UNIVERSITY OF CALIFORNIA SAN DIEGO

Expression and NMR Studies of Coronavirus Envelope and Spike Proteins

A Thesis submitted in partial satisfaction of the requirements for the degree Master of Science

in

Chemistry

by

Haley Siddiqi

Committee in Charge:

Professor Stanley J. Opella, Chair
Professor John Guatelli
Professor Patricia A. Jennings

2021

The thesis of Haley Siddiqi is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

Dedication

To my parents, Sadia and Khalid Siddiqi, my siblings Harris and Sonia Siddiqi and my fiancé, Nick Thakar for their endless love, support, and encouragement.

Acknowledgments

I would like to acknowledge Dr. Stanley Opella for his support and guidance throughout my undergraduate and master's studies. The confidence he has shown in me has had a profound influence on my outlook of my potential as a researcher. I appreciate Dr. Opella's help in broadening my understanding of biochemistry since my third year of undergraduate studies when he first afforded me the opportunity to pursue the high-level research that led me to this project.

I would like to express my sincere gratitude to Dr. Sang Ho Park for his patience in answering my many questions as a junior researcher and his guidance during experimental procedures. His rigor and dedication towards his research inspired me to continue asking myself questions and trying every possible option through weeks of failed experiments. I will always be grateful for his guidance.

I would also like to acknowledge Dr. Anna De Angelis for her suggestions and help in optimizing my experimental design as well as Yui Na and Jiaqian Wu for their patience and guidance in helping me understand and conduct experiments during my first year as a research student.

I would also like to thank my lab mate Daniela Castro in playing an essential role as a friend and collaborator during my research here at UCSD. I am grateful to have had her by my side to tackle experimental challenges, spend early mornings, and late evenings often repeating experiments in lab and learning to maintain the magnificent magnets that we have come to appreciate.

Figures 1, 12, 13, 14, 16 and 17 have been submitted for publication of the material as it may appear in PLOS pathogens, 2021, Park SH, Castro DV, Siddiqi H, De Angelis A, Oom A, Stoneham C, Lewinski M, Clark A, Croker B, Carlin A, Guatelli J, Opella SJ .2021. The thesis author was a co-author of these figures. [24]

TABLE OF CONTENTS

List of Abbreviations

SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2

SARS- CoV/SCV: Severe acute respiratory syndrome coronavirus

MERS-CoV: Middle east respiratory syndrome coronavirus

NMR: Nuclear magnetic resonance

E: Envelope protein

S: Spike protein

M: Membrane protein

N: Nucleocapsid protein

ACE2: Angiotensin-converting enzyme 2

RBD: Receptor-binding domain

MBD: Membrane-binding domain

Cryo-EM: Cryogenic electron microscopy

TMPRSS2: Transmembrane protease serine 2

KSI: Ketosteroid Isomerase

HPC: N-hexadecylphosphocholine OR fos-choline-16

SDS: Sodium dodecyl sulfate

TPC: N-tetradecylphosphocholine or fos-choline-14

DPC: Dodecylphosphocholine

HMA: Hexamethylene amiloride

SDS – PAGE: Sodium dodecyl sulfate polyacrylamide gel electrophoresis

MHV: Mouse hepatitis virus

HCoV-229E: Human coronavirus 229E

H/D: Hydrogen-deuterium

^{15}N: Nitrogen-15

^{1}H: Proton

Val: Valine

Leu: Leucine

CSP: Chemical shift perturbations

HSQC: Heteronuclear quantum coherence

CD: Circular Dichroism Spectroscopy

IPTG: Isopropyl beta-D-1-thiogalactopyranoside

Lys: Lysine

Asn: Asparagine

Glu: Glutamic acid

Gln: Glutamine

Ser: Serine

Arg: Arginine

Gly: Glycine

Phe: Phenylalanine

Thr: Threonine

EF: Full-length E protein

ET: N-terminal transmembrane domain of E protein

EC: C-terminal domain of E protein

S CTD: C-terminal domain of S protein

Cys: Cysteine

Ni-NTA: Nickel-nitrilotriacetic acid (affinity chromatography)

CMC: Critical micelle concentration

kDa: kilo Dalton

GST: Glutathione-S-transferase

N15A: Asparagine15 → Alanine

V25F: Valine25 → Phenylalanine

R102E: Arginine102 → Glutamic acid

DTT: Dithiothreitol

TCEP: Tris(2-carboxyethyl) phosphine

ER: Endoplasmic reticulum

ERGIC: Endoplasmic reticulum and Golgi intermediate complex

His: Histidine

E. coli: Escherichia coli

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

U: International Units of Thrombin

EDTA: Ethylenediamene tetracetic acid

LC-MS: Liquid chromatography – mass spectrometry

NaCl: Sodium Chloride

NaF: Sodium Fluoride

FIPV: Feline infectious peritonitis -inducing virus

LB: Luria-Bertani

TCEP-HCl: Tris(2-carboxyethyl) phosphine hydrochloride

Tris- HCl: Tris(hydroxymethyl)aminomethane hydrochloride

MES: 2-(N-morpholino) ethane sulfonic acid

DSS: Sodium trimethylsilylpropanesulfonate

List of Figures

ABSTRACT OF THE THESIS

Expression and NMR Studies of Coronavirus Envelope and Spike Proteins

by

Haley Siddiqi

Master of Science in Chemistry

University of California San Diego, 2021

Professor Stanley J. Opella, Chair

The spread of novel coronavirus, SARS-CoV-2, a lethal and infectious respiratory syndrome, has heightened urgency in antiviral drug development. The SARS-CoV-2 envelope (E) and spike (S) proteins play a significant role in viral lifecycle and pathogenicity. The immediate threat posed by this infectious disease necessitates structural and functional characterization of these key structural membrane proteins involved in viral infection and replication. E protein's transmembrane domain exhibits ion channel activity while its C-terminal domain is thought to participate in protein-protein interactions. S protein initiates viral infection through host-cell attachment and is proposed to participate in intermolecular interactions with E. Here, a bacterial expression and purification system is introduced to allow successful

expression and purification of membrane proteins through *E. coli.* Use of this system accelerates production of purified full-length E and S proteins for spectroscopic NMR studies. Purification optimization allowed for high-resolution NMR spectra leading to characterization of the full-length E protein secondary structure and dynamics through assessment of CSPs. Interactions between E and known ion-channel inhibitor hexamethylene amiloride (HMA) allowed for characterization of the drug binding site. Analysis of the N-terminal transmembrane domain indicated Asn15 crucial to preserving E protein's conformation. To evaluate potential intermolecular interactions between the C-terminal domains of E and S proteins, a truncated S protein construct was developed. Purification and expression of the S protein construct led to an initial $^{1}H/^{15}N$ HSQC spectrum, paving the way for future studies involving E and S protein-protein interactions integral to their structure and function in the pathogenicity of SARS-CoV-2.

Introduction

On March 11, 2020, the World Health Organization declared the infectious severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) a global pandemic [1]. The spread of this enveloped single-strand RNA virus affected communities across all six inhabited continents and contributed to over 3 million deaths [2,33]. Beginning with the SARS-CoV outbreak in 2002 and later with the 2012 outbreak of MERS-CoV coronaviruses have received increasing attention; the novel coronavirus SARS- CoV-2 relates to its predecessors as a member of the *Coronaviridae* family and betacoronavirus genus [1,2]. Due to the nature of its pathogenicity and transmission the development of therapeutics and vaccines against SARS-CoV-2 is imperative and largely relies on structural and dynamic insight of the viral proteins as well as their roles in viral infection and replication. Here, the optimization of a novel bacterial expression and purification system enabled the full-length, wild-type membrane protein to be prepared directly for NMR studies. Using purified protein obtained from this novel system, solution-state NMR spectroscopy was applied to the envelope (E) and spike (S) proteins, two of the four structural membrane proteins of SARS-CoV-2. With the focus on characterizing its drug binding site and assessing the effects of single-site mutations, interactions between various E constructs and hexamethylene amiloride were also evaluated through $^{1}H/^{15}N$ HSQC experiments. Furthermore, assessment of potential protein-protein interactions between E and S proteins, was initiated with S protein construct design and an initial solution NMR spectrum.

The large RNA genome of SARS-CoV-2 spans 30 kb encoding for 29 proteins [1]. Four structural proteins found to be most abundant in the virus: membrane (M), envelope (E), nucleocapsid (N), and spike (S) (Figure 1). M, E and S are integral membrane proteins embedded in the lipid bilayer of the viral envelope [1]. The integral membrane proteins are each found to be oligomeric under varying experimental conditions. M is found to be a dimer or dimer of dimers; E is thought to be a pentamer and S is a trimer [3-6].

The viral envelope surface has several trimeric glycoprotein spikes, aiding in evasion of immune detection and initiation of host cell receptor attachment [7]. The attachment of S protein to the host angiotensin-converting enzyme 2 (ACE2) initiates coronavirus infection. Following fusion to the endosomal membrane of the host cell, genomic RNA is released and replicated immediately. During its intracellular lifecycle, structural and accessory proteins are encoded and eventually secreted as virions through exocytosis [8] (Figure 2). S protein plays a major role in membrane fusion and viral entry of the host cell, making it a potential target for vaccine and therapeutic development. During envelope formation, the M and S proteins are thought to form homotypic interactions in the ER-Golgi intermediate compartment (ERGIC) allowing for the incorporation of S to the new virion [9].

During the virus replication cycle, E protein is abundantly expressed and important for viral assembly, release, and pathogenesis [7]. During viral replication, E has mostly been found at the site of intracellular trafficking through the ER, Golgi and ERGIC where it aids in the assembly and buddying of the new virion [9]. Weakened viral maturation and reduced propagation has been seen in recombinant coronaviruses lacking E protein [9]. The hydrophobic 75 residue amino acid sequence of SARS-CoV-2 E protein shares similarity with the E protein sequence of SARS-CoV [9, 32]. Due to the hydrophobic nature of E protein as a viral membrane miniprotein [10] its function as a viroporin is under investigation. The E protein has displayed ion-channel activity yet does not display sequence homology with known viroporins [4,5].

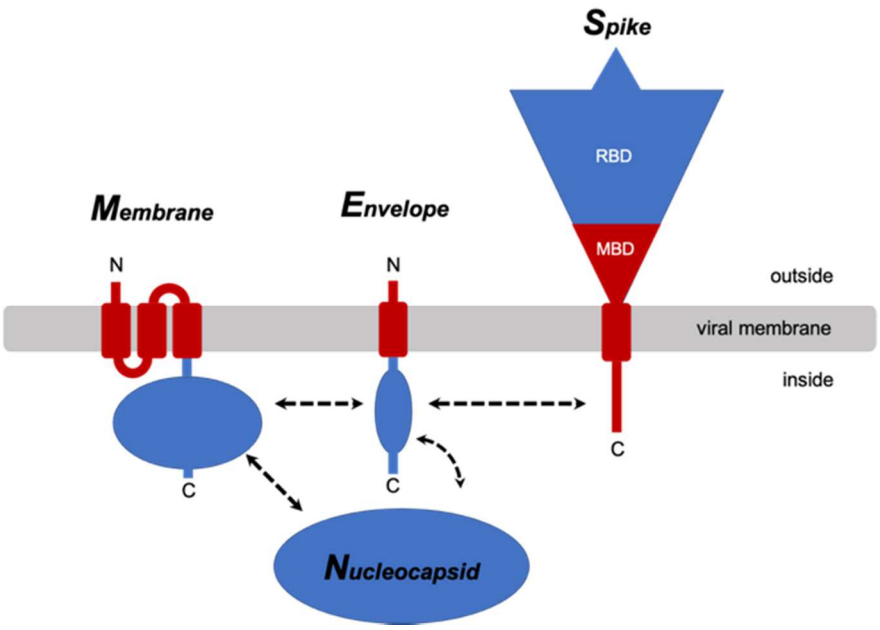

Figure 1. Pictorial representation of SARS-CoV-2 structural proteins. Structural membrane proteins are labeled membrane (M) protein, envelope (E) protein, and spike (S) protein. Shown in red are the membrane-associated regions of each protein while in blue the extra- and intra- cellular regions can be seen. Proposed intraviral protein-protein interactions are shown through dashed arrows. RBD refers to receptor-binding domain; MBD refers to membrane-binding domain. Nucleocapsid (N) is the fourth structural protein [24].

The E protein plays an essential role in viral replication and maturation as previously described [7] therefore it is an appealing target for therapeutic and antiviral drug development. To effectively design anti-viral drugs high-resolution structural data of the protein receptor in bound and free states are necessary. As a small membrane protein, E protein neither be an effective candidate for cryo-EM or X-ray crystallography as crystallization of native E protein in in liquid crystalline membrane bilayers is likely to be difficult. Therefore, considering the experimental methods and conditions necessary in developing a membrane-like environment to study the structure of E protein, NMR spectroscopy is the best suited approach[11]. When studying membrane proteins by NMR, caution is necessary in using micelle environments [12] due to the possibility of protein aggregation and structural distortions [13,14]. Through optimization of sample conditions, solution-state NMR has allowed for valuable data collection of structural membrane proteins. To ensure samples were integrated into an amphipathic membrane-like environment it

is essential to prepare membrane protein samples in micelles and obtain high-resolution NMR spectra. The resulting spectra give insight in the chemical purity of the sample and whether it is misfolded or aggregated before initiating the solid-state NMR studies in phospholipid bilayers. Solid-state NMR experiments require protein samples embedded into liquid crystalline, hydrated phospholipid bilayers at high lipid to protein ratios in order to ensure that experiments are performed under near-native conditions. Characterization of the protein and bilayers is necessary to ensure the protein is stable and in a biologically active conformation.

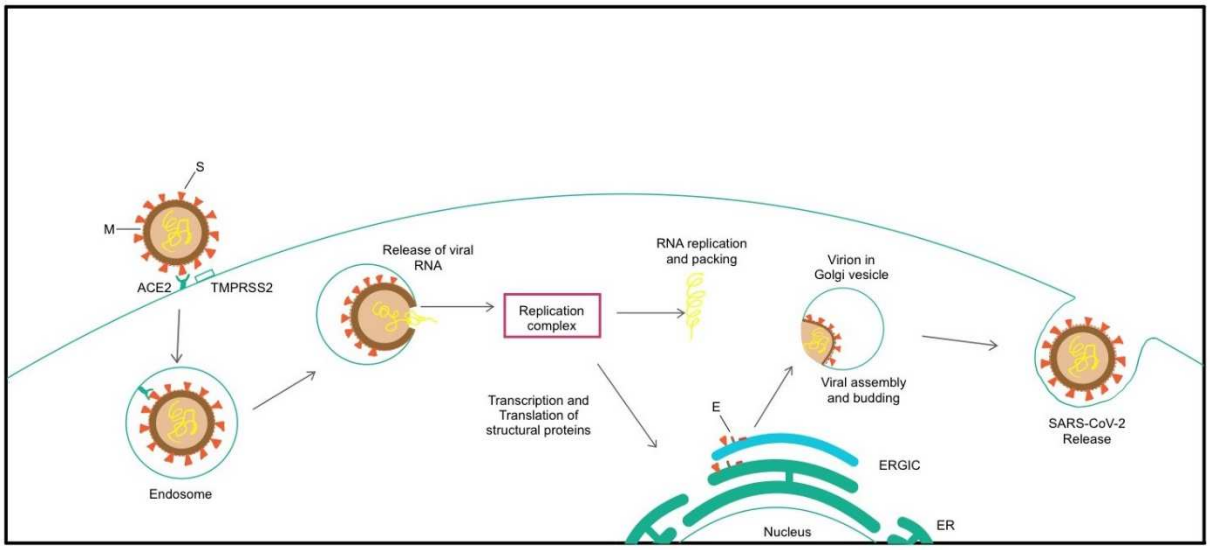

Figure 2. Cartoon of SARS-CoV-2 Infection and Lifecycle. The spike (S), membrane (M) and envelope (E) protein is labeled accordingly in the cartoon. During initial infection SARS-CoV-2 spike protein binds to the host ACE2 receptor while host factor TMPRSS2 promotes viral fusion into the cell. Following entry, viral RNA is released and uncoated for translation through several steps, a replication complex is formed where RNA and structural and accessory proteins are transcribed and translated. During viral assembly and budding, interactions between M and S in the ER-Golgi intermediate compartment (ERGIC) allow for the incorporation of the S protein to the virion envelope. E protein seen in the ERGIC aids in the final assembly and budding of the virion [8,9].

Previous NMR structural studies of coronavirus E protein have utilized short, truncated polypeptide sequences encompassing a portion of the N-terminal domain containing the transmembrane helix responsible for homo-oligomerization, ion channel formation, as well as the residues involved in drug binding [4,5,15,16]. However, currently no structural data have been published on the full-length coronavirus E protein. Therefore, the development of a reliable

expression and purification system is crucial in allowing for the advancement of full-length E protein studies as well as those of other membrane proteins. The three cysteine residues present in the C-terminal domain of the E protein are thought to be involved in intermolecular interactions with S protein, however no studies have yet investigated these potential interactions. Here a construct of the S protein containing all ten-cysteine residues intact was expressed and purified to initiate interaction studies with the C-terminal domain of E protein. The dynamics and structure of full-length E protein from SARS-CoV-2 was characterized in HPC micelles, enabling for a comprehensive study of the complete drug binding site as well as the native N- and C-terminus which are essential for structure-based drug discovery.

The E protein transmembrane helix is highly hydrophobic and considered to be responsible for homo-oligomerization and ion-channel activity [15,16]. Such channel activity has been thought to be involved in viral replicaiton [17]. Meanwhile the C-terminal E protein residues, DLLV, have been identified as a PDZ-binding motif which has been found to interact with the human tight junction-associated PALS1 protein [21]. The C-terminal domain of the E protein also consists of a triple cysteine motif which has been proposed to interact with the cysteine-rich C-terminal domain of the S protein through intramolecular disulfide linkages [18]. Additionally, the C-terminal domain of the E protein which resembles the bromodomain binding site of histone H3 which has been found to interact with bromodomains 2 and 4, involved in regulation of gene transcription, through acetylation of Lys63 [19]. These findings only further justify the significance of studying the coronavirus E protein as an attractive target for drug discovery.

As a channel blocker, hexamethylene amiloride (HMA), inhibits the conductance of the ion channel activity in E protein from HCoV-229E as well as MHV [20]. HMA has also been seen to interact with synthetic peptide sequences corresponding to sections of coronavirus E protein, confirming the presence of channel conductance activity, at least under some conditions, of both synthetic and expressed polypeptides containing the transmembrane region of E protein [15,16]. Prior studies have shown that there are interactions between HMA and E protein from

SARS-CoV however the NMR chemical shift perturbations of residues in the HMA binding site exhibited high variation among the different E protein constructs and sample conditions [4,5,15,16]. Here we introduce a novel bacterial expression and purification system which allows for membrane protein expression and purification in *E. coli*. The resulting purified protein samples are then used in solution-state NMR spectroscopy to study the full-length constructs of E and S proteins from SARS-CoV-2 in HPC micelles, to explore the interactions between HMA and wild-type and mutant E protein constructs and generate an initial spectrum of the S protein C-terminal domain.

Results

Chapter 1: SARS-CoV-2 Recombinant E Protein Expression and Purification

1.1 E protein Expression and Construct Design

To characterize the structure of the SARS-CoV-2 E protein, it was essential to design and implement an expression system designed for efficient protein expression and purification. Each polypeptide sequence and construct was based on the wild-type full-length E protein from SARS-CoV-2 isolate Wuhan-Hu-1 (NC_045512). A codon optimized gene for the complete amino acid sequence of the E protein was synthesized to improve protein expression in *E. coli*. This gene was inserted into a modified pET-31(b)+ vector to be expressed as a ketosteroid isomerase (KSI) fusion protein. The pET-31(b)+ vector was modified to allow for enzymatic cleavage, rather than the chemical cleavage system that was originally integrated into the vector. Following the KSI sequence, a 24-residue linker sequence was inserted which included a 10 His-tag and a 6-residue thrombin cleavage site (LVPRGS) (Figure 3A). The thrombin cleavage site between Arg and Gly residues resulted in a final post-cleavage product containing two extra N-terminal residues (Gly, Ser) in addition to the wild-type E protein sequence.

The three constructs designed to aid in the structural and dynamic investigation of the E protein included: full-length, wild-type E protein (EF) spanning residues 1-75, the N-terminal transmembrane domain of E protein (ET) spanning residues 1-39, and the C-terminal cytoplasmic domain of E protein spanning residues 36-75 (Figure 3B). The plasmid vectors containing the recombinant E protein constructs were transformed into C43(DE3) competent *E. coli* cells with expression induced through the addition of 1mM isopropyl beta-D-1-thiogalactopyranoside (IPTG), recombinant protein expression was directed to inclusion bodies (Figure 4A). The inclusion body pellets were then solubilized to prepare for fusion protein purification.

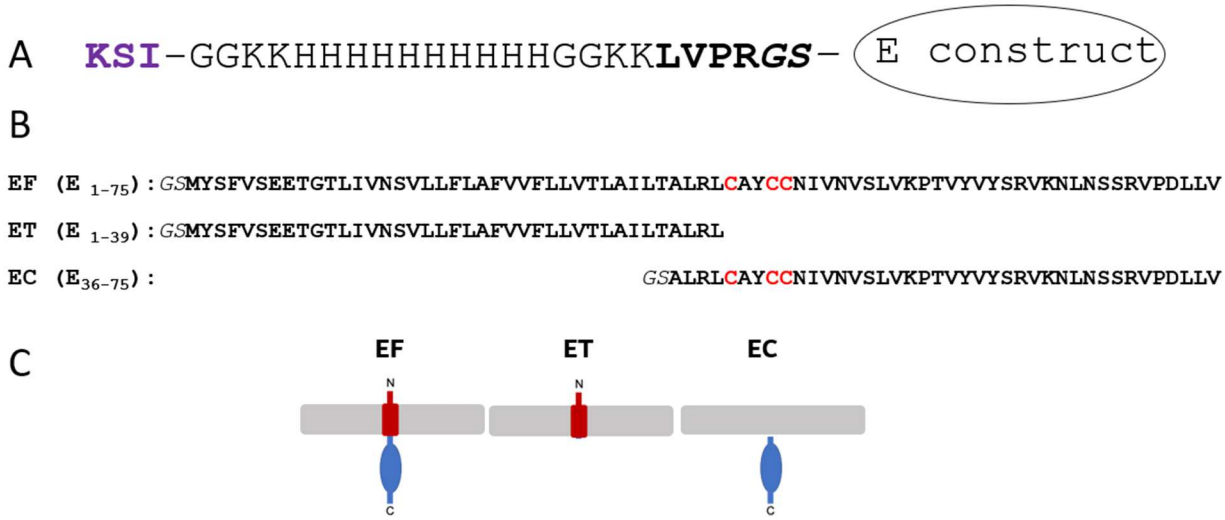

Figure 3. E protein Construct Design. The construct design is seen with the ketosteroid isomerase (KSI) sequence, followed by a 10 – his tag linker sequence and the thrombin cleavage site in bold (A) This construct design was used for all three E protein sequences. Each E protein construct post-enzymatic cleavage is shown with cysteine residues in red (B) with the GlySer residues in italics followed by the 75-residue full-length sequence (EF), the 40 residue trans-membrane sequence (ET) and the 39 residue C-terminal cytosolic sequence (EC). A cartoon representation of each E protein construct is shown with the transmembrane region in red and the cytosolic region in blue (C)

1.2 E Protein Purification Scheme and Optimization

Solubilization of the protein-containing inclusion bodies was performed through the addition of n-hexadecyl phosphocholine (HPC; fos-choline-16). Detergents such as sodium-dodecyl sulfate (SDS), dodecyl phosphocholine (DPC) and n-tetradecyl phosphocholine (TPC; fos-choline-14) were tested, however ultimately HPC was found to be the most efficient in thoroughly solubilizing E protein along with other hydrophobic membrane proteins found in the inclusion bodies. Although SDS was found to be effective in solubilizing E protein, SDS is a harsh detergent and would require detergent exchange for NMR sample preparation. In contrast, DPC and TPC did not solubilize the inclusion bodies thoroughly and would require detergent exchange for NMR sample preparation. HPC being a mild detergent with a relatively low critical micelle concentration (CMC) of 13 μM made it an excellent detergent for E protein purification and NMR sample preparation. HPC could be utilized throughout all steps of purification because it does not interfere with thrombin cleavage specificity (Figure 5B).

The HPC concentration utilized for Ni-affinity column chromatography was optimized through testing various low HPC concentrations (0.1%, 0.05% and 0.025% w/v as well as no HPC) for each E protein construct depicted in Figure 3B. Reviewing the first Ni-NTA purification results with SDS-PAGE, I did not find a significant difference in purity or protein concentration in the fusion protein eluate when varying HPC concentrations and the eluate without HPC (Figure 4), however after enzymatic cleavage and during the second Ni-NTA purification I found that with lower detergent concentration there was a significant decrease in overall purified protein yield, suggesting the possibility of on-column aggregation. Therefore, detergent-free purification was not optimal for high protein yield and purity, to minimize the detergent concentration while maximizing the protein yield all E protein purification steps were performed under 0.05% HPC.

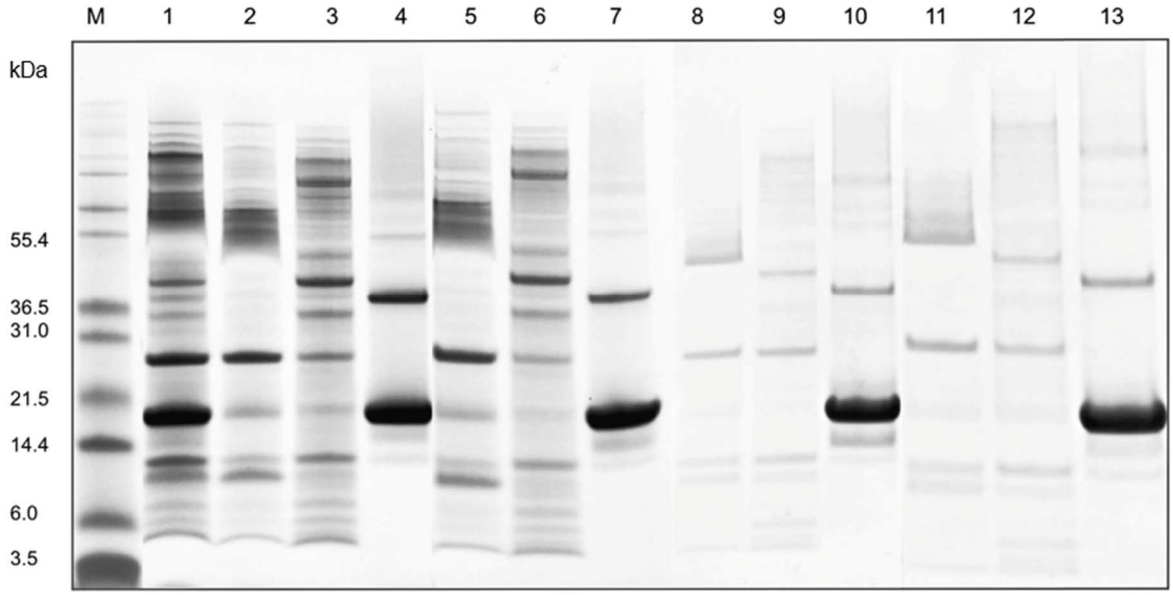

Figure 4. HPC concentration test for 1st Ni- Column Chromatography of EC. SDS-PAGE depicting first Ni-NTA column chromatography purification of EC using different HPC concentrations. Lanes 1-4 (0.1% HPC), Lanes 5-7 (0.05% HPC), Lanes 8-10 (0.025% HPC), Lanes 11-13 (No HPC). Lane M: marker, lane 1: HPC solubilized inclusion bodies containing KSI-EC fusion protein, lanes 2, 5, 8, 11: Ni-affinity column flowthrough, lanes 3, 6, 9, 12: Ni-affinity column wash with imidazole, lanes 4, 7, 10, 13: eluate of KSI-EC fusion protein

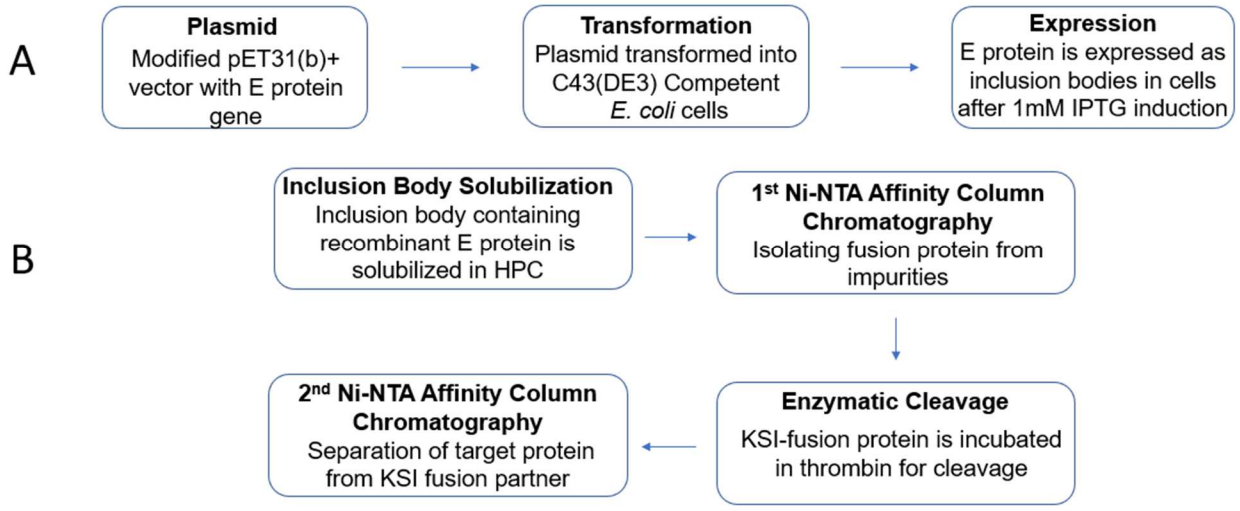

Figure 5. E protein expression and purification scheme. The plasmid vector containing recombinant E protein was transformed into *E. coli* competent cells then expressed through 1mM IPTG induction (A). Upon inclusion body preparation, the inclusion bodies were solubilized in HPC then prepared for Ni-NTA column chromatography where the KSI-fusion protein was isolated from most impurities, then enzymatically cleaved via thrombin. The purified target protein was then collected through second for Ni-NTA column chromatography (B).

Purification through Ni-NTA column chromatography requires a competitive ligand such as imidazole to elute the protein of interest as well as high salt concentrations to reduce non-specific protein binding. In order to improve the sample purity of the KSI-fusion protein eluate, an imidazole optimization test was performed. During purification, wash buffers containing different concentrations of imidazole (20 mM- 80mM) were compared. Although the higher imidazole concentrations improved purity of the fusion-protein eluate, there was an increase in protein loss (up to 50% fusion- protein loss) with the higher concentration imidazole buffers (60-80mM) in the wash fraction seen through SDS-PAGE. To minimize loss while washing away any non-specific bound proteins, higher washing volumes were used at a lower imidazole concentration (20mM), washing the column until absorbance at 280nm = 0.000 to ensure removal of impurities (Figure 6, lane 5).

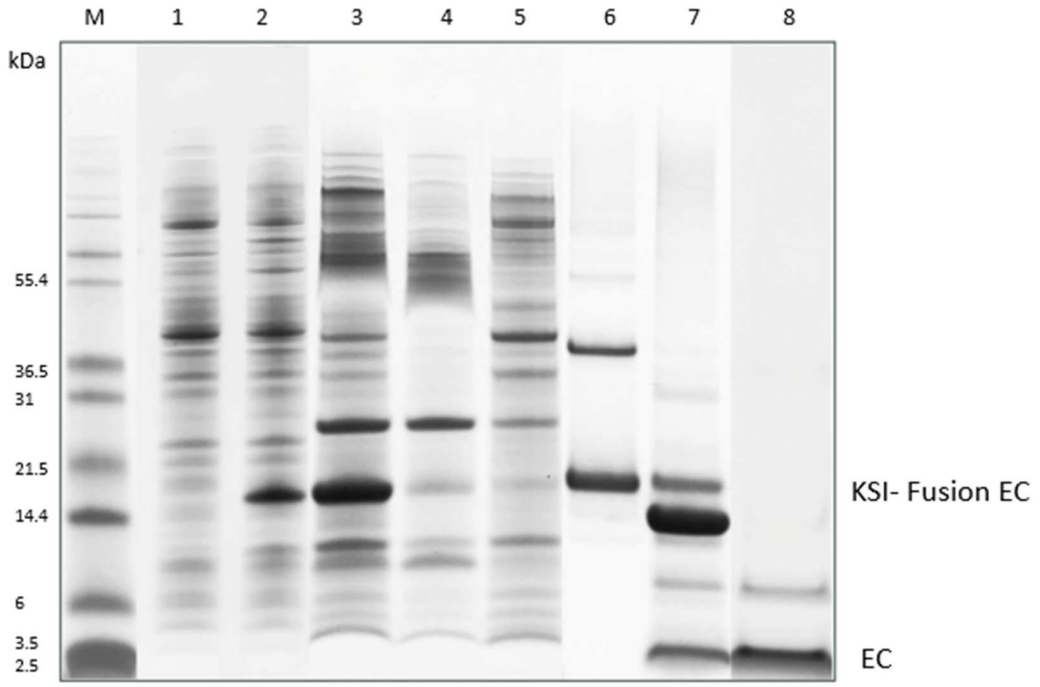

Figure 6. SDS-PAGE of C-terminal domain E protein (EC) expression and purification. Lane M: Mark 12 marker, lane 1: pre-induction cells, lane 2: post-induction cells (KSI-fusion EC is seen ~21.1 kDa), lane 3: HPC-solubilized inclusion bodies containing KSI-EC fusion protein, lane 4: Ni-affinity column flow through, lane 5: Ni-affinity column wash with imidazole, lane 6: eluate of the KSI-EC fusion protein, lane 7: post-enzymatic cleavage of the KSI-EC fusion protein, lane 8: purified EC (~4.6 kDa) with some impurities.

Upon the optimization of the first Ni-NTA purification scheme, the concentration of thrombin was optimized for improved cleavage efficiency. The full-length E protein construct (EF), N-terminal transmembrane construct (ET) and the C-terminal E protein construct (EC) was optimized for thrombin cleavage. Several thrombin concentration tests were performed to determine the optimal concentration for high cleavage efficiency. Concentrations tested for EC included 5U/mg fusion protein, 10U/mg, 20U/mg, and 40U/mg incubating the fusion protein sample overnight (Figure 7). Through optimizing thrombin cleavage for both the full-length E protein (EF) construct as well as the C-terminal cytosolic E protein (EC) construct, I found that with higher thrombin concentrations, there was an increase in low molecular weight impurities. Using thrombin at a concentration of 5U/mg of EC fusion protein provided near complete cleavage as well as minimal impurities. However, when running the fusion protein eluate through second Ni-NTA chromatography, the lower molecular weight impurities that appeared post-enzymatic cleavage were difficult to separate from the target protein, EC (Figure 6). These same impurities were found with the full-length E protein, EF post-enzymatic cleavage and second Ni-NTA chromatography (Figure 8). Since thrombin tends to cleave at Arg, these impurities present in through SDS-PAGE were thought to originate from over digestion of the KSI fusion partner (Figure 9A). Analyzing the KSI fusion partner sequence, two Arg residues, Arg72 and Arg102 are present, which following thrombin cleavage results in four possible short peptide fragments (Figure 10B). The impurities detected through SDS-PAGE were approximately 11.0 kDa and 7.7 kDa (Figure 6). As these short peptide fragments lack a His-tag, they are unable to bind to the column, resulting in decreased purity of the target protein in the eluate. To determine the identity of the small molecular weight impurities that remained after the second Ni-NTA purification an excision of an impurity band from the gel was taken and sent for liquid chromatography-mass spectrometry (LC-MS) analysis, which enabled an efficient assessment of the cleavage site of the impurities that likely resulted from KSI (Figure 6B).

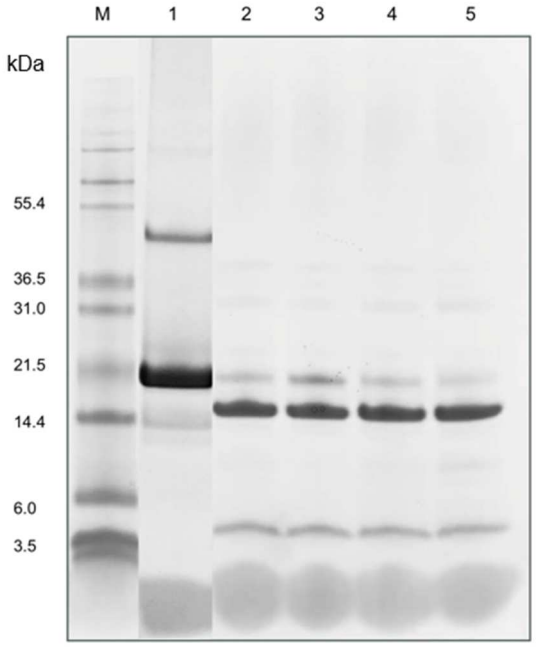

Figure 7. SDS-PAGE of Thrombin Optimization of C-terminal cytosolic E protein (EC). Lane M: mark 12 marker, lane 1: KSI-EC fusion protein pre-enzymatic cleavage, lane 2: KSI-EC post cleavage with 5U/mg thrombin, lane 3: KSI-EC post cleavage with 10U/mg thrombin, lane 4: KSI-EC post cleavage with 20U/mg thrombin, lane 5: KSI-EC post cleavage with 40U/mg thrombin.

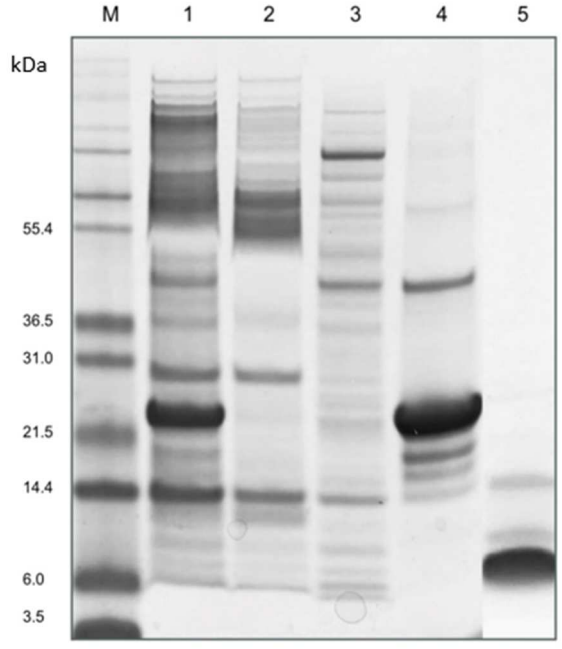

Figure 8. SDS-PAGE of first Ni-NTA column chromatography purification of wild-type full-length E protein purified in 0.05% HPC. Lane M, mark 12 marker; lane 1, HPC-solubilized inclusion bodies containing the KSI-EF fusion protein; lane 2, Ni-affinity column flow through; lane 3, Ni-affinity column wash with imidazole; lane 4, eluate of the KSI-EF fusion protein (~25 kDa) from the column; lane 5, purified EF (~8.5 kDa) with some impurities.

Results of the LC-MS analysis indicated that the fragments were from the KSI fusion partner. The top hits are shown in Figure 10A, with the corresponding KSI cleavage site. Since the most intense impurity band present in the gel was ~11.0 kDa, a mutation was introduced to the KSI sequence at Arg102. The KSI sequence originated from *Comamonas testosteroni* therefore to account for any conserved residues, a sequence alignment was performed against KSI sequences from various other species. Ultimately, Arg102 was mutated to Gly and the mutated KSI R102E EC mutant construct was inserted into the same pET-31(b)+ plasmid vector for *E. coli* expression (Figure 9). Growth and expression of the recombinant C-terminal E protein was not affected by this mutation and the first Ni-NTA column purification results were consistent with previous purification data. However, the post-enzymatic cleavage results of the KSI R102E mutant produced the same impurity band previously seen with the original C-terminal E protein construct. Therefore, these results suggest that the modified pET-31(b)+ vector was not well suited for enzymatic thrombin cleavage.

A KSI-EC Fusion Sequence

MHTPEHITAVVQRFVAALNAGDLDGIVALFADDATVEDPVGSEPRSGTAAIREFYANSL
KLPLAVELTQEVRAVANEAAFAFTVSFEYQGRKTVVAPIDHFRFNGAGKVVSIRALFGE
KNIHACQMLDPGGKKHHHHHHHHHHGGKKLVPRGSALRLCAYCCNIVNVSLVKPTVYVY
SRVKNLNSSRVPDLLV

B KSI-EC Mutant Sequence

MHTPEHITAVVQRFVAALNAGDLDGIVALFADDATVEDPVGSEPRSGTAAIREFYANSL
KLPLAVELTQEVRAVANEAAFAFTVSFEYQGRKTVVAPIDHFEFNGAGKVVSIRALFGE
KNIHACQMLDPGGKKHHHHHHHHHHGGKKLVPRGSALRLCAYCCNIVNVSLVKPTVYVY
SRVKNLNSSRVPDLLV

Figure 9. Sequence of wild-type and mutant KSI-EC fusion protein. The wild-type KSI-EC fusion protein sequence (A) and the mutant KSI-EC fusion protein sequence (B). The sequence of the KSI fusion partner is in bold, followed by the 24-residue linker sequence and then the C-terminal domain E protein sequence (residues 36-75). Residue in green is Arg72, a potential thrombin cleavage site in the KSI sequence. Residue in red of the wild-type KSI-EC fusion sequence is Arg102 which is mutated to Glu seen in the mutant KSI-EC fusion sequence.

A

Top Hits	P-Value	Cleavage Site
LPLAVELTQEVR	>10	QEVR///AV
KTVVAPIDHFR	>10	DHFR///FN

B

Cleavage at Arginine 72 (QEVR///AV):

MHTPEHITAVVQRFVAALNAGDLDGIVALFADDATVEDPVGSEPRSGTAAIREFYANSLKLPLAVELT**QEVR**
7.7 kDa

AVANEAAFAFTVSFEYQGRKTVVAPIDHFRFNGAGKVVSIRALFGEKNIHACQMLDPGGKKHHHHHHHHHHGGK
8.8 kDa

Cleavage at Arginine 102 (DHFR///FN):

MHTPEHITAVVQRFVAALNAGDLDGIVALFADDATVEDPVGSEPRSGTAAIREFYANSLKLPLAVELTQEVRAVA
NEAAFAFTVSFEYQGRKTVVAPI**DHFR**
11.0 kDa

FNGAGKVVSIRALFGEKNIHACQMLDPGGKKHHHHHHHHHHGGKKLVPR
5.5 kDa

Figure 10. Top hits from LC-MS trypsin digest. The top hits indicate the peptide sequences that are most likely the identity of the cleaved impurity. P-score >10 refers to high probability of sequence identity and cleavage site corresponds to the residues where the KSI fusion partner was likely cleaved for each corresponding peptide (A) Cleavage at Arg72 results in two possible peptide fragments with molecular weights of 7.7 kDa and 8.8 kDa. Cleavage at Arg102 results in two possible peptide fragments with molecular weights of 11.0 kDa and 5.5 kDa (B)

Chapter 2: Circular Dichroism Spectroscopy and Solution-State NMR Experiments

2.1 Circular Dichroism Experiments of E protein Constructs

Qualitative analysis of the secondary structure of each E protein construct was determined through circular dichroism (CD) spectroscopy. What makes CD spectroscopy unique is its ability to measure the difference in absorbance of right- and left- circularly polarized light by a molecule, as opposed to the common isotropic light absorbance seen in other forms of spectroscopy [23]. Analysis of CD spectra tends to fall between 180-260nm to characterize secondary structural features such as alpha helices or parallel and antiparallel beta sheet [23]. Therefore, CD spectroscopy can allow for a qualitative view of the alpha helix content of each E protein construct ultimately providing a sense of the secondary structure to give clarity in prospective solid-state NMR studies. Here CD spectroscopy was used to compliment structural data obtained through solution-state NMR spectroscopy. To reduce high background signal and noise prevalent in 190-200nm, CD wavelength scan experiments were measured in the range of 200 – 260nm for EF, ET and EC (Figure 7). The protein concentration and buffer conditions were kept constant for each construct, with 50 uM protein concentration, 20 mM HEPES, 50 mM NaCl at pH 6.5. To keep sample conditions consistent with NMR experiments, NaCl was used during CD sample preparation rather than the common substitute NaF. Therefore, to account for the high absorbance of chloride ions in the spectra, the wavelength range was 200-260nm.

The results of the CD wavelength scan experiments are shown in Figure 11, where EF, ET and EC have two local minima at 210 and 220nm, respectively. As expected, the CD spectra recorded a characteristic alpha helix pattern, for all three constructs (Figure 11). These results are indicative of the highly hydrophobic residues found in the transmembrane core, and the

distinct alpha helix pattern seen for EC is representative of the residues that compose the cytosolic region of the E protein.

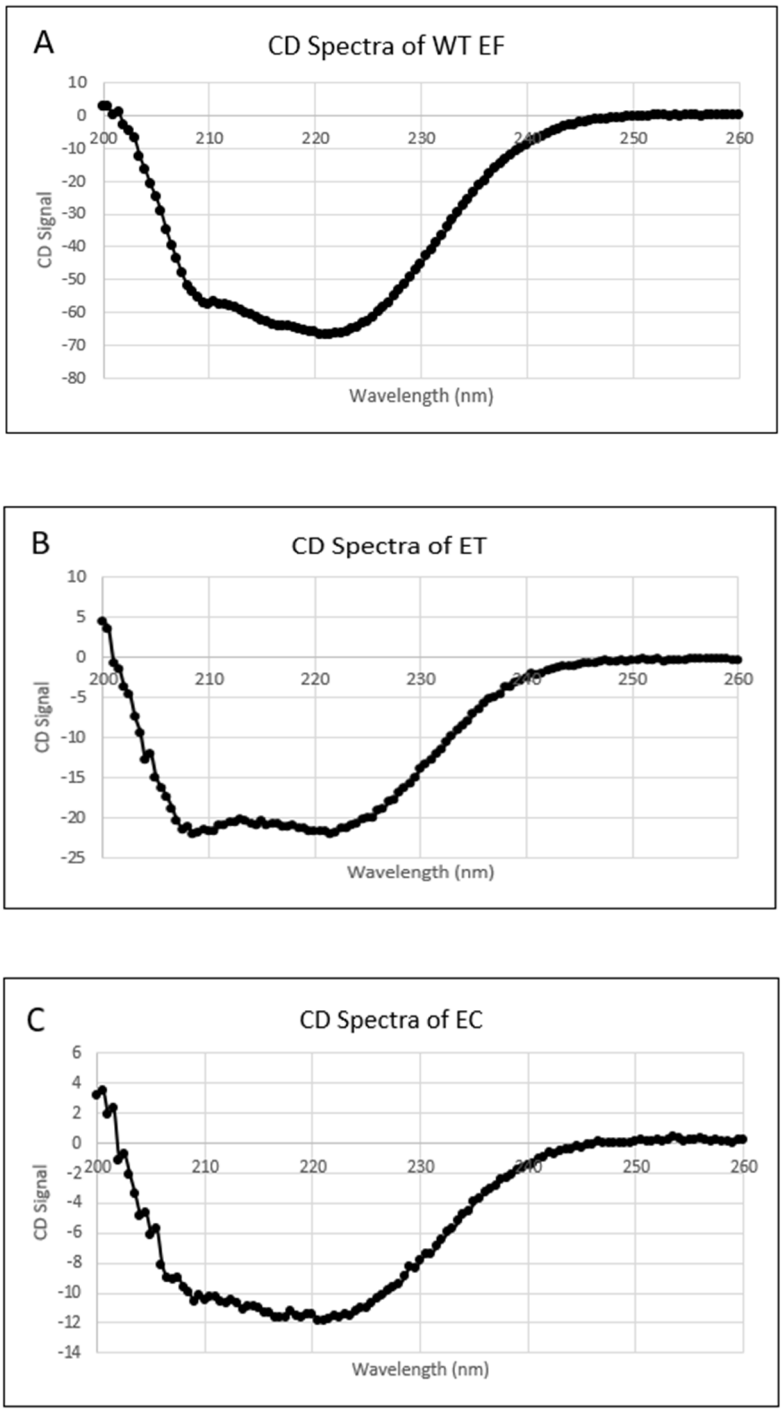

Figure 11. CD spectra of E protein constructs. Below the CD spectra is recorded for EF (A) ET (B) and EC (C) in the range of 200-260nm, respectively.

2.2 E protein Backbone Assignment and Domain Conservation through Solution NMR

To further characterize the transmembrane domain of the E protein, $^{1}H/^{15}N$ heteronuclear single quantum coherence (HSQC) spectra were acquired for each uniformly ^{15}N-labeled E protein construct in HPC micelles (Figure 13). Each protein construct was purified through Ni-NTA column chromatography and directly prepared for NMR experiments. Selectively ^{15}N-Val and ^{15}N-Leu labeled full length EF samples were prepared to aid in peak assignment of the full-length wild-type E protein (residues 1-75). Analysis of selectively labeled spectra revealed the expected number of signals with no sign of doublings or other spectroscopic ambiguities (Figure 12A and B). Therefore, these results validate the chemical purity of the full-length E protein sample and verifies the conformational homogeneity of the samples in HPC micelles. The amide chemical shift frequencies span < 2 ppm, yet because of the relatively narrow linewidths the resulting spectra were well-resolved and consistent with previous studies assessing transmembrane helical conformations [4,15,16].

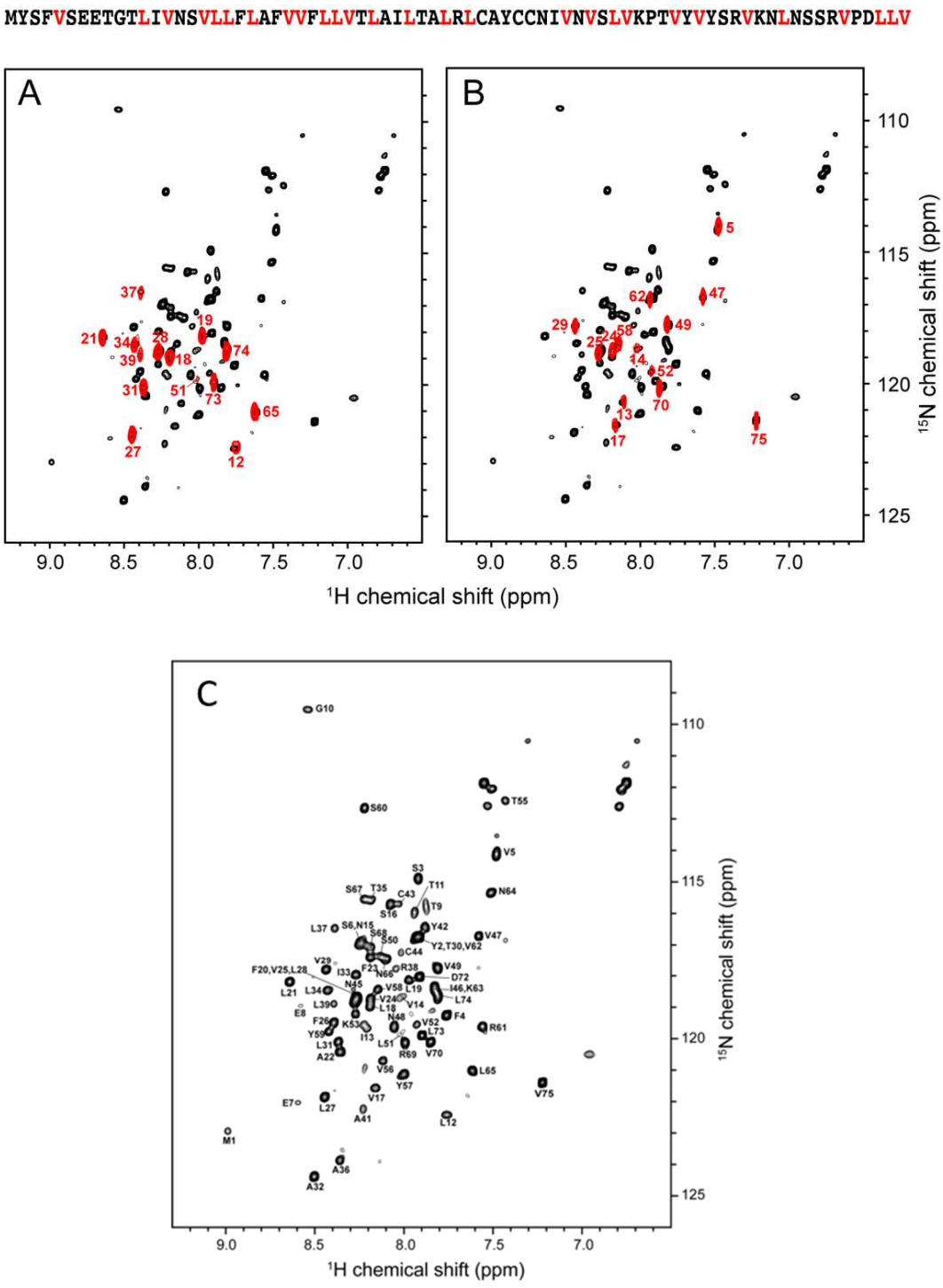

Figure 12. ^{1}H/^{15}N HSQC spectra of uniformly ^{15}N-labeled full-length E protein (EF) and selectively ^{15}N-Val and ^{15}N-Leu labeled EF. ^{15}N-Leu labeled EF (A) and ^{15}N-Val labeled EF (B) aided in the full backbone assignment of ^{15}N-labeled EF (C) The spectra of selectively labeled EF is depicted in red contours and superimposed on uniformly ^{15}N-labeled EF which is depicted in black contours. The resonance assignment of each selectively labeled spectra is labeled. The sequence of full-length E protein is seen above the selectively labeled samples, where all valine and leucine residues are marked in red [24].

Hydrogen/deuterium (H/D) exchange enabled identification of all residues in transmembrane helices. The spectrum of the full-length E protein was superimposed with those of the N-terminal transmembrane domain (ET) (spanning residues 1-39) and the C-terminal cytoplasmic domain (EC) (which spans residues 36-75) (Figure 13). Each E protein sample was first prepared in ~90% H_2O then incubated with >90% D_2O and placed in the spectrometer. The use of H/D exchange in $^{1}H/^{15}N$ HSQC aided in assessing whether the conformations of the E protein domains are stable. Results of H/D exchange indicated no amide signals from residues 36-75, which correspond to the cytoplasmic domain in the spectra of EF and EC (Figure 13). This suggests, that the amide hydrogens in the cytoplasmic domain are less accessible to the solvent. In contrast, strong signals from residues 19-35 and 19-33 in the spectra of EF and ET, verifying that these residues contribute to the stability of the transmembrane helix.

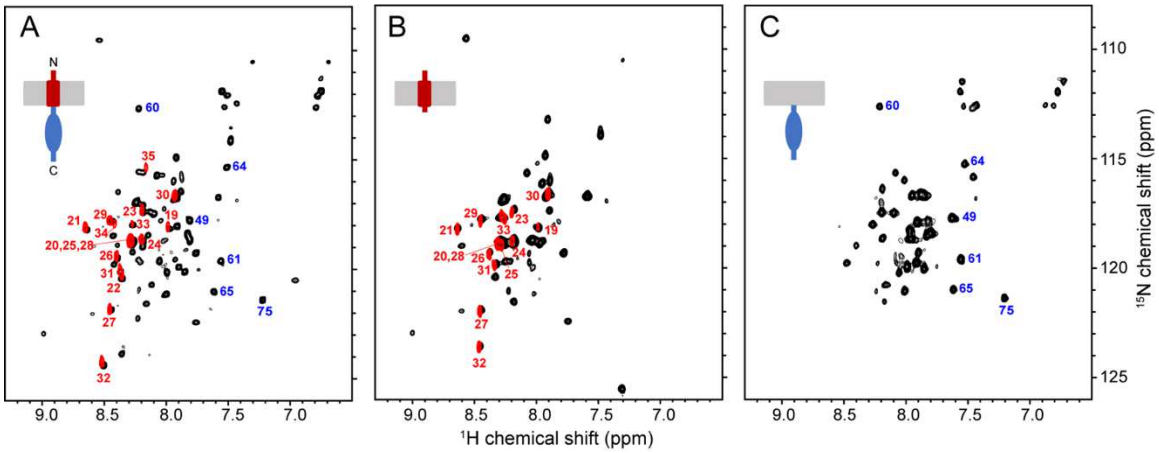

Figure 13. $^{1}H/^{15}N$ HSQC of uniformly ^{15}N-labeled full-length and truncated E protein constructs in HPC micelles in H_2O and D_2O. Full length 75 residue E protein (EF) is depicted in (A) N-terminal transmembrane domain of E protein (ET) (residues 1-39) is depicted in (B) C-terminal cytoplasmic domain of E protein (EC) (residues 36-75) is depicted in (C) Black contours represent sample in H_2O while red contours represent sample after D_2O exchange. Selected residues in the c-terminal cytoplasmic domain are labeled in blue and N-terminal transmembrane domain are labeled in red to distinguish signals of EC and ET [24].

Chapter 3: Characterization of E protein Drug Binding Site

3.1 Hexamethylene Amiloride Binding Interaction with Wild-type E protein Constructs

Hexamethylene amiloride (HMA), is a well-established channel blocker which has been shown to interact with E protein of SARS-CoV, however peptide sequences of E protein previously examined were truncated, excluding critical residues, i.e., the first 7 residues of the N-terminus and the last 37 residues of the C-terminus[15]. Here we examine HMA binding to each E protein construct, full-length E protein (EF) (1-75 residues), N-terminal transmembrane helix domain (ET) (1-38 residues) and the C-terminal cytosolic E protein (EC) (39-75 residues) (Figure 10). A ten-fold molar excess of HMA was added to each E protein NMR sample, then placed in the spectrometer to measure chemical shift perturbations (CSPs). The $^{1}H/^{15}N$ HSQC spectra of each E protein construct upon addition of HMA are shown in Figure 10. Red contours represent the signals in the presence of HMA and the black contours represent protein signals in the absence of HMA. Signals from select residues, primarily in the N-terminal region of the E protein, were found to be perturbed by HMA, as seen in the spectra of EF and ET (Figure 14A and B). The chemical shifts perturbations are plotted as a function of residue number in Figure 14D and E. Analysis of the spectra from the cytoplasmic domain of the E protein, EC, reveals no significant CSPs, which is consistent with the lack of chemical shift changes in the spectra of the C-terminal domain of the full-length E protein, EF (Figure 14A and E). Chemical shift perturbation occurred for residues 2-5 with the most perturbed signals from residues 6-18 in the N-terminal region of the transmembrane helix. These residues also correspond to the portion of the N-terminal end of the helix which was resistant to H/D exchange (Figure 13); therefore, along with the chemical shift perturbation data, we can qualitatively confirm interaction between HMA and the N-terminal region of E protein.

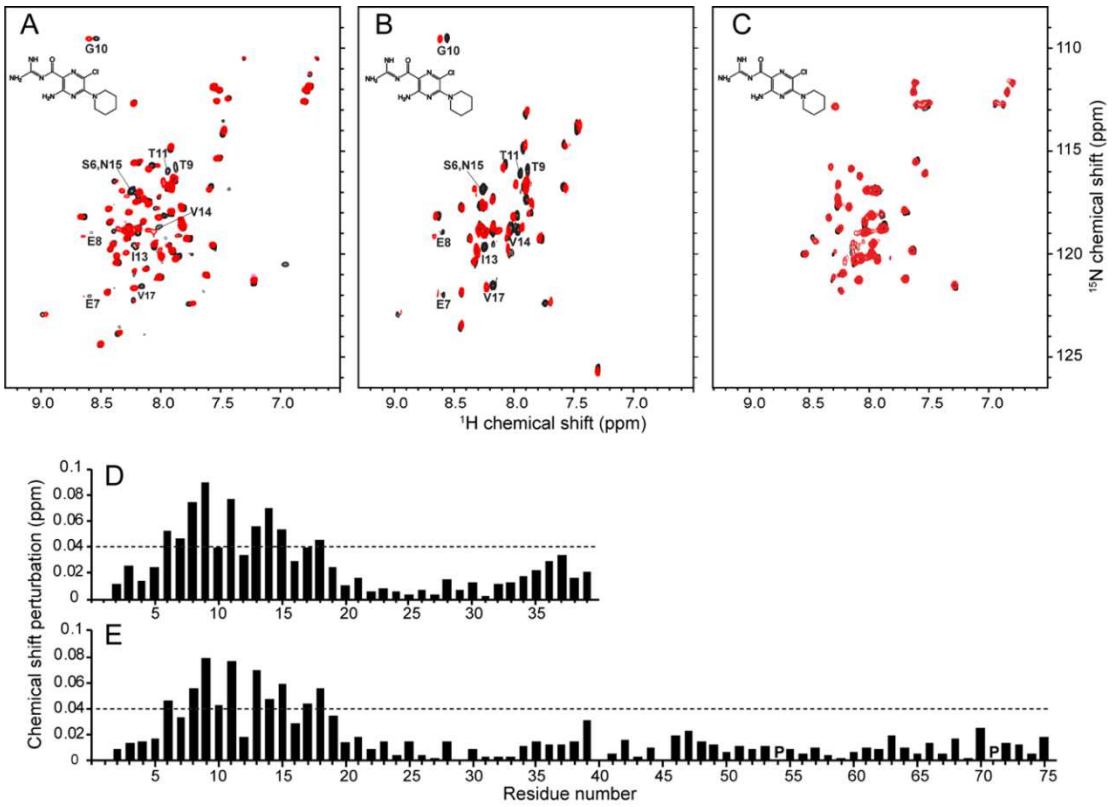

Figure 14. Chemical shift perturbation of E protein constructs interacting with HMA. $^{1}H/^{15}N$ HSQC spectra of ^{15}N-labeled EF, ET and EC shown in black contours, superimposed with the same spectra after addition of HMA, red contours. The full-length E protein (EF) (residue 1-75) (A) N-terminal transmembrane domain of E protein (ET) (residues 1-39) (B) C-terminal cytoplasmic domain of E protein (EC) (residues 36-75). Molar ratio of protein to HMA is 1:10, with the chemical structure of HMA present in each spectrum. The chemical shift perturbations of resonances after HMA binding are labeled with their assignments (A and B). Chemical shift perturbations as a function of residue number is seen for ET and EF in panel D and E, respectively. Horizontal dotted lines on panels D and E correspond to 1.5 times the average chemical shift perturbations induced by HMA binding to ET. Proline sites are marked as “P”

3.2 Hexamethylene Amiloride Binding Interaction with Mutant E protein Constructs

To assess the effect of mutations on the transmembrane domain of the SARS-CoV-2 E protein, we made two mutants of the full-length E protein, N15A, where Asn15 was mutated to Ala and V25F, where Val25 was mutated to Phe. The N15A mutation to the SARS-CoV-2 full length E protein construct (residues 1-75) resulted in, significant chemical shift perturbations from residues throughout the N-terminal domain of E protein, specifically those corresponding to Ser6, Glu7, Leu12 and Ser16 (Figure 16A and C). However, only minor chemical shift perturbations are observed for residues near the V25F mutation (Figure 16B and D). Comparison of the circular dichroism spectra of each mutant revealed no significant differences between the wild-type EF spectra and the N15A and V25F mutants (Figure 15), therefore the large chemical shift perturbations seen with the N15A mutation may be a result of intermolecular hydrogen bonding with the Asn15 sidechains [22]. Previous studies applied mutations to polypeptides consisting of the transmembrane helix of SARS-CoV E protein and found that in lipid bilayers point mutations can affect the ion channel activity of E protein. The interaction between the EF N15A and EF V25F mutants with HMA were studied to determine whether inducing single mutations in the transmembrane domain of the SARS-CoV-2 E protein affects drug binding. With the N15A mutant, a decrease in HMA binding was observed as there were no significant chemical shift changes (apart from Ser6) in the presence of HMA (Figure 17A and C). However, the V25F mutant did not affect HMA binding, as there were no changes in the signals of residues near the HMA binding site, with CSPs that closely resembled those of wild-type EF (Figure 17B and D). Therefore, these results suggest that Asn15 is integral to the conformational stability of E and its ability to bind HMA.

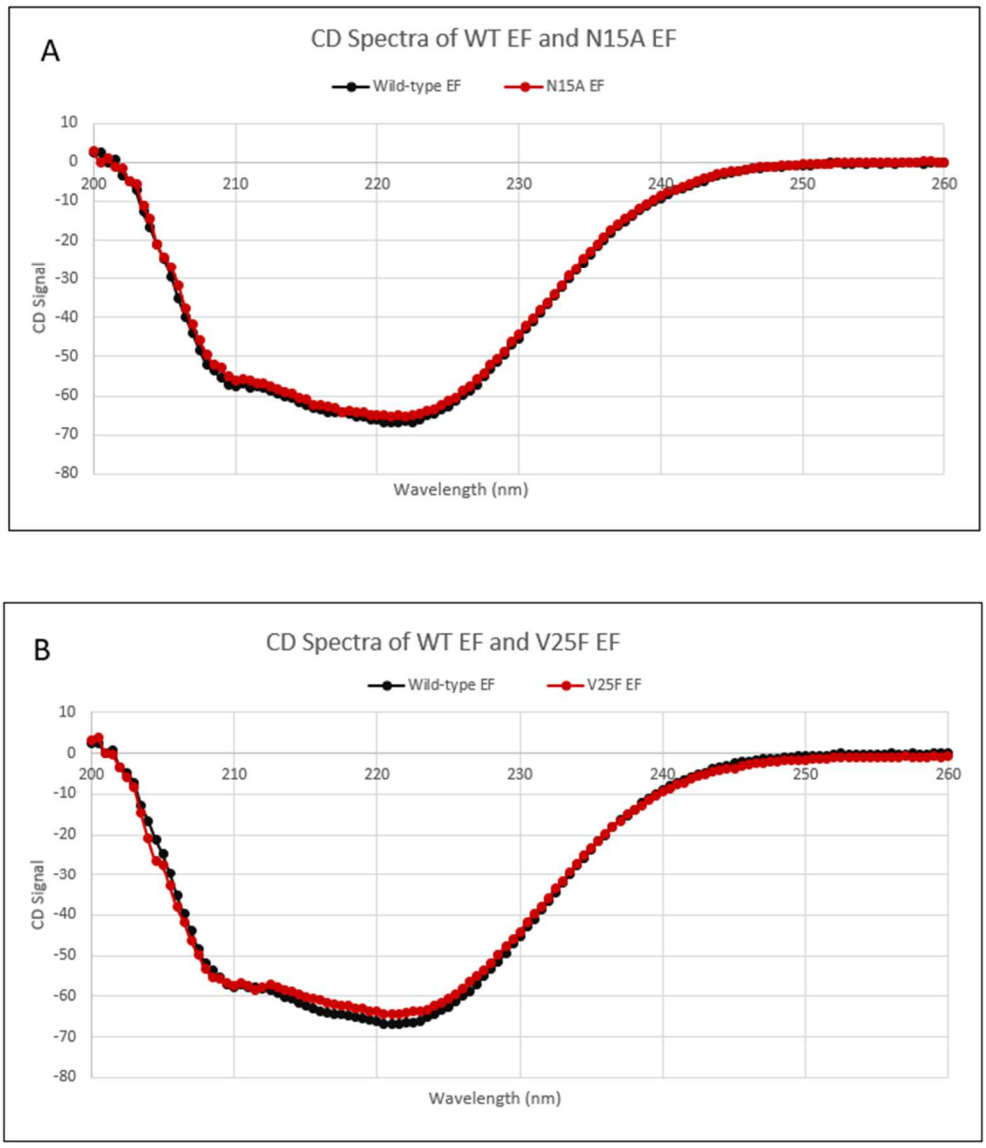

Figure 15. CD spectra of wild-type and mutant EF. CD spectra of wild-type full-length E protein (black) superimposed with mutant full-length E protein (red) recorded from 200-260nm. CD spectra of N15A EF overlayed on wild-type full-length EF (A) CD spectra of V25F EF overlayed on wild-type full-length EF (B).

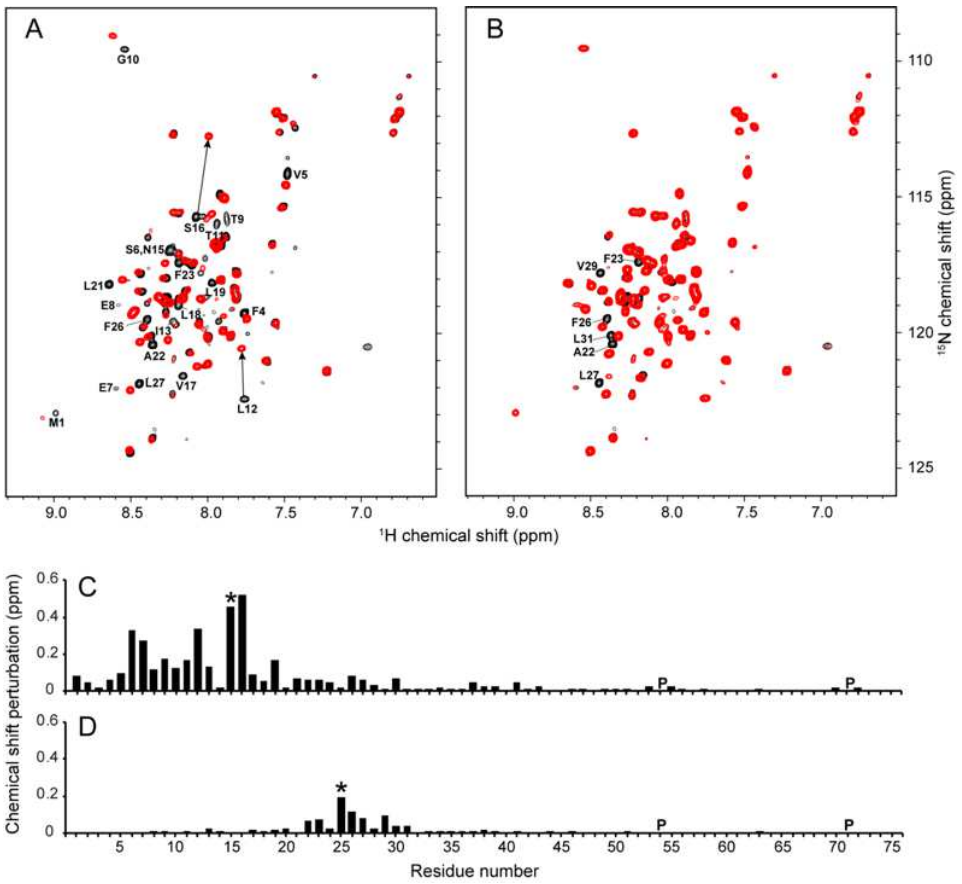

Figure 16. Comparison of ^{1}H/^{15}N HSQC of N15A and V25F EF mutants. The N15A mutant is seen in (A) while the V25F mutant is seen in (B) Wild type full-length E protein spectrum (EF) is seen in black contours while the N15A and V25F mutants' spectra are seen in red contours. Resonances that have significant chemical shift perturbation are assigned on each figure. The chemical shift perturbation plot of the N15A mutant (C) and the V25F mutant (D) is seen in the lower panels. Mutation sites are denoted with an asterisk while proline residues are marked as "P".

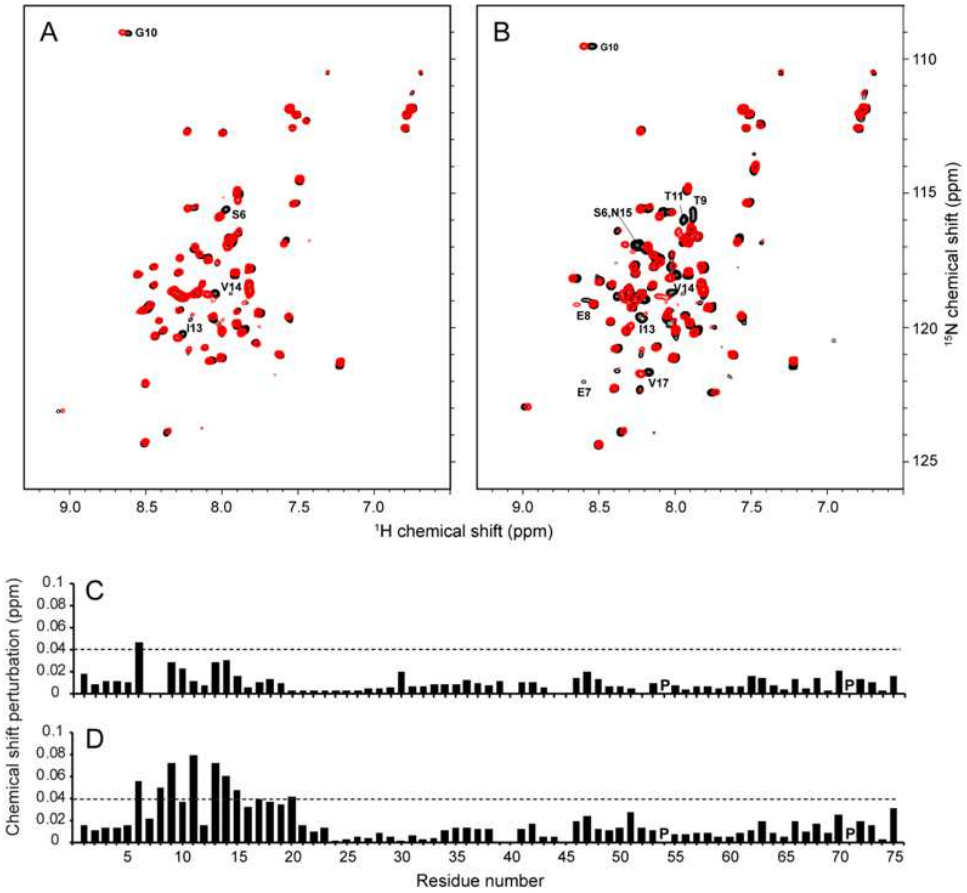

Figure 17. Comparison of ^{1}H/^{15}N HSQC of N15A and V25F EF mutants in the presence of HMA. The N15A mutant is seen in (A.) while the V25F mutant is seen in (B.) The spectra of the mutants in the absence of HMA is seen in black contours while the spectra of N15A and V25F mutants in HMA are seen in red contours. Resonances that have significant chemical shift perturbation by HMA are assigned on each figure. The chemical shift perturbation plot due to HMA binding to the N15A mutant (C.) and the V25F mutant (D.) is seen in the lower panels. The average chemical shift changes of V25F EF by HMA times 1.5 is denoted by the dotted lines. Proline residues are marked as "P".

Figures 1, 12, 13, 14, 16 and 17 have been submitted for publication of the material as it may appear in PLOS pathogens, 2021, Park SH, Castro DV, Siddiqi H, De Angelis A, Oom A, Stoneham C, Lewinski M, Clark A, Croker B, Carlin A, Guatelli J, Opella SJ. The thesis author was a co-author of these figures. [24]

Chapter 4: Spike Protein Construct Design, Purification and NMR experiments

In addition to structurally characterizing SARS-CoV-2 E protein, the spike protein has been of interest due to the potential intermolecular disulfide interactions between the C-terminal domain of the E protein and the cysteine rich, C-terminal domain of S protein [18]. The roles of E and S in viral replication, infection and pathogenicity affirm the importance of investigating their structure and functional capabilities individually and in tandem. To prepare for interaction studies between the two domains, a construct was designed for the C-terminal domain of SARS-CoV-2 S protein (SCTD) spanning residues 1201-1262 (Figure 18). The same pET-31(b)+ expression vector was utilized for SCTD fusion-protein expression and purification. The SCTD fusion protein was expressed in *E. coli* and purified following the same purification scheme used for KSI-E fusion protein constructs, with the use of 0.05% HPC throughout the entire purification process (Figure 5).

Due to the ten cysteine residues present in the C-terminal domain of S protein, purification of S was performed under reduced conditions. In assessing SDS-PAGE results, SCTD fusion protein was well expressed (seen ~24 kDa) when directed to inclusion bodies, in *E. coli* (Figure 18C). The binding efficiency of SCTD can be seen in the flowthrough fraction (Figure 18C, lane 4) where there is minimal fusion-protein loss. The washing fractions further demonstrate its tight column binding, as the impurities are shown to be washed away (Figure 18C, lanes 5 and 6). The resulting eluate depicts the KSI- SCTD fusion protein (~24 kDa) (Figure 18C, lane 7). Enzymatic cleavage of the fusion protein was achieved through overnight incubation in thrombin. After running the 2nd Ni-NTA column purification, the target SCTD protein is seen to be mostly pure (~7.2 kDa) (Figure 18C, lane 9).

Currently, no structural studies have characterized the full-length or C-terminal domain of S protein, however since it plays an integral role in viral infection and host-cell entry, understanding the dynamics and structure to function relationship of the S protein are likely to be crucial in the development of effective anti-viral therapeutics. Cryo-EM studies of the Spike

protein have previously characterized the S1 and S2 subunits trimming the transmembrane and C-terminal domains [28], however these regions are critical in probing potential intermolecular interactions between S and E protein. In order to assess the structural folding of SCTD in HPC micelles, a $^{1}H/^{15}N$ HSQC spectrum was obtained (Figure 19). To prevent S protein aggregation, 1 mM TCEP was added to keep all cysteine residues reduced. The spectrum was well-resolved and demonstrated early success in the optimization of purifying and preparing the S protein sample. In future studies of the protein-protein interactions between E and S, the reducing agent must be removed to allow for the cysteine sidechains of the C-terminal domain of E protein to interact with those of the C-terminal domain of the S protein.

A KSI- SCTD Fusion:

MHTPEHITAVVQRFVAALNAGDLDGIVALFADDATVEDPVGSEPRSGTAAIREFYA
NSLKLPLAVELTQEVRAVANEAAFAFTVSFEYQGRKTVVAPIDHFRFNGAGKVVSI
RALFGEKNIHACQMLDPGGKKHHHHHHHHHHGGKKLVPR**GSQELGKYEQYIKWPWY**
IWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSE

B SCTD (1201-1262):

GSQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCC
KFDEDDSE

C

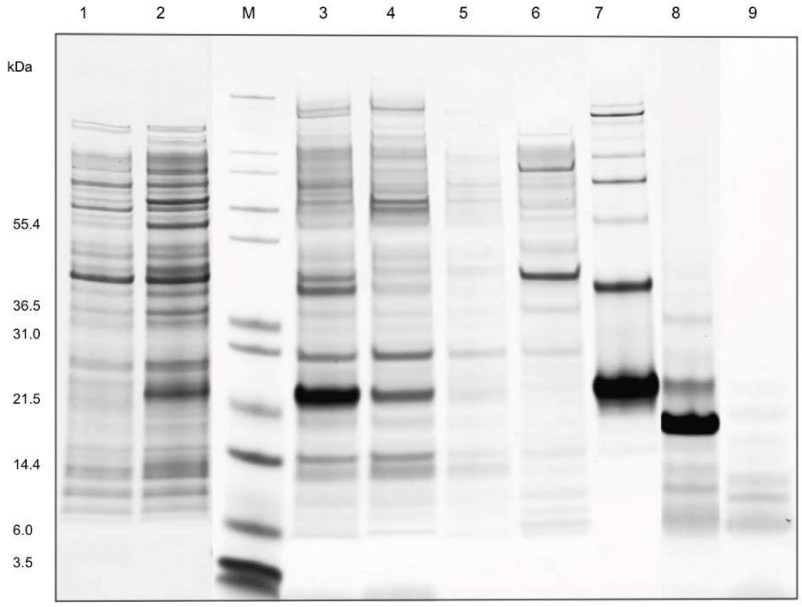

Figure 18. C-terminal domain S protein (SCTD) construct design and purification. The KSI- fusion sequence of S CTD is seen to be 215 amino acids (A.) The sequence post-enzymatic cleavage is seen to be 64 amino acids in length, spanning residues 1201-1262 (B.) The SDS-PAGE of the purification of S CTD is shown with the molecular weight of the marker listed in kDa to the left of lane 1. Lane 1, pre-induction cells; lane 2, post-induction cells (KSI-fusion S CTD is seen ~24 kDa); lane 3, HPC-solubilized inclusion bodies containing the KSI-S CTD fusion protein; lane 4, Ni-affinity column flow through; lane 5, Ni-affinity column wash; lane 6, Ni-affinity column wash with imidazole; lane 7, eluate of the KSI-S CTD fusion protein from the column; lane 8, after thrombin cleavage of the KSI-SCTD fusion protein; lane 9, purified S CTD (~7.2 kDa).

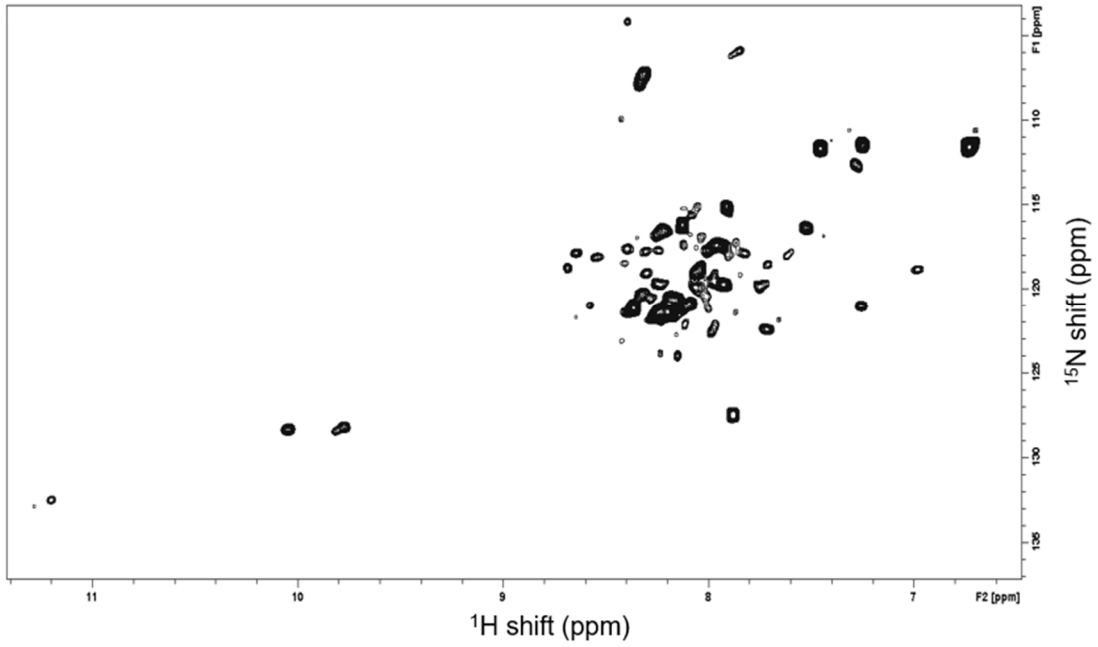

Figure 19. Initial SCTD spectrum. $^{1}H/^{15}N$ HSQC of C-terminal domain S protein construct (SCTD) (residues 1201-1262)

Discussion

The medical impact of SARS-CoV-2 has expedited the structural and functional studies of coronavirus proteins to develop antiviral therapeutics. There is particular interest in structural studies of the envelope (E) and spike (S) proteins as they are involved in pathogenesis and viral replication and infection (Figure 1 and 2)[23]. Here, I demonstrate the use of a novel bacterial expression and purification system that allows for direct membrane protein expression and purification through inclusion bodies in *E. coli* and Ni-column chromatography. The E and S protein samples were then investigated using solution-state NMR spectroscopy in HPC micelles. Ultimately, various aspects of the structure and dynamics of the full-length E protein as well as its suitability as a drug target should emerge. The full length 75 residue E protein, EF, is studied alongside the truncated constructs ET (residues 1-39) and EC (residues 36-75), the N-terminal transmembrane domain and the C-terminal cytosolic domain, respectively (Figure 3B).

To begin spectroscopic studies, the expression of isotopically labeled protein is required, however that can pose a challenge when working with viral membrane proteins. As with other intrinsic membrane proteins, the hydrophobic natures of E and S make them likely to aggregate, and their cysteine residues make them likely to form non-specific intermolecular or intramolecular disulfide linkages, especially when considering the cysteine rich region of the S protein and the closely spaced cysteines present in C-terminal domain of the E protein. Therefore, the development of a simple and efficient system where membrane protein expression is possible through *E. coli* and direct protein purification is possible through column chromatography makes the modified pET-31(b)+ vector a valuable contribution to the experimental system (Figure 5).

Advantages to expressing viral mammalian proteins through *E. coli* include the economic efficiency and high production yield of recombinant protein. The expression of full-length SARS-CoV-2 E protein was optimized through this novel bacterial expression and purification system. The KSI-E fusion protein was expressed in competent C43(DE3) *E. coli* cells; the key aspect to

this system is that the hydrophobic E protein expression was directed to the inclusion bodies, thereby preventing cytotoxicity to the cells. The expression system reveals a 24-residue linker sequence consisting of a ten-His tag and 6-residue thrombin cleavage sequence (LVPRGS) between the KSI fusion partner and the E protein construct of interest. This allowed the use of affinity column chromatography and enzymatic cleavage, which were not possible with the unmodified pET-31(b)+ vector. In purifying the expressed fusion protein, HPC, a mild detergent with a low CMC, was utilized throughout the entire process (Figure 6). From inclusion body solubilization to Ni-NTA affinity column chromatography, HPC was the only detergent present when purifying each E protein and S protein construct, which allowed for direct NMR sample preparation after obtaining the purified target protein. To maximize protein concentration and sample purity, concentrations of HPC and imidazole used during protein purification were optimized (Figure 4). This ensured that the fusion-protein was exposed to the lowest possible detergent and imidazole concentration to prevent on-column aggregation and protein loss, respectively.

Upon optimization of thrombin concentration (Figure 7), issues arose with the cleavage of the C-terminal cytosolic E protein construct (EC) as the KSI fusion partner was also cleaved, resulting in the presence of peptide fragments in the final target protein eluate (Figure 6). The impurities seen around 11.0 kDa and 5.5 kDa, were identified through LC-MS, with the cleavage site being within the KSI fusion protein between Arg102 and Phe103 (Figure 10A and C). To address this problem Arg102 of the KSI sequence was mutated to Glu (Figure 9), after ensuring that Arg102 was not conserved amongst the KSI sequences of several other species. Although this mutant fusion protein expressed well in *E. coli*, it produced the same peptide fragments upon thrombin cleavage, which remained in the final eluate during the second Ni-NTA column chromatography. This specific cleavage of the KSI fusion partner occurred for both the full-length E protein construct (Figure 8) and the truncated N-terminal transmembrane domain construct (ET) as well. Therefore, although this expression system was ideal for high yield

membrane protein expression in *E.coli* and utilized the simplicity of enzymatic cleavage, the KSI fusion partner was not an ideal candidate for enzymatic cleavage via thrombin. In future studies, to preserve the effective and efficient use of Ni-NTA column chromatography while avoiding over digestion of KSI and peptide fragments in the final target protein eluate, substitution of the KSI fusion partner for GST may be necessary. Since KSI was modified to allow for enzymatic cleavage, the use of GST may be preferable as it is designed from enzymatic cleavage, thereby simplifying the plasmid vector, and reducing impurities in the final target protein eluate.

To gain insight on the secondary structure of SARS-CoV-2 E protein, each of the E protein constructs, full-length (EF), N-terminal transmembrane domain (ET) and the C-terminal cytosolic (EC) as well as the EF mutant (N15A and V25F) constructs, were prepared for circular dichroism (CD) spectroscopy (Figure 11 and 15). CD spectroscopy allows for an estimation of the alpha helix and beta sheet content in a protein sample based on the spectral pattern, which gives qualitative insight of secondary structure or through algorithmic calculations a quantitative understanding can be determined. Upon assessing the spectral pattern for each construct, the full-length E protein construct EF, the N-terminal transmembrane construct ET and the C-terminal domain construct EC displayed a typical alpha helical pattern from 200-260nm, with two distinct local minima around 210nm and 220nm, indicative of their secondary structure. Therefore, the CD data suggests that the full-length E protein (EF), especially the N-terminal transmembrane domain (ET) and a region in the C-terminal domain (EC) have alpha helix content, which was found to be consistent with the $^{1}H/^{15}N$ NMR spectra of each construct.

To further validate the findings from CD spectroscopy, solution-state NMR spectroscopy was implemented along with hydrogen/deuterium exchange to assess the dynamics of the E protein (Figure 13). Solution NMR studies of the full-length SARS-CoV-2 E protein was performed in HPC micelles, which resulted in high-resolution spectra evident in the signal-to-noise ratios of the amide signals and their chemical shift frequencies in the $^{1}H/^{15}N$ HSQC spectrum of uniformly ^{15}N labeled EF (Figure 13). This resolution is also depicted in the fully

assigned backbone signals of the full-length EF spectrum as well as the resonances of the truncated ET (residues 1-39) and EC (residues 36-75) constructs, ultimately aiding in the characterization of E protein secondary structure and dynamics (Figure 12). The use of CD (Figure 14 and 15) in conjunction with solution-state NMR spectroscopy revealed that the SARS-CoV-2 E protein has a long 36-residue alpha-helix in the N-terminal transmembrane domain (residues 8-43). The C-terminal domain of E protein depicted an 8-residue alpha-helix, which is shorter than that of the transmembrane domain and was found to have a different orientation. Analysis of the hydrogen/deuterium exchange NMR spectra of each construct, confirms that most of the ^{1}H and ^{15}N chemical shifts of the resonances seen in the full-length (EF), N-terminal transmembrane domain (ET) and the C-terminal cytosolic (EC) constructs in the two-dimensional HSQC spectra overlap well. The structural conservation of each domain is evident whether these domains are seen individually as truncated constructs or seen in the full-length E protein construct.

The SARS-CoV-2 E protein exhibits ion channel activity in the full-length E protein and its truncated N-terminal transmembrane helix constructs. Hexamethylene amiloride (HMA), a well-established channel blocker was shown to inhibit the channel activity of E protein from several coronaviruses, MHV, HCoV-229E, SCV, and FIPV at low micromolar range [20,26]. The transmembrane domain of SARS-CoV E protein was found to bind HMA, based on NMR studies of the binding interaction with HMA and truncated E protein constructs [4,5,15,16]. Several drug binding sites were proposed upon assessing the chemical shift perturbations seen for 31- or 58-residue polypeptides. Here, the NMR spectra of full-length SARS-CoV-2 E protein in the presence of ten-fold molar excess HMA added to EF, ET, and EC constructs (Figure 14). The data collected from these experiments show the strongest chemical shift perturbations in the signals from in residues 2-18, demonstrating that the N-terminal residues of the E protein are affected by HMA binding. Hydrophilic residues (Ser6, Glu7, Glu8, Thr9, Thr11, and Asn15)[24] are also affected by HMA, as seen through the perturbed signals (Figure 14). The spectrum of the

C-terminal region of the transmembrane helix had only minor chemical shift perturbations in ET (Figure 14B), however this was not observed in spectra of the EF and EC constructs, therefore these CSPs are likely due to non-specific interactions. These results contrast with the CSPs seen in the truncated E protein constructs (residues 8-65) from previous studies, which depicted large CSPs for Val49 and Lys65.

The interaction between the mutant N15A and V25F full-length E protein in the presence of ten-molar HMA was observed through $^{1}H/^{15}N$ HSQC spectra. Comparison of the NMR spectra of N15A EF without HMA (Figure 16A) and that of N15A EF in the presence of HMA (Figure 17A) depicted a significant chemical shift perturbation in the N-terminal region of the E protein, whereas with the V25F mutation no significant changes were observed. With the EF N15A mutant, we can see that the drug binding site is affected and there is no interaction with HMA, indicating that the residue Asn15 is integral to the SARS-CoV-2 E protein conformation. This residue, Asn15, or Gln15 has been found to be highly conserved in both alpha and beta coronaviruses, with evidence suggesting that Gln aids in helix-helix associations in transmembrane domains through hydrogen bonding [27]. The Asn15 sidechain may be involved in intermolecular hydrogen bonds which allow for the N-terminal region to maintain its conformation and orientation [5]. The CSPs seen from the V25F mutation in the presence of HMA were small and local to the mutation site which suggests that the full-length E protein mutant retained its conformation, which is consistent with the data seen with the wild-type full-length E protein interacting with HMA.

Interest in intermolecular interactions between the SARS-CoV-2 spike (S) protein and E protein lead to designing the S C-terminal domain (SCTD) construct. With all ten cysteine residues of the spike c-terminal domain intact, SCTD (residues 1201-1262) was inserted into the KSI expression system and purified through Ni-NTA column chromatography using 0.05% HPC. Due to the cysteine rich c-terminal domain, preventing on-column aggregation during protein purification was a challenge, however, this concern was addressed by adding 1mM

TCEP to all buffers during purification and post-enzymatic cleavage via thrombin (Figure 18). Upon initial SDS-PAGE assessment, the fusion protein was well expressed through the use of a modified plasmid vector and the final purified protein eluate depicted minimal impurities which may require further optimization. From the purified sample, an initial $^{1}H/^{15}N$ HSQC spectrum was obtained for the SCTD construct under reduced conditions in HPC micelles. However, interaction studies have yet to be tested as the reducing agent, TCEP, must be removed from the sample to allow cysteine residues to interact with those of the full-length E protein. For prospective studies, the TCEP should be dialyzed out of the final SCTD sample then SCTD should be incubated overnight with the full-length E protein. A solution-state NMR sample should then be prepared, and the resulting spectrum compared to that of SCTD (Figure 19) to assess any potential CSPs.

Development of the KSI expression system and HPC purification scheme were essential for the study of membrane proteins. The development of the pET-31(b)+ vector and expression system, allowed for inclusion body directed membrane protein expression and enzymatic cleavage, although further fusion partner modifications may be needed to improve overall sample purity. This expression system also proved to be effective in expressing the C-terminal domain S protein construct, ultimately leading to an initial $^{1}H/^{15}N$ HSQC spectrum. The samples resulting from the expression and purification of various E protein constructs were used to characterize its secondary structure and drug binding site. Interaction experiments with HMA and E protein revealed key residues involved in drug binding (Ser6, Glu7, Glu8, Thr9, Thr11, and Asn15) within the N-terminal domain of the transmembrane helix. The interaction between HMA and the full-length E protein mutants (N15A and V25F) highlighted the necessity of Asn15 in preserving the conformation of the HMA binding site and the potential role of E protein's N-terminal domain in virus assembly and release. The secondary structure of E protein was also found to consist of a long transmembrane helix spanning residues 8-43 of the full-length E

protein N-terminal region, with a shorter helix seen between residues 53-60 of the E protein C-terminal domain. To gain a complete understanding of the biological function of SARS-CoV-2 E protein in near native conditions, an essential step forward would be determining the three-dimensional structure of E protein in phospholipid bilayers as well as testing potential intermolecular interaction with S protein. A comprehensive understanding of the SARS-CoV-2 E protein and any possible intermolecular interactions it may have with S protein will contribute to antiviral drug development and therapeutics.

Materials and Methods

SARS-CoV-2 E and S protein Construct Design

The E protein sequences were based of the 75-residue full-length wild-type E protein and the C-terminal domain of S protein from SARS-CoV-2 isolate Wuhan-Hu-1 (NC_045512). E protein expression in *E. coli* C43(DE3) was enhanced through the addition of a codon-optimized gene of the full-length E protein amino acid sequence. This gene was then inserted into a pET-31(b)+ vector which was modified through the addition of a twenty-four-residue linker sequence consisting of a 10 His-tag and followed by a 6-residue (LVPRGS) thrombin cleavage site. These modifications were introduced after the ketosteroid isomerase sequence, innate to the pET-31(b)+ vector (Figure 3B). This expression system was utilized for all E protein constructs including the full-length wild-type (EF) (residues 1-75), N-terminal transmembrane domain (ET) (residues 1-39) and the C-terminal cytoplasmic domain (EC) (residues 36-75) (Figure 3B). Mutant full-length E protein constructs, N15A and V25F, were made through a site-directed mutagenesis kit from NEB.com.

Fusion Protein Expression and Purification

The plasmid vector consisting of the target S or E protein constructs were transformed into competent *E. coli* strain C43(DE3) cells (www.lucigen.com). The cells were grown in minimal media and isotopically labeled [31] through the addition of 1 g/L $(^{15}NH_4)_2SO_4$. Selectively labeled ^{15}N-Val and ^{15}N-Leu samples were grown in minimal media consisting of unlabeled ammonium sulfate, 100 mg/L ^{15}N-labeled Valine or Leucine (www.isotope.com) and at 200-500 mg/L of the corresponding 19 amino acid residues. To begin fusion protein expression, a preculture consisting of 50 mL of Luria-Bertani (LB) broth was grown overnight, then 1% (v/v) overnight preculture was added to a baffled two-liter flask containing 500 mL of minimal medium. The culture was grown at 37 ºC in a 200 rpm shaking incubator until the cell's reached 0.5 optical density at 600nm. Fusion protein expression was induced upon the addition of

isopropyl beta-D-1-thiogalactopyranoside (IPTG) to 1mM final concentration. Cells were harvested 3 hours post- IPTG induction by centrifugation at 4 °C, 5,000 xg for 20 minutes, then stored overnight at -80°C.[24]

To cause cell lysis, 36 mL of buffer containing 20mM Tris-HCl, 500 mM NaCl at pH 8 with 50 μg/mL lysozyme, and 250 units Benzonase nuclease (www.sigmaaldrich.com) was added to 1L cell pellet and resuspended thoroughly. Then using the Fisher Scientific Sonic Dismembrator 550 with duty cycle 20% and output control 4, the cell lysate was sonicated on ice for 10 minutes. Then the sonicated lysate was incubated with gentle rotation for 1 hour at room temperature with 4 mL of 20% (v/v) Triton X-100 at a final concentration of 2% (v/v). The inclusion body pellet was obtained through centrifugation of the cell lysate for 30 minutes at 4°C, 20,000 xg. Then after discarding the supernatant, the pellet was solubilized for 2 hours at room temperature in buffer consisting of 30 mL 20 mM HEPES, 500 mM NaCl at pH 7.8, with 300 mg HPC (www.anatrace.com) added at 1% (w/v) final concentration and tris (2-carboxyethyl) phosphine hydrochloride (TCEP-HCl) added at 1mM final concentration. The inclusion body solution was then centrifuged for 30 minutes at 15°C, 40,000 xg and the resulting supernatant was loaded on a 10 mL Ni-NTA superflow column equilibrated with HPC binding buffer containing 0.05% HPC, 20 mM HEPES, 500 mM NaCl at pH 7.8. The column was then washed with 50 mL HPC binding buffer followed by 100 mL of HPC washing buffer containing 0.05% HPC, 20 mM HEPES, 500 mM NaCl, 20 mM imidazole at pH 7.8. The KSI-fusion protein was then eluted through the addition of 20 mL HPC elution buffer containing 0.05% HPC, 20 mM HEPES, 500 mM NaCl, 500 mM imidazole at pH 7.8.[24]

The fusion protein eluate was then cleaved through dialysis overnight in a 10 kDa MW cutoff dialysis membrane (www.spectrumchemical.com) against thrombin cleavage buffer 20 mM HEPES, 50 mM NaCl, 1 mM EDTA at pH 7.8. The post-cleavage suspension was then loaded to a Ni-NTA affinity column equilibrated with HPC binding buffer. The flowthrough was

collected as it contained the target protein (E or S depending on construct used). This purification protocol was used for all S and E protein constructs.[24]

Electrophoresis

NuPAGE 4-12% Bis-Tris precast gels from Invitrogen were used to perform SDS-PAGE in 2-(N-morpholino) ethane sulfonic acid (MES) buffer at room temperature. Coomassie blue staining was used to visualize protein bands. The 5 μg protein samples were prepared in HPC binding buffer mixed with NuPAGE LDS sample buffer (4X) and DTT, then vortexed and centrifuged at 12,000 xg and loaded into the gel.

CD and NMR Sample preparation

Each sample prepared for CD spectroscopic experiments had a concentration of 50 μM across all E protein constructs (EF, ET, EC, N15A EF and V25F EF). Samples were prepared in binding buffer containing 20 mM HEPES, 50 mM NaCl, pH 6.5 at a final volume of 400 μL for the 1 mM Quartz CD Cuvette. Fixed wavelength scans were performed at room temperature with data was collection from 200-260nm with CD signal collected every 0.5nm step on AVIV CD spectrometer model 215.

Solution NMR samples were prepared by concentrating proteins with Amicon Ultra-4 10K centrifugal filters (www.endmillipore.com). Uniformly ^{15}N-labeled and selectively ^{15}N-Leu and ^{15}N-Val labeled E protein samples at a final concentration of 0.5mM in 5% (w/v) (123 mM) HPC, 20 mM HEPES, 50 mM NaCl, 10% (v/v) $^{2}H_2O$, 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. NMR experiments were performed at 50^{o}C on triple-resonance Bruker Avance 800 and Avance 600 spectrometers. The chemical shifts of ^{1}H were referenced to 0 ppm for DSS and data was processed through Bruker Topspin 4 (www.bruker.com), NMRpipe/NMR Draw [30], and NMR View [29].

HMA binding

A stock solution containing 100mM 5-(N, N-hexamethylene)-amiloride (HMA) was prepared in deuterated dimethyl sulfoxide (DMSO-d_6). The samples for two-dimensional $^{1}H/^{15}N$ HSQC experiments samples were prepared with 0.2 mM uniformly ^{15}N-labeled E protein with and without 2 mM HMA. The chemical shift perturbations of each protein resonance 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. To determine the chemical shift changes as a function of HMA concentration for the N-terminal transmembrane domain E protein (ET), the $^{1}H/^{15}N$ HSQC spectra of 0.2 mM uniformly ^{15}N-labeled ET was recorded in the presence of 0, 0.25, 0.5, 1, 1.5, and 2 mM HMA.[24]

References

1. Valencia D (2020) Brief Review on COVID-19: The 2020 Pandemic Caused by SARS-CoV-2. Cureus. doi: 10.7759/cureus.7386

2. Dyer O. Covid-19: Study claims real global deaths are twice official figures BMJ 2021; 373 :n1188 doi:10.1136/bmj.n1188

3. Neuman, BW, Kiss G, Kunding AH, Bhella D, Baksh MF, Connelly S, Droese B, Klaus JP, Makino S, Sawicki SG, Siddell SG, Stamou DG, Wilson IA, Kuhn P, Buchmeier MJ. A structural analysis of M protein in coronavirus assembly and morphology. J Struct Biol. 2011;174(1):11-22. Epub 2010/12/07. doi: 10.1016/j.jsb.2010.11.021. PubMed PMID: 21130884; PubMed Central PMCID: PMCPMC4486061.

4. Surya W, Li Y, Torres J. Structural model of the SARS coronavirus E channel in LMPG micelles. Biochim Biophys Acta Biomembr. 2018;1860(6):1309-17. Epub 2018/02/24. doi: 10.1016/j.bbamem.2018.02.017. PubMed PMID: 29474890; PubMed Central PMCID: PMCPMC7094280.

5. Mandala VS, McKay MJ, Shcherbakov AA, Dregni AJ, Kolocouris A, Hong M. Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers. Nat Struct Mol Biol. 2020;27(12):1202-8. Epub 2020/11/13. doi: 10.1038/s41594-020-00536-8. PubMed PMID: 33177698; PubMed Central PMCID: PMCPMC7718435.

6. Walls AC, Tortorici MA, Bosch BJ, Frenz B, Rottier PJM, DiMaio F, Rey F, Veesler D. Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. Nature. 2016;531(7592):114-7. Epub 2016/02/09. doi: 10.1038/nature16988. PubMed PMID: 26855426; PubMed Central PMCID: PMCPMC5018210.

7. Singh Tomar, Prabhat Pratap, and Isaiah T Arkin. “SARS-CoV-2 E protein is a potential ion channel that can be inhibited by Gliclazide and Memantine.” *Biochemical and biophysical research communications* vol. 530,1 (2020): 10-14. doi:10.1016/j.bbrc.2020.05.206

8. V’kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. Coronavirus biology and replication: implications for SARS-CoV-2. Nature Reviews Microbiology 19:155–170 doi: 10.1038/s41579-020-00468-6.

9. Schoeman D, Fielding BC. Coronavirus envelope protein: current knowledge. Virol J. 2019;16(1):69. Epub 2019/05/28. doi: 10.1186/s12985-019-1182-0. PubMed PMID: 31133031; PubMed Central PMCID: PMCPMC6537279.

10. Opella SJ. Relating structure and function of viral membrane-spanning miniproteins. Curr Opin Virol. 2015;12:121-5. Epub 2015/06/10. doi: 10.1016/j.coviro.2015.05.006. PubMed PMID: 26057606; PubMed Central PMCID: PMCPMC4476644.

11. Opella SJ, Marassi FM. Applications of NMR to membrane proteins. Arch Biochem Biophys. 2017;628:92-101. Epub 2017/05/23. doi: 10.1016/j.abb.2017.05.011. PubMed PMID: 28529197; PubMed Central PMCID: PMCPMC5657258.

12. Cross TA, Opella SJ. Structural properties of fd coat protein in sodium dodecyl sulfate micelles. Biochem Biophys Res Commun. 1980;92(2):478-84. Epub 1980/01/29. doi: 10.1016/0006-291x(80)90358-7. PubMed PMID: 6986868.

13. Frey L, Lakomek NA, Riek R, Bibow S. Micelles, Bicelles, and Nanodiscs: Comparing the Impact of Membrane Mimetics on Membrane Protein Backbone Dynamics. Angew Chem Int Ed Engl. 2017;56(1):380-3. Epub 2016/11/25. doi: 10.1002/anie.201608246. PubMed PMID: 27882643; PubMed Central PMCID: PMCPMC6680326.

14. Chipot C, Dehez F, Schnell JR, Zitzmann N, Pebay-Peyroula E, Catoire LJ, Miroux B, Kunji ER, Veglia G, Cross TA, Schanda P. Perturbations of Native Membrane Protein Structure in Alkyl Phosphocholine Detergents: A Critical Assessment of NMR and Biophysical Studies. Chem Rev. 2018;118(7):3559-607. Epub 2018/03/01. doi: 10.1021/acs.chemrev.7b00570. PubMed PMID: 29488756; PubMed Central PMCID: PMCPMC5896743.

15. Pervushin K, Tan E, Parthasarathy K, Lin X, Jiang FL, Yu D, Vararattanavech A, Soong TW, Liu DX, Torres J. Structure and inhibition of the SARS coronavirus envelope protein ion channel. PLoS Pathog. 2009;5(7):e1000511. Epub 2009/07/14. doi: 10.1371/journal.ppat.1000511. PubMed PMID: 19593379; PubMed Central PMCID: PMCPMC2702000.

16. Li Y, Surya W, Claudine S, Torres J. Structure of a conserved Golgi complex-targeting signal in coronavirus envelope proteins. J Biol Chem. 2014;289(18):12535-49. Epub 2014/03/29. doi: 10.1074/jbc.M114.560094M114.560094 [pii]. PubMed PMID: 24668816; PubMed Central PMCID: PMC4007446.

17. Wilson L, McKinlay C, Gage P, Ewart G. SARS coronavirus E protein forms cation-selective ion channels. Virology. 2004;330(1):322-31. Epub 2004/11/06. doi: 10.1016/j.virol.2004.09.033. PubMed PMID: 15527857; PubMed Central PMCID: PMCPMC7111769.

18. Wu Q, Zhang Y, Lu 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 Ji, Bi S, Yang H. The E protein is a multifunctional membrane protein of SARS-CoV. Genomics Proteomics Bioinformatics. 2003;1(2):131-44. Epub 2005/01/01. doi: 10.1016/s1672-0229(03)01017-9. PubMed PMID: 15626343; PubMed Central PMCID: PMCPMC5172412.

19. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, O'Meara MJ, Rezelj VV, Guo JZ, Swaney DL, Tummino TA, Hüttenhain R, Kaake RM, Richards AL, Tutuncuoglu B, Foussard H, Batra J, Haas K, Modak M, Kim M, Haas P, Polacco B, Braberg H, Fabius JM, Eckhardt M, Soucheray M, Bennett MJ, Cakir M, McGregor MJ, Li Q, Meyer B, Roesch F, Vallet T, Mac Kain A, Miorin L, Moreno E, Naing ZZ, Zhou Y, Peng S, Shi Y, Zhang Z, Shen W, Kirby IT, Melnyk JE, Chorba JS, Lou K, Dai SA, Barrio-Hernandez I, Memon D, Hernandez-Armenta C, Lyu J, Mathy CJ, Percia T, Pilla KB, Ganesan SJ, Saltzberg DJ, Rakesh R, Liu X, Rosenthal SB, Calviello L, Venkataramanan S, Liboy-Lugo J, Lin Y, Huang X, Liu Y, Wankowicz SA, Bohn M, Safari M, Ugur FS, Koh C, Savar NS, Tran QD, Shengjuler D, Fletcher SJ, O'Neal MC, Cai Y, Chang JC, Broadhurst DJ, Klippsten S, Sharp PP, Wenzell NA, Kuzuoglu-Ozturk D, Wang H, Trenker R, Young JM, Cavero DA, Hiatt J, Roth TL, Rathore U, Subramanian A, Noack J, Hubert M, Stroud RM, Frankel AD, Rosenberg OS,

Verba KA, Agard DA, Ott M, Emerman M, Jura N, Von Zastrow M, Verdin E, Ashworth A, Schwartz O, D'Enfert C, Mukherjee S, Jacobson M, Malik HS, Fujimori DG, Ideker T, Craik CS, Floor SN, Fraser JS, Gross JD, Sali A, Roth BL, Ruggero D, Taunton J, Kortemme T, Beltrao P, Vignuzzi M, Garcia-Sastre A, Shokat KM, Shoichet BK, Krogan NJ. SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature. 2020;583(7816):459-68. Epub 2020/05/01. doi: 10.1038/s41586-020-2286-9. PubMed PMID: 32353859; PubMed Central PMCID: PMCPMC7431030.

20. Wilson L, Gage P, Ewart G. Hexamethylene amiloride blocks E protein ion channels and inhibits coronavirus replication. Virology. 2006;353(2):294-306. Epub 2006/07/04. doi: 10.1016/j.virol.2006.05.028. PubMed PMID: 16815524; PubMed Central PMCID: PMCPMC7111787.

21. Toto A, Ma S, Malagrino 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. 2020;29(10):2038-42. Epub 2020/08/22. doi: 10.1002/pro.3936. PubMed PMID: 32822073; PubMed Central PMCID: PMCPMC7461438.

22. Choma C, Gratkowski H, Lear JD, DeGrado WF. Asparagine-mediated self-association of a model transmembrane helix. Nat Struct Biol. 2000;7(2):161-6. Epub 2000/02/03. doi: 10.1038/72440. PubMed PMID: 10655620.

23. Wei Y, Thyparambil AA, Latour RA. Protein helical structure determination using CD spectroscopy for solutions with strong background absorbance from 190 to 230nm. *Biochem Biophys Acta*. 2014;1844(12):2331-2337. doi:10.1016/j.bbapap.2014.10.001

24. Park SH, Siddiqi H, Castro DV, De Angelis A, Oom A, Stoneham C, Lewinski M, Clark A, Croker B, Carlin A, Guatelli J, Opella SJ. (2021) Interactions of SARS-CoV-2 envelope protein with amilorides correlate with antiviral activity. PLOS Pathogens. doi: 10.1371/journal.ppat.1009519

25. Parthasarathy K, Lu H, Surya W, Vararattanavech A, Pervushin K, Torres J. Expression and purification of coronavirus envelope proteins using a modified beta-barrel construct. Protein Expr Purif. 2012;85(1):133-41. Epub 2012/07/24. doi: 10.1016/j.pep.2012.07.005. PubMed PMID: 22819936; PubMed Central PMCID: PMCPMC7129850.

26. Takano T, Nakano K, Doki T, Hohdatsu T. Differential effects of viroporin inhibitors against feline infectious peritonitis virus serotypes I and II. Arch Virol. 2015;160(5):1163-70. Epub 2015/02/24. doi: 10.1007/s00705-015-2370-x. PubMed PMID: 25701212; PubMed Central PMCID: PMCPMC7086594.

27. Gratkowski H, Lear JD, DeGrado WF. Polar side chains drive the association of model transmembrane peptides. Proc Natl Acad Sci U S A. 2001;98(3):880-5. Epub 2001/02/07. doi: 10.1073/pnas.98.3.880. PubMed PMID: 11158564; PubMed Central PMCID: PMCPMC14678.

28. Wrapp D, Wang N, Corbett KS, Goldsmith J, Hsieh C, Abiona O, Graham B, McLellan J. (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 367:1260–1263. doi: 10.1126/science.abb2507

29. Johnson BA, Blevins RA. NMR View: A computer program for the visualization and analysis of NMR data. J Biomol NMR. 1994;4(5):603-14. Epub 1994/09/01. doi: 10.1007/BF00404272. PubMed PMID: 22911360.

30. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995;6(3):277-93. Epub 1995/11/01. doi: 10.1007/BF00197809. PubMed PMID: 8520220.

31. Cross TA, DiVerdi JA, Opella SJ. Strategy for nitrogen NMR analysis of biopolymers. Journal of the American Chemical Society. 1982;104(6):1759-61. doi: 10.1021/ja00370a062.

32. Sarkar M, Saha S (2020) Structural insight into the role of novel SARS-CoV-2 E protein: A potential target for vaccine development and other therapeutic strategies. PLOS ONE. doi: 10.1371/journal.pone.0237300