NMR STUDIES OF THE GLYCOPROTEIN CYTOPLASMIC TAILS OF HANTAVIRUS AND CRIMEAN CONGO HEMORRHAGIC FEVER VIRUS

By

©2011
David Fernando Estrada

B.S., Kansas State University, Manhattan, KS, 1999

Submitted to the Department of Molecular Biosciences and the
Faculty of the Graduate School of the University of Kansas
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Committee members:

_____
Chairperson – Roberto N. De Guzman

_____
Mark Richter

_____
Kathy Supernant

_____
Audrey Lamb

_____
Jennifer Laurence

_____
David Davido

Date Defended: 20 April, 2011

The Dissertation Committee for David Fernando Estrada certifies that this is the approved version of the following dissertation:

NMR STUDIES OF THE GLYCOPROTEIN CYTOPLASMIC TAILS OF HANTAVIRUS AND CRIMEAN CONGO HEMORRHAGIC FEVER VIRUS

Committee members:

Chairperson – Roberto N. De Guzman

Mark Richter

Kathy Supernant

Audrey Lamb

Jennifer Laurence

David Davido

Date approved: 26 April, 2011

I dedicate this work to my loving and patient wife Amanda and our three children: Katrina, Isabella, and Xavier

Abstract

The Bunyaviridae family of viruses is a diverse grouping of approximately 350 members that is present throughout the world. Viruses range from the innocuous to the severely pathogenic and impact not only public health, but crop production and the live-stock industry as well. Collectively, Bunyaviruses represent a serious health and economic risk to not only the United States, but also to Europe, Asia, and Africa.

Humans infected by the most pathogenic Bunyaviruses, such as *Hantaviruses* and *Nairoviruses*, can have a collection of symptoms that include hemorrhagic fevers, pulmonary edema, severe ecchymosis of the extremities, and in the most serious cases, respiratory failure. Outbreaks of the *Hantavirus* in the U.S. in 1993 and the Crimean Congo Hemorrhagic Fever virus in Turkey in 2005 have had mortality rates as high as 40%.

Bunyaviruses are enveloped anti-sense RNA viruses that contain three genomic segments (Small, Medium, and Large). The M segment produces a glycoprotein precursor that is post-translationally cleaved into a Gn and Gc glycoproteins. Gn and Gc form a heterodimer with the ectodomain representing the surface spike proteins, and the a cytoplasmic domain that extends into the interior of the assembled virus. The Gn cytoplasmic tail in particular is present in a variety of lengths in different Bunyaviruses. Recent studies strongly suggest the Gn tail participates in assembly of the mature virion.

Prior to this dissertation, the atomic structure for a Bunyaviridae Gn tail was unknown. This work describes the solution structures of two such cytoplasmic domains in the Andes (*Hantavirus*) and the Crimean Congo Hemorrhagic Fever virus (*Nairovirus*), as well as characterize the structure of a third domain from the non-pathogenic Prospect Hill Virus (*Hantavirus*). As it turns out, a conserved repeating CCHC motif in these domains folds into a unique arrangement of back-to-back ββα type zinc fingers. Whereas classical ββα zinc fingers typically form an extended "beads on a string" arrangement, the Bunyavirus-type zinc fingers form a compact domain in which each zinc finger folds interdependently. Furthermore, while the core zinc finger fold is conserved in different Bunyaviruses, the distribution in the conserved surface electrostatics can vary between genera. Lastly, this work contains preliminary electrophoretic mobility shift assays that suggest that these cytoplasmic zinc finger domains of the Bunyaivurs Gn cytoplasmic tails may contribute to a protein-RNA interaction, thus providing further evidence for a role in binding the ribonucleocapsid complex during assembly.

The work presented in this dissertation has expanded our knowledge of the Bunyavirus life cycle by bringing attention to the structure and the function of these conserved Gn cytoplasmic domains. Understanding the behavior of these domains sheds light on critical assembly events that until know have remained largely unknown.

Acknowledgements

This work would not have been possible without the unwavering support and love of my family: my wife Amanda, our three children, my parents, brothers and sisters. Your encouragement has meant the world to me. Thank You.

I would also like to acknowledge the day-to-day contributions of the scientists, current and former, of the DeGuzman lab. In particular I would like to acknowledge the guidance and instruction of Dr. Dalian Zhong and Dr. Yu Wang. Their selfless contributions to my education and scientific training were crucial to the work presented in these Chapters. I would also like to mention the contributions of other faculty members and students who assisted with this work: Dr. Peter Gegenheimer for providing guidance during the RNA studies, Dr. Asokan Anbanandam for technical assistance during NMR data acquisition, and graduate student Mike Connor for his help in assigning the backbone of the Prospect Hill zinc finger.

This work began as a collaborative effort with the lab of Dr. Stephen St. Jeor of the University of Nevada at Reno. Therefore, I would like to acknowledge the intellectual contributions of both Dr. St. Jeor and his former student, Dr. Dan Boudreaux.

Finally, I would like particularly acknowledge my PhD advisor, Dr. Roberto N. DeGuzman for introducing me to the worlds of structural biology and NMR Spectroscopy. You provide an environment in your laboratory that encourages scientific debate and discussion. I will carry your teachings forward in future endeavors. Thank You.

Table of Contents

	Page
Abstract	iv
Acknowledgements	vi
Table of Contents	vii
List of Figures	ix
List of Tables	xiii
List of Abbreviations	xiv
Chapter 1. Introduction to Bunyaviridae Organization, Life Cycle, and Cytoplasmic Tails	1
A. Organization and Genome	2
B. Life Cycle	3
C. Gn Cytoplasmic Tails	4
D. The CCHC Motif	7
E. Conclusion	10
F. References	12
Chapter 2. Materials and Methods	
A. Protein Production and Purification	15
B. NMR Spectroscopy	21
C. Structure Calculation	22
D. *In Vitro* Transcription	23
E. Electrophoretic Mobility Shift Assay	26
F. CD Spectroscopy	27
G. References	28
Chapter 3. NMR Structure of the Andes Zinc Finger Domain	
A. Introduction	30
B. Protein Expression Results	32
C. Zn^{2+} EDTA Titration	33
D. NMR Structure Determination	34
E. Description of the Back-to-back CCHC ZF Fold	40
F. Mutagenesis of Zn^{2+} coordinating residues	42
G. Mutagenesis of non-Zn^{2+} coordinating residues	43
H. Hantaviral Gn Zinc Fingers Do Not Bind RNA	46
I. Implicatin of the Zinc Finger Structure in the Biology of Hantaviruses	47
J. Unique Properties of Hantaviral Zinc Fingers	49
K. Other Viral Zinc Fingers	50
L. References	51
Chapter 4. NMR Characterization of the Prospect Hill Gn Zinc Finger Domain	
A. Introduction	55

B. Protein Expression Results 57
C. PHV Gn Tail Also Relies on Zn^{2+} for Proper Folding 57
D. Backbone NMR Assignments 59
E. Modeling of the PHV Gn Zinc Finger Domain 60
F. Discussion 60
G. References 64

Chapter 5. NMR Structure of the Crimean Congo Hemorrhagic Fever Virus Gn Tail
A. Introduction 66
B. Protein Expression Results 67
C. Zn^{2+} is Required for Proper Folding of the CCHFV Gn Tail 69
D. NMR Structure Determination 69
E. Structure of CCHFV Gn Tail 729-805 75
F. The CCHFV Gn ZF Domain Contains Conserved Electrostatic Surfaces 77
G. The CCHFV Gn Tail Binds to RNA 78
H. Discussion 80
I. Key Differences Between Nairovirus and Hantavirus Gn Zinc Fingers 81
J. RNA Binding Properties of the CCHFV Gn 729-805 83
J. Proposed Role in Viral Assembly 84
K. References 85

Chapter 6. NMR Characterization of the Extended Andes ZF Domain (residues 534-610)
A. Introduction 88
B. Protein Production Results 89
C. NMR Assignments of Andes Gn 534-610 90
D. HNCA Derived Secondary Chemical Shift Profile 90
E. Backbone Dynamics of Andes Gn 534-610 93
F. Discussion 95
G. References 100

Chapter 7. RNA Binding Properties of the Bunyavirus Gn Cytoplasmic Tails
A. Introduction 101
B. Protein Production Results 102
C. Agarose EMSA Results of Andes Hantavirus Cytoplasmic Tail 104
D. Poly-Acrylamide EMSA Results of Andes Hantavirus Cytoplasmic Tail 106
E. Mutagenesis and EMSA Results of CCHFV Gn 729-805 107
F. Discussion 108
G. References 111

Chapter 8. Summary of NMR Studies of the Bunyaviridae Glycoprotein Cytoplasmic Tails 112

List of Figures

Figure	Page
1-1. Bunyaviridae Organization	2
1-2. Overview of Bunyaviridae Life Cycle	5
1-3A. Bunyaviridae Gn Tails	8
1-3B. Bunyaviridae Gn Tail Sequence Alignment	8
2-1. In Vitro Transcription of CCHFV M Segment Panhandle	25
3-1. Multiple Sequence Alignment of Hantavirus Gn Tails	32
3-2. CD Spectra of EDTA Titrations	33
3-3. ^{1}H-^{15}N HSQC Spectra of EDTA Titrations	35
3-4. Assigned ^{1}H-^{15}N HSQC Spectrum of the Andes Gn ZF	36
3-5. Secondary Chemical Shift Profile of the Andes Gn ZF	37
3-6. 2D 1H-13C HMQC Spectrum of the Andes Gn ZF	38
3-7. ^{1}H-^{15}N HMQC Spectrum of Histidine Side Chains	39
3-8A. Stereo Superposition of 20 lowest energy structures	41
3-8B. NMR Structure of the Andes Gn ZF	41
3-9A. Hydrogen-Deuterium Exchange Overlaid Spectra	42
3-9B. Hydrogen Bonding at the ZF Loop Apex	42
3-10. ^{1}H-^{15}N HSQC Spectra of Zn^{2+} Coordinating Residues	44
3-11. ^{1}H-^{15}N HSQC Spectra of non Zn^{2+} Coordinating Residues	45
3-12A. Electrophoretic Mobility Shift Assay using the Andes Gn ZF	46
3-12B. RT-PCR to Detect Viral S Segment for EMSA	46
3-13A. ^{13}C NOESY Spectra for Cys 548 Hβ's	48

3-13B. ^{13}C NOESY Spectra for Cys 573 Hβ's 48

3-14A. Zinc Binding Topology Diagram for the Andes Gn ZF Domain 50

3-14B. Zinc Binding Topology Diagram for a RING Domain 50

3-14C. Zinc Binding Topology Diagram for a LIM Domain 50

4-1. Gn Tail Sequence Alignment of Pathogenic and non-Pathogenic Hantaviruses 56

4-2. Assigned ^{1}H-^{15}N HSQC Spectrum of the PHV Gn ZF 58

4-3A. ^{1}H-^{15}N HSQC Spectra of EDTA Titration 59

4-3B. CD Spectra of EDTA Titrations 59

4-4. Secondary Cα Chemical Shift Profiles for PHV and Andes ZF Domains 61

4-5A. Modeled Structure of PHV Zinc Finger Domain 62

4-5B. Surface Electrostatics of Modeled PHV Zinc Finger Domain 62

5-1. Expression Summary of the CCHFV Gn Tail 68

5-2A. NMR Spectra for EDTA Titration of the CCHFV Gn Tail 70

5-2B. CD Spectra for EDTA Titration of the CCHFV Gn Tail 70

5-3. Assigned ^{1}H-^{15}N HSQC Spectrum of the CCHFV Gn ZF 71

5-4A. HNCA Strips of Residues 767-776 72

5-4B. HNCACB Strips of Residues 767-776 72

5-4C. CBCA(CO)NH Strips of Residues 767-776 72

5-5A. Secondary Chemical Shift Profile of the CCHFV Gn Tail 73

5-5B. [^{1}H]-^{15}N Heteronuclear NOE of the CCHFV Gn Tail 73

5-6. 2D ^{1}H-^{15}N HMQC Spectrum of Histidine Side Chains 74

5-7A. Stereo Superposition of 20 lowest energy structures 76

5-7B. NMR Structure of the CCHFV Gn Tail (729-805) 76

5-8A. Surface Electrostatics of the CCHFV Gn Tail 79

5-8B. Surface Electrostatics of the Andes Hantavirus ZF Domain 79

5-9A. M Segment Panhandle Diagrams for CCHFV and Andes Viruses 80

5-9B. Electrophoretic Mobility Shift Assay with the Andes and CCHFV Gn Tail ZF Domain 80

5-10A. Stereo Superposition of the Core ZF Domains of Andes and CCHFV 82

5-10B. Gn Tail Sequence Alignment of the Andes and CCHF Viruses 82

5-11. Packaging Model for the CCHFV 85

6-1. Diagram of the Extended Andes ZF Domain 89

6-2. Refolded and Native ^{1}H-^{15}N HSQC Spectra of Andes Gn 534-610 91

6-3. Assigned ^{1}H-^{15}N HSQC Spectrum of Andes Gn 534-610 91

6-4. HNCA 3D Strips of the Poly-Basic Flanking Sequences 92

6-5A. Secondary Cα Chemical Shift Profile of Andes Gn 543-599 94

6-5B. Secondary Cα Chemical Shift Profile of Andes Gn 534-610 94

6-5C. Secondary Cα Chemical Shift Profile of CCHFV Gn 729-805 94

6-6. [^{1}H]-^{15}N heteronuclear NOE Plot of Andes Gn 534-610 95

6-7. Helical Wheel Projection of Helix 3 97

6-8. [^{1}H]-^{15}N heteronuclear NOE Plots for Andes and CCHF Viruses Gn Tails 99

7-1. Bunyavirus Gn Tail Sequence Alignment Highlighting Conserved Basic Residues 102

7-2A. Agarose EMSA for the Andes Gn Tail (534-610) 104

7-2B. Agarose EMSA for the MBP-Andes Gn Tail (534-628) 104

7-3A. Agarose EMSA for the CCHFV Gn (729-805) 105

7-3B. Agarose EMSA for the Bunyamwera Gn (249-302) 105

7-4A. Acrylamide EMSA for the Andes Gn (534-610) Bound to the Andes Panhandle 106

7-4B. Reproduced EMSA for the Andes Gn (534-610) Bound to the Andes Panhandle 106

7-4C. EMSA for the Andes Gn (534-610) Under Denaturing Conditions 106

7-5A. Acrylamide EMSA for Wilde-type and Mutants of the CCHFV Gn (729-805) 108

7-5B. Cartoon Depiction of the Distribution of Mutated Residues 108

List of Tables

Table	Page
3-1. Restraints and Structural Statistics for the Andes Gn ZF	37
5-1. Restraints and Structural Statistics for the CCHFV Gn ZF	75

List of Abbreviations

CDC	Centers for Disease Control & Prevention
HSQC	Heteronuclear Single Quantum Coherence
HMQC	Heternuclear Multiple Quantum Coherence
NOE	Nuclear Overhauser Effect
NOESY	NOE Spectroscopy
CD	Circular Dichroism
IPTG	Isopropyl-β-D-thiogalactopyranoside
CYANA	Combined assignment and Dynamics Algorithm for NMR Applications
AMBER	Assisted Model Building with Energy Refinement
RMSD	Root-Mean-Square Deviation
CHARMM	Chemistry at HARvard Macromolecular Mechanics
N	Nucleocapsid protein
Gn	N-terminal Glycoprotein
RNP	Ribonucleoprotein
CCHFV	Crimean Congo Hemorrhagic Fever Virus
ITAM	Immunoreceptor Tyrosine-Based Activating Motif
Degron	Degradation signal
TRAF3	TNF Reseptor-Associated Factor 3
TNF	Tumor Necrosis Factor
TBK-1	TANK-Binding Kinase 1

Chapter 1. Introduction to Bunyaviridae Organization, Life Cycle, and Cytoplasmic Tails

Bunyaviridae is a large family of RNA viruses comprising approximately 350 diverse members [1]. Of these, perhaps the best known in the western hemisphere are the *Hantaviruses*. The family, however, contains other equally dangerous viruses that pose a threat to public health all over the world. The family is divided into five distinct genera (*Hantavirus*, *Nairovirus*, *Phlebovirus*, *Orthobunyavirus*, and *Tospovirus*) and can be further classified according to their natural reservoirs. Four of the five genera are carried by arthropods, such as mosquitoes, ticks, and thrips. The *Hantavirus*, which is rodent-borne, is the lone exception [1].

Human infection by many Bunyaviruses can result in a variety of illnesses ranging from very mild flu-like symptoms (Toscana virus, genus *Phlebovirus*) to very severe symptoms (Andes and Sin Nombre virus, genus *Hantavirus*) [2, 3]. In the most serious cases, infection can result in a hemorrhagic fever, myalgia, and nausea [2-4]. In rare cases (infection by the Crimeang Congo Hemorrhagic Fever, genus *Nairovirus*), the antiviral Ribavarin has been shown to be effective for reducing the symptoms of infection [5]. However, since many Bunyavirus infections are diagnosed late, mortality rates can climb as high as 40% (Andes virus, genus *Hantavirus*) [2-4].

Despite these risks, human infection does not represent the only concern. Infection by the Tomato Spotted Wilt Virus (genus Tospovirus) poses a serious threat to food crops and infection by the Rift Valley Fever virus (genus Phlebovirus) causes a devastating disease (near 100% mortality) in unborn livestock [6, 7]. All of this underlines the diverse nature and the potentially far-reaching social and economic impacts of Bunyaviruses.

Organization

All known Bunyaviruses share a similar organization. A lipid bilayer envelope surrounds three separate anti-sense RNA segments, packaged into ribonucleocapsids (**Figure 1-1**) [1]. The nucleocapsid proteins, which form a multimer in some bunyaviruses (*Hantaviruses*), encapsidate the RNA segments into distinct ribonucleocapsid complexes [1, 2]. The RNA polymerase completes the complex by also attaching to the RNA. One distinguishing characteristic of Bunyavirus RNA segments are conserved semi-complementary regions at the 5' and 3' ends, which are predicted to form a 'panhandle' like structure [1]. Two glycoproteins are present on the surface of the virus as part of a multimeric spike comples (**Figure 1-1**) [1]. The glycosylated ectodomains of the Gn/Gc heterodimer interact with integrins at the cell surface upon infection [1, 2, 8]. On the interior of the virus, Gn and Gc may have a cytoplasmic tail of varying length.

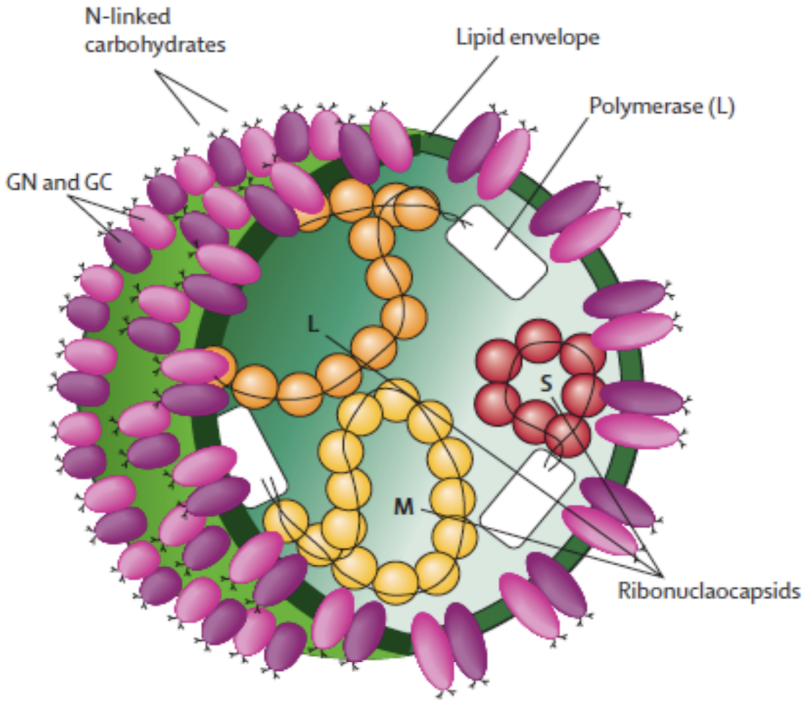

Figure 1-1. Bunyavirus Organization. Figure adapted from Ergonul et al., 2006 [9]

Bunyavirus Genome

The small segment (S, 1-3 kDa) encodes for the nucleocapsid protein, which is a key component of the ribonucleocapsid complex [1]. The medium segment (M, ~4 kDa) encodes a polyprotein that is post-translationally processed into the two key surface glycoprotiens: Gn and Gc (also known as G1 and G2) [1]. Here Bunyaviruses diverge considerably, with some family members producing no more than the two basic surface glycoproteins (*Hantaviruses*) [1, 2], while the preGn and preGc of other family members undergo further processing to produce as many as three additional proteins (Crimean Congo Hemorrhagic Fever virus, genus *Nairovirus*) [10, 11]. Lastly, the largest segment (L, ~6 kDa in most Bunyaviruses, ~12 kDa in *Nairoviruses*) encodes the viral RNA polymerase [1].

Bunyavirus Life Cycle

In addition to a common virion organization, most Bunyaviruses share a similar life cycle. Infection occurs either through inhalation of the virus (*Hantaviruses*) [2] or by direct mosquito, sandfly, or tick bite (remaining Bunyaviruses) [1]. Infection typically occurs in the endothelium. The virus enters via receptor mediated endocytosis following attachment to beta 3 integrins [2, 4, 8]. Changes in pH at the endosome are believed to initiate fusion of the viral envelope with the endosomal membrane, causing release of all three nucleocapsid complexes into the cytoplasm, which is also the site of viral replication [1, 2]. The RNA polymerase initiates transcription of mRNA for each segment (**Figure 1-2**). S and L segment mRNA transcripts are translated by free ribosomes, releasing nascent ribonucleocapsid protein and RNA

polymerase into the cytoplasm [1, 2]. M segment mRNA is translated by ER bound ribosomes to produce the membrane bound Gn and Gc glycoproteins. Subsequent replication of the antigenomic (+)RNA by the viral polymerase yields nascent genomic (-)RNA, which is free to become encapsidated by the cytosolic nucleocapsid protein [1, 2]. Meanwhile, the glycoprotein precursor is cleaved to yield the Gn and Gc heterodimer. For most Bunyaviruses, Gn/Gc is retained in the cis-golgi, which is also the site of virion assembly (one exception is the Sin Nombre virus, genus *Hantavirus*, which apparently buds at the plasma membrane) [1, 2].

Less is known regarding assembly of the virion. The Gn cytoplasmic tail likely plays an important role; it not only is present in all known Bunyaviruses but it is also in the proper orientation to participate in packaging of the ribonucleocapsid complexes into the budding virion. Because budding occurs into the lumen of the cis-golgi, the Gn cytoplasmic tail is converted from a cytoplasmic orientation (pre-budding) to an inward facing orientation in the mature virus (post-budding) (**Figure 1-2**).

Cytoplasmic Tails Are Crucial to Bunyavirus Assembly Events

In recent years, increasing evidence has reinforced a role for the Gn tail in virus assembly. Shi et al. (2007) demonstrated that deletions or mutagenesis of the Gn and Gc tails of Bunyamwera virus (genus Orthobunyavirus) prevented the packaging of a minigenome into virus like particles [12]. Furthermore, alanine scanning of the Gn tail in Uukuniemi virus (genus Phlebovirus) had the same results [13]. Most recently, Hepojoki et al. (2010) used point

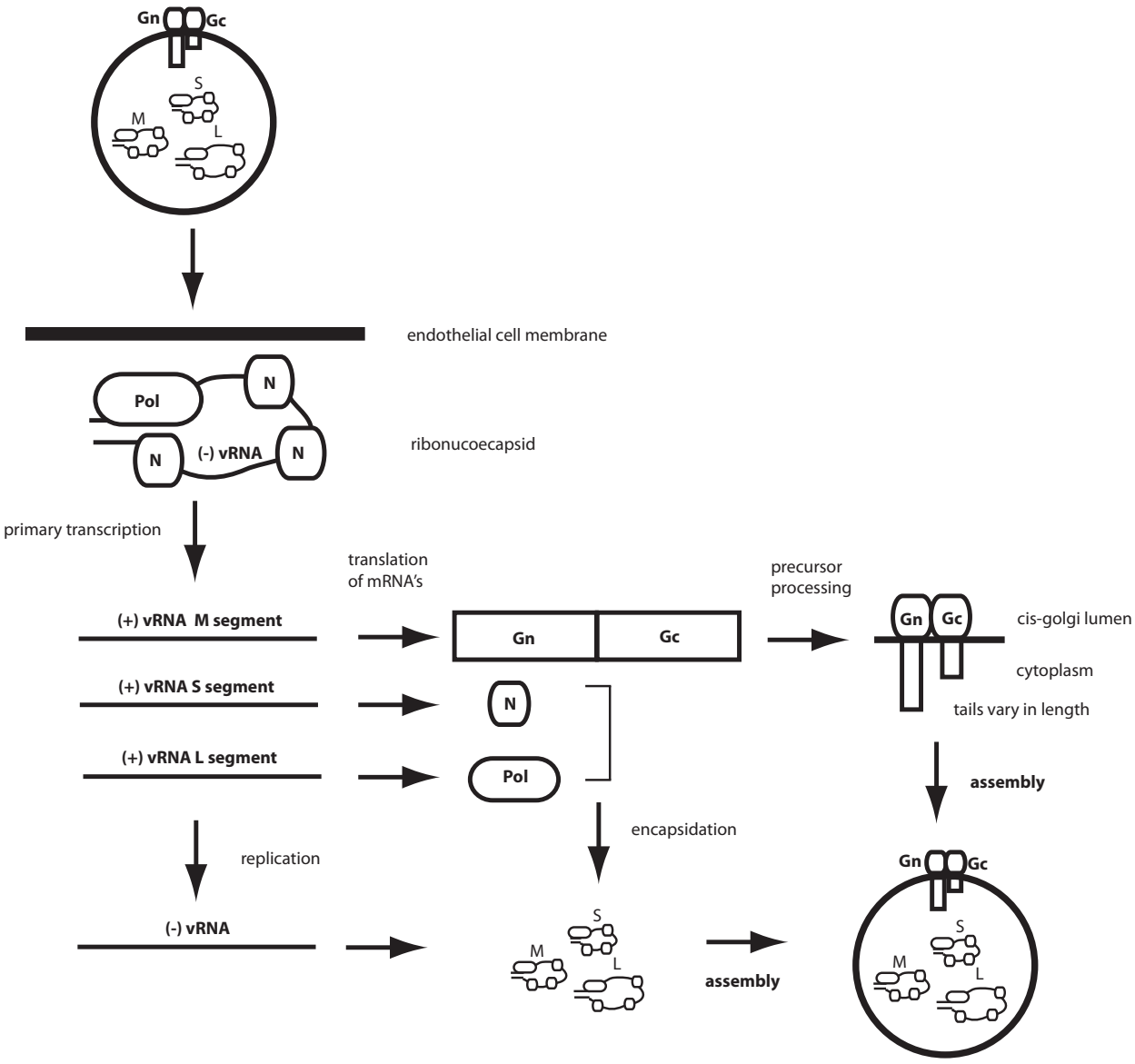

Figure 1-2. Overview of the Bunyaviridae life cycle in an infected endothelial cell. Replication occurs primarly in the cytoplasm with assembly occurring at the *cis*-golgi membrane.

mutagenesis and immunoprecipitation to demostrate that the Gn tail of Puumala virus (genus *Hantavirus*) containsmultiple interaction sites with the nucleocapsid protein [14], an integral component of ribonucleocapsid complexes. This finding was later supported by cryo-electron microscopy imaging of the intact Hantaan virus (genus *Hantavirus*), which showed that the ribonucleocapsid complex closely associates with the cytoplasmic side of the virus envelope [15]. Taken together, these data support a role for the Gn cytoplasmic tail in packaging the newly formed ribonucleocapsids into the budding virus.

Gn Cytoplasmic Tails of Hantaviruses also Perform an Immunogenic Function

Hantavirus Gn tails are on average 50 residues longer than other Bunyaviridae tails (**Fig. 1-3A**). Functional studies of the *Hantavirus* Gn tails strongly suggest a role for suppression of early immunogenic response in the infected cell. Two functional immunoreceptor tyrosine-based activation motifs (ITAMs) along with a degradation signal (Degron) located near the Gn tail's C-terminus (**Fig. 1-3A**) were shown to direct the ubiquitination and proteasomal degradation of the tail [16-18]. Degradation of the Gn tail has also been shown to be exclusive to pathogenic *Hantaviruses* [18]. Furthermore, the Gn tail of the pathogenic New York *Hantavirus* has been reported to regulate interferon-β response by disrupting TBK1-TRAF3 complex formation, a critical early step in the interferon response [19, 20]. These data suggest the Gn tail contributes at least partially to virulence in *Hantaviruses*.

Notably, the conserved ITAMS and degradation signal have not been reported in other Bunyavirus Gn tails. Perhaps not coincidentally, the Rift Valley fever virus (*Phlebovirus*) and the La Crosse virus (*Orthobunyavirus*) encode a non-structural protein from the S segment (NSs) that has been shown to drastically reduce interferon beta response in infected cells [21, 22]. Pathogenic *Hantaviruses* have not been reported to encode for such a non-structural protein, indicating that the Gn tail may partially substitute for the function of the non-structural protein.

Bunyavirus Gn Tails Have a Highly Conserved Dual CCHC Motif

Bunyavirus Gn tails vary in both length and composition (**Figure 1-3A**). A multiple sequence alignment of representative members of each of the five genera (**Figure 1-3B**) reveals a large amount of variation between different groupings. *Hantaviruses* and *Nairoviruses*, for example, share a mere 12% sequence identity (23% sequence similarity). Likewise, *Orthobunyaviruses* and *Nairoviruses* share 10% sequence identity (14% sequence similarity). Within each genus, however, Gn tails are relatively more conserved. The *Hantavirus* Gn tails, for instance, are approximately 40% identical (~50% similarity). Perhaps the greatest sequence variability within any particular genus occurs in the Phlebovirus, with a sequence identity of 22% (36% similarity).

Despite this sequence variability, analysis of the alignment reveals a conserved repeated cysteine and histidine motif (**Figure 1-3B**). This dual CCHC motif is present in all but one genera of Bunyaviruses, with Phleboviruses the lone exception. While the spacing of the motif

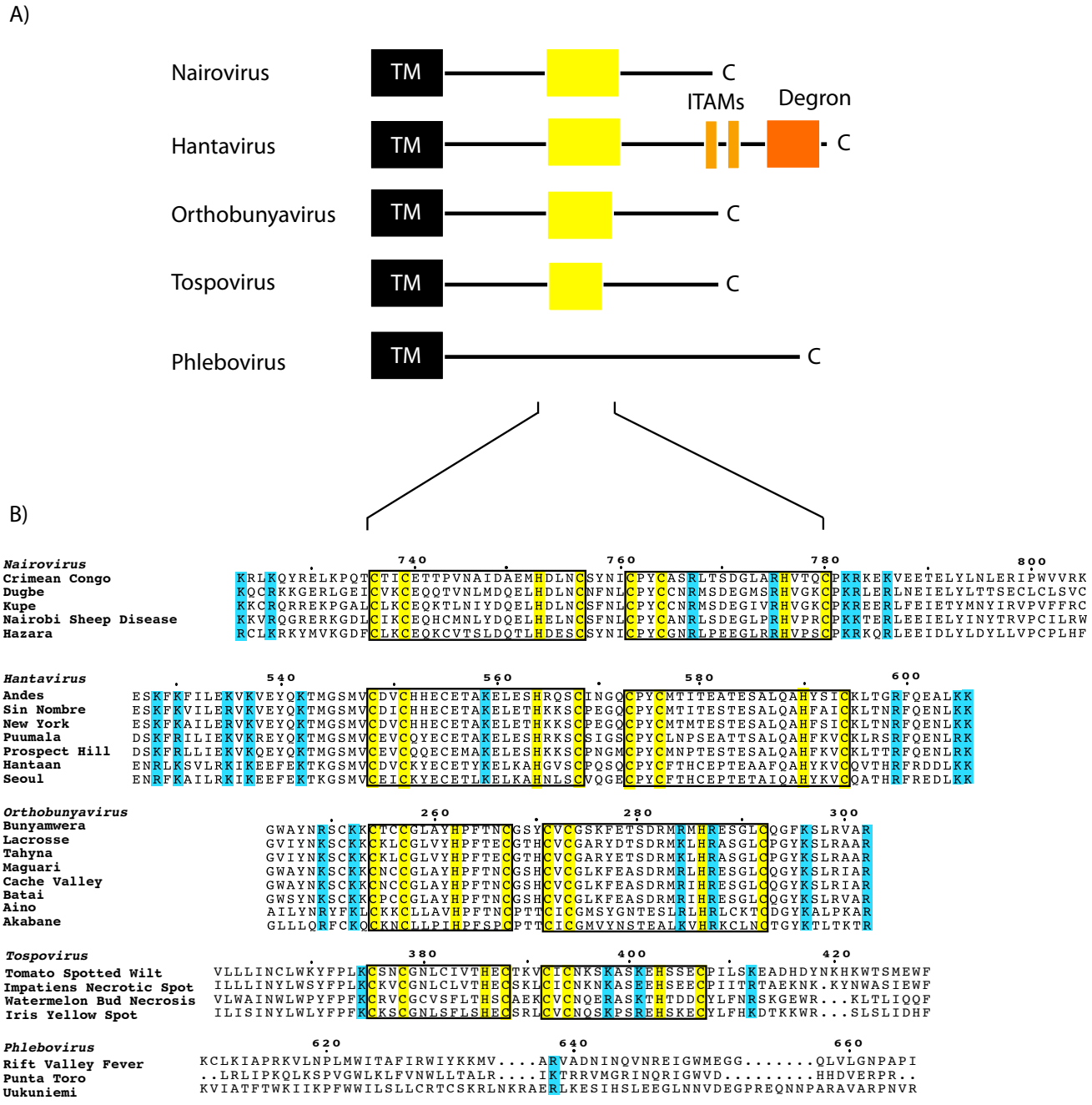

Figure 1-3. Gn tail sequence alignment of representative members of Bunyaviridae. A) Diagram indicating the location of the conserved dual CCHC motif (yellow box) in each genus. The immunoreceptor tyrosine-based activation motifs (ITAMs) and the degradation signal (Degron) of the *Hantavirus* Gn tail are represented by light orange and dark orange rectangles, respectively. B) The conserved dual CCHC motif is highlighted in yellow with a box drawn around each predicted zinc finger fold. Conserved basic residues are highlighted in blue.

is tightly conserved within each respective genera, it varies slightly between the different groups. The motif is more dispersed, for example, in *Hantaviruses* and in *Nairoviruses*, covering 46 residues of each Gn tail. This represents approximately one third the length of *Hantavirus* Gn tails and almost one half the length of *Nairovirus* Gn tails. The same motif is more tightly arranged, however, in the *Orthobunyavirus* (38 residues) and the *Tospovirus* (32 residues) (**Figure 1-3B**). Despite this, one feature that is common to all four genera is the presence of a very short linker sequence between the two motifs. Overall, the consensus motif can be written as (Cys-X_2-Cys-X_{4-12}-His-X_{1-4}-Cys) -X_{3-4}-(Cys-X_{1-4}-Cys-X_{8-12}-His-X_{3-5}-Cys), where X represents any amino acid.

The CCHC Motif in Structural Biology

Conserved cysteines and histidines frequently form part of metal binding domains [23]. While they can theoretically bind a number of metals, including cadmium and cobalt, zinc binding is most common [24]. Zinc finger proteins are the most ubiquitous structural motifs in the eukaryotic genome and perform a variety of functions, most frequently reported as a scaffold for DNA and RNA interactions [25, 26]. In recent years, the number of zinc fingers reported to be involved in mediating protein-protein interactions has increased [27].

The CCHC motif, along with the variation CCHH motif, is known to form a classical ββα zinc finger in which one zinc atom is coordinated between an anti-parallel beta hairpin and a short alpha helix [25, 27]. Perhaps the most well known CCHC type zinc finger occurs in the

HIV-1 nucleocapsid, where it mediates encapsidation of the RNA packaging signal in the virus [28, 29].

The Dual CCHC Motif in Bunyaviruses

Prior to this dissertation, the existence of zinc fingers in Bunyaviral proteins was unconfirmed. However, the presence of the conserved dual CCHC motif had given rise to multiple theories regarding its function. For example, knowledge that the *Hantavirus* Gn tail was ubiquitinated and degraded led to speculation that the conserved dual CCHC motif most likely formed a RING finger domain [18], because RING domains frequently function as E3 ubiquitin ligases [30]. Notably, RING finger domains have a characteristic overlapping zinc binding topology, the spacing of which closely resembles the Bunyavirus dual CCHC motif [30]. Meanwhile, in *Nairoviruses*, where the Gn tail has not been shown to be degraded, the high cysteine content of the Crimean Congo Hemorrhagic Fever Gn cytoplasmic tail led to speculation of extensive disulfide bond formation in the tail [31]. Indeed, the conserved and the widespread nature of the CCHC motif in Bunyaviruses suggests a general, albeit important, function in the virus life cycle.

Conclusion

The global distribution of Bunyaviruses, the increased spread of zoonotic disease, and the unusually high mortality rate of hemorrhagic fevers make the study of this particular family of

viruses important. The Centers for Disease Prevention has listed *Hantaviruses* as a potential bioterrorism agent due in part to a potential for high mortality [32]. Nonetheless, current challenges in Bunyavirus research have hampered progress. For example, a reverse genetics system for the study of *Hantaviruses* has proven to be elusive. The requirement for Bio Safety Level 4 facilities for the study of the Crimean Congo Hemorrhagic Fever virus is also a constraint. Consequently, many questions remain regarding the Bunyavirus life cycle. Specifically, we have yet to answer the questions: How does the Gn tail participate in virus assembly? What role, if any, is played by the conserved CCHC motif in life cycle of the virus?

In this dissertation, I hypothesize that the conserved dual CCHC motif in Bunyaviruses forms a dual zinc finger fold. Because the motif is also present in the Gn tails of non-pathogenic *Hantaviruses* that have now been shown to be exceptions to the ubiquitin mediate degradation of the tail [18] (hence obviating the need for a RING domain on the tail), I further hypothesize that this dual zinc finger plays a role in events leading to packaging of the ribonucleocapsid complexes into the intact virus.

The topics of the subsequent chapters include the high resolution structures via nuclear magnetic resonance spectroscopy (NMR) of the Gn folded domains of two viruses: the Andes virus and the Crimean Congo Hemorrhagic Fever virus, representing members of the *Hantavirus* and the *Nairovirus* genus respectively. The structures demonstrate that the dual CCHC motif does in fact fold into a compact double beta-hairpin and helix zinc finger. The two classically folded zinc fingers together, however, represent a novel arrangement of zinc fingers previously undescribed in the structural database. Furthermore, the domain contains back-to-back zinc

binding topology, eliminating the possibility of a RING domain in the Gn tail and dispelling this convention in the literature.

Furthermore, this work also identifies a distinct difference in the distribution of the conserved surface electrostatics between the two structures. This difference sheds light on the possible function of not only *Hantaviruses* and *Nairoviruses* but other members of Bunyaviridae as well. In addition, the results of preliminary mobility shift studies suggest a likely interaction between Gn cytoplasmic tails and the RNA component of the ribonucleocapsid complex, which is the first time this interaction has been reported. The results identify initial steps in assembly of Bunyaviruses.

References

1. Elliott, R.M., Bouloy, M., Calisher, C. H., Goldbach, R., Moyer, J. T., Nichol, S. T., Pettersson, R., Plyusnin, A., and Schmaljohn, C., *Bunyaviridae*, in *Virus Taxonomy: The classification and Nomenclature of Viruses. The Seventh Report of the International Committee on Taxonomy of Viruses.*, C.M. Fauquet, Mayo, M.A., Maniloff, J. Desselberger, U., and Ball, eds., Editor. 2005, Academic Press: San Diego.
2. Khaiboullina, S.F., S.P. Morzunov, and S.C. St Jeor, *Hantaviruses: molecular biology, evolution and pathogenesis.* Curr Mol Med, 2005. **5**(8): p. 773-90.
3. Overturf, G.D., *World arboviruses: the Bunyaviridae.* Pediatr Infect Dis J, 2009. **28**(11): p. 1014-5.
4. Flick, R. and C.A. Whitehouse, *Crimean-Congo hemorrhagic fever virus.* Curr Mol Med, 2005. **5**(8): p. 753-60.
5. Fisher-Hoch, S.P., et al., *Crimean Congo-haemorrhagic fever treated with oral ribavirin.* Lancet, 1995. **346**(8973): p. 472-5.
6. Hanssen, I.M., M. Lapidot, and B.P. Thomma, *Emerging viral diseases of tomato crops.* Mol Plant Microbe Interact, 2010. **23**(5): p. 539-48.
7. Hollidge, B.S., F. Gonzalez-Scarano, and S.S. Soldan, *Arboviral encephalitides: transmission, emergence, and pathogenesis.* J Neuroimmune Pharmacol, 2010. **5**(3): p. 428-42.
8. Mackow, E.R. and I.N. Gavrilovskaya, *Hantavirus regulation of endothelial cell functions.* Thromb Haemost, 2009. **102**(6): p. 1030-41.

9. Ergonul, O., *Crimean-Congo haemorrhagic fever.* Lancet Infect Dis, 2006. **6**(4): p. 203-14.
10. Sanchez, A.J., et al., *Crimean-congo hemorrhagic fever virus glycoprotein precursor is cleaved by Furin-like and SKI-1 proteases to generate a novel 38-kilodalton glycoprotein.* J Virol, 2006. **80**(1): p. 514-25.
11. Altamura, L.A., et al., *Identification of a novel C-terminal cleavage of Crimean-Congo hemorrhagic fever virus PreGN that leads to generation of an NSM protein.* J Virol, 2007. **81**(12): p. 6632-42.
12. Shi, X., et al., *Role of the cytoplasmic tail domains of Bunyamwera orthobunyavirus glycoproteins Gn and Gc in virus assembly and morphogenesis.* J Virol, 2007. **81**(18): p. 10151-60.
13. Overby, A.K., R.F. Pettersson, and E.P. Neve, *The glycoprotein cytoplasmic tail of Uukuniemi virus (Bunyaviridae) interacts with ribonucleoproteins and is critical for genome packaging.* J Virol, 2007. **81**(7): p. 3198-205.
14. Hepojoki, J., et al., *Cytoplasmic tails of hantavirus glycoproteins interact with the nucleocapsid protein.* J Gen Virol, 2010. **91**(Pt 9): p. 2341-50.
15. Battisti, A.J., et al., *Structural studies of Hantaan virus.* J Virol, 2011. **85**(2): p. 835-41.
16. Geimonen, E., et al., *Tyrosine residues direct the ubiquitination and degradation of the NY-1 hantavirus G1 cytoplasmic tail.* J Virol, 2003. **77**(20): p. 10760-868.
17. Geimonen, E., et al., *Hantavirus pulmonary syndrome-associated hantaviruses contain conserved and functional ITAM signaling elements.* J Virol, 2003. **77**(2): p. 1638-43.
18. Sen, N., A. Sen, and E.R. Mackow, *Degrons at the C terminus of the pathogenic but not the nonpathogenic hantavirus G1 tail direct proteasomal degradation.* J Virol, 2007. **81**(8): p. 4323-30.
19. Alff, P.J., et al., *The pathogenic NY-1 hantavirus G1 cytoplasmic tail inhibits RIG-I- and TBK-1-directed interferon responses.* J Virol, 2006. **80**(19): p. 9676-86.
20. Alff, P.J., et al., *The NY-1 hantavirus Gn cytoplasmic tail coprecipitates TRAF3 and inhibits cellular interferon responses by disrupting TBK1-TRAF3 complex formation.* J Virol, 2008. **82**(18): p. 9115-22.
21. Blakqori, G., et al., *La Crosse bunyavirus nonstructural protein NSs serves to suppress the type I interferon system of mammalian hosts.* J Virol, 2007. **81**(10): p. 4991-9.
22. Billecocq, A., et al., *NSs protein of Rift Valley fever virus blocks interferon production by inhibiting host gene transcription.* J Virol, 2004. **78**(18): p. 9798-806.
23. Iuchi, S. and N. Kuldell, *Zinc finger proteins : from atomic contact to cellular function*. Molecular biology intelligence unit. 2005, Georgetown, Tex.
New York: Landes Bioscience ;
Kluwer Academic/Plenum Publishers. 276 p.
24. Hartwig, A., *Zinc finger proteins as potential targets for toxic metal ions: differential effects on structure and function.* Antioxid Redox Signal, 2001. **3**(4): p. 625-34.
25. Laity, J.H., B.M. Lee, and P.E. Wright, *Zinc finger proteins: new insights into structural and functional diversity.* Curr Opin Struct Biol, 2001. **11**(1): p. 39-46.
26. Brown, R.S., *Zinc finger proteins: getting a grip on RNA.* Curr Opin Struct Biol, 2005. **15**(1): p. 94-8.
27. Gamsjaeger, R., et al., *Sticky fingers: zinc-fingers as protein-recognition motifs.* Trends Biochem Sci, 2007. **32**(2): p. 63-70.

28. Summers, M.F., et al., *High-resolution structure of an HIV zinc fingerlike domain via a new NMR-based distance geometry approach.* Biochemistry, 1990. **29**(2): p. 329-40.
29. De Guzman, R.N., et al., *Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element.* Science, 1998. **279**(5349): p. 384-8.
30. Deshaies, R.J. and C.A. Joazeiro, *RING domain E3 ubiquitin ligases.* Annu Rev Biochem, 2009. **78**: p. 399-434.
31. Sanchez, A.J., M.J. Vincent, and S.T. Nichol, *Characterization of the glycoproteins of Crimean-Congo hemorrhagic fever virus.* J Virol, 2002. **76**(14): p. 7263-75.
32. Moran, G.J., *Threats in bioterrorism. II: CDC category B and C agents.* Emerg Med Clin North Am, 2002. **20**(2): p. 311-30.

Chapter 2: Materials & Methods

Protein Expression and Purification

Andes Gn 534-610 and Prospect Hill Gn 548-602 — The cysteine/histidine-rich region (residues 543-599) of the Gn tail of the Andes virus (strain 23) and Prospect Hill virus were sub-cloned into a pET-21a (Novagen) vector using NdeI and BamHI as restriction sites. The vector was modified previously to a encode for a C-terminal fusion to a His_6-tagged GB1 domain separated by a TEV protease cleavage site. GB1 is the B1 immunoglobulin-binding domain of *Streptococcus* protein [1], and a GB1 expression plasmid (obtained from Peter E. Wright, Scripps Research Institute, La Jolla, California) was used in the subcloning. Isotopically (^{15}N or $^{15}N,^{13}C$) labeled protein was over-expressed in bacteria as follows: Freshly transformed *E. coli* BL21(DE3) were grown in 1 liter M9 minimal media supplemented with 0.1 mM $ZnSO_4$ before and after induction. Cells were grown at 37°C, induced with 1 mM isopropyl-β-D-thiogalacto-pyranoside at $A_{600} \sim 0.8$, and incubated at 15°C overnight, with gentle shaking at 150 rpm (to a final $A_{600} \sim 2.0$). Cells were harvested by centrifugation at 4,000 rpm for 10 minutes, resuspended in buffer A (20 mM Tris-HCl pH 8.0, 20 mM NaCl, 1 mM DTT, 0.1 mM $ZnSO_4$), and lysed by sonication on ice (5 minutes of 2 second-long pulses interspersed by 20 seconds of cooling). Cellular debris was removed by centrifugation at 10,000 rpm for 10 minutes and to the soluble fraction was added $^1/_{10}$ volume of 1% polyethyleneimine (pH 8) stock in order to precipitate the nucleic acids. Following centrifugation, the supernatant was applied to a FPLC 5 mL HiTrap Q column (GE Healthcare) and washed with 5 column volumes of additional buffer A at a flow rate of 2 ml per minutes. The bound protein was eluted with a 100 mL linear gradient of buffer B (20 mM Tris-HCl pH 8.0, 0.5 M NaCl, 1 mM DTT, 1 mM $ZnSO_4$). Peak fractions were analyzed by

SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) and were pooled and dialyzed against TEV-digestion buffer (50 mM Tris-HCl pH 8.0, 20 mM NaCl, 1 mM DTT, 1 mM $ZnSO_4$). TEV digestion was carried out at 25°C for 16 hrs with 0.16 mg of recombinant TEV protease [2] per 10 ml of fusion protein. The His_6-tagged GB1 domain was then removed by passing the digest through a charged 1 mL nickel affinity column (Sigma #I1408); purified $G1^{543\text{-}599}$ was recovered in the flow-through. Recombinant $G1^{543\text{-}599}$ zinc finger retained two extra N-terminal amino acids (Gly-His) resulting from the subcloning.

Site-directed mutagenesis was performed using the QuickChange kit (Stratagene). In total, 7 cysteine and 5 histidine residues (4 native His residues and the cloning artifact, His^{542}) were mutated individually to serine or phenylalanine, respectively and confirmed by DNA sequencing. Andes mutants H542F, H553F, H552F, H590F, and C594S expressed as soluble proteins and purified them by nickel affinity chromatography as previously described [3]. Mutants C548S, C551S, C555S, H564F, C568S, C573S, and C576S expressed as inclusion bodies (despite the presence of the GB1 solubility tag) and were purified as follows: inclusion bodies were resuspended at room temperature in solubilization buffer (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 8 M urea, 1 mM DTT, 0.1 mM $ZnSO_4$, and 0.1 mM phenylmethanesulphonyl fluoride). Overnight incubation in solubilization buffer resulted in near complete solubilized protein. Solubilized proteins were then dialyzed into urea-containing binding buffer (20 mM Tris-HCl pH 8.0, 0.5 M NaCl, 6 M urea, 1 mM DTT, and 0.1 mM $ZnSO_4$), loaded onto a 5 mL nickel affinity column, washed with five column volumes of urea-containing binding buffer, and eluted with urea-containing buffer (20 mM Tris-HCl pH 8.0, 20 mM NaCl, 6 M urea, 1 mM DTT, 0.1 mM $ZnSO_4$, 1 M imidazole). Eluted protein was then refolded using stepwise dialysis while in 3 kDa

dialysis tubing (refolding buffer consisted of 20 mM Tris (8.0), 100 mM NaCl, 1 mM DTT, 0.1 mM ZnSO4, and urea concentrations consisting of 6M, 4M, 2M, 1M, 0.5M, and 0.25M) to remove urea. None of the mutant fusion proteins were cleaved. The 2D ^{1}H-^{15}N HSQC of GB1 was used as a marker to determine the proper refolding of the fusion protein.

MBP Tagged Andes Gn 526-610 – Plasmids encoding the N-terminal MBP tagged Andes Gn 526-610 were obtained from the COBRE Protein Production Group, courtesy of Dr. Philip Gao. Additional plasmids encoding the MBP tagged full-length tail were toxic to *E. coli* and so were not utilized. However, MBP-Andes Gn 526-610 expressed as soluble protein and was utilized in early RNA binding experiments.

The above plasmid was transformed into E. coli BL21(DE3) and grown in 1L Lysogeny Broth (LB) media with gentle shaking at 37°C until ~0.8 optical density, at which point the cells were induced with induced with 1 mM isopropyl-β-D-thiogalactopyranoside and incubated overnight with gentle shaking (150 rpm) at 15°C. Cells were harvested by centrifugation at 4,000 rpm for 10 minutes and resuspended in IMAC resin binding buffer as described previously [3]. Lysis was achieved by sonication on ice (using 50% strength for 5 minutes total time of 2 second pulses alternating with 20 seconds cooling). Cell debris were removed by centrifugation at 10,000 rpm for 15 minutes and the soluble lysate applied to a 10 ml Nickel-charged IMAC resin column. Upon washing of the column with five volumes of IMAC binding buffer, the protein was eluted using standard elution buffer (20 mM Tris (8.0) 500 mM NaCl, 250 mM imidazole, 0.1 mM $ZnSO_4$). Upon confirmation of the peak fractions by SDS-PAGE, the fractions were pooled, concentrated to a final volume of 10 ml using 15ml centrifugal concentrators (Amicon),

and dialyzed into SEC (size exclusion chromatography) running buffer (20 mM Tris (8.0) 300 mM NaCl, 1 mM DTT, 0.1 mM $ZnSO_4$. The protein was applied to a 300 ml SEC column and run overnight at a flow rate of 1 ml/min. The elution trace indicated part of the protein was present in the void volume, and hence was most likely part of an aggregate. Upon analysis by SDS-PAGE, the peak fractions eluting after the void volume were found to include most of the recombinant protein. These fractions were dialyzed into RNA binding buffer (30 mM $NaPO_4$, 30 mM NaCl, 1 mM DTT, 0.1 mM $ZnSO_4$).

Attempts to cleave the MBP-Andes Gn 526-610 were not entirely successful. Overnight incubation with TEV only reached approximately 50% completion. Increasing the amount of protease and incubations at higher temperatures (37°C) failed to cleave the construct more efficiently. Furthermore, attempts to separate the cleaved MBP tag from the cleaved protein were entirely unsuccessful, as both proteins co-eluted in a SEC column, ion exchange, and nickel affinity columns. These results suggest the two proteins were still attached, although most likely non-covalently. Therefore, subsequent RNA binding experiments relied on proteins expressed as GB1 fusions instead.

Andes Gn 534-610 – The largest fragment of the *Hantavirus* Gn cytoplasmic that was successfully expressed and cleaved was from residues 534-610. However, overnight expression in *E. coli* BL21(DE3) resulted in only half of the protein yield in the soluble fraction, as determined by analysis with the BugBuster Protein Extraction Reagent (EMBD Biosciences, Cat. no. 70923-3) with the remainder present as an inclusion body. The soluble fraction was purified using standard nickel IMAC resin as described previously. Upon TEV mediated cleavage of the

GB1 tag, the mixture was passed back through a charged IMAC resin column and the cleaved protein was collected in the wash fractions. The insoluble fractions were resuspended in denaturing buffer (4 M urea, 100 mM NaCl, 15 mM DTT). Upon overnight solubilization, the urea was removed via step-wise dialysis as described previously for insoluble mutants of Andes Gn, followed by step-wise dialysis to bring the reducing conditions down to 1 mM DTT. Upon refolding, the protein cleaved efficiently by overnight 25°C incubation with TEV, and the tag separated as described above. The refolded Andes Gn 534-610 was kept separate than the natively expressed protein, because their 2D ^{1}H-^{15}N HSQC spectra showed minor differences (Chapter 6).

CCHFV Gn 729-805 – Various constructs of the CCHF virus (strain SPU103/87, genus *Nairovirus*) Gn cytoplasmic region (**Fig. 5-1**) were subcloned from a synthetic gene (GenScript) into the expression vectors pDZ1 and pDZ3 [4], which expressed His_6-tagged GB1 fusion proteins with TEV protease cleavage sites. For NMR structure determination, the soluble Gn construct spanning residues 729-805 ($Gn^{729\text{-}805}$) was expressed and purified under native conditions following closely the method reported above for the Andes *Hantavirus* zinc finger domain [5]. Briefly, ^{15}N- and $^{15}N/^{13}C$-labeled proteins were expressed in *E. coli* BL21(DE3) grown in 1 liter M9 minimal media supplemented with 0.1 mM $ZnSO_4$ before and after induction. Cells were grown at 37°C, induced with 1 mM isopropyl-β-D-thiogalactopyranoside at A_{600} ~ 0.8, and continued cell growth in a 15°C shaker overnight (to a final A_{600} ~ 2.0). Cells were then harvested by centrifugation at 4,000 rpm for 10 minutes, resuspended in buffer A (20 mM Tris-HCl pH 8.0, 20 mM NaCl, 1 mM DTT, 0.1 mM $ZnSO_4$), and lysed by sonication on ice (5 minutes total sonication time using 2 second pulses with 20 seconds cooling, probe at 50%

strength). Cellular debris was removed by centrifugation, and to the supernatant was added $^{1}/_{10}$ volume of 1% polyethyleneimine (pH 8) in order to precipitate the nucleic acids. Following centrifugation at 10,000 rpm for 10 minutes, supernatant was passed through a 40 mL Q column (GE Healthcare) , washd with three column volumes buffer A, and eluted with a 280 ml linear gradient of buffer B (20 mM Tris-HCl pH 8.0, 0.5 M NaCl, 1 mM DTT, 1 mM $ZnSO_4$). Eluted protein was detected using SDS-PAGE. For TEV protease digestion, fractions containing the fusion protein were pooled and dialyzed at 25°C overnight in buffer (50 mM Tris-HCl pH 8.0, 20 mM NaCl, 1 mM DTT, 1 mM $ZnSO_4$) with 0.16 mg recombinant TEV protease [2] per 10 ml of fusion protein. The TEV digestion mixture was then dialyzed back into buffer A and passed again through a 40 mL Q column (GE Healthcare). The GB1 tag (theoretical pI of 5.6) was retained on the column while $Gn^{729\text{-}805}$ (theoretical pI of 6.9) was present in the flow-through as it was not acidic enough to bind the column. The 50 mL flow-through fraction was concentrated to 3 ml using Ultra-15 centrifugal filters (Amicon) and dialyzed it into NMR buffer (10 mM $NaPO_4$ pH 7.0, 10 mM NaCl, 1 mM DTT, 0.1 mM $ZnSO_4$). The $Gn^{729\text{-}805}$ construct retained two residues (Gly-His) cloning artifacts at the N-terminus.

Site-directed mutagenesis was performed using the QuickChange kit (Stratagene). In total, three single and one double mutant were generated to correspond with the reversal of conserved basic charges on the protein surface. Expression plasmids containing the mutations K782D, R783D, K786D, and R767D/K786D were all confirmed by DNA sequencing. All mutants expressed in a soluble form and were purified using nickel affinity His_6 tag purification as described above. Briefly, cells were grown, harvested and sonicated. The soluble lysate was applied to an IMAC resin previously charged with 50 mM $NiSO_4$. Upon washing with additional

binding buffer, the protein was eluted using 20 mM Tris (8.0) 500 mM NaCl, 250 mM imidazole, 0.1 mM $ZnSO_4$. Eluted protein was then cleaved by overnight incubation with TEV protease as described above. Lastly, the GB1 tag was removed by passing the cleaved mixture back through a charged and equilibrated nickel resin column.

NMR Spectroscopy

NMR data were acquired at 25°C using a Bruker Avance 800 MHz spectrometer equipped with a cryoprobe, data was processed with NMRPipe [6], and analyzed with NMRView [7]. Backbone assignments were obtained for Andes Gn 543-599, Prospect Hill Gn 548-602, and CCHFV Gn 729-805 from 2D ^{1}H-^{15}N HSQC [8] and 3D HNCA [9], CBCA(CO)NH [9], and HNCACB [10]. Backbone assignments for Andes Gn 534-610 were obtained solely from the 3D HNCA dataset. Secondary structures were identified from the Cα, Cβ, and Hα chemical shifts [11]. Side chain assignments were obtained from 2D ^{1}H-^{13}C HMQC [12], 3D HBHA(CO)NH [13], and 3D ^{13}C-edited HMQC-NOESY [14] (t_{mix}=120 ms). The histidine ring nitrogen atoms coordinated to Zn^{2+} ions were identified from a 2D ^{15}N HMQC [15] using a nitrogen sweep width of 160-230 ppm. NOE (nuclear Overhauser effect) crosspeaks from 3D were identified ^{15}N-edited NOESY-HSQC [16] (t_{mix}=120 ms) and 3D ^{13}C-edited HMQC-NOESY [14] (t_{mix}=120 ms). The heteronuclear ^{15}N-^{1}H NOE data set was acquired as a pair of 2D datasets in an interleaved manner (where portions of each 2D were acquired sequentially until both datasets were completed) [17]. The first 2D dataset contained a 3 sec proton saturation (achieved with a series of 120° pulses) whereas the second 2D dataset contained a 3 sec delay. The heteronuclear

^{15}N-^{1}H NOE was then calculated as the ratio of the intensities for each peak in the two datasets. Each 2D dataset was acquired with 2048 (^{1}H) × 128 (^{15}N) complex points, 32 scans per point, and a 5 sec recycle delay. Error bars were estimated using the standard deviation of the background signal of each spectrum. T1 and T2 relaxation data were acquired as eight separate 2D datasets with randomly varying T1 (D7) and T2 (D20) delays [17]. Intensity decay was measured for each peak as a function of time in order to obtain the corresponding R1 and R2 values.

Structure Calculation

The protocol used for NMR structure calculation has been described previously [5]. Briefly, unique NOE distance restraints were classified into upper bounds of 2.7, 3.5, 4.5 and 5.5 Å and lower bound of 1.8 Å based on peak volumes. Backbone dihedral angles in the α-helical regions identified by the secondary Cα, Cβ, and Hα chemical shifts [11] were restrained to φ (-60±20°) and ψ (-40±20°). Initial structures were generated using CYANA [18], followed by molecular dynamics and simulated annealing in AMBER7 [19]; first *in vacuo*, then with the generalized Born (GB) potential. Initial structural calculations were performed in CYANA without the Zn^{2+} restraints to confirm that the zinc finger domain will fold from NOE-derived restraints only. Once the topology of the Zn^{2+}-coordinated residues were confirmed, subsequent CYANA structure calculations used distance restraints that imposed tetrahedral Zn^{2+}-coordination to Cys and His residues [14]. Iterative cycles of AMBER calculations were then performed followed by refinement of NMR-derived restraints until the structures converged with low restraint violations and good statistics in the Ramachandran plot. A family of twenty lowest energy structures were analyzed using PROCHECK [20] and molecular graphics generated using Pymol [21]. The

surface electrostatic potentials were calculated using APBS [22] and visualized them in Pymol [21].

In vitro Transcription

Overlapping PCR primer extension was used to assemble a dsDNA template representing the Andes and CCHFV M genomic segment panhandle (**Fig. 2-1A**) and used as a template for *in vitro* transcription (**Fig. 2-1B,C**). Briefly, forward and reverse primers corresponding to the dsDNA template were designed as a two-primer assembly with a 10-15 base overlap. The PCR reaction was prepared using equal molar concentrations of each primer and run for 30 cycles (Tm=95°C, Ta=52°C, and 1 minute extension time at 72°C). The assembled dsDNA template was detected using a 1.5% agarose gel stained with ethidium bromide. The DNA concentration was then measured by absorbance at 260nm. Due to the small size of the PCR product (approximately 50 nucleotides), PCR cleanup was not used prior to *in vitro* transcription.

In vitro transcription was carried out following manufacturer's protocol (MAXIscript Kit, Ambion). Briefly, a 20 uL reaction (2 ul 10X MAXIscript transcription buffer, 4 ul of dNTP mixture, 2 ul polymerase, 100ng template and ddH_2O up to 20 uL) was incubated for 1.5 hours (37°) and terminated by the addition of 2 uL 0.25 M EDTA and heating to 90° followed by rapid cooling on ice. Reaction mixtures were then treated with DNAse I (1 uL, MAXIscript Kit) and subjected to ethanol precipitation. Ethanol precipitation consisted of adding 1/10 volume 3M Na Acetate (pH 5.2), followed by two volumes of 100% ethanol. Incubation for three hours in

-20°C was followed by centrifugation at 13,000 rpm for 30 minutes at 4°C on a bench top microcentrifuge. The ethanol was gently siphoned out and the pellet washed with a solution of 70% ethanol, followed by further centrifugation at 13,000 rpm for five minutes. The ethanol solution was again gently siphoned out and the eppendorf tube opened and left on the bench for three hours or until all traces of ethanol had evaporated. The dry pellet was then resuspended in RNAse-free ddH_2O and analyzed for purity on a native 12% acrylamide gel stained with SYBR-GREEN II dye (Invitrogen).

EMSA experiments were conducted using both agarose and acrylamide matrices and using both Vero E6 ribosomal RNA and viral RNA transcripts as binding substrates. Initial experiments used only raw RNA extract from Vero E6 cells infected with the Andes virus (provided kindly by Dan Boudreaux from the lab of Dr. Stephen St. Joer, University of Reno, Nevada). Given the larger relative size of this RNA, a low percentage (.7%) agarose gel was used as the primary matrix. For later experiments, a native poly-acrylamide matrix (12%) was used for running smaller RNA transcripts.

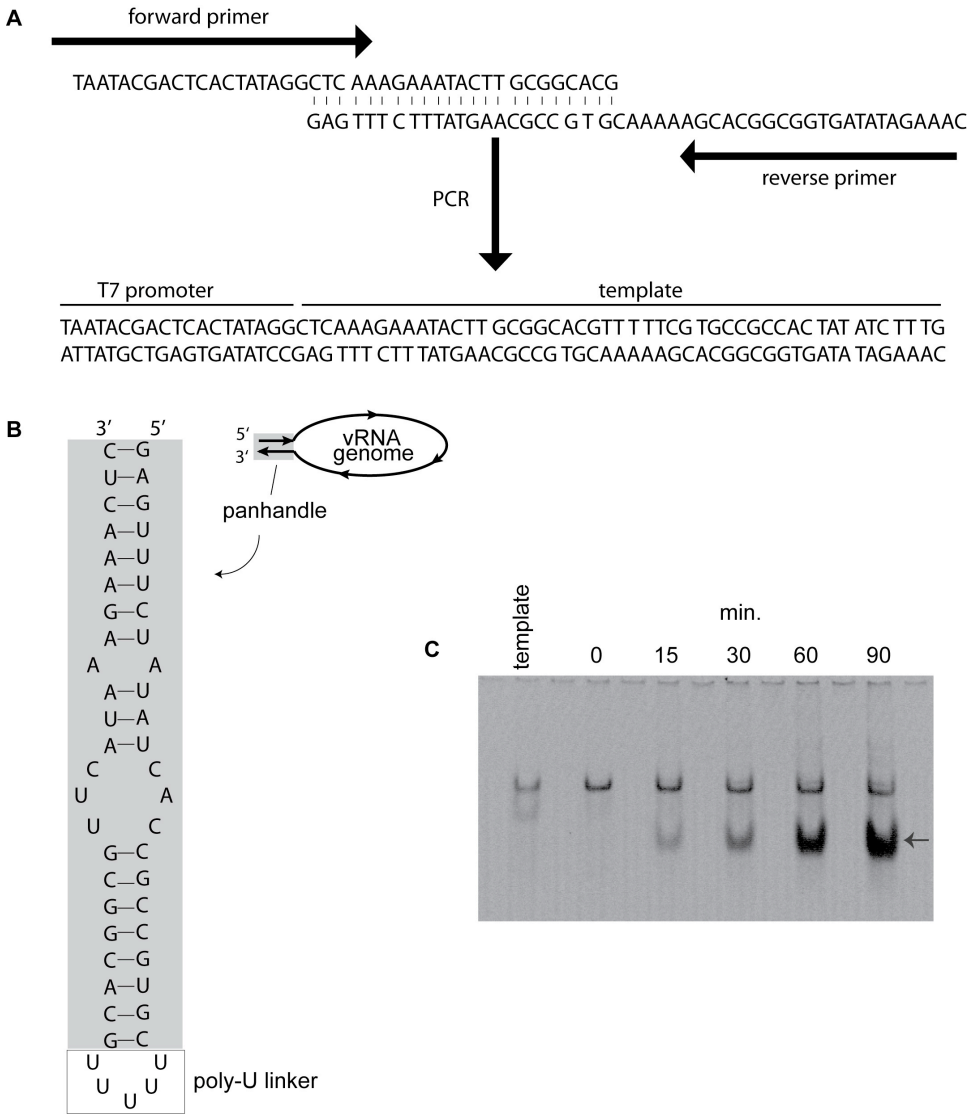

Figure 2-1. In vitro transcription of CCHF virus M segment panhandle RNA. **(A)** dsDNA template was PCR-amplified using a two primer overlap extension. The resulting template contained a T7 promoter sequence followed by the panhandle DNA template. **(B)** Predicted complimentary annealing of the 5' and 3' termini of the M segment produces a hairpin-like structure containing two mismatched bulges. The protein-encoding region of the panhandle segment was replaced by a poly-U linker. **(C)** Time scale measurement of a trial in vitro transcription reaction using the PCR generated template (arrow marks the band for the RNA product).

RNA Binding Assays

Andes Gn 534-610 - Andes virus (strain 23) was used to inoculate a T175 flask of confluent monolayer of Vero E6 cells at a multiplicity of infection of 0.1 pfu/cell and incubated for 14 days in a Biosafety Level 3 environment. Cells were harvested and total RNA was extracted using Trizol (Invitrogen #15596-018), ethanol precipitated as described above, and resuspended in water to a final concentration of 300 ng/uL. The presence of viral and cellular RNA was confirmed by RT-PCR. Total cellular and viral RNA was incubated at room temperature for 15 min with increasing amounts of Andes Gn zinc finger protein and a known RNA binding protein, PACT [23], (a gift from Dr. Gaya Amarasinghe, Iowa State University) in binding buffer (10 mM $NaPO_4$, 10 mM NaCl, pH 7.6). Samples were mixed with an equal volume 50% glycerol and loaded them in a 0.7% agarose Tris-Borate (TB) gel for electrophoretic mobility shift assay. The gel was run at 70 V for 50 min in TB buffer, pH 8.3. The gel was visualized by staining with SYBR Green II RNA specific dye (Invitrogen).

CCHFV Gn 729-805 – Wild-type and mutant forms of CCHFV Gn 729-805 were bound to smaller RNA transcripts corresponding to the semi-complementary 5' and 3' ends of the CCHFV RNA segments. The components of the poly-acrylamide gel for one gel cassette consisted of:

5 ml	2 X TB buffer (8.3)
4 ml	30 % (w/v) 29:1 Acrylamide : Bis-acrylamide Solution (National Diagnostics, order number EC-851)
1 ml	ddH_2O
100 ul	10 % APS
5 ul	TEMED
10 ml total volume	

The mixture was poured into an empty gel cassette (Invitrogen) and covered with a 12 well comb. Polymerization occurred after approximately 30 minutes, but a minimum of one hour was allowed prior to using the gel.

The polymerized gel was loaded onto a Novex XCell SureLock Electrophoresis Cell and pre-run it at 90 volts in 1 X Tris Borate (8.3) buffer for 1 hour. The intact apparatus was placed in a cooling water bath in order to avoid heating of the binding mixture during electrophoresis.

RNA transcripts (M segment CCHFV RNA panhandle for CCHFV Gn 729-805; and M segment Andes virus panhandle for Andes Gn 534-610) were incubated on ice (15 min.) with increasing amounts of respective protein. The final mixture components were adjusted to include one third by volume 50% glycerol and extra binding buffer (30 mM $NaPO_4$, 30 mM NaCl) was added to a final volume of 20 ul. The mixture was then loaded onto the gel along with a 100 bp DNA ladder (NEB, # NO467S). The gel was run for one hour at 100 volts, then stained the gel for 30 minutes using SYBR Green II RNA specific dye (Invitrogen).

CD Spectroscopy

CD spectra were collected in triplicate at 25° on a JASCO J-815 Spectro-polarimeter using a scanning speed of 50 nm/min. Protein concentrations were kept at approximately 1 μM in buffer (10 μM $NaPO_4$, 10 μM NaCl, 0.1 mM $ZnSO_4$). EDTA and $ZnSO_4$ titrations were applied to the same sample.

References

1. Huth, J.R., et al., *Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR.* Protein Sci., 1997. **6**(11): p. 2359-2364.
2. Geisbrecht, B.V., S. Bouyain, and M. Pop, *An optimized system for expression and purification of secreted bacterial proteins.* Protein Expr. Purif., 2006. **46**(1): p. 23-32.
3. Wang, Y., et al., *NMR structure of the N-terminal coiled coil domain of the Andes hantavirus nucleocapsid protein.* J. Biol. Chem., 2008: p. in press.
4. Chatterjee, S., et al., *The Crystal Structure of the Salmonella Type III Secretion System Tip Protein SipD in Complex with Deoxycholate and Chenodeoxycholate.* Protein Sci. in press, 2010.
5. Estrada, D.F., et al., *The Hantavirus Glycoprotein G1 Tail Contains Dual CCHC-type Classical Zinc Fingers.* J Biol Chem, 2009. **284**(13): p. 8654-60.
6. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* J. Biomol. NMR, 1995. **6**(3): p. 277-293.
7. Johnson, B.A., *Using NMRView to visualize and analyze the NMR spectra of macromolecules.* Methods Mol. Biol., 2004. **278**: p. 313-352.
8. Grzesiek, S. and A. Bax, *The importance of not saturating H_2O in protein NMR. Application to sensitivity enhancement and NOE measurements.* J. Am. Chem. Soc., 1993. **115**: p. 12593-12594.
9. Grzesiek, S., et al., *^{1}H, ^{13}C, and ^{15}N NMR backbone assignments and secondary structure of human interferon-gamma.* Biochemistry, 1992. **31**(35): p. 8180-8190.
10. Wittekind, M. and L. Mueller, *HNCACB, a high sensitivity 3D NMR experiment to correlate amide proton and nitrogen resonances with the alpha-carbon and beta-carbon resonances in proteins.* J. Magn. Reson., 1993. **101B** p. 201-205.
11. Wishart, D.S. and A.M. Nip, *Protein chemical shift analysis: a practical guide.* Biochem. Cell Biol., 1998. **76**(2-3): p. 153-163.
12. Tolman, J.R., J. Chung, and J.H. Prestegard, *Pure-phase heteronuclear multiple-quantum spectroscopy using field gradient selection.* J. Magn. Reson., 1992. **98** p. 462-467.
13. Grzesiek, S. and A. Bax, *Amino acid type determination in the sequential assignment procedure of uniformly $^{13}C/^{15}N$-enriched proteins.* J. Biomol. NMR, 1993. **3**(2): p. 185-204.
14. Hoffman, R.C., et al., *A simple method for the refinement of models derived from NMR data demonstrated on a zinc-finger domain from yeast ADR1.* J. Magn. Reson. Ser. B, 1993. **102**(1): p. 61-72.
15. Pelton, J.G., et al., *Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signal-transducing protein from Escherichia coli, using two-dimensional heteronuclear NMR techniques.* Protein Sci., 1993. **2**(4): p. 543-558.
16. Marion, D., et al., *Overcoming the overlap problem in the assignment of ^{1}H NMR spectra of larger proteins by use of three-dimensional heteronuclear ^{1}H-^{15}N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta.* Biochemistry, 1989. **28**(15): p. 6150-6156.

17. Stone, M.J., et al., *Backbone dynamics of the Bacillus subtilis glucose permease IIA domain determined from ^{15}N NMR relaxation measurements.* Biochemistry, 1992. **31**(18): p. 4394-4406.
18. Guntert, P., *Automated NMR structure calculation with CYANA.* Methods Mol. Biol., 2004. **278**: p. 353-378.
19. Case, D.A., et al., *AMBER7*. 2002, University of California, San Francisco.
20. Laskowski, R.A., et al., *AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR.* J. Biomol. NMR, 1996. **8**(4): p. 477-486.
21. DeLano, W.L., *The PyMOL Molecular Graphics System* 2002, San Carlos, California, U.S.A.: DeLano Scientific.
22. Baker, N.A., et al., *Electrostatics of nanosystems: application to microtubules and the ribosome.* Proc. Natl. Acad. Sci. U.S.A., 2001. **98**(18): p. 10037-10041.
23. Patel, R.C. and G.C. Sen, *PACT, a protein activator of the interferon-induced protein kinase, PKR.* EMBO J., 1998. **17**(15): p. 4379-90.

Chapter 3: NMR Structure of the Andes *Hantavirus* Gn Zinc Finger Domain

Introduction

Many viruses in the family Bunyaviridae, which consists of five genera (*Hantavirus, Orthobunyavirus, Nairovirus, Phlebovirus*, and *Tospovirus*), cause emerging zoonotic infections in humans [1]. Examples are the La Crosse Encephalitis Orthobunyavirus, Rift Valley Fever *Phlebovirus*, and the Crimean-Congo Hemorrhagic Fever *Nairovirus* (*Tospoviruses* are plant pathogens). *Hantaviruses* use rodents as their primary reservoir, and although some (*e.g.* Prospect Hill virus) are nonpathogenic to humans, others (*e.g.* Andes virus) can cause either the *Hantavirus* Cardiopulmonary Syndrome or the Hemorrhagic Fever with Renal Syndrome in humans [1]. Annually, over 150,000 cases of Hantaviral infections are reported worldwide [2] with mortality rates reaching as high as 40% [30].

Bunyaviridae viruses are enveloped and have three genomic RNA molecules: the small (S), medium (M), and large (L) segments. These segments encode for four viral proteins: the RNA polymerase, the nucleocapsid (N) protein, and the membrane glycoproteins, Gn and Gc [1]. The ectodomains of Gn and Gc are glycosylated, form a multimeric complex on the viral surface, and function as the viral spike proteins [1,50]. In Gn and G2, the N termini form the ectodomains, followed by single pass transmembrane helices, then the C termini or cytoplasmic tails project within the virions. Bunyaviridae viruses lack a matrix protein [2], which typically link the membrane to the ribonucleoprotein among enveloped viruses [3]. Based on this observation, it was suggested that the cytoplasmic tails of the viral glycoproteins might bind the viral ribonucleoprotein [33]. Indeed, recent results have

shown that the Gn tail binds the viral ribonucleoprotein in phlebovirus [12] and is required for packaging the genome in orthobunyavirus [11]. These data suggest that the Gn tail plays a critical role in viral assembly.

Other reports suggest that among *Hantaviruses*, the Gn tail is important in host-pathogen interaction. The Gn tail of human pathogenic *Hantaviruses* inhibits the cellular interferon response [18, 19] against viral infection by disrupting protein-protein interactions [19]. In nonpathogenic *Hantaviruses*, by contrast, the interferon response is activated [18, 34]. The Gn tail contains conserved immunoreceptor tyrosine-based activation motifs, which are involved in protein-protein interactions in the cellular immune response to viral infection [15, 16]. Furthermore, the Gn tail of pathogenic *Hantaviruses* is ubiquitinated and proteasomally degraded [15], which is thought to regulate the activity of the Gn tail [15], whereas the nonpathogenic *Hantavirus* Gn tail is stable and is not degraded.

The Gn tail varies in length from 78 residues in orthobunyaviruses to 142 residues among *Hantaviruses*. Sequence alignment shows a region of conserved cysteine and histidine residues in the Gn tail of Bunyaviridae. This region was predicted to form a RING finger motif in the Gn tail of *Hantavirus* [17]. Here, we show by NMR that the conserved cysteine/histidine region in the Gn tail of *Hantaviruses* forms a dual ββα-fold zinc fingers [35-38] and not a RING finger structure as suggested earlier [17]. We also discuss the implication of our structural findings of the *Hantavirus* Gn tail in the context of viral assembly and host pathogen interaction.

Results

Protein Expression—Sequence analysis of the Gn cytoplasmic tail of *Hantaviruses* revealed two highly conserved $CX_2CX_{12–13}$-HX_3C (where *X* is any amino acid) motifs, which suggested the presence of two CCHC-type zinc fingers (**Fig. 3-1**). Expression of the Andes virus Gn zinc fingers (residues 543–599) in *E. coli* resulted in cell death, with cell density reaching only A_{600} of ~0.9 after induction at A_{600} of ~0.8, suggesting that the zinc finger was toxic to *E. coli*. Thus, the zinc finger domain was expressed as a GB1 fusion protein. The GB1 tag contained His_6 for nickel affinity purification and a TEV protease cleavage site to recover the native Gn zinc finger domain. The fusion protein was over-expressed in soluble form in *E. coli*, purified under native conditions, and digested with TEV protease to obtain the Gn zinc finger domain.

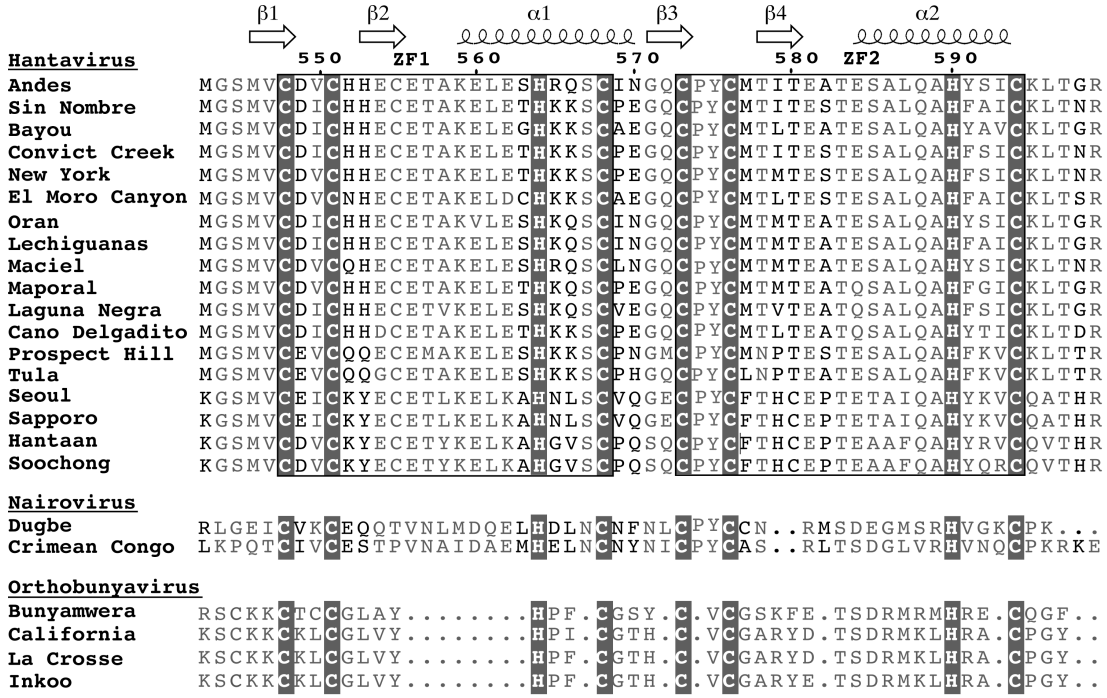

Figure 3-1. The Gn tail of *Hantaviruses*, *Nairoviruses*, and *Orthobunyaviruses* (genera of Bunyaviridae), contains a cysteine/histidine-rich region with two CCHC-arrays. Structure determination of the Andes virus dual CCHC-region revealed a novel zinc finger domain. Shown are the secondary structures (α-helices and β-strands), zinc-coordinating residues (*blocked*), the two CCHC motifs (*boxed*), conserved residues (*gray*), and residue numbers for the Andes virus Gn sequence. Sequence alignment was generated using CLUSTALW and formatted with ESPript 2.2 (Gouet *et al.*, *Bioinformatics*, 1999, 15, 305-308).

Zn^{2+} Is Required for Proper Folding—CD spectrum of the Andes virus Gn zinc finger showed a α-helical domain, with local minima at 209 and 222 nm (**Fig. 2-2**). Titration of EDTA to a final concentration of 1.25 mM caused a spectral shift to 205 nm, indicating a partial loss of helical structure. However, the minimum at 222 nm remained despite EDTA treatment, suggesting that although the global fold is disrupted by removal of zinc ion, some residual helical structure remained. Subsequently, titrating $ZnSO_4$ back into the solution resulted in partial recovery of α-helical content, suggesting restoration of the global fold.

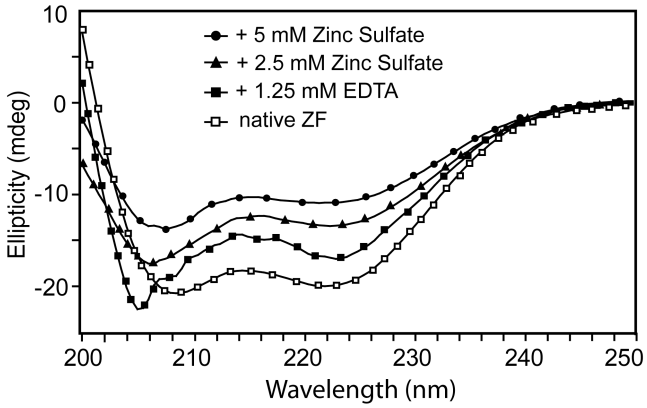

Figure 3-2. CD spectroscopy and titration with EDTA and $ZnSO_4$ of recombinant Andes virus Gn tail CCHC-region (residues 543-599), which was expressed and purified under native conditions, showed that Zn^{2+}-binding is required for the proper folding of this domain. Native Gn tail zinc finger domain showed a folded CD spectrum (*open squares*). Titration with an excess of EDTA resulted in a shifted minimum (at 204 nm) and reduced the helical minimum (at 222 nm) (*closed squares*). Addition of 2.5 and 5.0 mM $ZnSO_4$ yielded folded CD spectra (*triangles and circles*).

NMR data were also used to confirm the requirement for Zn^{2+} coordination on the proper folding of the zinc finger domain. The Andes virus and the Prospect Hill virus zinc-binding domains purified under native conditions showed well dispersed and sharp peaks in their two-dimensional ^{1}H-^{15}N HSQC spectra (**Figure 3-3**). After treatment with excess EDTA, peaks in the HSQC of the Andes virus zinc-binding domain deteriorated, showed residual peaks that differ markedly from each other with respect to peak intensities and sharpness, and displayed a collapse of the amide side chains (**Figure 3-3A**). In the presence of

excess EDTA, the HSQC spectrum of the Prospect Hill virus showed a collapse of the backbone and side chain amide peaks (**Figure 3-3B**), which indicated that the protein was unfolded. The Prospect Hill Virus Gn zinc finger domain is characterized further in Chapter 4.

NMR Structure Determination—We determined the NMR structure of the Andes virus zinc finger domain. The Andes virus Gn zinc finger domain showed an excellent well dispersed two-dimensional ^{1}H-^{15}N HSQC (**Fig. 3-4**). Nearly complete backbone assignments were obtained from three-dimensional HNCA, CBCA(CO)NH, HNCACB, and ^{15}N-edited NOESY-HSQC. Analysis of the C^{α}, H^{α}, C^{β}, and C′ secondary chemical shifts (**Fig. 3-5**) supported the presence of two short α-helices and two random coil regions flanking the central domain [32]. Side chain assignments were completed using two-dimensional ^{1}H-^{13}C HMQC (**Fig. 3-6**), three-dimensional HBHA-(CO)NH, and three-dimensional ^{13}C-edited HMQC-NOESY. There were four conserved histidines (at positions 552, 553, 564, and 590) that could potentially coordinate Zn^{2+} ion; however, long distance NOEs (between Cys^{548}-His^{564} and Cys^{573}-His^{590}) indicated that His^{564} and His^{590} were involved in Zn^{2+} coordination. A two-dimensional ^{15}N HMQC [33] spectrum showed Zn^{2+} coordination through the His^{564} $N^{\delta 1}$ and His^{590} $N^{\varepsilon 2}$ atoms (**Fig. 3-7**). Manual analysis of three-dimensional ^{15}N- and ^{13}C-edited NOESY spectra identified 859 unambiguous interproton distance restraints. The NOE restraints together with 24 phi and 15 psi dihedral angle restraints and zinc coordination restrains (**Table 3-1**) were used in structure calculation and refinement with CYANA [48] and AMBER [49]. The 20 lowest energy NMR structures converged into a family of structures (**Fig. 3-8A**) with low restraint violations and good Ramachandran plot statistics (**Table 3-1**).

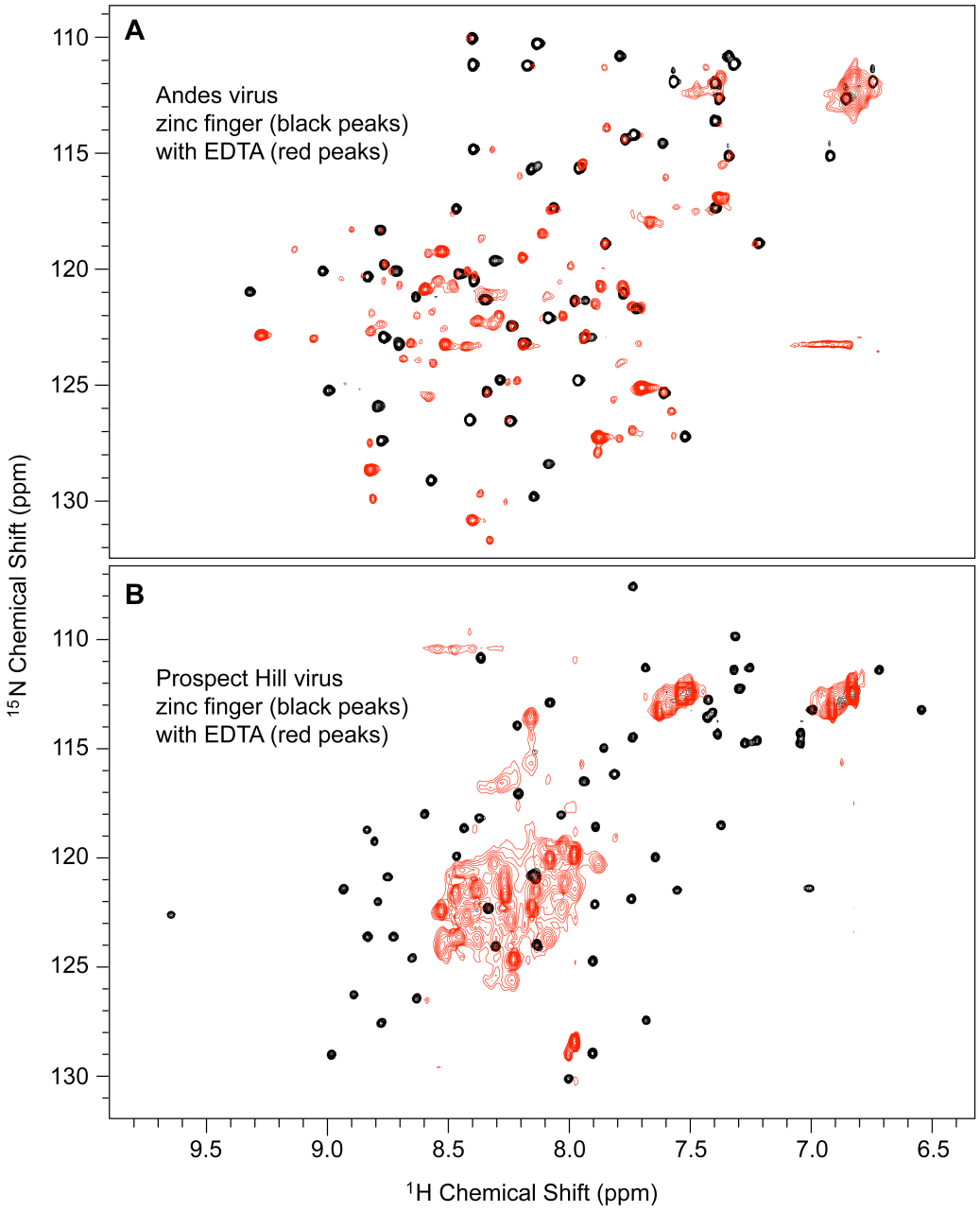

Fig. 3-3. The proper folding of the *Hantavirus* Gn tail zinc finger domain depends on Zn^{2+}-binding. ^{1}H-^{15}N HSQC spectra of **(A)** Andes virus and **(B)** Prospect Hill virus Gn zinc finger domains before (black peaks) and after (red peaks) addition of 4 mM EDTA. In the Andes zinc finger, addition of EDTA resulted in the deterioration of the HSQC peaks, with peaks that are sharp and other peaks with reduced intensities, as well as collapse of asparagine and glutamine side chain resonances. In the Prospect Hill virus, addition of EDTA collapsed the backbone and side chain peaks (at 7.8-8.6 and 6.8-7.6 ^{1}H ppm, respectively) into a characteristic HSQC of an unfolded protein. We determined the NMR structure of the Gn zinc finger domain from the Andes virus, which is pathogenic to humans (whereas Prospect Hill virus is not).

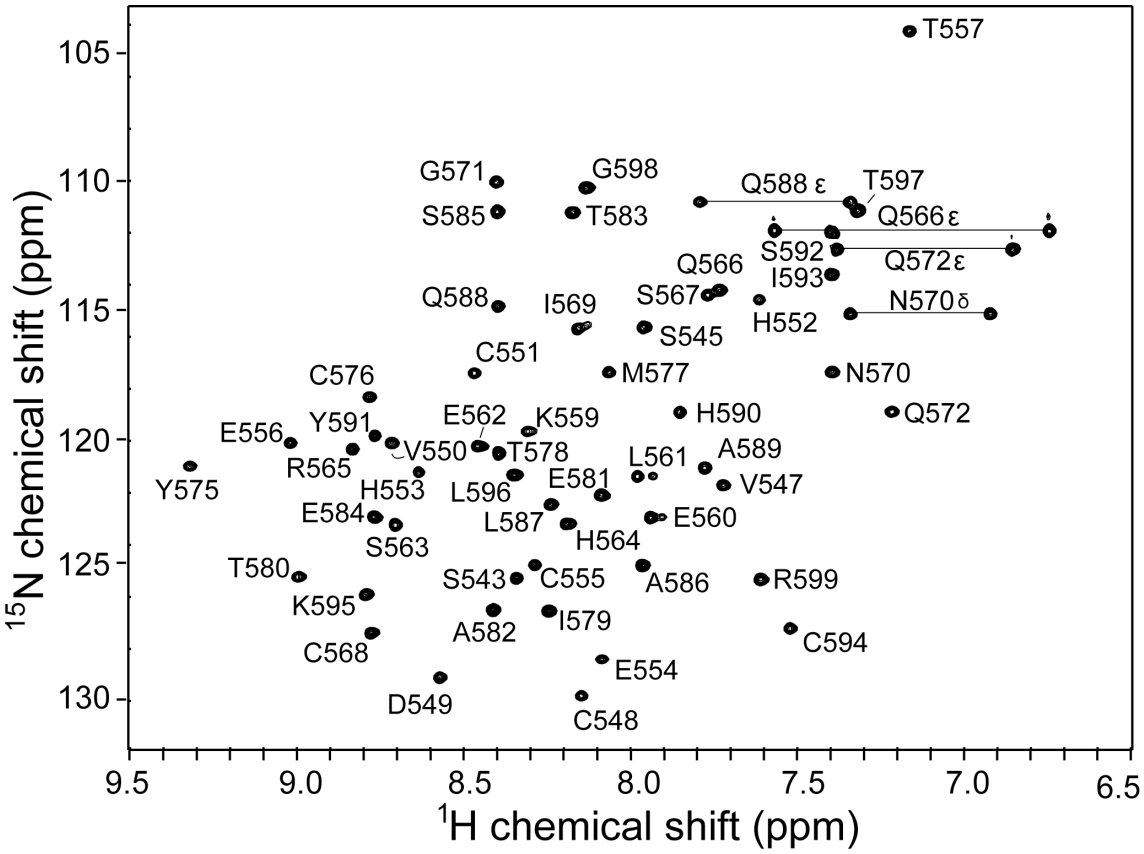

Figure 3-4. The Andes virus Gn tail zinc finger domain (residues 543-599) shows a well dispersed two-dimensional ^{1}H-^{15}N HSQC spectrum, which facilitated the acquisition of additional NMR datasets and allowed resonance assignments and NMR structure determination of this domain. Shown are the complete assignments for backbone and side chain amides.

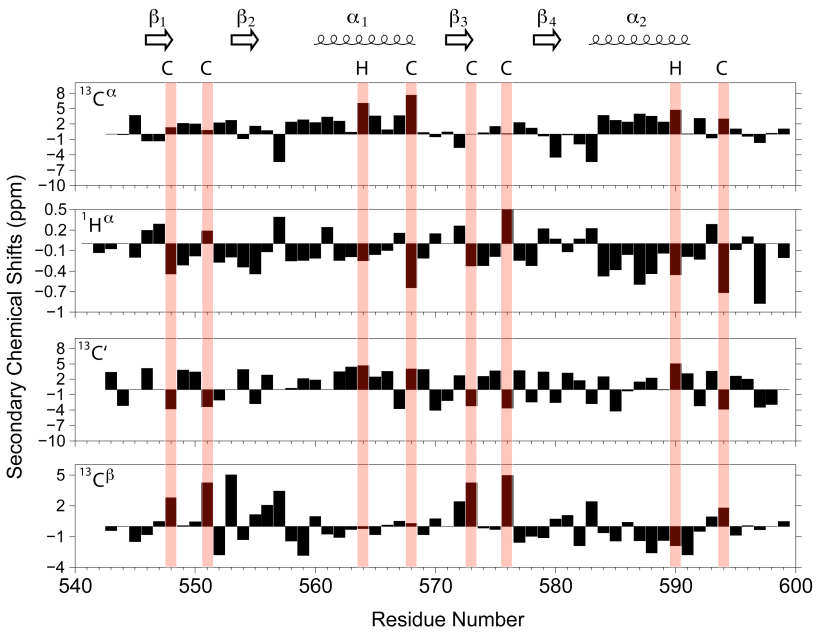

Figure 3-5. Secondary Cα, Hα, C', and Cβ chemical shifts of the Andes virus Gn zinc finger domain. Shown are the CCHC-zinc coordination ligands (red bars) and the secondary structures (β-strands and α-helices).

Table 3-1. Restraints and structural statistics for 20 NMR structures.

Total NOE-derived distance restraints	859
Intraresidue (i,i)	323
Sequential (i, i+1)	207
Medium Range (2≤=|i-j|≤=4)	161
Long Range (|i-j| > 4)	168
Total dihedral angle restraints	39
Phi ϕ	24
Psi ψ	15
RMS difference from mean structure	
Backbone atoms (N,Cα,C') (Å)	0.20
All heavy atoms (C,N,O) (Å)	0.64
Violation analysis	
Max distance violation (Å)	0.5
Max dihedral angle violation (°)	40
Energies	
mean GB-AMBER energy (kcal mol^{-1})	-2113
mean restraint energy (kcal mol^{-1})	13
Ramachandran plot	
Most favorable region	81.5%
Additionally allowed regions	18.0%
Generously allowed regions	0.3%
Disallowed regions	0.1%

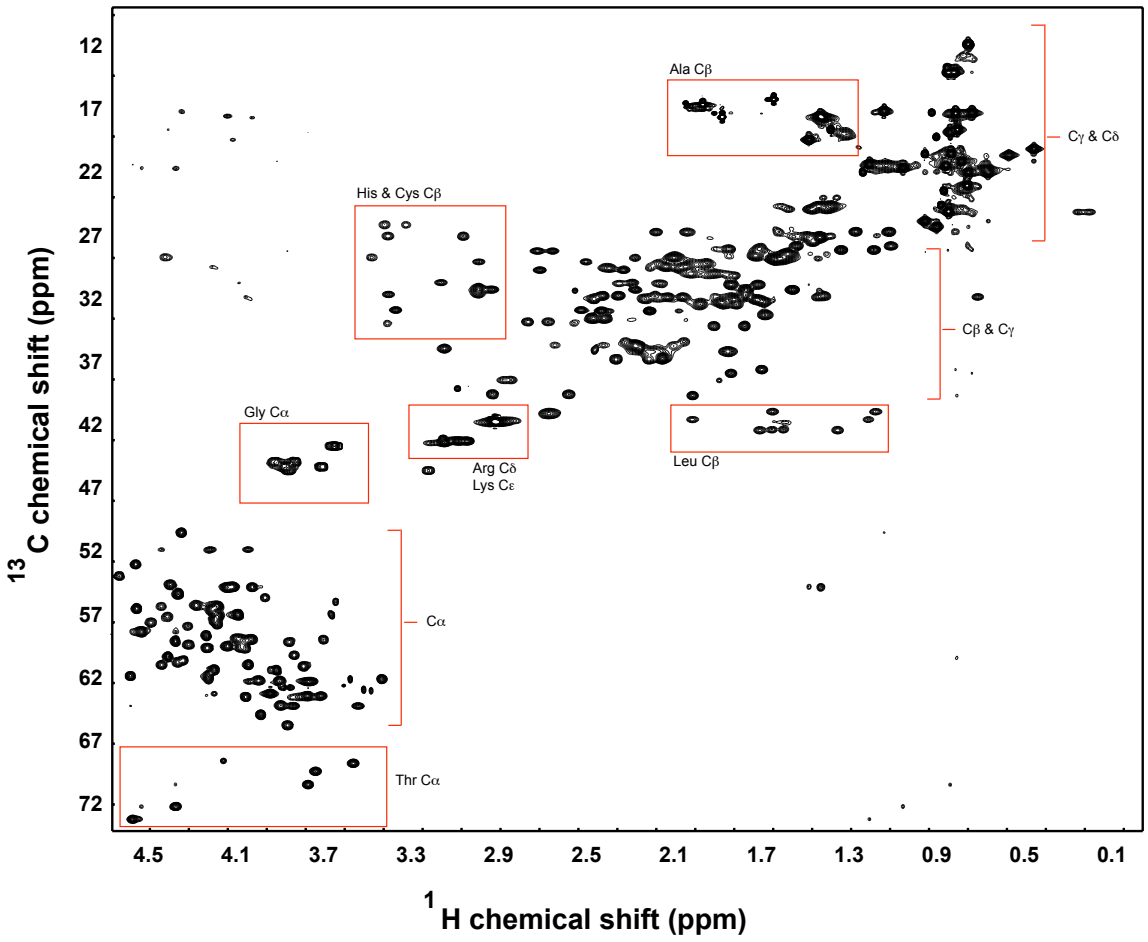

Figure 3-6. 2D ^{1}H-^{13}C HMQC spectrum of Andes Gn 534-599. Sample used was ^{15}N-^{13}C labeled protein lyophilized and resuspended in 100% D_2O. Excellent dispersion in both dimensions greatly facilitated side chain assignments.

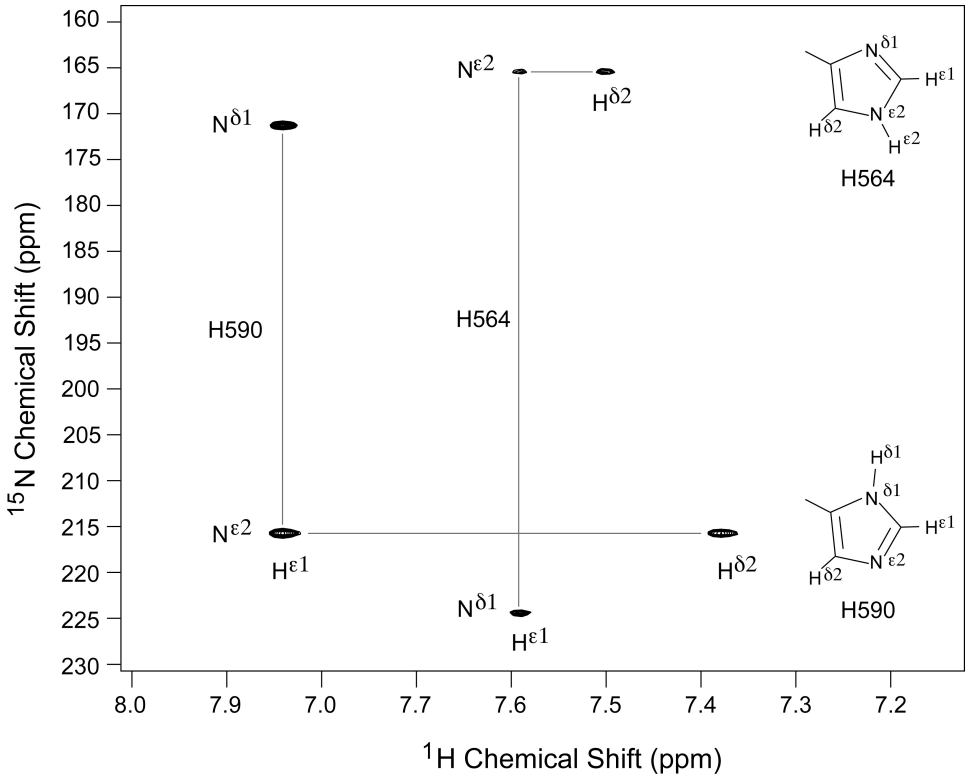

Fig. 3-7. The tautomeric states of the Zn^{2+}-coordinating histidines (His^{590} and His^{564}) were determined by 2D ^{1}H-^{15}N HMQC following the method of Pelton *et al.* (*Protein Sci.*, 1993, 2, 543-558). This spectrum was acquired using an ^{15}N-labeled protein in buffer (10 mM $NaPO_4$, pH 7.0, 10 mM NaCl, 1 mM DTT, 0.1 mM $ZnSO_4$) in 100% D_2O with the following acquisition parameters: ^{15}N carrier frequency (195 ppm), ^{15}N sweep width (70 ppm), number of scans (32), number of ^{15}N complex points (128).

Structure of Individual Zinc Finger—Each Gn zinc finger folded similarly to the ββα fold of classical zinc fingers. In the first CCHC array (ZF1), residues Met^{546}–Cys^{555} formed a β-hairpin that encompassed the first two coordinating cysteines (Cys^{548} and Cys^{551}; **Fig 3-8B, left zinc in Figure**). Asp^{549} and Val^{550} formed the loop apex with the coordinating cysteines on either side of the β-hairpin. The structured region terminated at Lys^{559}, where helix α_1 began, and folded back toward the β-hairpin and allowed the completion of ZF1 with His^{564} and Cys^{568} on the interior face of helix α_1.

In ZF2, the β-hairpin (Gly^{571}–Thr^{580}) contained the first two coordinating cysteines (Cys^{573} and Cys^{576}; **Fig. 3-8, B and C, right zinc in Figure**). The coordination site on the loop was partly formed by the positioning of Pro^{574} between the two. Strong Cys^{573} H^{α} to Pro^{574} H^{δ} NOEs indicated that Pro^{574} was in the *trans* configuration. A structured loop followed the β-hairpin and terminated at Glu^{584}, where helix α_2 began, and folded back toward the β-hairpin to complete ZF2 by coordinating His^{590} and Cys^{594} to the Zn^{2+} ion.

Hydrogen Bonding in the β hairpin – Each β hairpin turn is stabilized by a network of hydrogen bonding, as revealed by exchange (**Fig. 3-9**). The backbone amides of the second coordinating cysteines in each ZF (Cys^{551} and Cys^{576}, repsectively) and its immediately preceding residue fold back to form hydrogen bonds with the first coordinating cysteine side chain (Cys^{548} and Cys^{573}) in each zinc finger. The non-coordinating residues involved in hydrogen bonding are either highly conserved (Val^{550}) or completely conserved (Tyr^{575}) in *Hantaviruses* (**Fig. 3-1**).

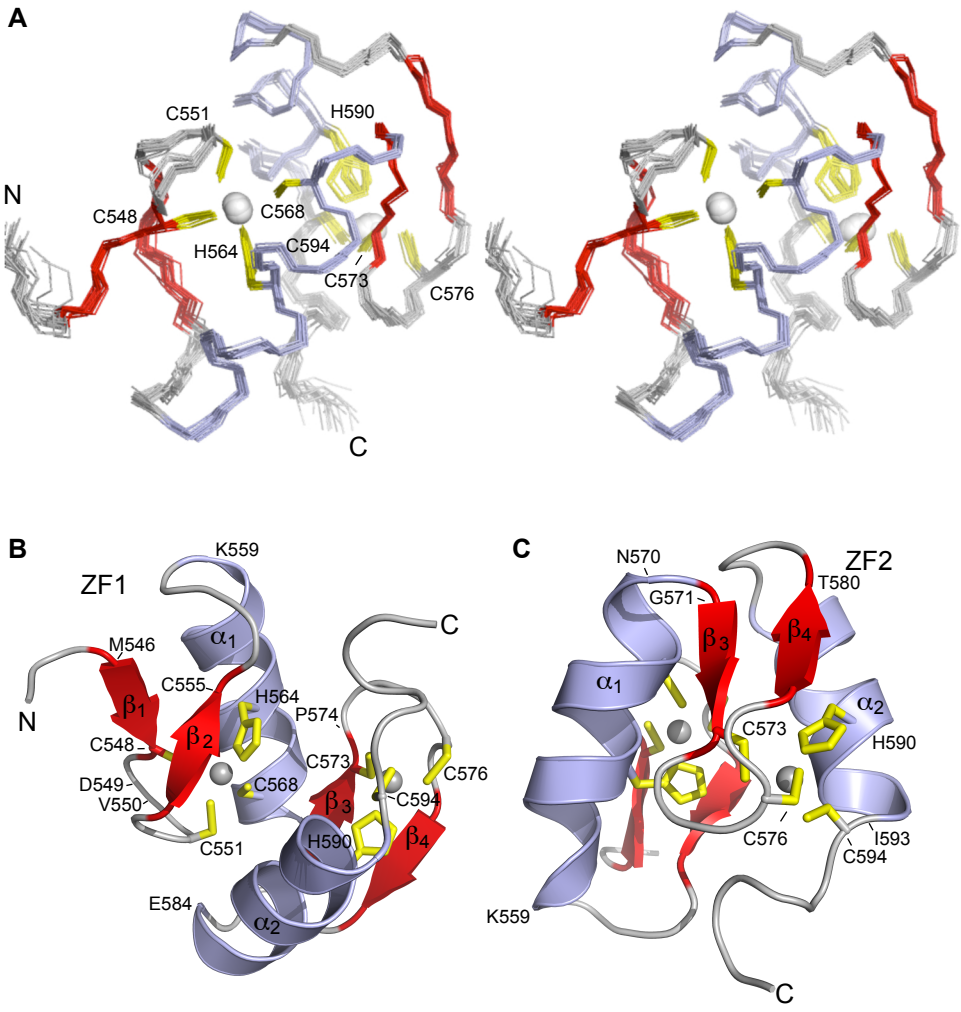

Figure 3-8. The NMR structure of the Andes virus Gn tail zinc binding domain reveals two classical ββα fold zinc fingers that are joined together. **(A)** Stereoview of the superposition of 20 lowest energy NMR structures, and ribbon structures of the lowest energy NMR structure showing the residues involved in the **(B)** first (ZF1 on left) and **(C)** second (ZF2 on right) zinc fingers. Shown are the cysteine and histidine residues (yellow) that coordinate Zn^{2+} ions (gray) as well as the secondary structures (α_1-α_2, β_1-β_2). The dual Hantaviral Gn zinc fingers interact with each other and form a single domain with a novel fold as revealed by DALI [7] and TM-align [8] structural homology searches.

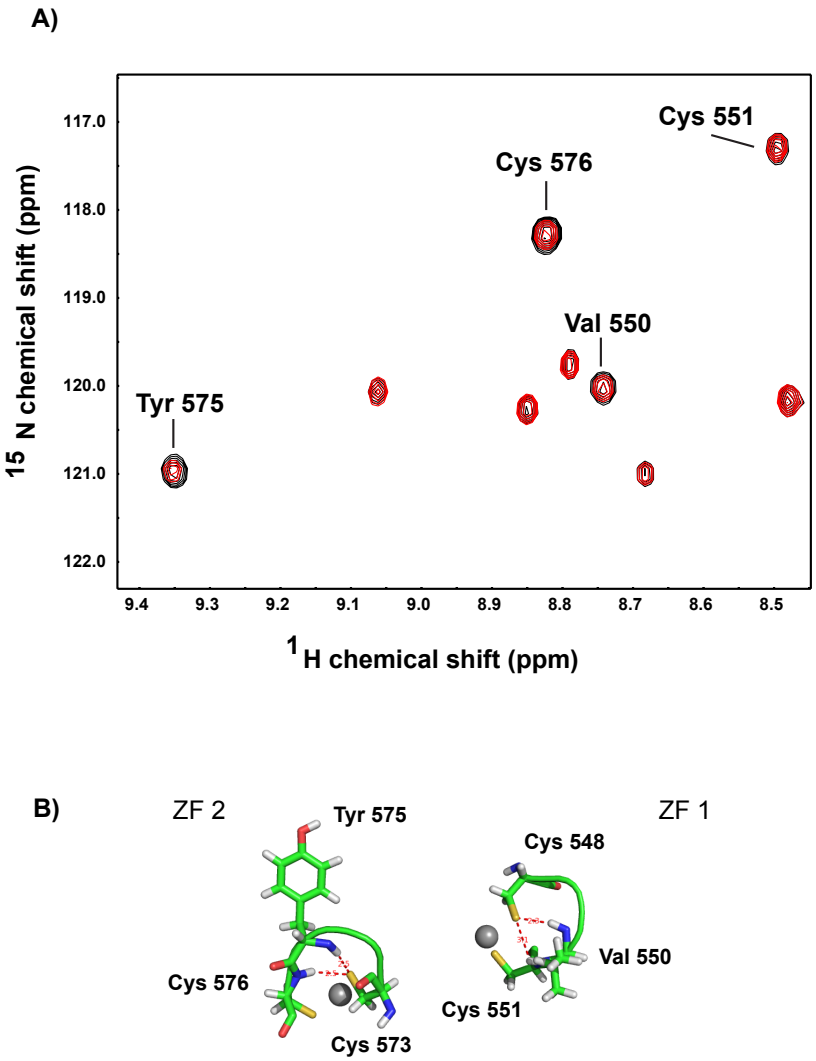

Fig. 3-9. Hydrogen bonding data acquired by hydrogen-deuterium exchange (**A**) reveals a pattern of hydrogen bond formation at the apex of each zinc finger ß hairpin. (**B**) The backbone amide protons of the second coordinating Cys and the preceding residues (Val550 & Cys551 in ZF 1 and Tyr575 & Cys576 in ZF 2), converge to form a network of hydrogen bonds (B, in red) with the side chain of the corresponding first coordinating residue (Cys548 and Cys573, respectively). This hydrogen bond network stabilizes the tight turn at each zinc finger apex.

Mutations of Zn^{2+}-Coordinating Residues—To confirm the Zn^{2+} coordination topology indicated by the NMR structure, we created conservative point mutants at each of the cysteine and histidine residues within the Andes virus zinc finger domain. In order to preserve the polarity and size of the side chains, serines were substituted for cysteines and

phenylalanines were substituted for histidines. Of the 8 residues expected to coordinate zinc, only C594S and H590F expressed as soluble proteins, the remaining constructs (C548S, C551S, H564F, C568S, C573S, and C576S) could be expressed only as inclusion bodies despite the presence of the GB1 solubility tag. This result suggested that zinc coordination was necessary for stabilizing the overall fold of the zinc finger domain. For further analysis, all of the inclusion bodies were solubilized overnight in 8 M urea, purified by nickel affinity chromatography, and refolded by stepwise dialysis to remove urea. Refolding of the zinc finger domain was determined by the proper refolding of the attached GB1 tag using two-dimensional ^{15}N HSQC, which served as a control to show that the refolding conditions would have properly refolded a native protein. The spectra of the mutated Gn zinc finger domain in the GB1 fusion proteins consisting of C548S, C551S, H564F, C573S, C576S, H590F, and C594S all showed narrowly dispersed spectra consistent with an unfolded domain (**Figure 3-10**). These results suggested that the two zinc fingers did not fold independently of each other (**Figure 3-10**). Of these eight positions, only C568S showed any peak dispersion at all (**Figure 3-10**). In each instance, the peaks corresponding to the attached GB1 tag were well dispersed, thus indicating that the fusion protein was refolded properly (**Figure 3-10**). These results suggested that, in the dual zinc finger domain, mutation of a Zn^{2+}-coordinating residue in either ZF1 or ZF2 lead to the unfolding of the entire domain.

Mutations of Non-Zn^{2+}-Coordinating Residues—The domain contains three histidines (His^{552}, His^{553}, and His^{542}, the cloning artifact) and a cysteine (Cys^{555}) (**Fig. 3-1**) that are not involved in Zn^{2+} coordination. To eliminate the possibility that Zn^{2+} could be coordinated to these other cysteine and histidine residues, we generated four additional point mutants corresponding to H542F, H552F, H553F, and C555S. Three of the four mutants (H552F, H553F, and H542F) gave a dispersed spectrum consistent with a folded domain (**Figure 3-11**). Only the C555S mutant gave an unfolded spectrum (**Figure 3-11**). Analysis of the structure reveals that the side chain of Cys^{555} was oriented toward the interior of the structure

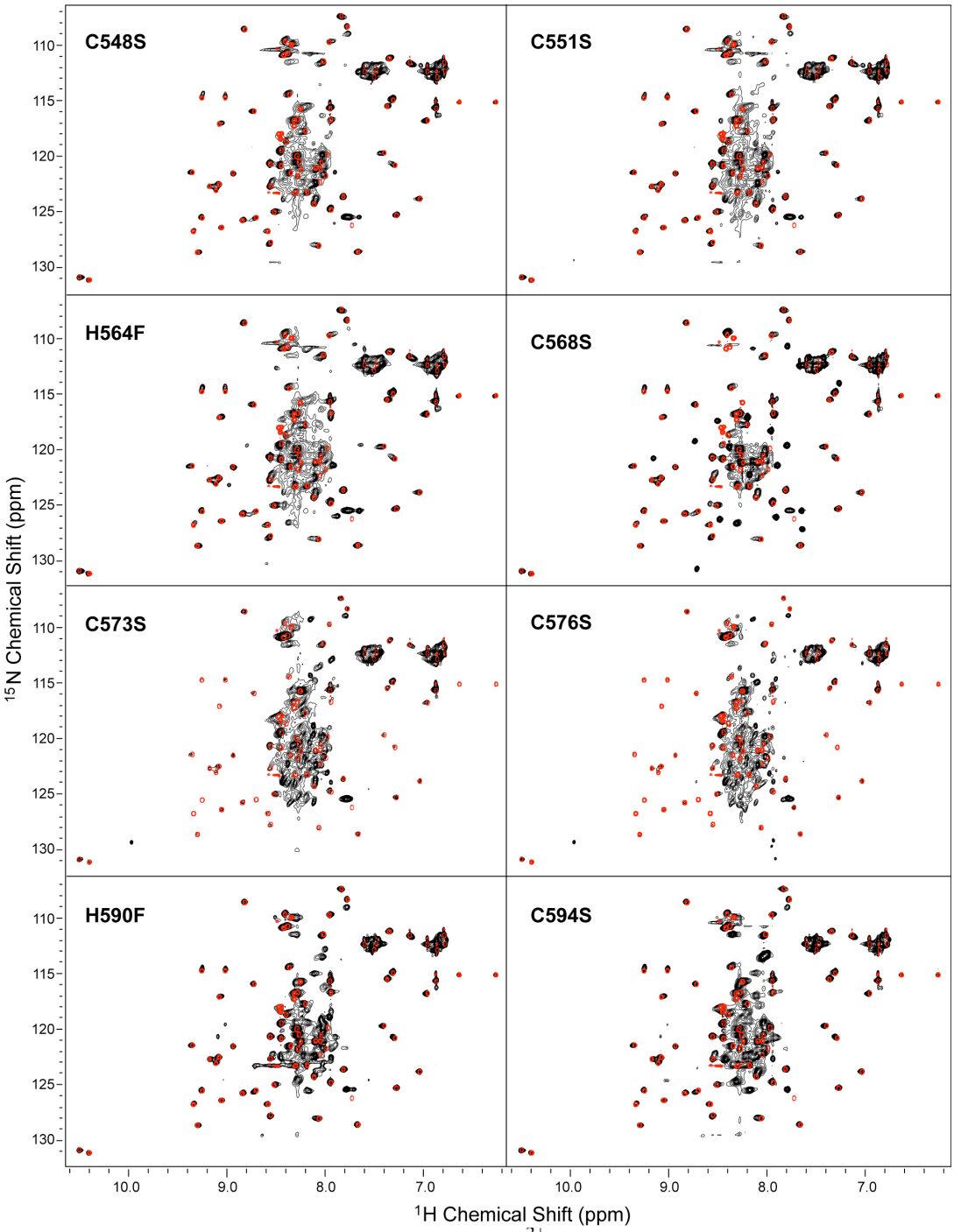

Fig. 3-10. Effect of point mutations in the Zn^{2+}-coordinating cysteine and histidine residues. ^{1}H-^{15}N HSQC spectra of GB1-zinc finger domain fusion proteins with point mutations in the eight Zn^{2+}-coordinating residues. Cysteine was mutated to serine and histidine to phenylalanine. Spectra of GB1 fusion proteins (black peaks) are overlayed with the spectrum of free GB1 tag (red peaks)

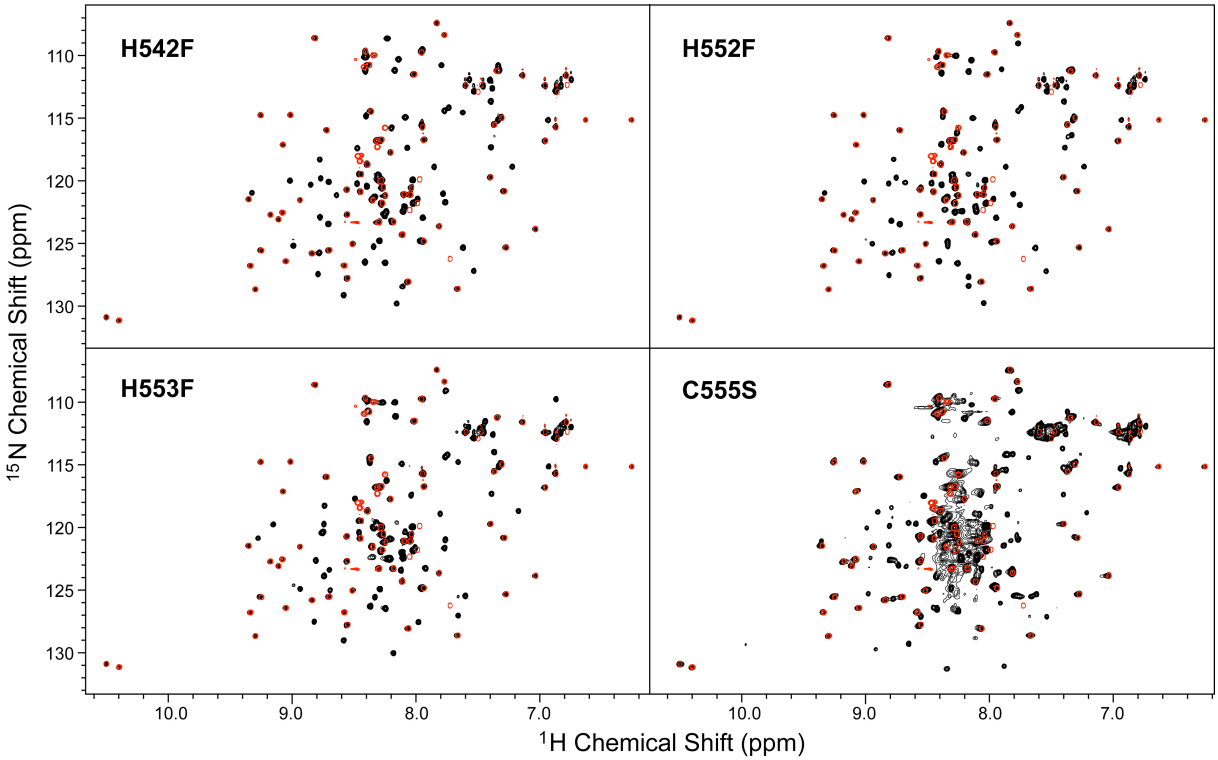

Fig. 3-11. Effect of point mutation in histidine and cysteine residues that do not coordinate Zn^{2+} ion. The 2D ^{1}H-^{15}N HSQC spectra of histidine point mutants (H542F, H552F, and H553F) showed the characteristic well dispersed and distinct peaks of a folded protein. The spectrum for the C555S mutant also showed well dispersed peaks, however, there is a region in the middle of the spectrum that is characteristic of an unfolded protein. C555 is a non-Zn^{2+}-coordinating residue, however, it forms part of the hydrophobic core, and the C555S mutation caused an unfolding in some regions of the zinc finger domain. Shown are the ^{1}H-^{15}N HSQC spectra of GB1-zinc finger fusion proteins (black peaks) overlayed with the spectrum of free GB1 tag (red peaks).

and therefore played a role in stabilizing the hydrophobic core of the overall domain. These data hence confirmed that Zn^{2+} was coordinated to the predicted zinc finger residues (Cys^{548}, Cys^{541}, Cys^{568}, Cys^{574}, Cys^{576}, and Cys^{594} and His^{564} and His^{590}).

Hantaviral Gn Zinc Fingers Does Not Bind RNA—Classical ββα-fold zinc fingers are well known nucleic acid-binding motifs [9-12]. However, our attempts to verify the ability of the Andes virus Gn zinc finger domain to bind RNA by electrophoretic mobility shift assay revealed that under the conditions used, the Gn zinc finger domain did not bind RNA obtained from Andes virus-infected Vero E6 cells (**Fig. 3-11**). Although a known RNA-binding protein PACT [13] showed smearing of the RNA bands, which suggested nonspecific PACT-RNA interaction, increasing amounts of the Andes virus Gn zinc finger failed to demonstrate even nonspecific binding of RNA (**Fig. 3-11**).

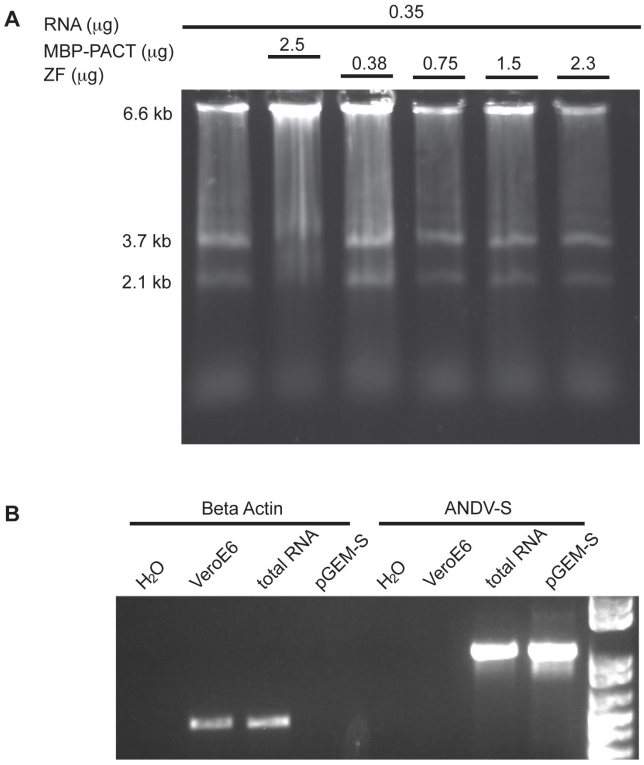

Figure 3-12. The *Hantaviral* zinc finger domain does not bind RNA. **(A)** Electrophoretic mobility shift assay of total RNA (from VeroE6 cells 14 days post-infection with Andes virus), a known RNA-binding protein (PACT) expressed as a fusion protein with the maltose binding protein (MBP), and the Andes virus zinc finger domain (ZF). The smearing of the RNA bands by MBP-PACT showed nonspecific RNA binding, while the RNA bands remained unchanged with the zinc finger domain. **(B)** RT-PCR using primers specific to beta actin and the S-segment of the Andes virus genome (ANDV-S) confirmed the presence of cellular and viral RNA. Controls were uninfected VeroE6 cells, water, and plasmid DNA with the S segment of the Andes virus genome (pGEM-S).

Discussion

The Gn tail of Bunyaviridae viruses is important in viral assembly [14, 15] and host-pathogen interaction [15-19]. Our results showed that a conserved cysteine/histidine-rich region in the *Hantavirus* Gn tail (**Fig. 3-1**) required Zn^{2+} binding to fold properly (**Fig. 3-2**). This region formed an independently folded domain that gave excellent NMR data (**Fig. 3-4**), and NMR structure determination revealed dual CCHC-type zinc fingers where Zn^{2+} ligands were sequential and nonoverlapping (**Fig. 3-8**). The folding of each Gn zinc finger is related to the classical ββα zinc finger fold [20], which is among the most abundant protein motifs in eukaryotic genomes (reviewed in Ref. [21]).

Implication of the Zinc Finger Structure in the Biology of Hantavirus—It has been suggested that the conserved cysteine-histidine region in the Gn tail of *Hantaviruses* forms a RING finger motif [22]. This assumption is based on the following observations: (i) the Gn tail is ubiquitinated and proteasomally degraded as part of the host-pathogen interaction of *Hantaviruses* [17], (ii) RING fingers are structural domains of ubiquitin ligases, which are part of the ubiquitin degradation pathway [23], and (iii) some viruses contain RING finger motifs that are involved in the ubiquitin degradation pathway as part of their host evasion mechanism [22].

Instead of forming a RING finger motif [22], however, our results showed that the conserved cysteine/histidine region in the Gn tail of *Hantaviruses* formed a dual classical type ββα-fold zinc fingers (NOE's supporting back-to-back topology (**Figure 3-13**). Classical zinc fingers are well known DNA- and RNA-binding domains [11, 12, 24]. Recent reports indicate that proteins containing classical zinc fingers are also involved in protein-protein interaction [21, 25-27]. Thus, instead of functioning as a domain of a ubiquitin ligase (as a RING finger), the classical ββα-fold of the Hantaviral zinc fingers suggests nucleic acid binding and/or protein-protein interaction. This is consistent with the observations that the Gn tail is important in binding the ribonucleoprotein during viral assembly [14, 28].

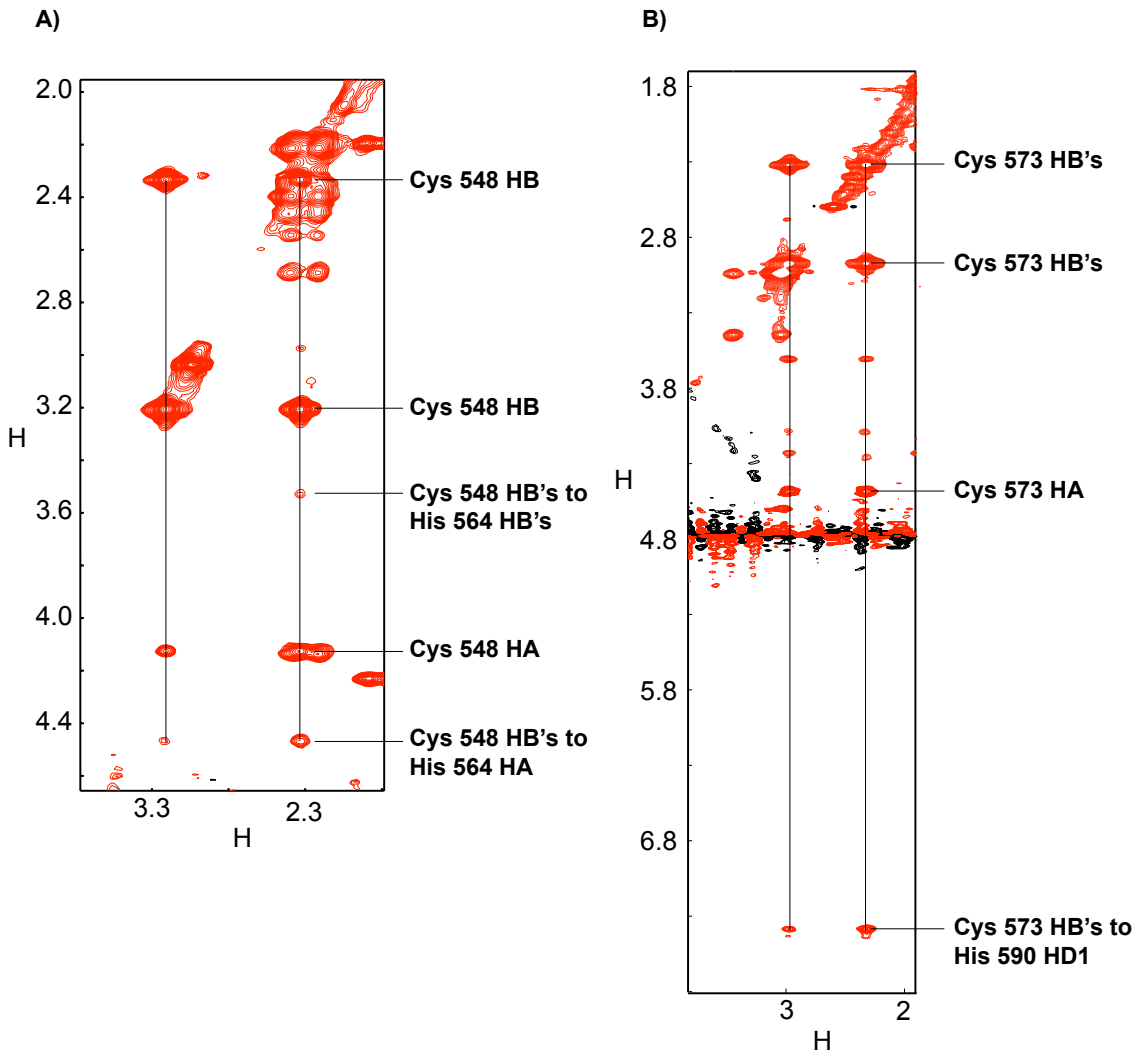

Fig. 3-13. Carbon NOESY spectra for Cys548 (A) and Cys573 (B). Long distance NOE's between Cys548 and His564, along with NOE's between Cys573 and His590 suggest the Andes Gn tail binds zinc in a back-to-back topology rather than an overlapping (RING-type) topology.

Thus, the ββα-fold implies that the *Hantaviral* zinc fingers may interact with the RNA genome or the protein component of the ribonucleoprotein during viral assembly. Our electrophoretic mobility shift assay showed that the *Hantaviral* zinc fingers did not interact with RNA (**Fig. 3-12**). This result is not surprising, since the theoretical pI of the domain

(5.8) is too acidic to be a nucleic acid-binding motif. Further, many of the dual zinc fingers that have been characterized to date (RING, MYND, and LIM) are involved in protein-protein interactions (reviewed in Ref. 25, **Fig. 3-14**). These observations seem to suggest a similar protein binding function for the core *Hantaviral* zinc finger. Interestingly, a more recent study published after the work presented here suggested a protein-protein binding interaction between short peptides of the *Hantaviral* Gn tail and the intact nucleocapsid protein [13]. However, this study goes on to map the interaction to three candidate binding sites: one site located on the N-terminal side of the zinc finger domain and two located at the C-terminal side. No binding sites were reported for the core zinc finger domain, suggesting involvement of the flanking sequences in the interaction instead (Chapter 6 includes the NMR characterizations of one of these reported binding sites and part of a second. Furthermore, Chapter 7 includes electrophoretic mobility shift assay results that suggest these same binding sites may also participate in a protein-RNA interaction). Taken together, these results appear to suggest a mere structural role for the core zinc finger domain, in which the domain serves as a kind of spacer to keep the flanking binding sites an appropriate distance apart. However, studies of the zinc finger domain of the related Crimean Congo Hemorrhagic Fever virus (Chapter 5), in which the structure is similar to *Hantaviruses* but the surface electrostatics are not, indicate the role of the core domain in Bunyaviruses is likely more involved than a simple spacer.

Unique Properties of Hantaviral Zinc Fingers—Although Hantaviral zinc fingers have classical ββα zinc finger fold (**Fig. 3-8**), they differ from classical zinc fingers in two aspects. First, the two Hantaviral zinc fingers fold together as a single domain, which is likely due to a short 4-residue linker between the two zinc fingers. Commonly, classical zinc fingers fold independently of each other, forming a beads-on-a-string configuration. However, multiple classical zinc fingers can interact with each other when bound to DNA [24] or RNA [11, 12]. Another example of a dual classical ββα zinc finger that folds together as one unit is the yeast Zap1 transcription factor [30]. Second, the folding of one Hantaviral zinc finger affects the folding of the other zinc finger. For example, mutations in cysteine and histidine residue

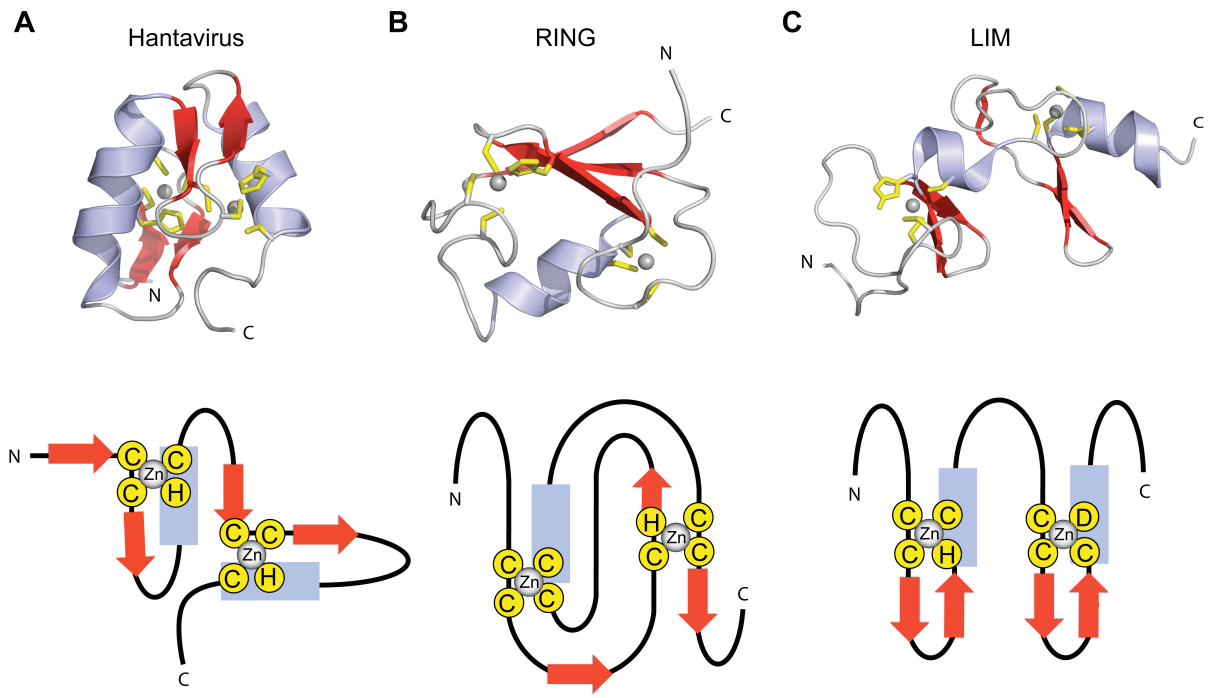

Figure 3-14. The structures (top panel) and zinc binding topology (bottom pane) of the **(A)** Andes virus Gn zinc finger domain is different from that of **(B)** RING domain (PDB 1CHC) and (C) LIM domain (PDB 1M3V). The cartoon depiction of zinc binding topology was adapted from Gamsjaeger *et al.* [25].

that disrupted the first or second zinc finger disrupted the folding of the entire dual zinc finger domain (**Figure 3-10**). Because classical zinc fingers fold independently of each other, disrupting the folding of one zinc finger domain does not affect the folding of the other zinc fingers. For example, in the case of LIM domains, the zinc fingers are only loosely packed together with minimal interdependence between them.

Other Viral Zinc Fingers—Among viruses, the CCHC zinc fingers of the nucleocapsid proteins of retroviruses [31] have been studied extensively because of their critical role in binding and packaging the RNA genomes. Examples of zinc fingers in viral glycoproteins, however, are scarce. Our structure presented here is the first atomic resolution structure of a

zinc finger domain from a viral glycoprotein. The sequence homology of the *Hantaviral* zinc finger region with other Bunyaviridae (**Fig. 3-1**) also suggests that that the Gn tail in *Nairoviruses* and *Orthobunyaviruses* will also form zinc finger motifs. This is confirmed by our description of the Crimean Congo Hemorrhagic Fever zinc finger domain in Chapter 5. Therefore, our results form the structural framework for future studies aimed at elucidating the precise role of the Gn tail in the viral assembly and immune evasion of Bunyaviridae.

References

1. Elliott, R.M., Bouloy, M., Calisher, C. H., Goldbach, R., Moyer, J. T., Nichol, S. T., Pettersson, R., Plyusnin, A., and Schmaljohn, C., *Bunyaviridae*, in *Virus Taxonomy: The classification and Nomenclature of Viruses. The Seventh Report of the International Committee on Taxonomy of Viruses.*, C.M. Fauquet, Mayo, M.A., Maniloff, J. Desselberger, U., and Ball, eds., Editor. 2005, Academic Press: San Diego.
2. Khaiboullina, S.F., S.P. Morzunov, and S.C. St Jeor, *Hantaviruses: molecular biology, evolution and pathogenesis.* Curr Mol Med, 2005. **5**(8): p. 773-90.
3. Overturf, G.D., *World arboviruses: the Bunyaviridae.* Pediatr Infect Dis J, 2009. **28**(11): p. 1014-5.
4. Flick, R. and C.A. Whitehouse, *Crimean-Congo hemorrhagic fever virus.* Curr Mol Med, 2005. **5**(8): p. 753-60.
5. Fisher-Hoch, S.P., *et al.*, *Crimean Congo-haemorrhagic fever treated with oral ribavirin.* Lancet, 1995. **346**(8973): p. 472-5.
6. Hanssen, I.M., M. Lapidot, and B.P. Thomma, *Emerging viral diseases of tomato crops.* Mol Plant Microbe Interact, 2010. **23**(5): p. 539-48.
7. Hollidge, B.S., F. Gonzalez-Scarano, and S.S. Soldan, *Arboviral encephalitides: transmission, emergence, and pathogenesis.* J Neuroimmune Pharmacol, 2010. **5**(3): p. 428-42.
8. Mackow, E.R. and I.N. Gavrilovskaya, *Hantavirus regulation of endothelial cell functions.* Thromb Haemost, 2009. **102**(6): p. 1030-41.
9. Sanchez, A.J., *et al.*, *Crimean-congo hemorrhagic fever virus glycoprotein precursor is cleaved by Furin-like and SKI-1 proteases to generate a novel 38-kilodalton glycoprotein.* J Virol, 2006. **80**(1): p. 514-25.
10. Altamura, L.A., *et al.*, *Identification of a novel C-terminal cleavage of Crimean-Congo hemorrhagic fever virus PreGN that leads to generation of an NSM protein.* J Virol, 2007. **81**(12): p. 6632-42.
11. Shi, X., *et al.*, *Role of the cytoplasmic tail domains of Bunyamwera orthobunyavirus glycoproteins Gn and Gc in virus assembly and morphogenesis.* J Virol, 2007. **81**(18): p. 10151-60.

12. Overby, A.K., R.F. Pettersson, and E.P. Neve, *The glycoprotein cytoplasmic tail of Uukuniemi virus (Bunyaviridae) interacts with ribonucleoproteins and is critical for genome packaging.* J Virol, 2007. **81**(7): p. 3198-205.
13. Hepojoki, J., *et al.*, *Cytoplasmic tails of Hantavirus glycoproteins interact with the nucleocapsid protein.* J Gen Virol, 2010. **91**(Pt 9): p. 2341-50.
14. Battisti, A.J., *et al.*, *Structural studies of Hantaan virus.* J Virol, 2011. **85**(2): p. 835-41.
15. Geimonen, E., *et al.*, *Tyrosine residues direct the ubiquitination and degradation of the NY-1 Hantavirus GN cytoplasmic tail.* J Virol, 2003. **77**(20): p. 10760-868.
16. Geimonen, E., *et al.*, *Hantavirus pulmonary syndrome-associated Hantaviruses contain conserved and functional ITAM siGn aling elements.* J Virol, 2003. **77**(2): p. 1638-43.
17. Sen, N., A. Sen, and E.R. Mackow, *Degrons at the C terminus of the pathogenic but not the nonpathogenic Hantavirus GN tail direct proteasomal degradation.* J Virol, 2007. **81**(8): p. 4323-30.
18. Alff, P.J., *et al.*, *The pathogenic NY-1 Hantavirus GN cytoplasmic tail inhibits RIG-I- and TBK-1-directed interferon responses.* J Virol, 2006. **80**(19): p. 9676-86.
19. Alff, P.J., *et al.*, *The NY-1 Hantavirus Gn cytoplasmic tail coprecipitates TRAF3 and inhibits cellular interferon responses by disrupting TBK1-TRAF3 complex formation.* J Virol, 2008. **82**(18): p. 9115-22.
20. Blakqori, G., *et al.*, *La Crosse bunyavirus nonstructural protein NSs serves to suppress the type I interferon system of mammalian hosts.* J Virol, 2007. **81**(10): p. 4991-9.
21. Billecocq, A., *et al.*, *NSs protein of Rift Valley fever virus blocks interferon production by inhibiting host gene transcription.* J Virol, 2004. **78**(18): p. 9798-806.
22. Hartwig, A., *Zinc finger proteins as potential targets for toxic metal ions: differential effects on structure and function.* Antioxid Redox SiGn al, 2001. **3**(4): p. 625-34.
23. Laity, J.H., B.M. Lee, and P.E. Wright, *Zinc finger proteins: new insights into structural and functional diversity.* Curr Opin Struct Biol, 2001. **11**(1): p. 39-46.
24. Brown, R.S., *Zinc finger proteins: getting a grip on RNA.* Curr Opin Struct Biol, 2005. **15**(1): p. 94-8.
25. Gamsjaeger, R., *et al.*, *Sticky fingers: zinc-fingers as protein-recoGn ition motifs.* Trends Biochem Sci, 2007. **32**(2): p. 63-70.
26. Summers, M.F., *et al.*, *High-resolution structure of an HIV zinc fingerlike domain via a new NMR-based distance geometry approach.* Biochemistry, 1990. **29**(2): p. 329-40.
27. De Guzman, R.N., *et al.*, *Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recoGn ition element.* Science, 1998. **279**(5349): p. 384-8.
28. Deshaies, R.J. and C.A. Joazeiro, *RING domain E3 ubiquitin ligases.* Annu Rev Biochem, 2009. **78**: p. 399-434.
29. Sanchez, A.J., M.J. Vincent, and S.T. Nichol, *Characterization of the glycoproteins of Crimean-Congo hemorrhagic fever virus.* J Virol, 2002. **76**(14): p. 7263-75.
30. Elliott, R.M., *Molecular biology of the Bunyaviridae.* J Gen Virol, 1990. **71 (Pt 3)**: p. 501-22.
31. Flint, S.J., *et al.*, *Principles of Virology: Molecular Biology, Pathogenesis, and Control*. 2000, Washington, D.C.: ASP Press.

32. Wishart, D.S. and A.M. Nip, *Protein chemical shift analysis: a practical guide.* Biochem. Cell Biol., 1998. **76**(2-3): p. 153-163.
33. Pelton, J.G., *et al.*, *Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a siGn al-transducing protein from Escherichia coli, using two-dimensional heteronuclear NMR techniques.* Protein Sci., 1993. **2**(4): p. 543-558.
34. Holm, L. and C. Sander, *Mapping the protein universe.* Science, 1996. **273**(5275): p. 595-603.
35. Zhang, Y. and J. Skolnick, *TM-aliGn : a protein structure aliGn ment algorithm based on the TM-score.* Nucleic Acids Res., 2005. **33**(7): p. 2302-9.
36. Wuttke, D.S., *et al.*, *Solution structure of the first three zinc fingers of TFIIIA bound to the coGn ate DNA sequence: determinants of affinity and sequence specificity.* J Mol Biol, 1997. **273**(1): p. 183-206.
37. Nolte, R.T., *et al.*, *Differing roles for zinc fingers in DNA recoGn ition: structure of a six-finger transcription factor IIIA complex.* Proc Natl Acad Sci U S A, 1998. **95**(6): p. 2938-43.
38. Lu, D., M.A. Searles, and A. Klug, *Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recoGn ition.* Nature, 2003. **426**(6962): p. 96-100.
39. Lee, B.M., *et al.*, *Induced fit and "lock and key" recoGn ition of 5S RNA by zinc fingers of transcription factor IIIA.* J Mol Biol, 2006. **357**(1): p. 275-91.
40. Patel, R.C. and G.C. Sen, *PACT, a protein activator of the interferon-induced protein kinase, PKR.* EMBO J., 1998. **17**(15): p. 4379-90.
41. Spiropoulou, C.F., *et al.*, *Andes and Prospect Hill Hantaviruses differ in early induction of interferon although both can downregulate interferon siGn aling.* J Virol, 2007. **81**(6): p. 2769-76.
42. Miller, J., A.D. McLachlan, and A. Klug, *Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes.* EMBO J., 1985. **4**(6): p. 1609-14.
43. Foster, M.P., *et al.*, *Domain packing and dynamics in the DNA complex of the N-terminal zinc fingers of TFIIIA.* Nat. Struct. Biol., 1997. **4**(8): p. 605-8.
44. Fox, A.H., *et al.*, *Transcriptional cofactors of the FOG family interact with GATA proteins by means of multiple zinc fingers.* EMBO J., 1999. **18**(10): p. 2812-22.
45. Westman, B.J., *et al.*, *Structural studies on a protein-binding zinc-finger domain of Eos reveal both similarities and differences to classical zinc fingers.* Biochemistry, 2004. **43**(42): p. 13318-27.
46. Kaukinen, P., A. Vaheri, and A. Plyusnin, *Hantavirus nucleocapsid protein: a multifunctional molecule with both housekeeping and ambassadorial duties.* Arch. Virol., 2005. **150**(9): p. 1693-1713.
47. Wang, Z., *et al.*, *Solution structure of a Zap1 zinc-responsive domain provides insights into metalloregulatory transcriptional repression in Saccharomyces cerevisiae.* J. Mol. Biol., 2006. **357**(4): p. 1167-83.
48. Guntert, P., *Automated NMR structure calculation with CYANA.* Methods Mol. Biol., 2004. **278**: p. 353-378.
49. Case, D.A., *et al.*, *AMBER7*. 2002, University of California, San Francisco.
50. Hepojoki, J., *et al.*, *Interactions and oligomerization of hantavirus glycoproteins.* J Virol, 2010. **84**(1): p. 227-42.

Chapter 4: NMR Characterization of the Prospect Hill Gn Zinc Finger Domain

Introduction

An intriguing aspect of *Hantavirus* infection is the wide range of hemorrhagic fevers resulting from transmission of the virus. European and Asian strains of the virus, for example, cause hemorrhagic fever with renal syndrome (HFRS), which is fatal in 1-15% of patients [1]. The American strains, by contrast, cause the more severe *Hantavirus* Cardio Pulmonary Syndrome (HCPS), which is more aggressive and carries a 40% mortality rate [1]. Perhaps even more interesting are the two well known asymptomatic strains; Tula virus (present in European common voles) and the Prospect Hill virus (present in the American meadow vole) [2].

Studies into the virulence of *Hantaviruses* reveal several notable differences in cells infected by pathogenic and non-pathogenic *Hantaviruses*. For example, while the non-pathogenic Prospect Hill virus (PHV) still manages to infect endothelial cells, the virus apparently does not successfully replicate in humans [2]. This suggests PHV does not manage to evade the host immunogenic response. This notion is supported by evidence that early infection by PHV causes the induction of up to 24 early interferon genes, as opposed to only three by the pathogenic New York and Hantaan viruses [3, 4].

Findings from numerous studies have now mapped *Hantavirus* virulence to the Gn cytoplasmic tail [5-7]. Recently, the Gn tail of PHV was found not to be degraded, as is the case for typical pathogenic *Hantaviruses* [8]. However, the mutations of four residues at the carboxyl

terminus of the Gn tail effectively targeted the PHV tail for proteasomal degradation [8]. The PHV tail also fails to co-precipitate (tumor necrosis factor) receptor-associated factor 3 (otherwise known as TRAF3), as is the case for the New York *Hantavirus* virus [6]. TRAF3 is a key component of the host cell's interferon response to viral infection.

The observation of a virulence contribution by the Gn tail raises the possibility of potentially important differences between the predicted dual CCHC-type zinc finger domain of PHV and the structure determined for the pathogenic Andes virus (Chapter 3). The two Gn tails overall are highly conserved between both viruses (75% identity, 84% similarity for the entire tail; 70% identity, 77% similarity for the zinc finger domain alone) (**Fig 4-1**).

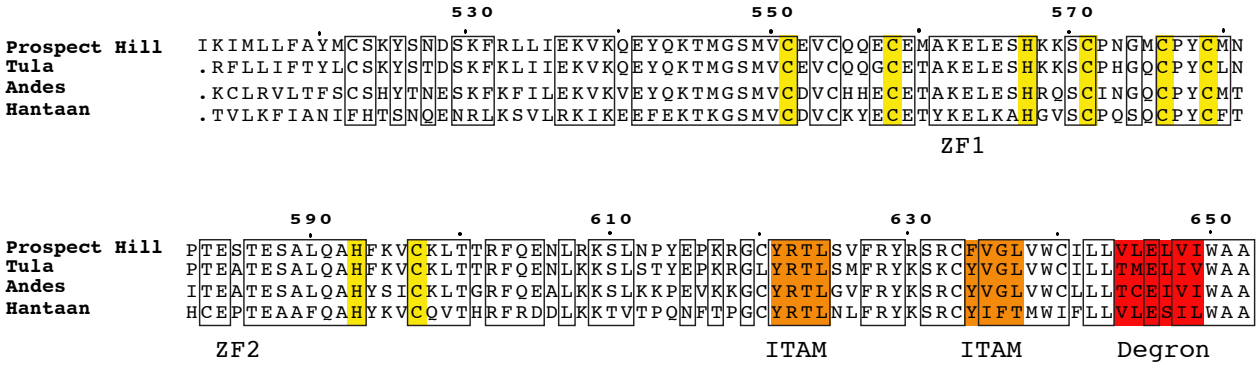

Fig. 4-1. Sequence alignment comparing the entire Gn tails of two non-pathogenic *Hantaviruses* (PHV, Tula) with two pathogenic *Hantaviruses* (Andes, HPS; Hantaan, HFRS). Immunoreceptor tyrosine-based activation motifs (ITAMS) are highlighted in orange and the degradation signal (Degron) is highlighted in red. Identical sequences between the Andes and Prospect Hill viruses are indicated by boxes.

In this study, 2D and 3D NMR experiments were used to assign the protein backbone of a construct corresponding to the Gn zinc finger domain of PHV. The NMR assignments suggest that, similar to the Andes virus, the dual CCHC motif forms an independently folded zinc binding domain. The Cα secondary chemical shifts of the PHV zinc finger domain are

remarkably similar to those of the Andes structure, suggesting there is no appreciable difference in the two structures. These findings further support reports that the virulence determinants are located further toward the C-terminal end of the cytoplasmic tails.

Results

Protein Expression - The PHV Gn tail ZF construct was designed similarly to the Andes construct [9] in Chapter 3, yielding a construct of residues 548-602. The protein expressed well in *E. coli* as an N-terminal GB1 fusion recombinant protein and was purified under native conditions via ion exchange chromatography. Since attempts to express the full-length Gn tail in other *Hantaviruses* have been unsuccessful, expression of the full-length Gn tail of PHV was not attempted. Upon TEV mediated cleavage of the GB1 tag, PHV Gn, residues 548-602, gave a well dispersed, well resolved ^{1}H-^{15}N HSQC spectrum of the PHV Gn tail, residues 548-602 (**Fix. 4-2**).

PHV Gn Tail Also Relies on Zn^{2+} for Proper Folding - As expected, treatment of the ^{1}H-^{15}N labeled NMR sample with excess EDTA resulted in a near complete collapse of the HSQC spectrum (**Fig. 4-3A**), suggesting a loss of tertiary structure. Similarly, the protein CD trace is also affected by the presence of EDTA (**Fig. 4-3B**). The addition of Zn^{2+} back into the sample is only partially able to recover the original helical trace.

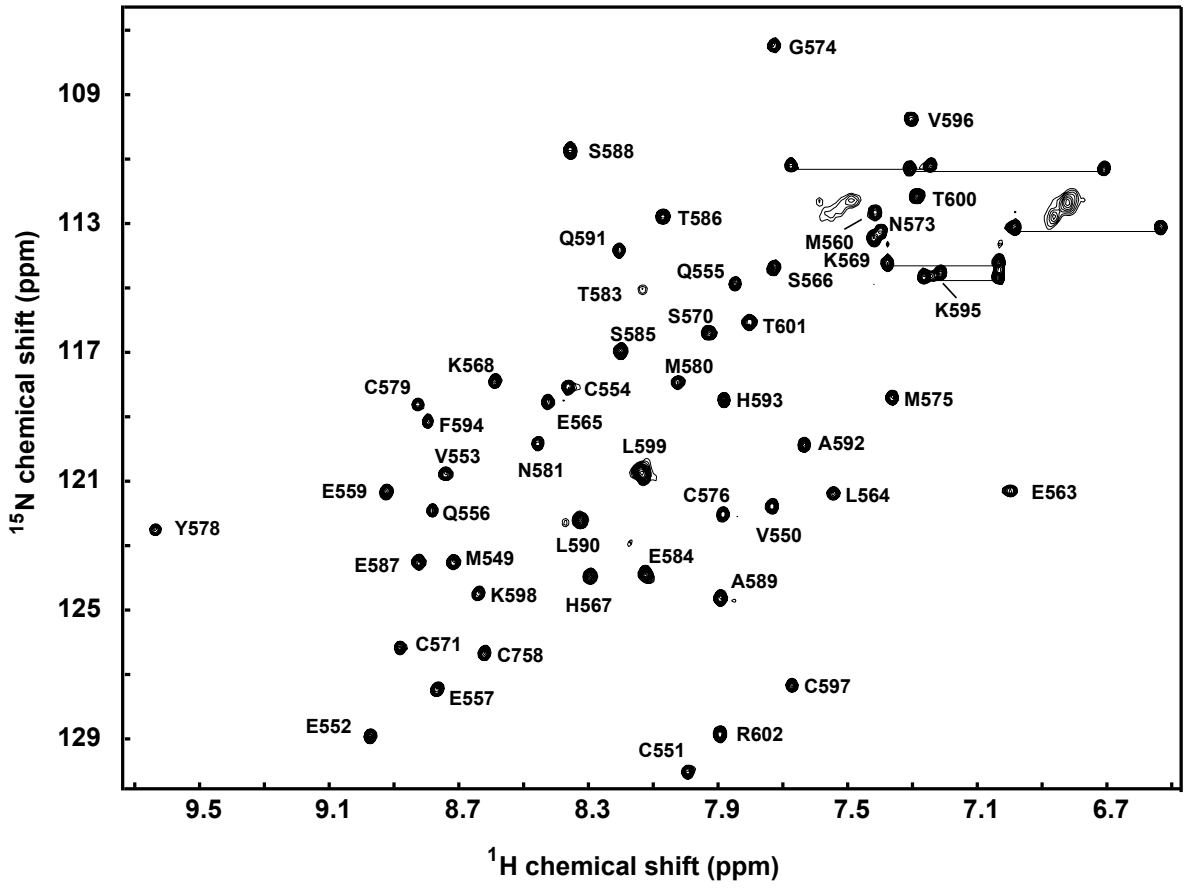

Figure 4-2. ^{1}H-^{15}N HSQC spectrum for PHV Gn 548-602 shows a well dispersed and well resolved spectrum.

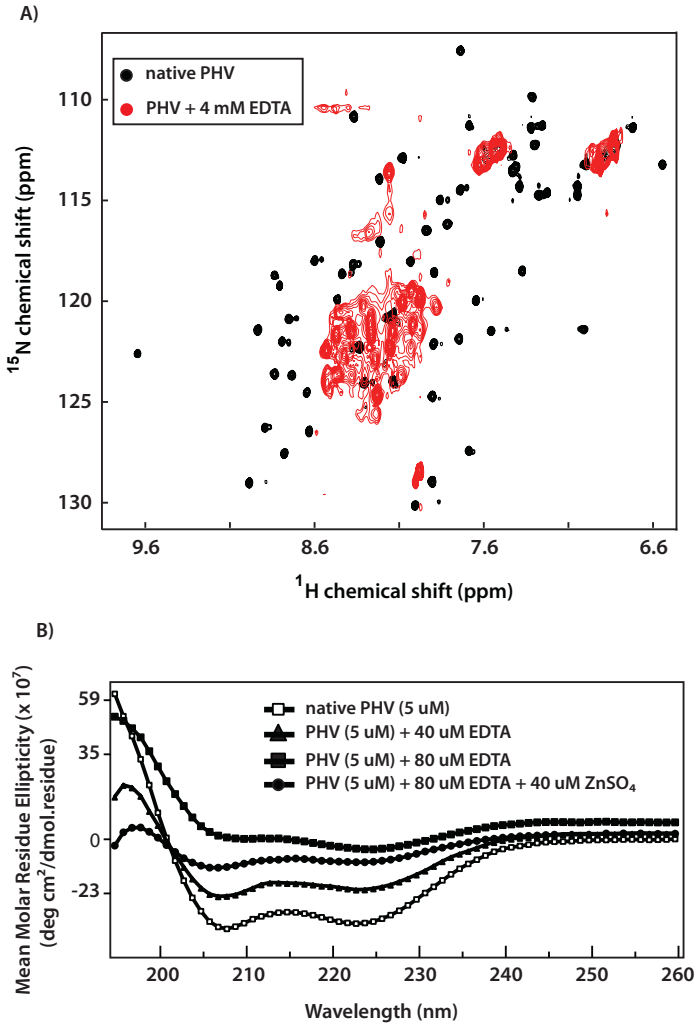

Figure 4-3. (A) EDTA titration of the 2D spectrum of PHV Gn 548-602 causes a dramatic narrowing of the spectrum, suggesting a loss of tertiary structure and (B) the effect on the protein CD trace.

Backbone NMR Assignments - PHV Gn 548-602 showed a well dispersed two dimensional 1H-15N HSQC (**Fig. 4-2**). Nearly complete backbone assignments were obtained from the three-dimensional HNCA, CBCA(CO)NH, and HNCACB experiments (**Fig. 4-2**). Analysis of the Cα secondary chemical shifts (**Fig. 4-4**) indicated the presence of two short α-helices and two random coil regions flanking the central domain [10]. This pattern closely resembles that of the Andes Gn zinc finger domain. Given the high degree of chemical

similarity between both structures (77% similarity) and the remarkable similarity of the Cα secondary chemical shift profile, full side chain assignments and structure determination of the construct was not continued.

Modeling the PHV Gn Zinc Fingers – The similarity between the secondary chemical shift profiles of the PHV and the Andes zinc fingers suggests the two domains have similar folds. Therefore, the Andes structure (PDB ID 2K9H) was used as a template for the I-TASSER [11] model of the PHV structure. The modeled structure is shown in Figure 4-5A (left panel). It closely resembles that of the Andes structure, with a backbone rmsd difference of 0.68 between the two (**Fig. 4-5A, right panel**). Notably, the surface electrostatics of the PHV model closely resemble those of the Andes structure (**Fig. 4-5B**) in that neither exhibit extensive clustering of charges.

Discussion

Recent studies have begun to map virulence in *Hantaviruses* to the Gn cytoplasmic tail [5-7, 11]. Given these findings, preliminary structural studies were conducted on the predicted Gn tail zinc finger domain of the non-pathogenic Prospect Hill virus for the purpose of comparing the structure to that of the pathogenic Andes virus. The 2D ^{1}N-^{15}N HSQC of the PHV Gn tail consisting of residues 548-602 showed a well dispersed spectrum (**Fig 4-2**), suggesting a monomeric and independently folded protein.

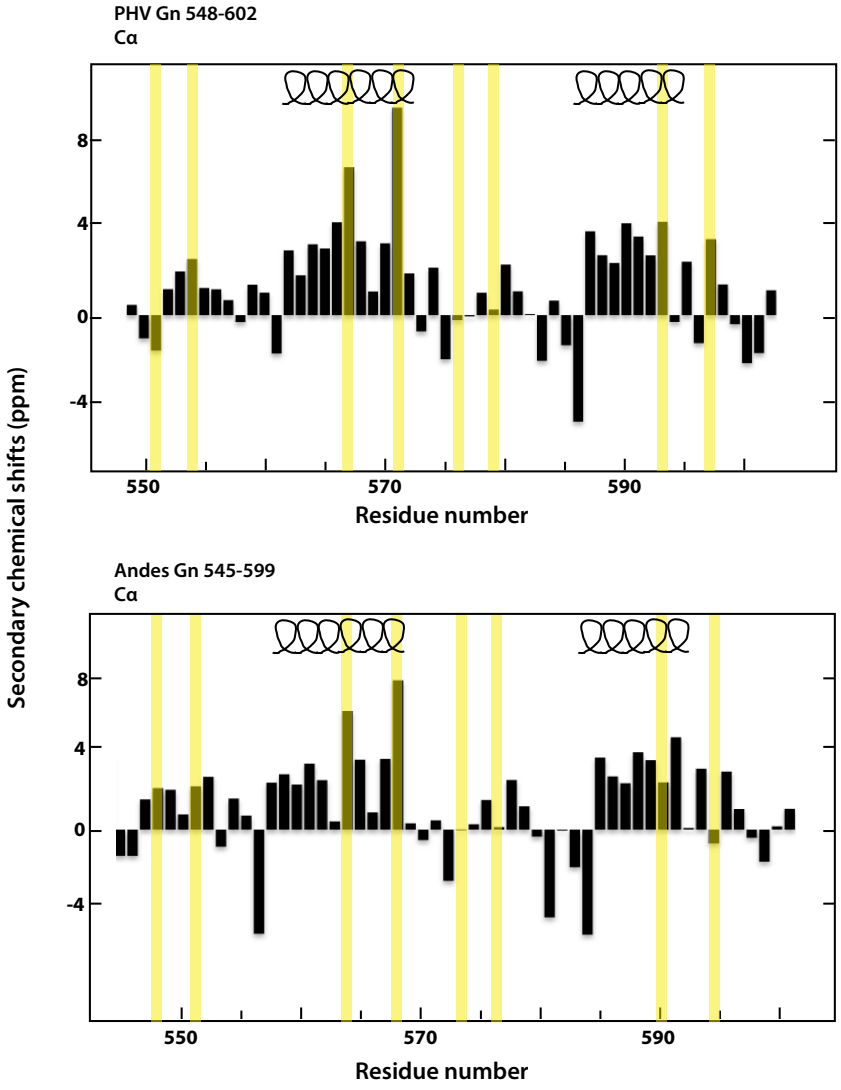

Figure 4-4. Secondary Cα chemical shift profile for both PHV and Andes Gn zinc finger structures. The zinc coordinating residues are highlighted in yellow.

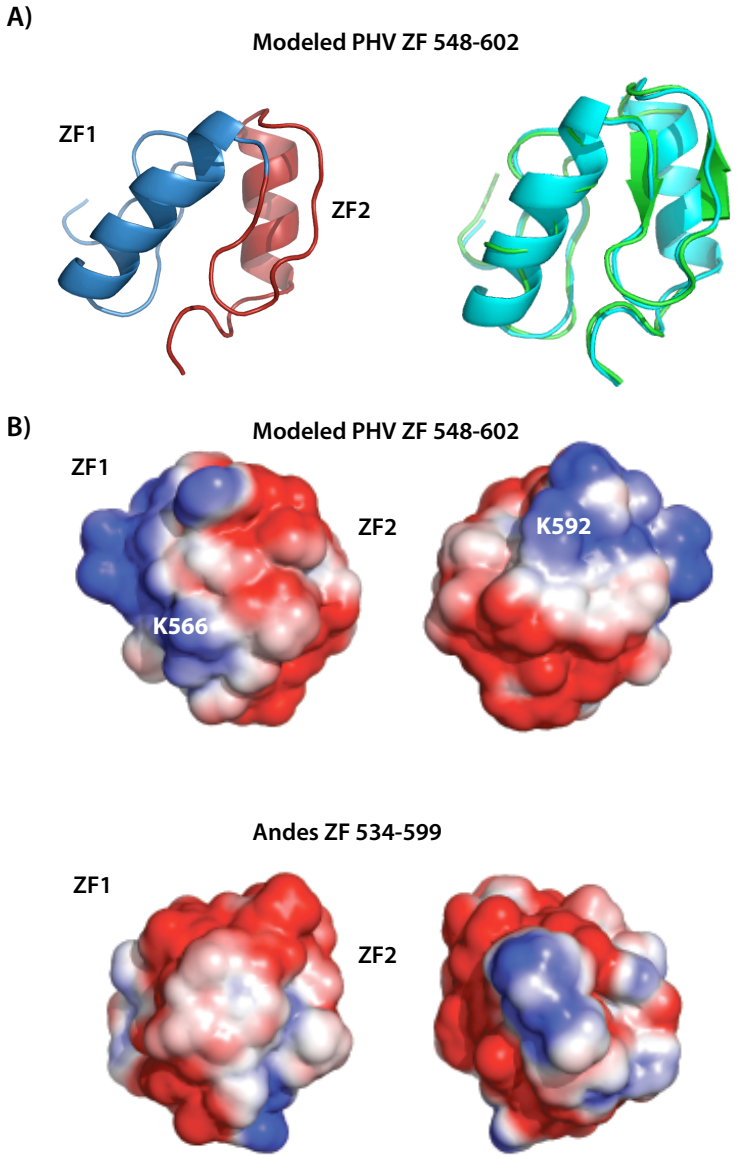

Figure 4-5. Model of PHV Gn 548-602 (A, left) and superimposed on the Andes template structure (A, right). The PHV model (cyan) matched the Andes structure (green) with a backbone rmsd of 0.68 for all Cα's. The electrostatic surface representations for each structure are shown in B. The PHV model was constructed using I-TASSER [12] using the Andes structure as a template. Surface electrostatics were calculated using APBS [13] and visualized using Pymol [14]. Structural alignment was carried out using I-TASSER TM-align [15].

We've demonstrated that both the NMR and CD spectra of PHV zinc finger are sensitive to EDTA chelation (**Fig. 4-3**), as was the case for the corresponding zinc fingers of the Andes and Sin Nombre viruses. This isn't surprising given the conserved nature of the dual CCHC motif in all known *Hantaviruses*.

The high quality of the 3D NMR data facilitated assignment of the protein backbone, yielding the Cα chemical shift profile. The profile of the PHV zinc finger domain is compared to that of the Andes virus in **Figure 4-4.** Given the high degree of sequence conservation, as well as conservation in the spacing of the dual CCHC motif (**Fig. 4-1**), it is not surprising to find that the profiles are remarkably similar. The only notable differences occur in the first and final coordinating cysteines (Cys^{551} and Cys^{597}, respectively). These differences may be due to the presence of two non-conserved residues, Glu^{552} and Lys^{598}, immediately following each cysteine.

Due to the apparent similarity of the structures, we did not solve the three dimensional structure of the PHV zinc finger domain. However, a model using I-TASSER [12] and the Andes structure as a template was constructed. The modeled PHV zinc finger domain contains two similar short α-helices in a relative orientation as seen in the Andes structure (**Fig. 4-5**). The surface electrostatics of the PHV structure also bear a resemblance to that of the Andes structure, with dispersed charges covering the zinc fingers. Only two non-conserved basic charges, K^{566} and K^{592}, are present which are not on the Andes structure. Despite these, the PHV structure is also acidic, with a theoretical isoelectric point of approximately 6.0, thus suggesting that differences in electrostatics of the core zinc finger domain are not likely to play a role in pathogenicity.

One interesting feature of the PHV Gn amino acid sequence is the presence of a proline (Pro^{572}) in the linker. The proline is somewhat conserved; it is present in Tula and Hantaan viruses but absent in the Andes virus, which has an isoleucine in the same position. Interestingly, a residue with a rigid peptide bond in this position has little effect on the Cα secondary chemical shift of Cys^{571}. The secondary chemical shift profile for this part of the ZF1 array is virtually identical to that of the ZF1 array in the Andes structure (**Fig. 4-4**), suggesting the tight turn between the β hairpin and helix of ZF1 is maintained despite the difference in residues in this region. This observation was an early indicator that the dual zinc finger fold described in the previous chapter is preserved despite large differences in chemical composition, as seen in Chapter 5.

Overall, the PHV Gn zinc finger domain appears to adopt the same dual zinc finger fold to that of the pathogenic Andes virus. There were no obvious differences in either the secondary chemical shift profile or the predicted surface electrostatics. Based on this data we conclude that, given the widespread nature of the dual zinc finger fold in *Hantaviruses*, its role is likely general in nature and apparently doesn't play a role in the determination of virulence. These conclusions are consistent with studies mapping virulence to the C-terminus of the Gn tails [4, 8].

References

1. Khaiboullina, S.F., S.P. Morzunov, and S.C. St Jeor, *Hantaviruses: molecular biology, evolution and pathogenesis.* Curr Mol Med, 2005. **5**(8): p. 773-90.
2. Mackow, E.R. and I.N. Gavrilovskaya, *Hantavirus regulation of endothelial cell functions.* Thromb Haemost, 2009. **102**(6): p. 1030-41.

3. Geimonen, E., et al., *Pathogenic and nonpathogenic hantaviruses differentially regulate endothelial cell responses.* Proc Natl Acad Sci U S A, 2002. **99**(21): p. 13837-42.
4. Spiropoulou, C.F., et al., *Andes and Prospect Hill hantaviruses differ in early induction of interferon although both can downregulate interferon signaling.* J Virol, 2007. **81**(6): p. 2769-76.
5. Alff, P.J., et al., *The pathogenic NY-1 hantavirus G1 cytoplasmic tail inhibits RIG-I- and TBK-1-directed interferon responses.* J Virol, 2006. **80**(19): p. 9676-86.
6. Alff, P.J., et al., *The NY-1 hantavirus Gn cytoplasmic tail coprecipitates TRAF3 and inhibits cellular interferon responses by disrupting TBK1-TRAF3 complex formation.* J Virol, 2008. **82**(18): p. 9115-22.
7. Geimonen, E., et al., *Hantavirus pulmonary syndrome-associated hantaviruses contain conserved and functional ITAM signaling elements.* J Virol, 2003. **77**(2): p. 1638-43.
8. Sen, N., A. Sen, and E.R. Mackow, *Degrons at the C terminus of the pathogenic but not the nonpathogenic hantavirus G1 tail direct proteasomal degradation.* J Virol, 2007. **81**(8): p. 4323-30.
9. Estrada, D.F., et al., *The Hantavirus Glycoprotein G1 Tail Contains Dual CCHC-type Classical Zinc Fingers.* J Biol Chem, 2009. **284**(13): p. 8654-60.
10. Wishart, D.S. and A.M. Nip, *Protein chemical shift analysis: a practical guide.* Biochem. Cell Biol., 1998. **76**(2-3): p. 153-163.
11. Geimonen, E., et al., *Tyrosine residues direct the ubiquitination and degradation of the NY-1 hantavirus G1 cytoplasmic tail.* J Virol, 2003. **77**(20): p. 10760-868.
12. Zhang, Y., *I-TASSER server for protein 3D structure prediction.* BMC Bioinformatics, 2008. **9**: p. 40.
13. Baker, N.A., et al., *Electrostatics of nanosystems: application to microtubules and the ribosome.* Proc. Natl. Acad. Sci. U.S.A., 2001. **98**(18): p. 10037-10041.
14. DeLano, W.L., *The PyMOL Molecular Graphics System* 2002, San Carlos, California, U.S.A.: DeLano Scientific.
15. Zhang, Y. and J. Skolnick, *TM-align: a protein structure alignment algorithm based on the TM-score.* Nucleic Acids Res., 2005. **33**(7): p. 2302-9.

Chapter 5: NMR Structure of the Crimean-Congo Hemorrhagic Fever Virus Gn Tail

Introduction

Recent outbreaks of the Crimean Congo Hemorrhagic Fever (CCHF) virus along with the reported ability of the virus to transfer between humans have raised concerns of a widespread pandemic (1). The virus is normally transmitted to humans by tick bite or by direct handling of infected animal meat or blood (1,2). Infection causes a hemorrhagic fever and myalgia resulting in mortality rates approaching 30% (1-3). The virus contains an antisense RNA genome divided into three segments, and named according to lengths as the S, M, and L (for Small, Medium, and Large) segments (4). The viral proteins are the nucleocapsid protein, two membrane glycoproteins Gn and Gc (also referred to as G1 and G2 in other *Bunyaviridae*) (5,6), a nonstructural protein (NSm) (7), and an RNA polymerase (4). In the mature virion, the Gn glycoprotein contains a 176 residue ectodomain followed by a 24 residue transmembrane region and terminates in a long cytoplasmic tail consisting of approximately 100 residues (5,7).

Recent results from other related *Bunyaviridae* viruses suggest a role of the Gn tail in viral assembly is to bind the ribonucleocapsid complex. For example, alanine mutagenesis of the cytoplasmic tails of Uukuniemi virus (genus *Phlebovirus*) (8) and Bunyamwera virus (genus *Orthobunyavirus*) (9) affect the ability of Virus Like Particles (VLPs) to effectively incorporate ribonucleoproteins, thus suggesting a role for Gn tails in genome packaging. More recently, the Gn tail of Puumala virus (genus *Hantavirus*) was shown to co-immunoprecipitate with the Puumala nucleocapsid protein (10). These results suggest that the CCHF virus Gn tail plays an equally important role in viral assembly of genus *Nairovirus*.

The sequence of the CCHF virus cytoplasmic tail is somewhat variable in *Nairoviruses* (~24% identity) and even more so when compared to other Bunyaviruses (12% identity with *Hantavirus* Gn tails). However, one characteristic feature present in four of the five genera of *Bunyaviridae* is a conserved dual **C**-X-**C**-X-**H**-X-**C** motifs of cysteine and histine residues with X representing any amino acid (**Fig. 1-3**). Others have suggested that the high cysteine content of the CCHF virus Gn tail could be due to extensive disulfide bonding (5). Recently, we reported that the cysteines in the Andes *hantavirus* Gn tail fold into a novel arrangement of back-to-back classical ββα zinc fingers (11). Despite low sequence identity between the Gn tail of *Nairoviruses* and *Hantaviruses*, the spacing of the dual CCHC motif in the CCHF virus most closely resembles that of *Hantaviruses*, suggesting the presence of a similar dual zinc finger structure. To test this hypothesis, the NMR structure of the CCHF virus Gn cytoplasmic tail was determined from residues 729-805. This study reports the first known atomic structure of any protein component of the CCHF virus and demonstrates that the high cysteine content of the Gn cytoplasmic tail is partly due to the presence of dual, back-to-back ββα-type zinc fingers similar to those found in *Hantaviruses*. Unlike Hantaviral zinc fingers, however, the electrostatic surface of the CCHF virus zinc finger reveals a clear distribution of conserved electrostatic charges. Moreover, we demonstrate using electrophoretic mobility shift assays (EMSA) that these conserved electrostatics may play a role in forming a surface for binding viral RNA. Together, these data provide insight into the role of the Gn tail in *Nairovirus* assembly.

Results

Protein Expression and Purification – Our previous work with *Hantavirus* glycoprotein cytoplasmic tails indicates expression of the tail is toxic to *E. coli* (11). Therefore, all constructs of the CCHF virus Gn cytoplasmic were expressed as GB1-fusion proteins (**Fig. 5-1**). The GB1

tag contained His_6 for nickel affinity purification and a TEV protease cleavage site to recover the native Gn zinc finger domain. The fusion protein was expressed in soluble form in *E. coli*, purified under native conditions, and digested with TEV protease to obtain the Gn zinc finger domain. Longer constructs comprising the entire predicted cytoplasmic tail ($Gn^{719\text{-}819}$) expressed as insoluble inclusion bodies. $Gn^{729\text{-}819}$, which was missing the first ten residues following the transmembrane region, expressed as soluble protein but with low yield. $Gn^{729\text{-}805}$ represented the longest construct containing the conserved **C**-X_2-**C**-$X_{11\text{-}12}$-**H**-X_3-**C** (where X is any amino acid) that also expressed in high enough yield to give high resolution NMR data.

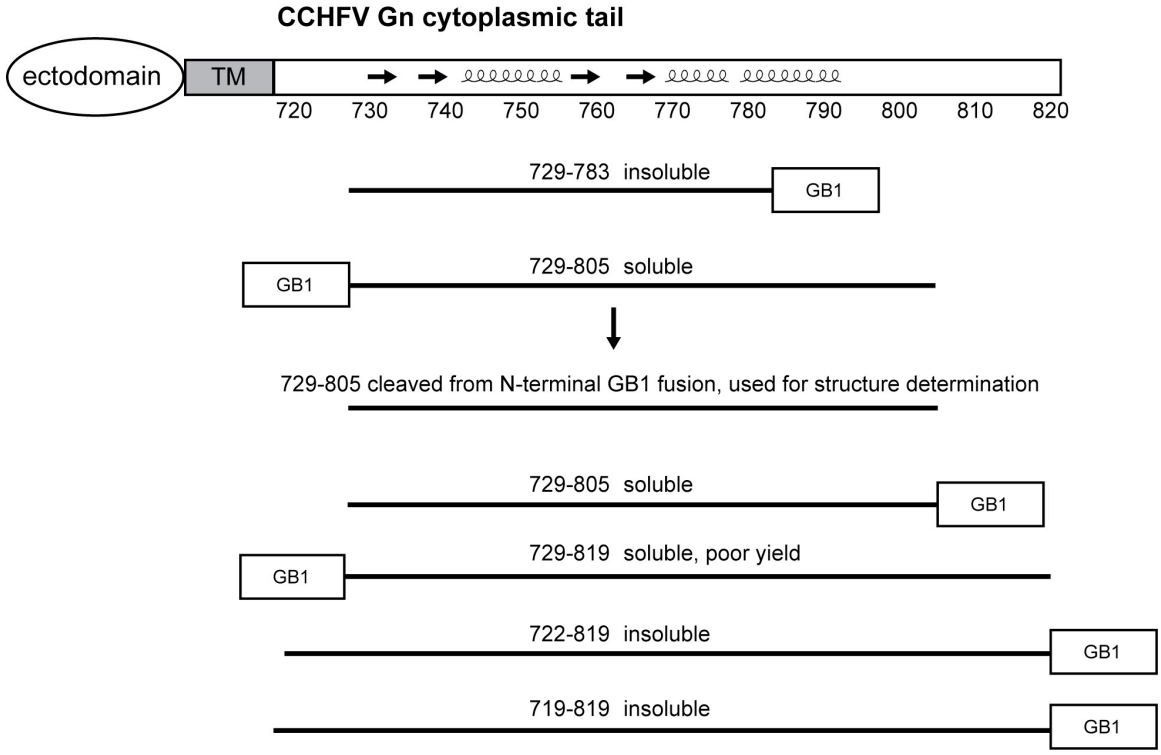

Figure 5-1. The Crimean Congo Hemorrhagic Fever virus Gn tail was expressed in various lengths as GB1 fusion proteins. Constructs representing the entire predicted tail (719-819) expressed as insoluble protein. The smallest construct was also insoluble, most likely due to the truncation of helix α3. The largest soluble construct from 729-819 expressed in low yields. Therefore, the most useful construct for NMR structure determination was 729-805.

Zn^{2+} is Required for Proper Folding – To examine the reliance of Zn^{2+}-coordination on the proper folding of the CCHF virus Gn tail, the 2D ^{15}N-HSQC of the $Gn^{729-805}$ was recorded in the presence of 4 mM EDTA (**Fig. 5-2A**). The spectrum in the presence of EDTA is collapsed between ppm values of 6.5 and 8.6, whereas the folded spectrum in the absence of EDTA is well dispersed between 6.5 and 9.3 ppm. Narrowing of the spectrum suggests a loss of tertiary structure upon removal of Zn^{2+}, indicating the requirement for Zn^{2+} binding in folding of the domain. A similar titration using circular dichroism (CD) spectroscopy demonstrates that the presence of EDTA causes a downward spectral shift, indicating a transition toward an unfolded protein (**Fig. 5-2B**). Here it was also demonstrate that the addition of Zn^{2+} ion back into the sample recovers the original native spectrum. Therefore, Zn^{2+} is required for proper folding of the CCHF virus Gn tail.

NMR Structure Determination – CCHF virus $Gn^{729-805}$ showed a well dispersed 2D ^{1}H-^{15}N HSQC (**Fig. 5-2**). Complete backbone assignments were obtained from 3D HNCA, CBCA(CO)NH, HNCACB, and ^{15}N-edited NOESY-HSQC (**Fig. 5-3,4**). The C^{α}, H^{α}, and C^{β} secondary chemical shifts (**Fig. 5-5A**) showed the presence of three short alpha helices with an intervening random coil region between the second and third helix. Two more regions in random coil orientations flanked the central sequence of the domain, as indicated by the heteronuclear ^{15}N-^{1}H NOE (**Fig. 5-5B).** Side chain assignments were completed using 2D ^{1}H-^{13}C HMQC, 3D HBHA(CO)NH, and 3D ^{13}C edited HMQC-NOESY. There were six invariant cysteine and two histidine residues (His^{752} and His^{776}) in $Gn^{729-805}$ (**Fig. 1-3**), all of which were involved in Zn^{2+} coordination. Long distance NOE's confirmed that His^{752} and His^{776} were involved in Zn^{2+} coordination.

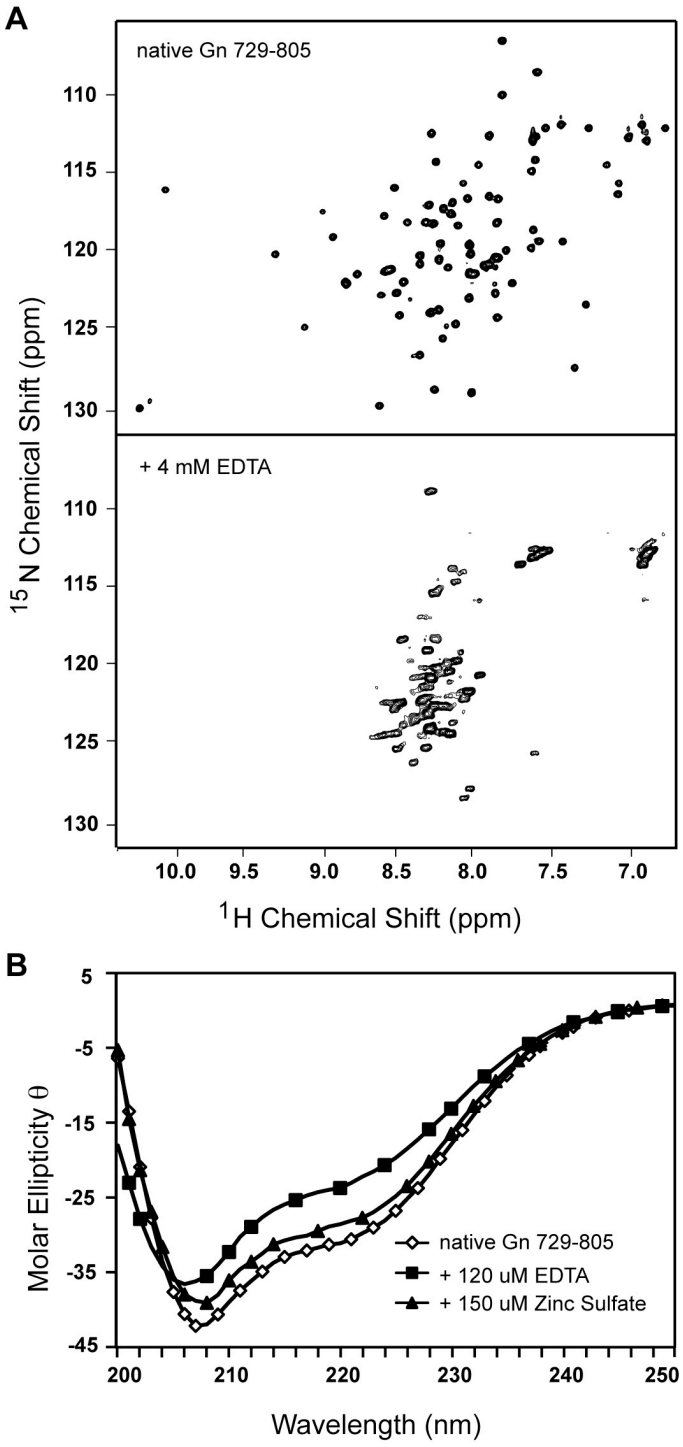

Figure 5-2. CCHF virus Gn 729-805 relies on Zn^{2+}-binding for proper folding. **(A)** Addition of 4 mM EDTA to a sample of ^{15}N-labeled Gn 729-805 effectively narrows the HSQC spectrum into a characteristic of an unfolded protein . **(B)** Likewise, addition of a metal chelator causes a downward shift at 208 nm in the CD spectra toward random coil (Y axis: molar ellipticity θ per residue, deg·cm^{2}dmol^{-1}residue-1 ×104). Titration of zinc sulfate back into the sample recovers the original CD trace.

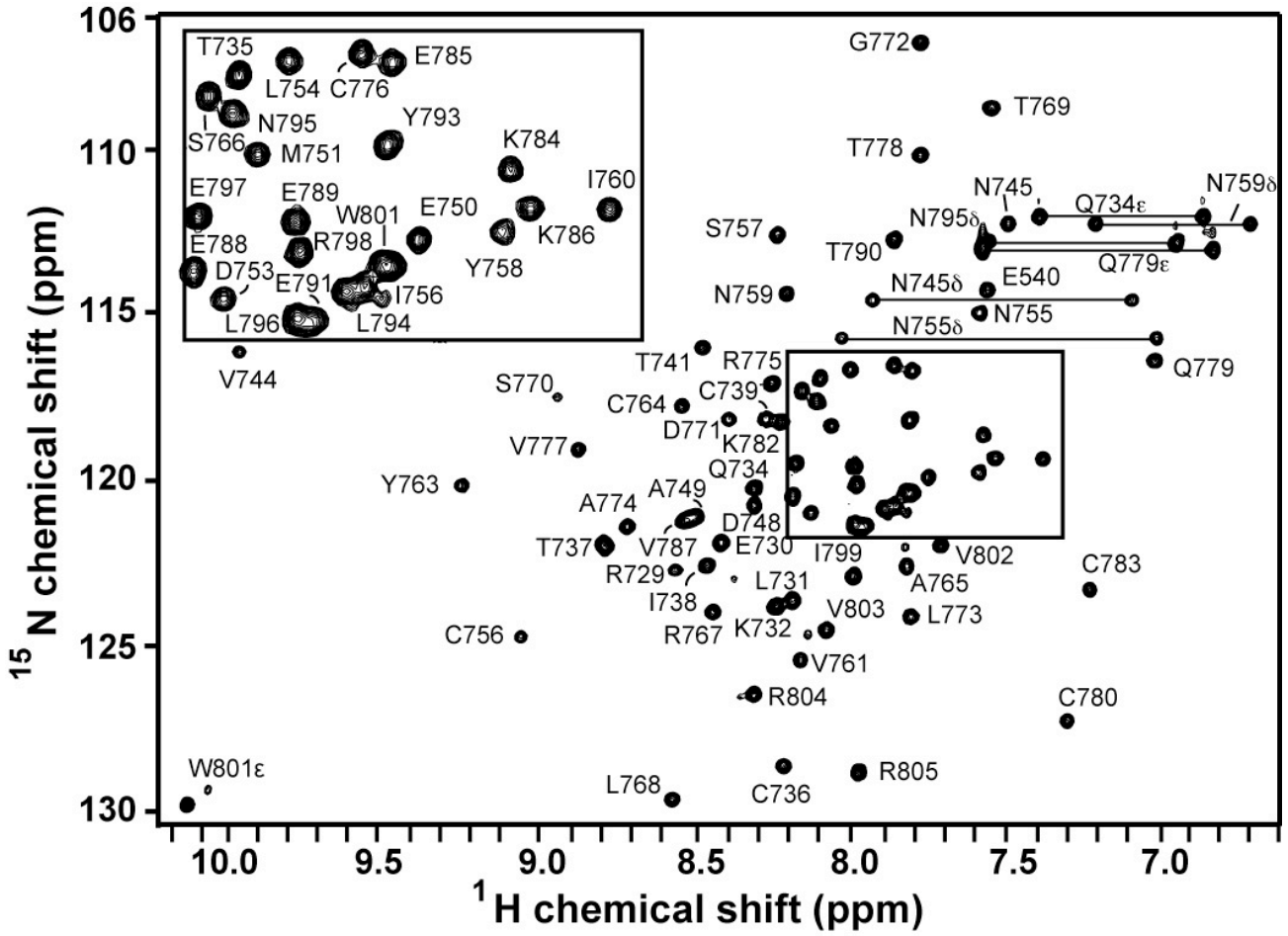

Figure 5-3. The CCHF virus Gn tail (residues 729-805) yielded a well dispersed 2D ^{1}H-^{15}N HSQC spectrum, shown with the complete assignments for backbone and side chain amides. The smaller peak in the tryptophan (W801) side-chain suggested a minor conformation of the tryptophan ring possibly due to ring flip-flop.

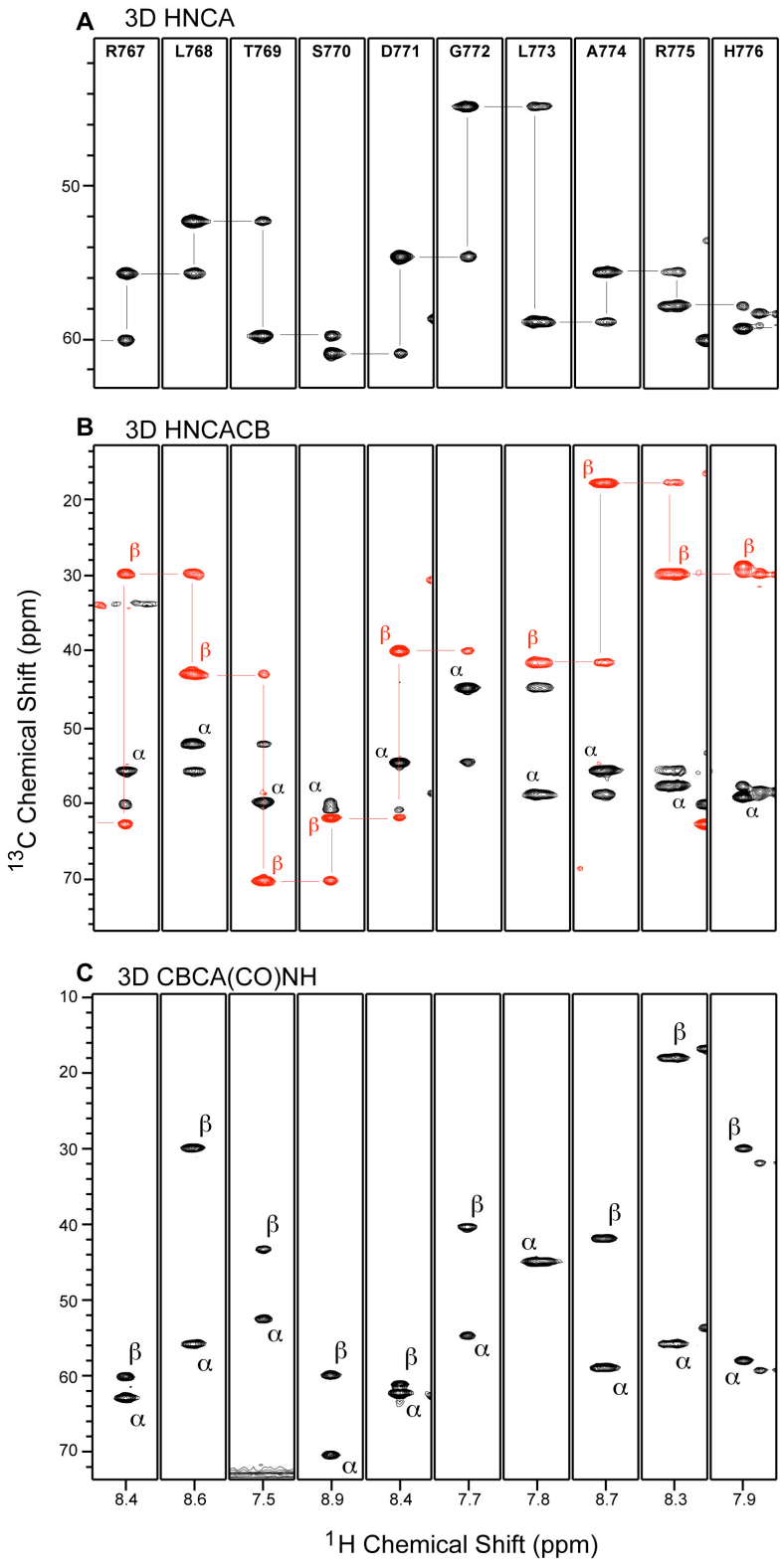

Figure 5-4. Representative strips taken at the ^{15}N planes of residues 767-776 of **(A)** 3D HNCA, **(B)** 3D HNCACB and **(C)** 3D CBCA(CO)NH used in the backbone assignments.

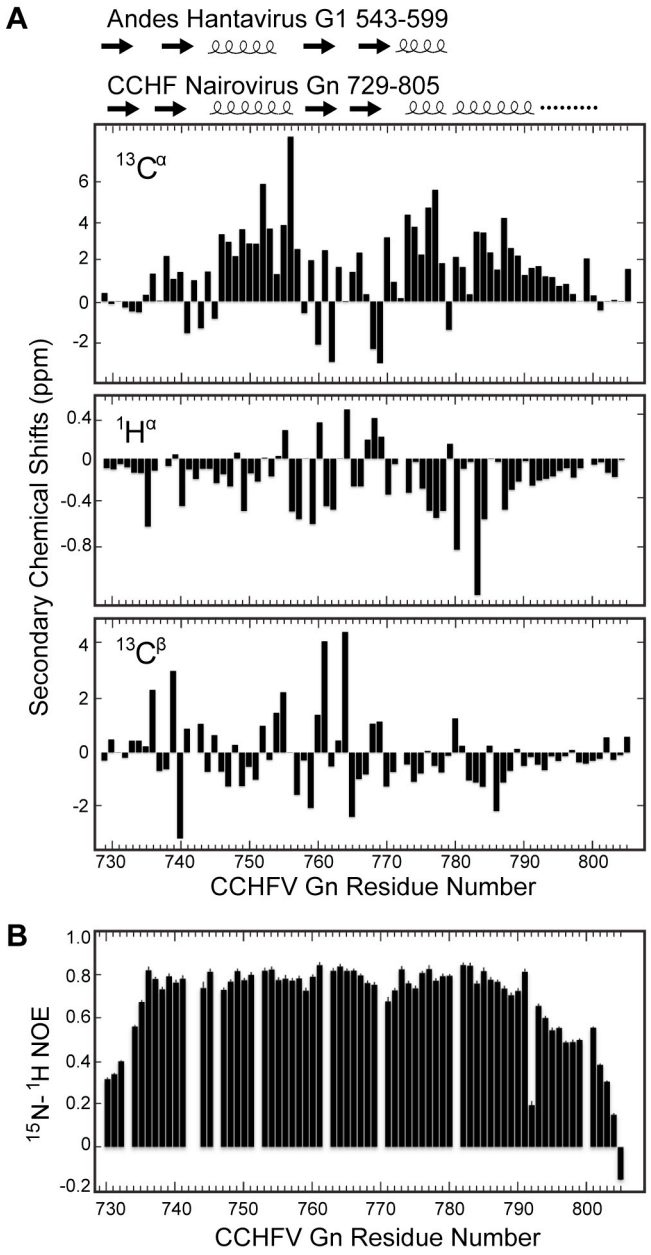

Figure 5-5. **(A)** Secondary chemical shifts for $^{13}C\alpha$, $^{1}H\alpha$, and $^{13}C\beta$ suggest the presence of three short α helices interspersed with two short β hairpins. **(B)** Heteronuclear ^{15}N-^{1}H NOE results indicate the structure is largely rigid, with more flexible termini.

Notably, His^{752} Hε1 shares an NOE with Cys^{736} Hβ's. Similarly, His^{776} Hε1 and Hδ2 share NOE'S with Cys^{761} and Cys^{780} Hβ's. A 2D ^{15}N HMQC (23) spectrum showed that His^{752} and His^{776} coordinated Zn^{2+} through the Nδ1 and Nε2 atoms (**Figure 5-6**), respectively. Manual analysis of 3D ^{15}N- and ^{13}C-edited NOESY spectra identified 1193 unambiguous interproton NOE distance restraints. The NOE restraints together with 26 φ and 26 ψ dihedral angle restraints and zinc coordination restraints (**Table 5-1**) were used in structure calculation and refinement in CYANA (35) and AMBER (36). The 20 low-energy NMR structures of Gn$^{729-805}$ converged into a family of structures **(Fig. 5-7A)** with low restraint violations and good Ramachandran plot statistics (**Table 5-1**).

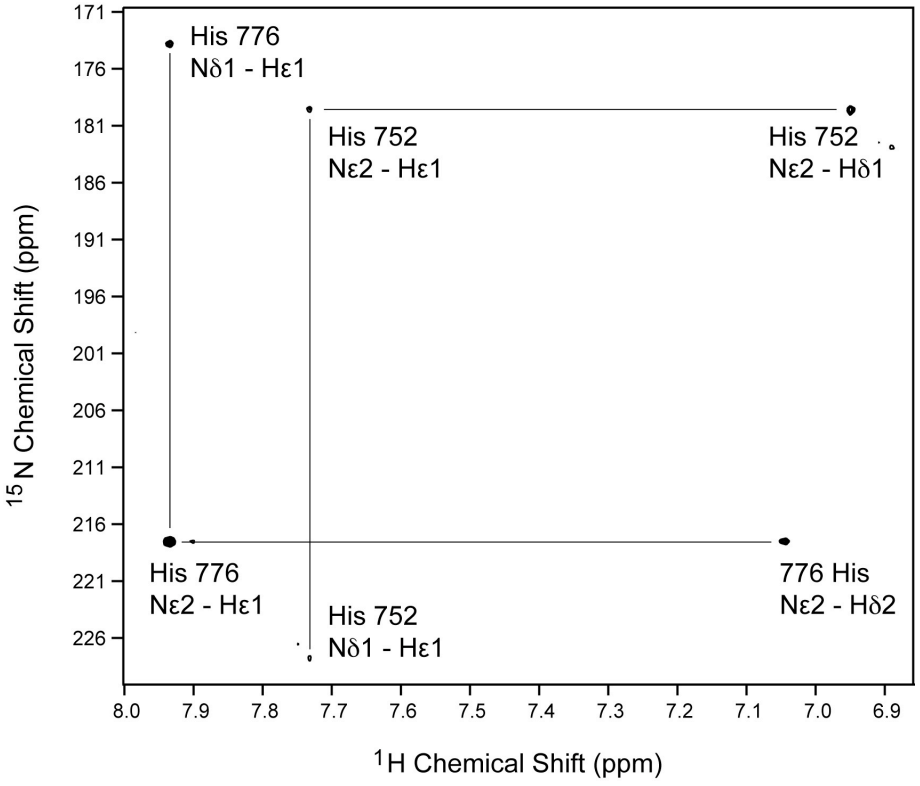

Figure 5-6. A 2D ^{1}H-^{15}N HMQC spectrum identified the tautomeric states of His^{752} and His^{776} and the the histidine ring nitrogen atoms that are Zn^{2+}-coordinated. The spectrum was acquired with 70 ppm ^{15}N sweep width centered at 195 ^{15}N ppm. The analysis was based on Pelton et al. (*Prot. Sci.*, 1993, 2, 543-58).

Table 5-1. Restraints and structural statistics for 20 NMR structures.

Total NOE-derived distance restraints	1193
Intraresidue (i, i)	246
Sequential (i, i +1)	407
Long Range (|i- j)| > 4	307
Total dihedral angle restraints	52
Phi ϕ	26
Psi ψ	26
Root mean square difference from mean structure	
Backbone atoms (N, Cα, C′) (Å)	0.25
All heavy atoms (C, N, O) (Å)	0.76
Violation analysis	
Max distance violation (Å)	0.47
Max dihedral angle violation (º)	5.3
Energies	
Mean generalized Born potential-AMBER energy (kcal mol^{-1})	-3359
Mean restraint energy (kcal mol^{-1})	79
Ramachandran plot (%)	
Most favorable region	79.2
Additionally allowed regions	20.1
Generously allowed regions	0.6
Disallowed regions	0.2

Structure of CCHF Virus Zinc Finger – The NMR structure of $Gn^{729-805}$ reveals a rigid, compact three-helix structure with four short beta strands (**Fig. 5-7A**). The structure contains a pair of tightly associated, back-to-back ββα zinc fingers connected by a short four-residue linker (Ser^{757}-Ile^{760}) (**Fig. 5-7B**). The first CCHC-zinc finger array (ZF1, right side) consists of a Zn^{2+} ion coordinated to residues Cys^{736}, Cys^{739}, $His^{752,}$ and Cys^{756} and forms the classical ββα zinc finger fold. Cys^{736} and Cys^{739} form part of a short β-hairpin. Thr^{737} and Ile^{738} form a loop with Cys^{736} and Cys^{739} on either side of the hairpin. The ZF1 structure contains helix α1 formed by

Ile[747] to Ser[757] that folds back toward the β-hairpin, forming the ββα zinc finger fold. Cys[756] forms the fourth Zn^{2+}-coordinating residue and is located on the same surface of helix α3 with His[752].

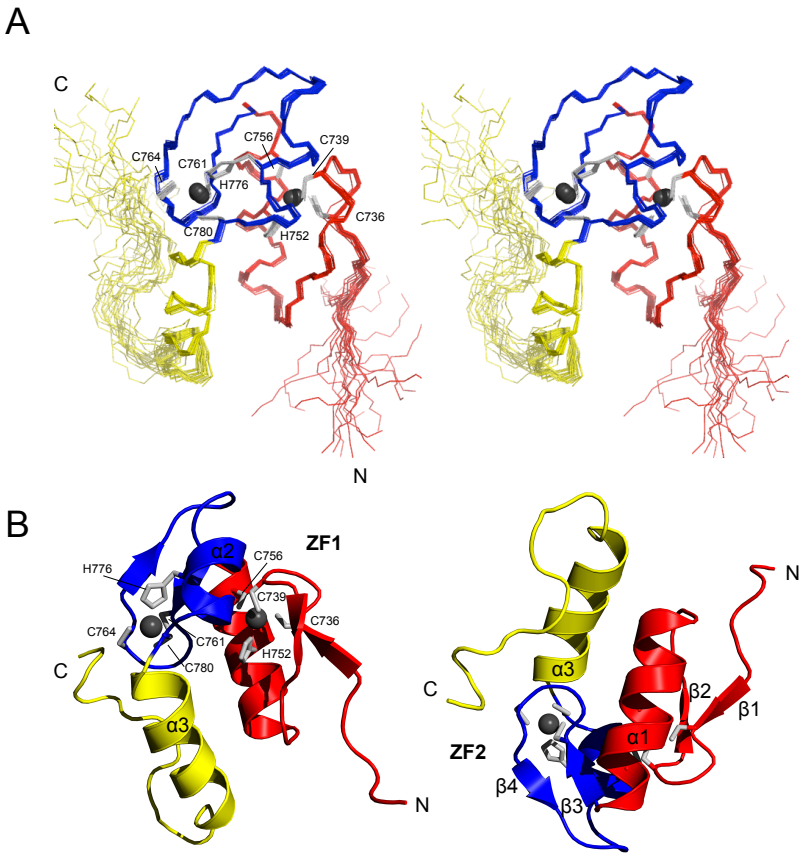

Figure 5-7. (A) Stereoview of the superposition of 20 lowest energy NMR structures of CCHF virus Gn zinc finger. **(B)** CCHF virus Gn zinc finger domain folds into a compact three-helix structure consisting of two back-to-back ββα zinc fingers with helix α3 pinned underneath the core zinc finger structure.

Similarly, the second CCHC zinc finger array (ZF2, left side) consists of a second Zn^{2+} ion coordinated to residues Cys[761], Cys[764], His[776], and Cys[780] into a classical ββα zinc finger fold. Cys[761] and Cys[764] are positioned on either side of a short β-hairpin. Pro[762] and Tyr[763] form a loop

between Cys^{761} and Cys^{764}. The structure is followed by a short helix α2 formed by residues 773-780 and folded back toward the β-hairpin, forming the ββα zinc finger fold. His^{776} is located toward the middle of helix α3, and the final coordinating cysteine (Cys^{780}) is located at the end of helix α3. Although ZF2 also resembles the classical ββα fold, a minor difference exists when compared to ZF1. The helix α2 of ZF2 is shorter than helix α1 by three residues. This is due to the presence of helix breakers Gly^{772} and Pro^{781} located at either end of helix α2.

In addition to the two classical ββα fold zinc fingers, the structure contains an additional helix, helix α3, formed by residues 782-791 that packs against the dual zinc finger fold. A hydrophobic interaction between Val^{744} of ZF1 and Val^{787} keeps helix α3 pinned to the core structure. The orientation of helix α3 to ZF2 is partially determined by Pro^{781}. Strong Cys Cα to Pro^{718} Cδ NOE's indicated a trans proline isomer. The C-terminal 13 residues (Leu^{792} - Lys^{805}) are primarily in a random coil conformation. The heteronuclear ^{15}N-^{1}H NOE values between 0.4 to 0.6 of all but the final three residues of this region (**Fig. 5-5B**) suggest flexibility. A NOE between Ile^{799} Cγ2 and Met^{751} Cγ indicate the C-terminal loop is pinned to the rest of the structure.

CCHF Virus Zinc Finger Contains Conserved Electrostatic Surfaces – Analysis of the surface electrostatics of the CCHF virus $Gn^{729\text{-}805}$ reveals clustering of positively and negatively charged surfaces on opposite faces of the structure (**Fig. 5-8**). Surface residues Glu^{740}, Glu^{750}, and Asp^{753} of ZF1 converge with surface residues Glu^{789} and Glu^{791} of helix α3 and Glu^{797} of the C-terminal tail to form a large, nearly contiguous negatively charged surface. Similarly, Arg^{767}

and Arg^{775} of ZF2 converge with Lys^{782}, Lys^{784}, and Lys^{786} of helix α3 and Arg^{798} of the C-terminal tail to form a large contiguous positively charged surface. Of the charged surface residues, Glu^{750}, Glu^{797}, Lys^{784}, and Arg^{798} are not conserved in *Nairoviruses*. Most of the surface electrostatics, therefore, are conserved features of the CCHF virus zinc finger domain.

CCHF Virus Gn Tail Binds RNA – RNA electrophoretic mobility shift binding assays were carried out using two different constructs: the zinc finger domain of CCHF virus consisting of $Gn^{729-805}$ and the core zinc finger domain $G1^{534-599}$ of the related Andes virus, genus *Hantavirus*. Each respective assay used a different sequence of RNA corresponding to the M segment panhandle sequence of either virus. The CCHF virus M segment panhandle transcripts migrated as two bands, with the higher band representing a likely RNA tetramer and the lower band migrating at a size consistent with monomeric hairpin-like panhandle. Incubation of the Andes *Hantavirus* structure with the Andes *Hantavirus* M segment panhandle (**Fig. 5-9A**) failed to alter the migration of RNA through a 12% native acrylamide gel (**Fig. 5-9B**). However, incubation with the CCHF construct notably altered the migration of RNA transcripts, as demonstrated by the presence of an additional band migrating just below the tetramer RNA band (**Fig. 5-9B**). The shifted band likely represents a complex between the RNA panhandle and the CCHFV protein. Our results indicate that the CCHF *Nairovirus* zinc finger binds RNA, whereas the core Andes *Hantavirus* does not, as reported earlier (Chapter 3).

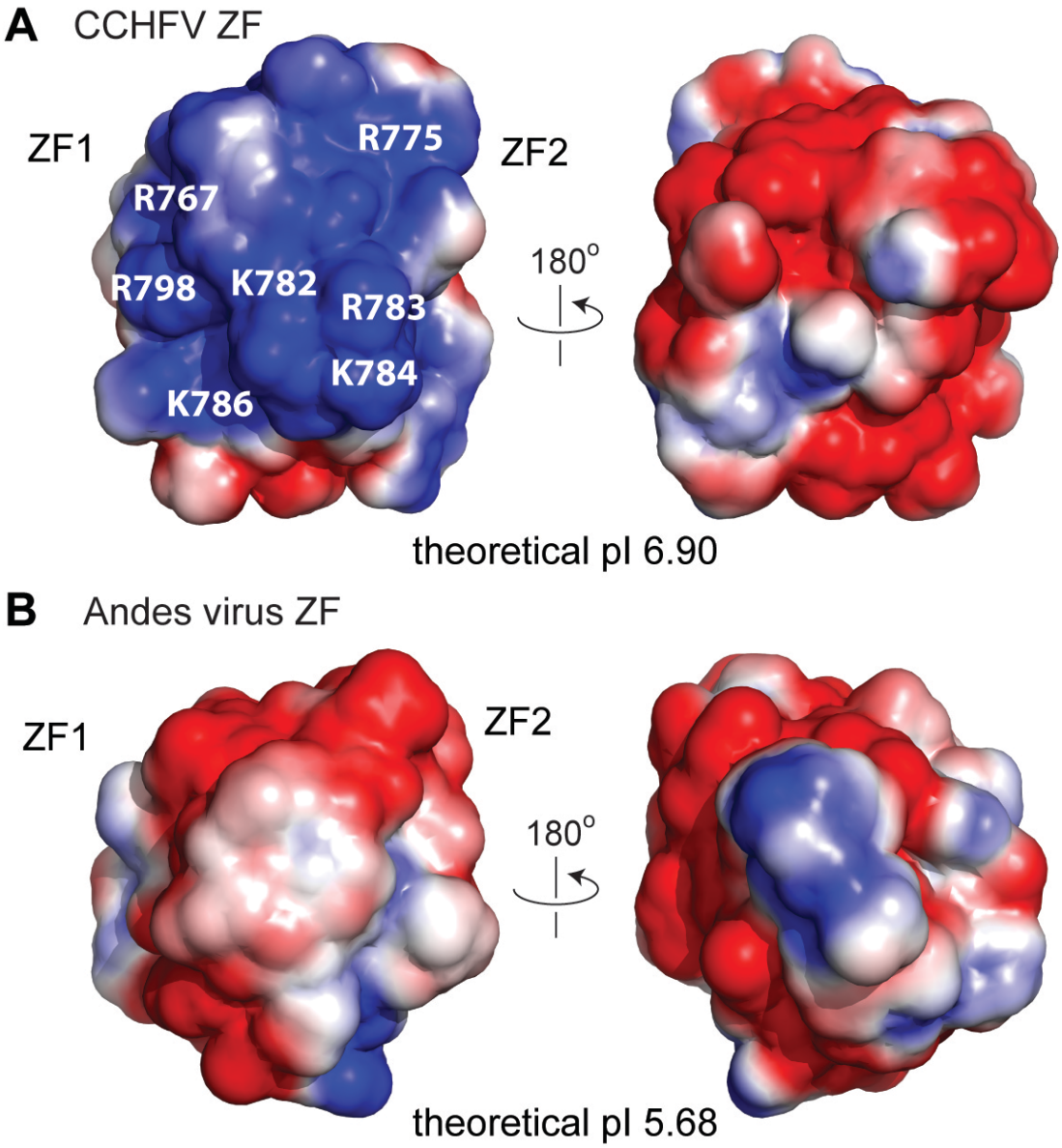

Figure 5-8. **(A)** Analysis of the surface electrostatics of the CCHF virus Gn zinc finger reveals a contiguous cluster of basic charges (colored blue) that cover the entire half of the structure. Rotating the structure 180° reveals an equally large cluster of acidic charges (colored red) on the opposite face. **(B)** Surface electrostatics of the Andes *Hantavirus* zinc fingers (PDB 2K9H) in the same orientation as the CCHF virus structure. Notably, the clustering of conserved basic surface in *Nairovirus* is absent in *Hantavirus*.

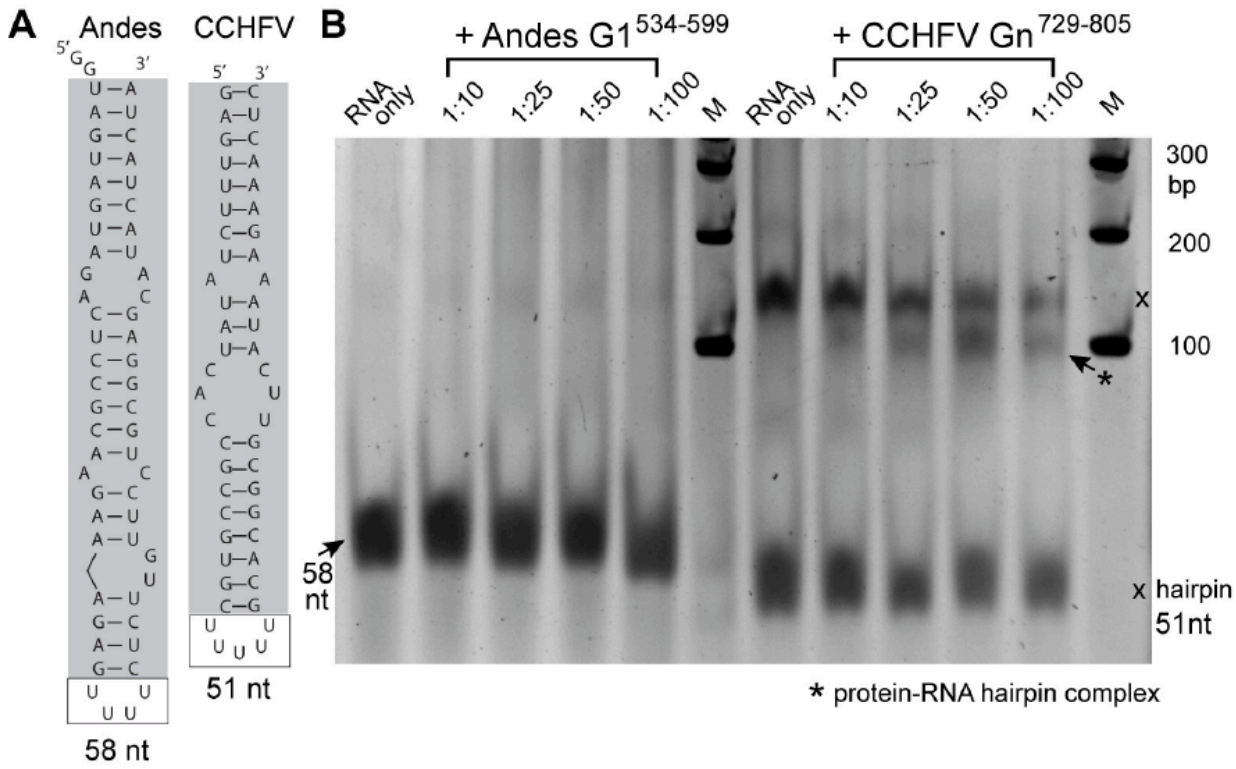

Figure 5-9. RNA electrophoretic mobility shift assay comparing the Andes virus and the CCHF virus zinc fingers. While the core Andes virus zinc finger fails to alter the mobility (B, left) of Andes M segment panhandle RNA (A, left), increasing the amount of CCHF virus zinc finger causes the appearance of a secondary band (B, right) for CCHF M segment panhandle RNA (A, right), thus suggesting a protein- RNA complex. Each lane contained 0.24 umol RNA, proteins came from 0.4 mM stock diluted accordingly.

Discussion

Zinc fingers have been known to be common viruses (in particular, the HIV-1 nucleocapsid protein zinc fingers (31) are critical in RNA packaging and viral assembly). However, zinc fingers are rarely found in viral envelope glycoproteins. Including the zinc finger in this report, there are currently only three known structures of viral envelope glycoprotein zinc fingers. First is the zinc finger of the *Hantavirus* G1 envelope glycoprotein (11), and second is the zinc finger of the Junin virus envelope glycoprotein (32) (also an RNA virus of family *Arenaviridae*). In all

three cases, the cytoplasmic tails of the envelope glycoproteins contain the zinc finger domains. The Junin virus zinc fingers (32) form a unique fold that do not show any structural or sequence similarity with the *Nairovirus* and *Hantavirus* zinc fingers. Although the *Nairovirus* and the *Hantavirus* zinc fingers (11), which both belong to family *Bunyaviridae*, show a similar global fold (**Fig. 5-7**, **Fig. 5-10A**), they also have significant structural differences and properties.

Key Differences Between Nairovirus and Hantavirus Gn Zinc Fingers – Examination of the surface electrostatics of the CCHF virus Gn tail reveals key differences when compared to the Andes *hantavirus* structure. Whereas the CCHF Gn tail displays sharp clustering of conserved charges that form a large contiguous swath on the protein surface, the Andes *Hantavirus* zinc fingers display charges that are predominately negative and appear randomly dispersed (**Fig. 5-8**). The variation in charge conservation is also evident in the sequence analysis (**Fig. 5-10B**). Whereas the spacing of the CCHC motif is mostly conserved, the spacing between conserved charges is highly variable. The CCHF virus displays conserved acidic residues on ZF1 and conserved basic residues on ZF2. The Andes *Hantavirus* sequence, however, displays clustering of conserved basic residues on the sequences flanking the negatively charged core zinc finger structure.

Moreover, the CCHF virus Gn tail contains a structural motif that is absent in the Andes *Hantavirus* structure. Helices $\alpha 2$ and $\alpha 3$, residues Leu^{773}-Leu^{792}, form a helix-kink-helix motif due the positioning of the helix breaker Pro^{780}. While not an uncommon motif, this structural aspect in the CCHF Gn tail forms the core scaffold for a large basic surface partly composed of the conserved residues Arg^{775}, Lys^{782}, and Lys^{786}. By contrast, the Andes *Hantavirus*

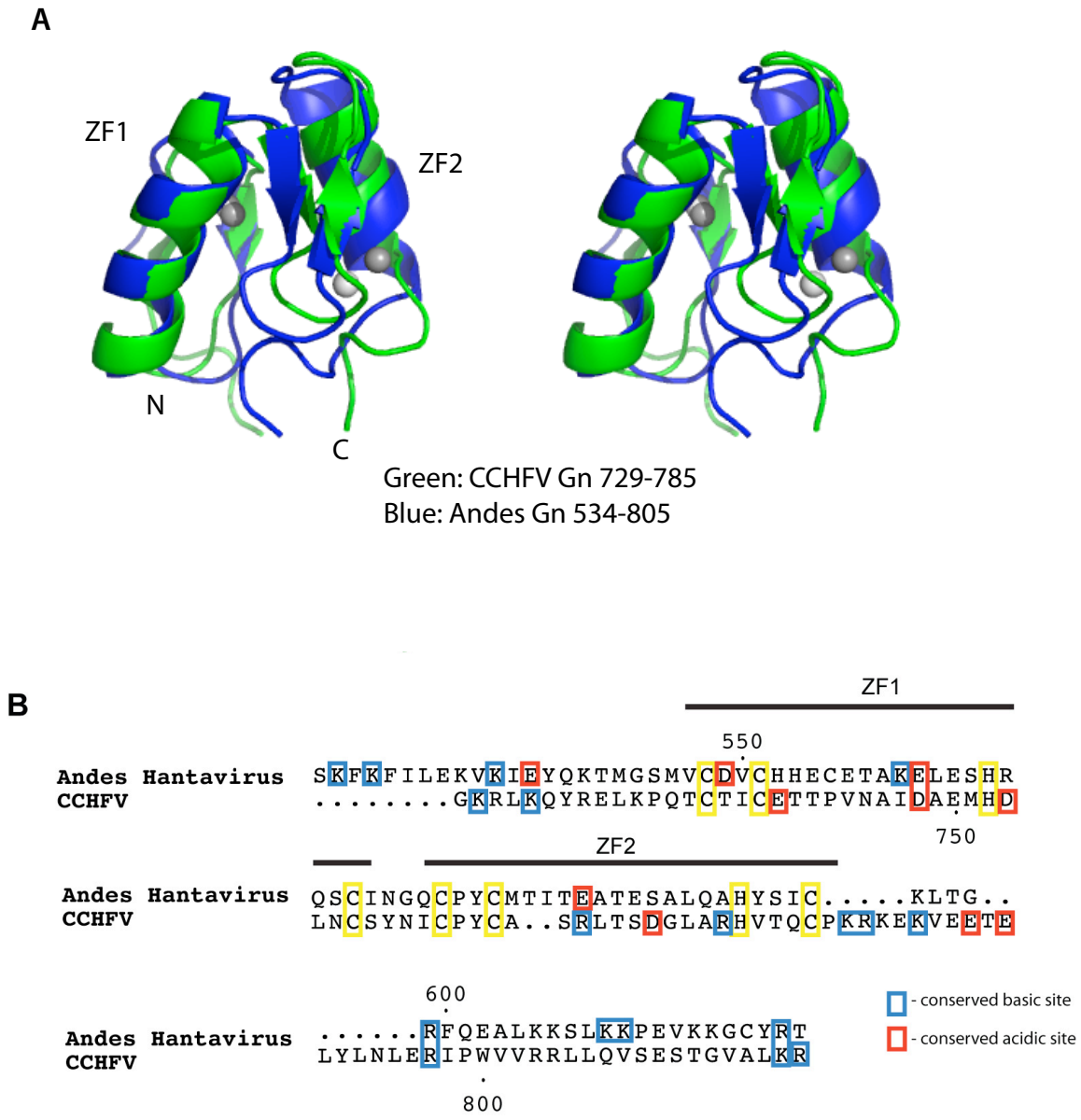

Figure 5-10. **(A)** Stereo superposition of the core zinc finger structures of the CCHF virus and the Andes virus reveals that both cytoplasmic tails contain a similar fold. The backbone Cα rmsd for the helical regions (20 residues) was approximately 0.8Å. **(B)** Remarkably, the sequence identity between the two structures is a mere 12%. Sequence variability is also evident when comparing the distinct distribution of conserved charges.

structure contains neither the corresponding proline nor the charges to support a similar motif (**Fig. 1-3**). Instead, it contains the non-conserved helix breaker Gly^{598} followed by conserved basic residues exclusively on the C-terminal end of ZF2. In this respect, the surface electrostatics of the CCHF virus Gn tail may more closely resemble that of *Orthobunyaviruses*, with conserved helix breakers flanked by conserved basic residues. Overall, the general preservation of the fold indicates that the dual zinc finger motif plays a general but important role in the life cycle of both *Nairoviruses* and *Hantaviruses*. However, the major differences in the surface electrostatics of the CCHF virus and *Hantavirus* cytoplasmic tail structures also suggest that while general (i.e., a function required by all Bunyaviruses), the function of the tail may be species specific.

RNA Binding Properties of the CCHFV Gn 729-805. The RNA binding properties of the CCHFV Gn tail provide additional clues regarding its function. While the core Andes zinc finger domain in this study once again failed to form a protein-RNA complex, subsequent studies described in Chapter 7 indicate that an extended form of the Andes Gn tail does have RNA binding ability. Therefore, the RNA binding properties of the CCHFV Gn tail also appear to be conserved in Hantaviruses, despite the obvious differences in the surface electrostatics between the two cytoplasmic tails. Notably, in the case of the CCHFV Gn tail (729-805), the conserved basic residues actually don't confer a net basic charge on the construct. The theoretical isolectric point of the construct is estimated at approximately 6.90, due in large part to the presence of an equal number of acidic charges. Therefore, the distribution of basic residues on the surface of the structure is not random; there is a clear and deliberate segregation of charged residues. This observation suggests a definite function for the electrostatics in the function of this domain.

Proposed Role in Viral Assembly. Given the data available regarding the Bunyaviridae Gn role in viral assembly (8-10), we propose that the surface electrostatics play an important role in assembly of the CCHF virus, presumably via direct interaction with some component of the ribonucleoprotein (**Fig. 5-11**). The large positively charged surface of the Gn tail would suggest RNA binding, as is the traditional role for zinc fingers in retroviruses (31,33). Our EMSA results suggest that this may be the case (**Fig. 5-9B**). Adding increasing amounts of $Gn^{729\text{-}805}$ resulted in the presence of an additional RNA band of increasing intensity. Because accelerated migration of RNA is rare in EMSA's, the shifted band most likely represents a slower migration of a protein-RNA of the monomeric panhandle molecule. The decreasing band intensity of the tetrameric RNA likely indicates the two species are in equilibrium, with the presence of the protein sequestering monomer and therefore causing the equilibrium to shift away from the high molecular weight species. These results suggest the possibility of an interaction between the Gn tail and the RNA component of the ribonucleoproteins (**Fig. 5-11**). A reverse genetics system for studying the CCHF virus has only recently been developed (34), thus allowing testing of this model in the future.

In summary, the NMR structure of a zinc finger domain in the Gn tail of the CCHF virus has been determined. Currently, this is the only available atomic structure for a protein component of the *Nairovirus* genus. The global fold of this zinc finger is similar to that of the *Hantavirus* zinc finger, which represents a unique fold of two classical-type ββα-zinc fingers that are structurally interdependent (in contrast, individual classical ββα-zinc fingers behave as independent domains, like beads-on-a-string). The CCHF virus Gn tail was also demonstrated to bind RNA *in vitro*, thus suggesting the possibility of an interaction between the Gn tail with the

viral RNA. These results contribute novel mechanistic insight toward understanding the CCHF virus life cycle.

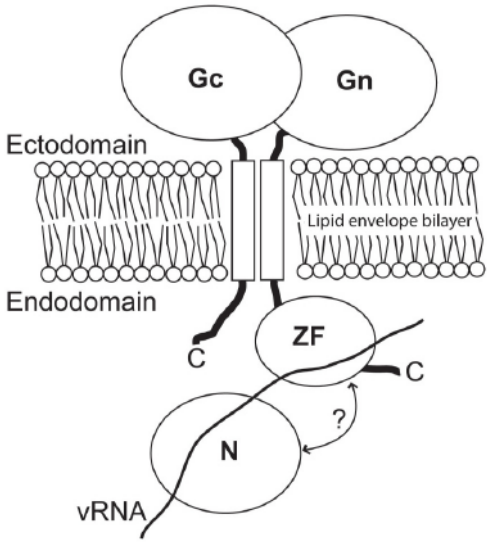

Figure 5-11. Surface electrostatics combine with EMSA results of the Gn tail provide mechanistic insight into RNP packaging. We propose a model in which packaging consists of a zinc finger-RNA complex. Studies in related viruses suggest a nucleocapsid – Gn tail interaction, presented here as a possible additional packaging site.

References

1. Ergonul, O. (2006) *Lancet Infect Dis* **6**, 203-214
2. Whitehouse, C. A. (2004) *Antiviral Res* **64**, 145-160
3. Elliott, R. M., Bouloy, M., Calisher, C. H., Goldbach, R., Moyer, J. T., Nichol, S. T., Pettersson, R., Plyusnin, A., and Schmaljohn, C. (2000) *Bunyaviridae*. in *Virus Taxonomy: The classification and Nomenclature of Viruses. The Seventh Report of the International Committee on Taxonomy of Viruses.* (Van Regenmortel, M. H. V., Fauquet, C. M., Bishop, D. H. L., Carsten, E. B., Estes, M. K., Lemon, S. M., Maniloff, J., Mayo, M. A., McGeoch, D. J., Pringle, C. R., and Wickner, R. B. eds.), Academic Press, San Diego. pp 599-621
4. Schmaljohn, C. S., and Hooper, J. W. (2001) Bunyaviridae: the viruses and their replication, 4th ed. in *Fields Virology* (Knipe, D. M., and Howley, P. M. eds.), Lippincott, Williams & Wilkins, Philadelphia. pp 1581-1602
5. Sanchez, A. J., Vincent, M. J., and Nichol, S. T. (2002) *J. Virol.* **76**, 7263-7275
6. Sanchez, A. J., Vincent, M. J., Erickson, B. R., and Nichol, S. T. (2006) *J Virol* **80**, 514-525

7. Altamura, L. A., Bertolotti-Ciarlet, A., Teigler, J., Paragas, J., Schmaljohn, C. S., and Doms, R. W. (2007) *J. Virol.* **81**, 6632-6642
8. Overby, A. K., Pettersson, R. F., and Neve, E. P. (2007) *J. Virol.* **81**, 3198-3205
9. Shi, X., Kohl, A., Li, P., and Elliott, R. M. (2007) *J. Virol.* **81**, 10151-10160
10. Hepojoki, J. M., Strandin, T., Wang, H., Vapalahti, O., Vaheri, A., and Lankinen, H. (2010) *J. Gen. Virol.* **91**, 2341-2350
11. Estrada, D. F., Boudreaux, D. M., Zhong, D., St Jeor, S. C., and De Guzman, R. N. (2009) *J. Biol. Chem.* **284**, 8654-8660
12. Chatterjee, S., Zhong, D., Nordhues, B. A., Battaile, K. P., Lovell, S. W., and De Guzman, R. N. (2010) *Protein Sci. in press*
13. Geisbrecht, B. V., Bouyain, S., and Pop, M. (2006) *Protein Expr. Purif.* **46**, 23-32
14. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR* **6**, 277-293
15. Johnson, B. A. (2004) *Methods Mol. Biol.* **278**, 313-352
16. Grzesiek, S., and Bax, A. (1993) *J. Am. Chem. Soc.* **115**, 12593-12594
17. Grzesiek, S., Dobeli, H., Gentz, R., Garotta, G., Labhardt, A. M., and Bax, A. (1992) *Biochemistry* **31**, 8180-8190
18. Wittekind, M., and Mueller, L. (1993) *J. Magn. Reson.* **101B** 201-205
19. Wishart, D. S., and Nip, A. M. (1998) *Biochem. Cell Biol.* **76**, 153-163
20. Tolman, J. R., Chung, J., and Prestegard, J. H. (1992) *J. Magn. Reson.* **98** 462-467
21. Grzesiek, S., and Bax, A. (1993) *J. Biomol. NMR* **3**, 185-204
22. Hoffman, R. C., Xu, R. X., Klevit, R. E., and Herriott, J. R. (1993) *J. Magn. Reson. Ser. B* **102**, 61-72
23. Pelton, J. G., Torchia, D. A., Meadow, N. D., and Roseman, S. (1993) *Protein Sci.* **2**, 543-558
24. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., and Clore, G. M. (1989) *Biochemistry* **28**, 6150-6156
25. Stone, M. J., Fairbrother, W. J., Palmer, A. G., III, Reizer, J., Saier, M. H., Jr., and Wright, P. E. (1992) *Biochemistry* **31**, 4394-4406
26. Guntert, P. (2004) *Methods Mol. Biol.* **278**, 353-378
27. Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham Iii, T. E., Wang, J., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Tsui, V., Gohlke, H., Radmer, R. J., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K., and Kollman, P. A. (2002) *AMBER7*, University of California, San Francisco
28. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR* **8**, 477-486
29. DeLano, W. L. (2002) *The PyMOL Molecular Graphics System* DeLano Scientific, San Carlos, California, U.S.A.
30. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10037-10041
31. De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N., and Summers, M. F. (1998) *Science* **279**, 384-388
32. Briknarova, K., Thomas, C. J., York, J., and Nunberg, J. H. (2010) *J. Biol. Chem.* **in press**
33. D'Souza, V., and Summers, M. F. (2005) *Nat. Rev. Microbiol.* **3**, 643-655
34. Flick, R., Flick, K., Feldmann, H., and Elgh, F. (2003) *J. Virol.* **77**, 5997-6006

35. Guntert, P., *Automated NMR structure calculation with CYANA*. Methods Mol. Biol., 2004. **278**: p. 353-378.

36. Case, D.A., *et al.*, *AMBER7*. 2002, University of California, San Francisco.

Chapter 6: NMR Characterization of the Extended Andes Zinc Finger Domain (534-610)

Introduction

The *Hantavirus* Gn cytoplasmic tail plays a probable role in virus assembly events at the *cis*-golgi membrane [1-4]. Most recently, Hepoki *et al*. (2010) reported that the *Hantavirus* Gn cytoplasmic tail is capable of interacting with the nucleocapsid (N) protein [1]. Specifically, the interaction was mapped to a series of possible binding sites located on either side of the core zinc finger domain of the Gn tail (**Figure 6-1**) [1].

Because viral RNA is also a component of the ribonuleocapsid complex, these findings raise the possibility of a further interaction between *Hantavirus* Gn tails and viral RNA. Preliminary electrophoretic mobility shift experiments (Chapter 7) using the Andes core zinc finger fold, residue 534-599, suggest the zinc fingers alone are insufficient to bind RNA. However, a larger construct of the Gn tail, comprised of residues 534-610, does show some RNA binding capability (Chapter 7).

Based on this data, it appears that the presence of the terminal flanking sequences (residues 534-543 and 599-610) may be important for an interaction between the Gn tail and RNA. Specifically, each flanking sequence or arm contains conserved basic residues (**Figure 6-1**), which may be important in mediating an interaction with vRNA, which is the most negatively charged component of the ribonucleocapsid complex.

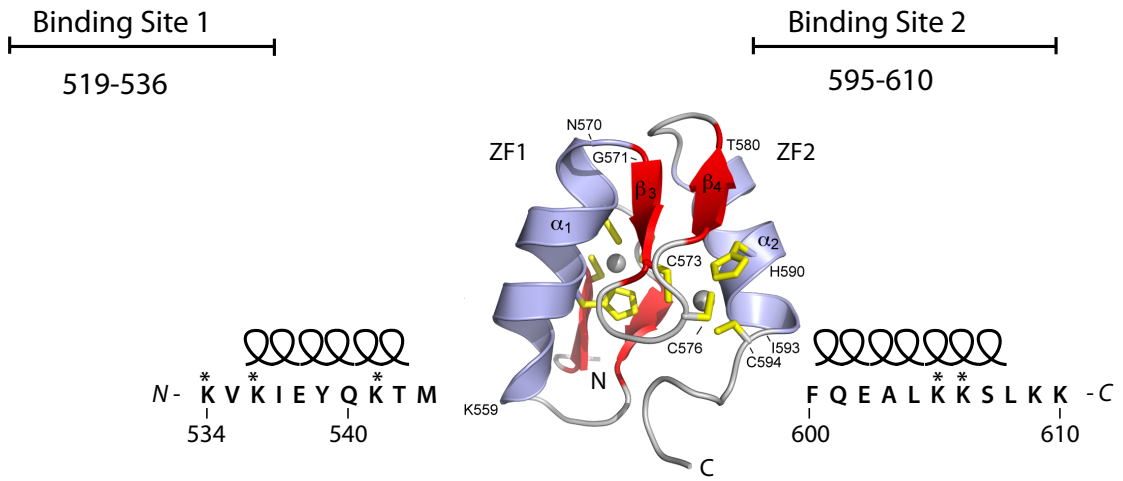

Figure 6-1. The core zinc finger domain of the Andes Gn tail (PDB ID 2K9H) with terminal flanking sequences. Secondary structure prediction for each flanking arm was derived from JPRED [5]. Conserved basic charges are marked with an asterisk. N protein binding sites 1 & 2 are as reported by Hepojoki *et al*., 2010 [1]. The extended Andes construct used in this analysis contains the entire Binding Site 2 and the final three residues (two of which are conserved basic residues) of Binding Site 1.

In this study, 2D and 3D NMR spectroscopy were used to characterize the behavior of these poly-basic arms. The NMR data suggests that, despite the predicted helical secondary structure at the amino terminus, this first poly-basic arm is largely disordered. The carboxyl terminal sequence, however, does contain a short α-helix that contains a conserved basic patch. Furthermore, neither arm appears to interact with the core zinc finger fold and are semi-flexible, which is consistent with motion independent of the rigid core structure. These results provide further insight into a possible *Hantavirus* Gn tail RNA binding motif that may be unique from that of other Bunyaviruses.

Results

Protein Production and NMR assignments – Given the toxicity of the full length Andes Gn tail and the poor solubility of larger constructs (Chapter 3), the protein used in this study,

consisting of residues 534-610, represents the largest construct (76 residues) of the Gn tail produced. Even still, half of the protein yield after overnight expression at 15°C is in the form of insoluble inclusion bodies. The ^{15}N-^{13}C labeled sample prepared for this study was natively expressed in the supernatant and pooled from two 1 L batches of *E. coli*. In an attempt to improve the overall yield, the insoluble fractions were solubilized in 6 M urea and 10 mM DTT, then refolded by step-wise dialysis to remove the denaturant. The refolded GB1 fusion protein was monomeric and cleavable, however, while the ^{1}H-^{15}N-HSQC spectrum of the refolded and cleaved protein showed a similar peak pattern for the core zinc finger domain, several of the additional peaks corresponding to the poly-basic arms had a different chemical shift than the natively expressed protein (**Fig. 6-2**). Therefore, the 3D NMR data used for our analysis was collected exclusively on the natively expressed Andes Gn 534-610.

Overall, Andes Gn 534-610 gave a well-resolved 2D $^{1}H^{15}N$-HSQC spectrum (**Fig. 6-3**). By transferring assignments from the core zinc finger domain (Chapter 3) and analyzing the 3D HNCA data set, the complete assignment of the flanking poly-basic arms was possible (**Figure 6-4**).

Secondary chemical shift profile – Based on the HNCA derived assignments of the backbone, we have constructed the secondary chemical shift profile of Andes Gn 534-610 (**Fig. 6-5, panel B**). Notably, analysis of the profile suggests the amino terminal sequence (Lys^{534}–Met^{543}) is completely disordered (**Fig. 6-5, panel B**). These results run contrary to the predicted presence of a short α-helix (**Fig. 6-1**). The α-helix predicted at the carboxyl terminus appears

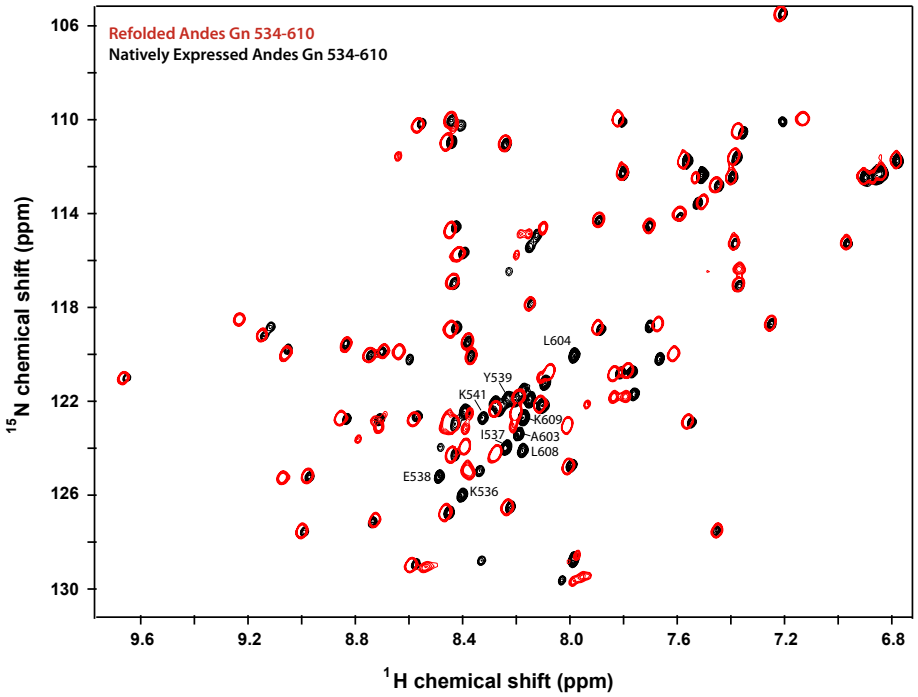

Figure 6-2. ^{1}H-^{15}N-HSQC spectral overlay of Andes Gn 534-610 from the natively expressed (black) fraction and the refolded (red) fraction. The peaks of the refolded spectrum are displayed with fewer contours in order to emphasize overlaid peaks.

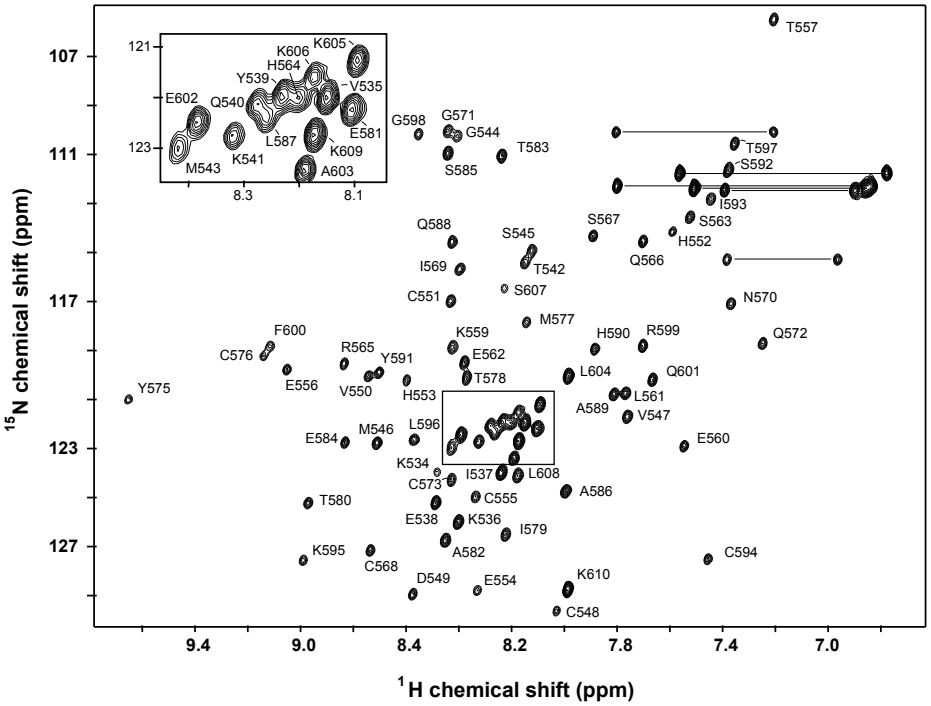

Figure 6-3. ^{1}H-^{15}N-HSQC of the assigned Andes Gn 534-610.

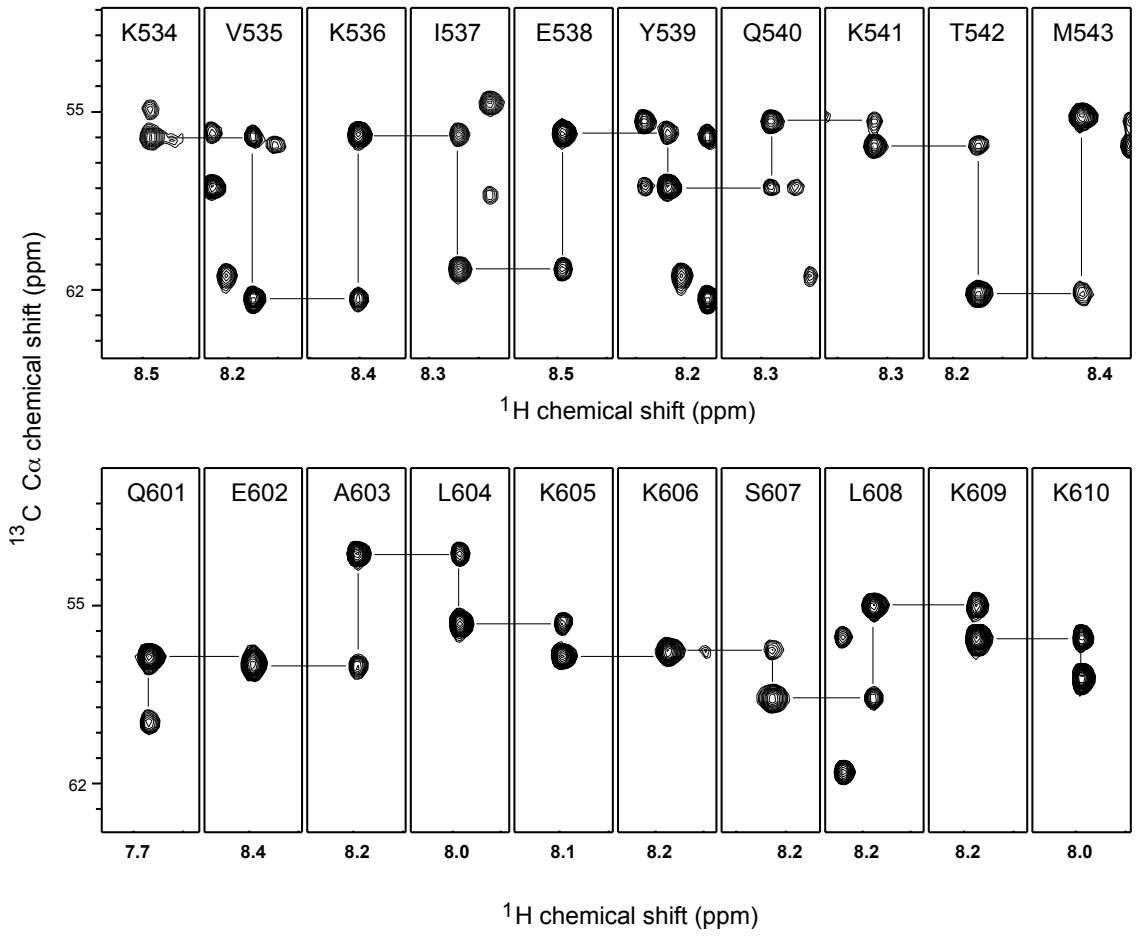

Figure 6-4. Backbone HNCA assignments of the Andes Gn 534-610 residues K534-M543 (top panel) and Q601-K610 (bottom panel). All residues consisting of the two poly-basic arms were assigned.

intact between residues Arg^{599} and Lys^{606}. Another interesting feature is that the intervening sequence corresponding to the core zinc finger domain appears largely unchanged from that of the protein without the flanking sequences (**Fig. 6-5, panels A & B**), indicating that the presence of the poly-basic arms does not affect the environment of the core domain backbone. This suggests the arms have minimal contact with the core domain. Furthermore, the similarity in the chemical shift profile of Andes Gn 534-610 with that of CCHFV Gn 729-805 (**Fig. 6-5B,C**) reinforces the notion that the core zinc finger domain contains a similar fold in these different viruses.

Backbone Dynamics of Andes Gn 534-610 – $[^1H]$-^{15}N heteronuclear NOE analysis of Andes Gn 534-610 suggests that the core zinc finger domain is largely a rigid structure (**Fig. 6-6**). This is consistent with the backbone dynamics of CCHFV Gn 729-805. However, the flanking sequences, shaded in gray in **Fig. 6-6**, are considerably more flexible. The disordered amino terminus is almost entirely flexible relative to the core structure as evident by the decreased peak intensity in this region of the $[^1H]$-^{15}N heteronuclear NOE profile. Notably, gradually increasing flexibility at the carboxyl terminus corresponds to the tapering of helix 3, with greatest flexibility at the disordered five terminal residues.

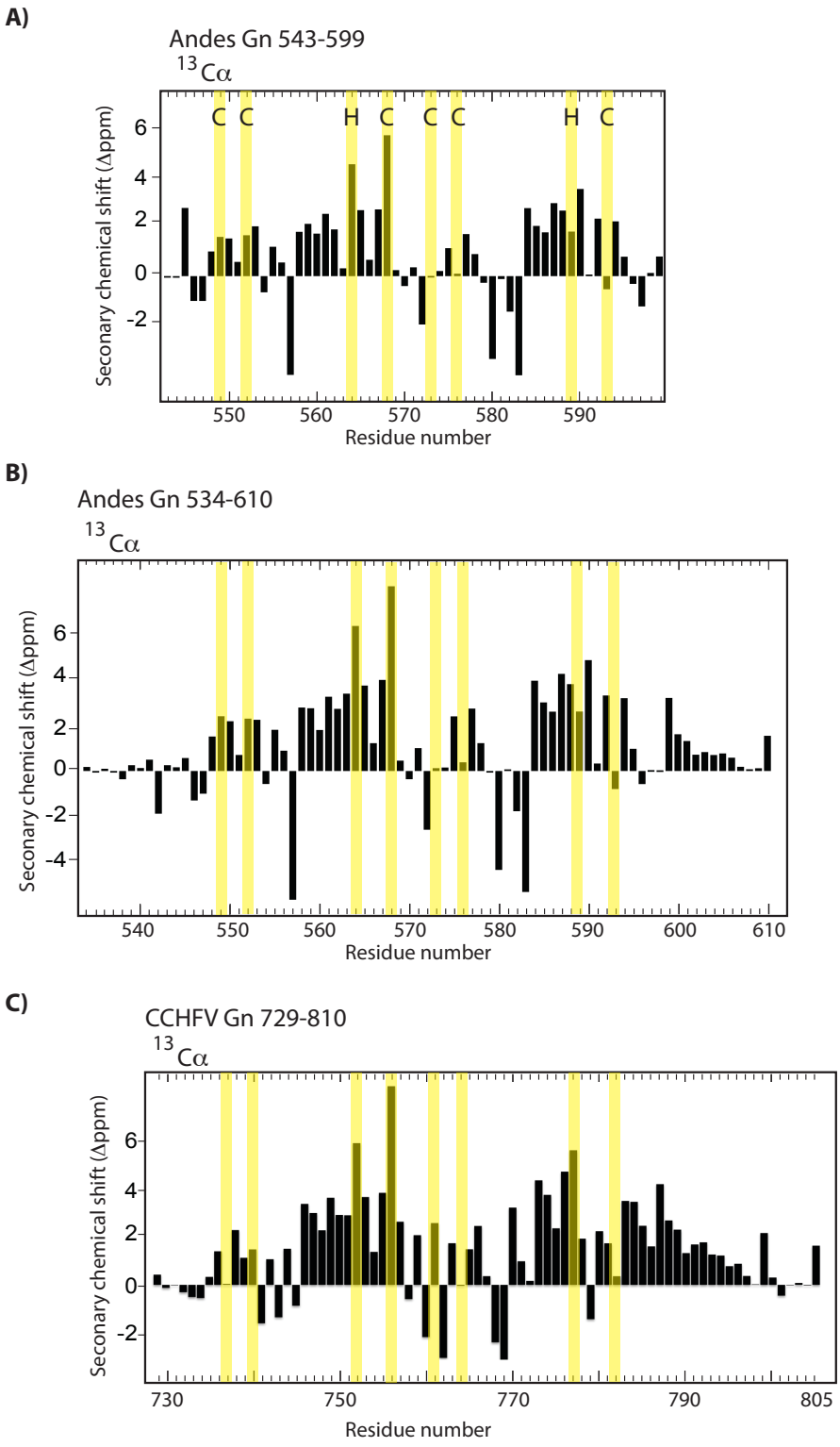

Figure 6-5. Secondary chemical shift profiles of the core Andes zinc finger domain (A), the extended Andes zinc finger domain (B), and the Crimean Congo Hemorrhagic Fever virus (CCHFV) Gn zinc finger domain. All plots are aligned along the core zinc coordinating residues, highlighted with yellow bars.

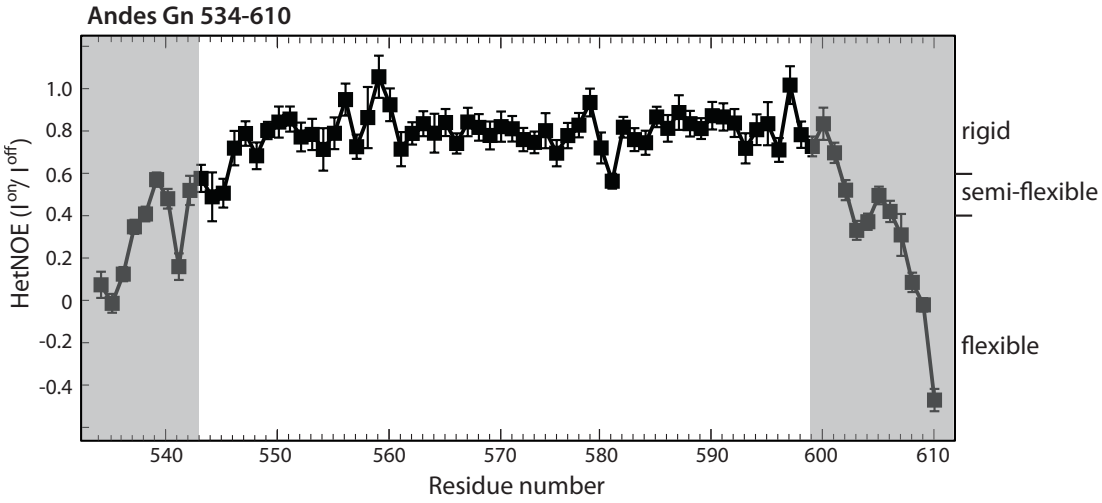

Figure 6-6. [^{1}H]-^{15}N heteronuclear NOE plot reflecting the Andes Gn 534-610 backbone dynamics. The flexible poly-basic sequences at each termini are highlighted in gray.

Discussion

The amino acid sequences flanking the Andes *Hantavirus* zinc finger domain have been shown to be important for a potential Gn-N protein interaction [5]. Specifically, short peptides representing these regions of the Gn tail were able to disrupt the Gn-N protein interaction as detected using a SPOT peptide array [5]. Our results using electrophoretic mobility shift assays suggest some of these same sequences may participate in a protein-RNA interaction. In this study, we have used a construct of the Andes Gn tail (residues 534-610) that contains one of the bindings sites reported by Hepojoki *et al*. (2010) and two conserved basic residues of a second. Namely, Cys^{519}-Lsy^{536} was reported as one bindings site, Lys^{595}-Lys^{610} was reported as another (**Figure 6-1**), and a third binding site was reported at the seven carboxyl residues of the tail [5]. Considering these data, these flanking sequences may have the potential to form multiple contacts during assembly events.

Early studies of the Andes Gn cytoplasmic tail were focused on the core zinc finger domain (Chapter 3) primarily due to relatively poor yield from the extended construct used here. However, since the less sensitive HNCACB and CBCA(CO)NH experiments were not required for these assignments, the low protein yield was overcome the complete backbone assignments, secondary structure, and the dynamic behavior of the flanking sequences are reported.

An overlay of the $^{1}H^{15}N$-HSQC spectrum of the extended Andes *Hantavirus* Gn zinc finger domain with that of the core Andes zinc finger domain (**Figure 6-3**) showed a high degree of correlation in the peak pattern, suggesting the core domain is similarly folded in the extended construct. Overall, there were 18 additional resolvable peaks present in the $^{1}H^{15}N$-HSQC spectrum of Andes Gn 534-610 when compared to the core domain. Of these, 17 were narrowly dispersed between 8.0 and 8.5 ppm in the ^{1}H dimension, suggesting the poly-basic arms may be disordered in the extended structure. However, upon assignment of the additional peaks (**Fig. 6-4**) it was determined that the carboxyl terminus does contain the short α-helix as predicted by JPRED. The amino terminal sequence, by contrast, does appear to be disordered (**Fig. 6-5, panel B**).

Interestingly, while the refolded protein had a similar zinc finger fold, the additional peaks representing the poly-basic arms had a different chemical shift than those in the natively expressed protein (**Fig. 6-2**). This likely represents two different conformations for the terminal sequences. While this is likely an artifact of refolding, it does suggest the arms can be arranged independent of the core zinc finger structure. This notion is supported by the $[^{1}H]$-^{15}N

heteronuclear NOE plot (**Fig. 6-6**), which suggests that much of the flanking sequence is either somewhat or fully flexible.

The short helix located near the carboxyl terminus, which we've termed helix 3 due to the presence of two short helices in the core structure, does contain three conserved basic residues. Arg^{599} is located at the beginning turn of the helix, while the lysine pair located at positions 605 and 606 approximate the end of the helix. A helical wheel projection for helix 3 (**Fig. 6-7**) indicates that Arg^{599} and Lys^{606} are likely oriented in the same direction and may form a conserved basic patch. Notably, the surface opposite these charges contains the two conserved hydrophobic residues Phe^{600} and Leu^{604}, suggesting the helix is amphipathic. Despite this, the helix does not appear to form part of the independently folded core domain described in Chapter 3.

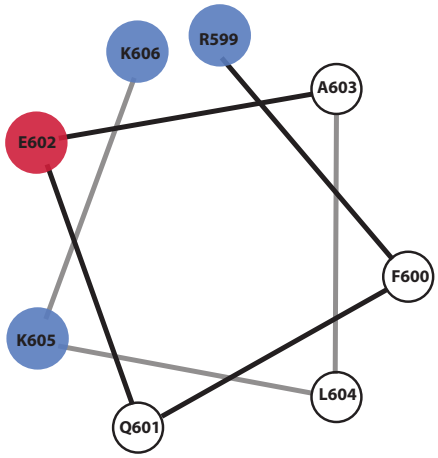

Figure 6-7. Helical wheel projection of Andes Gn 534-610, helix 3. Projection was created using the program at http://rzlab.ucr.edu/scripts/wheel/wheel.cgi, by Don Armstrong.

Helix 3 is not unique to *Hantavirus*es. The Crimean Congo Hemorrhagic Fever Virus (CCHFV) Gn zinc finger domain also contains a short helix in approximately the same position (**Fig. 6-5, panels B & C**). However, differences do exist. In the CCHFV Gn tail, the beginning of helix 3 contains three conserved basic residues (**Fig. 5-5**), whereas the Andes helix 3 begins with one, Arg^{599}, and ends with Lys^{605} and Lys^{606}. While the precise orientation of these basic residues relative to the rest of the molecule is unknown, they do appear to be distributed differently than the basic residues of the CCHFV helix 3.

Furthermore, the backbone dynamics of helix 3 in each structure appear to be slightly different. By comparing the $[^{1}H]$-^{15}N heteronuclear NOE intensity plots we can see that the Andes helix 3 is slightly more flexible than the corresponding helix 3 in the CCHFV (**Fig. 6-8**). This difference may be due to the fact that the CCHFV helix 3 is longer by five residues. The NMR derived structure of the CCHFV Gn tail (Chapter 5) shows helix 3 tucked underneath the two CCHC zinc fingers and held in place by a combination of hydrophobic contacts and a salt bridge. By comparison, the shorter Andes helix 3 may not be in position to form these same contacts and may therefore be, if only marginally, more flexible. Further, the fact that the core Andes domain is stable in the absence of helix 3 (Chapter 5), while the CCHFV Gn tail requires helix 3 to form a soluble construct (Chapter 5, **Figure 5-1**), also points toward a difference in the behavior of helix 3 in the two respective structures.

How these differences correlate to differences in the role played by the Gn tail in RNA binding or in a Gn-N protein interaction for each virus is still unclear. However, understanding

both the structure and the dynamics of the Andes Gn poly-basic arms provides valuable insight into what is likely an important intermolecular contact point during virus assembly.

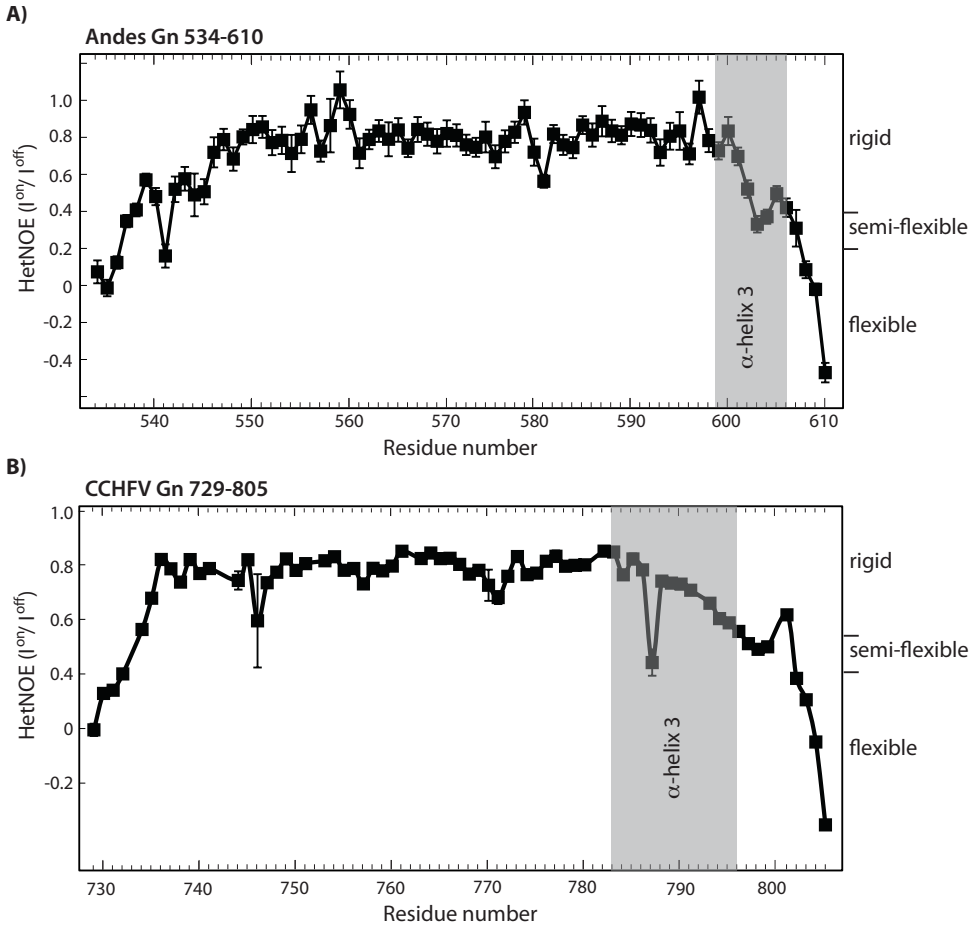

Figure 6-8. ^{15}N heteronuclear NOE intensity plots for the (A) Andes Gn 534-610 and (B) CCHFV Gn 729-805.

References

1. Hepojoki, J.M., *et al*., *The Cytoplasmic Tails of Hantavirus Glycoproteins Interact with the Nucleocapsid Protein.* J. Gen. Virol., 2010. **91**: p. 2341-50.
2. Overby, A.K., R.F. Pettersson, and E.P. Neve, *The glycoprotein cytoplasmic tail of Uukuniemi virus (Bunyaviridae) interacts with ribonucleoproteins and is critical for genome packaging.* J Virol, 2007. **81**(7): p. 3198-205.
3. Battisti, A.J., *et al*., *Structural studies of Hantaan virus.* J Virol, 2011. **85**(2): p. 835-41.
4. Shi, X., *et al*., *Role of the cytoplasmic tail domains of Bunyamwera orthobunyavirus glycoproteins Gn and Gc in virus assembly and morphogenesis.* J Virol, 2007. **81**(18): p. 10151-60.
5. Cole, C., J.D. Barber, and G.J. Barton, *The Jpred 3 secondary structure prediction server.* Nucleic Acids Res, 2008. **36**(Web Server issue): p. W197-201.

Chapter 7: RNA Binding Properties of Bunyavirus Gn Cytoplasmic Tails

Introduction

Given the pre and post budding positioning of the Gn cytoplasmic tails during the Bunyaviridae life cycle (**Fig. 1-2**), the Gn tails are candidates to play key roles in assembly events. Recent findings support this hypothesis [1-3]. Specifically, the core zinc finger domain of *Hantavirus* Gn protein has now been shown to bind the middle region of the nucleocapsid protein [1]. However, the RNA binding determinants of the nucleocapsid protein have also been mapped to the middle region of the protein [4]. This raises the possibility of a potential protein-RNA interaction between the cytoplasmic tails and the vRNA component of the ribonucleocapsid complex. However, to date there has been no published data to support this proposed interaction.

Zinc finger folds have traditionally been associated with nucleic acid binding [5]. However, an electropheretic mobility shift assay (EMSA) using the core Andes zinc finger domain, residues 544-599 (**Fig. 2-11**) demonstrated the absence of any kind of interaction with Vero cell RNA extract. Perhaps not surprisingly, the theoretical isoelectric point (pI) of this domain is low (5.7), and therefore not characteristic of an RNA binding protein. However, the theoretical pI of the Andes cytoplasmic tail increases considerably if we incorporate a larger sequence; residues 534-610 has a theoretical pI of 8.2 and the intact Gn tail, residues 520-650, would have a theoretical pI of approximately 8.7. This difference is largely due to the presence of conserved basic charges located in and around the dual CCHC motif within each of four zinc-finger containing Bunyaviridae families (**Fig. 7-1**). Interestingly, only the Phlebovirus genus,

which is also absent the conserved dual CCHC motif, is also apparently missing the same number of conserved basic sites on the Gn tail.

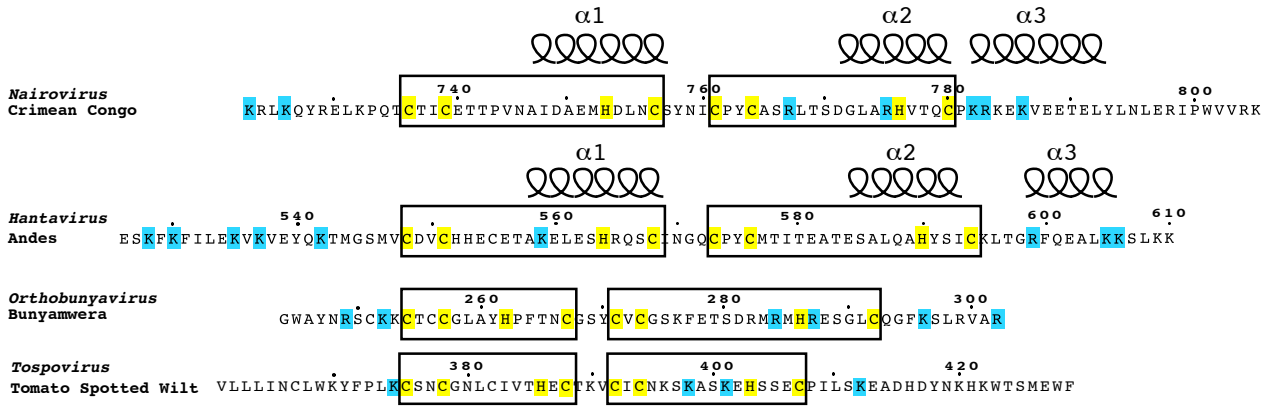

Figure 7-1. Sequences from representative members of Bunyaviridae families. Sequences are aligned along the dual CCHC zinc fingers. Conserved basic residues within each respective family are highlighted in blue.

In this study, we have used additional EMSA experiments of recombinantly expressed Gn tails in order to elucidate the RNA binding capabilities of these proteins. We demonstrate that the Gn tails of three members of the Bunyaviridae family appear to have the capability to interact weakly with RNA. We discuss the impact of these findings in regards to the role played by the Gn tails in virus assembly.

Results

Protein Production – The intact Gn tails of *Hantaviruses* are toxic to *E. coli* (Chapter 3). Therefore, with the assistance of Dr. Phillip Gao of the COBRE protein production group, we attempted to express the Andes Gn tail from a series of plasmids containing a variety of solubility enhancement tags (GB1, GST, MBP). None of these tags were able to bypass the

toxicity of the full length tail. However, we did produce two soluble constructs representing N-terminal MBP fusion proteins of regions of the Gn tail larger than the core zinc finger domain itself. MBP-Andes Gn 526-630 and MBP-Andes Gn 534-610 both expressed in good yield and soluble form. However, attempts to cleave MBP from the Gn tail constructs failed to achieve efficient removal of the tag. As a result, initial EMSA experiments were performed in the presence of the tag, with cleaved MBP as a negative control. After initial EMSA experiments, we managed to express and purify a cleavable N-terminal GB1 construct of residues 534-610 of the Andes Gn tail (described further in Chapter 6) that was also used to characterize binding to RNA.

In addition to RNA binding to *Hantavirus* cytoplasmic tails, we also report in this study a similar property in other Bunyaviruses; the CCHFV Gn 729-805 (described in detail in Chapter 5) and the Bunyamwera Gn 249-302. The Bunyamwera virus (genus *Orthobunyavirus*) expressed in low yield and appeared stable in 500 mM NaCl. However, upon dialysis into lower salt concentrations (30 mM NaCl) a significant amount of the prep (~90%) precipitated. The remaining portion was concentrated and used in this study as a GB1 fusion construct.

EMSA Experiments – Preliminary EMSA experiments were based on detecting an interaction between the Bunyavirus cytoplasmic tails and viral RNA extracted from Andes infected Vero E6 cells. However, although the harvested cells were two weeks post-infection and the presence of viral RNA verified by RT-PCR (Chapter 3), the predominant RNA bands visible on an agarose gel consisted of 18S and 28S ribosomal subunits. The viral RNA concentration was simply too weak to detect. Therefore, the early experiments actually detect

what is likely a non-specific interaction between the Andes virus cytoplasmic tail and Vero E6 ribosomal RNA (**Fig. 7-2**).

Chapter 3 describes an experiment in which the core Andes zinc finger domain failed to affect the migration of ribosomal RNA through an agarose matrix (**Fig. 7-2A**). In this study, a similar experiment was carried out using an N-terminal MBP fusion construct of the Andes Gn 534-628 (**Fig. 7-2B**). In contrast to the core domain, the extended cytoplasmic tail caused a markedly noticeable change in RNA migration. In the presence of large amounts of the protein, all of the RNA remained in the well rather than migrate. This early result hinted that not only did the Andes virus Gn cytoplasmic tail have the ability to bind RNA, but sequences outside of the core zinc finger domain were important for this observed interaction.

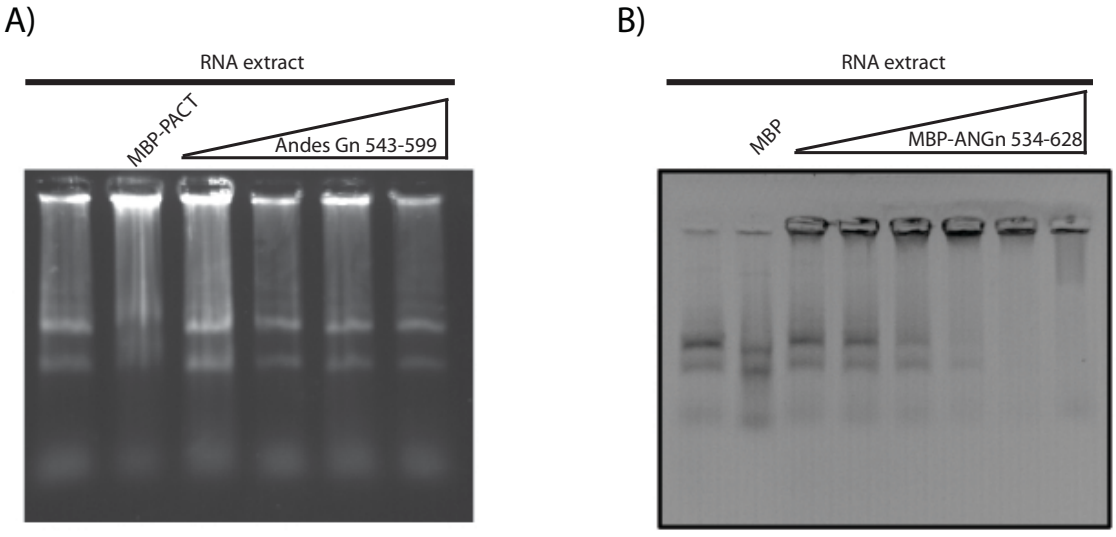

Figure 7-2. Agarose EMSA experiments for the (A) core Andes Gn zinc finger domain (534-599) and the (B) MBP fusion construct containing the extended Andes Gn (534-628). The perterbation of the ribosomal RNA bands in the presence of the exteneded construct indicated the extra flanking sequences were important for observed interaction.

Based on these early findings, other Bunyavirus cytoplasmic tails were also tested using the same agarose-based assay. The Bunyamwera Gn tail (genus *Orthobunyavirus*, residues 249-302, theoretical pI of ~9.4) and the CCHFV Gn tail (genus *Nairovirus*, residues 729-805, theoretical pI of ~7.8) also demonstrate the ability to affect the migration of ribosomal RNA through an agarose matrix (**Fig. 7-3 A &B**).

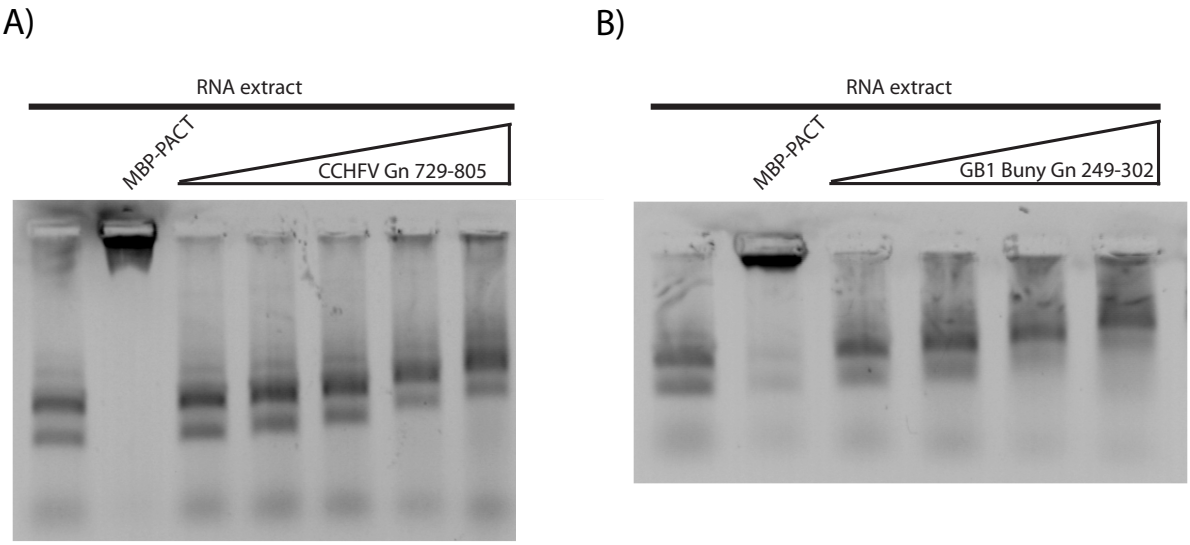

Figure 7-3. Agarose EMSA experiments for the A) CCHFV Gn 729-805 and B) GB1 fusion construct of Bunyamwera Gn 249-302. In both experiments, the decreased migration and enhanced smearing of the ribosomal RNA bands was directly correlated with the addition of increasing amounts of protein.

Since ribosomal RNA binding is most likely non-specific, it became necessary to use sequences representing vRNA. Consequently, we began transcribing the semi-complementary 3' and 5' ends of the viral RNA segments (as described in Chapter 2, *in vitro* transcription), otherwise known as the viral panhandle, in an attempt to detect a potential specific binding

interaction. Using much smaller RNA segments also allowed me to begin running EMSA experiments on a more conventional native poly-acrylamide matrix [6].

In order to further characterize the RNA binding properties of the Andes *Hantavirus*, we used a construct of the extended Andes Gn tail, residues 534-610 (described in detail in Chapter 6) to bind the Andes M segment panhandle (**Fig. 7-4A,** repeated in **Fig. 7-4B**). The disappearance of the free RNA band into the well in the presence of increasing amounts of Andes Gn 534-610 is suggestive of a protein-RNA interaction. In order to demonstrate that the effect is not attributable to RNA degradation, the mixture was also run under denaturing conditions (6 M urea) (**Fig. 7-4C**). The RNA signal is intact, suggesting the degradation of the panhandle was not an issue. Taken together with the previous set of experiments (**Fig. 7-2**), these data suggest the poly-basic sequences flanking the core Andes zinc finger domain are likely important for the binding interaction (the behavior of these poly-basic arms is described further in Chapter 6).

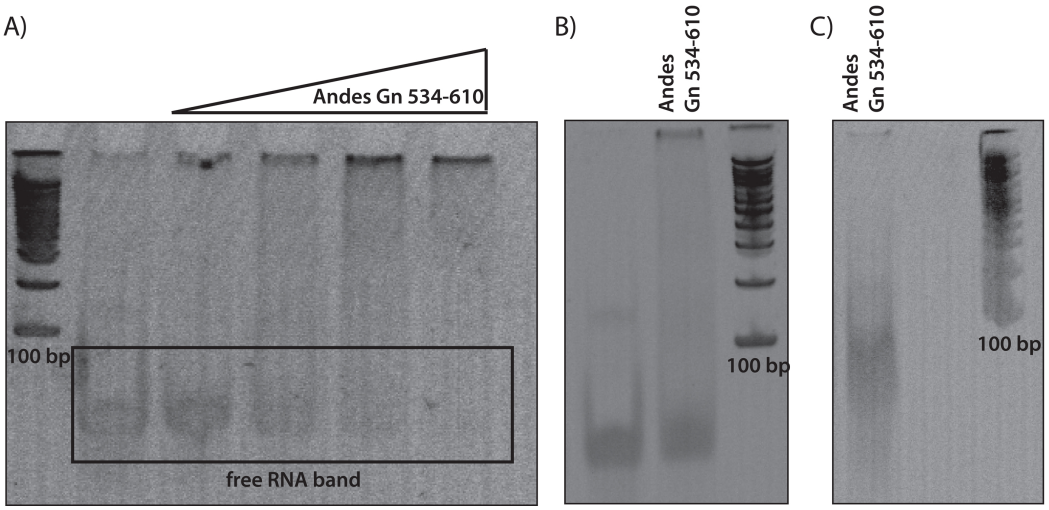

Figure 7-4. (A) EMSA depicting the binding of Andes Gn 534-610 to the Andes M segment RNA panhandle. Assays were run on a native 12% poly-acrylamde gel and gave reproducible results (B). Performing the same experiment under denaturing condictions (C) gave an intact RNA signal, thus indicating RNA degradation was not present in the native gel.

A similar acrylamide-based EMSA using the CCHFV and the CCHFV panhandle sequence (described in Chapter 2) gave a slightly different result (**Fig. 5-9A, 7-5A**). In this case, the free RNA appears as two bands, which given their migrations relative to the DNA ladder, likely represent the monomeric hairpin-like panhandle (~ 50 nt) and a tetrameric species (> 200 nt). In these experiments, addition of the CCHFV Gn cytoplasmic tail (729-805) causes the appearance of a new band at approximately 100 bp.

Previously, we described the surface of the CCHFV Gn 729-805 as having a large contiguous swath of conserved basic residues (**Fig. 5-8A**). In order to determine the role of these basic residues in the observable EMSA results, a series of three single mutants (K782D, R783D, K786D) and one double mutant (R767D/K786D) were generated, in which the focus of the mutations was to reverse the conserved positive charges located on the surface of the protein.

An acrylamide EMSA experiment using the mutant CCHFV Gn 729-805 (**Fig. 7-5A**) suggested that charge reversals at positions 767, 783, and 786 had little effect on the observed interaction, as these mutations all contained the shift band present with the wild-type protein. However, the mutation K782D did not show the same band and closely resembled the migration of free RNA in lane 1 (**Fig. 7-5A**), thus suggesting the positively charged side chain at position 782 is important for the interaction.

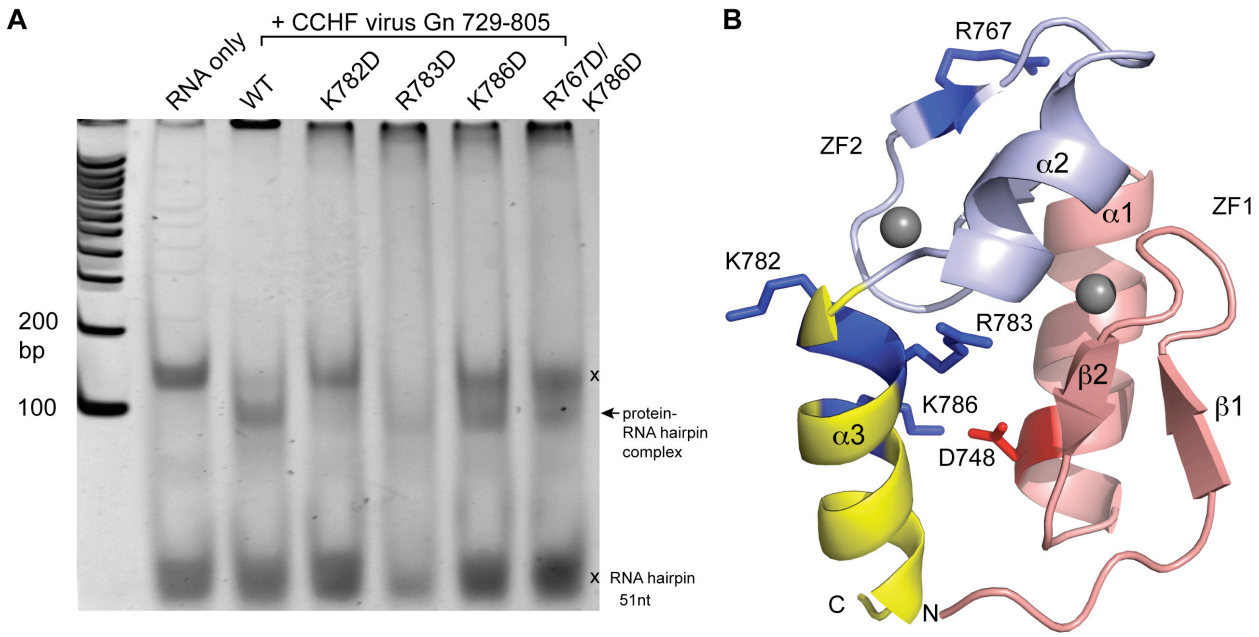

Figure 7-5. A) An poly-acrylamide based EMSA experiment for wild-type and mutant CCHFV Gn 729-805. Only the lane containing the K782D mutation resembled the mobility pattern observed for the RNA in the absence of protein. (B) A cartoon depicting the distribution of the conserved charges along with the salt bridge between K786 and D748.

Discussion

Prior to the work presented in this chapter, there had been no published data that provided evidence of a direct binding interaction between a Bunyavirus cytoplasmic tail and RNA. Given the orientation of the tails during the viral life cycle (**Fig. 1-2**), the proximity of the cytoplasmic domains to the ribonucleocapsid complex in the complete virus [3], as well as the number and distribution of conserved basic residues (**Fig. 7-1**), it is possible that such an interaction could play an important role in assembly events.

These preliminary EMSA studies demonstrate that such an interaction is possible between the cytoplasmic tails of three different families of Bunyaviridae and ribosomal RNA (**Fig. 7-2B & 7-3**), followed by interactions with small transcripts representing the viral

panhandle of the M segment for both the Andes and the CCHFV (**Fig . 7-4, 4-8A**). Lastly, mutagenesis of basic residues on the CCHFV Gn 729-805 (**Fig. 7-5A**) suggests that one conserved residue in particular, K782, may be important for the effect seen on a native poly-acrylamide gel (**Fig. 7-5A**).

The sequence of the RNA panhandle is highly conserved within each Bunyavirus genus. In *Hantaviruses*, the panhandles have been shown to associate specifically with the trimeric form of the nucleocapsid protein [7, 8]. Notably, there is no current model of *Hantavirus* assembly that would require an interaction between the Gn cytoplasmic tail and the panhandle. However, given the conserved nature of the panhandles and the enhanced resistance of dsRNA to ribonucleases, these hairpin-like sequences were chosen as candidate binding substrates for these initial RNA binding assays. The observed binding, therefore, is very likely not sequence specific. Despite this, the shift pattern observed with the CCHFV consisted of a well resolved band that traveled through the gel matrix with a different mobility than the other two bands (**Fig. 4-8A, 7-5A**). As described in Chapter 5, this new band likely represents a slower migration of monomeric RNA molecules in complex with protein, thus shifting the equilibrium away from the high molecular weight species.

It is unclear at this point what the binding determinant might be. However, it appears to be nearly completely disrupted by substituting a charge at K782 of helix 3 (**Fig. 7-5A**). K782 is located near the apex of the conserved helix-kink-helix structural motif (**Fig. 7-5B**). K782 is also located at the very center of the large patch of basic residues described in Chapter 5 (**Fig. 5-8A**) but also maintains a unique orientation distinct than that of the other basic residues in the

patch. This suggests that, while the surrounding basic residues may still compliment the interaction, therefore requiring conservation of these residues in different *Nairovirus*es, the key surface is likely to be the helix-kink-helix motif of the CCHFV cytoplasmic tail.

As a whole, there are key early observations than can be made regarding the RNA binding properties of Bunyaviridae Gn cytoplasmic tails. To begin with, it has become increasingly clear from this data that corresponding constructs of both the Andes (*Hantavirus*) and the CCHFV (*Nairovirus*) are required for RNA binding. Specifically, the inclusion of helix 3 (described in detail in Chapter 5 for the CCHFV and Chapter 6 for the Andes virus) is important in achieving the observed effect described in this study. Mutagenesis of this region in the CCHFV also suggests a lysine of the helix-kink-helix motif, composed of α2-Pro$_{781}$-α3, is somehow involved in the interaction (**Fig. 7-5A**). Interestingly, helix 3 varies between both viral species in not only length and flexibility (the Andes helix 3 is shorter and slightly more flexible, as described in Chapter 6) (**Fig. 7-1**), but also in the distribution of charged residues as well (**Fig. 7-1**).

Another key observation is that the apparent interaction also appears to be somewhat weak. For example, the amount of CCHFV Gn 729-805 required to generate these EMSA results approached a molar ratio of 1:50 (**Fig. 4-9**). Notably, Hepojoki *et al.* (2010) proposed a model of the *Hantavirus* spike complex formation in which Gn forms a homo-tetramer in which the cytoplasmic tails are in close proximity to one another [9]. While such a model has not been proposed for other Bunyaviruses, it does generate the possibility that a single cytoplasmic tail is insufficient for a tight or specific binding interaction. Furthermore, since there is mounting

evidence for a Gn tail/N protein interaction [1-3], it is also possible that the full RNA binding properties of the Gn tail are only evident when binding the intact ribonucleocapsid protein as a combined protein-protein and protein-RNA interaction.

References

1. Hepojoki, J., *et al*., *Cytoplasmic tails of hantavirus glycoproteins interact with the nucleocapsid protein.* J Gen Virol, 2010. **91**(Pt 9): p. 2341-50.
2. Overby, A.K., R.F. Pettersson, and E.P. Neve, *The glycoprotein cytoplasmic tail of Uukuniemi virus (Bunyaviridae) interacts with ribonucleoproteins and is critical for genome packaging.* J Virol, 2007. **81**(7): p. 3198-205.
3. Battisti, A.J., *et al*., *Structural studies of Hantaan virus.* J Virol, 2011. **85**(2): p. 835-41.
4. Severson, W., *et al*., *Essential amino acids of the hantaan virus N protein in its interaction with RNA.* J Virol, 2005. **79**(15): p. 10032-9.
5. Brown, R.S., *Zinc finger proteins: getting a grip on RNA.* Curr Opin Struct Biol, 2005. **15**(1): p. 94-8.
6. Gagnon, K.T. and E.S. Maxwell, *Electrophoretic mobility shift assay for characterizing RNA-protein interaction.* Methods Mol Biol, 2011. **703**: p. 275-91.
7. Mir, M.A. and A.T. Panganiban, *Trimeric hantavirus nucleocapsid protein binds specifically to the viral RNA panhandle.* J Virol, 2004. **78**(15): p. 8281-8.
8. Mir, M.A. and A.T. Panganiban, *The hantavirus nucleocapsid protein recognizes specific features of the viral RNA panhandle and is altered in conformation upon RNA binding.* J Virol, 2005. **79**(3): p. 1824-35.
9. Hepojoki, J., *et al*., *Interactions and oligomerization of hantavirus glycoproteins.* J Virol, 2010. **84**(1): p. 227-42.

Chapter 8: Summary of NMR Studies of the Bunyaviridae Glycoprotein Cytoplasmic Tails

Zinc fingers remain the most ubuiquitous structural motif in the eukaryotic genome. They perform a variety of functions ranging from mediating protein-nucleic acid or protein-protein interactions to serving as zinc ion sensors in cells [1, 2]. In retroviruses, zinc finger motifs are frequently located on the nucleocapsid protein and function to encapsidate viral RNA into a ribonucleocapsid complex [1, 3]. However, Bunyaviruses are a notable exception to this general rule, as their nucleocapsid proteins have not been reported to contain a zinc finger structure at all. Instead, we report in this dissertation that a zinc finger structure is located on the cytoplasmic tail of the Gn protein of the glycoprotein spike complex [4, 5]. Despite a mere 12% sequence identity between viruses, the structure is conserved in at least two Bunyavirus genera (*Hantavirus* and *Nairovirus*) and appears to be present in two more (*Orthobunyavirus* and *Tospovirus*). The general preservation of the fold in two diverse genera and the apparent conservation of the dual CCHC motif in four of the five Bunyavirus genera suggests the structure has an important, albeit general role in the viral life cycle.

The unusual location of the Bunyavirus-type zinc finger domain is made even more interesting by its unusual structure, which consists of a novel arrangement of back-to-back, classical ββα CCHC-type zinc fingers (Chapters 3 and 5). Significantly, the zinc binding topology is distinct from the overlapping topology of a RING domain, which had previously been speculated to form on the cytoplasmic tails [6]. Additionally, site mutagenesis of the conserved zinc binding residues confirmed that the two closely associated zinc fingers actually fold interdependently, thus further distinguishing the structure from that of other dual zinc

fingers such as LIM domains [1, 7]. Thus, the Bunyavirus-type zinc finger domain represents a novel arrangement of classical CCHC-type zinc fingers.

Many of the events leading to assembly of intact Bunyaviruses are still unclear. However, given the absence of any kind of matrix protein counterpart, as well as the pre-budding orientation of the Gn cytoplasmic tails in the cis-golgi (**Fig. 1-2**), these carboxyl sequences of glycoprotein are leading candidates for mediating an interaction between the viral spike complexes and the ribonucleocapsid complexes required to produce a pathogenic daughter virus. Recent studies in various Bunyavirus families support just such a role [8-10], thus making the importance of these tails increasingly evident. Notably, the Bunyavirus-type zinc finger domain makes up approximately one third of the overall length of *Hantavirus* Gn tails and nearly two thirds of the overall length of the *Nairovirus* Gn tail, thus intimating a potential role for this unique zinc finger fold in assembly events.

One early indicator of the function of the zinc finger domain is the distribution of conserved basic residues (**Fig. 1-3**). Three of the five genera (*Nairovirus*, *Orthobunyavirus*, and *Tospovirus*) contain a similar repeating pattern of arginines and lysines in which two basic residues overlap with zinc finger 2 (ZF2) in the linear sequence. A second grouping of conserved basic residues is located on the C-terminal side of ZF2 and a third set is located on the N-terminal side, thus potentially providing sets of flanking charges on either side of the zinc finger domain. *Hantavirus*es contain a similar pattern of flanking basic residues, differing only in that zinc finger 1 (ZF1) contains a conserved basic site instead of ZF2. While the Andes *Hantavirus* structure only contained one of these residues (Lys^{559}, which was surface exposed

near the apex of ZF1), the structure of the CCHFV Gn, residues 729-805, contains five of the seven conserved basic residues of the Gn tail (**Fig. 8-5B**). Of these, four are surface exposed (Arg^{767}, Arg^{775}, Lys^{782}, and Arg^{783}) and only one, Lys^{786}, forms a salt bridge with Asp^{784} of helix 1. These surface exposed basic residues form a conserved contiguous basic swath on the protein surface (**Fig. 5-8A**). Notably, this CCHFV Gn construct does not have a basic net charge. It actually contains as many acidic side chains (11 residues) as it does basic. According to PROTPARAM [11], the theoretical pI of the construct is predicted to be approximately 6.9.

Hence, the distribution pattern of charged residues in the Bunyavirus cytoplasmic tails is not random. Their placement in the linear sequence relative to the dual CCHC motif is largely conserved. Therefore, it is not unreasonable to hypothesize a likely role for these charges in the function of the cytoplasmic tail. The study published by Hepojoki et al., (2010) appears to suggest such an interaction in *Hantavirus*es, in which two of the three reported binding sites between the Gn tail and the nucleocapsid protein contain the same conserved basic residues outlined above [10]. While the poor solubility of constructs of the *Hantavirus* nucleocapsid protein prevented our examination of this protein-protein interaction, we were able to identify via electrophoretic mobility shift assays (EMSAs) an interaction between the Gn tail and the panhandle sequence of the viral RNA (Chapter 7). Perhaps not coincidentally, some of the same flanking sequences containing the conserved basic residues were also required for the observed protein-RNA interaction as were required in the protein-protein study [10], further reinforcing the notion that the distribution of basic residues in the Bunyavirus zinc finger domain is functionally significant.

An RNA binding motif consisting of a zinc finger fold flanked by conserved basic residues is not novel. The Moloney Murine Leukemia virus nucleocapsid protein, for example, contains a single CCHC-type zinc finger fold that is flanked by conserved lysine and arginine residues that participate in forming a complex with RNA [12, 13]. Therefore, given our RNA binding studies as well as the distribution pattern of conserved basic residues, it is not unreasonable to suggest that the Bunyavirus-type zinc finger domain forms part of an RNA binding motif.

In order to reconcile our results, which indicate the presence of an RNA binding motif, with that of the aforementioned study [10], which indicates a protein-protein binding role for the *Hantavirus* Gn tail, it's useful to examine the model proposed in the very same publication. In this model, the Hantaviral spike complex consists of a Gn tetramer that is further complexed with an outer ring of the Gc protein. A Gn tetramer would have four Gn cytoplasmic tails within a very small space. The model also depicts the cytoplasmic tails as independent domains (with the current data, it is not clear whether the intact cytoplasmic tails would complex together; the core zinc finger domains never dimerized in solution, however, the behavior of the intact Hantaviral tail is unknown). To these cytoplasmic domains is bound the assembled ribonucleocapsid complex consisting of repeating trimers of nucleocapsid protein bound to the viral RNA. In this crowded binding model, regardless of the organization of the four Gn tails, it is not difficult to imagine an interaction that is both multivalent and adaptable to either a protein-protein or protein-RNA interaction, conceivably with some Gn tails forming contacts with protein and others contacting RNA. Flanking poly-basic sequences that are flexible, as those

described in Chapter 6, would fit well with this model as they would allow for additional modularity in binding.

Perhaps another argument in favor of a conserved RNA binding motif is the observation that - for both the Andes virus as well as the CCHFV, RNA binding was only observed when using the same corresponding construct for each respective virus. For example, the core Andes virus Gn zinc finger domain did not bind RNA (**Fig. 3-11**), yet the extended form consisting of 76 residues (534-610), did bind RNA (**Fig. 7-4A**). Significantly, the CCHFV construct used for structure determination and consisting of residues 729-805 (also a 76 residue construct) also demonstrated the ability to bind RNA (**Fig. 5-9A**). The identical size of the two constructs is not a coincidence; both RNA binding constructs contain very similar features. Namely, both contain i) the conserved dual zinc finger fold, ii) a third helix located at the C-terminal side of the zinc fingers, and iii) three conserved basic residues located on helix 3. In the case of the CCHFV, one of these basic residues, Lys^{782}, was shown to be important for the RNA binding observed in the EMSA (**Fig. 7-5A**). The corresponding residue in the Andes virus would likely be Arg^{599}. While Arg^{599} was present in the structure of the core Andes zinc finger domain, it constituted the extreme carboxyl residue of the construct, thus was not in its native chemical environment. Clearly, mutagenesisis of Arg^{599} of the extended Andes Gn tail (534-610) is warranted to determine the importance of this position in binding RNA.

Future studies investigating the role of the Bunyavirus Gn cytoplasmic tials are warranted. Elucidating the specific protein-RNA or protein-protein contact sites between the cytoplasmic tails and their corresponding ribonucleocapsid complexes may lead to novel pharmacological targets designed to disrupt the Bunyavirus life cycle. Unfortunately, the

elusiveness of a reverse genetics system for *Hantavirus*es and the BSL-4 safety restrictions of working with the CCHFV often make this kind of research impractical. In spite of these limitations, the work presented in this study provides an important baseline in the search for novel therapeutic options.

References

1. Laity, J.H., B.M. Lee, and P.E. Wright, *Zinc finger proteins: new insights into structural and functional diversity.* Curr Opin Struct Biol, 2001. **11**(1): p. 39-46.
2. Gamsjaeger, R., et al., *Sticky fingers: zinc-fingers as protein-recognition motifs.* Trends Biochem Sci, 2007. **32**(2): p. 63-70.
3. De Guzman, R.N., et al., *Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element.* Science, 1998. **279**(5349): p. 384-8.
4. Estrada, D.F., et al., *The Hantavirus Glycoprotein G1 Tail Contains Dual CCHC-type Classical Zinc Fingers.* J Biol Chem, 2009. **284**(13): p. 8654-60.
5. Estrada, D.F. and R.N. De Guzman, *Structural characterization of the Crimean-Congo hemorrhagic fever virus Gn tail provides insight into virus assembly.* J Biol Chem, 2011.
6. Sen, N., A. Sen, and E.R. Mackow, *Degrons at the C terminus of the pathogenic but not the nonpathogenic hantavirus G1 tail direct proteasomal degradation.* J Virol, 2007. **81**(8): p. 4323-30.
7. Hammarstrom, A., et al., *Solution structure of a naturally-occurring zinc-peptide complex demonstrates that the N-terminal zinc-binding module of the Lasp-1 LIM domain is an independent folding unit.* Biochemistry, 1996. **35**(39): p. 12723-32.
8. Battisti, A.J., et al., *Structural studies of Hantaan virus.* J Virol, 2011. **85**(2): p. 835-41.
9. Overby, A.K., R.F. Pettersson, and E.P. Neve, *The glycoprotein cytoplasmic tail of Uukuniemi virus (Bunyaviridae) interacts with ribonucleoproteins and is critical for genome packaging.* J Virol, 2007. **81**(7): p. 3198-205.
10. Hepojoki, J., et al., *Cytoplasmic tails of hantavirus glycoproteins interact with the nucleocapsid protein.* J Gen Virol, 2010. **91**(Pt 9): p. 2341-50.
11. Gasteiger, E., et al., *ExPASy: The proteomics server for in-depth protein knowledge and analysis.* Nucleic Acids Res, 2003. **31**(13): p. 3784-8.
12. Dey, A., et al., *Composition and sequence-dependent binding of RNA to the nucleocapsid protein of Moloney murine leukemia virus.* Biochemistry, 2005. **44**(10): p. 3735-44.
13. Housset, V., et al., *Basic amino acids flanking the zinc finger of Moloney murine leukemia virus nucleocapsid protein NCp10 are critical for virus infectivity.* J Virol, 1993. **67**(5): p. 2537-45.