Hepatitis C virus NS5A is a zinc metalloprotein

The NS5A Protein of Hepatitis C Virus is a Zinc Metalloprotein

Timothy L. Tellinghuisen¹, Joseph Marcotrigiano¹, Alexander E. Gorbalenya², and Charles M. Rice¹*

1. Laboratory of Virology and Infectious Disease
   Center for the Study of Hepatitis C
   The Rockefeller University
   1230 York Avenue, Box 64
   New York, New York 10021
   Tel: +1 212 327 7046
   Fax: +1 212 327 7048
   e-mails: rice@rockefeller.edu
tellint@rockefeller.edu
marcoj@rockefeller.edu

2. Department of Medical Microbiology
   Center of Infectious Diseases
   Leiden University Medical Center
   PO Box 9600, E4-P, Room L04-036
   2300 RC Leiden
   The Netherlands
   Phone: **31-71-526-1436/4830 (voice)
   Phone: **31-71-526-1652 (MM Secretariat)
   Fax: **31-71-526-6761
e-mail: a.e.gorbalenya@lumc.nl

* corresponding author
Hepatitis C virus NS5A is a zinc metalloprotein

Running Title: Hepatitis C virus NS5A is a zinc metalloprotein

Summary:

The NS5A protein of hepatitis C virus is believed to be an integral part of the viral replicase. Despite extensive investigation, the role of this protein remains elusive. Only limited biochemical characterization of NS5A has been performed, with most research to date involving the myriad of host proteins and signaling cascades that interact with NS5A. The need for better characterization of NS5A is paramount for elucidating the role of this protein in the virus life cycle. Examination of NS5A using bioinformatics tools suggested the protein consisted of three domains and contained an unconventional zinc-binding motif within the amino-terminal domain. We have developed a method to produce NS5A and performed limited proteolysis to confirm the domain organization model. The zinc content of purified NS5A and the amino-terminal domain of NS5A was determined, and each of these proteins was found to coordinate one zinc atom per protein. The predicted zinc-binding motif consists of four cysteine residues, conserved among the Hepacivirus and Pestivirus genera, fitting the formula of Cx^{17}CxCx^{20}C. Mutation of any of the four cysteine components of this motif reduced NS5A zinc coordination and led to a lethal phenotype for HCV RNA replication, whereas mutation of other potential metal coordination residues in the amino-terminal domain of NS5A, but outside the zinc-binding motif, had little effect on zinc binding and, aside from one exception, were tolerated for replication. Collectively, these results indicate that NS5A is a zinc metalloprotein and that zinc coordination is likely required for NS5A function in the hepatitis C replicase.
Hepatitis C virus NS5A is a zinc metalloprotein

**Introduction:**

Since the discovery of the hepatitis C virus (HCV) as the causative agent of non-A, non-B hepatitis in 1989 (1), understanding the biology of this pathogen has been a major focus of both basic and clinical research. HCV is a significant global public health problem, with estimates placing nearly 3% of the world’s population, roughly 170 million people, as HCV infected (2). Despite recent advances, current HCV therapeutics are inadequate for the majority of patients.

HCV is a member of the *Flaviviridae* family of enveloped, positive-strand RNA viruses and is the sole member of the *Hepacivirus* genus (3). The HCV genome consists of a ~9.6 kb RNA molecule containing a large open reading frame flanked by structured 5' and 3' non-translated regions (NTR). Viral RNA is translated to generate a polyprotein precursor via an internal ribosome entry site within the 5' NTR. The ten HCV proteins are organized in the polyprotein in the order: NH$_2$-C-E1-E2-p7-NS2-NS3-NS4A-NS4B-NS5A-NS5B-COOH (4). An additional protein, F (for frameshift protein) or ARFP (alternate reading frame protein), generated from an overlapping reading frame in the core (C) protein coding sequence, has been proposed (5) (6,7). The polyprotein undergoes a series of membrane associated co- and post-translational cleavages by viral and host cell proteases to yield the mature forms of the individual HCV proteins (8). The structural proteins, C, E1, and E2 are directed to the ER-Golgi complex and processed by resident enzyme activities to generate the components for the assembly of progeny. The small hydrophobic p7 protein has been demonstrated to form ion channels in the host cell membrane, although the functional consequences of this observation are not clear (9). The non-structural (NS) proteins are processed by the two distinct viral protease activities
Hepatitis C virus NS5A is a zinc metalloprotein (8,10). The integral membrane NS2 protein, together with the amino terminal region of the NS3 protein, constitutes the NS2-3 cysteine proteinase that catalyzes the cleavage between NS2 and NS3 (8,10). The NS2 protein has no known function after this cleavage has occurred, and this protein is not required for RNA replication in cell culture (11,12). The NS3 protein, in conjunction with the NS4A cofactor, serves as a serine proteinase for the cleavage of the remaining non-structural proteins (10,13,14). Once cleaved, the NS proteins assemble into the membrane associated HCV RNA replicase complex. The putative HCV replicase complex has been imaged in replicon-containing cell lines in a structure deemed the membranous web (15). This complex has been shown to contain all of the HCV NS proteins, as well as actively replicating HCV RNA (15). Functional analysis of the HCV replicase using the HCV replicon system has defined NS3, NS4A, NS4B, NS5A, and NS5B as essential components of the replicase (11,12).

The functions of a number of the NS proteins have been at least partially defined. In addition to NS3’s aforementioned role in polyprotein processing, its carboxyl-terminal domain serves as an RNA helicase activity likely utilized during replication (16). The small NS4A protein serves as a cofactor for both the protease and helicase activities of NS3 (14,17). NS4B is an integral membrane protein that has a direct role in reorganization of cellular membranes to form the membranous web (18). Additionally, this protein has been proposed to function as a membrane bound scaffolding upon which the remaining replicase components assemble. The NS5B protein serves as the viral RNA dependent RNA polymerase (1). The function of NS5A in the replicase is not clear, but this protein appears to be an absolutely required component of the replicase and is required for HCV replication (11,12).
Hepatitis C virus NS5A is a zinc metalloprotein

NS5A is a large (56-58 kDa), hydrophilic phosphoprotein of unknown function. The ability of adaptive mutations in NS5A to greatly stimulate HCV replication in cell culture, and the association of this protein with other members of the putative replicase suggests NS5A plays an important role in RNA replication (11,12). The NS5A protein exists in multiple phosphorylation states, designated p56 (basal) and p58 (hyper), based on their migration on SDS-PAGE gels. Considerable effort has been placed in mapping NS5A phosphorylation sites and identifying the enzyme(s) responsible for this activity. Despite this, our knowledge of phosphate acceptor sites is rudimentary and the kinase(s) remains elusive (19-24). Although some NS5A adaptive mutations affect the phosphorylation state of NS5A in HCV replicons (11), the link between NS5A phosphorylation and replication is unknown. Additional roles for NS5A in modulating various cell signaling cascades have been proposed, perhaps based on the observed interaction of NS5A with cellular kinase activities, but the relevance of many of these to replication and pathogenesis remains to be demonstrated (see (25) for review).

Despite the vast number of NS5A related publications, surprisingly little effort has been devoted to the basic biochemical characterization of this protein. One area of recent attention has been the interaction of NS5A with cellular membranes. NS5A has been shown to be post-translationally associated with ER derived membranes via an amino-terminal amphipathic α-helix that has been proposed to be partially buried in one leaflet of the cellular membrane (26). Disruption of this helix leads to a diffuse cytoplasmic localization of NS5A and is lethal for HCV RNA replication (27). These data, combined with sub-cellular co-localization studies, replicase component interaction studies, and NS5A adaptive replicon mutants provide an increasing body of evidence
Hepatitis C virus NS5A is a zinc metalloprotein suggesting the importance of NS5A in the RNA replicase. One area that is clearly lacking in NS5A research is an identification and evaluation of the domain organization of the protein and a biochemical evaluation of potentially important sequence motifs.

In this study, a combination of bioinformatics analysis, biochemical assays with purified NS5A protein, and replication studies, with the replicon system, were used to characterize NS5A. We have generated a model for the domain organization of NS5A, which was confirmed by limited proteolysis and mass spectrometry. Analysis of the sequence of NS5A has suggested the presence of an unconventional, yet highly conserved, zinc metal ion coordination site within the amino-terminal domain, which was evaluated by a combination of reverse genetics in the replicon system and atomic absorption spectroscopy using purified NS5A protein. The data presented herein provides the first model of NS5A domain organization and defines the essential zinc binding properties of NS5A, observations that further our understanding of this enigmatic protein.

**Experimental Procedures:**

**Cloning and in vitro mutagenesis:**

All mutagenesis was performed using the Quick Change (Stratagene) *in vitro* mutagenesis system following the manufacturer’s standard protocols. Mutations were generated in the shuttle vector pSL1180GIT, which contained the *BsrG1-MfeI* fragment of the Con1/SG-neo GIT replicon encompassing the NS5A coding sequence and surrounding region. The Con1/SG-neo GIT replicon is identical to the subgenomic replicons described previously(11), with the exception of the presence of two adaptive mutations in NS3 (E1202G, T1280I) and one in NS4A (K1846T), thereby allowing
Hepatitis C virus NS5A is a zinc metalloprotein analysis of NS5A mutants on a wild type NS5A backbone (28). All mutations generated by Quick-Change mutagenesis were confirmed by DNA sequencing of the final replicon plasmids. Fragments of interest (e.g. NS5A coding sequence) were then subcloned into the appropriate vectors for protein expression or replicon experiments as described below. For replicon experiments, the BsrG1-MfeI fragment of pSL1180GIT containing the desired mutation was directly cloned back into the parental Con1/Sg-neo GIT replicon plasmid. For protein expression clones, pSL1180GIT DNA containing the desired mutation was used as a template for PCR amplification of the NS5A coding sequence corresponding to amino acids 25 to 448 of NS5A (NS5A-Δh). PCR primers were designed to create an in frame 5’ SacII restriction site for fusion of NS5A-Δh to the ubiquitin coding sequence cloned into the NdeI (ATG start codon) site of pET30b. The 3’ end of the NS5A coding sequence was altered to include a KpnI site for the in frame fusion with the linker, enterokinase cleavage site and polyhistidine tag of pET30b. Domain I expression constructs were constructed similarly, with the exception of the introduction of the 3’ KpnI site in a location to generate an NS5A coding sequence terminating at amino acid 215 with a fusion to the linker, enterokinase cleavage site and polyhistidine tag of pET30b. Clones generated in pET30b using PCR were confirmed by an additional round of DNA sequencing to ensure no PCR based mutations were present. Numbering of NS5A amino acid residues corresponds to the first serine residues of NS5A as residue number one. Mutations of amino acids in NS5A are designated by; single letter amino acid code of the parental sequence, the residue number in NS5A, and the altered amino acid present in the mutant construct, for example, a change of cysteine 141 of NS5A to a glycine residue would be described as C141G. Conversion of this
Hepatitis C virus NS5A is a zinc metalloprotein numbering scheme to the amino acid numbering scheme of the complete 1b genotype HCV polyprotein requires the addition of 1972 to the numbers used in this manuscript. HCV genotype sequences used for sequence alignments in this manuscript are designated by their corresponding GenBank database accession numbers.

**Cell Culture:**

Huh-7.5 (human hepatoma) cells were maintained in Dulbecco’s Modified Eagle Medium (DMEM; Invitrogen) supplemented with 10% fetal bovine serum, non-essential amino acids, 100 U/ml penicillin, and 100 µg/ml streptomycin at 37°C in 5% CO2. Cells were passaged every 3 days at a split ratio of 1:4 after treatment with 0.05% trypsin, 0.02% EDTA. All experiments were performed using cells between passage number 28-35 expanded from an original seed stock of the Huh7.5 cell clone described previously (29).

**In Vitro RNA Transcription:**

In vitro RNA transcription of HCV replicon RNAs was performed as described previously (11). Cesium chloride density gradient purified Con1/SG-neo GIT DNA was linearized with ScaI (New England Biolabs), followed by phenol-chloroform extraction and ethanol precipitation. RNA transcripts were synthesized at 37°C for 1.5 hours in a 50 µL reaction containing 40 mM Tris-HCl (pH 7.9), 10 mM NaCl, 12 mM MgCl2, 2 mM spermidine, 10 mM dithiothreitol, 3 mM each rNTP, 100 U of RNAsin (Roche), 100 U
Hepatitis C virus NS5A is a zinc metalloprotein

T7 RNA polymerase (Epicentre Technologies), and 3 μg of linearized DNA template. Following transcription, 10 U of DNase I (Ambion) was added and reactions were incubated at 37°C for 20 minutes to remove the DNA template. RNA was then purified using the Qiagen RNAeasy kit following the manufacturer’s instructions (Qiagen). The optional ‘on column’ DNase treatment described for the Qiagen RNAeasy kit was used to further remove residual DNA template. RNA yield was determined by quantification with a UV spectrophotometer. The quality of RNA was further confirmed by agarose gel electrophoresis.

Electroporation of HCV Replicons and G418 Selection:

Experiments using HCV replicons were performed essentially as described previously, with minor modifications (11). Briefly, subconfluent Huh-7.5 cells at were trypsinized, washed twice in ice cold phosphate buffered saline, and resuspended at a concentration of 2.5x10^7 cells/ml in ice cold RNase free phosphate buffered saline. A 400 µl aliquot of cells was then mixed with 1 µg of HCV replicon RNA and electroporated using a BTX square wavelength electroporator in a 0.2 mm gap cuvette (five 99 microsecond pulses, 900 volts, 1.1 second time interval). Following a 10 minute recovery period at room temperature, cells were diluted in 9.6 ml of complete medium and plated into 100 mm dishes. Cells were plated a 1x10^6, 1x10^5, and 1x10^4 cells at a fixed overall cell density of 1x10^6 cells/ml, with the remaining cell density consisting of cells electroporated with a replication deficient Con1/SG-neo GIT pol- RNA. Selection of replicon containing colonies was performed 2 days post electroporation by the addition of complete medium
Hepatitis C virus NS5A is a zinc metalloprotein supplemented with 1 mg/ml G418. Selection was carried out for a period of two weeks post-electroporation with fresh G418 containing media added to the cells every four days. Following selection, cells were fixed with 7% formaldehyde and stained with 1% crystal violet in 50% ethanol. G418 selection was performed for three weeks for the data presented in Figure 5, as one mutant had a small colony phenotype. G418 resistant colonies were then counted as used to calculate transduction efficiency in units of colony forming units (cfu) per microgram of input RNA. All experiments presented are the average of three independent RNA preparations delivered by three independent electroporations.

**NS5A-Δh Expression and Purification:**

For the expression of NS5A-Δh, 1 L of LB medium supplemented with 30 µg /L of kanamycin and 25 µg/L of chloramphenicol was inoculated from an overnight culture of *E. coli* BL21(de3) containing pET30ubiNS5A-Δh and the ubiquitin specific protease pCG1 plasmids such that the cell density was at approximately OD$_{600}$ of 0.05. Cells were grown at 37°C and 250 r.p.m until an OD$_{600}$ of 0.55 was reached. Cells were then chilled at 4°C for 30 minutes, induced with 1 mM IPTG, and incubated for 5 hours at 25°C and 250 r.p.m. Cells were then collected by centrifugation at 6,000 x g for 10 minutes and resuspended in 20 ml of buffer A (100 mM Tris-HCl (pH8.0), 150 mM NaCl, 50 mM Imidazole, 10 mM 2-mercaptoethanol) per liter. Cells were lysed by three passes through a cold Avestin air emulsifier at 15,000 p.s.i. Following lysis, cell extracts were clarified at 25,000xg for 30 minutes at 4°C. Clarified extracts were then loaded on a 5 ml bed
Hepatitis C virus NS5A is a zinc metalloprotein volume HiTrap IMAC column (Amersham/Pharmacia) equilibrated with buffer A at a flow rate of 2.5 ml/minute. Following extensive washing with buffer A and buffer A supplemented with 1M NaCl, NS5A-Δh was eluted with a 25 ml linear gradient of buffer A supplemented with 500 mM imidazole. Fractions containing NS5A-Δh were pooled and exchanged into buffer B (25 mM HEPES (pH 7.4), 25 mM NaCl) using a HiPrep 26/10 desalting column at a flow rate of 8 ml/min (Amersham/Pharmacia). Desalted protein was then concentrated via Amicon ultra centrifugal concentrators (Millipore). Concentration of purified NS5A-Δh was determined using the Bio-Rad protein assay (Bio-Rad). Protein yields were typically between 2 and 5 mg/L of bacterial culture.

Domain Mapping of NS5A by Limited Trypsinization:

Fifteen micrograms per sample of purified NS5A-Δh in 50 mM HEPES (pH 7.4), 1mM DTT, 0.5% NP40, and 30% glycerol were digested with 2 μl of a 50% slurry of agarose bead immobilized TPCK treated trypsin (Pierce Biotechnology) equilibrated in the same buffer. Digestion was allowed to proceed at 37°C for a total of 3 hours, with samples removed every thirty minutes for analysis. Once removed, these samples were centrifuged for 1 minute at 10,000 x g in a microcentrifuge to pellet the immobilized trypsin. Supernatants were removed, mixed with an equal volume of 2x SDS-PAGE reducing sample buffer and boiled for 2 minutes. These aliquots were then run on 12% polyacrylamide SDS-PAGE gels according to the method of Laemmli. Gels were electroblotted to a 0.2 micron pore PVDF membrane using the buffer, 10 mM CAPS (pH 11.0), 10% methanol. Following staining of membranes for 10 minutes with 0.5%
Hepatitis C virus NS5A is a zinc metalloprotein

Ponceau-S stain in 1% acetic acid and destaining in water, bands corresponding to the major proteolytic products were excised and submitted to the Rockefeller University Proteomics Resource Center for analysis by amino-terminal Edman degradation and MALDI/TOF mass spectrometry.

Production of HCV NS5A-Domain I-Δh:

For the expression of NS5A-Domain I-Δh, 1 L of LB medium supplemented with 30 µg/L of kanamycin and 25 µg/L of chloramphenicol was inoculated from an overnight culture of *E. coli* BL21(de3) containing pET30ubi-NS5A-Domain I-Δh and pCG1 plasmids such that the cell density was at approximately OD_{600} of 0.05. Cells were grown at 37°C and 250 r.p.m until an OD_{600} of 0.55 was reached. Cells were then chilled at 4°C for 30 minutes, induced with 1 mM IPTG, and incubated for 5 hours at 25°C and 250 r.p.m. Cells were then collected by centrifugation at 6,000 x g for 10 minutes and resuspended in 20 ml of buffer C (25 mM Tris-HCl (pH8.0), 25 mM NaCl) per liter.

Cells were lysed by three passes through a cold Avestin air emulsifier at 15,000 p.s.i. Following lysis, cell extracts were clarified at 25,000 x g for 30 minutes at 4°C. Clarified extracts were fractionated by ammonium sulfate precipitation, with the bulk of NS5A-Domain I-Δh present in the 5-25% saturation pellet. Ammonium sulfate pellets were resuspended in 10 ml of buffer C and applied to a HiPrep 26/10 desalting column at a flow rate of 8 ml/min (Amersham/Pharmacia). The desalted protein was then loaded onto a HiPrep 16/20 DEAE column equilibrated with buffer C at 2.5 ml/min and eluted with a linear 500 mM NaCl gradient. Pooled DEAE elution fractions were then loaded on a 5 ml
Hepatitis C virus NS5A is a zinc metalloprotein

bed volume HiTrap IMAC column (Amersham/Pharmacia) equilibrated with buffer C at a flow rate of 2.5 ml/min. Following extensive washing with buffer C and buffer C supplemented with 1 M NaCl, NS5A-Domain I-Δh was eluted with a 25 ml linear gradient of buffer C supplemented with 250 mM imidazole. Fractions containing NS5A-Domain I-Δh were pooled and exchanged into buffer B (25 mM HEPES (pH 7.4), 25 mM NaCl) using a HiPrep 26/10 desalting column at a flow rate of 8 ml/min (Amersham/Pharmacia). Desalted protein was then concentrated via Amicon ultra centrifugal concentrators (Millipore). Concentration of purified NS5A-Domain I-Δh was determined using the Bio-Rad protein assay (Bio-Rad). Protein yields were typically between 10 and 12 mg/L of bacterial culture.

Atomic Absorption Analysis:

Atomic absorption experiments were performed by the Trace Metal Core Facility, The Department of Environmental Sciences, Mailman School of Public Health, Columbia University, New York, N.Y. All experiments were performed according to the standard operating procedures of the core facility. Background zinc level of the protein analysis buffer was performed, and this value was subtracted from all values determined for protein samples. All values reported are the average of three independent measurements of two independently prepared aliquots of each sample. Zinc values determined were based on comparison with a zinc standard calibration curve. All glass and plastic ware used for the preparation of NS5A-Δh and NS5A-Domain I-Δh was acid washed overnight in 50% nitric acid followed by copious washing in ultrapure de-ionized water. All buffers
Hepatitis C virus NS5A is a zinc metalloprotein

and reagents were prepared using the highest quality grade of chemicals commercially available. Concentrations of NS5A-Δh and NS5A-Domain I-Δh used for the calculation of protein:zinc molar ratios were determined by Bio-Rad protein assay.

Computer-aided comparative sequence analyses.

Amino acid sequences were derived from the GenPeptides database maintained by the National Center for Biotechnology Information (NCBI). Sequence alignments were produced with the Gibbs sampler (30) and the Blossum62 scoring inter-residues table (31) using the Macaw workbench (32) and were processed for presentation using GeneDoc (33). The program SEG was used to identify potential low complexity sequences (LCS) within NS5A (34). Secondary structure predictions were performed using the JPRED server (35).

Results:

Prediction and Verification of the Domain Organization of HCV NS5A

Our investigation of the NS5A protein began with computer analysis of the protein sequence with a goal of identifying secondary structure content, inter-domain boundaries, and key sequence motifs. Using a variety of secondary structure prediction programs remarkably similar results for secondary structure composition were obtained. The amino-terminal half of NS5A appeared to contain a large number of predicted β-strand secondary structural elements, a small number of α-helical structures, and a minimum of random coil sequence. In contrast, the carboxyl-terminal half of the protein
Hepatitis C virus NS5A is a zinc metalloprotein was predicted to be primarily random coil structure, with small regions of predicted regular secondary structure, primarily that of α-helices, located near the center of the NS5A sequence. Examination of sequence conservation of various regions of NS5A across HCV genotypes indicates that the amino-terminal half of the protein is the most conserved region of NS5A, with lower sequence conservation seen in the carboxyl-terminal half of the protein. These data suggested that the amino-terminal half of NS5A formed a compact globular domain, with the remaining carboxyl-terminal half of the protein having a more variable/unfolded structure. The carboxyl-terminal region of NS5A has recently been proposed to be similar to the natively unfolded class of proteins (36). These data collectively suggested that NS5A is a multi-domain protein, with a globular amino-terminal domain, a largely helical central domain, and a mostly random coil carboxyl-terminal domain. Similar conclusions were drawn from examination of a number of HCV genotype NS5A protein sequences. The identification of the inter-domain boundaries connecting these regions of NS5A was an important aspect in developing a domain organization model of this protein.

Using the program SEG, an algorithm that identifies repetitive low-complexity sequences in proteins, two regions of potential low complexity sequences (LCS) were identified in NS5A (34). These regions were designated LCS I (residues 214-249) and LCS II (residues 343-355) (Figure 1A). As repetitive, low complexity sequences are often loop regions connecting independent domains of a protein, the identification of these regions allowed the construction of a hypothetical domain organization model for NS5A (Figure 1A). The model divides NS5A into three domains based on the location of the
Hepatitis C virus NS5A is a zinc metalloprotein blocks of low complexity sequence; domain I (amino acids 1-213), domain II (amino acids 250-342), and domain III (amino acids 356-447).

During the development of this NS5A domain organization model, a system for the generation of milligram quantities of purified NS5A from *E. coli* became available (kindly provided by Dr. Craig Cameron) (37). We modified this system by substituting the NS5A coding region with the Con1 1b genotype NS5A sequence for genetic compatibility with the replicon system. Additional modifications were performed to generate NS5A in a form for amenable to our experiments, including the deletion of the majority of the amino-terminal membrane anchor, thereby allowing production of soluble NS5A (NS5A-Δh) in the absence of any detergent.

With large quantities of relatively pure NS5A-Δh protein available, we set forth to evaluate the domain organization model of NS5A by limited proteolysis. As inter-domain connecting sequences are often surface exposed flexible loops, and therefore sensitive to protease activity under mild conditions, limited proteolysis is often a valuable tool in mapping domain structure of proteins. Limited protease digestion of NS5A-Δh with immobilized TPCK treated trypsin, combined with analysis of digests by matrix assisted laser desorption time of flight mass spectrometry (MALDI-TOF) and amino-terminal Edman degradation sequencing, identified two major protease resistant products, corresponding to proteins consisting of amino acids 25-215 and 25-355. The cleavage sites at amino acids 215 and 355 are located within the predicted LCS I and LCS II, respectively (Figure 1A, grey triangles). Similar results were obtained with other proteases (chymotrypsin, Glu-C, Arg-C, and V8 protease), with major cleavage sites located within LCS I and LCS II (TT, JM, CMR, unpublished results). These data
Hepatitis C virus NS5A is a zinc metalloprotein suggest that LCS I and LCS II represent exposed, protease sensitive regions of NS5A, and likely serve as connectors between the relatively protease insensitive domains.

Expression constructs corresponding to domain I (amino acids 25-215, NS5A-Domain I-Δh) and domain I+II (amino acids 25-355, NS5A-Domain I+II-Δh) were generated and methods for the purification of these proteins were developed. As NS5A-Δh and NS5A-Domain I-Δh are used extensively in this manuscript, an SDS-PAGE gel of purified samples of these proteins is presented in Figure 1, panel D. The solubility and preliminary analysis by circular dichroism spectroscopy of purified preparations of these proteins suggests they represent soluble, independently folding domains of NS5A (data not shown). Overall, these data correlate well with the model of the domain organization of NS5A presented in Figure 1A.

**Domain I of NS5A is Predicted to Contain a Conserved Zinc Binding Site**

In attempt to identify the most conserved and, consequently, functionally most important amino acid residues, we extended our comparative sequence analysis of NS5A to include the NS5A sequences encoded in the polyproteins of the closely related GB viruses and pestiviruses. Using Gibbs sampler (1) and Blosum62 scoring table (2) in the Macaw-based (3) analysis the sequence conservation in NS5A among these viruses was found to be limited to approximately 90 amino acid residues proximal to the amino-terminus. In this conserved region, 6 positions are occupied by invariant residues including 4 cysteines (C39, C57, C59 and C80 in HCV), and 2 glycines (G60 and G70) (Schematic representation of a portion of this alignment is shown in Figure 1B).

Cysteines are known to be among most reactive residues and are involved in a variety of
Hepatitis C virus NS5A is a zinc metalloprotein activities. The presence of four conserved cysteine residues in the close proximity to each other in an intracellular protein is commonly associated with a structural tetrahedral metal (zinc) binding site, although we are not aware about other predicted or proven metal-binding proteins with a constellation of cysteine residues similar to that identified in NS5A. Based on the results of this analysis, we concluded that the HCV NS5A domain I might contain a conserved zinc-binding site composed of residues C39, C57, C59 and C80. The relative position of these residues in the domain architecture of NS5A is indicated in Figure 1 A. The cysteine-proline-cysteine tripeptide, designated CPC (C57, P58, and C59), hypothesized as the central portion of the aforementioned putative zinc-binding site, is shown. As a number of cysteine and histidine residues, aside from those predicted to be part of the putative zinc-binding site, are conserved among the various HCV genotypes, and could potentially be involved in metal ion coordination, a simple evaluation of the importance of the CPC sequence in replication and metal binding seemed an appropriate course. Based on the zinc-binding site modeling, CPC should be essential for metal ion coordination, and a less conserved, but similar CPC sequence (designated CPC2, amino acids 140-142) should be dispensable, thereby serving as an ideal control, allowing the mutation of a conserved cysteine residue (cysteine 142) that is not part of the predicted metal binding site. We used the replicon system and in vitro biochemical methods to critically evaluate the hypothesis that NS5A is a zinc binding protein, and that zinc coordination is dependent on the presence of the four cysteine residues predicted to comprise the NS5A metal binding site.

Initial Analysis of Putative Zinc Binding Site Mutants Using the Replicon System
Hepatitis C virus NS5A is a zinc metalloprotein

The ability to generate site directed mutants in the HCV replicase and analyze the effects of these mutants on RNA replication using the replicon system is a powerful method to assess the importance of predicted sequence motifs. The relative ease of generation and analysis of mutants in this system made the replicon an ideal first step in characterizing the importance of the CPC tripeptide including C57 and C59 of the predicted zinc-binding site in HCV replication. CPC2, containing a conserved cysteine residue at the second position of this tripeptide (C142) believed to be independent of the metal binding site, serves as a control. Mutations altering the CPC of each tripeptide to CPG were generated and cloned individually into the Conl/SGneo GIT replicon. The Conl/SGneo GIT replicon contains a total of three adaptive mutations, two in NS3 (E1202G, T1280I) and one in NS4A (K1846T), thereby allowing assessment of NS5A mutations in an NS5A background lacking any adaptive mutations and avoiding any potential incompatibility of these directed mutations with any NS5A adaptive changes (28). The second cysteine of each tripeptide was chosen for the initial mutagenesis, as these residues, unlike the first cysteine in the CPC2 sequence, are conserved among all HCV genotypes. Replicon RNAs bearing mutations of CPC (C59G) and CPC2 (C142G) were generated and delivered to the adapted cell line Huh-7.5 by electroporation for the determination of G418 transduction efficiency as described in materials and methods. The data generated from the analysis of the CPC and CPC2 mutants is presented in Figure 2. The parental GIT replicon yielded a transduction efficiency of \(4.98 \times 10^5\) c.f.u./\(\mu\)g (\(+/-\ 0.05 \times 10^5\)). A GIT replicon substituting three essential active site residues of the NS5B polymerase (pol+) served as a negative control, and this RNA had no G418 transduction capability. The CPC site mutation also had no G418 transduction capability,
Hepatitis C virus NS5A is a zinc metalloprotein indicating RNAs bearing this mutation do not replicate. Attempts at isolating revertants that rescue this mutation were unsuccessful. The CPC2 site mutation had a G418 transduction efficiency of $2.35 \times 10^5$ c.f.u./µg ($\pm 0.01 \times 10^5$), indicating that alteration of this site had a very mild effect on RNA replication. A mutation of CPC2 altering this site to APG (C140A, C141G) was also generated and found to have a transduction efficiency similar to that of the CPG mutation of this sequence, indicating that neither cysteine in CPC2 was essential for RNA replication (data not shown). Although these data indicate CPC is important for HCV replication, and conversely that CPC2 is of little importance, matching the expected results based on the original model, they do little to clarify the issue of whether NS5A is actually a zinc-binding protein. As the CPC mutation could alter other aspects of the function of NS5A in the replicase, and possibly have nothing to do with metal ion coordination, a more direct method for determining the zinc content of the wild type and mutant NS5A proteins was needed.

**NS5A-Δh is a Zinc Binding Protein, and Zinc Coordination is Dependent on the CPC Tripeptide**

One method for determining the zinc binding capability of a protein is to directly measure the zinc content of a purified preparation of the protein by atomic absorption spectroscopy. The availability of large quantities of NS5A-Δh protein made atomic absorption spectroscopy an ideal method for evaluating NS5A zinc content. The zinc content of two independent preparations of wild type NS5A-Δh, the CPC site mutation (C59G), and the CPC2 site mutation (C142G) was determined using triplicate measurements of each sample. The zinc content of each sample (approximately 660 µg protein per measurement) was determined, and following subtraction of background
Hepatitis C virus NS5A is a zinc metalloprotein

metal contamination from a buffer only blank, values were converted to a ratio of moles of zinc to moles of protein. Typical raw data of zinc content were quite consistent, with variations of less than 2% between samples representing independent preparations of the same protein. As an example, wild type NS5A-Δh raw values for zinc content for approximately 660 µg of protein were; 987, 979, and 978 µg/L for preparation 1, and 973, 971, and 970 µg/L for preparation 2. The data of the ratios of moles zinc to moles protein determined for wild type NS5A-Δh, NS5A-Δh CPC (C59G), and NS5A-Δh CPC2 (C142G) are presented graphically in Figure 3. The values shown are calculated from the average zinc content from the six measurements collected for each protein. The ratios of moles of zinc per moles of protein determined were 0.897 for wild type NS5A-Δh, 0.170 for the NS5A-Δh CPC mutation, and 0.766 for the NS5A-Δh CPC2 mutation. These data indicate that the wild type NS5A-Δh protein is a zinc binding protein with one zinc atom coordinated per protein molecule. Similarly, the NS5A-Δh CPC2 site mutant coordinated approximately one zinc atom per protein molecule. The value measured for these proteins, 0.897 and 0.766, were very close to the ideal value of one mole of zinc per mole of protein, especially when the purity of the preparations of NS5A protein used for these analyses are considered (approximately 85%). Additional reasons for this variation from the ideal zinc:protein ratio of one can be imagined if the fact that little is known about the percentage of metal binding competent protein present in our preparations, with improperly folded, aggregated, or otherwise metal coordinating incompetent protein potentially in these samples. The NS5A-Δh CPC tripeptide mutant had a zinc:protein ratio of 0.170, indicating the point mutation of C59 to glycine impaired the zinc binding capability of NS5A. The value of 0.170 observed for the ratio of zinc:protein for the CPC
Hepatitis C virus NS5A is a zinc metalloprotein mutant, although clearly lower than that observed for non-mutant NS5A or CPC2 mutant NS5A, was not zero. Perhaps the CPC tripeptide mutant retains some ability to bind zinc, albeit at a much lower occupancy than non-mutant NS5A. Nonetheless, the data indicate the presence of a zinc ion associated with NS5A-Δh via the CPC tripeptide. Furthermore, when combined with the data from the replicon system (Figure 2), it is clear that zinc coordination is required for RNA replication. As the NS5A-Δh CPC mutation analyzed contains a single cysteine to glycine change, an additional three metal ion coordinating residues likely exist in NS5A allowing zinc binding in the classical tetrahedral fashion. Although our original model of the NS5A zinc-binding site suggested the importance of C39, C57, C59, and C80 in metal ion coordination, no experimental evidence existed to fully validate this model.

As the NS5A proteins used in these experiments contained a carboxy-terminal polyhistidine affinity purification tag, a known metal ion coordination sequence, experiments were repeated using proteins from which this tag was proteolytically removed following purification. The presence of the polyhistidine tag sequence did not alter the zinc content observed for NS5A-Δh (data not shown). Analysis of purified NS5A proteins with the amino-terminal amphipathic α helix did not affect the observed results (data not shown). Although mutant and wild type NS5A proteins were of similar solubility when expressed in E. coli, the possibility existed that mutation of the CPC site led to a global defect in NS5A folding, thereby eliminating zinc binding by an indirect method, as opposed to a more direct role for this sequence in metal ion coordination. Circular dichroism spectroscopy was used to confirm NS5A-Δh CPC had similar
Hepatitis C virus NS5A is a zinc metalloprotein secondary structure content to that observed for wild type NS5A-∆h and possessed no global defect in protein folding (data not shown).

**Localizing the NS5A Zinc Binding Site to Domain I**

The model of the NS5A metal coordination site (C39, C57, C59, and C80) suggested the zinc binding phenotype of NS5A should localize to the domain I region of NS5A. To verify this prediction, preparations of NS5A-Domain I-∆h were submitted to atomic absorption spectroscopy to determine the zinc content of this portion of NS5A. The data of the ratio of moles zinc to moles protein determined for NS5A-Domain I-∆h is presented graphically in Figure 4 with the ratios for wild type NS5A-∆h and the NS5A-∆h CPC mutant shown as positive and negative standards, respectively. The ratio of moles of zinc per moles of protein determined for NS5A-Domain I-∆h was 0.822. NS5A-Domain I-∆h, similar to the NS5A-∆h protein, coordinates one zinc atom per protein molecule. This experiment demonstrated that the zinc binding of NS5A is localized to domain I, with domains II and III having no observable effect on metal ion coordination.

**The NS5A Zn-binding Site is Vital for HCV Replication**

Based on the observation that the zinc-binding site of NS5A is localized to domain I, a search to identify the residues within this domain that were responsible for zinc coordination was conducted. The observation that a mutation that disrupted zinc binding was lethal for RNA replication suggested that site directed mutagenesis of possible metal coordinating residues and analysis with the replicon system could be used
Hepatitis C virus NS5A is a zinc metalloprotein as a method to determine the importance in replication of putative zinc binding site components. The model of the NS5A metal coordination site, although implicating C39, C57, C59, and C80 as important targets for mutagenesis, could not eliminate the potential importance of other amino acids for metal binding, and a more conservative mutagenesis approach altering all of the conserved classical zinc coordinating cysteine and histidine residues within domain I of NS5A was undertaken. Potential cysteine and histidine targets for mutagenesis were selected based on conservation of these residues in NS5A sequences among the six major HCV genotypes, as well as the location of these residues in relation to predicted secondary structure elements. The ten sites selected for analysis are listed in Figure 5. Of these mutants, four (H66, H85, C98, and C190) had transduction efficiencies close to that of the parental replicon. These mutants were not analyzed further. Two mutants (H158, and H207) were of reduced transduction efficiency compared to the parental replicon. The H158 mutant, in addition to having a lower transduction efficiency, generated smaller G418 resistant colonies than either the parental replicon or the other mutants. The basis of this phenotype was not explored further. Four of the mutations generated were lethal for RNA replication, with no G418 resistant colonies observed. These mutants are highlighted in bold text in Figure 5, and include C39, C57, C80, and H128. All of the cysteine residues in this set of lethal mutations, like previously characterized C59 residue (Fig. 2), are part of the predicted tetrahedral zinc-binding site. The basis of the lethal effect of the H128A mutation, not predicted to be a component of the NS5A zinc-binding site, is currently being studied. The obtained results are internally consistent with the predicted model of the NS5A zinc-binding site and further indicate that the zinc binding is vital for RNA replication. The identification of a
Hepatitis C virus NS5A is a zinc metalloprotein total of five replication essential cysteine and histidine residues within domain I, although strongly suggesting the predicted zinc-binding site was correct, could not unambiguously identify the residues responsible for the tetrahedral coordination of a zinc atom.

**Mutations at the Predicted Metal Coordination Residues Diminish NS5A Zinc Content.**

To analyze the NS5A zinc-binding site model further, we employed atomic absorption spectroscopy for the characterization of preparations of full-length NS5A-Δh proteins bearing mutations that were lethal or impaired for RNA replication. These included the individual C39A, C80G, H128A, H158A, H207G, and C57G mutations. The zinc content values determined for these proteins were converted to a ratio of moles of zinc to moles of protein (Figure 6). The ratio of moles of zinc to moles of protein for wild type NS5A-Domain I-Δh is shown as a positive control. The low zinc:protein ratio of the C39 mutation (0.131), C57, the first cysteine residue in the CPC tripeptide (0.134), C59, the second cysteine in the CPC tripeptide (data from figure 3), and the C80 mutation (0.208) indicated these residues were likely involved in metal ion coordination. It is important to note that mutation of these residues was lethal for RNA replication (see Figure 5). In contrast, the H128 mutation was lethal for RNA replication, but had a zinc:protein ratio (0.765) close to that of the wild-type NS5A-Δh control. The nature of the defect in RNA replication for this mutant is currently under investigation, but this data suggests this residue is not involved in zinc coordination. The two mutations that showed reduced RNA replication in figure 5, H158 and H207, had zinc:protein ratios similar to the wild type control (0.859 for H128, and 0.905 for H207), suggesting they
Hepatitis C virus NS5A is a zinc metalloprotein

have no role in zinc binding. Taken together, the RNA replication and zinc content data suggest that NS5A coordinates a single zinc atom via residues C39, C57, C59, and C80, and that this metal ion binding is likely required for RNA replication. These data provide genetic and biochemical evidence confirming the original prediction of the NS5A zinc-binding site.

**Discussion:**

The data presented herein provides the first domain organizational model for the HCV NS5A protein and represents one of the first views of NS5A as an individual protein and not merely an unknown component of the HCV replicase. The domain model of NS5A has led to the development of a number of protein expression constructs of small soluble domains of NS5A, allowing a more detailed analysis of the function of the various regions of this protein. Particularly, the presence of a conserved zinc-binding motif within domain I was evaluated by a combination of genetic and atomic absorption spectroscopy experiments. The data indicate that NS5A is a zinc metalloprotein, coordinating a single zinc atom per molecule of NS5A via four essential cysteine residues within domain I. Although the functional implications of zinc binding have not yet been evaluated, the presence of a novel zinc-binding motif in NS5A provides insight to this protein. In the following paragraphs we discuss our recent observations on NS5A in the context of the body of HCV literature to provide a critical interpretation of the data and put forth several important concepts, notably the domain organization of NS5A, the surface accessibility of potentially important sequences, the potential for inter-domain contacts, and the importance of a zinc atom in NS5A structure and function.
Hepatitis C virus NS5A is a zinc metalloprotein

**NS5A Domain Organization:**

Based on the results of comparative sequence analysis and limited proteolysis of purified protein, we have proposed a domain organization model for HCV NS5A that includes domain I (amino acids 1-213), domain II (amino acids 250-342), and domain III (amino acids 355-447).

The region of NS5A proposed to constitute domain I represents the most conserved portion of this protein among the HCV genotypes. Despite all appearances suggesting a folded, globular protein domain, homology based sequence searches have failed to identify any proteins related to NS5A (with the exception of the related pestivirus NS5A protein). Scanning domain I, and indeed the entire NS5A protein, for motifs that might imply a function for this enigmatic protein have been futile, with the exception of the putative zinc-binding site described in this manuscript. At least one definitive function has been attributed to sequences within domain I, that of a membrane anchor (26,27). Although the bulk of the work described in this manuscript used NS5A proteins lacking the amino-terminal membrane anchoring helix, similar results have been observed with constructs that retain this sequence. These data suggest that the amino-terminal helix is likely a separate entity from the largely globular, hydrophilic domain I region. The membrane anchor sequence is mentioned in this discussion of domain I because the proximity of this helix to domain I suggests that the amino-terminal region of NS5A is likely close to the membrane. Interactions with karyopherin β3 (38), apolipoprotein A1 (39), phosphoinositide-3-kinase p85 subunit (40), hTAFII32 (41), homeodomain protein PTX1 (42), and the La autoantigen (43) have been mapped to
Hepatitis C virus NS5A is a zinc metalloprotein

various portions of domain I. The proposed proximity of domain I to the cellular membrane is compatible with the properties of a number of these domain I interacting proteins that are either membrane associated or involved in membrane trafficking. Many of these interactions are intriguing, but are of unknown relevance to HCV biology. The goal of future research will be to identify a function for NS5A, and it seems likely that the answer would lie in the conserved, well-ordered region of the protein such as domain I.

The proposed domain II region of NS5A is of lower sequence conservation among HCV genotypes than domain I. This region contains a number of predicted secondary structural elements, primarily α-helical, that suggest it has a defined structure. Although no function has been attributed to this region, domain II and the preceding LCS I region have an important impact on RNA replication efficiency as a large number of NS5A adaptive mutations map to this area (11,12). Although the prediction is that the majority of these adaptive mutation sites are within LCS I, one adaptive mutant of particular interest, the Δ47 deletion described previously (deletion of amino acids 235-282) (11), clearly extends into the predicted domain II region of NS5A. This adaptive mutant suggests that at least the amino-terminal portion of domain II may play a role in replication. Whether the adaptive phenotype of this mutant is due to alteration of the LCS I sequence or that of domain II remains to be evaluated. Domain II contains a region of NS5A that has been referred to as the interferon sensitivity determining region (ISDR, amino acids 237-276), so named for a possible correlation between hyper-mutation in this sequence and patient response to interferon therapy (44). This region and 26 amino acids carboxy terminal to the ISDR have been shown to be essential for interaction with the
Hepatitis C virus NS5A is a zinc metalloprotein

IFN-induced double stranded RNA dependent protein kinase (PKR) (45-51; see (52) for review). Another recent publication has described an interaction of residues within domain II with the Bax protein and postulated the importance of this interaction in anti-apoptotic pathways (53). The observation that domain II may be involved in interferon response, PKR dependent cellular signaling, and apoptosis is interesting, but unfortunately, the complex interplay between NS5A and the interferon pathways is not yet clear enough to understand the role of domain II in these processes.

The region of NS5A hypothesized to comprise domain III is the most variable among HCV genotypes. This region contains large regions predicted to be random coil structures by various secondary structure prediction programs. The ability to generate functional NS5A proteins containing in frame fusions with other proteins (i.e. green fluorescent protein) within domain III in the replicon system highlights the sequence flexibility of this region (54). This region does contain several regions of sequence that are absolutely conserved among all HCV sequences suggesting some selective evolutionary pressure exists for maintenance of this region. A careful evaluation of the regions of domain III required for HCV RNA replication is warranted.

One area of particular interest in evaluating the domain model of NS5A is the proposed surface exposure of the inter domain connector sequences (LCS I and LCS II) in relation to the HCV literature. It is important to note that designating a sequence “low complexity” does not imply that this sequence has no function. Indeed, examination of sequence alignments of various genotypes of HCV shows these low complexity sequence blocks tend to have considerable sequence conservation, thereby suggesting some importance to NS5A function and HCV biology. LCS I overlaps the region where
Hepatitis C virus NS5A is a zinc metalloprotein

numerous NS5A cell culture replicon adaptive mutations, some capable of increasing RNA replication 10,000 fold, have been described (11). It is enticing to hypothesize that the exposed residues comprising the LCS I region are a loop region connecting domain I to domain II, and that this loop region is involved in interaction with viral and cellular components of the replicase in a regulatory fashion or modulating contacts between the domains of NS5A. Similarly to LCS I, the sequence at LCS II is conserved among HCV isolates, suggesting a selective pressure for the maintenance of this region. Several publications suggest this proline-rich region of NS5A contains a class II polyproline motif like sequence (amino acids 350-356) involved in interaction with Grb2 and downstream cellular signaling cascades (55-57). This sequence has also been shown to interact with the cellular protein amphiphysin II (58). The demonstration LCS I and LCS II are exposed on the surface of NS5A, and available for potential interactions with viral and cellular factors, is an important observation. These ideas clearly need to be evaluated critically, but the data provides some intriguing hints at NS5A function and interactions.

**NS5A is a Zinc Metalloprotein:**

Comparative sequence analysis of NS5A suggested this protein might contain an unusual zinc coordination motif. Initial analysis of mutations of the central CPC tripeptide sequence of the HCV NS5A zinc-binding site using the replicon system indicated that the conserved CPC was required for RNA replication. Atomic absorption analysis of purified NS5A-Δh demonstrated that this protein contained one zinc atom per molecule, and that this zinc binding was dependent on cysteine residues within the CPC tripeptide.
Hepatitis C virus NS5A is a zinc metalloprotein sequence. Through further use of the replicon system and atomic absorption spectroscopy, the likely residues comprising the HCV zinc-binding site have been verified. These residues include C39, C57, C59, and C80, amino acids originally predicted to comprise the zinc-binding site. Mutation of any of these residues led to a lethal phenotype in the replicon system and resulted in a loss of zinc ion coordination. These residues are conserved across all of the HCV genotypes (Figure 1, panel C) with the exception of the cysteine at position 39 in the published genomic sequence for genotype 4a (tryptophan is present at this position in 4a, accession number Y11604). However, a recent analysis of several independent NS5A clones generated from RT-PCR products from genotype 4a infected chimpanzee serum demonstrated a cysteine residue at this position, suggesting tryptophan at this position in the Y11604 4a sequence may be a sequencing error (Hongqing Fan and CMR, unpublished results). As no infectious clone exists for HCV genotype 4a, it is difficult to interpret this disparity in sequence by evaluating a known infectious clone sequence.

Analysis of the mapped HCV zinc-binding site in relation to the secondary structure predictions of NS5A provides some interesting insights into the possible structure of this region of NS5A. The secondary structure predictions for the region of NS5A encompassing the zinc-binding motif indicate the presence of four short β-strand structures (35). These secondary structure elements appear to be conserved among all HCV genotype NS5A sequences. The arrangement of the four cysteine residues shown to be involved in zinc binding in relation to the predicted β-strand structures is also conserved among NS5A sequences. The first cysteine of the zinc-binding motif, C39, lies at the amino terminus of the first predicted β-strand (β-strand 1, amino acids 43-48). The
Hepatitis C virus NS5A is a zinc metalloprotein
two cysteines of the CPC tripeptide (C57 and C59) lie in a short loop connecting β-
strands 2 (amino acids 51-56) and 3 (amino acids 62-78). The final cysteine of the zinc
motif is located at the carboxyl end of β-strand 4 (amino acids 71-76). The position of
these cysteines relative to predicted secondary structure elements is shown in Figure 7,
panel A. Analysis of numerous crystal structures containing structural zinc ions suggests
the cysteine side chain sulfur atom to zinc ion distance is approximately 2.3 Å (59). This
sulfur to zinc distance places constraints on the arrangement of the predicted secondary
structure elements in the zinc-binding region of NS5A. This relatively fixed distance,
combined with the short inter-strand loop regions, strongly suggests the four β-strands
comprising the zinc-binding motif are arranged in manner to generate a 4-stranded anti-
parallel β-sheet structure.

A cartoon schematic of our hypothetic model of this region of NS5A is shown in
Figure 7, panel B. These observations provide the first tertiary contacts, and constraints,
demonstrated for NS5A structure. The data generated to date fit the model
quite well and the conservation of the proposed zinc coordinating cysteines and
secondary structure elements surrounding these cysteines suggests that these
components of the zinc-binding domain are essential. The validity of this model is
perhaps best supported by the convergence of the original bioinformatics model of the
zinc-binding site and the results of the subsequent genetic and biochemical analysis.
Nonetheless, since this model relies heavily on predicted secondary structures it may not
accurately represent the real structure of NS5A. Although we cannot formally exclude the
possibility exists that one or more ligands coordinating the zinc ion may be water-bridged
interactions, with our lethal cysteine mutants representing some other global folding
Hepatitis C virus NS5A is a zinc metalloprotein defect. Preliminary circular dichroism analysis of NS5A with and without mutations in the CPC tripeptide suggested both proteins were folded, but this analysis is yet to be performed on all of the possible zinc-coordinating ligand mutants (data not shown).

Involvement of other amino acids, such as glutamate and aspartate residues, that were not analyzed in this study, in the zinc binding activity of NS5A remain a possibility, although this seems unlikely as these residues are not conserved in the NS5A sequence.

An important question that remains to be answered is in what aspect of NS5A function is zinc binding involved. Unfortunately, no function has yet been attributed to the NS5A protein, complicating the evaluation of the importance of zinc binding. The analysis of the role of zinc binding for NS5A function is further complicated by the large number of functions attributed to zinc binding in proteins. Zinc coordination in proteins has been demonstrated to be involved in a wide range of enzymatic activities, nucleic acid interactions, metabolic regulatory functions, protein-protein interactions, and structural fold maintenance (60). Sequence database and motif database searches were unsuccessful at identifying proteins that contain a zinc coordination motif similar to that mapped for HCV NS5A. The HCV zinc-binding site does not appear to posses the characteristics of any of the documented classes of enzymatic functions associated with zinc binding proteins (61). Additionally, this site does not appear to be similar to any of the known classes of zinc dependent nucleic acid interaction motifs, although the idea of NS5A interacting with viral RNA via a zinc motif is attractive (62,63). The coordination of the NS5A zinc atom by four cysteine residues strongly suggests this is a structural metal ion necessary for NS5A folding or stability (61). Limited CD data suggests that NS5A proteins bearing mutations at the CPC site are of lower thermal stability than non
Hepatitis C virus NS5A is a zinc metalloprotein mutant proteins, although a more detailed analysis of this observation has yet to be performed. A brief review of well-characterized four cysteine structural metal-binding sites has also failed to find a motif similar to that observed for NS5A. The majority of these sites fit the general template CxxC_{17-30}CxxC which is inconsistent with the NS5A zinc site, C_{17}CPC_{20}C. Structural metal sites tend to have lower sequence conservation and are more variable than zinc sites involving a catalytic function, making sequence based comparisons outside the context of a known protein structure/fold complicated. Focusing on the central CPC portion of the HCV zinc-binding site during the search for similar metal binding sites was somewhat more successful. A well-documented CxC motif exists in the HCV NS3 protease domain (62). This NS3 zinc-binding site is a structural metal required for protease activity that is similar to the zinc-binding site in the poliovirus 2A protease. Unfortunately, this site is also quite unlike the HCV zinc-binding site, with a sequence of CxC_{35}C_{3}H. The overall spacing and presence of a histidine metal coordinating zinc via a water molecule bridge makes direct comparisons between this site and the NS5A site questionable. The zinc-binding site mapped for HCV NS5A is clearly unlike other published zinc coordination motifs, suggesting this site may represent a novel four β-strand metal-ion coordination fold. The NS5A zinc-coordination motif is somewhat reminiscent of the common zinc finger three-stranded antiparallel sheet β-ribbon architecture, although the presence of four predicted β-strands and an altered cysteine residue spacing in NS5A is clearly different from this class of proteins (63). The determination of the three-dimensional structure of NS5A will likely be essential for the classification of its Zn-binding domain.
Hepatitis C virus NS5A is a zinc metalloprotein

In summary, we have presented a domain organization model of the HCV NS5A protein based on sequence analysis and limited proteolysis of purified protein. Comparative sequence analysis of NS5A suggested the presence of a zinc ion coordination site, and this site has been verified using a combination of the HCV replicon system and atomic absorption spectroscopy. This zinc-binding site appears to be a novel four cysteine motif that is common to the NS5A proteins of the Hepacivirus and Pestivirus genera, thus supporting the relationship of these genera in the Flaviviridae family. The coordination of a zinc ion, combined with secondary structure elements surrounding the coordinating residues, imparts important structural constraints on NS5A. Unfortunately, no function could be attributed to the zinc ion coordinated by NS5A, largely due to the lack of a defined function for this protein. The challenge for future research clearly lies in the determination of a function for NS5A and/or the determination of a high-resolution structure of this protein.

Acknowledgements:

The authors wish to thank Dr. Craig Cameron for the original NS5A bacterial expression system reagents and methods, Ms. Vesna Slavkovich for atomic absorption method development, Mr. Jack Hietpas for general laboratory support, and Mr. Hongqing Fan for NS5A genotype 4a sequencing. Additionally, we wish to thank Dr. Matthew Evans, Dr. Brett Lindenbach, Dr. Shihyun You, Dr. Ivo Lorenz, and Dr. Annick Gauthier for scientific discussions related to this manuscript. T.T. was supported, in part, by fellowships from the Charles H. Revson Foundation for Biomedical Research and the National Institutes of Health Ruth L. Kirschstein National Research Service Award (5F32
Hepatitis C virus NS5A is a zinc metalloprotein

AI51820-03) granted through the National Institute of Allergy and Infectious Disease. J.M. was supported as a Merck Fellow of the Life Sciences Research Foundation. Additional financial support for this work came from grant 5 R01 CA57973-12 from the National Institutes of Health and the Greenberg Medical Research Institute (C.M.R.).

Figure Legends:

Figure 1. Domain Organization Model of NS5A and Identification of Potential Zinc Coordination Motif. Panel A. Model of the domain organization of NS5A. The NS5A protein is divided into three domains based on the presence of two low complexity sequence blocks (designated LCS I and II) predicted to be inter-domain connecting loops. The region believed to comprise domain I (amino acids 1-213) contains the amino-terminal membrane anchoring helix, as well as a potential metal ion coordination motif. Domain II (amino acids 250-342) and domain III (amino acids 356-447) comprise the carboxyl-terminal half of NS5A. Trypsin cleavage sites from the limited proteolysis of NS5A-∆h are shown as grey triangles within the LCS regions. These cleavages correspond to digestion of NS5A-∆h at residues 215 and 355, generating stable protease resistant products consisting of domain I and domain I + domain II. Sequence analysis suggests the presence of a zinc-binding motif in domain I, designated CPC (amino acids 39, 57-59, and 80). Panel B. Putative zinc-binding site is conserved in NS5A of HCV, HGBV and pestiviruses. Shown is an alignment of a portion of the most conserved part of NS5A of a representative set of HCV, HGBV and pestiviruses. According to the Macaw workbench (see Methods), an alignment block around the CXC tripeptide is the most conserved in NS5A. Residues shaded in gray represent the putative zinc coordinating
Hepatitis C virus NS5A is a zinc metalloprotein residues of the zinc-binding site. Conserved glycine residues within this region are shown with asterisks. Numbering corresponds to the amino acid sequence of the Con1 isolate of HCV. Virus names and respective NCBI protein IDs or Genebank accession numbers: HCV-Con 1, Hepatitis C virus 1b genotype sequence from replicons and proteins used in this manuscript (AJ238799); HCV-H77, Hepatitis C virus 1a H77 infectious clone (AF009606); HGBV-B, Human GB virus B (9628102); HGBV-C, Human GB virus C/hepatitis G virus (4426796); HCV-H, HCV isolate H (329738); BVDV-N, BVDV strain NADL (9626650); CSFV-B, classical swine fever virus, strain Brescia (130457). Panel C. Conservation of the putative HCV zinc coordination site residues across the major HCV genotypes. The following sequences were used for alignments; 1a H77 infectious clone (AF009606), 1b Con1 replicon sequence (AJ238799), 2a infectious clone (AF177036), 2b genome sequence (AF238486), 3b genome sequence (D4937), 4a NS5A sequences (internal lab sequencing, Hongqing Fan, unpublished data), 5a genome sequence (Y13184), and 6a genome sequence (Y12083). Panel D. Gel representing purified NS5A-Δh and NS5A-Domain I-Δh used in this manuscript.

Figure 2. Analysis of Mutations of Potential Zinc-Coordination Motif Using the HCV Replicon System. Transduction efficiencies of wild type (parental), zinc coordination site mutant (C59G, CPC) and the control predicted non-zinc binding cysteine mutant (C142G, CPC2) following electroporation into Huh7.5 cells and selection with G418 for a period of two weeks. Cells were plated in a log series dilution on replication deficient (pol-) feeder cells as described in material and methods. Transduction efficiencies, in units of colony forming units per microgram of RNA, are shown. These values represent
Hepatitis C virus NS5A is a zinc metalloprotein that average of three independent electroporations of three independent preparations of the RNAs tested, and the calculated standard deviations are shown in parentheses. The parental replicon RNA, Con1/SGneo GIT is shown (GIT). The amino acid substitution present in each of the mutants is indicated, for example, CPC contained a mutation of cysteine 59 to glycine (C59G). A replication defective RNA (pol-) is shown as a negative control.

Figure 3. Determination of the Zinc Content of NS5A-Δh and CPC tripeptide Mutants by Atomic Absorption Spectroscopy. Graphical representation of data from the analysis of the zinc content of NS5A-Δh and NS5A-Δh bearing mutations of the CPC (C59G) and CPC2 (C142G) sites by atomic absorption spectroscopy. The data presented is the average of three measurements from two independent preparations of each protein, for total of six values. The raw data (see text section for representative values) was converted to moles of zinc per mole of protein. Errors were too small to represent graphically as error bars, with maximum standard deviations of the calculated ratios of less than 0.002.

Figure 4. Determination of the Zinc Content of NS5A-Domain I-Δh by Atomic Absorption Spectroscopy. Graphical representation of data from the analysis of the zinc content of NS5A-Domain I-Δh by atomic absorption spectroscopy. Data collection and analysis was performed as described for Figure 3. The zinc:protein ratios of NS5A-Δh and NS5A-Δh bearing the CPC mutation (C59G) are shown as positive and negative controls, respectively.
Hepatitis C virus NS5A is a zinc metalloprotein

Figure 5. Analysis of Replication Fitness of Potential Zinc Coordinating Residues. Transduction efficiencies of wild-type (parental) and potential zinc-coordination residue mutant RNAs following electroporation into Huh7.5 cells and selection with G418 for a period of three weeks. Cells were plated in a log series dilution on replication deficient (pol-) feeder cells as described in material and methods. Transduction efficiencies, in units of colony forming units per microgram of RNA, are shown. These values represent that average of three independent electroporations of three independent preparations of the RNAs tested, and the calculated standard deviations are shown in parentheses. Replication defective mutant RNAs are shown in bold-face type. The first cysteine residue of the CPC tripeptide is also shown (CPC, C57G). The amino acid substitution present in each of the mutants is indicated in parentheses as described in figure 2, for example, C39A contained a mutation of cysteine 39 to alanine (C39A). A replication defective RNA (pol-) is shown as a negative control.

Figure 6. Evaluation of the Zinc Content of Lethal and Replication Impaired Mutants of Potential Zinc Coordinating Residues. Graphical representation of data from the analysis of the zinc content of NS5A-Δh putative zinc coordination site mutants by atomic absorption spectroscopy. The data presented is the average of three measurements from each protein. The raw data was converted to moles of zinc per mole of protein. Errors were too small to represent graphically as error bars, with maximum standard deviations of the calculated ratios of less than 0.002. The zinc:protein ratios of NS5A-Δh and NS5A-Δh bearing the CPC mutation (cysteine 59 to glycine) are shown as positive and negative
Hepatitis C virus NS5A is a zinc metalloprotein

controls, respectively. Proteins that did not coordinate zinc to wild-type levels included
cysteine 39, cysteine 57, cysteine 59, and cysteine 80.

Figure 7. Panel A. Secondary structure prediction of region of NS5A surrounding zinc-binding site. HCV, 1b genotype Con1 isolate NS5A sequence (AJ238799) from amino acids 35 to 85 are shown. Cysteine residues comprising the zinc-binding site are shaded in gray. Black arrows represent the JPRED (35) predicted β-strands within this region of NS5A, which are numbered according to the discussion section of the manuscript. Panel B. Hypothetical model of the organization of the HCV zinc coordination site in relation to surrounding predicted secondary structure elements. The relatively fixed distance of approximately 2.3 Å for cysteine sulfur atom to the coordinated zinc atom seen in structural metal binding sites suggests an anti-parallel arrangement of β-strands. The individual β-strands have been numbered to correlate with their discussion in the text and panel A.
Hepatitis C virus NS5A is a zinc metalloprotein

References:

Hepatitis C virus NS5A is a zinc metalloprotein

37. Luyun Huang, E. V. S., Michelle R.S. Hargittai, Suresh D. Sharma, Mehul Suthar, Kevin D. Raney, and Craig E. Cameron. (2004) *Protein Expression and Purification In Press*
Hepatitis C virus NS5A is a zinc metalloprotein

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Dilution Series (cells plated)</th>
<th>Transduction (cf.u./ug)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parental</td>
<td><img src="image" alt="Cells" /></td>
<td>4.98 x 10^5 (+/- 0.05)</td>
</tr>
<tr>
<td>C59G (CPC)</td>
<td><img src="image" alt="Cells" /></td>
<td>0</td>
</tr>
<tr>
<td>C142G (CPC2)</td>
<td><img src="image" alt="Cells" /></td>
<td>2.35 x 10^5 (+/- 0.01)</td>
</tr>
<tr>
<td>Pol-</td>
<td><img src="image" alt="Cells" /></td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 3

The graph shows the moles of zinc per moles of protein for different samples:

- **NS5A-Δh**: A significantly higher value, indicating a high ratio of zinc to protein.
- **C59G CPC**: A lower value compared to NS5A-Δh, showing a moderate ratio.
- **C142G CPC2**: The highest value among the three, indicating a high ratio of zinc to protein.

The y-axis represents the ratio of moles of zinc to moles of protein, ranging from 0 to 1.
Figure 4

The graph shows the ratio of moles of zinc to moles of protein for different samples. The samples are:

- **NS5A-Δh**
- **C59G CPC**
- **NS5A-Δh Domain I**

The graph indicates that the ratio for **NS5A-Δh** is significantly higher than for the other two samples.
<table>
<thead>
<tr>
<th>Mutation</th>
<th>Dilution Series (cells plated)</th>
<th>Transduction (c.f.u./ug)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parental</td>
<td>10^6</td>
<td>10^5</td>
</tr>
<tr>
<td>C39A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C57G (CPC)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H66G</td>
<td></td>
<td>10^6</td>
</tr>
<tr>
<td>C80G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H85A</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>C98A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H128A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H158A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C190G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H207G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pol -</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 6

The bar chart shows the ratio of moles of zinc to moles of protein for various proteins and mutations:

Figure 7
The NS5A protein of hepatitis C virus is a zinc metalloprotein
Timothy L. Tellinghuisen, Joseph Marcotrigiano, Alexander E. Gorbalenya and Charles M. Rice

J. Biol. Chem. published online August 31, 2004

Access the most updated version of this article at doi: 10.1074/jbc.M407787200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at
http://www.jbc.org/content/early/2004/08/31/jbc.M407787200.citation.full.html#ref-list-1