviviruses, as NS2B3 from JEV was unable to perform the same cleavage events (Yu et al., 2015). This specificity may contribute to the unique pathogenesis of some flaviviruses. Congenital ZIKV infection is associated with the development of neurological and ocular abnormalities, which are not observed with other flaviviruses (Roach and Alcendor, 2017; Carod-Artal, 2018). It is possible that perturbation of mitochondrial processes by viruses are especially potent in these tissues, as metabolic demands are high and these tissues are very sensitive to mitochondrial dysfunction (Pacheu-Grau et al., 2018; Norat et al., 2020) In ZIKV-infected neural stem cells (NSCs) mitochondria numbers and size are significantly decreased, associated with concomitant decreases in MFN2 protein expression, whereas the other fusion/fission proteins (MFN1, OPA1, DRP1, and FIS1) were unchanged (Yang et al., 2020). ZIKV had similar effects on the mitochondria of retinal pigment epithelial (RPE) cells, with mitochondria appearing more fragmented and punctate in nature (García et al., 2020). In both cases, ZIKV-associated morphology changes involved the loss of mitochondrial membrane potential, resulting in diminished ATP production and mitochondrial function. Whether ZIKV NS2B3 performs similar cleavage of MFN2 as the DENV protease or if ZIKV relies on other unique interactions requires further experimentation.

ANTAGONISM OF HOST IMMUNITY BY FLAVIVIRUS-HOST PPIS

While some interactions between viral proteins and host proteins associated with the immune system restrict flavivirus replication and pathogenesis (Sun et al., 2013a; Douradinha et al., 2014; Wessel et al., 2021), flaviviruses have evolved numerous mechanisms to sabotage the host innate immune response *via* interactions with host proteins. Here, we review the major mechanisms of antagonism associated with IFN production, IFN signaling, and the complement system (**Figure 4**).

IFN Production

Pattern-recognition receptors (PRRs) sense flaviviruses upon entry. The major PRRs relevant for flaviviruses are TLR3 and TLR7/8, which are located primarily in endosomal vesicles and recognize viral RNA of incoming virions (Heil et al., 2004; Tsai et al., 2009); retinoic-acid inducible gene-I (RIG-I), and myeloma differentiation factor 5 (MDA5), which recognize cytosolic RNA (Yoneyama et al., 2005); and cyclic GMP-AMP synthase (cGAS) which recognize cytosolic DNA (Sun et al., 2013b). Mechanistically, activated TLR3 and TLR7 recruited adaptor protein MyD88 and TRIF to initiate further changes regulating the expression of cytokines, chemokines, and type I IFNs (Kawasaki and Kawai, 2014). After sensing viral RNA, RIG-I and MDA5 move from the cytosol to mitochondria and interact with their adaptor, mitochondrial antiviral signaling protein (MAVS), to continue further downstream signaling that activates IRF3 and NF-kβ (Chiang et al., 2014). cGAS activation after recognizing cytosolic DNA catalyzes the synthesis of cyclic GAMP (cGAMP) which activates STING, which subsequently activates IFN expression (Cai et al., 2014). The signaling induced by these sensors converges on a common cascade that induces the production of IFN and downstream genes stimulated by IFN called interferon-stimulated genes (ISGs; Nasirudeen et al., 2011).

Flaviviruses have evolved different strategies to interfere with the host production of IFNs. A major mechanism for flaviviruses is to disrupt double-stranded RNA-sensing pathway. NS3 from DENV and ZIKV binds 14-3-3ε, an important protein in antiviral immunity (Liu et al., 2012), *via* a conserved phosphomimetic motif on NS3 and prevent the translocation of RIG-I to mitochondria, and consequently IFN production (Chan and Gack, 2016; Riedl et al., 2019). During DENV infection, DENV NS4A also physically interacts with MAVS to prevent RIG-I from forming complexes with MAVS in mitochondria-associated endoplasmic reticulum membranes (MAMs), leading to the disruption of RIG-I-induced IRF3 activation and subsequently suppression of IFN production (He et al., 2016).

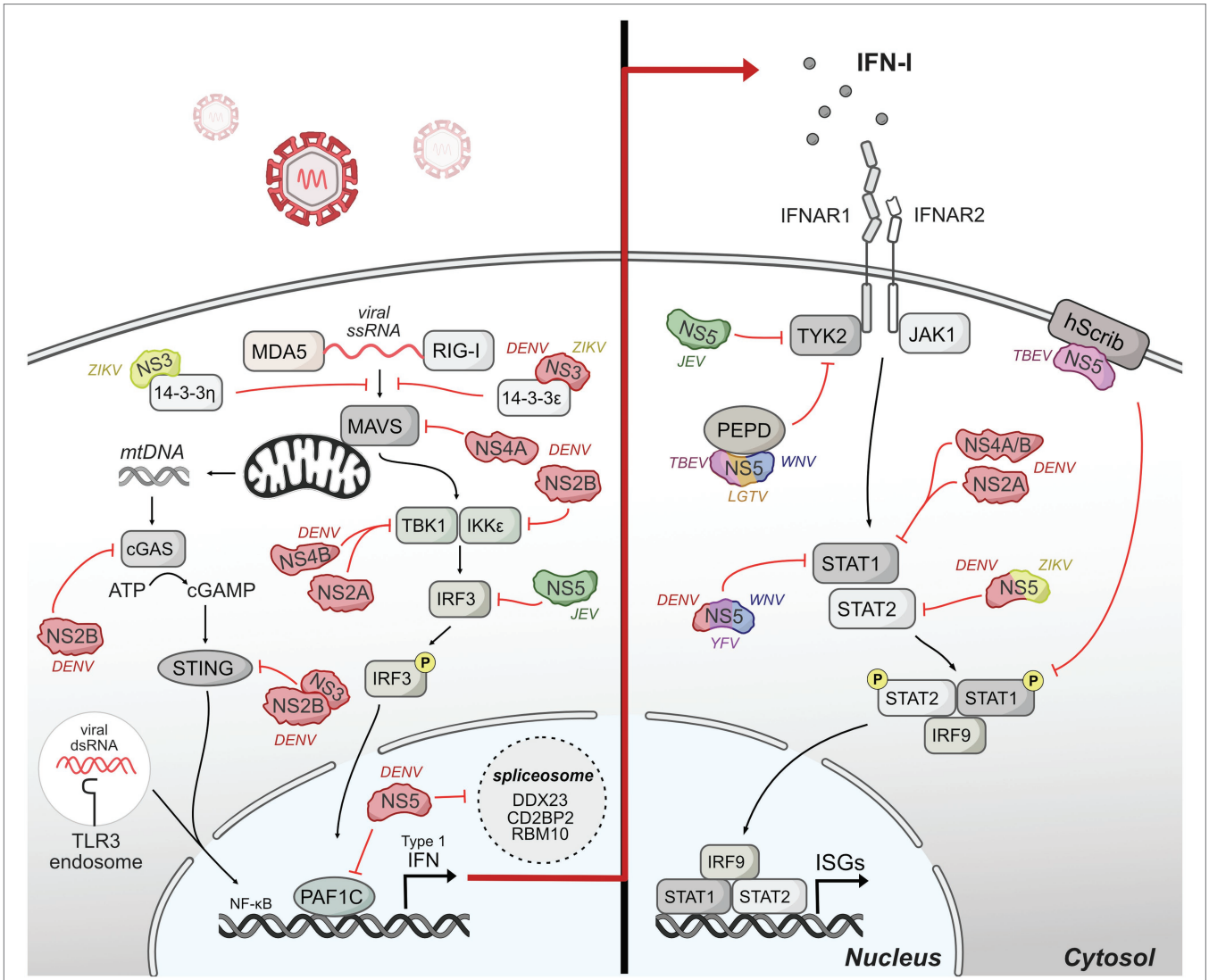


FIGURE 4 | Host innate immune response is antagonized by flavivirus protein interactions. Upon entry, flaviviruses are sensed by different pattern-recognition receptors (PRRs) such as TLR-3 and RIG-I. The signaling induced by these sensors converges on a common cascade that induces the production of interferon (IFN) and downstream genes stimulated by IFN called interferon-stimulated genes (ISGs). Flaviviruses have evolved invasive strategies to interfere with host immune response by antagonizing different protein components of innate immune signaling pathways associated with IFN production and IFN signaling. The DENV NS2B3 protease is shown as NS2B and NS3 interacting together to antagonize STING.

Further down in the RIG-I-induced type I IFN pathway, DENV serotypes 1, 2, and 4 (DENV1, DENV2, and DENV4) NS2A and NS4B proteins inhibit RIG-I/MDA5-regulated interferon beta (IFN-β) induction by blocking TBK1/IRF3 (Dalrymple et al., 2015). DENV NS2B3 interacts with IKKe to prevent IRF3 phosphorylation (Angleró-Rodríguez et al., 2014) potentially changing the activation of multiple antiviral genes including type I IFN. JEV inhibits IFN-β production by suppressing IFR3 and NF-kβ (Ye et al., 2017). Mechanistically, JEV NS5 interacts with nuclear transport proteins KPN2, KPN3, and KPN4 and blocks their interaction with IRF3 and P65, therefore preventing nuclear translocation of IRF3 and NF-kβ and reducing type I IFN production.

Even as RNA viruses, flaviviruses also antagonize IFN production by interfering with cytosolic DNA-sensing pathway by cGAS and its adaptor STING. Specifically, as DENV infection triggers innate immune response through mtDNA sensing by the DNA sensor cGAS (Sun et al., 2017), DENV NS2B targets cGAS for degradation to prevent the detection of mitochondrial DNA released during DENV infection, blocking the activation of cGAS/STING pathway and the induction of type I IFN (Aguirre et al., 2017). DENV NS2B3 also physically interacts with and cleaves STING to inhibit type I IFN production in species-specific manner (Aguirre et al., 2012; Yu et al., 2012). During ZIKV infection, cGAS is targeted and cleaved by NS1-stablized caspase-1, leading to enhanced NLRP3 inflammation activation and reduced type I induction to benefit the infection (Zheng et al., 2018). The multiple mechanisms by which flaviviruses antagonize DNA sensing suggest that this is an important

mechanism of immune evasion for RNA viruses and represents a new frontier of investigation for virus-host interactions.

Flavivirus NS5 proteins also antagonize innate immunity upstream of IFN production and signaling, primarily through interactions with host gene expression machinery in the nucleus. The extent of NS5 localization to the nucleus varies depending on flavivirus species, yet some nuclear localization and nuclear/cytoplasmic shuttling appears to occur for nearly all NS5s, with the exception of duck Tembusu virus (Hannemann et al., 2013; Zhao et al., 2015; Grant et al., 2016; Kumar et al., 2016; Tay et al., 2016; Duan et al., 2019; Cheng et al., 2021; Petit et al., 2021). NS5 protein interactions with host gene expression machinery have been noted through several unbiased screens, including mass spectrometry and yeast-two-hybrid screens (Khadka et al., 2011; Le Breton et al., 2011; De Maio et al., 2016; Coyaud et al., 2018; Scaturro et al., 2018; Shah et al., 2018; Li et al., 2019). Several of these protein interactions have been linked back to NS5 perturbation of host gene expression. For example, a proteomic study of DENV NS5 during infection revealed interactions with CD2BP2 and DDX23, core components of U5 small nuclear ribonucleoprotein particles (U5 snRNPs) that ultimately interfere splicing efficiency (De Maio et al., 2016). DENV NS5 also dysregulates host splicing by physically interacting with RBM10, a splicing factor that regulates spermidine/spermine-N1-acetyltransferase (SAT1) splicing and promoting RBM10 proteasomal degradation. The interaction potentially restricts RBM10 from its proinflammatory function and benefits DENV replication (Pozzi et al., 2020). In our studies on DENV-human protein interactions, we identified an interaction with PAF1C (Shah et al., 2018), which regulates the transcription elongation of many immune response genes (Marazzi et al., 2012; Parnas et al., 2015). Our recent work dissecting the NS5-PAF1C interaction demonstrated that PAF1C regulates immune response genes upstream of type I IFN production, including the RIG-I/DDX58 signaling axis. Breaking the NS5-PAF1C interaction through mutagenesis of NS5 rescued PAF1-dependent gene expression, underlining the importance of this protein interaction (Petit et al., 2021).

The recurring theme of NS5 nuclear localization and interactions with nuclear proteins has led to much speculation in the field regarding why the polymerase and methyltransferase of a cytoplasmic RNA virus would have such behavior. There is mounting evidence that flavivirus NS5 protein can perturb host gene expression, both through dissection of virus-host protein interactions, and through more generalized gene expression studies. For example, independent studies of ZIKV, WNV, and DENV NS5 all point to overall inhibition of immune gene expression (López-Denman et al., 2021; Petit et al., 2021; Zhao et al., 2021). On the other hand, studies involving infection have not revealed *in vitro* or *in vivo* phenotypes for DENV mutants that reduce NS5 nuclear localization (Kumar et al., 2016; Tay et al., 2016; Cheng et al., 2021), resulting in skepticism regarding the biological significance of NS5 nuclear localization. However, for studies involving DENV serotype 2, it should be noted that there are two distinct nuclear localization signals (NLSs) that contribute to nuclear localization. In fact, mutation of a single NLS still results in substantial NS5 nuclear localization (~1:1 nuclear:cytoplasmic ratio). In our own studies, we show that only mutation of both NLSs truly excludes NS5 from the nucleus and disrupts the NS5-PAF1C interaction (Petit et al., 2021). Thus, a modest amount of NS5 nuclear localization may be sufficient for its role in perturbing host gene expression. Using a double NLS mutant to study protein interactions and virus replication phenotypes will be essential to understanding the true function of nuclear NS5. Creating similar NLS mutants for other flavivirus NS5 will also strengthen the evidence supporting a role for nuclear NS5 in general.

IFN Signaling

Although flaviviruses actively exploit various strategies to suppress the production of IFN by infected cells, secreted IFN can still bind to the heterodimeric IFN receptor, IFNR1 and IFNR2 that are present on most cells. Binding of IFN to the receptors triggers the activation of JAK1 and Tyk2 to phosphorylate cytoplasmic STAT1 and STAT2 (Darnell et al., 1994). The phosphorylated STAT1 and STAT2 form a heterotrimeric complex with interferon regulatory factor 9 (IFR9) known as IFN-stimulated gene factor 3 (ISGF3), which translocate to the nucleus and binds to interferon-stimulated response element (ISRE) to regulate the transcription of IFN-stimulated genes (ISGs), many of which are antiviral (Haller et al., 2006).

Multiple studies have shown that flaviviruses use different ways to manipulate IFN signaling. DENV NS2A, NS4A, and NS4B block the IFN-induced transduction cascade in human A549 cells by interfering STAT1 phosphorylation, resulting decreased IFN-induced ISRE-promoter activation and enhanced DENV2 virus replication (Muñoz-Jordán et al., 2003, 2005). NS5 also inhibits IFN signaling *via* multiple mechanisms that appear to be virus-specific (Best et al., 2005; Guo et al., 2005; Lin et al., 2006; Park et al., 2007; Laurent-Rolle et al., 2010; Grant et al., 2016), and we will highlight mechanisms for which the role of NS5-host protein interactions has been dissected. NS5 of Langat virus (LGTV), a member of tick-born encephalitis complex of viruses, also interacts with IFN-a/b receptor subunit (IFNAR2) and IFN-g receptor subunit (IFNGR1) to block Jak1 and Tyk2 phosphorylation (Best et al., 2005; Park et al., 2007). TBEV NS5 protein also interacts with hScrib, a protein is expressed at the membrane of mammalian cells and controls cell-to-cell contact, resulting in impaired pSTAT1 formation in response to IFN-a/b and IFN-g (Werme et al., 2008). TBEV and WNV NS5 can also inhibit IFNAR1 mutation and accumulation at the cell surface through an interaction with PEPD (Lubick et al., 2015). DENV and ZIKV NS5 interact with and target STAT2 for proteasome-mediated degradation to inhibit IFN-induced signaling (Ashour et al., 2009; Mazzon et al., 2009; Morrison et al., 2013; Grant et al., 2016). Interestingly, DENV and ZIKV NS5 can target and degrade human STAT2 (hSTAT2) but not mouse STAT2 (Ashour et al., 2010; Grant et al., 2016). Thus, STAT2 is a species-specific target of a flaviviral nonstructural protein, similar to STING. Given the many emerging flaviviruses that circulate in non-human reservoirs, exploring the biophysical and biochemical differences

underlying species-specific virus restriction could help predict the constraints to emergence.

Complement System

The complement system is an important part of innate immunity to control early infection. It contains more than 50 plasma proteins and membrane proteins expressed on cell surface (Merle et al., 2015a). Complement activities occur in plasma, in tissues, or within cells (Kolev et al., 2014). The activation of complement happens through three distinct target-dependent pathways: classical, lectin, and alternative pathways. The classical pathway is initiated by the direct binding of C1q to the pathogen surface or antigen–antibody complexes; the lectin pathway is activated when mannose binding lectin (MBL), a serum protein, binds to mannose-containing carbohydrates on pathogens; and the alternative pathway is active when a spontaneously activated complement component binds to the pathogen surface (Merle et al., 2015a). Each pathway has its own protease to target and process different antigens, but they all generate a protease called C3 convertase and share common terminal outcomes after C3 cleavage: pathogen opsonization, regulation of inflammation, and clearance of immune complexes and cell debris. Complement activation is a bridge linking innate immune response and adaptive response by B cells and T cells in viral infection (Mehlhop and Diamond, 2006; Merle et al., 2015b).

Flaviviruses have evolved mechanisms to antagonize this part of the innate immune system. Among NS proteins from flaviviruses, NS1 has been recognized as an immune invasion protein that interferes with the complement system. A study by Avirutnan showed that NS1 from DENV, WNV, and YFV reduced complement activation pathways by interacting and forming a complex with C4 and C1s, leading to reduced classical pathway C4b deposition and C3 convertase (C4b2a) activity and consequent protection of DENV from complement-regulated neutralization (Avirutnan et al., 2010). In another study, NS1 from these flaviviruses directly binds to C4b binding protein (C4BP), a regulatory plasma protein of the classical and lectin pathway, to inactivate C4b in both cell surface and fluid; thereby protecting the viruses from complement attack (Avirutnan et al., 2011). WNV NS1 also binds to factor H (fH), a key regulator of the alternative pathway, and facilitates factor I-mediated cleavage of C3b. Additionally, cell surface-associated NS1 recruits fH and reduces C3b deposition and C5b–9 membrane attack complexes on cell surfaces, reducing the recognition of infected cells by complement system (Chung et al., 2006). For the lectin pathway, insect-derived DENV NS1 not only binds to human C1s, C4, and C4b-binding protein to suppress classical pathway of complement activation but also binds to mannose binding lection (MBL) to disrupt neutralization by the lectin pathway (Thiemmeca et al., 2016). DENV NS1 was also reported to interact with other proteins and interfere with the terminal pathway of complement activation. Specifically, NS1 interacts with complement regulator vitronectin (VN) and inhibits membrane attack complex (MAC) formation, suggesting a role of NS1 in antagonizing complement activation (Conde et al., 2016).

FLAVIVIRUS-HOST PPIS INVOLVED IN DISEASE

While identifying virus-host PPIs is important to inform on the fundamental mechanisms driving their replication, they can also be critical to understanding pathogenesis. Indeed, in recent years, several individual interactions have sparked interest in how flaviviruses alter host cellular behavior to cause disease. ZIKV has deservedly received significant research attention due to its unique association CZS. The most notable presentation of CZS is microcephaly, a condition in which head and brain size are dramatically reduced at birth (Moore et al., 2017). Many studies have provided insight into the mechanisms by which ZIKV causes CZS. Notably, however, is that many of these studies focus on virus strain/variants, placental damage, and the innate immune response *in utero* (Miner et al., 2016; Jagger et al., 2017; Rosenfeld et al., 2017; Yuan et al., 2017; Platt et al., 2018; Yockey et al., 2018). Here, we review how flavivirus-host PPIs directly dysregulate important developmental pathways to cause CZS.

In our own global proteomics screen, we identified an interaction between ZIKV NS4A and host ANKLE2 (Shah et al., 2018), mutations in which are known to be associated with hereditary microcephaly in humans and small-brain phenotypes in *Drosophila melanogaster* (flies; Yamamoto et al., 2014). NS4A expression alone in flies is sufficient to induce similar brain size defects in an ANKLE2-dependent manner. Further investigation revealed ANKLE2 is critical for spindle pole alignment during asymmetric division of fly neuroblasts, akin to mammalian neuroprogenitor cells that are targeted by ZIKV, and expression of NS4A results in similar division defects. Elegant fly genetics were used to demonstrate that NS4A inhibits the ANKLE2 pathway specifically (Link et al., 2019). Together, this demonstrates NS4A interacts with and disrupts ANKLE2 function, which in susceptible neuroblasts can disturb brain development. The extent to which this specific interaction impacts vertebrate brain development requires further investigation. Additionally, these studies bring to light the interplay between host genetics and viral pathogenesis. In flies, Ankle2 mutation heterozygosity results in normal brain development. However, NS4A expression in these flies is dramatically more severe than in wild-type flies (Shah et al., 2018; Link et al., 2019), suggesting that host genetics can pre-dispose an organism to disease that may be associated with virus-host PPIs. Another example of this phenomenon involves the previously discussed host-factor TMEM41B. Naturally occurring single nucleotide polymorphisms that lead to Ile266Val/Leu substitutions are prevalent in certain human populations but fail to rescue flavivirus replication in TMEM41B KO cells. This suggests these variants cannot be utilized by flaviviruses for the function they require to effectively replicate (Hoffmann et al., 2021).

One of the most common clinical findings associated with CZS is intracranial calcifications (Pool et al., 2019). A recent study explored how ZIKV induces these calcifications through the specific interaction between the viral protease NS3 and host bone morphogenic protein 2 (BMP2). BMP2 is an

essential signaling protein in the process of osteogenesis, inducing the expression of downstream genes that ultimately facilitate bone growth. BMP2 normally must be cleaved by furin-type proteases prior to secretion, where it then induces these signaling cascades. Infection with ZIKV leads to increased expression of BMP2 and downstream genes, and subsequent calcification *in vitro* and *in vivo*. In fact, the expression of NS3 alone is sufficient to induce these phenotypes in U2OS osteosarcoma epithelial cells, but not a protease-defective mutant, suggesting that ZIKV NS3 cleavage of BMP2 initiates osteogenesis in the brain, leading to intracranial calcifications (Chen et al., 2021).

Beyond these examples, several other ZIKV-host interactions have been found to impact brain development and may ultimately play a role in human pathogenesis. Even while the ZIKV epidemic was ongoing it was shown that expression of ZIKV NS4A and NS4B specifically impaired the growth of NSCs by perturbing autophagy, while corresponding DENV proteins did not (Liang et al., 2016). This is not the only example of ZIKV-specific effects. Expression of ZIKV NS2A *in vivo* disrupts neurogenesis through physical interactions with adherens junctions in radial glial cells (Yoon et al., 2017). In another recent study, systematic proteomics in NSCs revealed the interaction between ZIKV Capsid and Dicer, a pivotal protein in the host RNAi pathway with implications in neurodevelopment (Zeng et al., 2020). Among flaviviruses, this interaction is also unique to ZIKV. Mechanistically, Dicer is a host restriction factor and ZIKV Capsid interaction inhibits this antiviral function, as infection with a H41R mutation, which ablates this interaction leads to less viral burden *in vivo*. Indeed, even ZIKV Capsid expression alone, dependent on its interaction with Dicer, is sufficient to induce severe defects in brain development.

Together, these studies highlight how certain aspects of pathogenesis may be uniquely derived from single virus-host PPIs. However, given the incredible complexity of human development, it is not likely that any single interaction during infection is solely responsible for disease outcome. More realistically in the case of ZIKV, it is the culmination of these perturbations and dysregulation of brain development by multiple mechanisms that results in CZS. Intriguingly, the intersection of flavivirus-host PPIs and disease, including host factors implicated in hereditary disease, opens the door to the possibility of host genetics being a major and overlooked contributing factor to susceptibility to CZS. For example, loss-of-function variants for host factors like ANKLE2, which are haplo-sufficient for their role in development in the absence of a virus-host PPI, but haplo-insufficient in the context of a virus-host PPI, could tip the balance in the favor of disease in an otherwise healthy individual. Future studies exploring this concept are warranted.

CONCLUDING REMARKS

Flaviviruses are arthropod-borne viruses that cause significant human disease worldwide. Their limited genome requires them to co-opt host proteins through physical interactions during infection to properly replicate. Some of these interactions appear to be broadly conserved among flaviviruses, while other unique interactions contribute to observed differences in host tropism and pathogenesis. Flaviviruses employ a wide range of host receptors utilized for entry into host cells. Replication within the ER involves vast remodeling into a microenvironment well-suited to the generation of viral progeny. Even outside the ER viral proteins orchestrate modulation of host cell systems. This includes physical interaction with other cellular pathways and organelles critical to virus replication and with different protein components of the host innate immune system. These virus-host PPIs can be influential in the development of pathogenesis. Thus, understanding these mechanisms is essential for creating new therapeutics to alleviate human disease caused by flaviviruses.

AUTHOR CONTRIBUTIONS

AF, OP, NB, MK, and PS wrote manuscript. MK, AF, OP, and NB designed and generated the figures. PS secured funding for work. All authors contributed to the article and approved the submitted version.

FUNDING

Funding was provided by the W. M. Keck Foundation to PS and a NIH T32 Fellowship (2T32AI060555-16) to AF.

REFERENCES

Acosta, E. G., Castilla, V., and Damonte, E. B. (2008). Functional entry of dengue virus into Aedes albopictus mosquito cells is dependent on clathrin-mediated endocytosis. *J. Gen. Virol.* 89, 474–484. doi: 10.1099/vir.0.83357-0

Adams, D. R., Ron, D., and Kiely, P. A. (2011). RACK1, A multifaceted scaffolding protein: structure and function. *Cell Commun. Signal* 9:22. doi: 10.1186/1478-811X-9-22

Aguirre, S., Luthra, P., Sanchez-Aparicio, M. T., Maestre, A. M., Patel, J., Lamothe, F., et al. (2017). Dengue virus NS2B protein targets cGAS for degradation and prevents mitochondrial DNA sensing during infection. *Nat. Microbiol.* 2:17037. doi: 10.1038/nmicrobiol.2017.37

Aguirre, S., Maestre, A. M., Pagni, S., Patel, J. R., Savage, T., Gutman, D., et al. (2012). DENV inhibits type I IFN production in infected cells by cleaving human STING. *PLoS Pathog.* 8:e1002934. doi: 10.1371/journal.ppat.1002934

Aktepe, T. E., Liebscher, S., Prier, J. E., Simmons, C. P., and Mackenzie, J. M. (2017). The host protein reticulon 3.1A is utilized by Flaviviruses to facilitate membrane remodelling. *Cell Rep.* 21, 1639–1654. doi: 10.1016/j.celrep.2017.10.055

Allison, S. L., Schalich, J., Stiasny, K., Mandl, C. W., Kunz, C., and Heinz, F. X. (1995). Oligomeric rearrangement of tick-borne encephalitis virus envelope proteins induced by an acidic pH. *J. Virol.* 69, 695–700. doi: 10.1128/jvi.69.2.695-700.1995

Alvarez, D. E., De Lella Ezcurra, A. L., Fucito, S., and Gamarnik, A. V. (2005). Role of RNA structures present at the 3'UTR of dengue virus on translation, RNA synthesis, and viral replication. *Virology* 339, 200–212. doi: 10.1016/j.virol.2005.06.009

Anand, S. K., and Tikoo, S. K. (2013). Viruses as modulators of mitochondrial functions. *Adv. Virol.* 2013:738794. doi: 10.1155/2013/738794

Angleró-Rodríguez, Y. I., Pantoja, P., and Sariol, C. A. (2014). Dengue virus subverts the interferon induction pathway via NS2B/3 protease-IκB kinase ε interaction. *Clin. Vaccine Immunol.* 21, 29–38. doi: 10.1128/CVI.00500-13

Ashour, J., Laurent-Rolle, M., Shi, P.-Y., and García-Sastre, A. (2009). NS5 of dengue virus mediates STAT2 binding and degradation. *J. Virol.* 83, 5408–5418. doi: 10.1128/JVI.02188-08

Ashour, J., Morrison, J., Laurent-Rolle, M., Belicha-Villanueva, A., Plumlee, C. R., Bernal-Rubio, D., et al. (2010). Mouse STAT2 restricts early dengue virus replication. *Cell Host Microbe* 8, 410–421. doi: 10.1016/j.chom.2010.10.007

Avirutnan, P., Fuchs, A., Hauhart, R. E., Somnuke, P., Youn, S., Diamond, M. S., et al. (2010). Antagonism of the complement component C4 by flavivirus nonstructural protein NS1. *J. Exp. Med.* 207, 793–806. doi: 10.1084/jem.20092545

Avirutnan, P., Hauhart, R. E., Somnuke, P., Blom, A. M., Diamond, M. S., and Atkinson, J. P. (2011). Binding of Flavivirus nonstructural protein NS1 to C4b binding protein modulates complement activation. *J. Immunol.* 187, 424–433. doi: 10.4049/jimmunol.1100750

Barajas, D., Xu, K., Sharma, M., Wu, C.-Y., and Nagy, P. D. (2014). Tombusviruses upregulate phospholipid biosynthesis via interaction between p33 replication protein and yeast lipid sensor proteins during virus replication in yeast. *Virology* 471–473, 72–80. doi: 10.1016/j.virol.2014.10.005

Barbier, V., Lang, D., Valois, S., Rothman, A. L., and Medin, C. L. (2017). Dengue virus induces mitochondrial elongation through impairment of Drp1-triggered mitochondrial fission. *Virology* 500, 149–160. doi: 10.1016/j.virol.2016.10.022

Barrows, N. J., Anglero-Rodriguez, Y., Kim, B., Jamison, S. F., Le Sommer, C., McGee, C. E., et al. (2019). Dual roles for the ER membrane protein complex in flavivirus infection: viral entry and protein biogenesis. *Sci. Rep.* 9:9711. doi: 10.1038/s41598-019-45910-9

Bento, C. F., Renna, M., Ghislat, G., Puri, C., Ashkenazi, A., Vicinanza, M., et al. (2016). Mammalian autophagy: how does it work? *Annu. Rev. Biochem.* 85, 685–713. doi: 10.1146/annurev-biochem-060815-014556

Best, S. M., Morris, K. L., Shannon, J. G., Robertson, S. J., Mitzel, D. N., Park, G. S., et al. (2005). Inhibition of interferon-stimulated JAK-STAT signaling by a tick-borne Flavivirus and identification of NS5 as an interferon antagonist. *J. Virol.* 79, 12828–12839. doi: 10.1128/JVI.79.20.12828-12839.2005

Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., et al. (2013). The global distribution and burden of dengue. *Nature* 496, 504–507. doi: 10.1038/nature12060

Bogachek, M. V., Protopopova, E. V., Loktev, V. B., Zaitsev, B. N., Favre, M., Sekatskii, S. K., et al. (2008). Immunochemical and single molecule force spectroscopy studies of specific interaction between the laminin binding protein and the West Nile virus surface glycoprotein E domain II. *J. Mol. Recognit.* 21, 55–62. doi: 10.1002/jmr.866

Brady, O. J., Gething, P. W., Bhatt, S., Messina, J. P., Brownstein, J. S., Hoen, A. G., et al. (2012). Refining the global spatial limits of dengue virus transmission by evidence-based consensus. *PLoS Negl. Trop. Dis.* 6:e1760. doi: 10.1371/journal.pntd.0001760

Byk, L. A., Iglesias, N. G., Maio, F. A. D., Gebhard, L. G., Rossi, M., and Gamarnik, A. V. (2016). Dengue virus genome uncoating requires ubiquitination. *MBio* 7, e00804–e00816. doi: 10.1128/mBio.00804-16

Cai, X., Chiu, Y.-H., and Chen, Z. J. (2014). The cGAS-cGAMP-STING pathway of cytosolic DNA sensing and signaling. *Mol. Cell* 54, 289–296. doi: 10.1016/j.molcel.2014.03.040

Cao, Y.-Q., Yuan, L., Zhao, Q., Yuan, J.-L., Miao, C., Chang, Y.-F., et al. (2019). Hsp40 protein DNAJB6 interacts with viral NS3 and inhibits the replication of the Japanese encephalitis virus. *Int. J. Mol. Sci.* 20:E5719. doi: 10.3390/ijms20225719

Cao-Lormeau, V.-M., Blake, A., Mons, S., Lastère, S., Roche, C., Vanhomwegen, J., et al. (2016). Guillain-Barré syndrome outbreak associated with Zika virus infection in French Polynesia: a case-control study. *Lancet* 387, 1531–1539. doi: 10.1016/S0140-6736(16)00562-6

Carod-Artal, F. J. (2018). Neurological complications of Zika virus infection. *Expert Rev. Anti-Infect. Ther.* 16, 399–410. doi: 10.1080/14787210.2018.1466702

Carpp, L. N., Rogers, R. S., Moritz, R. L., and Aitchison, J. D. (2014). Quantitative proteomic analysis of host-virus interactions reveals a role for Golgi brefeldin A resistance factor 1 (GBF1) in dengue infection. *Mol. Cell. Proteomics* 13, 2836–2854. doi: 10.1074/mcp.M114.038984

Chan, Y. K., and Gack, M. U. (2016). A phosphomimetic-based mechanism of dengue virus to antagonize innate immunity. *Nat. Immunol.* 17, 523–530. doi: 10.1038/ni.3393

Chandramouli, S., Joseph, J. S., Daudenarde, S., Gatchalian, J., Cornillez-Ty, C., and Kuhn, P. (2010). Serotype-specific structural differences in the protease-cofactor complexes of the dengue virus family. *J. Virol.* 84, 3059–3067. doi: 10.1128/JVI.02044-09

Chao, L. H., Klein, D. E., Schmidt, A. G., Peña, J. M., and Harrison, S. C. (2014). Sequential conformational rearrangements in flavivirus membrane fusion. *elife* 3:e04389. doi: 10.7554/eLife.04389

Chatel-Chaix, L., Cortese, M., Romero-Brey, I., Bender, S., Neufeldt, C. J., Fischl, W., et al. (2016). Dengue virus perturbs mitochondrial morphodynamics to dampen innate immune responses. *Cell Host Microbe* 20, 342–356. doi: 10.1016/j.chom.2016.07.008

Chavali, P. L., Stojic, L., Meredith, L. W., Joseph, N., Nahorski, M. S., Sanford, T. J., et al. (2017). Neurodevelopmental protein Musashi-1 interacts with the Zika genome and promotes viral replication. *Science* 357, 83–88. doi: 10.1126/science.aam9243

Chen, S., Desai, T., McNew, J. A., Gerard, P., Novick, P. J., and Ferro-Novick, S. (2015). Lunapark stabilizes nascent three-way junctions in the endoplasmic reticulum. *Proc. Natl. Acad. Sci. U. S. A.* 112, 418–423. doi: 10.1073/pnas.1423026112

Chen, W., Foo, S.-S., Hong, E., Wu, C., Lee, W.-S., Lee, S.-A., et al. (2021). Zika virus NS3 protease induces bone morphogenetic protein-dependent brain calcification in human fetuses. *Nat. Microbiol.* 6, 455–466. doi: 10.1038/s41564-020-00850-3

Chen, Y., Maguire, T., Hileman, R. E., Fromm, J. R., Esko, J. D., Linhardt, R. J., et al. (1997). Dengue virus infectivity depends on envelope protein binding to target cell heparan sulfate. *Nat. Med.* 3, 866–871. doi: 10.1038/nm0897-866

Cheng, C. X., Alvin Tan, M. J., Chan, K. W. K., Watanabe, S., Wang, S., Choy, M. M., et al. (2021). In vitro and in vivo stability of P884T, a mutation that relocalizes dengue virus 2 non-structural protein 5. *ACS Infect. Dis.* 7, 3277–3291. doi: 10.1021/acsinfecdis.1c00441

Chiang, J. J., Davis, M. E., and Gack, M. U. (2014). Regulation of RIG-I-like receptor signaling by host and viral proteins. *Cytokine Growth Factor Rev.* 25, 491–505. doi: 10.1016/j.cytogfr.2014.06.005

Choi, Y., Bowman, J. W., and Jung, J. U. (2018). Autophagy during viral infection—a double-edged sword. *Nat. Rev. Microbiol.* 16, 341–354. doi: 10.1038/s41579-018-0003-6

Chotiwan, N., Andre, B. G., Sanchez-Vargas, I., Islam, M. N., Grabowski, J. M., Hopf-Jannasch, A., et al. (2018). Dynamic remodeling of lipids coincides with dengue virus replication in the midgut of Aedes aegypti mosquitoes. *PLoS Pathog.* 14:e1006853. doi: 10.1371/journal.ppat.1006853

Chu, J. J., and Ng, M.-L. (2004). Interaction of West Nile virus with αvβ3 integrin mediates virus entry into cells*. *J. Biol. Chem.* 279, 54533–54541. doi: 10.1074/jbc.M410208200

Chung, K. M., Liszewski, M. K., Nybakken, G., Davis, A. E., Townsend, R. R., Fremont, D. H., et al. (2006). West Nile virus nonstructural protein NS1 inhibits complement activation by binding the regulatory protein factor H. *PNAS* 103, 19111–19116. doi: 10.1073/pnas.0605668103

Conde, J. N., da Silva, E. M., Allonso, D., Coelho, D. R., Andrade, I. D. S., de Medeiros, L. N., et al. (2016). Inhibition of the membrane attack complex by dengue virus NS1 through interaction with vitronectin and terminal complement proteins. *J. Virol.* 90, 9570–9581. doi: 10.1128/JVI.00912-16

Cortese, M., Goellner, S., Acosta, E. G., Neufeldt, C. J., Oleksiuk, O., Lampe, M., et al. (2017). Ultrastructural characterization of Zika virus replication factories. *Cell Rep.* 18, 2113–2123. doi: 10.1016/j.celrep.2017.02.014

Coyaud, E., Ranadheera, C., Cheng, D., Gonçalves, J., Dyakov, B. J. A., Laurent, E. M. N., et al. (2018). Global Interactomics uncovers extensive Organellar targeting by Zika virus. *Mol. Cell. Proteomics* 17, 2242–2255. doi: 10.1074/mcp.TIR118.000800

da Silva Voorham, J. M. (2014). A possible fifth dengue virus serotype. *Ned. Tijdschr. Geneeskd.* 158:A7946.

Dalrymple, N. A., Cimica, V., and Mackow, E. R. (2015). Dengue virus NS proteins inhibit RIG-I/MAVS signaling by blocking TBK1/IRF3 phosphorylation: dengue virus serotype 1 NS4A is a unique interferon-regulating virulence determinant. *MBio* 6, e00553–e00515. doi: 10.1128/mBio.00553-15

Damas, N. D., Fossat, N., and Scheel, T. K. H. (2019). Functional interplay between RNA viruses and non-coding RNA in mammals. *Noncoding RNA* 5:7. doi: 10.3390/ncrna5010007

Darnell, J. E., Kerr, I. M., and Stark, G. R. (1994). Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science* 264, 1415–1421. doi: 10.1126/science.8197455

Das, S., Laxminarayana, S. V., Chandra, N., Ravi, V., and Desai, A. (2009). Heat shock protein 70 on Neuro2a cells is a putative receptor for Japanese encephalitis virus. *Virology* 385, 47–57. doi: 10.1016/j.virol.2008.10.025

de Araújo, T. V. B., Rodrigues, L. C., de Alencar Ximenes, R. A., de Barros Miranda-Filho, D., Montarroyos, U. R., de Melo, A. P. L., et al. (2016). Association between Zika virus infection and microcephaly in Brazil, January to may, 2016: preliminary report of a case-control study. *Lancet Infect. Dis.* 16, 1356–1363. doi: 10.1016/S1473-3099(16)30318-8

de la Fuente, J., Antunes, S., Bonnet, S., Cabezas-Cruz, A., Domingos, A. G., Estrada-Peña, A., et al. (2017). Tick-pathogen interactions and vector competence: identification of molecular drivers for tick-borne diseases. *Front. Cell. Infect. Microbiol.* 7:114. doi: 10.3389/fcimb.2017.00114

De Maio, F. A., Risso, G., Iglesias, N. G., Shah, P., Pozzi, B., Gebhard, L. G., et al. (2016). The dengue virus NS5 protein intrudes in the cellular spliceosome and modulates splicing. *PLoS Pathog.* 12:e1005841. doi: 10.1371/journal.ppat.1005841

DeBiasi, R. L., and Tyler, K. L. (2006). West Nile virus meningoencephalitis. *Nat. Clin. Pract. Neurol.* 2, 264–275. doi: 10.1038/ncpneuro0176

Dejarnac, O., Hafirassou, M. L., Chazal, M., Versapuech, M., Gaillard, J., Perera-Lecoin, M., et al. (2018). TIM-1 ubiquitination mediates dengue virus entry. *Cell Rep.* 23, 1779–1793. doi: 10.1016/j.celrep.2018.04.013

Denks, K., Vogt, A., Sachelaru, I., Petriman, N.-A., Kudva, R., and Koch, H.-G. (2014). The sec translocon mediated protein transport in prokaryotes and eukaryotes. *Mol. Membr. Biol.* 31, 58–84. doi: 10.3109/09687688.2014.907455

Diaz, A., Wang, X., and Ahlquist, P. (2010). Membrane-shaping host reticulon proteins play crucial roles in viral RNA replication compartment formation and function. *Proc. Natl. Acad. Sci. U. S. A.* 107, 16291–16296. doi: 10.1073/pnas.1011105107

Douradinha, B., McBurney, S. P., Soares de Melo, K. M., Smith, A. P., Krishna, N. K., Barratt-Boyes, S. M., et al. (2014). C1q binding to dengue virus decreases levels of infection and inflammatory molecules transcription in THP-1 cells. *Virus Res.* 179, 231–234. doi: 10.1016/j.virusres.2013.11.007

Duan, Y., Zeng, M., Zhang, W., Liu, P., Yang, C., Wang, M., et al. (2019). Expression and purification of the truncated duck DTMUV NS5 protein and the subcellular localization of NS5 in vitro. *Poult. Sci.* 98, 2989–2996. doi: 10.3382/ps/pez117

Echavarria-Consuegra, L., Smit, J. M., and Reggiori, F. (2019). Role of autophagy during the replication and pathogenesis of common mosquito-borne flavi- and alphaviruses. *Open Biol.* 9:190009. doi: 10.1098/rsob.190009

Egloff, M.-P., Benarroch, D., Selisko, B., Romette, J.-L., and Canard, B. (2002). An RNA cap (nucleoside-2'-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. *EMBO J.* 21, 2757–2768. doi: 10.1093/emboj/21.11.2757

Falgout, B., Pethel, M., Zhang, Y. M., and Lai, C. J. (1991). Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. *J. Virol.* 65, 2467–2475. doi: 10.1128/JVI.65.5.2467-2475.1991

Funk, A., Truong, K., Nagasaki, T., Torres, S., Floden, N., Balmori Melian, E., et al. (2010). RNA structures required for production of subgenomic flavivirus RNA. *J. Virol.* 84, 11407–11417. doi: 10.1128/JVI.01159-10

Galluzzi, L., Baehrecke, E. H., Ballabio, A., Boya, P., Bravo-San Pedro, J. M., Cecconi, F., et al. (2017). Molecular definitions of autophagy and related processes. *EMBO J.* 36, 1811–1836. doi: 10.15252/embj.201796697

García, C. C., Vázquez, C. A., Giovannoni, F., Russo, C. A., Cordo, S. M., Alaimo, A., et al. (2020). Cellular organelles reorganization During Zika virus infection of human cells. *Front. Microbiol.* 11:1558. doi: 10.3389/fmicb.2020.01558

Gatica, D., Lahiri, V., and Klionsky, D. J. (2018). Cargo recognition and degradation by selective autophagy. *Nat. Cell Biol.* 20, 233–242. doi: 10.1038/s41556-018-0037-z

Germi, R., Crance, J.-M., Garin, D., Guimet, J., Lortat-Jacob, H., Ruigrok, R. W. H., et al. (2002). Heparan sulfate-mediated binding of infectious dengue virus type 2 and yellow fever virus. *Virology* 292, 162–168. doi: 10.1006/viro.2001.1232

Gillespie, L. K., Hoenen, A., Morgan, G., and Mackenzie, J. M. (2010). The endoplasmic reticulum provides the membrane platform for biogenesis of the Flavivirus replication complex. *J. Virol.* 84, 10438–10447. doi: 10.1128/JVI.00986-10

Giraldo, M. I., Xia, H., Aguilera-Aguirre, L., Hage, A., van Tol, S., Shan, C., et al. (2020). Envelope protein ubiquitination drives entry and pathogenesis of Zika virus. *Nature* 585, 414–419. doi: 10.1038/s41586-020-2457-8

Golubeva, V. A., Nepomuceno, T. C., de Gregoriis, G., Mesquita, R. D., Li, X., Dash, S., et al. (2020). Network of interactions between ZIKA virus non-structural proteins and human host proteins. *Cell* 9:E153. doi: 10.3390/cells9010153

Goyal, U., and Blackstone, C. (2013). Untangling the web: mechanisms underlying ER network formation. *Biochim. Biophys. Acta* 1833, 2492–2498. doi: 10.1016/j.bbamcr.2013.04.009

Grant, A., Ponia, S. S., Tripathi, S., Balasubramaniam, V., Miorin, L., Sourisseau, M., et al. (2016). Zika virus targets human STAT2 to inhibit type I interferon signaling. *Cell Host Microbe* 19, 882–890. doi: 10.1016/j.chom.2016.05.009

Gubas, A., and Dikic, I. (2021). A guide to the regulation of selective autophagy receptors. *FEBS J.* 289, 75–89. doi: 10.1111/febs.15824

Guo, J.-T., Hayashi, J., and Seeger, C. (2005). West Nile virus inhibits the signal transduction pathway of alpha interferon. *J. Virol.* 79, 1343–1350. doi: 10.1128/JVI.79.3.1343-1350.2005

Hackett, B. A., and Cherry, S. (2018). Flavivirus internalization is regulated by a size-dependent endocytic pathway. *Proc. Natl. Acad. Sci. U. S. A.* 115, 4246–4251. doi: 10.1073/pnas.1720032115

Hafirassou, M. L., Meertens, L., Umaña-Diaz, C., Labeau, A., Dejarnac, O., Bonnet-Madin, L., et al. (2017). A global Interactome map of the dengue virus NS1 identifies virus restriction and dependency host factors. *Cell Rep.* 21, 3900–3913. doi: 10.1016/j.celrep.2017.11.094

Haller, O., Kochs, G., and Weber, F. (2006). The interferon response circuit: induction and suppression by pathogenic viruses. *Virology* 344, 119–130. doi: 10.1016/j.virol.2005.09.024

Hannemann, H., Sung, P.-Y., Chiu, H.-C., Yousuf, A., Bird, J., Lim, S. P., et al. (2013). Serotype-specific differences in dengue virus non-structural protein 5 nuclear localization. *J. Biol. Chem.* 288, 22621–22635. doi: 10.1074/jbc.M113.481382

Hanners, N. W., Eitson, J. L., Usui, N., Richardson, R. B., Wexler, E. M., Konopka, G., et al. (2016). Western Zika virus in human fetal neural progenitors persists Long term with partial cytopathic and limited immunogenic effects. *Cell Rep.* 15, 2315–2322. doi: 10.1016/j.celrep.2016.05.075

He, Z., Zhu, X., Wen, W., Yuan, J., Hu, Y., Chen, J., et al. (2016). Dengue virus subverts host innate immunity by targeting adaptor protein MAVS. *J. Virol.* 90, 7219–7230. doi: 10.1128/JVI.00221-16

Heaton, N. S., Moshkina, N., Fenouil, R., Gardner, T. J., Aguirre, S., Shah, P. S., et al. (2016). Targeting viral Proteostasis limits influenza virus, HIV, and dengue virus infection. *Immunity* 44, 46–58. doi: 10.1016/j.immuni.2015.12.017

Heaton, N. S., and Randall, G. (2010). Dengue virus-induced autophagy regulates lipid metabolism. *Cell Host Microbe* 8, 422–432. doi: 10.1016/j.chom.2010.10.006

Heil, F., Hemmi, H., Hochrein, H., Ampenberger, F., Kirschning, C., Akira, S., et al. (2004). Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. *Science* 303, 1526–1529. doi: 10.1126/science.1093620

Hoffmann, H.-H., Schneider, W. M., Rozen-Gagnon, K., Miles, L. A., Schuster, F., Razooky, B., et al. (2021). TMEM41B is a Pan-flavivirus host factor. *Cell* 184, 133–148.e20. doi: 10.1016/j.cell.2020.12.005

Hu, J., Shibata, Y., Zhu, P.-P., Voss, C., Rismanchi, N., Prinz, W. A., et al. (2009). A class of dynamin-like GTPases involved in the generation of the tubular ER network. *Cell* 138, 549–561. doi: 10.1016/j.cell.2009.05.025

Hu, D., Wang, Y., Li, A., Li, Q., Wu, C., Shereen, M. A., et al. (2021). LAMR1 restricts Zika virus infection by attenuating the envelope protein ubiquitination. *Virulence* 12, 1795–1807. doi: 10.1080/21505594.2021.1948261

Huang, Y.-J. S., Higgs, S., Horne, K. M., and Vanlandingham, D. L. (2014). Flavivirus-mosquito interactions. *Viruses* 6, 4703–4730. doi: 10.3390/v6114703

Huang, D., Xu, B., Liu, L., Wu, L., Zhu, Y., Ghanbarpour, A., et al. (2021). TMEM41B acts as an ER scramblase required for lipoprotein biogenesis and lipid homeostasis. *Cell Metab.* 33, 1655–1670.e8. doi: 10.1016/j.cmet.2021.05.006

Jagger, B. W., Miner, J. J., Cao, B., Arora, N., Smith, A. M., Kovacs, A., et al. (2017). Gestational stage and IFN-λ signaling regulate ZIKV infection in utero. *Cell Host Microbe* 22, 366–376.e3. doi: 10.1016/j.chom.2017.08.012

Jonikas, M. C., Collins, S. R., Denic, V., Oh, E., Quan, E. M., Schmid, V., et al. (2009). Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science* 323, 1693–1697. doi: 10.1126/science.1167983

Jordan, T. X., and Randall, G. (2017). Dengue virus activates the AMP kinase-mTOR axis to stimulate a proviral lipophagy. *J. Virol.* 91, e02020–e03016. doi: 10.1128/JVI.02020-16

Jouvenet, N., Neil, S. J. D., Zhadina, M., Zang, T., Kratovac, Z., Lee, Y., et al. (2009). Broad-spectrum inhibition of retroviral and filoviral particle release by tetherin. *J. Virol.* 83, 1837–1844. doi: 10.1128/JVI.02211-08

Junjhon, J., Pennington, J. G., Edwards, T. J., Perera, R., Lanman, J., and Kuhn, R. J. (2014). Ultrastructural characterization and three-dimensional architecture of replication sites in dengue virus-infected mosquito cells. *J. Virol.* 88, 4687–4697. doi: 10.1128/JVI.00118-14

Kaufusi, P. H., Kelley, J. F., Yanagihara, R., and Nerurkar, V. R. (2014). Induction of endoplasmic reticulum-derived replication-competent membrane structures by West Nile virus non-structural protein 4B. *PLoS One* 9:e84040. doi: 10.1371/journal.pone.0084040

Kawasaki, T., and Kawai, T. (2014). Toll-Like receptor signaling pathways. *Front. Immunol.* 5:461. doi: 10.3389/fimmu.2014.00461

Ke, P.-Y. (2018). The multifaceted roles of autophagy in Flavivirus-host interactions. *Int. J. Mol. Sci.* 19:3940. doi: 10.3390/ijms19123940

Keenan, R. J., Freymann, D. M., Stroud, R. M., and Walter, P. (2001). The signal recognition particle. *Annu. Rev. Biochem.* 70, 755–775. doi: 10.1146/annurev.biochem.70.1.755

Khadka, S., Vangeloff, A. D., Zhang, C., Siddavatam, P., Heaton, N. S., Wang, L., et al. (2011). A physical interaction network of dengue virus and human proteins. *Mol. Cell. Proteomics* 10:012187. doi: 10.1074/mcp.M111.012187

Kobayashi, S., Yoshii, K., Phongphaew, W., Muto, M., Hirano, M., Orba, Y., et al. (2020). West Nile virus capsid protein inhibits autophagy by AMP-activated protein kinase degradation in neurological disease development. *PLoS Pathog.* 16:e1008238. doi: 10.1371/journal.ppat.1008238

Koh, C., Islam, M. N., Ye, Y. H., Chotiwan, N., Graham, B., Belisle, J. T., et al. (2020). Dengue virus dominates lipid metabolism modulations in Wolbachia-coinfected Aedes aegypti. *Commun. Biol.* 3, 518–514. doi: 10.1038/s42003-020-01254-z

Kolev, M., Friec, G. L., and Kemper, C. (2014). Complement—tapping into new sites and effector systems. *Nat. Rev. Immunol.* 14, 811–820. doi: 10.1038/nri3761

Kozik, P., Hodson, N. A., Sahlender, D. A., Simecek, N., Soromani, C., Wu, J., et al. (2013). A human genome-wide screen for regulators of Clathrin-coated vesicle formation reveals an unexpected role for the V-ATPase. *Nat. Cell Biol.* 15, 50–60. doi: 10.1038/ncb2652

Krishnan, M. N., Ng, A., Sukumaran, B., Gilfoy, F. D., Uchil, P. D., Sultana, H., et al. (2008). RNA interference screen for human genes associated with West Nile virus infection. *Nature* 455, 242–245. doi: 10.1038/nature07207

Kumar, A., Hou, S., Airo, A. M., Limonta, D., Mancinelli, V., Branton, W., et al. (2016). Zika virus inhibits type-I interferon production and downstream signaling. *EMBO Rep.* 17, 1766–1775. doi: 10.15252/embr.201642627

Lafourcade, C., Sobo, K., Kieffer-Jaquinod, S., Garin, J., and van der Goot, F. G. (2008). Regulation of the V-ATPase along the endocytic pathway occurs through reversible subunit association and membrane localization. *PLoS One* 3:e2758. doi: 10.1371/journal.pone.0002758

Lahon, A., Arya, R. P., and Banerjea, A. C. (2021). Dengue virus dysregulates master transcription factors and PI3K/AKT/mTOR signaling pathway in megakaryocytes. *Front. Cell. Infect. Microbiol.* 11:715208. doi: 10.3389/fcimb.2021.715208

Laurent-Rolle, M., Boer, E. F., Lubick, K. J., Wolfinbarger, J. B., Carmody, A. B., Rockx, B., et al. (2010). The NS5 protein of the virulent West Nile virus NY99 strain is a potent antagonist of type I interferon-mediated JAK-STAT signaling. *J. Virol.* 84, 3503–3515. doi: 10.1128/JVI.01161-09

Le Breton, M., Meyniel-Schicklin, L., Deloire, A., Coutard, B., Canard, B., de Lamballerie, X., et al. (2011). Flavivirus NS3 and NS5 proteins interaction network: a high-throughput yeast two-hybrid screen. *BMC Microbiol.* 11:234. doi: 10.1186/1471-2180-11-234

Lee, C. M., Xie, X., Zou, J., Li, S.-H., Lee, M. Y. Q., Dong, H., et al. (2015). Determinants of dengue virus NS4A protein oligomerization. *J. Virol.* 89, 6171–6183. doi: 10.1128/JVI.00546-15

Lemasson, M., Caignard, G., Unterfinger, Y., Attoui, H., Bell-Sakyi, L., Hirchaud, E., et al. (2021). Exploration of binary protein-protein interactions between tick-borne flaviviruses and Ixodes ricinus. *Parasit. Vectors* 14:144. doi: 10.1186/s13071-021-04651-3

Lennemann, N. J., and Coyne, C. B. (2017). Dengue and Zika viruses subvert reticulophagy by NS2B3-mediated cleavage of FAM134B. *Autophagy* 13, 322–332. doi: 10.1080/15548627.2016.1265192

Li, M., Johnson, J. R., Truong, B., Kim, G., Weinbren, N., Dittmar, M., et al. (2019). Identification of antiviral roles for the exon-junction complex and nonsense-mediated decay in flaviviral infection. *Nat. Microbiol.* 4, 985–995. doi: 10.1038/s41564-019-0375-z

Li, Y., Li, Q., Wong, Y. L., Liew, L. S. Y., and Kang, C. (2015). Membrane topology of NS2B of dengue virus revealed by NMR spectroscopy. *Biochim. Biophys. Acta* 1848, 2244–2252. doi: 10.1016/j.bbamem.2015.06.010

Li, M., Wang, P., Zheng, Z., Hu, K., Zhang, M., Guan, X., et al. (2017). Japanese encephalitis virus counteracts BST2 restriction via its envelope protein E. *Virology* 510, 67–75. doi: 10.1016/j.virol.2017.07.008

Li, Y., Wong, Y. L., Lee, M. Y., Li, Q., Wang, Q.-Y., Lescar, J., et al. (2016). Secondary structure and membrane topology of the full-length dengue virus NS4B in micelles. *Angew. Chem. Int. Ed. Eng.* 55, 12068–12072. doi: 10.1002/anie.201606609

Liang, Q., Luo, Z., Zeng, J., Chen, W., Foo, S.-S., Lee, S.-A., et al. (2016). Zika virus NS4A and NS4B proteins deregulate Akt-mTOR signaling in human fetal neural stem cells to inhibit neurogenesis and induce autophagy. *Cell Stem Cell* 19, 663–671. doi: 10.1016/j.stem.2016.07.019

Lin, R.-J., Chang, B.-L., Yu, H.-P., Liao, C.-L., and Lin, Y.-L. (2006). Blocking of interferon-induced Jak-Stat signaling by Japanese encephalitis virus NS5 through a protein tyrosine phosphatase-mediated mechanism. *J. Virol.* 80, 5908–5918. doi: 10.1128/JVI.02714-05

Lin, D. L., Inoue, T., Chen, Y.-J., Chang, A., Tsai, B., and Tai, A. W. (2019). The ER membrane protein complex promotes biogenesis of dengue and Zika virus non-structural multi-pass Transmembrane proteins to support infection. *Cell Rep.* 27, 1666–1674.e4. doi: 10.1016/j.celrep.2019.04.051

Link, N., Chung, H., Jolly, A., Withers, M., Tepe, B., Arenkiel, B. R., et al. (2019). Mutations in ANKLE2, a ZIKA virus target, disrupt an asymmetric cell division pathway in drosophila neuroblasts to cause microcephaly. *Dev. Cell* 51, 713–729.e6. doi: 10.1016/j.devcel.2019.10.009

Liu, H. M., Loo, Y.-M., Horner, S. M., Zornetzer, G. A., Katze, M. G., and Gale, M. (2012). The mitochondrial targeting chaperone 14-3-3ε regulates a RIG-I translocon that mediates membrane association and innate antiviral immunity. *Cell Host Microbe* 11, 528–537. doi: 10.1016/j.chom.2012.04.006

Liu, Y., Luo, S., He, S., Zhang, M., Wang, P., Li, C., et al. (2015). Tetherin restricts HSV-2 release and is counteracted by multiple viral glycoproteins. *Virology* 475, 96–109. doi: 10.1016/j.virol.2014.11.005

Lodeiro, M. F., Filomatori, C. V., and Gamarnik, A. V. (2009). Structural and functional studies of the promoter element for dengue virus RNA replication. *J. Virol.* 83, 993–1008. doi: 10.1128/JVI.01647-08

López-Denman, A. J., Tuipulotu, D. E., Ross, J. B., Trenerry, A. M., White, P. A., and Mackenzie, J. M. (2021). Nuclear localisation of West Nile virus NS5 protein modulates host gene expression. *Virology* 559, 131–144. doi: 10.1016/j.virol.2021.03.018

Lu, Z.-Y., Cheng, M.-H., Yu, C.-Y., Lin, Y.-S., Yeh, T.-M., Chen, C.-L., et al. (2020). Dengue nonstructural protein 1 maintains autophagy through retarding caspase-mediated cleavage of Beclin-1. *Int. J. Mol. Sci.* 21:E9702. doi: 10.3390/ijms21249702

Lubick, K. J., Robertson, S. J., McNally, K. L., Freedman, B. A., Rasmussen, A. L., Taylor, R. T., et al. (2015). Flavivirus antagonism of type I interferon signaling reveals Prolidase as a regulator of IFNAR1 surface expression. *Cell Host Microbe* 18, 61–74. doi: 10.1016/j.chom.2015.06.007

Ma, H., Dang, Y., Wu, Y., Jia, G., Anaya, E., Zhang, J., et al. (2015). A CRISPR-based screen identifies genes essential for West-Nile-virus-induced cell death. *Cell Rep.* 12, 673–683. doi: 10.1016/j.celrep.2015.06.049

Malygin, A. A., Bondarenko, E. I., Ivanisenko, V. A., Protopopova, E. V., Karpova, G. G., and Loktev, V. B. (2009). C-terminal fragment of human laminin-binding protein contains a receptor domain for venezuelan equine

encephalitis and tick-borne encephalitis viruses. *Biochemistry (Mosc)* 74, 1328–1336. doi: 10.1134/s0006297909120050

Marazzi, I., Ho, J. S. Y., Kim, J., Manicassamy, B., Dewell, S., Albrecht, R. A., et al. (2012). Suppression of the antiviral response by an influenza histone mimic. *Nature* 483, 428–433. doi: 10.1038/nature10892

Marceau, C. D., Puschnik, A. S., Majzoub, K., Ooi, Y. S., Brewer, S. M., Fuchs, G., et al. (2016). Genetic dissection of Flaviviridae host factors through genome-scale CRISPR screens. *Nature* 535, 159–163. doi: 10.1038/nature18631

Mazzon, M., Jones, M., Davidson, A., Chain, B., and Jacobs, M. (2009). Dengue virus NS5 inhibits interferon-α signaling by blocking signal transducer and activator of transcription 2 phosphorylation. *J. Infect. Dis.* 200, 1261–1270. doi: 10.1086/605847

McLean, J. E., Wudzinska, A., Datan, E., Quaglino, D., and Zakeri, Z. (2011). Flavivirus NS4A-induced autophagy protects cells against death and enhances virus replication. *J. Biol. Chem.* 286, 22147–22159. doi: 10.1074/jbc.M110.192500

Meertens, L., Carnec, X., Lecoin, M. P., Ramdasi, R., Guivel-Benhassine, F., Lew, E., et al. (2012). The TIM and TAM families of phosphatidylserine receptors mediate dengue virus entry. *Cell Host Microbe* 12, 544–557. doi: 10.1016/j.chom.2012.08.009

Mehlhop, E., and Diamond, M. S. (2006). Protective immune responses against West Nile virus are primed by distinct complement activation pathways. *J. Exp. Med.* 203, 1371–1381. doi: 10.1084/jem.20052388

Merle, N. S., Church, S. E., Fremeaux-Bacchi, V., and Roumenina, L. T. (2015a). Complement system part i – molecular mechanisms of activation and regulation. *Front. Immunol.* 6:262. doi: 10.3389/fimmu.2015.00262

Merle, N. S., Noe, R., Halbwachs-Mecarelli, L., Fremeaux-Bacchi, V., and Roumenina, L. T. (2015b). Complement system part II: role in immunity. *Front. Immunol.* 6:257. doi: 10.3389/fimmu.2015.00257

Miller, J. L., de Wet, B. J. M., deWet, B. J. M., Martinez-Pomares, L., Radcliffe, C. M., Dwek, R. A., et al. (2008). The mannose receptor mediates dengue virus infection of macrophages. *PLoS Pathog.* 4:e17. doi: 10.1371/journal.ppat.0040017

Miller, S., Kastner, S., Krijnse-Locker, J., Bühler, S., and Bartenschlager, R. (2007). The non-structural protein 4A of dengue virus is an integral membrane protein inducing membrane alterations in a 2K-regulated manner. *J. Biol. Chem.* 282, 8873–8882. doi: 10.1074/jbc.M609919200

Miner, J. J., Cao, B., Govero, J., Smith, A. M., Fernandez, E., Cabrera, O. H., et al. (2016). Zika virus infection during pregnancy in mice causes placental damage and fetal demise. *Cell* 165, 1081–1091. doi: 10.1016/j.cell.2016.05.008

Monel, B., Compton, A. A., Bruel, T., Amraoui, S., Burlaud-Gaillard, J., Roy, N., et al. (2017). Zika virus induces massive cytoplasmic vacuolization and paraptosis-like death in infected cells. *EMBO J.* 36, 1653–1668. doi: 10.15252/embj.201695597

Monel, B., Rajah, M. M., Hafirassou, M. L., Sid Ahmed, S., Burlaud-Gaillard, J., Zhu, P.-P., et al. (2019). Atlastin endoplasmic reticulum-shaping proteins facilitate Zika virus replication. *J. Virol.* 93, e01047–e01019. doi: 10.1128/JVI.01047-19

Moore, C. A., Staples, J. E., Dobyns, W. B., Pessoa, A., Ventura, C. V. Fonseca, E. B. et al, (2017). Characterizing the pattern of anomalies in congenital Zika syndrome for pediatric clinicians. *JAMA Pediatr.* 171, 288–295. doi:10.1001/jamapediatrics.2016.3982

Moretti, F., Bergman, P., Dodgson, S., Marcellin, D., Claerr, I., Goodwin, J. M., et al. (2018). TMEM41B is a novel regulator of autophagy and lipid mobilization. *EMBO Rep.* 19:e45889. doi: 10.15252/embr.201845889

Morita, K., Hama, Y., and Mizushima, N. (2019). TMEM41B functions with VMP1 in autophagosome formation. *Autophagy* 15, 922–923. doi: 10.1080/15548627.2019.1582952

Morrison, J., Laurent-Rolle, M., Maestre, A. M., Rajsbaum, R., Pisanelli, G., Simon, V., et al. (2013). Dengue virus co-opts UBR4 to degrade STAT2 and antagonize type I interferon signaling. *PLoS Pathog.* 9:e1003265. doi: 10.1371/journal.ppat.1003265

Muñoz-Jordán, J. L., Laurent-Rolle, M., Ashour, J., Martínez-Sobrido, L., Ashok, M., Lipkin, W. I., et al. (2005). Inhibition of alpha/Beta interferon signaling by the NS4B protein of Flaviviruses. *J. Virol.* 79, 8004–8013. doi: 10.1128/JVI.79.13.8004-8013.2005

Muñoz-Jordán, J. L., Sánchez-Burgos, G. G., Laurent-Rolle, M., and García-Sastre, A. (2003). Inhibition of interferon signaling by dengue virus. *PNAS* 100, 14333–14338. doi: 10.1073/pnas.2335168100

Mustafa, M. S., Rasotgi, V., Jain, S., and Gupta, V. (2015). Discovery of fifth serotype of dengue virus (DENV-5): a new public health dilemma in dengue control. *Med. J. Armed Forces India* 71, 67–70. doi: 10.1016/j.mjafi.2014.09.011

Nasirudeen, A. M. A., Wong, H. H., Thien, P., Xu, S., Lam, K.-P., and Liu, D. X. (2011). RIG-I, MDA5 and TLR3 synergistically play an important role in restriction of dengue virus infection. *PLoS Negl. Trop. Dis.* 5:e926. doi: 10.1371/journal.pntd.0000926

Neil, S. J. D., Zang, T., and Bieniasz, P. D. (2008). Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* 451, 425–430. doi: 10.1038/nature06553

Neufeldt, C. J., Cortese, M., Scaturro, P., Cerikan, B., Wideman, J. G., Tabata, K., et al. (2019). ER-shaping atlastin proteins act as central hubs to promote flavivirus replication and virion assembly. *Nat. Microbiol.* 4, 2416–2429. doi: 10.1038/s41564-019-0586-3

Ngo, A. M., Shurtleff, M. J., Popova, K. D., Kulsuptrakul, J., Weissman, J. S., and Puschnik, A. S. (2019). The ER membrane protein complex is required to ensure correct topology and stable expression of flavivirus polyproteins. *elife* 8:e48469. doi: 10.7554/eLife.48469

Niu, J., Jiang, Y., Xu, H., Zhao, C., Zhou, G., Chen, P., et al. (2018). TIM-1 promotes Japanese encephalitis virus entry and infection. *Viruses* 10:630. doi: 10.3390/v10110630

Norat, P., Soldozy, S., Sokolowski, J. D., Gorick, C. M., Kumar, J. S., Chae, Y., et al. (2020). Mitochondrial dysfunction in neurological disorders: exploring mitochondrial transplantation. *NPJ Regen. Med.* 5, 1–9. doi: 10.1038/s41536-020-00107-x

Offerdahl, D. K., Dorward, D. W., Hansen, B. T., and Bloom, M. E. (2017). Cytoarchitecture of Zika virus infection in human neuroblastoma and Aedes albopictus cell lines. *Virology* 501, 54–62. doi: 10.1016/j.virol.2016.11.002

Ooi, Y. S., Dubé, M., and Kielian, M. (2015). BST2/Tetherin inhibition of Alphavirus exit. *Viruses* 7, 2147–2167. doi: 10.3390/v7042147

Ooi, Y. S., Majzoub, K., Flynn, R. A., Mata, M. A., Diep, J., Li, J. K., et al. (2019). An RNA-centric dissection of host complexes controlling flavivirus infection. *Nat. Microbiol.* 4, 2369–2382. doi: 10.1038/s41564-019-0518-2

Pacheu-Grau, D., Rucktäschel, R., and Deckers, M. (2018). Mitochondrial dysfunction and its role in tissue-specific cellular stress. *Cell Stress* 2, 184–199. doi: 10.15698/cst2018.07.147

Pan, X.-B., Han, J.-C., Cong, X., and Wei, L. (2012). BST2/Tetherin inhibits dengue virus release from human hepatoma cells. *PLoS One* 7:e51033. doi: 10.1371/journal.pone.0051033

Park, G. S., Morris, K. L., Hallett, R. G., Bloom, M. E., and Best, S. M. (2007). Identification of residues critical for the interferon antagonist function of Langat virus NS5 reveals a role for the RNA-dependent RNA polymerase domain. *J. Virol.* 81, 6936–6946. doi: 10.1128/JVI.02830-06

Parnas, O., Jovanovic, M., Eisenhaure, T. M., Herbst, R. H., Dixit, A., Ye, C. J., et al. (2015). A genome-wide CRISPR screen in primary immune cells to dissect regulatory networks. *Cell* 162, 675–686. doi: 10.1016/j.cell.2015.06.059

Pereira, L. H. S., de Souza, T. P. P., Camargos, V. N., de Oliveira Barbosa, L. A., Taranto, A. G., Junior, M. C., et al. (2019). Assays with recombinant soluble isoforms of DC-SIGN, a dengue virus ligand, show variation in their ability to bind to mannose residues. *Arch. Virol.* 164, 2793–2797. doi: 10.1007/s00705-019-04377-9

Perera, R., Riley, C., Isaac, G., Hopf-Jannasch, A. S., Moore, R. J., Weitz, K. W., et al. (2012). Dengue virus infection perturbs lipid homeostasis in infected mosquito cells. *PLoS Pathog.* 8:e1002584. doi: 10.1371/journal.ppat.1002584

Petit, M. J., Kenaston, M. W., Pham, O. H., Nagainis, A. A., Fishburn, A. T., and Shah, P. S. (2021). Nuclear dengue virus NS5 antagonizes expression of PAF1-dependent immune response genes. *PLoS Pathog.* 17:e1010100. doi: 10.1371/journal.ppat.1010100

Platt, D. J., Smith, A. M., Arora, N., Diamond, M. S., Coyne, C. B., and Miner, J. J. (2018). Zika virus-related neurotropic flaviviruses infect human placental explants and cause fetal demise in mice. *Sci. Transl. Med.* 10:eaao7090. doi: 10.1126/scitranslmed.aao7090

Ponia, S. S., Robertson, S. J., McNally, K. L., Subramanian, G., Sturdevant, G. L., Lewis, M., et al. (2021). Mitophagy antagonism by ZIKV reveals Ajuba as a regulator of PINK1 signaling, PKR-dependent inflammation, and viral invasion of tissues. *Cell Rep.* 37:109888. doi: 10.1016/j.celrep.2021.109888

Pool, K.-L., Adachi, K., Karnezis, S., Salamon, N., Romero, T., Nielsen-Saines, K., et al. (2019). Association between neonatal neuroimaging and clinical outcomes in Zika-exposed infants from Rio de Janeiro, Brazil. *JAMA Netw. Open* 2:e198124. doi: 10.1001/jamanetworkopen.2019.8124

Pozzi, B., Bragado, L., Mammi, P., Torti, M. F., Gaioli, N., Gebhard, L. G., et al. (2020). Dengue virus targets RBM10 deregulating host cell splicing and innate immune response. *Nucleic Acids Res.* 48, 6824–6838. doi: 10.1093/nar/gkaa340

Protopopova, E. V., Konavalova, S. N., and Loktev, V. B. (1997). Isolation of a cellular receptor for tick-borne encephalitis virus using anti-idiotypic antibodies. *Vopr. Virusol.* 42, 264–268.

Pujhari, S., Brustolin, M., Macias, V. M., Nissly, R. H., Nomura, M., Kuchipudi, S. V., et al. (2019). Heat shock protein 70 (Hsp70) mediates Zika virus entry, replication, and egress from host cells. *Emerg. Microbes Infect.* 8, 8–16. doi: 10.1080/22221751.2018.1557988

Ramanathan, H. N., Zhang, S., Douam, F., Mar, K. B., Chang, J., Yang, P. L., et al. (2020). A sensitive yellow fever virus entry reporter identifies valosin-containing protein (VCP/p97) as an essential host factor for flavivirus uncoating. *MBio* 11, e00467–e00420. doi: 10.1128/mBio.00467-20

Ray, D., Shah, A., Tilgner, M., Guo, Y., Zhao, Y., Dong, H., et al. (2006). West Nile virus 5′-cap structure is formed by sequential guanine N-7 and ribose 2'-O methylations by nonstructural protein 5. *J. Virol.* 80, 8362–8370. doi: 10.1128/JVI.00814-06

Reyes-Del Valle, J., Chávez-Salinas, S., Medina, F., and Del Angel, R. M. (2005). Heat shock protein 90 and heat shock protein 70 are components of dengue virus receptor complex in human cells. *J. Virol.* 79, 4557–4567. doi: 10.1128/JVI.79.8.4557-4567.2005

Richard, A. S., Shim, B.-S., Kwon, Y.-C., Zhang, R., Otsuka, Y., Schmitt, K., et al. (2017). AXL-dependent infection of human fetal endothelial cells distinguishes Zika virus from other pathogenic flaviviruses. *Proc. Natl. Acad. Sci. U. S. A.* 114, 2024–2029. doi: 10.1073/pnas.1620558114

Riedl, W., Acharya, D., Lee, J.-H., Liu, G., Serman, T., Chiang, C., et al. (2019). Zika virus NS3 mimics a cellular 14-3-3-binding motif to antagonize RIG-I- and MDA5-mediated innate immunity. *Cell Host Microbe* 26, 493–503.e6. doi: 10.1016/j.chom.2019.09.012

Rismanchi, N., Soderblom, C., Stadler, J., Zhu, P.-P., and Blackstone, C. (2008). Atlastin GTPases are required for Golgi apparatus and ER morphogenesis. *Hum. Mol. Genet.* 17, 1591–1604. doi: 10.1093/hmg/ddn046

Roach, T., and Alcendor, D. J. (2017). Zika virus infection of cellular components of the blood-retinal barriers: implications for viral associated congenital ocular disease. *J. Neuroinflammation* 14:43. doi: 10.1186/s12974-017-0824-7

Rodrigues, R., Danskog, K., Överby, A. K., and Arnberg, N. (2019). Characterizing the cellular attachment receptor for Langat virus. *PLoS One* 14:e0217359. doi: 10.1371/journal.pone.0217359

Roosendaal, J., Westaway, E. G., Khromykh, A., and Mackenzie, J. M. (2006). Regulated cleavages at the West Nile virus NS4A-2K-NS4B junctions play a major role in rearranging cytoplasmic membranes and Golgi trafficking of the NS4A protein. *J. Virol.* 80, 4623–4632. doi: 10.1128/JVI.80.9.4623-4632.2006

Rosenfeld, A. B., Doobin, D. J., Warren, A. L., Racaniello, V. R., and Vallee, R. B. (2017). Replication of early and recent Zika virus isolates throughout mouse brain development. *Proc. Natl. Acad. Sci. U. S. A.* 114, 12273–12278. doi: 10.1073/pnas.1714624114

Routhu, N. K., Lehoux, S. D., Rouse, E. A., Bidokhti, M. R. M., Giron, L. B., Anzurez, A., et al. (2019). Glycosylation of Zika virus is important in host-virus interaction and pathogenic potential. *Int. J. Mol. Sci.* 20:E5206. doi: 10.3390/ijms20205206

Ruzek, D., Avšič Županc, T., Borde, J., Chrdle, A., Eyer, L., Karganova, G., et al. (2019). Tick-borne encephalitis in Europe and Russia: review of pathogenesis, clinical features, therapy, and vaccines. *Antivir. Res.* 164, 23–51. doi: 10.1016/j.antiviral.2019.01.014

Sahoo, B. R., Pattnaik, A., Annamalai, A. S., Franco, R., and Pattnaik, A. K. (2020). Mechanistic target of Rapamycin signaling activation antagonizes autophagy to facilitate Zika virus replication. *J. Virol.* 94, e01575–e01520. doi: 10.1128/JVI.01575-20

Sakoonwatanyoo, P., Boonsanay, V., and Smith, D. R. (2006). Growth and production of the dengue virus in C6/36 cells and identification of a laminin-binding protein as a candidate serotype 3 and 4 receptor protein. *Intervirology* 49, 161–172. doi: 10.1159/000089377

Sakuma, T., Noda, T., Urata, S., Kawaoka, Y., and Yasuda, J. (2009). Inhibition of Lassa and Marburg virus production by tetherin. *J. Virol.* 83, 2382–2385. doi: 10.1128/JVI.01607-08

Sánchez-González, G., Belak, Z. R., Lozano, L., and Condé, R. (2021). Probability of consolidation constrains novel serotype emergence in dengue fever virus. *PLoS One* 16:e0248765. doi: 10.1371/journal.pone.0248765

Sarkar, R., Sharma, K. B., Kumari, A., Asthana, S., and Kalia, M. (2021). Japanese encephalitis virus capsid protein interacts with non-lipidated MAP1LC3 on replication membranes and lipid droplets. *J. Gen. Virol.* 102:1508. doi: 10.1099/jgv.0.001508

Savidis, G., McDougall, W. M., Meraner, P., Perreira, J. M., Portmann, J. M., Trincucci, G., et al. (2016). Identification of Zika virus and dengue virus dependency factors using functional genomics. *Cell Rep.* 16, 232–246. doi: 10.1016/j.celrep.2016.06.028

Scaturro, P., Stukalov, A., Haas, D. A., Cortese, M., Draganova, K., Płaszczyca, A., et al. (2018). An orthogonal proteomic survey uncovers novel Zika virus host factors. *Nature* 561, 253–257. doi: 10.1038/s41586-018-0484-5

Schmidt, K., Keller, M., Bader, B. L., Korytář, T., Finke, S., Ziegler, U., et al. (2013). Integrins modulate the infection efficiency of West Nile virus into cells. *J. Gen. Virol.* 94, 1723–1733. doi: 10.1099/vir.0.052613-0

Sessions, O. M., Barrows, N. J., Souza-Neto, J. A., Robinson, T. J., Hershey, C. L., Rodgers, M. A., et al. (2009). Discovery of insect and human dengue virus host factors. *Nature* 458, 1047–1050. doi: 10.1038/nature07967

Shah, P. S., Link, N., Jang, G. M., Sharp, P. P., Zhu, T., Swaney, D. L., et al. (2018). Comparative Flavivirus-host protein interaction mapping reveals mechanisms of dengue and Zika virus pathogenesis. *Cell* 175, 1931–1945.e18. doi: 10.1016/j.cell.2018.11.028

Sharma, K. B., Chhabra, S., Aggarwal, S., Tripathi, A., Banerjee, A., Yadav, A. K., et al. (2021). Proteomic landscape of Japanese encephalitis virus-infected fibroblasts. *J. Gen. Virol.* 102, 1–17. doi: 10.1099/jgv.0.001657

Shives, K. D., Beatman, E. L., Chamanian, M., O'Brien, C., Hobson-Peters, J., and Beckham, J. D. (2014). West Nile virus-induced activation of mammalian target of Rapamycin complex 1 supports viral growth and viral protein expression. *J. Virol.* 88, 9458–9471. doi: 10.1128/JVI.01323-14

Shue, B., Chiramel, A. I., Cerikan, B., To, T.-H., Frölich, S., Pederson, S. M., et al. (2021). Genome-wide CRISPR screen identifies RACK1 as a critical host factor for Flavivirus replication. *J. Virol.* 95:e0059621. doi: 10.1128/JVI.00596-21

Shurtleff, M. J., Itzhak, D. N., Hussmann, J. A., Schirle Oakdale, N. T., Costa, E. A., Jonikas, M., et al. (2018). The ER membrane protein complex interacts cotranslationally to enable biogenesis of multipass membrane proteins. *elife* 7:e37018. doi: 10.7554/eLife.37018

Solomon, T., Dung, N. M., Kneen, R., Gainsborough, M., Vaughn, D. W., and Khanh, V. T. (2000). Japanese encephalitis. *J. Neurol. Neurosurg. Psychiatry* 68, 405–415. doi: 10.1136/jnnp.68.4.405

Stadler, K., Allison, S. L., Schalich, J., and Heinz, F. X. (1997). Proteolytic activation of tick-borne encephalitis virus by furin. *J. Virol.* 71, 8475–8481. doi: 10.1128/JVI.71.11.8475-8481.1997

Sun, B., Sundström, K. B., Chew, J. J., Bist, P., Gan, E. S., Tan, H. C., et al. (2017). Dengue virus activates cGAS through the release of mitochondrial DNA. *Sci. Rep.* 7:3594. doi: 10.1038/s41598-017-03932-1

Sun, L., Wu, J., Du, F., Chen, X., and Chen, Z. J. (2013b). Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* 339, 786–791. doi: 10.1126/science.1232458

Sun, E., Zhao, J., TaoYang, N., Xu, Q., Qin, Y., Wang, W., et al. (2013a). Antibodies generated by immunization with the NS1 protein of West Nile virus confer partial protection against lethal Japanese encephalitis virus challenge. *Vet. Microbiol.* 166, 145–153. doi: 10.1016/j.vetmic.2013.05.026

Tan, M. J. A., Brown, N. G., Chan, K. W. K., Jin, J. Y., Zu Kong, S. Y., and Vasudevan, S. G. (2020). Mutations in the cytoplasmic domain of dengue virus NS4A affect virus fitness and interactions with other non-structural proteins. *J. Gen. Virol.* 101, 941–953. doi: 10.1099/jgv.0.001462

Tan, B. H., Fu, J., Sugrue, R. J., Yap, E. H., Chan, Y. C., and Tan, Y. H. (1996). Recombinant dengue type 1 virus NS5 protein expressed in Escherichia coli exhibits RNA-dependent RNA polymerase activity. *Virology* 216, 317–325. doi: 10.1006/viro.1996.0067

Tang, W.-F., Yang, S.-Y., Wu, B.-W., Jheng, J.-R., Chen, Y.-L., Shih, C.-H., et al. (2007). Reticulon 3 binds the 2C protein of enterovirus 71 and is required for viral replication. *J. Biol. Chem.* 282, 5888–5898. doi: 10.1074/jbc.M611145200

Tay, M. Y. F., Smith, K., Ng, I. H. W., Chan, K. W. K., Zhao, Y., Ooi, E. E., et al. (2016). The C-terminal 18 amino acid region of dengue virus NS5 regulates its subcellular localization and contains a conserved arginine residue essential for infectious virus production. *PLoS Pathog.* 12:e1005886. doi: 10.1371/journal.ppat.1005886

Teo, C. S. H., and Chu, J. J. H. (2014). Cellular vimentin regulates construction of dengue virus replication complexes through interaction with NS4A protein. *J. Virol.* 88, 1897–1913. doi: 10.1128/JVI.01249-13

Thiemmeca, S., Tamdet, C., Punyadee, N., Prommool, T., Songjaeng, A., Noisakran, S., et al. (2016). Secreted NS1 Protects Dengue Virus from

Mannose-Binding Lectin–Mediated Neutralization. *J. Immunol.* 197, 4053–4065. doi: 10.4049/jimmunol.1600323

Thongtan, T., Wikan, N., Wintachai, P., Rattanarungsan, C., Srisomsap, C., Cheepsunthorn, P., et al. (2012). Characterization of putative Japanese encephalitis virus receptor molecules on microglial cells. *J. Med. Virol.* 84, 615–623. doi: 10.1002/jmv.23248

Tio, P. H., Jong, W. W., and Cardosa, M. J. (2005). Two dimensional VOPBA reveals laminin receptor (LAMR1) interaction with dengue virus serotypes 1, 2 and 3. *Virol. J.* 2:25. doi: 10.1186/1743-422X-2-25

Tran, P.-T.-H., Asghar, N., Johansson, M., and Melik, W. (2021). Roles of the endogenous Lunapark protein during Flavivirus replication. *Viruses* 13:1198. doi: 10.3390/v13071198

Tsai, Y.-T., Chang, S.-Y., Lee, C.-N., and Kao, C.-L. (2009). Human TLR3 recognizes dengue virus and modulates viral replication in vitro. *Cell. Microbiol.* 11, 604–615. doi: 10.1111/j.1462-5822.2008.01277.x

van der Schaar, H. M., Rust, M. J., Chen, C., van der Ende-Metselaar, H., Wilschut, J., Zhuang, X., et al. (2008). Dissecting the cell entry pathway of dengue virus by single-particle tracking in living cells. *PLoS Pathog.* 4:e1000244. doi: 10.1371/journal.ppat.1000244

Wang, S., Tukachinsky, H., Romano, F. B., and Rapoport, T. A. (2016). Cooperation of the ER-shaping proteins atlastin, lunapark, and reticulons to generate a tubular membrane network. *elife* 5:e18605. doi: 10.7554/eLife.18605

Wang, Y., Zhang, S., Tang, Y., and Diao, Y. (2019). Screening of duck Tembusu virus NS3 interacting host proteins and identification of its specific interplay domains. *Viruses* 11:E740. doi: 10.3390/v11080740

Welsch, S., Miller, S., Romero-Brey, I., Merz, A., Bleck, C. K. E., Walther, P., et al. (2009). Composition and three-dimensional architecture of the dengue virus replication and assembly sites. *Cell Host Microbe* 5, 365–375. doi: 10.1016/j.chom.2009.03.007

Werme, K., Wigerius, M., and Johansson, M. (2008). Tick-borne encephalitis virus NS5 associates with membrane protein scribble and impairs interferon-stimulated JAK-STAT signalling. *Cell. Microbiol.* 10, 696–712. doi: 10.1111/j.1462-5822.2007.01076.x

Wessel, A. W., Doyle, M. P., Engdahl, T. B., Rodriguez, J., Crowe, J. E., and Diamond, M. S. (2021). Human monoclonal antibodies against NS1 protein protect against lethal West Nile virus. *Infection* 12:e0244021. doi: 10.1128/mBio.02440-21

Westermann, B. (2010). Mitochondrial fusion and fission in cell life and death. *Nat. Rev. Mol. Cell Biol.* 11, 872–884. doi: 10.1038/nrm3013

Xie, X., Gayen, S., Kang, C., Yuan, Z., and Shi, P.-Y. (2013). Membrane topology and function of dengue virus NS2A protein. *J. Virol.* 87, 4609–4622. doi: 10.1128/JVI.02424-12

Yamamoto, S., Jaiswal, M., Charng, W.-L., Gambin, T., Karaca, E., Mirzaa, G., et al. (2014). A drosophila genetic resource of mutants to study mechanisms underlying human genetic diseases. *Cell* 159, 200–214. doi: 10.1016/j.cell.2014.09.002

Yang, S., Gorshkov, K., Lee, E. M., Xu, M., Cheng, Y.-S., Sun, N., et al. (2020). Zika virus-induced neuronal apoptosis via increased mitochondrial fragmentation. *Front. Microbiol.* 11:598203. doi: 10.3389/fmicb.2020.598203

Ye, J., Chen, Z., Li, Y., Zhao, Z., He, W., Zohaib, A., et al. (2017). Japanese encephalitis virus NS5 inhibits type I interferon (IFN) production by blocking the nuclear translocation of IFN regulatory factor 3 and NF-κB. *J. Virol.* 91, e00039–e00017. doi: 10.1128/JVI.00039-17

Yi, Z., Yuan, Z., Rice, C. M., and MacDonald, M. R. (2012). Flavivirus replication complex assembly revealed by DNAJC14 functional mapping. *J. Virol.* 86, 11815–11832. doi: 10.1128/JVI.01022-12

Yockey, L. J., Jurado, K. A., Arora, N., Millet, A., Rakib, T., Milano, K. M., et al. (2018). Type I interferons instigate fetal demise after Zika virus infection. *Sci. Immunol.* 3:eaao1680. doi: 10.1126/sciimmunol.aao1680

Yoneyama, M., Kikuchi, M., Matsumoto, K., Imaizumi, T., Miyagishi, M., Taira, K., et al. (2005). Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity. *J. Immunol.* 175, 2851–2858. doi: 10.4049/jimmunol.175.5.2851

Yoon, K.-J., Song, G., Qian, X., Pan, J., Xu, D., Rho, H.-S., et al. (2017). Zika-virus-encoded NS2A disrupts mammalian cortical neurogenesis by degrading adherens junction proteins. *Cell Stem Cell* 21, 349–358.e6. doi: 10.1016/j.stem.2017.07.014

Yu, C.-Y., Chang, T.-H., Liang, J.-J., Chiang, R.-L., Lee, Y.-L., Liao, C.-L., et al. (2012). Dengue virus targets the adaptor protein MITA to subvert host innate immunity. *PLoS Pathog.* 8:e1002780. doi: 10.1371/journal.ppat.1002780

Yu, C.-Y., Liang, J.-J., Li, J.-K., Lee, Y.-L., Chang, B.-L., Su, C.-I., et al. (2015). Dengue virus impairs mitochondrial fusion by cleaving mitofusins. *PLoS Pathog.* 11:e1005350. doi: 10.1371/journal.ppat.1005350

Yu, I.-M., Zhang, W., Holdaway, H. A., Li, L., Kostyuchenko, V. A., Chipman, P. R., et al. (2008). Structure of the immature dengue virus at low pH primes proteolytic maturation. *Science* 319, 1834–1837. doi: 10.1126/science.1153264

Yuan, L., Huang, X.-Y., Liu, Z.-Y., Zhang, F., Zhu, X.-L., Yu, J.-Y., et al. (2017). A single mutation in the prM protein of Zika virus contributes to fetal microcephaly. *Science* 358, 933–936. doi: 10.1126/science.aam7120

Yung, C.-F., Lee, K.-S., Thein, T.-L., Tan, L.-K., Gan, V. C., Wong, J. G. X., et al. (2015). Dengue serotype-specific differences in clinical manifestation, laboratory parameters and risk of severe disease in adults, Singapore. *Am. J. Trop. Med. Hyg.* 92, 999–1005. doi: 10.4269/ajtmh.14-0628

Zeng, J., Dong, S., Luo, Z., Xie, X., Fu, B., Li, P., et al. (2020). The Zika virus capsid disrupts Corticogenesis by suppressing dicer activity and miRNA biogenesis. *Cell Stem Cell* 27, 618–632.e9. doi: 10.1016/j.stem.2020.07.012

Zhang, J., Lan, Y., Li, M. Y., Lamers, M. M., Fusade-Boyer, M., Klemm, E., et al. (2018). Flaviviruses exploit the lipid droplet protein AUP1 to trigger lipophagy and drive virus production. *Cell Host Microbe* 23, 819–831.e5. doi: 10.1016/j.chom.2018.05.005

Zhang, X., Liang, C., Wang, H., Guo, Z., Rong, H., Pan, J., et al. (2022). T-cell immunoglobulin and Mucin domain 1 (TIM-1) is a functional entry factor for tick-borne encephalitis virus. *MBio* e0286021. doi:10.1128/mbio.02860-21 [Epub ahead of print]

Zhang, R., Miner, J. J., Gorman, M. J., Rausch, K., Ramage, H., White, J. P., et al. (2016). A CRISPR screen defines a signal peptide processing pathway required by flaviviruses. *Nature* 535, 164–168. doi: 10.1038/nature18625

Zhao, Y. G., Chen, Y., Miao, G., Zhao, H., Qu, W., Li, D., et al. (2017). The ER-localized transmembrane protein EPG-3/VMP1 regulates SERCA activity to control ER-isolation membrane contacts for autophagosome formation. *Mol. Cell* 67, 974–989.e6. doi: 10.1016/j.molcel.2017.08.005

Zhao, Y., Soh, T. S., Zheng, J., Chan, K. W. K., Phoo, W. W., Lee, C. C., et al. (2015). A crystal structure of the dengue virus NS5 protein reveals a novel inter-domain Interface essential for protein flexibility and virus replication. *PLoS Pathog.* 11:e1004682. doi: 10.1371/journal.ppat.1004682

Zhao, Z., Tao, M., Han, W., Fan, Z., Imran, M., Cao, S., et al. (2021). Nuclear localization of Zika virus NS5 contributes to suppression of type I interferon production and response. *J. Gen. Virol.* 102:001376. doi: 10.1099/jgv.0.001376

Zheng, Y., Liu, Q., Wu, Y., Ma, L., Zhang, Z., Liu, T., et al. (2018). Zika virus elicits inflammation to evade antiviral response by cleaving cGAS via NS1-caspase-1 axis. *EMBO J.* 37:e99347. doi: 10.15252/embj.201899347

Zheng, A., Yuan, F., Kleinfelter, L. M., and Kielian, M. (2014). A toggle switch controls the low pH-triggered rearrangement and maturation of the dengue virus envelope proteins. *Nat. Commun.* 5:3877. doi: 10.1038/ncomms4877

Zurek, N., Sparks, L., and Voeltz, G. (2011). Reticulon short hairpin transmembrane domains are used to shape ER tubules. *Traffic* 12, 28–41. doi: 10.1111/j.1600-0854.2010.01134.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 15 April 2022
doi: 10.3389/fmicb.2022.870938

Assessing the Mobility of Severe Acute Respiratory Syndrome Coronavirus-2 Spike Protein Glycans by Structural and Computational Methods

Soledad Stagnoli[1†], Francesca Peccati[2†], Sean R. Connell[1,3,4], Ane Martinez-Castillo[1], Diego Charro[1], Oscar Millet[5], Chiara Bruzzone[5], Asis Palazon[3,6], Ana Ardá[3,7], Jesús Jiménez-Barbero[3,7], June Ereño-Orbea[3,7], Nicola G. A. Abrescia[1,3,8] and Gonzalo Jiménez-Osés[2,3]**

[1] Structure and Cell Biology of Viruses Lab, Center for Cooperative Research in Biosciences (CIC bioGUNE), Basque Research and Technology Alliance (BRTA), Derio, Spain, [2] Computational Chemistry Lab, CIC bioGUNE, Basque Research and Technology Alliance (BRTA), Derio, Spain, [3] IKERBASQUE, Basque Foundation for Science, Bilbao, Spain, [4] Structural Biology Unit, BioCruces Bizkaia Health Research Institute, Barakaldo, Spain, [5] Precision Medicine and Metabolism Laboratory, CIC bioGUNE, Basque Research and Technology Alliance (BRTA), Derio, Spain, [6] Cancer Immunology and Immunotherapy Lab, CIC bioGUNE, Basque Research and Technology Alliance (BRTA), Derio, Spain, [7] Chemical Glycobiology Laboratory, CIC bioGUNE, Basque Research and Technology Alliance (BRTA), Derio, Spain, [8] Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Instituto de Salud Carlos III, Madrid, Spain

Two years after its emergence, the coronavirus disease-2019 (COVID-19) pandemic caused by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) remains difficult to control despite the availability of several vaccines. The extensively glycosylated SARS-CoV-2 spike (S) protein, which mediates host cell entry by binding to the angiotensin converting enzyme 2 (ACE2) through its receptor binding domain (RBD), is the major target of neutralizing antibodies. Like to many other viral fusion proteins, the SARS-CoV-2 spike protein utilizes a glycan shield to thwart the host immune response. To grasp the influence of chemical signatures on carbohydrate mobility and reconcile the cryo-EM density of specific glycans we combined our cryo-EM map of the S ectodomain to 4.1 Å resolution, reconstructed from a limited number of particles, and all-atom molecular dynamics simulations. Chemical modifications modeled on representative glycans (defucosylation, sialylation and addition of terminal LacNAc units) show no significant influence on either protein shielding or glycan flexibility. By estimating at selected sites the local correlation between the full density map and atomic model-based maps derived from molecular dynamics simulations, we provide insight into the geometries of the α-Man-(1→3)-[α-Man-(1→6)-]-β-Man-(1→4)-β-GlcNAc(1→4)-β-GlcNAc core common to all *N*-glycosylation sites.

Keywords: cryo-EM, molecular dynamics, glycans, SARS-CoV-2, spike

OPEN ACCESS

Edited by:
Mercè Llabrés, University of the Balearic Islands, Spain

Reviewed by:
Filippo Prischi, University of Essex, United Kingdom
Terra Sztain, Freie Universität Berlin, Germany
Lorenzo Casalino, University of California, San Diego, United States

***Correspondence:**
Nicola G. A. Abrescia
nabrescia@cicbiogune.es
Gonzalo Jiménez-Osés
gjoses@cicbiogune.es

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *07 February 2022*
Accepted: *25 March 2022*
Published: *15 April 2022*

Citation:
Stagnoli S, Peccati F, Connell SR, Martinez-Castillo A, Charro D, Millet O, Bruzzone C, Palazon A, Ardá A, Jiménez-Barbero J, Ereño-Orbea J, Abrescia NGA and Jiménez-Osés G (2022) Assessing the Mobility of Severe Acute Respiratory Syndrome Coronavirus-2 Spike Protein Glycans by Structural and Computational Methods. Front. Microbiol. 13:870938. doi: 10.3389/fmicb.2022.870938

INTRODUCTION

Since the emerging of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) more than 900 cryo-EM related entries have been deposited in the Electron Microscopy Database (EMDB); about 600 targeted spike protein alone or in complex with ligands and/or antibodies. It has been shown structurally that the interaction with the cellular receptor ACE2 is mainly protein-protein and not strictly mediated by glycans. However, glycosylation of the SARS-CoV-2 spike is emerging as playing an important role in early attachment across different cell types (Wang et al., 2021). The interaction of the receptor binding domain (RBD) glycans with different human lectins expressed in different organs and tissues that may be affected during the infection process, has been shown, and recent studies demonstrated how lectins enhance SARS-CoV-2 infection (Lenza et al., 2020; Lempp et al., 2021). Also, density and type of glycosylation (so called the "glycan shield") critically impact on the antibody neutralization mechanism and on the development of suitable vaccines (Casalino et al., 2020). The difficulty to over-express the S protein and the RBD in different cell-types has biased all current structural research to three major mammalian cell lines HEK293, CHO, and Vero and one insect BL9 type (Allen et al., 2021). Efforts have been made to show that current recombinant and vaccine-produced S protein mimic the native viral glycoprotein, but further research is needed (Allen et al., 2021; Wang et al., 2021; Watanabe et al., 2021). Other glycomics studies of the spike protein purified from sputum of infected patients have shown that the glycosylation pattern is similar to the recombinantly produced S protein (Tian et al., 2021). The possibility of alternative glycan sequences for the same site challenges the understanding of the influence of glycosylation in physio-pathological processes. Furthermore, the fact that different tissues and cell types possess their own glycosylation signatures generates a combinatorial complexity to the glycans display that it is challenging to untangle. The advances in cryo-EM have indeed enormously impacted on our structural and dynamic understanding of the viral spike (Stuart et al., 2016; Subramaniam, 2020; Rapp et al., 2022). Recent studies based on the chemical structures of the most populated *N*-glycans derived from glycoanalytic data, have combined computational simulations with existing cryo-EM maps in order (*i*) to develop a fully glycosylated SARS-CoV-2 spike protein model; (*ii*) to map epitopes not shielded by the highly flexible glycans, (*iii*) to reveal their role in modulating the RBD dynamics (Casalino et al., 2020; Woo et al., 2020; Sikora et al., 2021) and (*iv*) to analyze their contribution to interaction with host cell receptors (Kapoor et al., 2021). Although the detailed determination of the glycan structure by cryo-EM (and X-ray crystallography) remains challenging due to the glycans' mobility (Atanasova et al., 2020), when visible the corresponding densities are the result of the averaging of a "predominant" glycan conformation. The local mechanical flexibility of a carbohydrate polymer is dictated by its primary and secondary structures at the single-linkage level (Anggara et al., 2021) a fact that amplifies the relevance of tissue-specific glycosylation and the structural location of the attachment site in modulating the infection process (not only in SARS-CoV-2).

In this study we sought to use our cryo-EM map at 4.1 Å resolution, reconstructed from a limited number of particles, as a means to explore the implications of oligosaccharide modifications on glycan flexibility as derived from molecular dynamics (MD) simulations. Apart from showing that our low resolution cryo-EM map is informative on the glycans presence and dynamics, we found that in MD trajectories the glycans modeled using the prevalent structure observed in previous studies for defined sites in the S2 domain of the S protein (Casalino et al., 2020) are more congruent with the cryo-EM map than when the same glycans are modified by defucosylation, sialylation and addition of terminal LacNAc units. This finding suggests that the original glycosylation pattern might be dominant in our recombinantly expressed and structurally characterized spike protein.

METHODOLOGY

The Ectodomain of the S Protein Production and Purification

The ectodomain of the S protein (ecto-S; BEI construct NR-52394) was expressed by transient transfection of HEK293F suspension cells and purified from clarified cell supernatants 7 days post-transfection using a nickel affinity column and size-exclusion chromatography as previously described (Stadlbauer et al., 2020).

The Ectodomain of the S Protein Sample Preparation and Cryo-EM Data Collection

0.06 mg/mL of ecto-S in saline buffer (20 mM Tris-HCl, pH 8.0 and 300 mM NaCl) was deposited onto a 300 mesh, R 1.2/1.3 Quantifoil grid with continuous carbon support (Electron Microscopy Sciences) that had been previously glow-discharged (37 s at 8 mA). The samples were vitrified using a Vitrobot (Mark III), after blotting for 3 s at 14°C and 100% humidity. Movies were collected on a Titan Krios operating at 300 kV using the EPU automated data collection software and recorded on a Gatan K3 Summit direct electron detector operating in counting mode. Images were recorded at a nominal magnification of 105,000× (super-resolution 0.41 Å/pixel) with a defocus range of −0.8 to −2.3 μm. Two dataset were collected; the first dataset with a dose of 1.24 e-/Å^2/frame which resulted in 49.6 e-/Å^2 total dose over 40 frames while the second one with a dose of 1.04 e-/Å^2/frame which resulted in 52 e-/Å^2 total dose over 50 frames (**Supplementary Table 1**).

Image Processing, 3D Reconstruction, and Model Refinement

Frames were motion corrected using MotionCorr2.1 and the defocus of the resulting individual micrographs was estimated using CTFFIND4 (Rohou and Grigorieff, 2015; Zheng et al., 2017). Initially, the processing of the two datasets was carried

out independently. Particles were picked in a reference-free manner using Topaz and crYOLO software (Wagner et al., 2019; Bepler et al., 2020). Particles were extracted, binned ×4 (1.66 Å/pixel), and subjected to iterative rounds of reference-free 2D and then 3D classification to identify differential class averages in RELION-3 (Zivanov et al., 2018) (as reference map we used EMD-21452 filtered to 60 Å resolution). The resulting analysis of the individual datasets led to their merging (65,372 particles). 3D classifications were performed with C1 and C3 symmetries and followed up with 3D refinement routines as we could detect both states: (*i*) one RBD up and (*ii*) three RBD down. We continued the processing of the ecto-S with the three RBD in down conformation in cryoSPARC in C3 symmetry importing 27,668 particles from a previous RELION run (Punjani et al., 2017). An *ab-initio* model was generated in cryoSPARC and a 3D classification with two classes resulted in one "junk" class (4,814 particles) and one "good" class (22,854 particles). Particles in the "good" class were refined first using the homogenous refinement protocol (4.39 Å; Bfactor −150.5 Å^2) and subsequently using a non-uniform refinement (Punjani et al., 2020) (4.10 Å; Bfactor −157.5 Å^2; 4.06 Å with auto tightening; default settings) (**Supplementary Figure 1**). The model PDB ID 6XR8 was chosen for the fitting as it presented a more complete description of the experimental sugar *versus* the corresponding density (Cai et al., 2020). This was initially fitted as a rigid-body into our cryo-EM density and minimized with NCS and secondary structure restraints (three cycles) leading to an overall CC = 70% (**Supplementary Table 1**), then the downstream analysis focused on the glycan density and their structures.

Comparison Across the Ectodomain of the S Protein Cryo-EM Maps

To inspect the interpretability of the density corresponding to the glycans in our map, different B-factors were applied: −78.5 and −100 Å^2. Then, to compare our cryo-EM density with the available higher resolution maps, the power spectra of these maps were adjusted to our map using RELION-3 (**Figure 1**). In the case of the ecto-S map whose protein was expressed in insect cells this was directly compared since it reached 4.4 Å resolution.

Model Building for Molecular Dynamics Simulations

Four fully glycosylated spike protein models in the closed conformation were built from PDB structure 6XR8 (trimer with all receptor binding domains in RBD-down conformation) (Cai et al., 2020). This reference structure presents four missing loops, corresponding to residues H69-K77, L244-S254, T618-W633, and T676-S689. These missing loops were taken from a fully glycosylated structure available from the CHARMM-GUI archive (6VSB model 1_1_1) (Jo et al., 2008; Woo et al., 2020) and grafted onto the reference structure to obtain complete chains from residue Q14 to P1162. These models are based on the cryo-EM structure identified by the PDB code 6VSB. This cryo-EM structure has played a fundamental role in the development of computational approaches to understand dynamic and structural implications of glycosylation, most notably by Amaro et al. (Casalino et al., 2020; Sztain et al., 2021). The initial 13 residues at the flexible N-terminal were not modeled. Glycans were built and added to the 57 glycosylation positions (19 for each chain of the trimer) using the GLYCAM-web glycoprotein builder (Woods, 2005). Missing hydrogens, protonation states of titratable residues and histidine tautomers were added and assigned with the *tleap* tool in AMBER (Götz et al., 2014) at neutral pH. This model was named M0 and constituted our glycan reference structure.

Then, modifications were introduced: defucosylation at positions N616, N1098, N1134, sialylation at position N657, and addition of terminal LacNAc to high-mannose glycans at positions N603, N709, N717, N801, and N1074 to generate models M1, M2, and M3, respectively (**Supplementary Figure 2**). These positions are located in the S1 (N603, N616, and N657) and S2 (N709, N717, N801, N1074, N1098, and N1134) subunits of the spike protein trimer (**Figure 2A**). The selection of these sugar sites was based on the available information on the conformational flexibility of the spike protein and glycosylation pattern density. The head region of the S protein (S1) undergoes conformational transitions at the RBD domains, that can assume two different conformations ("up" and "down"), and shows a higher glycosylation density than the S2 domain (Casalino et al., 2020; Wrapp et al., 2020). It was reasoned that focusing the comparison with the cryo-EM maps on the final region of the S1 domain and the S2 domain, would allow analyzing flexibility effects originating purely from the glycans, and not from conformational transitions of the underlying protein, such as the "up" to "down" switch of the RBD splendidly described by Amaro et al. (Sztain et al., 2021). Also, it was reasoned that focusing on a less densely glycosylated region would allow analyzing intrinsic glycan flexibility with less interference from glycan-glycan contacts.

Molecular Dynamics Simulations

Molecular dynamics simulations were carried out with AMBER 20 suite (Götz et al., 2014) using the ff14SB force field for the protein (Maier et al., 2015) and GLYCAM 06j-1 for glycans (Kirschner et al., 2008). Glycosylated spike protein models were immersed in a water box with an 8 Å buffer distance from the solute of TIP3P water molecules (Jorgensen et al., 1983) and neutralized by adding explicit Na^+ counterions. The choice of the size of the buffer distance was dictated by the need of computational efficiency, as the solvated system amounts to more than 630,000 atoms. The total charge of the models before neutralization is −6 for variants M0, M1, and M2, and −9 for M3 due to the presence of sialylated glycans in the latter. A two-stage geometry optimization approach was performed. The first stage minimizes only the positions of solvent molecules and ions, and the second stage is an unrestrained minimization of all the atoms in the simulation cell. The systems were then heated by incrementing the temperature from 0 to 300 K under a constant pressure of 1 atm and periodic boundary conditions. Harmonic restraints of 10 kcal mol^{-1} Å^{-2} were applied to the solute, and the Andersen temperature coupling scheme (Andersen, 1980) was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages, allowing

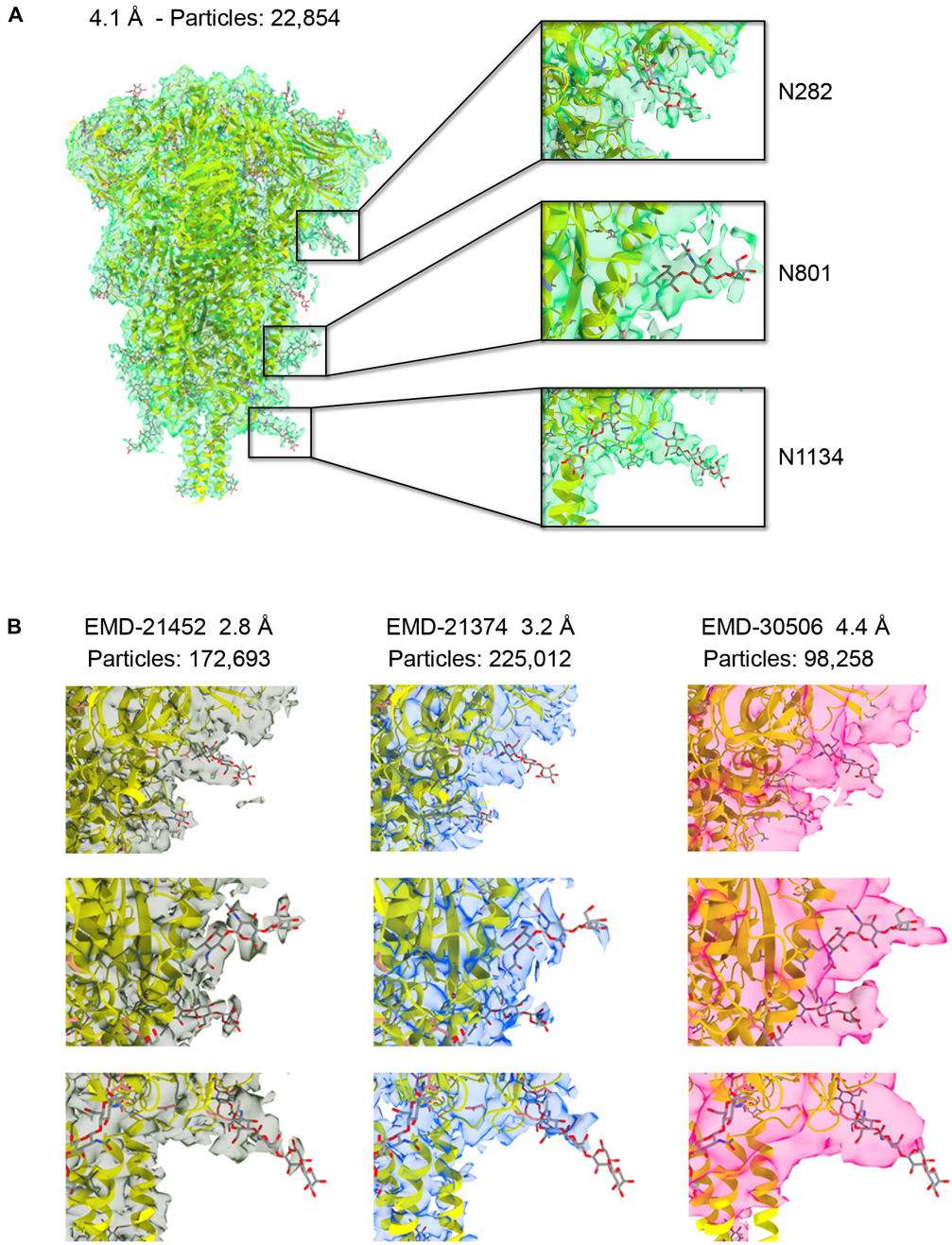


FIGURE 1 | (A) The 4.1 Å resolution cryo-EM map of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) spike ectodomain shown as transparent green isosurface depicted with a sdLevel = 3 in Chimera X (Pettersen et al., 2021) fitted with the refined structure of the closed, prefusion trimer (PDB ID 6XR8), with protein in cartoon and glycan in sticks. Insets show selected *N*-glycosylation sites along the trimeric spike with different densities features displayed at a sdLevel = 3 (not to scale). **(B)** Comparison of the density features for the same sugars shown in the insets **(A)** for available maps at higher resolution and at similar resolution to ours (not to scale).

potential inhomogeneities to self-adjust. The SHAKE algorithm was employed for further equilibration and production with a 2 fs time step (Miyamoto and Kollman, 1992). Long-range electrostatic effects were modeled using the particle mesh Ewald method (Darden et al., 1993). A cutoff of 8 Å was applied to Lennard-Jones interactions. Each system was equilibrated for 2 ns at constant volume and temperature of 300 K. To prevent substantial structural deviation from the reference structure, harmonic restraints of 10 kcal mol^{-1} $Å^{-2}$ were imposed on the protein for the whole simulation, except for the four flexible loops that are not resolved in the reference cryo-EM structure, corresponding to residues H69-K77, L244-S254, T618-W633,

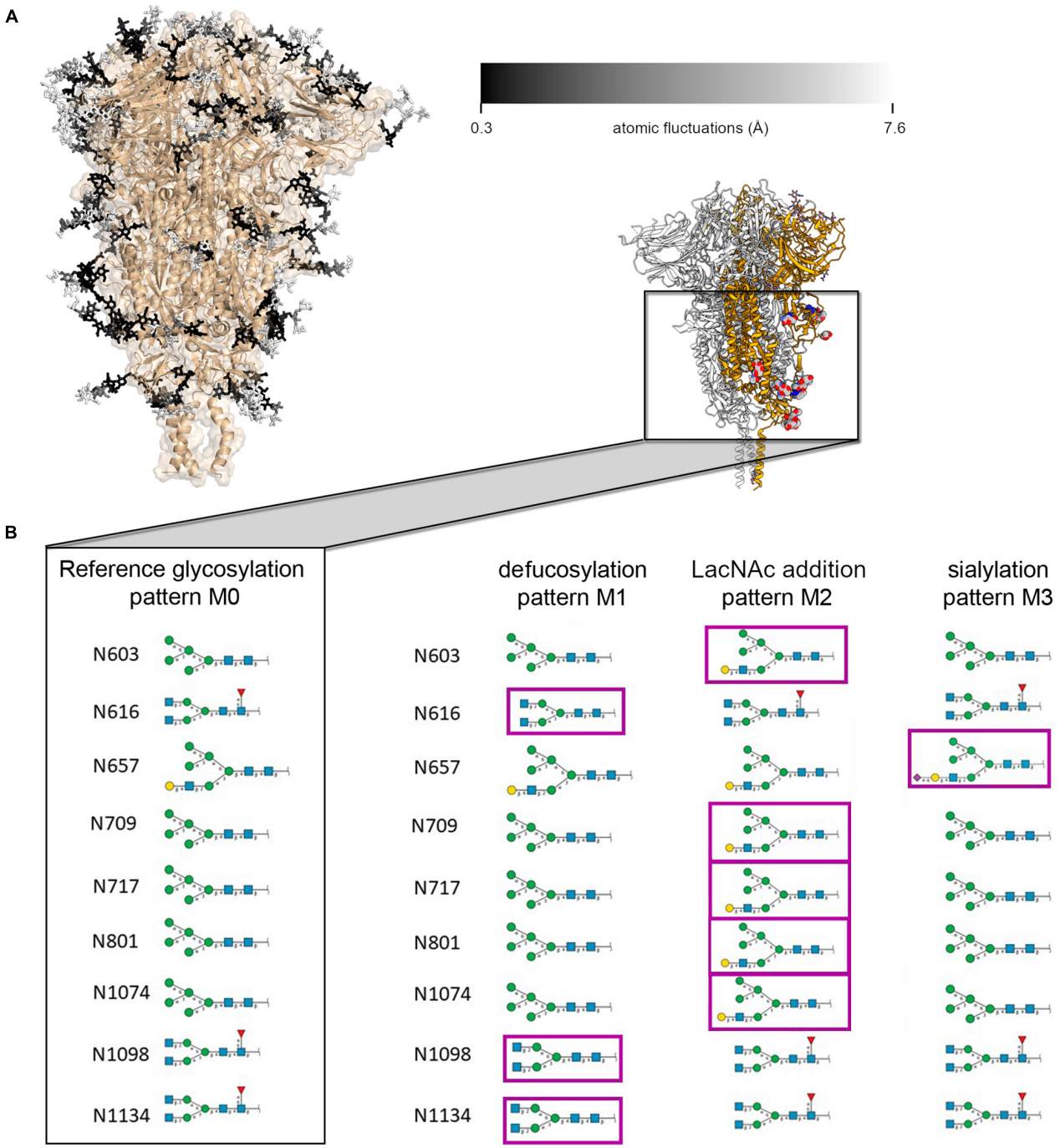


FIGURE 2 | (A) Representation of ecto-*S* glycan mobility in model M0 along a 100 ns MD simulation. Glycans (in sticks) are color-coded according to the average mobility (RMSF) of each individual carbohydrate, dark to light corresponding to rigid (0.3 Å) to flexible (>7.6 Å). **(B)** Schematic representation of the glycan structures at each site on the spike protein S2 domain for the model M0 and the modifications introduced in M1, M2, and M3 models; the inset shows the localization on the spike protein of the modified glycans (in balls and sticks).

and T676-S689. Production simulations were run as a single 100 ns trajectories for each model.

To verify that the spike protein models do not exceed the simulation box owing to glycan flexibility, their size along the three Cartesian coordinates was estimated for selected frames of the MD simulation of the four models (M0–M3) and compared it to the size of the simulation box (which is fixed in the production in the NVT ensemble). The restraints applied to protein atoms

guarantee that the orientation of the model is constant along the trajectory. To estimate the size of each glycoprotein, the distances between all pairs of C1 carbons of terminal glycans were computed along each coordinate for 20 evenly sampled frames (every 5 ns) from the production simulation and compared the maximum value with the corresponding box side (**Supplementary Tables 2–5**). In all cases, the estimated size of the system is lower than the box size by at least 15 Å, which ensures that the model stays within the simulation box along the whole trajectory.

To assess how the conformation of the glycosidic linkages at the glycans common core are affected by the interplay of glycan-glycan and protein-glycan interactions, MD simulations of the free glycan core in solution were performed. Thus, a model of the common α-Man-(1→3)-[α-Man-(1→6)-]β-Man-(1→4)-β-GlcNAc(1→4)-β-GlcNAc core linked to a single asparagine residue capped with N-terminal COMe and C-terminal NHMe groups was built. The same simulation parameters employed for the MD simulations of the spike protein trimers were used, with the exception of the size of the solvent box (a 10 Å buffer of TIP3P water molecules was used), the lack of positional restraints in the production run, and simulation time (500 ns). MD breaks the three-fold symmetry of the spike protein, as every chain moves independently of each other. The fact that MD does not introduce dramatic distortions of the trimer structure in the equilibration step, at which the system undergoes significant volume variations, was ensured by computing the RMSD of the models before and after the equilibration (**Supplementary Table 6**).

Analysis of Glycan Molecular Dynamics Trajectories Across the M0, M1, M2, and M3 Glycan Models

The flexibility of each carbohydrate unit along the MD simulations was evaluated by computing the atomic positional fluctuation (also known as root-mean-square fluctuations, RMSF) using the *cpptraj* tool in AMBER (Roe and Cheatham, 2013):

$$AtomFluct_i \text{ or } RMSF_i = \sqrt{\langle (x_i - \langle x_i \rangle)^2 \rangle}$$

where x_i are the atomic positions and averaging is over the considered frames. Atomic positional fluctuations were combined in per-residue mass-weighted averages:

$$\langle Fluct \rangle = \frac{\sum_i AtomFluct + Mass_i}{\sum_i Mass_i}$$

This flexibility analysis was performed on the whole trajectory: 5,000 frames sampled with an even stride (every 0.02 ns) from the 100 ns production simulation were considered and aligned to the first one prior to RMSF calculation. The solvent accessible surface area (SASA) was evaluated with the *surf* module of *cpptraj* on a subset of 100 frames sampled along the whole trajectory with an even stride (every 1 ns) using a probe radius of 1.4 Å. Glycan shielding percentage was computed as follows: for each MD simulation frame, the SASA value of its protein residues was subtracted from the SASA of a reference spike protein structure with the same aminoacidic composition but not glycosylated. In this way, an instantaneous glycan shielding percentage is computed for each frame, instead of accumulating glycan coverage from multiple frames.

For the cross-correlation analyses (*vide infra*) another subset of 1,000 frames sampled along the whole trajectory with an even stride (every 0.1 ns) was generated. Due to the three-fold symmetry of the cryo-EM structure, the coordinates for each protein monomer were further extracted from the 1,000 MD frames to generate a total set of 3,000 geometries per atomic model.

Glycan Dynamics in the Context of the Ectodomain of the S Protein Cryo-EM Density

The individual frames composing the trajectories derived from the MD simulation of the individual M0, M1, M2, and M3 models were fitted into our density map using Chimera X (Pettersen et al., 2021). The Electron Microscopy Data Analytical Toolkit (EMDA) was used for estimating the local correlation between the full map and the atomic model-based map at positions corresponding to the atoms in the glycan sugars (Warshamanage et al., 2022). Cross-correlation values were averaged over all non-hydrogen atoms in the individual glycans or over the entire glycan chain.

RESULTS AND DISCUSSION

Cryo-EM Map of the Ectodomain of the S Protein at 4.1 Å Resolution

The cryo-EM density of the ectodomain of SARS-CoV-2 spike protein was reconstructed from 22,854 particles and reached 4.1 Å resolution (**Figure 1A** and **Supplementary Figure 1**). This number of particles represents about 10% of those contributing to EMD-21374 at 3.2 Å, about 13% of those contributing to EMD-21452 at 2.80 Å and about 2.5% of those contributing to EMD-22251 at 2.4 Å resolution (Walls et al., 2020; Wrapp et al., 2020; Zhou et al., 2020). Until recently and to the best of our knowledge – monitoring the Electron Microscopy Data Bank (EMDB) entries only for the S protein structure in closed conformation – our map was the one reconstructed with lowest number of contributing particles. And yet, despite the limited resolution, the map is informative and descriptive of the major structural features noted in higher resolution structures. The fitting and refinement of the atomic model PDB ID 6XR8 (Cai et al., 2020) into the electron density led to a good map-model agreement (CC = 70%) (**Figure 1A** and **Supplementary Table 1**). Notably, the density for some of the nineteen glycans in the construct was readily visible in particular for glycans located in the S2 region (**Supplementary Figure 3**). While the interpretability of critical regions of cryo-EM maps at high-resolution recapitulates the global and local application of B-factor/sharpening as shown in previous test cases also with SARS-CoV-2 structures (Kaur et al., 2021; Sanchez-Garcia et al., 2021), we show that the assessment of the glycan density and their mobility on the spike surface can be derived from a medium-resolution map obtained from a streamlined number of

contributing particles. This might serve as further evidence for the implementation of a cryo-EM pipeline for the evaluation of the glycan shielding of targeted protein constructs.

Comparison of the Densities Corresponding to Glycans Across Cryo-EM Maps of the Ectodomain of the S Protein

The glycan shield in viral proteins greatly influences the recognition of epitopes by the immune system and in some cases when the shield itself is highly dense, as in the case of glycoprotein gp120 on the surface of HIV-1, carbohydrates can be targeted for antibody recognition (Trkola et al., 1996; Kunert et al., 1998). Therefore, rapid structural assessment of the glycan shield can foretell the challenges ahead for an antibody-based therapeutic strategy. To compare the interpretability of the density corresponding to the glycans across our cryo-EM map and the deposited original high-resolution maps EMD-21452 (2.8 Å), –21374 (3.2 Å) we first adjusted the power spectra amplitudes of the latter two to our map (Scheres, 2012). The spike protein samples used for the above cryo-EM reconstructions were produced in the mammalian cell system HEK293. We also compared our map with a fourth one, EMDB-30506 at 4.4 Å resolution (98,258 contributing particles) whose sample was produced in *Trichoplusia ni* insect cells. Differently than mammalian cells, insect cells have a limited capacity to produce glycans with terminal sialic acid (Marchal et al., 2001). Sialyation is most abundantly found at sites N17, N74, N165, N331, N1098, and N1194 within the spike (Watanabe et al., 2020). In our medium-resolution map glycans linked to N709, N717, N801, N1074, N1098, and N1134 were those that showed the most order in their corresponding cryo-EM density, in particular they show good clarity for two GlcNAc and a mannose before bifurcation (**Figure 1A** and **Supplementary Figure 3**). We noted, however, that density of the corresponding glycans across the four maps and in correspondence of the S1 and S2 regions were all consistent and with a gradient of increasing mobility from the connector domain (CD) toward the N-terminal domain (NTD) and receptor binding domain RBD (**Figure 1B**). This obviously reflects a convolution effect of protein dynamics with the intrinsic carbohydrate flexibility, but it also prompts that independently derived spike cryo-EM maps snapshot glycan conformations that are statistically more favorable in time and space than others.

Influence of Glycan Identity on Mobility

X-ray and cryo-EM derived structures (each in its own way) provide average static snapshots of the targeted macromolecule. In this averaging process either *in cristallo* or through 3D reconstruction methods, flexible parts (carbohydrates even more) become blurred. On the other hand, MD provides a dynamic picture of the glycan shield around the ecto-S with important implications for antibody neutralization (Casalino et al., 2020; Sikora et al., 2021). We first analyzed the flexibility of each carbohydrate constituting the glycan shield at each glycosylation position as described by Amaro and co-workers (Casalino et al., 2020). In order to maintain consistency with the three-fold symmetry (C3) of the cryo-EM map and facilitate further analysis, we modeled the same glycan at each given position of the three protomers (**Figure 2A**). Besides this original model (M0), we introduced plausible modifications on selected glycans which could be compatible with available mass spectrometry data reporting on the heterogeneity of spike protein glycosylation (**Figure 2B**; Watanabe et al., 2020). Hence, in variant M1 the glycans linked to residues N616, N1098, and N1134 were defucosylated. In variant M2, *N*-acetyllactosamine (LacNAc) units were added at the end of the α3-branch of the high mannose glycan at positions N603, N709, N717, N801, and N1074. Finally, the α3-branch at position N657 was sialylated (variant M3). In all models, the glycosylation pattern was identical for the three protomers. Overall, the flexibility of the glycans at each glycosylation position was interrogated 12 times through MD simulations (three glycoprotein chains for each of the four M variants). In this way, we obtain a cumulated sampling time of 1.2 μs for each glycosylation position, thus covering a wide variety of glycan-glycan and protein-glycan interactions and terminal glycan modifications. This strategy of analyzing frames accumulated from different glycan identities and trimer chains permits reducing the impact of the intrinsically low conformational sampling achieved by our short individual simulations, by exploring a variety of local environments, and averaging glycan conformations from multiple interaction contexts allowing for different degrees of mobility (**Supplementary Figures 4–7**). We calculated a similar dynamic glycan shielding of around 13% for the four spike protein variants at each simulation frame (**Supplementary Table 7**, see "Methodology" section for details on the calculation of glycan shielding). A general observation from our simulations was that, irrespective of the nature of the glycans or the chain they are located, local glycan-glycan and glycan-protein interactions affect flexibility along the whole glycan chain. Given their length, branching and intrinsic flexibility, glycans tend to randomly form local clusters and adhere to the protein surface, thus affecting their local mobility from one simulation to another. This result reveals the need for an efficient sampling strategy when addressing glycan flexibility in large, complex and densely glycosylated proteins, which still poses a significant computational challenge.

No clear trend was observed regarding the effect of glycan terminal modifications on the flexibility of the glycosylation core (i.e., the five common carbohydrates through which the glycan is attached to the protein, see **Figure 3** and **Supplementary Figures 4–7**), suggesting that for a densely coated glycoprotein, the chemical identity of the glycans terminal region affects glycan core flexibility only indirectly by establishing multiple transient interactions. **Figure 3** shows the cumulated mobility of each individual core carbohydrate at each glycosylation position computed from the MD simulations of M0, M1, M2, and M3. Comparison across different glycosylation positions reveals large differences in mobility depending on the morphology of the underlying protein and the glycan density of the surroundings. We detected that positions N234 and N717 are especially rigid. Of note, position N234 has been attributed a structural role (Casalino et al., 2020). Similarly, the large flexibility at N74 reflects the fact that this glycosylation position is located on a flexible loop.

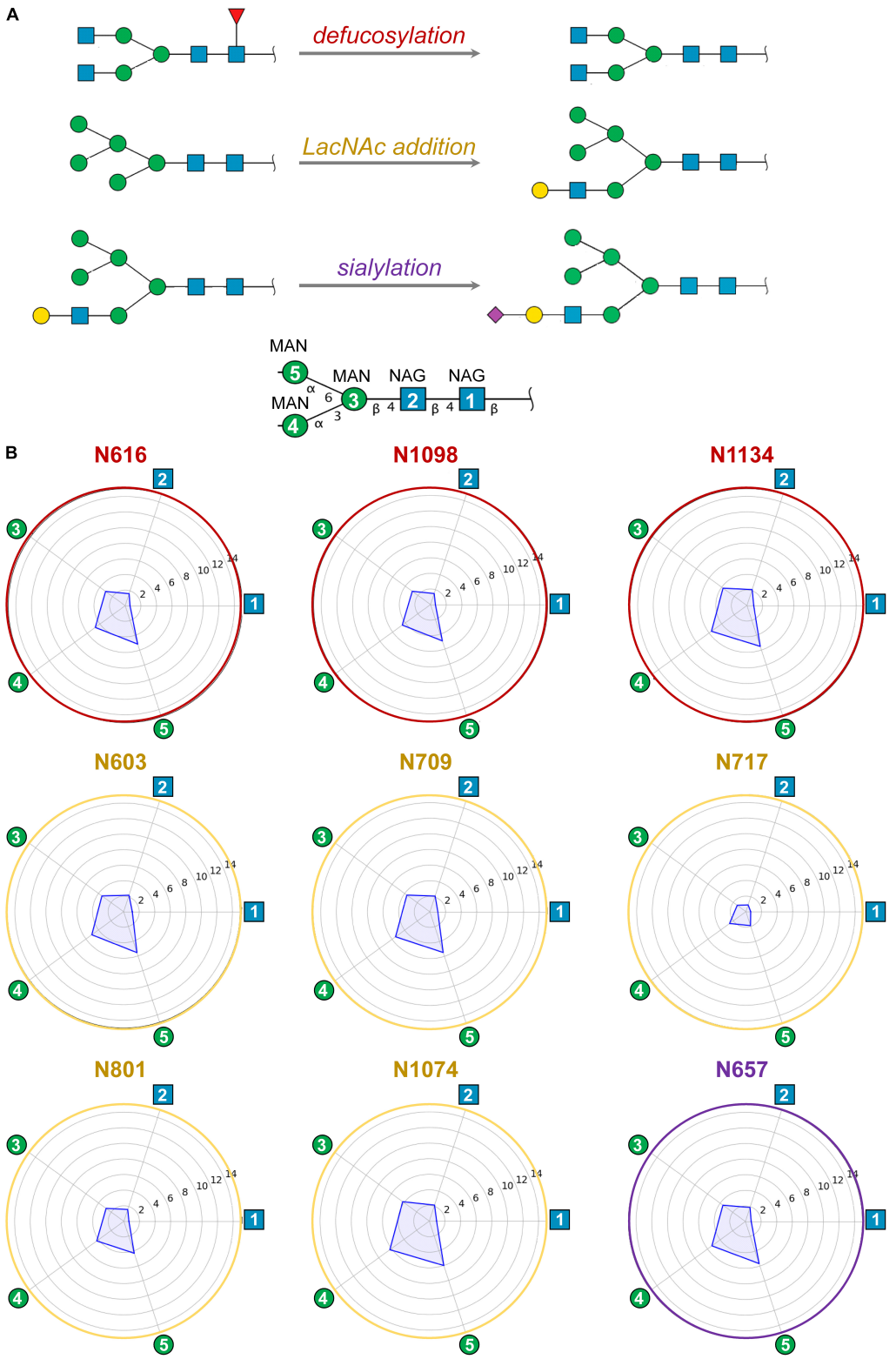


FIGURE 3 | (A) Modifications of selected ecto-S glycans with respect to model M0 analyzed by MD simulations: defucosylation (positions N616, N1098, and N1134; model M1), addition of terminal LacNAc (positions N603, N709, N717, N801, and N1074; model M2) and sialylation (position N657; model M3). **(B)** Computed mobility of glycans from MD simulations. For each glycosylation position, the mobility of core carbohydrates is computed as the atomic positional fluctuation (RMSF) in Å. The first two *N*-acetylglucosamines (NAG) and three mannoses (MAN) – i.e., the glycan core – have been considered as they are shared among all glycans. Atomic fluctuations are presented as radial plots. The mobility of the five carbohydrate units is represented anticlockwise starting from the first NAG.

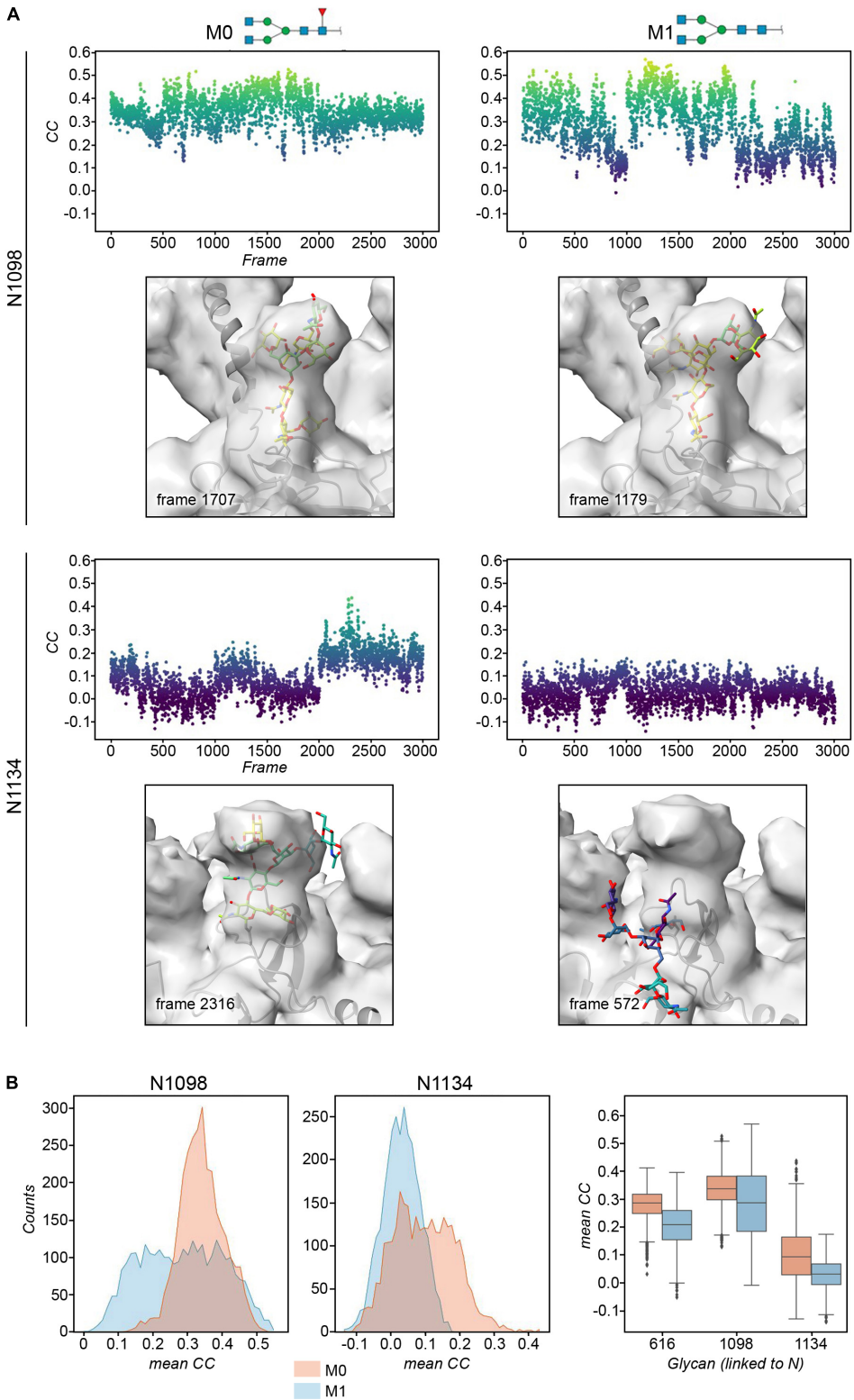


FIGURE 4 | (A) Plots of the real-space correlation (CC) of the glycan at site N1098 (*top*) and N1134 (*bottom*) across all the 3,000 frames (each monomer is considered independent) for models M0 (*left*) and M1 (*right*, defucosylated). Below the plot, the glycan structure with the highest real-space cross correlation value in the trajectory is shown within the cryo-EM map. **(B)** Histograms and box plots of the above cross correlation values for the glycans linked to N1098 and N1134 for the M0 and M1 constructs. See also **Supplementary Figures 8, 9**.

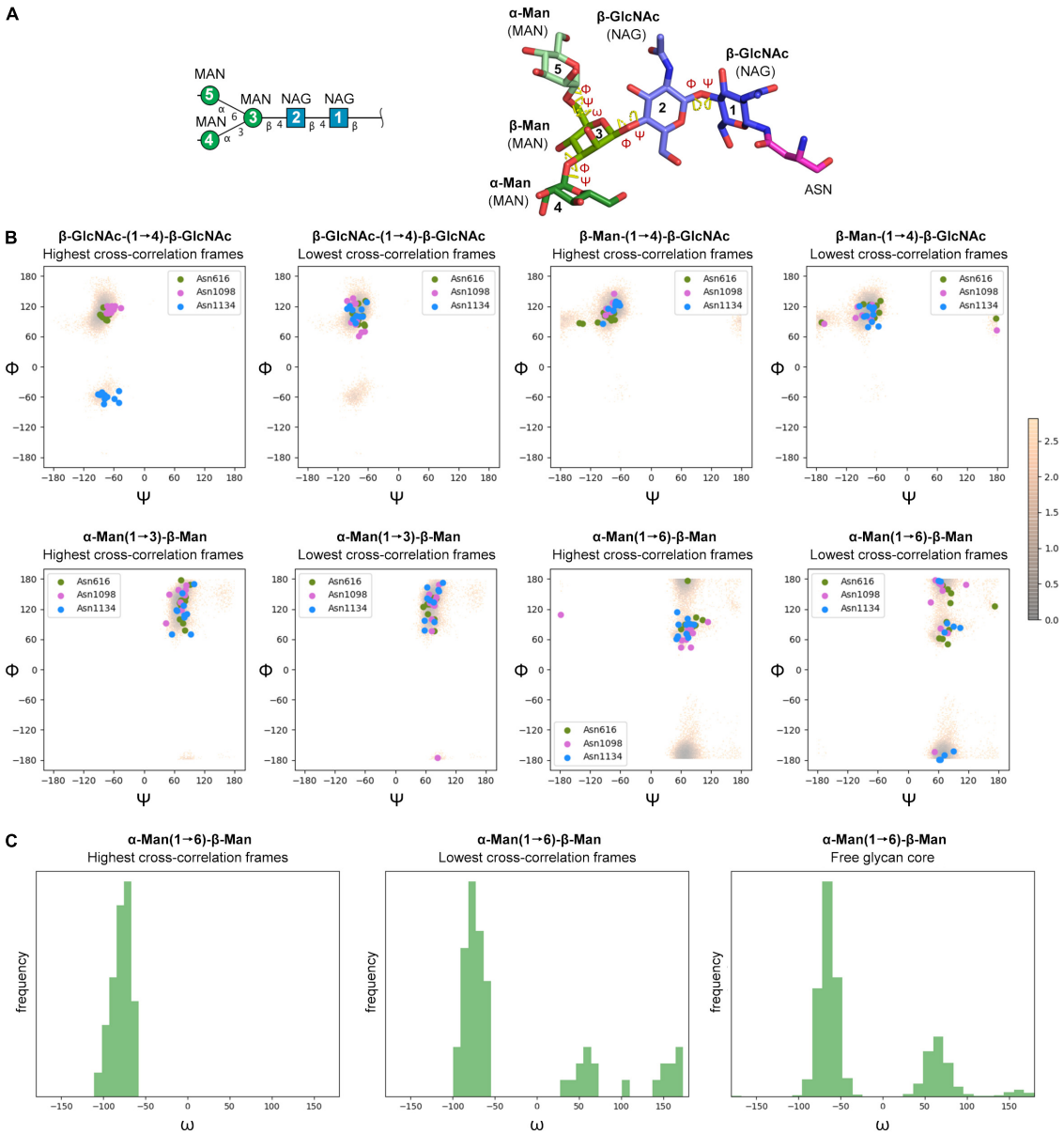


FIGURE 5 | (A) Symbol Nomenclature for Glycans (SNFG) and three-dimensional representations of the glycan core analyzed by MD simulations. **(B)** Φ/Ψ plots for the glycosidic bonds in the glycan core of selected glycosylation positions (N616, N1098, and N1134) corresponding to MD snapshots that scored the highest (n = 10) and the lowest (n = 10) real-space correlation (CC) with cryo-EM density (colored circles), and the free glycan core in solution (density map color-coded according to conformational energy in kcal mol^{-1} derived from population analysis though a Boltzmann distribution at 25°C). **(C)** Histograms for the ω dihedral angle in the branching mannose of the glycan core of selected glycosylation positions (N616, N1098, and N1134) corresponding to MD snapshots that scored the highest (n = 10) and the lowest (n = 10) real-space correlation (CC) with cryo-EM density, and the free glycan core in solution.

Reconstruction of Plausible Glycan Conformations From Cryo-EM and MD Simulations

The cryo-EM maps of the ecto-S unequivocally show that the density corresponding to the glycans of the S2 domain (N709, N717, N801, N1074, N1098, and N1134) is more visible than the others – most of their tubular shaped density projects outward radially from the protein backbone (**Figure 1A**). This density has been modeled with two GlcNAc and a mannose moiety in the different deposited atomic models and it was indeed inferred that these glycan subunits are quite rigid. Fitting in density of additional monosaccharides is challenging as the density becomes weaker as one moves away and yet in the 4.1 Å resolution map the density further protrudes for defined sites (**Supplementary Figure 3**). So, opposite to the real-space refinement procedure which optimizes the fitting of a "model" of a protein/glycan to an electron density map, we sought to

filter the glycan conformations explored by the MD simulations using the constraints offered by the experimental cryo-EM map to grasp possible glycan chemical signatures that would favor the adopted conformation in the map. This filtering was performed by extracting the individual snapshots from the M0, M1, M2, and M3 trajectories and estimating the real-space correlation between each carbohydrate conformation at a site of interest and the corresponding cryo-EM density by means of the EMDA software (see Methods; Warshamanage et al., 2022). We analyzed all nine glycosylation positions, grouping them according to the type of chemical modification. The real-space correlation at sites N1098 and N1134 (those sugars for which a stronger density is visible) is shown in **Figure 4**, while the corresponding analysis for the rest of the glycosylation positions is shown in **Supplementary Figures 8, 9**. For N1098 and N1134, of all the conformations explored only a small selection would fulfill the map. While general rules cannot be drawn it is clear that the presence of a fucose alters the mobility of the glycan. In the case of site N1098 the fucose provides a narrow distribution of poses around the highest CC (**Figure 4B**, left) while in the case of site N1134 it practically leads to a broad distribution of glycan conformations that, however, on average stays in density more than in the case of the defucosylated variant (**Figure 4B**, center and right). In the case of M2, the model-map correlation is also lower than for the original glycans in the M0 model while for the M3, the sialylation practically does not alter the average presence of the glycan in the map (**Supplementary Figures 8, 9**). We cannot exclude possible protein and glycan neighboring effects modulating the adopted conformation along the trajectories. In the case of M0 we also investigated those snapshots that scored the highest (n = 10) and the lowest (n = 10) real-space correlation and analyzed their glycosidic linkage torsions to assess whether low correlation can arise from violations of the canonical, low energy conformations of individual glycosidic linkages as a result of strong glycan-glycan or glycan-protein interactions (**Figure 5**). Results show no significant difference between lowest and highest cross-correlation frames in terms of the conformations assumed by the individual glycosidic linkages, that are all canonical, low energy ones. Thus, when enough density is available such as in positions N1098 and N1134, correlation with the cryo-EM map originates from the global glycan shape resulting from the combination of said linkage conformations, and the relative orientation of the glycan branches dictated by the geometry around the α-Man-(1→6)-β-Man bond, particularly the β-Man ω angle. The cryo-EM preferred value for ω is $\sim -60°$, corresponding to the most populated *gg* conformation for the free glycan core in solution (**Figure 5C**), which is therefore the value assumed by this angle in the pool of highest real-space correlation frames. The pool of lowest real-space correlation frames presents a population distribution among the three possible conformations (*gg*, *tg*, and *gt*), with an increased presence of the *tg* conformation that is virtually absent in the free glycan. Results for the other types of chemical modifications (**Supplementary Figures 10, 11**) are in line with those shown in **Figure 5**. When the glycans are less visible, the density does not carry information on the orientation of the branch connected to the α-Man-(1→6)-β-Man bond, and highest CC frames show a distribution of ω angles similar to that of the free glycan. Finally, the geometrical similarity of the highest and lowest cross-correlation (CC) frames was characterized by calculating the heavy-atom RMSD of the glycan core of each frame using as reference the global cross-correlation maximum (**Supplementary Figure 12**). The highest correlation frames lie in a narrow RMSD range over the global maximum, while the lowest correlation frames fall at larger (often > 10 Å) RMSD value, demonstrating that cross-correlation is a function of the glycan core geometry.

CONCLUSION

Glycan mobility has several implications in many physiological processes. In the attempt to reconcile the apparently more ordered glycans on the S2 domain of SARS-CoV-2 spike protein provided by the cryo-EM density with their dynamics derived from MD simulations, we investigated the glycan poses that best would fit the density and assessed the effect of sugar chemical modifications on the model-map correlation. While those poses that best fit the map are as energetically viable as those that do not fit the map, according to the geometry of the core glycosidic bonds, we noted that the best fits are characterized by sugars showing very similar and canonical geometries around the α-Man-(1→6)-β-Man bond. This observation may suggest that the best model-map fits may recapitulate those poses for which the movements of the most external carbohydrates are more geometrically restricted, and it may explain why some of those sugars are visible in the cryo-EM map beyond the fact that the proteinaceous region of attachment is rigid.

In conclusion, integrating static and dynamic views remains challenging particularly in the case of glycoproteins for which experimentally derived densities are difficult to interpret and rationalize and computer modeling is limited by insufficient conformational sampling. This study, however, shows that experimental and computational tools combined can provide valuable insights on the conformational preferences of inherently flexible and complex glycoconjugates.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

NA and GJ-O conceived the study. AM-C, DC, OM, CB, AP, and JE-O helped throughout the biochemical tasks. SS produced and purified biological and prepared sample for cryo-EM and together with SC and NA collected cryo-EM data. SS, SC, and NA processed the images and interpreted cryo-EM maps. FP and GJ-O performed MD simulations, and with AA and JJ-B interpreted MD results. SS, SC, FP, NA, and GJ-O analyzed and contextualized all data and wrote the manuscript. All authors approved the final findings presented.

FUNDING

This study was supported by the *Agencia Estatal de Investigacion* (Spain) for projects RTI2018-095700-B-I00 (NA), RTI2018-099592-B-C22 (GJ-O), CTQ2017-82222-R (SC), RTI2018-094751-B-C21 (JJ-B), RTI2018-101269-B-I00 (OM), PID2019-107956RA-I00 and RYC2018-024183-I (AP), by the Basque *Departamento de Educación, Política Lingüística y Cultura* for project PRE_2018_1_0102 (AM-C and NA), by the Basque Government COVID-19 CO-2020/00001 (NA, OM, and JJ-B) and by the European Research Council for projects ERC-2017-AdG, 788143-RECGLYCANMR (JJ-B) and ERC-2018-StG, 804236-NEXTGEN-IO (AP). FP thanks the Ministerio de Economía y Competitividad for a Juan de la Cierva Incorporación (IJC2020-045506-I) research contract. We thank MINECO for the Severo Ochoa Excellence Accreditation to the CIC bioGUNE (SEV-2016-0644). We acknowledge Diamond Light Source for access and support of the cryo-EM facilities at the United Kingdom's national Electron Bio-imaging Centre (eBIC) [under proposal BI23872 and COVID-19 Diamond Rapid Access Proposal], funded by the Wellcome Trust, MRC, and BBRSC.

ACKNOWLEDGMENTS

We thank the EM platform at the CIC bioGUNE for support in cryo-grid sample preparation and in-house cryo-EM imaging. We are also grateful to eBIC (Diamond Light Source, United Kingdom) staff for assistance during high-resolution cryo-EM data collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.870938/full#supplementary-material

REFERENCES

Allen, J. D., Chawla, H., Samsudin, F., Zuzic, L., Shivgan, A. T., Watanabe, Y., et al. (2021). Site-specific steric control of SARS-CoV-2 Spike Glycosylation. *Biochemistry* 60, 2153–2169. doi: 10.1021/acs.biochem.1c00279

Andersen, H. C. (1980). Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* 72, 2384–2393. doi: 10.1063/1.439486

Anggara, K., Zhu, Y., Fittolani, G., Yu, Y., Tyrikos-Ergas, T., Delbianco, M., et al. (2021). Identifying the origin of local flexibility in a carbohydrate polymer. *Proc. Natl. Acad. Sci.* 118:e2102168118. doi: 10.1073/pnas.2102168118

Atanasova, M., Bagdonas, H., and Agirre, J. (2020). Structural glycobiology in the age of electron cryo-microscopy. *Curr. Opin. Struct. Biol.* 62, 70–78. doi: 10.1016/j.sbi.2019.12.003

Bepler, T., Kelley, K., Noble, A. J., and Berger, B. (2020). Topaz-Denoise: general deep denoising models for cryoEM and cryoET. *Nat. Commun.* 11:5208. doi: 10.1038/s41467-020-18952-1

Cai, Y., Zhang, J., Xiao, T., Peng, H., Sterling, S. M., Walsh, R. M., et al. (2020). Distinct conformational states of SARS-CoV-2 spike protein. *Science* 369, 1586–1592. doi: 10.1126/science.abd4251

Casalino, L., Gaieb, Z., Goldsmith, J. A., Hjorth, C. K., Dommer, A. C., Harbison, A. M., et al. (2020). Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* 6, 1722–1734. doi: 10.1021/acscentsci.0c01056

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092. doi: 10.1063/1.464397

Götz, A. W., Clark, M. A., and Walker, R. C. (2014). An extensible interface for QM/MM molecular dynamics simulations with AMBER. *J. Comput. Chem.* 35, 95–108. doi: 10.1002/jcc.23444

Jo, S., Kim, T., Iyer, V. G., and Im, W. (2008). CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* 29, 1859–1865. doi: 10.1002/jcc.20945

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi: 10.1063/1.445869

Kapoor, K., Chen, T., and Tajkhorshid, E. (2021). Post-Translational Modifications Optimize the Ability of SARS-CoV-2 Spike for Effective Interaction with Host Cell Receptors. *bioRxiv* doi: 10.1101/2021.12.02.470852

Kaur, S., Gomez-Blanco, J., Khalifa, A. A. Z., Adinarayanan, S., Sanchez-Garcia, R., Wrapp, D., et al. (2021). Local computational methods to improve the interpretability and analysis of cryo-EM maps. *Nat. Commun.* 12:1240. doi: 10.1038/s41467-021-21509-5

Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., et al. (2008). GLYCAM06: A generalizable biomolecular force field. *Carbohyd. J. Comput. Chem.* 29, 622–655. doi: 10.1002/jcc.20820

Kunert, R., Rüker, F., and Katinger, H. (1998). Molecular Characterization of Five Neutralizing Anti-HIV Type 1 Antibodies: Identification of Nonconventional D Segments in the Human Monoclonal Antibodies 2G12 and 2F5. *AIDS Res. Hum. Retrovir.* 14, 1115–1128. doi: 10.1089/aid.1998.14.1115

Lempp, F. A., Soriaga, L. B., Montiel-Ruiz, M., Benigni, F., Noack, J., Park, Y.-J., et al. (2021). Lectins enhance SARS-CoV-2 infection and influence neutralizing antibodies. *Nature* 598, 342–347. doi: 10.1038/s41586-021-03925-1

Lenza, M. P., Oyenarte, I., Diercks, T., Quintana, J. I., Gimeno, A., Coelho, H., et al. (2020). Structural characterization of N-Linked glycans in the receptor binding domain of the SARS-CoV-2 Spike protein and their interactions with human lectins. *Angew. Chemie. Int. Ed.* 59, 23763–23771. doi: 10.1002/anie.202011015

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11, 3696–3713. doi: 10.1021/acs.jctc.5b00255

Marchal, I., Jarvis, D. L., Cacan, R., and Verbert, A. (2001). Glycoproteins from insect cells: sialylated or not? *Biol. Chem.* 382, 151–159. doi: 10.1515/BC.2001.023

Miyamoto, S., and Kollman, P. A. (1992). Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* 13, 952–962. doi: 10.1002/jcc.540130805

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., et al. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* 30, 70–82. doi: 10.1002/pro.3943

Punjani, A., Rubinstein, J. L., Fleet, D. J., and Brubaker, M. A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* 14, 290–296. doi: 10.1038/nmeth.4169

Punjani, A., Zhang, H., and Fleet, D. J. (2020). Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* 17, 1214–1221. doi: 10.1038/s41592-020-00990-8

Rapp, M., Shapiro, L., and Frank, J. (2022). Contributions of single-particle cryoelectron microscopy toward fighting COVID-19. *Trends Biochem. Sci.* 47, 117–123. doi: 10.1016/j.tibs.2021.10.005

Roe, D. R., and Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* 9, 3084–3095. doi: 10.1021/ct400341p

Rohou, A., and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* 192, 216–221. doi: 10.1016/j.jsb.2015.08.008

Sanchez-Garcia, R., Gomez-Blanco, J., Cuervo, A., Carazo, J. M., Sorzano, C. O. S., and Vargas, J. (2021). DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun. Biol.* 4, 874. doi: 10.1038/s42003-021-02399-1

Scheres, S. H. W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* 180, 519–530. doi: 10.1016/j.jsb.2012.09.006

Sikora, M., von Bülow, S., Blanc, F. E. C., Gecht, M., Covino, R., and Hummer, G. (2021). Computational epitope map of SARS-CoV-2 spike protein. *PLoS Comput. Biol.* 17:e1008790. doi: 10.1371/journal.pcbi.1008790

Stadlbauer, D., Amanat, F., Chromikova, V., Jiang, K., Strohmeier, S., Arunkumar, G. A., et al. (2020). SARS-CoV-2 seroconversion in humans: a detailed protocol for a serological assay, antigen production, and test setup. *Curr. Protoc. Microbiol.* 57:e100. doi: 10.1002/cpmc.100

Stuart, D. I., Subramaniam, S., and Abrescia, N. G. A. (2016). The democratization of cryo-EM. *Nat. Methods* 13, 607–608. doi: 10.1038/nmeth.3946

Subramaniam, S. (2020). COVID-19 and cryo-EM. *IUCRJ* 7, 575–576. doi: 10.1107/S2052252520008799

Sztain, T., Ahn, S.-H., Bogetti, A. T., Casalino, L., Goldsmith, J. A., Seitz, E., et al. (2021). A glycan gate controls opening of the SARS-CoV-2 spike protein. *Nat. Chem.* 13, 963–968. doi: 10.1038/s41557-021-00758-3

Tian, Y., Parsons, L. M., Jankowska, E., and Cipollo, J. F. (2021). Site-specific glycosylation patterns of the SARS-CoV-2 spike protein derived from recombinant protein and viral WA1 and D614G Strains. *Front. Chem.* 9:76744. doi: 10.3389/fchem.2021.767448

Trkola, A., Purtscher, M., Muster, T., Ballaun, C., Buchacher, A., Sullivan, N., et al. (1996). Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *J. Virol.* 70, 1100–1108. doi: 10.1128/jvi.70.2.1100-1108.1996

Wagner, T., Merino, F., Stabrin, M., Moriya, T., Antoni, C., Apelbaum, A., et al. (2019). SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* 2:218. doi: 10.1038/s42003-019-0437-z

Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 181, 281—-292. doi: 10.1016/j.cell.2020.02.058

Wang, Y., Wu, Z., Hu, W., Hao, P., and Yang, S. (2021). Impact of expressing cells on glycosylation and glycan of the SARS-CoV-2 Spike Glycoprotein. *ACS Omega* 6, 15988–15999. doi: 10.1021/acsomega.1c01785

Warshamanage, R., Yamashita, K., and Murshudov, G. N. (2022). EMDA: a python package for electron microscopy data analysis. *J. Struct. Biol.* 214:107826. doi: 10.1016/j.jsb.2021.107826

Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S., and Crispin, M. (2020). Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* 369, 330–333. doi: 10.1126/science.abb9983

Watanabe, Y., Mendonça, L., Allen, E. R., Howe, A., Lee, M., Allen, J. D., et al. (2021). Native-like SARS-CoV-2 Spike Glycoprotein Expressed by ChAdOx1 nCoV-19/AZD1222 Vaccine. *ACS Cent. Sci.* 7, 594–602. doi: 10.1021/acscentsci.1c00080

Woo, H., Park, S.-J., Choi, Y. K., Park, T., Tanveer, M., Cao, Y., et al. (2020). Developing a fully glycosylated full-length SARS-CoV-2 spike protein model in a viral membrane. *J. Phys. Chem. B* 124, 7128–7137. doi: 10.1021/acs.jpcb.0c04553

Woods, R. J. (2005). *GLYCAM Web. Complex Carbohydrate Research Center*. Athens, GA: University of Georgia.

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., et al. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367, 1260–1263.

Zheng, S. Q., Palovcak, E., Armache, J.-P., Verba, K. A., Cheng, Y., and Agard, D. A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* 14, 331–332. doi: 10.1038/nmeth.4193

Zhou, T., Tsybovsky, Y., Gorman, J., Rapp, M., Cerutti, G., Chuang, G.-Y., et al. (2020). Cryo-EM Structures of SARS-CoV-2 Spike without and with ACE2 Reveal a pH-dependent switch to mediate endosomal positioning of receptor-binding domains. *Cell Host Microb.* 28, 867–879. doi: 10.1016/j.chom.2020.11.004

Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J. H., Lindahl, E., et al. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* 7:e42166. doi: 10.7554/eLife.42166

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

PERSPECTIVE
published: 15 April 2022
doi: 10.3389/fmicb.2022.842976

Deep Learning-Powered Prediction of Human-Virus Protein-Protein Interactions

*Xiaodi Yang[1], Shiping Yang[2], Panyu Ren[1], Stefan Wuchty[3,4,5] and Ziding Zhang[1]**

[1] State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing, China, [2] State Key Laboratory of Plant Physiology and Biochemistry, College of Biological Sciences, China Agricultural University, Beijing, China, [3] Department of Computer Science, University of Miami, Miami, FL, United States, [4] Department of Biology, University of Miami, Miami, FL, United States, [5] Sylvester Comprehensive Cancer Center, University of Miami, Miami, FL, United States

Identifying human-virus protein-protein interactions (PPIs) is an essential step for understanding viral infection mechanisms and antiviral response of the human host. Recent advances in high-throughput experimental techniques enable the significant accumulation of human-virus PPI data, which have further fueled the development of machine learning-based human-virus PPI prediction methods. Emerging as a very promising method to predict human-virus PPIs, deep learning shows the powerful ability to integrate large-scale datasets, learn complex sequence-structure relationships of proteins and convert the learned patterns into final prediction models with high accuracy. Focusing on the recent progresses of deep learning-powered human-virus PPI predictions, we review technical details of these newly developed methods, including dataset preparation, deep learning architectures, feature engineering, and performance assessment. Moreover, we discuss the current challenges and potential solutions and provide future perspectives of human-virus PPI prediction in the coming post-AlphaFold2 era.

Keywords: human-virus protein-protein interactions, machine learning, deep learning, transfer learning, prediction

OPEN ACCESS

Edited by:
Gorka Lasso Cabrera, Albert Einstein College of Medicine, United States

Reviewed by:
Mu Gao, Georgia Institute of Technology, United States
Stéphane Téletchéa, Université de Nantes, France

****Correspondence:***
Ziding Zhang
zidingzhang@cau.edu.cn

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *24 December 2021*
Accepted: *25 March 2022*
Published: *15 April 2022*

Citation:
Yang X, Yang S, Ren P, Wuchty S and Zhang Z (2022) Deep Learning-Powered Prediction of Human-Virus Protein-Protein Interactions. Front. Microbiol. 13:842976. doi: 10.3389/fmicb.2022.842976

INTRODUCTION

Currently, viral infection is a major factor threatening human health and global economic development (Qiu et al., 2017; Rasul, 2020; Lu and Peng, 2021). For instance, the current pandemic disease of novel coronavirus pneumonia, induced by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has caused nearly 280 million confirmed cases and more than 5 million deaths worldwide by the end of 2021.[1] Viruses invade host cells and complete their own life cycle by exploiting the host's molecular machinery, which is largely determined by virus-host protein-protein interactions (PPIs) (Jean Beltran et al., 2017). Therefore, systematic characterization of human-virus protein interactions can help to decipher viral infection mechanisms and provide new leads for antiviral drug discovery and vaccine development. Experimental techniques [e.g., yeast two-hybrid (Y2H) assays (Calderwood et al., 2007; Tripathi et al., 2010; Rozenblatt-Rosen et al., 2012) and affinity purification coupled with mass spectrometry (AP-MS) (Shah et al., 2018; Gordon et al., 2020; Li et al., 2021; Stukalov et al., 2021)] have determined a great amount of human-virus protein interactions. Despite such tremendous progress

[1]https://covid19.who.int/

in the last decades, human-virus interactomes are still far from complete, while existing interaction data usually focus on some well-studied virus species (Lian et al., 2021).

To complement experimental methods, many computational methods have been developed to automatically predict PPIs between human host and various viruses. Existing prediction methods include interolog mapping (Yu et al., 2004; Yang et al., 2021a), domain-domain/motif interaction-based inference (Dyer et al., 2007; Evans et al., 2009; Chiang et al., 2017; Zhang et al., 2017), structure-informed method (de Chassey et al., 2013; Lasso et al., 2019) and machine learning (ML)-based prediction (Dyer et al., 2011; Barman et al., 2014; Yang et al., 2020). For more information on these computational methods, see the reviews (Mariano and Wuchty, 2017; Lian et al., 2021). With the accumulation of experimental PPIs, ML-based methods have been increasingly popular to predict human-virus PPIs. Briefly, ML-based methods train a binary classifier using known human-virus PPI data to predict interacting protein pairs from query samples. Traditional ML methods, such as support vector machines and random forests, have been used extensively and achieved reasonable performance (Emamjomeh et al., 2014). As an important branch of ML, deep learning (DL) has been successfully applied to predict intra-species protein interactions (Du et al., 2017; Hashemifar et al., 2018; Li et al., 2018; Chen et al., 2019). Very recently, several DL architectures have been developed to predict human-virus PPIs with favorable performance compared to traditional ML methods (Lanchantin et al., 2021; Liu-Wei et al., 2021; Tsukiyama et al., 2021; Yang et al., 2021b). In this review, we provide an overview of dataset construction, model architectures, feature engineering and performance assessment of DL in human-virus PPI identification (**Figure 1A**). In particular, we also discuss the technical challenges and future directions of this exciting topic in the coming era of post-AlphaFold2 (Jumper et al., 2021).

DATASET CONSTRUCTION OF HUMAN-VIRUS PROTEIN-PROTEIN INTERACTION PREDICTION

Positive Sample Selection and Filtering

The construction of training/test datasets, including positive and negative samples, is the first important step in developing a DL-based predictor. Generally, positive samples are experimentally determined human-virus PPIs, which can be collected from public database resources such as HPIDB (Ammari et al., 2016) and HVIDB (Yang et al., 2021a), or directly adopted from literature. Considering that experimental results may contain false positives, the obtained positive data should be further filtered according to various strategies. Both LSTM-PHV (Tsukiyama et al., 2021) and DeepViral (Liu-Wei et al., 2021) downloaded human-virus PPIs from HPIDB and only retained interactions with a significant MI score (a confidence score of molecular interactions) (Villaveces et al., 2015). In our previous works [i.e., TransPPI (Yang et al., 2021b) and doc2vec + RF (Yang et al., 2020)], we excluded interactions from large-scale MS experiments that have been experimentally detected only once to obtain a high-quality positive dataset. DeepVHPPI (Lanchantin et al., 2021) directly used the compiled dataset of our previous doc2vec + RF method (Yang et al., 2020). Still, the selection of high-confidence interactions is usually met with a tradeoff strategy between training data set size and quality as a perfect scoring system for assessing the reliability of experimental human-virus PPIs is still not available. While large known virus-host PPI data allow us to filter interactions with strict criteria, we can only adopt loose filtering criteria when only scarce interaction data are available to ensure that the retaining data size and quality are sufficient for training.

Negative Sampling

In the absence of a gold standard for negative sample selection, random sampling is probably the most commonly used method (Dyer et al., 2011; Barman et al., 2014). For example, DeepViral randomly samples pairs of human and viral proteins that do not occur in the positive dataset (Liu-Wei et al., 2021). However, random sampling may inevitably introduce false-negative data points in the compiled training sets, prompting the development of a different negative sampling method called "Dissimilarity-Based Negative Sampling" (Eid et al., 2016; Yang et al., 2020, 2021a,b; Tsukiyama et al., 2021). The core idea is that if a viral protein *A* is similar to a viral protein *B* that interacts with human protein *C* (i.e., *B*-*C* is a positive sample), then the virus-host protein pair *A*-*C* cannot be a negative sample.

Another open issue related to negative sampling is the ratio of positive to negative samples. Often, a simple balanced ratio (i.e., 1:1) is used for many prediction tasks. However, it will cause the overestimation of model performance if the number of negative samples is obviously larger than that of positive samples in the real world (e.g., the issue of PPI prediction). An extremely unbalanced ratio will also yield biased results by over-predicting false negatives since negative samples are over-represented in the training set. Although a perfect solution for the ratio of positive-to-negative does not exist, an imbalanced ratio (e.g., 1:10) has been proven reasonable to predict human-virus PPIs (Yang et al., 2020; Liu-Wei et al., 2021; Tsukiyama et al., 2021).

DEEP LEARNING IN HUMAN-VIRUS PROTEIN-PROTEIN INTERACTION PREDICTION

Classification Model Construction Through Supervised Deep Learning

In contrast to traditional ML methods, DL approaches are flexible in allowing the known labels to relate to the input feature vectors (Wainberg et al., 2018). However, the large number of trainable parameters in DL creates more challenges to avoid model overfitting (i.e., lose the generalization to new data) compared to traditional ML techniques. To deal with this issue, early stopping mechanisms by monitoring loss on the training and validation sets, regularization of the model, or dropout techniques are often adopted. As flexible architectures are a main feature of DL approaches, some dominant DL architectures such as convolutional neural network (CNN), recurrent neural

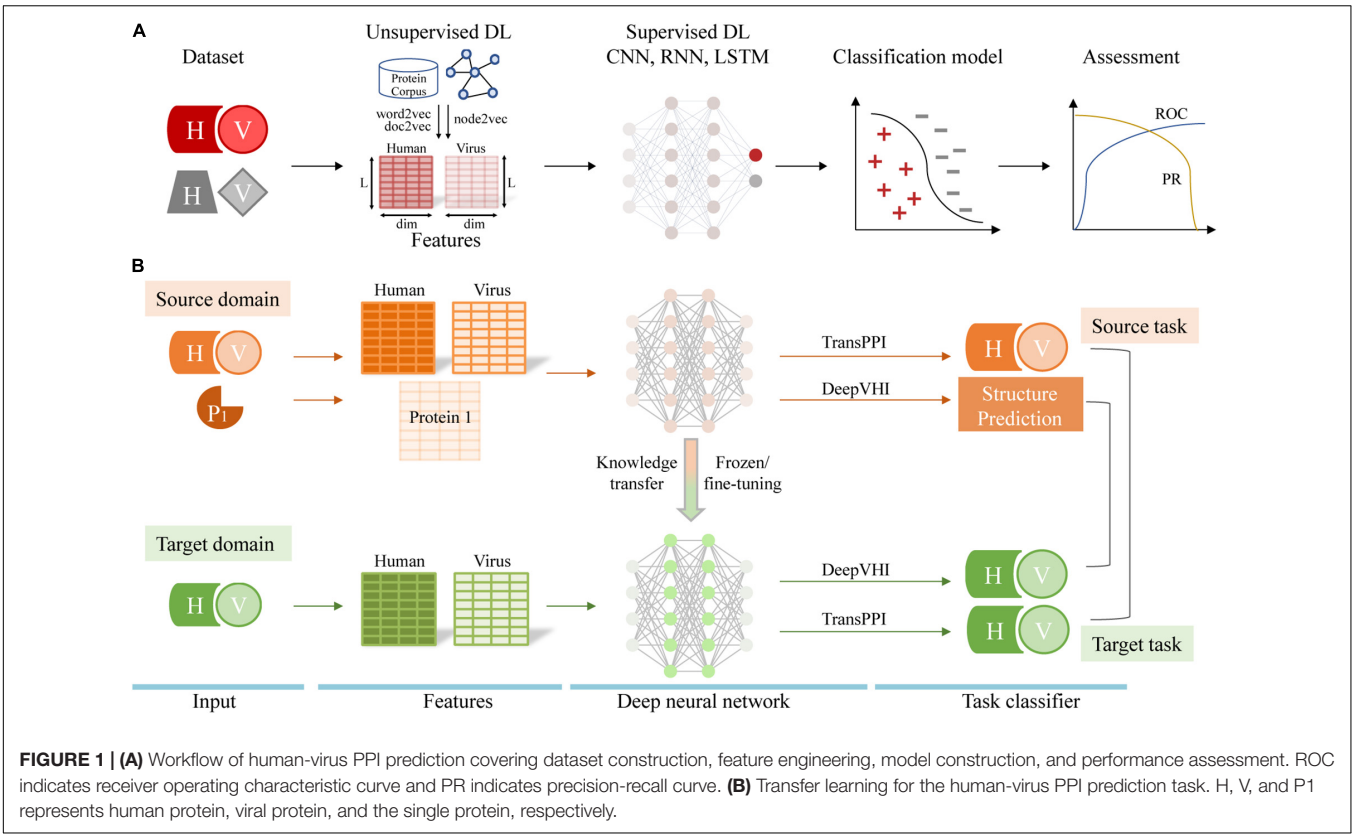


FIGURE 1 | (A) Workflow of human-virus PPI prediction covering dataset construction, feature engineering, model construction, and performance assessment. ROC indicates receiver operating characteristic curve and PR indicates precision-recall curve. **(B)** Transfer learning for the human-virus PPI prediction task. H, V, and P1 represents human protein, viral protein, and the single protein, respectively.

network (RNN), long-short term memory (LSTM) have been used to predict human-virus protein interactions (**Table 1**). Such DL architectures can be considered feature extractors, which usually connect fully connected layers–also called Multi-layer perceptron (MLP)–to provide end-to-end binary classifiers for PPI prediction. After such supervised learning steps, trained models can be used to predict interactions from query human-virus protein pairs.

Convolutional Neural Networks

Deep neural networks with one or more convolutional and pooling layers (i.e., CNNs) are usually applied to process image data to capture local pixelated features (Krizhevsky et al., 2012). In recent years, CNNs have been widely used to capture protein features in bioinformatics studies (Hashemifar et al., 2018), allowing the effective detection of local motif features of proteins that mediate protein interactions while following pooling layers reduce the dimensions of feature maps. Our previous work applied a sequence-based siamese one-dimensional (1D) CNN architecture to train a human-virus PPI classifier and achieved better performance than traditional ML methods especially in relatively large datasets (Yang et al., 2021b). In particular, we employed the siamese network (Bromley et al., 1993) to learn complex interaction relationships between human and viral proteins. The core idea of the siamese network is parameter sharing between two identical subnetworks (i.e., the human and virus protein input subnetworks) that can effectively capture the mutual influence of protein pairs (Chen et al., 2019). Liu-Wei et al. (2021) employed 16 1D-convolutional layers with a pooling layer and several dense layers to predict human-virus PPIs. Moreover, Lanchantin et al. (2021) applied a convolutional layer with multiple convolutional filters for neural network training. Different architectures of CNNs in these publications further demonstrate the flexibility of DL.

Recurrent Neural Networks and Long Short-Term Memory

The main application of RNNs is in natural language processing, such as machine translation (Sutskever et al., 2014) and speech recognition (Graves et al., 2013). In particular, recurrent layers allow the handling and integration of complex long-range sequential information. Like convolutional layers, recurrent layers also scan the input sequential data element by element but preserve previous output value (i.e., a memory of the earlier state) that are combined with the current input value to output a value of the current state. RNNs are useful to convert variable-length data to fixed-size representations as the inputs to the next fully connected layers for prediction tasks (Greener et al., 2022). In particular, the more advanced bidirectional gated recurrent unit (GRU) variant of RNNs has been used to predict intraspecies protein interactions, showing excellent performance in combination with a CNN (RCNN) (Chen et al., 2019). Yet, this deep learning framework did not allow more favorable predictions of human-virus PPIs compared to simple CNNs (Liu-Wei et al., 2021). Gradient explosion and disappearance will occur when RNNs propagate backward since there are long-term

TABLE 1 | Existing deep learning prediction methods of human-virus PPIs.

Method	Virus species	Input information	Embedding approach	Model architecture	Number of positive/negative samples	Negative sampling	URL
TransPPI (Yang et al., 2021b)	Multiple viruses	Protein sequences	PSSM	CNN + MLP + transfer learning	31,381/313,810	Dissimilarity-based negative sampling	https://github.com/XiaodiYangCAU/TransPPI/
DeepViral (Liu-Wei et al., 2021)	14 viral families	Protein sequences, functions, and disease phenotypes	one-hot and node2vec	CNN + MLP	24,678/246,780	Random sampling	https://github.com/bio-ontology-research-group/DeepViral/
LSTM-PHV (Tsukiyama et al., 2021)	All viruses	Protein sequences	word2vec	LSTM + MLP	22,383/223,830	Dissimilarity-based negative sampling	http://kurata35.bio.kyutech.ac.jp/LSTM-PHV/
DeepVHPPI (Lanchantin et al., 2021)	Multiple viruses	Protein sequences	one-hot	CNN + MLP + transfer learning	22,653/226,530	Dissimilarity-based negative sampling	https://github.com/QData/DeepVHPPI/
MTT (Dong et al., 2021)	Multiple viruses	Protein sequences	mLSTM	MLP + transfer learning	Multiple settings	Multiple settings	https://git.l3s.uni-hannover.de/dong/multitask-transfer/

dependencies over the sequential series (Sun et al., 2020). As an advanced architecture of RNN, LSTM introduces the concept of cells and gates (an input gate, an output gate and a key forget gate) (Gers et al., 2000). LSTM cells can store long-term information while these gates regulate the information into cells. Recently, Tsukiyama et al. (2021) employed two LSTM subnetworks to transform the human and viral proteins-embedding matrixes into two fixed-length vectors as the input to subsequent fully connected layers to predict human-virus PPIs. The LSTM architecture mitigates the gradient explosion and disappearance problems of RNNs, effectively preserving long-term memory information of protein sequences.

Feature Engineering in Deep Learning

Protein feature vectors used in DL models are often inferred from protein sequences, including simple residue position information, physicochemical properties, and evolutionary information of residues, such as one-hot encoding and position-specific scoring matrix (PSSM) (**Table 1**). Briefly, the one-hot method encodes each amino acid as a vector of length n that corresponds to the set of amino acid categories, allowing us to represent a protein sequence of length L as a $L \times n$ matrix with 0 and 1 entries. As a more fine-grained method to present protein features PSSMs capture evolutionary relationships between proteins. In particular, each amino acid (n) in the protein sequence of length L has a specific score, allowing an alternative representation of a protein sequence as a $L \times n$ matrix. Furthermore, some word embedding techniques from natural language processing have been adapted to represent proteins, which can automatically convert k-mer amino acids or proteins to fixed-dimensional feature vectors. Here, we mainly focus on these embedding techniques and their applications in the DL-based prediction of human-virus PPIs.

Word2vec and Doc2vec

Word2vec is a word embedding technique derived from natural language processing to obtain distributed representations of words through model training. Word2vec uses two-layer shallow neural networks to obtain feature vectors of words by using linguistic contexts, where two architecture choices including continuous bag-of-words (CBOW) and skip-gram (Le and Mikolov, 2014; Kimothi et al., 2016) are often used. Briefly, the CBOW model predicts the current word by using surrounding context words while skip-gram uses the current word to predict the surrounding words. In particular, a textual corpus is generally used to train the word2vec model to assign fixed-dimensional vectors to words, enabling that the words sharing common contexts and semantics in the training corpus are embedded close to each other (Kimothi et al., 2016). Such an embedding approach has been applied to represent protein sequences in several bioinformatics tasks. For instance, the unsupervised word2vec model trained from a corpus containing non-redundant proteins in the Swiss-Prot database and the resulting feature vectors of human and viral proteins were further used to train human-virus PPI prediction models (Tsukiyama et al., 2021). In this work, k-mers (i.e., k consecutive residues) in each sequence were regarded as single words, representing each protein sequence through multiple k-mers. The authors employed the CBOW architecture to train the word2vec model and optimally set k to 4. As a result, 128-dimensional embedding vectors for multiple k-mers were retrieved and further concatenated to obtain embedding feature matrixes of proteins. Additionally, domains or motifs in proteins can also be treated as words in documents. Similar to the word2vec model, protein sequences can therefore be represented by feature vectors based on their domains or motifs (i.e., domain or motif embeddings). In Pan et al. (2021), the authors employed the skip-gram model to pre-train domain embeddings and averaged multiple domain embeddings in a protein sequence to construct the corresponding protein feature vector. The resulting protein feature vectors were further used to predict protein toxicity. Considering that human-virus protein interactions are generally mediated by domain-domain/motif interactions, the feature representation strategy of domain/motif embeddings should be informative in predicting human-virus PPIs.

As an extension of word2vec, doc2vec adds the whole document as another word. Doc2vec considers the context information of words and the whole document. When applied

to protein sequences, each sequence is regarded as a document, in which *k*-mers are defined as the corresponding words (Yang et al., 2018). Subsequently, the doc2vec model is trained to learn the feature vector representation of each protein sequence in the corpus by using similar model architectures in word2vec. In our previous work (Yang et al., 2020), we successfully employed the doc2vec model to pre-train the embeddings of proteins based on the Swiss-Prot corpus. We further used the obtained low-dimensional feature vectors of human and viral proteins as input to train an RF classifier to predict human-virus PPIs (i.e., doc2vec + RF) and achieved better performance than other sequence-based traditional ML algorithms.

Node2vec

Graphs, also known as networks, have been widely used to represent biological entries (i.e., nodes) and their relations (i.e., edges). A series of graph embedding methods have been developed to automatically learn low-dimensional feature representation for each node in the graph (Grover and Leskovec, 2016; Ou et al., 2016; Wang et al., 2016). Such low-dimensional feature representations preserve the network structure information of the graph, which can be employed to train ML models to tackle node classification or link prediction problems (Yue et al., 2020). As one of the most commonly used graph embedding methods, node2vec firstly adopts a flexible random walk process to generate node sequences (multiple word lists), which are subsequently fed to the word2vec model to obtain node embedding features (i.e., node representations) (Grover and Leskovec, 2016). In the field of bioinformatics, node2vec is often used in node classification tasks such as identifying essential proteins based on a PPI network (Zeng et al., 2019) and detecting tissue-specific cellular functions through multi-layer PPI networks (Zitnik and Leskovec, 2017). Additionally, node2vec has been employed to obtain protein features based on the network consisting of proteins, Gene Ontology (GO) terms, and their associations called GO2Vec (Zhong et al., 2019). Further, these network embeddings were used to predict protein interactions (Zhong and Rajapakse, 2020; Liu-Wei et al., 2021). In particular, Liu-Wei et al. (2021) employed their DL2Vec method (node2vec variant) to embed human and viral proteins by using GO and cross-species phenotype ontology annotations. Such embeddings were then used as input to train a neural network, allowing the reliable prediction of human-virus PPIs, suggesting that node embedding is informative in recognizing human-virus PPIs.

MODEL GENERALIZATION THROUGH TRANSFER LEARNING

Since data available for training human-virus PPI prediction models of novel or rarely investigated virus species are often limited, the lack of sufficient labeled data is a major obstacle to ML-based PPI identification. Transfer learning is a good solution for processing relatively scarce data and improving prediction performance. The core idea of transfer learning is to leverage informative prior knowledge learning from other related tasks to enable learning of a target task with small-scale data. In the context of DL, deep transfer learning is becoming a promising method in generalizing a DL-based human-virus PPI prediction model.

Our recent work, TransPPI, employed two transfer learning approaches to accurately predict human-virus PPIs (Yang et al., 2021b). Specifically, we trained a CNN (i.e., the feature extractor) as well as fully connected layers (i.e., the MLP classifier) with multiple large-scale human-virus PPI datasets. In the next step, we retrained the model on the target human-virus PPI dataset through two types of transfer learning. (i) In the "frozen" approach, we kept learned parameters of CNN layers unchanged and retrained MLP layers with a target dataset; (ii) In the "fine-tuning" approach we retrained both CNN parameters and MLP layers with a target dataset. In general, the above transfer learning strategies effectively utilized prior knowledge from a "source" (e.g., human-HIV PPIs) to train in a target task domain (e.g., human-SARS-CoV-2 PPIs), allowing us to improve the performance and generalization of models based on small-scale data (**Figure 1B**).

In a different approach, Lanchantin et al. (2021) adopted a new transfer learning strategy to predict human-virus PPIs for a novel virus without any experimental known interaction data. The proposed architecture called DeepVHPPI first pre-trained supervised structure prediction (i.e., secondary structure prediction, residue contact prediction and remote homology detection) models as source tasks. Then, their approach fine-tunes the entire neural network on the target task (human-virus PPI prediction) by transferring information from source tasks (**Figure 1B**). Finally, DeepVHPPI showed promising prediction performance when determining interactions with human-SARS-CoV-2. While the underlying principle is based on the assumption that both source and target learning objectives share similar statistical properties, allowing to share similar model parameters, the transfer learning strategies of TransPPI and DeepVHPPI are different. In particular, TransPPI transfers model parameters that were learned from a source, large-scale human-virus PPI data set to predict interactions in a different target human-virus setting. In comparison, DeepVHPPI transfers feature representations of protein structures that were learned from a source data set to predict human-virus PPIs, assuming that the sequence-structure relationship of interacting proteins is similar, regardless of the considered organisms.

Dong et al. (2021) employed a multi-task transfer learning method called MultiTask Transfer (MTT) to construct human-virus PPI prediction model for novel viruses. Using a pre-trained UNIREP model (Alley et al., 2019) based on multiplicative LSTM (mLSTM) human and viral protein embeddings were obtained to predict human-specific and human-virus PPIs based on known PPI data from various benchmark datasets. In particular, such an approach makes the implicit assumption that the underlying statistical characteristics of amino acid composition of interacting proteins are generally similar. Although viral proteins try to mimick human interaction partners to bind to a specific host protein (Mariano and Wuchty, 2017), the

number of human interaction partners a virus usually interacts with is rather limited. As human interaction partners hardly cover the whole human proteome, human PPIs potentially introduce a training bias, overpowering the specificity of human-virus interactions.

STRENGTH AND WEAKNESS OF EXISTING DEEP LEARNING-BASED HUMAN-VIRUS PROTEIN-PROTEIN INTERACTION PREDICTION METHODS

In **Table 1**, we summarize recently developed deep learning-based human-virus PPI prediction methods (see Sections "Dataset Construction of Human-Virus Protein-Protein Interaction Prediction" to "Model Generalization Through Transfer Learning" for details of methods) to further analyze the strengths and weaknesses of these methods. LSTM-PHV employed word2vec + LSTM + MLP framework to train the human-virus PPI prediction model, where word2vec effectively captures context semantic information of *k*-mer amino acids. Furthermore, LSTM mitigates the explosion and disappearance of gradients in RNNs, enabling long-range sequential learning. Notably, other methods mainly adopt CNNs in their model architecture, better capturing local features of protein sequences, such as linear binding motifs that mediate human-virus PPIs compared to RNN/LSTM-based methods.

As the main innovation, DeepViral learns protein representations, that account for GO and disease phenotype ontology information as additional features to simple sequence information using a node2vec approach. Although such an approach allows a better representation of proteins compared to a simple one-hot sequence representation, this feature encoding method comparatively relies on functional and disease phenotype data of human and viral proteins. Such a dependence on auxiliary data may be limiting the method applicability to host-virus domains where virus specific information is missing.

The highlight of TransPPI and DeepVHPPI is the application of transfer learning techniques, that can improve model performance and generalization ability when available training data of novel or rarely investigated virus species are limited. In contrast to DeepVHPPI, TransPPI taps similarities of sequence composition of interacting human and viral proteins, potentially leading to better prediction performance. In particular, DeepVHPPI trains on a human-all virus PPI set, which is finally used to predict human-specific virus PPIs. In contrast, TransPPI requires that the target virus species has a small number of known human-virus PPI data. DeepVHPPI does not have this requirement, making this approach applicable to host-virus pairs where no experimental data is available. Another transfer learning method MTT mainly employs a multi-task learning strategy by considering human-specific PPIs as well. While such auxiliary training data improves model generalizability, such PPIs also introduce host-specific interaction characteristics that may impair the specificity to detect host-virus interactions.

DISCUSSION

Deep learning is playing an increasingly important role in human-virus PPI prediction. Although existing DL methods have outperformed traditional ML methods in predicting human-virus PPIs, much room for improvement remains. First, more DL architectures and feature representations should be used. The optimal combination of the DL architecture and feature engineering should be sought to maximize prediction performance. Existing DL methods may supplement previous human-virus PPI prediction methods. Thus, the integration of different prediction methods can often result in a more accurate and robust predictor. Moreover, model interpretability received a wide concern for ML-based methods. Usually, the way DL architectures end up with their predictions and predictive features are unknown, prompting the call for more explainable DL methods. In some bioinformatics tasks (Pan et al., 2021; Zhu et al., 2021), the prediction models have generally been simply interpreted by using t-distributed stochastic neighbor embedding (t-SNE) (van der Maaten and Hinton, 2008) to visualize the learned high-dimensional feature representations in 2D space. Note, that such t-SNE-based visualization can merely demonstrate the general effectiveness of the feature embedding methods, while the contributive features are not highlighted. Recently, attention mechanisms have provided a new direction for interpreting black-box DL models (Choi et al., 2017; Zhou et al., 2018), which should be introduced to interpret the DL models of human-virus PPI predictions as well.

Similar to other bioinformatics prediction tasks, rigorous and fair performance comparison of different human-virus PPI prediction methods is crucial. Generally, the performance of a newly developed human-virus PPI prediction is evaluated by using test sets that are specifically compiled or commonly used (Barman et al., 2014; Eid et al., 2016). Considering that such datasets were constructed based on different criteria, the performance comparison of different methods will inevitably yield biased results. To allow a more comprehensive method comparison, community-wide efforts should be taken. First, some comprehensive human-virus PPI data sets with strict reliability and quality controls should be compiled, which is fundamental for comparing different methods. Second, the developers should make their methods freely accessible to the community either through the construction of web servers or the release of source codes. Third, third-party teams should be encouraged to conduct a critical assessment of different prediction methods to obtain more unbiased comparison results. Last but not least, regular community-wide competition is also helpful to boost the improvement of human-virus PPI prediction. To this end, we should follow the successful experience of the Critical Assessment of protein Structure Prediction (CASP) experiments.[2]

Currently, dramatic progress in protein structure prediction has been made by AlphaFold2, a DL-powered method developed by the research team of DeepMind, and its high-accuracy performance has been reported in the

[2]https://www.predictioncenter.org/

CASP14 experiment (Jumper et al., 2021). Undoubtedly, the coming post-AlphaFold2 era will provide an unprecedented opportunity for the protein bioinformatics community, suggesting that many prediction methods can be significantly improved and upgraded by incorporating accessible and accurate structural information, including the prediction of human-virus PPIs. First, structural information has been widely used in previous human-virus PPI prediction methods. For instance, the P-HIPSTer model developed by Lasso et al. (2019) relied on the structural similarity of query human-virus protein pairs to known structural domain-domain/motif interactions to quantify the interaction possibility of query protein pairs. Although P-HIPSTer provided accurate prediction results, coverage of the predicted interactome is insufficient, mainly as a consequence of limited available 3D structures. With more accurate structural predictions from AlphaFold2, prediction coverage of such structure-informed human-virus PPI prediction method can be significantly increased. Second, the available structural information can contribute rich feature representations to develop DL-based prediction models. For instance, residue-level structural features can be easily introduced into the established DL architectures. 3D structures of proteins can also be converted into graphs, allowing the application of more effective DL architectures such as graph convolutional neural networks. Last but not least, highly accurate protein structures will not only propel binary PPI predictions but also predict interaction details from binding regions/residues to 3D conformational dynamics of two interacting proteins. Indeed, Baek et al. (2021) have taken the initiative to employ two DL-based structure prediction methods (i.e., RoseTTAFold and AlphaFold2) to systematically detect PPIs and construct accurate 3D models of protein complexes within the yeast proteome (Humphreys et al., 2021), which will be used for human-virus PPI prediction as well in the future. Very recently, Gao et al. (2022) developed a DL-based protein complex prediction method termed as AF2Complex, in which AlphaFold2 monomer models were employed to predict the structures of multimeric protein complexes and metrics for predicting direct PPIs between arbitrary protein pairs were also introduced. Considering AF2Complex does not rely on paired multiple sequence alignments, it could be suitable for addressing human-virus PPIs. Taken together, we are fast approaching the development of successful methods to predict human-virus PPIs empowered by DL and AlphaFold2, unveiling the secrets of human-virus relationships.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

XY wrote the draft of the manuscript. ZZ supervised the work and significantly revised the manuscript. SW, SY, and PR revised the final version of manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Key Research and Development Program of China (2017YFC1200205).

REFERENCES

Alley, E. C., Khimulya, G., Biswas, S., AlQuraishi, M., and Church, G. M. (2019). Unified rational protein engineering with sequence-based deep representation learning. *Nat. Methods* 16, 1315–1322. doi: 10.1038/s41592-019-0598-1

Ammari, M. G., Gresham, C. R., McCarthy, F. M., and Nanduri, B. (2016). HPIDB 2.0: a curated database for host-pathogen interactions. *Database* 2016:baw103. doi: 10.1093/database/baw103

Baek, M., Dimaio, F., Anishchenko, I., Dauparas, J., Lee, G. R., Wang, J., et al. (2021). Accurate prediction of protein structures and interactions using a 3-track neural network. *Science* 373, 871–876. doi: 10.1126/science.abj8754

Barman, R. K., Saha, S., and Das, S. (2014). Prediction of interactions between viral and host proteins using supervised machine learning methods. *PLoS One* 9:e112034. doi: 10.1371/journal.pone.0112034

Bromley, J., Bentz, J. W., Bottou, L., Guyon, I., Lecun, Y., Moore, C., et al. (1993). Signature verification using a "Siamese" time delay neural network. *Int. J. Pattern Recognit. Artif. Intell.* 7, 669–688. doi: 10.1142/s0218001493000339

Calderwood, M. A., Venkatesan, K., Xing, L., Chase, M. R., Vazquez, A., Holthaus, A. M., et al. (2007). Epstein-Barr virus and virus human protein interaction maps. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7606–7611. doi: 10.1073/pnas.0702332104

Chen, M., Ju, C. J. T., Zhou, G., Chen, X., Zhang, T., Chang, K. W., et al. (2019). Multifaceted protein-protein interaction prediction based on Siamese residual RCNN. *Bioinformatics* 35, i305–i314. doi: 10.1093/bioinformatics/btz328

Chiang, A. W. T., Wu, W. Y. L., Wang, T., and Hwang, M. J. (2017). Identification of entry factors involved in hepatitis C virus infection based on host-mimicking short linear motifs. *PLoS Comput. Biol.* 13:e1005368. doi: 10.1371/journal.pcbi.1005368

Choi, E., Bahadori, M. T., Song, L., Stewart, W. F., and Sun, J. (2017). "GRAM: graph-based attention model for healthcare representation learning," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY: Association for Computing Machinery), 787–795.

de Chassey, B., Meyniel-Schicklin, L., Aublin-Gex, A., Navratil, V., Chantier, T., André, P., et al. (2013). Structure homology and interaction redundancy for discovering virus-host protein interactions. *EMBO Rep.* 14, 938–944. doi: 10.1038/embor.2013.130

Dong, T. N., Brogden, G., Gerold, G., and Khosla, M. (2021). A multitask transfer learning framework for the prediction of virus-human protein–protein interactions. *BMC Bioinformatics* 22:572. doi: 10.1186/s12859-021-04484-y

Du, X., Sun, S., Hu, C., Yao, Y., Yan, Y., and Zhang, Y. (2017). DeepPPI: boosting prediction of protein-protein interactions with deep neural networks. *J. Chem. Inf. Model.* 57, 1499–1510. doi: 10.1021/acs.jcim.7b00028

Dyer, M. D., Murali, T. M., and Sobral, B. W. (2007). Computational prediction of host-pathogen protein-protein interactions. *Bioinformatics* 23, i159–i166. doi: 10.1093/bioinformatics/btm208

Dyer, M. D., Murali, T. M., and Sobral, B. W. (2011). Supervised learning and prediction of physical interactions between human and HIV proteins. *Infect. Genet. Evol.* 11, 917–923. doi: 10.1016/j.meegid.2011.02.022

Eid, F.-E., Elhefnawi, M., and Heath, L. S. (2016). DeNovo: virus-host sequence-based protein-protein interaction prediction. *Bioinformatics* 32, 1144–1150. doi: 10.1093/bioinformatics/btv737

Emamjomeh, A., Goliaei, B., Zahiri, J., and Ebrahimpour, R. (2014). Predicting protein-protein interactions between human and hepatitis C virus via an ensemble learning method. *Mol. Biosyst.* 10, 3147–3154. doi: 10.1039/c4mb00410h

Evans, P., Dampier, W., Ungar, L., and Tozeren, A. (2009). Prediction of HIV-1 virus-host protein interactions using virus and host sequence motifs. *BMC Med. Genomics* 2:27. doi: 10.1186/1755-8794-2-27

Gao, M., An, D. N., Parks, J. M., and Skolnick, J. (2022). Predicting direct physical interactions in multimeric proteins with deep learning. *bioRxiv* [Preprint]. doi: 10.1101/2021.11.09.467949

Gers, F. A., Schmidhuber, J., and Cummins, F. (2000). Learning to forget: continual prediction with LSTM. *Neural Comput.* 12, 2451–2471. doi: 10.1162/089976600300015015

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583, 459–468. doi: 10.1038/s41586-020-2286-9

Graves, A., Mohamed, A., and Hinton, G. (2013). "Speech recognition with deep recurrent neural networks," in *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (Piscataway, NJ: IEEE), 6645–6649. doi: 10.1109/icassp.2013.6638947

Greener, J. G., Kandathil, S. M., Moffat, L., and Jones, D. T. (2022). A guide to machine learning for biologists. *Nat. Rev. Mol. Cell Biol.* 23, 40–55. doi: 10.1038/s41580-021-00407-0

Grover, A., and Leskovec, J. (2016). "node2vec: scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY: Association for Computing Machinery), 855–864. doi: 10.1145/2939672.2939754

Hashemifar, S., Neyshabur, B., Khan, A. A., and Xu, J. (2018). Predicting protein-protein interactions through sequence-based deep learning. *Bioinformatics* 34, i802–i810. doi: 10.1093/bioinformatics/bty573

Humphreys, I. R., Pei, J., Baek, M., Krishnakumar, A., Anishchenko, I., Ovchinnikov, S., et al. (2021). Computed structures of core eukaryotic protein complexes. *Science* 374:eabm4805. doi: 10.1126/science.abm4805

Jean Beltran, P. M., Federspiel, J. D., Sheng, X., and Cristea, I. M. (2017). Proteomics and integrative omic approaches for understanding host–pathogen interactions and infectious diseases. *Mol. Syst. Biol.* 13:922. doi: 10.15252/msb.20167062

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi: 10.1038/s41586-021-03819-2

Kimothi, D., Soni, A., Biyani, P., and Hogan, J. M. (2016). Distributed representations for biological sequence analysis. *arXiv* [Preprint]. arXiv:1608.05949

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 25, 1097–1105.

Lanchantin, J., Weingarten, T., Sekhon, A., Miller, C., and Qi, Y. (2021). Tranfer learning for predicting virus–host protein interactions for novel virus sequences. *bioRxiv* [Preprint]. doi: 10.1101/2020.12.14.422772

Lasso, G., Mayer, S. V., Winkelmann, E. R., Chu, T., Elliot, O., Patino-Galindo, J. A., et al. (2019). A structure-informed atlas of human-virus interactions. *Cell* 178, 1526–1541. doi: 10.1016/j.cell.2019.08.005

Le, Q., and Mikolov, T. (2014). "Distributed representations of sentences and documents," in *Proceedings of the 31st International Conference on Machine Learning* 32, 1188–1196.

Li, H., Gong, X. J., Yu, H., and Zhou, C. (2018). Deep neural network based predictions of protein interactions using primary sequences. *Molecules* 23:1923. doi: 10.3390/molecules23081923

Li, J., Guo, M., Tian, X., Wang, X., Yang, X., Wu, P., et al. (2021). Virus-host interactome and proteomic survey reveal potential virulence factors influencing SARS-CoV-2 pathogenesis. *Med* 2, 99–112. doi: 10.1016/j.medj.2020.07.002

Lian, X., Yang, X., Yang, S., and Zhang, Z. (2021). Current status and future perspectives of computational studies on human–virus protein–protein interactions. *Brief. Bioinform.* 22:bbab029. doi: 10.1093/bib/bbab029

Liu-Wei, W., Kafkas, Ş, Chen, J., Dimonaco, N. J., Tegnér, J., and Hoehndorf, R. (2021). DeepViral: prediction of novel virus–host interactions from protein sequences and infectious disease phenotypes. *Bioinformatics* 37, 2722–2729. doi: 10.1093/bioinformatics/btab147

Lu, C., and Peng, Y. (2021). Computational viromics: applications of the computational biology in viromics studies. *Virol. Sin.* 36, 1256–1260. doi: 10.1007/s12250-021-00395-7

Mariano, R., and Wuchty, S. (2017). Structure-based prediction of host–pathogen protein interactions. *Curr. Opin. Struct. Biol.* 44, 119–124. doi: 10.1016/j.sbi.2017.02.007

Ou, M., Cui, P., Pei, J., Zhang, Z., and Zhu, W. (2016). "Asymmetric transitivity preserving graph embedding," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY: Association for Computing Machinery), 1105–1114. doi: 10.1145/2939672.2939751

Pan, X., Zuallaert, J., Wang, X., Shen, H.-B., Campos, E. P., Marushchak, D. O., et al. (2021). ToxDL: deep learning using primary structure and domain embeddings for assessing protein toxicity. *Bioinformatics* 36, 5159–5168. doi: 10.1093/bioinformatics/btaa656

Qiu, W., Rutherford, S., Mao, A., and Chu, C. (2017). The pandemic and its impacts. *Heal. Cult. Soc.* 9, 1–11. doi: 10.5195/hcs.2017.221

Rasul, I. (2020). The economics of viral outbreaks. *AEA Pap. Proc.* 110, 265–268. doi: 10.1257/pandp.20201016

Rozenblatt-Rosen, O., Deo, R. C., Padi, M., Adelmant, G., Calderwood, M. A., Rolland, T., et al. (2012). Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. *Nature* 487, 491–495. doi: 10.1038/nature11288

Shah, P. S., Link, N., Jang, G. M., Sharp, P. P., Zhu, T., Swaney, D. L., et al. (2018). Comparative flavivirus-host protein interaction mapping reveals mechanisms of dengue and Zika virus pathogenesis. *Cell* 175, 1931–1945. doi: 10.1016/j.cell.2018.11.028

Stukalov, A., Girault, V., Grass, V., Karayel, O., Bergant, V., Urban, C., et al. (2021). Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. *Nature* 594, 246–252. doi: 10.1038/s41586-021-03493-4

Sun, L., Wang, Y., He, J., Li, H., Peng, D., and Wang, Y. (2020). A stacked LSTM for atrial fibrillation prediction based on multivariate ECGs. *Heal. Inf. Sci. Syst.* 8:19. doi: 10.1007/s13755-020-00103-x

Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Adv. Neural Inf. Process. Syst.* 2014, 3104–3112.

Tripathi, L. P., Kataoka, C., Taguwa, S., Moriishi, K., Mori, Y., Matsuura, Y., et al. (2010). Network based analysis of hepatitis C virus core and NS4B protein interactions. *Mol. Biosyst.* 6, 2539–2553. doi: 10.1039/c0mb00103a

Tsukiyama, S., Hasan, M. M., Fujii, S., and Kurata, H. (2021). LSTM-PHV: prediction of human-virus protein–protein interactions by LSTM with word2vec. *Brief. Bioinform.* 22:bbab228. doi: 10.1093/bib/bbab228

van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.

Villaveces, J. M., Jiménez, R. C., Porras, P., Del-Toro, N., Duesbury, M., Dumousseau, M., et al. (2015). Merging and scoring molecular interactions utilising existing community standards: tools, use-cases and a case study. *Database* 2015:bau131. doi: 10.1093/database/bau131

Wainberg, M., Merico, D., Delong, A., and Frey, B. J. (2018). Deep learning in biomedicine. *Nat. Biotechnol.* 36, 829–838. doi: 10.1038/nbt.4233

Wang, D., Cui, P., and Zhu, W. (2016). "Structural deep network embedding," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY: Association for Computing Machinery), 1225–1234. doi: 10.1145/2939672.2939753

Yang, K. K., Wu, Z., Bedbrook, C. N., and Arnold, F. H. (2018). Learned protein embeddings for machine learning. *Bioinformatics* 34, 2642–2648. doi: 10.1093/bioinformatics/bty178

Yang, X., Lian, X., Fu, C., and Wuchty, S. (2021a). HVIDB: a comprehensive database for human-virus protein-protein interactions. *Brief. Bioinform.* 22, 832–844. doi: 10.1093/bib/bbaa425

Yang, X., Yang, S., Li, Q., Wuchty, S., and Zhang, Z. (2020). Prediction of human-virus protein-protein interactions through a sequence embedding-based machine learning method. *Comput. Struct. Biotechnol. J.* 18, 153–161. doi: 10.1016/j.csbj.2019.12.005

Yang, X., Yang, S., Lian, X., Wuchty, S., and Zhang, Z. (2021b). Transfer learning via multi-scale convolutional neural layers for human-virus protein-protein interaction prediction. *Bioinformatics* 37, 4771–4778. doi: 10.1093/bioinformatics/btab533

Yu, H., Luscombe, N. M., Lu, H. X., Zhu, X., Xia, Y., Han, J. D. J., et al. (2004). Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. *Genome Res.* 14, 1107–1118. doi: 10.1101/gr.1774904

Yue, X., Wang, Z., Huang, J., Parthasarathy, S., Moosavinasab, S., Huang, Y., et al. (2020). Graph embedding on biomedical networks: methods, applications and evaluations. *Bioinformatics* 36, 1241–1251. doi: 10.1093/bioinformatics/btz718

Zeng, M., Li, M., Wu, F. X., Li, Y., and Pan, Y. (2019). DeepEP: a deep learning framework for identifying essential proteins. *BMC Bioinformatics* 20:506. doi: 10.1186/s12859-019-3076-y

Zhang, A., He, L., and Wang, Y. (2017). Prediction of GCRV virus-host protein interactome based on structural motif-domain interactions. *BMC Bioinformatics* 18:145. doi: 10.1186/s12859-017-1500-8

Zhong, X., and Rajapakse, J. C. (2020). Graph embeddings on gene ontology annotations for protein–protein interaction prediction. *BMC Bioinformatics* 21:560. doi: 10.1186/s12859-020-03816-8

Zhong, X., Kaalia, R., and Rajapakse, J. C. (2019). GO2Vec: transforming GO terms and proteins to vector representations via graph embeddings. *BMC Genomics* 20:918. doi: 10.1186/s12864-019-6272-2

Zhou, M., Huang, M., and Zhu, X. (2018). An interpretable reasoning network for multi-relation question answering. *arXiv* [Preprint]. arxiv:1801.04726,

Zhu, Y., Li, F., Xiang, D., Akutsu, T., Song, J., and Jia, C. (2021). Computational identification of eukaryotic promoters based on cascaded deep capsule neural networks. *Brief. Bioinform.* 22:bbaa299. doi: 10.1093/bib/bbaa299

Zitnik, M., and Leskovec, J. (2017). Predicting multicellular function through multi-layer tissue networks. *Bioinformatics* 33, i190–i198. doi: 10.1093/bioinformatics/btx252

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

MINI REVIEW
published: 21 April 2022
doi: 10.3389/fmicb.2022.849781

The Intricacy of the Viral-Human Protein Interaction Networks: Resources, Data, and Analyses

Deeya Saha[1], Marta Iannuccelli[2], Christine Brun[1,3], Andreas Zanzoni[1†] and Luana Licata[2*†]*

[1] Aix-Marseille Univ., Inserm, TAGC, UMR_S1090, Marseille, France, [2] Department of Biology, University of Rome Tor Vergata, Rome, Italy, [3] CNRS, Marseille, France

Viral infections are one of the major causes of human diseases that cause yearly millions of deaths and seriously threaten global health, as we have experienced with the COVID-19 pandemic. Numerous approaches have been adopted to understand viral diseases and develop pharmacological treatments. Among them, the study of virus-host protein-protein interactions is a powerful strategy to comprehend the molecular mechanisms employed by the virus to infect the host cells and to interact with their components. Experimental protein-protein interactions described in the scientific literature have been systematically captured into several molecular interaction databases. These data are organized in structured formats and can be easily downloaded by users to perform further bioinformatic and network studies. Network analysis of available virus-host interactomes allow us to understand how the host interactome is perturbed upon viral infection and what are the key host proteins targeted by the virus and the main cellular pathways that are subverted. In this review, we give an overview of publicly available viral-human protein-protein interactions resources and the community standards, curation rules and adopted ontologies. A description of the main virus-human interactome available is provided, together with the main network analyses that have been performed. We finally discuss the main limitations and future challenges to assess the quality and reliability of protein-protein interaction datasets and resources.

Keywords: protein-protein interactions, virus-host protein-protein interaction databases, virus-human interactomes, molecular interaction data standards, SARS-CoV-2, emerging viruses

OPEN ACCESS

Edited by:
Gorka Lasso Cabrera, Albert Einstein College of Medicine, United States

Reviewed by:
Gregory Caignard, INRA UMR 1161 Virologie, France
Mohamed Ghadie, University Health Network, Canada
Cristina P. Vieira, Universidade do Porto, Portugal

***Correspondence:**
Andreas Zanzoni
andreas.zanzoni@univ-amu.fr
Luana Licata
luana.licata@uniroma2.it

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *06 January 2022*
Accepted: *11 March 2022*
Published: *21 April 2022*

Citation:
Saha D, Iannuccelli M, Brun C, Zanzoni A and Licata L (2022) The Intricacy of the Viral-Human Protein Interaction Networks: Resources, Data, and Analyses. Front. Microbiol. 13:849781. doi: 10.3389/fmicb.2022.849781

INTRODUCTION

Infectious diseases, including respiratory viral infections, are among the top 10 causes of death worldwide accounting for millions of fatalities every year, especially in low-income countries (World Health Organization, 2020). Moreover, the increasing incidence of (re-)emerging infectious diseases is posing serious global health threats (Jones et al., 2008; Cui et al., 2019; Pierson and Diamond, 2020), as exemplified by the COVID-19 pandemic (Morens and Fauci, 2020).

The development of effective antiviral pharmacological treatments relies on an in-depth understanding of the virus biology and the host response (Eckhardt et al., 2020). In the last decades, protein-protein interaction (PPI) discovery experiments have gained momentum among the different approaches to study viral diseases (de Chassey et al., 2014; Goodacre et al., 2020). Indeed, the systematic mapping of interactions between viral and host proteins can provide a better understanding of the molecular mechanisms of viral infections and identify viral perturbations

underlying disease phenotypes, thus suggesting novel potential targets of therapeutic intervention (Cakir et al., 2021).

Over the years, these interaction maps described in the scientific literature have been systematically captured into several publicly available molecular interaction databases (e.g., Guirimand et al., 2015; Calderone et al., 2020; Del Toro et al., 2021; Oughtred et al., 2021). The interaction data is organized in structured formats (Orchard et al., 2007; Porras et al., 2020), that can be easily processed and exploited to perform downstream computational and network analyses (Porras et al., 2020).

In this review, we discuss the state-of-the-art of available PPI resources and in particular those dedicated to viruses and the human host. A brief description of the available datasets is provided along with the developed community standards, curation rules, and strategies, adopted ontologies and controlled vocabularies, quality control procedures and scoring systems. We also give an overview of the largest available viral-human interactomes with a particular focus on the recently generated interaction maps between SARS-CoV-2 and human proteins, as well as those of other (re-)emerging viruses like Zika and Dengue, outlining common and virus-specific interaction and host-cell perturbation patterns.

We discuss how these interaction networks can provide novel mechanistic insights on viral infection biology and can suggest novel pharmacological strategies. Finally, we review the main limitations of molecular interaction resources and datasets and their future challenges.

PUBLIC RESOURCES COLLECTING VIRUS-HOST PROTEIN-PROTEIN INTERACTION DATA

Virus-host molecular interactions, mostly PPIs, detected from high-throughput studies, together with those identified in hundreds of biochemical and biophysical low-throughput studies, have been gathered in distinct public databases using structured formats (Licata and Orchard, 2016; Goodacre et al., 2020).

These public resources can be divided in: (*i*) primary databases that collect only manually curated molecular interactions extracted from peer-reviewed journals and related to different viruses and their relative hosts, such as MINT (Calderone et al., 2020), IntAct (Del Toro et al., 2021), and BioGRID (Oughtred et al., 2021); (*ii*) metadatabases integrating data from primary resources, such as VirusMentha (Calderone et al., 2015) and APID (Alonso-López et al., 2019); (*iii*) databases combining experimental interaction data with predicted PPIs, such as virusSTRING (Szklarczyk et al., 2021), human-virus PPI database (HVIDB) (Yang et al., 2021) and the pathogen-host interaction search tool PHISTO (Durmuş Tekir et al., 2013); (*iv*) databases, such as VirHostnet3.0 database (Guirimand et al., 2015), which are both primary resources collecting manually annotated PPIs and metadatabases integrating data from other molecular interaction databases; and (*v*) databases collecting information only related to a specific virus-host interactome, such as DenHunt (Karyala et al., 2016) and DenvInt (Dey and Mukhopadhyay, 2017) for the Dengue virus, the HIV-1 Human Interaction Database (Ako-Adjei et al., 2015) and the Hepatitis C Virus Protein Interaction Database (HCVpro) (Kwofie et al., 2011).

Despite the large amount of data accumulated over the years in these resources, the early data collection did not follow common criteria in terms of data curation and standardization. This discrepancy in dataset formats and curation strategies is sometimes the cause of heterogeneous data generation, which is difficult to filter, use and analyze without data loss and a time-consuming scrupulous work by bioinformaticians. With this in mind, several years ago, the Molecular Interaction working group of the HUPO-Proteomics Standards Initiative (HUPO-PSI) has developed standards, tools and Controlled Vocabularies (CVs) that have allowed life science communities to combine and analyze datasets collected and stored in different molecular interaction databases (Kerrien et al., 2007; Deutsch et al., 2017). In 2007, the working group defined the minimum information required for reporting a molecular interaction experiment (MIMIx), which enables the systematic capture and the access to interaction data in different resources (Orchard et al., 2007). Several databases have adopted this standard over the years (e.g., BIOGRID, IntAct, MINT, VirHostNet), thus enabling seamless integration of distinct interaction datasets at the minimum level of interaction details, such as interaction detection and participant detection methods.

For instance, the integration of virus-human PPIs from the main resources collecting virus-host interactions (e.g., MINT, IntAct, VirHostnet 3.0, and BIOGRID, data fetched in August 2021), generates a very large set of 54,237 interactions between viral and human proteins (**Figure 1A**). Notably, the overlap between them is very small and mainly consists of the large-scale virus-human interactomes (**Figure 1B**), suggesting that the different resources may use complementary strategies to mine the available literature.

INTERNATIONAL MOLECULAR EXCHANGE DATABASES, THEIR CURATION STRATEGIES AND ADOPTED STANDARDS

Starting from 2012, some of the major resources collecting PPI data, agreed to unify their curation efforts to obtain a shared and non-redundant dataset, which is annotated using the same curation rules and common export standards.

The result of this coordination is the International Molecular Exchange (IMEx) consortium,[1] whose members (such as IntAct, MINT, DIP, UniProtKB) have agreed to curate only experimental interaction data coming from peer-reviewed papers.

The consortium members are all professional bio-curators, employing a common detailed curation guideline and up-to-date controlled vocabularies that allow high accuracy of quality control procedures. For instance, interaction data is checked twice before its release, and specific tools are used (e.g., the

[1] http://www.imexconsortium.org/

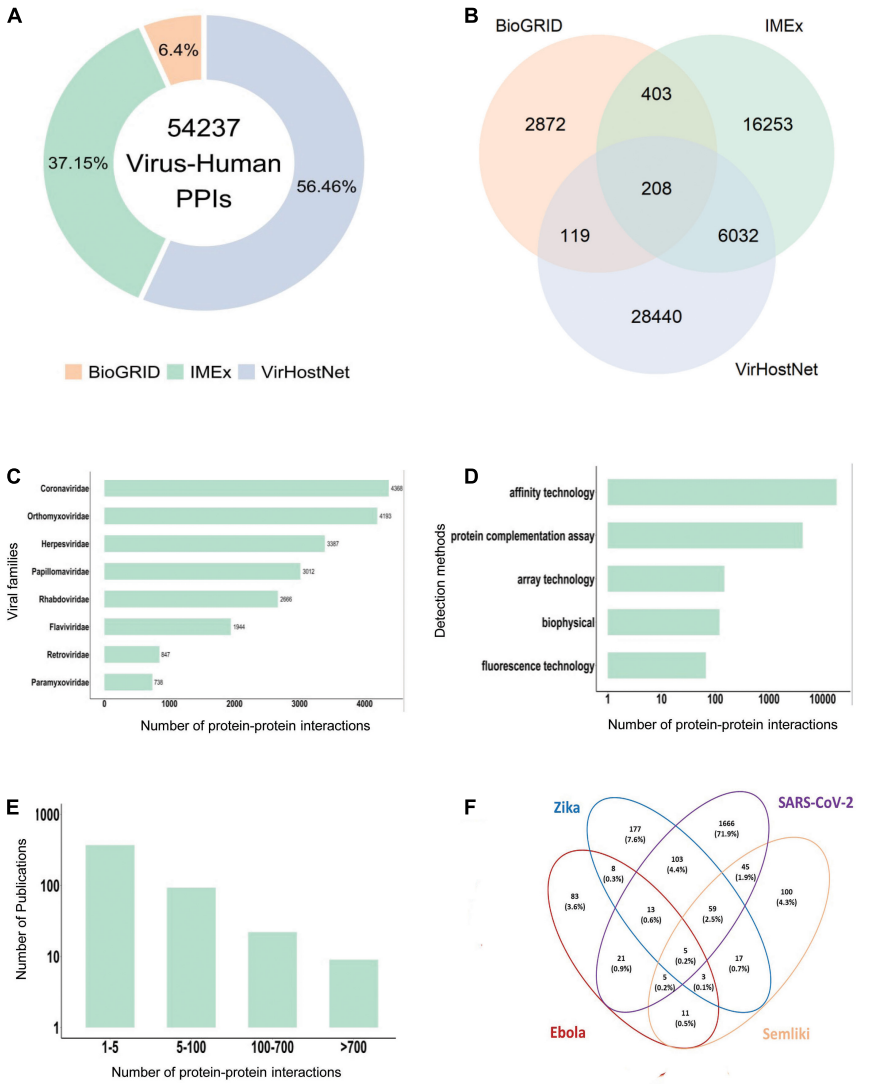


FIGURE 1 | Virus-human protein-protein interaction data statistics. **(A)** Virus-Human PPI data from the three major primary interaction resources (August 2021, BioGRID: 3,943 interactions, IMEx: 22,896 interactions; VirHostNet: 34,799 interactions). **(B)** PPI data overlap among BioGRID, IMEx and VirHostNet databases. **(C)** Number of PPIs in the IMEx dataset for the most representative viral families. **(D)** Number of PPIs in the IMEx dataset according to the experimental methods used for the interaction detection. Methods were grouped in broad categories. For instance, yeast two-hybrid is considered a "protein complementation assay," and pull-down and coimmunoprecipitation belong to the "affinity technology" category. **(E)** Number of detected PPIs in each paper curated in the IMEx dataset. Most papers describe less than 5 interactions and very few contain more than 100 interactions. The Y-axis is log-transformed. **(F)** Human targets overlap in the PPI network of four emerging viruses. Percentages are computed over the union of all the interactors.

PSI-MI semantic validator; Montecchi-Palazzi et al., 2009) to automatically check for potential errors and discrepancies related to the PSI-MI ontology of all the entries.

All IMEx entries are annotated with a wealth of details, such as the role played by the participant within the experiment (e.g., bait, prey, neutral), host organism information, cell line or tissue where the experiment was carried out, and several other features related to the interaction, such as binding sites, mutation effect, construct tags, parameters and stoichiometry (Porras et al., 2020).

IMEx datasets can be filtered according to the MIscore, a scoring system that measures the quality of a PPI based on the number of manuscripts reporting the interaction, the type of interaction and the experimental methods used to detect the given interaction (Villaveces et al., 2015).

All IMEx virus-host datasets are available at the IntAct download page.[2] As of August 2021, the IMEx virus-human dataset consists of 22,896 PPIs (**Supplementary Table 1**). Coronaviridae, Orthomyxoviridae, Papillomaviridae, and Herpesviridae are among the most represented families (**Figure 1C**). The datasets contain not only virus-human host interactions, but also interactions involving proteins from other

[2]https://www.ebi.ac.uk/intact/download/datasets

animal viruses and hosts. A dedicated COVID-19 dataset is as well available (Perfetto et al., 2020).

MAIN LIMITATIONS OF PROTEIN-PROTEIN INTERACTION DATA AND THEIR IMPACT ON INTERACTOME ANALYSIS

Studies of virus-host interactomes have turned out to be extremely powerful to identify the main host target proteins and the biological processes perturbed during a viral infection, but also to predict new potential therapeutic targets and drugs repurposing candidates (Bouhaddou et al., 2020; Gordon et al., 2020b).

However, the quality and reliability of these analyses are biased by several factors, such as the interactome coverage, the methodologies applied to generate the PPI network, the lack of low throughput validation studies, true negative PPIs and the types of standards adopted to annotate those data (Braun et al., 2009; Venkatesan et al., 2009).

Most of the available molecular interaction data is associated with the frequently studied viral families due to their impact on public health and global economy (**Figure 1C**). Their interactomes are often the result of large-scale yeast two-hybrid or AP-MS screens (de Chassey et al., 2008; Shapira et al., 2009; Tripathi et al., 2010; Muller et al., 2012; Dolan et al., 2013; Wang et al., 2017; Gordon et al., 2020a; Li et al., 2021; Stukalov et al., 2021; **Supplementary Table 2** and **Figure 1D**).

According to the methodology applied, different subsets of PPIs and different interaction types (direct or indirect) can be detected, and this partially explains the poor overlap often observed between large-scale PPI datasets (Braun, 2012). Furthermore, these differences are often related to the strategies employed by researchers during the selection of high confident interactors and the removal of spurious interactors (Walhout and Vidal, 1999; Hein et al., 2015; Choi et al., 2019).

As an example, the three main high-throughput experimental screens to map the interactome between SARS-CoV-2 and human proteins employed similar AP-MS methodologies (Gordon et al., 2020a; Li et al., 2021; Stukalov et al., 2021). However, Gordon et al., 2020a and Li et al. (2021) used HEK293T cells, while Stukalov et al. (2021) used A549 cells. Despite the use of the same technique and in two cases of the same cell line, the three screens detected a different number of interactions and showed a poor overlap in terms of human targets. However, pathway enrichment analyses revealed commonalities in the biological processes and cellular pathways targeted by viral proteins, such as cell cycle and response to stress (Perfetto et al., 2020).

This variability can be further amplified by different experimental conditions, tissues or cell lines used or experimental and participant modifications (e.g., use of chemicals or drugs, use and position of a tag, protein mutations).

Ammari et al. (2018) showed that the use of rich datasets, such as the ones provided by IMEx resources, allows performing more comprehensive network analysis whose output can differ greatly depending on the biological context or methodology used. For example, the host interacting partners of HCV proteins change depending on the cell line used to perform the experiments (e.g., Huh7 vs. HEK293) and consequently the cellular processes in which they are involved (Ammari et al., 2018).

All these aspects must be considered before selecting, merging and analyzing PPI datasets. The choice of a dataset containing information on the biological context (Porras et al., 2020) in which the interactions have occurred, can allow more sophisticated analysis and reliable outcomes.

Another important aspect that can strongly impact the evaluation of the quality of a virus-host interactome is the use of small-scale biochemical and biophysical studies that can validate and confirm the interactions found in large-scale experiments. A detailed analysis of the available validated virus-host interactions has been presented in a recent review (Goodacre et al., 2020).

In-silico approaches based on sequence (e.g., Eid et al., 2016; Liu-Wei et al., 2021) and structural similarity (e.g., de Chassey et al., 2013; Lasso et al., 2019), as well as protein docking (Wierbowski et al., 2021), have been also used to predict virus-host protein-protein interactions. The recent advent of deep-learning methods to predict protein structures (Senior et al., 2020; Baek et al., 2021) as well as protein macromolecular complexes (Baek et al., 2021; Bryant et al., 2021; Evans et al., 2021), can be a useful complementary strategy to identify or validate the molecular determinants of virus-host protein interactions identified in experimental assays.

Finally, negative PPIs can be extremely important for validating interaction data or to assess the quality of interaction prediction methods. To our knowledge, the Negatome Database 2.0 is the only available resource collecting valuable negative interaction data (Blohm et al., 2014). Indeed, the database lists experimentally verified non-interacting proteins identified either by manual curation from literature (2,171 negative interactions, 75 of which involve at least one viral protein) or derived by the analysis of the protein structures from the PDB (4,397 negative interactions, only two involve at least one viral protein).

The IMEx consortium databases also collect negative interactions (Porras et al., 2020). However, the size of the dataset is still small (∼1,000 PPIs) and only 18 of those are negative virus-host interactions, suggesting that, on one hand, researchers should systematically provide the negative interaction data coming from their experiments, and on the other hand, additional curation effort is needed to extract this information from the scientific literature.

VIRAL-HUMAN INTERACTOMES: FROM NETWORK PERTURBATION TO DYSREGULATED BIOLOGICAL PROCESSES IN DISEASE

Over the past two decades, several high-throughput techniques, such as yeast two-hybrid and affinity purification coupled to mass spectrometry (AP-MS), have been developed to map model

organism interactomes in order to decipher the dynamics and complexity of interaction networks (Snider et al., 2015). These methodologies have also been applied to chart the interactome between several viruses and the human host (**Supplementary Table 2** and **Figures 1D,E**).

The first virus-host interaction maps that have been deciphered (EBV, HCV) revealed that viral proteins preferentially target highly connected proteins (hubs) among their host proteins (Calderwood et al., 2007; de Chassey et al., 2008). As these hub proteins are relatively close in the network to a large number of proteins involved in different cellular processes, this could represent a virus strategy to subvert the cellular processes at its own benefit (Bösl et al., 2019).

Early structural bioinformatics analyses showed that human-cell hijacking by viral proteins can be achieved through interface mimicry of endogenous interactions (i.e., interaction between host proteins) (Franzosa and Xia, 2011; Garamszegi et al., 2013). Notably, they estimated that up to one-third of the viral-human interactions studied can be related to this phenomenon, in particular through the mimicry of non-globular protein interaction elements known as short linear motifs (SLiMs), which are short stretches of contiguous amino acids residues that often mediate transient PPIs (Davey et al., 2012) and have emerged through convergent evolution (Davey et al., 2011). Viral abuse of SLiMs is widespread (Davey et al., 2011; Hagai et al., 2014; Via et al., 2015), and the pervasiveness of interface mimicry provides potential connections between infectious agents and human diseases (Chen and Xia, 2019; Lasso et al., 2021).

Indeed, the targeted and consequently perturbed processes by human viruses encompass different and relevant signaling pathways: TGFbeta for SARS-CoV-2 and Hepatitis C Virus (HCV) (de Chassey et al., 2008; Stukalov et al., 2021); JAK/STAT for HCV (de Chassey et al., 2008); Notch for Epstein-Barr Virus (EBV), Human Papillomavirus (HPV), Polyoma Virus (PyV), and Adenovirus (Ad5) (Fossum et al., 2009); Wnt for Influenza A Virus (IAV-H1N1) (Shapira et al., 2009), and cellular processes such as autophagy (SARS-CoV-2) (Stukalov et al., 2021), apoptosis (EBV, HPV, PyV, and Ad5) (Fossum et al., 2009), focal adhesion (HCV) (de Chassey et al., 2008) or non-sense-mediated mRNA decay [Semliki Forest Virus (SFV); Contu et al., 2021]. The identification of targeted cellular functions is usually performed using computational tools for functional enrichment analysis such as g:Profiler (Raudvere et al., 2019) and Metascape (Zhou et al., 2019).

The blockade of some key factors through interactions is also often observed from PPI analysis. Whereas SARS-CoV-2 proteins perturb the NF-kB-repressing factor (NKRF), therefore potentially contributing to the host inflammatory response by acting on the IL-8-mediated chemotaxis of neutrophils (Li et al., 2021), the Ebola virus increases its own transcription and replication by interfering with an ubiquitin ligase (RBBP6) (Batra et al., 2018). Zika and Dengue viruses suppresses interferon-stimulated genes by inhibiting the recruitment of the transcription complex PAF1C (Shah et al., 2018), and HIV protects its replication by cleaving EIF3D, a subunit of eukaryotic translation initiation factor 3, able to inhibit HIV replication (Jäger et al., 2012). Conversely, interactome analysis also allows discovering host proteins that protect against infection such as Plakophilin 2 (PKP2), a natural inhibitor of IAV polymerase complex (Wang et al., 2017).

In addition, interaction analysis can explain disease phenotypes and unravel pathogenic mechanisms. The Zika virus (ZIKV) can cause neurodevelopmental defects (Platt et al., 2018). The viral NS4A protein interacts with a gene linked to hereditary microcephaly in humans (hANKLE2) (Shah et al., 2018). Strikingly, the ubiquitous expression of NS4A in wild type Drosophila phenocopies microcephaly that, in turn, is rescued by the expression of hANKLE2 and or its ortholog in Drosophila (dAnkle2) (Shah et al., 2018). Virus-host PPI mapping therefore provides biological insights and unveils potential pathogenic mechanisms.

Finally, although beyond the scope of this review, in the case of vector-borne diseases such as Dengue and Zika fever, the comparison between the virus-vector and the virus-host interaction maps (i.e., Shah et al., 2018) can reveal promising drug target candidates or treatment strategies to reduce the risk of viral resistance.

VIRAL-HOST INTERACTOMES OF THE EMERGING VIRUSES: COMMONALITIES AND SPECIFICITIES

Viruses have evolved sophisticated strategies to enter and evade host-cell defense and accelerate viral replication by perturbing a variety of cellular functions. Several integrated network analyses revealed that some of these strategies are virus-specific whereas others perturb common cellular pathways (Pichlmair et al., 2012; Shah et al., 2018; Bösl et al., 2019).

In this section, we focus on four emerging viruses (SARS-CoV-2, Ebola virus, ZIKV, and SFV), for which a repertoire of PPIs with human proteins in the IMEx consortium databases is available. As shown previously (Bösl et al., 2019), the four viruses show both common and specific human protein interactors (**Figure 1F**) as well as targeted biological processes. For instance, among the commonly targeted cellular functions, the most represented are related to protein translation and RNA processing (**Supplementary Table 3**), in agreement with the biology of RNA viruses. Indeed, around one quarter of the known ∼2,000 human RNA binding proteins (RBPs) has been shown to play a critical role during viral infection (Garcia-Moreno et al., 2018).

Interestingly, only five human interactors are shared by all the four viruses, and four out of five are RNA binding proteins or RBPs. One of them is the prohibitin (PHB1), which is known for its role in cell-to-cell transmission of herpes virus (Watanabe et al., 2021) and plays a pivotal role during other viral infections like that of Enterovirus and HCV (Liu et al., 2015; Too et al., 2018). Interestingly, RBPs that are commonly targeted by ZIKV, SFV, and SARS-CoV-2 are not only involved in mRNA translation but in many other immunoregulatory processes. Fifty-nine proteins are commonly targeted by ZIKV,

SFV and SARS-CoV-2 (**Figure 1F**). All of them have RNA binding activity and some of them also take active part in immune regulation. For instance, DDX21, an RNA helicase, acts in innate immune response as positive regulator of NF-kB signaling (Zhang et al., 2011; Chen et al., 2014; Abdullah et al., 2021) and as antiviral factor (Chen et al., 2014). In addition, many RBPs commonly targeted by the three viruses are associated with ubiquitin mediated protein degradation pathways (e.g., RPS7, RPL11, RPS2, RPL5) and regulation of apoptotic processes (e.g., SERBP1, RSL1D1, RPS7), thus underlying the key role of RBPs in virus-host interactions. Do these emerging viruses strategically target RBPs, as also shown for IAV-H1N1 (Shapira et al., 2009)? If this is the case, what are the consequences of the hijacking of RBPs on host defense response upon infection? These are still open questions. However, recent studies highlight the antiviral or immune related function of RBPs (Newman et al., 2015; Díaz-Muñoz and Turner, 2018; Garcia-Moreno et al., 2019) and their implication in viral processes (Embarc-Buh et al., 2021; Kamel et al., 2021).

Among SARS-CoV-2 specific human targets, there are 23 proteins linked to ER-associated protein degradation pathways members, such as BAG6 and STUB1. Recently, a study has shown that ER stress inducer thapsigargin inhibits coronavirus replication (Shaban et al., 2021). Moreover, coronaviruses, including SARS-CoV-2, suppress ER quality control processes or ER associated degradation which is re-activated by the drug thapsigargin (Shaban et al., 2021). Hence, targeting of ER-associated degradation pathways (ERAD) pathways by SARS-CoV-2 or other coronaviruses could be a unique strategy to evade host defense and facilitate viral replication within the host.

ZIKV specific human targets are mainly involved in mitochondrial translation. Recent studies show that ZIKV infection impairs mitochondrial functions (Yang et al., 2020; Yau et al., 2021). On the other hand, SFV specific interactors are involved in non-sense mediated mRNA decay (NMD) (Contu et al., 2021). Indeed, SFV inhibits NMD, which in turn helps the stabilization of the viral genomic RNA within the host cell (Contu et al., 2021).

Altogether, a quick scrutiny of the human interactors of these four emerging viruses sheds light on some of the common as well as specific strategies to subvert host cellular machinery. Further and deeper investigation of these common and specific human proteins can therefore generate testable hypotheses on the infection biology of emerging and re-emerging diseases.

CONCLUSION AND FUTURE CHALLENGES

PPI databases are important resources to gather and organize in structured formats virus-host PPI datasets useful for further network analysis. A better coverage of the curated virus-host PPIs together with the complete annotation of the experimental feature details, such as the biological context of an interaction, are necessary to perform more sophisticated network analysis. Indeed, network analysis has been proved to be fundamental to understand the perturbed cellular machinery by viruses.

Reverse genetic systems are used to manipulate virus genomes in order to understand genotypic variation or to investigate specific gene functions (Messer et al., 2012; V'kovski et al., 2021). These technologies can be also useful to contextualize virus-host PPIs during the virus life cycle and to gain important information on virus pathological processes at the molecular level.

Furthermore, the integration of interactome data with available proteomic, genetic, structural and clinical data can give a more comprehensive picture of the biological process perturbed during viral infection, paving the way to the identification of novel drug targets and successful treatments (Bouhaddou et al., 2020; Gordon et al., 2020b; Wierbowski et al., 2021).

AUTHOR CONTRIBUTIONS

AZ and LL wrote the first draft of the manuscript. DS, MI, and CB wrote sections of the manuscript. DS prepared the figure. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was supported by the European Union's Horizon 2020 Research and Innovation Programme (Project ID 101003633, RiPCoN; CB and AZ). LL acknowledges The Italian Association for Cancer Research (AIRC IG 2017 no. 20322) for financial support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.849781/full#supplementary-material

REFERENCES

Abdullah, S. W., Wu, J., Zhang, Y., Bai, M., Guan, J., Liu, X., et al. (2021). DDX21, a host restriction factor of FMDV IRES-dependent translation and replication. *Viruses* 13:1765. doi: 10.3390/v13091765

Ako-Adjei, D., Fu, W., Wallin, C., Katz, K. S., Song, G., Darji, D., et al. (2015). HIV-1, human interaction database: current status and new features. *Nucleic Acids Res.* 43, D566–D570. doi: 10.1093/nar/gku1126

Alonso-López, D., Campos-Laborie, F. J., Gutiérrez, M. A., Lambourne, L., Calderwood, M. A., Vidal, M., et al. (2019). APID database: redefining protein-protein interaction experimental evidences and binary interactomes. *Database (Oxford)* 2019:baz005. doi: 10.1093/database/baz005

Ammari, M., McCarthy, F., and Nanduri, B. (2018). Leveraging experimental details for an improved understanding of host-pathogen interactome. *Curr. Protoc. Bioinformatics* 61, 8.26.1–8.26.12. doi: 10.1002/cpbi.44

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., et al. (2021). Accurate prediction of protein structures and interactions using

a three-track neural network. *Science* 373, 871–876. doi: 10.1126/science.abj8754

Batra, J., Hultquist, J. F., Liu, D., Shtanko, O., Von Dollen, J., Satkamp, L., et al. (2018). Protein interaction mapping identifies RBBP6 as a negative regulator of ebola virus replication. *Cell* 175, 1917–1930.e13. doi: 10.1016/j.cell.2018.08.044

Blohm, P., Frishman, G., Smialowski, P., Goebels, F., Wachinger, B., Ruepp, A., et al. (2014). Negatome 2.0: a database of non-interacting proteins derived by literature mining, manual annotation and protein structure analysis. *Nucleic Acids Res.* 42, D396–D400. doi: 10.1093/nar/gkt1079

Bösl, K., Ianevski, A., Than, T. T., Andersen, P. I., Kuivanen, S., Teppor, M., et al. (2019). Common nodes of virus-host interaction revealed through an integrated network analysis. *Front. Immunol.* 10:2186. doi: 10.3389/fimmu.2019.02186

Bouhaddou, M., Memon, D., Meyer, B., White, K. M., Rezelj, V. V., Correa Marrero, M., et al. (2020). The Global phosphorylation landscape of SARS-CoV-2 infection. *Cell* 182, 685–712.e19. doi: 10.1016/j.cell.2020.06.034

Braun, P. (2012). Interactome mapping for analysis of complex phenotypes: insights from benchmarking binary interaction assays. *Proteomics* 12, 1499–1518. doi: 10.1002/pmic.201100598

Braun, P., Tasan, M., Dreze, M., Barrios-Rodiles, M., Lemmens, I., Yu, H., et al. (2009). An experimentally derived confidence score for binary protein-protein interactions. *Nat. Methods* 6, 91–97. doi: 10.1038/nmeth.1281

Bryant, P., Pozzati, G., and Elofsson, A. (2021). Improved prediction of protein-protein interactions using AlphaFold2. *bioRxiv* [Preprint] bioRxiv: 2021.09.15.460468, doi: 10.1101/2021.09.15.460468

Cakir, M., Obernier, K., Forget, A., and Krogan, N. J. (2021). Target discovery for host-directed antiviral therapies: application of proteomics approaches. *mSystems* 6:e0038821. doi: 10.1128/mSystems.00388-21

Calderone, A., Iannuccelli, M., Peluso, D., and Licata, L. (2020). Using the MINT database to search protein interactions. *Curr. Protoc. Bioinformatics* 69:e93. doi: 10.1002/cpbi.93

Calderone, A., Licata, L., and Cesareni, G. (2015). VirusMentha: a new resource for virus-host protein interactions. *Nucleic Acids Res.* 43, D588–D592. doi: 10.1093/nar/gku830

Calderwood, M. A., Venkatesan, K., Xing, L., Chase, M. R., Vazquez, A., Holthaus, A. M., et al. (2007). Epstein-Barr virus and virus human protein interaction maps. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7606–7611. doi: 10.1073/pnas.0702332104

Chen, G., Liu, C.-H., Zhou, L., and Krug, R. M. (2014). Cellular DDX21 RNA helicase inhibits influenza a virus replication but is counteracted by the viral NS1 protein. *Cell Host Microbe* 15, 484–493. doi: 10.1016/j.chom.2014.03.002

Chen, Y. F., and Xia, Y. (2019). Convergent perturbation of the human domain-resolved interactome by viruses and mutations inducing similar disease phenotypes. *PLoS Comput. Biol.* 15:e1006762. doi: 10.1371/journal.pcbi.1006762

Choi, S. G., Olivet, J., Cassonnet, P., Vidalain, P.-O., Luck, K., Lambourne, L., et al. (2019). Maximizing binary interactome mapping with a minimal number of assays. *Nat. Commun.* 10:3907. doi: 10.1038/s41467-019-11809-2

Contu, L., Balistreri, G., Domanski, M., Uldry, A.-C., and Mühlemann, O. (2021). Characterisation of the semliki forest virus-host cell interactome reveals the viral capsid protein as an inhibitor of nonsense-mediated mRNA decay. *PLoS Pathog.* 17:e1009603. doi: 10.1371/journal.ppat.1009603

Cui, J., Li, F., and Shi, Z.-L. (2019). Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* 17, 181–192. doi: 10.1038/s41579-018-0118-9

Davey, N. E., Travé, G., and Gibson, T. J. (2011). How viruses hijack cell regulation. *Trends Biochem. Sci.* 36, 159–169. doi: 10.1016/j.tibs.2010.10.002

Davey, N. E., Van Roey, K., Weatheritt, R. J., Toedt, G., Uyar, B., Altenberg, B., et al. (2012). Attributes of short linear motifs. *Mol. Biosyst.* 8, 268–281. doi: 10.1039/c1mb05231d

de Chassey, B., Meyniel-Schicklin, L., Aublin-Gex, A., Navratil, V., Chantier, T., André, P., et al. (2013). Structure homology and interaction redundancy for discovering virus-host protein interactions. *EMBO Rep.* 14, 938–944. doi: 10.1038/embor.2013.130

de Chassey, B., Meyniel-Schicklin, L., Vonderscher, J., André, P., and Lotteau, V. (2014). Virus-host interactomics: new insights and opportunities for antiviral drug discovery. *Genome Med.* 6:115. doi: 10.1186/s13073-014-0115-1

de Chassey, B., Navratil, V., Tafforeau, L., Hiet, M. S., Aublin-Gex, A., Agaugué, S., et al. (2008). Hepatitis C virus infection protein network. *Mol. Syst. Biol.* 4:230. doi: 10.1038/msb.2008.66

Del Toro, N., Shrivastava, A., Ragueneau, E., Meldal, B., Combe, C., Barrera, E., et al. (2021). The IntAct database: efficient access to fine-grained molecular interaction data. *Nucleic Acids Res.* 50, D648–D653. doi: 10.1093/nar/gkab1006

Deutsch, E. W., Orchard, S., Binz, P.-A., Bittremieux, W., Eisenacher, M., Hermjakob, H., et al. (2017). Proteomics standards initiative: fifteen years of progress and future work. *J. Proteome Res.* 16, 4288–4298. doi: 10.1021/acs.jproteome.7b00370

Dey, L., and Mukhopadhyay, A. (2017). DenvInt: a database of protein-protein interactions between dengue virus and its hosts. *PLoS Negl. Trop. Dis.* 11:e0005879. doi: 10.1371/journal.pntd.0005879

Díaz-Muñoz, M. D., and Turner, M. (2018). Uncovering the role of RNA-binding proteins in gene expression in the immune system. *Front. Immunol.* 9:1094. doi: 10.3389/fimmu.2018.01094

Dolan, P. T., Zhang, C., Khadka, S., Arumugaswami, V., Vangeloff, A. D., Heaton, N. S., et al. (2013). Identification and comparative analysis of hepatitis C virus-host cell protein interactions. *Mol. Biosyst.* 9, 3199–3209. doi: 10.1039/c3mb70343f

Durmuş Tekir, S., Çakır, T., Ardiç, E., Sayılırbaş, A. S., Konuk, G., Konuk, M., et al. (2013). PHISTO: pathogen-host interaction search tool. *Bioinformatics* 29, 1357–1358. doi: 10.1093/bioinformatics/btt137

Eckhardt, M., Hultquist, J. F., Kaake, R. M., Hüttenhain, R., and Krogan, N. J. (2020). A systems approach to infectious disease. *Nat. Rev. Genet.* 21, 339–354. doi: 10.1038/s41576-020-0212-5

Eid, F.-E., ElHefnawi, M., and Heath, L. S. (2016). DeNovo: virus-host sequence-based protein-protein interaction prediction. *Bioinformatics* 32, 1144–1150. doi: 10.1093/bioinformatics/btv737

Embarc-Buh, A., Francisco-Velilla, R., and Martinez-Salas, E. (2021). RNA-binding proteins at the host-pathogen interface targeting viral regulatory elements. *Viruses* 13:952. doi: 10.3390/v13060952

Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., et al. (2021). Protein complex prediction with AlphaFold-Multimer. *bioRxiv* [Preprint]. doi: 10.1101/2021.10.04.463034

Fossum, E., Friedel, C. C., Rajagopala, S. V., Titz, B., Baiker, A., Schmidt, T., et al. (2009). Evolutionarily conserved herpesviral protein interaction networks. *PLoS Pathog.* 5:e1000570. doi: 10.1371/journal.ppat.1000570

Franzosa, E. A., and Xia, Y. (2011). Structural principles within the human-virus protein-protein interaction network. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10538–10543. doi: 10.1073/pnas.1101440108

Garamszegi, S., Franzosa, E. A., and Xia, Y. (2013). Signatures of pleiotropy, economy and convergent evolution in a domain-resolved map of human-virus protein-protein interaction networks. *PLoS Pathog.* 9:e1003778. doi: 10.1371/journal.ppat.1003778

Garcia-Moreno, M., Järvelin, A. I., and Castello, A. (2018). Unconventional RNA-binding proteins step intothe virus-host battlefront. *Wiley Interdiscip. Rev. RNA* 9:e1498. doi: 10.1002/wrna.1498

Garcia-Moreno, M., Noerenberg, M., Ni, S., Järvelin, A. I., González-Almela, E., Lenz, C. E., et al. (2019). System-wide profiling of RNA-binding proteins uncovers key regulators of virus infection. *Mol. Cell* 74, 196–211.e11. doi: 10.1016/j.molcel.2019.01.017

Goodacre, N., Devkota, P., Bae, E., Wuchty, S., and Uetz, P. (2020). Protein-protein interactions of human viruses. *Semin. Cell Dev. Biol.* 99, 31–39. doi: 10.1016/j.semcdb.2018.07.018

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., et al. (2020b). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583, 459–468. doi: 10.1038/s41586-020-2286-9

Gordon, D. E., Hiatt, J., Bouhaddou, M., Rezelj, V. V., Ulferts, S., Braberg, H., et al. (2020a). Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science* 370:eabe9403. doi: 10.1126/science.abe9403

Guirimand, T., Delmotte, S., and Navratil, V. (2015). VirHostNet 2.0: surfing on the web of virus/host molecular interactions data. *Nucleic Acids Res.* 43, D583–D587. doi: 10.1093/nar/gku1121

Hagai, T., Azia, A., Babu, M. M., and Andino, R. (2014). Use of host-like peptide motifs in viral proteins is a prevalent strategy in host-virus interactions. *Cell Rep.* 7, 1729–1739. doi: 10.1016/j.celrep.2014.04.052

Hein, M. Y., Hubner, N. C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., et al. (2015). A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* 163, 712–723. doi: 10.1016/j.cell.2015.09.053

Jäger, S., Cimermancic, P., Gulbahce, N., Johnson, J. R., McGovern, K. E., Clarke, S. C., et al. (2012). Global landscape of HIV-human protein complexes. *Nature* 481, 365–370. doi: 10.1038/nature10719

Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., et al. (2008). Global trends in emerging infectious diseases. *Nature* 451, 990–993. doi: 10.1038/nature06536

Kamel, W., Noerenberg, M., Cerikan, B., Chen, H., Järvelin, A. I., Kammoun, M., et al. (2021). Global analysis of protein-RNA interactions in SARS-CoV-2-infected cells reveals key regulators of infection. *Mol. Cell* 81, 2851–2867.e7. doi: 10.1016/j.molcel.2021.05.023

Karyala, P., Metri, R., Bathula, C., Yelamanchi, S. K., Sahoo, L., Arjunan, S., et al. (2016). DenHunt – a comprehensive database of the intricate network of dengue-human interactions. *PLoS Negl. Trop. Dis.* 10:e0004965. doi: 10.1371/journal.pntd.0004965

Kerrien, S., Orchard, S., Montecchi-Palazzi, L., Aranda, B., Quinn, A. F., Vinod, N., et al. (2007). Broadening the horizon–level 2.5 of the HUPO-PSI format for molecular interactions. *BMC Biol.* 5:44. doi: 10.1186/1741-7007-5-44

Kwofie, S. K., Schaefer, U., Sundararajan, V. S., Bajic, V. B., and Christoffels, A. (2011). HCVpro: hepatitis C virus protein interaction database. *Infect. Genet. Evol.* 11, 1971–1977. doi: 10.1016/j.meegid.2011.09.001

Lasso, G., Honig, B., and Shapira, S. D. (2021). A sweep of earth's virome reveals host-guided viral protein structural mimicry and points to determinants of human disease. *Cell Syst.* 12, 82–91.e3. doi: 10.1016/j.cels.2020.09.006

Lasso, G., Mayer, S. V., Winkelmann, E. R., Chu, T., Elliot, O., Patino-Galindo, J. A., et al. (2019). A Structure-informed atlas of human-virus interactions. *Cell* 178, 1526–1541.e16. doi: 10.1016/j.cell.2019.08.005

Li, J., Guo, M., Tian, X., Wang, X., Yang, X., Wu, P., et al. (2021). Virus-host interactome and proteomic survey reveal potential virulence factors influencing SARS-CoV-2 pathogenesis. *Med (N Y)* 2:99–112.e7. doi: 10.1016/j.medj.2020.07.002

Licata, L., and Orchard, S. (2016). The MIntAct project and molecular interaction databases. *Methods Mol. Biol.* 1415, 55–69. doi: 10.1007/978-1-4939-3572-7_3

Liu, S., Wang, W., Brown, L. E., Qiu, C., Lajkiewicz, N., Zhao, T., et al. (2015). A novel class of small molecule compounds that inhibit hepatitis C virus infection by targeting the prohibitin-craf pathway. *EBioMedicine* 2, 1600–1606. doi: 10.1016/j.ebiom.2015.09.018

Liu-Wei, W., Kafkas, Ş, Chen, J., Dimonaco, N. J., Tegnér, J., and Hoehndorf, R. (2021). DeepViral: prediction of novel virus-host interactions from protein sequences and infectious disease phenotypes. *Bioinformatics* 37, 2722–2729. doi: 10.1093/bioinformatics/btab147

Messer, W. B., Yount, B., Hacker, K. E., Donaldson, E. F., Huynh, J. P., de Silva, A. M., et al. (2012). Development and characterization of a reverse genetic system for studying dengue virus serotype 3 strain variation and neutralization. *PLoS Negl. Trop. Dis.* 6:e1486. doi: 10.1371/journal.pntd.0001486

Montecchi-Palazzi, L., Kerrien, S., Reisinger, F., Aranda, B., Jones, A. R., Martens, L., et al. (2009). The PSI semantic validator: a framework to check MIAPE compliance of proteomics data. *Proteomics* 9, 5112–5119. doi: 10.1002/pmic.200900189

Morens, D. M., and Fauci, A. S. (2020). Emerging pandemic diseases: how we got to COVID-19. *Cell* 182, 1077–1092. doi: 10.1016/j.cell.2020.08.021

Muller, M., Jacob, Y., Jones, L., Weiss, A., Brino, L., Chantier, T., et al. (2012). Large scale genotype comparison of human papillomavirus E2-host interaction networks provides new insights for e2 molecular functions. *PLoS Pathog.* 8:e1002761. doi: 10.1371/journal.ppat.1002761

Newman, R., McHugh, J., and Turner, M. (2015). RNA binding proteins as regulators of immune cell biology. *Clin. Exp. Immunol.* 183, 37–49. doi: 10.1111/cei.12684

Orchard, S., Salwinski, L., Kerrien, S., Montecchi-Palazzi, L., Oesterheld, M., Stümpflen, V., et al. (2007). The minimum information required for reporting a molecular interaction experiment (MIMIx). *Nat. Biotechnol.* 25, 894–898. doi: 10.1038/nbt1324

Oughtred, R., Rust, J., Chang, C., Breitkreutz, B.-J., Stark, C., Willems, A., et al. (2021). The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.* 30, 187–200. doi: 10.1002/pro.3978

Perfetto, L., Pastrello, C., Del-Toro, N., Duesbury, M., Iannuccelli, M., Kotlyar, M., et al. (2020). The IMEx coronavirus interactome: an evolving map of Coronaviridae-host molecular interactions. *Database (Oxford)* 2020:baaa096. doi: 10.1093/database/baaa096

Pichlmair, A., Kandasamy, K., Alvisi, G., Mulhern, O., Sacco, R., Habjan, M., et al. (2012). Viral immune modulators perturb the human molecular network by common and unique strategies. *Nature* 487, 486–490. doi: 10.1038/nature11289

Pierson, T. C., and Diamond, M. S. (2020). The continued threat of emerging flaviviruses. *Nat. Microbiol.* 5, 796–812. doi: 10.1038/s41564-020-0714-0

Platt, D. J., Smith, A. M., Arora, N., Diamond, M. S., Coyne, C. B., and Miner, J. J. (2018). Zika virus-related neurotropic flaviviruses infect human placental explants and cause fetal demise in mice. *Sci. Transl. Med.* 10:eaao7090. doi: 10.1126/scitranslmed.aao7090

Porras, P., Barrera, E., Bridge, A., Del-Toro, N., Cesareni, G., Duesbury, M., et al. (2020). Towards a unified open access dataset of molecular interactions. *Nat. Commun.* 11:6144. doi: 10.1038/s41467-020-19942-z

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., et al. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 47, W191–W198. doi: 10.1093/nar/gkz369

Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., et al. (2020). Improved protein structure prediction using potentials from deep learning. *Nature* 577, 706–710. doi: 10.1038/s41586-019-1923-7

Shaban, M. S., Müller, C., Mayr-Buro, C., Weiser, H., Meier-Soelch, J., Albert, B. V., et al. (2021). Multi-level inhibition of coronavirus replication by chemical ER stress. *Nat. Commun.* 12:5536. doi: 10.1038/s41467-021-25551-1

Shah, P. S., Link, N., Jang, G. M., Sharp, P. P., Zhu, T., Swaney, D. L., et al. (2018). Comparative flavivirus-host protein interaction mapping reveals mechanisms of dengue and Zika virus pathogenesis. *Cell* 175, 1931–1945.e18. doi: 10.1016/j.cell.2018.11.028

Shapira, S. D., Gat-Viks, I., Shum, B. O. V., Dricot, A., de Grace, M. M., Wu, L., et al. (2009). A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. *Cell* 139, 1255–1267. doi: 10.1016/j.cell.2009.12.018

Snider, J., Kotlyar, M., Saraon, P., Yao, Z., Jurisica, I., and Stagljar, I. (2015). Fundamentals of protein interaction network mapping. *Mol. Syst. Biol.* 11:848. doi: 10.15252/msb.20156351

Stukalov, A., Girault, V., Grass, V., Karayel, O., Bergant, V., Urban, C., et al. (2021). Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. *Nature* 594, 246–252. doi: 10.1038/s41586-021-03493-4

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49, D605–D612. doi: 10.1093/nar/gkaa1074

Too, I. H. K., Bonne, I., Tan, E. L., Chu, J. J. H., and Alonso, S. (2018). Prohibitin plays a critical role in Enterovirus 71 neuropathogenesis. *PLoS Pathog.* 14:e1006778. doi: 10.1371/journal.ppat.1006778

Tripathi, L. P., Kataoka, C., Taguwa, S., Moriishi, K., Mori, Y., Matsuura, Y., et al. (2010). Network based analysis of hepatitis C virus core and NS4B protein interactions. *Mol. Biosyst.* 6, 2539–2553. doi: 10.1039/c0mb00103a

V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., and Thiel, V. (2021). Coronavirus biology and replication: implications for SARS-CoV-2. *Nat. Rev. Microbiol.* 19, 155–170. doi: 10.1038/s41579-020-00468-6

Venkatesan, K., Rual, J.-F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., et al. (2009). An empirical framework for binary interactome mapping. *Nat. Methods* 6, 83–90. doi: 10.1038/nmeth.1280

Via, A., Uyar, B., Brun, C., and Zanzoni, A. (2015). How pathogens use linear motifs to perturb host cell networks. *Trends Biochem. Sci.* 40, 36–48. doi: 10.1016/j.tibs.2014.11.001

Villaveces, J. M., Jiménez, R. C., Porras, P., Del-Toro, N., Duesbury, M., Dumousseau, M., et al. (2015). Merging and scoring molecular interactions utilising existing community standards: tools, use-cases and a case study. *Database (Oxford)* 2015:bau131. doi: 10.1093/database/bau131

Walhout, A. J., and Vidal, M. (1999). A genetic strategy to eliminate self-activator baits prior to high-throughput yeast two-hybrid screens. *Genome Res.* 9, 1128–1134. doi: 10.1101/gr.9.11.1128

Wang, L., Fu, B., Li, W., Patil, G., Liu, L., Dorf, M. E., et al. (2017). Comparative influenza protein interactomes identify the role of plakophilin 2 in virus restriction. *Nat. Commun.* 8:13876. doi: 10.1038/ncomms13876

Watanabe, M., Arii, J., Takeshima, K., Fukui, A., Shimojima, M., Kozuka-Hata, H., et al. (2021). Prohibitin-1 contributes to cell-to-cell transmission of herpes simplex virus 1 via the MAPK/ERK signaling pathway. *J. Virol.* 95, e1413–e1420. doi: 10.1128/JVI.01413-20

Wierbowski, S. D., Liang, S., Liu, Y., Chen, Y., Gupta, S., Andre, N. M., et al. (2021). A 3D structural SARS-CoV-2-human interactome to explore genetic and drug perturbations. *Nat. Methods* 18, 1477–1488. doi: 10.1038/s41592-021-01318-w

World Health Organization (2020). *The Top 10 Causes Of Death*. Geneva: World Health Organization.

Yang, S., Gorshkov, K., Lee, E. M., Xu, M., Cheng, Y.-S., Sun, N., et al. (2020). Zika virus-induced neuronal apoptosis via increased mitochondrial fragmentation. *Front. Microbiol.* 11:598203. doi: 10.3389/fmicb.2020.598203

Yang, X., Lian, X., Fu, C., Wuchty, S., Yang, S., and Zhang, Z. (2021). HVIDB: a comprehensive database for human-virus protein-protein interactions. *Brief. Bioinform.* 22, 832–844. doi: 10.1093/bib/bbaa425

Yau, C., Low, J. Z. H., Gan, E. S., Kwek, S. S., Cui, L., Tan, H. C., et al. (2021). Dysregulated metabolism underpins Zika-virus-infection-associated impairment in fetal development. *Cell Rep.* 37:110118. doi: 10.1016/j.celrep.2021.110118

Zhang, Z., Kim, T., Bao, M., Facchinetti, V., Jung, S. Y., Ghaffari, A. A., et al. (2011). DDX1, DDX21, and DHX36 helicases form a complex with the adaptor molecule TRIF to sense dsRNA in dendritic cells. *Immunity* 34, 866–878. doi: 10.1016/j.immuni.2011.03.027

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 10:1523. doi: 10.1038/s41467-019-09234-6

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

REVIEW
published: 15 July 2022
doi: 10.3389/fmicb.2022.827742

The Landscape of Virus-Host Protein–Protein Interaction Databases

*Gabriel Valiente**

Algorithms, Bioinformatics, Complexity and Formal Methods Research Group, Department of Computer Science, Technical University of Catalonia, Barcelona, Spain

Knowledge of virus-host interactomes has advanced exponentially in the last decade by the use of high-throughput screening technologies to obtain a more comprehensive landscape of virus-host protein–protein interactions. In this article, we present a systematic review of the available virus-host protein–protein interaction database resources. The resources covered in this review are both generic virus-host protein–protein interaction databases and databases of protein–protein interactions for a specific virus or for those viruses that infect a particular host. The databases are reviewed on the basis of the specificity for a particular virus or host, the number of virus-host protein–protein interactions included, and the functionality in terms of browse, search, visualization, and download. Further, we also analyze the overlap of the databases, that is, the number of virus-host protein–protein interactions shared by the various databases, as well as the structure of the virus-host protein–protein interaction network, across viruses and hosts.

Keywords: protein–protein interaction, virus-host protein–protein interaction, protein–protein interaction database, virus-host protein–protein interaction database, overlap

OPEN ACCESS

Edited by:
Gorka Lasso Cabrera, Albert Einstein College of Medicine, United States

Reviewed by:
Haiqing Zhao, Columbia University Irving Medical Center, United States
Chiara Pastrello, University Health Network (UHN), Canada

***Correspondence:**
Gabriel Valiente
gabriel.valiente@upc.edu

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *02 December 2021*
Accepted: *17 January 2022*
Published: *15 July 2022*

Citation:
Valiente G (2022) The Landscape of Virus-Host Protein–Protein Interaction Databases. Front. Microbiol. 13:827742. doi: 10.3389/fmicb.2022.827742

1. INTRODUCTION

Knowledge of virus-host interactomes has advanced exponentially in the last decade by the use of high-throughput screening technologies to obtain a more comprehensive landscape of virus-host protein–protein interactions (de Chassey et al., 2014; Sharma et al., 2015). Beyond physical methods such as affinity chromatography and coimmunoprecipitation (Phizicky and Fields, 1995), the development of mass spectrometric methods such as the yeast two-hybrid system (Fields and Sternglanz, 1994) and affinity purification combined with mass spectrometry (Kim et al., 2010) has fostered the high-throughput identification and characterization of protein–protein interactions (Börnke, 2008), computationally predicted and experimentally validated using these techniques, for protein–protein interactions within single bacteria, viruses, and small and large eukaryotes (Zhang, 2009) and also for interactions between viral proteins and proteins of the host they infect (Brito and Pinney, 2017).

In this article, we present a systematic review of the available virus-host protein–protein interaction database resources. The resources covered in this review are seven generic virus-host protein–protein interaction databases: EBI-GOA-nonIntAct (Huntley et al., 2015), BioGRID (Oughtred et al., 2021), VirusMentha (Calderone et al., 2015), IntAct (Orchard et al., 2014), VirHostNet (Navratil et al., 2009; Guirimand et al., 2015), HPIDB (Kumar and Nanduri, 2010), and

Viruses.STRING (Cook et al., 2018), as well as one database of protein–protein interactions for a specific virus, HCVpro (Kwofie et al., 2011), and three databases of protein–protein interactions for those viruses that infect a particular host, VirusMINT (Chatr-aryamontri et al., 2009), PHISTO (Tekir et al., 2013), and HVIDB (Yang et al., 2021).

The databases are reviewed on the basis of the specificity for a particular virus or host, the number of virus-host protein–protein interactions included, and the functionality in terms of browse, search, visualization, and download. Further, we also analyze the overlap of the databases, that is, the number of virus-host protein–protein interactions shared by the various databases, as well as the structure of the virus-host protein–protein interaction network, across viruses and hosts.

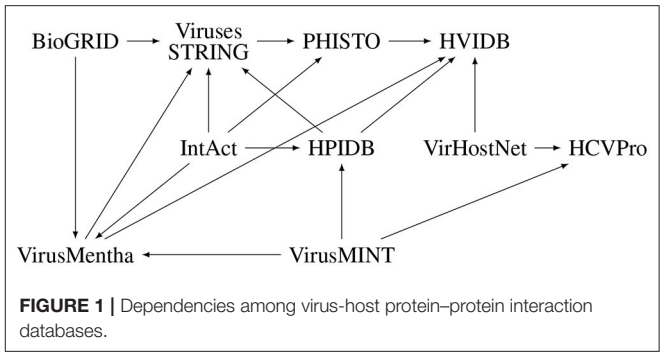


FIGURE 1 | Dependencies among virus-host protein–protein interaction databases.

2. METHODS AND RESULTS

2.1. Databases

For all the generic databases, we downloaded the virus-host protein–protein interaction data. The current (October 2021) release of EBI-GOA-nonIntAct, downloaded from http://www.ebi.ac.uk/Tools/webservices/psicquic/view/, contained 18,468 unknown, 105 virus-virus, 1,009 virus-host, and 77,852 host-host protein–protein interactions. Release 4.4.202 of BioGRID, downloaded from https://downloads.thebiogrid.org/BioGRID/Release-Archive/BIOGRID-4.4.202/, contained 702 virus-virus, 28,473 virus-host, and 2,256,186 host-host protein–protein interactions. The August 2021 update of VirusMentha, downloaded from https://virusmentha.uniroma2.it/, contained 10,907 virus-host protein–protein interactions. The current (October 2021) release of IntAct, downloaded from http://ftp.ebi.ac.uk/pub/databases/intact/current/psimitab/intact-micluster.zip, contained 18,468 unknown, 2,680 virus-virus, 26,443 virus-host, and 621,788 host-host protein–protein interactions. The March 2021 release of VirHostNet, downloaded from https://virhostnet.prabi.fr/, contained 4,442 virus-virus, 35,405 virus-host, and 158 host-host protein–protein interactions. The current (August 2021) release of HPIDB, downloaded from https://hpidb.igbb.msstate.edu/, contained 51,216 virus-host and 18,571 host-host protein–protein interactions. Last, release 10.5 of Viruses.STRING, downloaded from http://viruses.string-db.org/, contained 12,420 virus-virus, 330,136 virus-host, and 650,750,772 host-host protein–protein interactions. The ETE3 toolkit (Huerta-Cepas et al., 2016) version 3.1.2 was used to map the taxonomic identifiers for the proteins to the NCBI Taxonomy (Schoch et al., 2020) in order to determine their classification as virus or host proteins.

We also downloaded the virus-host protein–protein interaction data for all the virus-specific and host-specific databases. The current (October 2021) release of HCVpro, downloaded from https://www.cbrc.kaust.edu.sa/hcvpro/, contained 621 virus-host protein–protein interactions. The current (October 2021) release of VirusMINT, from https://maayanlab.cloud/Harmonizome/dataset/Virus+MINT+Protein-Viral+Protein+Interactions, contained 1,036 virus-host protein–protein interactions. The current (October 2021) release of PHISTO, downloaded from https://phisto.org/, contained one unknown and 52,976 virus-host protein–protein interactions. The current (October 2021) release of HVIDB, downloaded from http://zzdlab.com/hvidb/, contained 48,643 virus-host protein–protein interactions.

EBI-GOA-nonIntAct, BioGRID, IntAct, VirHostNet, HPIDB, and VirusMINT contain interactions derived from literature curation which are, in most cases, experimentally validated virus-host protein–protein interactions, while VirusMentha, HPIDB, Viruses.STRING, HCVpro, PHISTO, and HVIDB essentially integrate virus-host protein–protein interactions from other databases. In fact, VirusMentha takes virus-host protein–protein interactions from VirusMINT, IntAct, DIP (Salwinski et al., 2004), MatrixDB (Chautard et al., 2011), and BioGRID; HPIDB takes interactions from BIND (Alfarano et al., 2005), VirusMINT, PIG (Driscoll et al., 2009), GeneRIF (Jimeno-Yepes et al., 2013), Reactome (Croft et al., 2011), and IntAct; Viruses.STRING takes interactions from BioGRID, IntAct, DIP, HPIDB, and VirusMentha; HCVpro takes interactions from BIND, VirusMint, and VirHostNet; PHISTO takes interactions from APID (Prieto and De Las Rivas, 2006), IntAct, DIP, VirusMINT, iRefIndex (Razick et al., 2008), Viruses.STRING, MPIDB (Goll et al., 2008), BIND, and Reactome; and HVIDB takes virus-host protein–protein interactions from VirusMentha, VirHostNet, HPIDB, PHISTO, and PDB (Rose et al., 2017). Despite these dependencies among the databases, further illustrated in **Figure 1**, there is not much overlap among them, as discussed in Section 2.5.

These databases were chosen by means of a comprehensive literature search, and complemented with suggestions by the reviewers. P-HIPSTer (Lasso et al., 2019) was discarded because, unfortunately, the 282,528 computationally predicted viral-human protein–protein interactions therein are not available for download. ViRBase (Li et al., 2015) was discarded because the virus-host interactions therein are ncRNA-associated interactions, not protein–protein interactions and, in fact, none of the 44,276 gene symbols or 56,678 miRBase identifiers in ViRBase version 3.0 could be mapped to UniProtKB-AC unique identifiers.

2.2. Datasets

In order to be able to analyze the overlap of the databases, we mapped all virus and host protein identifiers to UniProtKB-AC unique identifiers, using the programmatic access to the database

TABLE 1 | Virus-host protein–protein interaction datasets with UniProtKB-AC unique identifiers.

Database	Viruses	Hosts	Viral proteins	Host proteins	Interactions
EBI-GOA-nonIntAct	77	26	173	455	534
BioGRID	13	6	50	2,101	5,157
VirusMentha	114	8	627	3,624	10,626
IntAct	197	68	1,062	8,102	22,727
VirHostNet	128	6	984	7,361	28,132
HPIDB	205	36	1,387	7,570	33,906
Viruses.STRING	186	61	1,703	52,440	242,784
HCVpro	1	1	7	138	140
VirusMINT	28	1	287	287	372
PHISTO	182	1	1,700	6,520	39,010
HVIDB	146	1	1,313	7,060	40,132

identifier mapping service at https://www.uniprot.org/mapping/. Host protein identifiers in the Viruses.STRING database were also mapped to UniProtKB-AC unique identifiers using the mapping files available at https://version-10-5.string-db.org/mapping_files/uniprot_mappings/. Apart from discarding any virus-virus and host-host protein–protein interactions, some of the virus-host protein–protein interactions had to be discarded as well, because the corresponding virus or host protein identifiers could not be mapped to UniProtKB-AC in a unique way. We have also raised viral strains to the species (virus) level, in order to facilitate comparison of virus-host protein–protein interactions across the databases. The resulting virus-host protein–protein interaction datasets are summarized in **Table 1** and further detailed below.

The 1,009 virus-host protein–protein interactions in the EBI-GOA-nonIntAct database contained 628 UniProtKB AC/ID identifiers, all of which were mapped to UniProtKB-AC in a unique way. This resulted in 534 unique virus-host protein–protein interactions among 173 unique proteins from 77 viruses and 455 unique proteins from 26 hosts.

The 28,473 virus-host protein–protein interactions in the BioGRID database contained 4 UniProtKB AC/ID identifiers, all of which were mapped to UniProtKB-AC in a unique way; 2 BioGRID identifiers, which could not be mapped to UniProtKB-AC; and 6,589 Entrez Gene (GeneID) identifiers, 3,007 of which were mapped to UniProtKB-AC in a unique way. This resulted in 5,157 unique virus-host protein–protein interactions among 50 unique proteins from 13 viruses and 2,101 unique proteins from 6 hosts.

The 10,907 virus-host protein–protein interactions in the VirusMentha database contained 4,347 UniProtKB AC/ID identifiers, 4,332 of which were mapped to 4,313 UniProtKB-AC in a unique way. This resulted in 10,626 unique virus-host protein–protein interactions among 627 unique proteins from 114 viruses and 3,624 unique proteins from 8 hosts.

The 26,443 virus-host protein–protein interactions in the IntAct database contained 10,282 UniProtKB AC/ID identifiers, 10,047 of which were mapped to UniProtKB-AC in a unique way. This resulted in 22,727 unique virus-host protein–protein interactions among 1,062 unique proteins from 197 viruses and 8,102 unique proteins from 68 hosts.

The 35,405 virus-host protein–protein interactions in the VirHostNet database contained 10,049 protein identifiers: 9,868 UniProtKB AC/ID identifiers, 9,717 of which were mapped to UniProtKB-AC in a unique way; 180 RefSeq Protein identifiers, 169 of which were mapped to UniProtKB-AC in a unique way; and one EMBL/GenBank/DDBJ identifier, which could not be mapped to UniProtKB-AC. This resulted in 28,132 unique virus-host protein–protein interactions among 984 unique proteins from 128 viruses and 7,361 unique proteins from 6 hosts.

The 51,216 virus-host protein–protein interactions in the HPIDB database contained 19,784 protein identifiers: 16,465 UniProtKB AC/ID identifiers, 16,295 of which were mapped to UniProtKB-AC in a unique way; 3,106 Entrez Gene (GeneID) identifiers, 1,928 of which were mapped to UniProtKB-AC in a unique way; 110 RefSeq Protein identifiers, 86 of which were mapped to UniProtKB-AC in a unique way; four EMBL/GenBank/DDBJ identifiers, one of which was mapped to UniProtKB-AC in a unique way; two Ensembl Protein identifiers, one of which was mapped to UniProtKB-AC in a unique way; one Ensembl Genomes Protein identifier, which was mapped to UniProtKB-AC in a unique way; and 96 IntAct identifiers, none of which could be mapped to UniProtKB-AC in a unique way. This resulted in 33,906 unique virus-host protein–protein interactions among 1,387 unique proteins from 205 viruses and 7,570 unique proteins from 36 hosts.

The 330,136 virus-host protein–protein interactions in the Viruses.STRING database contained 41,490 protein identifiers: 29,236 Ensembl Protein identifiers, 29,093 of which were mapped to UniProtKB-AC in a unique way; 1,371 Ensembl Genomes Protein identifiers, 1,212 of which were mapped to UniProtKB-AC in a unique way; and 131 UniProtKB AC/ID identifiers, all of which were mapped to UniProtKB-AC in a unique way. None of the remaining 10,752 identifiers could be mapped to UniProtKB-AC in a unique way. However, using the aforementioned mapping files, 37,395 host protein identifiers were mapped to UniProtKB-AC in a unique way. Combining the two approaches, this resulted in 242,784 unique virus-host

protein–protein interactions among 1,703 unique proteins from 186 viruses and 52,440 unique proteins from 61 hosts.

The 621 virus-host protein–protein interactions in the virus-specific HCVpro database contained 487 protein identifiers, 145 of which were mapped to UniProtKB-AC in a unique way. This resulted in 140 unique virus-host protein–protein interactions among 7 unique Hepatitis C virus proteins and 138 unique human proteins.

The 1,036 virus-host protein–protein interactions in the host-specific VirusMINT database contained 706 gene identifiers and 706 protein identifiers. Only 993 of the 1,412 gene and protein identifiers were mapped to UniProtKB-AC in a unique way. This resulted in 391 unique virus-host protein–protein interactions among 287 unique proteins from 43 viruses and 287 unique human proteins.

The 52,976 virus-host protein–protein interactions in the host-specific PHISTO database contained 8,212 UniProtKB AC/ID identifiers, 8,167 of which were mapped to UniProtKB-AC in a unique way. This resulted in 39,010 unique virus-host protein–protein interactions among 1,700 unique proteins from 182 viruses and 6,520 unique proteins from one host.

Finally, the 48,643 virus-host protein–protein interactions in the host-specific HVIDB database contained 9,900 protein identifiers, 9,699 of which were mapped to UniProtKB-AC in a unique way. This resulted in 44,590 unique virus-host protein–protein interactions among 1,939 unique proteins from 737 viruses and 7,437 unique human proteins.

2.3. Functionality of the Databases

All the databases support, to some extent, browsing, searching, visualization, and download. While EBI-GOA-nonIntAct, BioGRID, VirusMentha, IntAct, and HPIDB only allow for browsing search results, VirHostNet allows for browsing the database by virus lineage (Baltimore class, family, species, and taxon) and by UniProtKB keyword annotation, and Viruses.STRING has no browsing facilities, although it allows for searching by virus or host name.

EBI-GOA-nonIntAct allows for searching over the entire database using a query language based on the PSI-MITAB format (Kerrien et al., 2007), using the PSICQUIC web service (del Toro et al., 2013). BioGRID allows for searching by gene name, publication identifier, and full text search using a simple query language. IntAct allows for searching by gene name, UniProtKB identifier, taxon identifier, publication identifier, and Gene Ontology terms. VirusMentha allows for searching by gene name, UniProtKB identifier, and keyword annotation, over the entire database or for a specific virus family or host. VirHostNet allows for searching by UniProtKB identifier, name, keyword annotation, virus lineage (species or taxon), and PubMed identifier (PMID), and also allows for BLASTP (Altschul et al., 1990) searches in a database of interacting protein sequences. HPIDB allows for regular expression searching by protein accession number or name, species or taxon identifier or name, PubMed identifier (PMID) or author name, and interaction type. Viruses.STRING allows for searching by protein, virus, and host name.

For the virus-specific and the host-specific databases, HCVpro allows for browsing by virus (Hepatitis C) protein name or host (human) protein name or chromosome, virus protein identifier, interaction type, and PMID, as well as for searching by host protein name or gene identifier. VirusMINT has no browse, search, or visualization facilities, as the resource at http://mint.bio.uniroma2.it/virusmint/ is no longer available. PHISTO allows for browsing by virus family and species, and searching by taxon identifier, virus name, virus or host protein name or UniProtKB identifier, experimental method, and PMID. HVIDB allows for browsing by viral family, and searching by UniProtKB identifier, UniProtKB entry name, gene identifier, gene name, protein name, and keyword annotation.

EBI-GOA-nonIntAct, BioGRID, VirusMentha, IntAct, VirHostNet, HPIDB, Viruses.STRING, PHISTO, and HVIDB all allow for visualization of search results using a graphics applet, Cytoscape.js (Franz et al., 2016) in the case of EBI-GOA-nonIntAct and VirHostNet. HCVpro has no such visualization facilities.

Download facilities differ among the various databases. For the generic databases, EBI-GOA-nonIntAct allows for downloading a single tab-separated (TSV) text file with all the interactions stored in the database, as the result of a query to the PSICQUIC web service. BioGRID allows for downloading a single text file, in PSI-MITAB format, with all the interactions stored in the database. VirusMentha allows for downloading a zip file containing a single semicolon-separated text file for each of the 8 hosts and for each of the 25 families of viruses covered in the database, and these zip files are updated every week. IntAct also allows for downloading a single text file in PSI-MITAB format with all the interactions stored in the database. VirHostNet also allows for downloading a single tab-separated text file with all the interactions stored in the database. HPIDB also allows for downloading a single text file in PSI-MITAB format with all the interactions stored in the database. Viruses.STRING allows for downloading a tar-gzip-compressed folder containing a single space-separated text file with either all the interactions stored in the database, or only those for a particular virus or host. On the other hand, for the virus-specific and the host-specific databases, all of them allow for downloading a single comma-separated (CSV) (for PHISTO) or tab-separated (for HCVpro, VirusMINT, and HVIDB) text file with all the virus-host interactions stored in the corresponding database. The main features of the various databases are summarized in **Table 2**.

2.4. Structure of the Virus-Host Protein–Protein Interaction Networks

The structure of biological networks in general, and protein–protein interaction networks in particular, can be analyzed by means of topological measures (Börnke, 2008; Steuer and López, 2008; Zhang and Hwang, 2009; Gaudelet and Pržulj, 2019; Hauschild et al., 2019). We show next that, under several of these topological measures, virus-host protein–protein interaction networks do not differ much from other protein–protein interaction networks.

Protein–protein interaction networks usually consist of a large component that fills most of the network, with the rest of

TABLE 2 | Main features of the virus-host protein–protein interaction databases.

Database	Browse	Search	Visualization	Download	Update frequency
EBI-GOA-nonIntAct	No	Yes	Cytoscape	TSV	Monthly
BioGRID	No	Yes	Yes	PSI-MITAB	Monthly
VirusMentha	No	Yes	Yes	CSV (semicolon)	Weekly
IntAct	No	Yes	Yes	PSI-MITAB	Every 8 weeks
VirHostNet	Yes	Yes	Cytoscape	TSV	Every 8 weeks
HPIDB	No	Yes	Yes	PSI-MITAB	Every 3 months
Viruses.STRING	No	Yes	Yes	CSV (space)	12 Aug 2021
HCVpro	Yes	Yes	No	TSV	Every 6 months
VirusMINT	No	No	No	TSV	26 Oct 2012
PHISTO	Yes	Yes	Yes	CSV	Monthly
HVIDB	Yes	Yes	Yes	TSV	25 Jun 2020

Date of the last update is shown when the update frequency is unknown.

TABLE 3 | Structure of the virus-host protein–protein interaction networks.

Network	Nodes	Edges	Components			Average path length
			Number	Size	Count	
EBI-GOA-nonIntAct	628	534	116	2–13	115	1.260108
				254	1	
BioGRID	2,151	5,157	5	2–3	4	1.636054
				2,141	1	
VirusMentha	4,252	10,625	69	2–45	68	1.273371
				4,022	1	
IntAct	9,164	22,677	145	2–55	144	1.306846
				8,585	1	
VirHostNet	8,345	28,132	35	2–8	33	1.208920
				118	1	
				8,147	1	
HPIDB	8,958	33,752	92	2–56	90	1.234933
				118	1	
				8,496	1	
Viruses.STRING	54,146	242,784	104	2–80	100	1.437420
				139	1	
				250	1	
				868	1	
				52,248	1	
HCVpro	145	140	5	2–4	4	1.366622
				134	1	
VirusMINT	659	372	287	2–8	287	1.073096
PHISTO	8,220	39,010	52	2–8	51	1.157549
				8,097	1	
HVIDB	8,373	40,132	26	2–7	25	1.293151
				8,304	1	

the network divided into a large number of small components disconnected from the rest. Within each component, the average path length is the average length of the shortest paths for all pairs of nodes in the component. The average path length of a network is the average over all components of the average path length of each component, and average path lengths are usually small in biological networks (Newman, 2018).

Table 3 shows the size (number of nodes and edges), the number of connected components, the distribution of component sizes, and the average path length for the generic,

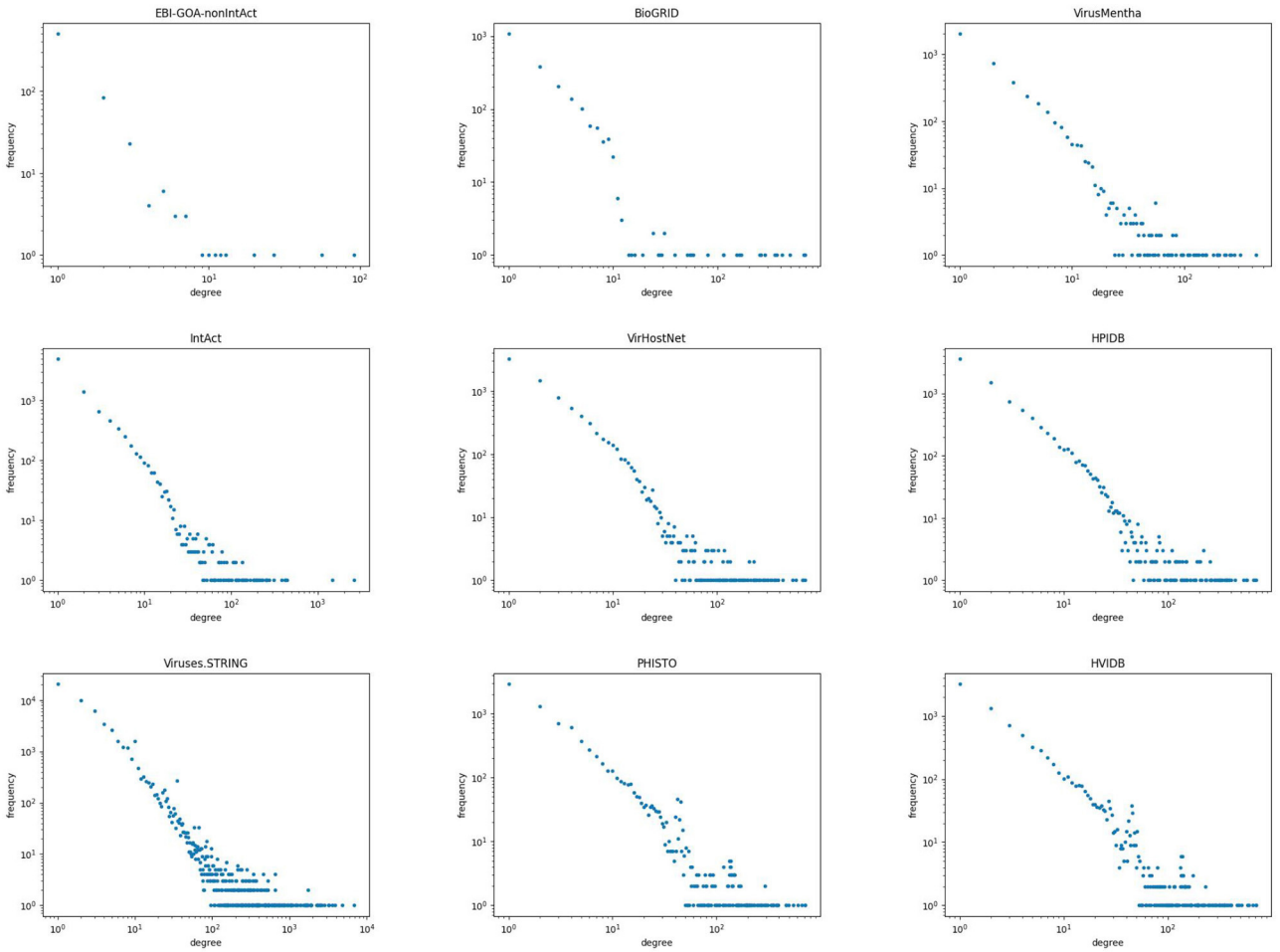


FIGURE 2 | Degree distribution of the virus-host protein–protein interaction networks.

virus-specific, and host-specific virus-host protein–protein interaction networks. These data show that virus-host protein–protein interaction networks also consist of a large component and a large number of small components, all of small average path length.

The degree of a node in a network is the number of edges attached to it, and the degree distribution of a network is the fraction p_k of the nodes that have degree k, for every k. Thus, p_k is the probability that a randomly chosen node in the network has degree k, and the degree distribution measures the frequency with which nodes of different degrees appear in the network (Newman, 2018).

Biological networks tend to have degree distributions that follow a power law of the form $p_k \sim k^{-\gamma}$ for some positive constant γ, that is, a straight line with a negative slope. **Figure 2** shows a scatter plot of the degree distribution, in logarithmic scale, for all but the two smallest virus-host protein–protein interaction networks. As can be seen therein, the degree distribution of virus-host protein–protein interaction networks follows a power law, that is, they are scale-free networks. The same behavior has been observed in other protein–protein interaction networks (Jeong et al., 2001; Barabási and Oltvai, 2004).

These structural properties of virus-host protein–protein interaction networks also characterize the networks for a specific virus or for the viruses that infect a specific host. **Table 4** shows the size (number of nodes and edges), the number of connected components, the distribution of component sizes, and the average path length of the virus-host protein–protein interaction network for the *Influenza A* virus. This virus-specific network also consists of a large component and a large number of small components, all of small average path length, although the number of small components is smaller and the average path length is larger than in the whole virus-host protein–protein interaction networks.

2.5. Overlap of the Datasets

Most of the databases contain interactions derived from literature curation and from the other databases and thus, their overlap in terms of common proteins and

TABLE 4 | Structure of the *Influenza A* virus-host protein–protein interaction networks.

Network	Nodes	Edges	Components			Average path length
			Number	Size	Count	
VirusMentha	563	1,325	5	2	4	1.562567
				555	1	
IntAct	1,737	4,141	9	2	5	1.479075
				3	1	
				4	1	
				6	1	
				1,714	1	
VirHostNet	2,620	7,921	2	118	1	2.753600
				2,502	1	
HPIDB	3,230	10,920	7	2	2	1.719102
				3	1	
				4	1	
				6	1	
				118	1	
				3,095	1	
Viruses.STRING	4,183	6,831	1	4,183	1	3.161478
PHISTO	2,943	10,416	5	2	2	1.735121
				4	2	
				2,931	1	
HVIDB	3,215	11,408	6	2	1	1.782689
				3	2	
				4	1	
				11	1	
				3,192	1	

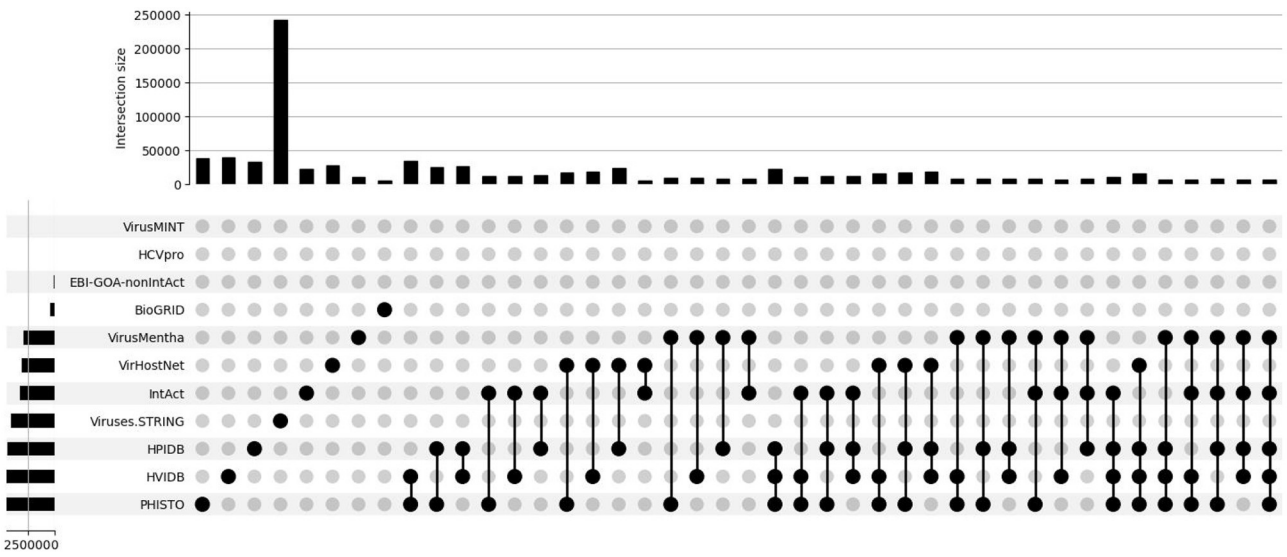


FIGURE 3 | Overlap of the virus-host protein–protein interaction databases.

interactions could be expected to be large. However, the overlap of each pair of datasets is rather small, especially with Viruses.STRING: only 35 of the 534 interactions in EBI-GOA-nonIntAct, 235 of the 5,157 interactions in BioGRID, 4,424 of the 10,625 interactions in VirusMentha, 3,801 of the 22,677 interactions in IntAct, 79 of the 28,132 interactions in VirHostNet, 306 of the 33,752 interactions in HPIDB, 4,669 of

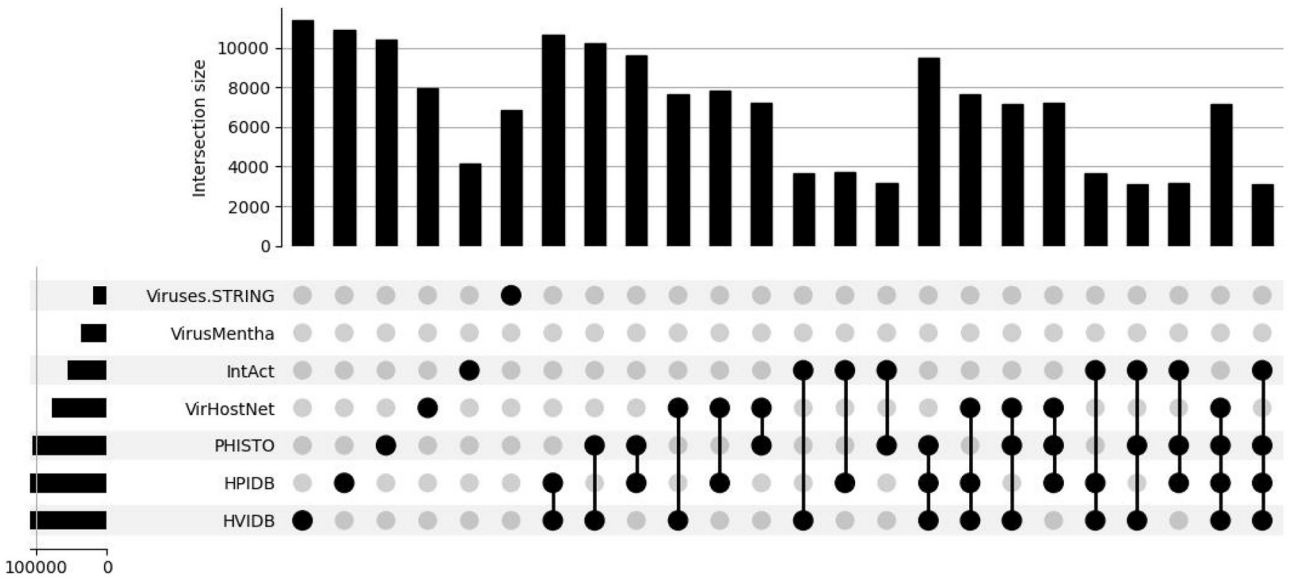


FIGURE 4 | Overlap of the virus-host protein–protein interaction databases for the *Influenza A* virus.

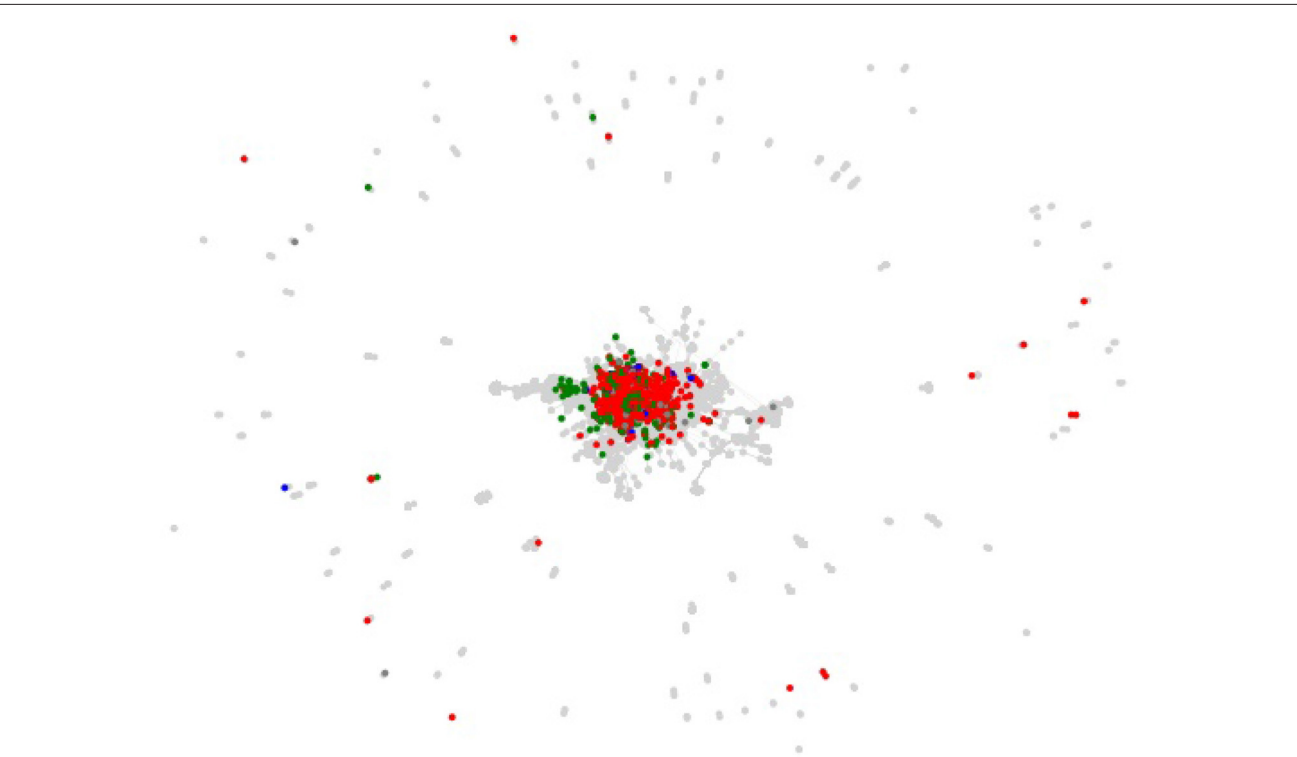


FIGURE 5 | Structure of the virus-host protein–protein interaction network for the Viruses.STRING dataset and overlapping proteins and interactions in the EBI-GOA-nonIntAct, BioGRID, VirusMentha, IntAct, VirHostNet, HPIDB, HCVpro, VirusMINT, PHISTO, and HVIDB datasets (gray). Proteins overlapping only with IntAct are shown in red, proteins overlapping only with VirHostNet in green, and proteins overlapping only with BioGRID in blue.

the 39,010 interactions in PHISTO, and 4,665 of the 40,132 interactions in HVIDB are also in the Viruses.STRING dataset.

The overlap among each three or more generic datasets is even smaller. For example, while 8,505 of the 43,944 interactions in VirusMentha, IntAct, and HPIDB are shared by the three

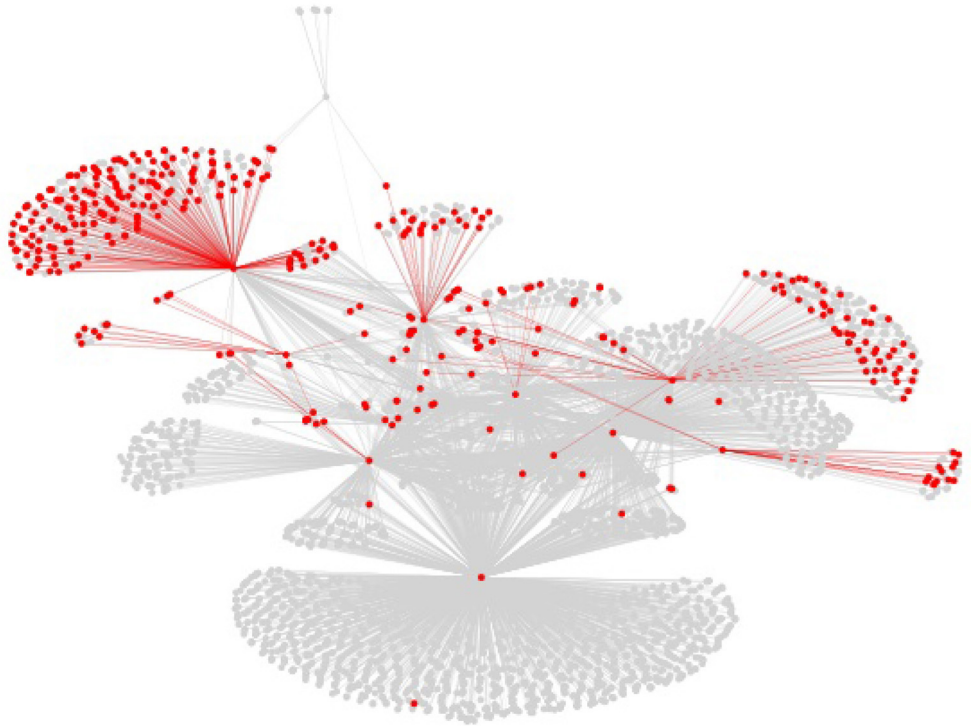


FIGURE 6 | Structure of the virus-host protein–protein interaction network for the *Influenza A* virus in the Viruses.STRING dataset (gray) and overlapping proteins and interactions in the VirusMentha dataset (red).

datasets, only 3,617 of the 281,942 interactions in VirusMentha, IntAct, HPIDB, and Viruses.STRING are shared by the four datasets, only 1,180 of the 285,650 interactions in VirusMentha, IntAct, VirHostNet, HPIDB, and Viruses.STRING are shared by the five datasets, and only 38 of the 289,406 interactions in BioGRID, VirusMentha, IntAct, VirHostNet, HPIDB, and Viruses.STRING are shared by the six datasets. Further, none of the 289,753 interactions in EBI-GOA-nonIntAct, BioGRID, VirusMentha, IntAct, VirHostNet, HPIDB, and Viruses.STRING are shared by the seven generic datasets.

This is all summarized in the set intersection diagram shown in **Figure 3**, which were obtained using a Python implementation of the UpSet tool (Lex et al., 2014). The overlap across the datasets is also small in the virus-host protein–protein interaction networks for the *Influenza A* virus, as shown in the set intersection diagram in **Figure 4**.

The centrality of proteins and interactions in the virus-host protein–protein interaction networks can also be studied by means of topological measures, in order to establish whether the networks overlap on central or on peripheral proteins and interactions. For example, the centrality of a virus-host protein–protein interaction can be measured by means of the betweenness centrality of the corresponding edge in the virus-host protein–protein interaction network, which is the sum of the fraction of all-pairs shortest paths in the network that contain the edge (Brandes, 2008). However, visual inspection of the virus-host protein–protein interaction networks, as shown in **Figure 5** for the Viruses.STRING dataset along with all the other datasets, suffice to determine that they overlap on peripheral, as opposed to central, interactions. The overlap on peripheral proteins and interactions is even more clear in the virus-host protein–protein interaction networks for the *Influenza A* virus in the Viruses.STRING and VirusMentha datasets, shown in **Figure 6**.

3. DISCUSSION

Central to the comparative review of the available virus-host protein–protein interaction database resources is the mapping of the virus and host protein identifiers used in each of the databases to unique proteins identifiers. The reader may be familiar with the good old six-symbol unique identifiers found in the UniProtKB-AC database (The UniProt Consortium, 2017). There are about 30 million 6-symbol and about 200 million 8-symbol identifiers stored therein now, what comes as a surprise since unique identifiers made up of six letters and digits would suffice to store over two billion proteins. Nevertheless, the comparative analysis of virus-host protein–protein interaction databases requires mapping proteins to unique protein identifiers such as those in UniProtKB-AC.

While some of the databases include such a mapping, it is in general neither complete nor up-to-date. The mapping problem is not trivial, as the virus and host protein identifiers used in the databases do not always map to unique proteins identifiers. Moreover, some of the databases even include proteins annotated to multiple organisms, such as HVIDB, which has 552 unique

proteins in 10,689 interactions annotated to multiple organisms, often along the same lineage. Thus, the identifier mapping problem can only be partially solved, and about 25% of the proteins in the generic, virus-specific, and host-specific databases had to be discarded because they could not be mapped to unique UniProtKB-AC identifiers.

Overall, the generic, virus-specific, and host-specific databases have very good search and visualization facilities. However, when it comes to downloading protein–protein interaction data for further use, most of the databases have their own protein identifiers and include only partial, if any, unique mappings to UniProtKB-AC. Indeed, once the protein identifiers in the various databases have been mapped to UniProtKB-AC identifiers, the resulting datasets have a rather small overlap. For example, while 14.27% of the interactions in BioGRID, 31.84% of the interactions in EBI-GOA-nonIntAct, 61.90% of the interactions in IntAct, 84.60% of the interactions in VirHostNet, and 84.71% of the interactions in VirusMentha are also found in HPIDB, only 4.55% of the interactions in BioGRID, 5.30% of the interactions in VirHostNet, 6.55% of the interactions in EBI-GOA-nonIntAct, 12.41% of the interactions in HPIDB, 16.76% of the interactions in IntAct, and 41.64% of the interactions in VirusMentha are also found in Viruses.STRING.

Further, the structural analysis of the virus-host protein–protein interaction networks showed that the databases overlap mostly on peripheral interactions, and the central interactions in the networks are not shared among the databases. This comes as a surprise, because essential proteins are known to have higher centrality in a protein–protein interaction network than the network average (Jeong et al., 2001; Raman et al., 2014) and thus, central proteins and interactions are more widely studied and more likely to be reflected in virus-host protein–protein interaction databases than peripheral proteins and interactions. The structural analysis of the virus-host protein–protein interaction network for the *Influenza A* virus, on the other hand, showed that it has a smaller number of small components and a larger average path length than the other virus-host protein–protein interaction networks, which can be explained by *Influenza A* being a widely studied virus, with a larger fraction of the virus-host protein–protein interactions reflected in the databases.

DATA AVAILABILITY STATEMENT

The datasets generated for this study (virus-host protein-protein interactions) are available in the **Supplementary Material**.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This research was partially supported by the Spanish Ministry of Science and Innovation, and the European Regional Development Fund, through project PID2021-126114NB-C44 (FEDER/MICINN/AEI).

ACKNOWLEDGMENTS

We thank Damian Szklarczyk for assistance in mapping Viruses. STRING protein identifiers to unique UniProtKB-AC identifiers.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.827742/full#supplementary-material

REFERENCES

Alfarano, C., Andrade, C. E., Anthony, K., Bahroos, N., Bajec, M., Bantoft, K., et al. (2005). The Biomolecular interaction network database and related tools: 2005 update. *Nucl. Acids Res.* 33, D418–D424. doi: 10.1093/nar/gki051

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Barabási, A.-L., and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113. doi: 10.1038/nrg1272

Börnke, F. (2008). "Protein interaction networks," in *Analysis of Biological Networks*, Ch. 9, eds B. H. Junker and F. Schreiber (Hoboken, NJ: John Wiley & Sons), 207–232.

Brandes, U. (2008). On variants of shortest-path betweenness centrality and their generic computation. *Soc. Netw.* 30, 136–145. doi: 10.1016/j.socnet.2007.11.001

Brito, A. F., and Pinney, J. W. (2017). Protein-protein interactions in virus-host systems. *Front. Microbiol.* 8,1557. doi: 10.3389/fmicb.2017.01557

Calderone, A., Licata, L., and Cesareni, G. (2015). VirusMentha: a new resource for virus-host protein interactions. *Nucl. Acids Res.* 43, D588–D592. doi: 10.1093/nar/gku830

Chatr-Aryamontri, A., Ceol, A., Peluso, D., Nardozza, A., Panni, S., Sacco, F., et al. (2009). VirusMINT: a viral protein interaction database. *Nucl. Acids Res.* 37, D669–D673. doi: 10.1093/nar/gkn739

Chautard, E., Fatoux-Ardore, M., Ballut, L., Thierry-Mieg, N., and Ricard-Blum, S. (2011). MatrixDB, the extracellular matrix interaction database. *Nucl. Acids Res.* 39, D235–D240. doi: 10.1093/nar/gkq830

Cook, H. V., Doncheva, N. T., Szklarczyk, D., von Mering, C., and Jensen, L. J. (2018). Viruses.STRING: a virus-host protein-protein interaction database. *Viruses* 10,519. doi: 10.3390/v10100519

Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., et al. (2011). Reactome: a database of reactions, pathways and biological processes. *Nucl. Acids Res.* 39, D691–D697. doi: 10.1093/nar/gkq1018

de Chassey, B., Meyniel-Schicklin, L., Vonderscher, J., André, P., and Lotteau, V. (2014). Virus-host interactomics: new insights and opportunities for antiviral drug discovery. *Gen. Med.* 6,115. doi: 10.1186/s13073-014-0115-1

del Toro, N., Dumousseau, M., Orchard, S., Jimenez, R. C., Galeota, E., Launay, G., et al. (2013). A new reference implementation of the PSICQUIC web service. *Nucl. Acids Res.* 41, W601–W606. doi: 10.1093/nar/gkt392

Driscoll, T., Dyer, M. D., Murali, T. M., and Sobral, B. W. (2009). PIG: The pathogen interaction gateway. *Nucl. Acids Res.* 37, D647–D650. doi: 10.1093/nar/gkn799

Fields, S., and Sternglanz, R. (1994). The two-hybrid system: an assay for protein-protein interactions. *Trends Gen.* 10, 286–292. doi: 10.1016/0168-9525(90)90012-u

Franz, M., Lopes, C. T., Huck, G., Dong, Y., Sumer, O., and Bader, G. D. (2016). Cytoscape.js: A graph theory library for visualisation and analysis. *Bioinformatics* 32, 309–311. doi: 10.1093/bioinformatics/btv557

Gaudelet, T., and Pržulj, N. (2019). "Introduction to graph and network theory," in *Analyzing Network Data in Biology and Medicine: An Interdisciplinary Textbook for Biological, Medical and Computational Scientists*, chapter 3, ed N. Pržulj (Cambridge: Cambridge University Press), 111–150.

Goll, J., Rajagopala, S. V., Shiau, S. C., Wu, H., Lamb, B. T., and Uetz, P. (2008). MPIDB: The microbial protein interaction database. *Bioinformatics* 24, 1743–1744. doi: 10.1093/bioinformatics/btn285

Guirimand, T., Delmotte, S., and Navratil, V. (2015). VirHostNet 2.0: surfing on the web of virus/host molecular interactions data. *Nucl. Acids Res.* 43, D583–D587. doi: 10.1093/nar/gku1121

Hauschild, A.-C., Pastrello, C., Kotlyar, M., and Jurisica, I. (2019). "Protein-protein interaction data, their quality, and major public databases," in *Analyzing Network Data in Biology and Medicine: An Interdisciplinary Textbook for Biological, Medical and Computational Scientists*, chapter 4, ed N. Pržulj (Cambridge: Cambridge University Press), 151–192.

Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* 33, 1635–1638. doi: 10.1093/molbev/msw046

Huntley, R. P., Sawford, T., Mutowo-Meullenet, P., Shypitsyna, A., Bonilla, C., Martin, M. J., et al. (2015). The GOA database: gene ontology annotation updates for 2015. *Nucl. Acids Res.* 43, D1057–D1063. doi: 10.1093/nar/gku1113

Jeong, H., Mason, S. P., Barabási, A.-L., and Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature* 411, 41–42. doi: 10.1038/35075138

Jimeno-Yepes, A. J., Sticco, J. C., Mork, J. G., and Aronson, A. R. (2013). GeneRIF indexing: Sentence selection based on machine learning. *BMC Bioinformatics* 14,171. doi: 10.1186/1471-2105-14-171

Kerrien, S., Orchard, S., and Hermjakob, H. (2007). Broadening the horizon — level 2.5 of the HUPO-PSI format for molecular interactions. *BMC Biol.* 5,44. doi: 10.1186/1741-7007-5-44

Kim, E. D. H., Sabharwal, A., Vetta, A. R., and Blanchette, M. (2010). Predicting direct protein interactions from affinity purification mass spectrometry data. *Algorithms Mol. Biol.* 5:34. doi: 10.1186/1748-7188-5-34

Kumar, R., and Nanduri, B. (2010). HPIDB: A unified resource for host-pathogen interactions. *BMC Bioinf.* 11,S16. doi: 10.1186/1471-2105-11-S6-S16

Kwofie, S. K., Schaefer, U., Sundararajan, V. S., Bajic, V. B., and Christoffels, A. (2011). HCVpro: hepatitis C virus protein interaction database. *Infect. Genet. Evol.* 11, 1971–1977. doi: 10.1016/j.meegid.2011.09.001

Lasso, G., Mayer, S. V., Winkelmann, E. R., Tim Chu and, O. E., Patino-Galindo, J. A., Park, K., et al. (2019). A structure-informed atlas of human-virus interactions. *Cell* 178, 1526–1541. doi: 10.1016/j.cell.2019.08.005

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., and Pfister, H. (2014). UpSet: visualization of intersecting sets. *IEEE Trans. Visual. Comput. Graph.* 20, 1983–1992. doi: 10.1109/TVCG.2014.2346248

Li, Y., Wang, C., Miao, Z., Bi, X., Wu, D., Jin, N., et al. (2015). ViRBase: a resource for virus-host ncRNA-associated interactions. *Nucl. Acids Res.* 43, D578–D582. doi: 10.1093/nar/gku903

Navratil, V., de Chassey, B., Meyniel, L., Delmotte, S., Gautier, C., Andrá, P., et al. (2009). VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus-host interaction networks. *Nucl. Acids Res.* 37, D661–D668. doi: 10.1093/nar/gkn794

Newman, M. E. J. (2018). *Networks*. 2nd Ed (Oxford: Oxford University Press).

Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., et al. (2014). The MIntAct project: IntAct as a common curation platform for 11 molecular interaction databases. *Nucl. Acids Res.* 42, D358–D363. doi: 10.1093/nar/gkt1115

Oughtred, R., Rust, J., Chang, C., Breitkreutz, B.-J., Stark, C., Willems, A., et al. (2021). The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.* 30, 187–200. doi: 10.1002/pro.3978

Phizicky, E. M., and Fields, S. (1995). Protein-protein interactions: methods for detection and analysis. *Microbiol. Rev.* 59, 94–123. doi: 10.1128/mr.59.1.94-123.1995

Prieto, C., and De Las Rivas, J. (2006). APID: agile protein interaction data analyzer. *Nucl. Acids Res.* 39, W298–W302. doi: 10.1093/nar/gkl128

Raman, K., Damaraju, N., and Joshi, G. K. (2014). The organisational structure of protein networks: revisiting the centrality-lethality hypothesis. *Syst. Synth. Biol.* 8, 73–81. doi: 10.1007/s11693-013-9123-5

Razick, S., Magklaras, G., and Donaldson, I. M. (2008). iRefIndex: A consolidated protein interaction database with provenance. *BMC Bioinf.* 9,405. doi: 10.1186/1471-2105-9-405

Rose, P. W., Prlić, A., Altunkaya, A., Bi, C., Bradley, A. R., Christie, C. H., et al. (2017). The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucl. Acids Res.* 45, D271–D281. doi: 10.1093/nar/gkw1000

Salwinski, L., Miller, C. S., Smith, A. J., Pettit, F. K., Bowie, J. U., and Eisenberg, D. (2004). The database of interacting proteins: 2004 update. *Nucl. Acids Res.* 32, D449–D451. doi: 10.1093/nar/gkh086

Schoch, C. L., Ciufo, S., Domrachev, M., Hotton, C. L., Kannan, S., Khovanskaya, R., et al. (2020). NCBI taxonomy: a comprehensive update on curation, resources and tools. *Database* 2020, 1–21. doi: 10.1093/database/baaa062

Sharma, D., Priyadarshini, P., and Vrati, S. (2015). Unraveling the web of viroinformatics: Computational tools and databases in virus research. *J. Virol.* 89, 1489–1501. doi: 10.1128/JVI.02027-14

Steuer, R., and López, G. Z. (2008). "Global network properties," in *Analysis of Biological Networks*, chapter 3, eds B. H. Junker and F. Schreiber (Hoboken, NJ: John Wiley & Sons), 31–63.

Tekir, S. D., Çakir, T., Ardiç, E., Sayilirbaş, A. S., Konuk, G., Konuk, M., et al. (2013). PHISTO: pathogen-host interaction search tool. *Bioinformatics* 29, 1357–1358. doi: 10.1093/bioinformatics/btt137

The UniProt Consortium (2017). Uniprot: The universal protein knowledgebase. *Nucl. Acids Res.* 45, D158–D169. doi: 10.1093/nar/gkh131

Yang, X., Lian, X., Fu, C., Yang, S., and Zhang, Z. (2021). HVIDB: a comprehensive database for human-virus protein-protein interactions. *Briefings Bioinf.* 22, 832–844. doi: 10.1093/bib/bbaa425

Zhang, A. (2009). *Protein Interaction Networks: Computational Analysis* (New York, NY: Cambridge University Press).

Zhang, A., and Hwang, W.-C. (2009). "Topological analysis of protein interaction networks," in *Protein Interaction Networks: Computational Analysis*, chapter 6, ed A. Zhang (New York, NY: Cambridge University Press), 63–108.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership