UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Probing Pandemic Virus Biology: Approaches from Molecular Biology to Systems-Level Biology for Severe Acute Respiratory Syndrome Coronavirus 2 and Human Immunodeficiency Virus 1

Permalink

https://escholarship.org/uc/item/33r5j3vb

Author

Oom, Aaron Louis

Publication Date

2021

Supplemental Material

https://escholarship.org/uc/item/33r5j3vb#supplemental

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA SAN DIEGO

Probing Pandemic Virus Biology:
Approaches from Molecular Biology to Systems-Level Biology for Severe Acute Respiratory Syndrome Coronavirus 2 and Human Immunodeficiency Virus 1

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Biomedical Sciences

by

Aaron Louis Oom

Committee in charge:

Professor John Guatelli, Chair
Professor Shane Crotty
Professor David Gonzalez
Professor Tracy Handel
Professor Deborah Spector

2021

The dissertation of Aaron Louis Oom is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

In recognition of the amazing support I have received for years from family and friends. I want to particularly thank my parents who have nourished my love of science since I was a young child. In addition, a tremendous thank you to my many friends that have supported this endeavor, especially Ryan and Mike; I will always have Appleton no matter where I am.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Acquired Immunodeficiency Syndrome (AIDS), Bayesian analysis of differential localization experiments (BANDLE), blue native (BN), chemical shift index (CSI), chemical shift perturbation (CSP), Coronavirus Disease 2019 (COVID-19), concentration for 50% cytotoxicity (CC_{50}), concentration for 50% inhibition (IC_{50}), critical micelle concentration (CMC), dimethyl amiloride (DMA), 1,2-dimyristoyl-sn-glycero-phosphocholine (DMPC), dodecylphosphocholine (DPC), Dynamic Organellar Mapping (DOM), ethyl isopropyl amiloride (EIPA), feline infectious peritonitis virus (FIPV), heteronuclear single quantum coherence (HSQC), hexamethylene amiloride (HMA), high pressure liquid chromatography (HPLC), human coronavirus 229E (HCoV-229E), human immunodeficiency virus 1 (HIV-1), Human Protein Atlas (HPA), hydrogen/deuterium (H/D), hyperplexed localization of organelle proteins by isotope tagging (hyperLOPIT), infectious bronchitis virus (IBV), in-phase anti-phase (IPAP), ketosteroid isomerase (KSI), localization of organelle proteins by isotope tagging (LOPIT), localization of organelle proteins by isotope tagging using differential centrifugation (LOPIT-DC), magic angle spinning (MAS), mass spectrometry (MS), Middle East Respiratory Syndrome Coronavirus (MERS-CoV), multiplicity of infection (MOI), murine hepatitis virus (MHV), n-hexadecylphosphocholine (HPC; fos-choline-16), nuclear magnetic resonance (NMR), nuclear Overhauser effect (NOE), nuclear Overhauser effect spectroscopy (NOESY), oriented sample (OS), perfluorooctanoic acid (PFO), Polarization Inversion Spin Exchange at the Magic Angle (PISEMA), residual dipolar coupling (RDC), Severe Acute Respiratory Syndrome coronavirus (SARS-CoV or SCV), Severe Acute Respiratory Syndrome coronavirus 2 (SARS-CoV-2), Sodium trimethylsilylpropanesulfonate (DSS), translocation analysis of spatial proteomics (TRANSPIRE), Tris (2-Carboxyethyl) phosphine (TCEP), virus-like particle (VLP)

LIST OF SUPPLEMENTAL FILES

Chapter 1 Supplemental Files: Chap1Supplemental.zip

Chapter 2 Supplemental Files: Chap2Supplemental.zip

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to acknowledge Drs. John Guatelli, Charlotte Stoneham, and Mary Lewinski for their support and counsel throughout my dissertation. Additionally, I would like to acknowledge Professors Hannah Carter, Max Qian, and Satish Pillai for consultation on the computational methods found in Chapter 1 of this dissertation.

Chapter 1, in full, has been submitted for publication of the material as it may appear in Molecular and Cellular Proteomics 2021. This material was co-authored with Stoneham, Charlotte A., Lewinski, Mary K., Richards, Alicia, Wozniak, Jacob M., Shams-Ud-Doha, K., Gonzalez, David J., Krogan, Nevan J., and Guatelli, John. The dissertation author was the principal author of this chapter.

Chapter 2, in full, is a reprint of the material[1] as it appears in PLoS Pathogens 2021. This material was co-authored with Park, Sang Ho, Siddiqi, Haley, Castro, Daniela V., De Angelis, Anna A., Stoneham, Charlotte A., Lewinski, Mary K., Clark, Alex E., Croker, Ben A., Carlin, Aaron F., Guatelli, J., and Opella, Stanley J. The dissertation author was a co-author of this chapter.

VITA

2016 Bachelor of Science, Michigan State University

2021 Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Oom, A.L., Stoneham, C.A., Lewinski, *et al.* Proteomic Organellar Mapping Reveals Modulation of Peroxisomes by HIV-1. bioRxiv. (2021).

Park, S.H., Siddiqi, H., Castro, D., *et al.* Interactions of SARS-CoV-2 envelope protein with amilorides correlate with antiviral activity. PLoS Path. (2021).

Martin-Sancho, L., Lewinski, M.K., Pache, L., *et al.* Functional Landscape of SARS-CoV-2 Cellular Restriction. Mol Cell. (2021).

Gauglitz, J.M., Aceves, C.M., Aksenov, A.A., *et al.* Untargeted mass spectrometry-based metabolomics approach unveils molecular changes in raw and processed foods and beverages. Food Chemistry, 302: 125290 (2020).

Oom, A.L., Smith, D., and Akrami, K. Identification of putative unique immunogenic ZIKV and DENV1-4 peptides for diagnostic cellular based tests. Scientific Reports, 7: article number 6218 (2017).

Pauthner, M., Havenar-Daughton, C., Sok, D., *et al.* Elicitation of Robust Tier 2 Neutralizing Antibody Responses in Nonhuman Primates by HIV Envelope Trimer Immunization Using Optimized Approaches. Immunity, 46(6): 1073-1088.e6 (2017).

Oom, A.L., Humphries, B.A., and Yang, C. MicroRNAs: novel players in cancer diagnosis and therapies. BioMed Research International. Article ID 959461, doi:10.1155/2014/959461 (2014).

Humphries, B., Wang, Z., Oom, A.L., *et al.* MicroRNA-200b targets protein kinase Cα and suppresses triple negative breast cancer tumor metastasis. Carcinogenesis, 35(10): 2254-2263 (2014).

FIELDS OF STUDY

Major Field: Biomedical Sciences

Studies in virology
Dr. John Guatelli

ABSTRACT OF THE DISSERTATION

Probing Pandemic Virus Biology:
Approaches from Molecular Biology to Systems-Level Biology for Severe Acute Respiratory Syndrome Coronavirus 2 and Human Immunodeficiency Virus 1

by

Aaron Louis Oom

Doctor of Philosophy in Biomedical Sciences

University of California San Diego, 2021

Professor John Guatelli, Chair

Viral pandemics have followed humans throughout their evolutionary history. These pandemics occur when a virus to which humans have no pre-existing immunity enters the human population and impact societies on a global scale. Recent instances, such as the Human Immunodeficiency Virus 1 (HIV-1)/Acquired Immunodeficiency Syndrome (AIDS) pandemic and

the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) pandemic, have demonstrated the ruinous impact of these events on both developing and developed nations. While studies of HIV-1 have been ongoing since the start of the HIV-1/AIDS pandemic in 1981, there still exists no cure to or vaccine against the virus. Likewise, the SARS-CoV-2 pandemic, which started in late 2019, will likely challenge the world for several years despite a concerted global vaccine rollout due to a lack of highly effective treatments that can be deployed in advance of local vaccination campaigns. The work completed here illustrates the value of both macroscopic and microscopic analyses in probing pandemic virus biology.

As systems biology approaches to virology have become more tractable, highly studied viruses such as HIV can now be analyzed in new, unbiased ways, including spatial proteomics. Spatial proteomics is the study of the distribution of proteins across cellular organelles. We employed here a differential centrifugation protocol to fractionate an inducible model of HIV infection for proteomic analysis by mass spectrometry. Using these proteomics data, we evaluated the merits of several publicly available machine learning pipelines for classification of the spatial proteome. From these analyses we have found that the performance of different classification methods varies by organelle and with the expression of HIV.

On the other hand, traditional, reductionist methods that isolate a selection of proteins from a virus can be of particular use when faced with a new viral threat such as SARS-CoV-2. Here we utilized structural nuclear magnetic resonance studies to characterize the envelope (E) protein of SARS-CoV-2. These studies highlighted a transmembrane channel formed by E that can be targeted by amiloride derivatives to significantly inhibit SARS-CoV-2 replication in cell culture.

Taken together, these findings illustrate the varied approaches that can be employed in the study of pandemic viruses.

INTRODUCTION

Viruses have long plagued humankind as the species has grown and spread across the globe. Noteworthy viruses in human history have included variola virus (cause of smallpox), various pandemic strains of influenza A virus, human immunodeficiency virus 1 (HIV-1), and Severe Acute Respiratory Syndrome coronavirus 2 (SARS-CoV-2). While smallpox was declared eradicated in 1980 by the World Health Organization, by that point it had already killed an estimated 300 million people in the 20th century alone[2]. Similarly, accounting for both pandemic and seasonal strains of influenza A virus, it is estimated that influenza deaths totaled just over 100 million in the 20th century[3]. When the ongoing pandemics of HIV-1 and SARS-CoV-2 are considered, it becomes abundantly clear that viral pandemics are not just a historic nuisance but a fixture of human societies.

Human Immunodeficiency Virus 1

HIV is a lentivirus that primarily infects CD4+ T cells. The virus exists as two primary types, HIV-1 and HIV-2, with HIV-1 (henceforth referred to simply as HIV) being the driver of nearly all pandemic cases. Following infection with HIV, the virus can be found in a variety of bodily fluids including semen, blood, breast milk, and genital secretions. Subsequent transmission can occur through direct blood contact (i.e. blood transfusion or needle sharing), vertical transmission from mother to child, or sexual contact. Patients will typically progress to an immunodeficient state over several years as the virus kills CD4+ T cells, leaving the patient vulnerable to opportunistic infections, i.e. infections that are normally controlled by immunocompetent individuals. This immunodeficient state is known as Acquired Immunodeficiency Syndrome (AIDS).

HIV was first officially recognized in 1981 following a cluster of 5 otherwise young and healthy patients with immunodeficiencies in Los Angeles[4]. However, subsequent analyses of

primate reservoirs of simian immunodeficiency virus, the ancestral precursor to HIV-1, suggest that the virus actually entered into the human population on the African continent sometime around the turn of the 20th century but only began to spread more broadly in the 1970s[5]; the lag in surveillance of the disease being due in part to the years long progression to AIDS following infection. Since the start of the pandemic, there have been nearly 80 million people infected with HIV with almost 35 million of those individuals ultimately dying from AIDS-related illnesses according to 2021 estimates from the Joint United Nations Programme on HIV/AIDS[6].

The replication cycle of HIV begins when its envelope protein (Env) binds to CD4 molecules on the surface of target cells as well as a co-receptor, either CXCR4 or CCR5. Following binding, Env goes through a set of conformational changes that facilitate the fusion of the viral particle, or virion, to the target cell[7]. This fusion event allows for the core of the virus to enter the cell, where the viral RNA genome is reverse transcribed into DNA[8] and subsequently integrated into the host genome[9]. With the integration of HIV into the host genome, the virus is maintained within the infected individual with no current treatment options able to remove these integrated DNA forms known as proviruses. The integrated viral genome is transcribed and translated similarly to other genes in the genome with HIV co-opting various host factors to carry out the remainder of the replication cycle[10]. As part of this co-opting, HIV encodes in its genome a set of accessory proteins that dramatically increase the ability of the virus to replicate and infect[11–13]. One such protein is Negative Factor, or Nef, which is capable of hijacking host cellular membrane trafficking machinery and redirecting multiple host proteins from their typical subcellular location; this redirection often leads to degradation of the host protein[11]. Notable examples of Nef targets include the cell surface proteins CD4 and class I major histocompatibility complex molecules.

Current treatment options for HIV rely on inhibitors of various steps in the viral replication cycle including viral entry, reverse transcription, integration, and proteolytic cleavage of viral polyproteins[14]. While treatment keeps the virus in a latent form that does not produce clinically

detectable levels of virus, thereby rendering the infection untransmissible, these therapeutics do not cure infected individuals and must be taken every day for the remainder of the patient's life. It is worth noting, however, that in spite of the need for lifelong treatment, many people living with HIV have a close to typical life expectancy when adherent to treatment[15,16].

Systems Biology Approaches to Determine Protein Localization

Within a cell, proteins are distributed across a range of membranous organelles. This subcellular localization of a given protein will often augment its function as the localization determines which proteins and other biomolecules are nearby. There are a variety of techniques that can be used to study protein localization, with methods ranging from those originally designed to study a single protein or a small selection of proteins (i.e. microscopy, co-immunoprecipitation, and proximity labeling) to high-throughput methods intended to characterize the protein content of a single organelle or even an entire cell (i.e. organellar purification and whole cell spatial proteomics). I will provide here a brief discussion of each of these methods to give the reader an overview of the field; the reader is referred elsewhere for an extensive review of these methods[17].

Microscopy

The most straightforward method for determining a protein's subcellular localization is to use immunofluorescence or fluorescence microscopy. If there is an existing, high-quality antibody against the chosen protein(s), then the preferred option would be immunofluorescence microscopy. In this method, the researcher takes a primary antibody that has been raised against the protein typically in a host animal, such as a mouse, rabbit, or goat. Either this primary antibody ("direct staining") or an additional secondary antibody against the host animal's antibodies ("indirect staining") will be conjugated to a fluorescent protein or molecule. The fluorescence from the antibody can then be visualized using light microscopy. From these

images it is generally possible with sufficient experience to identify the localization of the protein. Confirmation of this localization is accomplished by co-staining for proteins known to localize to a single organelle which serve as markers of that organelle.

For those proteins that do not have existing, high-quality antibodies or for which an antibody cannot be produced, a recombinant form of the protein can be expressed in cells with either an epitope tag or a fluorescent protein attached. In the case of epitope tagged proteins, this epitope tag can be detected using similar methods as above. For fluorescent proteins, there is no need for antibody staining as the fluorescent protein is readily detectable by light microscopy.

While this technique was originally developed for the detection of a single protein or a small number of proteins, advances in high-throughput screening and image analysis have allowed for much of the human proteome to be characterized in this manner. By far, the most notable example of such an effort is the Human Protein Atlas (HPA)[18]. A sprawling collaboration across multiple research groups, the HPA started in 2003 with the goal of characterizing the subcellular localization of every human protein via immunofluorescence microscopy. The HPA aims to characterize each protein across a variety of human cell lines in order to generate a consensus profile for each protein. The entire HPA database is publicly available online.

Co-immunoprecipitation

Although typically used to determine the protein makeup of a complex, co-immunoprecipitation can additionally be used to examine subcellular localization. An antibody against a chosen protein can be used to gently precipitate the protein along with proteins in complex. In the case where an antibody against the protein is not readily available, a recombinant version of the protein with an epitope tag can alternatively be used. Following precipitation, the complexed proteins can be analyzed by mass spectrometry (MS) to give a list of interactors. Examination of these interactors can identify proteins of known subcellular

localization, which in turn informs the researcher of their protein of interest's localization. Using this method, the interaction network of each HIV protein has been characterized[19]. This approach can also be turned towards mapping the network of proteins in a given organelle by immunoprecipitating proteins known to localize to that organelle[20].

Proximity Labeling

As co-immunoprecipitation relies on careful sample preparation, it is possible to lose parts of protein complexes or to completely miss transient interactions. Proximity labeling was developed as a method to avoid these potential pitfalls. Typically, a recombinant protein is attached to an enzyme such as BioID[21,22], APEX[23,24], or TurboID[25] which will add a biotin molecule to nearby proteins. Biotinylated proteins can be precipitated using streptavidin beads and the protein slurry can be fully characterized by MS to yield a list of interactors and proximal proteins. Similar to co-immunoprecipitation studies, interactors and proximal proteins can be analyzed for proteins of known localization to determine the potential localization of a protein of interest. One notable study added a temporal dimension to track the location and interactors over time of a G-protein coupled receptor following internalization from the cell surface[26]. Fusing the chosen labeling enzyme to proteins resident to a single organelle also allows for a more comprehensive characterization of that organelle's proteome as compared to the respective co-immunoprecipitation studies for those proteins[24].

Organellar Purification

The methods discussed thus far rely on selecting a protein that is known to localize to a single organelle for the purpose of mapping out other proteins similarly localized. For a more unbiased approach, it is possible in certain circumstances to purify a single organelle from cell lysates, thereby allowing for analysis of the full organelle. The nucleus and mitochondria are both well-suited for these approaches as they can be separated from other organelles with

greater ease than other membranous organelles[27–29]. However, this means the method is limited in its ability to analyze the remaining organelles as they will often co-purify. The best known example of this approach is the MitoCarta database[27,30] which stands as the most comprehensive description of mitochondrial proteins to date. In a slight variation on typical organellar purification methods, there have also been studies that have purified away membranous organelles for the purpose of examining the contents of the cytosol and small vesicles[31,32].

Whole Cell Spatial Proteomics

The final method discussed here is of particular relevance to the studies described in this thesis. Due to the co-purifying nature of many organelles, more recent studies have focused on characterizing the whole proteome of a cell and classifying proteins into a defined set of organelles. Samples are generally prepared by one of two subcellular fractionation methods: differential centrifugation or density gradient fractionation. Following fractionation, cellular fractions are analyzed by mass spectrometry to determine the abundance of proteins across all fractions. Researchers can then cross-reference their detected proteins with a variety of resources to identify proteins that can serve as markers of chosen organelles and be used in subsequent machine learning-based classification of non-markers.

For density gradient fractionation, there are two main protocols developed by Kathryn Lilley's group: localization of organelle proteins by isotope tagging (LOPIT)[33] and the significant overhaul of LOPIT, hyperplexed LOPIT (hyperLOPIT)[34]. For differential centrifugation fractionation, there are also two main protocols: Dynamic Organellar Mapping (DOM) from Georg Borner's group[28,35] and an expansion on the DOM method from the Lilley group, LOPIT using differential centrifugation (LOPIT-DC)[36]. There are two major differences between density gradient and differential centrifugation fractionation that significantly impact studies: 1) differential centrifugation protocols are generally quicker, require less starting material, and are

technically easier to perform than density gradient protocols[28,34], and 2) density gradient protocols generally yield better resolution between organelles than differential centrifugation protocols[37].

The data from any of these methods can be analyzed by a variety of machine learning algorithms to determine the localization of detected proteins. Common algorithms used for the classification of proteins include support vector machine[28,29], neural networks[38], and Bayesian models[39]. Each algorithm can generate a confidence score for each of its predictions, but only Bayesian methods allow for a determination of an actual probability.

It is also possible to use any of these methods for comparative analyses, such as following perturbation of a cellular receptor[28]. In such a case, researchers can examine proteins that are considered to have moved following cellular perturbation. Methods for determining protein translocations include (from least to most complex): label-based movement[38], movement-reproducibility plots[28], translocation analysis of spatial proteomics (TRANSPIRE)[40], and Bayesian analysis of differential localization experiments (BANDLE)[41].

A current limitation of the field is that while each of these methods has been explored in the context of various cell models, there has only been a single study applying spatial proteomics to viral replication[38]. As the choice of computational method can impact the outcomes of spatial proteomic studies[37], it is thus important to first consider which classification and translocation identification methods are best suited for studies of HIV. A comparison of several of these methods as applied to HIV follows in Chapter 1 of this dissertation. From this comparison, one can identify the optimal method for detecting cellular changes that have been previously overlooked in more focused studies of HIV.

Severe Acute Respiratory Syndrome Coronavirus 2

SARS-CoV-2 is a coronavirus that primarily infects the respiratory tract, although autopsy studies of patients that have died from Coronavirus Disease 2019 (COVID-19), the

disease caused by SARS-CoV-2, suggests other potential organs that can support limited viral replication[42–44]. The pandemic first came to international attention in late December 2019, but evolutionary genetics studies suggest that spread among humans likely started sometime between mid-October and mid-November 2019 in Hubei Province, China[45]. Early in the SARS-CoV-2 pandemic the R_0 of the virus was estimated to be between 2 and 3, i.e. an infected individual will go on to infect 2-3 other people on average in a population with no immunity[46,47].

The most common symptoms of COVID-19 include cough, fever, headaches, and sore throat[48–50]. On average, patients will begin experiencing symptoms 5 days post-exposure, but an overwhelming majority of patients, ~80%, will have no or mild symptoms; much of the transmission also appears to take place in this incubation period[48]. Elderly individuals are at a much higher risk of hospitalization and death with current estimates from the Centers for Disease Control and Prevention showing patients 85 years and older are 15 times more likely to be hospitalized and 570 times more likely to die as compared to 18-29 year olds[51]. In these more severe cases, patients appear to go through a viral symptom phase in the first week following symptom onset, followed by an inflammatory period in subsequent weeks[52]. Patients with mild to moderate symptoms can typically manage symptoms from home[53], although patients at high risk of hospitalization are recommended to receive treatment with monoclonal antibodies[54]. Following hospitalization with severe COVID-19—cases that generally require supplemental oxygen—patients will receive the antiviral remdesivir, the steroid dexamethasone, the Janus kinase inhibitor baricitinib, and/or the interleukin-6 receptor inhibitor tocilizumab, depending on clinical presentation[55].

SARS-CoV-2 is a betacoronavirus, the same genus as two other Severe Acute Respiratory Syndrome coronavirus (SARS-CoV) and Middle East Respiratory Syndrome coronavirus (MERS-CoV). All betacoronaviruses have four structural proteins: membrane (M), envelope (E), nucleocapsid (N), and spike (S)[56]. Previous betacoronavirus research has shown that loss of the E protein significantly attenuates viral replication which suggests that E is not

strictly required for replication[57,58], however it is unclear how generalizable this finding is for coronaviruses as the alphacoronavirus transmissible gastroenteritis virus has no replication in the absence of E[59]. These findings led researchers to question whether E-deficient coronaviruses could serve as live attenuated vaccine candidates. Animal studies using an E-deficient SARS-CoV as a live attenuated vaccine found the attenuated virus generated significant protection against subsequent infection challenges[60–62]. This was recapitulated using SARS-CoV with an intact E gene that contained partial deletions of either the N- or C-terminus of the protein and animals were again protected[63]. Further studies have suggested that E protein may be a driver of immunopathology in hosts with animal studies showing a dependency on E function for inflammatory cytokine production[64] and *ex vivo* studies identifying E protein as a ligand for TLR2 signaling in macrophages[65]. Additionally, the coronavirus E protein has ion channel activity that can be targeted using channel blockers[66]. This function is particularly compelling in light of the early effectiveness of the antivirals, amantadine and ramantadine, which target the ion channel activity of the influenza A M2 protein[67]. Taken together, this body of work highlights the importance of E in viral pathogenesis and replication, as well as the possibility of targeting it for vaccine and antiviral development. Chapter 2 of this dissertation presents work on the SARS-CoV-2 E protein that explores the structure and function of the protein in addition to its potential as a drug target.

CHAPTER 1: COMPARATIVE ANALYSIS OF T CELL SPATIAL PROTEOMICS AND THE INFLUENCE OF HIV EXPRESSION

Introduction

Spatial proteomics is a methodologically diverse and rapidly growing field within mass spectrometry (MS) that aims to understand the subcellular localization of the human proteome[28,29,31,34,35,38,68]. While initial efforts focused on establishing techniques and reference maps for various cell lines, recent work by the Cristea group expanded the field to understand the whole-cell effects of viral infection using human cytomegalovirus (HCMV) as a prototype[38]. This work led to novel findings on the importance of peroxisomes in herpesvirus infectivity[69], exemplifying the power of these methods for uncovering new viral biology. However, as this was a first in its class study, how different methodologies might impact the results of viral studies using spatial proteomics is unclear. Using the well-characterized HIV-1 as a model virus system, we aimed to compare the output of several published spatial proteomic analysis pipelines[39–41,70] as a survey of established methods.

To model HIV expression, we used a Jurkat T cell line that harbors a doxycycline-regulated HIV-1 genome. These cells were previously developed by our group to generate nearly homogenous HIV-positive cell populations for MS analysis[71]. As an additional biological comparator, we examined both wild-type (WT) virus and a virus lacking the accessory gene *nef* (ΔNef). Nef is a small (27 kDa), myristoylated membrane-associated accessory protein expressed early during the viral replication cycle[72,73]. Nef increases viral growth-rate and infectivity[74], and it dysregulates the trafficking of cellular membrane proteins such as CD4, class I MHC, and proteins involved in T cell activation such as CD28[75] and p56-Lck[76]. Some of these activities enable the virus to evade immune detection[11,77]. Here we use inducible Jurkat T cell lines containing either WT or ΔNef HIV-1_{NL4-3} provirus and compare the spatial proteome of uninduced cells to cells post-induction with doxycycline. To fractionate the cells, we used a

modified version of the Dynamic Organellar Mapping protocol[28,35] with additional centrifugation steps[31] to enhance organellar resolution, then analyzed the fractions by MS using TMT multiplexing.

Following the generation and processing of MS data, two broad steps are required for spatial proteomics: classification and hit determination. For classifying detected proteins into cellular organelles we compared two methods from pRoloc, an R software package developed by the Lilley lab[70]. The first was support vector machine (SVM) classification which outputs a label for each protein and an algorithm specific confidence score that can be used to threshold assignments[29]. The second was a Bayesian approach called t-augmented Gaussian mixture modeling with *maximum a posteriori* estimates (TAGM-MAP) which outputs a label for each protein and an actual probability of assignment[39]. To gauge the quality of these classifications, we compared the two methods using the QSep metric developed by the Lilley group[37], which quantifies the separation, or resolution, of the organelles in question. We additionally cross-referenced our organellar assignments to existing organellar proteome databases[18,30,78,79].

After classification, data were analyzed for translocating proteins following HIV expression. We compared three different methods for determining protein translocations: label-based movement, translocation analysis of spatial proteomics (TRANSPIRE)[40], and Bayesian analysis of differential localization experiments (BANDLE)[41]. Label-based movement relies strictly on identifying proteins that are consistently classified in one organelle prior to a cellular perturbation, then consistently classified in another organelle following the perturbation; this method was employed by the Cristea group in their HCMV study[38]. TRANSPIRE is a refined methodology from the Cristea lab that relies on generating synthetic translocations from proteins of known localization and uses Bayesian analysis to determine the likelihood of proteins of unknown localization behaving in a manner consistent with anticipated translocations following a cellular perturbation[40]. Lastly, BANDLE is another method developed by the Lilley group that takes replicated data, both with and without a perturbation, and uses Bayesian analysis to yield

a ranked list of possible translocations with their associated likelihood of occurrence[41]. We compared the hits from these various methods by cross-referencing hits with a previous study of the HIV interactome[19] as well as the more broad NIH HIV-1 Human Interaction Database[80].

From these comparisons we found that the performance of different classifiers is organelle-dependent and shows varied effects from HIV expression. As determined by agreement with previously published organellar proteomes, classification with TAGM-MAP showed increased accuracy in mitochondrial and ER-classified proteins, while SVM outperformed TAGM-MAP with nuclear, cytosolic, and plasma membrane-classified proteins. We also observed generally higher performance for protein translocation using BANDLE on SVM-classified data when compared to the HIV interactomes. BANDLE analysis of both WT and ΔNef data identified known Nef interactors involved in T cell activation and the coatomer complex. Finally, we found that SVM classification showed higher consistency and was less sensitive to HIV-dependent noise. These findings illustrate the complexities in choosing a computational method for spatial proteomics study and serve as a foundation for additional studies.

Experimental Procedures

Experimental design and statistical rationale

All fractionation experiments with mass spectrometric analysis were performed in technical triplicate for each condition (uninduced and induced), with two biological replicates for wild-type and ΔNef NL4-3 Jurkat cells (Fig. 1.1A). This yielded a total of 6 uninduced and 6 induced technical replicates for each virus type. Biological replicates were prepared on separate days and analyzed by mass spectrometry on separate days. Western blotting and flow cytometry were performed on each technical replicate. Analyses for QSep (Fig. 1.2B and C) used Welch's t-test to determine statistical significance.

Cell culture

The doxycycline-inducible NL4-3 HIV-1 and NL4-3 ΔNef Jurkat cell lines were previously described[71,81]. The replication-incompetent genome used was based on pNL4-3 but lacked most of the 5' U3 region, encoded a self-inactivation deletion in the 3' LTR, and contained the V3 region from the R5-tropic 51-9 virus[82] to prevent the cell-cell fusion of the Jurkat T cells used herein, which do not express CCR5. Inducible cells were cultured in RPMI 1640 media supplemented with penicillin/streptomycin (pen/strep) and 10% Tet-free fetal bovine serum (FBS), as well as puromycin (1 µg/mL) and G418 (200 µg/mL) to maintain persistence of the tetracycline trans-activator and the inducible genome. Cells were passaged every two days to keep concentrations between $3.5x10^5$ and $1x10^6$ cells/mL. Cells were maintained at 37°C, 5% CO_2, and 95% humidity.

Doxycycline induction and fractionation

On the day prior to fractionation, $2.016x10^9$ cells were plated at $6x10^5$ cells/mL in T75 flasks at a total volume of 40 mL/flask. Half of these cells were induced to express HIV-1/HIV-1ΔNef with doxycycline (1 µg/mL) for 18 hours, while the other half remained uninduced. Following induction, cells of each condition, i.e. uninduced and induced, were split into three technical replicates, and then centrifuged at 500x*g* for 5 min at 4°C. Each technical replicate was pooled into a single 50 mL tube using ice cold 1X PBS, then counted by hemocytometer. From each technical replicate, $3x10^8$ cells were fractionated. Two aliquots of cells were taken from each technical replicate for whole cell western blots and testing induction by flow cytometry.

The fractionation protocol used here is derived from the Dynamic Organellar Maps method[28] with additional centrifugation steps[31] and TMT-based MS analysis rather than SILAC[35]. Cells for fractionation were centrifuged at 500x*g* for 5 min at 4°C then resuspended in ice-cold PBS and incubated for 5 min on ice. Cells were again centrifuged at 500x*g* for 5 min at

4°C, then resuspended in ice-cold hypotonic lysis buffer (25 mM Tris-HCl (pH 7.5), 50 mM sucrose, 0.5 mM $MgCl_2$, and 0.2 mM EGTA in water) and incubated for 5 min on ice. Using a 7 mL Dounce homogenizer, cells were homogenized with 20 full strokes of the tight pestle. Cell homogenates were then immediately transferred to a 13 mL (14x89 mm) ultracentrifuge tube with sufficient ice-cold hypertonic sucrose buffer (1.25 M sucrose, 25 mM Tris-HCl (pH 7.5), 0.5 mM $MgCl_2$, and 0.2 mM EGTA in water) to restore 250 mM sucrose concentration. All replicates were then centrifuged at 1,000x*g* for 10 min at 4°C in a Beckman Coulter ultracentrifuge (SW-41 Ti rotor), balancing each tube with balance buffer (250 mM sucrose, 25 mM Tris-HCl (pH 7.5), 0.5 mM $MgCl_2$, and 0.2 mM EGTA in water). Supernatants were transferred to a fresh ultracentrifuge tube, balanced with balance buffer, then fractionated using the following differential centrifugation protocol: 3,000x*g* for 10 min, 5,400x*g* for 15 min, 12,200x*g* for 20 min, 24,000x*g* for 20 min, 78,400x*g* for 30 min, 110,000x*g* for 35 min, and 195,500x*g* for 40 min All centrifugation steps were performed at 4°C with pellets from each spin being resuspended in SDS buffer (2.5% SDS and 50 mM Tris-HCl (pH 8.0) in water). Fractions were then heated for 10 minutes at 72°C. Protein content of each fraction was quantified in triplicate using a bicinchoninic acid (BCA) protein assay (Thermo-Fisher).

Confirmatory western blots and p24 flow cytometry

Prior to mass spectrometric analysis of fractions, induction and fractionation were evaluated by flow cytometry and western blotting (Fig. 1.1B and C). For p24 flow cytometry, an aliquot of $2x10^6$ cells from each technical replicate were pelleted at 500x*g* for 5 min at 4°C then resuspended in ice-cold FACS buffer (2% FBS and 0.1% sodium azide in 1X PBS). The cells were again pelleted at 500x*g* for 5 min at 4°C then resuspended in Cytofix/Cytoperm reagent (BD Biosciences) and incubated on ice for 30 min Following fixation/permeabilization, cell suspensions were diluted with wash buffer and pelleted at 500x*g* for 5 min at 4°C. Cells were resuspended in p24 primary antibody solution (1:100 dilution of p24-FITC antibody clone KC57

(Beckman Coulter) diluted in perm/wash buffer) and incubated on ice for 30 min in darkness. Ice-cold FACS buffer was added to each sample and cells were pelleted at 500x*g* for 5 min at 4°C. The intracellular p24 was analyzed using an Accuri C6 flow cytometer (BD Biosciences). Uninduced cells had an average p24+ population of 0.27% (S.D. = 0.20) and live cell population of 85.78% (S.D. = 3.37). Induced cells had an average p24+ population of 94.85% (S.D. = 1.23) and live cell population of 79.25% (S.D. = 4.35).

An aliquot of $1x10^7$ cells from each technical replicate was lysed in SDS buffer and probe sonicated on ice until no longer viscous. 3,000x*g* fractions were also probe sonicated. The samples were mixed with 4X loading buffer (200 mM Tris-HCl (pH 6.8), 8% SDS, 40% glycerol, 200 mM tris(2-carboxyethyl)phosphine-HCl (TCEP), and 0.04% bromophenol blue in water) and proteins were then separated on 10% SDS-PAGE gels at a constant 70V. Proteins were transferred to polyvinylidene difluoride (PVDF) membranes for 1 hour using the Trans-Blot turbo (BioRad) system using standard conditions. The membranes were blocked in 5% milk in 1X PBS-T for 30 min at room temperature prior to incubation with primary antibodies diluted in 1% milk and 0.05% sodium azide in 1X PBS-T: sheep anti-Nef (gift from Celsa Spina, diluted 1:3,000), mouse anti-p24 (Millipore, diluted 1:500), Chessie8 (mouse anti-gp41, NIH AIDS Research and Reference Reagent program[83], diluted 1:10,000), rabbit anti-Vpu (NIH AIDS Research and Reference Reagent program ARP-969, contributed by Dr. Klaus Strebel, diluted 1:1,000), and mouse anti-GAPDH (GeneTex, diluted 1:5,000). The blots were washed and probed with either horseradish peroxidase-conjugated goat anti-mouse, HRP-goat anti-rabbit, or HRP-rabbit anti-sheep secondary (BioRad) diluted 1:3,000, incubating for 1 hour at room temperature on a shaker. Apparent molecular mass was estimated with PageRuler protein standard (Thermo Scientific). Blots were imaged using Western Clarity detection reagent (BioRad) before detection on a BioRad Chemi Doc imaging system with BioRad Image Lab v5.1 software.

Sample digestion for mass spectrometry

Disulfide bonds were reduced with 5 mM TCEP at 30°C for 60 min and cysteines were subsequently alkylated (carbamidomethylated) with 15 mM iodoacetamide (IAA) in the dark at room temperature for 30 min Proteins were then precipitated with 9 volumes of methanol, pelleted and resuspended in 1M urea, 50 mM ammonium bicarbonate. Following precipitation, protein concentration was determined using a BCA protein assay. A total of 0.2 mg of protein was subjected to overnight digestion with 8.0 µg of mass spec grade Trypsin/Lys-C mix (Promega). Following digestion, samples were acidified with formic acid (FA) and subsequently 150 ug peptides were desalted using AssayMap C18 cartridges mounted on an Agilent AssayMap BRAVO liquid handling system, C18 cartridges were first conditioned with 100% acetonitrile (ACN), followed by 0.1% FA. The samples were then loaded onto the conditioned C18 cartridge, washed with 0.1% FA, and eluted with 60% MeCN, 0.1% FA. Finally, the organic solvent was removed in a SpeedVac concentrator prior to LC-MS/MS analysis.

TMT Labeling

Peptide concentration was determined using a Nanodrop, and a total of 15 µg of peptide was then used for TMT labeling, each replicate serving as a multiplex. Briefly, dried peptide sample was resuspended in 200 mM HEPES (pH 8) and incubated for 1 h at room temperature with one of the TMT10-plex reagents (ThermoFisher) solubilized in 100% anhydrous ACN. Reactions were quenched using a 5% hydroxylamine solution at 1-2 µl per 20 µl TMT reagent. The multiplexed samples were then pooled and dried in a SpeedVac. The labeled peptides were resuspended in 0.1% FA. After sonication for 1min, the sample was desalted manually using SepPak; the column was first conditioned with 100% ACN, followed by 0.1% FA. Sample was loaded, then washed with 0.1% FA and eluted in a new vial with 60% ACN, 0.1% FA. Finally, the organic solvent was removed using a SpeedVac concentrator prior to fractionation.

High pH Reverse-Phase Fractionation

Dried samples were reconstituted in 20mM ammonium formate (pH ~10) and fractionated using a Waters ACQUITY CSH C18 1.7 µm 2.1 × 150 mm column mounted on a MClass Ultra Performance Liquid Chromatography (UPLC) system (Waters corp., Milford, MA) at a flow rate of 40 µl/min with buffer A (20 mM ammonium formate pH 10) and buffer B (100% ACN). Absorbance values at 215 nm and 280 nm were measured on a Waters UV/Vis spectrophotometer, using a flowcell with a 10 mm path length. Peptides were separated by a linear gradient from 5% B to 25% B in 62.5 min followed by a linear increase to 60% B in 4.5 min and 70% in 3 min and maintained for 7 min before increasing to 5% in 1 min Twenty-four fractions were collected and pooled in a non-contiguous manner into twelve total fractions. Pooled fractions were dried to completeness in a SpeedVac concentrator.

LC-MS^3 Analysis

Dried samples were reconstituted with 0.1% FA and analyzed by LC-MS/MS on an Orbitrap Fusion Lumos mass spectrometer (Thermo) equipped with an Easy nLC 1200 ultra-high pressure liquid chromatography system interfaced via a Nanospray Flex nanoelectrospray source (Thermo). Samples were injected on a C18 reverse phase column (25 cm x 75 um packed with Waters BEH 1.7 um particles) and separated over a 120-min linear gradient of 2-28% solvent B at a flow rate of 300nL/min The mass spectrometer was operated in positive data-dependent acquisition mode.

Parameter settings were set as follows: FT MS1 resolution (120 000) with AGC target of 1e6, ITMS2 isolation window (0.4 m/z), IT MS2 max. inject time (120 ms), IT MS2 AGC (2E4), IT MS2 CID energy (35%), SPS ion count (up to 10), FT MS3 isolation window (0.4 m/z), FT MS3 max. inject time (150 ms), FT MS3 resolution (50 000) with AGC target of 1e5. A TOP10 method was used where each FT MS1 scan was used to select up to 10 precursors for

interrogation by CID MS2 with readout in the ion trap. Each MS2 was used to select precursors (SPS ions) for the MS3 scan which measured reporter ion abundance.

Mass spectrometry spectra identification

Raw files were analyzed using Proteome Discoverer v2.3 (Thermo Fisher Scientific). MS/MS spectra were searched against a concatenated database containing Uniprot human and HIV-1 proteins (downloaded 02/03/20) and reverse decoy sequences using the Sequest algorithm[84]; the database contained 20,367 total entries. Mass tolerance was specified at 50 ppm for precursor ions and 0.6 Da for MS/MS fragments. Static modifications of TMT 10-plex tags on lysine and peptide n-termini (+229.162932 Da) and carbamidomethylation of cysteines (+57.02146 Da), and variable oxidation of methionine (+15.99492 Da) were specified in the search parameters. Data were filtered to a 1% false discovery rate at the peptide and protein level through Percolator[85] using the target-decoy strategy[86]. TMT reporter ion intensities were extracted from MS3 spectra within Proteome Discoverer to perform quantitative analysis.

Computational analysis

Matching biological replicates were combined (i.e. WT biological replicate 1 and 2), then analyzed using the various pipelines described. The *Homo sapiens* (“hsap”) marker set from pRoloc was used in all cases. For classification and hit generation, only the proteins commonly detected across matched biological replicates were analyzed to allow for consistency in comparing methods on the same data set.

The pRoloc implementation of SVM[70] was performed on row-normalized data sets, while the pRoloc implementation of TAGM-MAP[39] required PCA transformation and no row-normalization with the first four principal components carried forward. The PCA transformation was used because of floating point arithmetic errors that arose because of highly correlated features. Default parameters for algorithms were used excepting the following:

SVM hyperparameter classification: 10 times 10-fold cross-validation

SVM classification threshold: median algorithm score for each organelle

TAGM-MAP model training: 200 iterations

BANDLE: 6 chains

TRANSPIRE was run on averaged row-normalized datasets, i.e., technical replicates were row-normalized then values for each feature were averaged for each protein across matched technical replicates. Organelles were combined into 5 groups: 1) Golgi apparatus/plasma membrane/endoplasmic reticulum/peroxisomes/lysosomes, 2) cytosol/actin cytoskeleton/proteasome, 3) nucleus, 4) mitochondria, and 5) 40S/60S ribosome. The number of inducing points and the kernel function were chosen from amongst the suggested values in the TRANSPIRE documentation. For these datasets, 75 inducing points and the squared exponential kernel performed best and were used in the analysis.

The average distribution of proteins across organelles was calculated by determining the average organellar distribution for a single technical replicate, then averaging the values of matched technical replicates. Marker profiles were generated by averaging the behavior of markers for a given organelle within a technical replicate, then averaging those values across technical replicates for each organelle. Organellar QSep scores were calculated by averaging the individual QSep scores between two organelles across all matched technical replicates, then plotting the distribution of those averages.

Comparisons to the Human Protein Atlas (HPA) were completed by combining several HPA subcellular localization annotations to align with the organelles used by pRoloc:

1. Nuclear membrane, nucleoli fibrillar center, nucleoli rim, nucleoli, kinetochore, mitotic chromosome, nuclear bodies, nuclear speckles, and nucleoplasm: Nucleus
2. Actin filaments and focal adhesion sites: Actin Cytoskeleton
3. Plasma membrane and cell junctions: Plasma Membrane

Remaining designations within the HPA beyond the above and those in common with pRoloc's "hsap" markers were not considered. The 40S Ribosome, 60S Ribosome, and Proteasome classes from the SVM and TAGM-MAP classified data were collapsed into the Cytosol label.

Thresholds for Figures 1.6 and S1.8 were determined by dividing the size of the Jäger HIV interactome[19], 453 proteins, or the NIH HIV interactome[80], 4,628 proteins, by the predicted human proteome size of 19,773 proteins[87]. G.O. analysis for Figure 1.6B was conducted using the STRING database[88].

Results

Doxycycline-inducible HIV-1_{NL4-3} Jurkat T cells are a scalable and uniform system for subcellular fractionation and proteomic studies

The WT HIV-1 inducible cells used here were previously generated and used for whole-cell quantitative- and phospho-proteomics[71]. To avoid the formation of syncytia, which could alter the subcellular fractionation and subsequent spatial proteomic data, the inducible HIV-1_{NL4-3} genomes were modified with a CCR5-tropic Env protein to avoid cell-cell fusion between the CCR5-negative Jurkat cells. Due to the high induction rates of HIV-1 expression and the scalability of this culture system, we reasoned that it would be amenable to subcellular fractionation by differential centrifugation with subsequent MS analysis (Fig. 1.1A). To determine the optimal time-point for analysis following induction of HIV-1 expression, cells were treated with doxycycline for 0, 4, 8, 12, 16, and 18 hours, and the expression of HIV-1 proteins was detected by western blotting and flow cytometry (Fig. 1.1B and C). WT cells began to express detectable Nef by 4 hours post-induction, and both WT and ΔNef cells expressed p55 Gag precursor (the precursor protein for virion structural proteins) by 8 hours and gp160 (the envelope glycoprotein precursor) by 12 hours. By 18 hours, viral proteins were robustly expressed; about 90-95% of both WT and ΔNef cells were positive by flow cytometry for p24 capsid (a proteolytic product of p55).

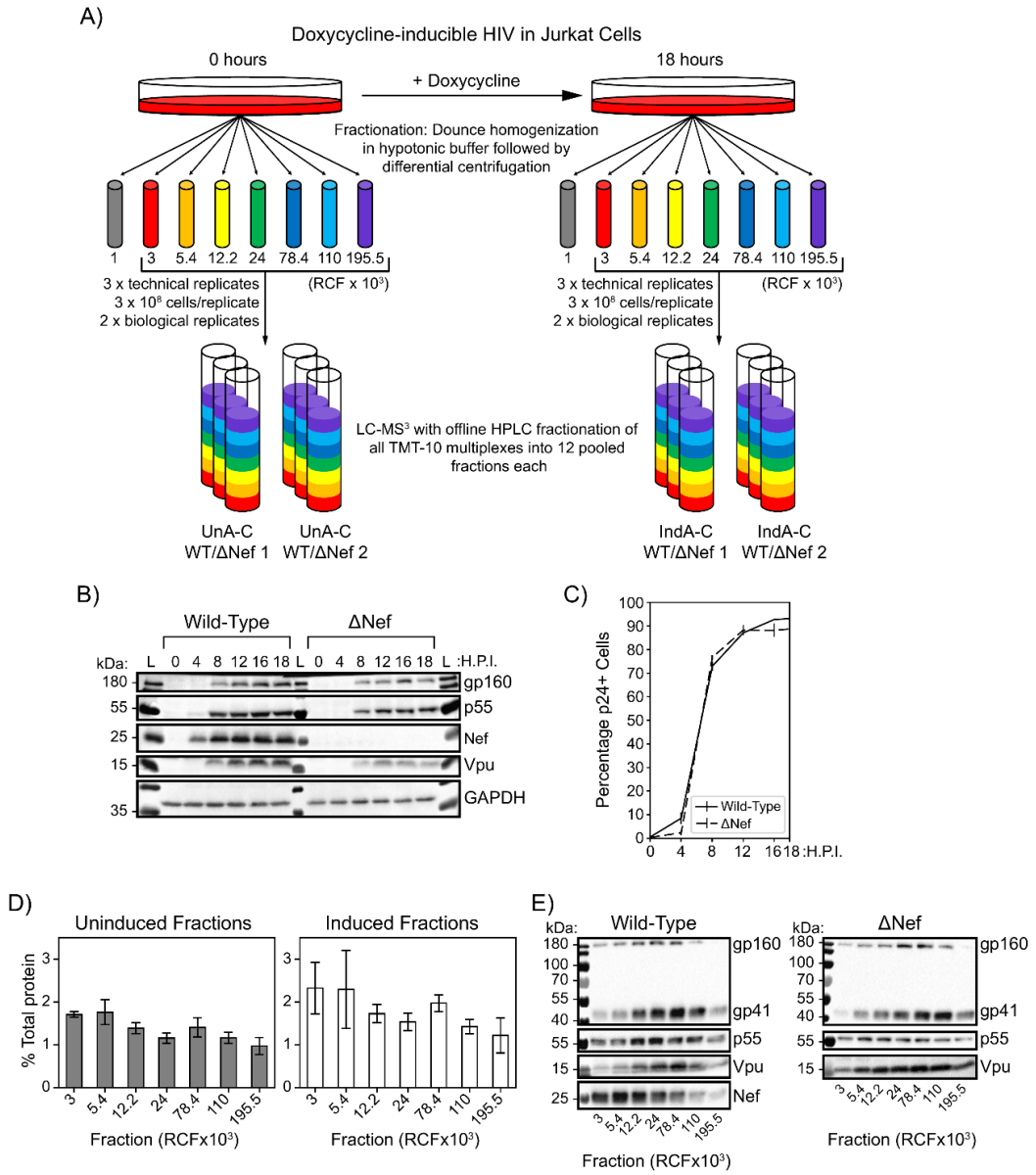

Figure 1.1. Inducible HIV-1 Jurkat cell lines yield a near pure population of HIV-expressing cells suitable for fractionation by differential centrifugation.

A) Equal numbers of doxycycline-inducible wild-type and ΔNef HIV Jurkat cells were induced or left uninduced for 18 hours then fractionated by Dounce homogenization in a hypotonic lysis buffer. Cell homogenates were put through a differential centrifugation protocol, discarding the nuclear pellet (1,000x*g*) and lysing remaining pellets in 2.5% SDS buffer. Fractions were labeled for TMT-10 multiplexing and further offline HPLC fractionation. All multiplexes were run for 3 hours on LC-MS^3. B) Western blot showing induction of HIV p55, gp160, gp41, Nef, and Vpu with a GAPDH loading control. Cells were induced for 0, 4, 8, 12, 16, and 18 hours, lysed, then a portion of these cell lysates was run on 10% SDS-PAGE gels. C) Flow cytometry analysis of remaining sample from 1B. HIV-1 expression peaked at ~95% of cells p24+ by 18 hours. D) Average percentage of total cellular protein detected in each fraction by BCA protein assay. Bars represent the mean value for a given fraction based on the average from each biological replicate. Error bars are one standard deviation. All BCA assays were performed in technical triplicate on 10-fold dilutions for each biological replicate. E) Western blots for cell fractions of inducible wild-type HIV Jurkat cells (left) and ΔNef HIV Jurkat cells (right), 18 hours post-induction. Blots shown are representative of both biological replicates.

Subcellular fractionation was performed 18 hours post-induction; the cells were mechanically ruptured with a Dounce homogenizer in hypotonic solution, then subjected to a differential centrifugation protocol before preparation for quantitative, multiplexed MS analysis. Uninduced and induced cells were handled in technical triplicate for each biological replicate (n=2). We used a modified version of the Dynamic Organellar Mapping (D.O.M.) protocol[28,35] with additional fractions generated at 110,000x*g* and 195,500x*g* to increase the resolution of the classification analysis; a similar method of expanded differential centrifugation fractionation was previously described by the Lilley group[36]. As a quality control before MS, protein yields were quantified for each fraction (Fig. 1.1D). The post-nuclear fractions accounted for only ~10-15% of total cellular protein, presumably because nuclear proteins and soluble cytoplasmic proteins that failed to pellet at 195,500x*g* were discarded, leaving primarily membranous organelles or organellar fragments and large, cytoplasmic complex proteins in the fractions analyzed. We also observed decreasing protein yields across the fractions, with an increase in the 78,400x*g* fraction, consistent with the original D.O.M. study using HeLa cells[28]. In further support of differential fractionation, varied abundances of viral proteins across the fractions in cells expressing either the WT or ΔNef genomes were observed by western blotting (Fig. 1.1E). Following confirmation of differential fractionation, we analyzed all fractions by LC-MS^3 with TMT-10 multiplexing (Fig. 1.1A).

To determine the consistency of the MS analysis we used unsupervised hierarchical clustering by Spearman correlation coefficient for the individual fractions. We found that for both the WT and ΔNef data the fractions clustered by *g*-force rather than biological replicate (SuppFile 1.1 and 1.2), suggesting consistent quantification values. Because the WT and ΔNef Jurkat cell lines represent individual clones for each, we also compared the uninduced fractions of the WT and ΔNef data to each other. This comparison showed that fractions still clustered by *g*-force rather than HIV genome (SuppFile 1.3).

SVM shows greater organellar resolution than TAGM-MAP even with stringent thresholds of classification for TAGM-MAP

To classify the fractionation data and identify translocating proteins, we employed a variety of previously published methods (Fig. 1.2A). As several resources detail known HIV interactors[19,80], we primarily focused on comparing classification and translocation identification methods using our WT data. In subsequent analyses, we examined the ΔNef data to determine the power of various methods in identifying Nef-specific effects.

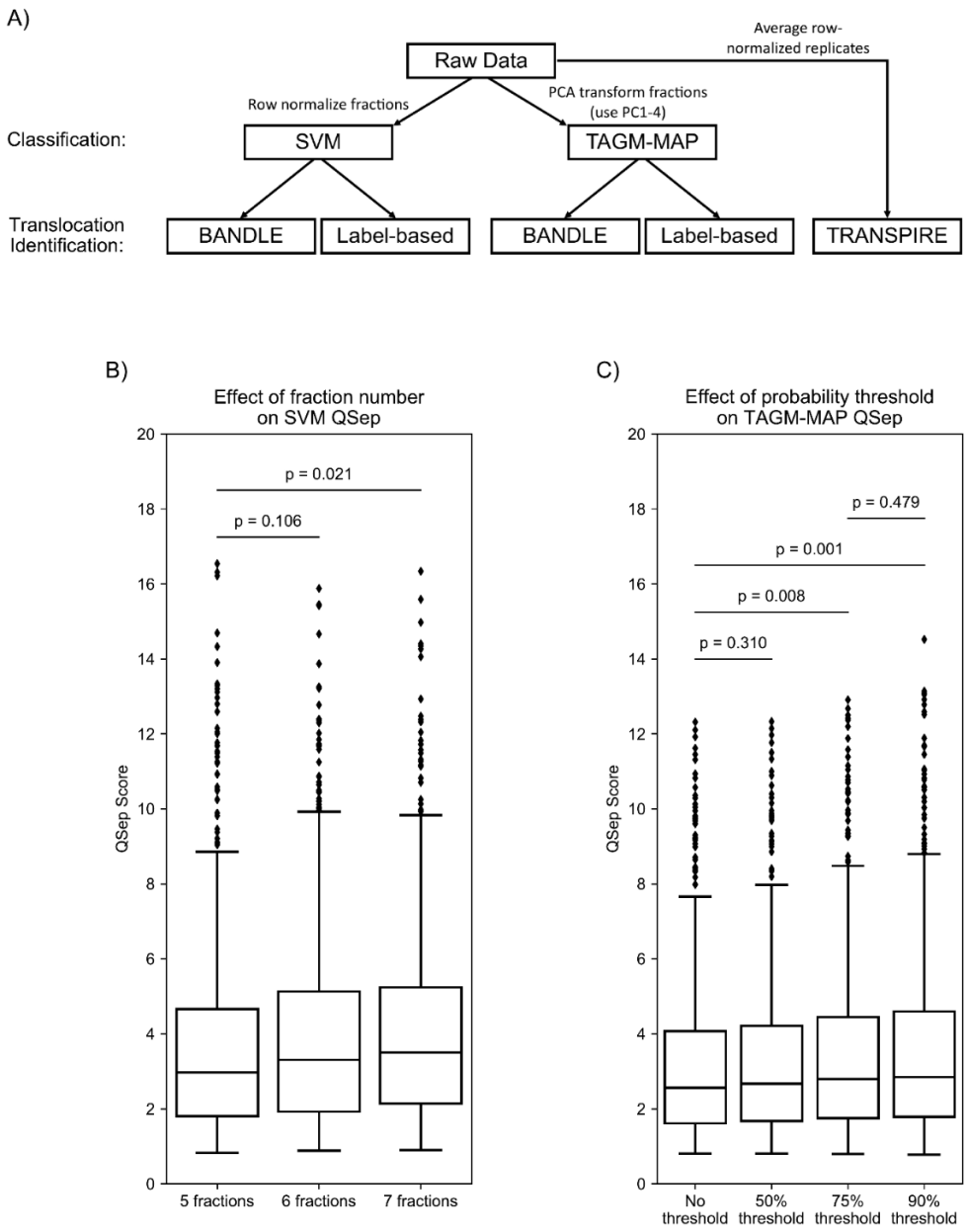

Figure 1.2. Analysis of fractionation data reveals increased organellar resolution from added fractions and thresholding TAGM-MAP data.

A) Diagram of the computational methods used here. For SVM classification, the raw data of individual technical replicates were row normalized. For TAGM-MAP classification, the raw data of individual technical replicates were PCA transformed, with the first four principal components (PC1-4) carried forward for analysis. Both SVM and TAGM-MAP classified data were fed into BANDLE or label-based movement analysis. Lastly, for analysis with TRANSPIRE, individual technical replicates were row normalized then averaged together. B) QSep scores for SVM analysis of WT uninduced samples using the original 5 fractions described by Itzhak et al.[28], adding a 110,000x*g* fraction (6 fractions), or adding both a 110,000x*g* and a 195,500x*g* fraction (7 fractions). C) QSep scores for TAGM-MAP analysis of WT uninduced samples comparing using no threshold for remaining classified, a 50% chance of classification, a 75% chance of classification, or a 90% chance of classification. Statistical significance is calculated using a two-sided, independent Student's t-test with Welch's correction for unequal variance. Boxplots show median, not mean, line.

For classification, proteins were classified using either the pRoloc implementation of SVM or TAGM-MAP. As the differential centrifugation protocol employed here is a modified version of the D.O.M. method which generates only 5 fractions[28], we first examined whether our two additional fractions improved organellar resolution. The D.O.M. method classifies proteins with SVM, so we compared the resolution of organelles with the QSep analysis[37] using the first 5 fractions for SVM classification, then the first 6 fractions, and finally all 7 fractions (Fig. 1.2B). We found that while the addition of the 110,000x*g* spin alone had no significant effect on organellar resolution as compared to the original method, the subsequent addition of the 195,500x*g* spin yielded a significant increase from a mean QSep score of 3.74 to 4.05 (median scores 2.97 and 3.50, respectively). In light of this, all subsequent analyses on the SVM data were performed on the full 7 fractions.

To determine if an alternate method for classification would perform better than SVM, we also tested the pRoloc implementation of TAGM-MAP. The outputs from TAGM-MAP give both a localization and a probability that the given protein is located in that organelle. These probabilities allowed us to test the effect of different probability thresholds on TAGM-MAP's QSep scores. While using a 50% threshold, i.e. converting all proteins with a probability of localization lower than 50% to an "unknown" designation, showed no significant effect, 75% and 90% thresholds both showed significant gains over no thresholding (Fig. 1.2C). A 90% threshold showed no significant increase in QSep scores over the 75% threshold, so subsequent analyses employed the 75% threshold for TAGM-MAP classification. Of importance, we observed that the QSep scores from SVM classification were on average higher than those from TAGM-MAP even when comparing TAGM-MAP's highest condition (90% probability threshold, average score of 3.55) to SVM's lowest condition (5 fractions, average score of 3.74).

SVM classifies proteins more consistently than TAGM-MAP

We next wanted to understand how the SVM and TAGM-MAP methods compared for consistency of classification across WT replicates (Fig. 1.3A and B). Both SVM (Fig. 1.3A) and TAGM-MAP (Fig. 1.3B) showed a low percentage (~10-15%) of proteins that were classified identically in 6 out of 6 technical replicates for either WT uninduced or induced. However, allowing for a majority of replicates, i.e. 4 out of 6, gave ~70-75% of proteins as classified consistently by SVM (Fig. 1.3A). This compared to ~50-55% of proteins classified to a similar consistency by TAGM-MAP (Fig. 1.3B). HIV expression modestly decreased the consistency of both SVM and TAGM-MAP (~5% difference), suggesting an increase in experimental noise from HIV expression.

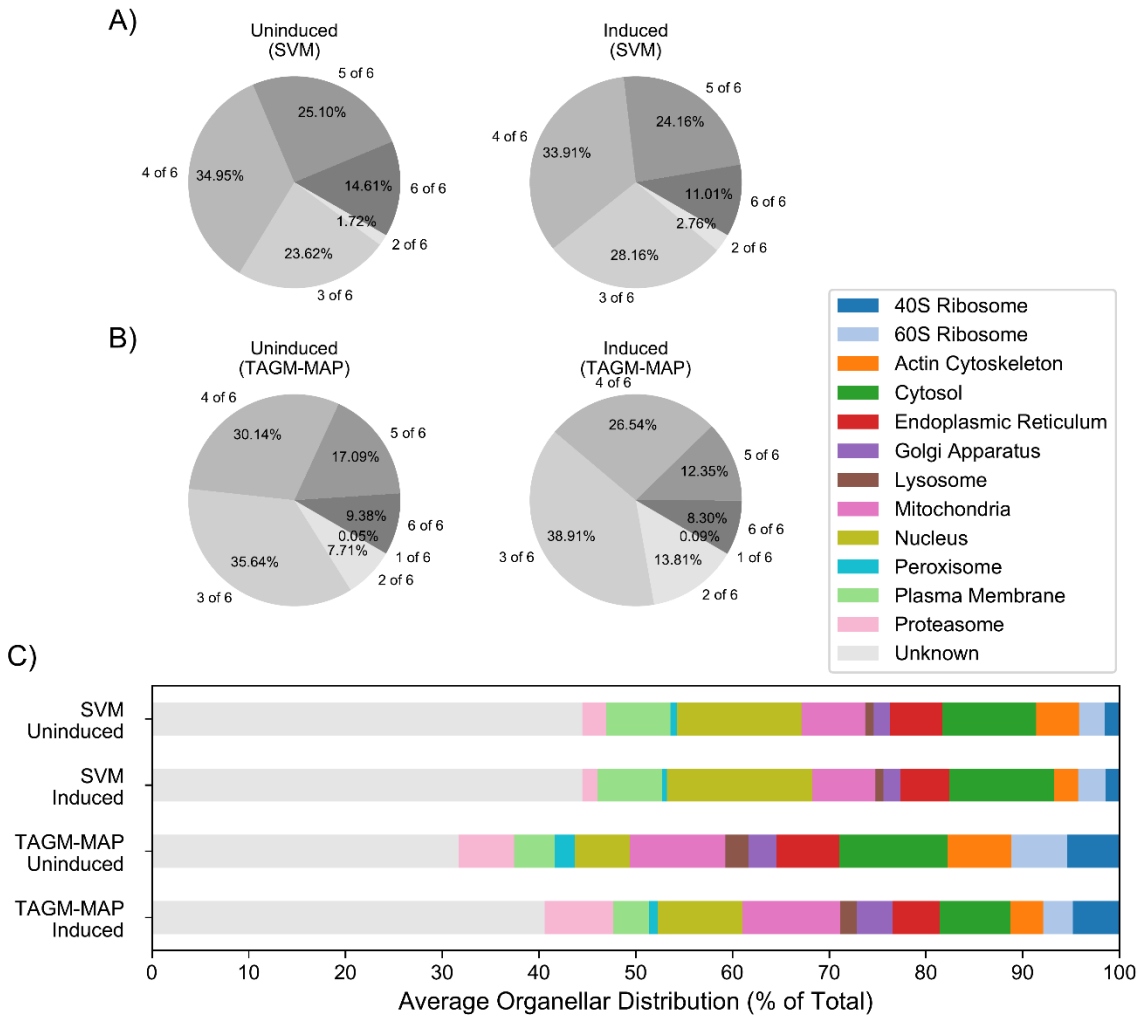

Figure 1.3. Classification with SVM shows greater consistency than TAGM-MAP classification. A) Proteins were classified by SVM and the most frequent organellar classification was identified along with its frequency, i.e. number of technical replicates classified as such. Left pie chart shows consistency of classification for WT uninduced replicates and right pie chart shows WT induced replicates. B) Same as A), but classification by TAGM-MAP. C) Average distribution of proteins across organelles for each indicated condition. All charts consider the same common proteins found across all WT replicates (4,765 proteins).

Looking at the average distribution of proteins across organelles, we found that SVM yielded a higher percentage of proteins that reverted to an unknown designation (Fig. 3C, 44% of proteins); this may partly explain the higher QSep scores generally seen for SVM compared to TAGM-MAP (Fig. 1.2). However, this percentage is stable between WT uninduced and induced replicates, while the lower percentage of unknown proteins (32% for uninduced and 41% for induced) for TAGM-MAP is more sensitive to HIV expression. Similar trends were seen

within the ΔNef data (SuppFile 1.4); marker behavior for WT (SuppFile 1.5) and ΔNef (SuppFile 1.6) is also similar, which likely explains the consistent trends. These data show a greater consistency for SVM classification and additionally suggest that SVM is less susceptible to noise introduced into data by HIV expression.

Agreement between SVM and TAGM-MAP classification is organelle-dependent and is variably affected by HIV expression

To determine the concordance of SVM and TAGM-MAP for classification, we examined all proteins that were classified consistently in at least 4 of 6 WT replicates for both SVM and TAGM-MAP. We found more such proteins for the uninduced replicates (Fig. 1.4A), 1,863 proteins, as compared to the induced replicates (Fig. 1.4B) with 1,448 proteins. This difference may be attributable to the decrease in classification consistency caused by HIV expression for both SVM and TAGM-MAP, which would be accentuated by the increased susceptibility of TAGM-MAP to HIV-dependent noise. Of these consistently classified proteins, HIV expression minimally affected classifier agreement; 65% agreed between SVM and TAGM-MAP for WT uninduced and 69% agreed between SVM and TAGM-MAP for induced replicates (see diagonal of heatmaps). However, HIV expression increased the proportion of proteins that were consistently designated unknown by both SVM and TAGM-MAP: in uninduced cells, 40% of proteins agreed upon by the two methods were designated unknown (Fig. 1.4A), while 71% of agreed upon proteins were designated unknown from induced cells (Fig. 1.4B). This shift seems primarily driven by the increase in unknown designations for TAGM-MAP following HIV expression: in uninduced replicates, 52% of proteins designated unknown by SVM agreed with TAGM-MAP, but in induced replicates, 81% of these proteins agreed with TAGM-MAP. Matching trends were seen in ΔNef data (SuppFile 1.7). Taken together, these data suggest that while HIV expression has relatively little effect on the proportion of consistently classified proteins that are agreed upon by the two classifiers, the proportion of these proteins that are

designated unknown increased, and the overall number of consistently classified proteins is decreased.

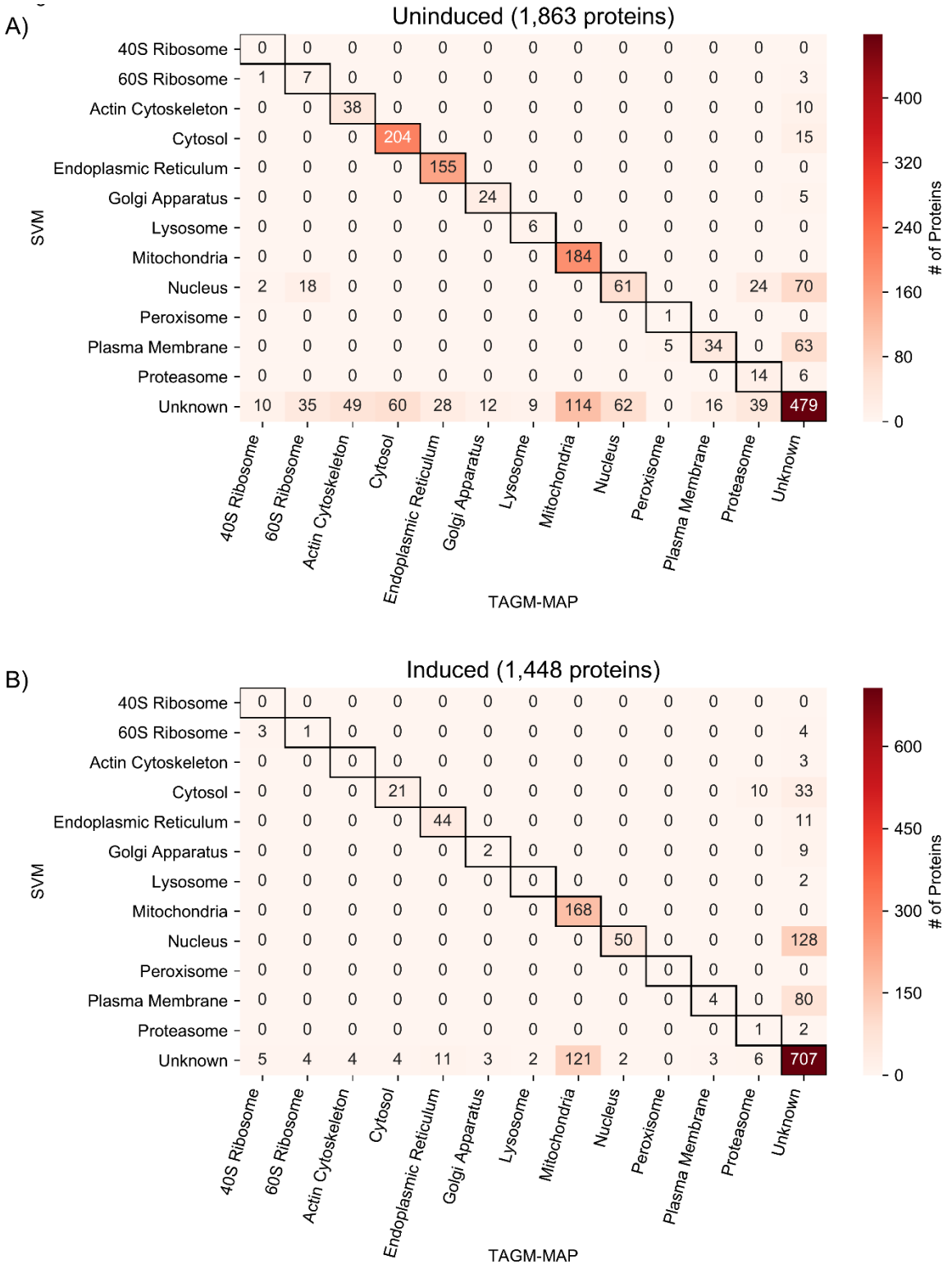

Figure 1.4. Concordance of SVM and TAGM-MAP classifications depends on organelle and expression of HIV.

A) Heat map of common proteins that were consistently classified (proteins classified consistently in at least 4 of 6 replicates) by both SVM and TAGM-MAP for uninduced condition. Annotations indicate number of proteins in a given scenario. B) Same as A) for induced condition.

We found that proteins from the cytosol, ER, and mitochondria were the most frequent among consistently classified proteins. These three organelles also showed the best agreement between SVM and TAGM-MAP for uninduced replicates (Fig. 1.4A, SuppFile 1.7A). However, HIV expression decreased the proportion of cytosolic proteins and ER proteins in agreement between SVM and TAGM-MAP: 73% of all proteins classified as cytosolic and 85% of all proteins classified as ER agreed for WT uninduced replicates, but only 31% of cytosolic proteins and 67% of ER proteins agreed for induced replicates. This decrease was smaller for mitochondrial proteins: 62% for uninduced and 58% for induced. Similar trends for cytosolic and mitochondrial proteins were seen in ΔNef data, but ER proteins showed little change (SuppFile 1.7). These data show a clear organelle-dependent trend in classifier agreement that is variably affected by HIV expression.

TAGM-MAP classification yields higher agreement than SVM classification with reported ER and mitochondria proteomes, but lower agreement in other organelles

To gauge the quality of our classifications, we compared those proteins that were consistently classified, i.e. 4 out of 6 replicates, for WT uninduced to several published spatial proteomes: MitoCarta2.0 database[30], a study of the mitochondrial matrix proteome[78], and a review of lysosome proteomic studies[79] (Fig. 1.5A). Examining those proteins from each study that were detected in our datasets, we found that TAGM-MAP consistently out-performed SVM for mitochondria but performed less well for lysosomes. We also compared only those proteins that received an organellar classification, i.e. we excluded consensus unknown designations, to see if a focus on only proteins that remained classified would change the performance of SVM (orange bars) or TAGM-MAP (dark orange bars). SVM was more responsive to the exclusion of unknown proteins compared to TAGM-MAP, which is likely due to the lower proportion of unknown proteins in the TAGM-MAP uninduced condition.

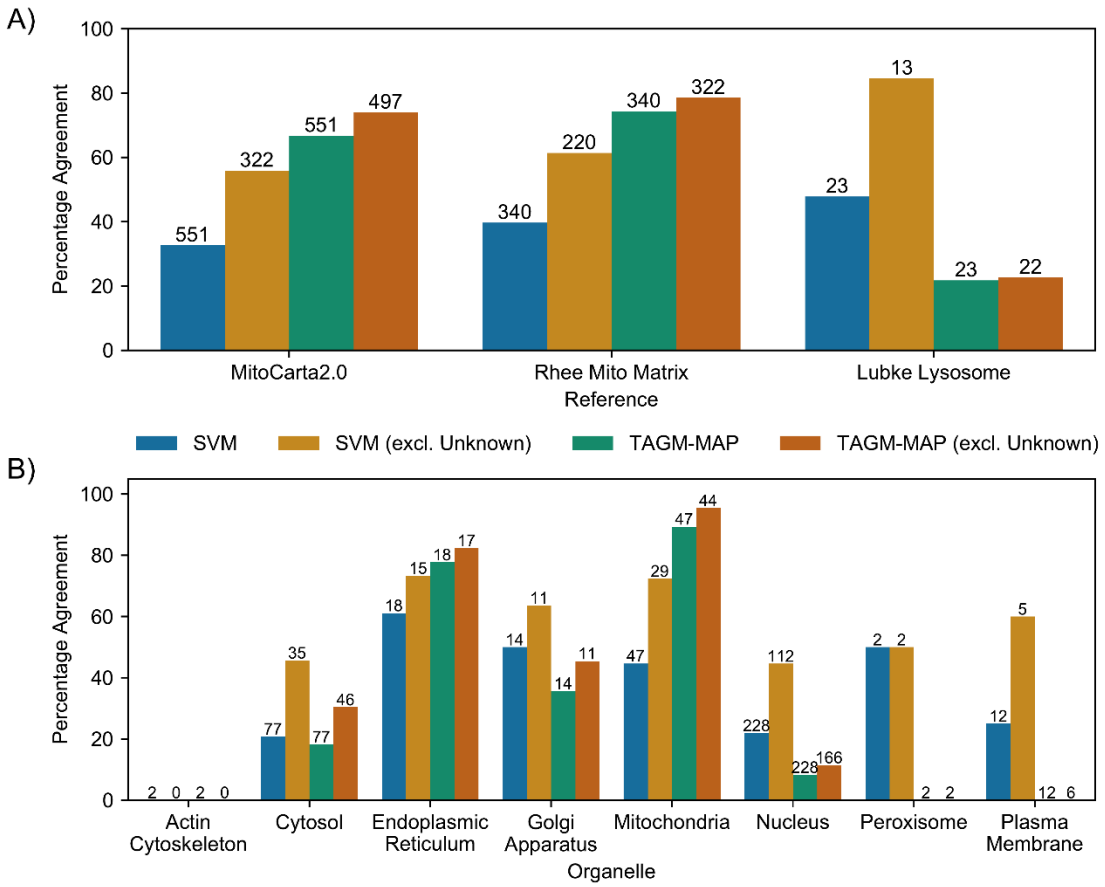

Figure 1.5. Validation of protein classification reveals better performance for ER and mitochondria using TAGM-MAP, but better performance for Golgi apparatus, nucleus, and plasma membrane using SVM. A) Percentage of detected proteins from MitoCarta2.0 database[30], Rhee et al. mitochondrial matrix study[78], or Lubke lysosome proteome[79] that were consistently classified (proteins classified consistently in at least 4 of 6 replicates) in line with the respective reference. Numbers above bars indicate the total number of proteins from that reference that were detected and classified for a given method. B) Proteins classified by SVM or TAGM-MAP were cross-referenced against the Human Protein Atlas and any protein considered to be singularly localized with an Enhanced rating was kept. The percentage of these proteins that were consistently classified by SVM or TAGM-MAP into the HPA-designated organelle is shown. Numbers above bars indicate the number of HPA proteins considered for each organelle. For conditions with Unknown proteins excluded, those proteins that were consistently classified as Unknown were removed from the analysis.

We did a similar analysis for additional organelles by comparing to the Human Protein Atlas (HPA)[18]. To give a baseline to our analysis, we focused on those proteins considered by the HPA to be localized to a single organelle with high confidence (enhanced rating). Of those proteins, we then plotted the percentage that were similarly classified by SVM or TAGM-MAP (Fig. 1.5B). Again, we found that TAGM-MAP outperformed SVM for mitochondrial proteins, and

we saw a similar trend for ER proteins, albeit to a lesser degree. Conversely, SVM outperformed TAGM-MAP in the Golgi apparatus, nucleus, peroxisomes, and plasma membrane, although only two proteins were considered for the peroxisome. Similar to our observations above, the exclusion of unknown proteins yielded a larger increase in percentage agreement for SVM (orange vs blue bars) than TAGM-MAP (dark orange vs green bars); this exclusion also increased the performance in the cytosol for SVM over TAGM-MAP. These data correspond well to those of Figure 1.4A where 114 proteins designated as unknown by SVM were classified as mitochondrial by TAGM-MAP. Similar trends were found within ΔNef data (SuppFile 1.8). Taken together, this suggests that at least in this cell system and using these fractionation methods, TAGM-MAP is better suited for spatial proteomic studies focused on the mitochondria and the ER, while SVM is better suited for studies of the Golgi, nucleus, and plasma membrane. This finding was surprising as we observed higher average QSep scores for the mitochondria and ER in WT replicates using SVM as compared to TAGM-MAP (SuppFile 1.9), with less of a difference in ΔNef replicates (SuppFile 1.10), which may indicate an imperfect correlation between QSep scores and general accuracy for certain organelles.

SVM-based BANDLE of WT replicates yielded the best agreement of HIV-dependent translocations with known HIV interactomes; partial overlap with ΔNef translocation hits

Following our analysis of classifiers, we examined various pipelines for identifying protein translocations. We inputted our SVM and TAGM-MAP classified data into BANDLE[41] and a basic label-based analysis[38], and inputted unclassified data into TRANSPIRE[40] (Fig. 1.2A). For TRANSPIRE, we combined the organelles into 5 groups: 1) Golgi apparatus/plasma membrane/endoplasmic reticulum/peroxisomes/lysosomes, 2) cytosol/actin cytoskeleton/proteasome, 3) nucleus, 4) mitochondria, and 5) 40S/60S ribosome. This is in line with the authors' recommendation to combine similarly behaving organelles to increase translocation confidence[40], although in our case we lose the ability to identify proteins moving

between the membranous organelles most likely to be affected by Nef, i.e. secretory organelles. To compare the performance of these five methods, we cross-referenced their hits against an HIV interactome derived from affinity purification-mass spectrometry (AP-MS)[19] as well as the NIH HIV interactome[80]. The AP-MS study is more stringent since it includes only those proteins that directly complex with HIV proteins, while the NIH HIV interactome includes proteins that are affected by HIV even in the absence of evidence for a direct interaction. We found that the percentage of hits from each method that were in the interactomes was consistently above the threshold expected by chance (Fig. 1.6A, dashed line). Comparing the methods, the top 50 hits from the BANDLE analysis of SVM-classified data performed best for both interactomes with 20% and 84% of hits in the Jäger et al study (direct interactome by AP-MS) and NIH HIV interactome (functional as well as direct interactors), respectively. Of note, ~1,500 proteins were considered to be translocation hits by the BANDLE analysis of SVM, i.e. greater than 95% probability of translocation. The validity of this value is difficult to gauge, but it is much higher than the ~50 proteins from the BANDLE analysis of TAGM-MAP-classified data with a similar probability of translocation. We conducted a similar hit analysis on our ΔNef inducible line and found that SVM-based BANDLE was still the highest performer for the NIH HIV interactome, but was only 3rd best for the AP-MS interactome (SuppFile 1.11).

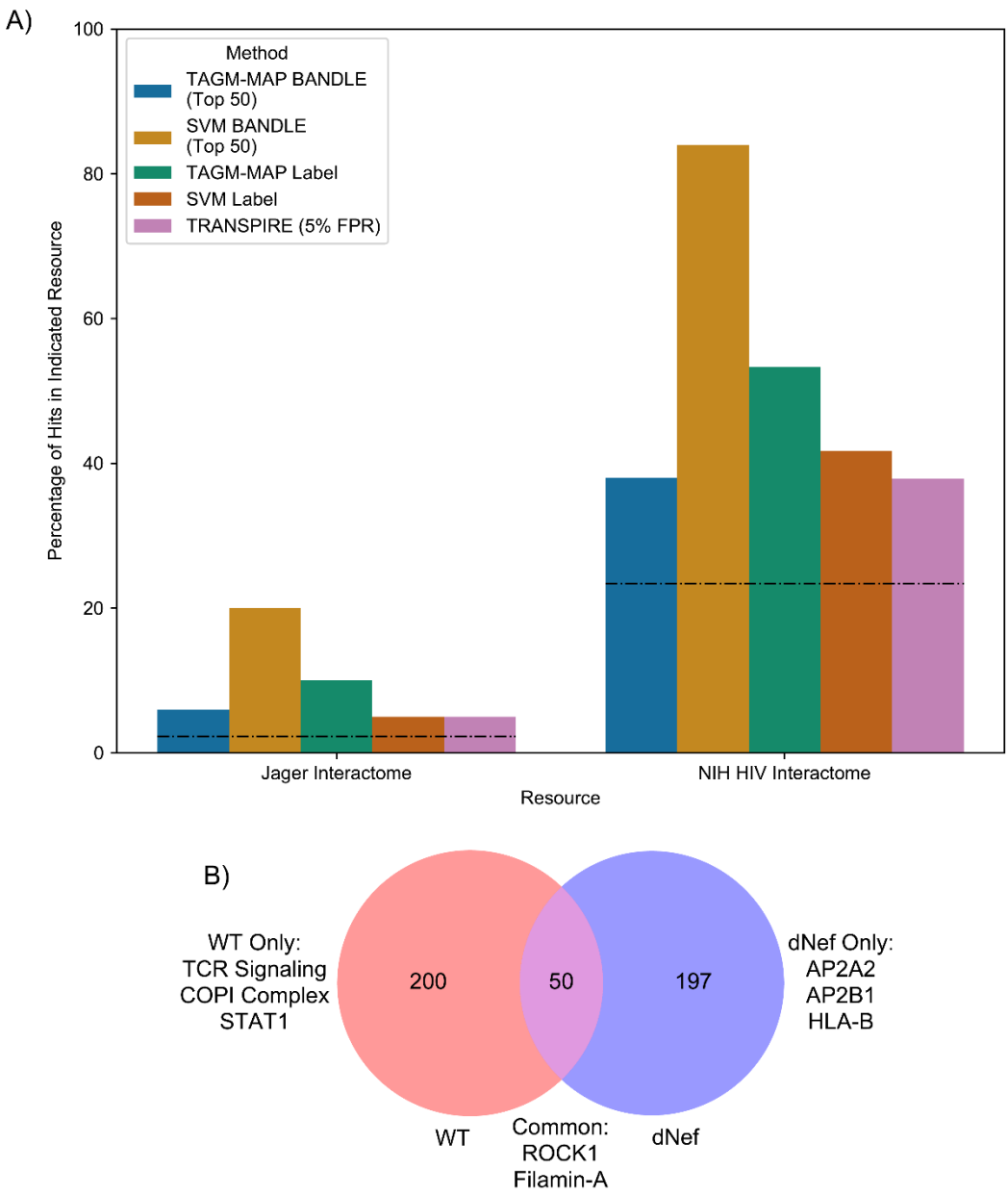

Figure 1.6. Detection of protein translocations by BANDLE analysis of SVM-classified data shows the highest rate of identifying known HIV interactors.
A) The percentage of hits from each method that are in the Jäger HIV interactome[19] (left bars) or the NIH HIV interactome[80] (right bars) is shown. Dashed lines indicate the percentage of hits that would be expected by chance based on the proportion of the human proteome represented in each interactome. B) Venn diagram of top 250 hits from SVM-based BANDLE for WT and ΔNef replicates. Three of the hits from the ΔNef analysis were not detected by MS in WT replicates and were thus removed from consideration.

The top 250 hits from SVM-based BANDLE for WT and ΔNef were compared to see if the method could identify Nef-dependent translocations (Fig. 1.6B); hits that were detected by

MS in only WT or ΔNef replicates were removed to avoid detection bias. Of those hits found only for WT, we observed several known Nef targets and cofactors: ZAP70 (ref.[89]), Lck[76,90], STAT1 (ref.[91]), and coatomer complex I (COPI complex)[92,93]. Five separate proteins in the COPI complex appear together as well as three proteins from the T cell signaling pathway, suggesting high coverage of perturbed complexes. For commonly shared hits, proteins involved in cytoskeleton organization were enriched. Disruption of the cytoskeleton following infection with HIV has been attributed to Nef among other viral proteins, but the enriched proteins here lacked known targets of Nef but instead included ROCK1, an interactor of HIV Tat, and filamin-A, an interactor of HIV Gag[94]. We were surprised to see two components of the AP2 complex, known interactors of Nef[95], and HLA class B, a known target of Nef[96,97], in the ΔNef only translocations. The SVM classification for these select proteins and STRING diagrams of the full protein sets are shown in the Supplemental Figures (SuppFile 1.12-15). Notably, the SVM classifications rarely provided definitive organellar translocations for the hits identified by BANDLE (SuppFile 1.12). In some cases, this was due to the majority of replicates becoming unclassified in the induced condition. An interesting exception is Filamin-A: although a translocation hit in both WT and ΔNef cells by BANDLE (Fig. 1.6B), by SVM classification Filamin-A moves from the actin cytoskeleton to the cytosol in cells expressing WT but not ΔNef (SuppFile 1.12K). While the basis for such analytic discrepancies is unclear, taken together these data suggest potential value in identifying novel HIV cofactors, targets, and interactors via BANDLE analysis of spatial proteomics data.

Discussion

We have detailed here a discussion of computational methods within the field of spatial proteomics as an example and guide for researchers hoping to use these methods to better understand viral infection and replication. Extensive work in the field, particularly by the Lilley[34,36,37,39,98], Cristea[38,40,69], and Borner groups[28,31,32,35] along with their collaborators, offers a

variety of established choices for fractionation, classification, and translocation identification methods. To build off of the work of the Cristea group with HCMV, we chose to examine HIV-1 as a model virus due to the existing wealth of knowledge on HIV-dependent protein interactions and translocations. We found in our T cell line model and using differential centrifugation for cell fractionation that the choice of computational method for classification is organelle-dependent: TAGM-MAP offered an advantage for mitochondrial and ER proteins, while SVM performed better for the Golgi apparatus, nucleus, and plasma membrane. For identifying translocations, BANDLE gave the highest agreement with known HIV biology (i.e. published interactome data), particularly when coupled with SVM-classified data.

The model of inducible HIV in Jurkat T cells used here has advantages and disadvantages. One advantage is that the system provides a highly homogenous population of HIV-expressing cells suitable for mass spectrometric analysis[71]. A homogenous population is particularly important in spatial proteomic studies as mixed populations of cells might yield erroneous classifications of proteins due to mixtures of different states[98]. Another advantage is scalability. These experiments required just over $3x10^8$ cells for each technical replicate, or over $1x10^9$ cells for a single biological replicate, to be induced. In our initial attempts with fewer cells, centrifugation at higher RCF (110,000x*g* and 195,500x*g*) yielded insufficient protein mass for quality control and mass spectrometry (data not shown). This highlights an advantage of using this T cell line compared to using primary CD4+ T-cells[99], which in principle would be more relevant but would require at least $2x10^9$ cells and extraordinary viral inocula to achieve a high-multiplicity, synchronized infection. A disadvantage of using this T cell system is that the cytoplasmic volume of the cell is relatively small. We required an order of magnitude more cells for each technical replicate here than were used in the D.O.M. studies of Itzhak et al., who used HeLa cells with larger cytoplasm.

In addition to these technical considerations for modeling viral infection/expression, the choice of fractionation method has practical and computational implications. The use of

differential centrifugation here and by Itzhak et al. requires the downstream analysis of fewer fractions than gradient fractionation methods and is far less time-, resource-, and labor-intensive[34]. On the other hand, gradient fractionation methods seem to show increased resolution of protein classification[37]. In an attempt to increase the organellar resolution of the D.O.M method we used additional high-speed centrifugation steps to those described in the D.O.M. method of Itzhak et al. and found a significant increase in overall organellar resolution using seven fractions as compared to the original five (Fig. 1.2). Previous work by the Lilley group comparing differential centrifugation and gradient-based methods for fractionation revealed comparable downstream results for the two methods using U-2 OS cells with differential centrifugation having a slight advantage in resolving the cytosol and proteasome[36], but whether this trend would hold in different cell types after viral infection or gene-expression is unclear. Generalizable rules for spatial proteomics might require comparisons of various fractionation and computational methods in multiple systems, or perhaps more likely, the specific experimental system and questions asked might be best addressed by a specific method. For example, to investigate translocations caused by HIV-1 Nef, better separation of membranous organelles (see SuppFile 1.5 and 1.6) might have yielded more Nef-specific translocations.

Our findings on classification consistency and accuracy might influence the choice of classifier, at least for this model system. We found that SVM yielded higher consistency in classification than TAGM-MAP, although both suffered similar losses in consistency following HIV expression. In cases where infection or viral expression is expected to introduce greater noise to data, as seems to be the case here, SVM may be the better option as it yielded a higher starting point for consistency. If lower tolerance to noise is acceptable, TAGM-MAP offers an advantageous alternative for both the mitochondria and ER. TAGM-MAP also suffered less loss of protein classification to unknown designations for uninduced replicates, perhaps due in part to the threshold used here for retaining SVM classification. While we used a basic median

SVM algorithm score threshold for each organelle[34] to allow for raw comparisons of classifiers to existing spatial proteomes, this might have been overly stringent for certain organelles, which would explain the higher number of unknown designated proteins for SVM. An alternative method would be to introduce an organelle-dependent threshold that would cap false positives by comparing classifier outputs to gene ontology analysis and published spatial proteomes; this method was employed previously by the Lilley group[29,36]. We further note the fact that while SVM showed generally higher QSep scores for the mitochondria and ER it still underperformed compared to TAGM-MAP for these organelles. This suggested to us that organellar resolution as measured by QSep might be an imperfect measure of classification accuracy for a given organelle, a hypothesis that will need further examination.

Lastly, the choice of translocation identification method requires consideration of several factors, the first of which is the experimental design. Part of BANDLE's power comes from its ability to factor multiple replicates of a condition into hit determination. Indeed, we saw a generally higher predictive power for BANDLE compared to other methods. The ranked list of output is also useful in cases where resources are limited and only a few hits can be pursued. TRANSPIRE seemed to have poorer performance compared to other methods, but this might reflect our need to combine similarly fractionated organelle groups to reduce computational demand and increase resolution. In cases where individual organellar resolution is greater, TRANSPIRE might yield higher quality hits. Notably, both BANDLE and TRANSPIRE require intensive computational resources, with TRANSPIRE requiring supercomputer access for larger, more complex datasets. In cases where computational power is limited, label-based methods would be suitable. Indeed, this method was employed by the Cristea group for their HCMV study with success[38].

A challenge not addressed here is how to handle changes in whole-organellar behavior within spatial proteomics, such as might be induced by viruses. Indeed, we observed such a change within our data: peroxisomal marker proteins shifted in their fractionation behavior (peak

abundance occurring at a higher *g*-force) when WT HIV was induced, becoming very similar in their behavior to marker proteins of the ER (SuppFile 1.5). This effect was not observed for ΔNef (SuppFile 1.6). A previous discussion of this issue by the Lilley group[41] highlighted the various possible causes of whole-organellar changes—e.g. differences in organelle protein content, lipid composition, morphology, etc.—as potentially problematic for the movement-reproducibility method of translocation identification[28], but how these types of biochemical changes would affect translocation detection methods or classifiers is not obvious. In our preliminary analyses of the average distance between organellar clusters based on pairwise distances, we found that peroxisomes alone shifted in relation to other organelles following the induction of WT HIV (but not ΔNef). However, analyses using QSep, which additionally considers the average intracluster distance (i.e., the dispersal of the cluster that defines the organelle), gave a less clear picture, with the potential for multiple relative movements among organelles (data not shown). These observations suggest that computational methodology will affect conclusions about organellar behavior as a whole. While the uniform shift of all markers for a given organelle should have only a minor impact on classification, how likely such a shift is in the context of viral gene expression probably depends on the specific virus and the type of cytopathic effect it induces. Indeed, the greater sensitivity of TAGM-MAP to HIV expression for classifier consistency could be a manifestation of subtle changes in organelle behavior. Careful examination of marker proteins used as well as the integration of pre-existing knowledge on the cytopathic effects of the virus under study are doubtlessly important for interpretation of whole-organellar changes.

With these considerations in mind, our findings underscore that studies of spatial proteomics require careful consideration of the question at hand to inform the choice of methodology. Our work and that of others highlights the potential differences in organellar resolution that can result from the choice of fractionation and analytical methods. Interest in a particular organelle and in specific types of translocations will factor into the choice of methods.

Our findings offer a reference point for studies of viral infection by spatial proteomics, for general studies of the spatial proteome, and for the study of additional gene dropout mutants of HIV-1.

Data availability

Mass spectrometry data (.RAW files and peptide identification tables) can be found on the ProteomeXchange database using project accession number PXD024716.

Supplemental data

This article contains supplemental data.

Author contributions

A.L.O., C.A.S., M.K.L., and J.G. conceived the design and scope of this study. M.K.L. created the inducible cell lines used in this study. A.L.O. and C.A.S. performed the fractionation experiments. A.L.O., J.M.W., and D.G. performed pilot studies for the mass spectrometry analysis. A.R., K.S., and N.J.K. performed the mass spectrometry analysis. A.L.O. performed the computational analysis. A.L.O. and J.G. wrote the manuscript and all authors reviewed and edited the manuscript.

Declaration of interests: none

Acknowledgements

The authors would like to thank the NIH AIDS Reagent Program and The Pendleton Charitable Trust. The authors would like to also thank Hannah Carter, Max Qian, and Oliver

Crook for productive discussions on the computational analysis. This work was partially supported by a UC San Diego Center for AIDS Research (UCSD CFAR) Developmental Award, an NIH funded program (P30AI036214). A.LO. is supported by NIAID F31AI141111. C.A.S. is supported by a UCSD CFAR Developmental Award (P30AI036214). M.K.L. was supported by K08AI112394. A.R. and N.J.K. are funded by a grant to the HARC Center from the National Institutes of Health (P50AI150476 to N.J.K.). J.M.W. is supported by NIH/NIGMS Grants T32 GM007752 and NIH/NIAMS T32 AR064194. D.G. is supported by The Collaborative Center of Multiplexed Proteomics at UCSD. J.G. is supported by NIAID R01AI129706. All funding sources had no involvement in the design and execution of this study nor the final manuscript.

Chapter 1, in full, has been submitted for publication of the material as it may appear in Molecular and Cellular Proteomics 2021. This material was co-authored with Stoneham, Charlotte A., Lewinski, Mary K., Richards, Alicia, Wozniak, Jacob M., Shams-Ud-Doha, K., Gonzalez, David J., Krogan, Nevan J., and Guatelli, John. The dissertation author was the principal author of this chapter.

CHAPTER 2: INTERACTIONS OF SARS-COV-2 ENVELOPE PROTEIN WITH AMILORIDES CORRELATE WITH ANTIVIRAL ACTIVITY

PLoS Pathogens 2021

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has garnered attention as the causative agent of the disease COVID-19. It is an enveloped RNA virus classified as a beta coronavirus[100] similar to the previously studied SARS-CoV[101] and MERS-CoV[102] viruses. While rapid progress has been made in analyzing the SARS-CoV-2 genome[103] and the development of protective vaccines[104,105], the discovery of therapeutics has lagged, largely due to the lack of structures of the viral proteins and information about their specific roles in infection, replication, and propagation. Here we apply NMR spectroscopy to the envelope (E) protein, one of the structural membrane proteins of SARS-CoV-2, in order to characterize its secondary structure, drug binding site, and effects of selected single-site mutations on its structure and binding of amiloride compounds. To accomplish these goals, the results from NMR on E protein are augmented by those from virological experiments on infected cells[106] as well as the measurement of antiviral activities of amiloride compounds.

The approximately 30kb RNA genome of SARS-CoV-2 encodes for 29 proteins (www.ncbi.nlm.nih.gov/nuccore/NC_045512). The most abundant are four structural proteins, membrane (M), envelope (E), nucleocapsid (N), and spike (S), of which M, E, and S are integral membrane proteins embedded in the lipid bilayer of the viral envelope (Fig 2.1). Each of these proteins exists as a homo-oligomer under some experimental conditions: a dimer or dimer of dimers for M[107], a pentamer for E[108,109], and a trimer for S[110]. The biological relevance of E protein comes from its involvement in key aspects of the virus lifecycle, including infection, replication, assembly, budding, and pathogenesis[111]. Furthermore, recombinant coronaviruses lacking E protein exhibit significantly reduced viral titers, crippled viral maturation, and yield

propagation incompetent progeny[57,58,112]. SARS-CoV-2 E protein is a hydrophobic 75-residue protein with an amino acid sequence nearly identical to that of SARS-CoV E protein (SuppFile 2.1)[111]. Since E protein is a viral membrane-spanning miniprotein[113], a recurring question is whether it is a viroporin. Although ion-channel activity has been detected in a variety of preparations it lacks sequence homology with any of the well-established viroporins, and there is a notable absence of charged sidechains on the interior of a pore formed by pentamers of the protein in membrane environments[108,109,114,115].

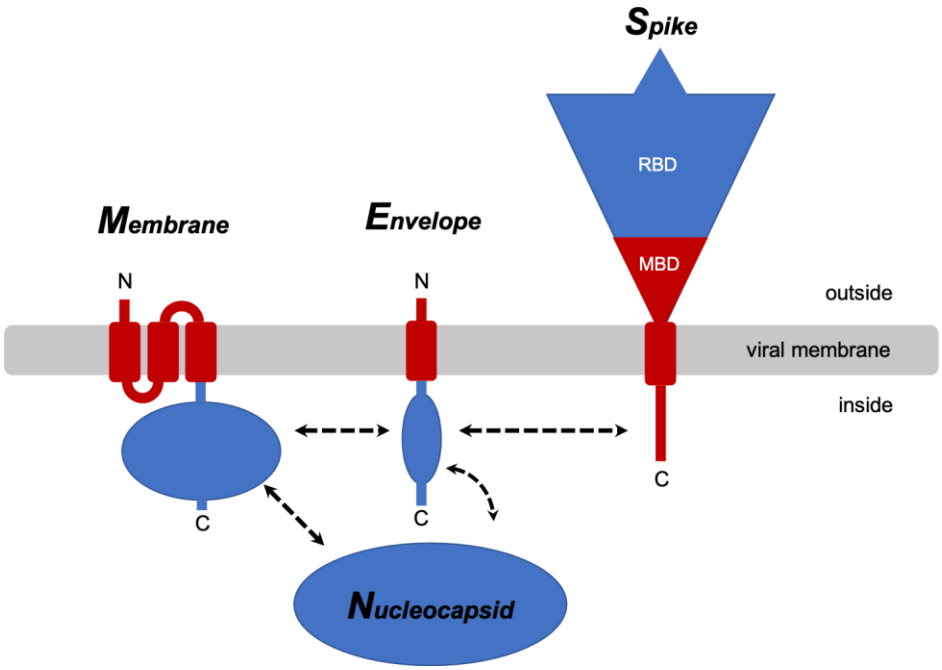

Figure 2.1. Cartoon representations of the four structural proteins of SARS-CoV-2. The membrane-associated portions of the membrane (M) protein, envelope (E) protein, and spike (S) protein are shown in red, and the extra- and intra- cellular portions are shown in blue. Proposed intraviral protein-protein interactions are indicated by the dashed arrows. RBD: receptor-binding domain; MBD: membrane-binding domain. Nucleocapsid (N) is the fourth structural protein.

The importance of E protein for viral replication and maturation is well established, making it an attractive target for antiviral drugs. Drug design requires high-resolution structures of the protein receptor in its bound and free states. Small membrane proteins are notoriously difficult to crystallize in their native states in liquid crystalline membrane bilayers for X-ray

crystallography and are too small for cryoEM to be effective. While generally suited for NMR spectroscopy, careful consideration of the membrane-like environment of the samples and the NMR experimental methods are essential[116]. Even the earliest NMR studies of membrane proteins showed that caution is called for when using micelle environments[117], because of the potential for aggregation and structural distortions[118,119]. Nonetheless, careful optimization of sample conditions has enabled solution NMR to provide valid structural information about membrane proteins that could be obtained in no other way. Moreover, we have found it essential to prepare samples of membrane proteins in micelles that yield high-resolution solution NMR spectra in order to verify that they integrate into an amphipathic membrane-like environment, are chemically pure, not mis-folded, and not aggregated before initiating significantly more demanding solid-state NMR studies of phospholipid bilayer samples. In order to ensure that solid-state NMR experiments are performed under near-native conditions, both the protein and the bilayers must be fully characterized to ensure that the protein is in its biologically active conformation and stably embedded in liquid crystalline, fully hydrated phospholipid bilayers at high lipid to protein ratios.

Here we describe solution NMR studies of full-length SARS-CoV-2 E protein and several truncated and mutated constructs in highly optimized n-hexadecylphosphocholine (HPC; fos-choline-16) micelles. Because our novel purification scheme starts by using HPC to solubilize the protein-containing inclusion bodies and HPC is present during all subsequent steps, the polypeptides are never exposed to any other detergent or lipid, which would require exchanges, or to any organic solvent, which would require refolding. The success of the HPC-based protein purification and sample preparation scheme results in the well-resolved solution NMR spectra presented in the Figures. Furthermore, this scheme leads directly to the preparation of magnetically aligned bilayer samples that are well-suited for protein structure determination by oriented sample (OS) solid-state NMR[120].

Previous structural studies of coronavirus E protein, especially by NMR, have been simplified by using relatively short polypeptides with sequences corresponding to a substantial portion of the N-terminal domain containing the transmembrane helix that forms ion channels through homo-oligomerization as well as residues responsible for drug binding[108,109,114,115]. To date, no structural data have been presented for any full-length coronavirus E protein. Structures of a 31-residue synthetic polypeptide (residues 8–38)[114] and of a longer 58-residue expressed polypeptide (residues 8–65) containing three Cys to Ala mutations and non-native 23-residues in its N-terminus[108,115] have been described for sequences from the SARS-CoV E protein. They are highly relevant to studies of the SARS-CoV-2 E protein because the amino acid sequences of these two proteins are identical between residues 1 and 68. The partial E protein structures determined for these polypeptides in micelles by solution NMR have been modeled as pentamers[108,114,115]. In addition, an expressed 31-residue polypeptide with the same sequence as residues 8–38 of SARS-CoV-2 E protein has been studied by magic angle spinning (MAS) solid-state NMR in the presence of phospholipids, and its structure has also been modeled as a pentamer[109]. There are significant differences between the conclusions derived from these earlier studies of relatively small polypeptides missing the N-terminal seven residues and those presented here based on studies of full-length protein (residues 1–75) and two overlapping constructs encompassing the N-terminal domain (residues 1–39) and the C-terminal domain (residues 36–75). Notably, the wild-type N-terminal 39-residues are present in both the full-length and C-terminal truncated proteins.

While the transmembrane helix of E protein is thought to be largely responsible for homo-oligomerization and ion-channel activity[114,115], its highly hydrophobic nature makes modeling a channel similar to those of other miniproteins difficult[113]. Intraviral interactions between E and M proteins have been shown to involve the C-terminal domains of both proteins[121,122]. The triple cysteine motif (C40, C43, and C44) in E protein has been proposed to associate with the cysteine-rich C-terminal region of S protein by forming intermolecular

disulfide bonds[123]. E protein also interacts with host proteins[111]. The C-terminal four residues, DLLV, have been identified as a PDZ-binding motif that interacts with the tight junction-associated PALS1 protein[124]. The C-terminal region of E protein that resembles the bromodomain binding site of histone H3 interacts with bromodomains 2 and 4 via acetylated Lys63, which is involved in the regulation of gene transcription[125]. These studies provide strong biological and mechanistic justification for considering coronavirus E protein as a potential drug target. Extending structural studies to samples of the full-length protein that include the complete drug binding site as well as the native N-terminus, C-terminus, and other features is essential for structure-based drug discovery. Equally important is the correlation of structural features of the protein with specific biological activities of the virus as it reproduces in human cells.

The channel activity of E protein has been suggested to play a role in viral replication[126]. A well-characterized channel blocker, hexamethylene amiloride (HMA), inhibits ion channel conductance of E proteins from HCoV-229E and MHV as well as virus replication in cultured cells[66]. HMA also inhibits the channel conductance of transmembrane-containing synthetic and expressed polypeptides from the SARS-CoV E protein[114,115]. Although interactions of HMA with E protein of SARS-CoV have been detected in prior studies, the residues in the HMA binding site identified by NMR chemical shift perturbations varied quite a bit depending upon the specific E protein constructs and experimental conditions[108,109,114,115].

Here we characterize the secondary structure of full-length E protein from SARS-CoV-2 in HPC micelles. We also map out the complete binding sites of amiloride and three amiloride derivatives (dimethyl amiloride (DMA), ethyl isopropyl amiloride (EIPA), and HMA) and compare their binding properties. Importantly, the antiviral potency of the amiloride derivatives against SARS-CoV-2 infection of Vero E6 cells correlates well with their strength of binding as observed in the NMR experiments. The N15A and V25F mutations of the SARS-CoV-2 E protein have very different effects on the NMR spectra of the protein; the N15A mutation causes greater

chemical shift perturbations over a larger region of the protein than the V25F mutation, which causes only minor changes near the site of the amino acid substitution. These mutations affect production of virus-like particle (VLP) and, in the case of N15A, the binding of HMA.

Results

Preparation of full-length SARS-CoV-2 E protein

In order to apply NMR spectroscopy to full-length E protein of SARS-CoV-2 it was essential to develop and implement an entirely new sample preparation scheme. We were unable to overcome the difficulties inherent in dealing with hydrophobic membrane proteins in the case of E protein using approaches that we had previously applied successfully to viral, bacterial, and human membrane proteins with between one and seven transmembrane helices[127–132]. These preparative difficulties may be among the reasons that prior NMR studies of E protein have been limited to N- and C- terminal truncated constructs with only 31 or 58 residues, which are notably missing the seven N-terminal residues that our data show to be essential components of the drug-binding site.

The ketosteroid isomerase (KSI) fusion partner facilitated the expression of high levels of three different E protein constructs, including the full-length protein (residues 1–75), as inclusion bodies in *E. coli*[128,133]. A ten-residue His-tag followed by a six-residue thrombin cleavage site, LPVRGS, inserted between the KSI and the E protein sequences enabled purification by Ni-affinity chromatography and efficient enzymatic cleavage (Fig 2.2C). The resulting E protein sequence differs from that of the SARS-CoV-2 isolate Wuhan-Hu-1 (NC_045512) only by the presence of two additional residues (GlySer) at the N-terminus that were originally part of the thrombin cleavage site (Fig 2.2B).

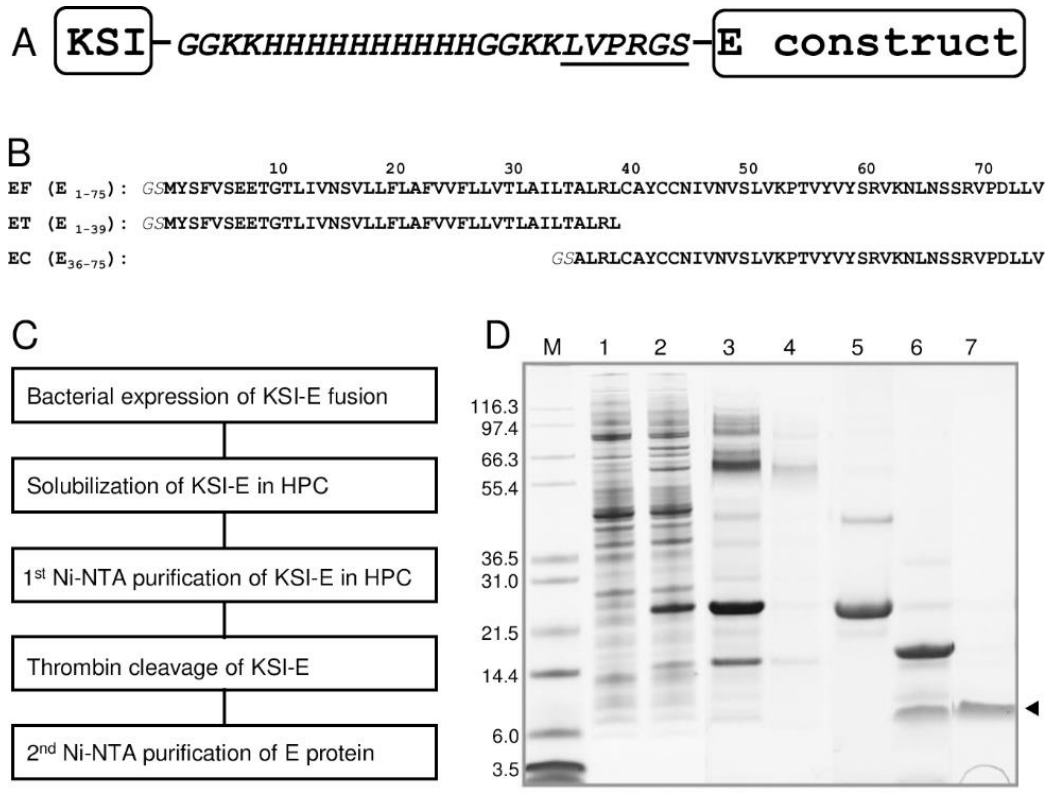

Figure 2.2. Heterologous expression and purification of the full-length SARS-CoV-2 E protein and truncated protein constructs.

(A) Design of SARS-CoV-2 E protein and KSI fusion protein construct utilized for efficient bacterial expression and purification. The six residues (LVPRGS) that define the thrombin cleavage site are underlined. (B) Amino acid sequences of the polypeptides used here: full-length E protein (EF) (residues 1–75), the N-terminal transmembrane domain of E protein (ET) (residues 1–39), and the C-terminal cytoplasmic domain of E protein (EC) (residues 36–75). Two additional residues, GlySer, are present at the N-termini of all E protein constructs. (C) Block diagram of the expression and purification protocols applied to the polypeptide sequences shown in part B. (D) Example of SDS-PAGE at various stages of the expression and purification of EF: lane 1, pre-induction cells; lane 2, post-induction cells; lane 3, HPC-solubilized inclusion bodies containing the KSI-EF fusion protein; lane 4, Ni-affinity column flow through; lane 5, eluate of the KSI-EF fusion protein from the column; lane 6, after thrombin cleavage of the KSI-EF fusion protein; lane 7, arrow marks the single band of purified EF used in samples for the NMR experiments.

Our scheme for the expression and purification of the full-length wild-type and mutated, and truncated constructs of E protein is outlined in Fig 2.2C. They were all expressed in E. coli as fusion proteins and sequestered in inclusion bodies (Fig 2.2D). After screening many detergents informed by our extensive experience with solution NMR studies of membrane proteins[117,127,134–137] and thorough literature reviews[119], we found that the highest resolution spectra were obtained when E protein was solubilized in hexadecylphosphocholine (HPC, fos-

choline-16) micelles. Chemically similar to the commonly used dodecylphosphocholine (DPC)[138], HPC has been previously considered for, but, to our knowledge, not used to study membrane proteins by NMR[139]. HPC is notable for its low critical micelle concentration (CMC, 13 μM) (www.anatrace.com). It is able to solubilize E protein and other hydrophobic membrane proteins, is effective with Ni-affinity chromatography, and at a low concentration of 0.05% w/v (1.23 mM) does not interfere with specific thrombin cleavage. This approach is highly efficient and, significantly, obviates the need for detergent exchanges or exposure to organic solvents at any stage of the isolation and purification process. Purified full-length E (EF) in HPC micelles runs as a monomer (~ 8.5 kDa) with a narrow band on SDS-PAGE (Fig 2.2D). By contrast, as observed by others[140], it runs as a broad ill-defined band on PFO (perfluorooctanoic acid)-PAGE that may demonstrate the presence of an oligomeric species generally assumed to be a pentamer consistent with its viroporin-like properties[108,109,114,115,140,141]. All samples used in the NMR experiments were prepared directly from protein solubilized in HPC from start to finish. The resulting NMR spectra are well-resolved with narrow resonance linewidths. The samples exhibit excellent long-term stability at 50°C (Fig 2.3).

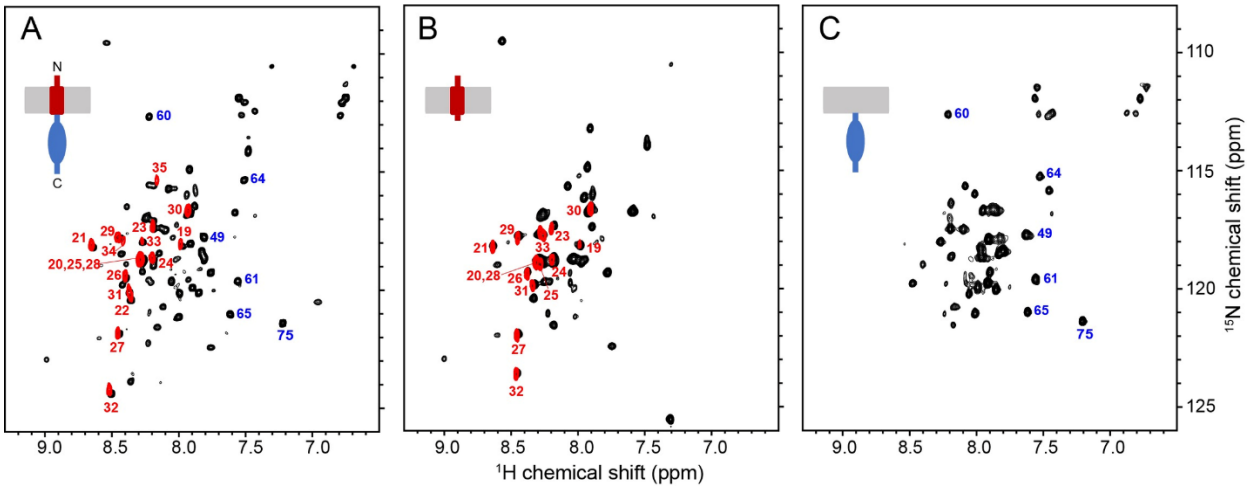

Figure 2.3. Comparison of $^{1}H/^{15}N$ HSQC spectra of uniformly ^{15}N-labeled E protein constructs in HPC micelles in H_2O (black contours) and D_2O (red contours).
(A) Full-length E protein (EF) (residues 1–75). (B) N-terminal transmembrane domain of E protein (ET) (residues 1–39). (C) C-terminal cytoplasmic domain of E protein (EC) (residues 36–75). For reference, cartoons of each construct are shown. The assignments of selected resonances are marked to distinguish among signals from ET (red numbers) and EC (blue numbers).

Conformations of SARS-CoV-2 E protein domains are preserved

Fig 2.3 compares $^{1}H/^{15}N$ HSQC spectra of three E protein constructs in HPC micelles. The spectra are well-resolved despite the relatively narrow span of ^{1}H amide chemical shift frequencies (< 2 ppm) consistent with the predominantly helical conformations observed previously[108,114,115]. The backbone resonances of full-length E protein have been assigned and their chemical shifts deposited in the Biological Magnetic Resonance Data Bank (accession number: 50813) (SuppFile 2.3 and SuppFile 2.8). Notably, the observation of the expected number of resonances, with no evidence of doublings or unusual line shapes from the selectively ^{15}N-Leu and ^{15}N-Val labeled samples (SuppFile 2.4), where there are no ambiguities due to spectral overlap, confirms chemical purity and conformational homogeneity of the full-length protein in HPC under the experimental conditions. Any evidence of detergent-induced structural perturbations or heterogeneous aggregation detected in these spectra would call for

further sample optimization before moving forward with solution NMR experiments or the initiation of the preparation of bilayer samples for solid-state NMR experiments.

The spectra of the N-terminal transmembrane helix-containing domain (ET) (residues 1–39) (Fig 2.3B) and the C-terminal cytoplasmic domain (EC) (residues 36–75) (Fig 2.3C) are superimposable on the spectrum of the full-length protein (EF) (residues 1–75) (Fig 2.3A), with the exception of signals from residues proximate to the newly formed C-terminus of ET and N-terminus of EC (SuppFile 2.5). These results demonstrate that the folded structures of the domains are not perturbed by separation from each other, which suggests an absence of inter-domain interactions and possibly independence of their biological activities, which remains to be demonstrated *in vivo*.

Hydrogen/deuterium (H/D) exchange is an effective way to identify residues in transmembrane helices of membrane proteins[142]. When samples of the E protein constructs were prepared in >90% D_2O instead of ~ 90% H_2O, no amide signals from residues 36–75 in the cytoplasmic domain were observable in the spectra of EF or EC; by contrast, strong signals from residues 19–35 and 19–33 were present in the spectra of EF and ET, respectively (Figs 2.3 and 2.4C), demonstrating that these residues contribute to the stable core of its unusually long trans-membrane helix. Truncation at residue 39 enhances solvent exchange at residues 34 and 35 of ET, which are 5 and 6 residues distal to its C-terminus, respectively, due to structural changes reflected in chemical shift changes of the resonances from the nine terminal residues (SuppFile 2.4C).

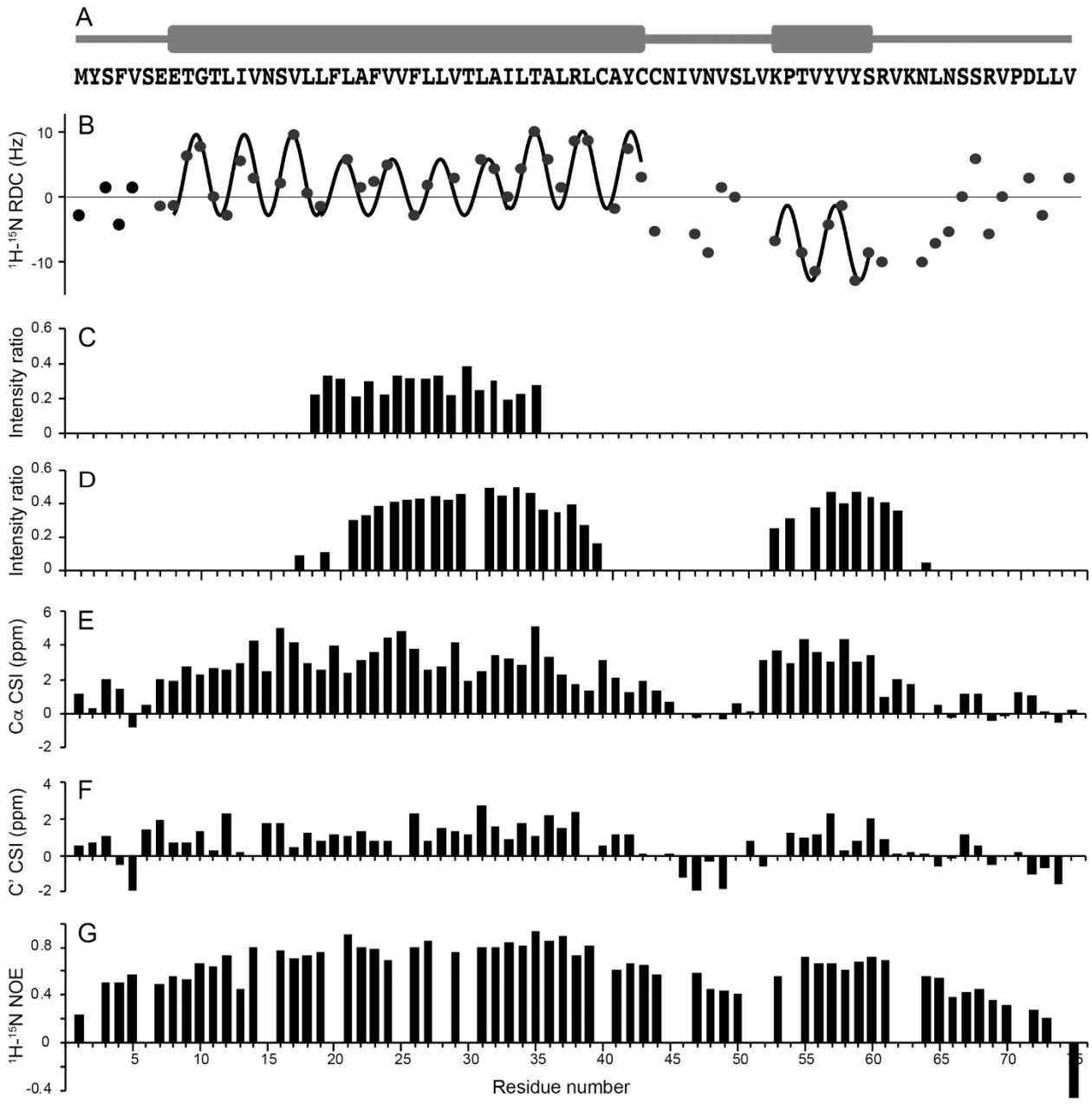

Figure 2.4. Summary of NMR data obtained on full-length E protein in HPC micelles at 50°C. (A) Schematic representation of the distribution of helical segments (thick bars) above the corresponding amino acid residues of E protein. (B) Plot of residual ^{1}H-^{15}N residual dipolar couplings as a function of residue number. Fits to sine waves with a periodicity of 3.6 reveal the dipolar waves characteristic of alpha helical secondary structure. The residual dipolar couplings were measured on a weakly-aligned sample as shown in SuppFile 2.6. (C) Ratios of resonance intensities with the protein in D_2O compared to those in H_2O solution. (D) Ratios of resonance intensities in the presence and absence of $MnCl_2$. (E and F) Chemical shift index plots of alpha (E) and carbonyl (F) carbon resonances, respectively. (G) Plot of ^{1}H-^{15}N heteronuclear NOEs as a function of residue number.

Secondary structure and dynamics of full-length E protein in HPC micelles

In previous studies, evidence has been presented that the predominant secondary structure of E protein is α-helix. However, the lengths of the proposed helical segments varied widely, depending upon which residues were included in the polypeptide constructs, the types of samples, and the experimental conditions[108,109,114,115]. Here we describe the secondary structure of full-length E protein in HPC micelles by analyzing the chemical shifts of backbone ^{13}C resonances[143] and amide $^{1}H/^{15}N$ residual dipolar couplings (RDCs)[144,145]. Further support comes from H/D exchange, manganese titration, and heteronuclear $^{1}H/^{15}N$ NOE measurements backed up by preliminary solid-state NMR spectra of protein-containing phospholipid bilayers. Complementary results have been obtained from samples of EF, ET, and EC.

Both the ^{13}C chemical shift index (CSI) plots (Fig 2.4C and 2.4D) and the ^{1}H-^{15}N dipolar wave plot (Fig 2.4B) demonstrate that full-length E protein has a long 36-residue transmembrane helix and a separate short 8-residue cytoplasmic helix. None of these backbone data indicate the presence of regular secondary structure in residues 43–52 located in the region linking the two helices. Although the RDCs have significant amplitudes, as expected for a structured region, the $^{1}H/^{15}N$ heteronuclear NOE data (Fig 2.4E) suggests that this well-defined internal region of the protein undergoes modest amplitude/frequency motions that are not present in the helical regions. The $^{1}H/^{15}N$ heteronuclear NOE data also shows that residues 2–7, before the start of the N-terminal helix, and residues 61–75 following the end of the C-terminal helix exhibit gradients of increasing motion towards the termini, although even the terminal residues do not appear to be highly mobile and unstructured, as is sometimes the case in this class of proteins[128,146].

The sinusoidal waves that fit best to the magnitudes and signs of the measured RDCs as a function of residue number have a periodicity of 3.6 residues per turn, proving with a very high level of confidence that the protein has segments of regular α-helix secondary structure[144,145,147]. The addition or subtraction of a single residue at either end of the helical

segments significantly degrades the quality of the fit, providing a clear demarcation of the length of the helical segments. The different average amplitudes of the two distinct dipolar waves in Fig 2.4B show that the two helices have different orientations relative to the direction of molecular alignment. Another notable feature is that the dipolar wave for the core region of the transmembrane helix (residues 19–34) is best fit by a sine wave with a somewhat smaller amplitude than for the rest of the long helical region (residues 8–18 and 35–43), suggesting that the 36-residue helix is not completely uniform throughout its length.

To assess the orientation of the C-terminal helix and possible interactions of the cytoplasmic domain with the hydrophilic headgroups of HPC, we examined the effects of adding paramagnetic manganese ions to samples of full-length E protein. Broadening of many $^{1}H/^{15}N$ HSQC resonances was observed as a function of increasing the concentration of $MnCl_2$. Significantly, the signals from residues 1–16, 40–51, and 64–75 were broadened beyond detection at a concentration of 5 mM $MnCl_2$, while the signals from residues 17–39 and 52–63 remained readily observable. These signals correspond almost exactly to the residues in the core of the long hydrophobic helix (Fig 2.4B–2.4D) and the short cytoplasmic helix, with the later suggesting that the cytoplasmic helix may interact with the membrane surface.

Interactions of SARS-CoV-2 E protein with amilorides

The chemical shift perturbations (CSPs) in the $^{1}H/^{15}N$ HSQC spectra of the three E protein constructs (EF, ET, and EC) caused by the addition of a ten-fold molar excess of hexamethylene amiloride (HMA) to the samples are illustrated in Fig 2.5. The black contours represent the protein signals in the absence and the red contours in the presence of HMA. For the constructs that include the N-terminal portion of the protein, EF and ET, the chemical shifts of the corresponding residues were perturbed in the same directions and to a similar extent, as illustrated in the plots of the chemical shift changes as a function of residue number in Fig 2.5D and 2.5E. By contrast, no significant chemical shift changes were observed in the resonances

from the cytoplasmic domain (EC) alone (Fig 2.5E) or as part of the full-length protein (EF) (Fig 2.5A). Although there is evidence that residues 2–5 are affected by drug binding, the most strongly perturbed signals are associated with residues 6–18 at the N-terminal end of the long helix and extending to the core portion distinguished by its resistance to H/D exchange and broadening by manganese ions, as well as the reduced amplitude of its dipolar wave. Qualitatively, the data in Fig 2.5 confirm that HMA interacts with the N-terminal domain of E protein.

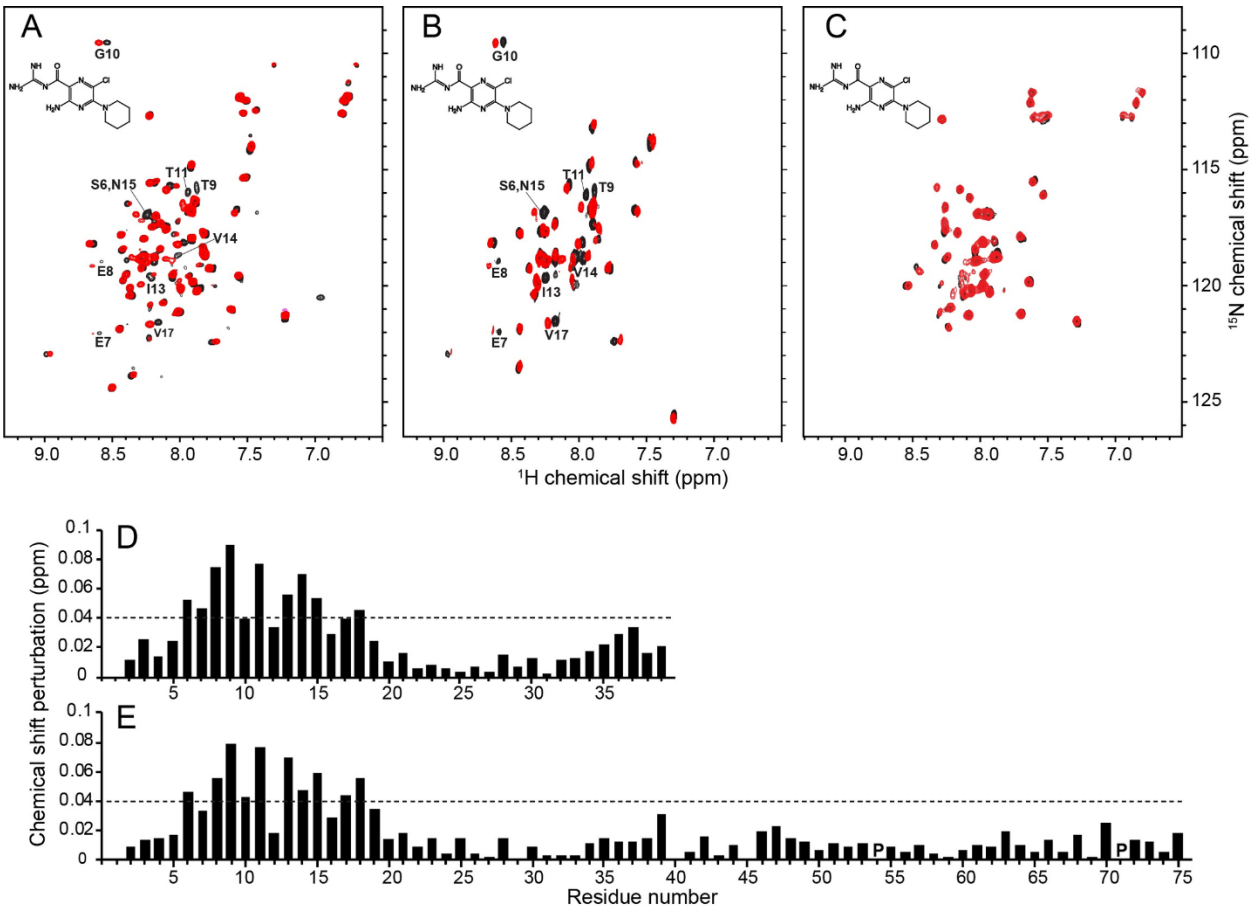

Figure 2.5. Chemical shift perturbations resulting from HMA binding to E protein constructs in HPC micelles.

(A-C) Superposition of $^{1}H/^{15}N$ HSQC NMR spectra of uniformly ^{15}N-labeled E protein constructs in the absence (black contours) and presence (red contours) of HMA. (A) Full-length E protein (EF) (residues 1–75). (B) N-terminal transmembrane domain of E protein (residues 1–38) (ET). (C) C-terminal cytoplasmic domain of E protein (residues 39–75) (EC). The molar ratio of protein to HMA is 1:10. The chemical structure of HMA is shown in each spectrum. The resonances perturbed by binding HMA are labeled with their assignments. (D and E) Plots of chemical shift perturbations as a function of residue number of ET (D) and EF (E) derived from the NMR spectra in B. and A., respectively. The horizontal dotted lines represent 1.5 times the average chemical shift perturbations induced by HMA binding to ET. Proline sites are marked as “P”.

The EF and ET constructs were designed to include all N-terminal residues, and these data show that the binding site definitely includes residues 6, 7, an 8, and likely residues 2, 3, 4 and 5, none of which were present in the previously studied constructs, and extends to residue 18. Nearly all of the residues that constitute the binding site belong to the highly regular helix, until it abruptly changes tilt angles at residue 18, the start of the core region. The residues between Ser6 and Leu18 are perturbed by binding HMA and undergo facile H/D exchange:

resistance to H/D exchange starts with residue 18. Notably, signals from four hydrophilic residues (Glu8, Thr9, Thr11, and Asn15) as well as Ile13 are most perturbed by HMA binding. Smaller CSPs observed in the C-terminal region of ET were not present with EF, which may be due to non-specific HMA binding to the unnatural exposed C-terminal region of ET. Titration experiments demonstrate that HMA binding occurs in fast exchange on the timescales defined by the chemical shift differences.

To compare the binding sites and affinities, we added increasing amounts of amiloride and two amiloride derivatives, dimethyl amiloride (DMA) and ethyl isopropyl amiloride (EIPA), to samples of full-length E protein (EF) and monitored their two-dimensional $^{1}H/^{15}N$ HSQC spectra (Fig 2.6). Notably, the same residues of EF were affected by all of the amiloride derivatives albeit with different magnitudes of CSPs, indicating that they all utilize the same binding site but with different binding affinities. No significant changes were observed to the EF spectrum upon addition of amiloride (Fig 2.6A–2.6D), while the largest changes were observed with EIPA (Fig 2.6C–2.6F), DMA induced moderate changes and the magnitudes of its CSPs lie between those of amiloride and HMA (Fig 2.6B–2.6E). The magnitudes of the CSPs indicate that the order of binding affinities to E protein is EIPA ≈ HMA > DMA >> amiloride.

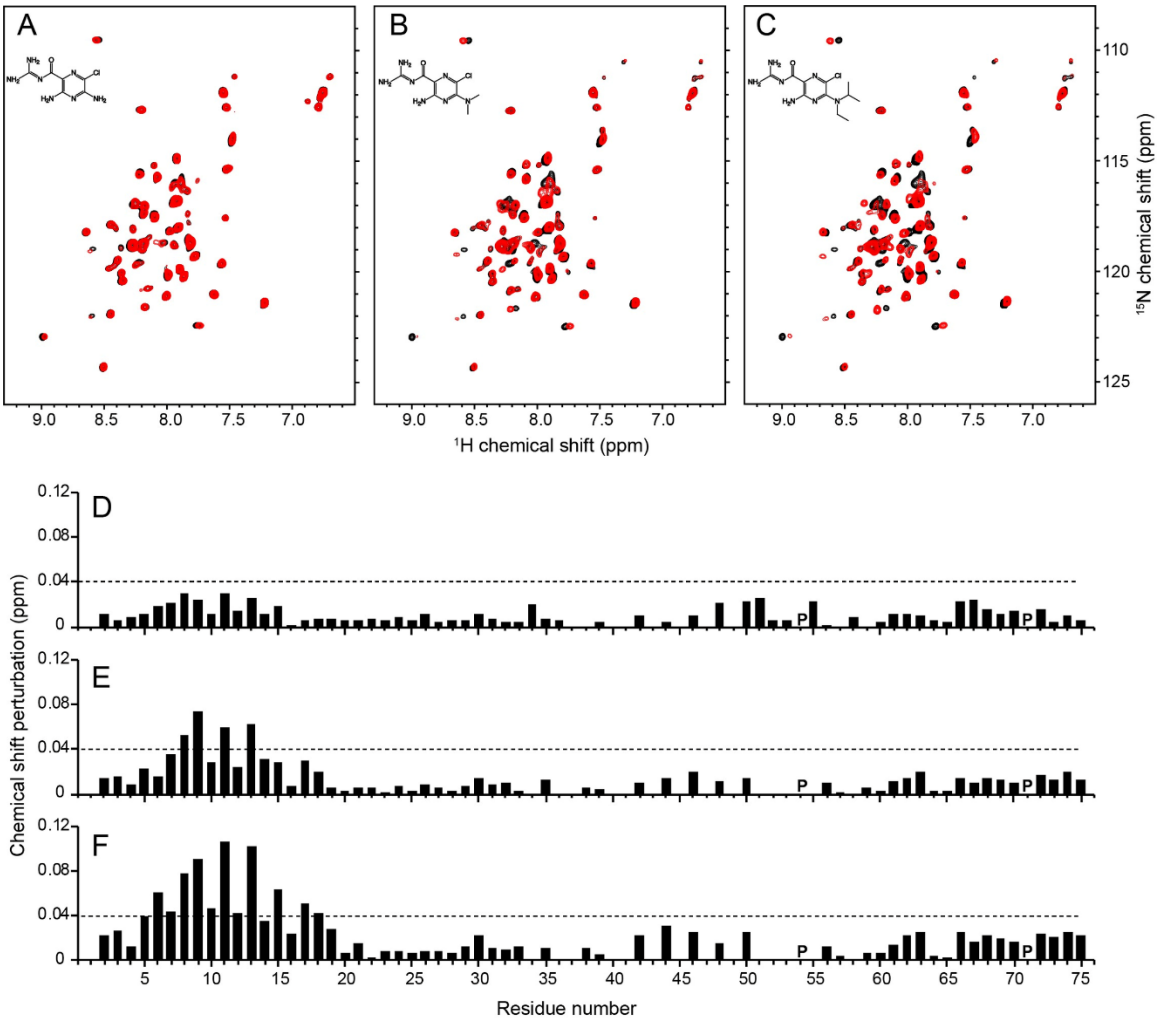

Figure 2.6. Comparison of interactions of E protein with amiloride compounds.
(A-C) ^{1}H/^{15}N HSQC NMR spectra of uniformly ^{15}N-labeled full length E protein (EF) in the absence (black contours) and presence (red contours) of (A) amiloride, (B) DMA, and (C) EIPA. The molar ratio of EF to each compound is 1:10. Chemical structures of the drugs are shown in the spectra. (D-F) Plots of chemical shift perturbations (CSPs) in the presence of (D) amiloride, (E) DMA, and (F) EIPA as a function of residue number. The dotted lines indicate 1.5 times the average chemical shift changes of EF by EIPA. Proline sites are marked as "P".

Antiviral activity of amilorides against SARS-CoV-2

The amiloride derivatives were tested for their ability to inhibit replication of SARS-CoV-2 in Vero E6 cells. Mirroring the NMR binding data of E protein in Figs 2.5 and 2.6, the compounds with bulkier aliphatic or aromatic substituents at the 5' pyrazine ring (EIPA and HMA) showed the strongest inhibition, with sub-micromolar IC_{50} values, while the compounds

with smaller substituents were less effective inhibitors (Fig 2.7). The similarity of the trends for inhibition of replication and of binding to E protein suggests that this protein may very well be a target for the antiviral activity of amiloride compounds. Of note, the most active compound examined here, HMA, shows considerable cytotoxicity (therapeutic index = 21.23), therefore, EIPA may be a better choice for potential therapeutic use (therapeutic index = 84.83) or as a starting point for further drug development.

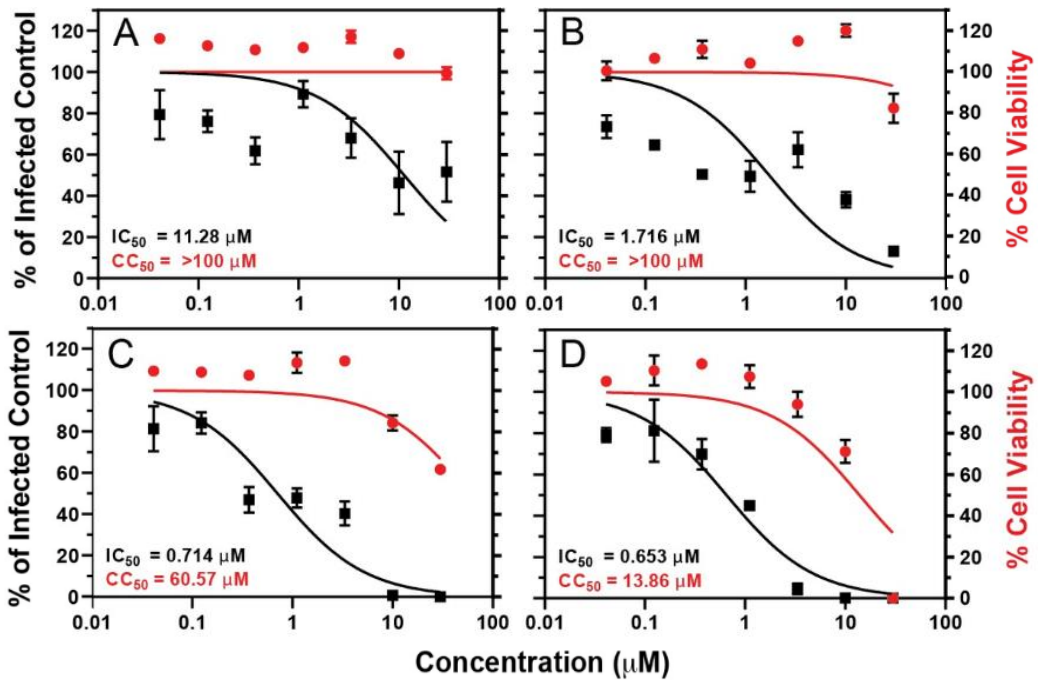

Figure 2.7. Inhibition of SARS-CoV-2 infection by amiloride compounds in Vero E6 cells infected at low MOI and incubated for 48 hours.
IC_{50} and CC_{50} curves of (A) amiloride, (B) DMA, (C) EIPA, (D) HMA. The compounds were added at the indicated concentrations to Vero E6 cells simultaneously with authentic SARS-CoV-2 virus (MOI 0.1) and incubated for 48 hours. Inhibition of infection (solid squares and curves in black) was measured by high-content imaging for intracellular SARS-CoV-2 N protein and is relative to a DMSO-treated infected control. Cytotoxicity (solid circles and curves in red) was measured similarly using a nuclei stain and quantifying cell numbers relative to the DMSO-treated infected control. The curves were calculated using the nonlinear regression analysis in GraphPad Prism 9. IC_{50} and CC_{50} values for each compound are indicated in the plots.

In order to identify the stage of the viral replication cycle affected by the amiloride compounds, their antiviral activity was reevaluated at a high multiplicity of infection (MOI of 1.0

infectious units per cell) and after a relatively brief incubation (Fig 2.8, 18 hours). The same ranking of antiviral activity among the compounds was observed, with HMA and EIPA the most active. The observed IC_{50} values were higher than those measured in the experiments summarized by the data shown in Fig 2.7. This was not unexpected since the antiviral assay of Fig 2.8 was done at an MOI ten-fold higher than that of Fig 2.7, and the time of incubation allowed for only one or two replication cycles (18 hours in Fig 2.8 compared to 48 hours in Fig 2.7). We observed microscopically that EIPA and HMA decreased the number of cells in each infected focus in the monolayer (Fig 2.8C and 2.8D). This effect was especially striking for HMA; most foci contained only one or two cells, suggesting that the spread of infection to adjacent cells in the foci was inhibited. To determine whether the amiloride compounds were inhibiting only this cell-cell spread or were also affecting the infectivity of the inoculum, we enumerated both the number of infected cells and the number of infected-cell-foci (containing one or more cells) and compared the IC_{50} values obtained with each (Fig 2.8). The IC_{50} values obtained using the number of infected cells (Fig 2.8E–2.8H) were less than those obtained using the number of infected-cell-foci (Fig 2.8I–2.8L). These data suggest that the amiloride compounds act late in the viral replication cycle and affect the spread of virus from cell-to-cell, although they do not exclude the possibility of a modest effect on the establishment of infection by cell-free virus.

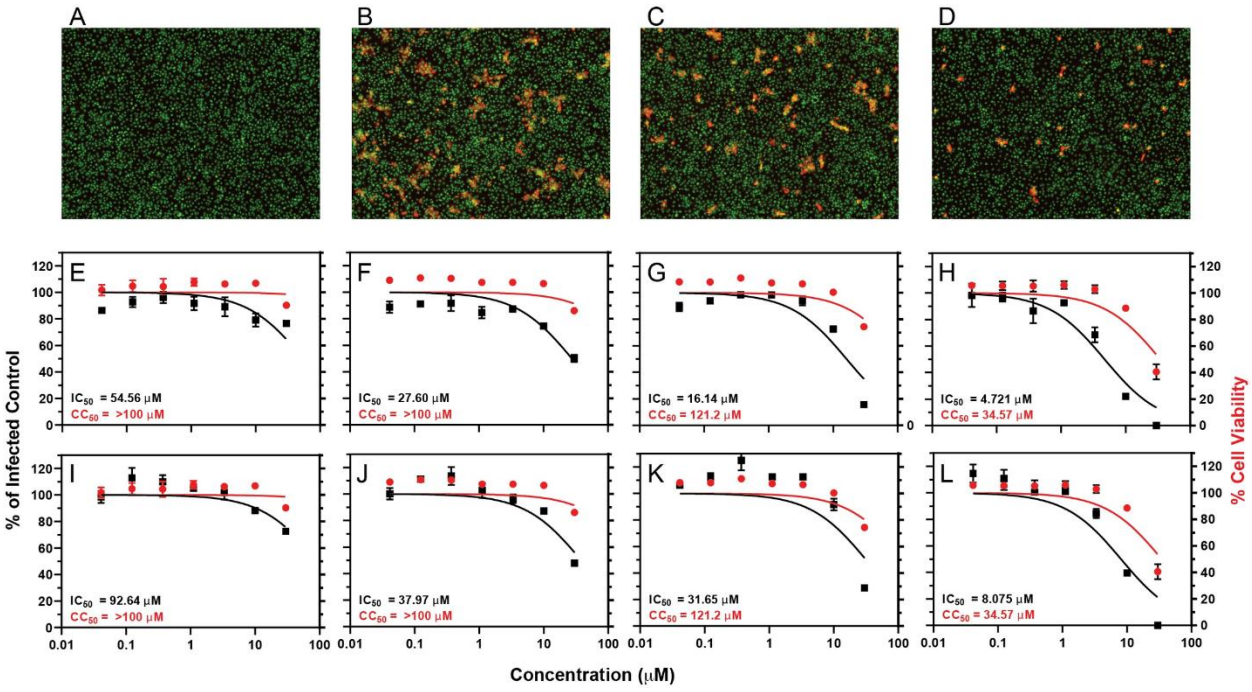

Figure 2.8. Inhibition of SARS-CoV-2 infection by amiloride compounds in Vero E6 cells infected at high MOI and incubated for 18 hours.
(A-D) Images of cells. Green: nuclear stain (Sytox green); red: stain for nucleocapsid (N) using an antibody conjugated to AlexFluor594. (A) Uninfected cells. (B-D) Infected with SARS-CoV-2 and treated with (B) 0.1% DMSO, (C) 10 µM EIPA, and (D) 10 µM HMA. (E-L) IC_{50} and CC_{50} curves of (E and I) amiloride, (F and J) DMA, (G and K) EIPA, and (H an L) HMA for total infected cells (E-H) and foci of infection (I-L). The compounds were added at the indicated concentrations to Vero E6 cells simultaneously with SARS-CoV-2 (MOI 1) and incubated for 18 hours. Inhibition of infection (solid squares and curves in black) was measured by high-content imaging for intracellular SARS-CoV-2 N protein and is relative to a DMSO-treated infected control. Cytotoxicity (solid circles and curves in red) was measured similarly using a nuclei stain and quantifying cell numbers relative to the DMSO-treated infected control. The curves were calculated using the nonlinear regression analysis in GraphPad Prism 9. IC_{50} and CC_{50} values for each compound are indicated in the plots.

N15A and V25F mutations of E protein affect VLP production

Co-expression of the structural proteins M and N of SARS-CoV-2 in HEK293T cells results in a modest release of N-containing virus-like particles (VLPs) judging by the intensity of the N protein band in western blots of culture supernatants after centrifugation though a sucrose cushion. Notably, the added expression of wild-type E protein greatly stimulated the release of VLPs (Fig 2.9A and 2.9B).

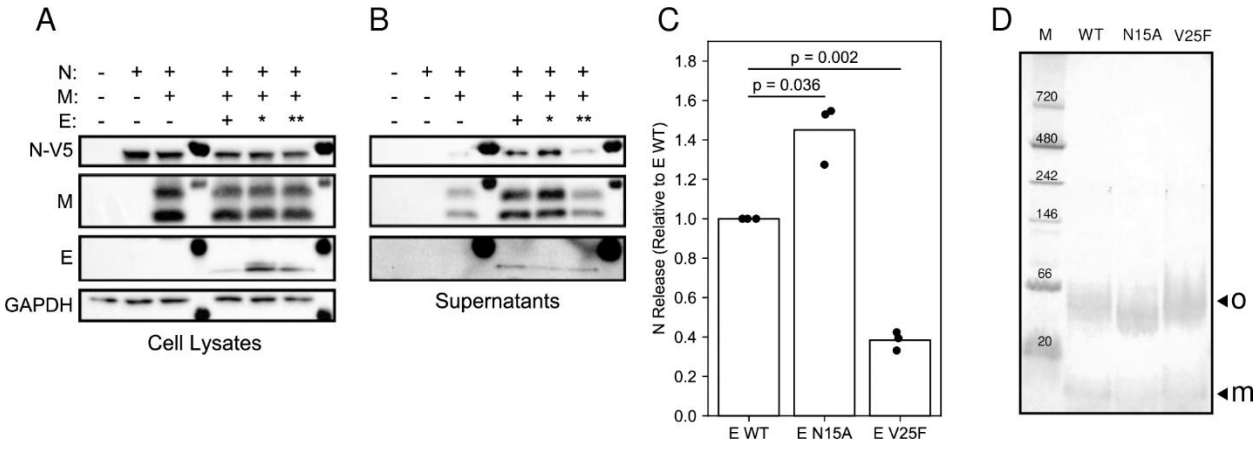

Figure 2.9. Comparison of virus-like particle (VLP) production among wild-type and two mutant E proteins. (A and B) Representative western blots of HEK293T cell lysates and sucrose cushion-purified supernatants following co-transfection with SARS-CoV-2 M, E, and N protein sequences, respectively. For E protein, * indicates the N15A mutant and ** indicates the V25F mutant. (C) Densitometry of the N protein band in purified supernatants from three independent western blot experiments. Each has M+N and the indicated E protein. The relative change over M+N without E protein is plotted for each condition. (D) PFO-PAGE of wild-type, N15A mutant, and V25F mutant E proteins. Monomer (m) and oligomer (o) bands are marked with arrows.

Previous studies of E protein of Infectious Bronchitis Virus (IBV), a gamma coronavirus, showed that mutations within the transmembrane domain altered the ability of VLPs to assemble[148]. To determine the impact of similar mutations in SARS-CoV-2 E protein, two mutants were generated, N15A and V25F, which are analogous to IBV E protein residues Thr16 and Ala26, respectively (SuppFile 2.1). The N15A and V25F mutations in SARS-CoV-2 E protein increased their expression compared to the wild-type protein in HEK293T cell lysates (Fig 2.9A). Similar to the T16A mutation in IBV E protein, the N15A mutation in SARS-CoV-2 E protein increased VLP production by approximately 40% compared to the wild-type E protein, while the V25F mutation decreased VLP production by 60% compared to wild-type E protein, similar to the effect of the A26F mutation on the IBV E protein (Fig 2.9B and 2.9C).

The mutations T16A and A26F in the IBV E protein have been shown to affect its oligomeric state[148]. However, the analogous mutations in the SARS-CoV-2 E protein do not appear to affect its oligomerization *in vitro* under our experimental conditions; in PFO-PAGE, both of these mutant E proteins ran as oligomers with only slightly different migration patterns

compared to the wild-type protein (Fig 2.9D). As expected, both of the mutant proteins ran as monomers in SDS-PAGE with their apparent molecular weights similar to that of the wild-type protein.

Effects of N15A and V25F mutations on structure and HMA binding of E protein

The N15A mutation results in significant chemical shift perturbations of resonances from residues throughout the N-terminal region of E protein, especially for the signals from Ser6, Glu7, Leu12, and Ser16 (Fig 2.10A–2.10C). In contrast, only minor perturbations were observed for signals from residues adjacent to the mutation site in the V25F mutant E protein (Fig 2.10B–2.10D). Since no significant differences were observed among the circular dichroism spectra from wild-type E protein and these two mutant proteins, the relatively large and wide spread chemical shift perturbations by the N15A mutation may result from changes in intermolecular hydrogen bonding involving Asn15 side chains[149].

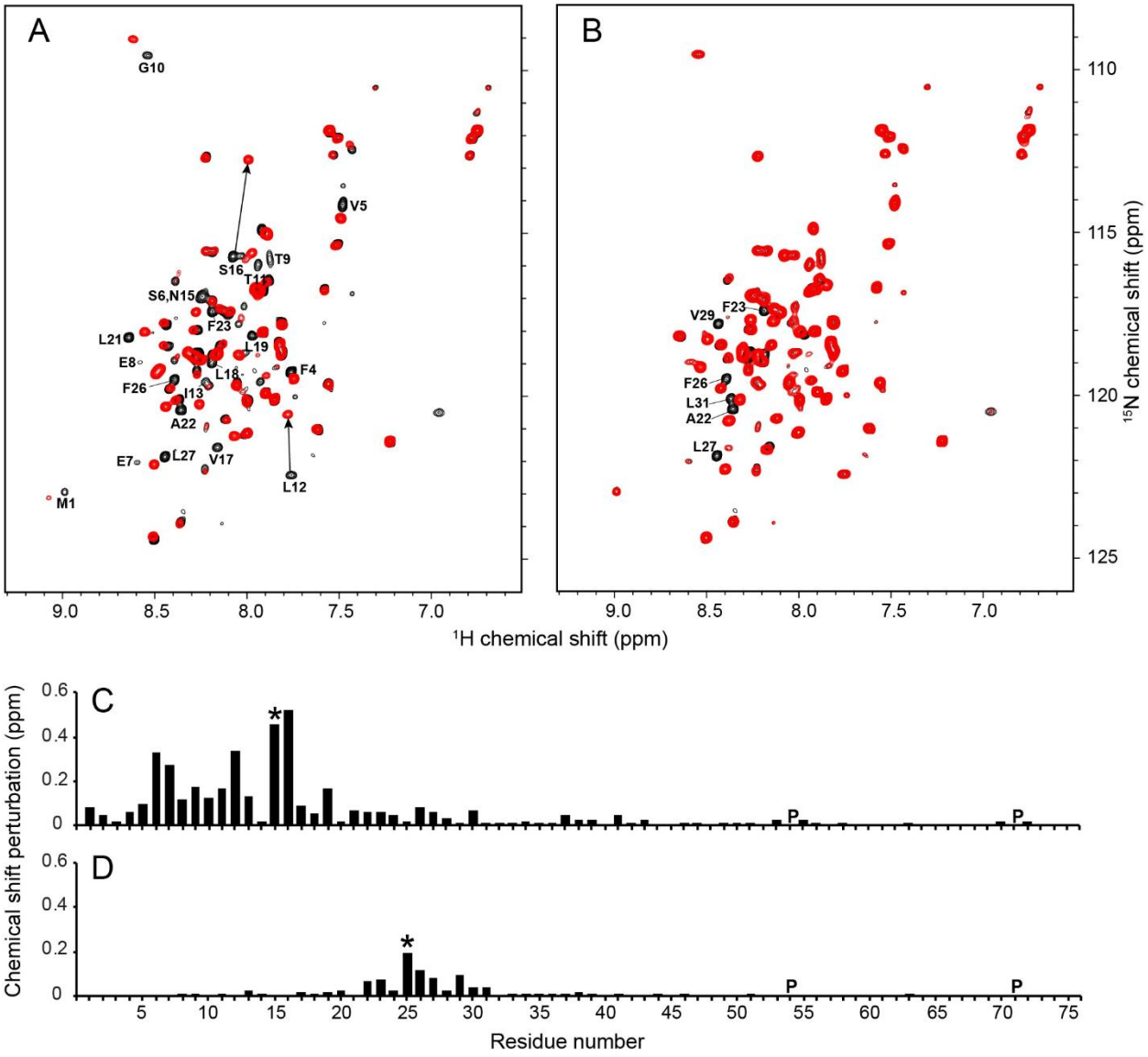

Figure 2.10. Comparison of NMR data of N15A and V25F mutants of E protein.
(A and B) $^{1}H/^{15}N$ HSQC NMR spectra of N15A and V25F mutant E proteins (red contours) superimposed on those from the wild-type E protein (black contours), respectively. The resonances that are significantly perturbed by the mutations are labeled with their assignments. (C and D) Chemical shift perturbation plots for the N15A and V25F mutants of E protein, respectively. The mutation sites are indicated with asterisks and the proline sites are marked as "P".

Previous mutational studies of polypeptides containing the transmembrane helix of SARS-CoV E protein have shown that a single mutation, e.g., N15A or V25F, can disrupt ion channel activity in lipid bilayers[150,151]. We found that the N15A mutation of SARS-CoV-2 E protein decreased HMA binding, since no significant chemical shift changes are observed in the presence of HMA, with the exception of Ser6 (Fig 2.11A–2.11C). By contrast, HMA binding was

not affected by the V25F mutation, since the chemical shifts of signals from residues near the HMA binding site were unchanged and their CSPs were identical to those observed for wild-type E protein (Fig 2.11B–2.11D). Based on these results, it appears that Asn15 is essential for maintaining the conformation of E protein required for binding HMA.

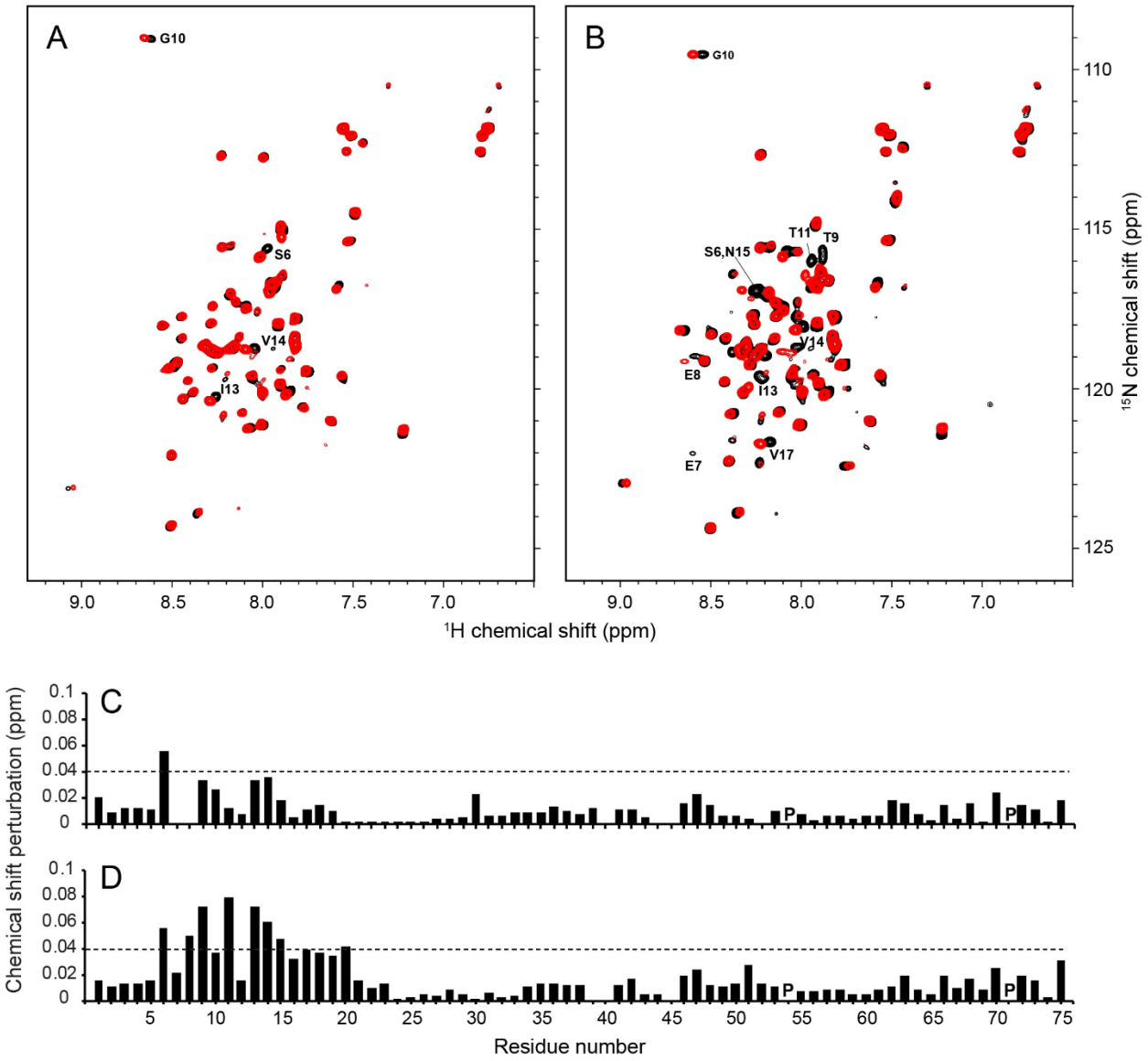

Figure 2.11. Comparisons of the effects of HMA binding on the NMR spectra of N15A and V25F mutants of E protein.

(A and B) $^{1}H/^{15}N$ HSQC NMR spectra of N15A and V25F mutants of E protein in the absence (black contours) and presence (red contours) of HMA, respectively. Significantly perturbed resonances by HMA are labeled with their assignments. (C and D) Chemical shift perturbation plots of the effects of HMA binding to the N15A and V25F mutants of E protein, respectively. The dotted lines indicate 1.5 times the average chemical shift changes of V25F EF by HMA. Proline sites are marked as “P”.

Discussion

The coronavirus SARS-CoV-2 presents formidable challenges to human health, virology, and structural biology. Structural and functional studies of the envelope (E), spike (S), and

membrane (M) proteins are especially challenging because, as shown in Fig 2.1, significant portions of these proteins reside within the phospholipid bilayer of the viral envelope. Here we combine the results of solution NMR spectroscopic studies and virologic studies of SARS-CoV-2 E protein to evaluate its potential as a drug target. We focused our studies on the 75-residue full-length E protein and two overlapping truncated constructs corresponding to the N-terminal transmembrane domain (residues 1–39), that includes a long hydrophobic helix, and the C-terminal cytoplasmic domain (residues 36–75), that includes a short helix and three cysteine residues (Fig 2.2B).

Heterologous expression of viral membrane proteins in *E*. *coli*, the most convenient system for the preparation of milligram amounts of isotopically labeled proteins, is generally problematic. Hydrophobic membrane proteins are prone to aggregation, likely from non-specific hydrophobic intermolecular interactions or possibly incorrect intra- and/or inter- molecular disulfide linkages, the latter of which is especially pertinent for SARS-CoV-2 E protein because it has three closely spaced cysteine residues in its C-terminal domain. The expression of full-length E protein from SARS-CoV, whose sequence is nearly identical to that of the 75-residue protein from SARS-CoV-2, has been reported[140]; however, a modified β-barrel construct was used as the expression tag along with urea to solubilize the inclusion bodies, followed by chemical cleavage and HPLC purification. The resulting protein in DPC or mixed DPC/SDS micelles did not yield NMR spectra suitable for structural studies. Essentially all previous NMR studies of SARS-CoV-2 E protein[108,109,114,115] were carried out on substantially smaller polypeptides with either 31 or 58 residues. In addition to the smaller number of residues in the polypeptides and the missing N- and C- terminal amino acids, the prior studies differ from those described here in several other substantial ways, including the expression system, choice of fusion protein, method of protein expression and purification, choice of micelle-forming detergent, and other experimental parameters. Not surprisingly, there are many significant differences between the findings of the previous NMR studies and those described here.

Moreover, we carried out spectroscopic and virologic studies in parallel, with the results of both serving as controls, suggesting subsequent experiments and guiding the interpretation of the findings.

Nothing could be done without the preparation of isotopically labeled E protein samples suitable for NMR spectroscopy. This formidable barrier required the design and implementation of a novel bacterial expression and purification system (Fig 2.2). There are three notable aspects to our approach: 1) The KSI-E protein fusion protein expression system in C43(DE3) *E*. *coli* cells boosts expression levels and circumvents cytotoxicity by sequestering the overexpressed hydrophobic E protein in inclusion bodies; 2) Insertion of a 24-residue linker, which includes a ten-His tag and a 6-residue (LVPRGS) thrombin cleavage site, between the sequences of the KSI and E proteins, facilitates affinity chromatography purification and enzymatic cleavage because thrombin retains specificity and activity at low detergent concentrations; 3) A single “mild” detergent, HPC, with low CMC, is used to solubilize the protein throughout all steps of isolation, purification, and sample preparation. This eliminates the need for detergent or lipid exchanges and is applicable to full-length, truncated, and mutated constructs of E protein (Figs 2.3 and 2.9). Moreover, this approach to sample preparation may be generally applicable to other membrane proteins. We have already used it to prepare samples of several constructs of the membrane binding domain of the SARS-CoV-2 Spike protein (Fig 2.1) that yield high-resolution NMR spectra. In addition, this approach provides an expedient starting point for the preparation of samples of E protein and potentially other membrane proteins in liquid crystalline phospholipid bilayers at the high lipid to protein ratios required for solid-state NMR spectroscopy under near-native conditions. Initial comparisons between results obtained in HPC micelles by solution NMR and those obtained in phospholipid bilayers by oriented sample solid-state NMR (SuppFile 2.7A) provide assurance that the protein structure is not strongly affected by HPC. This is significant because there have been no previous NMR studies of membrane proteins in HPC micelles. The feasibility of solution NMR

studies of full-length SARS-CoV-2 E protein is demonstrated by the high-resolution and signal-to-noise ratios of resonances from individual amide sites in the two-dimensional $^{1}H/^{15}N$ NMR spectrum of a uniformly ^{15}N labeled sample (Fig 2.3A). This includes the assignment of backbone resonances using standard triple-resonance methods[152] on a uniformly ^{13}C- and ^{15}N-double-labeled sample. Following the sequential assignment of all backbone resonances in the spectra of three overlapping E protein constructs, EF (residues 1–75), ET (residues 1–39), and EC (residues 36–75), it was straightforward to characterize the overall organization, secondary structure, and local dynamics of E protein in HPC micelles using the set of experimental data aligned by residue number in Fig 2.4. The most striking feature to emerge is that E protein has a very long 36-residue α-helix (residues 8–43) in the N-terminal transmembrane domain. There is also a shorter 8-residue α-helix (residues 53–60) in the C-terminal domain that has a different orientation in the protein than the long helix. Since the ^{1}H and ^{15}N chemical shifts of the vast majority of resonances present in the two-dimensional HSQC spectra (Fig 2.3) of the full-length and truncated constructs overlap nearly exactly, the conformations of the N-terminal and C-terminal domains are the same whether alone or as part of the intact protein. The conservation of domain structures, also observed for the small membrane protein Vpu from HIV-1 (references[127,153]), suggests that each domain of E protein has separate roles in the virus life cycle, although this remains to be shown in future *in vivo* experiments.

Prior NMR studies have shown E protein to be largely helical. However, the polypeptides used in the experiments and the model membrane environments differ so much that it is premature to provide a comprehensive analysis of why the lengths, locations, and distortions of the helical segments differ so drastically among various reports[108,109,114,115]. As an example, in 2009 Pervushin et al.[114] found by NMR that all 31 residues of a synthetic polypeptide with a sequence corresponding to residues 8–38 of E protein participated in a continuous α-helix in the presence of DPC. By contrast, in a 2020 report Mandala et al.[109] found by NMR a 21- or 25-

residue helix, with a substantial local distortion, in the same polypeptide prepared by bacterial expression, in the presence of DMPC instead of DPC.

Here we make direct comparisons between our results on the 75-residue full-length E protein and those in the most recent report, cited above, on the widely used 31-residue doubly truncated polypeptide[109]. We find that the 36-residue transmembrane helix is quite long (residues 8–43 of the full-length E protein) compared to the more typical 21- or 25- residue transmembrane helix found by Mandala et al.[109] The results also differ regarding the distortion of this helix. We find it to be continuous and straight, with the exception of the 17-residue core (residues 19–35), identified by resistance to H/D solvent exchange and broadening by the presence of paramagnetic Mn^{2+} in the solution, and most definitively by the dipolar wave analysis that shows that this segment is also straight albeit with a detectably different tilt angle than the co-linear N- and C-terminal portions of the helix (residues 8–18 and 36–43). Instead, Mandala et al.[109] describe a singular 4-residue distortion at residues 20–23. Application of a PISA-Wheel based analysis[154,155] to oriented sample solid-state NMR data (SuppFile 2.7A) shows that the membrane-spanning helix has a large tilt angle (approx. 45^{o}) as necessitated by hydrophobic matching with the 14-carbon methylene chains of DMPC bilayers[156]. By contrast, Mandala et al.[109] interpret their MAS solid-state NMR data to show that this helix has a very small tilt angle in the presence of DMPC. On the one hand, our solid-state NMR experiments were performed on a uniaxially aligned sample with the bilayer normal perpendicular to the direction of the magnetic field (SuppFile 2.7A), therefore the protein must be undergoing rapid rotational diffusion at 35°C in order to yield spectra with narrow single-line resonances. On the other hand, Mandala et al.[109] state that the protein does not undergo fast rigid-body uniaxial rotation at high temperatures. The differences between the structural findings in the two studies may arise from a number of possible sources, such as the difference in the lengths of the polypeptides (75 vs. 31 residues), properties of the membrane-like environments produced by HPC and DMPC, and the use of different NMR approaches (primarily solution NMR

complemented by a contribution from OS solid-state NMR vs. MAS solid-state NMR). The one-dimensional NMR spectrum in SuppFile 2.7 as well as complementary two-dimensional PISEMA spectra foreshadow the structure determination of SARS-CoV-2 E protein in phospholipid bilayers under near-native conditions.

Outstanding questions about E protein include whether it has *in vivo* ion channel activity, and whether this activity is responsible for essential biological functions. Channel activity has been observed for full-length E protein as well as N-terminal constructs containing its principal helix. This has been used as evidence that it forms a pentamer with a central pore characteristic of viroporins. Since it is small viral protein with 75 residues, it is classified as a miniprotein[113]. If its ion channel activity does indeed result from forming a defined oligomer, then it can be categorized as a viroporin. However, its primary and secondary structures differ dramatically from proteins previously described as viroporins. Most notably, the 36-residue helix of the E protein is much longer than the trans-membrane helices identified in archetypical viroporins like Vpu from HIV-1 (reference[128]) and M2 from influenza virus[157], whose shorter transmembrane helices have 18- and 25- residues, respectively.

A hallmark of viroporins is that they form homo-oligomers in the host membranes and their amphipathic transmembrane domain is essential for ion channel activity. The full-length E protein has a monomeric molecular weight of 8.5 kDa. It migrates as an oligomer with an apparent molecular weight of about 50 kDa with a minor band of monomers in PFO-PAGE (Fig 2.8D). The protein has three cysteines (C40, C43, and C44) and at least two of them are conserved across α/β coronaviridae (SuppFile 2.1). The presence of reducing agents does not affect the PFO-PAGE or the NMR spectra of E protein, suggesting that cysteines are not involved in oligomerization or aggregation. Its existence in pentamers is primarily attributed to results from detergent micelle-based analytical ultracentrifugation and BN-PAGE and PFO-PAGE analysis[115,140]. IBV E protein has also been shown to exist as both monomers and oligomers during transient expression and infection by sucrose gradient analysis, and its

oligomers have been proposed to correlate with stimulation of VLP production[148]. As with most other miniproteins, with the notable exception of M2 from influenza, the definition of E protein as a viroporin remains controversial.

Ion-channel activity invites the use of established channel blocking compounds as experimental probes. HMA has exhibited inhibitory activity against E protein ion channels from various coronaviruses, including MHV, HCoV-229E, SCV and FIPV, with a low micromolar range of EC_{50} (references[66,158]) as well as Vpu from HIV-1 (reference[159]) and p7 from HCV[160]. Interactions of HMA with the transmembrane domain of SARS-CoV E protein have been previously examined by NMR[108,109,114,115], and different drug binding sites have been proposed based on the chemical shift perturbations observed for different truncated E protein constructs and experimental conditions. The chemical shift perturbations we observe in spectra of 75-residue full-length SARS-CoV-2 E protein in Fig 2.5 provide a more complete picture of its interactions with HMA than is possible with 31- or 58- residue polypeptides. In addition, the comparison of chemical shift perturbations of three SARS-CoV-2 E protein constructs, EF, ET, and EC, in the presence of HMA clearly demonstrates that N-terminal residues 2–18 are affected by binding HMA. In our spectra, signals from hydrophilic residues (S6, E7, E8, T9, T11, and N15) are strongly affected by HMA. Minor perturbations of signals from residues in the C-terminal end of the trans-membrane helix (residues 35–37) were observed in ET but not in EF and EC and are likely due to nonspecific interactions from the truncated site. This contrasts with a prior result obtained on a truncated E protein construct with residues 8–65 that showed large CSPs for V49 and L65 (reference[115]). A dramatic illustration that caution must be used when drug binding sites are mapped using truncated constructs is an early study of M2 (reference[161]).

A ten-fold molar excess of amiloride and its derivatives DMA, HMA, and EIPA affect resonances from the same set of amino acid residues in SARS-CoV-2 E protein, demonstrating that they utilize the same binding site. With different CSP magnitudes, they display different binding affinities. Amiloride itself did not induce any significant changes, DMA induced modest

changes, and HMA and EIPA induced the largest changes. Notably, the order of affinity of the compounds, EIPA ≈ HMA > DMA >> amiloride, correlates well with their partition coefficient (logP) values: EIPA, 1.3; HMA, 1.3; DMA, 0.1, and amiloride, -0.7 (https://pubchem.ncbi.nlm.nih.gov). Therefore, introduction of bulky aliphatic or aromatic moieties in the 5' position of the amiloride pyrazine ring, which increases the lipophilicity of the compounds, appears to increase their binding affinity for E protein. Most significantly, these findings correlate well with the antiviral activities observed for these compounds in cultures of Vero E6 cells infected with SARS-CoV-2. These data provide additional support for E protein being the likely *in vivo* target of these compounds, and they suggest that inhibition of the ion channel activity may suppress virus replication. Notably, the activity of EIPA and HMA for the Na+/K+ ATPase is higher than of amiloride[162]. Therefore, structure-based optimization of target-selectivity will be necessary in order to develop an amiloride-based drug aimed at E protein.

E protein not only stimulates viral assembly and release but also alters the secretory pathway of the cell in a manner that preserves the function of the Spike protein[163]. Consequently, while attributing the antiviral activity of the amilorides to assembly and release functions is tempting, these compounds might also impair the infectivity of virions by inhibiting the "S-preserving" function of E protein. The activities of the amilorides shown here under the conditions of high-multiplicity infection suggest that much of their antiviral action is at a late-event in the replication cycle, consistent with a block to assembly and release. Nonetheless, we have not fully excluded an effect, albeit modest, on the infectivity of cell-free virus. Such an effect would be consistent with a partial loss of S activity in mediating viral entry into target cells. High-order oligomerization of IBV E protein has been proposed as a requirement for virus assembly[148]. Although we do not observe changes in oligomeric states of N15A or V25F mutant E proteins under our experimental conditions, comparisons of their CSPs suggests that the N15A mutation but not the V25F mutation causes a significant change in the N-terminal region. The N15A mutation affects the entire binding site and abolishes the interaction with HMA,

demonstrating that residue N15 plays a key role in maintaining SARS-CoV-2 E protein's native conformation and its ability to interact with HMA. N15 (or Q15) is highly conserved in alpha and beta coronavirus E proteins. Moreover, a single Gln can mediate helix-helix associations through intermolecular hydrogen bonding within transmembrane domains[164]. Intermolecular hydrogen bonds involving the sidechain of N15 may be essential for maintaining the conformation and orientation of the N-terminal region, a conclusion also suggested by the pentameric model of the E protein oligomer[109]. The small CSPs induced by the V25F mutation are localized near the mutation site indicating that the conformation of the mutant E protein is preserved, which is consistent with its response to HMA being identical to that of the wild-type E protein.

Interest in the structure and function of SARS-CoV-2 E protein motivated the development of an efficient new approach to the expression and purification of membrane proteins so that the full-length protein could be studied. We demonstrate that HMA and EIPA bind to the N-terminal region of the E protein and exhibit antiviral activity against SARS-CoV-2. We also found that residue N15 plays an important role in maintaining the conformation of the HMA binding site, providing insight that might be helpful in the design of drugs targeting E protein. Changes associated with the N15A and V25F mutations are suggestive of involvement of E protein's N-terminal domain in virus assembly and/or release. These biological activities can be correlated with the secondary structure of E protein, which consists of a long hydrophobic transmembrane helix with a large tilt angle between residues 8–43 separated by a slightly dynamic but still structured linker region to a second shorter helix between residues 53–60 with a significantly different tilt angle. Determination of the three-dimensional structure of E protein in phospholipid bilayers is an essential next step that should provide the structural information required to not only understand the protein's biological functions more fully, but also optimize interactions with compounds that have the potential to be developed into antiviral drugs.

Materials and Methods

Design of SARS-CoV-2 E protein constructs

All of the studies described here utilized polypeptides with sequences based on that of the wild-type 75-residue full-length E protein from the SARS-CoV-2 isolate Wuhan-Hu-1 (NC_045512) (Fig 2.2B). To enhance the expression of the viral E protein in *E. coli*, a codon-optimized gene for its amino acid sequence was synthesized using the codons of highly expressed *E. coli* genes (SuppFile 2.2) (www.idtdna.com). The codon-optimized gene was inserted into a modified pET-31b(+) vector (www.emdmillipore.com) and expressed as a ketosteroid isomerase (KSI)-fusion protein. A twenty-four-residue linker sequence incorporating a 10 His-tag and a 6-residue (LVPRGS) thrombin cleavage site was inserted between the KSI and E protein sequences (Fig 2.2A). The same expression and purification system was used with two truncated constructs of E protein, the N-terminal transmembrane domain (ET) (residues 1–39) and the C-terminal cytoplasmic domain (EC) (residues 36–75) (Fig 2.2B). Two EF mutant proteins, N15A EF and V25F EF were generated using a site-directed mutagenesis kit (www.neb.com).

Protein expression and purification

E. coli strain C43(DE3) (www.lucigen.com) cells transformed with the plasmid vectors carrying the target E protein constructs were grown in minimal medium with 1 g/L $(^{15}NH_4)_2SO_4$ as the sole nitrogen source for producing uniformly ^{15}N-labeled samples[165] and with 2 g/L $^{13}C_6$ D-glucose as the carbon source for uniformly $^{13}C/^{15}N$- double-labeled proteins. For selectively (by residue type) ^{15}N-labeled samples, the minimal medium with unlabeled ammonium sulfate was supplemented with 100–500 mg/L of each of 19 amino acid residues and 100 mg/L of the ^{15}N-labeled amino acid. The isotopically labeled compounds were obtained from Cambridge Isotope Laboratories (www.isotope.com). A preculture was grown overnight in

50 mL of Luria-Bertani (LB) broth, then a 1% (v/v) aliquot of the preculture was added to 500 mL of the minimal medium in a two-liter flask. The culture was maintained at 37°C with shaking at 200 rpm until a cell density with an OD_{600} of 0.5 was reached. Expression of the KSI-E protein fusion proteins was induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. After growth for 3 hr (Fig 2.2C lane 2) the cells were harvested by centrifugation at 5,000 xg for 20 min at 4°C. The cell pellet was stored at -80°C overnight.

The cell pellet was resuspended in 72 mL of a solution containing 20 mM Tris-HCl, 500 mM NaCl, pH 8 with 50 µg/mL lysozyme, and 250 units Benzonase nuclease (www.sigmaaldrich.com) per liter of culture. The cell lysate was sonicated (duty cycle 20%, output control 4, Sonic Dismembrator 550, Fisher Scientific) for 10 minutes on ice. 8 mL of 20% (v/v) Triton X-100 was added to the cell lysate to a final concentration of 2% (v/v) and incubated with gentle rotation for one hr at room temperature. The cell lysate was then centrifuged at 20,000 xg for 30 min at 4°C. The supernatant was discarded, and the pellet containing the inclusion bodies was resuspended in 40 mL of 20 mM HEPES, 500 mM NaCl, pH 7.8. 400 mg of n-hexadecylphosphocholine (HPC, fos-choline 16, www.anatrace.com) was added to the suspension at a 1% (w/v) final concentration and Tris (2-carboxyethyl) phosphine hydrochloride (TCEP-HCl) at a final concentration of 1 mM; it was incubated with stirring for 2 hr at room temperature or until the inclusion bodies were completely dissolved. The solubilized inclusion bodies were centrifuged at 40,000 xg for 30 min at 15°C. The supernatant was loaded onto a Ni-NTA superflow (www.qiagen.com) column equilibrated with HPC binding buffer (0.05% HPC, 20 mM HEPES, 500 mM NaCl, pH 7.8) (Fig 2.2C lane 3). The column was washed with five-bed volumes of HPC binding buffer and then 10-bed volumes of HPC washing buffer (0.05% HPC, 20 mM HEPES, 500 mM NaCl, 20 mM imidazole, pH 7.8) (Fig 2.2C lane 4). The KSI-E protein fusion proteins were eluted with two-bed volumes of HPC elution buffer (0.05% HPC, 20 mM HEPES, 500 mM NaCl, 500 mM imidazole, pH 7.8).

The fractions containing the fusion protein were pooled and dialyzed overnight against thrombin cleavage buffer (20 mM HEPES, 50 mM NaCl, 1 mM EDTA, pH 7.8) in a 10 kDa MW cutoff dialysis membrane (www.spectrumchemical.com). Approximately 50 mg of uniformly ^{15}N labeled KSI-E protein fusion protein was obtained from 1L of culture (Fig 2.2C lane 5). 10 units of high-purity thrombin (www.mpbio.com) per mg of fusion protein were added to the dialyzed solution and incubated overnight at room temperature with gentle rotation (Fig 2.2C lane 6). Importantly, thrombin retains its specificity and protease activity in the presence of dilute HPC. The mixture of thrombin-cleaved polypeptides was loaded onto a Ni affinity column equilibrated with HPC binding buffer and the flowthrough containing the target E protein was pooled. Typically, a yield of 10 mg of highly pure ^{15}N-uniformly labeled E protein was obtained from one liter of cell culture (Fig 2.2C lane 7). The TM domain of E (ET), the cytoplasmic domain of E (EC), and the single-site mutants of EF, N15A and V25F, were all prepared following essentially the same protocol and resulted in similar yields.

Electrophoresis

SDS-PAGE was performed using NuPAGE 4–12% Bis-Tris gels in 2-(N-morpholino)ethane sulfonic acid (MES) buffer at room temperature. The protein bands were visualized by Coomassie blue staining (Fig 2.2C). PFO (perfluorooctanoic acid)-PAGE was performed as previously described[128,166] using Novex 4–20% Tris-Glycine gels without SDS. The NuPAGE and Novex precast gels were obtained from Invitrogen (www.thermofisher.com). 5 µg protein samples in HPC binding buffer were mixed with the same volume of the PFO sample buffer (100 mM Tris base, 4% (w/v) NaPFO (www.alfa.com), 20% (v/v) glycerol, 0.05% bromophenol blue, pH 8.0), vortex-mixed, centrifuged for five minutes at 12,000 xg and then applied to the gel. PFO-PAGE was performed with a precooled PFO running buffer (25 mM Tris base, 192 mM glycine, 0.5% (w/v) PFO, pH 8.5) at 120 V for 3.5 hours in a cold room at 4°C. The protein bands were visualized by Coomassie blue staining (Fig 2.8D).

Sample preparation and NMR experiments

Samples for solution NMR experiments were prepared by concentrating the purified proteins with Amicon Ultra-4 10K centrifugal filters (www.endmillipore.com). Samples of 0.5 mM uniformly ^{15}N-labeled and selectively ^{15}N-Leu and ^{15}N-Val labeled E protein in 5% (w/v) (123 mM) HPC, 20 mM HEPES, 50 mM NaCl, 10% (v/v) D_2O, 1 mM DSS, pH 6.5 were used for the two-dimensional ^{1}H/^{15}N HSQC, ^{1}H/^{15}N HSQC-NOESY, ^{1}H/^{15}N heteronuclear NOE, and ^{1}H/^{15}N IPAP-HSQC experiments[167]. Samples of 1 mM uniformly ^{13}C,^{15}N-double labeled proteins in 7% (w/v) (172 mM) HPC, 20 mM HEPES, 50 mM NaCl, 10% (v/v) D_2O, 1 mM DSS, pH 6.5 were used for the three-dimensional HNCA, HN(CO)CA, HNCO, and HN(CA)CO experiments[152]. TCE P-HCl was added to the EF and EC samples at a final concentration of 10 mM.

The HPC concentration in the E protein NMR samples was estimated by comparison of the ^{1}H NMR signal intensity from the HPC acyl chains with that from a 5% (w/v) (123 mM) HPC reference sample. HPC with its low critical micelle concentration and high aggregation number did not pass through the membrane filter with a 10 kDa molecular weight cut-off. Although the HPC concentration varied slightly batch to batch after filter concentration, no significant changes in the protein NMR spectra were observed for HPC:E protein monomer molar ratios between about 100:1 and 250:1. Nevertheless, in order to maintain a consistent ratio of HPC to E protein in the samples, the HPC concentration was adjusted to 5% (w/v) for 0.5 mM E protein and 7% (w/v) for 1 mM E protein.

For H/D exchange experiments, the 90% D_2O NMR samples were prepared by 9-fold dilution of the samples in 90% H_2O/10% D_2O with a 100% D_2O NMR buffer followed by concentration with a 10 kDa molecular weight cut-off filter.

^{15}N-^{1}H residual dipolar couplings (RDCs) were measured by comparison of the $^{1}J_{NH}$ couplings of isotropic and weakly aligned EF samples. Weak alignment was induced

and maintained by addition of Y21M fd bacteriophage to the protein-containing micelle solutions at a final concentration of 20 mg/mL[168].

The NMR experiments were performed on triple-resonance Bruker Avance 800 and Avance 600 spectrometers at 50°C. The two-dimensional $^{1}H/^{15}N$ HSQC-NOESY data were obtained using 100 ms and 200 ms mix times. $^{1}H/^{15}N$ heteronuclear NOE data were obtained with a recycle delay of 4 sec. ^{1}H chemical shifts were referenced to 0 ppm for DSS. The NMR data were processed and analyzed using the computer programs Bruker Topspin 4 (www.bruker.com), NMRpipe/NMR Draw[169], and NMR View[170].

Drug binding

100 mM stock solutions of amiloride, 5'-(N, N-dimethyl)-amiloride (DMA), 5-N-ethyl-N-isopropyl amiloride (EIPA), and 5-(N, N-hexamethylene)-amiloride (HMA) (www.caymanchem.com) were prepared by dissolving the appropriate amount of solid material in deuterated dimethyl sulfoxide (DMSO-d_6).

To observe the chemical shift perturbations of protein resonances by these compounds, two-dimensional $^{1}H/^{15}N$ HSQC spectra were obtained from samples containing 0.2 mM uniformly ^{15}N-labeled protein in the absence and presence of 2 mM amiloride, DMA, HMA, or EIPA. $^{1}H/^{15}N$ HSQC spectra of uniformly ^{15}N-labeled EC in the absence and presence of 2 mM HMA were also obtained. $^{1}H/^{15}N$ HSQC spectra of 0.2 mM uniformly ^{15}N-labeled ET with 0, 0.25, 0.5, 1, 1.5, and 2 mM HMA present in the solution were obtained in order to track the chemical shift changes as a function of concentration. All samples used in the binding experiments contained 2% (v/v) DMSO-d_6 at pH 6.5 to ensure the absence of artifacts. The chemical shift perturbations were calculated using the equation $CSP = [(\Delta\delta_H)^2+(0.2\Delta\delta_N)^2]^{1/2}$, where $\Delta\delta_H$ is the change in the backbone amide ^{1}H chemical shift and $\Delta\delta_N$ is the change in backbone amide ^{15}N chemical shift of an individual resolved and assigned resonance.

SARS-CoV-2 antiviral test of amilorides

Vero E6 and Caco-2 were obtained from ATCC and grown in DMEM (www.corning.com) with 10% FBS, 10mM HEPES, and Penicillin-Streptomycin (www.thermofisher.com). SARS-CoV-2 isolate USA-WA1/2020 (www.beiresources.org) was propagated on Caco-2 cells and infectious units quantified by focus forming assay using Vero E6 (ATCC) cells. Approximately 10e4 Vero E6 cells per well were seeded in a 96-well plate and incubated overnight. Compounds or controls were added at the indicated concentrations with addition of SARS-CoV-2 at a multiplicity of infection (MOI) equal to 1 or 0.1 as indicated in Figs 2.7 and 2.8. After incubation for 18 hr for MOI 1 or 48 hr for MOI 0.1 at 37°C and 5% CO_2, the medium was removed, and the cells were incubated in 4% formaldehyde for 30 minutes at room temperature. Formaldehyde fixed cells were washed with phosphate buffered saline and permeabilized for immunofluorescence using 0.1% Triton X-100 in PBS with 1% bovine serum albumin (BSA) fraction V (www.emdmillipore.com) and stained for SARS-CoV-2 with a primary anti-Nucleocapsid antibody (www.genetex.com GTX135357) labeled with AlexaFluor 594. Cells were washed twice in PBS, and the nuclei were stained with Sytox Green. Four to five images per well were obtained at 10x magnification using an Incucyte S3 (Sartorius). The percent infected cells, nuclei count, and infected foci count were calculated using built-in image analysis tools for the Incucyte S3. Foci were categorized as multi-cell foci or single infected cell by repeating the analysis with area size restrictions in the red (nucleocapsid) channel. IC_{50} and CC_{50} were determined using the nonlinear regression analysis in GraphPad Prism 9 with the bottom and top parameters constrained to 0 and 100, respectively. All work with authentic SARS-CoV-2 was conducted under Biosafety Level-3 conditions at the University of California San Diego. The reagent, SARS-Related Coronavirus 2, Isolate USA-WA1/2020, NR-52281 was deposited by the Centers for Disease Control and Prevention and obtained through BEI Resources, NIAID, NIH.

VLP assays

For SARS-CoV-2 proteins, dsDNA gene fragments (gBlocks) encoding human-codon optimized sequences for M, E, and N-V5, corresponding to those of the SARS-CoV-2 Wuhan-Hu-1 isolate (genbank MN908947.3), were synthesized by Integrated DNA Technologies (www.idtdna.com). The gene fragments were inserted into the pcDNA3.1(-) plasmid backbone between the NotI and EcoRI restriction sites using In-Fusion Cloning (www.takarabio.com). The mutations in the transmembrane region of the E protein were generated from the wild-type E protein construct using the QuikChange site-directed mutagenesis kit (www.agilent.com) and verified by Sanger sequencing (www.genewiz.com). HEK293T cells were cultured in complete Dulbecco's modified Eagle medium containing 10% Fetal Bovine Serum and penicillin-streptomycin.

HEK293T cells were seeded in 6 well plates at a density of 250,000 cells/mL/well in complete Dulbecco's modified Eagle medium (DMEM). The cells were transfected the following day with 500 ng each of plasmids encoding the selected viral proteins and pcDNA2.3 plasmid backbone, using Lipofectamine 2000 (www.thermofisher.com), according to the manufacturer's protocol (3,200 ng total plasmid/well). Twenty-four hr after transfection, the supernatant from each well was clarified by centrifugation at 1,000 x*g* for 5 min at 4°C. Clarified supernatants were then pelleted through 20% sucrose for 1 hr at 23,500 x*g* and 4°C. Pelleted VLPs and cells were lysed in 1X TSDS-PAGE sample buffer containing TCEP 1X Laemmli buffer with 50 mM Tris(2-carboxyethyl) phosphine (www.sigmaaldrich.com) substituted for 2-mercaptoethanol. Cell lysates were boiled for 5 min prior to use. Proteins in VLP and cell lysates were separated on 10% SDS-PAGE gels, transferred to PVDF membranes, and immunoblotted with the following antibodies (Fig 2.7A): mouse monoclonal anti-V5 tag (www.thermofisher.com, #R960-25), rabbit polyclonal anti-SARS M (generous gift of C. Machamer[171]), rabbit polyclonal anti-SARS E (generous gift of C. Machamer[172]), and mouse monoclonal anti-GAPDH (www.genetex.com, #GTX627408). Primary antibodies were detected using horseradish peroxidase (HRP)-

conjugated goat anti-mouse IgG (www.bio-rad.com) or HRP-donkey anti-rabbit IgG (www.bio-rad.com) and Western Clarity detection reagent (www.bio-rad.com). Apparent molecular masses were estimated using a commercial protein standard (www.thermofisher.com, PageRulePlus). Chemiluminescence was detected using a Bio-Rad Chemi Doc imaging system and analyzed using Bio-Rad Image Lab v5.1 software. Densitometry was performed using the Image Lab software (www.bio-rad.com) and statistical significance was determined with Welch's *t*-test.

Data deposition

Backbone NMR resonances of full-length E protein was deposited in the BRMB (accession number: 50813).

Acknowledgements

We thank Francesca M. Marassi and Ye Tian for helpful discussions on the experimental results, and Xuemei Huang and Jinghua Yu for assistance with the instrumentation.

Chapter 2, in full, is a reprint of the material[1] as it appears in PLoS Pathogens 2021. This material was co-authored with Park, Sang Ho, Siddiqi, Haley, Castro, Daniela V., De Angelis, Anna A., Stoneham, Charlotte A., Lewinski, Mary K., Clark, Alex E., Croker, Ben A., Carlin, Aaron F., Guatelli, J., and Opella, Stanley J. The dissertation author was a co-author of this chapter.

REFERENCES

1. Park, S. H., Siddiqi, H., Castro, D. V., Angelis, A. A. De, Oom, A. L., Stoneham, C. A., Lewinski, M. K., Clark, A. E., Croker, B. A., Carlin, A. F., Guatelli, J. & Opella, S. J. Interactions of SARS-CoV-2 envelope protein with amilorides correlate with antiviral activity. *PLOS Pathog.* **17**, e1009519 (2021).

2. Yang, Z., Gray, M. & Winter, L. Why do poxviruses still matter? *Cell Biosci. 2021 111* **11**, 1–8 (2021).

3. Piret, J. & Boivin, G. Pandemics Throughout History. *Front. Microbiol.* **11**, 3594 (2021).

4. Gottlieb, M. S., Schanker, H. M., Fan, P. T., Saxon, A., Weisman, J. D. & Pozalski, I. Pneumocystis Pneumonia --- Los Angeles. *CDC Morbidity and Mortality Weekly Report* https://www.cdc.gov/mmwr/preview/mmwrhtml/june_5.htm (1981).

5. Sharp, P. M. & Hahn, B. H. Origins of HIV and the AIDS Pandemic. *Cold Spring Harb. Perspect. Med.* **1**, a006841 (2011).

6. Global HIV & AIDS statistics — Fact sheet | UNAIDS. https://www.unaids.org/en/resources/fact-sheet.

7. Wilen, C. B., Tilton, J. C. & Doms, R. W. HIV: Cell Binding and Entry. *Cold Spring Harb. Perspect. Med.* **2**, a006866 (2012).

8. Hu, W.-S. & Hughes, S. H. HIV-1 Reverse Transcription. *Cold Spring Harb. Perspect. Med.* **2**, a006882 (2012).

9. Craigie, R. & Bushman, F. D. HIV DNA Integration. *Cold Spring Harb. Perspect. Med.* **2**, a006890 (2012).

10. Guerrero, S., Batisse, J., Libre, C., Bernacchi, S., Marquet, R. & Paillart, J.-C. HIV-1 Replication and the Cellular Eukaryotic Translation Apparatus. *Viruses 2015, Vol. 7, Pages 199-218* **7**, 199–218 (2015).

11. Ramirez, Sharma, Singh, Stoneham, Vollbrecht & Guatelli. Plasma Membrane-Associated Restriction Factors and Their Counteraction by HIV-1 Accessory Proteins. *Cells* **8**, 1020 (2019).

12. Strebel, K. HIV accessory proteins versus host restriction factors. *Curr. Opin. Virol.* **3**, 692–699 (2013).

13. Malim, M. H. & Emerman, M. HIV-1 Accessory Proteins—Ensuring Viral Survival in a Hostile Environment. *Cell Host Microbe* **3**, 388–398 (2008).

14. Arts, E. J. & Hazuda, D. J. HIV-1 Antiretroviral Drug Therapy. *Cold Spring Harb. Perspect. Med.* **2**, a007161 (2012).

15. Smiley, C. L., Rebeiro, P. F., Cesar, C., Belaunzaran-Zamudio, P. F., Crabtree-Ramirez, B., Padgett, D., Gotuzzo, E., Cortes, C. P., Pape, J., Veloso, V. G., McGowan, C. C. & Castilho, J. L. Estimated life expectancy gains with antiretroviral therapy among adults

with HIV in Latin America and the Caribbean: a multisite retrospective cohort study. *Lancet HIV* **8**, e266–e273 (2021).

16. Edwards, J. K., Cole, S. R., Breger, T. L., Rudolph, J. E., Filiatreau, L. M., Buchacz, K., Humes, E., Rebeiro, P. F., D'Souza, G., Gill, M. J., Silverberg, M. J., Mathews, W. C., Horberg, M. A., Thorne, J., Hall, H. I., Justice, A., Marconi, V. C., Lima, V. D., Bosch, R. J., Sterling, T. R., Althoff, K. N., Moore, R. D., Saag, M. & Eron, J. J. Mortality Among Persons Entering HIV Care Compared With the General U.S. Population. *https://doi.org/10.7326/M21-0065* (2021) doi:10.7326/M21-0065.

17. Lundberg, E. & Borner, G. H. H. Spatial proteomics: a powerful discovery tool for cell biology. *Nat. Rev. Mol. Cell Biol. 2018 205* **20**, 285–302 (2019).

18. Thul, P. J., Akesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L. M., Bäckström, A., Danielsson, F., Fagerberg, L., Fall, J., Gatto, L., Gnann, C., Hober, S., Hjelmare, M., Johansson, F., Lee, S., Lindskog, C., Mulder, J., Mulvey, C. M., Nilsson, P., Oksvold, P., Rockberg, J., Schutten, R., Schwenk, J. M., Sivertsson, A., Sjöstedt, E., Skogs, M., Stadler, C., Sullivan, D. P., Tegel, H., Winsnes, C., Zhang, C., Zwahlen, M., Mardinoglu, A., Pontén, F., Von Feilitzen, K., Lilley, K. S., Uhlén, M. & Lundberg, E. A subcellular map of the human proteome. *Science (80-.).* **356**, (2017).

19. Jäger, S., Cimermancic, P., Gulbahce, N., Johnson, J. R., McGovern, K. E., Clarke, S. C., Shales, M., Mercenne, G., Pache, L., Li, K., Hernandez, H., Jang, G. M., Roth, S. L., Akiva, E., Marlett, J., Stephens, M., D'Orso, I., Fernandes, J., Fahey, M., Mahon, C., O'Donoghue, A. J., Todorovic, A., Morris, J. H., Maltby, D. A., Alber, T., Cagney, G., Bushman, F. D., Young, J. A., Chanda, S. K., Sundquist, W. I., Kortemme, T., Hernandez, R. D., Craik, C. S., Burlingame, A., Sali, A., Frankel, A. D. & Krogan, N. J. Global landscape of HIV–human protein complexes. *Nat. 2011 4817381* **481**, 365–370 (2011).

20. Mackinder, L. C. M., Chen, C., Leib, R. D., Patena, W., Blum, S. R., Rodman, M., Ramundo, S., Adams, C. M. & Jonikas, M. C. A Spatial Interactome Reveals the Protein Organization of the Algal CO2-Concentrating Mechanism. *Cell* **171**, 133-147.e14 (2017).

21. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* **196**, 801–810 (2012).

22. Kim, D. I., Jensen, S. C., Noble, K. A., KC, B., Roux, K. H., Motamedchaboki, K. & Roux, K. J. An improved smaller biotin ligase for BioID proximity labeling. *https://doi.org/10.1091/mbc.E15-12-0844* **27**, 1188–1196 (2016).

23. Martell, J. D., Deerinck, T. J., Sancak, Y., Poulos, T. L., Mootha, V. K., Sosinsky, G. E., Ellisman, M. H. & Ting, A. Y. Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nat. Biotechnol. 2012 3011* **30**, 1143–1148 (2012).

24. Hung, V., Udeshi, N. D., Lam, S. S., Loh, K. H., Cox, K. J., Pedram, K., Carr, S. A. & Ting, A. Y. Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2. *Nat. Protoc.* **11**, 456–75 (2016).

25. Branon, T. C., Bosch, J. A., Sanchez, A. D., Udeshi, N. D., Svinkina, T., Carr, S. A., Feldman, J. L., Perrimon, N. & Ting, A. Y. Efficient proximity labeling in living cells and organisms with TurboID. *Nat. Biotechnol.* **36**, 880–887 (2018).

26. Lobingier, B. T., Hüttenhain, R., Eichel, K., Miller, K. B., Ting, A. Y., von Zastrow, M. & Krogan, N. J. An Approach to Spatiotemporally Resolve Protein Interaction Networks in Living Cells. *Cell* **169**, 350-360.e12 (2017).

27. Pagliarini, D. J., Calvo, S. E., Chang, B., Sheth, S. A., Vafai, S. B., Ong, S. E., Walford, G. A., Sugiana, C., Boneh, A., Chen, W. K., Hill, D. E., Vidal, M., Evans, J. G., Thorburn, D. R., Carr, S. A. & Mootha, V. K. A Mitochondrial Protein Compendium Elucidates Complex I Disease Biology. *Cell* **134**, 112–123 (2008).

28. Itzhak, D. N., Tyanova, S., Cox, J. & Borner, G. H. H. Global, quantitative and dynamic mapping of protein subcellular localization. *Elife* **5**, 1–36 (2016).

29. Christoforou, A., Mulvey, C. M., Breckels, L. M., Geladaki, A., Hurrell, T., Hayward, P. C., Naake, T., Gatto, L., Viner, R., Arias, A. M. & Lilley, K. S. A draft map of the mouse pluripotent stem cell spatial proteome. *Nat. Commun.* **7**, 1–12 (2016).

30. Calvo, S. E., Clauser, K. R. & Mootha, V. K. MitoCarta2.0: An updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res.* **44**, D1251–D1257 (2016).

31. Borner, G. H. H., Hein, M. Y., Hirst, J., Edgar, J. R., Mann, M. & Robinson, M. S. Fractionation profiling: a fast and versatile approach for mapping vesicle proteomes and protein-protein interactions. *Mol. Biol. Cell* **25**, 3178–3194 (2014).

32. Davies, A. K., Itzhak, D. N., Edgar, J. R., Archuleta, T. L., Hirst, J., Jackson, L. P., Robinson, M. S. & Borner, G. H. H. AP-4 vesicles contribute to spatial control of autophagy via RUSC-dependent peripheral delivery of ATG9A. *Nat. Commun.* **9**, 3958 (2018).

33. Sadowski, P. G., Dunkley, T. P., Shadforth, I. P., Dupree, P., Bessant, C., Griffin, J. L. & Lilley, K. S. Quantitative proteomic approach to study subcellular localization of membrane proteins. *Nat. Protoc. 2006 14* **1**, 1778–1789 (2006).

34. Mulvey, C. M., Breckels, L. M., Geladaki, A., Britovšek, N. K., Nightingale, D. J. H., Christoforou, A., Elzek, M., Deery, M. J., Gatto, L. & Lilley, K. S. Using hyperLOPIT to perform high-resolution mapping of the spatial proteome. *Nat. Protoc.* **12**, 1110–1135 (2017).

35. Itzhak, D. N., Davies, C., Tyanova, S., Mishra, A., Williamson, J., Antrobus, R., Cox, J., Weekes, M. P. & Borner, G. H. H. A Mass Spectrometry-Based Approach for Mapping Protein Subcellular Localization Reveals the Spatial Proteome of Mouse Primary Neurons. *Cell Rep.* **20**, 2706–2718 (2017).

36. Geladaki, A., Britovšek, N. K., Breckels, L. M., Smith, T. S., Vennard, O. L., Mulvey, C. M., Crook, O. M., Gatto, L. & Lilley, K. S. Combining LOPIT with differential ultracentrifugation for high-resolution spatial proteomics. *Nat. Commun. 2019 101* **10**, 1–15 (2019).

37. Gatto, L., Breckels, L. M. & Lilley, K. S. Assessing sub-cellular resolution in spatial proteomics experiments. *Curr. Opin. Chem. Biol.* **48**, 123–149 (2019).

38. Jean Beltran, P. M., Mathias, R. A. & Cristea, I. M. A Portrait of the Human Organelle Proteome In Space and Time during Cytomegalovirus Infection. *Cell Syst.* **3**, 361–373 (2016).

39. Crook, O. M., Breckels, L. M., Lilley, K. S., Kirk, P. D. W. & Gatto, L. A Bioconductor workflow for the Bayesian analysis of spatial proteomics. *F1000Research 2019 8446* **8**, 446 (2019).

40. Kennedy, M. A., Hofstadter, W. A. & Cristea, I. M. TRANSPIRE: A Computational Pipeline to Elucidate Intracellular Protein Movements from Spatial Proteomics Data Sets. *J. Am. Soc. Mass Spectrom.* **31**, 1422–1439 (2020).

41. Crook, O. M., Davies, C. T. R., Gatto, L., Kirk, P. D. W. & Lilley, K. S. Inferring differential subcellular localisation in comparative spatial proteomics using BANDLE. *bioRxiv* 2021.01.04.425239 (2021) doi:10.1101/2021.01.04.425239.

42. Puelles, V. G., Lütgehetmann, M., Lindenmeyer, M. T., Sperhake, J. P., Wong, M. N., Allweiss, L., Chilla, S., Heinemann, A., Wanner, N., Liu, S., Braun, F., Lu, S., Pfefferle, S., Schröder, A. S., Edler, C., Gross, O., Glatzel, M., Wichmann, D., Wiech, T., Kluge, S., Pueschel, K., Aepfelbacher, M. & Huber, T. B. Multiorgan and Renal Tropism of SARS-CoV-2. *https://doi.org/10.1056/NEJMc2011400* **383**, 590–592 (2020).

43. Murgolo, N., Therien, A. G., Howell, B., Klein, D., Koeplinger, K., Lieberman, L. A., Adam, G. C., Flynn, J., McKenna, P., Swaminathan, G., Hazuda, D. J. & Olsen, D. B. SARS-CoV-2 tropism, entry, replication, and propagation: Considerations for drug discovery and development. *PLOS Pathog.* **17**, e1009225 (2021).

44. Liu, J., Li, Y., Liu, Q., Yao, Q., Wang, X., Zhang, H., Chen, R., Ren, L., Min, J., Deng, F., Yan, B., Liu, L., Hu, Z., Wang, M. & Zhou, Y. SARS-CoV-2 cell tropism and multiorgan infection. *Cell Discov. 2021 71* **7**, 1–4 (2021).

45. Pekar, J., Worobey, M., Moshiri, N., Scheffler, K. & Wertheim, J. O. Timing the SARS-CoV-2 index case in Hubei province. *Science (80-.).* **372**, 412–417 (2021).

46. D'Arienzo, M. & Coniglio, A. Assessment of the SARS-CoV-2 basic reproduction number, R0, based on the early phase of COVID-19 outbreak in Italy. *Biosaf. Heal.* **2**, 57–59 (2020).

47. Petersen, E., Koopmans, M., Go, U., Hamer, D. H., Petrosillo, N., Castelli, F., Storgaard, M., Al Khalili, S. & Simonsen, L. Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *Lancet Infect. Dis.* **20**, e238–e244 (2020).

48. Wiersinga, W. J., Rhodes, A., Cheng, A. C., Peacock, S. J. & Prescott, H. C. Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. *JAMA* **324**, 782–793 (2020).

49. Chow, E. J., Schwartz, N. G., Tobolowsky, F. A., Zacks, R. L. T., Huntington-Frazier, M., Reddy, S. C. & Rao, A. K. Symptom Screening at Illness Onset of Health Care Personnel

With SARS-CoV-2 Infection in King County, Washington. *JAMA* **323**, 2087–2089 (2020).

50. Young, B. E., Ong, S. W. X., Kalimuddin, S., Low, J. G., Tan, S. Y., Loh, J., Ng, O.-T., Marimuthu, K., Ang, L. W., Mak, T. M., Lau, S. K., Anderson, D. E., Chan, K. S., Tan, T. Y., Ng, T. Y., Cui, L., Said, Z., Kurupatham, L., Chen, M. I.-C., Chan, M., Vasoo, S., Wang, L.-F., Tan, B. H., Lin, R. T. P., Lee, V. J. M., Leo, Y.-S., Lye, D. C. & Team, for the S. 2019 N. C. O. R. Epidemiologic Features and Clinical Course of Patients Infected With SARS-CoV-2 in Singapore. *JAMA* **323**, 1488–1494 (2020).

51. CDC. Risk for COVID-19 Infection, Hospitalization, and Death By Age Group | CDC. https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-age.html (2021).

52. Griffin, D. O., Brennan-Rieder, D., Ngo, B., Kory, P., Confalonieri, M., Shapiro, L., Iglesias, J., Dube, M., Nanda, N., In, G. K., Arkfeld, D., Chaudhary, P., Campese, V. M., Hanna, D. L., Sawcer, D., Ehresmann, G., Peng, D., Smorgorzewski, M., Amstrong, A., Vinjevoll, E. H., Dasgupta, R., Sattler, F. R., Mussini, C., Mitjà, O., Soriano, V., Peschanski, N., Hayem, G., Piccirillo, M. C., Lobo-Ferreira, A., Rivero, I. B., Hung, I. F. H., Rendell, M., Ditmore, S., Varon, J. & Marik, P. The Importance of Understanding the Stages of COVID-19 in Treatment and Trials. *AIDS Rev.* **23**, 40–47 (2021).

53. NIH. Nonhospitalized Patients: General Management | COVID-19 Treatment Guidelines. https://www.covid19treatmentguidelines.nih.gov/management/clinical-management/nonhospitalized-patients--general-management/ (2021).

54. NIH. Anti-SARS-CoV-2 Monoclonal Antibodies | COVID-19 Treatment Guidelines. https://www.covid19treatmentguidelines.nih.gov/therapies/anti-sars-cov-2-antibody-products/anti-sars-cov-2-monoclonal-antibodies/ (2021).

55. NIH. Hospitalized Adults: Therapeutic Management | COVID-19 Treatment Guidelines. https://www.covid19treatmentguidelines.nih.gov/management/clinical-management/hospitalized-adults--therapeutic-management/ (2021).

56. Perlman, S. & Netland, J. Coronaviruses post-SARS: Update on replication and pathogenesis. *Nature Reviews Microbiology* vol. 7 439–450 (2009).

57. DeDiego, M. L., Álvarez, E., Almazán, F., Rejas, M. T., Lamirande, E., Roberts, A., Shieh, W.-J., Zaki, S. R., Subbarao, K. & Enjuanes, L. A Severe Acute Respiratory Syndrome Coronavirus That Lacks the E Gene Is Attenuated In Vitro and In Vivo. *J. Virol.* **81**, 1701–1713 (2007).

58. Kuo, L., Hurst, K. R. & Masters, P. S. Exceptional Flexibility in the Sequence Requirements for Coronavirus Small Envelope Protein Function. *J. Virol.* **81**, 2249–2262 (2007).

59. Ortego, J., Escors, D., Laude, H. & Enjuanes, L. Generation of a Replication-Competent, Propagation-Deficient Virus Vector Based on the Transmissible Gastroenteritis Coronavirus Genome. *J. Virol.* **76**, 11518–11529 (2002).

60. Lamirande, E. W., DeDiego, M. L., Roberts, A., Jackson, J. P., Alvarez, E., Sheahan, T., Shieh, W.-J., Zaki, S. R., Baric, R., Enjuanes, L. & Subbarao, K. A Live Attenuated

Severe Acute Respiratory Syndrome Coronavirus Is Immunogenic and Efficacious in Golden Syrian Hamsters. *J. Virol.* **82**, 7721–7724 (2008).

61. Netland, J., DeDiego, M. L., Zhao, J., Fett, C., Álvarez, E., Nieto-Torres, J. L., Enjuanes, L. & Perlman, S. Immunization with an attenuated severe acute respiratory syndrome coronavirus deleted in E protein protects against lethal respiratory disease. *Virology* **399**, 120–128 (2010).

62. Fett, C., DeDiego, M. L., Regla-Nava, J. A., Enjuanes, L. & Perlman, S. Complete Protection against Severe Acute Respiratory Syndrome Coronavirus-Mediated Lethal Respiratory Disease in Aged Mice by Immunization with a Mouse-Adapted Virus Lacking E Protein. *J. Virol.* **87**, 6551–6559 (2013).

63. Regla-Nava, J. A., Nieto-Torres, J. L., Jimenez-Guardeño, J. M., Fernandez-Delgado, R., Fett, C., Castaño-Rodríguez, C., Perlman, S., Enjuanes, L. & DeDiego, M. L. Severe Acute Respiratory Syndrome Coronaviruses with Mutations in the E Protein Are Attenuated and Promising Vaccine Candidates. *J. Virol.* **89**, 3870–3887 (2015).

64. Nieto-Torres, J. L., DeDiego, M. L., Verdiá-Báguena, C., Jimenez-Guardeño, J. M., Regla-Nava, J. A., Fernandez-Delgado, R., Castaño-Rodriguez, C., Alcaraz, A., Torres, J., Aguilella, V. M. & Enjuanes, L. Severe Acute Respiratory Syndrome Coronavirus Envelope Protein Ion Channel Activity Promotes Virus Fitness and Pathogenesis. *PLoS Pathog.* **10**, e1004077 (2014).

65. Zheng, M., Karki, R., Williams, E. P., Yang, D., Fitzpatrick, E., Vogel, P., Jonsson, C. B. & Kanneganti, T.-D. TLR2 senses the SARS-CoV-2 envelope protein to produce inflammatory cytokines. *Nat. Immunol.* **22**, 829–838 (2021).

66. Wilson, L., Gage, P. & Ewart, G. Hexamethylene amiloride blocks E protein ion channels and inhibits coronavirus replication. *Virology* **353**, 294–306 (2006).

67. Nicholson, K. G. & Wiselka, M. J. Amantadine for influenza A. *Br. Med. J.* **302**, 425–426 (1991).

68. Nightingale, D. J., Geladaki, A., Breckels, L. M., Oliver, S. G. & Lilley, K. S. The subcellular organisation of Saccharomyces cerevisiae. *Curr. Opin. Chem. Biol.* **48**, 86–95 (2019).

69. Jean Beltran, P. M., Cook, K. C., Hashimoto, Y., Galitzine, C., Murray, L. A., Vitek, O. & Cristea, I. M. Infection-Induced Peroxisome Biogenesis Is a Metabolic Strategy for Herpesvirus Replication. *Cell Host Microbe* **24**, 526-541.e7 (2018).

70. Gatto, L., Breckels, L. M., Wieczorek, S., Burger, T. & Lilley, K. S. Mass-spectrometry-based spatial proteomics data analysis using pRoloc and pRolocdata. *Bioinformatics* **30**, 1322–1324 (2014).

71. Lapek, J. D., Lewinski, M. K., Wozniak, J. M., Guatelli, J. & Gonzalez, D. J. Quantitative Temporal Viromics of an Inducible HIV-1 Model Yields Insight to Global Host Targets and Phospho-Dynamics Associated with Protein Vpr. *Mol. Cell. Proteomics* **16**, 1447–1461 (2017).

72. Guy, B., Rivière, Y., Dott, K., Regnault, A. & Kieny, M. P. Mutational analysis of the HIV nef protein. *Virology* **176**, 413–425 (1990).

73. Yu, G. & Felsted, R. L. Effect of myristoylation on p27nef subcellular distribution and suppression of HIV-LTR transcription. *Virology* **187**, 46–55 (1992).

74. Chowers, M. Y., Spina, C. A., Kwoh, T. J., Fitch, N. J., Richman, D. D. & Guatelli, J. C. Optimal infectivity in vitro of human immunodeficiency virus type 1 requires an intact nef gene. *J. Virol.* **68**, 2906–14 (1994).

75. Pawlak, E. N., Dirk, B. S., Jacob, R. A., Johnson, A. L. & Dikeakos, J. D. The HIV-1 accessory proteins Nef and Vpu downregulate total and cell surface CD28 in CD4+ T cells. *Retrovirology* **15**, 6 (2018).

76. Thoulouze, M. I., Sol-Foulon, N., Blanchet, F., Dautry-Varsat, A., Schwartz, O. & Alcover, A. Human Immunodeficiency Virus Type-1 Infection Impairs the Formation of the Immunological Synapse. *Immunity* **24**, 547–561 (2006).

77. Sugden, S. M., Bego, M. G., Pham, T. N. Q. & Cohen, É. A. Remodeling of the host cell plasma membrane by HIV-1 nef and Vpu: A strategy to ensure viral fitness and persistence. *Viruses* **8**, 1–30 (2016).

78. Rhee, H., Zou, P., Udeshi, N. D., Martell, J. D., Mootha, V. K., Carr, S. A. & Ting, A. Y. Proteomic Mapping of Mitochondria. *Science (80-.).* **339**, 1328 (2013).

79. Lübke, T., Lobel, P. & Sleat, D. E. Proteomics of the lysosome. *Biochimica et Biophysica Acta - Molecular Cell Research* vol. 1793 625–635 (2009).

80. Fu, W., Sanders-Beer, B. E., Katz, K. S., Maglott, D. R., Pruitt, K. D. & Ptak, R. G. Human immunodeficiency virus type 1, human protein interaction database at NCBI. *Nucleic Acids Res.* **37**, D417–D422 (2009).

81. Tokarev, A., Stoneham, C., Lewinski, M. K., Mukim, A., Deshmukh, S., Vollbrecht, T., Spina, C. A. & Guatelli, J. Pharmacologic Inhibition of Nedd8 Activation Enzyme Exposes CD4-Induced Epitopes within Env on Cells Expressing HIV-1. *J. Virol.* **90**, 2486–2502 (2016).

82. Chesebro, B., Wehrly, K., Nishio, J. & Perryman, S. Macrophage-tropic human immunodeficiency virus isolates from different patients exhibit unusual V3 envelope sequence homogeneity in comparison with T-cell-tropic isolates: definition of critical amino acids involved in cell tropism. *J. Virol.* **66**, 6547–54 (1992).

83. Abacioglu, Y. H., Fouts, T. R., Laman, J. D., Claassen, E., Pincus, S. H., Moore, J. P., Roby, C. A., Kamin-Lewis, R. & Lewis, G. K. Epitope Mapping and Topology of Baculovirus-Expressed HIV-1 gp160 Determined with a Panel of Murine Monoclonal Antibodies. *AIDS Res. Hum. Retroviruses* **10**, 371–381 (1994).

84. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).

85. Käll, L., Canterbury, J. D., Weston, J., Noble, W. S. & MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods 2007 411* **4**, 923–925 (2007).

86. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods 2007 43* **4**, 207–214 (2007).

87. Adhikari, S., Nice, E. C., Deutsch, E. W., Lane, L., Omenn, G. S., Pennington, S. R., Paik, Y.-K., Overall, C. M., Corrales, F. J., Cristea, I. M., Van Eyk, J. E., Uhlén, M., Lindskog, C., Chan, D. W., Bairoch, A., Waddington, J. C., Justice, J. L., LaBaer, J., Rodriguez, H., He, F., Kostrzewa, M., Ping, P., Gundry, R. L., Stewart, P., Srivastava, S., Srivastava, S., Nogueira, F. C. S., Domont, G. B., Vandenbrouck, Y., Lam, M. P. Y., Wennersten, S., Vizcaino, J. A., Wilkins, M., Schwenk, J. M., Lundberg, E., Bandeira, N., Marko-Varga, G., Weintraub, S. T., Pineau, C., Kusebauch, U., Moritz, R. L., Ahn, S. B., Palmblad, M., Snyder, M. P., Aebersold, R. & Baker, M. S. A high-stringency blueprint of the human proteome. *Nat. Commun. 2020 111* **11**, 1–16 (2020).

88. Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J. & Von Mering, C. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).

89. Hung, C.-H., Thomas, L., Ruby, C. E., Atkins, K. M., Morris, N. P., Knight, Z. A., Scholz, I., Barklis, E., Weinberg, A. D., Shokat, K. M. & Thomas, G. HIV-1 Nef Assembles a Src Family Kinase-ZAP-70/Syk-PI3K Cascade to Downregulate Cell-Surface MHC-I. *Cell Host Microbe* **1**, 121–133 (2007).

90. Río-Iñiguez, I. del, Vázquez-Chávez, E., Cuche, C., Bartolo, V. Di, Bouchet, J. & Alcover, A. HIV-1 Nef Hijacks Lck and Rac1 Endosomal Traffic To Dually Modulate Signaling-Mediated and Actin Cytoskeleton–Mediated T Cell Functions. *J. Immunol.* **201**, 2624–2640 (2018).

91. Federico, M., Percario, Z., Olivetta, E., Fiorucci, G., Muratori, C., Micheli, A., Romeo, G. & Affabris, E. HIV-1 Nef activates STAT1 in human monocytes/macrophages through the release of soluble factors. *Blood* **98**, 2752–2761 (2001).

92. Benichou, S., Bomsel, M., Bodéus, M., Durand, H., Douté, M., Letourneur, F., Camonis, J. & Benarous, R. Physical interaction of the HIV-1 Nef protein with beta-COP, a component of non-clathrin-coated vesicles essential for membrane traffic. *J. Biol. Chem.* **269**, 30073–30076 (1994).

93. Piguet, V., Gu, F., Foti, M., Demaurex, N., Gruenberg, J., Carpentier, J.-L. & Trono, D. Nef-Induced CD4 Degradation: A Diacidic-Based Motif in Nef Functions as a Lysosomal Targeting Signal through the Binding of beta-COP in Endosomes. *Cell* **97**, 63–73 (1999).

94. Stella, A. O. & Turville, S. All-Round Manipulation of the Actin Cytoskeleton by HIV. *Viruses 2018, Vol. 10, Page 63* **10**, 63 (2018).

95. Chaudhuri, R., Lindwasser, O. W., Smith, W. J., Hurley, J. H. & Bonifacino, J. S.

Downregulation of CD4 by Human Immunodeficiency Virus Type 1 Nef Is Dependent on Clathrin and Involves Direct Interaction of Nef with the AP2 Clathrin Adaptor. *J. Virol.* **81**, 3877–3890 (2007).

96. Craig, H. M., Pandori, M. W. & Guatelli, J. C. Interaction of HIV-1 Nef with the cellular dileucine-based sorting pathway is required for CD4 down-regulation and optimal viral infectivity. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 11229–34 (1998).

97. Jia, X., Singh, R., Homann, S., Yang, H., Guatelli, J. & Xiong, Y. Structural basis of evasion of cellular adaptive immunity by HIV-1 Nef. *Nat. Struct. Mol. Biol. 2012 197* **19**, 701–706 (2012).

98. Gatto, L., Breckels, L. M., Burger, T., Nightingale, D. J. H., Groen, A. J., Campbell, C., Nikolovski, N., Mulvey, C. M., Christoforou, A., Ferro, M. & Lilley, K. S. A Foundation for Reliable Spatial Proteomics Data Analysis. *Mol. Cell. Proteomics* **13**, 1937–1952 (2014).

99. Naamati, A., Williamson, J. C., Greenwood, E. J. D., Marelli, S., Lehner, P. J. & Matheson, N. J. Functional proteomic atlas of HIV infection in primary human CD4+ T cells. *Elife* **8**, (2019).

100. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* **5**, 536–544 (2020).

101. Aronin, S. I. & Sadigh, M. Severe acute respiratory syndrome. *Conn. Med.* **68**, 207–15 (2004).

102. de Groot, R. J., Baker, S. C., Baric, R. S., Brown, C. S., Drosten, C., Enjuanes, L., Fouchier, R. A. M., Galiano, M., Gorbalenya, A. E., Memish, Z. A., Perlman, S., Poon, L. L. M., Snijder, E. J., Stephens, G. M., Woo, P. C. Y., Zaki, A. M., Zambon, M. & Ziebuhr, J. Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. *J. Virol.* **87**, 7790–2 (2013).

103. Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C. & Zhang, Y.-Z. A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).

104. Li, Y.-D., Chi, W.-Y., Su, J.-H., Ferrall, L., Hung, C.-F. & Wu, T.-C. Coronavirus vaccine development: from SARS and MERS to COVID-19. *J. Biomed. Sci.* **27**, 104 (2020).

105. Heaton, P. M. The Covid-19 Vaccine-Development Multiverse. *N. Engl. J. Med.* **383**, 1986–1988 (2020).

106. Ogando, N. S., Dalebout, T. J., Zevenhoven-Dobbe, J. C., Limpens, R. W. A. L., van der Meer, Y., Caly, L., Druce, J., de Vries, J. J. C., Kikkert, M., Bárcena, M., Sidorov, I. & Snijder, E. J. SARS-coronavirus-2 replication in Vero E6 cells: replication kinetics, rapid adaptation and cytopathology. *J. Gen. Virol.* **101**, 925–940 (2020).

107. Neuman, B. W., Kiss, G., Kunding, A. H., Bhella, D., Baksh, M. F., Connelly, S., Droese, B., Klaus, J. P., Makino, S., Sawicki, S. G., Siddell, S. G., Stamou, D. G., Wilson, I. A.,

Kuhn, P. & Buchmeier, M. J. A structural analysis of M protein in coronavirus assembly and morphology. *J. Struct. Biol.* **174**, 11–22 (2011).

108. Surya, W., Li, Y. & Torres, J. Structural model of the SARS coronavirus E channel in LMPG micelles. *Biochim. Biophys. Acta - Biomembr.* **1860**, 1309–1317 (2018).

109. Mandala, V. S., McKay, M. J., Shcherbakov, A. A., Dregni, A. J., Kolocouris, A. & Hong, M. Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers. *Nat. Struct. Mol. Biol.* **27**, (2020).

110. Walls, A. C., Tortorici, M. A., Bosch, B. J., Frenz, B., Rottier, P. J. M., DiMaio, F., Rey, F. A. & Veesler, D. Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature* **531**, 114–117 (2016).

111. Schoeman, D. & Fielding, B. C. Coronavirus envelope protein: Current knowledge. *Virol. J.* **16**, (2019).

112. Ortego, J., Ceriani, J. E., Patiño, C., Plana, J. & Enjuanes, L. Absence of E protein arrests transmissible gastroenteritis coronavirus maturation in the secretory pathway. *Virology* **368**, 296–308 (2007).

113. Opella, S. J. Relating structure and function of viral membrane-spanning miniproteins. *Curr. Opin. Virol.* **12**, 121–125 (2015).

114. Pervushin, K., Tan, E., Parthasarathy, K., Lin, X., Jiang, F. L., Yu, D., Vararattanavech, A., Tuck, W. S., Ding, X. L. & Torres, J. Structure and inhibition of the SARS coronavirus envelope protein ion channel. *PLoS Pathog.* **5**, (2009).

115. Li, Y., Surya, W., Claudine, S. & Torres, J. Structure of a conserved golgi complex-targeting signal in coronavirus envelope proteins. *J. Biol. Chem.* **289**, 12535–12549 (2014).

116. Opella, S. J. & Marassi, F. M. Applications of NMR to membrane proteins. *Arch. Biochem. Biophys.* **628**, 92–101 (2017).

117. Cross, T. A. & Opella, S. J. Structural properties of fd coat protein in sodium dodecyl sulfate micelles. *Biochem. Biophys. Res. Commun.* **92**, 478–484 (1980).

118. Frey, L., Lakomek, N. A., Riek, R. & Bibow, S. Micelles, Bicelles, and Nanodiscs: Comparing the Impact of Membrane Mimetics on Membrane Protein Backbone Dynamics. *Angew. Chemie - Int. Ed.* **56**, 380–383 (2017).

119. Chipot, C., Dehez, F., Schnell, J. R., Zitzmann, N., Pebay-Peyroula, E., Catoire, L. J., Miroux, B., Kunji, E. R. S., Veglia, G., Cross, T. A. & Schanda, P. Perturbations of Native Membrane Protein Structure in Alkyl Phosphocholine Detergents: A Critical Assessment of NMR and Biophysical Studies. *Chem. Rev.* **118**, 3559–3607 (2018).

120. Opella, S. J. NMR and membrane proteins. *Nat. Struct. Biol.* **4 Suppl**, 845–8 (1997).

121. Lim, K. P., Xu, H. Y. & Liu, D. X. Physical interaction between the membrane (M) and envelope (E) proteins of the coronavirus avian infectious bronchitis virus (IBV). *Adv. Exp.*

Med. Biol. **494**, 595–602 (2001).

122. Corse, E. & Machamer, C. E. The cytoplasmic tails of infectious bronchitis virus E and M proteins mediate their interaction. *Virology* **312**, 25–34 (2003).

123. Wu, Q., Zhang, Y., Lü, H., Wang, J., He, X., Liu, Y., Ye, C., Lin, W., Hu, J., Ji, J., Xu, J., Ye, J., Hu, Y., Chen, W., Li, S., Wang, J., Wang, J., Bi, S. & Yang, H. The E protein is a multifunctional membrane protein of SARS-CoV. *Genomics, proteomics Bioinforma. / Beijing Genomics Inst.* **1**, 131–144 (2003).

124. Toto, A., Ma, S., Malagrinò, F., Visconti, L., Pagano, L., Stromgaard, K. & Gianni, S. 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 (2020).

125. Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., O'Meara, M. J., Rezelj, V. V., Guo, J. Z., Swaney, D. L., Tummino, T. A., Huettenhain, R., Kaake, R. M., Richards, A. L., Tutuncuoglu, B., Foussard, H., Batra, J., Haas, K., Modak, M., Kim, M., Haas, P., Polacco, B. J., Braberg, H., Fabius, J. M., Eckhardt, M., Soucheray, M., Bennett, M. J., Cakir, M., McGregor, M. J., Li, Q., Meyer, B., Roesch, F., Vallet, T., Mac Kain, A., Miorin, L., Moreno, E., Naing, Z. Z. C., Zhou, Y., Peng, S., Shi, Y., Zhang, Z., Shen, W., Kirby, I. T., Melnyk, J. E., Chorba, J. S., Lou, K., Dai, S. A., Barrio-Hernandez, I., Memon, D., Hernandez-Armenta, C., Lyu, J., Mathy, C. J. P., Perica, T., Pilla, K. B., Ganesan, S. J., Saltzberg, D. J., Rakesh, R., Liu, X., Rosenthal, S. B., Calviello, L., Venkataramanan, S., Liboy-Lugo, J., Lin, Y., Huang, X. P., Liu, Y. F., Wankowicz, S. A., Bohn, M., Safari, M., Ugur, F. S., Koh, C., Savar, N. S., Tran, Q. D., Shengjuler, D., Fletcher, S. J., O'Neal, M. C., Cai, Y., Chang, J. C. J., Broadhurst, D. J., Klippsten, S., Sharp, P. P., Wenzell, N. A., Kuzuoglu, D., Wang, H. Y., Trenker, R., Young, J. M., Cavero, D. A., Hiatt, J., Roth, T. L., Rathore, U., Subramanian, A., Noack, J., Hubert, M., Stroud, R. M., Frankel, A. D., Rosenberg, O. S., Verba, K. A., Agard, D. A., Ott, M., Emerman, M., Jura, N., von Zastrow, M., Verdin, E., Ashworth, A., Schwartz, O., D'Enfert, C., Mukherjee, S., Jacobson, M., Malik, H. S., Fujimori, D. G., Ideker, T., Craik, C. S., Floor, S. N., Fraser, J. S., Gross, J. D., Sali, A., Roth, B. L., Ruggero, D., Taunton, J., Kortemme, T., Beltrao, P., Vignuzzi, M., García-Sastre, A., Shokat, K. M., Shoichet, B. K. & Krogan, N. J. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).

126. Wilson, L., Mckinlay, C., Gage, P. & Ewart, G. SARS coronavirus E protein forms cation-selective ion channels. *Virology* **330**, 322–331 (2004).

127. Ma, C., Marassi, F. M., Jones, D. H., Straus, S. K., Bour, S., Strebel, K., Schubert, U., Oblatt-Montal, M., Montal, M. & Opella, S. J. Expression, purification, and activities of full-length and truncated versions of the integral membrane protein Vpu from HIV-1. *Protein Sci.* **11**, 546–557 (2002).

128. Park, S. H., Mrse, A. A., Nevzorov, A. A., Mesleh, M. F., Oblatt-Montal, M., Montal, M. & Opella, S. J. Three-dimensional structure of the channel-forming trans-membrane domain of virus protein 'u' (Vpu) from HIV-1. *J. Mol. Biol.* **333**, 409–424 (2003).

129. Howell, S. C., Mesleh, M. F. & Opella, S. J. NMR structure determination of a membrane protein with two transmembrane helices in micelles: MerF of the bacterial mercury

detoxification system. *Biochemistry* **44**, 5196–5206 (2005).

130. Cook, G. A., Stefer, S. & Opella, S. J. Expression and purification of the membrane protein p7 from hepatitis C virus. *Biopolymers* **96**, 32–40 (2011).

131. Park, S. H., Casagrande, F., Chu, M., Maier, K., Kiefer, H. & Opella, S. J. Optimization of purification and refolding of the human chemokine receptor CXCR1 improves the stability of proteoliposomes for structure determination. *Biochim. Biophys. Acta - Biomembr.* **1818**, 584–591 (2012).

132. Amin, A., Sarwar, A., Saleem, M. A., Latif, Z. & Opella, S. J. Expression purification of transmembrane protein MerE from Mercury-Resistant Bacillus cereus. *J. Microbiol. Biotechnol.* **29**, 274–282 (2019).

133. Kuliopulos, A., Nelson, N. P., Yamada, M., Walsh, C. T., Furie, B., Furie, B. C. & Roth, D. A. Localization of the affinity peptide-substrate inactivator site on recombinant vitamin K-dependent carboxylase. *J. Biol. Chem.* **269**, 21364–70 (1994).

134. Schiksnis, R. A., Bogusky, M. J., Tsang, P. & Opella', S. J. Structure and Dynamics of the Pfl Filamentous Bacteriophage Coat Protein in Micelles. *Biochemistry* **26**, 1373–1381 (1987).

135. McDonnell, P. A., Shon, K., Kim, Y. & Opella, S. J. Fd coat protein structure in membrane environments. *J. Mol. Biol.* **233**, 447–463 (1993).

136. Lee, S., Mesleh, M. F. & Opella, S. J. Structure and dynamics of a membrane protein in micelles from three solution NMR experiments. *J. Biomol. NMR* **26**, 327–334 (2003).

137. Berkamp, S., Park, S. H., De Angelis, A. A., Marassi, F. M. & Opella, S. J. Structure of monomeric Interleukin-8 and its interactions with the N-terminal Binding Site-I of CXCR1 by solution NMR spectroscopy. *J. Biomol. NMR* **69**, 111–121 (2017).

138. Kallick, D. A., Tessmer, M. R., Watts, C. R. & Li, C. Y. The Use of Dodecylphosphocholine Micelles in Solution NMR. *J. Magn. Reson. Ser. B* **109**, 60–65 (1995).

139. Jaremko, Ł., Jaremko, M., Giller, K., Becker, S. & Zweckstetter, M. Conformational Flexibility in the Transmembrane Protein TSPO. *Chem. - A Eur. J.* **21**, 16555–16563 (2015).

140. Parthasarathy, K., Lu, H., Surya, W., Vararattanavech, A., Pervushin, K. & Torres, J. Expression and purification of coronavirus envelope proteins using a modified β-barrel construct. *Protein Expr. Purif.* **85**, 133–141 (2012).

141. Parthasarathy, K., Ng, L., Lin, X., Ding, X. L., Pervushin, K., Gong, X. & Torres, J. Structural flexibility of the pentameric SARS coronavirus envelope protein ion channel. *Biophys. J.* **95**, (2008).

142. Veglia, G., Zeri, A. C., Ma, C. & Opella, S. J. Deuterium/hydrogen exchange factors measured by solution nuclear magnetic resonance spectroscopy as indicators of the structure and topology of membrane proteins. *Biophys. J.* **82**, 2176–2183 (2002).

143. Wishart, D. S. & Sykes, B. D. The 13C Chemical-Shift Index: A simple method for the identification of protein secondary structure using 13C chemical-shift data. *J. Biomol. NMR* **4**, 171–180 (1994).

144. Mesleh, M. F., Veglia, G., DeSilva, T. M., Marassi, F. M. & Opella, S. J. Dipolar waves as NMR maps of protein structure. *J. Am. Chem. Soc.* **124**, 4206–4207 (2002).

145. Mesleh, M. F., Lee, S., Veglia, G., Thiriot, D. S., Marassi, F. M. & Opella, S. J. Dipolar waves map the structure and topology of helices in membrane proteins. *J. Am. Chem. Soc.* **125**, 8928–8935 (2003).

146. Cook, G. A. & Opella, S. J. Secondary structure, dynamics, and architecture of the p7 membrane protein from hepatitis C virus by NMR spectroscopy. *Biochim. Biophys. Acta - Biomembr.* **1808**, 1448–1453 (2011).

147. Mesleh, M. F. & Opella, S. J. Dipolar waves as NMR maps of helices in proteins. *J. Magn. Reson.* **163**, 288–299 (2003).

148. Westerbeck, J. W. & Machamer, C. E. A Coronavirus E Protein Is Present in Two Distinct Pools with Different Effects on Assembly and the Secretory Pathway. *J. Virol.* **89**, 9313–9323 (2015).

149. Choma, C., Gratkowski, H., Lear, J. D. & DeGrado, W. F. Asparagine-mediated self-association of a model transmembrane helix. *Nat. Struct. Biol.* **7**, 161–166 (2000).

150. Torres, J., Maheswari, U., Parthasarathy, K., Ng, L., Liu, D. X. & Gong, X. Conductance and amantadine binding of a pore formed by a lysine-flanked transmembrane domain of SARS coronavirus envelope protein. *Protein Sci.* **16**, 2065–2071 (2007).

151. Verdiá-Báguena, C., Nieto-Torres, J. L., Alcaraz, A., DeDiego, M. L., Torres, J., Aguilella, V. M. & Enjuanes, L. Coronavirus E protein forms ion channels with functionally and structurally-involved membrane lipids. *Virology* **432**, 485–494 (2012).

152. Grzesiek, S., Bax, A., Döbeli, H., Gentz, R., Garotta, G. & Labhardt, A. M. 1H, 13C, and 15N NMR Backbone Assignments and Secondary Structure of Human Interferon-γ. *Biochemistry* **31**, 8180–8190 (1992).

153. Marassi, F. M., Ma, C., Gratkowski, H., Straus, S. K., Strebel, K., Oblatt-Montal, M., Montal, M. & Opella, S. J. Correlation of the structural and functional domains in the membrane protein Vpu from HIV-1. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 14336–14341 (1999).

154. Wang, J., Denny, J., Tian, C., Kim, S., Mo, Y., Kovacs, F., Song, Z., Nishimura, K., Gan, Z., Fu, R., Quine, J. R. & Cross, T. A. Imaging Membrane Protein Helical Wheels. *J. Magn. Reson.* **144**, 162–167 (2000).

155. Marassi, F. M. & Opella, S. J. A Solid-State NMR Index of Helical Membrane Protein Structure and Topology. *J. Magn. Reson.* **144**, 150–155 (2000).

156. Sang, H. P. & Opella, S. J. Tilt angle of a trans-membrane helix is determined by hydrophobic mismatch. *J. Mol. Biol.* **350**, 310–318 (2005).

157. Nishimura, K., Kim, S., Zhang, L. & Cross, T. A. The closed state of a H+ channel helical bundle combining precise orientational and distance restraints from solid state NMR. *Biochemistry* **41**, 13170–13177 (2002).

158. Takano, T., Nakano, K., Doki, T. & Hohdatsu, T. Differential effects of viroporin inhibitors against feline infectious peritonitis virus serotypes I and II. *Arch. Virol.* **160**, 1163–1170 (2015).

159. Ewart, G. D., Mills, K., Cox, G. B. & Gage, P. W. Amiloride derivatives block ion channel activity and enhancement of virus-like particle budding caused by HIV-1 protein Vpu. *Eur. Biophys. J.* **31**, 26–35 (2002).

160. Premkumar, A., Wilson, L., Ewart, G. D. & Gage, P. W. Cation-selective ion channels formed by p7 of hepatitis C virus are blocked by hexamethylene amiloride. *FEBS Lett.* **557**, 99–103 (2004).

161. Cady, S. D., Schmidt-Rohr, K., Wang, J., Soto, C. S., Degrado, W. F. & Hong, M. Structure of the amantadine binding site of influenza M2 proton channels in lipid bilayers. *Nature* **463**, 689–692 (2010).

162. David, P., Mayan, H., Cragoe, E. J. & Karlish, S. J. D. Structure-activity relations of amiloride derivatives, acting as antagonists of cation binding on Na+/K+-ATPase. *BBA - Biomembr.* **1146**, 59–64 (1993).

163. Westerbeck, J. W. & Machamer, C. E. The Infectious Bronchitis Coronavirus Envelope Protein Alters Golgi pH To Protect the Spike Protein and Promote the Release of Infectious Virus. *J. Virol.* **93**, (2019).

164. Gratkowski, H., Lear, J. D. & DeGrado, W. F. Polar side chains drive the association of model transmembrane peptides. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 880–885 (2001).

165. Cross, T. A., DiVerdi, J. A. & Opella, S. J. Strategy for Nitrogen NMR of Biopolymers. *J. Am. Chem. Soc.* **104**, 1759–1761 (1982).

166. Ramjeesingh, M., Huan, L. J., Garami, E. & Bear, C. E. Novel method for evaluation of the oligomeric structure of membrane proteins. *Biochem. J.* **342 (Pt 1**, 119–23 (1999).

167. Ottiger, M., Delaglio, F. & Bax, A. Measurement of J and Dipolar Couplings from Simplified Two-Dimensional NMR Spectra. *J. Magn. Reson.* **131**, 373–378 (1998).

168. Sang, H. P., Woo, S. S., Mukhopadhyay, R., Valafar, H. & Opella, S. J. Phage-induced alignment of membrane proteins enables the measurement and structural analysis of residual dipolar couplings with dipolar waves and λ-maps. *J. Am. Chem. Soc.* **131**, 14140–14141 (2009).

169. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).

170. Johnson, B. A. & Blevins, R. A. NMR View: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603–614 (1994).

171. McBride, C. E., Li, J. & Machamer, C. E. The Cytoplasmic Tail of the Severe Acute Respiratory Syndrome Coronavirus Spike Protein Contains a Novel Endoplasmic Reticulum Retrieval Signal That Binds COPI and Promotes Interaction with Membrane Protein. *J. Virol.* **81**, 2418–2428 (2007).

172. Cohen, J. R., Lin, L. D. & Machamer, C. E. 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 (2011).