uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Lisette Haddad

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.Sc. (Biochemistry)

GRADE / DEGREE

Department of Biochemistry, Microbiology and Immunology

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Towards Structure Determination of the Hepatitis C Virus p7 Protein

TITRE DE LA THÈSE / TITLE OF THESIS

Natalie Goto

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Yves Aubin

Kathryn Wright

Gary W. Slater

Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Towards Structure Determination of the Hepatitis C Virus p7 Protein

Lisette Haddad

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the M.Sc. degree in Biochemistry

Department of Biochemistry, Microbiology and Immunology
Faculty of Medicine
University of Ottawa

Your file Votre référence
ISBN: 978-0-494-50884-8
Our file Notre référence
ISBN: 978-0-494-50884-8

Abstract

The hepatitis C virus (HCV) affects 3% of the world's population and is a major cause of liver cirrhosis and cancer. The p7 protein of HCV has recently gained much attention due, in part, to its essential role in viral infectivity. This small integral membrane protein containing two transmembrane helices is thought to perform this function by forming a heptameric cation-specific ion channel. In order to gain insight into its channel function I have initiated nuclear magnetic resonance (NMR) experiments that can start to reveal the structure of this protein. For this purpose a C-terminally His_6-tagged p7 was expressed in *E. coli* as a C-terminal fusion protein to glutathione-S-transferase (GST). This fusion protein was purified under denaturing conditions using nickel affinity chromatography and cleaved using thrombin to separate p7 from GST. Greater than 90% pure His-tagged p7 was obtained using reversed-phase high-performance liquid chromatography. ^{15}N, ^{13}C-labeled samples were reconstituted in detergent solutions to allow analysis by solution NMR. A limited series of sample optimization experiments were performed to identify conditions that would give rise to folded p7 in a detergent-micelle complex that would be amenable for chemical shift assignment. Of the conditions tested, p7 solubilized in the lysophospholipid LMPG at 40°C was found to give rise to the most favorable spectrum with a chemical shift dispersion indicative of a helical protein. A suite of three-dimensional triple resonance experiments were then performed and assignment of backbone chemical shifts was initiated. Based on these experiments it was possible to make preliminary chemical shift assignments for over half of the observed resonances. Secondary chemical shift calculations performed with these chemical shifts are consistent with expectations for α-

helical structure in the TM segments. These results also show that p7 possesses a third helix outside of the putative TM segments. This structure may be important for the proposed genotype-specific function of this part of p7. Overall, the development of protocols to obtain isotope-labeled p7 samples, along with the initiation of NMR analysis, has opened the door to the ultimate goal of p7 structure determination.

Acknowledgements

First and foremost, I would like to thank Dr Natalie Goto for her constant support and advice. Not only have I been able to expand my scientific knowledge, but it is in large part due to my research experience that I have been to achieve my long-term goal of entering medical education. I also would like to thank Dr Thierry Ducat for all the advice and suggestions he has given me, but also for answering my never-ending questions. And finally in the "Dr" category, thank you to Dr John Pezacki for his help and support in this project.

Of course, the Goto lab experience would not have been complete without the GLF crew. I have to first thank Jenny Cheng (a.k.a. "Master GFL") for all her help. She taught me so much when I first started as a fourth year student and she continued to help me even when I kept repeating myself. Heydude, you rock! As for the rest of the "old" crew, Steve (a.k.a. Dr Zoidberg, Purple cheeks, Big YA), Dennis (a.k.a. Papa smurf, D-Rod, Rainbow), Allie (a.k.a. Mooshy) and Tatjana (a.k.a. Toots), I would not have been able to do any of this without you! Fred, Tabussom, Asma, Benson, you guys have been amazing! Finally, a big thanks to Selena Sagan for her work (and her help) in this project. I'm truly glad to have had the privilege of pursuing my master's degree in the most awesome research lab ever!

Table of contents

CHAPTER 3 – RESULTS

CHAPTER 4 – DISCUSSION

List of Abbreviations

BCA	bicinchoninic acid
BLM	black lipid membrane
CMC	critical micelle concentration
CP	crude protein
DNA	deoxy-ribonucleic acid
DPC	Dodecylphosphocholine
DSS	2,2-dimethyl-2-silapentane-5-sulfonate
ER	endoplasmic reticulum
GST	glutathione-S-transferase
HIV	human immunodeficiency virus
HPLC	high-performance liquid chromatogaphy
HSQC	heteronuclear single quantum correlation
IFN	Interferon
IPTG	isopropyl-β-D-thiogalactopyranoside
IRES	internal ribosomal entry site
kb	kilo base
LB	luria-bertani
LMPG	1-myristoyl-2-hydroxy-sn-glycero-3-[phospho-RAC-(1-glycerol)]
LPPG	1-palmitoyl-2-hydroxy-sn-glycero-3-[phospho-RAC-(1-glycerol)]
MW	molecular weight
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
NS	non-structural
NTA	nitriloacetic acid
OD	optical density
PAGE	polyacrylamide gel electrophoresis
pI	Isoelectric point
PVDF	polyvinylidene difluoride
RNA	ribonucleic acid
RP	reversed-phase
SDS	sodium dodecyl sulfate
SF	soluble fraction
SVR	sustained virologic response
TBS	tris-buffered saline
TEM	transmission electron microscopy
TEV	tobacco-etch virus
TFA	trifluoroacetic acid
TM	transmembrane
Tris	tris(hydroxymethyl)aminomethane
TTBS	tween-20 tris-buffered saline

List of Figures

List of Tables

Chapter 1 INTRODUCTION

1.1 Hepatitis C

Hepatitis C is a worldwide problem caused by the hepatitis C virus (HCV) affecting approximately 170 million people or 3 % of the world's population (2). Although HCV infection is widespread, countries in Africa and Asia have a higher incidence rate while countries in North America and western Europe are affected to a lesser degree (3). The virus is primarily transmitted by blood, with transmission by sexual contact or from mother to child being inefficient (2, 4). The site of HCV infection is primarily the liver, but extrahepatic tissues such as kidney and skin cells can also be infected (5). While the exact mechanism of HCV entry into these sites has yet to be determined, several candidate receptors, for example CD-81, a member of the tetraspanin family, have been identified (6). Most HCV infections result in chronic hepatitis leading to continuous destruction of hepatocytes, causing disorders such as cirrhosis, where the liver becomes extensively scarred leading to functional loss, and/or hepatocellular carcinoma (5). Both will often lead to the death of the infected patients. The virus is also the leading cause of liver transplantation in developed countries (7).

In developed countries, hepatitis C is usually treated with a combination of pegylated-interferon-α and ribavirin (3, 4, 8). Successful treatment results in a sustained virologic response (SVR) which is defined as the absence of HCV RNA in serum for at least six months after the end of treatment (8). While there are several types of interferon (IFN) that exhibit some degree of activity against HCV, IFN types α2a, α2b and

consensus IFN are the ones commercially available for treatment as they are effective at yielding high SVR rates in treated patients (8). IFN works by activating hundreds of interferon-stimulated genes which stimulate the production of proteins that can act against viral infection, such as nucleases specific for double-stranded RNA and inhibitors of viral protein translation (8, 9). Ribavirin is a guanosine analogue which possesses activity against many RNA and DNA viruses (10). The exact mechanism of action is unknown, but several possibilities have been suggested, including direct inhibition of HCV replication by misincorporation of ribavirin into viral RNA (8). Even when successful, the combination IFN-ribavirin therapy is often associated with difficult side-effects such as IFN-induced depression, possibly caused by a decrease in serotonin levels (11), and ribavirin-induced anemia, perhaps due to the drug's toxic accumulation in erythrocytes (12). An additional disadvantage to this therapy is that it only provides approximately 50 % of patients with a SVR, for reasons that have not yet been elucidated (8). For these reasons, it is of great importance to find more effective drugs that will improve the health outcome of hepatitis C-infected patients.

HCV is similar to the human immunodeficiency virus (HIV) in that the development of an effective vaccine is not practical because of the virus' ability to rapidly mutate which continuously generates virion mutants that escape immunity (2, 13). Error-prone replication by the viral RNA polymerase has given rise to a high degree of genetic diversity that reflects the evolution of HCV in different populations around the world. This has led to the classification of HCV genomes into 6 different genotypes each which are associated with a specific geographical range, and differ in approximately 31-34 % of their nucleotide sequence (4). Subtypes or strains of these genotypes (identified

by lower-case letters e.g. 1a, 1b, 1c, etc...) have also been classified, differing by approximately 20-23 % of their nucleotide sequence (4) and may arise in response to population differences in immune responses to infection (2). An additional level of genetic diversity also tends to accumulate within an infected patient over time, giving rise to strains that are known as quasispecies.

1.2 The HCV Proteome

HCV is a small, enveloped RNA virus that belongs to the family *Flaviviridae*, genus *Hepacivirus* (6). Its genome is an approximately 9.6 kb single-stranded positive sense RNA molecule which encodes a polyprotein precursor of about 3 000 amino acids. This polyprotein is co- and post-translationally cleaved by host and viral proteases into 10 different proteins (Figure 1.1A) (4, 6). The first three proteins at the N-terminal side of the polyprotein are released by cellular signal peptidases within the ER (14). These proteins are the core (C) and envelope glycoproteins E1 and E2 which make up the structure of the virus (Figure 1.1B). The core protein is involved in nucleocapsid formation while E1 and E2 are embedded in the host-derived lipid bilayer that forms the viral envelope. The envelope glycoproteins are thought to be essential for entry of the virus by binding to receptors and inducing fusion of the viral envelope with the host cell membrane (4) although the exact mechanism of viral entry is still unknown.

The remainder of the polyprotein contains non-structural (NS) proteins that are involved in several different processes important for viral replication. Specifically, the NS2 and NS3 proteins contain protease activity and are responsible for processing the remainder of the polyprotein (14). NS3 also contains a helicase domain that has multiple functions including RNA unwinding. Although the exact role of the helicase in the viral

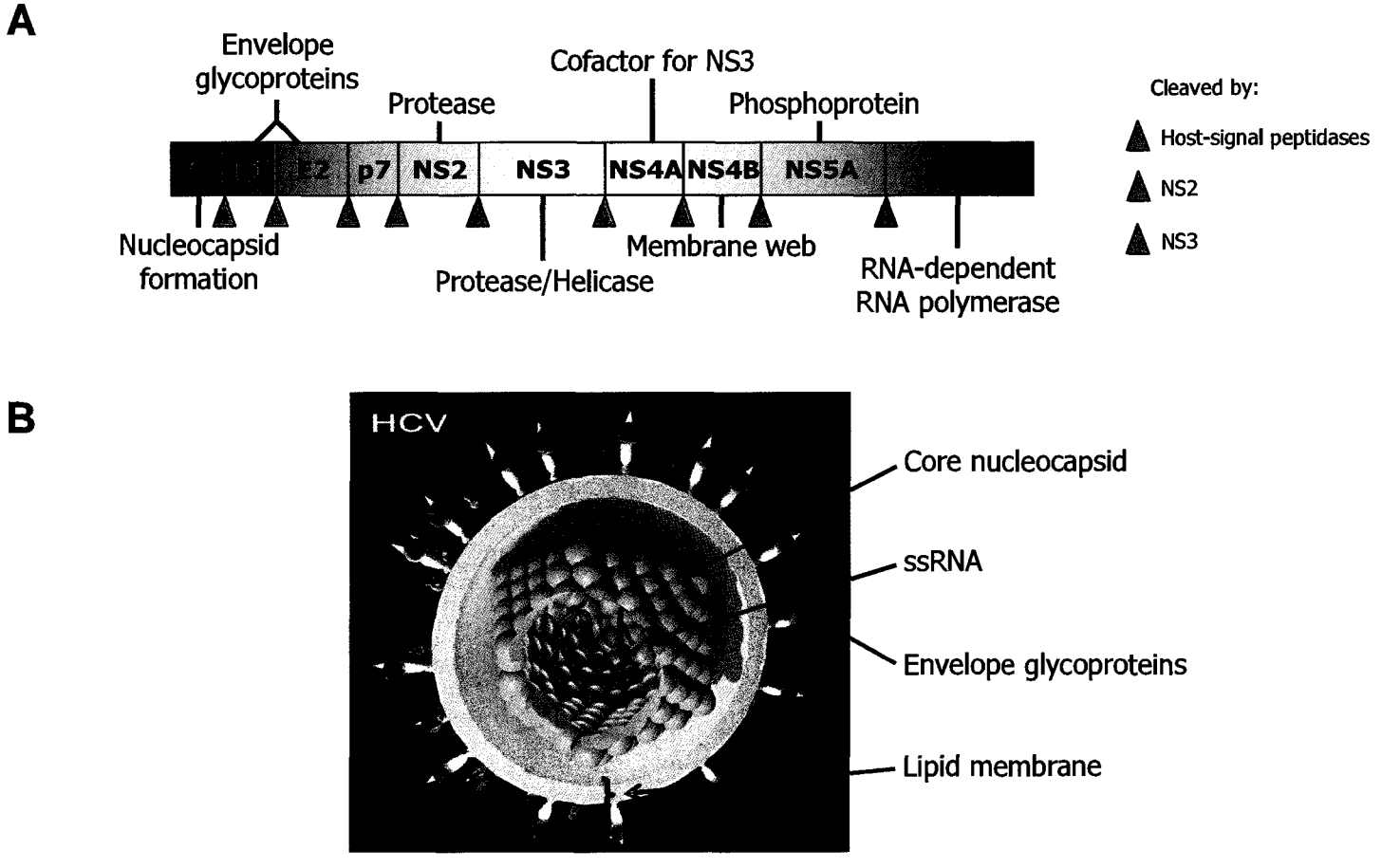


Figure 1.1 (A) Polyprotein encoded by the HCV genome. The first three proteins (C, E1 and E2) are known as the structural proteins as they make up the structure of the virus. The "NS" proteins are non-structural and are involved in various processes including proteolytic cleavage and RNA replication. Colored triangles indicate which proteins are responsible for cleavage at particular sites of the polyprotein. **(B) HCV structure.** Diagram taken and adapted from http://www.med-ars.it/galleries/virus_1.htm.

life cycle is unknown, it is thought to be used for initiation and/or completion of HCV replication and also to aid in maintaining the HCV RNA in a structureless form to facilitate protein synthesis (13). The functions of proteins NS4B and NS5A are not yet known, although NS4B may be involved in the formation of an ER-derived membranous web that contains all HCV proteins as well as replicating RNA (15, 16). In contrast, NS5B has been well characterized at the biochemical (17, 18) and structural level (19, 20) as an RNA polymerase that is responsible for the replication of viral RNA.

One protein that has not yet been defined as a structural or non-structural protein is located in the HCV polyprotein at the junction of the structural and non-structural proteins. As described in the following sections, this protein is known as p7, and is the subject of particular interest due to its potential to be a target for new drug development.

1.3 The HCV p7 Protein

p7 is a 63 amino acid-long integral membrane protein with an amino acid sequence that is ~ 68 - 83% conserved between different HCV genotypes (Figure 1.2A). This level of sequence conservation is higher than that observed over the whole HCV genome, highlighting the functional importance of this sequence. In addition, its overall hydropathic character is conserved across the genotypes (Figure 1.2B), with hydropathy profiles predicting two transmembrane helices separated by a short loop (Figure 1.3) (1). Experimental confirmation of this topology was obtained through antibody accessibility studies with various p7 fusions to a Myc epitope (1) which also showed that the loop localizes to the cytoplasmic side of the membrane. In addition, this study showed that

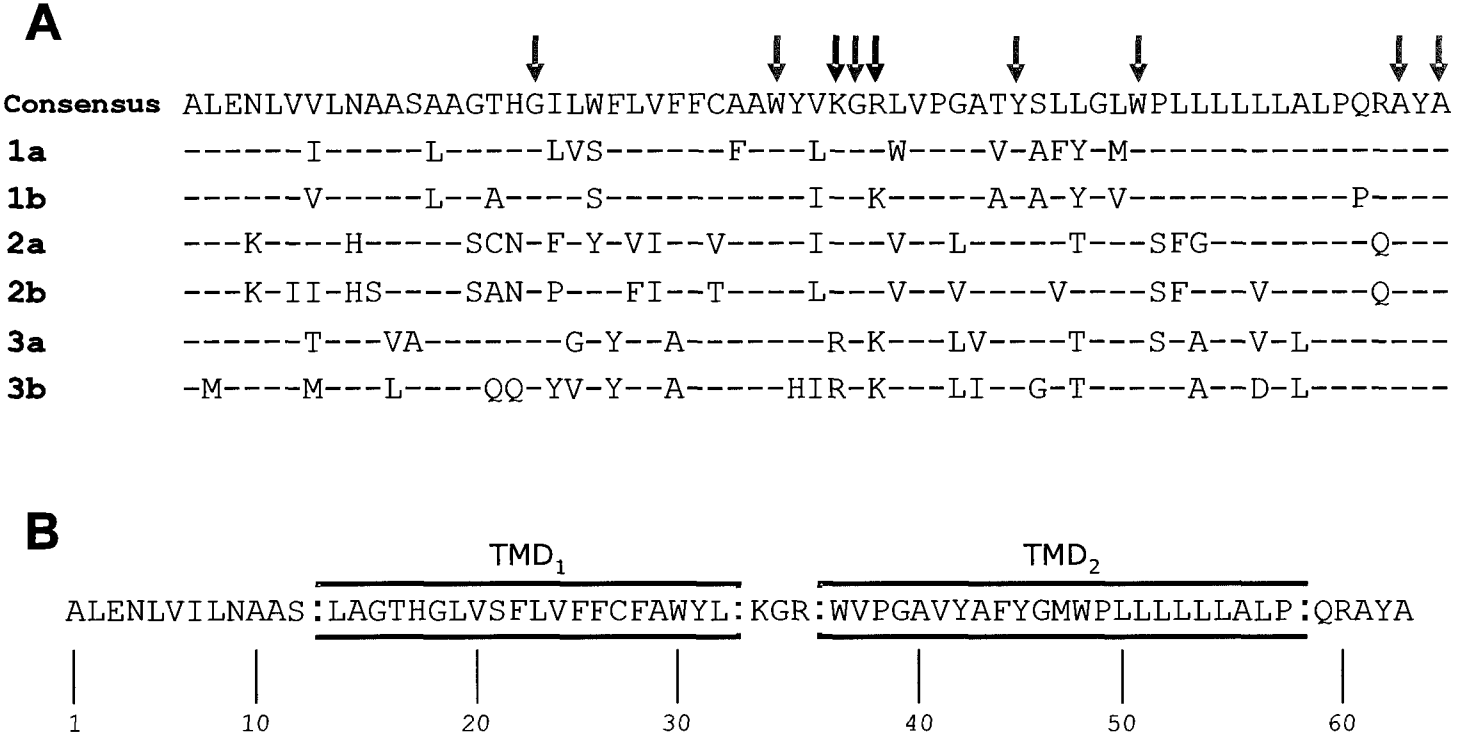


Figure 1.2 (A) Alignment of p7 sequences representing the principal subtypes of genotypes 1 to 3 (figure adapted from Carrère-Kremer *et al.* (1)). The consensus p7 amino acid sequence over these genotypes is represented at the top. Pink arrows show the residues that are strictly conserved throughout all HCV genotypes. Green arrows point to the positions showing conservation of positively-charged residues in the cytoplasmic loop. **(B) Consensus amino acid sequence of the genotype 1a HCV p7 protein.** Prediction methods such as TMHMM and TopPred2 (21, 22), suggest that p7 to contains 2 α-helical transmembrane segments, the first spanning residues 12 or 13 to 32 and the second spanning residues ~37 to 58.

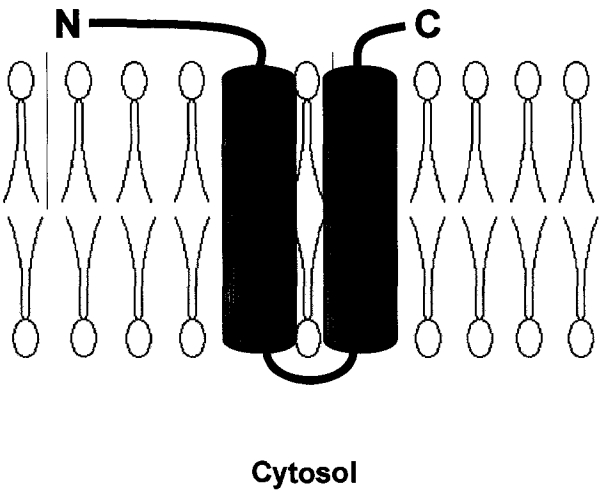


Figure 1.3 HCV p7 topology as determined by Carrère-Kremer *et al.* (1).

p7 localizes primarily to the ER, a result that has since been independently confirmed in more physiological *in vitro* replication systems (23).

The relatively high level of conservation of the p7 sequence and hydropathy characteristics suggests that this protein is a functionally important component of HCV. However, its existence as a distinct protein product of the HCV polyprotein was only recognized just over a decade ago when two distinct forms of the E2 glycoprotein were observed; one containing E2 sequence fused to p7 and one with E2 alone (24). Subsequent expression of a series of C-terminally truncated polyproteins allowed p7 to be identified as a separate protein product (24). However, until recently, knowledge about the role of p7 in the HCV life cycle has been hindered by the absence of cell culture systems that can allow replication of the full-length HCV genome. Low amounts of viral RNA, high mutation rates and high instability of the cloned genome in plasmid vectors have greatly hindered the development of a fully replicating HCV cell culture system (25). Consequently, one of the only methods to study HCV and replicate the virus was to use the chimpanzee as an *in vivo* model of HCV infection. However, a more convenient alternative was generated by the creation of an HCV replicon in which the core, E1, E2, p7 and NS2 proteins were replaced by a marker gene as well as a heterologous internal ribosomal entry site (IRES) (25). Transfection of the Huh-7 human hepatoma cell line with this replicon yielded high amounts of self-replicating HCV RNAs allowing the study of the function of the remaining NS proteins and the effect of their replication on the host cell (18). However, this system did not allow production of infectious viral particles, making it difficult to understand how all proteins contribute to HCV pathogenicity *in vivo*. Fortunately, intense interest in the study of full-length HCV has led to the

generation of full-length replication of the HCV genome in 2005 and production of virus particles, although the specific conditions that are required to allow this to occur are still not understood (25-29). Nonetheless, these systems are just now yielding new insights into the functional importance of HCV proteins such as p7.

Recent evidence obtained utilizing a full-length replication system has suggested that p7 acts at an early stage of virion morphogenesis prior to the assembly and release of virus particles (30). This was suggested based on studies evaluating levels of intracellular infectious virus accumulation where an HCV replication system lacking the p7 sequence did not produce detectable intracellular virus levels as compared to the wild-type genome (30). In contrast, other studies, also utilizing a full-length replication system, have suggested that p7 actually acts at the point of assembly and release of virions (31). Mutations to conserved residues within the p7 sequence suppressed release of particles (possibly due to assembly or release problems) while entry into liver cells was unaffected. To distinguish between inhibitory effects on viral assembly versus viral release, levels of intra- and extracellular infectious virions were measured and were found to indicate an essential role for p7 in both processes (31). While new insights into the role of p7 have been obtained through these studies, the molecular mechanism by which the protein acts is still unclear. Furthermore, it is still not known whether this protein is a component of the HCV virion itself.

Earlier work with the HCV replicons established that p7 is not necessary for viral genome replication since replicons which lack the p7 gene replicate efficiently (18). While it has been suggested that E2-p7, a fusion protein resulting from incomplete processing at the E2 $\leftrightarrow$ p7 junction, may have a role in the HCV life cycle, recent studies

with the full-length replication system *in vitro* have shown that E2-p7 is not required for infectivity of HCV (30). Conversely, these studies have shown that p7 is essential for infectivity of HCV as full-length replication systems lacking the p7 sequence failed to produce detectable levels of infectious virus (30, 31). This is in accordance with earlier studies showing the importance of p7 in *in vivo* infectivity of HCV as chimpanzees transfected with RNA transcripts lacking the p7 sequence did not cause infection of the animal (32). Overall these results have established that p7 is crucial for proper viral function, making p7 the focus of intense interest for laboratories seeking to understand, and develop active compounds against HCV.

Additional insight into the functional role of p7 has been obtained from mutational studies targeting highly conserved residues in the p7 sequence. For example, when W30 and Y42 were mutated into phenylalanine, decreased viral infectivity was observed (31). Furthermore, a double mutant in the cytoplasmic loop (K33I/R35S) abolished viral infectivity *in vivo* (32). Similarly, mutation of these residues into alanine caused a large decrease in the *in vitro* levels of virus production (30). Interesting observations have also been made with chimeric p7 proteins composed of sequences from different genotypes of HCV. Specifically, when the N- and C-termini of p7 from the 1a genotype were replaced by those of the 2a genotype, the mutant version of the 1a virus was found to be non-viable in chimpanzees (32). In contrast, when all of the p7 1a sequence except for these terminal sequences was replaced by the 2a sequence, viral infectivity was not affected (32, 33). Only 4 amino acids are different between the 1a and 2a sequences in these terminal regions of p7 (specifically N4K, I7V, N9H and R60Q), indicating that one or more of these amino acids is important for genotype-specific

interactions, potentially with p7 itself. All of these mutational studies indicate an important functional role for p7 in the generation of infective HCV particles and also point to a special role for specific hydrophilic residues in its function.

1.3.1 p7 as a Viroporin

Although the mechanism by which p7 acts to promote viral assembly and release has not yet been identified, several studies have demonstrated a propensity for p7 to form cation-specific channels in lipid membranes, raising the possibility that this protein belongs to a family of proteins known as viroporins. The primary evidence for this functional assignment comes from single channel conductance measurements of purified p7 introduced into black lipid membranes (BLM) (34-37). In addition, p7 is similar to other viroporins since they are usually small (~ 60 - 120 amino acids) and highly hydrophobic viroproteins with at least one transmembrane α-helix (38). Like p7, some viroporins have additional hydrophobic domains that may interact with the membrane as well as a stretch of basic amino acids which may be involved in membrane destabilization (38). This membrane-destabilizing function may explain why the presence of positively charged residues at positions 33 and 35 of p7 appears to be required for normal viral viability. Another similarity between p7 and some viroporins is the presence of a region that is rich in aromatic residues within the transmembrane helix (38). The interaction of these residues with the bilayer has been suggested to be important for membrane destabilization (38).

A number of viroporins from different viruses have been identified and characterized, including the Vpu protein from HIV and the M2 protein from the influenza A virus. While it is generally thought that these different viroporins share similar

functions in the life cycle of viruses, the exact role of viroproteins in the life cycle of a virus has yet to be identified (35). However, one feature that appears to be common between them is that once inserted into the membrane they oligomerize to form a hydrophilic pore with an aggregation number that appears to be protein dependent (39-41). For example, Vpu is thought to form a pentameric ion channel (42) while the influenza A virus M2 channel, possesses a tetrameric structure (as established through cross-linking and sedimentation studies) (39, 43, 44). p7 also appears to form oligomeric structures in membranes, and has been shown by cross-linking studies as well as transmission electron microscopy (TEM) to be heptameric in size (37). These channels may be primarily formed by the first TM helix of p7, as C-terminal deletion constructs ending at the N-terminal side of the second TM helix still gave rise to ion channel activity in BLM (33).

Of particular interest for anti-HCV drug development was the observation that p7 ion channel activity in BLM systems could be abrogated by amantadine, a drug that has been shown to increase the number of patients with a sustained viral response when used to treat HCV-infected patients who had not responded to conventional interferon/ribavirin-based therapies (45-49). Amantadine was previously shown to be effective against the influenza virus and exert an inhibitory effect on M2 channel function. In the case of M2, it has been shown that amantadine binds to His-37, an amino acid that along with Trp-41 is important for proton conductance. According to a model of this channel generated with restraints obtained by solid-state NMR (shown in Figure 1.4) both these residues are in the single M2 TM segment and line the polar interior of the pore. Based on the M2 model it has been proposed that inhibition of the p7

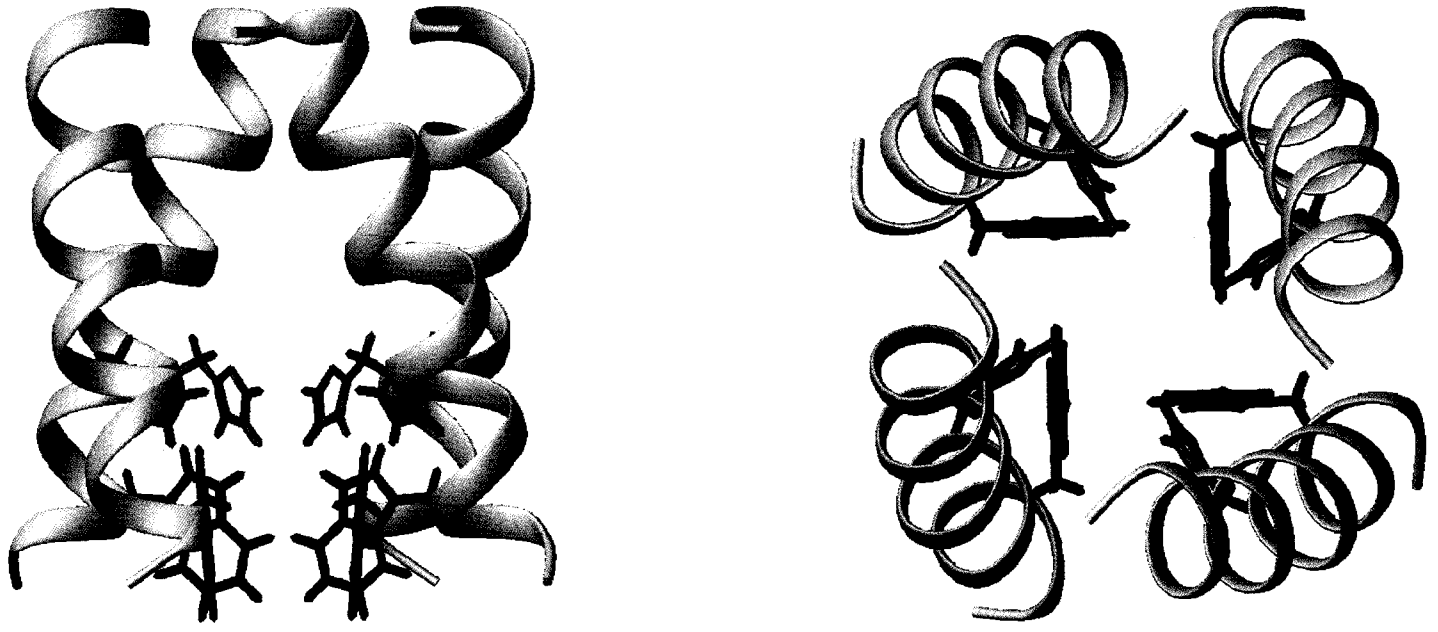


Figure 1.4 Model of the tetrameric form of the influenza A virus M2 proton channel based on constraints from solid-state NMR (50). A ribbon diagram of the 25 residue transmembrane helix from each subunit is shown with functionally important side-chains of His-37 (green) and Trp-41 (red).

channel by amantadine may occur through interactions with His-17 in the first TM segment (51). This residue is not strictly conserved throughout all HCV genotypes, which may explain why amantadine was not effective in all patients (52). Although it has recently been shown, using a full-length HCV replication system, that amantadine has no effect on *in vitro* viral proliferation, other p7-channel blocking compounds (iminosugar derivatives) were able to clear infectious HCV from cell culture. These results support the possibility that compounds that can inhibit p7 channel activity may be good leads in the development of new HCV-specific anti-viral compounds (53). Since high-resolution structural information can help to characterize the binding site and aid in future drug design, there is significant interest in determining a structure of for p7, which is a current goal in the Goto laboratory.

Although x-ray crystallography has provided high-resolution structures for an increasing number of membrane proteins, it can be difficult to crystallize small oligomeric membrane proteins. In contrast, solution NMR spectroscopy has the potential to elucidate structures for these types of membrane protein systems. For this reason, in this thesis a solution NMR approach to determining a structure for the p7 protein was pursued. The methods used for protein sample preparation as well as background behind this technique are, therefore, provided in the following sections.

1.4 Membrane Protein Sample Preparation

The first and often most time-consuming step in structure determination is the production of milligram quantities of highly pure protein samples. When proteins that are larger than ~10 kDa are being studied by solution NMR, it is usually necessary to uniformly incorporate stable isotopes that are able to give rise to a signal in an NMR

spectrum. In the case of p7, it is expected that solubilization of the protein in a detergent micelle will create a large complex, therefore making it necessary to incorporate ^{15}N and ^{13}C in the protein. Although it is possible to produce smaller proteins, including p7, using solid-phase synthetic approaches (35, 36), uniform incorporation of ^{15}N-, ^{13}C-labeled amino acids would be prohibitively expensive. For this reason, study of p7 by solution NMR would be greatly facilitated by the development of a high-yield biosynthetic method that can allow cost-effective production of uniformly isotope-labeled p7. As is usually done in NMR studies of proteins, *Escherichia coli* can be used as an expression host since heterologous protein expression in this bacteria is usually straightforward, and has the potential to produce large quantities of recombinant protein in one or two liters of media (54-56).

When membrane proteins are expressed at high levels in *E. coli*, they usually either insert into the membrane or form insoluble aggregates of misfolded proteins known as inclusion bodies. While expression to the membrane may be favored because of its potential physiological relevance, low protein yields are often observed due to cell toxicity (55). Although refolding may be required after purification, inclusion body-targeting can be advantageous since these aggregates tend to be more resistant to proteolysis during expression. Furthermore, since the overexpressed protein is usually the main component of inclusion bodies, this targeting method also has the potential to simplify protein purification (56).

Higher amounts of protein can often be obtained by expressing the protein of interest as a fusion to the C-terminus of another protein that has a tendency to be successfully overexpressed in *E. coli* (56). Among the most widespread fusion partners

are thioredoxin, and glutathione-S-transferase (GST) (57). Even though both thioredoxin and GST are often used for the production of soluble fusion proteins, they can also be used to promote expression of proteins in inclusion bodies (58, 59). Previously in our laboratory, it was found that GST fused to another small membrane protein gave the highest quantity of fusion protein which localized to inclusion bodies when compared to other fusion partners tested (60). In addition, work done by a previous graduate student in the Goto laboratory (Ms. Jenny Cheng) showed that high levels of expression could also be obtained when GST was fused to the HCV p7 protein, with all of the protein localizing to the inclusion body fraction (Cheng, unpublished results).

Following expression of the fusion protein, it is necessary to purify it from native *E. coli* proteins. This is often done using nickel affinity chromatography, one of the most popular recombinant protein purification methods (61-63). Typically, a hexa-histidine (His) tag is fused to either the N- or C-terminus of the protein. The imidazole side-chains of the adjacent histidine residues can then form coordination bonds with the nickel ions on the column. One particular advantage of this method is that it can be conducted under denaturing conditions. This is imperative when expressing to inclusion bodies, as resolubilization of the expressed proteins requires strong denaturants.

Since a protein fusion partner such as GST is significant in size and may interfere with future structural and functional characterization of the protein of interest, it is usually necessary to cleave it from the fusion protein. Cleavage can be achieved by chemical methods (e.g. cyanogen bromide (62)), however, the protein or peptide must be stable in relatively harsh solvent conditions and cleavage site specificity is usually poor (64). Enzymatic methods can allow a high degree of cleavage site specificity to be

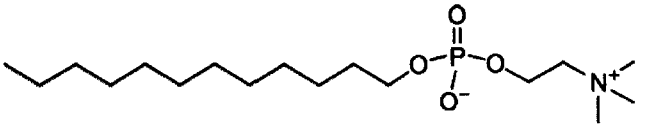
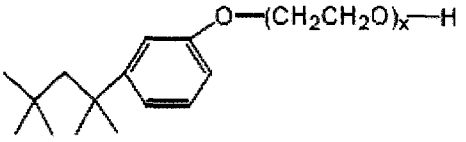
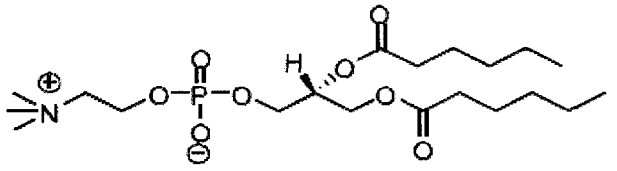
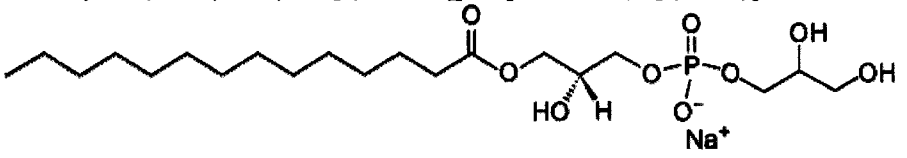
achieved by introducing consensus cleavage sites between the carrier and target proteins. This approach has been widely used with enzymes such as the tobacco-etch virus (TEV) protease (65, 66), enterokinase (67), factor Xa (68, 69) or thrombin (70, 71). Once cleavage has been achieved, the protein of interest can be separated from the cleavage mix using conventional chromatographic techniques. These general approaches can be used to generate milligram quantities of protein samples that are >90% pure and therefore suitable for solution NMR investigations.

As previously discussed (section 1.4.2) sample conditions must be found that produce spectra with intense, relatively well-resolved signals while maintaining the protein in a folded form. In the case of membrane proteins this can be particularly challenging since the native lipid bilayer environment is not compatible with solution-phase techniques. In order to benefit from the high resolution provided by solution-phase experiments, it is therefore necessary to find conditions where the membrane protein is in a soluble form that exposes it to a reasonable mimic of the lipid bilayer environment. Organic solvents are sometimes used for this purpose as phase separation can occur on a microscopic level in a way that sometimes supports a folded conformation in membrane proteins. However, only a small number of membrane proteins appear to be amenable to this type of environment (72). On the other hand, detergents spontaneously self-assemble to form micelles with a hydrophobic core with a polar exterior (73, 74). These structures provide an amphipathic membrane-like environment for the accommodation of a membrane protein that allows the membrane protein to remain in a functionally competent folded form. In fact, a number of membrane proteins have been shown to

maintain their biological activities in detergent micelles (62, 75-77) supporting the membrane-mimetic role of these detergents.

While no single detergent is ideal for structural studies of all membrane proteins a large number of detergents are available for protein NMR. Indeed, the choice of detergent is one of the most critical aspects of membrane protein sample preparation that can greatly affect the quality of the NMR spectrum. An ideal detergent is one that is mild enough to prevent protein denaturation while being an effective solubilizing agent that can prevent aggregation at the high protein concentrations required for NMR. In addition, a detergent must allow the protein to retain a folded conformation, and form detergent-protein complexes that are relatively small (ideally <40 kDa). As shown in Table 1.1, two of the most commonly used detergents for structural studies (sodium dodecyl sulfate (SDS) and dodecylphosphocholine (DPC)) form relatively small micelles (73, 78). The smaller size of these micelles arises due to the relatively small size of the detergent molecule itself, and its tendency to form micelles with smaller aggregation numbers (the number of detergent molecules within a micelle). Lysolipids also form relatively small micelles, and have been suggested to be more suitable for obtaining NMR spectra of membrane proteins (73), with lysophosphatidylglycerol (LPPG) being particularly favorable. However, this is not necessarily applicable for all membrane proteins, since in a subsequent study performed by Columbus *et al.*, LPPG, provided the poorest quality spectra of all the detergents tested (79). Overall, these studies indicate the importance of testing a wide variety of detergents for solution NMR of membrane proteins.

Table 1.1 Physicochemical properties of commonly used detergents for solution NMR

Detergent Name and Structure	MW (Da)	CMC* (mM)	Protein-Free Micelle MW (Da)§
Sodium Dodecyl Sulfate - SDS 	288	8	18 000
Dodecylphosphocholine - DPC 	352	1	19 700
Triton X-100 	647	0.2	90 000
1,2-Dihexanoyl-phosphatidylcholine - DHPC 	454	14	9 000
1-myristoyl-2-hydroxy-sn-glycero-3-[phospho-RAC-(1-glycerol)] - LMPG 	479	0.05	26 300

* CMC, critical micelle concentration; obtained from the literature: SDS (80), DPC (81), TX-100 (75), DHPC (75), LMPG (82)

§ Micelle molecular weights were calculated by multiplying the aggregation number by the detergent monomer molecular weight. Aggregation numbers obtained from the literature: SDS (83), DPC (81), TX-100 (75), DHPC (75), LMPG (84)

1.5 Solution NMR Spectroscopy

1.5.1 Basic Principles of NMR

NMR spectroscopy utilizes the intrinsic property of some nuclei known as angular momentum or spin (I). Of particular interest for biomolecular NMR spectroscopists are nuclei with a spin of ½; specifically ^{1}H, ^{15}N, ^{13}C and ^{31}P. For these nuclei the application of an external magnetic field (from the spectrometer) will induce a microscopic dipole that adopts only two orientations with respect to this external field (B_0): either a low energy state (usually parallel to B_0) or a high energy state (opposite in orientation to the low energy state). If these nuclei are subjected to a radiofrequency pulse with an energy that is equal to the energy difference between these two spin states, its energy will be absorbed leading to transitions between the states. However, since this absorption can only be detected for the small excess of nuclei that are in the lower energy state (as dictated by the Boltzmann distribution) NMR is a relatively insensitive technique. In addition, sensitivity also depends on the size of the molecule being studied, with larger molecules that have slow solution tumbling times showing significant decreases in sensitivity relative to small, rapidly tumbling molecules (73). For this reason, the study of proteins and other macromolecules by solution NMR requires significant quantities of pure, isotopically labeled samples, typically on the order of 0.5 mM or higher.

The data produced by solution NMR spectroscopy typically show peaks occurring at the frequency at which absorption occurs, with a magnitude that is proportional to the number of atoms giving rise to the signal (Figure 1.5). The frequency of absorption is reported as a shift from a standard frequency in units of parts per million (ppm), and is usually referred to as the chemical shift since it contains contributions from the local

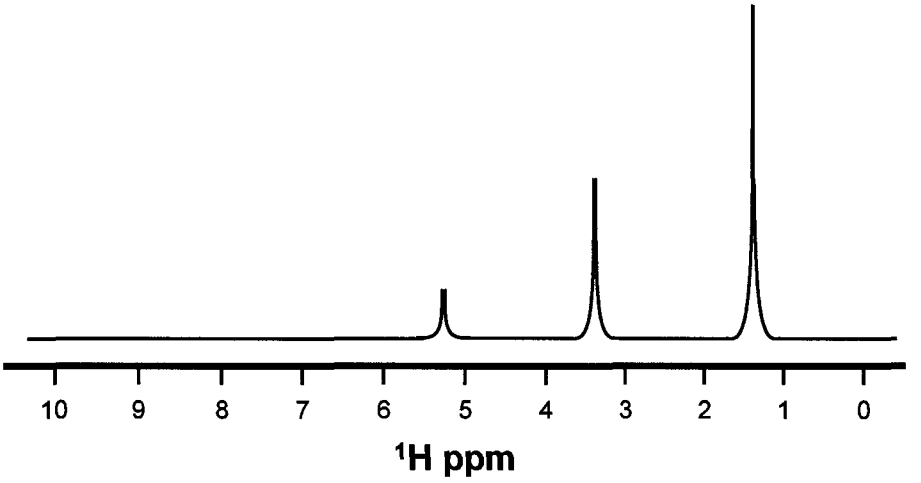


Figure 1.5 Theoretical 1D proton NMR spectrum of a simple molecule.

chemical environment. In the case of molecules containing a small number of NMR-active spins, one-dimensional spectra such as the one shown in Figure 1.5 are simple to analyze, as few peaks would be present. However, proteins are very large and proton-rich, giving rise to very complex ^{1}H spectra with a large number of overlapping peaks. Since protons are the most sensitive NMR-active nuclei present in proteins and are required for detection of NOEs, multidimensional NMR experiments must be performed to provide a simplified view of these peaks and assign them to specific atoms in the molecule being studied. This is greatly facilitated by uniform incorporation of the NMR-active isotopes ^{15}N and ^{13}C in order to increase the amount of through-bond (J-coupling) correlations that can be made within the polypeptide backbone. These interactions can be used to acquire 3D spectra that allow the connectivity of backbone atoms to be straightforwardly determined (described in section 1.5.4).

1.5.2 The $^{1}H - ^{15}N$ HSQC

One of the most commonly used types of multi-dimensional experiment used to provide the first view of a protein spectrum is known as the $^{1}H - ^{15}N$ heteronuclear single quantum coherence (HSQC) experiment (62, 85). The $^{1}H - ^{15}N$ HSQC spectrum is sometimes referred to as a 'fingerprint' spectrum since each polypeptide gives rise to a unique pattern of peaks. In this two-dimensional spectrum, peaks are obtained for amide protons bonded to ^{15}N-labeled nitrogen atoms with the position of each peak reflecting the chemical shift of these atoms. Therefore, an ideal spectrum of a protein should show a single peak for each non-proline residue, corresponding to its backbone amide group. Therefore, one of the first qualities of a $^{1}H - ^{15}N$ HSQC spectrum that is evaluated for a new protein sample is the number of backbone amide peaks that are detected. If a

significant number of the backbone amide peaks cannot be observed, then sample conditions must be altered to increase the number of peaks that are detected.

Another feature of the $^{1}H - ^{15}N$ HSQC spectrum that is also evaluated for new protein samples is the extent of proton chemical shift dispersion, since this can provide information regarding the status of the protein's structure. Usually a folded protein will give rise to peaks with a wide range of amide proton chemical shifts (~ >1 ppm) while an unfolded protein will have a dispersion that generally ranges from 0.5 – 1 ppm. In addition, the unfolded portions of a protein will tend to exhibit much greater intensity than stably folded parts of the protein due to the greater mobility exhibited by highly flexible polypeptide segments. Therefore, an ideal NMR spectrum will contain a significant proportion of peaks with a wide proton chemical shift dispersion and a similar peak intensity that would suggest that the protein is in a folded conformation under NMR sample conditions. When this is not the case, sample conditions (e.g. pH, temperature, ionic strength) can be altered to promote a folded state for the protein under investigation.

1.5.3 Triple Resonance Multidimensional NMR Spectroscopy

Once optimal NMR sample conditions have been determined a series of three-dimensional experiments can be recorded in order to assign resonances to backbone atoms. In these experiments the data is represented in a 3-dimensional box with the chemical shift of three correlated atoms recorded in each dimension. An example is shown in Figure 1.6 which shows a schematic diagram of an HNCO spectrum. In this type of experiment the amide proton and nitrogen of each amino acid is also correlated with the carbonyl carbon from the preceding residue. As is the case for many three-dimensional NMR experiments, the HNCO is named for the atoms that are detected in

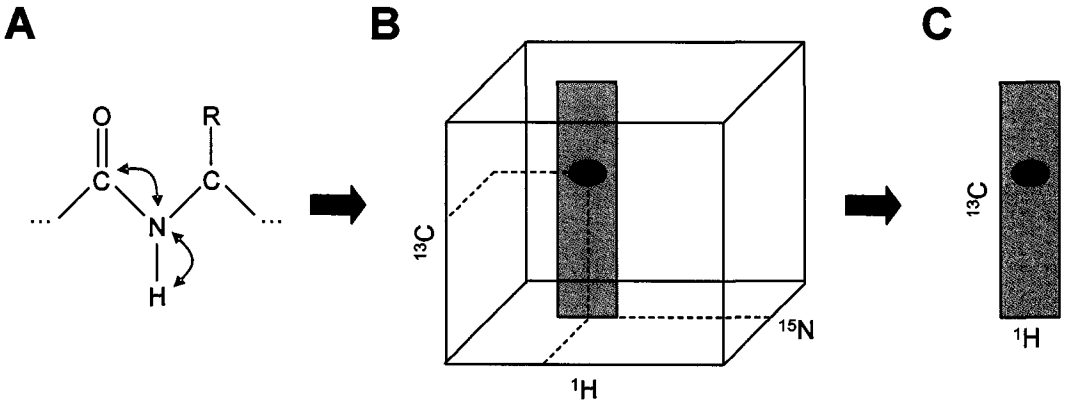


Figure 1.6 Overview of an HNCO experiment. **(A)** Magnetization is transferred from the amide proton via its directly attached nitrogen to the directly attached carbonyl carbon of the previous residue. **(B)** Schematic view of the spectrum that would be obtained for a ^{13}C, ^{15}N-labeled dipeptide and, **(C)** the ^{1}H-^{13}C strip extracted from the plane corresponding to the ^{15}N chemical shift of the dipeptide peak.

the experiment (i.e. amide proton (H), nitrogen (N) and carbonyl carbon (CO)). Similarly, an experiment that correlates the amide proton (H) with its directly attached nitrogen (N) and the alpha carbon (CA) would be called an HNCA.

As shown in Figure 1.6C, to simplify visualization of a 3D spectrum, two-dimensional strips can be extracted using analysis software such as NMRView (86, 87). For the purpose of backbone assignment, strips are usually taken such that the x-axis represents the proton dimension and the y-axis represents the carbon dimension, all at a specific ^{15}N chemical shift.

1.5.4 Backbone Chemical Shift Assignment

Two of the most important experiments for backbone assignment of proteins are the HNCACB and the CBCA(CO)NH (85, 88, 89). The HNCACB correlates the resonances of the amide proton and nitrogen atoms to the inter-residue and intra-residue alpha and beta carbon atoms while the CBCA(CO)NH does so for the inter-residue carbons only (Figure 1.7A). When two-dimensional strips are extracted at a specific ^{15}N chemical shift, the CBCA(CO)NH will reveal two peaks (inter-residue $^{13}C_\alpha$ and $^{13}C_\beta$) while the HNCACB will contain four (inter-residue $^{13}C_\alpha$ and $^{13}C_\beta$ and intra-residue $^{13}C_\alpha$ and $^{13}C_\beta$) (Figure 1.7B). Conveniently, the $^{13}C_\alpha$ and $^{13}C_\beta$ signals in the HNCACB experiment are of opposite phase, thus making the assignment of alpha versus beta carbons unambiguous. Furthermore, the peaks arising from the inter-residue correlations are generally weaker in intensity than the intra-residue correlation peaks and can be confirmed in the CBCA(CO)NH spectrum. Using the inter- and intra-residue correlations made available by these two spectra, sequential residues can be identified by matching

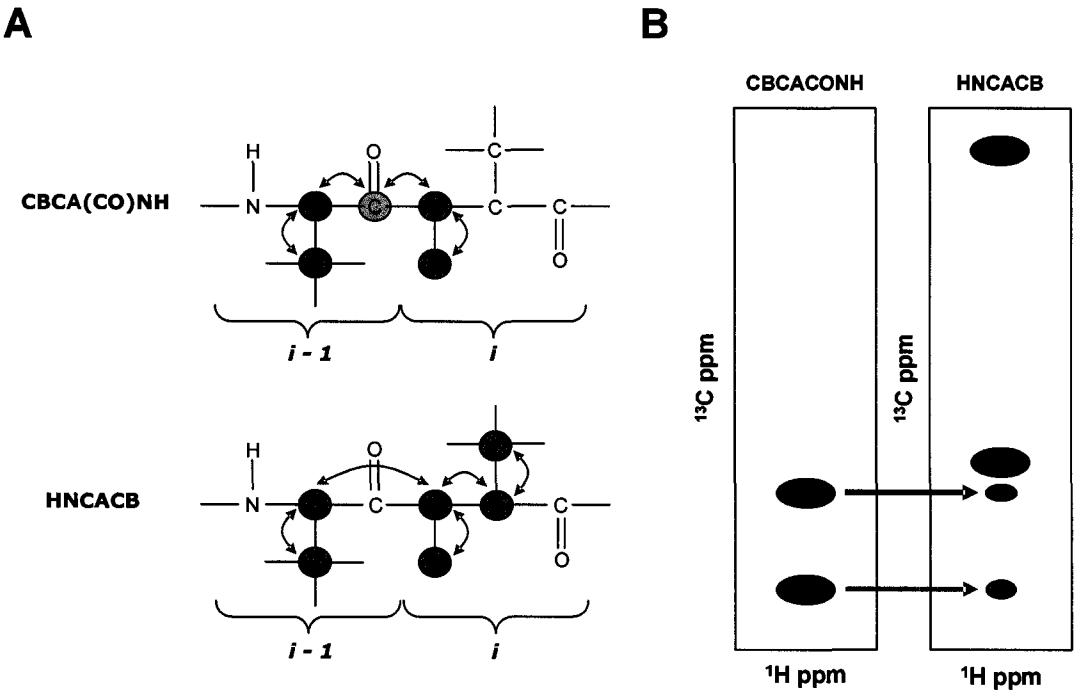


Figure 1.7 Overview of the triple-resonance experiments used for backbone assignment in this thesis. **(A)** Magnetization transfer pathways are illustrated for CBCA(CO)NH and HNCACB experiments. **(B)** Schematic diagram of the corresponding ^{1}H-^{13}C strip plots. In the HNCACB strip peaks arising from β carbons (red) are opposite in phase from Cα peaks (black) and inter-residue correlations give rise to smaller peaks than intra-residue correlations. The CBCA(CO)NH strip only shows correlations with carbon atoms from the preceding residue.

inter-residue carbon chemical shifts from one residue with intra-residue carbon shift from another residue.

In order to illustrate the assignment process a schematic diagram of a series of HNCACB strips from three sequential residues is shown in Figure 1.8, arranged as they would be after assignment. The first residue in the sequence (*i* - *1*) shows two intense peaks in the HNCACB strip that correspond to its $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ atoms. The matching HNCACB strip from the next residue (*i*) shows lower-intensity peaks corresponding to inter-residue correlations at the same carbon chemical shift (y-axis) positions as these major peaks in the (*i* – *1*) HNCACB strip. The match between inter- and intra-residue carbon chemical shifts between the two HNCACB strips confirms that these residues are adjacent in the sequence. A similar match is present between the inter-residue correlations in the strip for the next residue (*i* + *1*) and intra-residues correlations in the strip for residue (*i*). In addition to this strip-matching process, it is also necessary to determine the part of the protein amino acid sequence that this series of strips corresponds to. For this purpose the carbon chemical shift values are compared to database-derived average shifts observed for particular amino acid types (90). For example, the beta-carbon of alanine has a low characteristic chemical shift value while this shift is much higher in serine and threonine residues. Overall, this information can be coupled with the identification of sequential residues to allow complete assignment of backbone ^{15}N, ^{13}C and $^{1}H^{N}$ chemical shifts.

1.5.5 Secondary Structure Determination from Assigned Backbone Chemical Shifts

Using the chemical shift data obtained from backbone assignment, secondary structure prediction based on a relatively simple method originally described by Wishart

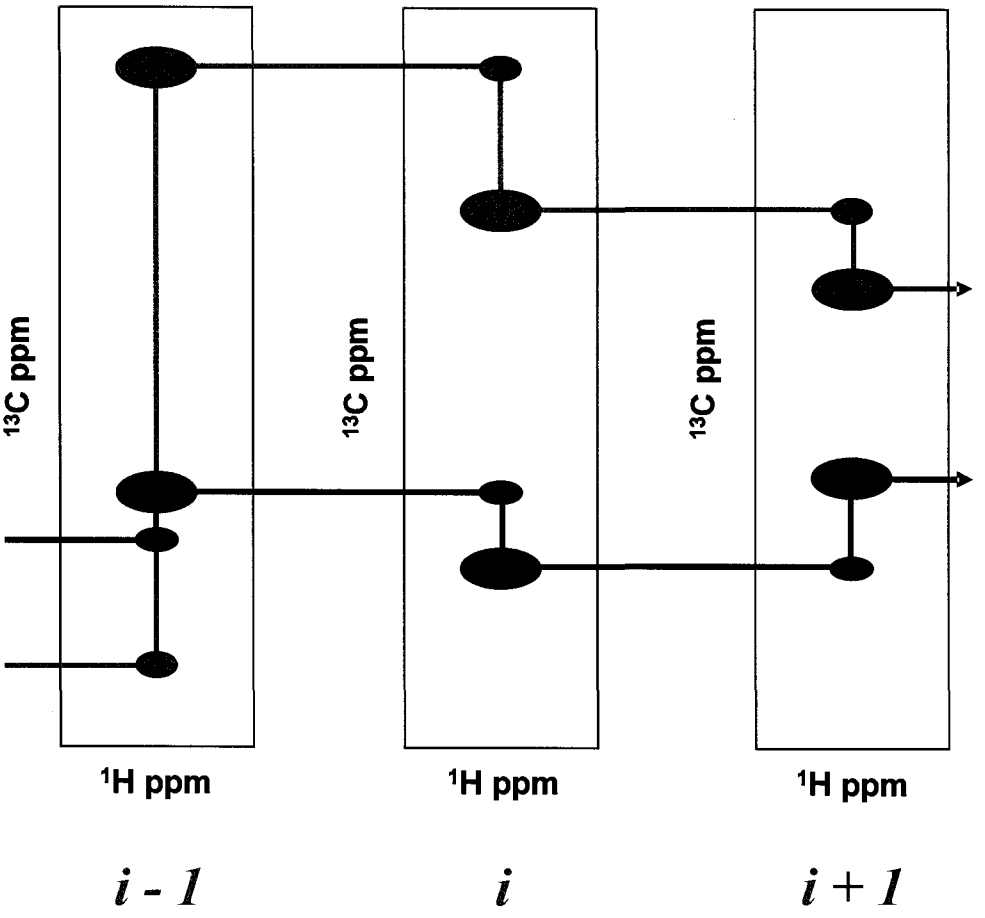


Figure 1.8 Example of HNCACB strips from 3 sequential residues. Sequential connectivities for C_α atoms are indicated with black lines and for C_β atoms with red lines.

and Sykes (91, 92) can be performed. For each residue in the sequence, the chemical shift that would be expected for that residue if it were in a random coil conformation is subtracted from the experimentally determined value to obtain the secondary chemical shift. The secondary chemical shifts are directly correlated with specific secondary structure types allowing an accurate secondary structure prediction to be derived from these values (93). For example, $^{13}C_{\alpha}$ atoms experience a downfield shift relative to random coil values when they are situated in helices and an upfield shift in β-strands. Therefore secondary $^{13}C_{\alpha}$ chemical shifts will be positive for α-helical structures and negative for β-sheets. Continuous stretches of strongly positive or negative secondary shifts will therefore provide a rapid indication of a protein's secondary structure since several adjacent residues which experience the same type of secondary shift will be a strong predictor of an α-helix or a β-sheet, respectively.

As outlined in Figure 1.9, this process of chemical shift assignment is just one part of the protein structure determination process by solution NMR. As previously mentioned, a pure, concentrated protein sample must first be generated and conditions optimized to allow high quality NMR spectra to be obtained. Once chemical shift assignments are done for the backbone, they must be completed for side-chain atoms using similar types of experiments. Information regarding the three-dimensional structure of the protein must then be obtained through measurement of the nuclear Overhauser effect (NOE) between protons that are close together in space (94). Identification of those atoms that are within 5 Å of each other can then be used, along with other sources of structural information to calculate a structure for this protein that is consistent with all of these structural restraints (95). In general, the greater the number

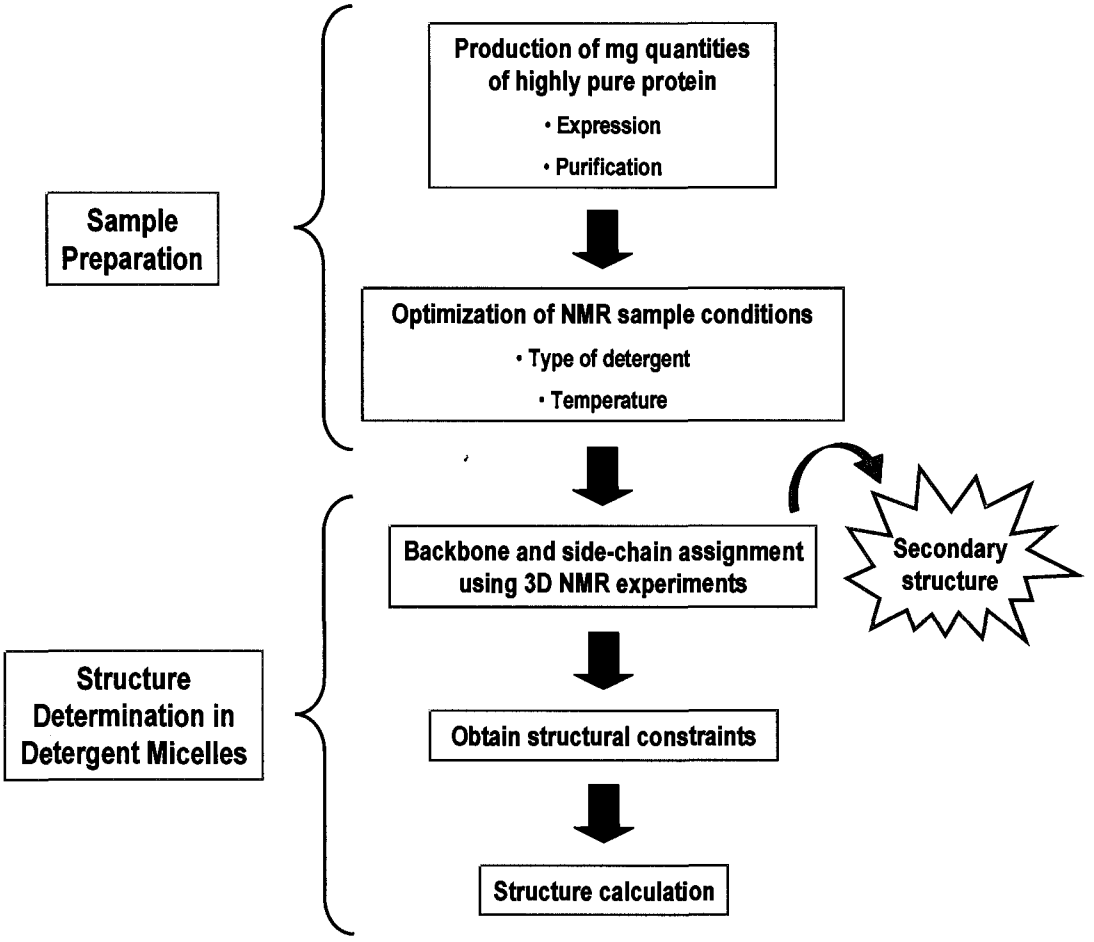


Figure 1.9 General outline of the process of structure determination by solution NMR spectroscopy.

of restraints that are used to calculate this structure, the more precise and accurate the structure will be (96). In this thesis I have focused on the initial stages of this process, namely NMR sample preparation and preliminary backbone chemical shift assignment as I worked toward the long-term goal of determining a high-resolution structure for p7.

1.5.6 Progress in Membrane Protein Structure Determination by Solution NMR

Structure determination of membrane proteins via solution NMR can be challenging due to difficulties in obtaining sufficient quantities of isotope-labeled samples. In addition, when detergents are used to solubilize these samples the size of the micelle-protein complex can be large (i.e. >30 kDa) which can lead to NMR sensitivity problems. However, much progress has been made in the field of NMR allowing the structure determination of larger protein complexes than ever before. For example, improvements in spectrometer hardware, the development of special isotope-labeling schemes and relaxation-optimized experiments (e.g. ^{1}H-^{15}N TROSY-HSQC (97)) have together allowed the structure determination of much larger protein complexes than previously possible, including the structure of a malate synthase protein that is >70 kDa in size (97-102). Using these types of advances the structures of detergent-solubilized membrane proteins from the β-barrel class were also solved (e.g. PagP in β-OG, OmpX in DHPC and OmpA in DPC (63, 103, 104)), highlighting the potential of the technology to provide new insights into membrane protein structure.

In spite of these impressive advances in solution NMR of membrane proteins, structures of polytopic α-helical membrane proteins have proven to be more difficult to access through this technique. One difficulty is that the chemical shift dispersion observed for these types of proteins is narrow, complicating chemical shift assignment. It

is also more difficult to obtain long-range NOE data that provide information about how the helices are arranged with respect to each other for polytopic α-helical membrane proteins. In fact, there are no reliable three dimensional structures for single-chain polytopic helical membrane proteins in detergent micelles yet determined by solution NMR. However, a small number of structures have been determined for α-helical membrane proteins with single TM helices that form oligomers. For example, the structure of the dimer formed by the single TM segment of glycophorin A in DPC micelles was determined to atomic resolution by solution NMR (23). More recently the solution NMR structure of the homopentamer formed by the single TM segment of phospholamban in DHPC micelles was also determined (105). In both cases the relatively high affinity of the self-interaction in detergent conditions made it possible to gain new insights into how these TM segments form the functionally relevant oligomeric state. Since p7 also forms oligomeric structures in detergents (37), solution NMR may be able to reveal similar types of insights into p7 function and structure.

1.6 Rationale and Objectives

p7 is essential for virus production and infectivity and therefore may be a target for the development of new HCV-specific antiviral compounds. A high-resolution structure has the potential to provide great insight into p7 function, as well as help in potential future drug design. Since no structural information is yet available for the HCV p7 protein I have initiated NMR sample development and solution NMR studies to study the structure of this protein. Therefore the objectives in this thesis are:

(i) Develop a biosynthetic protocol for the production of pure p7 protein sample suitable for NMR structural studies.

(ii) Determine the secondary structure of HCV p7 in a micellar environment using secondary backbone chemical shifts.

Successful achievement of these objectives will open the door to future elucidation of a high-resolution structure for HCV p7.

Chapter 2 MATERIALS AND METHODS

2.1 Competent Cells

E. coli were transformed using the $CaCl_2$ method adapted from Sambrook (106). To make competent cells a frozen stock of *E. coli* bacteria (DH5α (Invitrogen) or BL21(DE3) (Novagen)) was plated on non-selective agar and incubated overnight at 37°C. A single colony was inoculated into 50 ml of Luria-Bertani (LB) broth (Sigma) and shaken O/N at 220 rpm, 37°C. 0.5 ml of the O/N culture was used to inoculate 50 ml of LB and shaken at 37°C until the optical density at 600 nm (OD) reached approximately 0.3-0.4. The culture was centrifuged at 6 000 rpm at 4°C for 10 min and the pellet was resuspended with 12.5 ml of 100 mM $MgCl_2$. The cells were pelleted at 5 000 rpm at 4°C for 10 min and resuspended in 25 ml of ice cold 100 mM $CaCl_2$. The cells were incubated on ice for 30 min and then centrifuged at 5 000 rpm at 4°C for 10 min. The pellet was resuspended in 1 ml 85 mM $CaCl_2$ in 15% w/v glycerol solution, frozen in liquid nitrogen and stored at –80°C.

2.2 Plasmid DNA Minipreps

One colony from bacteria that had been transformed with the desired plasmid was inoculated into 2 ml LB medium containing 100 μg/ml of the antibiotic ampicillin and shaken overnight at 220 rpm, 37°C. 1.5 ml of the overnight culture was centrifuged at 12 000 g for 30 seconds (sec) at 4°C. The pellet was resuspended in 350 μl of STET (0.1 M NaCl, 10 mM Tris·Cl pH 8.0, 1 mM EDTA, 5% Triton X-100) to which was added 25 μl

of fresh lysozyme solution (10 mg/ml in 10 mM Tris·Cl pH 8.0). Mixing occurred by vortexing and the tube was placed in boiling water for 40 sec. The bacterial lysate was centrifuged at 12 000 g for 10 min at room temperature from which the pellet was removed. 40 µl of 2.5 mM sodium acetate and 420 µl of isopropanol were added to the supernatant, vortexed and incubated at room temperature for 5 min. The DNA was recovered by centrifugation at 12 000 g for 5 min at 4°C to which was added 1 ml of 70% ethanol. The pellet was once again recovered by centrifugation at 12 000 g for 2 min at 4°C and allowed to air dry until all traces of ethanol seemed to be removed. The nucleic acids were then re-dissolved in 50 µl of TE (pH 8.0) containing DNAase-free pancreatic RNAase (20 µg/ml) and stored at –20°C.

2.3 Construction of Fusion Recombinant Plasmids

The following section was performed by Sylvie Bélanger and Selena Sagan from Dr. John Pezacki's laboratory at the National Research Council of Canada. Briefly, the amino acid consensus sequence of the HCV genotype 1a p7 protein was codon optimized for bacterial expression, then synthesized and cloned into the vector pDrive by DNA 2.0. p7 was PCR amplified from the pDrive-p7 construct and cloned into pGEX-KT (Amersham Biosciences) digested with *Bam*HI and *Eco*RI (New England BioLabs), thus generating the plasmid pGEX-KT-p7.

After receiving this construct, Jenny Cheng subjected the plasmid to site-directed mutagenesis (Stratagene) according to the manufacturer's instructions to introduce a hexa-histidine tag at the C-terminus of p7, thus forming the plasmid pGEX-KT-p7-His, and at the C-terminus of GST, thus forming the plasmid pGEX-KT-His-p7. Both plasmids were verified by DNA sequencing.

2.4 GST-p7 fusion protein expression

The *E. coli* strain BL21(DE3) was transformed with a plasmid containing the p7 sequence (Appendix 1) as a C-terminal fusion to GST with a hexahistadine tag either at the N-terminal or C-terminal side of p7 (pGEX-KT-His-p7 or pGEX-KT-p7-His, respectively), as provided by Jenny Cheng (Goto lab) and Selena Sagan (Pezacki lab). Typically, one colony was used to inoculate 100 mL of M9 minimal media (7.0 g/L Na_2HPO_4, 3.0 g/L KH_2PO_4, 0.5 g/L NaCl, 11 mg/L $CaCl_2$, 120 mg/L $MgSO_4$, 50 mg/L thiamine, 1% v/v LB, 3 g/L D-glucose, 1g/L NH_4Cl) supplemented with 100 μg/ml ampicillin and shaken overnight at 37°C. The next morning, the pellet from the overnight culture was transferred into 1 L of fresh M9 and the cells were further grown until an OD of ~0.5 was reached. Protein expression was induced by addition of isopropyl-β-D-thiogalactopyranoside (IPTG). These cells were harvested the next morning by centrifugation at 4 800 g for 10 min at 4°C. They were then stored at -20°C until further use.

2.5 Fusion protein purification

Cell pellets from a 1 L expression were resuspended in 100 ml of lysis buffer (0.5 M NaCl, 0.02 M Tris pH 8.0) and divided into four 50 ml Corning tubes. Cell lysis was achieved by sonication on ice using a Sonic Dismembrator Model 500 (Fisher Scientific) at 50% amplitude for 1 min, with 1 sec on and 1 sec off pulses, for 2 rounds at 5°C. The cells lysates were then centrifuged at 20 000 g for 30 min at 4°C. Once the supernatant was removed, the inclusion body-containing pellets were each resuspended at 5°C, O/N on a rotating platform, in 10 ml of Bind buffer (8 M urea, 0.1 M Na_2HPO_4 pH 8.0). Crude protein extract was centrifuged at 30 000 g for 30 min at 4°C. Purification was

achieved by passing the supernatant of this sample through a 10 mL nickel nitriloacetic acid (NTA) resin (Novagen) that had been equilibrated with Bind buffer. The first flow-through was then reapplied to the column to maximize binding. For optimized purification of the fusion proteins, a total of 100 ml Wash buffer (8 M urea, 0.1 M Na_2HPO_4 pH 8.0, 0.03 M imidazole) was passed through the column. The His-tagged fusion protein was finally eluted with 15 ml of Elution buffer (8 M urea, 0.1 M Na_2HPO_4 pH 8.0, 0.4 M imidazole). Eluted protein was then immediately dialyzed against 4 L of 0.02 % SDS at 5°C overnight.

2.6 Thrombin Cleavage

For trial thrombin cleavage reactions, protein was precipitated by dialysis against water and was resuspended in 8 M urea pH 8.0. To all reactions, thrombin cleavage buffer (20 mM Tris-HCl pH 8.4, 150 mM NaCl, 2.5 mM $CaCl_2$) was added. Various thrombin and urea concentrations were tested, as were different incubation temperatures. After optimization, cleavage was done with samples that were dialyzed against 0.02 % SDS, keeping p7-His in solution. The sample was concentrated using an Amicon Ultra-15 centrifugal device with 10 kDa molecular weight cut-off (Millipore) down to approximately 1 ml at which point thrombin cleavage buffer (same as earlier) was added. 6 U of thrombin per mg of fusion protein was added and allowed to cleave the fusion protein overnight at room temperature. The approximate protein concentration was then determined using the BCA™ Protein Assay Kit (Pierce) according to the manufacturer's instructions.

2.7 Protein Purification – Reversed-phase High Performance Liquid Chromatography (RP-HPLC)

Initially, trial purifications were performed on a Waters 600E Multisolvent Delivery System with a Waters Delta Pak C4 column on an analytical scale. Typically, about 100 μg of cleaved protein was injected onto the column with flow rates ranging between 0.5 and 1 ml/min with various water/organic gradients (summarized in Table 3.1). Following these purification attempts, analytical scale purifications were performed on an Agilent 1100 Series with a 4.6 x 250 mm Zorbax 300SB-C3 column (Agilent Technologies). Approximately 100 μg of protein was injected onto the column and various water/acetonitrile gradients were attempted. For purifications performed on a preparative scale, 1 - 2 mg of protein was injected onto a 9.4 x 250 mm Zorbax 300SB-C3 column (Agilent Technologies) was used on a Gilson 806 Manometric Module. Typically, samples from the overnight cleavage reaction were injected onto the semi-prep C3 column which had been equilibrated with 60 % solution A (water, 0.1 % trifluoroacetic acid (TFA)) / 40 % B (20 % water, 80 % acetonitrile, 0.1 % TFA). A gradient of 40-100 % solvent B over 50 min was applied at a flow rate of 4 ml/min. p7 eluted at approximately 53 min, corresponding to approximately 54 % acetonitrile.

2.8 SDS-PAGE and Western Blotting

Fusion protein expression and purification was assessed by SDS-PAGE. Typically ~15 μl of sample in loading buffer was loaded either onto a 12 % Tris-Glycine acrylamide gel or onto a pre-cast 4-20 % gradient Tris-Glycine acrylamide gel (Invitrogen). The apparatus was set at 125 V for 110 min for the pre-cast gel or at 200 V

for 47 min for the 12 % gel in 500 ml 1X Tris-glycine buffer containing 0.1% w/v SDS, 1.44% w/v glycine, and 0.3% w/v Tris base.

For Western blot analysis, the transfer apparatus was set up to allow a freshly run SDS-PAGE gel to be transferred onto a polyvinylidene difluoride (PVDF) membrane (BIO-RAD) which was activated in methanol. The transfer was run at 25 V for 90 min in a buffer containing 0.36% w/v glycine, 0.072% w/v Tris base and 10% methanol. After transfer, a milk-based blocking solution was added to the membrane and incubated overnight at 4°C. The membrane was then transferred into blocking solution containing mouse anti-(6X)His monoclonal antibody (Clonetech) diluted 1:2000 and incubated for 90 min. The membrane was then washed twice with Tween-20 Tris-buffered saline (0.05% Triton X-100, 20 mM Tris, 0.5 M NaCl (TTBS)) for 10 min followed by incubation with Goat anti-mouse IgG polyclonal antibody conjugated with alkaline phosphatase (BIO-RAD) diluted 1:2000 in TTBS. The membrane is washed twice for 10 min with TTBS, then with TBS for 5 min. The blot was finally developed using the Alkaline Phosphatase Conjuguate Substrate Kit (BIO-RAD) according to the manufacturer's instructions.

2.9 NMR Spectroscopy

The eluted p7 fraction, now contained in a mixture of acetonitrile and water, was lyophilized overnight using a FreeZone® Plus™ 6 Freeze Dry System (Labconco). The protein powder was then resuspended in various detergent buffers (0.2 M detergent, 0.025 M NaCl, 0.025 M Na_2HPO_4 pH 6.5). 10% D_2O was added to this liquid sample (typically containing approximately 0.2-0.3 mM protein) which was then introduced into a Shigemi NMR tube. All initial detergent and temperature optimization HSQC

experiments were recorded on a Varian Inova500, processed by NMRPipe and analyzed with NMRView. Backbone chemical shift assignments were made on the C-terminally His-tagged p7 at 40°C, pH 6.5 based on HNCACB and HN(CO)CACB experiments which were recorded at the Quebec/Eastern Canada High Field NMR Facility on the cryoprobe equipped Inova500 (see Table 2.1 for spectral parameters). ^{1}H, ^{15}N and ^{13}C chemical shifts were referenced using a 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) internal standard. Secondary chemical shifts from peptide-derived random coil values (92) were calculated using NMRView.

Table 2.1 Spectral Data Set

Experiment	Transients (nt)	^{1}H Spectral Width (sw, Hz)	^{1}H Complex Points (np)	^{15}N Spectral Width (sw1, Hz)	^{15}N Complex Points (ni)	^{13}C Spectral Width (sw2, Hz)	^{13}C Complex Points (ni2)
^{15}N HSQC	256	7017.5	1024	2030.4	64	-	-
CBCA(CO)NH	64	8000	800	8175.3	48	1469.9	30
HNCACB	64	8000	800	8175.3	48	1469.9	30
^{15}N-edited NOESY	16	13008.1	1666	9596.1	100	2147.8*	28*

*F2 is for ^{1}H in this experiment

Chapter 3 RESULTS

3.1 GST-p7 Fusion Protein Expression: Background

p7 was originally cloned into pGEX-KT as a C-terminal fusion to GST by Selena Sagan, a graduate student working in the Pezacki laboratory. A former graduate student in the Goto lab, Ms. Jenny Cheng, subsequently investigated the expression properties of this fusion protein, and found that it localized to inclusion bodies. Resolubilization of GST-p7 from these misfolded aggregates required the use of strong chemical denaturants which unfolded the fusion protein. This prevented straightforward purification by glutathione affinity chromatography since this technique requires that GST be in a folded state. Since it would be difficult to find refolding conditions that are compatible with both water-soluble GST and the highly hydrophobic p7 protein, a hexahistidine (His) tag was introduced into the fusion sequence to allow purification by nickel affinity chromatography, a technique that is compatible with the use of strong denaturing conditions. For this purpose two new constructs were made by Ms. Cheng: one with the His-tag at the C-terminus (GST-p7-His) and one with the His-tag on the N-terminal side of p7 (GST-His-p7) (Figure 3.1).

3.2 Optimization of Protein Expression

In order to determine whether the new His-tagged GST-p7 fusion proteins would be expressed, expression trials were performed using the *E. coli* strain BL21(DE3). Coomassie-stained SDS-PAGE gels of whole cell lysates from these initial trials showed

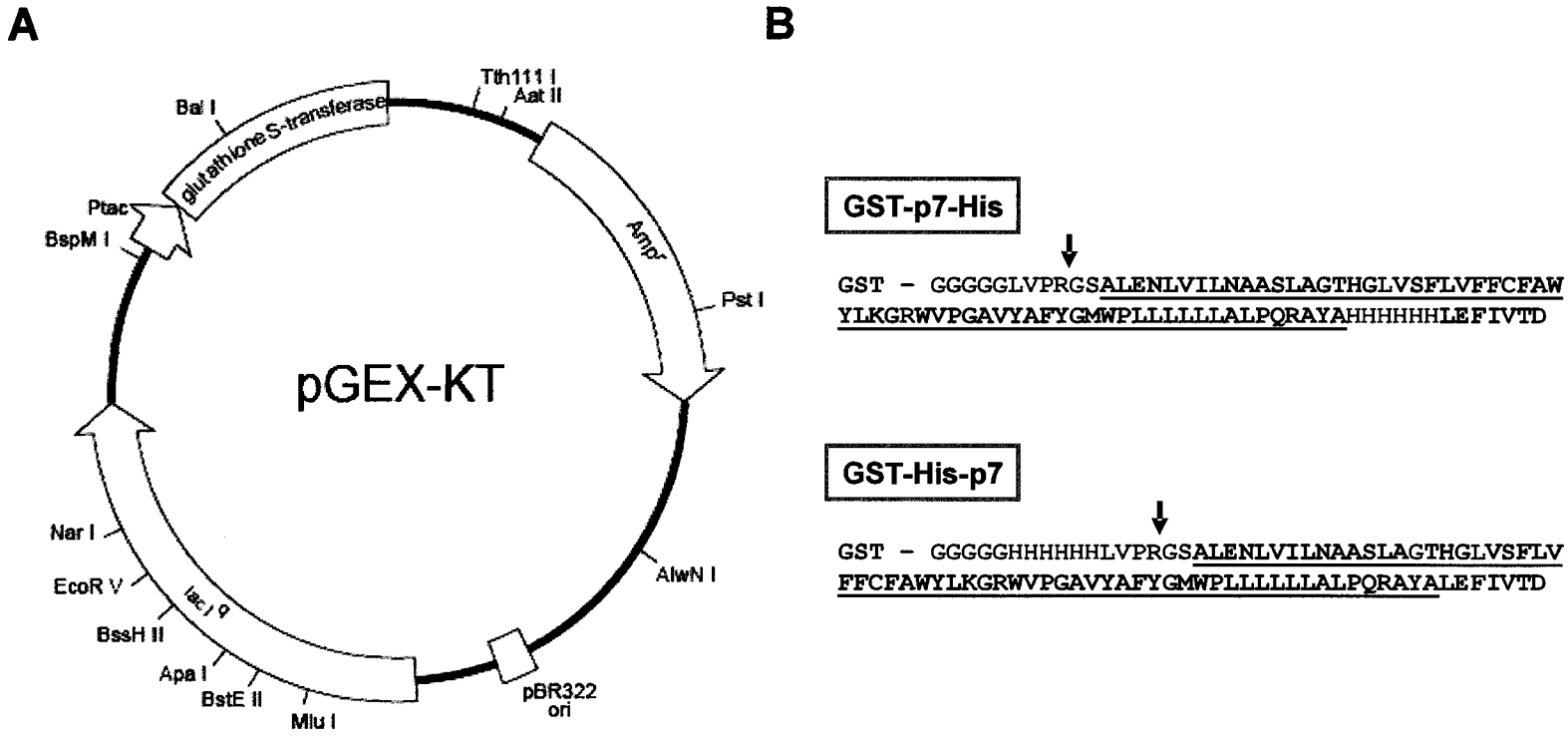


Figure 3.1 Expression plasmid and sequence of p7 fusion proteins. (**A**) Map of pGEX-KT expression vector used to construct the p7 fusion protein to GST. GST fusion protein expression is under the control of a *tac* promoter (Ptac) which can be further controlled by expression of the $lacI^q$ gene. This plasmid also confers resistance to the antibiotic ampicillin through expression of β-lactamase (Amp^R). A codon-optimized sequence for p7 was introduced to the C-terminal coding region of GST to give the protein sequences shown in (**B**). In these sequences GST is followed by a five-glycine linker (blue) and a consensus sequence for thrombin cleavage (green). The black arrow indicates the thrombin cleavage site while the underlined amino acids indicate the p7 amino acid sequence. A His tag (red) was inserted at the C-terminus of p7 (GST-p7-His) or at its N-terminus (GST-His-p7). (Diagram in (A) adapted and modified from http://www4.amersham biosciences.com/pdfs /970004m2-01.pdf)

unique protein bands arising upon induction that were more intense than background total *E. coli* protein bands (Figure 3.2A). These species migrated with a molecular weight of approximately 34 kDa corresponding to the expected molecular weight of both GST-p7 fusion proteins. The presence of the His tag was verified by Western Blot analysis with an anti-His_6 antibody (Figure 3.2B). In the case of the centrally His-tagged construct, degradation products were also observed in the Western blot.

Induction temperature and extent of growth prior to induction are common parameters that can be investigated to optimize protein yields (56). Therefore, a range of these conditions were tested for both GST-p7-His and GST-His-p7. Typically, expression was performed at 37°C since it is a standard bacterial growth temperature. When other induction temperatures were tested (15, 20, and 30°C) to try and improve protein yields, they were found to decrease the amount of fusion protein expressed leaving 37°C as the best induction temperature. However, when induction was performed at an optical density at 600 nm (OD_{600}) of 0.5 instead of 0.7 (Figure 3.2), a significant improvement was observed (Figure 3.2). Therefore, induction at 37°C at an OD_{600} of 0.5, with 0.5 mM IPTG in BL21(DE3) *E. coli* cells provided the highest level of fusion protein expression.

3.3 Fusion Protein Purification Using Nickel Affinity Chromatography

As expected, after cell lysis the His-tagged GST-p7 fusion proteins did not localize to the soluble fraction, as had previously been observed by Ms. Jenny Cheng with the non-His-tagged version of this fusion protein. I found that strong denaturants (6 M guanidine or 8 M urea) were required to solubilize the insoluble fraction containing the p7 fusion proteins. Purification via nickel affinity chromatography was attempted in the

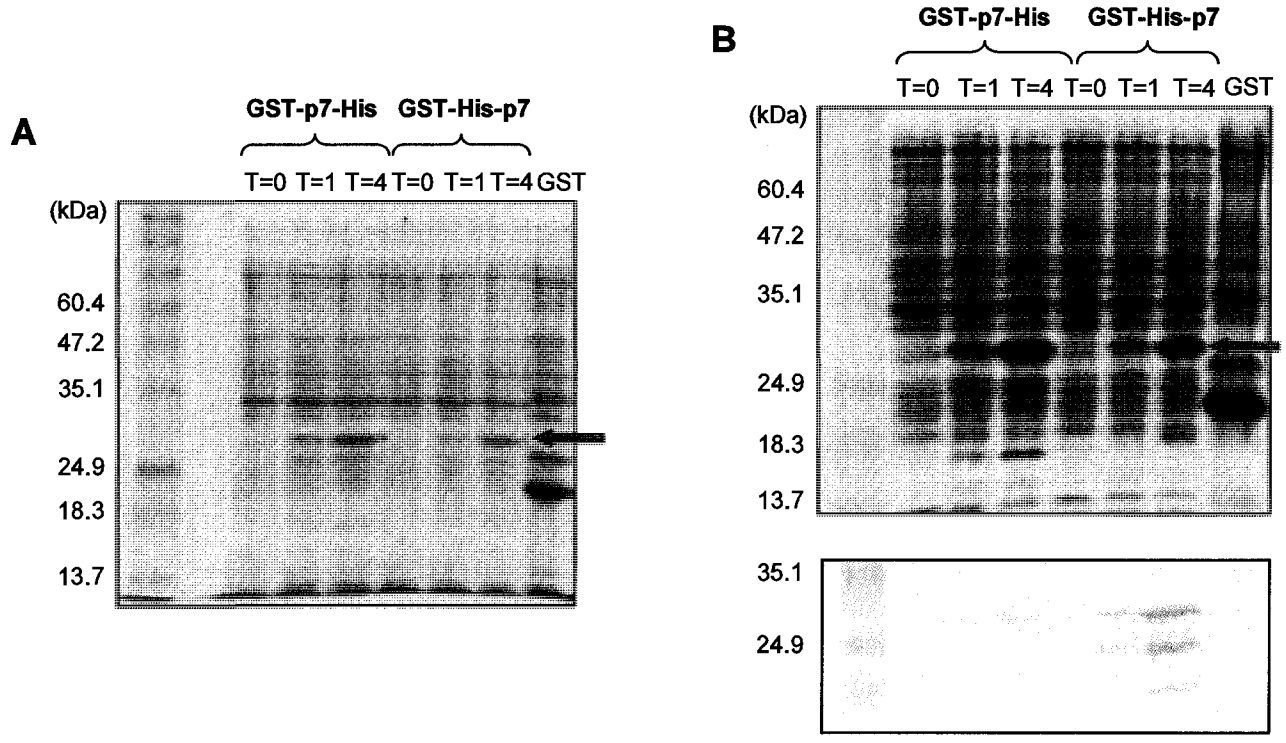


Figure 3.2 Expression of GST-p7-His and GST-His-p7 before (A) and after (B) optimization of expression monitored by SDS-PAGE (top) and anti-His tag Western blot (bottom). BL21(DE3) *E. coli* whole cell lysates from 75 μL of cell culture were obtained before (T=0), 1 hour after (T=1) and 4 hours after (T=4) induction with 0.5 mM IPTG at 37°C. Cells were induced at an OD_{600} of approximately 0.7 **(A)** or 0.5 **(B)**. A positive control lane (GST) has also been included to show the mobility of the GST protein (26 kDa) relative to the fusion proteins with p7 (~ 34 kDa).

presence of 8 M urea using standard protocols according the manufacturer's instructions. As shown in Figure 3.3, this resulted in approximately 50 % of the His-tagged fusion protein binding to the nickel resin, with the remainder being recovered in the flow-through. This significant amount of fusion protein loss resulted in poor purification yields, making it necessary to optimize nickel column binding conditions.

Guanidine is a chemical denaturant that is considered to be a more effective solubilizing agent than urea (107) and therefore was tested for its potential to improve accessibility of the His tag to the resin-bound nickel. Yet, when guanidine was used instead of urea, it did not improve binding for either fusion protein (Figure 3.3A). It has also been observed that addition of the detergent SDS to urea solutions can improve the nickel resin-binding of GST fusions to membrane proteins, again by potentially allowing better access of the His tag to the nickel ions (60). However, this did not aid in the binding of the GST-p7 fusions, despite the use of 0.02 % (Figure 3.3A) or 0.1 % SDS (data not shown). In contrast, significant improvements to binding were obtained when longer exposure times to the nickel resin were employed. Most efficient binding was achieved by incubation of the crude protein with the resin at room temperature for at least one hour, followed by slow elution from the column (flow rate < 1 mL/min) (Figure 3.3A). The purification protocol was also improved by increasing the concentration of imidazole during subsequent wash steps. The results of these modifications to the standard protocol allowed isolation of approximately 3 - 4 mg of fusion protein to be obtained from 1 L of culture with a purity of > 80 % (Figure 3.4).

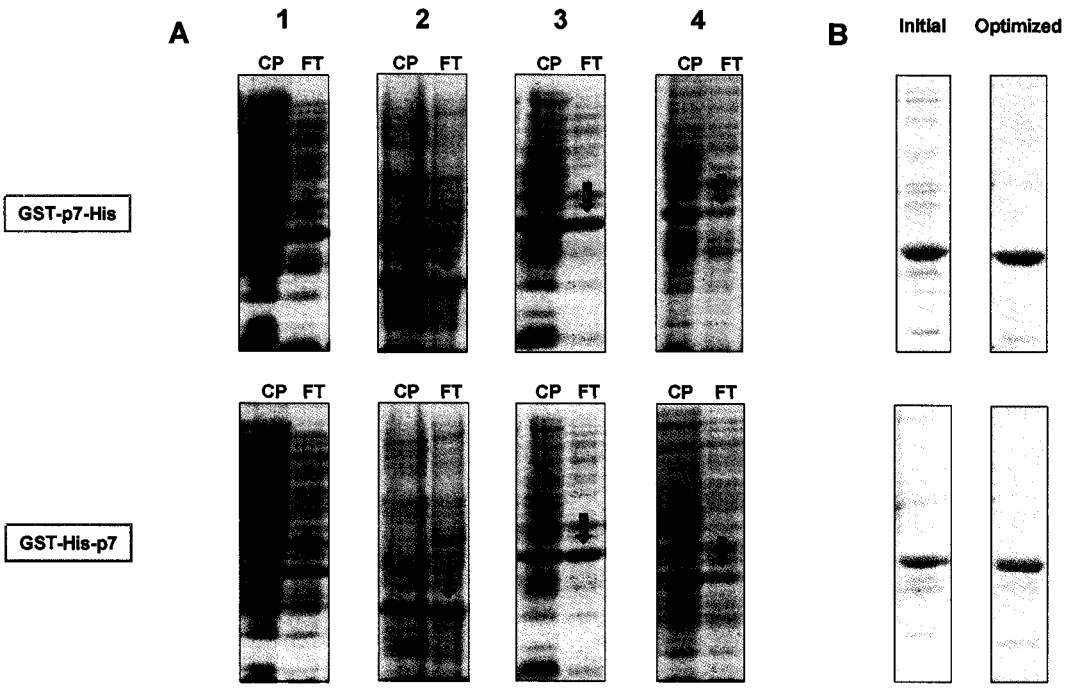


Figure 3.3 SDS-PAGE gels of GST-p7-His and GST-His-p7 trial purifications. **(A)** Investigation of the effect of buffer composition on nickel affinity column binding. The arrows indicate the migration of the fusion proteins on the SDS-PAGE gels. The "CP" lane represents the crude protein extract resulting from the extraction of induced cells solubilized in the corresponding buffer. The "FT" (flow-through) lane represents the collected fraction after having passed the CP once through the column. Buffer contains 0.1 M sodium phosphate pH 8.0 with **(1)** 8 M urea, 0.01 M Tris-Cl; **(2)** 6 M guanidine, 0.01 M Tris-Cl; **(3)** 8 M urea, 0.02 % SDS; **(4)** 8 M urea. **(B)** Examples of the elution fractions from nickel affinity chromatograpy purifications from p7 fusion proteins before and after optimization of the imidazole wash step.

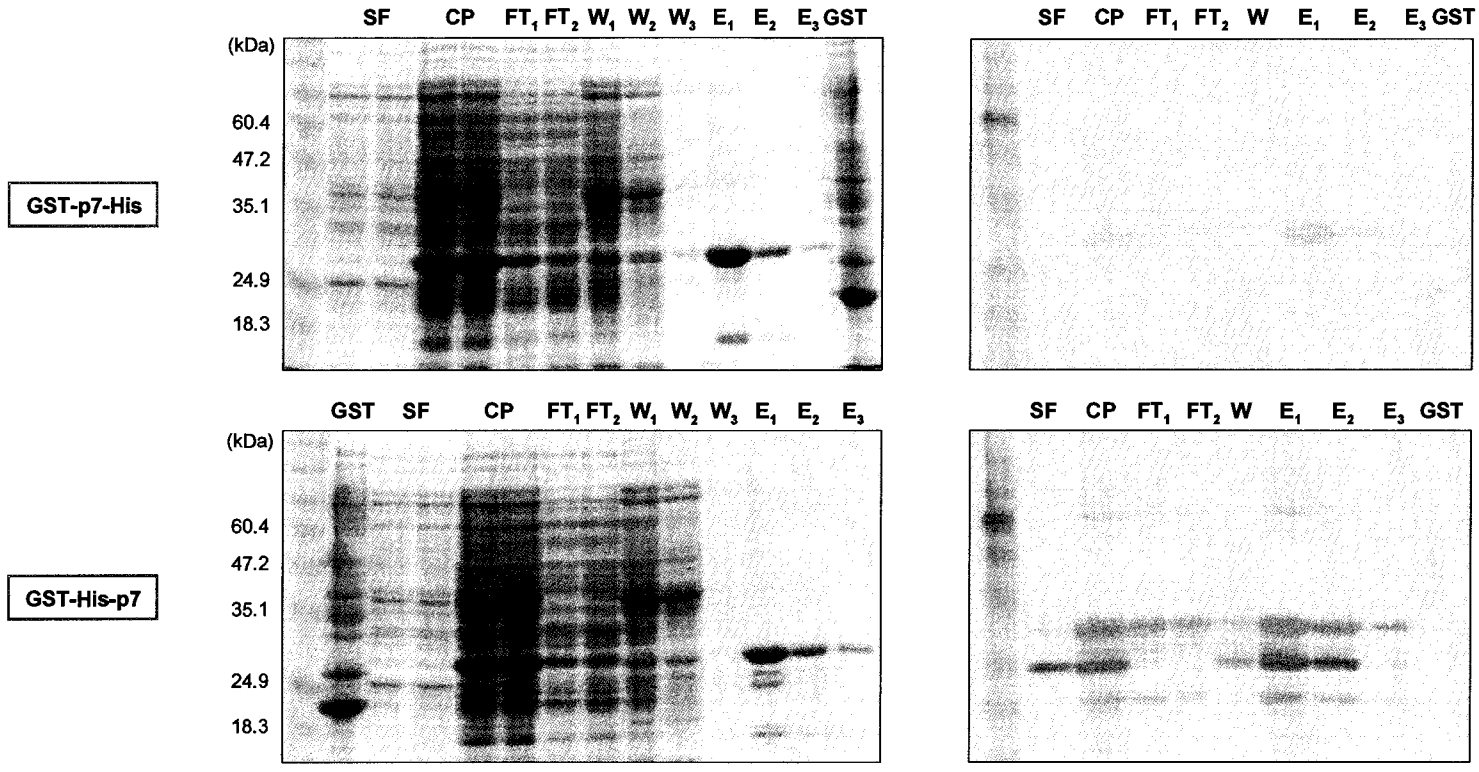


Figure 3.4 Coomassie stained SDS-PAGE gels (left) and anti-His tag Western blots (right) of fractions from optimized nickel affinity chromatography purifications of GST-p7-His and GST-His-p7. Soluble protein fractions (SF) were removed, and then the inclusion body-containing pellet was sonicated in 0.02 M Tris-Cl pH 8.0, 0.5 M NaCl, then centrifuged. The left-over pellet was resolubilized in bind buffer (8 M urea, 0.1 M sodium phosphate buffer pH 8.0) to give the crude protein extract (CP). The extract was loaded onto the column to give the first flow-through fraction (FT_1) which was then loaded again to give a second flow-through fraction (FT_2). W_1, W_2, and W_3 denote fractions eluted with wash buffer (8 M urea, 0.1 M sodium phosphate buffer pH 8.0, 0.03 M imidazole), and E_1, E_2 and E_3 indicate fractions eluted with elution buffer (8 M urea, 0.1 M sodium phosphate buffer pH 8.0, 0.3 M imidazole).

Although this purification was sufficient for subsequent thrombin cleavage reactions, the yields obtained for the centrally His-tagged construct were consistently lower than for the C-terminally His-tagged constructs. In addition, there were significant levels of impurities that retained the His-tag as shown by Western blot analysis (Figure 3.4) that were not very different in molecular weight from the desired product. Therefore, further work in p7 sample preparation was performed with the C-terminally His-tagged construct alone.

3.4 Thrombin Cleavage of GST-p7-His

Once GST-p7-His was purified, it was necessary to optimize conditions for thrombin cleavage which would allow separation of the GST carrier from p7. Since p7 is a membrane protein, conditions were required that would keep the fusion protein soluble while maintaining efficient thrombin protease activity. Therefore, several different cleavage reactions containing a range of urea concentrations were screened. As shown in Figure 3.5A, <50% of the fusion protein was cleaved when there was 1 - 2 M denaturant present (lanes 2 and 3) with cleavage being absent at higher urea concentrations (lanes 4 - 6). Under these conditions cloudiness was detected in the reaction mixture, indicating that the lower cleavage yields were possibly due to incomplete solubilization. To improve solubilization, sub-micellar concentrations of SDS were added to the reaction mixtures and found to improve reaction yields (Figure 3.5A lane 7). Ultimately it was found that >95% cleavage could be obtained in 0.02% SDS in the absence of any other chemical denaturant (Figure 3.5B). In addition, sample solubilization appeared to be complete under these conditions since precipitate was no longer observed. Other reaction conditions such as temperature, cleavage time and enzyme concentration were similarly

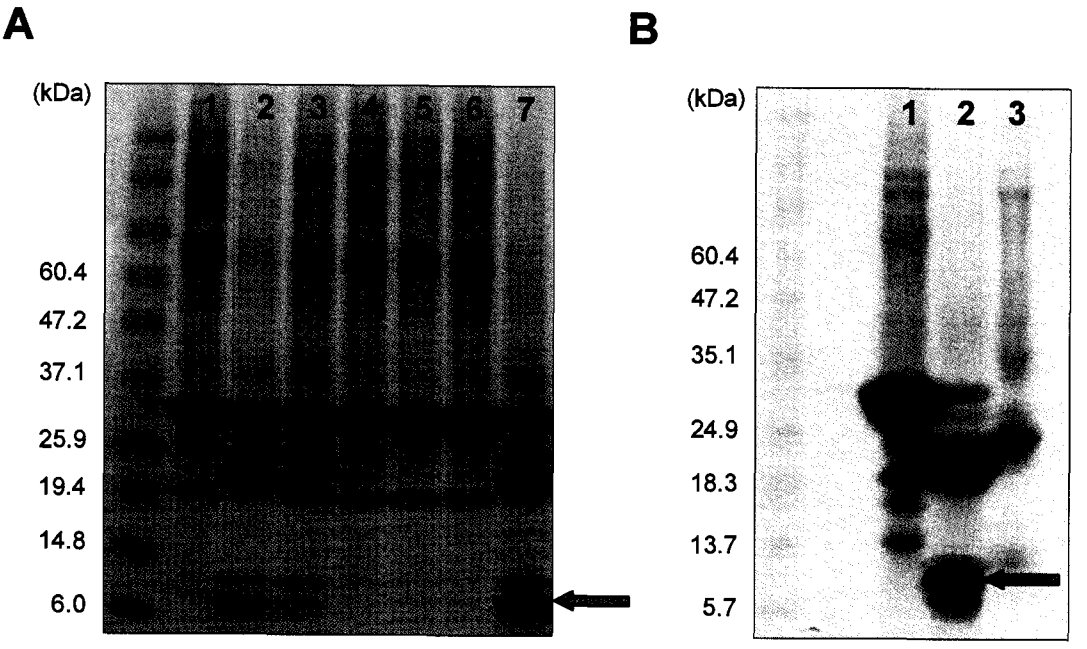


Figure 3.5 SDS-PAGE analysis of trial (A) and optimized (B) thrombin cleavage reactions of GST-p7-His. The arrow indicates the position of His-tagged p7. All thrombin cleavage reactions shown here were performed overnight at room temperature while varying the type and concentration of resolubilizing reagent. **(A)** Lane 1: Fusion protein alone, no thrombin. Lanes 2-6: Cleavage reactions containing 1-5 M urea respectively. Lane 7: Cleavage reaction with 1 M urea and 0.01 % SDS. **(B)** Lane 1: Fusion protein alone, no thrombin. Lane 2: Optimized cleavage reaction performed at room temperature, overnight in 0.02 % SDS. Lane 3: GST alone, no thrombin.

tested and showed that an overnight cleavage reaction with 6 units of thrombin per milligram of fusion protein at room temperature could be used to obtain nearly complete cleavage yields. The resulting cleavage mixture contained a dominant species in the SDS-PAGE gel with an apparent molecular weight of the expected p7-His protein (~8 kDa).

3.5 p7-His Purification

3.5.1 Nickel Affinity Chromatography

Following thrombin cleavage of GST-p7-His, it was necessary to purify p7-His from the cleavage mixture in order to obtain a pure sample suitable for structural studies. Again, the hydrophobicity of p7-His required the use of purification techniques that are compatible with denaturing conditions, ruling out the possibility that glutathione affinity chromatography could be straightforwardly employed to remove GST. Instead, since p7 is C-terminally His-tagged while GST is not, nickel affinity chromatography was investigated using a purification protocol similar to that used to purify the fusion protein. As shown in Figure 3.6, although some p7-His did appear to bind to the column, it was clear from the many flowthrough fractions that a significant amount was lost even when binding times were increased. An additional problem was that after column washing, a significant amount of GST co-eluted with p7-His. Similar results were obtained when more stringent washing procedures were used suggesting that GST may have been retained by non-specific interactions with p7-His. Although the presence of SDS in the denaturing buffer should have helped to prevent these types of non-specific interactions, it was apparently not sufficient to allow purification of p7-His by this method.

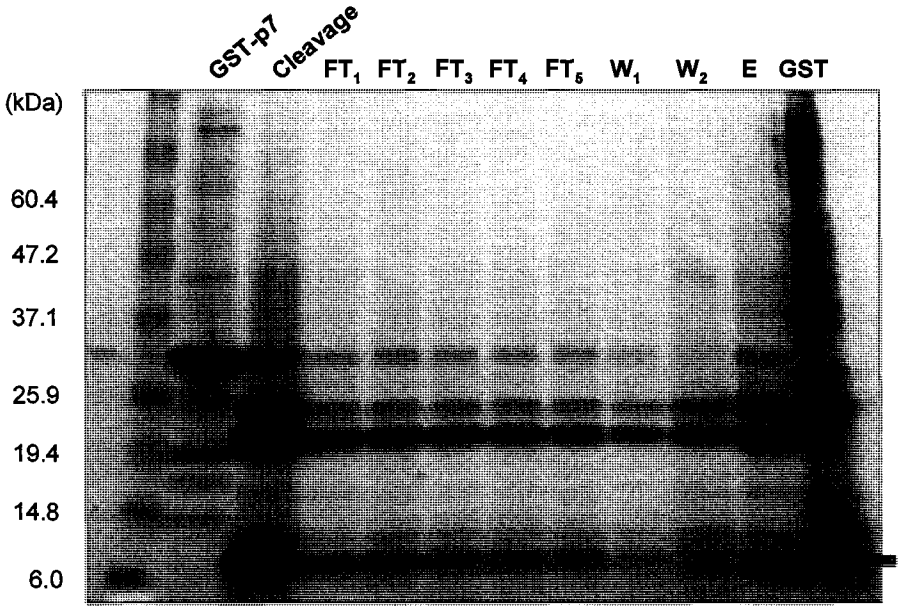


Figure 3.6 Nickel affinity chromatography of the cleaved fusion protein mixture. The arrow indicates the position of the His-tagged p7 protein. Intact fusion protein (GST-p7) was cleaved according to optimized conditions and bind buffer (8 M urea, 0.1 M sodium phosphate buffer pH 8.0) was then added to the cleavage reaction (Cleavage) and loaded onto a Ni-NTA chromatography column. The resulting flow-through (FT_1) was reloaded and collected 4 more times (FT_{2-5}). The column was then washed twice (W_1, W_2) with wash buffer (8 M urea, 0.1 M sodium phosphate buffer pH 8.0, 0.03 M imidazole). Bound proteins were finally eluted (E) with elution buffer (8 M urea, 0.1 M sodium phosphate buffer pH 8.0, 0.3 M imidazole).

3.5.2 Size Exclusion Chromatography

As nickel affinity chromatography was not successful in isolating p7 from the GST protein, other standard chromatographic methods that have previously been used to purify membrane proteins from proteolytic cleavage mixtures (62, 108) were also explored. One of these methods is size exclusion chromatography since it has been successfully used in denaturing solutions such as with 8 M urea (62, 109-112). Under these conditions, all proteins should be unfolded to a similar extent, potentially allowing separation based on differences in size and mobility properties of the different molecular weight species. As shown in Figure 3.7A, when the crude cleavage mixture was applied to the column in 8 M urea, all the p7-His eluted with GST in a single peak, approximately at the void volume of the column. SDS-PAGE analysis of peaks from this fraction confirmed that both GST and p7-His eluted in this fraction (Figure 3.7B), suggesting that the apparent molecular weight of these species exceeded 200 kDa. Although another peak appeared at a later elution point corresponding to a smaller species, SDS-PAGE analysis of fractions from this peak suggested that this was not a proteinaceous component of the mixture, since it was not visible in the gel (Figure 3.7B, lane 18). These results suggested that size exclusion chromatography with this column would not be a useful method for this purification.

3.5.3 Ion Exchange Chromatography

Another standard chromatographic method that has been used in membrane protein purification under fully denaturing conditions is cation exchange chromatography (62, 110, 113). Since the theoretical isoelectric point (pI) of GST is ~ 6 while that of His-

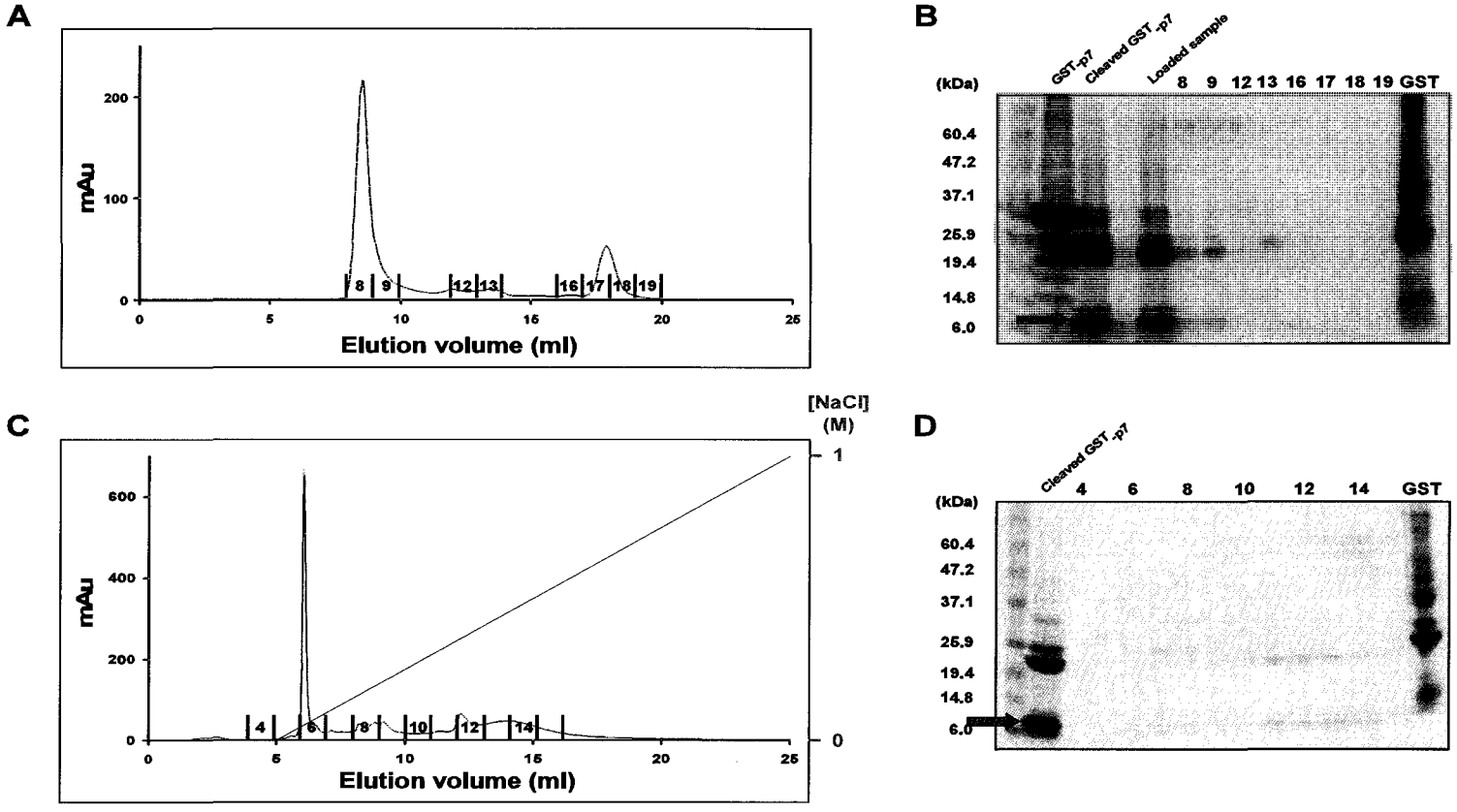


Figure 3.7 Size exclusion (A, B) or ion exchange (C, D) chromatography of the cleaved fusion protein mixture. Absorbance at 280 nm monitored as a function of elution volume (A, C) and the indicated eluted fractions analyzed by SDS-PAGE (B, D). For size exclusion chromatography the thrombin cleavage mixture was applied to a Superdex 200® 10/300 GL column at a flow-rate of 0.5 ml/min in 8 M urea, 0.05 M sodium phosphate buffer pH 8.0, 0.05 M NaCl. Ion exchange chromatography was performed using a Mono S^{TM} 5/50 GL column at a flowrate of 0.5 ml/min in 8M urea, 0.2 M sodium acetate buffer pH 5.0, with a 0 - 1 M NaCl gradient (purple line) run over a 20 mL volume.

tagged p7 is ~9 (calculated from the amino acid sequence using ProtParam on the ExPASy Server, http://www.expasy.ch/tools/pi_tool.html) (114), cation exchange chromatography was chosen to separate these two species. In neutral pH solutions, GST would not be expected to interact with an anionic resin, while p7-His should still have significant binding affinity. Inclusion of urea should allow solubilization of all components of the cleavage mixture and would not be expected to interact with the resin since it is a neutral molecule. However, initial purification attempts at pH 7.0 did not give rise to any protein binding to the column (data not shown). In order to improve binding the purification was also attempted at pH 5.0 since this would increase the positive charge in the proteins. As shown in Figure 3.7C, binding under these conditions did occur, and elution of difference species was achieved through the use of a salt gradient. SDS-PAGE analysis of fractions obtained from this purification showed bands corresponding to p7-His eluting at approximately 300 mM NaCl (Fractions 11 - 14 in Figure 3.7D). However, these fractions also contained significant amounts of GST. Since this is similar to the results obtained with nickel affinity and size exclusion chromatography, this co-elution of GST and p7-His may be occurring due to non-specific interactions that persist in the presence of 8 M urea. Therefore, use of an alternative purification method was required that could use a different denaturing environment that might more effectively disrupt GST – p7 aggregates.

3.5.4 Reversed-Phase High-Performance Liquid Chromatography

Reversed-phase high-performance liquid chromatography (RP-HPLC) is a purification method that has become very popular for the purification of membrane proteins (37, 105, 108, 115-118). This method is based on reversible hydrophobic

interactions with the reversed-phase column that are disrupted by organic solvents which can offer a partially, if not fully, denaturing environment for the proteins. Since an organic solvent may be more effective at disrupting the GST - p7 aggregate, this method was investigated for its ability to produce pure p7-His.

Initial purification attempts were performed on a C4 RP-HPLC column using acetonitrile as the organic solvent in the mobile phase. However, elution profiles obtained under these conditions lacked separation, with all proteins eluting over a single broad peak (Figure 3.8A) that is typical of aggregated proteins (119). SDS-PAGE analysis of fractions taken from this profile showed that both GST and p7-His, as well as other contaminating proteins, were present in all fractions, although the relative concentrations of these species did differ over the profile (Figure 3.8B). Of particular interest are the fractions taken at higher acetonitrile concentrations (approximately >45%) where p7 became a more prominent species as shown by the appearance of a thicker band migrating at approximately 8 kDa.

Although GST was still present in the p7-enriched samples obtained by RP-HPLC, this purification attempt had actually provided the best purification of p7 observed to that point. Therefore, a large number of different conditions were tested in order to improve this purification. Specifically, different flow rates, gradients and loading conditions were tested for their effect on the profile (Table 3.1). These studies suggested that shallower gradients could lead to modest improvements in p7-His purity (data not shown). More importantly, however, were the loading conditions, since lower concentrations of protein should reduce the possibility of non-specific interactions

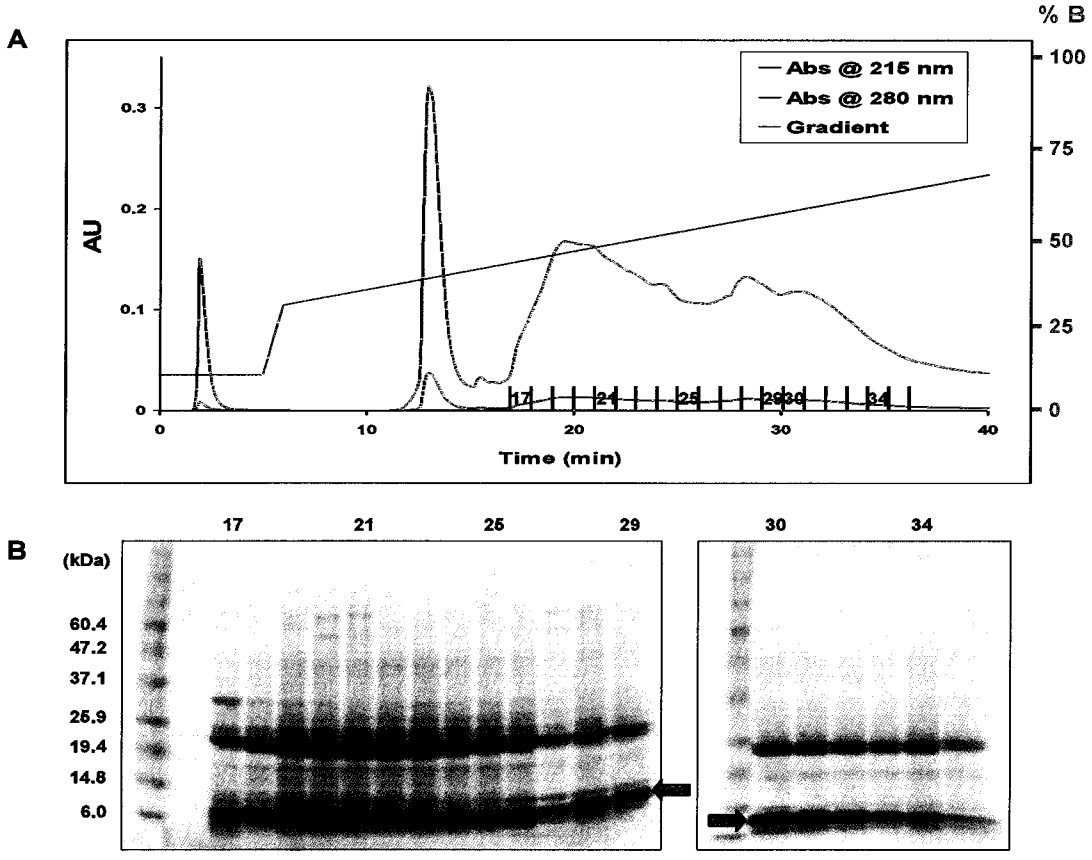


Figure 3.8 RP-HPLC purification using a C4 reversed-phase column. (A) Elution profile for the purification monitored at 215 and 280 nm (blue and green lines, respectively) after application of GST-p7-His onto a Waters C4 column. Purification was performed using conditions listed under Trial # 2 in Table 3.1. The pink line indicates the concentration of the organic solvent used (90% acetonitrile, 0.01% TFA). Fractions indicated on the x-axis of the elution profile were collected and analyzed by SDS-PAGE **(B)**. Pink arrows point towards the protein band assigned as p7.

Table 3.1 Summary of general conditions attempted on C4 RP-HPLC column

Trial #	Injection Conditions	Protein Concentration‡	Method	Length of Run†	Flow Rate
1	10% B for 5 min	1 μg/μl	Gradient: 10-100% B	45 min	1 ml/min
2	10% B for 5 min	1 μg/μl	Gradient: 30-80% B	50 min	1 ml/min
3	10% B for 5 min	1 μg/μl	Gradient: 30-80% B	50 min	0.5 ml/min
4	10% B for 5 min	1 μg/μl	Gradient: 30-80% B	100 min	1 ml/min
5	40% B for 30 min	10 ηg/μl	Gradient: 40-55% B	30 min	1 ml/min
6	40% B for 20 min	10 ηg/μl	Gradient: 40-60% B	20 min	1 ml/min
7	40% B for 20 min	10 ηg/μl	Isocratic: 50% B	10 min	1 ml/min
8	40% B for 15 min	10 ηg/μl	Isocratic: 55% B	10 min	1 ml/min

* All trials performed with solvents A (water) and B (90 % acetonitrile), both containing 0.1% trifluoroacetic acid (TFA)

‡ Approximate protein concentration at injection

† Length of run of method used (excludes injection time)

occurring prior to column binding. The best results were obtained when the cleavage mixture was diluted in a large volume of solvent and loaded under isocratic conditions at an acetonitrile concentration that greatly reduced binding of less hydrophobic proteins to the column. As shown in Figure 3.9A, when the crude mixture was loaded at a concentration of 10 μg/mL in 40% acetonitile the majority of the proteins eluted in a broad peak (~ 5-17 min). The acetonitrile concentration was then increased to 50 % leading to the elution of a second peak (Figure 3.9A, ~28 min). SDS-PAGE analysis revealed that the dominant species in this peak was p7-His, with a purity of approximately 80%, a significant improvement over previous RP-HPLC purifications (Figure 3.9B). Application of a gradient in the elution of this second peak did not improve purity levels, suggesting that this was the limit to the extent of purification that could be obtained with this column.

Although 80 % purity was an excellent result for this purification compared to previous methods, a higher level of purity was required for structure determination purposes. For this reason a less hydrophobic column was tested since unique selectivity can be obtained that could improve the p7-His purification and help prevent non-specific aggregation onto the column (120). For this purpose a C3 column was tested with a standard 2 % per min acetonitrile gradient. As show in Figure 3.10A, a broad, poorly resolved peak did appear in the centre of the purification profile, along with distinct well-resolved peaks throughout the gradient. SDS-PAGE analysis of the last distinct peak which eluted at ~50% acetonitrile showed that it contained p7-His with a purity exceeding 90 % (Figure 3.10B). Western blot analysis with an anti-His antibody confirmed that this species contained a His-tag as would be expected for this fragment

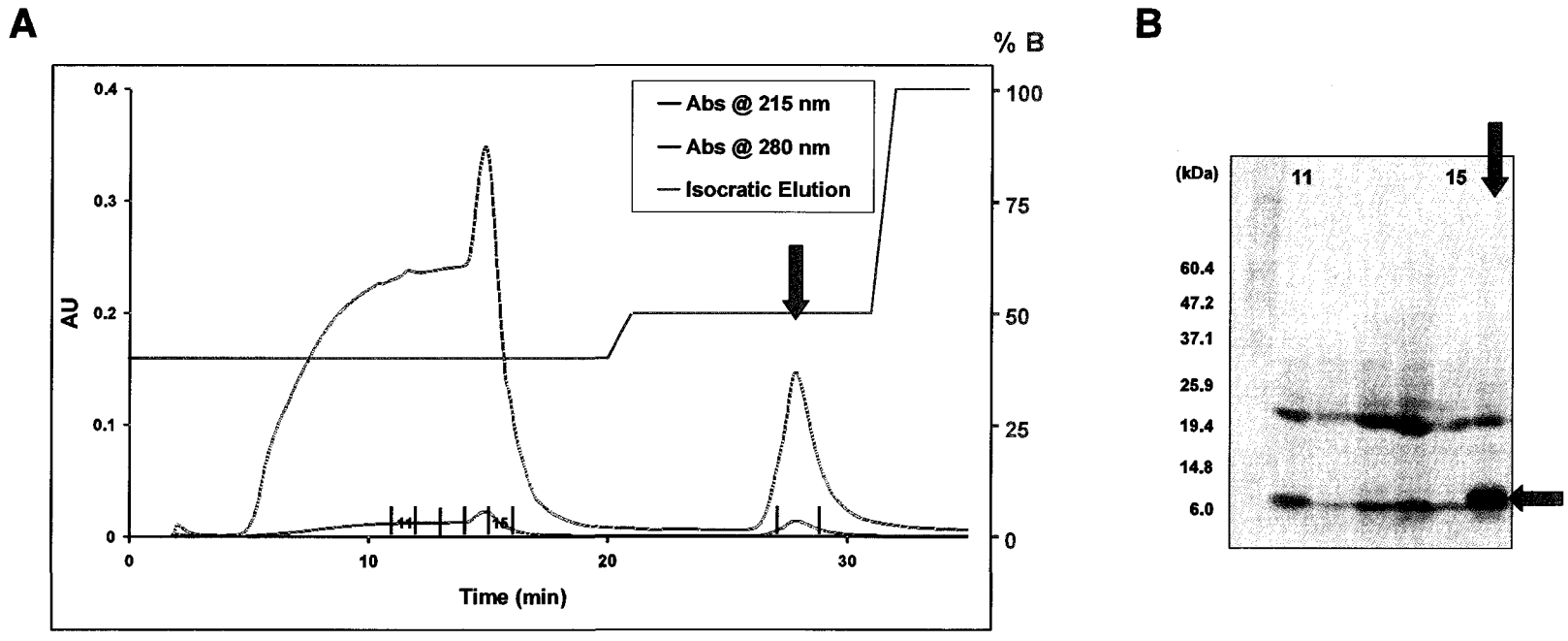


Figure 3.9 Optimized RP-HPLC purification on a C4 reversed-phase column. Figure labels are as described in Figure 3.7. The green arrows correspond to the fraction taken at approximately 28 min. The pink arrow indicates the position of the His-tagged p7 protein on the gel. The purification was performed under Trial # 7 conditions listed in Table 3.1.

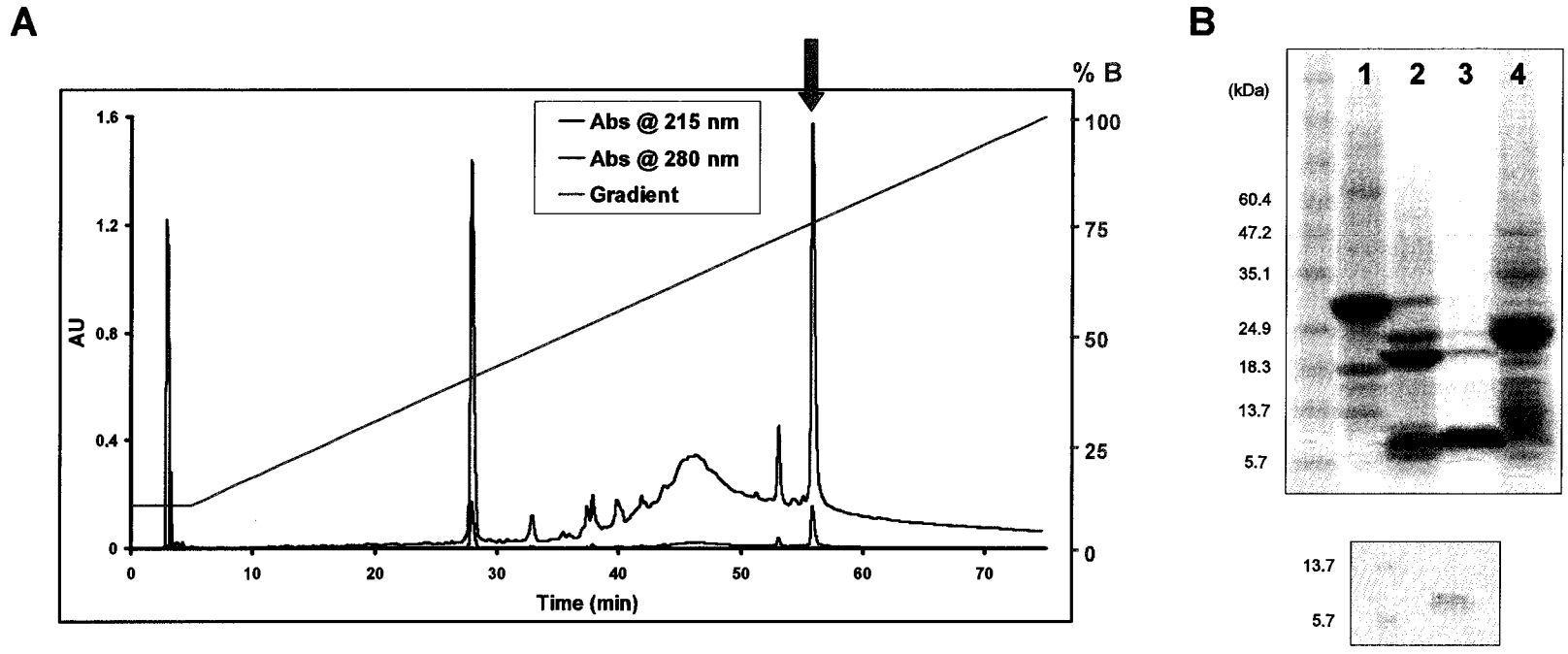


Figure 3.10 RP-HPLC purification on a C3 reversed-phase column. **(A)** Elution profile obtained with a Zorbax 300-SB C3 column using a gradient from 10 % to 100 % organic solvent phase (80 % acetonitrile, 0.1 % TFA). The pink arrow corresponds to the fraction taken at approximately 56 min containing purified p7-His. **(B)** Top - SDS-PAGE of purified p7-His. Lane 1: Uncleaved fusion protein. Lane 2: Cleavage reaction mixture. Lane 3: RP-HPLC purified fraction indicated by the pink arrow in A. Lane 4: GST protein alone. Bottom - Anti-His-tag Western blot performed on the RP-HPLC purified fraction indicated by the pink arrow in A.

(Figure 3.10B). MALDI-MS was performed on this purified fraction and showed that the molecular weight of this species was 8814 Da, which agrees closely with the expected molecular weight of 8803 Da, confirming the identity of this sample as the expected p7-His protein.

Based on the protocols established during the course of these studies, it was now possible to generate approximately 0.5 mg of >95% pure p7-His from 1 L of bacteria cultured in minimally defined media. In order to generate the quantities of protein required to prepare 1 300 μL sample for NMR (~ 0.2 - 0.5 mM) approximately 2 L of media is required. Overall, these yields and purity levels are sufficient to permit solution NMR investigations into the conformation of this protein, which was the next goal of this work.

3.6 Screening Sample Conditions for NMR Spectroscopy of p7-His

In order to study the structure of p7 by solution NMR, it was necessary to find conditions that could maintain this membrane protein in a folded conformation in solution. Detergents have been widely used to simulate the native membrane environment for this purpose (121-123) and, for this reason, were chosen to resolubilize RP-HPLC purified p7-His after lyophillization. Three detergents that have been successfully used for NMR analysis of membrane proteins were used to solubilize p7-His: SDS, DPC and LMPG (73).

When ~ 0.2 mM p7-His was solubilized in 200 mM SDS at pH 6.5, the sample was stable at room temperature for over 30 days. ^{1}H - ^{15}N HSQC spectra of p7-His at 40°C showed approximately 60 out of the expected 74 peaks (Figure 3.11). Also three peaks were observed with proton and nitrogen chemical shifts characteristic of tryptophan

indole amide protons, as would be expected from the primary sequence of this construct. Similarly, it was possible to identify 3 pairs of peaks corresponding to asparagine and/or glutamine side chain peaks as would be expected for this protein. However, in the region of the spectrum that normally shows backbone amide resonances for glycine residues only 4 of the expected 5 were observed. Furthermore, several peaks within the central area of the spectrum were highly overlapped, which would complicate resonance assignment. Although the proton chemical shift dispersion was greater than 1 ppm, providing evidence that p7 is folded in SDS, the missing peaks and poor spectral resolution suggested that this detergent may not be suitable for further structural studies.

Another ^{15}N labeled sample of p7 was prepared in 200 mM DPC, a commonly used detergent with the same alkyl chain as SDS, but with a zwitterionic choline head group. This sample exhibited the best stability out of the three detergents, persisting in solution for several months after its preparation. The HSQC spectrum showed that p7 is also folded in DPC, with a proton chemical shift dispersion characteristic of a primarily α-helical protein (Figure 3.11). However, while the expected number of side chain peaks were present, one glycine backbone peak was also missing from the spectrum. In fact, approximately the same number of peaks was obtained in DPC relative to SDS. Due the absence of a significant number of peaks another detergent that was more similar to physiological lipids was tested. Lysolipids are one type of physiological detergent, with lipid headgroups that are identical to those contained in biological membranes and was therefore the next type of detergent screened for solution NMR.

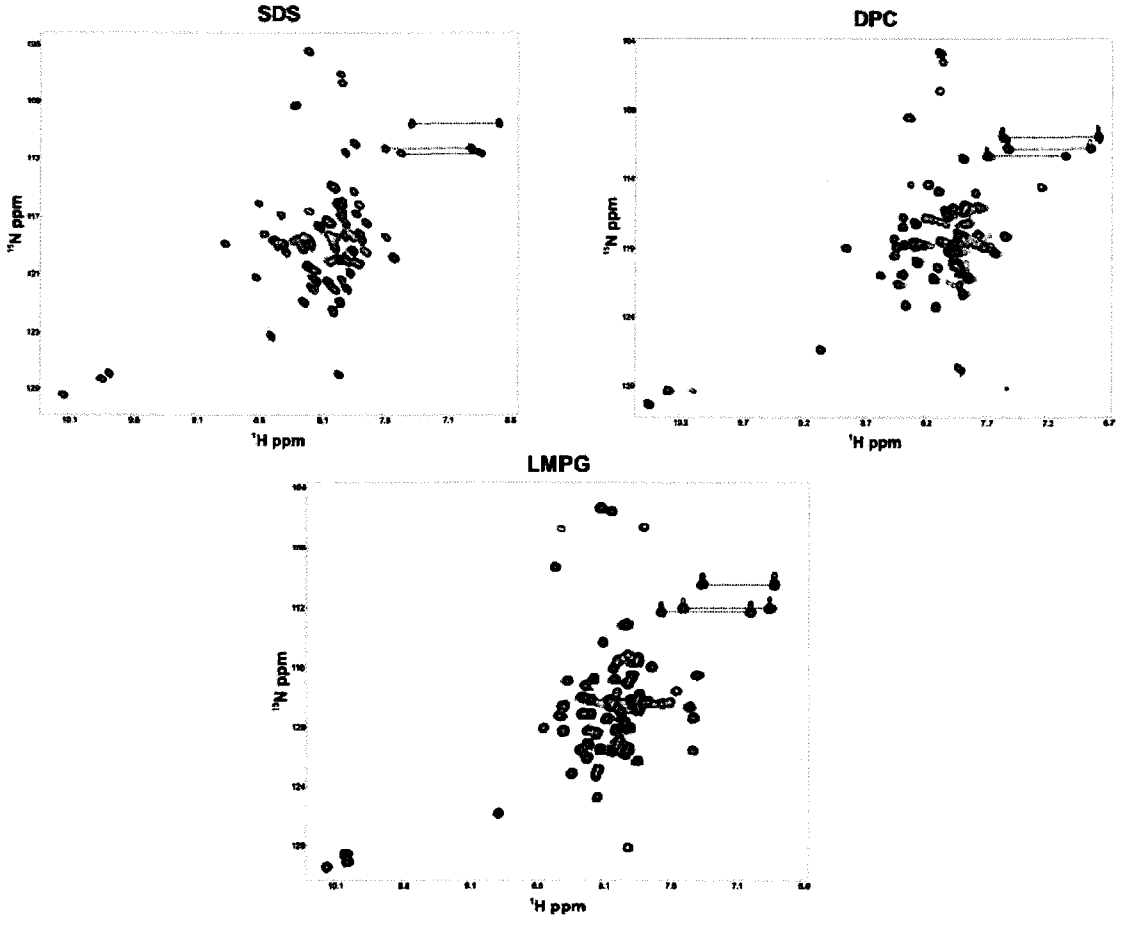


Figure 3.11 ^{1}H – ^{15}N HSQC spectra of p7-His solubilized in SDS, DPC or LMPG. The ~ 0.2 mM p7-His samples were analyzed in 0.2 M detergent, 0.025 M NaCl and 0.025 M sodium phosphate buffer pH 6.5. Spectra were collected with 64 transients (SDS) or 256 (DPC and LMPG) 64 complex points with a total recording time of 3 hours for SDS and 12 hours for DPC and LMPG. Experiments were recorded at 40°C on a Varian INOVA500.

Another ^{15}N-labeled sample of p7-His was made in a 200 mM solution of the anionic lysolipid LMPG and a ^{1}H-^{15}N HSQC spectrum acquired as was done for previous samples. As with the other two detergents, the chemical shift dispersion in the hydrogen dimension indicated that p7 likely adopts a largely helical conformation in LMPG (Figure 3.11). In this sample approximately 69 out of 74 expected peaks were obtained, which presented a small improvement over previous spectra obtained in SDS or DPC. In addition the expected number of Trp and Asn/Gln side-chain correlations, as well as the expected number of glycine backbone peaks were also observed. For these reasons, LMPG was chosen as the solvent for further structural analysis of p7-His.

For each sample of ^{15}N-labeled p7-His, the temperature was also investigated to see if improvements in spectral quality could be obtained. In general, ^{1}H – ^{15}N HSQC spectra tended to improve in sensitivity at higher temperatures as would be expected from the concomitant increase in molecular tumbling rates (73). An example of this is shown in Figure 3.12 which illustrates ^{1}H - ^{15}N HSQC spectra plotted at the same level for p7-His in 200 mM LMPG at three different temperatures. As temperature was increased from 30°C to 40°C one peak disappeared, but several others including two glycine peaks began to appear clearly favoring the higher temperature. When the temperature was increased from 40°C to 50°C, no additional peaks were observed. However, signals within the central area of the spectrum began to broaden and merge (pink star in Figure 3.12). For these reasons, 40°C was chosen as the optimal temperature for the recording of subsequent NMR experiments.

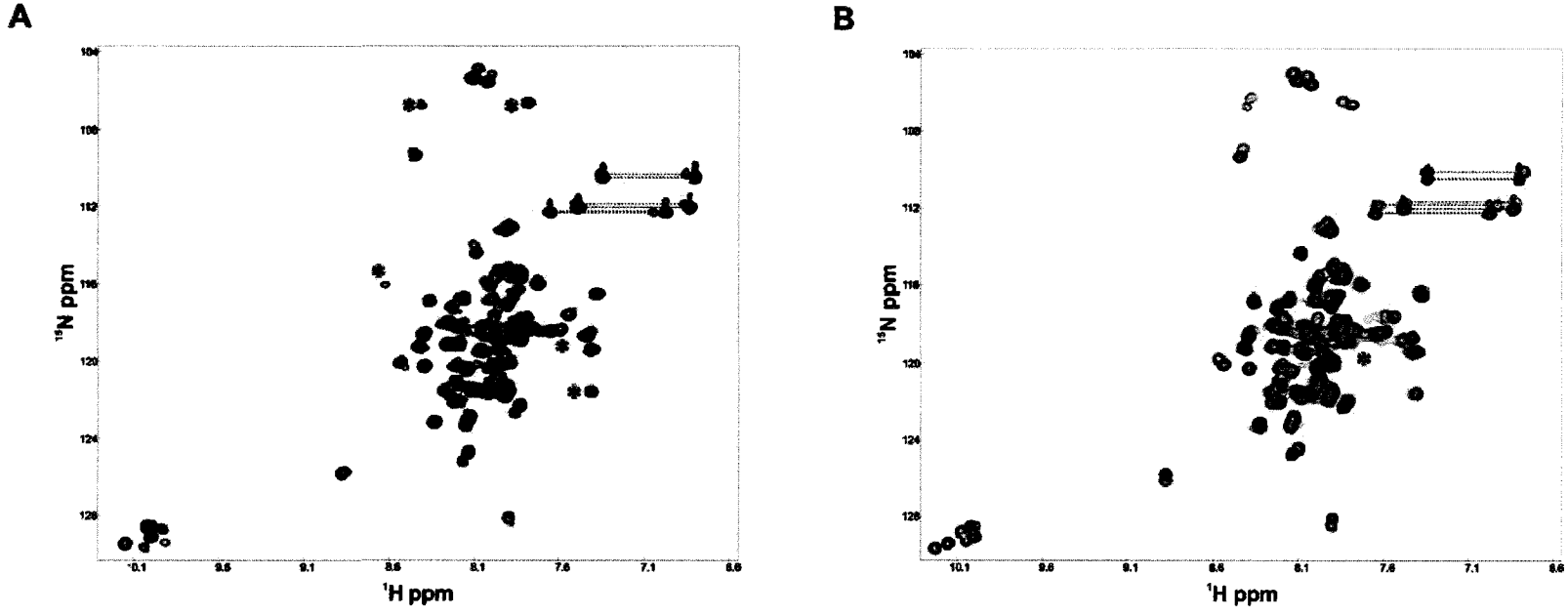


Figure 3.12 ^{1}H – ^{15}N HSQC spectra of LMPG-solubilized p7-His recorded at 30°C (pink), 40°C (black) and 50°C (blue). **(A)** Spectrum at 30°C superimposed over spectrum at 40°C. Blue asterisks indicate peaks that are present at 40°C, but not at 30°C. Green asterisk indicates a peak present at 30°C, but not at 40°C. **(B)** Spectrum at 50°C superimposed over spectrum at 40°C. Pink asterisk indicates peaks that begin to overlap at 50°C, but not at 40°C. The sample was analyzed in ~ 0.2 M LMPG, 0.025 M NaCl and 0.025 M sodium phosphate buffer pH 6.5. Spectra were collected with same parameters as in Fig 3.11 for LMPG.

3.7 Backbone Chemical Shift Assignment

Once sample conditions for p7-His had been optimized, a standard set of triple resonance NMR experiments were recorded to allow chemical shifts to be assigned for backbone ^{15}N, $^{1}H_N$, $^{13}C_\alpha$ and $^{13}C_\beta$ atoms. For this purpose 3D HNCACB (88) and CBCA(CO)NH (124) spectra were recorded on an NMR spectrometer equipped with a cryoprobe (to increase sensitivity) at McGill University in Montreal. As outlined in the introduction, the CBCA(CO)NH provides correlations linking the chemical shifts of the amide proton and nitrogen atoms with shifts from the $^{13}C_\alpha$ and $^{13}C_\beta$ atoms from the preceding residue, while the HNCACB does this for the intra-residue $^{13}C_\alpha$ and $^{13}C_\beta$ atoms as well. Following the strategy previously outlined, strips from these spectra were ordered according to matches in carbon chemical shifts linking inter- and intra-residue correlations to obtain backbone chemical shift assignments (Figures 3.13 to 3.16).

The HSQC spectrum of LMPG-solubilized p7-His showed approximately 69 peaks out of the expected 74. However, only 34 of these peaks provided us with complete $^{13}C_\alpha$ and $^{13}C_\beta$ correlations in both spectra. 21 others showed correlations in one but not the other while the 14 remaining peaks provided no correlations in either the HNCACB or CBCA(CO)NH spectra. Based on the quality of these data it should be possible to obtain assignment for approximately 34 p7 residues. In fact, tentative assignments were made for 33 p7 residues as well as 7 N- or C-terminal non-native residues that were included in the cloning of this construct (see Figure 3.1).

Assigned regions of p7 included a large portion from the N-terminus, including the beginning of the first putative TMD, and the C-terminal half of the second predicted

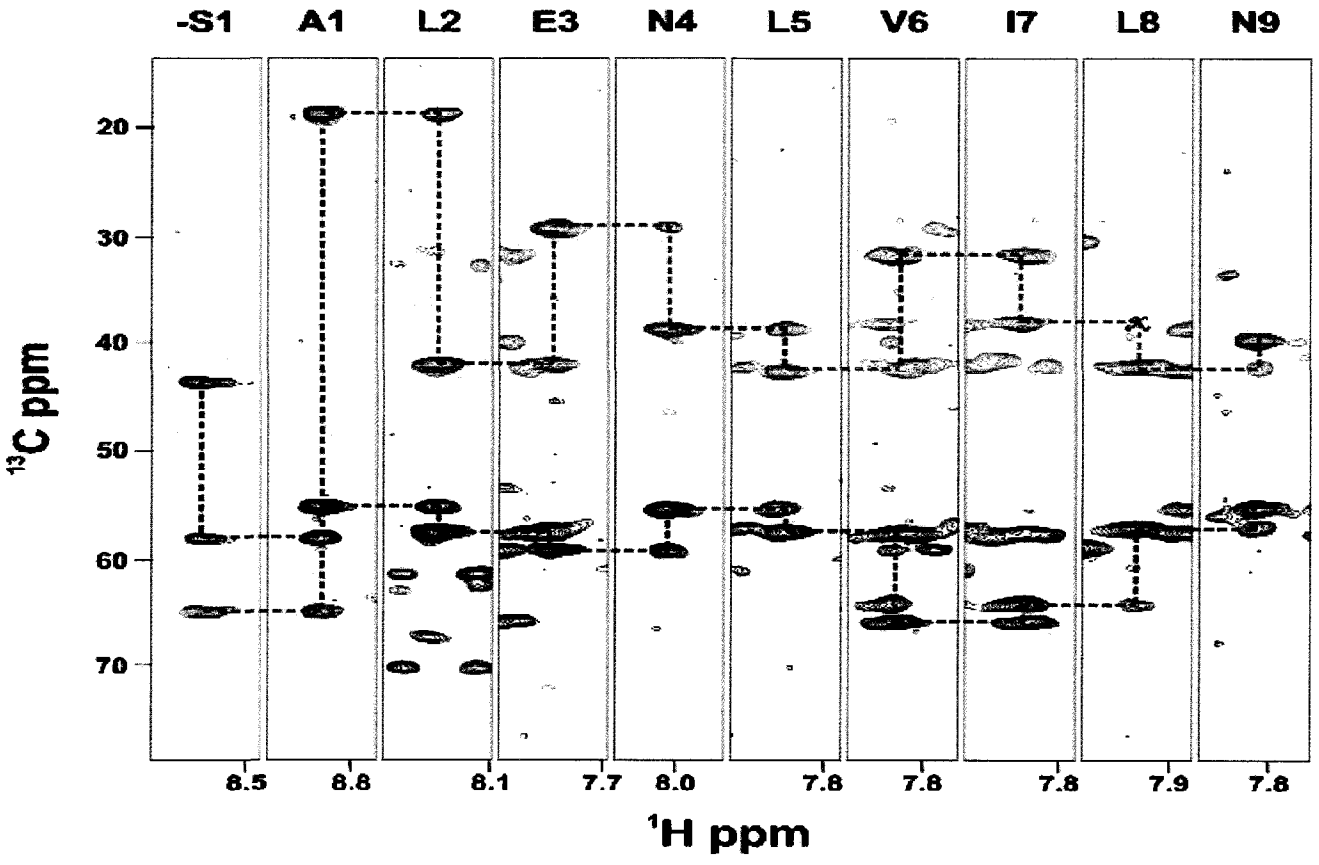


Figure 3.13 Assigned HNCACB strip plot for the N-terminus of p7-His. The HNCACB $^{1}H - ^{13}C$ strips are shown for N-terminal amide protons -1 to 9. Positive intensities are indicated by black contours and negative intensities by red contours. Red dotted lines outline the sequential connectivities between the inter- and intra-residue $^{13}C_{\beta}$ correlations while the black dotted lines show $^{13}C_{\alpha}$ connectivities. One peak position resulting from an inter-residue correlation was absent in the HNCACB, but observed in the CBCA(CO)NH spectrum and was marked in the strip with an X.

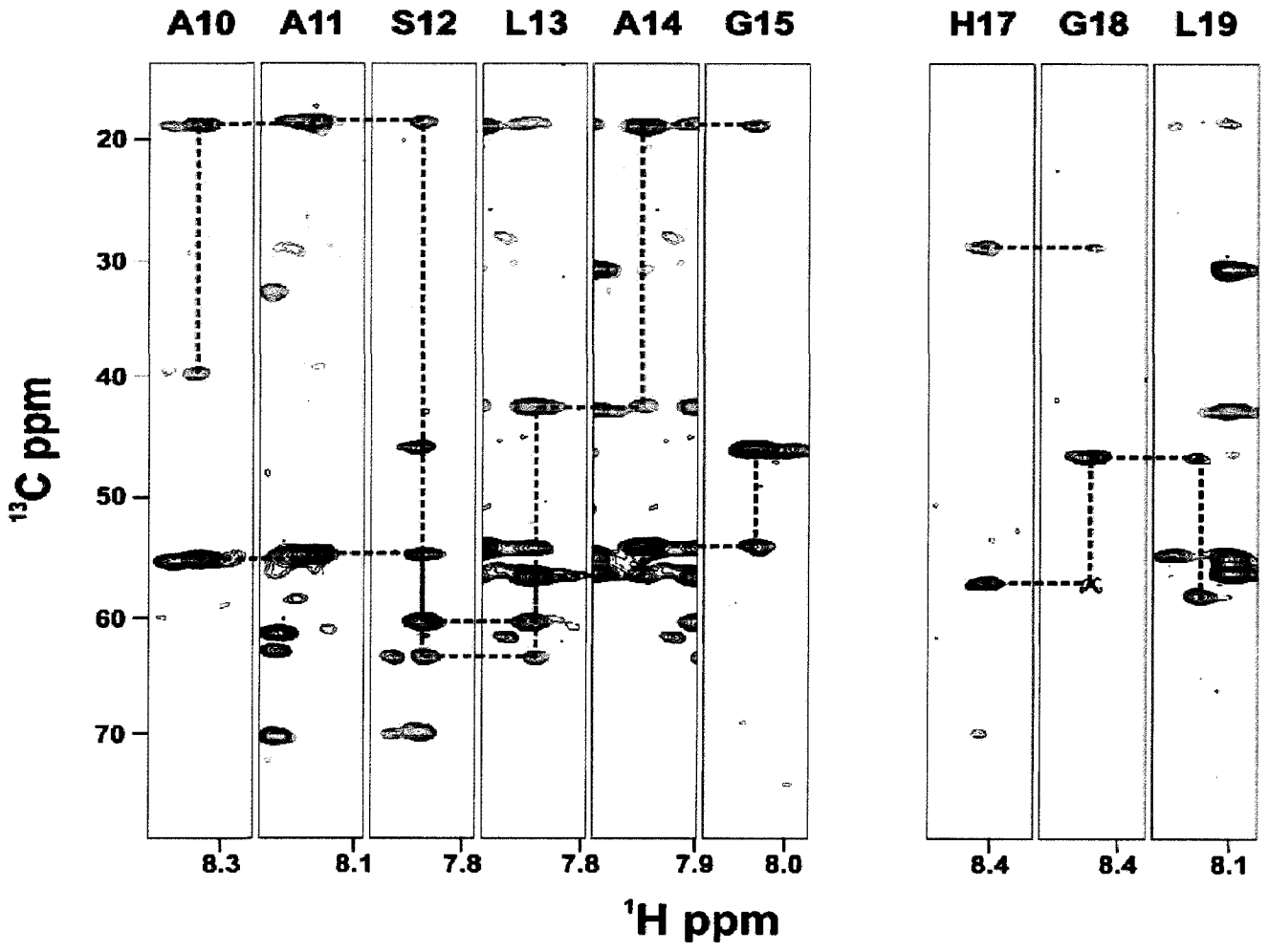


Figure 3.14 Assigned HNCACB strip plot for the N-terminal/first putative TMD regions of p7-His.

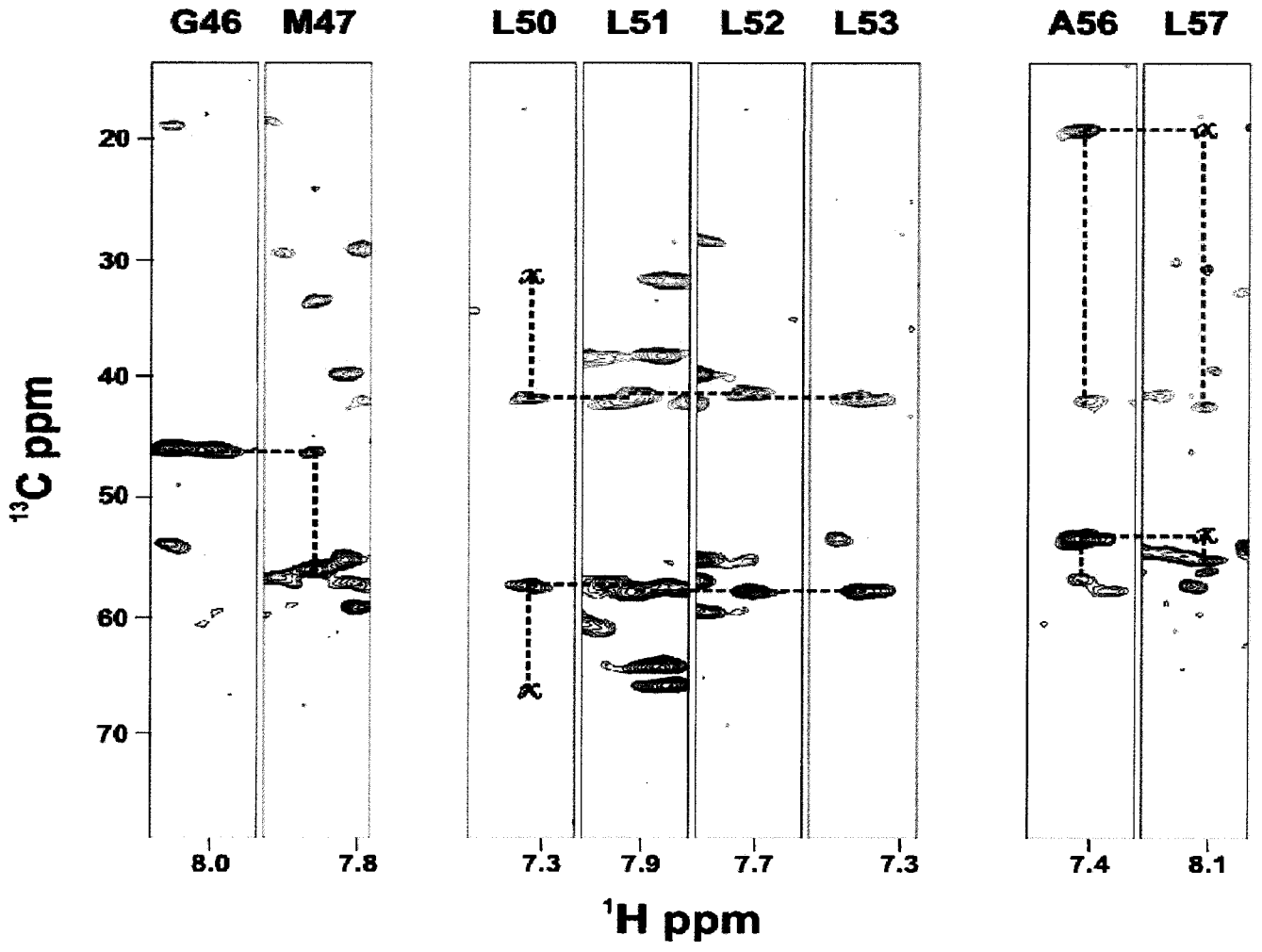


Figure 3.15 Assigned HNCACB strip plot for the second TMD region of p7-His.

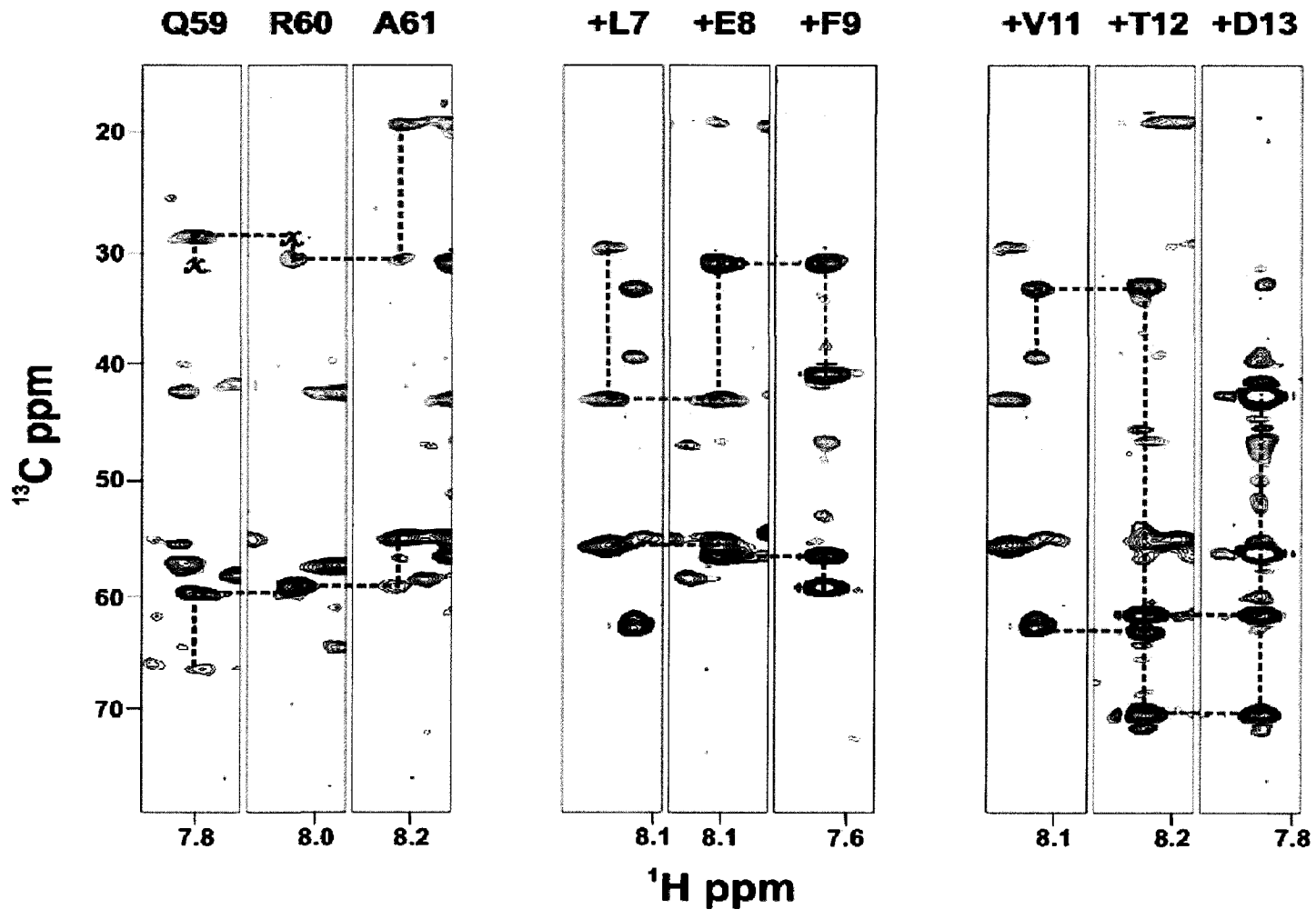


Figure 3.16 Assigned HNCACB strip plot for the C-terminus of p7-His.

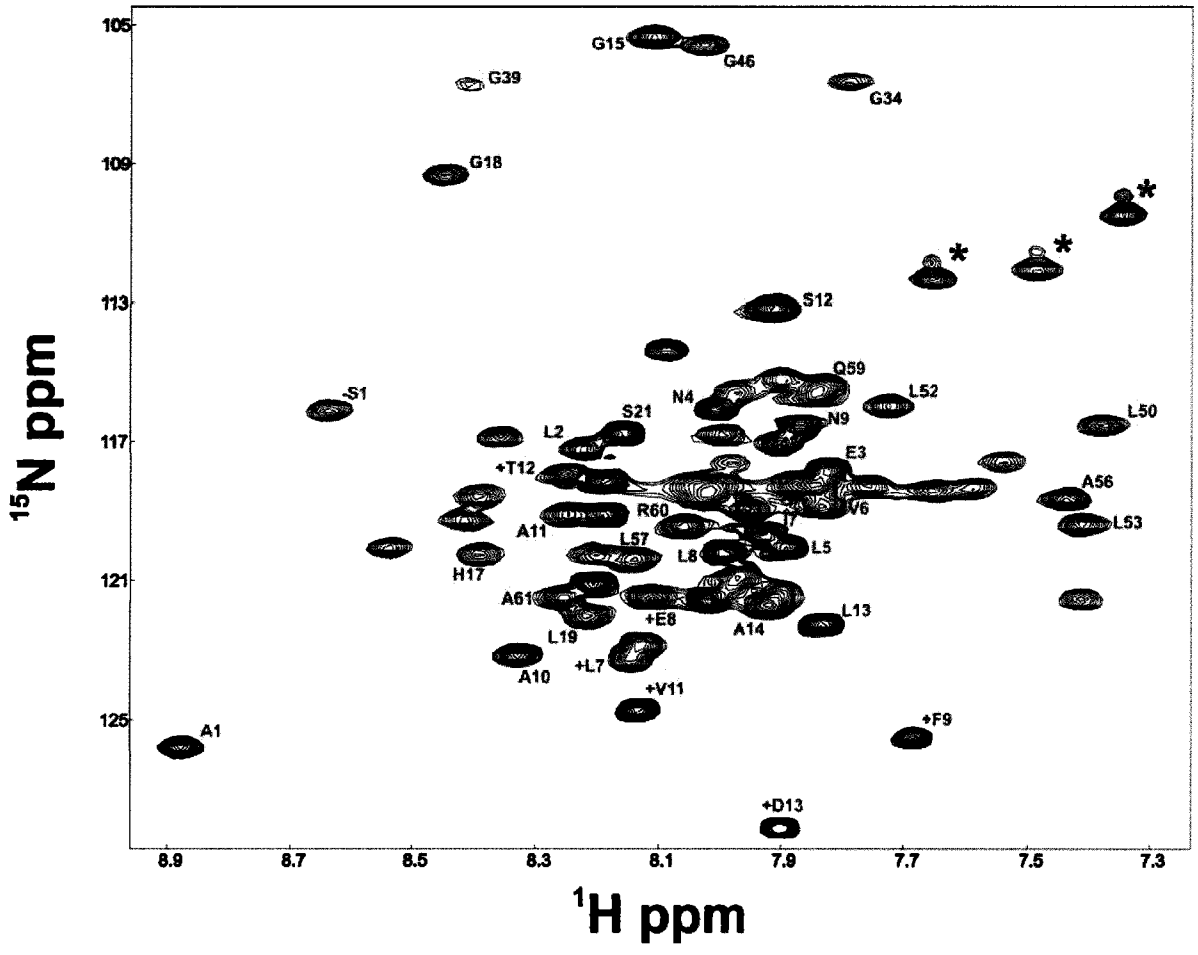


Figure 3.17 Assigned ^{1}H – ^{15}N HSQC spectrum of p7-His in 200 mM LMPG. Red asterisks denote one of the resonances arising from Asn/Gln side chain amides. Residues are numbered according to the p7 protein amino acid sequence.

TMD along with a few flanking periplasmic residues. As summarized by the labeled HSQC spectrum in Figure 3.17, approximately 52 % of the p7 sequence was assigned from these experiments. Although more experiments will need to be run in the future to try to achieve a higher degree of assignment completion, using these chemical shifts it is possible to obtain early insight into the secondary structure of assigned regions.

3.8 Secondary Structure Prediction Based on Secondary Shifts from Random Coil Values

Using the available backbone chemical shifts for p7-His, the difference between the experimental and the random coil values (91) was calculated for the $^{13}C_{\alpha}$ atoms to characterize the secondary structure of these residues. In this analysis a positive secondary chemical shift indicates an α-helical conformation while a negative value points toward a β-sheet structure. As seen in Figure 3.18, positive values were observed for most of p7 while N- and C-terminal linker residues were both positive and negative in intensity. This is not surprising as the two putative TMDs are predicted to be α-helical while the terminal non-native residues would not be expected to display significant structure. However, it is interesting to note the presence of an α-helical conformation in the N-terminus leading to the possibility of a third helix in the secondary structure of p7.

3.9 Confirmation of Assignments and Structure for Helical Regions of p7-His with the ^{15}N-edited NOESY

In order to confirm the presence of α-helical structure in assigned regions of p7, a 3D ^{15}N-edited NOESY was also acquired on this sample in LMPG on the cryoprobe-equipped 500 MHz spectrometer at McGill University. Strips from assigned residues were then analyzed for the presence of strong NOE cross-peaks between amide protons of

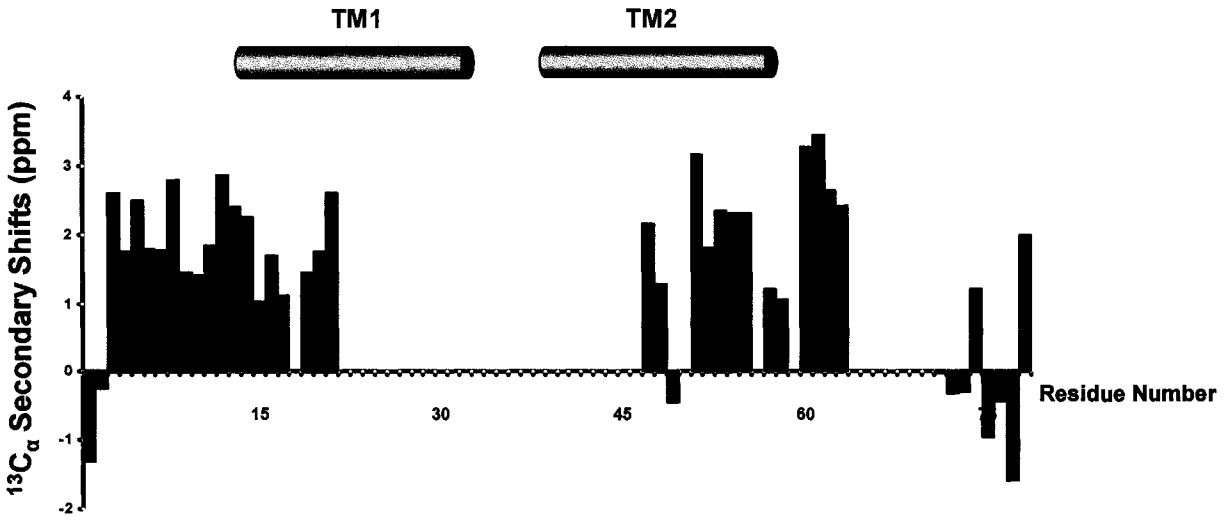


Figure 3.18 Secondary structure of p7-His determined from C_α chemical shifts. Shown are the differences between experimental C_α chemical shifts and published random coil values for p7-His. Unassigned residues have been given a secondary shift value of zero. The putative TM helices determined from TopPredII and TMHMM are drawn schematically at the top. The positive secondary shifts suggest that the assigned portions of p7-His are mainly α-helical in structure.

sequential residues as would be expected for an α-helical secondary structure. As seen in Figure 3.19, strong NOE cross-peaks were observed between sequential residues for the first nine residues in the p7 sequence, further supporting the idea of an N-terminal helix. Sequential amide NOEs were also observed from residues 10 – 14 indicating that the first TMD may begin around residue 10 (Figure 3.19).

Similar to the CBCA(CO)NH and the HNCACB spectra obtained on this sample the NOESY data showed poor sensitivity for residues within the central region of the amino acid sequence of p7-His. However, the NOESY spectrum could verify the presence of an α-helix between residues 50 to 53 and 59 to 61, both of which lie within the second putative TMD (Figure 3.20). Furthermore, the NOESY data confirmed that the C-terminal non-native residues were not α-helical in structure since strong NOEs were not observed for these amino acids (Figure 3.20).

3.10 Results Summary

The objectives of this research project were to develop a biosynthetic protocol for the production of milligram quantities of the HCV p7 protein and begin preliminary NMR structural studies on p7. During the course of this work the following results were obtained:

- Expression and purification protocols were developed for His-tagged GST fusion proteins to p7.
- Protocols were established to isolate p7-His from GST-p7-His using thrombin cleavage followed by RP-HPLC.

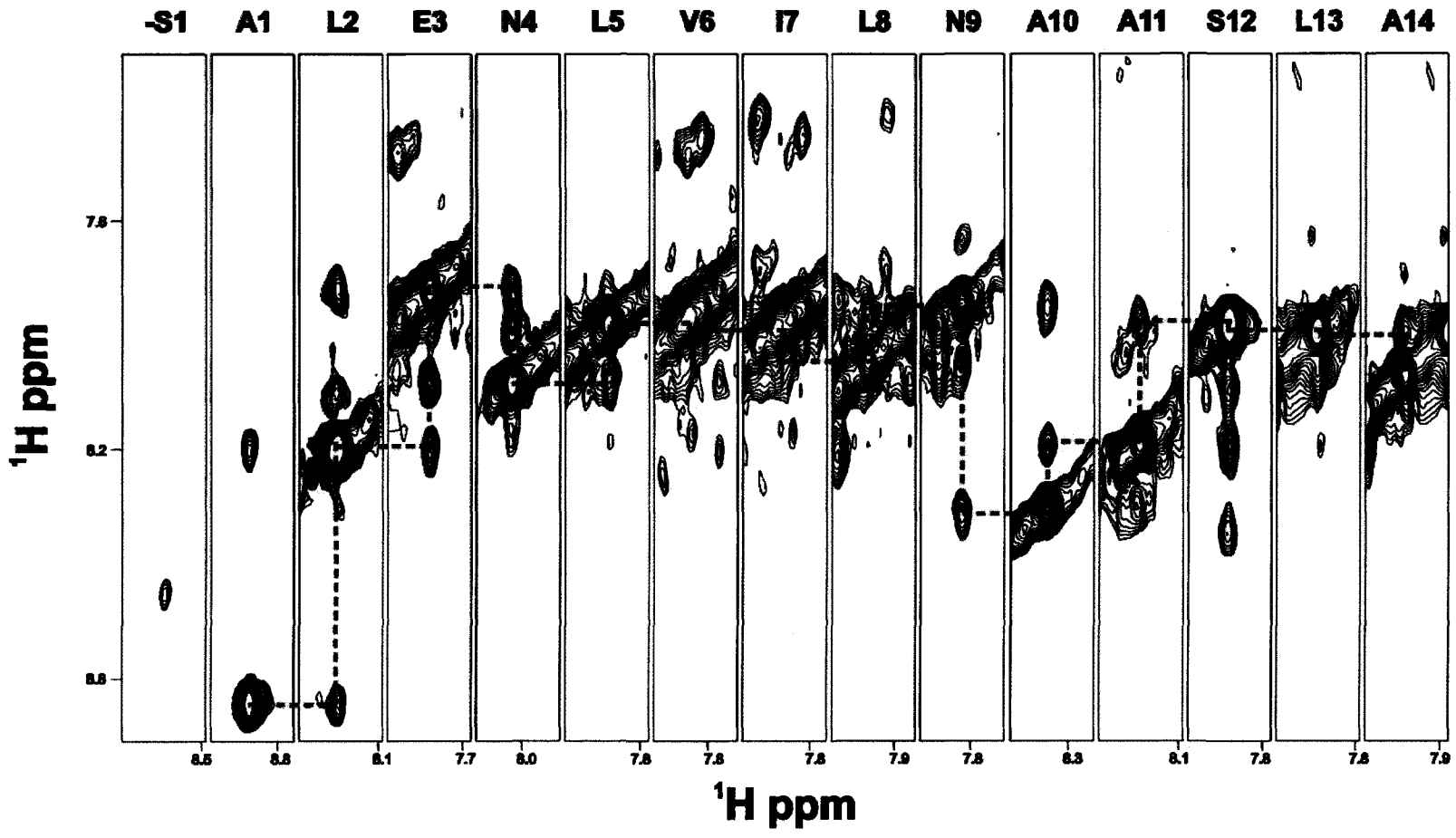


Figure 3.19 ^{1}H – ^{1}H strips from the ^{15}N-edited NOESY for the N-terminus of p7-His. Independent confirmation of backbone assignments could be obtained by strong sequential $^1H^N$ - $^1H^N$ NOEs for the helical N-terminus of p7-His.

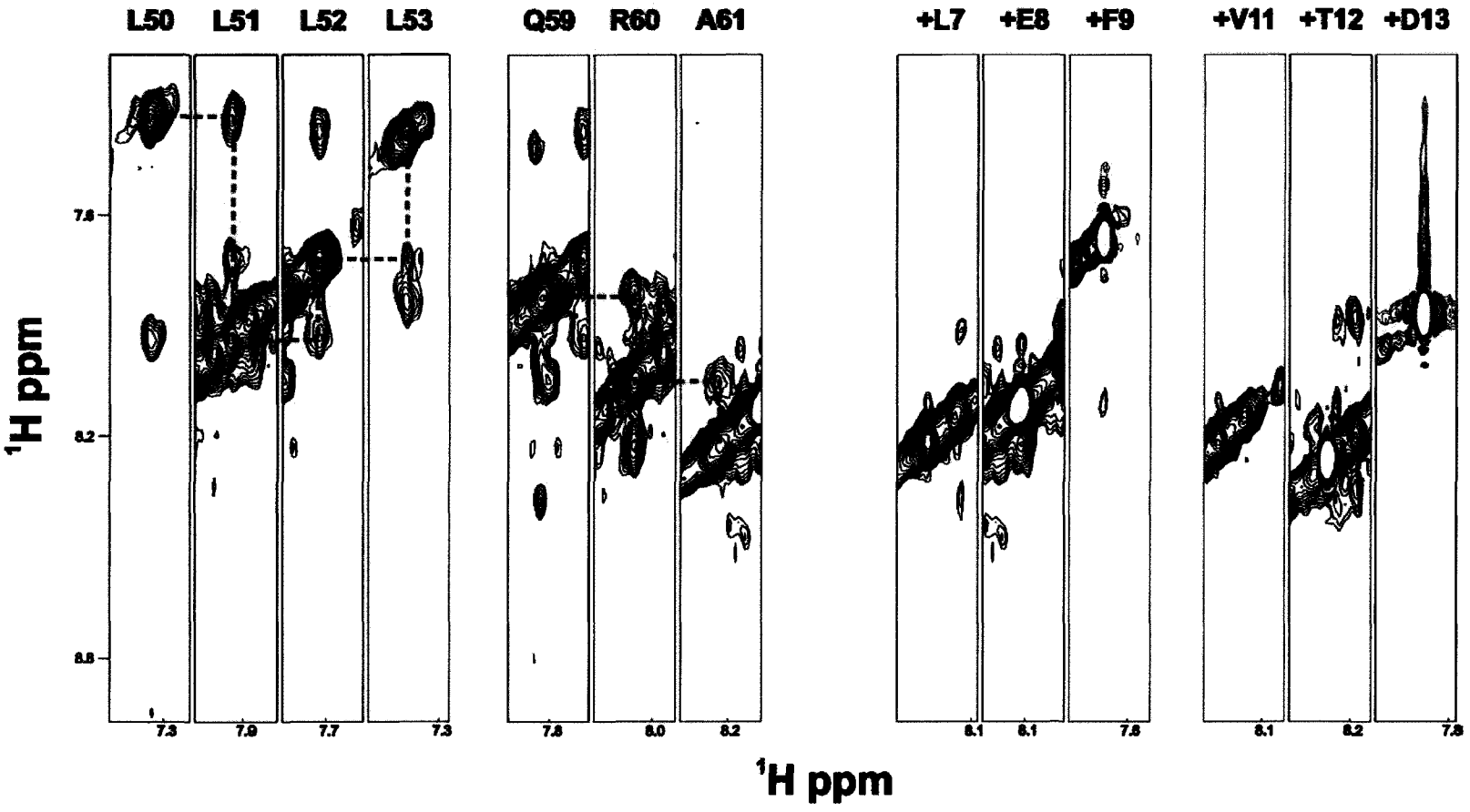


Figure 3.20 ^{1}H – ^{1}H strips from the ^{15}N-edited NOESY for the C-terminus of p7-His. Independent confirmation of backbone assignments could be obtained by strong sequential $^1H^N$ - $^1H^N$ NOEs for the helical C-terminus of p7-His. Confirmation that the C-terminal non-native residues are not helical in structure is shown by the lack of sequential NOEs.

- ^{1}H-^{15}N HSQC spectra of reasonable quality could be obtained for p7-His in LMPG at 40°C and proved slightly superior to spectra obtained in SDS or DPC.
- Triple resonance NMR experiments were used to determine chemical shift assignments for backbone atoms from 33 p7 residues.
- Secondary structure analysis showed that p7-His is, as expected, mainly α-helical in structure, but also revealed the presence of an N-terminal α-helix.

Chapter 4 DISCUSSION

4.1 Factors Affecting NMR Spectral Quality

Several factors can affect spectral quality, of which the type of detergent is one of the most important. The two detergents that are most popular for solution NMR studies of membrane proteins are SDS and DPC, which were also used in the current study (74). Their 12 carbon-long alkyl chains are relatively small which minimizes the size of the micelle, making them more suitable for solution NMR studies. Furthermore, the alkyl chain is almost long enough to cover the hydrophobic area of a membrane protein much like a phospholipid in a bilayer (78). However, SDS is known as one of the harshest detergents since it is anionic and has a small head group. It is a relatively well-known denaturant used in several applications such as SDS-PAGE. In spite of its reputation as a denaturant, it has been successfully used for NMR studies of oliogomeric membrane proteins such as the dimeric form of glycophorin A (23) or the tetrameric KcsA potassium channel (125). On the other hand, DPC has a zwitterionic headgroup that more closely resembles headgroups found in phospholipids and is therefore considered to be less denaturing. Like SDS, it has been used in NMR structure determinations of a number of membrane proteins, including the β-barrels ompA (104) and PagP (63) as well as the monomeric (126) and pentameric (105) forms of the 1 TM-helix protein phospholamban.

In the case of p7-His, both SDS and DPC-solubilized samples produced NMR spectra with a similar degree of proton chemical shift dispersion and number of observed

resonances. The absence of peaks in these spectra may be indicative of poorly resolved signals or flexibility in some regions of the protein that give rise to dynamics on a nanosecond-to-millisecond timescale. In the case of SDS, the small anionic headgroup may have led to unfavorable interactions that destabilized parts of the protein, leading to these spectroscopically unfavorable dynamic properties. Similar types of destabilizing interactions may have also been present between DPC and p7-His, although at this stage it is not possible to definitively understand the reason behind the similar quality spectra obtained.

Based on the appearance of the HSQC spectra, the lysophosphatidylglycerol LMPG appeared to be most suitable for solution NMR studies of p7-His of the 3 detergents examined. While it is not possible to know the origins of this spectral improvement, favorable interactions between the native LMPG headgroup and glycerol backbone may have occurred in a way that stabilized the protein. Furthermore, the alkyl chain of this detergent more closely mimics the phospholipid alkyl chain in length. It has been suggested that the lysophosphatidylglycerols improve spectral quality by decreasing conformational heterogeneity and conformational exchange (73) which may have also been the case in the current study. Overall, the results with p7-His reinforce the importance of screening different detergent solvents for suitability in solution NMR.

Based on the trend observed with p7-His, it is probable that other detergents in the group of the lysophosphatidylglycerols could also provide good quality spectra of this protein. For example, LPPG (1-palmitoyl-2-hydroxy-sn-glycero-3-[phospho-RAC-(1-glycerol)]) has a slightly longer alkyl chain than LMPG (16 carbon atoms instead of 14), potentially providing a closer mimic of the length of the native membrane phospholipid

alkyl chain. In fact, LPPG has been suggested to be one of the best detergents for solution NMR studies of several membrane proteins (73). On the other hand, the three-dimensional structure of other membrane proteins has been successfully determined in detergents like DHPC for OmpX (103) and *n*-octyl-β-D-glucoside for PagP (63) and therefore, it might be useful to investigate their utility for solution NMR of p7-His as well.

Another aspect of sample conditions that was not explored in the current study is the effect of detergent concentration on spectral quality (62). This aspect is particularly important for proteins which undergo oligomerization via interactions between their TM helices since the ratio of monomeric vs. oligomeric protein depends on the amount of detergent that is present relative to the amount of protein (127). In general, higher concentrations of detergents promote the monomeric form, while lower concentrations can shift the equilibrium towards the oligomeric form. Although the molecular weight of the detergent-protein complex in these NMR samples was not determined, it is likely that these NMR spectra correspond to the monomeric form of p7 since a relatively high concentration of detergent was used (protein:micellar detergent ratio of 1:1000). In addition, the predicted molecular weight of a heptameric oligomer associated with a small (~30 kDa) micelle is on the order of ~90 kDa, a size which would be too large for the detection of signals in the NMR spectrum. Therefore, it is likely that the sample that was studied in the current work is primarily monomeric. However, the potential for p7 to oligomerize in detergents may account for the lower signal-to-noise obtained for parts of the TM helices since these could be involved in self-interaction.

4.2 Issues in Assigning p7-His Spectra

Through several topology prediction methods, p7 was predicted to contain two TM helices (1, 128) although experimental confirmation of this structure has not been available thus far. In the current work, preliminary insight into the p7 backbone conformation in LMPG through the partial backbone chemical shift assignment of p7-His was obtained. Unfortunately, it was not possible to obtain a complete assignment owing to relatively poor sensitivity exhibited in the NMR spectra. As a result, several residues did not give rise to any correlations in the triple-resonance experiments making it impossible to assign their chemical shifts. Furthermore, there was extensive peak overlap for some regions of the spectra leading to ambiguities in the assignment. In order to help overcome these problems in the future, performing other triple-resonance NMR experiments that are more sensitive than the ones run in my work may help obtain more complete backbone assignment. For example, the HNCA and the HN(CO)CA are the most sensitive experiments that can be used for backbone chemical shift assignment (129), and may therefore allow some correlations to be observed where in the other experiments there were none. Furthermore, since the HNCA and HN(CO)CA provide complementary information, this should help to decrease the ambiguities in the assignment process. Conventional ways to improve signal-to-noise should also be explored, such as trying to increase the amount of protein in the sample or running these experiments on a higher field spectrometer since this may allow the detection of lower intensity resonances. Other ways to enhance resolution and sensitivity are to partially deuterate the sample and/or to use transverse-relaxation optimized spectroscopy (TROSY) both of which significantly reduce transverse relaxation (72, 130).

4.3 p7 in LMPG is Mostly α-Helical

From the backbone $^{13}C_{\alpha}$ secondary shift calculation performed using the available chemical shift data, along with the assignment of sequential amide NOEs in the NOESY, the presence of helical structure was confirmed, as would be expected for transmembrane segments. However, an unexpected finding was that the N-terminus of p7-His also appears to be α-helical in structure, even though it is not predicted to be associated with the membrane. Nonetheless these results are in agreement with secondary structure predictions which suggest the presence of an N-terminal helix (Figure 4.1A). Moreover, this finding also agrees with previous suggestions that this region possesses stable secondary structure due to the conserved pattern of occurrence of hydrophilic and hydrophobic residues (Figure 1.2) (1). This pattern could give rise to a partially amphipathic helical structure (Figure 4.1B) which was proposed to interact with the surface of the lipid bilayer (1). However, the helical wheel projection for the first 18 residues in the N-terminus of p7 does not necessarily show a clear hydrophobic face, raising the possibility that this helix does not interact with the membrane. This possibility is supported by the quality of the NMR signals obtained for this region of the protein since they were generally more intense than peaks from the TM segments. The difference in spectral quality could arise if the N-terminus was not integrated into the micelle, leading to additional mobility that would improve the tumbling characteristics of this part of the polypeptide. Further experiments that look at solvent protection can be done in the future to help identify the parts of the helix that are not embedded in the micelle. In addition, it will also be interesting to determine whether this helix is present when p7 is in different types of membrane-mimetic environments (e.g. different

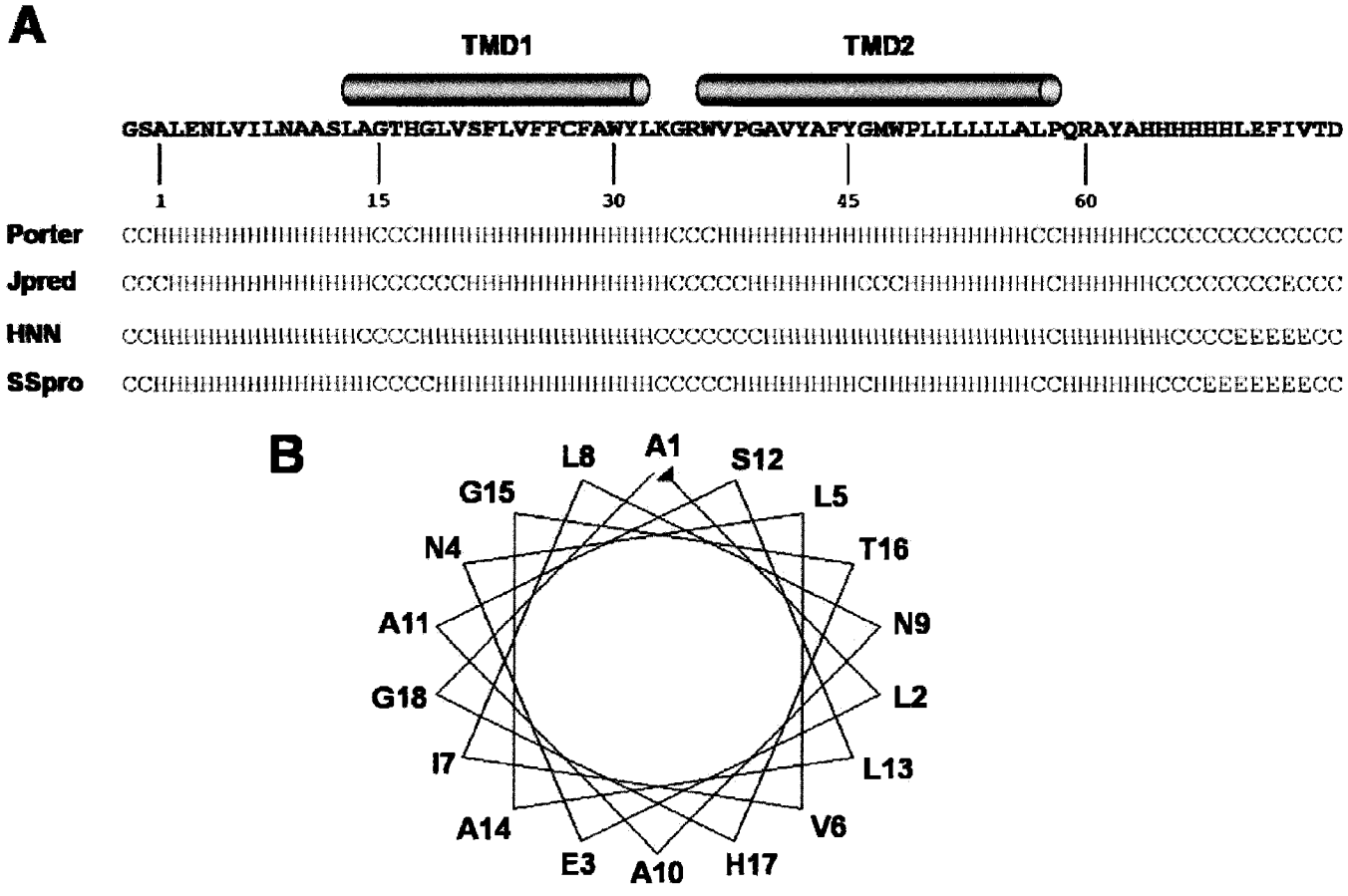


Figure 4.1 (A) Secondary structure prediction of the p7-His protein. Predictions performed by the methods shown on the left available from the ExPASy server online at http://expasy.org/tools/. C: random coil; H: helix; E: beta sheet. The predicted TM helices are shown above. **(B) N-terminal residues 1 to 18 projected along a helical wheel.** Blue residues are polar while black ones are hydrophobic.

detergents, bicelles, liposomes). Overall, this structure could be functionally significant since this region was implicated in genotype-specific functions (32). This N-terminal helix may therefore be important for interactions with other proteins in the HCV genome, or even interactions with itself.

4.4 p7 as an Oligomeric Ion Channel

In order to form a functional channel p7 is known to form oligomers in the lipid bilayer. Chemical cross-linking as well as TEM studies have suggested that this oligomeric state is heptameric, with the first helix lining the channel pore (37). While the role of the second TM is not yet known, it may be important for interactions with the first TM helix either within, or between, p7 subunits. Results from computer modeling methods have also led to the proposal that the second TM helix interacts with the membrane phospholipids and does not play a direct role in channel formation (51). Further structural studies will be required to distinguish between these possibilities.

Ultimately the Goto lab is interested in determining the structure of the p7 channel, particularly for the development of new anti-HCV therapies in the future. In order to do so, sample conditions will have to be investigated in greater detail in order to identify those that favor the oligomeric form of the protein. One of the easiest ways to achieve this is to decrease the detergent concentration. Since p7 can only exist in the detergent phase, decreasing the concentration of detergent increases the effective concentration of p7 (127). The detergent concentration used in our study was high with a protein: detergent ratio of approximately 1:1000. Based on an aggregation number of 55 for LMPG (Table 1.1), there would have been approximately 20 micelles available for every molecule of p7-His. However, if the detergent concentration is decreased, several

protein molecules could be forced into a single micelle thus increasing the chances of oligomer formation. As this occurs, some peaks within the spectra may shift and/or broaden due to participation in oligomer formation. Meanwhile, circular dichroism (CD) spectra of these samples should help determine whether the changes observed in the spectra are due to a change in secondary structure content or arise from the formation of the oligomeric species. In these experiments it will be necessary to distinguish between oligomerization due to specific interactions, versus a non-specific aggregation arising from the forced encapsulation of several protein molecules within a single micelle. This can be done by measuring the standard energy of interaction which should be larger in magnitude than values associated with non-specific interactions, as has been described for other oligomeric helical membrane proteins (127, 131, 132).

Once sample conditions are found that favor the oligomeric state of the protein, it will be necessary to address problems arising from NMR analysis of a large protein-detergent complex. For example, uniform deuteration of the protein of interest, combined with TROSY-based NMR experiments can allow the study of proteins approaching 100 kDa in size. Fortunately, it is likely that p7 forms a symmetrical oligomer, making it unlikely that spectral complexity (*i.e.* the number of peaks in the spectrum) will increase for the oligomeric form of p7.

4.5 Future Prospects for Structure Determination of p7

This thesis work has led to the development of a protocol for the production of a pure p7 sample that can be used to for structural analysis by solution NMR. Preliminary NMR data was acquired with this sample that has allowed partial backbone chemical shift assignments to be made for p7-His. With the use of other NMR experiments, a complete

backbone assignment of the p7-His protein should be possible. This can then be followed by side chain chemical shift assignment followed by assignment of NOEs for future structure calculations. This work is currently being undertaken by other members of the Goto lab in order to determine the tertiary structure of the p7-His monomer. Although we are ultimately interested in understanding the structure of the oligomeric channel formed by p7, a structure of the monomeric form will help to identify α-helical segments and to determine the relative positioning of helices. These studies also have the potential to help identify the parts of the protein that are interacting with the detergent. Ultimately this work should help to reveal the molecular mechanism behind p7 channel activity, potentially facilitating future development of lead compounds for the treatment of HCV infections.

References

1. Carrere-Kremer, S., C. Montpellier-Pala, L. Cocquerel, C. Wychowski, F. Penin, and J. Dubuisson. 2002. Subcellular localization and topology of the p7 polypeptide of hepatitis C virus. *J Virol* 76:3720-3730.
2. Simmonds, P. 2004. Genetic diversity and evolution of hepatitis C virus--15 years on. *J Gen Virol* 85:3173-3188.
3. Shepard, C.W., L. Finelli, and M.J. Alter. 2005. Global epidemiology of hepatitis C virus infection. *Lancet Infect Dis* 5:558-567.
4. Penin, F., J. Dubuisson, F.A. Rey, D. Moradpour, and J.M. Pawlotsky. 2004. Structural biology of hepatitis C virus. *Hepatology* 39:5-19.
5. Galossi, A., R. Guarisco, L. Bellis, and C. Puoti. 2007. Extrahepatic manifestations of chronic HCV infection. *J Gastrointestin Liver Dis* 16:65-73.
6. Cocquerel, L., C. Voisset, and J. Dubuisson. 2006. Hepatitis C virus entry: potential receptors and their biological functions. *J Gen Virol* 87:1075-1084.
7. Stauber, R.E., and V. Stadlbauer. 2006. Novel approaches for therapy of chronic hepatitis C. *J Clin Virol* 36:87-94.
8. Feld, J.J., and J.H. Hoofnagle. 2005. Mechanism of action of interferon and ribavirin in treatment of hepatitis C. *Nature* 436:967-972.
9. Huang, Y.P., S.L. Zhang, J. Cheng, L. Wang, J. Guo, Y. Liu, Y. Yang, L.Y. Zhang, G.Q. Bai, X.S. Gao, D. Ji, S.M. Lin, Y.W. Zhong, and Q. Shao. 2005. Screening of genes of proteins interacting with p7 protein of hepatitis C virus from human liver cDNA library by yeast two-hybrid system. *World J Gastroenterol* 11:4709-4714.
10. Hayashi, N., and T. Takehara. 2006. Antiviral therapy for chronic hepatitis C: past, present, and future. *J Gastroenterol* 41:17-27.
11. Horsmans, Y. 2005. Chronic hepatitis C, depression and interferon. *J Hepatol* 42:788-789.
12. Russmann, S., I. Grattagliano, P. Portincasa, V.O. Palmieri, and G. Palasciano. 2006. Ribavirin-induced anemia: mechanisms, risk factors and related targets for future research. *Curr Med Chem* 13:3351-3357.
13. Qureshi, S.A. 2007. Hepatitis C virus--biology, host evasion strategies, and promising new therapies on the horizon. *Med Res Rev* 27:353-373.
14. De Francesco, R. 1999. Molecular virology of the hepatitis C virus. *J Hepatol* 31 Suppl 1:47-53.
15. Egger, D., B. Wolk, R. Gosert, L. Bianchi, H.E. Blum, D. Moradpour, and K. Bienz. 2002. Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex. *J Virol* 76:5974-5984.
16. Gosert, R., D. Egger, V. Lohmann, R. Bartenschlager, H.E. Blum, K. Bienz, and D. Moradpour. 2003. Identification of the hepatitis C virus RNA replication complex in Huh-7 cells harboring subgenomic replicons. *J Virol* 77:5487-5492.
17. Behrens, S.E., L. Tomei, and R. De Francesco. 1996. Identification and properties of the RNA-dependent RNA polymerase of hepatitis C virus. *Embo J* 15:12-22.

18. Lohmann, V., F. Korner, J. Koch, U. Herian, L. Theilmann, and R. Bartenschlager. 1999. Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science* 285:110-113.
19. Lesburg, C.A., M.B. Cable, E. Ferrari, Z. Hong, A.F. Mannarino, and P.C. Weber. 1999. Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. *Nat Struct Biol* 6:937-943.
20. Bressanelli, S., L. Tomei, A. Roussel, I. Incitti, R.L. Vitale, M. Mathieu, R. De Francesco, and F.A. Rey. 1999. Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Proc Natl Acad Sci U S A* 96:13034-13039.
21. Claros, M.G., and G. von Heijne. 1994. TopPred II: an improved software for membrane protein structure predictions. *Comput Appl Biosci* 10:685-686.
22. Krogh, A., B. Larsson, G. von Heijne, and E.L. Sonnhammer. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305:567-580.
23. Haqshenas, G., J.M. Mackenzie, X. Dong, and E.J. Gowans. 2007. Hepatitis C virus p7 protein is localized in the endoplasmic reticulum when it is encoded by a replication-competent genome. *J Gen Virol* 88:134-142.
24. Lin, C., B.D. Lindenbach, B.M. Pragai, D.W. McCourt, and C.M. Rice. 1994. Processing in the hepatitis C virus E2-NS2 region: identification of p7 and two distinct E2-specific products with different C termini. *J Virol* 68:5063-5073.
25. Bartenschlager, R. 2006. Hepatitis C virus molecular clones: from cDNA to infectious virus particles in cell culture. *Curr Opin Microbiol* 9:416-422.
26. Lindenbach, B.D., M.J. Evans, A.J. Syder, B. Wolk, T.L. Tellinghuisen, C.C. Liu, T. Maruyama, R.O. Hynes, D.R. Burton, J.A. McKeating, and C.M. Rice. 2005. Complete replication of hepatitis C virus in cell culture. *Science* 309:623-626.
27. Wakita, T., T. Pietschmann, T. Kato, T. Date, M. Miyamoto, Z. Zhao, K. Murthy, A. Habermann, H.G. Krausslich, M. Mizokami, R. Bartenschlager, and T.J. Liang. 2005. Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nat Med* 11:791-796.
28. Yi, M., R.A. Villanueva, D.L. Thomas, T. Wakita, and S.M. Lemon. 2006. Production of infectious genotype 1a hepatitis C virus (Hutchinson strain) in cultured human hepatoma cells. *Proc Natl Acad Sci U S A* 103:2310-2315.
29. Zhong, J., P. Gastaminza, G. Cheng, S. Kapadia, T. Kato, D.R. Burton, S.F. Wieland, S.L. Uprichard, T. Wakita, and F.V. Chisari. 2005. Robust hepatitis C virus infection in vitro. *Proc Natl Acad Sci U S A* 102:9294-9299.
30. Jones, C.T., C.L. Murray, D.K. Eastman, J. Tassello, and C.M. Rice. 2007. Hepatitis C virus p7 and NS2 proteins are essential for infectious virus production. *J Virol*
31. Steinmann, E., F. Penin, S. Kallis, A.H. Patel, R. Bartenschlager, and T. Pietschmann. 2007. Hepatitis C Virus p7 Protein Is Crucial for Assembly and Release of Infectious Virions. *PLoS Pathog* 3:e103.
32. Sakai, A., M.S. Claire, K. Faulk, S. Govindarajan, S.U. Emerson, R.H. Purcell, and J. Bukh. 2003. The p7 polypeptide of hepatitis C virus is critical for infectivity and contains functionally important genotype-specific sequences. *Proc Natl Acad Sci U S A* 100:11646-11651.

33. Haqshenas, G., X. Dong, G. Ewart, S. Bowden, and E.J. Gowans. 2007. A 2a/1b full-length p7 inter-genotypic chimeric genome of hepatitis C virus is infectious in vitro. *Virology* 360:17-26.
34. Griffin, S.D., L.P. Beales, D.S. Clarke, O. Worsfold, S.D. Evans, J. Jaeger, M.P. Harris, and D.J. Rowlands. 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.
35. Pavlovic, D., D.C. Neville, O. Argaud, B. Blumberg, R.A. Dwek, W.B. Fischer, and N. Zitzmann. 2003. The hepatitis C virus p7 protein forms an ion channel that is inhibited by long-alkyl-chain iminosugar derivatives. *Proc Natl Acad Sci U S A* 100:6104-6108.
36. Premkumar, A., L. Wilson, G.D. Ewart, and P.W. Gage. 2004. Cation-selective ion channels formed by p7 of hepatitis C virus are blocked by hexamethylene amiloride. *FEBS Lett* 557:99-103.
37. Clarke, D., S. Griffin, L. Beales, C.S. Gelais, S. Burgess, M. Harris, and D. Rowlands. 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.
38. Gonzalez, M.E., and L. Carrasco. 2003. Viroporins. *FEBS Lett* 552:28-34.
39. Pinto, L.H., and R.A. Lamb. 2006. The M2 proton channels of influenza A and B viruses. *J Biol Chem* 281:8997-9000.
40. Agirre, A., A. Barco, L. Carrasco, and J.L. Nieva. 2002. Viroporin-mediated membrane permeabilization. Pore formation by nonstructural poliovirus 2B protein. *J Biol Chem* 277:40434-40441.
41. Melton, J.V., G.D. Ewart, R.C. Weir, P.G. Board, E. Lee, and P.W. Gage. 2002. Alphavirus 6K proteins form ion channels. *J Biol Chem* 277:46923-46931.
42. Grice, A.L., I.D. Kerr, and M.S. Sansom. 1997. Ion channels formed by HIV-1 Vpu: a modelling and simulation study. *FEBS Lett* 405:299-304.
43. Sugrue, R.J., and A.J. Hay. 1991. Structural characteristics of the M2 protein of influenza A viruses: evidence that it forms a tetrameric channel. *Virology* 180:617-624.
44. Holsinger, L.J., and R.A. Lamb. 1991. Influenza virus M2 integral membrane protein is a homotetramer stabilized by formation of disulfide bonds. *Virology* 183:32-43.
45. Brillanti, S., F. Levantesi, L. Masi, M. Foli, and L. Bolondi. 2000. Triple antiviral therapy as a new option for patients with interferon nonresponsive chronic hepatitis C. *Hepatology* 32:630-634.
46. Berg, T., B. Kronenberger, H. Hinrichsen, T. Gerlach, P. Buggisch, E. Herrmann, U. Spengler, T. Goeser, S. Nasser, K. Wursthorn, G.R. Pape, U. Hopf, and S. Zeuzem. 2003. Triple therapy with amantadine in treatment-naive patients with chronic hepatitis C: a placebo-controlled trial. *Hepatology* 37:1359-1367.
47. Deltenre, P., J. Henrion, V. Canva, S. Dharancy, F. Texier, A. Louvet, S. De Maeght, J.C. Paris, and P. Mathurin. 2004. Evaluation of amantadine in chronic hepatitis C: a meta-analysis. *J Hepatol* 41:462-473.
48. Mangia, A., G. Leandro, B. Helbling, E.L. Renner, M. Tabone, L. Sidoli, S. Caronia, G.R. Foster, S. Zeuzem, T. Berg, V. Di Marco, N. Cino, and A. Andriulli. 2004. Combination therapy with amantadine and interferon in naive

patients with chronic hepatitis C: meta-analysis of individual patient data from six clinical trials. *J Hepatol* 40:478-483.

49. Smith, J.P. 1997. Treatment of chronic hepatitis C with amantadine. *Dig Dis Sci* 42:1681-1687.
50. Wang, J., S. Kim, F. Kovacs, and T.A. Cross. 2001. Structure of the transmembrane region of the M2 protein H(+) channel. *Protein Sci* 10:2241-2250.
51. Patargias, G., N. Zitzmann, R. Dwek, and W.B. Fischer. 2006. Protein-protein interactions: modeling the hepatitis C virus ion channel p7. *J Med Chem* 49:648-655.
52. Castelain, S., D. Bonte, F. Penin, C. Francois, D. Capron, S. Dedeurwaerder, P. Zawadzki, V. Morel, C. Wychowski, and G. Duverlie. 2007. Hepatitis C Virus p7 membrane protein quasispecies variability in chronically infected patients treated with interferon and ribavirin, with or without amantadine. *J Med Virol* 79:144-154.
53. Steinmann, E., T. Whitfield, S. Kallis, R.A. Dwek, N. Zitzmann, T. Pietschmann, and R. Bartenschlager. 2007. Antiviral effects of amantadine and iminosugar derivatives against hepatitis C virus. *Hepatology* 46:330-338.
54. Drew, D., L. Froderberg, L. Baars, and J.W. de Gier. 2003. Assembly and overexpression of membrane proteins in Escherichia coli. *Biochim Biophys Acta* 1610:3-10.
55. Laage, R., and D. Langosch. 2001. Strategies for prokaryotic expression of eukaryotic membrane proteins. *Traffic* 2:99-104.
56. Cunningham, F., and C.M. Deber. 2007. Optimizing synthesis and expression of transmembrane peptides and proteins. *Methods* 41:370-380.
57. Davis, G.D., C. Elisee, D.M. Newham, and R.G. Harrison. 1999. New fusion protein systems designed to give soluble expression in Escherichia coli. *Biotechnol Bioeng* 65:382-388.
58. Kiefer, H., J. Krieger, J.D. Olszewski, G. Von Heijne, G.D. Prestwich, and H. Breer. 1996. Expression of an olfactory receptor in Escherichia coli: purification, reconstitution, and ligand binding. *Biochemistry* 35:16077-16084.
59. Montigny, C., F. Penin, C. Lethias, and P. Falson. 2004. Overcoming the toxicity of membrane peptide expression in bacteria by upstream insertion of Asp-Pro sequence. *Biochim Biophys Acta* 1660:53-65.
60. Cheng, J. 2006. Membrane Protein Overproduction in Escherichia coli: Investigation of Factors Affecting Membrane Protein Expression and Purification. In Biochemistry, Microbiology and Immunology. University of Ottawa, Ottawa. 120.
61. Crowe, J., H. Dobeli, R. Gentz, E. Hochuli, D. Stuber, and K. Henco. 1994. 6xHis-Ni-NTA chromatography as a superior technique in recombinant protein expression/purification. *Methods Mol Biol* 31:371-387.
62. Ma, C., F.M. Marassi, D.H. Jones, S.K. Straus, S. Bour, K. Strebel, U. Schubert, M. Oblatt-Montal, M. Montal, and S.J. Opella. 2002. Expression, purification, and activities of full-length and truncated versions of the integral membrane protein Vpu from HIV-1. *Protein Sci* 11:546-557.
63. Hwang, P.M., W.Y. Choy, E.I. Lo, L. Chen, J.D. Forman-Kay, C.R. Raetz, G.G. Prive, R.E. Bishop, and L.E. Kay. 2002. Solution structure and dynamics of the

outer membrane enzyme PagP by NMR. *Proc Natl Acad Sci U S A* 99:13560-13565.

64. Jenny, R.J., K.G. Mann, and R.L. Lundblad. 2003. A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. *Protein Expr Purif* 31:1-11.
65. Eisenmesser, E.Z., R.B. Kapust, J.P. Nawrocki, M.J. Mazzulla, L.K. Pannell, D.S. Waugh, and R.A. Byrd. 2000. Expression, purification, refolding, and characterization of recombinant human interleukin-13: utilization of intracellular processing. *Protein Expr Purif* 20:186-195.
66. Chelius, D., M.A. Baldwin, X. Lu, and E.M. Spencer. 2001. Expression, purification and characterization of the structure and disulfide linkages of insulin-like growth factor binding protein-4. *J Endocrinol* 168:283-296.
67. MacDonald, L.M., A. Armson, R.C. Thompson, and J.A. Reynoldson. 2001. Expression of Giardia duodenalis beta-tubulin as a soluble protein in Escherichia coli. *Protein Expr Purif* 22:25-30.
68. Hanck, T., R. Stricker, U.M. Krishna, J.R. Falck, Y.T. Chang, S.K. Chung, and G. Reiser. 1999. Recombinant p42IP4, a brain-specific 42-kDa high-affinity Ins(1,3,4,5)P4 receptor protein, specifically interacts with lipid membranes containing Ptd-Ins(3,4,5)P3. *Eur J Biochem* 261:577-584.
69. Panayotova-Heiermann, M., D.W. Leung, B.A. Hirayama, and E.M. Wright. 1999. Purification and functional reconstitution of a truncated human Na(+)/glucose cotransporter (SGLT1) expressed in E. coli. *FEBS Lett* 459:386-390.
70. Moss, J., R.S. Haun, S.C. Tsai, C.F. Welsh, F.J. Lee, S.R. Price, and M. Vaughan. 1994. Activation of cholera toxin by ADP-ribosylation factors: 20-kDa guanine nucleotide-binding proteins. *Methods Enzymol* 237:44-63.
71. McDonald, O.B., W.J. Chen, B. Ellis, C. Hoffman, L. Overton, M. Rink, A. Smith, C.J. Marshall, and E.R. Wood. 1999. A scintillation proximity assay for the Raf/MEK/ERK kinase cascade: high-throughput screening and identification of selective enzyme inhibitors. *Anal Biochem* 268:318-329.
72. Sanders, C.R., and F. Sonnichsen. 2006. Solution NMR of membrane proteins: practice and challenges. *Magn Reson Chem* 44 Spec No:S24-40.
73. Krueger-Koplin, R.D., P.L. Sorgen, S.T. Krueger-Koplin, I.O. Rivera-Torres, S.M. Cahill, D.B. Hicks, L. Grinius, T.A. Krulwich, and M.E. Girvin. 2004. An evaluation of detergents for NMR structural studies of membrane proteins. *J Biomol NMR* 28:43-57.
74. Prive, G.G. 2007. Detergents for the stabilization and crystallization of membrane proteins. *Methods* 41:388-397.
75. Vinogradova, O., F. Sonnichsen, and C.R. Sanders, 2nd. 1998. On choosing a detergent for solution NMR studies of membrane proteins. *J Biomol NMR* 11:381-386.
76. Sanders, C.R., 2nd, and G.C. Landis. 1995. Reconstitution of membrane proteins into lipid-rich bilayered mixed micelles for NMR studies. *Biochemistry* 34:4030-4040.

77. Jastrzebska, B., T. Maeda, L. Zhu, D. Fotiadis, S. Filipek, A. Engel, R.E. Stenkamp, and K. Palczewski. 2004. Functional characterization of rhodopsin monomers and dimers in detergents. *J Biol Chem* 279:54663-54675.
78. Tamm, L.K., and B. Liang. 2006. NMR of membrane proteins in solution. *Prog Nucl Magn Res Spectrosc* 48:201-210.
79. Columbus, L., J. Lipfert, H. Klock, I. Millett, S. Doniach, and S.A. Lesley. 2006. Expression, purification, and characterization of Thermotoga maritima membrane proteins for structure determination. *Protein Sci* 15:961-975.
80. Chattopadhyay, A., and E. London. 1984. Fluorimetric determination of critical micelle concentration avoiding interference from detergent charge. *Anal Biochem* 139:408-412.
81. Lauterwein, J., C. Bosch, L.R. Brown, and K. Wuthrich. 1979. Physicochemical studies of the protein-lipid interactions in melittin-containing micelles. *Biochim Biophys Acta* 556:244-264.
82. Klammt, C., D. Schwarz, K. Fendler, W. Haase, V. Dotsch, and F. Bernhard. 2005. Evaluation of detergents for the soluble expression of alpha-helical and beta-barrel-type integral membrane proteins by a preparative scale individual cell-free expression system. *Febs J* 272:6024-6038.
83. Lichtenberg, D., R.J. Robson, and E.A. Dennis. 1983. Solubilization of phospholipids by detergents. Structural and kinetic aspects. *Biochim Biophys Acta* 737:285-304.
84. Duchardt, E., A.B. Sigalov, D. Aivazian, L.J. Stern, and H. Schwalbe. 2007. Structure induction of the T-cell receptor zeta-chain upon lipid binding investigated by NMR spectroscopy. *Chembiochem* 8:820-827.
85. Kanelis, V., J.D. Forman-Kay, and L.E. Kay. 2001. Multidimensional NMR methods for protein structure determination. *IUBMB Life* 52:291-302.
86. Johnson, B.A., and R.A. Blevins. 1994. NMR View: A computer program for the visualization and analysis of NMR data. *J Biomol NMR* 4:603-614.
87. Johnson, B.A. 2004. Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* 278:313-352.
88. Wittekind, M., and Mueller, L. 1993. HNCACB, a high sensitivity 3D NMR experiment to correlate amide proton and nitrogen resonances with the alpha and beta carbon resonances in proteins. *J. Magn. Reson. Ser. B* 101:201-205.
89. Sattler, M., Schleucher, J., and Griesinger, C. 1999. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Magn Res Spectrosc* 34:93-158.
90. Seavey, B.R., E.A. Farr, W.M. Westler, and J.L. Markley. 1991. A relational database for sequence-specific protein NMR data. *J Biomol NMR* 1:217-236.
91. Wishart, D.S., and B.D. Sykes. 1994. The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. *J Biomol NMR* 4:171-180.
92. Wishart, D.S., C.G. Bigam, A. Holm, R.S. Hodges, and B.D. Sykes. 1995. 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J Biomol NMR* 5:67-81.

93. Wishart, D.S., B.D. Sykes, and F.M. Richards. 1991. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J Mol Biol* 222:311-333.
94. Baleja, J.D. 2001. Structure determination of membrane-associated proteins from nuclear magnetic resonance data. *Anal Biochem* 288:1-15.
95. Brunger, A.T., P.D. Adams, and L.M. Rice. 1997. New applications of simulated annealing in X-ray crystallography and solution NMR. *Structure* 5:325-336.
96. Clore, G.M., and A.M. Gronenborn. 1994. Structures of larger proteins, protein-ligand and protein-DNA complexes by multi-dimensional heteronuclear NMR. *Prog Biophys Mol Biol* 62:153-184.
97. Salzmann, M., K. Pervushin, G. Wider, H. Senn, and K. Wuthrich. 1998. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc Natl Acad Sci U S A* 95:13585-13590.
98. Foster, M.P., C.A. McElroy, and C.D. Amero. 2007. Solution NMR of large molecules and assemblies. *Biochemistry* 46:331-340.
99. Tugarinov, V., R. Muhandiram, A. Ayed, and L.E. Kay. 2002. Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase g. *J Am Chem Soc* 124:10025-10035.
100. Rosen, M.K., K.H. Gardner, R.C. Willis, W.E. Parris, T. Pawson, and L.E. Kay. 1996. Selective methyl group protonation of perdeuterated proteins. *J Mol Biol* 263:627-636.
101. Sattler, M., and S.W. Fesik. 1996. Use of deuterium labeling in NMR: overcoming a sizeable problem. *Structure* 4:1245-1249.
102. Wider, G., and K. Wuthrich. 1999. NMR spectroscopy of large molecules and multimolecular assemblies in solution. *Curr Opin Struct Biol* 9:594-601.
103. Fernandez, C., C. Hilty, G. Wider, P. Guntert, and K. Wuthrich. 2004. NMR structure of the integral membrane protein OmpX. *J Mol Biol* 336:1211-1221.
104. Arora, A., F. Abildgaard, J.H. Bushweller, and L.K. Tamm. 2001. Structure of outer membrane protein A transmembrane domain by NMR spectroscopy. *Nat Struct Biol* 8:334-338.
105. Oxenoid, K., and J.J. Chou. 2005. The structure of phospholamban pentamer reveals a channel-like architecture in membranes. *Proc Natl Acad Sci U S A* 102:10870-10875.
106. Sambrook, J., and D.W. Russell. 2001. Molecular Cloning: A Laboratory Manual Cold Spring Harbor Laboratory Press, New York.
107. Dempsey, C.E., T.J. Piggot, and P.E. Mason. 2005. Dissecting contributions to the denaturant sensitivities of proteins. *Biochemistry* 44:775-781.
108. Welling, G.W., R. van der Zee, and S. Welling-Wester. 1987. Column liquid chromatography of integral membrane proteins. *J Chromatogr* 418:223-243.
109. Kita, K., H. Murakami, H. Oya, and Y. Anraku. 1985. Quantitative determination of cytochromes in the aerobic respiratory chain of Escherichia coli by high-performance liquid chromatography and its application to analysis of mitochondrial cytochromes. *Biochem Int* 10:319-326.
110. Soutschek, E., B. Hoflacher, and M. Motz. 1990. Purification of a recombinantly produced transmembrane protein (gp41) of HIV I. *J Chromatogr* 521:267-277.

111. Tempst, P., D.D. Woo, D.B. Teplow, R. Aebersold, L.E. Hood, and S.B. Kent. 1986. Microscale structure analysis of a high-molecular-weight, hydrophobic membrane glycoprotein fraction with platelet-derived growth factor-dependent kinase activity. *J Chromatogr* 359:403-412.
112. Matson, R.S., and S.C. Goheen. 1986. Use of high-performance size exclusion chromatography to determine the extent of detergent solubilization of human erythrocyte ghosts. *J Chromatogr* 359:285-295.
113. Shire, S.J., L. Bock, J. Ogez, S. Builder, D. Kleid, and D.M. Moore. 1984. Purification and immunogenicity of fusion VP1 protein of foot and mouth disease virus. *Biochemistry* 23:6474-6480.
114. Gasteiger E., H.C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols and Handbook. J.M. Walker, editor Humana Press, Totowas. 571-607.
115. Fassina, G., S. Merli, S. Germani, G. Ciliberto, and G. Cassani. 1994. High yield expression and purification of human endothelin-1. *Protein Expr Purif* 5:559-568.
116. Sato, T., T. Kawakami, K. Akaji, H. Konishi, K. Mochizuki, T. Fujiwara, H. Akutsu, and S. Aimoto. 2002. Synthesis of a membrane protein with two transmembrane regions. *J Pept Sci* 8:172-180.
117. Lee, R.P., S.W. Doughty, K. Ashman, and J. Walker. 1996. Purification of hydrophobic integral membrane proteins from Mycoplasma hyopneumoniae by reversed-phase high-performance liquid chromatography. *J Chromatogr A* 737:273-279.
118. Dunbar, A.J., I.K. Priebe, M.P. Sanderson, and C. Goddard. 2001. Purification and molecular characterization of recombinant rat betacellulin. *J Mol Endocrinol* 27:239-247.
119. Tomich, J.M., L.W. Carson, K.J. Kanes, N.J. Vogelaar, M.R. Emerling, and J.H. Richards. 1988. Prevention of aggregation of synthetic membrane-spanning peptides by addition of detergent. *Anal Biochem* 174:197-203.
120. Carr, D. 2002. The Handbook of Analysis and Purification of Peptides and Proteins by Reversed-Phase HPLC. GraceVydac, California, USA.
121. Page, R.C., J.D. Moore, H.B. Nguyen, M. Sharma, R. Chase, F.P. Gao, C.K. Mobley, C.R. Sanders, L. Ma, F.D. Sonnichsen, S. Lee, S.C. Howell, S.J. Opella, and T.A. Cross. 2006. Comprehensive evaluation of solution nuclear magnetic resonance spectroscopy sample preparation for helical integral membrane proteins. *J Struct Funct Genomics* 7:51-64.
122. Damberg, P., J. Jarvet, and A. Graslund. 2001. Micellar systems as solvents in peptide and protein structure determination. *Methods Enzymol* 339:271-285.
123. Fernandez, C., and K. Wuthrich. 2003. NMR solution structure determination of membrane proteins reconstituted in detergent micelles. *FEBS Lett* 555:144-150.
124. Delaglio, F., S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277-293.
125. Chill, J.H., J.M. Louis, C. Miller, and A. Bax. 2006. NMR study of the tetrameric KcsA potassium channel in detergent micelles. *Protein Sci* 15:684-698.

126. Zamoon, J., A. Mascioni, D.D. Thomas, and G. Veglia. 2003. NMR solution structure and topological orientation of monomeric phospholamban in dodecylphosphocholine micelles. *Biophys J* 85:2589-2598.
127. Wu, J., Shish, S.C. and N.K. Goto. 2007. Probing the structure of the Ff bacteriophage major coat protein transmembrane helix dimer by solution NMR. *Accepted manuscript*
128. Griffin, S.D., R. Harvey, D.S. Clarke, W.S. Barclay, M. Harris, and D.J. Rowlands. 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. *J Gen Virol* 85:451-461.
129. Ramos, D. 2006. Conformational Studies of Cell Division Regulator MinE by Nuclear Magnetic Resonance and Circular Dichroism Spectroscopy. In Chemistry. University of Ottawa, Ottawa. 105.
130. Fernandez, C., and G. Wider. 2003. TROSY in NMR studies of the structure and function of large biological macromolecules. *Curr Opin Struct Biol* 13:570-580.
131. Fleming, K.G. 2002. Standardizing the free energy change of transmembrane helix-helix interactions. *J Mol Biol* 323:563-571.
132. Fleming, K.G., and D.M. Engelman. 2001. Specificity in transmembrane helix-helix interactions can define a hierarchy of stability for sequence variants. *Proc Natl Acad Sci U S A* 98:14340-14344.

Appendix

DNA sequence of constructs used in this study

GST-His-p7

gnggcgacctncnnccaaatcggnnnctgtctggtggtggtggtggtcatcatcatcatcatcatctggttccgcgtggatccgc
gctggaaaatctggtgattctgaatgcggcgagcctggcgggtacccatggtctggtgagctttctggtgtttttttgctttgcgtgg
tatctgaaaggtcgttgggtgccaggtgcggtgtatgcgttttatggtatgtggccactgctgctgctgctgctggcgctgccaca
gcgtgcgtatgcggaattcatcgtgactgactgacgatctgcctcgcgcgtttcggtgatgacggtgaaaacctctgacacatgc
agctcccggagacggtcacagcttgtctgtaagcggatgccgggagcagacaagcccgtcagggcgcgtcagcgggtgttg
gcgggtgtcggggcgcagccatgacccagtcacgtagcgatagcggagtgtaaattcttgaagacgaaagggcctcgtgatac
gcctatttttataggttaatgtcatgataataatggtttcttagacgtcaggtggcacttttcggggaaatgtgcgcggaaccccctatttt
gtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagta
tgagtattcaacatttccgtgtcgcccttattccctttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagta
aaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgcc
ccgaagaacgttttncaatgatgagcactttnaagttctgctatgtggcgcggtattatcccgtgttgacgccgggcaagagcaac
tcggtcgccccatacactattctaaaatgacttggntgagtactcccagtccagaaagcatcttacggatggcatgacngtaaga
nattatgcagtgctgcctaccatgagtgaaacctgngccanttactntgaaacgatcggaggacgangagctaccgtttttgcca
catggggatattaactcccttntcgtggnaccgacngatganctncaacanngnngacccentgctnnaatggnnacntgcc
anntantggnaattnctttntcngnnaaaaaacggtggggnnagtcngccttngcngccccgtggngtttgtnaatgncngn
ngggtnngtnttgnctgccangnacccctttttttnccgggnnnntnnaaaaaancngaggccctnantgacgccntnnnctt
nntantttagtggacttatncncncggttcngccnaaanttnnttnnntaaaccnggngaacttnggnnaattn

GST-p7-His

gnggcgacctncnnccaannegtttnctgtctggtcggtggtggtggtctggttccgcgtggatccgcgctggaaaatctggtg
attctgaatgcggcgagcctggcgggtacccatggtctggtgagctttctagtgtttttttgctttgcgtggtatctgaaaggtcgttg
ggtgccaggtgcggtgtatgcgttttatggtatgtggccactgctgctgctgctgctggcgctgccacagcgtgcgtatgcgcatc
accaccatcatcacttagaattcatcgtgactgactgacgatctgcctcgcgcgtttcggtgatgacggtgaaaacctctgacacat
gcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcagacaagcccgtcagggcgcgtcagcgggtgtt
ggcgggtgtcggggcgcagccatgacccagtcacgtagcgatagcggagtgtaaattcttgaagacgaaagggcctcgtgat
acgcctatttttataggttaatgtcatgataataatggtttcttagacgtcaggtggcacttttcggggaaatgtgcgcggaacccctа
tttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagag
tatgagtattcaacatttccgtgtcgcccttattccctttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaag
taaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgc
cccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtgttgacgccgggcaagagca
actcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaacatcttacggatggcatgacagt
aagagaatatgcagtgctgcntaccatgagtgataccctgcggcanttactntnacacgatcggaggacgaaganctaccgtttt
tnccaaatggggatatgtactcncttgatnttggaccgnctgatgacctccaacanagcggaaccgatgctnncatgnacacgt
nccaacatactgggannnttttantcngncaatanaactgtgngngnaagtncgncttgcccgcctcgggctgttttgnnaat
nnccggacnggttcgttntnnctggcaagacccctntttnccggntgcntggaaaannnngaggccntacngagcncntn
ntntatgtantttagtggaactattnnnctggttcgnacaaaanttatttnnnnnaacccggtgnaantnggnnaanttgnnc

N.B. Blue: Thrombin cleavage site, Red: p7 sequence, Purple: Hexahistidine tag