UNIVERSITY OF CALIFORNIA, SAN DIEGO

The NMR Structure and Oligomerization State of the Membrane Associated p7 Protein from Hepatitis C Virus

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

in

Chemistry

by

Lindsay A. Dawson

Committee in Charge:

Professor Stanley J. Opella, Chair
Professor John Guatelli
Professor Patricia A. Jennings
Professor Elizabeth A. Komives
Professor Mark H. Thiemens

2015

The Dissertation of Lindsay A. Dawson is approved, and it is acceptable in quality and form for publication on microfilm and electronically

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2015

DEDICATION

To my husband, Kyle for being an

endless support and to my mother, Diedre,

for her love of life, science, curiosity, and for inspiring

me to accomplish whatever I put my mind to do.

EPIGRAPH

Learn from yesterday, live for today, hope for tomorrow. The important thing is not to stop questioning.

-Albert Einstein

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

^{1}H	Proton
^{2}H	Deuterium
^{13}C	Carbon-13
^{15}N	Nitrogen-15
AMS	Ammonium sulfate
BS^{3}	bis[sulfosuccinimidyl] suberate
CNBr	Cyanogen bromide
CSA	Chemical shift anisotropy
CP	Cross polarization
DARR	Dipolar assisted rotational resonance
DC	Dipolar coupling
DHPC	1,2-dihexanoy-sn-glycero-3-phosphocholine
DMPC	1,2-dimyristoyl-sn-glycero-3-phosphocholine
DMSO	Dimethyl Sulfoxide
DPC	Dodecylphosphocholine
DSP	Dithiobis(succinimidyl propionate)
DSS	Disuccinimidyl surbrate
DTT	Dithiothreitol
E1	Envelope glycoprotein 1
E2	Envelope glycoprotein 2
EDTA	Ethylenediamine tetracetic acid
FPLC	Fast protein liquid chromatography

HCV	Hepatitis C Virus
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HETCOR	Heteronuclear correlation
HnNCa	^{1}H-^{15}N dipolar coupling / ^{15}N chemical shift / $^{13}C\alpha$ chemical shift
HnNCo	^{1}H-^{15}N dipolar coupling / ^{15}N chemical shift / ^{13}CO chemical shift
HPLC	High performance liquid chromatography
HSQC	Heteronuclear single quantum coherence
IGEPAL	*tert*-Octylphenoxy poly(oxyethylene)ethanol
INEPT	Insensitive nuclei enhanced polarization transfer
IPTG	Isopropyl β-thiogalactoside
kD	Kilo-Dalton
LB	Lysogeny broth
MAS	Magic angle spinning
MSP	Membrane Scaffolding Protein
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
NS2-NS5B	Nonstructural HCV proteins
NTA	Nickelnitrilotriacetic acid
NTR	Non-translated region
ORF	Open reading frame
PDB	Protein data bank
PDSD	Proton-driven spin diffusion

PFO	Perfluorooctanoic acid
PFO-PAGE	Perfluorooctanoic acid- polyacrylamide gel electrophoresis
PMSF	Phenylmethanesulfonylfluoride
PRE	Paramagnetic Relaxation Enhancement
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SLF	Separated local field
SPECIFIC	Spectrally induced filtering in combination with *cross polarization*
RDC	Residual dipolar coupling
RMSD	Root mean squared deviation
TFE	2,2,2-Trifluoroethanol
TFA	Trifluoroacetic acid
TM	Transmembrane
TX-100	Triton X-100

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to my thesis advisor, Professor Stanley Opella for giving me the opportunity to conduct research in his laboratory. This graduate thesis work would not be possible without his guidance and the laboratory research environment that is intellectually challenging. I would also like to thank other members in my thesis committee, Professor Patricia Jennings, Professor Betsy Komives, Professor Mark Thiemens and Professor John Guatelli, for their guidance, supportiveness and openness to the discussion of scientific questions and life itself.

I would like to give my thanks to the members who have come in and out of the Opella research group over the past five years. They have given me endless guidance and support both professionally and personally. Special acknowledgement must be given to Dr. Gabriel Cook for not only accepting me as a member of the p7 project but also for teaching me invaluable information and practically everything I know about membrane protein purification and spectroscopic techniques. I am also deeply indebted to his optimism and for teaching me to have a more positive outlook in research and in life. I am eternally grateful for your help in shaping me into the scientist I am today. I also value the friendship and mentorship you've shown me outside of lab. Other members of the lab that have taught me invaluable life and laboratory lessons include Dr. Henry Nothnagel, who provided me with stimulating discussions and showed me various biochemical techniques, Dr. Fabio Casagrande, Dr. Dongtao Cui who both gave me priceless words of wisdom on several occasions. To Dr. George Lu who I greatly admire as one of the most dedicated researchers and with whom I have had invaluable scientific

and life debates with on several occasions. A special thank you to Dr. Hua Zhang and Dr. Eugene Lin who both assisted me in developing my research projects and helped me in pursing results when I was unsure of where to turn for information and interpretation. I would like to thank Dr. Ye Tian for assisting in the calculation of p7 structure and showing me some computational techniques. I would like to deeply thank Dr. Anna De Angelis for her guidance in science and life and for including me in the service outreach projects the NMR resource provide to the greater San Diego area. Thank you to Dr. Sang Ho Park, Dr. Anna Pavlova, Sabrina Berkamp, Mignon Chu, Zheng Long, Jasmina Radiocic, Vivian Wang and Mitchell Zhao have shared their immeasurable support and wet lab experiences and biochemical knowledge with me on a daily base. Dr. Xuemei Huang and Dr. Anthony Mrse have provided important training about solution NMR spectroscopy and many stimulating discussions about science. I would also like to thank Dr. Bibhuti Das for teaching me a tremendous amount of knowledge about NMR spectroscopy. Dr. Chin Wu, Dr. Chris Grant at the Center for NMR Spectroscopy and Imaging of Proteins, "The Bubble", have also taught me many solid-state NMR techniques and theories, shared and encouraged my study on spectroscopy, and provided important assistance in solving practical instrumentation problem whenever needed. Last but not the least, I would like to thank Elena Vitoshka-tarasov for her AMAZING job as the administrative assistant of the laboratory and life, assisting me from time to time on fulfilling various UCSD administrative processes.

I would like to thank Professor Jeanne Hardy who was my undergraduate research advisor and mentor. She opened the door of my scientific career towards structural biology and gave me the confidence to pursue a doctorate degree.

Chapter 2, in full, is a reprint of the material as it appears in Biochemistry, "Three-Dimensional Structure and Interaction Studies of Hepatitis C Virus p7 in 1,2-Dihexanoyl-*sn*-glycero-3-phosphocholine by Solution Nuclear Magnetic Resonance" by G.A. Cook, Dawson, LA, Tian Y, and Opella SJ, 2013. The thesis author was the secondary author of the paper.

Chapter 3, in part is currently being prepared for submission for publication of the material. G.A. Cook & L.A. Dawson, Das, B. B., Tian, Y., and Opella SJ. The thesis author was the co-first author and investigator of this material.

VITA

2010	Bachelor of Science, Chemistry, University of Massachusetts, Amherst
2010-2012	Teaching Assistant, Department of Chemistry and Biochemistry, University of California, San Diego
2012	Master of Science, Chemistry, University of California, San Diego
2015	Doctor of Philosophy, Chemistry, University of California, San Diego

PUBLICATIONS

GA Cook, Dawson LA, Tian Y, Opella SJ (2013) The Three Dimensional Structure and Interaction Studies of Hepatitis C Virus in 1,2-Dihexanoyl-*sn*-glycero-3-phosphocholine by Solution Nuclear Magnetic Resonance. *Biochem.* 52(31):5295-5303.

ABSTRACT OF THE DISSERTATION

The NMR Structure and Oligomerization State of the Membrane Associated p7 Protein from Hepatitis C Virus

by

Lindsay A. Dawson

Doctor of Philosophy in Chemistry

University of California, San Diego, 2015

Professor Stanley J. Opella, Chair

The Hepatitis C viral (HCV) protein p7 has been of interest in the HCV community as a potential drug target for the treatment of hepatitis for over two decades. p7 is considered to be a viroporin, shown to oligomerize and induce channel activity, allowing channel-blocking compounds to target the porin as an effective antiviral treatment plan. High-resolution structural data is absolutely vital to a rational drug design. This includes determining not only preferred oligomerization state but also

determining a biologically relevant, high-resolution structure of p7. The backbone structure of p7 was determined in detergent micelles using solution NMR and conveys condensed structure due to the limited breadth of the membrane environment in detergent micelles. Here, both the solution and the solid-state monomeric structure of p7 coupled with the elucidation of the oligomerization state are the focus of our studies. The solution structure was determined with the help of mutational studies to dampen dynamics of unresolved regions of the protein structure. A single point mutation of an absolutely conserved residue, W48A, is used to help refine the solution structure. Additionally, interaction studies are observed for the transmembrane region of p7 and NS2, a nonstructural p7-interactting HCV protein. Drug-binding studies are conducted to map specific residues involved in binding known channel-blocking compounds, amantadine, N*N*-DNJ and HMA. The solid-state backbone structure is determined for p7 reconstituted in proteoliposomes using rotationally aligned (RA) solid-state NMR. Various methods of chemical cross-linking are carried out to establish an optimal method for determining oligomerization state. p7 is cross-linked in proteoliposomes using an amine-reactive cross-linker and products were visualized using silver-stained SDS-PAGE. A modified native PAGE technique using perfluorooctanoic acid (PFO) is applied to study the oligomerization state in a mild detergent native-like state. Unmodified p7 is incorporated into nanodiscs and a light sensitive chemical cross-linker is used to characterize preferred oligomerization state. In addition to these studies, the expression and purification of a separate genotype, 2a, subtype JFH-1 p7 is optimized for a structural comparison to J4 subtype.

Chapter 1: Introduction NMR and Hepatitis C Virus Protein p7

1.1 Introduction to Membrane Proteins Structural Studies

Membrane proteins comprise approximately one third of the prokaryotic and eukaryotic genomes of an organism and represent approximately 60% of approved drug targets. [1, 2] Their functions are vast and vital for cellular function and stability. They comprise integral members of the lipid bilayer that lead to structural integrity of the cell and commonly poses cell surface functions, which play essential roles in signal transduction. Additionally, they help to regulate and control the flow of ions and solute material and often possess enzymatic activity. Many viral genomes are comprised of membrane proteins that are directly involved in virus particle production and virus propagation. Moreover, defects in our own membrane proteins are affiliated with countless diseases, directly relating to the 60% of all drug targets. In order to design new compounds that are highly specified for these proteins, it becomes essential to not only understand the three-dimensional structure but also be able to interpret and predict their functions from the structure.

The ~25-30% portion that membrane proteins make up of our genome translates to approximately 6,000 integral membrane proteins. The first high-resolution (3 Å) structure of a membrane protein *Rhodopseudomonas viridis*, a photosynthetic reaction center, was determined in 1985 by Deisenhofer and coworkers and paved the way for the atomic-level interpretation of biophysical data for structure determination and confirmed the existence of transmembrane (TM) α-helices. Currently, there are just over 500

UNIQUE (defined according to membrane protein database) structures deposited into the Protein Data Bank (PDB) according to the membrane protein database (http://blanco.biomol.uci.edu/mpstruc/#tableDescription). This value is slightly lower than the theoretically predicted number of membrane protein structure determination based on soluble structure determination prediction. This, in large part, is due to the substantial difficulties inherent to membrane protein expression, purification, reconstitution and collection of highly resolved structural data. This presents a clear need to continue to develop new technologies for sample preparation and structure determination of membrane proteins.

1.2 Solution and Solid-State NMR

Various methods have been developed over the years for the application and study of structural biology, which aims to understand the biochemical makeup and functional mechanism of a biomolecule at the molecular level. These methods provide an in-depth understanding of atomic positioning and dynamics within the structure of the interest. The primary methods used today include X-Ray crystallography, electron microscopy (EM) and Nuclear magnetic resonance (NMR). Whereas both X-ray and EM make use of electrons either through the use of x-ray scattering or an accelerated electron beams, NMR relies on the nuclear interactions by examining their quantum spin states used to derive structural coordinates and dynamics of individual atoms.

NMR itself can be further divided into two significant disciplines: solution or liquid-state NMR and solid-state NMR. Solution NMR typically provides highly resolved

spectra as the molecules being studied undergo fast isotropic tumbling. This tumbling, known as the correlation time or τ_c, is dependent on the size of the molecule. The larger the molecule, the longer the correlation time becomes larger which leads to significant broadening of the observed line widths in the spectra. This limitation leads to a pitfall of solution NMR wherein proteins above 30 kDa become exceeding difficult to obtain resolved structural data on. Additionally, when a protein of a significantly large protein can lead to spectral crowding, making it difficult to isolate specific amino acid chemical shifts.

Solid-state NMR has nearly no size restriction with the exception of larger proteins that over-crowd the spectrum. However, in order to attain the same resolution in solid-state as observed in solution, high-powered pulses are required to remove undesired spin interactions that lead to line broadening. Additional measures need to be taken to attain high-resolution signals from these samples. These manipulations can be mechanical or intrinsic to the sample itself and include fast spinning at the magic angle or macroscopic sample alignment coupled with decoupling/recoupling pulse sequences.

Specifically, in the application of NMR to membrane proteins, the size limitation of solution NMR becomes more apparent as the proteins require solubilization in a membrane mimetic environment. This environment adds to the size of the protein, which in turn effects the correlation time driving the use of detergent micelles when studying membrane proteins in solution NMR. This issue with the use of these detergent micelles arises from their lack of resemblance to that of a biological lipid bilayer. Several examples of membrane proteins have been analyzed and shown to contain significant

distortions when the micelle structures are compared directly to the same structure determined in lipid bilayers.

In order to structurally characterize these proteins in their native environment we must obtain the necessary structural restraints in solid-state NMR, as the complexes are too large to undergo the fast isotropic tumbling for solution NMR. [3-7] When examining membrane proteins in the solid-state they are reconstituted in a proteoliposomes, or bilayer environments. In the proteoliposome, the proteins are positioned in every possible direction and are said to be unoriented. These samples give extremely broad signals in both chemical shift and heteronuclear dipolar couplings dimensions called powder patterns. [3-11] Powder samples are observed to be composed of millions of crystals with different orientations that result in a broad powder-pattern line shapes with characteristic features. Each nuclei possesses its own magnetic dipole moment within a small localized magnetic field that interacts with dipole moments of proximal nuclei. This interaction is a through space interaction known as dipolar coupling which serve as the bases for correlation spectra and distance measurements. [12]

Three major groups of obtained structural restraints are necessary for a solid-state NMR, restraint-guided molecular dynamic (MD) structural calculation. These include distance restraints or the distance between two nuclei, dihedral angle or the angle between adjacent chemical bonds (derived from isotropic chemical shift) and angular restrains or the angle between a chemical bond and a molecular or external axis. Our lab has focused on obtaining angular restraints for structure determination, as the distance restraints can be extremely difficult, if not impossible to deconvolute for helical membrane proteins. Additionally, dipolar couplings are crucial for determining a high-

resolution structure. Finally, long-range dipolar couplings that occur between non-bonded atoms provide crucial inter-helical distances; however, they are difficult to observe. This highlights the importance of obtaining the short-range dipolar couplings between bonded atoms, as they are the largest contributor to the angular restraints used in determining structure.

It should also be mentioned that both ^{1}H-^{15}N and ^{1}H-^{13}C possess exceedingly similar powder patterns as a result of not only powder averaging but also because they have highly similar bond lengths. It is nearly impossible to obtain structural information from these powder patterns, which resulted in the development of a number of different strategies to overcome these limitations. One widespread technique is to mechanically rotate the sample at fast spinning speeds to averages, over time, the anisotropic interactions of nuclei with spin 1/2 as well as the heteronuclear dipolar interactions to zero while preserving the isotropic, observable interactions. In order to satisfy these conditions, the fast spinning must occur at an angle close to 54.7° with respect to the magnetic field, known as the magic angle.

In addition to this mechanical rotation, we also rely on a sample intrinsic property of taking advantage of the proteins fast rotational diffusion (~10^{6} s) about the bilayer normal [10]. This allows for both the chemical shift tensors and the dipolar coupling perpendicular to the bilayer normal to average and reduce the broad powder pattern to narrow signals. These averaged powder patterns are completely dependent on the bilayer normal so we can obtain angular restraints for the protein undergoing the fast rotational diffusion. This methodology has appropriately been named rotationally aligned (RA) solid-state NMR and is the methodology we use to solve physiologically relevant

structures of membrane proteins. Finally, a series of rotor synchronized radio frequency pulse sequences designed to reintroduce the chemical shift anisotropy and the heteronuclear dipolar couplings are applied to allow us to measure chemical shifts for our assignments, obtained using the NCOCX and NCACX experiments along with the Dipolar assisted assignment protocol (DAAP) [13].

Taken together, these factors highlight how solid-state NMR is most suitable method for high-resolution, three-dimensional structure determination of unmodified, non-engineered, and full-length membrane proteins being in near native conditions with respect to phospholipid bilayers, pH and temperature.

1.3 Biological Significance of Hepatitis C Viral Protein, p7

Hepatitis C virus, first identified in 1989, infects approximately 80 million people worldwide. [14] The virus is a member of the *Hepacivirus* genus within the Flaviviridae family. The viral genome is a positive 9.6 kb single strand open reading frame (ORF) RNA encoding a 3000 amino acid polyprotein co- and post-translationally processed by cellular and viral proteases to produce ten structural and nonstructural proteins. The structural proteins consist of E1 and E2 or enveloped glycoproteins that are targets of host antibody response. The core protein, viral nucloeocapsid, interacts with progeny viral genomes for the assembly of virus particles. Nonstructural proteins assist in forming complexes with viral RNA in initiate viral replication in cytoplasmic membranous structure, endoplasmic reticulum (ER) or the golgi apparatus. [15] Finally, the protein separating the structural and nonstructural protein, our protein of interest for these studies is the small 63-residue two TM protein, p7.

There are several known HCV genotypes with well over 100 subtypes. [14] High replicative activity coupled with the lack of a proofreading action of the viral RNA-dependent RNA polymerase (NS5A) provides the basis for this high genetic variability of HCV. These properties, similar to that of HIV, warrant a strong rationale for the development of antiviral combination therapies. HCV virion particles typically target hepatocytes; however, B cells, dendritic cells and other cell types have been shown to be susceptible to infection. CD81, a tetraspanin protein found on most cell type surfaces as well as scavenger receptor class B type I (SR-BI), LDL receptor (LDLR) and claudin-1 have all been proposed to work in concert as receptors for HCV virus particles. Both CD81 and SR-B1 are required but not sufficient for HCV entry as CD81 and SR-B1 expressed in non-hepatocyte derived cell lines does not confer susceptibility to HCV infection. Additional hepatocyte-specific factors are required for HCV entry. HCV enters via a clathrin-mediated endocytosis with transit through endosomal membrane fusion.

The Hepatitis C viral (HCV) protein p7 has been of interest in the HCV community as a potential drug target for the treatment of hepatitis for the past two decades. p7 is composed of 63 amino acid residues with two TM helical regions with a fully conserved di-basic loop motif. [16] It is predicted to serve two main functions: oligomerize and form either hexa- or heptameric pores[16-19] capable modifying membrane permeability by inducing cation channel activity [20], giving it key characteristics of a viroporin and to engage in functional protein-protein interactions. P7 is involved in interacting with NS2 to recruit core to the ER, or the site of capsid assembly.[21-23] [24] The cation permeability is affiliated as a pH modulation of intracellular vesicles during virus packaging and egress. [25]

The channel activity and specifically the proton permeation ability has been verified by several labs *in vitro* using lipid bilayer/proton flux assays as well as in *in vivo* using Xenopus oocytes expressing p7.[26] The channel activity has been shown to be inhibited by known channel blocking compounds rimantadine, hexamethyl amiloride, and immunosugar derivative N*N*-DNJ [20, 27-29] to varying degrees of inhibition based on the different genotypes tested for each compound.[30] Moreover, several solution NMR structures [31, 32] have recently been published including two solution-state monomers and one oligomeric structure. [26] Each structure was solved in a different solubilized medium including 100% methanol, 90% 2,2,2-trifluoro- ethanol (TFE) mixed with asolectin in chloroform/methanol and dodecylphosphocholine (DPC). Each of these structures were solved using a different genotype, J4, HCV-J and EUH1480, respectively and maintain three or four transmembrane α-helical regions each of which present wildly different structures.

P7 is vital to the production of infectious virus particle as Sakai and coworkers [33] showed mutation of the conserved di-basic loop motif residues or truncation of the first TM region renders the virus inactive. It is not, however, essential for the replication of the viral RNA genome. It has been shown to localize to the ER and oligomerize to varying degrees depending on the examined genotype. [17, 19, 34] The oligomerization propensity of this protein has also made it extremely difficult to solve a biologically relevant structure of the protein. To help elucidate both the oligomeric state and the structure in this work, different lipid environments, cross-linking techniques and biophysical characterizations are applied to the protein in biologically relevant conditions.

Largely due to a non-standard approach to structural characterization of p7, different protein modifications and lipid environments have also led to the difference in observed oligomeric complexes. Recently, Chandler et al. used a monomeric structure elucidated by Montserret et al. to construct oligomeric models with any where from four to seven subunits evaluated through the use of MD simulations in a POPC bilayer. These experiments were run to not only to determine the optimal number of monomers but also identify the role of certain pore-lining residues in channel gating. Finally, they evaluated the adaptability of the structure to different lipid environments. They determined there is a certain level of compatibility with the co-existence of multiple oligomeric states and that the monomeric structure of p7 itself is flexible in that it is capable of bending to fit whatever lipid environment it is placed in. Additionally Whitfield et al. showed how the lipid environment directly impacts the ion-channel activity of p7 by impacting the length and size of the channel opening based on the type of lipid used. Chandler et al. identify hexa- or heptameric complexes to be the most energetically favorable.

The high-resolution structural data is absolutely vital to a rational drug design. This includes determining not only the preferred oligomerization state but also determining the most biologically relevant model of the p7 monomeric structure. By combining the oligomeric state with a refined monomeric structure, we can obtain a relevant model that may be used to screen compounds capable of blocking the p7 function. The atomic-resolution structure of p7 is an ideal candidate to solve using solid-state Nuclear Magnetic Resonance (NMR) spectroscopy, since we have the unique advantage of maintaining a physiologically relevant environment as the protein resides in a near-native lipid bilayer at physiological pH. These conditions minimize distortions in

the structure that can be caused by non-native environments. The backbone structure of p7 was determined in detergent micelles using solution NMR and conveys a compacted structure due to the limited breadth of the membrane environment in detergent micelles. Here, both the solution and the solid-state monomeric structure of p7 coupled with the elucidation of the oligomerization state are the focus of our studies.

The solution structure was determined with the help of mutational studies to dampen the dynamics of unresolved regions of the protein structure. A single point mutation of an absolutely conserved residue, W48A, is used to help refine the solution structure. Additionally, interaction studies are observed for the TM region of p7 and NS2, a nonstructural HCV protein thought to interact with p7 to recruit HCV core protein to the ER, the site where viral packaging occurs. Drug-binding studies are conducted to map specific residues involved in binding known channel-blocking compounds, amantadine, N*N*-DNJ and HMA. The solid-state backbone structure is determined by reconstituting p7 into proteoliposomes using rotationally aligned (RA) solid-state NMR from which high-sensitivity and high-resolution spectra are obtained. Chemical shifts for ^{15}N and ^{13}C as well as ^{1}H-^{15}N, ^{1}H-^{13}C dipolar couplings are measured and converted to appropriate structural restraints.

Various methods of chemical cross-linking are carried out to establish an optimal method for determining oligomerization state. P7 is cross-linked in proteoliposomes using an amine-reactive cross-linker and cross-linked species were visualized using silver-stained SDS-PAGE. A modified native PAGE technique using perfluorooctanoic acid (PFO) is applied to study the oligomerization state in a mild detergent native-like state. Unmodified p7 is incorporated into nanodiscs and a light sensitive chemical cross-

linker is used to characterize preferred oligomerization state of p7. In addition to this study, the expression and purification of a separate genotype, 2a, subtype JFH-1 p7 is optimized to structurally compare to the J4 subtype.

Chapter 2. The Three Dimensional Structure of Hepatitis C Virus p7 by Solution NMR

2.1 Abstract

Hepatitis C Virus (HCV) protein p7 plays an important role in the assembly and release of mature virus particles. This small 63-residue membrane protein has been shown to induce channel activity, which may contribute to its functions. p7 is highly conserved throughout the entire range of HCV genotypes, contributing to making p7 a potential target for anti-viral drugs. Here we describe the three-dimensional structure of p7 in short chain phospholipid (DHPC) micelles, which provide a reasonably effective membrane-mimicking environment that is compatible with solution NMR experiments. Using a combination of chemical shifts and residual dipolar couplings we determined the structure of p7 using an implicit membrane potential combining both CS-Rosetta decoys and Xplor-NIH refinement. The final set of structures has a backbone RMSD of 2.18 Å. Molecular dynamic simulations in NAMD indicate that several side chain interactions might be taking place, and that these could affect the dynamics of the protein. In addition to probing the dynamics of p7, several drug-protein and protein-protein interactions were evaluated. Established channel-blocking compounds such as amantadine, hexamethylene amiloride (HMA), and long alkyl-chain iminosugar derivatives inhibit the ion channel activity of p7. It has also been shown that the protein interacts with the HCV non-structural protein 2 (NS2) at the endoplasmic reticulum, and that this interaction may be important for the infectivity of the virus. Changes in the chemical shift frequencies of solution NMR spectra identify the residues taking part in these interactions.

2.2 Introduction

Hepatitis C virus (HCV) infects over 150 million humans, and causes over 350,000 deaths per year (World Health Organization 2012, www.who.int). It is a member of the *Hepacivirus* genus within the Flaviviridae family[35]. Currently, no vaccines exist for HCV and the treatment regime of interferon-α and ribavirin is poorly tolerated, costly, and only effective in approximately 50% of patients. Recent trials of drugs that inhibit viral enzymes have been disappointing because of the rapid development of resistance by the virus[28, 36, 37]. Due to the high variability in HCV genotypes, new treatments in all likelihood will require combinations drugs targeted to multiple viral proteins in order to minimize the development of resistance. The membrane protein described here, p7, provides a novel target that may function in such a setting.

HCV is an enveloped virus with a positive single-strand of RNA genome of approximately 9.6 kb, which is translated into a 3000-amino acid polyprotein that is cleaved by intracellular and viral proteases to yield 10 mature proteins[33, 38]. These proteins are divided into two classes, structural and nonstructural (NS) proteins. The structural proteins core and envelope glycoproteins, E1 and E2, comprise the virion. The nonstructural proteins NS2, NS3, NS4A, NS4B, NS5A and NS5B modulate host metabolism and replication of viral RNA. Notably, the nucleotide sequence for a small 63-amino acid hydrophobic transmembrane protein, p7, is located between the sequences that code for the structural and the nonstructural proteins[38].

p7 has previously been shown to be essential for efficient virus particle assembly and release, but not required for RNA replication[33]. p7 is localized to the endoplasmic reticulum. Under the conditions used for electron microscopy the protein appears to

oligomerize as heptamers or hexamers[39, 40]. However, SDS gels and narrow NMR linewidths are consistent with the protein being a monomer in micelles. The oligomers may account for the apparent channel activity of p7, which has been demonstrated in phospholipid bilayers[18, 41]. However, the role of the ion channel activity in the viral lifecycle remains uncertain, it can be blocked by a variety of compounds including amantadine, hexamethylene amiloride (HMA), and a number of iminosugar derivatives demonstrating the oligomeric form is a potential drug target[18, 20, 41-43].

The overall structure of p7 consists of two trans-membrane (TM) segments, TM1 and TM2, connected by a short, conserved inter-helical loop[44, 45]. However, the structure is more complex than this. As indicated in Figure 1C, residual dipolar couplings (RDCs) have identified seven distinct structural elements in the protein including four helical regions that constitute the two trans-membrane segments[46]. The first segment (TM1) spans from residue 6 to residue 26, and the second segment (TM2) from residue 38 to residue 58. The protein's termini face the lumen of the ER with the inter-helical loop region protruding into the cytosol[44, 45]. Previous NMR studies have also shown that the helices are tilted about 10° and 25° from the bilayer normal[47]. The magnitudes of the RDCs also revealed the presence of substantial backbone motions in the first and third helical regions of the protein. In particular, tryptophan 48 and proline 49 appear to be highly dynamic. Interestingly this tryptophan residue is strictly conserved throughout the genotypes of p7.

Recent results suggest that p7 is critical for functions in virus assembly unrelated to its channel activity[48]. Furthermore, p7 may act in unison with the HCV proteins E1, E2 and NS2, suggesting that its biological functions and possibly its ion channel activity are

regulated by specific protein-protein interactions[49]. A goal of our structural studies is to determine how specific protein-protein interactions, such as between p7 and NS2, impact channel activity, virion assembly, and the release of viral progeny.

It is important to acknowledge the uncertainties associated with studying membrane proteins in any membrane-mimicking environment other than phospholipid bilayers. However, we have extensive experience with these types of samples[50], and have found that the short chain phospholipid, DHPC, is a generally good choice as a minimally perturbing detergent to solubilize p7 for solution NMR spectroscopy. Except for the length of its hydrocarbon chains, its chemical properties are identical to those of the phospholipids that assemble into bilayers, in particular it has two hydrophobic chains connected to a phosphocholine headgroup.

RDCs are highly reliable sources of structural information about helical membrane proteins because they report on the regular patterns of secondary structure as well as the backbone dynamics of the protein. Moreover, RDCs are important because there are major limitations in measuring 'long-range' nuclear Overhauser enhancements (NOEs) among hydrogens in helical membrane proteins in micelles[51]. Tryptophan 48 was investigated not only due to its anomalous location, tryptophan residues are typically located at the interface between headgroups and hydrophobic residues of phospholipids, but also to follow-up on some preliminary calculation-based evidence that the dynamics of the third helical region was attenuated by mutating this to a non-aromatic residue. Additionally, other interactions of this residue were investigated. Specifically, the interactions of tryptophan 48 with the proximal tyrosine residues 31, 42 and 45 were examined.

Additional structural characterization of p7 was obtained through the observation of paramagnetic relaxation effects (PRE) from manganese ions in solution. This allowed for the identification of, with some degree of certainty, the residues that are accessible to the aqueous solution. The experiments also provided information about the effect of binding by the channel blocking compounds amantadine, iminosugar derivative *N*-nonyl-deoxynojirimycin (*N*N-DNJ), and 5-(N,N-Hexamethylene)amiloride (HMA). Previously, Griffin et al. demonstrated that amantadine could block p7 ion channel activity *in vitro*[42]. Steinmann et al. discovered that the effects of amantadine *in vivo* were associated with specific genotypes. They also showed that deoxynojirimycin (DNJ)-containing iminosugars, such as *N*N-DNJ, had inhibitory affects *in vitro*[28]. These effects not only interfered with HCV entry into host cells but also significantly inhibited virus assembly and release. Premkumar et al. also demonstrated the ability of HMA to block p7 ion channel activity[20]. By observing the chemical shifts of individual residues it is possible to identify which residues in a protein are involved in binding, possibly assisting the design of future compounds that exploit these known interactions.

The non-structural protein 2 (NS2) interacts with several other HCV proteins including p7[48, 52]. The topology of the first TM of NS2 and location of these interactions were investigated using a combination of isotopically labeled and unlabeled trans-membrane domains of both p7 and NS2. This method provides a way of looking at the effect on each of the proteins without complicating the spectra. These experiments provide a starting point for examining the interactions of the full-length versions of both proteins.

2.3 Materials and Methods

2.3.1 Protein Expression and Purification (p7, p7W48A and NS2TM1)

Samples of p7 from genotype J4 were prepared as previously described[53]. The mutant W48A was prepared using a Quikchange Lightning Site-Directed Mutagenesis Kit (Agilent, www.agilent.com) and primers synthesized (Allele) to replace the tryptophan at position 48 with alanine. The two primers used were GCTGCTTACGCTTTCTACGGTGTTGCGCCGCTGCTGCTGCTGCTGC and GCAGCAGCAGCAGCAGCGGCGCAACACCGTAGAAAGCGTAAGCAGC. The codons in the mutation site are underlined. The DNA was transformed into XL10 Gold Competent cells (Agilent) and plated on LB agar. The mutation was confirmed by sequence analysis (Eton Biosciences, www.etonbio.com) of individual colonies. The DNA was purified from an overnight LB growth with the cells containing the correct sequence and transformed into BL21(DE3) (Agilent) cells for protein expression. The protein was expressed and purified the same way as the wild-type protein using nickel affinity chromatography, chemical cleavage from the fusion protein and reverse-phase HPLC (Waters, www.waters.com).

The same expression vector and purification protocol as used for p7 were applied to a construct corresponding to the first trans-membrane helix of the non-structural protein NS2 of HCV. NS2 TM1 was cloned using genotype J4, the same genotype that was used for p7. Oligonucleotides were purchased from Allele (www.allelebiotech.com) to clone the amino acid sequence LDREMAAASSGGAVLVGLVFLTLSPYYK. Forward primer CGGGGAAGCTTATGCTTGACCGTGAAGCTGCTGCTTCTTCTGGTGGTGCTGTT

CTGGTTGGTCTGGTTTTCCTGACCCTGTCTCCG and reverse primer CCCCGGATCCTCATCATTTTTTTTTTTTTTTTTGTAGTACGGAGACAGGGTCA GGAAAACCAGACCAACCAGAACAGCACCACC were elongated to obtain the insert product. The DNA was cut and ligated into the pHLV TrpLE vector using the restriction enzymes HindIII and BamHI (New England Bioloabs, www.neb.com). The insert included a codon for a methionine at the N-terminus of the protein for chemical cleavage and an additional 5 lysines at the C-terminus to assist with the solubility of the protein during purification and sample preparation. Milligram amounts of the labeled polypeptide were prepared for NMR experiments.

2.3.2 Sample preparation

Samples for solution NMR experiments were prepared as described previously[53]. Briefly, the purified protein, dried following HPLC purification, was dissolved in 400 mM DHPC (Avanti Polar Lipids Inc., www.avantilipids.com) and diluted to an aqueous solution with a final concentration of 125 mM DHPC. D_2O was added to 10% to provide a signal for the spectrometer lock, and the final protein concentration was 0.5 mM. The pH was adjusted to 4.0 for all samples to ensure NMR spectra were well resolved and directly comparable to those analyzed previously[53]. The protein sample was transferred to a standard 5mm tube for the NMR experiments.

For RDC measurements, the isotropic sample was transferred to a Shigemi (Shigemi Inc., www.shigeminmr.com) tube containing a dried 6% neutral polyacrylamide gel that was originally cast in a 3 mm NMR tube and cut to 3 cm in length. The plunger of the NMR tube was set to a length of 2.1 cm to allow the gel to absorb the isotropic

protein-containing sample in 125 mM DHPC. The compressed gel resulted in weak alignment of the protein in the NMR sample.

The samples for PRE measurements were prepared by adding chelated manganese to unaligned samples of p7-containing DHPC in aqueous solution. The chelated metal was prepared by adding manganese sulfate to ethylenediaminetetraacetic acid (EDTA) (0.5 M) at pH 8.0 and allowed sufficient time for a chelated precipitate to form. The precipitate was centrifuged and the pellet was washed twice with methanol and once with ethanol. The washed complex was dried by lyophilization and a stock solution of 200 mM was prepared in 1M HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid buffer. The solution was added to the NMR sample at a final Mn^{2+}-EDTA concentration of 10 mM. To make sure that the addition of the buffer did not affect the isotropic signals, a sample with just the buffer added was used as a control.

2.3.3 NMR Experimental Measurements

The NMR experiments were performed on a 600 MHz Bruker spectrometer (www.bruker-biospin.com) using a triple resonance cryoprobe (^{1}H, ^{13}C, and ^{15}N) with three-axis pulsed gradients. The chemical shift frequencies were referenced to the proton resonance of water set to 4.70 ppm. The experiments were performed at 50°C. Most of the ^{1}H-^{15}N correlation experiments were performed using the fast-HSQC pulse sequence[54].Typical experiments involved the acquisition of 1024 t2 points for 256 t1 increments.

The backbone amide J couplings were measured using a modified ^{1}H-^{15}N HSQC experiment based on the in-phase/anti-phase (IPAP) measurement of JD-splitting. The

RDCs were measured by taking the difference in JD coupling between an isotropic and a weakly aligned sample of p7-containing DHPC micelles.

The NMR data were processed using NMRPipe[55], and the figures were prepared using Sparky (T.D. Goddard and D.G. Kneller, SPARKY 3, University of California, San Francisco). RDCs were analyzed using Matlab (Mathworks, www.mathworks.com). A sliding window method was used to fit sinusoidal waves of a periodicity of 3.6 residues to the data to predict helical regions, as described previously[46].

2.3.4 Structure Calculations

A two-stage combined protocol[6, 9] was employed to calculate the structure of p7 in DHPC micelles. The determination of the structure is illustrated in Figure 2.2. In the first stage, 10,000 initial structures were generated in CS-Rosetta[56]. Fragment candidates were based on 58 HN, 61 N and 63 Cα chemical shift frequencies. All of the structures were relaxed under the influence of the of NH RDC data using Rosetta. In both steps, the implicit membrane potential flag was on in order to mimic a membrane-like environment. After relaxation, the 1,000 lowest energy structures were chosen from the most populated cluster, which was calculated from 10,000 structures. Forty percent of the 10,000 relaxed structures fell into the most populated cluster, while the second most favored cluster contained 22% of the structures. The lowest energy structure identified in the subset of 1,000 selected structures was used as the initial structural model of p7. Average dihedral angles and their deviations were calculated from the group of 1,000 selected low energy structures for further refinement using Xplor-NIH.

In the second stage, structure refinement was performed using a simulated annealing protocol with Xplor-NIH torsion angle molecular dynamics and the experimental restraints. The temperature was reduced from the initial value of 1000K to 50K in steps of 12.5 K. k_{ta} = 200 kcal•mol^{-1}•rad^{-2}, k_{rdc} was gradually increased from 2 kcal•mol^{-1}•rad^{-2} to 5 kcal•mol^{-1}•rad^{-2}. The calculation also includes the Xplor-NIH potential for knowledge-based torsion angles with ramped force constants of 0.002 kcal•mol^{-1}•rad^{-2} to 1 kcal•mol^{-1}•rad^{-2}. k_{vdw} and s_{vdw} were geometrically increased from 0.004 kcal•mol^{-1}•Å^{-4} to 4 kcal•mol^{-1}•Å^{-4}, and 0.9 kcal•mol^{-1}•Å^{-4} to 0.8 kcal•mol^{-1}•Å^{-4}, respectively. A total of 100 structures were calculated, and the 10 lowest energy structures were accepted for analysis.

2.3.5 Molecular Dynamics Simulations

1-Palmitoyl-2-oleoylphosphatidylcholine (POPC) bilayer generated and solvated with TIP3P waters in VMD[57] was equilibrated for 1 ns. The lowest energy structure from the Xplor-NIH refinement was selected. After embedding p7 into the *de novo* bilayer, the overlapped water and lipid molecules were removed. This resulted in 80 lipid molecules and 12669 water molecules. A 50 ns simulation was performed after 10000 steps of minimization at 310 K and a constant pressure in NAMD[58] with CHARMM27 force field.

2.3.6 Drug Binding Studies

Drug binding studies were performed using amantadine (Sigma Aldrich, www.sigmaaldrich.com), hexylmethyl amiloride (HMA, Sigma Aldrich) and N-nonyl-

deoxynojirimycin (*N*N-DNJ, Toronto Chemical Company, www.trc-canada.com). Experiments were performed by adding the appropriate compound, either in water or DMSO, at a concentration that was 20X greater (10mM) than that of the protein (0.5 mM). The pH of the sample was measured to ensure that the sample conditions were consistent with the control experiment. A ^{1}H-^{15}N HSQC experiment was performed using the same parameters as the control. The chemical shifts of the protein resonances were measured. Difference plots were made that utilized both ^{1}H and ^{15}N shifts using the equation $[(\Delta\delta H)^2+(\Delta\delta N/5)^2]^{1/2}$.

2.4 Results

2.4.1 Chemical Shifts and Residual Dipolar Couplings

The amide proton and nitrogen shifts were measured from the two-dimensional $^{1}H/^{15}N$ HSQC spectra, while the alpha carbon chemical shifts were measured from the three-dimensional HNCA spectra that were used to assign the protein resonances. In this experiment the magnetization is transferred from the amide proton to the amide nitrogen, and then to the alpha carbons of the same residue and the preceding residue. The magnetization pathway provides the chemical shifts of the carbon nuclei and assigns the backbone resonances. The dispersion of resonances in the HSQC spectra demonstrated that the protein is folded in the DHPC micelles, and a plot of the Cα chemical shifts as a function of residue number gives a preliminary indication of the secondary structure being composed of mostly α-helices based on the upfield chemical shifts from random coil values.

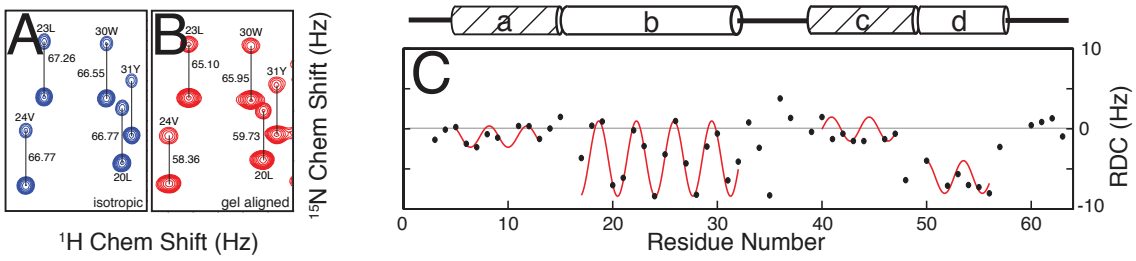

Figure 2.1: Measurement and analysis of RDCs performed on p7 in 1,2-dihexanoyl-sn-glycero-3-phosphocholine and aligned in a polyacrylamide gel.
A. & B. are spectra showing examples of measurements of RDCs using an in-phase/antiphase (IPAP) HSQC experiment on an isotropic sample (A) and a sample that is weakly aligned in a 6% charged and compressed polyacrylamide gel (B). C. Measured RDCs plotted as a function of residue number. Dipolar waves were fit to the four helical regions shown in the plot (a−d). RDCs for helices a and c are near 0, while those for helices b and d vary drastically, suggesting that the aforementioned are more dynamic.

As described under "Experimental Methods" and in previous publications, p7 could be weakly aligned in a compressed polyacrylamide gel for RDC measurements[59]. The residual dipolar couplings were measured for all of the backbone amide sites using an IPAP ^{1}H-^{15}N HSQC experiment on an isotropic sample and a weakly aligned gel sample (Figure 2.1A and B). The differences in the measured splittings, which ranged from +4 Hz to -8 Hz, were plotted as a function of residue number (Figure 2.1C). This was possible because all of the resonances have been assigned to specific backbone sites in the protein[46]. Sinusoidal waves were fit to the data using a periodicity of 3.6 residues per turn, appropriate for an α-helix, to identify those residues in helical regions of the protein[60]. Four separate waves resulted from the fitting procedure, as shown in Figure 2.1C. The amplitudes of the waves suggest that the helical segments labeled 'a' and 'c' have internal motions due to generally low values of residual dipolar couplings. Helices

'b' and 'd' appear to be fully structured on the NMR timescales. This result is in general agreement with the previous analysis of secondary structure of p7[47, 59].

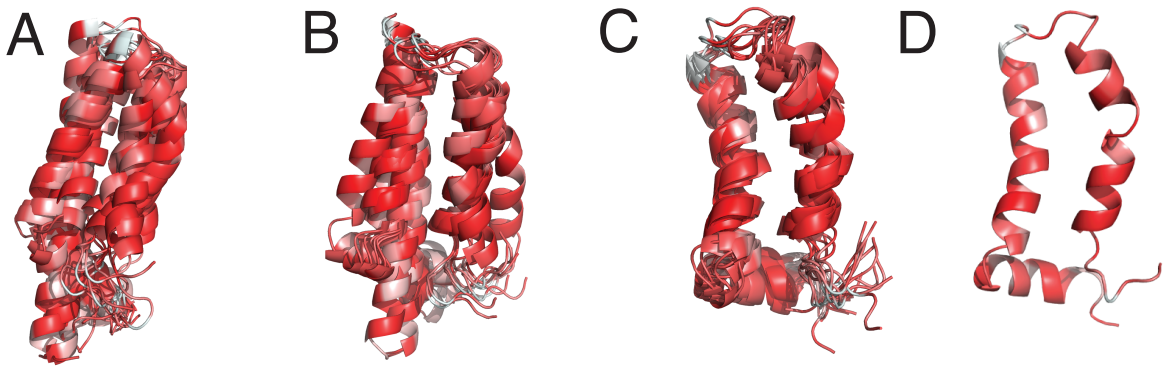

Figure 2.2: Characterization of the stepwise method for determining the structure of p7 from the measured NMR data
A. The chemical shifts (Cα, HN, and N) were used as the sole constraint in the CS-Rosetta calculation. The cluster shown consists of the 10 lowest-energy structures from 10000 decoys that were generated. B. Ten lowest- energy structures of 10000 relaxed initial structures in the presence of residual dipolar couplings as angular constraints. This set of structures converges over a larger portion of the backbone, including an overlay of helix a, which is a result of including orientation information provided by residual dipolar couplings. C. Ten lowest-energy structures calculated using the Xplor-NIH refinement protocol using the representative structures from the Rosetta calculation. The backbone rmsd for this ensemble is 2.18 Å. Although the rmsd is relatively low for a membrane protein, there is still variation in the mobile region, particularly the start of helix c. D. A single average structure from the Xplor-NIH refinement cluster was calculated.

2.4.2 Structure Calculations

For the structure calculations of p7, the chemical shift associated Molecular Fragment Replacement (MFR) method was used to obtain a starting structure that could be refined using RDC data as constraints. The initial set of structures calculated from Rosetta (Figure 2.2A) utilized the primary sequence of p7. The set shows convergence of the polypeptide backbone structure. An interesting feature, confirmed by the structure calculations, is that W48, a residue that is normally at the interface of the polar

headgroups and the hydrophobic sidechains of the lipids, is buried within the lipid bilayer in p7. When the RDC data were added as constraints to the calculation, the largest change was observed in the orientation of helical segment ‘a’ (Figure 2.2B). Consequently, the relative orientation of the two transmembrane segments changed. The refinement of these structures in Xplor-NIH using the same constraints caused the structures to converge with a RMSD of the backbone atoms of 2.18 Å (Figure 2.2C). An average structure from the cluster of 10 lowest energy structures was calculated (Figure 2.2D).

The average structure is shown in Figure 2.3 from three different perspectives. From these views the orientation of the four helices can be seen. Helix ‘a’ has a different orientation than that of helix ‘b’ and resides almost perpendicular to the membrane spanning segments. Helices ‘c’ and ‘d’ are clearly separated by several residues that include Trp48 and Pro49.

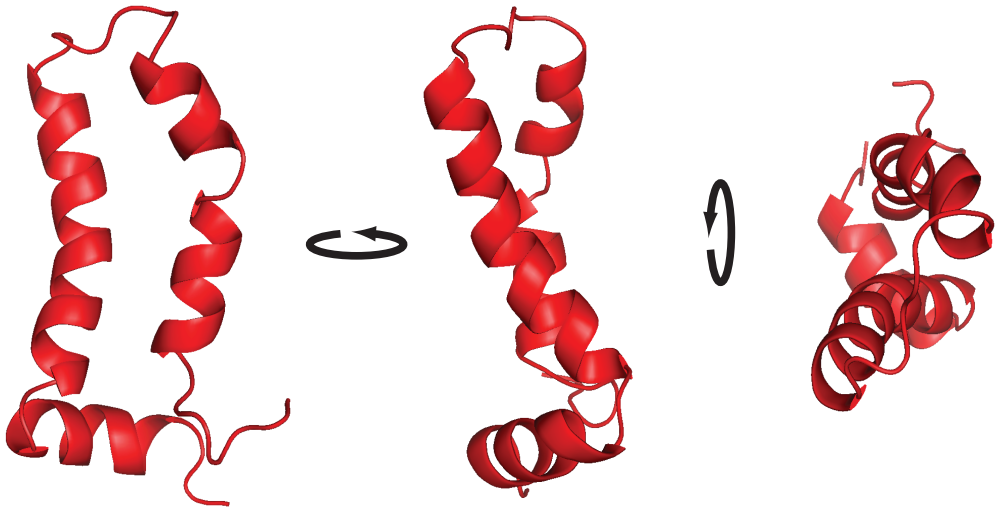

Figure 2.3: Average calculated structure of p7 from the Xplor-NIH refinement
The structure is the average structure from the lowest-energy cluster. The three views show the orientation of the four helices.

2.4.3 Manganese Data Compared with Structure

The PRE data[46] were plotted in the form of measured peak intensity as a function of residue number (Figure 2.4A). Resonance intensities that were reduced by at least 50% are highlighted on the structure of p7 (Figure 2.4B). The residues that are exposed to the aqueous solvent are strongly affected by the addition of the EDTA chelated paramagnetic metal. The intensities of the region between the first and second helical segments, 'a' and 'b' (residues 15-18), are reduced, suggesting that these residues are accessible to the chelated manganese in solution. This also suggests that the first helical segment is somewhat shielded from the effects of the chelated manganese within the polar head groups of the lipids. As illustrated in Figure 2.4, the loop region (residues 28-36) and the C-termini (residues 58-63) are also exposed to aqueous solution.

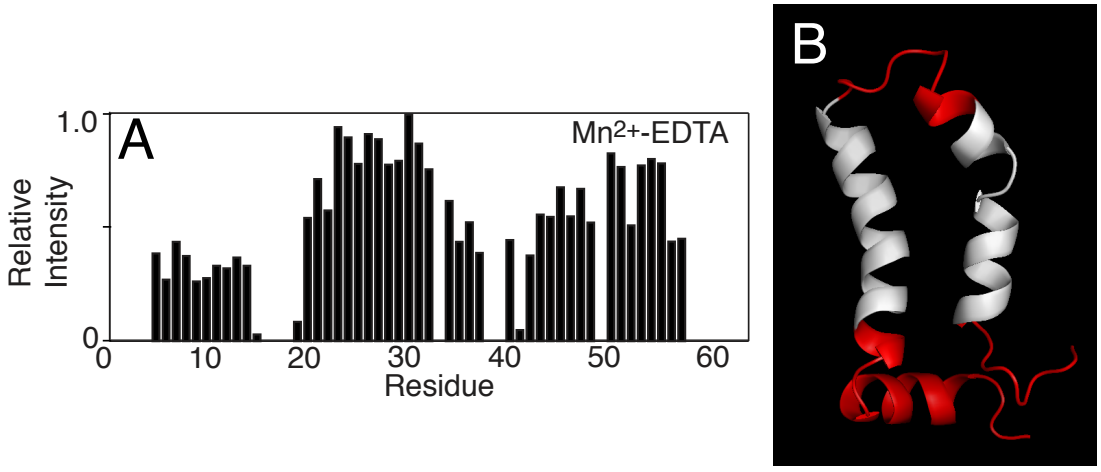

Figure 2.4: Paramagnetic relaxation effect of manganese on p7 characterized using differential signal intensities measured from 1H−15N HSQC spectra.
A. Spectra of p7 measured before and after the addition of Mn2+-EDTA. The relative intensity, or the ratio of the intensities, is plotted vs residue number. The signals with significantly reduced intensities indicate that the associated residues are solvent- exposed. B. Residues with a >50% reduction in signal intensity are colored red. From this representation, it is obvious which regions of the protein are buried within the hydrophobic region of the DHPC micelle.

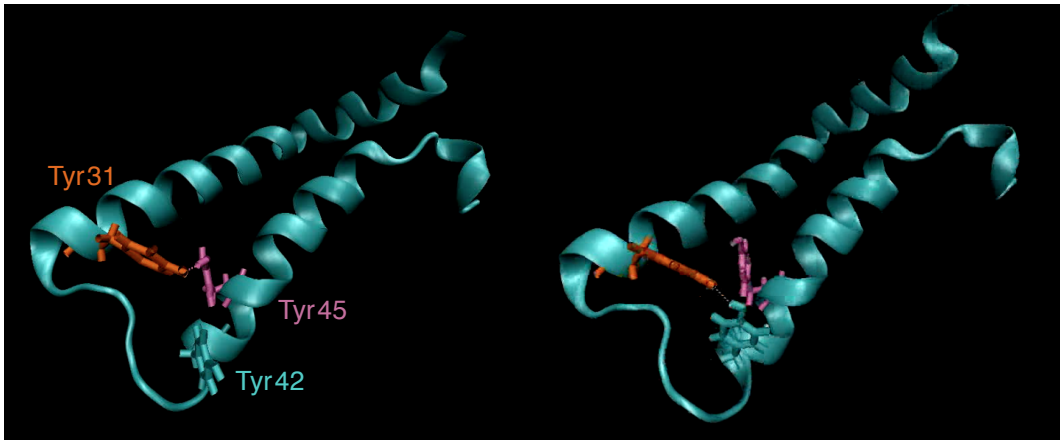

Figure 2.5: Simulations show evidence of side chain interactions of p7 in an implicit POPC bilayer.
Two "snapshots" were taken from a 50 ns MD simulation. The figure shows the hydrogen bonding that takes place between the side chains of tyrosine residues at the interface of the transmembrane segments, Y31 from TM1 and Y42 and Y45 from TM2. The structure on the left shows hydrogen bonding between Y31 and Y45, and the structure on the right shows that hydrogen bonding has switched from Y31 to Y42.

2.4.4 Simulations

To examine the dynamics of p7 in a lipid environment, the refined structure of p7 was placed in a *de novo* membrane made of POPC, and allowed to relax over a period of 50 ns. Due to the chain length difference of the lipids used for NMR experiments (6 carbons) and the simulations (16 and 18) some differences in the overall structure are anticipated. These differences do not necessarily affect the dynamics or the intramolecular interactions that take place within the protein. The simulations are intended to provide confirmation of the experimental findings described in this article, and to set the stage for more detailed study of these characteristics of the protein in phospholipid environments. During the simulation several side chain interactions were observed. The most interesting of these were among tyrosine residues near the interhelical loop, as illustrated in Figure 2.5. Simulations suggest that hydrogen bonding

between tyrosine 31 of the first trans-membrane helix and tyrosines 42 and 45 of the second trans-membrane helix occurs. During the simulation, a switch of the bond from one tyrosine to the other was observed. This specific switch may be associated with the intramolecular motions observed in helical segment ‘c’.

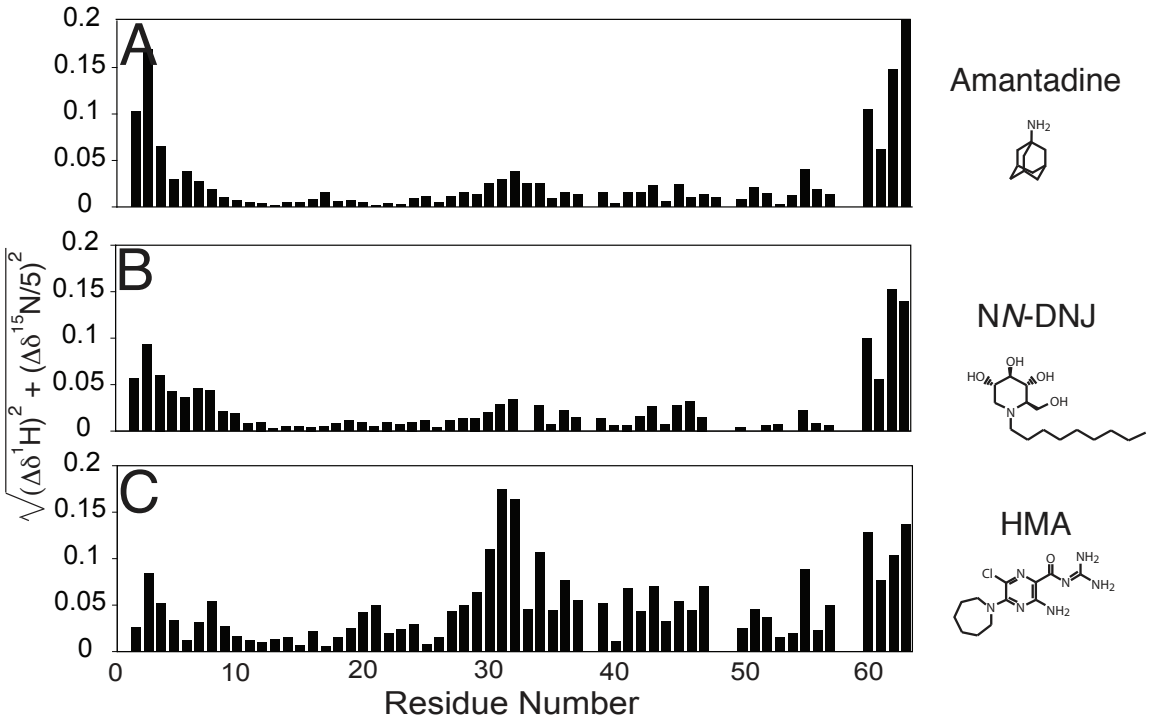

Figure 2.6: Chemical shift differences plotted as a function of residue number show the changes induced when different channel-blocking compounds are added to the NMR sample of p7. T
he chemical shifts of the backbone amide proton and nitrogen are taken into account and plotted vs the primary sequence. A. Amantadine and the iminosugar derivative. B. NN-DNJ appears to interact at the terminal regions, while C. hexylmethyl amiloride has a stronger effect on the loop region. These three plots show how compounds interact with p7 in different regions of the protein.

2.4.5 Drug Binding

Several known channel-blocking compounds have been shown to alter the ion channel activity of p7. These commercially available compounds were added directly to the NMR samples. As shown in Figure 2.6, when the ^{1}H and ^{15}N chemical shift changes

are plotted as a function of residue number, there is evidence of site-specific interactions between the drugs and residues on the protein. It is apparent that the three compounds, amantadine, *N*N-DNJ and HMA, interact with the protein differently. Amantadine and *N*N-DNJ have a larger effect on the chemical shifts of the terminal regions, while HMA alters the chemical shifts of residues in the loop region. Caution is required in this interpretation because of differences in their hydrophobicity and hence access to the residues in the protein that are within the hydrocarbon core of the micelles.

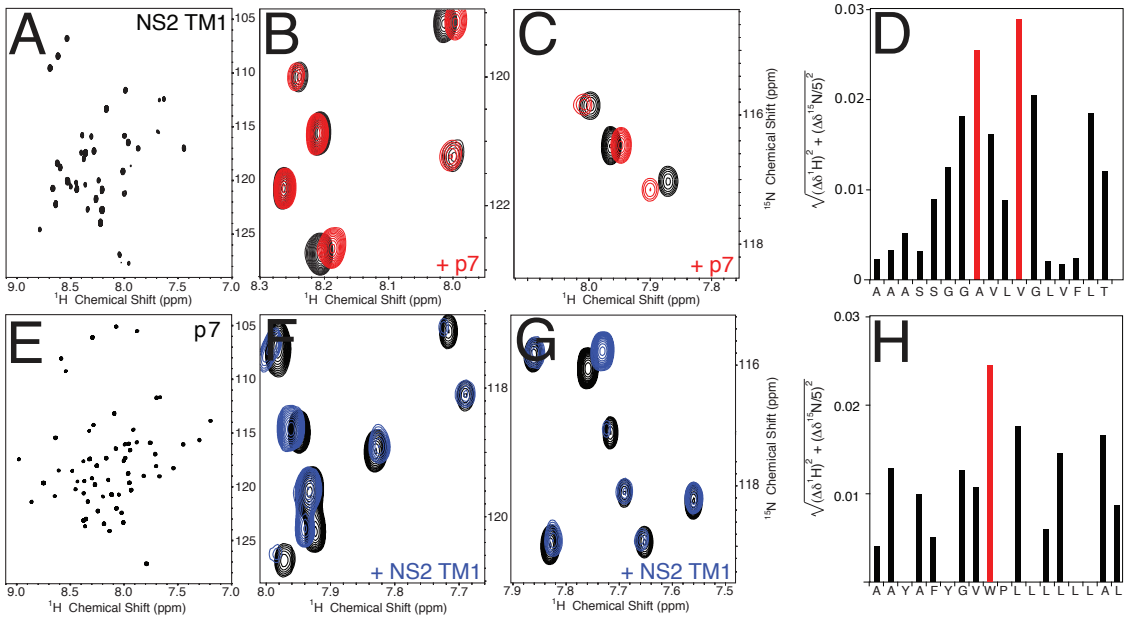

Figure 2.7: Protein−protein interactions between J4 p7 and NS2 TM1 in DHPC measured using chemical shift changes.
A. & E. Baseline 1H−15N HSQC spectra of NS2 TM1 and p7, respectively. B. & C. Regions from the NS2 TM1 HSQC spectra measured after the addition of unlabeled p7 (red) to labeled NS2 (black) shown to highlight some of the chemical shift perturbations to the NS2 TM1 protein. F. &G. Similar portions of the p7 HSQC spectra measured after the addition of unlabeled NS2 (blue) to labeled p7 (black). The proteins were added at a 1:1 concentration ratio. These spectra clearly show chemical shift changes indicating structural perturbations and interactions in both proteins. D. & H. Plot of the chemical shift changes as a function of residue showing that the interactions are likely taking place within the transmembrane regions of both proteins.

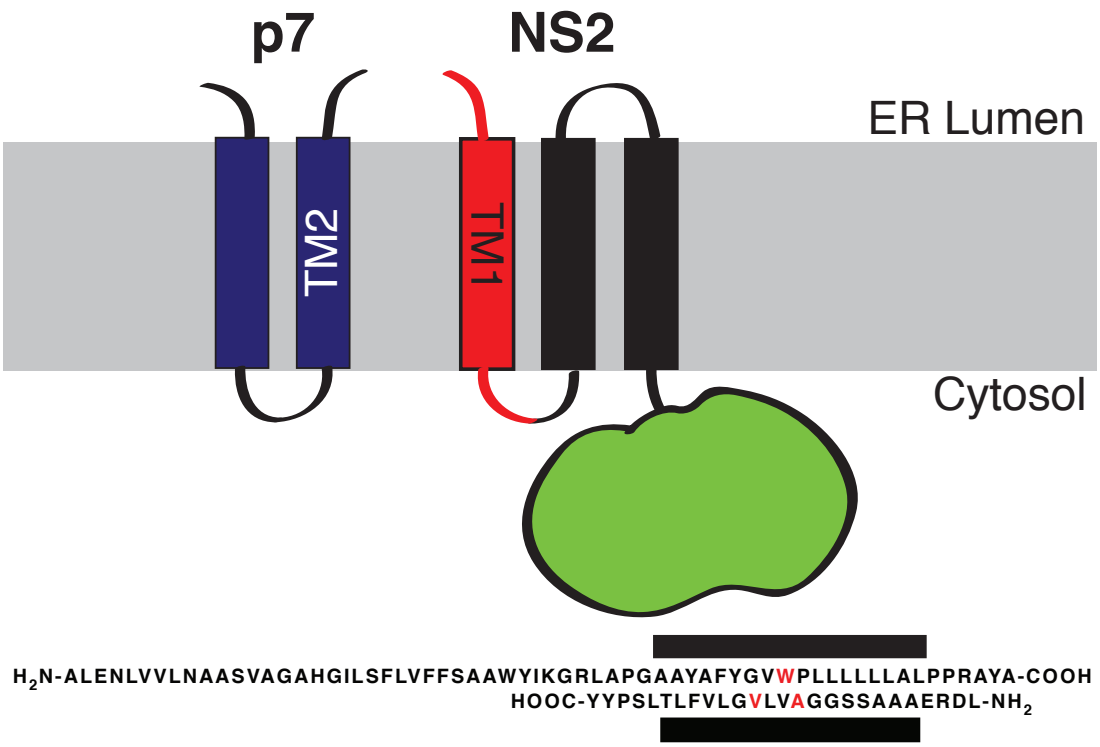

Figure 2.8: Cartoon illustration representing p7 and NS2 in the membrane of the endoplasmic reticulum.
The antiparallel arrangement of TM2 of p7 and TM1 of NS2 is shown in the diagram. Below the diagram is an alignment of the respective p7 and NS2 TM1 sequences with a black rectangle corresponding to the TM regions of the proteins. The residues that have the largest chemical shift changes in Figure 2.6 are colored red in the sequence.

2.4.6 NS2TM1 Expression and Purification

NS2 and p7 have been shown to co-localize in the ER membrane, and are believed to be involved in virus assembly[48], and may interact with each other within the membrane. In order to determine possible interactions between the two proteins the NS2 construct derived from the same genotype, J4, as p7, was expressed and purified. To simplify the identification of involved residues and accelerate the research a construct corresponding to the first TM of NS2, thought to be the crucial segment for these interactions, was used [61]. NS2TM1 was successfully expressed as a fusion protein, identical to the p7 construct,

and was directed to the inclusion bodies of the cells. Using a similar purification protocol to p7, the inclusion bodies were isolated during lysis, cleaved and HPLC purified. Milligram amounts of isotopically labeled ^{15}N protein that could be incorporated into DHPC micelles were obtained. These samples appeared to be well behaved in the same conditions that the p7 samples were measured in.

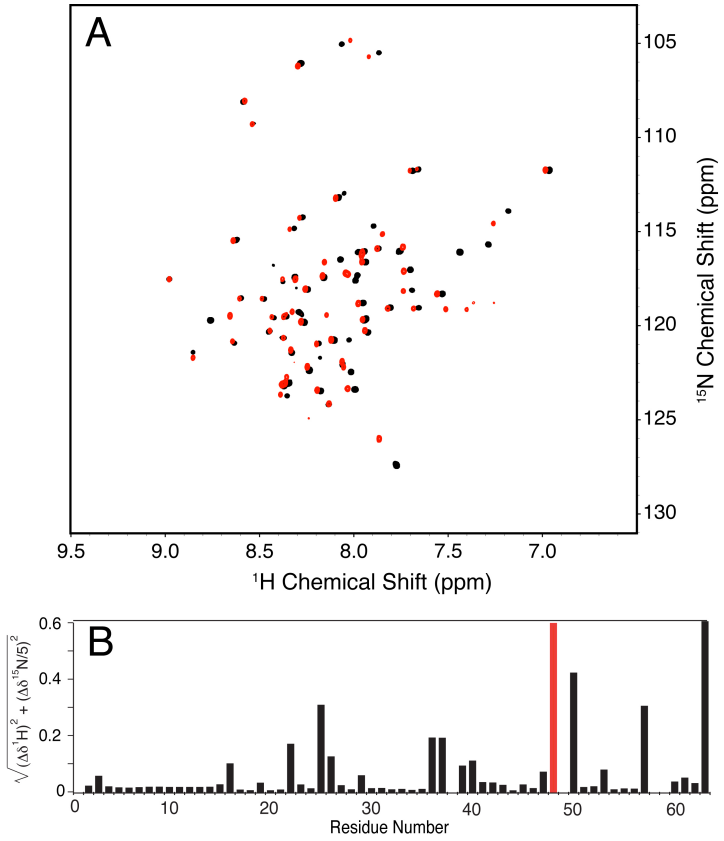

Figure 2.9: Overlay of HSQC spectra of wild-type p7 (black) and mutant p7 W48A (red) showing chemical shift differences.
A. The HSQC spectra of wild-type p7 and p7 W48A were recorded at 600 MHz, 50 °C, and pH 4.0 in 125 mM DHPC micelles (Experimental Procedures). B. Chemical shift changes are clearly evident in the second TM region where the mutation was made. Several residues directly across from residue 48 also show a degree of shift change, including three phenylalanine residues (Phe22, Phe25, and Phe26).

The resonances in the two-dimensional HSQC spectrum of the NS2 construct (Figure 2.7A) are well dispersed, and 85% of the backbone signals were assigned using a single two-dimensional ^{1}H-^{15}N HSQC NOESY experiment. HSQC spectra of the trans-membrane domain of NS2 in DHPC micelles were measured before and after the addition of unlabeled p7 to the sample. Several of the NS2 resonances were obviously shifted in the spectra, indicating that p7 was interacting with specific residues (Figure 2.7B and C). A plot of the chemical shift changes as a function of residue number upon the addition of p7 shows that the residues affected the most are those that are at the center of the trans-membrane helix, indicating that these interactions are taking place within the membrane (Figure 2.7D). Ala12 and Val15, the two residues that are most affected, are three residues apart, which is consistent with their residing on the same face of the helix.

The effect of addition of the NS2 construct on p7 spectra was also measured. When unlabeled NS2 construct was added to a sample of uniformly ^{15}N-labeled p7 protein spectral changes are observed. Interestingly, the residue that displays the largest change in chemical shift frequencies is Trp48. This residue is also predicted to be in the center of the membrane, as discussed above.

2.5 Discussion

The schematic drawing of p7 and NS2 in Figure 2.8A shows anti-parallel orientations of the transmembrane segments within the bilayer. The sequence alignment of the two proteins (Figure 2.8B) shows that the Trp48 residue of p7 and Ala12 and Val15 of NS2 are aligned near the center of the membrane when boundaries of the trans-membrane helices are aligned. Since significant chemical shift changes were limited to

one residue of p7 and two residues of NS2, the helices are likely to have different tilt angles within the bilayer. The tilt of the second trans-membrane helix of p7 was previously found to be approximately 25° relative to the bilayer normal in a DMPC bilayer.

Along with the protein-protein interactions, drug-protein interactions were examined using several compounds with p7. Unlike the NS2 studies, in which both components were incorporated into the lipid environment, the drug compounds were added to the solution. This limited the observable interactions to the loops and terminal regions of the protein since the drugs, with the exception of amantadine, are not hydrophobic. Although specific amino acids could be identified as interacting with the drugs, the results highlight the need for performing these studies in a bilayer environment where the protein can form the oligomers that have been observed in cryo-EM studies.

The simulation studies presented here provide an initial attempt to understand the reason p7 is so dynamic. The switching of the hydrogen bond at the tyrosine side chains during the simulations is an interesting finding that requires close analysis of other residues at the interface of the trans-membrane segments. It is clear that the mobile section of TM2 ends at the position of Trp48-Pro49. The tryptophan appears to result in steric hindrance that maintains a separation of the two segments. In order to examine this hypothesis a point mutation was made in p7 at the tryptophan residue. The mutant protein, p7 W48A, was successfully expressed and purified. The HSQC spectrum of the mutant protein (Figure 2.9A) indicates that it is well behaved in the same sample conditions as the wild type protein. All of the signals could be assigned by comparison with the spectra of the wild-type protein. A plot of the chemical shift changes as a

function of residue number indicates that the structural changes to the backbone are not limited to TM2 in the region where the mutation was made (Figure 2.9B). Although the changes appear to be localized to several short regions of the protein, suggesting that the global structure of p7 remains the same, one of these regions is in the middle of TM1. The residues most affected are phenylalanines 22, 25 and 26. This indicates the mutation has an interhelical effect. A plot of the relative intensity indicates that the third helical region is less dynamic as a result of the residue change which may suggest that the placement of a less bulky alanine at site 48 allows the two helices of the protein to move closer together leading to a possible stacking of the tyrosine rings and a more stable structure. Another region where the chemical shifts have changed is between the loop and helix 'c'. This might indicate that this is a hinge point that is responsible for accommodating the movement of TM2 to a position that is closer to TM1.

p7 is a remarkably complex protein considering that it has only 63 residues and its secondary structure is dominated by two hydrophobic transmembrane helices. Another NMR study of p7 in DPC micelles in solution suggests that it exists as a hexamer, and the resulting channel is more likely to be a dominant mechanism of biological activity[40]. In contrast, the structural and dynamic features identified with our NMR experiments on monomers of p7 in DHPC micelles in solution indicate that its biological roles may be more numerous and multifaceted than simply acting as a channel. Indeed, its channel activity may be secondary to its principal biological activities, a notion reinforced by the position of its gene between those for the non-structural and structural proteins of HCV. At least part of its functions are likely to involve interactions with other proteins, such as those shown in the example of p7 interacting with NS2 in Figure 2.8. Protein-protein

interactions, as well as possible channel functions, are likely to be affected by its structure and changes in it in response to the lifecycle of the virus and the addition of drugs.

Chapter 2, in full, is a reprint of the material "The Three Dimensional Structure and Interaction Studies of Hepatitis C Virus in 1,2-Dihexanoyl-*sn*-glycero-3-phosphocholine by Solution Nuclear Magnetic Resonance" as it appears in Biochemistry. 52(31):5295-303 by Cook GA, Dawson LA, Tian, Y, and Opella SJ. The thesis author was the secondary author of the paper.

Chapter 3. Three-dimensional structure of the p7 Protein of Hepatitis C Virus in Phospholipid Bilayers by Solid-State NMR

3.1 Abstract

The p7 protein from Hepatitis C virus (HCV) plays important roles in the assembly and release of virus particles from infected cells. This small 63-residue membrane protein appears to oligomerize and induce channel or pore activity under some conditions, and therefore is classified as a viroporin. p7 differs from other widely studied viroporins, such as M2 of Influenza and Vpu of HIV-1, in having two trans-membrane helices instead of one. p7 is highly conserved throughout the range of HCV genotypes, making it a potential target for anti-viral drugs. However, like other hydrophobic membrane proteins it presents challenges for structure determination. The structure of p7 has been previously characterized in two different organic solvents and two different detergent environments by solution NMR. These structures differ from each other and, notably, from the structure reported here, which was determined in phospholipid bilayers by solid-state NMR. Except for a few residues at both the N- and C- termini, the entire polypeptide shows an absence of local dynamics, including within the inter-helical loop, which results in a well-defined structure of the protein monomer in bilayers.

3.2 Introduction

Hepatitis C virus (HCV) is a small RNA virus [15, 62, 63] that infects about 100 million people throughout the world [14]. Although some of those are able to clear the virus; left untreated, the majority are chronically infected and at risk for severe health

problems, principally from liver diseases such as cirrhosis and hepatocellular carcinoma. Until recently, treatment consisted of the non-specific anti-viral agents pegylated-interferon and ribavirin, which have limited effectiveness and serious side effects. However, the therapeutic landscape is changing rapidly with the availability of direct-acting antivirals that are specific inhibitors of viral proteins[64], including the viral RNA polymerase (sofosbuvir) and the viral proteases NS3/4A (simeprevir, telaprevir)[65]. Nonetheless, with the potential for development of resistance[66] and the existence of multiple genotypes[14], the structures of other HCV proteins[62, 67] are of interest in order to develop drugs for additional targets.

Besides the therapeutic implications, HCV's small genome means that the details of molecular and structural biology of its proteins are of considerable importance to understanding the viral lifecycle. HCV is an enveloped virus with a positive single-strand RNA genome of approximately 9.6 kb. The RNA is translated into a 3000-amino acid polyprotein that is subsequently cleaved by intracellular and viral proteases to yield 10 mature proteins [15, 62, 63]. These proteins include structural proteins, core and envelope glycoproteins, E1 and E2 as well as nonstructural proteins NS2, NS3, NS4A, NS4B, NS5A and NS5B. In infected cells these proteins affect host metabolism, and are involved in the replication and packaging of viral RNA.

Notably, the nucleotide sequence for the small 63-residue membrane protein, p7, is located between those that code for the structural and the nonstructural proteins. p7 has been shown to be essential for efficient virus particle assembly and release, but not RNA replication[33, 68-70]. It is localized to the endoplasmic reticulum where there is evidence that the monomers oligomerize to form heptamers or hexamers[19, 26, 34, 71] that have

channel or pore activity. As a result, p7 is generally categorized as a viroporin[72, 73]. However, as is the case for many other small hydrophobic membrane proteins[74] its defining channel activity has not yet been related directly to a biological function, although it has been suggested that the channel activity helps to regulate the pH of lipid vesicles[75].

Understanding the biological functions of proteins necessarily begins with their three-dimensional structures, which are determined by both their primary structure and their surrounding chemical environment[76]. Although equally true for all proteins, the influence on protein structure by its environment is most evident in membrane proteins because they are embedded in highly asymmetric, liquid crystalline phospholipids bilayers[77], rather than an isotropic aqueous solution or a regular lattice. Local and global dynamics are also highly relevant to understanding the roles of p7.

With the exception of a 21-residue peptide corresponding to the transmembrane channel-forming domain of the M2[78, 79] protein of influenza no partial or intact viroporin has been crystallized. Consequently nearly all of the experimental structural information about this important class of proteins has come from NMR spectroscopy. [80 81] Computational modeling studies of p7 and other viroporins have contributed to our understanding of these proteins[82-86]. There are several prior NMR studies that have reported structures of p7[26, 31, 32, 47, 53, 81, 87, 88]; however in these previous examples, the sample environment has been that of organic solvents[31, 32] or detergent micelles in order for the protein to reorient in aqueous solution fast enough for solution NMR methods to be applied. Of particular relevance to the present study, two different structures of p7 have been recently determined by solution NMR in two different types of detergent

micelles[26, 87]. In contrast, here we present the three-dimensional backbone structure of p7 determined in phospholipid bilayers. The different local and global structures exhibited by p7 in various membrane environments may very well be associated with its multiple biological activities.[87]

Membrane proteins are effectively 'immobilized' on the timescales of the relevant nuclear spin-interactions (~10^4 Hz) by their interactions with the phospholipids in bilayers. Consequently, solid-state NMR methods are required to obtain high-resolution spectra of membrane proteins in phospholipid bilayers. p7 like many other membrane proteins undergoes fast rotational diffusion about the bilayer normal at temperatures above the gel to liquid crystalline transition of the phospholipids. Although this motion does not result in significant overall narrowing of the resonances, since it is only along a single axis, it does enable the application of rotationally aligned (RA) solid-state NMR[3, 6, 89, 90]. These experiments require that the protein in the sample be extremely pure, properly folded, and in a detergent-free phospholipid bilayer environment. Even small amounts of detergents left over from the sample preparation can destabilize and distort the structure of a membrane protein. This is part of the larger issue of the influence of the sample environment on the structures of membrane proteins[77].

The three-dimensional structure of the p7 monomer described here provides only part of the story. The protein may function *in vivo* as both a monomer and in several oligomeric states, and there are open questions about the roles of the oligomer-related channel activity, as is the case for most other viroporins, with the exception of M2 of Influenza[74, 91]. In 2003, Griffin et al. described the oligomerization state of p7[71]. A p7 construct was chemically cross-linked using a lipid-soluble cross-linking agent; the cross-

linked protein was subject to analysis using sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE), immunoblotting and transmission electron microscopy (TEM). A 42 kDa band was observed in the anti-p7 immunoblotting. A non-cross-linked glutathione S-transferase (GST)-tagged p7 construct showing ring structures that contained six subunits was identified in the TEM studies. Following this report, Clarke et al. published a study on a FLAG-tagged p7 construct[19] where they conducted cross-linking studies and observed a predominately heptameric complex in the immunoblot.

Subsequently, Luik et al. showed that the p7 reconstituted in DHPC detergent micelles cross-linked with a homobifunctional cross linker[34] observed up to 6 bands in the SDS-PAGE analysis, suggesting a hexameric oligomer. They also analyzed the non-cross-linked p7-DHPC complex using PAGE and size exclusion chromatography (SEC) but could not differentiate between a hexameric and heptameric complex. Chandler and coworkers[92] used molecular dynamics simulations to identify the possibility of p7 forming a hexameric complex, which can form transiently open channels for conducting ions. Their work described the potential for the existence of oligomers with between four and seven monomers, which is consistent with both experimental and other simulation results. Complementary calculations by Fischer and coworkers[83-85] further demonstrated the conformational and oligomeric plasticity of p7. There is also evidence that these properties are affected by the lipid environments in which they reside[93].

3.3 Material and methods

3.3.1 p7 purification and proteoliposome reconstitution

Expression and purification of genotype J4 p7 protein were carried out using previously described methods[53]. Following SEC-FPLC purification with a HiPrep Sephacryl S-200 gel filtration column (GE Healthcare Life Sciences, www.gelifesciences.com), fractions containing pure isotopically labeled p7 protein were concentrated using an EMD Millipore Series 8000 Stirred cell with an Ultracel Ultrafiltration PL3 (www.emdmillipore.com). The concentrated fractions containing 4 mg of the polypeptide were then reconstituted with 24 mg of DMPC (1,2-dimyristoyl-sn-glycero-3-phosphocholine) powder (Avanti Polar Lipids, www.avantilipids.com) to give a protein:lipid molar ratio of 1:60. The SDS was removed by first dialyzing against 20 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), pH 7.3 and 20 mM KCl for 48-72 hours followed by 20 mM HEPES, pH 7.3, containing 3 g of methyl-β-cyclodextrin at a final concentration of 0.5 mM for 24 hours. A final dialysis step was performed using 20 mM HEPES, pH 7.3 to remove any residual KCl or cyclodextrin. The p7-containing proteoliposomes were then pelleted by ultracentrifugation at 390,000 x g, in a Beckman Ti 70.1 rotor at 15°C for three hours.

3.3.2 NMR Spectroscopy

In rotationally aligned solid-state NMR, the anisotropic interactions arising from the dipole-dipole and chemical shielding are motionally averaged in a defined way according to the angle between the bond or principal axis and the axis of rotation[3, 90]. The resulting powder patterns are axially symmetric. The observed reduction in the frequency breadth depends upon the orientation of each molecular site with respect to the axis of rotation, which in the case of membrane proteins is the bilayer normal. p7 undergoes fast

rotational diffusion about the bilayer normal when reconstituted in phospholipid bilayers. This was monitored by measuring the $^{13}C'$ CSA powder pattern at slow spinning speeds as a function of temperature[94], as well as by measuring ^{1}H-^{15}N and ^{1}H-^{13}C dipolar coupling frequencies at moderate spinning speeds. We observed that the proteins undergo uniaxial rotational diffusion at temperatures above ~20°C in DMPC phospholipid bilayers, and the motion was not present at temperatures below ~15°C. The ability to turn 'on' and 'off' the axial diffusion about the bilayer normal provides assurance that the protein is not aggregated and is uniformly inserted into the bilayer [6].

Isotropic chemical shifts associated with individual residue types are readily observed in the two-dimensional ^{13}C/^{13}C homonuclear and ^{13}C/^{15}N heteronuclear correlation spectra shown in Figure 3.1. Single resonance signals resolved in three-dimensional experiments, notably NCACX and NCOCX, provided the foundation for sequential resonance assignments. Examples of assigned strip plots for residues 41 - 44 are shown in Figure 3.3. Angular restraints used for structure calculation were measured using three-dimensional separated local field experiments. RA solid-state NMR in lipid bilayers provides equivalent angular restraints to those observed in oriented sample solid-state NMR[90 47]. Three-dimensional experiments correlating ^{1}H-^{15}N dipolar coupling with ^{13}C and ^{15}N isotropic shifts were used to make ^{1}H-^{15}N dipolar coupling measurements[95, 96].

Experiments were carried out on a 700 MHz spectrometer equipped with a Bruker Avance II console and a home-built 3.2 mm solenoid coil MAS probe with a stator from Revolution NMR (www.revolutionnmr.com). The spinning rate was controlled to 12,626 Hz using a Revolution NMR MAS control system. 100 kHz radiofrequency (RF)

irradiation was used for ^{1}H decoupling during data acquisition and for double cross-polarization. Swept frequency two-pulse phase modulation was used for ^{1}H decoupling[97]. Cross-polarization (CP) from ^{1}H to ^{13}C and ^{15}N[96] was optimized using 50% amplitude modulated RF irradiation on the ^{1}H channel.

Multidimensional NMR experiments for sequential assignment were carried out using pulse schemes described previously[3, 6, 9, 90, 98]. Two-dimensional homonuclear $^{13}C/^{13}C$ correlation data were acquired with 4 ms and 12 ms evolution periods during indirect and direct data acquisition, respectively. The two-dimensional homonuclear correlation spectrum in Figure 3.1A was further processed using covariance spectroscopy[99-102] to obtain a symmetrized spectrum with a higher signal-to-noise ratio than a Fourier transformed spectrum. The HETCOR spectrum correlating $^{13}C\alpha$ and ^{15}N isotropic shifts was acquired with 7 ms for ^{15}N and 12 ms for ^{13}C chemical shift evolution. Spin-diffusion among homo- and hetero- nuclei was carried out under dipolar assisted rotational resonance (DARR)[103] and spectrally induced filtering in combination with *cross polarization (*SPECIFIC) CP[98] conditions. A mixing time of 40 ms for DARR and 3.6 ms for SPECIFIC CP with ^{13}C and ^{15}N RF irradiations set to 1.5 ν_r and 2.5 ν_r were used in all multidimensional experiments, unless otherwise noted. Three-dimensional SLF experiments correlating ^{1}H-^{15}N dipolar couplings with $^{13}C\alpha/^{15}N$ isotropic shifts were obtained with 1.12 ms of dipolar evolution. Equal numbers of data points were linear predicted in all dimensions prior to Fourier transformation.

Apodization of the free induction decays utilized a sine bell squared function with 60° shift. All data were processed using either Bruker Topspin 3.1 (http://www.bruker.com/) or NMRPipe[104] and visualized using Sparky[105].

3.3.3 Structure Calculation

In the first stage of the structure calculation, initial coarse-grained structures were generated in Rosetta. Molecular fragments were obtained from the Robetta server[106] based on the primary sequence. 10,000 decoys were calculated using the membrane protocol[107] in the Rosetta software package (version 3.4) on a local workstation. All 10,000 decoys were relaxed in all-atom mode following the coarse-grained mode. Then the relaxed structures were refined against 36 ^{1}H-^{15}N dipolar coupling restraints (Table 3.1) obtained from solid-state NMR experiments in the all-atom mode. After the all-atom refinement, clusters were calculated with a cluster radius of 7 Å for residues 6 through 57. A total of 36 clusters were formed with 36% of the structures forming the most populated cluster. The 100 lowest energy structures from the most populated cluster were accepted to extract dihedral angle restraints.

The simulated annealing protocol in Xplor-NIH was employed to calculate high-resolution protein structures using experimental restraints. The routinely used folding protocol in Xplor-NIH was carried out in four stages, starting from an extended structure: (1) torsion angle dynamics at high-temperature (3,500 K) for a period of 15 ps or 10,000 time steps. (2) torsion angle dynamics with simulated annealing, where the temperature is reduced from the initial high temperature value to 50 K in steps of 12.5 K, for a period of 0.8 ps or 500 time steps per temperature step, (3) 500 steps of Powell torsion angle minimization, and (4) 500 steps of Powell Cartesian minimization. The protein structure with lowest energy was accepted for further refinement.

Dihedral angle restraints and dipolar coupling restraints were applied with respective force constants of k_{CDIH}=200 kcal•mol^{-1}•rad^{-2} and k_{dc}=0.01 kcal•mol^{-1}•rad^{-2}, then ramped to 300 kcal•mol^{-1}•rad^{-2} and 3 kcal•mol^{-1}•rad^{-2} for dihedral and dipolar coupling restraints, respectively, during the annealing stage. The axial alignment parameter (D_a) was set to 10.52 and rhombicity (Rh) was fixed at 0 to reflect the physics of the samples in solid-state NMR experiments[108]. The torsionDB statistical torsion angle potential[109] was implemented with a force constant set to k_{tDB}=0.02 kcal•mol^{-1}•rad^{-2} in the high temperature stage and ramped geometrically from 0.02 to 2 kcal•mol^{-1}•rad^{-2} during simulated annealing. Atomic overlap was prevented by limiting allowed repulsions to those between atoms separated by four or more covalent bonds (nbxmod=4).

Refinement utilized similar steps, except starting with the lowest energy structure from the folding step, and using higher force constants for dihedral angle and dipolar coupling restraints. The refinement starts from 3000 K, with k_{CDIH}=400 kcal•mol^{-1}•rad^{-2} and k_{dc}=0.05 kcal•mol^{-1}•rad^{-2}. k_{dc} is ramped to 5 kcal•mol^{-1}•rad^{-2} during simulated annealing. D_a is initially set to 10.52 based on NH bond length of 1.05 Å, and allowed to vary during refinement, while Rh was fixed at 0.

100 structures were calculated in both folding and refinement, and the 10 lowest energy structures from refinement were accepted as the representative structure ensemble for p7 in DMPC liposomes. The backbone RMSD of the 10 lowest energy structures (shown in Figure 3.4A) is 1.73 ± 0.68 Å and the all-heavy atom RMSD is 2.60 ± 0.74 Å for residues 6 through 57. The terminal residues are deleted from the analysis because

they are mobile and unstructured on the 10^4 Hz timescale of the solid-state NMR experiments. The averaged D_a value is 10.36 kHz.

3.4 Results

3.4.1 NMR experimental measurements

All of the experimental NMR results shown in the Figures were obtained at a ^1H resonance frequency of 700 MHz on proteoliposome samples containing approximately 4 mg of uniformly ^{13}C and ^{15}N labeled p7 in 24 mg of DMPC. The unoriented protein-containing phospholipid bilayers were concentrated by centrifugation such that full hydration was maintained. They were packed into 3.2 mm outer diameter rotors for magic angle spinning (MAS) necessary for obtaining isotropic chemical shifts[110] as well as implementing rotationally aligned solid-state NMR as a method of structure determination[3]. Data were obtained at two temperatures. The protein was static, but not frozen, within the bilayer environment at -10°C, and undergoes rotational diffusion about the bilayer normal at 25°C at a rate that is fast compared to the relevant NMR timescale (~10^4 Hz). This was verified by observation of the breadth of the powder patterns[94] and the associated spinning sideband patterns of the $^{13}C'$ resonances. A family of sidebands consistent with a span of about 170 ppm was observed at -10°C and essentially only a center band with minimal sideband intensities was observed at 25°C with a MAS frequency of 5 kHz on a spectrometer with a ^1H resonance frequency of 700 MHz. We have found these sideband patterns to be definitive evidence of the fast rotational diffusion of a highly helical membrane protein about the bilayer normal at temperatures

above the gel-to-liquid crystalline phase transition temperature of the protein-containing phospholipids[6].

The correlation spectra in Figure 3.1 were obtained at -10°C where the protein does not undergo rotational diffusion. Figure 3.1A contains a two-dimensional $^{13}C/^{13}C$ homonuclear correlation spectrum, and Figure 3.1B contains a two-dimensional $^{13}C/^{15}N$ heteronuclear correlation spectrum. The spectra obtained at -10°C and 25°C display a high degree of overlap of chemical shift frequencies, which implies that there is little or no difference in the protein structures below and above the lipid phase transition temperature. In practice, this means that resonance assignments made more conveniently at the lower temperature can be transferred to the spectra obtained at the higher temperature where the key structural measurements are made.

The partially resolved backbone resonances in the spectra in Figure 3.1A and B were nearly fully resolved and assigned in three-dimensional spectra. Significantly, the only resonances that are genuinely missing (rather than possibly overlapped) after a careful combined analysis of the two- and three- dimensional spectra are those from five amino acids at the N- terminus (residues 1–5) and from three amino acids at the C-terminus (residues 61–63). Also, as expected, resonances from the imino acids Pro 59 and Pro 60 are also not seen in these spectra because their nitrogen does not have a directly bonded hydrogen. For the terminal residues, cross-polarization is ineffective because they undergo local motions that are rapid enough to average the dipolar couplings required for cross-polarization[96], even when the protein is not undergoing rotational diffusion. Thus, the terminal residues are classified as 'mobile' relative to the ~10^4 Hz timescale, and are not restrained by experimental data in the calculations.

In Figure 3.1A, the off-diagonal signals arise from homonuclear spin-exchange due to dipole-dipole interactions among proximate ^{13}C nuclei. With a mixing time of 150 msec, it is likely that all of the signals in the spectrum result from relatively short-range intra-residue interactions. The signals from the carbonyl carbons occur between 170 ppm and 180 ppm and those from the aliphatic carbons, including sidechains, occur between 10 ppm and 70 ppm.

In Figure 3.1B, each correlation resonance is from a single residue, and is characterized by two isotropic chemical shift frequencies, the ^{15}N frequency from the amide nitrogen and the ^{13}C frequency from the a-carbon. The partial resolution observed in this two-dimensional spectrum is sufficient to translate to complete resolution in three-dimensional spectra.

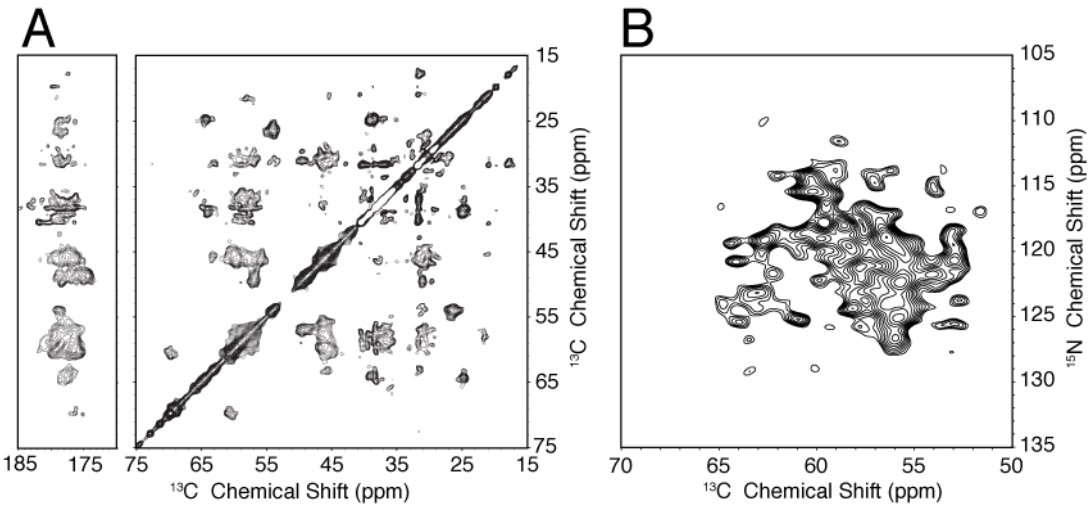

Figure 3.1: Two-dimensional solid-state NMR spectra of uniformly ^{13}C and ^{15}N labeled p7 in DMPC bilayers at -10^{o}C.
A. Homonuclear $^{13}C/^{13}C$ correlation spectrum. B. Heteronuclear $^{15}N/^{13}C$ correlation spectrum.

The data shown in Figure 3.2 were obtained from the same sample used to obtain the data in Figure 3.1 with the exception of the temperature being 25°C. As a result, the protein was undergoing fast rotational diffusion about the bilayer normal. Two-dimensional ^{1}H-^{15}N dipolar coupling / ^{13}C chemical shift planes taken at three different ^{15}N chemical shift frequencies are shown as examples of the quality of the experimental data obtained from the three-dimensional data set. The spectral planes demonstrate variations in ^{1}H-^{15}N dipolar coupling frequencies that reflect different orientations of ^{1}H-^{15}N amide bond vectors relative to the bilayer normal, the axis of protein rotation. The spectral plane in Figure 3.2C highlights the single-site resolution available in these experiments. All of the resonances in Figure 3.2A and B are assigned, although not all of them have their maximum intensity (along the ^{15}N frequency dimension) at the selected plane. Thus, while the ^{1}H-^{15}N dipolar coupling and $^{13}C\alpha$ isotropic chemical shift frequencies measured from the resonances in the Figure are accurate some of the ^{15}N isotropic chemical shift frequencies, associated with the signal maxima, are slightly different. The accurate measurements for each signal are used in the structure calculations. The single signal in Figure 3.2C is at its maximum intensity, therefore all three frequencies that can be measured by inspection of the spectral plane are accurate within the experimental error.

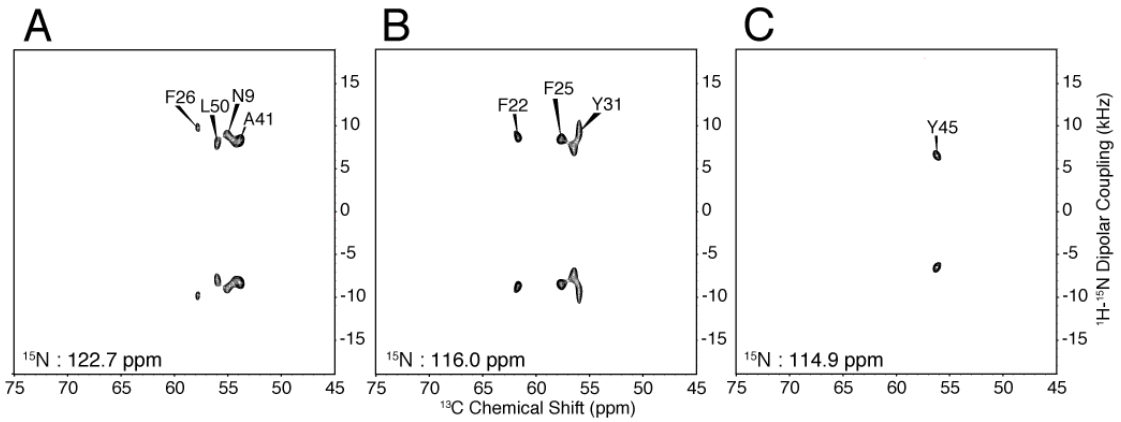

Figure 3.2: ^{1}H-^{15}N heteronuclear dipolar couplings / ^{13}Cα chemical shift correlation planes at three representative ^{15}N chemical shift frequencies from a three-dimensional data set.
A. Plane selected at ^{15}N shift of 122.7 ppm. B. Plane selected at 116.0 ppm. C. Plane selected at 114.9 ppm.

A representative selection of data used to assign the resonances for residues A41 through A43 is shown in Figure 3.3 in the form of alternating slices from intra-residue correlation (NCACX) and inter-residue correlation (NCOCX) spectra. The assignment of individual resonances to specific residues was performed by correlating the *i* to *i*-1 residue signals for the carbonyl and α-carbon backbone sites for the structured residues of p7 in phospholipid bilayers. In combination with the two-dimensional correlation spectra in Figure 3.1, these data enabled the assignment of the ^{15}N, ^{13}CO, and ^{13}Cα sites in the protein backbone.

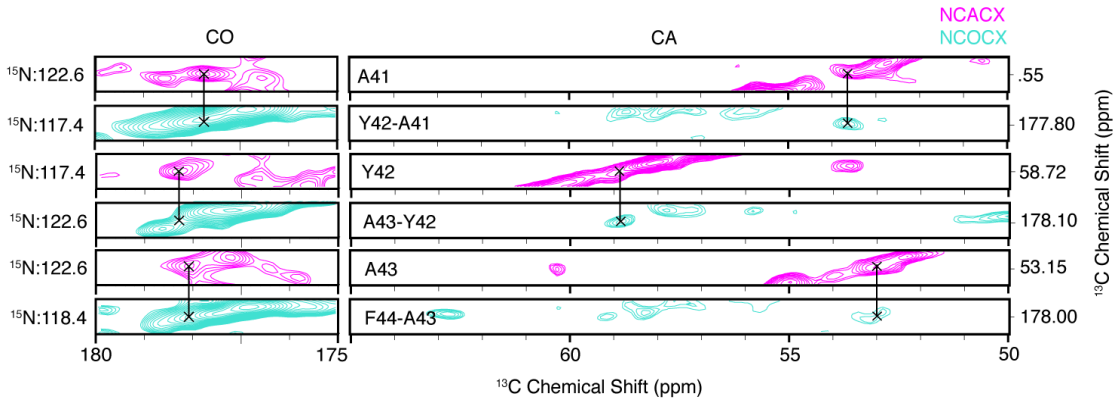

Figure 3.3: Representative data used in the resolution and assignment of the solid-state NMR spectra in Figures 3.1 and 3.2.
Alternating slices from intra-residue correlation (NCACX) and inter-residue correlation (NCOCX) spectra are shown. The two-dimensional ^{13}C-^{13}C chemical shift correlation slices were taken from three-dimensional experiments where the third dimension is the ^{15}N chemical shift frequency The assignment of resonances was performed by correlating the signals from i to i-1 residues for the carbonyl and alpha carbons as is shown here for residues A41 through A43.

Due to spectral overlap, a few dipolar-coupling frequencies could not be confidently measured and are not listed. Notably, there are two columns of dipolar coupling frequencies listed in Table 1. One marked “Exp.” contains the experimentally measured values from spectra such as those in Figure 3.2. The other, marked “Calc”, represents the back-calculated values from the final structure shown in Figure 3.4B.

Table 3.1: Summary of NMR data.

Residue	Chemical Shift (ppm)			Dipolar Coupling (kHz)	
	CO	C_α	N	Exp.	Calc.
6 V	171.9	59.03	126.4	3.76	3.91
7 V	171.4	58.89	126.4	17.47	17.28
8 L	172.9	53.57	130.0		
9 N	171.6	51.07	128.3	18.35	18.05
10 A	171.7	50.83	129.9	7.99	8.15
11 A	173.0	48.28	122.4		
12 S	172.6	55.78	120.4		
13 V	172.8	58.69	126.2	14.75	14.67
14 A	169.9	47.92	125.5	15.52	15.51
15 G	172.9	47.92	125.5	17.81	17.88
16 A	174.7	48.51	124.8	11.19	11.28
17 H	172.6	55.94	120.7	16.75	16.74
19 I	172.0	58.01	124.2	19.56	19.53
20 L	173.9	53.05	124.1	15.52	15.56
21 S	170.5	54.69	118.8	15.70	15.90
22 F	176.0	56.88	120.8	17.83	18.02
23 L	172.1	52.82	116.7	16.76	16.97
24 V	173.9	59.96	129.4		
25 F	171.4	52.85	120.7	17.35	17.40
26 F	170.7	55.57	122.6	19.90	19.60
27 S	170.9	55.82	119.5		
28 A	175.7	47.91	121.9	13.84	14.06
29 A	172.9	50.17	126.8	16.18	15.96
30 W	172.2	54.21	126.9	8.96	9.17
31 Y	174.9	51.79	122.1	19.06	18.98
32 I	173.0	58.74	125.7	14.32	14.28
33 K	171.9	55.94	123.3		
35 R	171.5	55.95	131.8	14.64	14.80
36 L	173.4	54.21	125.7		
37 A	172.3	54.25	123.6	15.34	15.33
38 P	172.7	58.65	131.6		
39 G	171.5	48.15	106.2		
41 A	172.8	48.70	127.6	16.87	16.99
42 Y	173.3	53.86	122.4	10.40	10.30
43 A	173.0	48.09	129.4	10.61	10.52
44 F	170.6	51.10	123.4	15.78	15.88
45 Y	174.1	52.11	119.8	13.22	13.31
47 V	169.2	60.03	130.5		
48 W			125.2	20.49	20.41
49 P	171.6	58.69	132.5		
50 L	172.1	51.08	126.8	16.11	16.13
51 L		51.71	126.5	20.14	18.66
52 L		52.24	126.9	19.43	19.43
53 L		51.89	125.5	14.98	15.08
54 L		52.06	126.2	18.55	19.91
55 L		51.75	126.2	19.41	19.38
57 L		51.78	125.8	16.01	15.96

3.4.2 Protein Structure Calculations

The structure calculation for p7 was carried out in several stages. First, sequence-based molecular fragments were generated on the Robetta server. Then, the experimentally measured ^{1}H-^{15}N dipolar couplings were applied for refinement. After

clustering and selecting 100 lowest energy structures, dihedral angle restraints were extracted and a high-resolution structure was calculated using the simulated annealing protocol in Xplor-NIH with experimental restraints. The refinement process was carried out by applying the dihedral angle and dipolar-coupling restraints to obtain the 100 lowest energy structures from the folding step. Finally, 100 lowest energy structures were calculated from both folding and refinement whereby the 10 lowest energy refined structures were accepted to represent the structural ensemble for p7 in DMPC lipid bilayers. These ten structures are overlaid in Figure 3.4A.

The lowest energy protein backbone structure is shown in Figure 3.4B. A CPK (Corey, Pauling, and Koltun) representation of the structure including added side chains is shown in Figure 3.4C. With the exception of the unrestrained mobile residues at the N- and C- termini, the entire structure, notably including the antiparallel helices, is tightly packed. Schematic drawings of the structure where the helices are shown as tubes are presented in Figure 3.4D – 3.4F. Each of the trans-membrane helices contains a substantial kink towards the middle of the bilayer. The kinks are observed in the calculated structures, and especially in the dipolar-coupling data plotted as a function of residue number to form the Dipolar Waves (Figure 3.5B). The horizontal black bars aligned with the sequence designate the helical segments derived from the protein structure calculation. The two hydrophobic helices match well with what would be expected from the hydropathy plot provided for comparison in Figure 3.5A.

The accuracy of the lowest energy structure shown in Figure 3.4 was verified by the correlation plot in Figure 3.5C. This is a plot of the measured ^{1}H-^{15}N heteronuclear dipolar couplings (“Exp.” in Table 3.1) versus those back-calculated from the structure

("Calc." in Table 3.1). The ability to back-calculate the measured spectral parameters from the structure is an important advantage of oriented sample NMR methods. In this case, the correlation coefficient is 0.993.

Significantly, there is no evidence of internal backbone motions in the inter-helical loop region where several large heteronuclear dipolar couplings are observed. These couplings would be consistently reduced in amplitude if any motional averaging were present. The fully structured backbone is essential for the calculation of the protein structure based on the methods used here.

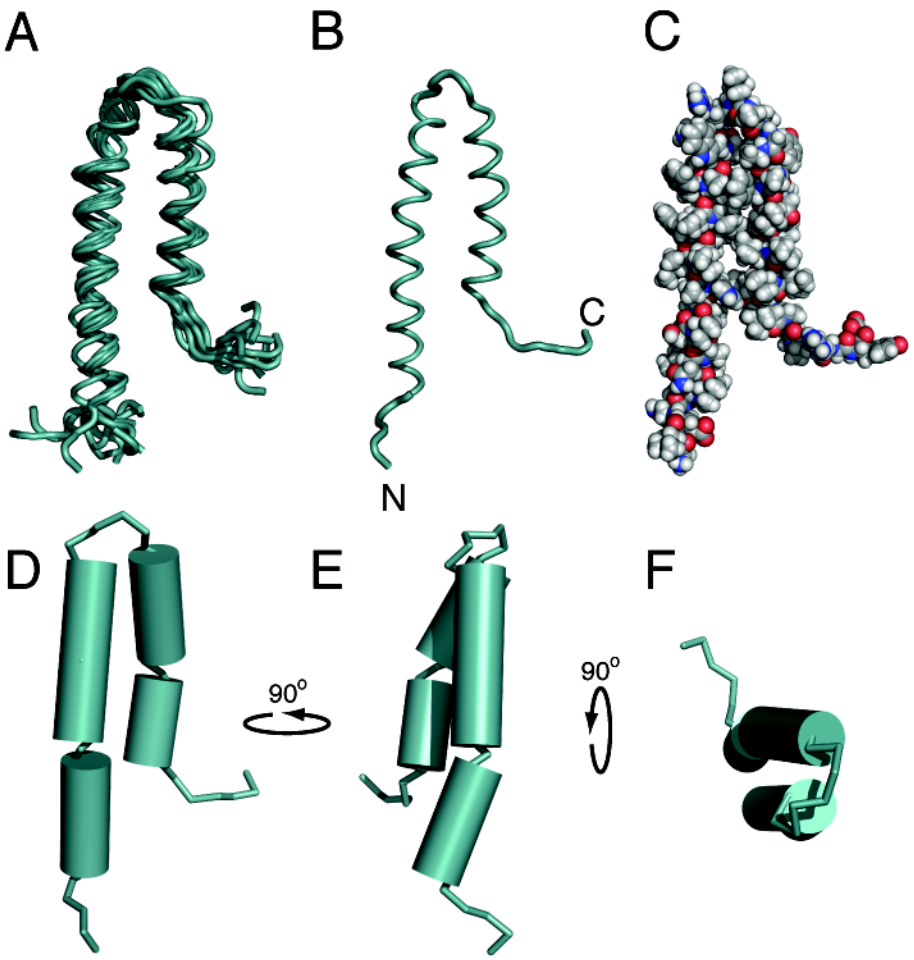

Figure 3.4: Structure of p7 in DMPC bilayers determined by rotationally aligned solid-state NMR.
A. Ensemble of the 10 lowest energy structures calculated from Xplor-NIH. B. Lowest energy structure. C. CPK representation of the lowest energy structure showing side chains. Panels D – F. show three orientations of the lowest energy structure of p7 in DMPC bilayers.

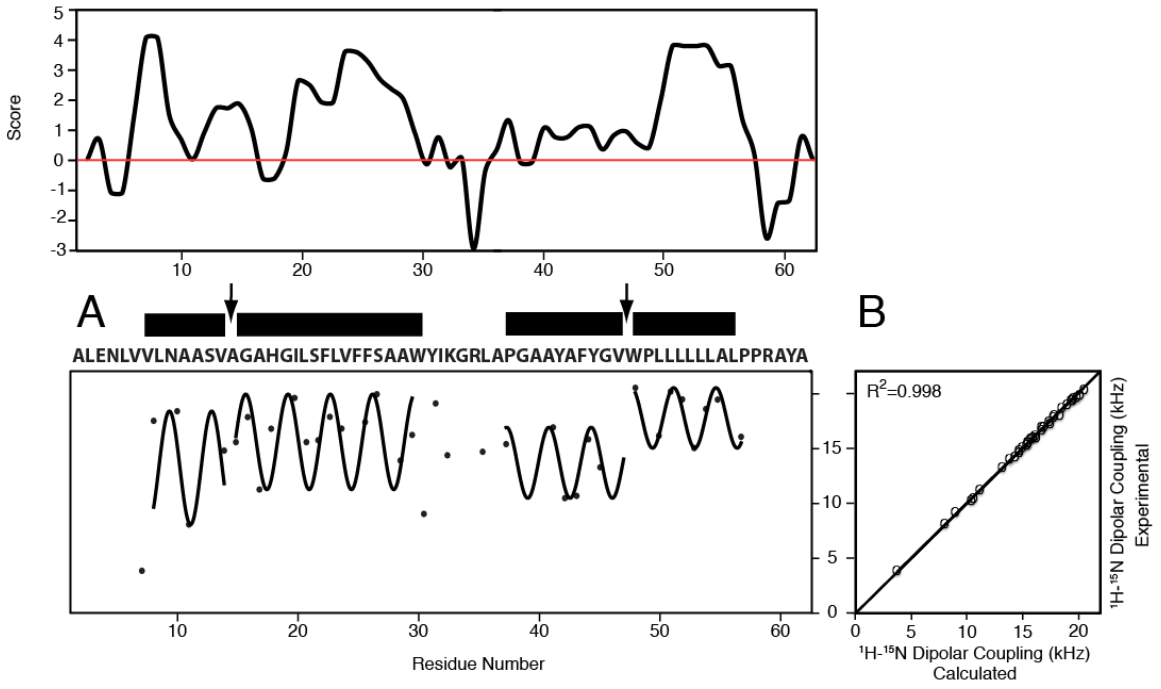

Figure 3.5: Analysis of the ^{1}H-^{15}N dipolar couplings of p7 in DMPC bilayers at 25°C. A. Kyte-Doolittle hydropathy plot of the amino acid sequence for the J4 strain of p7 showing a quantitative degree of hydrophobicity of each residue. B. Dipolar Waves. The magnitudes of the measured ^{1}H-^{15}N dipolar couplings are plotted as a function of residue number. Sinusoidal waves were fit to the data using a periodicity of 3.6 residues per turn, appropriate for an a-helix[111]. The dark bars on the top of the plot indicate the regions of a-helix in the structure. C. Correlation plot of the experimental dipolar couplings versus the back-calculated values from the lowest energy structure is shown. The correlation coefficient is 0.993.

3.5 Discussion

As reviewed by DiMaio[74] and by Nieva, Madan, and Carrasco[91], a wide range of viruses express small membrane proteins, called viral miniproteins or viroporins, depending on the functional perspective. Among many others, the classes of viruses that express one of these proteins include lentiviruses, influenza viruses, and coronaviruses. The small membrane proteins from these viruses have been studied to varying extents.

Here we are primarily concerned with the best-studied examples because, for the most part, these proteins have structural studies to complement the genetic and biological studies.

The small hydrophobic proteins with substantial structural studies include the M2 protein from influenza[112, 113], Vpu from HIV-1[114, 115], SARS-CoV E[116], protein from the human respiratory syncytial virus[117], and the protein that targets Golgi complex in coronavirus[118]. All of these examples have a single trans-membrane helix. This is the major difference with p7 from HCV, which has two trans-membrane helices connected by a short, structured inter-helical loop (Figure 3.4).

One of the most important considerations in the design and execution of structural studies by NMR spectroscopy is the environment of the protein. In order to use solution NMR methods the protein must undergo rapid global reorientation. The structure of p7 has been determined in four separate efforts by solution NMR, twice in organic solvents and twice in detergent micelles. There have been a number of studies seeking optimal detergents for solubilizing membrane proteins[119 77, 120-122]. In addition to the chemical nature of the detergent, the concentration has important effects on the properties of the membrane protein[123].

The structure of p7 has been reported five times, including the structure shown in Figure 3.4. All of the structures differ from each other, some substantially so, reflecting the impact of the lipid environment on a membrane protein, and perhaps the intrinsic plasticity of p7. Saint et al[124] synthesized p7, and they were able to obtain well-resolved NMR spectra in TFE/water. Montserrat et al [31] determined the secondary structure elements of the monomeric form of p7 in TFE/water, suggesting the presence of a helix-

loop-helix "hairpin" structure. Cook et al performed initial NMR studies on expressed p7 in phospholipid bilayers and micelles[47, 88]. From these studies it was determined that the protein contained helical segments tilted approximately 10^{o} and 25^{o} relative to the bilayer normal. Cook et al [87] also determined the three-dimensional structure of p7 in DHPC micelles by solution NMR. OuYang et al[26] determined the structure of p7 in DPC micelles by solution NMR; the structure was found to be a hexameric assembly, where the individual p7 monomers not only interact with their immediate neighbors but were also shown to associate with more distant neighbors. The structure is surprising in that it contains three rather than two helical segments, in contrast to the prediction of the hydrophobicity plot in Figure 3.5A, and differs from the other structures of p7 in other ways. These differences may be attributed to sequence variations between the two different p7 proteins with an identity of 54%, although their hydropathy plots are qualitatively similar. Recently, Foster et al [32] determined the structure of the monomer in methanol and used it for assembling a model of the oligomer.

We have determined the structure of p7 in DHPC micelles by solution NMR[87], and here in DMPC bilayers by solid-state NMR (Figure 3.4). There are some similarities between the two structures, especially in the midsection of the protein, which includes most of the N-terminal helix, the inter-helical loop, and the C-terminal helix. However, there is a large difference between the structures with the N-terminal helix bending over in the DHPC sample, apparently to stay within the hydrophobic environment of the micelle and avoid the aqueous solution. Importantly, the structure in Figure 3.4 was determined in liquid crystalline phospholipid bilayers, with no detergents or organic solvents present, under physiological conditions of hydration and pH. The structure is

consistent with the measured experimental data. The correlation coefficient for the comparison of the experimental measurements and the angular measurements from the final, calculated protein structure (Figure 3.5C) is 0.993.

The determination of the HCV p7 protein in near-native conditions provides a starting point for the design of channel-blocking and other types of anti-viral drugs [85]. Considering the invariant residues from the various genotypes of HCV and the accurate structural depiction of the p7 ion channel, one would have the capability to design specific compounds to target the channel architecture of this viroporin. The three-dimensional structure of the monomer is advantageous for analysis of and comparison to protein-protein complexes that may form within infected cells.

3.6 Conclusions

The three-dimensional backbone structure of p7 of HCV was determined in phospholipid bilayers. With two trans-membrane helices connected by a fully structured inter-helical loop, it forms a stable protein structure. While p7 is not involved in replication, it is an essential contributor to the assembly and release of new virus particles. A variety of structural, biophysical, and functional evidence indicates that p7 monomers can oligomerize in various preparation to yield channels or pores. Reconciling the monomer structures found under various conditions with the channel activities is a challenge for research in this area, as it is for all proteins that might be categorized as viroporins.

p7 adopts a tightly packed structure in liquid crystalline phospholipid bilayers. For such a small protein, it is remarkably complex. There are a few mobile residues at

both the N- and C- termini. Both of the trans-membrane helices have a single well-defined kink. The inter-helical loop shows no evidence of internal motions and has a complex tertiary structure. This suggests that the monomer is capable of recognizing and binding to other proteins as part of its function. Its ability to oligomerize adds another potential opportunity for involvement in the lifecycle of the virus.

Chapter 3, in part, is in preparation for submission for publication of the material by G.A. Cook & L.A. Dawson, Tian Y, Das BB, Opella SJ.

Chapter 4. Characterization of Oligomerization State of HCV p7

4.1 Abstract

There have been a multitude of oligomeric state characterization experiments concerning p7. These studies have led to a confusing debate as to what the preferred state of the functionally intracellular p7 prefers due to experimental inconsistencies including lipid environment, genotype, and epitope or cross-linker selection. As it becomes increasingly important to study the structure of the protein in its' native environment and to understand the preferred structural state, we must establish a defined methodology for characterizing preferred oligomeric states of channel proteins. It becomes clear that the variables stated: lipid environment, cross-linking agent, epitope tag modifications and, possibly, the genotype of p7 being studied all contribute to the type of oligomer that is observed. Here we describe a more refined method to characterize the oligomeric state of J4 p7 reconstituted in PFO, DMPC proteoliposomes and nanodiscs. We used an atypical cross-linking technique, Photo-Induced Cross-linking of Unmodified Proteins, PICUP, to isolate a potential preferred oligomerization state. The analysis was carried out using SDS-PAGE with silver staining to elucidate predominate oligomeric species in direct comparison to the non-cross-linked samples. The debate continues over the oligomeric state of p7 due to the lack of standardization of experimental procedures.

4.2 Introduction

The three-dimensional structure of the monomer of p7 provides only part of the story. The protein may function *in vivo* as both a monomer and/or an oligomer, and there

are open questions about the role of the oligomer-related channel activity. In 2003, Griffin et al. described the state of oligomerization state of p7. A HepG2-cellularly expressed His-tagged p7 construct was chemically cross-linked using the lipid-soluble cross-linking agent dithiobis[succinimidylpropionate] (DSP). The cross-linked protein was subject to analysis using non-reducing SDS-PAGE, immunoblotting and transition electron microscopy (TEM). A 42 kDa band was observed in the anti-p7 immunoblot as well as a non-cross linked GST-tagged p7 construct showing ring structures comprised of six subunits visible from the TEM studies.

Following this report, in 2006 Clarke et al. published a study on cross-linking FLAG-tagged p7 construct using DSP in artificially synthesized PA:PC unilamellar vesicles. The predominate species observed was hexameric complexes imaged in an α-FLAG immunoblot. In 2009, Luik et al. presented the full-length viroporin, p7 subtype JFH-1 solubilized in 1,2-diheptanoyl-*sn*-glycero-3-phosphocholine (DHPC) detergent, cross-linked using the homobifunctional cross-linker: 1,11-bis(maleimido)triethylene dlycol [$BM(PEO)_3$] and imaged using single-particle electron microscopy (EM). The $BM(PEO)_3$ cross-linker takes advantage of free cysteine residues inherent in the JFH-1 p7 subtype eliminating the need for the FLAG epitope tag allowing the authors to study the unmodified protein. Up to 6 bands were observed in SDS-PAGE, suggesting a preferred hexameric oligomer state. Additionally, the authors analyzed a non-cross-linked complex using BlueNative (BN) PAGE and size exclusion chromatography (SEC) but could not discern between the existence of either a hexameric or heptameric complex.

Combining all of the conflicting results from these oligomeric studies, it becomes abundantly clear that lipid environment, cross-linking agent, epitope tag modifications and, possibly, the genotype of p7 being or any combination of the aforementioned all contribute to the type of oligomer that is observed. The oligomeric state of J4 p7 reconstituted in DMPC proteoliposomes, DHPC micelles, DPC micelles, and LMPG micelles, used in this study, is investigated using a not only a FLAG-p7 construct reconstituted but also of an unmodified p7 construct using an atypical techniques that uses a photo-inducible chemical cross-linker, $Ru(III)bpy_3^{2+}$. Finally, a modified native gel using perfluorooctanoic acid, a very mild detergent was run in comparison to several proteins known to run as monomer or oligomer to gage what size p7 migrates to under native-like conditions. The debate over the oligomeric state of p7 is still active not only because this protein is too small to be analyzed by atypical instruments used to analyze oligomerization state, but also most likely due to a lack of standardization of experimental conditions.

4.3 Materials and Methods

4.3.1 Chemical Cross-linkers Used to Characterize FLAG-p7 Oligomerization State

Cross-linkers chosen for chemical cross-linking were selected based on membrane permeability. All membrane permeable cross-linkers are homobifunctional N-hydroxysuccinimide esters (NHS ester), which react most efficiently with primary amino groups in a pH range of 7-9. Dithiobis[succinimidylpropionate] (DSP) and Disuccinimidyl suberate (DSS) are non-water soluble cross-linking agents that first must be dissolved in DMSO. Bis[sulfosuccinimidy] (BS^3) is not membrane permeable and

does not need to be dissolved in DMSO prior to use, this was used as a negative control as we should not observe cross-linking.

To reconstitute the FLAG construct of p7 (**DYKDDDDK**ALENLVVLNAASVAGAHGILSFLVFFSAAWYIKGRLAPGAAYAFY GVWPLLLLLLALPPRAYA) in DMPC liposomes 2-4 mg of SEC purified protein (described previously) was concentrated in ~4% SDS. A 1:6 lipid to protein ratio was found to be most optimal for solid-state NMR studies (described previously) and the appropriate amount of solid DMPC was added to the concentrated protein sample and rocked overnight. The next day the sample was dialyzed in 20 mM HEPES pH 7.4 with constant buffer changing. For the following three days the protein was dialyzed in 30 mM potassium chloride and 20 mM HEPES pH 7.4 with constant changing until a majority of SDS was removed from the sample. Finally, the sample was dialyzed for one day in methyl-β-cyclodextrin for three changes and dialyzed in 20 mM HEPES, pH 7.4 overnight. Alternative lipid environments were chosen based on the solution and solid-state sample reconstitution protocol lipids, DHPC and DMPC, as well as lipids frequently used in the literature for p7 structural studies: 1-myristoyl-2-hydroxy-sn-glycero-3-[phosphor-RAC-(1-glycerol)] (LMPG), n-dodecylphosphocholine (DPC) and Soybean Polar Lipid Extract made up of 45.7% phosphatidylcholine (PC), 22% phosphatidylethanolamine (PE), 18.4 % phosphatidylinositol (PI), 6.9% phosphatidic acid (PA), and 6.9% unknown. The soybean extract was chosen based on having the closest make-up similar to that of the endoplasmic reticulum (55.2% PC, 21.1% PE, 12.7% PI/PS, 2.5% sphingolipids and 8.5% unknown). The remaining micelle preparations of the FLAG-p7 protein and lipids were preformed by dissolving 2-4 mg

lyophilized protein in .4 M stock concentrations and diluted to a final concentration of 300 mM lipid and the sample pH was adjusted to 7.0 using 0.1*N* NaOH. Solution NMR spectra of 300 mM LMPG/0.5 mM p7 and 300 mM DHPC/0.5 mM p7 at pH 7.0 were acquired on a Varian 500 MHz spectrometer equipped with a triple-resonance probe with three-axis field gradients and a deuterium lock channel. Experiments were preformed at 50°C using a 1.5-s recycle delay. A ^{1}H-^{15}N fast heteronuclear single quantum coherence (Farrow et al. Biochemistry 1994) (HSQC) pulse sequence was used with 1024 points in *t*2 and 256 points in *t*1. The spectra were referenced to water at 4.70 ppm. These spectra were acquired for direct comparison to the already well-resolved 125 mM DHPC/ 0.5 mM p7 at pH 4.0 spectrum used in solving the solution structure of p7.

To conduct the chemical cross-linking experiment, both DSP and DSS were dissolved in DMSO and BS^3 was dissolved in either 20 mM phosphate or 20 mM HEPES buffer, pH 7.3. Protein concentration was kept between 50 and 100 nM and a 30-fold molar excess of cross-linker was added to samples. As a control, DMSO was added at similar volumes to the non-cross-linked samples. Samples were rocked at room temperature for 3 hours then quenched with 1 M Tris/HCl pH 7.3 to a final concentration of 20-50 mM Tris and incubated for 15 minutes. All samples were transferred to 1-kDa dialysis tubing and dialyzed in 20 mM HEPES, pH 7.3 overnight to remove excess cross-linker. After dialysis, 30 µL of sample was added to 10 µL of 4X LDS buffer and run on Bis-Tris 4-12% gradient SDS-PAGE (Life Technologies, https://www.lifetechnologies.com/us/en/home.html) and visualized with silver staining.

4.3.2 Purification of MSP1D1-ΔH5 and Assembly of Protein-containing Nanodiscs

A his-tagged construct of MSP1D1-ΔH5 with a sequence of GHHHHHHHDYDIPTTENLYFQGSTFSKLREQLGPVTQEFWDNLEKETEGLRQEM SKDLEEVKAKVQPYLDDFQKKWQEEMELYRQKVEPLGEEMRDRARAHVDALR THLAPYSDELRQRLAARLEALKENGGARLAEYHAKATEHLSTLSEKAKPALEDL RQGLLPVLESFKVSFLSALEEYTKKLNTQ wherein Helix 5 has been deleted to restrict the size the nanodiscs can attain. The construct was expressed in *E. coli* LB overnight starter culture containing 30 mg/L kanamycin (kan). A 500 mL flask containing TB media and 30 mg/L kan was inoculated the following morning with 3 mL of overnight culture. The cells are grown at 37 °C until optical density at 600 nm (OD_{600}) = 1.00 or more at which point they are induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and grown for 4 hours at 37 °C with low shaking speed ~ 200 rpm. Cells were then harvested at 6000 rpm for 10 min at 4 °C and resuspended in 35 mL of lysis buffer (20 mM phosphate buffer pH 7.4) along with 1 mM phenylmethanesulfonylfluoride (PMSF), 1% Triton X-100 (TX-100), and 1 mg DNaseI. Cells were then sonicated a total of 5 minutes with 5 seconds on and 10 seconds off. Sonicated fractions were spun down at 12,000 rpm for 30 minutes at 4 °C. The supernatant was then loaded onto an equilibrated Ni-NTA column and sample was washed with 10-20 column volumes of three wash buffers: 1) 40 mM Tris/HCl pH 8.0, 300 mM NaCl, 1% TX-100, 2) 40 mM Tris/HCl pH 8.0, 300 mM NaCl, 50 mM Na-cholate, 20 mM imidazole, 3) 40 mM Tris/HCl pH 8.0, 300 mM NaCl. Following the wash steps, MSP is eluted with elution buffer (40 mM Tris/HCl pH 8.0, 300 mM NaCl, 400 mM imidazole) and dialyzed overnight in distilled deionized water (ddH_2O) at room

temperature. His-tag is cleaved by adding 1:100 TEV to the protein along with 1 mM DTT. Cleavage is carried out at room temperature overnight in dialyzing buffer (20 mM Tris/HCl pH 7.4, 100 mM NaCl, 0.5 mM EDTA, 1 mM DTT). After cleavage the protein is dialyzed several times in dialysis buffer without DTT to remove all remaining DTT form the solution. Dialyzed fraction was run over Ni-NTA column to isolate cleaved MSP product. Finally, MSP is dialyzed in end buffer (20 mM Phosphate buffer pH 7.5, 100 mM NaCl, 1 mM EDTA) overnight and stored at -20 °C until used.

Sodium-cholate is added to purified MSP1D1-ΔH5 to a final concentration of 20 mM. Following the preparation of unmodified J4 p7 reconstituted in DMPC liposomes (previously described), the MSP/cholate mixture is added to a 15 mL conical tube and allowed to rock for one to three hours at room temperature. The cloudiness observed in the original liposome sample disappears immediately upon addition of scaffolding protein if the ratio of protein:lipid:MSP is correctly calculated. Following shaking, 1 g of washed Biobeads SM-2 (Biorad) was added per 1 mL of assembly mixture and shaken for 4-12 hours or overnight. Biobeads are removed by centrifugation and nanodisc-containing supernatant is collected, pooled and concentrated (if needed).

4.3.3 Characterization using Photo-induced Cross-linking of Unmodified Proteins (PICUP)

Cross-linking was carried out using long wavelength light on the unmodified protein, meaning there were no affinity or expression tags fused to the purified J4 p7 sequence with a single-site C27S mutation. The reactions were carried out in near-native conditions at pH 7.4 in 15 mM phosphate buffer. The reaction was carried out in a total

volume of 50 μL in a buffer comprised of 15 mM sodium phosphate (pH 7.4), 150 mM NaCl and 0.125 tris-bipyridylruthenium (II) dication or Ru(III)bpy_3^{2+}. Protein concentrations were maintained between 0.01 and 10 μM (consistent throughout one trial). Ammionum persulfate (APS) was added to a final concentration of 2.5 mM just before irradiating the samples. The solutions were placed in a 1.7 mL clear eppendorf tube parallel to the beam of light focused to a distance of ~5 mm from a 38 mW Xenon arc lamp. The light was filtered through 1 cm of distilled water then through a IR/UV filter. Samples were irradiated for 30 seconds (including controls) and immediately placed in "quenching buffer" or 7 μL of 4X LDS loading buffer (Life Technologies). Samples were immediately run on a 4-12% Bis-Tris SDS polyacrylamide gel and visualized using silver staining technique.

4.3.4 Analysis by Native PFO-PAGE

This method employs a similar method for SDS-PAGE. Commercially available Tris-Glycine and Native Bis-Tris gels were purchased (Life Technologies) without any SDS. Lyopholized protein samples were directly dissolved in a sample buffer containing 50 mM Tris-base, 4% (w/v) Na-perfluorooctanoic acid (PFO), 10% (v/v) glycerol, 0.0025% Bromophenol Blue with adjusted pH of 8.0 using NaOH. The running buffer contained 25 mM Tris, 192 mM glycine, and 0.5% (w/v) PFO pH adjusted to 8.5 with NaOH. Electrophoresis was carried out at room temperature and gels were visualized using coomassie staining.

4.4 Results

Thee ultimate desire of this project was to examine the oligomerization state of p7 reconstituted in the proteoliposomes, being the same sample type used for our solid-state NMR studies. To ease the difficulties associated with cross-linking membrane proteins and ensure we would be able to carry out cross-linking individual monomers to one another, a FLAG epitope tag (Figure 4.3A) was expressed as a fusion partner to the J4 subtype of the p7 protein. In addition to analysis of our NMR sample evaluation, we were interested in how the protein behaved in other common lipid environments used to classify the oligomerization state in the authored studies previously mentioned. These include DHPC, used in the Cryo-EM studies to model a hexameric structure, LMPG used as detergent control, DPC used to solve the oligomeric structure in solution NMR, and DMPC used in our studies. As mentioned, the DMPC cross-linked samples were prepared in the same manner as the protocol for preparing the solid-state sample to ensure they resembled one another as closely as possible. The principal difference between the samples, aside from the FLAG-tag, was the addition of the cross-linking agent (in DMSO) and the quenching reagent Tris-HCl. Both DMSO and Tris-HCl were added to the control sample as well. A multitude of protein oligomers were visualized in the gel in the DMPC liposome sample. SDS-PAGE analysis of the cross-linked FLAG-p7 sample (Figure 4.1) indicated that bands representing dimer through what appears to be nonamer were present in the sample. As we can see from the gel, oligomers were present in all of the various lipid environments with and without crosslinking reagents. All samples display at least a monomer and a dimer (lanes 2-9). Of particular note is the presence of a

higher order complex present in the reconstituted soy sample (Figure 4.1, lane 3) with a distinguishable band around 66.3 kDa.

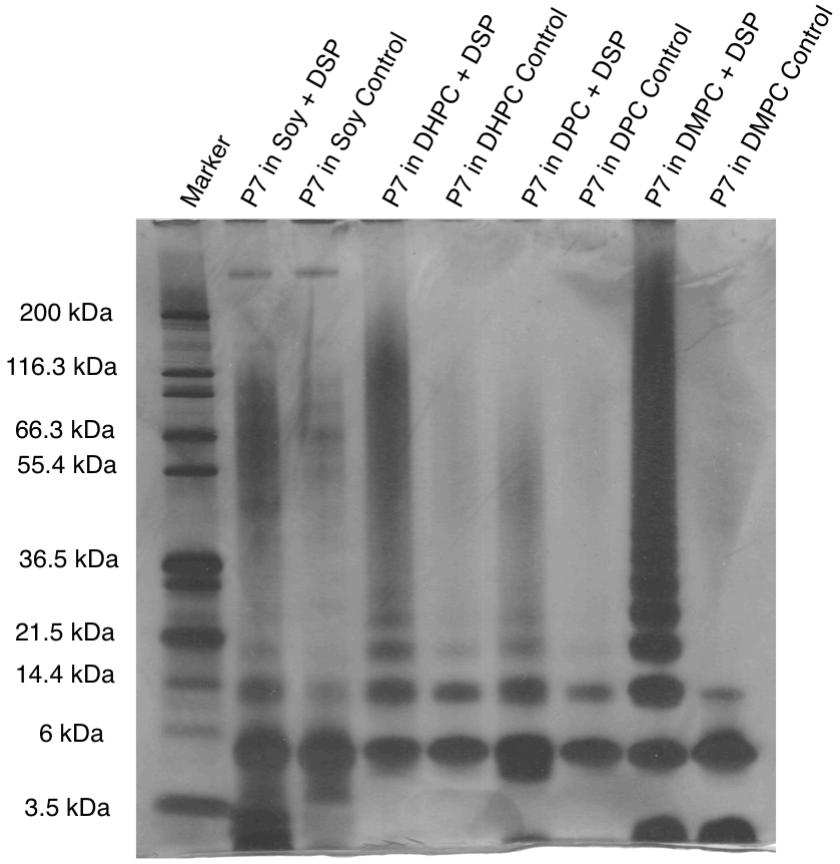

Figure 4.1 Silver-stained SDS-PAGE Analysis of chemically cross-linked p7 in various lipid environments.
Various lipid environments were selected from published studies to verify cross-linking reactions helped isolate a preferred oligomerization state. Soy extract was chosen due to its concentration resemblance to the endoplasmic reticulum lipid bilayer makeup.

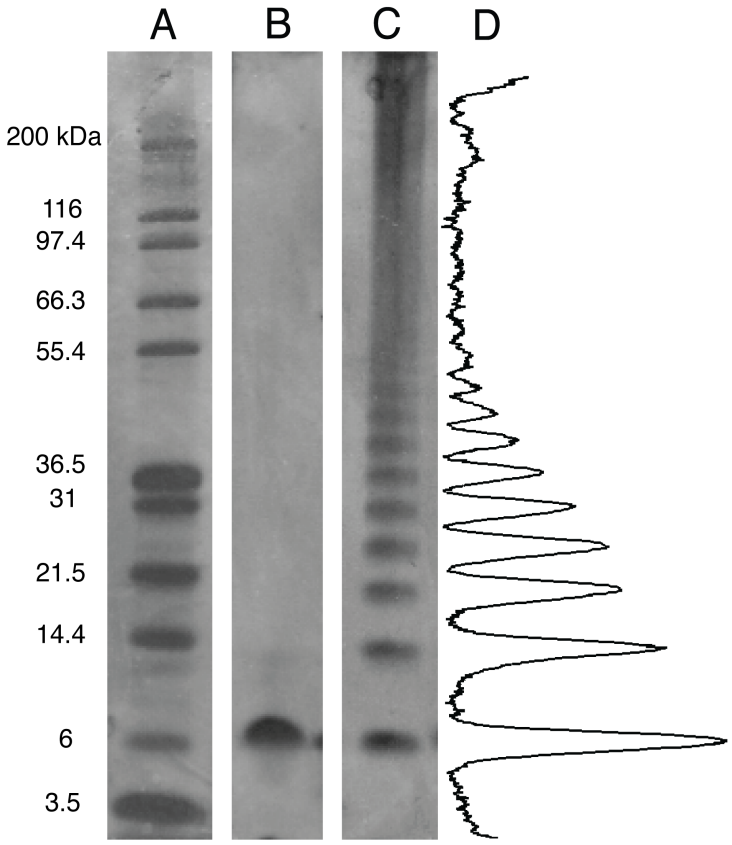

Figure 4.2: Silver-stained SDS-PAGE analysis of the FLAG-p7 construct reconstituted in DMPC liposomes with and without cross-linking.
A. The protein standard Mark 12 was used as the marker since the protein bands range from 2.5 kDa up to 200 kDa. B. FLAG-p7 in DMPC liposomes without cross-linking run as a control to show the lack of formation of higher order oligomers in liposomes. C. FLAG-p7 in DMPC liposomes after treatment with 10-fold excess of the DSP cross-linker. D. The silver-stained gel in Panel C was analyzed using a MATLAB script designed to correct the baseline and measure intensity.

To further highlight the presence of the nonamer in the DMPC sample, a lower concentration of sample was analyzed by SDS-PAGE analysis. The gel band intensities were then analyzed by converting the intensities to a readable signal generated from the MATLAB script, observed in Figure 4.2C. The baseline of the plot was compensated for due to the presence of the larger aggregated complexes, seen at the top of the gel (Figure

4.2B), which were incapable of deconvolution. We assume these larger complexes are not only higher order oligomers but also liposome-protein complexes that are too large to be separated by SDS-PAGE. These larger complexes were present even after testing several concentrations of protein and cross-linking agent in order to obtain the highest resolution of oligomers in the SDS-PAGE analysis. In contrast to existing literature cross-linking studies, we obtained a distribution of higher order oligomeric species indicating that it would be difficult to isolate a single preferred oligomeric population for the p7 protein.

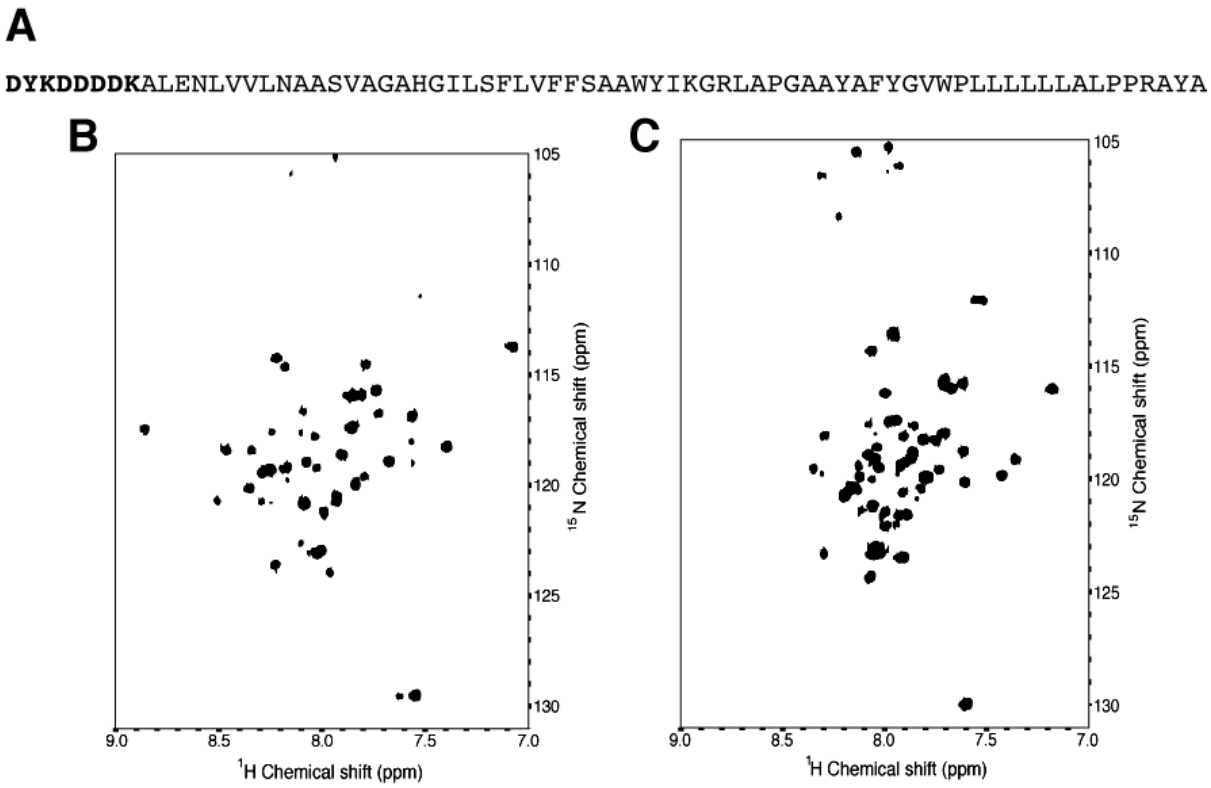

Figure 4.3: Comparison of lipid environments DHPC and LMPG at pH 7.0
A. Sequence of the J4 genotype of p7 with the FLAG-tag motif expressed on the N-terminal of the protein. B. 1H-15N HSQC of 1 mg uniformly 15N labeled p7 reconstituted in 300 mM DHPC with pH adjusted to 7.0. 128 scans were recorded on a 500 MHz magnet. C. Same as B in different lipid environment, LMPG.

Following the gel analysis, we desired to see if a change in pH would alter the HSQC spectrum of the protein. This idea was a direct result of the oligomeric structure that was published by OuYang *at al.* solved in DPC at pH 6.5. All of the solution data previously collected on our structure of p7 was in DHPC at pH 4.0. We ran HSQC spectrum of both the DHPC sample and the LMPG sample at pH 7.0. It should also be noted that both samples maintain the FLAG-tag motif as we were conducting cross-linking experiments.

We can see from the well-resolved DHPC spectrum (Figure 4.3B) that several residues are absent. All five glycine residues (^{15}N shit 110-105ppm) are gone including 20 other residues that are observed in the pH 4.0 spectrum. This is due, in part, to the slower exchange of the amide protons at higher pH values. The LMPG sample (Figure 4.3C) shows more amide-proton peak correlations than the DHPC sample indicating it as a more suitable lipid environment for physiological solution NMR studies, similar to the DPC sample. Some peaks are still missing in comparison to the HSQC spectrum taken at pH4.0 (Figure 2.9A). However, both spectra display narrow line-widths as they both maintain well-defined cross peaks, indicating the presence on a homogeneous monomeric protein species. If the sample was in the form of a hexamer reconstituted in a micelle, the molecular weight would most likely exceed 60 kDa leading to an extremely long τ_C, producing very broad peaks beyond the detection of our 500 MHz. The detection of the 800 MHz fitted with a cryoprobe, would likely be surpassed as well. We are still left wondering, even with deuteration of lipids, and segmented labeling of the proteins assembling the hexamer, how a 60 kDa complex produces highly-resolved individual peaks a 900 MHz NMR.

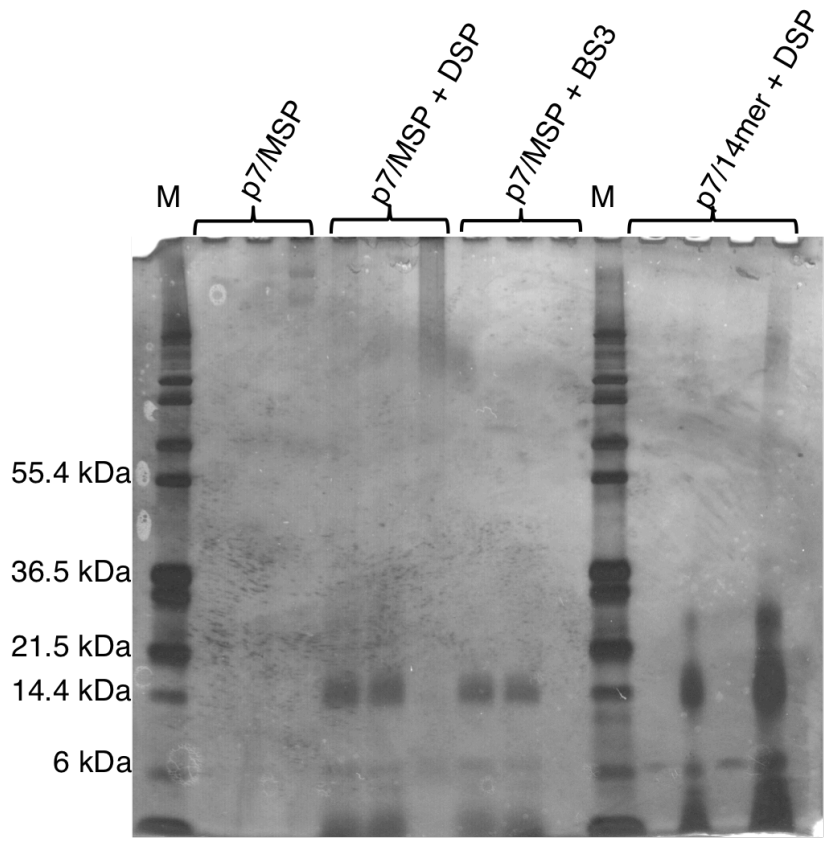

Figure 4.4: Silver-stained SDS-PAGE analysis of chemical cross-linking preformed in nanodiscs
Roughly 5 μg of p7 protein was loaded into each lane with each set of three lanes containing the suspension, supernatant and the pellet of the spun down samples. Following the second marker lane are the p7/14 mer nanodics at various concentration with and without DSP cross-linker.

After unsuccessful attempts were made to isolate a preferred, or highest attainable oligomeric complex of p7 in the solid-state NMR sample, we moved to a more isolated system, through the use of nanodiscs to help control the cross-linking reaction. As outlined in the section 4.3, the discs were prepared using the same method to prepare the liposome sample. The differentiation came about with the addition of either the membrane scaffolding protein, MSP1D1ΔH5, or a small 14 residue peptide, 14-mer.

Upon the addition of either peptide, following the appropriate calculation for having a 1:50 ratio of MSP:DMPC or a molar ratio of 1.67 of DMPC to the 14-mer peptide.

Initially, chemical cross-linkers were added to the nanodisc samples and SDS-PAGE analysis was carried out in the same fashion as the former experiments. Upon analysis of the gel (Figure 4.4) we can observe that we have an isolated reaction. In these experiments, the nanodiscs were prepared in such a way so as to attain one protein per two nanodiscs. We should have an empty disc for every protein reconstituted in one disc. We can see that we have obtained a dimeric species, with a distinguishable band at 14.4 kDa; however, a caution is warned for lanes 12-15 as the bands at 14.4 kDa may be cross-linked peptide. It becomes even more skeptical when we notice the p7-nanodisc complex, lacking cross-linker, was run (Fig 4.4 lanes 2 and 3, in which the MSP1D1ΔH5 band (~21 kDa) does not show up. Clearly this experiment requires additional controls to help identify exactly what is being cross linked.

We wanted to take an extra step and eliminate the FLAG-tag motif having the belief that this may be impeding the native oligomerization state of the protein. Since we removed the tag, we needed an alternative way to carry out the cross-linking since all of the chemical cross-linkers used thus far have been primary amine reactive with the FLAG-tag motif attached to p7. A less common method of cross-linking via photo-induced cross-linking of unmodified proteins or PICUP was selected for these experiments. The way in which this cross-linking works is that the ruthenium(II) tris-bipyridyl dication [$Ru(II)bpy_3^{2+}$], an efficient light-harvesting molecule, generates a reactive intermediate in the presence of ammonium persulfate. This intermediate is known to produce an excited state able to donate an electron to the persulfate, resulting in

cleavage of the O-O bond. The resulting products, Ru(II) one-electron oxidant, the sulfate radical and a sulfate anion act in conjunction to generate a tyrosyl radical as an initiating step wherein the subsequent removal of the hydrogen vicinal to the hydroxide group is carried out by the sulfate radical thereby consummating the reaction.

This concept allows us to take advantage of tyrosine residues already present in the amino acid sequence. For controls, empty nanodiscs were made for testing the crosslinking reaction on them in addition to having a non-hydrophobic protein, GST, as a positive control to ensure the reaction was working properly. We can see form the SDS-PAGE analysis (Figure 4.5) that our reaction does in fact proceed properly. For reference purposes, a concentrated fraction of MSP1D1ΔH5 was run for size comparison (Figure 4.5 lane 10). We can see the reaction produced a distinct band around 36.5 kDa but that this band is also present in the empty nanodiscs. We can were there are higher order oligomers in almost all of the samples, including the non-cross linked p7/DMPC sample observed between 36.5 kDa and 55.4 kDa (very faint) indicating that we have oligomers, just that there is most likely a broad distribution.

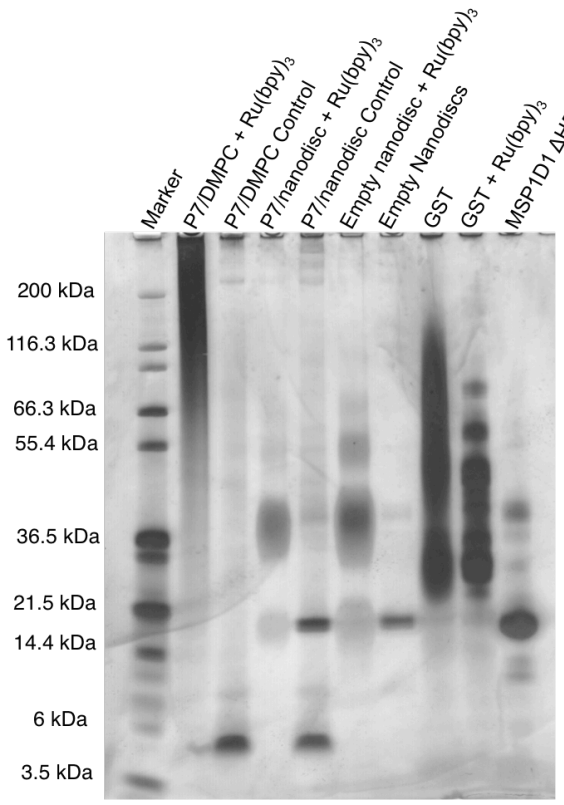

Figure 4.5: Silver-stained SDS-PAGE analysis of PICUP cross-linking method carried out in MSP1D1ΔH5/p7/DMPC nanodiscs

Crosslinking conditions were preformed at 15 mM phosphate buffer, 150 mM sodium chloride, 2.5 mM APS and 0.125 mM Ru(III)bpy_3^{2+}. Samples were cross linked upon exposure to high wavelength xenon light and quenched by add the reducing 4X LDS gel loading dye and immediately analyzed in SDS-PAGE.

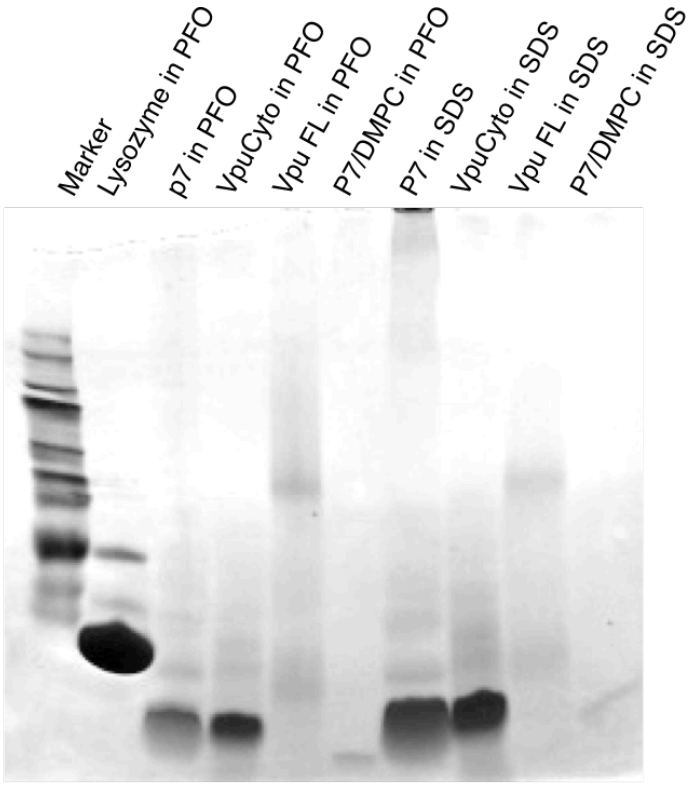

Figure 4.6: Coomassie stained native PFO-PAGE analysis of p7 reconstitued in PFO, DMPC and SDS
Samples were prepared immediately before running the gel. Lysozyme, p7, Vpu cyto, and Vpu full-length were all dissolved in either PFO or SDS to run on the Native Tris glycine gel with Tris glycine running buffer containing 1.5% PFO at pH 8.0. Gel was visualized using coomassie staining.

One final experiment was to try and run the samples on a native gel to see if we could obtain answers without the use of cross-linkers. Many conditions were attempted before finding the ease in the use of the per-flouorooctanoic acid or PFO-PAGE developed by Ramjeesingh and coworkers in 1999. The PFO acts as an extremely mild detergent allowing proteins to stay in near native conditions. The samples are dissolved in the PFO loading buffer and run on native gels using PFO-Tris glycine buffer. In order to have some comparison as to the size the p7 protein was running at we selected several controls. The first was lysozyme since we know it runs at about 14 kDa. The next logical

selection was the Vpu full-length protein as we know it runs as a tetramer in native and SDS-PAGE analysis. Vpu cyto was run as a size comparison for p7 as weights in close to 7 kDa. Finally, the samples were prepared in both PFO loading buffer and SDS loading buffer. From the gel (Figure 4.6) we can see that we run into essentially the same issue as with the cross-linkers by observing multiple oligomerization states present within the same sample (Figure 4.6 lane 3 and 7).

4.5 Discussion and Conclusion

From the analysis of all of these methods, it becomes clear that this protein is extremely hydrophobic leading to a highly promiscuous oligomeric formation. In light of these studies, we can see that there needs to be a standardized method for evaluating the oligomerization state of such small complexes. Current techniques used for a multitude of other oligomerization characterization include analytical ultracentrifugation, cryo-electron microscopy and SEC. Due to the small size difference between different oligomers all of these methods become extremely difficult to apply to the analysis of a preferred state. Our lab also attempted cryo-EM studies with little success. It appeared that we had the existence of tetrameric or pentameric complexes reconstituted in the DMPC liposome samples; however, the results were less then definitive and poorly processed which is why they are not included in this work.

As for observation of oligomerization states in other proteins, the definitive characteristics were analyzed in structural studies. This is seen in the case of M2 were two non-equivalent chemical shift peaks were observed for the isoleucine residue eluding to the fact that the M2 oligomer existed as a dimer of dimers. (REF) This, however, is a

convenience that is unfortunately not the case for the p7 protein. It is entirely possible that there is the existence of multiple oligomeric states for p7 within the cell that serve as different functions throughout the virus lifecycle. This is well demonstrated in the MD simulation analysis carried out by Chandler et al. wherein they identified that the plasticity of the protein as well as it's minimalist approach makes it an ideal candidate for the coexistence of multiple oligomeric states. This has been observed in other oligomeric ion channels including Vpu of HIV (Lu *et al.* 2010 & Montal 2003) and the antimicrobial peptide alamethicin (Tieleman *et al.* 2002). This implication stems from the idea that because the p7 monomers are not covalently bound, but instead interact using van der Waals and hydrogen-bonding, it becomes clear that it is easily possible to have the existence of multiple oligomeric arrangement for which the inter-subunit contact is sufficient enough to support a stable structure. This idea directly supports our finding that with and without the addition of cross-linkers, we observe a multitude of oligomerization species.

Chapter 5. Expression and Purification of HCV Genotype 2a Isolate JFH-1 p7

5.1 Abstract

Being that there are seven major genotypes and over 100 relevant subtypes, there becomes a need to investigate the structure and dynamic differences among the best-studied genotypes of p7. Importantly, different HCV genotypes display different treatment response rates indicating the clinical course of infection drastically differs for each patient. In addition to host factors, viral factors play a crucial role in the outcome of treatment for the infection. The comparison of well-characterized genotypes of p7 provides a basis for the relation between specific structural differences and different inhibitory effects of channel-blocking drugs, a well-observed trend across the literature. We need high-resolution structural data in order to identify key structural differences among the most highly variant sequences in order to make rational decisions on how to potentially inhibit the viral packaging properties of this protein. It has previously been shown that specific mutations to p7 result in a significant loss of channel activity, indicating that this function of p7 is sequence specific.

5.2 Introduction

In order to understand why different genotypes of the p7 protein confer different resistance of channel-blocking compounds, we need to probe both sequence and structural variation that may contribute to this anomaly. First, we need to understand structural differences that exist between the most divergent genotypes of p7. We can accomplish this by comparing the J4, EUH1480, and JFH-1 strains. Both EUH1480 and

JFH-1 strain maintain a 54% and 59% identity and 79% similarity to the J4 strain, giving these two genotypes the highest degree of sequence separation from J4. In order to compare the J4 strain with the JFH-1 strain the purification and sample preparation must also be optimized. The structural characterization can begin once the protein is successfully purified and sample preparation is established. Since the EUH1480 monomer structure is already published, comparison with this protein can also help elucidate structural discrepancies among the genotypes. In addition to these structural comparisons, a comparison of known channel blocking compounds can be evaluated for each genotype. Taken together, these studies have the potential to elucidate why genotypes have varying activity and drug binding affinities.

5.3 Materials and Methods

5.3.1 JFH-1 Gene synthesis and Vector Cloning

Initial efforts were made to synthesize the JFH-1 genotype of p7, sequence: ALEKLVVLHAASAANSHGLLYFAIFFVAAWHIRGRVVPLTTYSLTGLWPFSLLL MALPRQAYA. In the final sequence, the cysteine residues were mutated to serine and the methionine was mutated to an alanine so that cyanogen bromide cleavage could be carried out in the purification process. Synthesized oligonucleotides obtained from Allel, corresponding to the full-length sequence with modifications, were used to amplify products that could be ligated into the expression vector pHLV containing the fusion protein TrpΔLE used to drive the protein expression into inclusion bodies. The forward primers included a restriction site for HindIII endonuclease and a methionine codon for cyanogen bromide cleavage. The reverse primer included two stop codons at the end of

the JFH-1 coding region and a BamHI endonuclease restriction site. Altogether each primer was approximately 200 base pairs (bp) with 80 bp overlap. PCR was carried out using a variety of DNA polymerases including Pfu turbo, Vent and Tac as well as a variety of thermocycle protocols.

A PCR product was never obtained and the designed DNA sequence was ordered from Genewiz with the previously listed modifications. After obtaining the DNA, supplied in an empty plasmid, the sequence and pHLV plasmid were digested with the restriction enzymes Hind II and BamHI at 37°C for 1 hour and the products were separated on a 1.5% agarose gel. The band corresponding to the digested DNA and cut plasmid size were excised and purified using a gel extraction kit.

The DNA insert was ligated into the cut plasmid using Quick T4 DNA ligase (New England Biolabs). The reaction was incubated at room temperature for 5 minutes, chilled on ice and transformed into DH5α cells (Life Technologies) and plated on LB/carbenicillin agar media. Colonies were selected for growth in 5 mL LB cultures at 37°C until OD_{600} reached 0.6-0.8. The plasmids from the cells were purified using a mini-prep kit (Qiagen) and confirmed by sequencing (Eton Bioscience).

Following successful isolation of pHLV-p7 plasmid, the DNA sequence coding for the full-length JFH-1 p7 construct was modified twice. The first modification was to insert a Factor Xa cleavage (sequence: IEGR) site in place of the methionine residue in order to carry out on-the-column cleavage and avoid the inefficient cyanogen bromide cleavage. The second modification was to insert an N-terminal BamHI restriction site, a C-terminal Factor Xa cleavage site followed by a 6X histadine tag and a C-terminal Aat II restriction site. The new construct was excised and purified as previously described and

cloned into the pGEX-6P-1 vector containing the fusion partner GST with an N-terminal PreScission protease cleavage site (GE).

5.3.2 Expression Optimization

Plasmids with the proper DNA sequence were transformed into either BL21(DE3) pLysS, C41(DE3), or C43(DE3) cells depending on expression checks. Transformed cells were plated on LB medium supplemented with carbenicillin and placed at 37°C overnight. Cell-stock solutions were made by inoculating 5 mL of LB/carbenicillin with a single colony, grown to OD_{600} of 0.4 and sterilized glycerol was added to a final concentration of 17%. Fifty microliters of the cells were aliquoted into sterile 1.7 mL cryo tubes (Thermo Fischer) situated in a dry ice and isopropanol bath and placed in the -80°C freezer until needed for expression growth.

Starter cultures were prepared by adding 10 μL of the cell stocks to 5 mL LB/carbenicillin. The pre-culture was placed in the shaker for 2-4 hours at 270 rpm, 37°C and either 400 μL or 1 mL was inoculated into a 100 mL or 500 mL sterile M9 (40 mM disodium hydrogen phosphate, 20 nM potassium dihydrogen phosphate, 9 mM sodium chloride, and 1 mM $MgSO_4$, 0.1 mM $CaCl_2$ and 1% LB in a 2-L baffled flask) overnight flask, respectively. The flask was placed in a shaker for 16 hours at the same conditions listed previously. The culture was then transferred to either a 1 L flask or a 4.5 L of M9 media in either a 2-L baffled flask or batch/continuous bioreactor (New Brunswick Scientific, Bioflo), respectively. Specifically for the bioreactor growth, the pH was maintained at 7.0 by the addition of either 10 *N* NaOH or 3 M HCl, agitated at 500 rpm and the temperature was maintained at 37°C until the cells reached OD_{600} of 0.8. Protein

expression was induced by the addition of 0.5 mM IPTG. The temperature was then lowered to 25°C and the cells were allowed to grow for 18 hours. The cells were then harvested by transferring them to 1 L centrifuge bottles, spun at 6500 rpm for 20 minutes at 4°C. Cells were stored at -80°C until cell lysis is carried out.

5.3.3 Purification by Affinity and Fast Protein Liquid Chromatography

Cell lysis was performed by disrupting the cell pellet using a spatula and 30 mL of resuspension buffer I (1 mM EDTA, 50 mM Tris/HCl pH 8.0, 1 mM NaN_3, and 15% (v/v) glycerol). The resuspended cell pellet was transferred to a 50 mL polypropylene centrifuge tube and lysed using sonication. (550 Sonic Dismembrator) on ice for 5 minutes, 5 seconds on 10 seconds off to prevent overheating. The cells were then centrifuged at 17000 rpm spinning speed at 4°C for 30 minutes. The supernatant is decanted and the remaining pellet is resuspended in 30 mL of resuspension II buffer (1 mM EDTA, 50 mM Tris/HCl pH 8.0, 1 mM NaN_3, 1%(w/v) deoxycholic acid, and 1% IGEPAL CA-630). The suspension was again sonicated for 5 minutes, spun down at 17,000 rpm at 4°C for 30 minutes and the supernatant was decanted. The pHLV-JFH-1 p7 pellet, containing the isolated inclusion bodies, is flash frozen in liquid N_2 and lyophilized until completely dry. The pGEX-6P-1-JFH-1 p7 pellet is broken up in 30 mL binding buffer (1% (w/v) SDS, 50 mM Tris/HCl pH 8.0, 500 mM NaCl) and rocked at room temperature overnight.

The pHLV modified with a Factor Xa cleavage site was purified via Ni-NTA in guanidine denaturing conditions. Briefly, following lysis, the inclusion bodies were solubilized in 30 mL binding buffer (6 M guanidine hydrochloride (Gdn/HCl), 0.5 M

sodium chloride (NaCl), 20 mM Tris/HCl pH 8.0, 5 mM imidazole) and rocked overnight. Inclusion bodies were sonicated on ice as described and loaded onto an equilibrated Ni-NTA column (30 mL) with two volumes of binding buffer. After the protein was loaded, nonspecific bound proteins were eluted from the column using four column volumes of wash buffer (6 M Gdn/HCl, 0.5 M NaCl, 20 mM Trsi/HCl pH 8.0, 50 mM imidazole). Following the wash, the column was buffer exchanged with Factor Xa cleavage buffer containing 0.1% DPC, 250 mM NaCl, 20 mM Tris/HCl pH 7.4). Factor Xa was added at a 10 molar excess and cleavage pursued overnight at room temperature on a rocker 678. Following cleavage the elution contained the cleaved fractions. SEC FPLC was used to clean up the cleavage product.

For the pGEX-6P-1 construct, the partially dissolved inclusion bodies were sonicated on ice as described and filtered using a Steriflip filter unit. The protein solution was then loaded onto a Ni-NTA (Qiagen) column (30 mL) that was equilibrated with two column volumes of binding buffer. The solution was batch-bound to the column for 3 hours or overnight before washing the column with 3 column volumes of Wash buffer (1% SDS, 50 mM Tris/HCl pH 8.0, 500 mM NaCl, 5 mM Imidazole). Following the was step the column was equilibrated with 10 column volumes of PreScission protease cleavage buffer (1mM EDTA, 50 mM Tris/HCl pH 7.0, 150 mM NaCl, 1 mM DTT, 0.05% (v/v) Triton X-100 chilled to 4°C) On-column GST cleavage is carried out by adding 2 units enzyme/100 μg of bound GST-p7 protein and rocked at 4°C for 12-16 hours. Cleaved His-tagged p7 remains on the column while uncleaved and fusion protein GST elutes out as flow-through. His-tagged p7 is eluted using 3 column volumes of Elution buffer (50 mM Tris/HCl pH 8.0, 500 mM NaCl, 1% SDS, 250 mM Imidazole).

The eluted protein is concentrated using a 3-kDa cutoff Millipore Amnicon Ultra-15 concentrator spun at 4500 rcf at 23°C until volume reached 5 mL.

The fusion TrpΔLE of the pHLV construct is cleaved from target protein by treatment with cyanogen bromide. Approximately 10 mg/ml of the fusion construct was dissolved in 80% formic acid, and 100 mg/mL cyanogen bromide was added to the solution in a lyopholizing flask. The light-sensitive reaction is parafilmed and covered with foil proceeded for 3 hours at room temperature with constant shaking. The reaction is quenched by adding 3-times starting volume of 1*N* NaOH and rocked for 10 additional minutes. The cleavage product is then transferred to a 1-kDa dialysis bag and dialyzed in ddH_2O with continuous water changes until the solution becomes neutral. The solution was then transferred back to the lyopholizing flask, frozen in liquid N_2 and placed on the lyopholizer overnight.

Both constructs were purified using size exclusion chromatography as a final step in the purification procedure. The dry pHLV protein was dissolved in 2% SDS (20 mg/mL) and placed in a bath sonicator for 10 minutes. The concentration was diluted by adding FPLC buffer (20 mM sodium phosphate, 4 mM SDS, 1 mM EDTA, 1 mM NaN_3, pH 8.2) and loaded onto a Hi-Prep 26/60 sephacryl S-200 sizing column. The elution [15] profile was monitored by the absorbance at 280 nm and fractions containing pure p7 were collected and observed by SDS-PAGE and Mass Spectrometry. The pooled pure fractions were either dialyzed in a 1-kDa-size cutoff dialysis bag or reconstitiued in DMPC liposomes using a 1:6 protein to lipid (w/w) ratio. Precipitated protein was collected by centrifugation and lyophilized under vacuum overnight.

5.3.4 Mass Spectrometry and NMR Spectroscopy Sample Preparation

Mass Spectrometry analysis was carried out using matrix-assisted laser desorption ionization time-of-flight spectrometry preformed on a Voyager DE-STR instrument. The purified, dialyzed and lyophilized protein was dissolved in dissolved in 50% acetonitrile and 0.1% trifluoroacectic acid and mixed 1:1 with a matrix solution of 3,5-dimethoxy-4-hydroxycinnamic acid dissolved in 30% acetonitrile and 0.1% trifluoroacetic acid.

NMR samples were prepared by dissolving the protein powder in a 0.4 M 1,2-dihexyl-1-sn-glycero-3-phosphocholine (DHPC) solution with vortexing. Final concentration for the NMR sample was approximately 0.5 mM protein, 125 mM DHPC and 10% D_2O. The pH of the sample was adjusted to 4.0 with the addition of small amounts of 0.1 *N* NaOH or 0.1 *N* HCl. Using along transfer pipet, the sample was transferred to a 5 mm ×180 mm NMR tube.

Solution NMR spectra of uniformly ^{15}N-labeled p7 constructs were acquired on a Varian 500 MHz spectrometer equipped with a triple-resonance probe with three-axis field gradients and a deuterium lock channel. Experiments were preformed at 50°C using a 1.5-s recycle delay. A ^{1}H-^{15}N fast heteronuclear single quantum coherence (Farrow et al. Biochemistry 1994) (HSQC) pulse sequence was used with 1024 points in *t*2 and 256 points in *t*1. The spectra were referenced to water at 4.70 ppm.

5.4 Results and Discussion

The cloning initially carried out warranted an intervention by way of corporate purchase of a synthesized gene. A few months were spent on trying to synthesize the gene with three different methods of gene synthesis. After a significant amount of time

failing, it was fiscally responsible of us to simply order the synthesized gene. The gene came supplied in an empty pUC17 vector in which we simply needed to cut and amplify our gene in order to have significant concentration needed to carry out ligation into our desired expression vector. A successful digest was run using the restriction enzymes of BamHI HF and HindIII (Figure 5.1B). Following the digest, the gene was extracted and purified from the 1.5% agarose gel (Figure 5.1A). The pHLV vector was also successfully digested with the same enzymes and extracted and purified from a 1.5% agarose gel (Figure 5.1C). Both the cut DNA and vector were mixed in 1X quick ligation reaction buffer (NEB, www.neb.com) and quick T4 DNA ligase was added. The reaction was incubated for 5 minutes at room temperature and transformed into DH5α cells. Colonies were selected the next morning for LB culture growths from which the DNA was extracted and sent for sequencing.

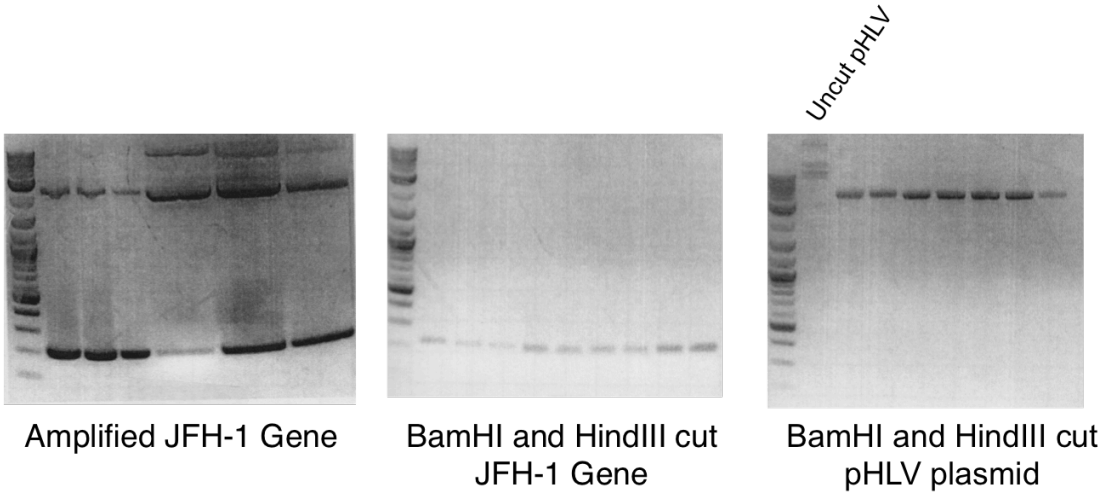

Figure 5.1: 1.5% agarose DNA gels for the analysis of restriction enzyme cleavage products.
During the molecular cloning of JFH-1 gene into the pHLV plasmid, several DNA gels were run in order to separate our gene and cut plasmid to carry out the ligation.

A positive result was obtained the first time so we could begin expression tests. The cell lines that are most amenable to membrane protein cell growth consist of BL21(DE3) plysS, C41(DE3), and C43(DE3) cells as these cells have been optimized for expressing toxic and membrane proteins effectively. All three cell lines are tested for expression checks wherein the cells are induced at OD_{600} 0.600 and grown for four hours. After each hour a sample is taken and concentration is normalized for all four samples before running an SDS_PAGE analysis. We can see from Figure 5.2 that we have effective and strong expression grown in C43 cells.

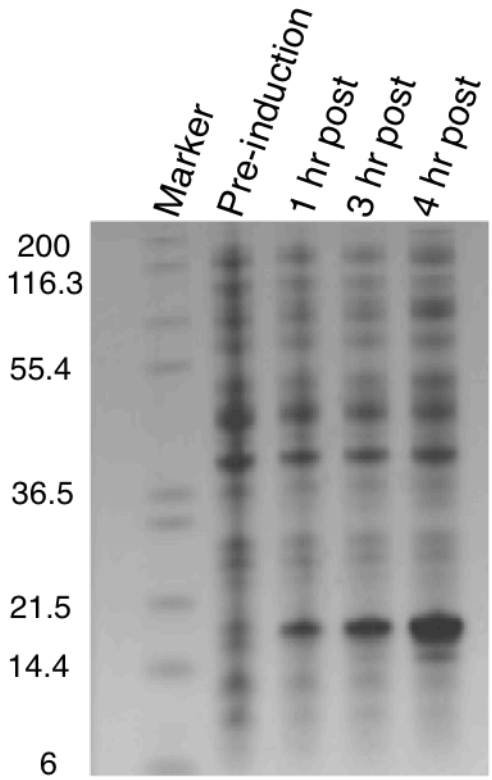

Figure 5.2: SDS-PAGE analysis of expression check for pHLV JFH-1 p7 protein
The protein was overexpressed in C43(DE3) cells and grown until it reached an OD_{600} of 0.6 at which point it was induced with 1 mM IPTG. Gel samples we collected for four hours at 37°C post-induction and normalized concentrations were loaded on the gel to ensure that our expression was increasing over the four hour growth.

Following optimized expression checks, we begin to optimize the growth and purification methods. Initially a similar protocol for growing and purifying the J4 genotype was followed, assuming this protein would behave similarly. Initially we scaled our growths up to a bioreactor system to assist in improving our yield and make our growths more cost-effective. A bioreactor maintains significant advantage over conventional cell growths. This is due directly to the controlled growth environment, which maintains temperature, pH, agitation rate, and dissolved oxygen. Since *E. coli* are sensitive to small temperature changes and labeling of the proteins increases sensitivity towards pH, precise control of these factors becomes necessary for an optimal growth. In comparison to a standard 1 L growth the cell density has the potential to be increased from OD_{600} 1.9 to 2.7 when growing in a bioreaction. Additionally, the increase in purified protein yield is increase more than 3-fold, resulting in 1.5 mg/L in comparison to 0.4 mg/L in a double-labeled growth 2 L flask.

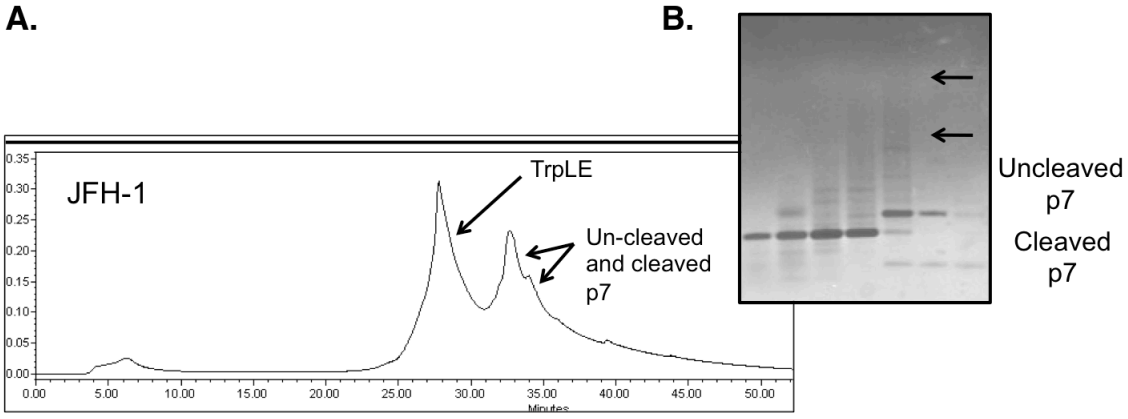

Figure 5.3: HPLC Purification of pHLV JFH-1 p7 construct
A. HPLC chromatogram with the following parameters: sample dissolved in 20:20 HFIP:TFE loaded onto semi-prep diphenyl column (300 mm) with a flow rate of 2 mL/min and a gradient of 10% to 90% B (90% ACN, 10% H_2O, 0.01% TFA) over 70 minutes. B. SDS-PAGE analysis of HPLC fractions with cleaved p7 co-eluting with uncleaved p7 shown in the last three lanes.

Once the growth was optimized, the removal of the TrpΔLE fusion protein following CNBr cleavage is our first and final purification step. This proved to be significantly more difficult than expected. Over 100 HPLC conditions were attempted to assist in the separation of the fusion tag from the pure JFH-1 construct. We can see from the gel and HPLC chromatogram displayed in Figure 5.3 that the three products co-elute as a mixture of cleaved JFH-1 p7 product with the residual uncleaved and TrpΔLE fusion protein. After multiple attempts to separate the cleaved and uncleaved forms of the p7 expression construct, it was decided that the protein needed an alternate cleavage method based on the SDS-PAGE analysis (Fig 5.3B) highlighting the inefficient cleavage of p7 from the fusion protein following CNBr chemical cleavage.

We elected to insert a Factor Xa cleavage site in between the TrpΔLE and the p7 protein (Figure 5.4A). Several protocols have been established for on the column cleavage of membrane proteins using this the Factor Xa enzyme. This approach is relatively straightforward for obtaining pure JFH-1 p7 involving a simple affinity column purification and cleavage. The genetic modification was made simply through the use of the Qaigen QuikChange Lightning kit and protocol. The protein expression and growth (Figure 5.4B) were duplicated from the original optimization and the protein was purified over the Ni-NTA column where our on the column cleavage was facilitated. The purification and cleavage of the construct worked beautifully (Figure 5.4 C) with the exception of one caveat: the protein appeared to dimerize in the cleavage buffer it was eluting off the column in as observed in lanes 4-7 in Figure 5.4C. The cleavage buffer for

the Factor Xa cleavage is made with an extremely mild detergent, HPC that would in deed support a more native fold of the protein in comparison to .1% SDS.

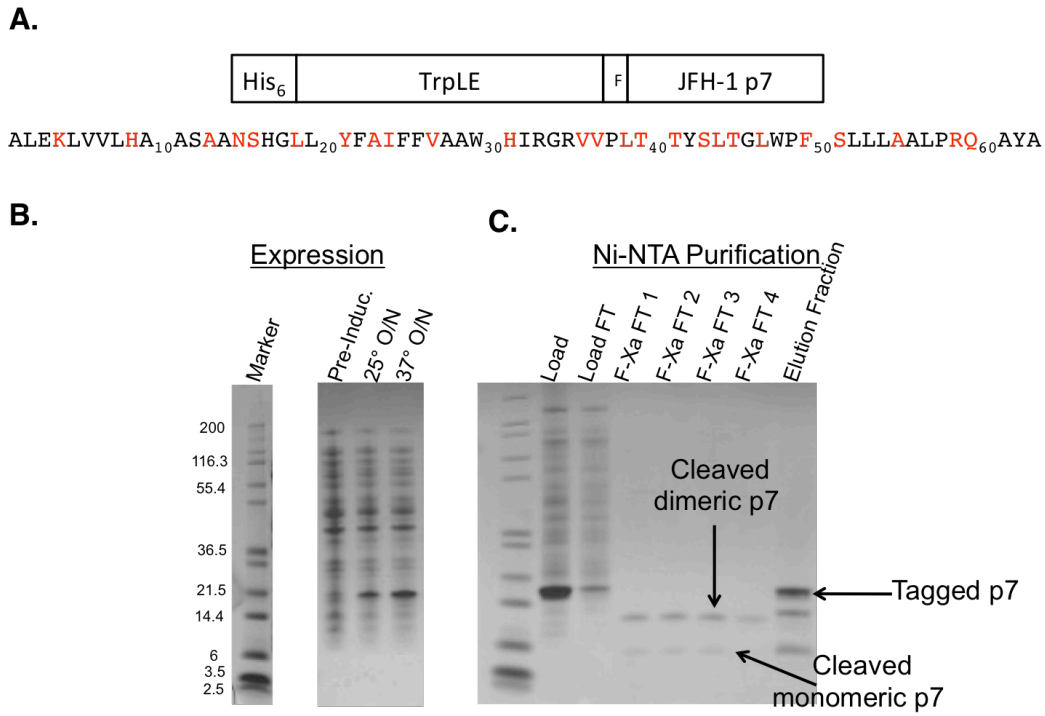

Figure 5.4: Design of Factor Xa and accompanying SDS-PAGE of expression and initial affinity Ni-NTA purification
A. Newly designed construct showing the insertion of the Factor Xa cleavage site B. Expression and growth conditions showing approximately the same amount of protein in the overnight growth at 25°C compared to four hours at 37°C C. Ni-NTA purification and successful cleavage of the TrpΔLE- p7 construct.

We were unsure if the protein was in fact dimerizing or if it was still co-eluting with the fusion protein, TrpΔLE. We decided to run a mass spectrometry (MS) to see if we could definitively say if it was a dimer or the TrpΔLE fusion protein, both running around 14 kDa on the SDS-PAGE analysis. As observed in the MS, Figure 5.5, we can see that we are in fact co-eluting the protein with the fusion protein as the mass of 14.2 kDa exactly matches the mass of the TrpΔLE. This was grossly disappointing but we decided to run an HSQC to see how the protein behaved in 125 mM DHPC conditions at

pH 4.0. The sequence is displayed above the spectrum to provide some residue reference (Figure 5.6.A). It becomes abundantly clear that there are a significantly larger number of peaks than what would typically be observable for a 63 amino acid residue protein. The initial spectrum also fortified our interpretation of the MS data showing that we were in fact co-eluting with the fusion protein and a new expression system should be evaluated to attain a pure JFH-1 sample.

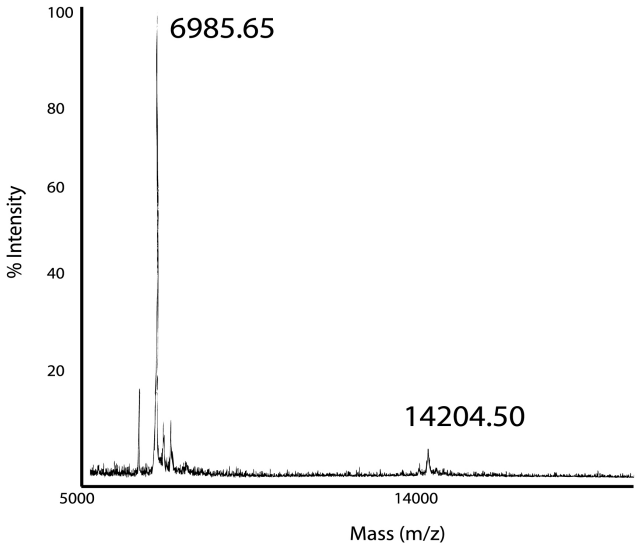

Figure 5.5: Mass Spectrometry analysis of Factor Xa Cleavage product of pHLV JFH-1 TrpΔLE-p7 construct.

A.

ALEKLVVLHAASAANSHGLLYFAIFFVAAWHIRGRVVPLTTYSLTGLWPFSLLLAALPRQAYA

B.

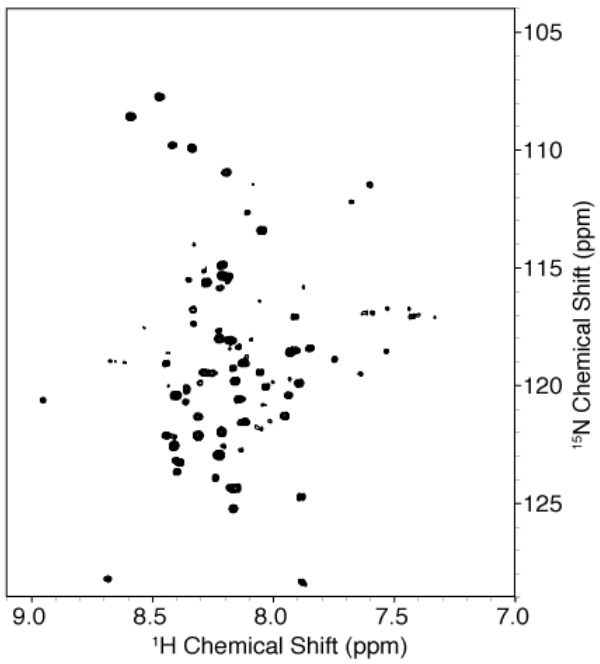

Figure 5.6: Fist viable HSQC spectrum taken of pHLV JFH-1 p7 post Factor-Xa cleavage The relatively pure sample from the on the column cleavage using Factor Xa was run in the NMR to check if we actually had the presence of dimeric complex as precieved from the gel. The sample was dissolved in 125 mM DHPC, pH 4.0 with 128 scans.

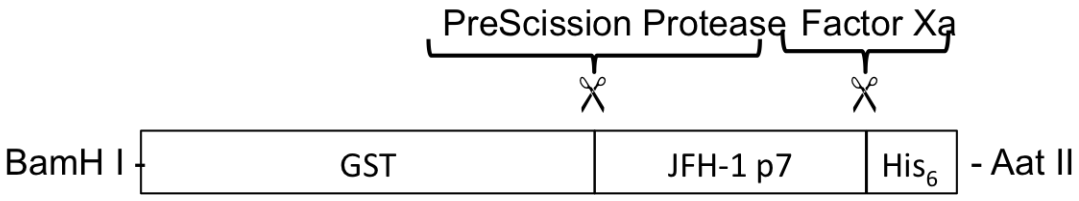

Figure 5.7: General construct of the new expression plasmid for the JFH-1 p7 genotype The JFH-1 construct was excised from the pHLV plasmid and cloned into the new pGEX-6P-1 plasmid with the addition of a C-terminal Factor Xa cleavage site and a C-terminal 6X His tag.

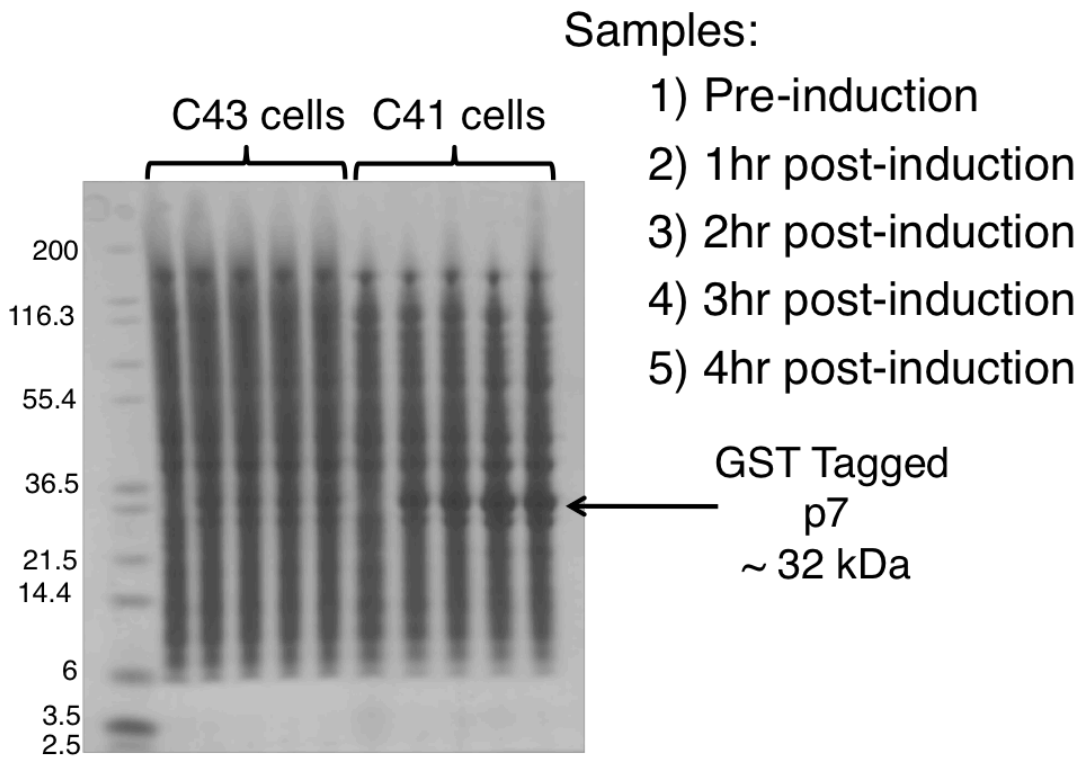

Figure 5.8: SDS-PAGE analysis of expression check for newly obtained JFH-1 pGEX-6p-1 p7 construct.
Cells grew best and express the most protein in the C43(DE3) cell line (seen in lanes 8-11).

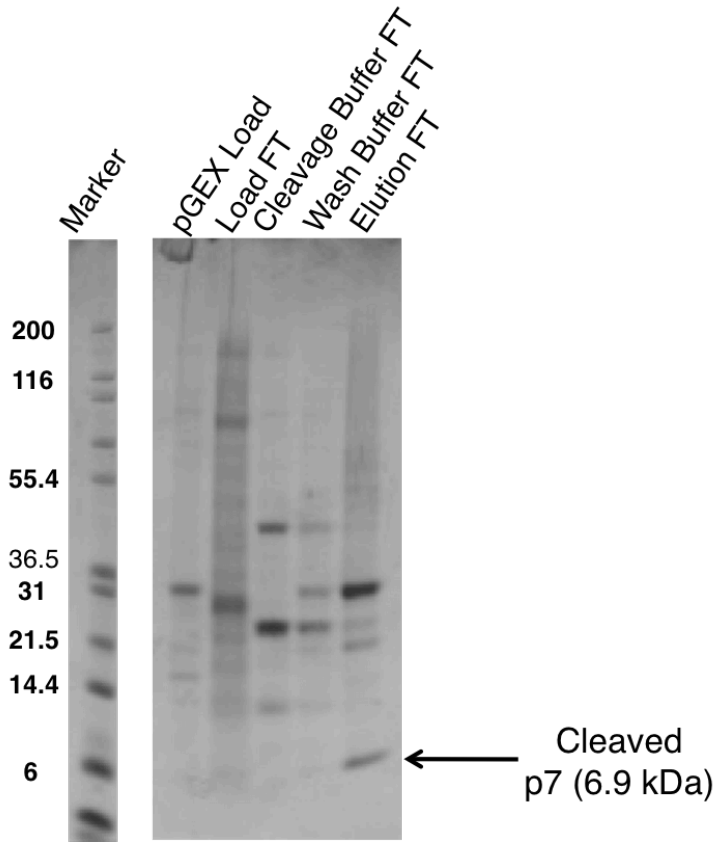

Figure 5.9: SDS-PAGE analysis of PP on the column cleavage of the pGEX-6P-1 JFH-1 p7 construct.
Analysis consisted of the fractions obtained upon purifying the protein using affinity Ni-NTA purification wherein the PP cleavage reaction took place overnight at 4°C on the column. It is quite apparent the cleavage was efficiently carried out but that additional purification methods are needed to get a highly pure sample for NMR studies.

As the fusion protein partner posed as are biggest issue in achieving high-quality separation suitable for NMR studies, we wanted to ease the separation and purification process by cloning the gene into an entirely different expression vector. The GPCR, CXCR1, expression and purification had already been successfully accomplished using a GST-expression vector. We quickly cloned our construct into the GE pGEX-6P-1 plasmid, which maintained a PreScission Protease (PP) cleavage site between the GST

tag and expression construct (Figure 5.7). We elected to place a Factor-Xa cleavage site on the c-terminus of the sequence in the case that the 6X His tag impeded our NMR studies and was required to be removed. Expression, growth and optimization was carried our as previously described. Notably, we attain the highest expression levels in the C41(DE3) cells (Figure 5.8). A similar protocol to the Factor-Xa purification and cleavage was followed using purchased and in-lab purified PP.

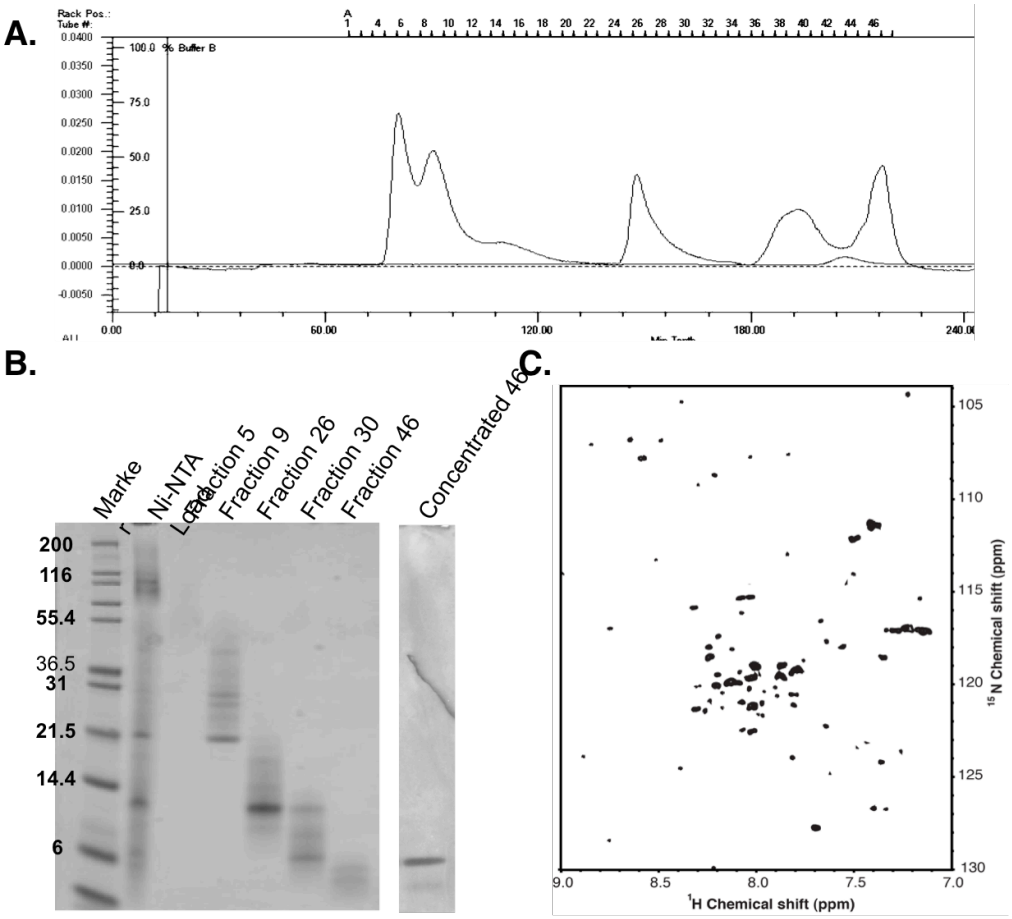

Figure 5.10 FPLC SEC chromatogram with accompanying SDS-PAGE and ^{1}H-^{15}N-HSQC

A. Size-exclusion chromatography was carried out as the final purification step for the JFH-1 construct, which eluted at fraction 46 (B). C. After dialyzing, powder sample was reconstituted in DHPC micelles, pH 4.0 and run for 128 scans on the 500 MHz.

From the elution profile, visualized using SDS-PAGE analysis, we can see that we have effective and efficient cleavage as we only left the column cleavage proceed for four hours. A final step in purification of the cleaved JFH-1 construct involved SEC fast purification liquid chromatography (FPLC). The efficiency of this method is highlighted in the highly resolved SEC chromatogram seen in Figure 5.10A. This level of separation is verified in the SDS-PAGE analysis and a concentrated fraction is then reconstituted in 300 mM DHPC with a pH adjusted to 4.0 and run in an HSQC experiment. Unfortunately, the apparent purity of the sample was fictitious as we see from the correlation spectrum (Figure 5.10C) that there are a vast number of impurities in this sampler that are interfering with the labeled and pure JFH-1 p7 sample.

5.5 Conclusion

While the sample preparation has yet to be established, we can see from the clean band on the protein gel following FPLC that we have pure sample. The sample preparation needs modification and several rounds of optimization. It could have taken several more years to establish a high-throughput purification protocol and sample preparation for this protein. While it would be nice if we could just apply everything we already know about the J4 subtype purification and sample preparation, the JFH-1 subtype behaves dramatically different from the J4 and so this application is simply not possible.

One of the largest hurdles to overcome was the issue of JFH-1 p7 “sticking” to its fusion partner, TrpΔLE. This is highly evidenced in the HPLC run and the subsequent efforts for alternative cleavage reactions, which still enabled the proteins to co-elute in

every experiment. The expression level attained in the GST-tagged construct was highly desirable. Additionally, if the method had been further optimized a highly pure sample could have been attained and analyzed via the HSQC NMR experiment. Finally, one additional optimization method could be to add the methionine cleavage sites back in to efficiently cleave both the GST and 6X His tag and separate the pure JFH-1 protein using SEC. It would be very interesting to be able to compare not only the behavior of this protein in the proteoliposome environment used to carry out solid-state NMR experiments but also to compare the potential structural variations that exist. It is extremely likely the structures could be significantly different just based off the protein' behavioral distinctions as the JFH-1 construct requires a completely different expression and purification protocol.

Chapter 6. Conclusion

p7 is a remarkably complex protein considering that it has only 63 residues and its secondary structure is dominated by two hydrophobic transmembrane helices. Other NMR studies of p7, including a particular study in DPC micelles in solution, suggest the protein exists as a hexamer, and the resulting channel is more likely to be a dominant mechanism of biological activity. In contrast, the structural and dynamic features identified with the solution NMR experiments on monomers of p7 in DHPC micelles in solution support the idea that its biological roles may be more numerous and multifaceted than simply acting as a channel. Indeed, its channel activity may be secondary to its principal biological activities, a notion reinforced by the position of its gene between those for the non-structural and structural proteins of HCV. At least part of its functions are likely to involve interactions with other proteins, such as those shown in the example of p7 interacting with NS2 in Figure 2.8. The protein-protein interactions, as well as possible channel functions, are likely to be affected by its structure and changes in response to the lifecycle of the virus and the addition of drugs

From the determined three-dimensional backbone structure of p7 of HCV in phospholipid bilayers we observe a clear difference between the restricted DHPC micelle structure (Figure 2.3 and 3.4) further confirming the notion of the distortions conveyed on the proteins from the micelle environment. The solid-state NMR p7 structure adopts a tightly packed structure in liquid crystalline phospholipid bilayers. A variety of structural, biophysical, and functional evidence indicate p7 monomers can oligomerize in various preparation to yield channels or pores. Reconciling the monomer structures found under various conditions with the channel activities is a challenge for research in this area, as it

is for all proteins that might be categorized as viroporins.

It is extremely important to consider the lipid environment when characterizing the structure of membrane proteins. This is highly evidenced in the case for p7 structure determination not only in comparison of our own structures but also in comparison to structures determined outside of our studies. The DHPC micelles favor a spatially limited monomeric structure whereas the DPC detergent micelles allow for a complex oligomeric structure. The liposome environment, closest to that of biologically relevant membranes in width, curvature and charge, allows for an undetermined oligomeric structure as evidenced by the difficulty in obtaining high-sensitivity, high-resolution data.

As for observation of oligomerization states in other proteins, the definitive characteristics were analyzed in structural studies. This is seen in the case of M2 were two non-equivalent chemical shift peaks were observed for the isoleucine residue eluding to the fact that the M2 oligomer existed as a dimer of dimers. (REF) This, however, is a convenience that is unfortunately not the case for the p7 protein. It is entirely possible that there is the existence of multiple oligomeric states for p7 within the cell that serve as different functions throughout the virus lifecycle. This is well demonstrated in the MD simulation analysis carried out by Chandler et al. wherein they identified that the flexibility of the protein as well as it's minimalist approach makes it an ideal candidate for the coexistence of multiple oligomeric states. This has been observed in other oligomeric ion channels including Vpu of HIV (Lu *et al.* 2010 & Montal 2003) and the antimicrobial peptide alamethicin (Tieleman *et al.* 2002). This implication stems from the idea that because the p7 monomers are not covalently bound, but instead interact using van der Waals and hydrogen-bonding, it becomes clear that it is easily possible to have

the existence of multiple oligomeric arrangement for which the inter-subunit contact is sufficient enough to support a stable structure. This idea directly supports our finding that with and without the addition of cross-linkers, we observe a multitude of oligomerization species. Moreover, the existence of these multiple oligomeric states could potentially be related to the multiple functions associated with the p7 protein.

Finally, while the sample preparation has yet to be established, we can see from the clean band on the protein gel following FPLC that we have a pure sample. The sample preparation needs modification and several rounds of optimization. It could have taken several more years to establish a high-throughput purification protocol and sample preparation for this protein. While it would be nice if we could just apply everything we already know about the J4 subtype purification and sample preparation, the JFH-1 subtype behaves dramatically different from the J4 and so this application is simply not possible.

Bibliography

[1] Overington, J. P., Al-Lazikani, B., and Hopkins, A. L. (2006) How many drug targets are there?, *Nature reviews. Drug discovery 5*, 993-996.

[2] Yildirim, M. A., Goh, K. I., Cusick, M. E., Barabasi, A. L., and Vidal, M. (2007) Drug-target network, *Nature biotechnology 25*, 1119-1126.

[3] Opella, S. J. (2013) Structure determination of membrane proteins in their native phospholipid bilayer environment by rotationally aligned solid-state NMR spectroscopy, *Accounts of chemical research 46*, 2145-2153.

[4] Opella, S. J., and Marassi, F. M. (2004) Structure determination of membrane proteins by NMR spectroscopy, *Chemical reviews 104*, 3587-3606.

[5] Radoicic, J., Lu, G. J., and Opella, S. J. (2014) NMR structures of membrane proteins in phospholipid bilayers, *Q Rev Biophys 47*, 249-283.

[6] Das, B. B., Nothnagel, H. J., Lu, G. J., Son, W. S., Tian, Y., Marassi, F. M., and Opella, S. J. (2012) Structure determination of a membrane protein in proteoliposomes, *Journal of the American Chemical Society 134*, 2047-2056.

[7] Das, B. B., Lin, E. C., and Opella, S. J. (2013) Experiments optimized for magic angle spinning and oriented sample solid-state NMR of proteins, *J Phys Chem B 117*, 12422-12431.

[8] Marassi, F. M., Das, B. B., Lu, G. J., Nothnagel, H. J., Park, S. H., Son, W. S., Tian, Y., and Opella, S. J. (2011) Structure determination of membrane proteins in five easy pieces, *Methods 55*, 363-369.

[9] Park, S. H., Das, B. B., Casagrande, F., Tian, Y., Nothnagel, H. J., Chu, M., Kiefer, H., Maier, K., De Angelis, A. A., Marassi, F. M., and Opella, S. J. (2012) Structure of the chemokine receptor CXCR1 in phospholipid bilayers, *Nature 491*, 779-783.

[10] Park, S. H., Mrse, A. A., Nevzorov, A. A., De Angelis, A. A., and Opella, S. J. (2006) Rotational diffusion of membrane proteins in aligned phospholipid

bilayers by solid-state NMR spectroscopy, *Journal of magnetic resonance 178*, 162-165.

[11] Das, B. B., Park, S. H., and Opella, S. J. (2015) Membrane protein structure from rotational diffusion, *Biochim Biophys Acta 1848*, 229-245.

[12] McDermott, A. (2009) Structure and dynamics of membrane proteins by magic angle spinning solid-state NMR, *Annual review of biophysics 38*, 385-403.

[13] Das, B. B., Park, S. H., and Opella, S. J. (2014) Membrane protein structure from rotational diffusion., *BBA-Biomembranes*, http://dx.doi.org/10.1016/j.bbamem.2014.1004.1002.

[14] Gower, E., Estes, C., Hindman, S., Razavi-Shearer, K., and Razavi, H. (2014) Global epidemiology and genotype distribution of the hepatitis C virus, *J Hepatol doi: 10.1016/j.jhep.2014.07.027*.

[15] Major, M. E., and Feinstone, S. M. (1997) The molecular virology of hepatitis C, *Hepatology 25*, 1527-1538.

[16] Griffin, S. D., Harvey, R., Clarke, D. S., Barclay, W. S., Harris, M., and Rowlands, D. J. (2004) A conserved basic loop in hepatitis C virus p7 protein is required for amantadine-sensitive ion channel activity in mammalian cells but is dispensable for localization to mitochondria, *The Journal of general virology 85*, 451-461.

[17] Griffin, S. (2003) The p7 protein of hepatitis C virus fomrs an ion channeel that is blocked by the antiviral drug amantadine, *FEBS Lett. 535*, 34-38.

[18] Pavlovic, D., Neville, D. C., Argaud, O., Blumberg, B., Dwek, R. A., Fischer, W. B., and Zitzmann, N. (2003) The hepatitis C virus p7 protein forms an ion channel that is inhibited by long-alkyl-chain iminosugar derivatives, *Proceedings of the National Academy of Sciences of the United States of America 100*, 6104-6108.

[19] Clarke, D., Griffin, S., Beales, L., Geais, C. S., Burgess, S., Harris, M., and Rowlands, D. (2006) Evidence for the formation of a heptameric ion channel complex by the hepatitis C virus p7 protein in vitro, *J. Biol. Chem. 281*, 37057-37068.

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

[21] Yi, M., Ma, Y., Yates, J., and Lemon, S. M. (2007) Compensatory mutations in E1, p7, NS2, and NS3 enhance yields of cell culture-infectious intergenotypic chimeric hepatitis C virus, *J Virol 81*, 629-638.

[22] Pietschmann, T., Kaul, A., Koutsoudakis, G., Shavinskaya, A., Kallis, S., Steinmann, E., Abid, K., Negro, F., Dreux, M., Cosset, F. L., and Bartenschlager, R. (2006) Construction and characterization of infectious intragenotypic and intergenotypic hepatitis C virus chimeras, *Proceedings of the National Academy of Sciences of the United States of America 103*, 7408-7413.

[23] Jones, C. T., Murray, C. L., Eastman, D. K., Tassello, J., and Rice, C. M. (2007) Hepatitis C virus p7 and NS2 proteins are essential for production of infectious virus, *J Virol 81*, 8374-8383.

[24] Gentzsch, J., Brohm, C., Steinmann, E., Friesland, M., Menzel, N., Vieyres, G., Perin, P. M., Frentzen, A., Kaderali, L., and Pietschmann, T. (2013) hepatitis c Virus p7 is critical for capsid assembly and envelopment, *PLoS pathogens 9*, e1003355.

[25] Wozniak, A. L., Griffin, S., Rowlands, D., Harris, M., Yi, M., Lemon, S. M., and Weinman, S. A. (2010) Intracellular proton conductance of the hepatitis C virus p7 protein and its contribution to infectious virus production, *PLoS pathogens 6*, e1001087.

[26] OuYang, B., Xie, S., Berardi, M. J., Zhao, X., Dev, J., Yu, W., Sun, B., and Chou, J. J. (2013) Unusual architecture of the p7 channel from hepatitis C virus, *Nature 498*, 521-525.

[27] Griffin, J. L., Sang, E., Evens, T., Davies, K., and Clarke, K. (2002) Metabolic profiles of dystrophin and utrophin expression in mouse models of Duchenne muscular dystrophy, *FEBS letters 530*, 109-116.

[28] Steinmann, E., Whitfield, T., Kallis, S., Dwek, R. A., Zitzmann, N., Pietschmann, T., and Bartenschlager, R. (2007) Antiviral effects of amantadine and iminosugar derivatives against hepatitis C virus, *Hepatology 46*, 330-338.

[29] Pavlovic, D., Fischer, W., Hussey, M., Durantel, D., Durantel, S., Branza-Nichita, N., Woodhouse, S., Dwek, R. A., and Zitzmann, N. (2005) Long alkylchain iminosugars block the HCV p7 ion channel, *Advances in experimental medicine and biology 564*, 3-4.

[30] Griffin, S., Stgelais, C., Owsianka, A. M., Patel, A. H., Rowlands, D., and Harris, M. (2008) Genotype-dependent sensitivity of hepatitis C virus to inhibitors of the p7 ion channel, *Hepatology 48*, 1779-1790.

[31] Montserret, R., Saint, N., Vanbelle, C., Salvay, A. G., Simorre, J.-P., Ebel, C., Sapay, N., Renisio, J.-G., Bockmann, A., Steinmann, E., Pietschmann, T., Dubuisson, J., Chipot, C., and Penin, F. (2010) NMR structure and ion channel activity of the p7 protein from hepatitis C virus, *J. Biol. Chem. 285*, 31446-31461.

[32] Foster, T., Thompson, G., Kalverda, A., Kankanala, J., Bentham, M., Wtherill, L., Thompson, J., Barker, A., Clarke, D., Noerenberg, M., Pearson, A., Rowlands, D., Homans, S., Harris, M., Foster, R., and Griffin, S. (2014) Structure-guided design affirms inhibitors of hepatitis C virus p7 as a viable class of antivirals targeting virion release, *Hepatology 59*, 408-422.

[33] Sakai, A., Claire, M. S., Faulk, K., Govindarajan, S., Emerson, S. U., Purcell, R. H., and Bukh, J. (2003) The p7 polypeptide of hepatitis C virus is critical for infectivity and contains functionally important genotype-specific sequences, *Proceedings of the National Academy of Sciences of the United States of America 100*, 11646-11651.

[34] Luik, P., Chew, C., Aittoniemi, J., Chang, J., Wentworth, P., Dwek, R. A., Biggin, P. C., Venien-Bryan, C., and Zitzmann, N. (2009) The 3-dimensional structure of a hepatitis C virus p7 ion channel by electron microscopy, *Proceedings of the National Academy of Sciences of the United States of America 106*, 12712-12716.

[35] Robertson, B., Myers, G., Howard, C., Brettin, T., Bukh, J., Gaschen, B., Gojobori, T., Maertens, G., Mizokami, M., Nainan, O., Netesov, S., Nishioka, K., Shin i, T., Simmonds, P., Smith, D., Stuyver, L., and Weiner, A. (1998) Classification, nomenclature, and database development for hepatitis C virus (HCV) and related viruses: proposals for standardization. International Committee on Virus Taxonomy, *Archives of virology 143*, 2493-2503.

[36] Chemello, L., Cavalletto, L., Noventa, F., Bonetti, P., Casarin, C., Bernardinello, E., Pontisso, P., Donada, C., Casarin, P., Belussi, F., and et al. (1995) Predictors of

sustained response, relapse and no response in patients with chronic hepatitis C treated with interferon-alpha, *Journal of viral hepatitis 2*, 91-96.

[37] Tomei, L., Altamura, S., Paonessa, G., De Francesco, R., and Migliaccio, G. (2005) HCV antiviral resistance: the impact of in vitro studies on the development of antiviral agents targeting the viral NS5B polymerase, *Antiviral chemistry & chemotherapy 16*, 225-245.

[38] Major, M. E., and Feinstone, S. M. (1997) The molecular virology of hepatitis C, *Hepatology 25*, 1527-1538.

[39] Luik, P., Chew, C., Aittoniemi, J., Chang, J., Wentworth, P., Jr., Dwek, R. A., Biggin, P. C., Venien-Bryan, C., and Zitzmann, N. (2009) The 3-dimensional structure of a hepatitis C virus p7 ion channel by electron microscopy, *Proceedings of the National Academy of Sciences of the United States of America 106*, 12712-12716.

[40] OuYang, B., Xie, S., Berardi, M. J., Zhao, X., Dev, J., Yu, W., Sun, B., and Chou, J. J. (2013) Unusual architecture of the p7 channel from hepatitis C virus, *Nature 498*, 521-525.

[41] Clarke, D., Griffin, S., Beales, L., Gelais, C. S., Burgess, S., Harris, M., and Rowlands, D. (2006) Evidence for the formation of a heptameric ion channel complex by the hepatitis C virus p7 protein in vitro, *The Journal of biological chemistry 281*, 37057-37068.

[42] Griffin, S. D., Beales, L. P., Clarke, D. S., Worsfold, O., Evans, S. D., Jaeger, J., Harris, M. P., and Rowlands, D. J. (2003) The p7 protein of hepatitis C virus forms an ion channel that is blocked by the antiviral drug, Amantadine, *FEBS Lett 535*, 34-38.

[43] Steinmann, E., Penin, F., Kallis, S., Patel, A. H., Bartenschlager, R., and Pietschmann, T. (2007) Hepatitis C virus p7 protein is crucial for assembly and release of infectious virions, *PLoS pathogens 3*, e103.

[44] Carrere-Kremer, S., Montpellier-Pala, C., Cocquerel, L., Wychowski, C., Penin, F., and Dubuisson, J. (2002) Subcellular localization and topology of the p7 polypeptide of hepatitis C virus, *Journal of virology 76*, 3720-3730.

[45] Patargias, G., Zitzmann, N., Dwek, R., and Fischer, W. B. (2006) Protein-protein interactions: modeling the hepatitis C virus ion channel p7, *J Med Chem 49*, 648-655.

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

[47] Cook, G. A., and Opella, S. J. (2010) NMR studies of p7 protein from hepatitis C virus, *Eur Biophys J 39*, 1097-1104.

[48] Popescu, C. I., Callens, N., Trinel, D., Roingeard, P., Moradpour, D., Descamps, V., Duverlie, G., Penin, F., Heliot, L., Rouille, Y., and Dubuisson, J. (2011) NS2 protein of hepatitis C virus interacts with structural and non-structural proteins towards virus assembly, *PLoS pathogens 7*, e1001278.

[49] Chandler, D. E., Penin, F., Schulten, K., and Chipot, C. (2012) The p7 protein of hepatitis C virus forms structurally plastic, minimalist ion channels, *PLoS computational biology 8*, e1002702.

[50] De Angelis, A. A., and Opella, S. J. (2007) Bicelle samples for solid-state NMR of membrane proteins, *Nature protocols 2*, 2332-2338.

[51] Tjandra, N., and Bax, A. (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium, *Science 278*, 1111-1114.

[52] Vieyres, G., Brohm, C., Friesland, M., Gentzsch, J., Wolk, B., Roingeard, P., Steinmann, E., and Pietschmann, T. (2013) Subcellular localization and function of an epitope-tagged p7 viroporin in hepatitis C virus-producing cells, *Journal of virology 87*, 1664-1678.

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

[54] Mori, S., Abeygunawardana, C., Johnson, M. O., and van Zijl, P. C. (1995) Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation, *Journal of magnetic resonance. Series B 108*, 94-98.

[55] Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *Journal of biomolecular NMR 6*, 277-293.

[56] Shen, Y., Lange, O., Delaglio, F., Rossi, P., Aramini, J. M., Liu, G., Eletsky, A., Wu, Y., Singarapu, K. K., Lemak, A., Ignatchenko, A., Arrowsmith, C. H., Szyperski, T., Montelione, G. T., Baker, D., and Bax, A. (2008) Consistent blind protein structure generation from NMR chemical shift data, *Proceedings of the National Academy of Sciences of the United States of America 105*, 4685-4690.

[57] Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics, *Journal of molecular graphics 14*, 33-38, 27-38.

[58] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD, *Journal of computational chemistry 26*, 1781-1802.

[59] Cook, G., and Opella, S. J. (2010) Secondary structure, dynamics, and architectue of the p7 membrane protein of hepatitis C virus by NMR spectroscopy, *Biochimica Biophysica Acta 1808*, 1448-1453.

[60] Mesleh, M. F., and Opella, S. J. (2003) Dipolar Waves as NMR maps of helices in proteins, *Journal of magnetic resonance 163*, 288-299.

[61] Gouklani, H., Beyer, C., Drummer, H., Gowans, E. J., Netter, H. J., and Haqshenas, G. (2013) Identification of specific regions in hepatitis C virus core, NS2 and NS5A that genetically interact with p7 and co-ordinate infectious virus production, *Journal of viral hepatitis 20*, e66-71.

[62] Moradpour, D., and Penin, F. (2013) Hepatitis C virus proteins: from structure to function, *Curr Top Microbiol Immunol 369*, 113-142.

[63] Reed, K., and Rice, C. (2000) Overview of hepatitis C virus genome structure, polyprotein processing, and protein properties, 55-84.

[64] deLemos, A., and Chung, R. (2014) Hepatitis C treatment: an incipient therapeutic revolution, *Trends Mol Medicine 20*, 315-321.

[65] Reardon, S. (2013) United States to approve potent oral drugs for hepatitis C, *Nature doi: 10.1038/nature.2013.14059*.

[66] Poveda, E., Wyles, D., Mena, A., Pedreira, J., Castro-Iglesias, A., and Cachay, E. (2014) Update on hepatitis C virus resistance to direct-acting antiviral agents, *Antiviral Research 108*, 181-191.

[67] Ploss, A., and Dubuisson, J. (2014) New advances in the molecular biology of hepatitis C virus infection: towards the identification of new treatment targets, *Gut 61(Suppl 1)*, i25-i35.

[68] Carrere-Kremer, S., Pontpellier-Pala, C., Cocquerel, L., Wychowski, C., Penin, F., and Dubuisson, J. (2002) Subcellular localization and topology of the p7 polypeptide of hepatitis C virus, *J Virol 76*, 3720-3730.

[69] Atoom, M., Taylor, N. G., and Russell, R. S. (2014) The elusive function of the hepatitis C virus p7 protein, *Virology doi: 10.1016/j.virol.2014.04.018*.

[70] Steinmann, E., and Pietschmann, T. (2010) Hepatitis C virus p7 - A viroporin crucial for virus assembly and an emerging target for antiviral therapy, *Viruses 2*, 2078-2095.

[71] Griffin, S. D. C., Beales, L. P., Clarke, D. S., Worsfold, O., Evans, S. D., Jaeger, J., Harris, M. P. G., and Rowlands, D. J. (2003) The p7 protein of hepatitis C virus fomrs an ion channeel that is blocked by the antiviral drug amantadine, *FEBS Lett. 535*, 34-38.

[72] Gonzalez, M. E., and Carrasco, L. (2003) Viroporins, *FEBS Lett. 552*, 28-34.

[73] Fischer, W., and Sansom, M. (2002) Viral ion channels: structure and function, *Biochim Biophys Acta 1561*, 27-45.

[74] DiMaio, D. (2014) Viral miniproteins, *Annu Rev Microbiol 68*, 21-43.

[75] Wozniak, A. L., Griffin, S., Rowlands, D., Harris, M., Yi, M., Lemon, S. M., and Weinman, S. A. (2010) Intracellular proton conductance of the hepatitis C virus p7 protein and its contribution to infelctions production, *PLoS Pathog. f6*, e1001087.

[76] Anfinsen, C. B. (1973) Principles that govern the folding of protein chains, *Science 181*, 223-230.

[77] Cross, T. A., Sharma, M., Myunggi, Y., and Zhou, H. X. (2011) Influence of solubilizing environments on membrane protein structures, *Trends Biochemical Sciences 36*, 117-125.

[78] Acharya, R., Carnevale, V., Fiorin, G., Levine, G. G., Polishchuk, A. L., Balannik, V., Smish, I., Lamb, R. A., Pinto, L. H., Degrado, W. F., and Klein, M. L. (2010) Structure and mechanism of proton transport throught the transmembrane tetrameric M2 proein bundle of the influenza A virus, *Proc Natl Acad Sci 107*, 15075-15080.

[79] Stouffer, A. L., Acharya, R., Salom, D., Levine, A. S., Di Costanzo, L., Soto, C. S., Tereshko, V., Nanda, V., Stayrook, S., and DeGrado, W. F. (2008) Structural basis for the function and inhibition of an influenza virus proton channel, *Nature 451*, 596-599.

[80] Bauer, C. M., Pinto, L. H., Cross, T. A., and Lamb, R. A. (1999) The influenza virus M2 ion channel protein: probing the structure of the transmembrane domain in intact cells by using engineered disulfide cross-linking, *Virology 254*, 196-209.

[81] Cook, G. A., Zhang, H., Park, S. H., Wang, Y., and Opella, S. J. (2011) Comparative NMR studies demonstrate profound differences between two viroporins: p7 of HCV and Vpu of HIV-1, *Biochimica et biophysica acta 1808*, 554-560.

[82] Patargias, G., Zitzmann, N., Dwek, R. A., and Fischer, W. B. (2006) Protein-protein interactions: modeling the hepatitis C virus ion channel p7, *J. Med. Chem. 49*, 648-655.

[83] Wang, Y. T., Schilling, R., Fink, R. H., and Fischer, W. B. (2014) Ion-dynamics in hepatitis C virus p7 helical transmembrane domains-a molecular dynamics simulation study, *Biophys. Chem. 192*, 33-40.

[84] Fischer, W. B., Li, L. H., Mahato, D. R., Wang, Y. T., and Chen, C. P. (2014) Viral channel proteins in intrcellular protein-protein communication: Vpu of HIV-1, E5 of HPV16, and p7 of HCV, *Biochim Biophys Acta 1838*, 1113-1121.

[85] Wang, Y. T., Hsu, H. J., and Fischer, W. B. (2013) Computational modeling of the p7 monomer from HCV and its interaction with small molecule drugs, *Springerplus doi: 10.1186/2193-1801-2-324*.

[86] Schindler, C., and Fischer, W. B. (2012) Sequence alignment of viral channel proteins with cellular ion channels, *J. Comput. Biol. 19*, 1060-1072.

[87] Cook, G. A., Dawson, L. A., Tian, Y., and Opella, S. J. (2013) Three-dimensional structure and interaction studies of hepatitis C virus p7 in 1,2-dihexanoyl-sn-glycero-3-phosphocholine by solution nuclear magnetic resonance, *Biochemistry 52*, 5295-5303.

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

[89] Das, B. B., Park, S. H., and Opella, S. J. (2015) Membrane protein structure from rotational diffusion, *Biochim Biophys Acta 1848*, 229-245.

[90] Park, S., Das, B. B., DeAngelis, A. A., Scrima, M., and Opella, S. J. (2010) Mechanically, magnetically, and 'rotationally aligned' membrane proteins in phospholipid bilayers give equivalent angular constraints for NMR structure determination, *J Phys Chem B 114*, 13995-13003.

[91] Nieva, J., Madan, V., and Carrasco, L. (2012) Viroporins: structure and biological functions, *Nat. Rev. Microbiol. 10*, 563-574.

[92] Chandler, D., Penin, F., Schulten, K., and Chipot, C. (2012) The p7 protein of hepatitis C virus forms structurally plastic, minimalist ion channels, *PLOS Comp Biol 8*, e1002702.

[93] Whitfield, T., Miles, A., Scheinost, J., Offer, J., Wentworth, P., Dwek, R. A., Wallce, B., Biggin, P., and Zitzmann, N. (2011) The influence of diferent lipid environments on the structure and function of the hepatitis C p7 ion channel protein, *Mol Membr Biol 28*, 254-264.

[94] Lewis, B. A., Harbison, G. S., Herzfeld, J., and Griffin, R. G. (1985) NMR structural analysis of a membrane protein: bacteriorhodopsin peptide backbone orientation and motion, *Biochemistry 24*, 4671-4679.

[95] Tang, W., and Nevzorov, A. A. (2011) Repetitive cross-polarization contacts via equilibration-re-equilibration of the proton bath: Sensitivity enhancement for NMR of membrane proteins reconstituted in magnetically aligned bicelles, *Journal of magnetic resonance 212*, 25-248.

[96] Pines, A., Gibby, M., and Waugh, J. (1973) Proton-enhanced NMR of dilute spins in solids, *J Chem Phys 59*, 569-590.

[97] Thaukur, R., Kurur, N., and Madhu, P. (2006) Swept-frequency two-pulse phase modulation for heteronuclear dipolar decoupling in solid-state NMR, *Chem Phys Lett 426*, 459-463.

[98] Baldus, M., Petkova, A., Herzfeld, J., and RG, G. (1998) Cross polarization in the tilted frame: assignmenet and spectral simplification in heteronuclear spin systems, *Mol Phys 95*, 1197-1207.

[99] Lin, E., and Opella, S. J. (2014) Covariance spectroscopy in high-resolution multi-dimensional solid-state NMR, *Journal of magnetic resonance 239*, 57-60.

[100] Hu, B., Amoureux, J., Trebosc, J., Deschamps, M., and Tricot, G. (2008) Solid-state NMR covariance of homonuclear correlation spectra, *J Chem Phys 128*, 134502.

[101] Zhang, F., and Bruschweiler, R. (2004) Indirect covariance NMR spectdroscopy, *J Amer Chem Soc 126*, 13180-13181.

[102] Bruschweiler, R., and Zhang, F. (2004) Covariance nuclear magnetic resonance spectroscopy, *J Chem Phys 120*, 5253-5260.

[103] Takegoshi, K., Nakamura, S., and Terao, T. (2001) 13C-1H dipolar-assisted rotational resonance in magic-angle-spinning NMR, *Chem Phys Lett 344*, 631-637.

[104] Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processisng system based on UNIX pipes, *Journal of biomolecular NMR 6*, 277-293.

[105] Goddard, T., and Kneller, D. SPARKY3, *University of California, San Francisco*.

[106] Kim, D. E., Chivian, D., and Baker, D. (2004) Protein structure prediction and analysis using the Robetta server, *Nucleic Acids Research 32*, W526-W531.

[107] Yarov-Yarovoy, V., Schonbrun, J., and Baker, D. (2006) Multipass membrane protein structure prediction using Rosetta, *Proteins-Structure Function and Bioinformatics 62*, 1010-1025.

[108] Tian, Y., Schwieters, C. D., Opella, S. J., and Marassi, F. M. (2012) AssignFit: a program for simultaneous assignment and structure refinement from solid-state NMR spectra, *J Magn Reson 214*, 42-50.

[109] Bermejo, G. A., Clore, G. M., and Schwieters, C. D. (2012) Smooth statistical torsion angle potential derived from a large conformational database via adaptive kernel density estimation improves the quality of NMR protein structures, *Protein Science 21*, 1824-1836.

[110] Schaefer, J., and Stejskal, E. O. (1976) C-13 Nuclear Magnetic Resonance of Polymers Spinning At the Magic Angle, *Journal of the American Chemical Society 98*, 1031-1032.

[111] Mesleh, M. F., Lee, S., Veglia, G., Thiriot, D. S., Marassi, F. M., and Opella, S. J. (2003) Dipolar waves map the structure and topology of helices in membrane proteins, *Journal of the American Chemical Society 125*, 8928-8935.

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

[113] Sharma, M., Yi, M., Dong, H., Qin, H., Peterson, E., Busath, D., Zhou, H., and Cross, T. A. (2010) Insight into the mechanism of the influenza A proton channel from a structure in a lipid bilayer, *Science 330*, 509-512.

[114] Park, S. H., Mrse, A. A., Nevzorov, A. A., Mesleh, M. F., Oblatt-Montal, M., Montal, M., and Opella, S. J. (2003) Three-dimensional structure of the channel-forming trans-membrane domain of virus protein "u" (Vpu) from HIV-1, *Journal of molecular biology 333*, 409-424.

[115] Wittlich, M., Thiagarajan, P., Koening, B., Hartmann, R., and Willbold, D. (2010) NMR structure of the transmembrane and cytoplasmic domains of human CD4 in micelles, *BBA 1798*, 122-127.

[116] Pervushin, K., Tan, E., Parthasarathy, K., Lin, X., Jiang, F., Yu, D., Vararattanavech, A., Soong, T., Liu, D., and Torres, J. (2009) Structure and inhibition of the SARS coronavirus envelope protein ion channel, *PLOS Pathogens 5*, e1000511.

[117] Gan, S.-W., Tan, E., Lin, X., Yu, D., Wang, J., Tan, G., Vararattanavech, A., Yeo, C., Soon, C., Soong, T., Pervushin, K., and Torres, J. (2012) The small hydrophobic protein of the human respiratory syncytial virus forms pentameric ion channels, *J Biol Chem 287*, 24671-24689.

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

[119] Catoire, L., Warnet, X., and Warschawski, D. (2014) Micelles, bicelles, amphipols, nanodiscs, liposomes, or intact cells: The hitchhiker's guide to the study of membane proteins by NMR, *Membrane Proteins Production for Structural Analysis (I. Mus-Veteau (ed))*, 315-345.

[120] Warschawski, D. E., Arnold, A. A., Beaugrand, M., Gravel, A., Chartrand, E., and Marcotte, I. (2011) Choosing membrane mimetics for NMR structural studies of transmembrane proteins, *Biochim Biophys Acta 1808*, 1957-1974.

[121] Tulumello, D. V., and Deber, C. M. (2012) Efficiency of detergents at maintaining membrane protein structures in their biologically relevant forms, *Biochim Biophys Acta 1818*, 1351-1358.

[122] Page, R. C., Moore, J. D., Nguyen, H. B., Sharma, M., Chase, R., Gao, F. P., Mobley, C. K., Sanders, C. R., Ma, L., Sonnichsen, F. D., Lee, S., Howell, S. C., Opella, S. J., and Cross, T. A. (2006) Comprehensive evaluation of solution nuclear magnetic resonance spectroscopy sample preparation for helical integral membrane proteins, *Journal of structural and functional genomics 7*, 51-64.

[123] Mcdonnell, P. A., and Opella, S. J. (1993) Effect of Detergent Concentration on Multidimensional Solution Nmr-Spectra of Membrane-Proteins in Micelles, *J Magn Reson Ser B 102*, 120-125.

[124] Saint, N., Montserret, R., Chipot, C., and Penin, F. (2009) Structural and functional analysis of the HCV p7 protein, *Methods Mol Biol 510*, 125-143.