

Copyright  
by  
Keerthi Gottipati  
2012

**The Dissertation Committee for Keerthi Gottipati Certifies that this is the approved  
version of the following dissertation:**

**STRUCTURAL AND BIOCHEMICAL CHARACTERIZATION OF  
N-TERMINAL PROTEASE OF CLASSICAL SWINE FEVER VIRUS**

**Committee:**

---

Kyung H. Choi, Ph.D., Supervisor

---

Werner Braun, Ph. D.

---

Robert O. Fox, Ph.D.

---

Jóse M. Barral, M.D., Ph.D.

---

Nicolas Ruggli, D.V.M., Ph.D.

---

---

Dean, Graduate School

**STRUCTURAL AND BIOCHEMICAL CHARACTERIZATION OF  
N-TERMINAL PROTEASE OF CLASSICAL SWINE FEVER VIRUS**

**by**

**Keerthi Gottipati, M.S.**

**Dissertation**

Presented to the Faculty of the Graduate School of

The University of Texas Medical Branch

in Partial Fulfillment

of the Requirements

for the Degree of

**Doctor of Philosophy**

**The University of Texas Medical Branch**

**August, 2012**

## **Dedication**

To the three wonderful women who taught me, supported me, and nurtured me through the years with love, kindness and patience: my mother, my sister, and my late grandmother.

## **Acknowledgements**

The last four years have been one of the most rewarding phases of my life and I owe it in no small part to the unwavering support and guidance of my mentor Dr. Kay Choi. She is one of the few people whose clarity of thought, patience, and optimism have been a true inspiration to me over the years, qualities that I hope to emulate in my own life and career. I would like to thank her for believing in me, for teaching me and giving me the opportunity and freedom to explore science in my own imperfect and often rather inelegant ways. Apart from my mentor, I've also had the pleasure of interacting with an amazing individual, Dr. Marc Morais who always encouraged me to aim high and 'think outside the box'. For the numerous lessons in the science and application of macromolecular crystallography, I am indebted to Dr. Mark White. His attention to detail, and his ever-willingness to engage me in discussion however trivial made learning a true joy. I would also like to thank Mengyi Ye, for not only training me in bench work but also for her companionship and the love and kindness that make her such a wonderful person to work with.

In the course of the six years that I have been associated with the department of Biochemistry and Molecular Biology, one individual bore all my shenanigans with great panache and patience. I cannot thank Ms. Debora Botting enough for all that she has done for me, and for her wise words and kind deeds. I would also like to thank Shirley Broz,

Angelina Johnson and Lisa Pippet for all their help and cheerful tête-à-têtes over the years. They have been the people behind the curtain who always ensured that the show went on. I would like to thank Dr. James Lee and Dr. Rodrigo Maillard for lessons in life and biochemistry; for laying the foundation during my early years in the graduate program that would later help me realize my goals. I also want to thank Dr. Jason Vertrees who taught me everything I know about Pymol and beer.

Few indeed, have been blessed with such friends as I have been. I would like to thank Jeeba Kuriakose, Larry Bellot and Aditya Hindupur for being there for me through the ups and downs of grad school and for having my back all these years. I am a better person for having known them and I will always cherish the memories. For keeping my sanity in check and for the quick laughs and great conversations, I want to thank Varsha Shridhar and Pradeep Kota, who have remained steadfast friends since my college days. I want to thank Gavri Fernandez for the candid discussions and practical advice; He taught me the difference between the right decision and the popular one. I am grateful for the love and friendship of Angirasa Acharya; he has been a constant companion and really just a beautiful person.

Above all, I am deeply indebted to my family, especially my parents who never stopped encouraging me even when I veered off the beaten path. Unlike most parents, they didn't just school me but gave me an education, for which I am ever so grateful. I am also thankful for a brother who always watched over me, loved and unquestioningly supported me; for a sister too, who has never failed to put a smile on my face, someone I admire and love fiercely and whose passion for life and single-mindedness never ceases to amaze me. It was the love of my family, the support of my friends and the innumerable small yet significant contributions of numerous others that has enabled all that I have accomplished thus far and I sincerely believe that my journey has only just begun.

# **Structural and Biochemical Characterization of N-terminal Protease of Classical Swine Fever Virus**

Publication No. \_\_\_\_\_

Keerthi Gottipati, Ph.D.

The University of Texas Medical Branch, 2012

Supervisor: Kyung H. Choi

Pestiviruses are single-strand positive sense RNA viruses that belong to the flaviviridae family. They express their genome as a single polypeptide that is subsequently cleaved into individual proteins by host- and virus-encoded proteases. N-terminal protease N<sup>pro</sup> is the first translated protein of the pestivirus genome. N<sup>pro</sup> is a cysteine auto-protease that cleaves between its own C-terminus (Cys168) and the N-terminus of the core protein (Ser169). N<sup>pro</sup> is essential for initiation of viral polyprotein processing and generation of a viable core protein for viral assembly. The predicted catalytic triad of N<sup>pro</sup> is Glu22, His49 and Cys69, which differs from the known catalytic triads in either serine or cysteine proteases. Due to its unique sequence and catalytic site, N<sup>pro</sup> forms its own cysteine protease family C53. After the initial self-cleavage, N<sup>pro</sup> is no longer active as a protease and no *trans* activity has ever been observed. N<sup>pro</sup> also plays a critical role in subverting the host's innate immune response by targeting interferon regulatory factor-3 (IRF3) for proteasomal degradation, and thus preventing the transcriptional activation of interferon- $\alpha/\beta$  genes.

We determined the crystal structure of wild type N<sup>pro</sup> and that of a cleavage site mutant (N<sup>pro</sup>-C168A) to 1.6 Å resolution. The structures show that N<sup>pro</sup> is structurally

distinct from other known cysteine proteases and has a novel “clam shell” fold consisting of two domains, the protease domain carrying the active site of N<sup>pro</sup> and the zinc-binding domain that harbors a conserved metal binding sequence motif viz., the TRASH motif formed by Cys112-Cys134-Asp136-Cys138. The structure conclusively shows that the catalytic site of N<sup>pro</sup> is a dyad formed by His49 and Cys69 and does not include Glu22. Rather, Glu22 has a key role in maintaining the structural integrity of the protein. The C-terminus is not only bound in the active site, but it also contributes a strand to the beta-sheet that makes up the active site. Thus, this structure explains the autocatalytic, and subsequent auto-inhibition mechanism of N<sup>pro</sup> (i.e., why there is no *trans* activity), and furthermore provides insight into the interaction of N<sup>pro</sup> with IRF3 and its role in subversion of host immune response.

## TABLE OF CONTENTS

List of Tables .....	xii
List of Figures .....	xiii
List of Illustrations .....	xv
<b>CHAPTER-1:</b>	<b>1</b>
Introduction to Pestiviral N-terminal Protease (N <sup>pro</sup> ) .....	1
1.1 Pestiviruses .....	1
1.2 Prevention and Control of Pestiviral Diseases .....	3
1.3 Molecular Biology of Pestiviruses .....	5
1.4 Pestivirus Genome .....	7
1.5 Pestivirus N-terminal Protease (N <sup>pro</sup> ) .....	11
1.6 Specific Aims of Current Study .....	15
1.6.1 Aim 1: To determine the molecular structure of N-terminal protease of CSFV .....	16
1.6.2 Aim 2: To determine the molecular mechanism of auto-catalysis and auto-inhibition of CSFV N <sup>pro</sup> .....	16
1.6.3 Aim 3: To determine the molecular basis of interaction of CSFV N <sup>pro</sup> with porcine IRF3 .....	17
1.7 Significance of Current Study .....	17
<b>CHAPTER-2</b>	<b>19</b>
Molecular Structure of Wild Type N <sup>pro</sup> of CSFV .....	19
2.1 Introduction .....	19
2.2 Experimental Design .....	21
2.2.1 Expression and purification of wild type N <sup>pro</sup> .....	22
2.2.2 Crystallization trials with full length CSFV N <sup>pro</sup> .....	24
2.2.3 Crystallization trials with individual domains of CSFV N <sup>pro</sup> .....	26
2.2.4 Crystallization of N-terminal deletion mutant of N <sup>pro</sup> .....	29

2.2.5 Data collection and structure determination of Wild type N <sup>pro</sup> using sulfur-SAD .....	31
2.3 Results .....	32
2.3.1 Overall structure of wild-type N <sup>pro</sup> .....	32
2.3.2 Cysteine protease of N <sup>pro</sup> has a catalytic dyad .....	35
2.2.2 Zinc-binding domain of N <sup>pro</sup> .....	38
<b>CHAPTER-3</b>	<b>41</b>
Crystal Structure of N <sup>pro</sup> -Cys168Ala .....	41
3.1 Introduction .....	41
3.2 Experimental Design .....	42
3.3 Materials and Methods .....	44
3.3.1 Plasmid constructs .....	44
3.3.2 Overexpression and purification of N <sup>pro</sup> -Cys168Ala .....	44
3.3.3 Crystallization of N <sup>pro</sup> -C168A .....	44
3.3.4 Data collection .....	45
3.3.5 Structure determination .....	45
3.4 Results .....	48
3.4.1 Overall structure of N <sup>pro</sup> -C168A .....	48
3.4.2 Cysteine protease of N <sup>pro</sup> .....	51
3.4.3 Substrate binding site of N <sup>pro</sup> .....	56
3.4.4 N <sup>pro</sup> is a unique viral cysteine protease .....	62
3.4.5 Mechanism of intra-molecular inhibition of cysteine protease of N <sup>pro</sup> .....	64
<b>CHAPTER-4</b>	<b>68</b>
Catalytic Activity of the Cysteine Protease of N <sup>pro</sup> .....	68
4.1 Introduction .....	68
4.2 <i>Trans</i> Activity of N <sup>pro</sup> .....	68
4.2.1 Experimental design .....	68
4.2.2 Results .....	70
4.3 <i>In vitro</i> Proteolytic Activity of N <sup>pro</sup> .....	71

4.3.1 Materials and methods .....	73
4.3.2 Results.....	75
<b>CHAPTER-5</b>	<b>84</b>
Interaction of N <sup>pro</sup> with Interferon Regulatory Factor-3 .....	84
5.1 Introduction.....	84
5.1.1 Map of proposed IRF3 binding interface on N <sup>pro</sup> .....	85
5.2 Experimental Design.....	87
5.2.1 Expression and purification of soluble recombinant porcine IRF3 .....	88
5.2.2 Purification of full length recombinant porcine IRF3 from <i>E. coli</i> inclusion bodies .....	91
5.3 Future Directions .....	92
<b>CHAPTER-6</b>	<b>94</b>
Summary.....	94
Appendix A.....	96
Bibliography .....	97
Vita .....	110

## List of Tables

Table 2.1:	Data collection, phasing and refinement statistics of N <sup>pro</sup> -Δ17N. ....	34
Table 3.1:	List of designed catalytic and cleavage site mutants of cysteine protease of N <sup>pro</sup> .....	43
Table 3.2:	Scaling and refinement statistics for N <sup>pro</sup> -C168A native and single wavelength anomalous dispersion (SAD) data on zinc. ....	50

## List of Figures

Figure 1.1:	Clinical signs of BVDV and CSFV.....	4
Figure 1.2:	World-wide distribution of BVDV and CSFV.....	6
Figure 1.3:	Model of flavivirus virion. ....	8
Figure 1.4:	Lifecycle of flaviviruses.....	10
Figure 1.5:	Constituents of pestiviral polyprotein. ....	11
Figure 1.6:	Polyprotein composition of <i>Flaviviridae</i> . ....	12
Figure 1.7:	Sequence alignment of pestivirus N <sup>pro</sup> . ....	14
Figure 1.8:	IFN- $\alpha/\beta$ mediated innate immune mechanism. ....	15
Figure 2.1:	Purification of full-length wild-type CSFV N <sup>pro</sup> . ....	23
Figure 2.2:	Lysine methylation. ....	26
Figure 2.3:	Limited proteolysis of CSFV N <sup>pro</sup> . ....	28
Figure 2.4:	Crystallization of N <sup>pro</sup> - $\Delta$ 17N. ....	33
Figure 2.5:	Structure of wild-type N <sup>pro</sup> (N <sup>pro</sup> - $\Delta$ 17N). ....	36
Figure 2.6:	The protease domain of N <sup>pro</sup> . ....	37
Figure 2.7:	Zinc-binding domain of N <sup>pro</sup> . ....	40
Figure 3.1:	Diffraction from N <sup>pro</sup> -C168A crystals. ....	46
Figure 3.2:	Stereo image of structure of N <sup>pro</sup> -C168A. ....	49
Figure 3.3:	Crystal packing and zinc coordination by N <sup>pro</sup> -C168A. ....	52
Figure 3.4:	Active site of cysteine protease of N <sup>pro</sup> compared to that of papain. ....	57
Figure 3.5:	Substrate binding site of papain. ....	58
Figure 3.6:	Substrate binding site of N <sup>pro</sup> . ....	60
Figure 3.7:	Intramolecular product inhibition of <i>trans</i> activity in viral auto-proteases. ....	67

Figure 4.1: <i>In vitro</i> protease activity of N <sup>pro</sup> .....	79
Figure 4.2: Catalytic domain construct of N <sup>pro</sup> .....	82
Figure 4.3: Catalytic domain of N <sup>pro</sup> .....	83
Figure 5.1: Residues of N <sup>pro</sup> involved in the subversion of interferon response. ....	87
Figure 5.2: Multiple sequence alignment of IRF3. ....	89
Figure 5.3: Purification of IRF3 (N-terminal 6xHis-tag).....	90
Figure 5.4: Purification of recombinant IRF3 under denaturing conditions.....	93

## **List of Illustrations**

Illustration 3.1:	Mechanism of action of cysteine proteinase.....	54
-------------------	---	----

## CHAPTER-1:

### Introduction to Pestiviral N-terminal Protease (N<sup>pro</sup>)

#### 1.1 PESTIVIRUSES

Pestiviruses are small RNA viruses belonging to the *Flaviviridae* family that cause highly contagious diseases in livestock such as pigs, cattle, sheep and in their wild counterparts such as wild-boar, deer etc. [1]. Members of this genus include classical swine fever virus (CSFV) that infects pigs and wild boar, bovine viral diarrhoea virus (BVDV) seen in cattle, and border disease virus (BDV) in sheep, although there have been numerous incidences of cross-infection in hosts of these viruses [2–4]. Classical swine fever (formerly called hog cholera) was first described in the 1800s and BVD along with the associated mucosal disease (MD) have been studied since the 1940s [5–7]. Pestivirus diseases are economically debilitating to the dairy and meat industry as most countries follow a ‘stamp out’ strategy to deal with the disease outbreak leading to extensive culling of infected and non-infected animals [6,8,9].

All pestiviruses are highly transmissible and transmission can take place through direct contact between animals or with their bodily fluids, and through contaminated surroundings such as animal feed or instruments. Most importantly the viruses can undergo trans-placental transmission to the progeny [10–13]. The severity of disease and symptoms vary widely depending on the age of the animal, the virus biotype, strain virulence, and super-infections. These viruses, based on laboratory observations, are divided into cytopathic (cp) or non-cytopathic (ncp) forms depending on whether the virus can cause cell lysis or not, respectively. There is no direct correlation between the

cytopathogenicity and virulence of these viruses. Classical swine fever (CSF) is characterized by anorexia, severe depression, diarrhoea, high fever, conjunctivitis, superficial and internal hemorrhages that often lead to death in the acute and sub-acute forms. Symptoms are milder with prolonged and intermittent disease periods in the chronic form [6,11]. In case of an intrauterine infection the disease could lead to stillbirth or mummification of fetus and alternately if the virus infects a pregnant sow in its first trimester it could give rise to persistently infected animals that show stunted growth and shed the virus throughout their life without showing apparent symptoms of the disease. However, these persistently infected animals eventually die of CSF in a majority of the cases [14].

BVDV differs from CSFV in that majority of infected animals show none to mild clinical signs of the disease with rapid recovery. Hence the animals can remain seropositive for BVDV antibodies without any noticeable clinical disease. In other instances, the symptoms could range from none to high fever with respiratory infection, lethargy, depression, oral ulcers, diarrhea and in some cases death (Figure-1.1). BVDV infection during pregnancy can result in abortion, still birth, emergence of persistently infected seronegative animals depending on the stage of pregnancy at the time of infection [5,10,14,15]. Persistently (continuously) infected animals are the result of the BVDV infection during pregnancy (between 40 and 120 days of gestation) of an antibody-negative mother. The virus is transferred to the fetus whose developing immune system recognizes the virus as part of self, therefore replicating the virus throughout its life. PI animals continuously shed virus in large quantities infecting other healthy animals in the herd but appear normal making it hard to detect and eradicate them. Therefore, epidemiologically (for the spread of the virus) PI animals are extremely important. PI

animals can also encounter mucosal disease. Mucosal disease which is characterized by severe bloody diarrhea and mucosal erosion is caused when ncpBVDV mutates into its cytopathic form or due to the super-infection of a cpBVDV strain on a persistently infected animal [16–18]. In PI animals typically after 2-3 years of infection the already existing ncpBVDV strain mutates suddenly to a cp biotype and results in mucosal disease. BDV shares similar disease characteristics as bovine viral diarrhea virus including the incidence of mucosal disease following cpBDV infection. BDV is also transmitted trans-placentally leading to persistent infection in sheep [7,19].

## **1.2 PREVENTION AND CONTROL OF PESTIVIRAL DISEASES**

The World Organization for Animal Health (O.I.E.) classifies CSF as notifiable since a suspected outbreak can have significant impact on international trade of animal meat products. Increased oversight and vaccination strategies have led to the eradication of CSF in domestic pigs in North America and most of western and central Europe [20]. However, there is a high risk of re-introduction of the virus in these regions from viruses that survive in the wild [11,21]. CSF is also highly prevalent in Asia, Latin America, East Europe and former USSR area. In countries where classical swine fever is endemic, vaccines may be used to protect animals. Although effective live attenuated vaccines are available for CSFV, the inability to discriminate antibodies induced by attenuated vaccines from those elicited following a natural infection has necessitated generation of emergency marker vaccines and other fast-responding anti-viral strategies [22,23]. In case of BVD/MD due to lack of extensive surveillance, almost every country in the world is positive, except those who implemented successful eradication programs (eg. Sweden).

However, the real challenge for BVDV is the detection and eradication of persistently infected animals (those that indeed can eventually make MD).



Figure 1.1: Clinical signs of BVDV and CSFV.

(a) Calf persistently infected with BVDV showing depression and general lethargy. (b) Pig with CSFV showing cyanotic (purplish discoloration) lesions on the skin.

Strict control measures, similar to the ones used in control of CSF outbreaks have been implemented in many countries in Europe for the control and spread of BVDV infections in cattle [8,24]. The ease of homologous recombination of the virus with the host genome, the emergence of new strains along with the limited effectiveness of vaccines, the need for regular booster doses and presence of seemingly healthy persistently infected animals have made it difficult to eradicate the disease [8,13,25]. There is no cure for either CSF or BVD disease, and treatment only alleviates complications from symptoms. Despite continued efforts to control these diseases, outbreaks occur intermittently worldwide. Figure-1.2 shows the distribution of CSF and BVD as reported by the O.I.E during the second half of the year 2011. Any reported outbreaks are dealt with by isolating the herds, slaughtering infected animals and preventive culling of other animals in the herd.

### **1.3 MOLECULAR BIOLOGY OF PESTIVIRUSES**

Pestiviruses belong to the *Flaviviridae* family of RNA viruses which includes important human pathogens such as West Nile virus, Japanese encephalitis virus, dengue virus etc. (flavivirus genus) and hepatitis C virus (hepacivirus genus). Most of the current knowledge on *Flaviviridae* comes from the study of flaviviruses. For pestiviruses and HCV, little is known about the virion structure and assembly. These viruses typically have virions with a diameter of ~50nm [26,27]. Mature virions are enveloped, while the envelope composition varies between the three genera. In pestiviruses, the envelope comprises the three proteins E<sup>ms</sup>, E1 and E2 while the other two genera only have the E1 and E2 glycoproteins arranged in an icosahedral symmetry. In flaviviruses the genome is encapsidated by the capsid protein forming the nucleo-capsid inside the virion [2,28,29] (Figure-1.3). Recent studies however suggested that the pestivirus core (C) protein has a

disordered structure [30] and its role in RNA packaging and virion morphogenesis remains to be clearly ascertained [31]

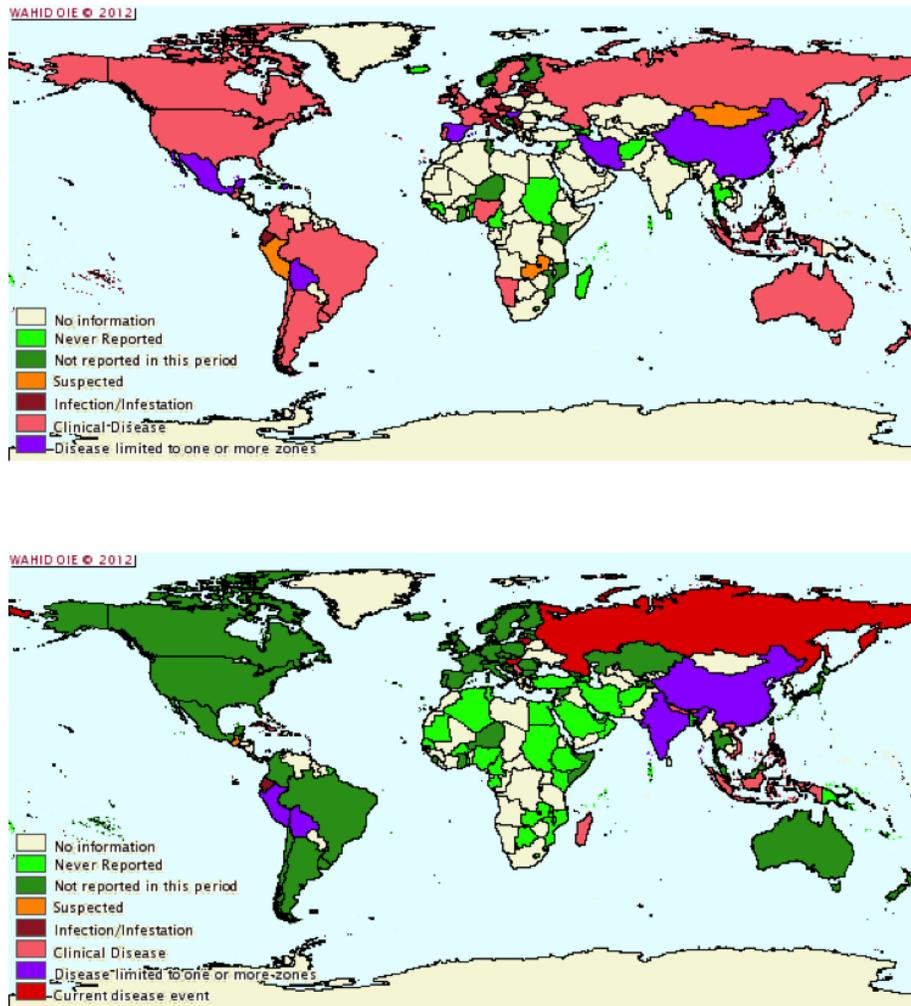


Figure 1.2: World-wide distribution of BVDV and CSFV.

World-wide reported cases of BVDV (a) and CSFV (b) in the period between July and December 2011 as reported by O.I.E. Clinical cases of BVDV in cattle have been reported in the US and Western Europe in this period. Although no new CSFV cases have been reported in the last two years in the US and Western Europe, there is a constant threat of an outbreak and emergence of new viral subspecies in these areas.

The typical lifecycle of viruses in the *Flaviviridae* family begins with the attachment of the virus to the host cell receptors following which the virus is endocytosed into the host cell. Facilitated by the acidic endosomal pH, viral membrane fuses with the endosomal membrane, which in turn causes release of the viral genomic RNA into the host cell cytoplasm. Since the genome of pestiviruses is a positive sense ssRNA, it acts as the viral messenger RNA. Upon release into the cytoplasm the ssRNA is translated by the host ribosomes into a single polyprotein. The polyprotein is processed co- and post-translationally into its constituent structural and non-structural proteins by both viral and host-cell proteases. The viral genome is then copied into a negative-sense complementary ssRNA by the viral RNA-dependent RNA polymerase (RdRp) via dsRNA intermediate. The new negative-sense ssRNA acts as a template for the generation of new genomic positive sense ssRNA. Following translation and replication, the genome is encapsidated by the core protein and virion is assembled by the structural proteins. Once progeny viruses are assembled at the endoplasmic reticulum, virions are transported to the cell membrane via the trans-ER-golgi network and finally exocytosed (Figure-1.4) [1,30].

#### **1.4 PESTIVIRUS GENOME**

Pestivirus genome is a continuous un-segmented (monopartite) single-strand positive sense RNA. The genome is approximately 12.5kb long. Unlike the genome of flaviviruses, pestivirus genome (and that of hepaciviruses) is not capped at the 5' terminus and not polyadenylated at the 3' terminus. Instead, the 3' UTR contains a short poly-(C) tract which helps in replication initiation whereas the 5' UTR carries the Internal Ribosome Entry Site or IRES required for ribosomal binding and initiation of genome translation [29,31,32].

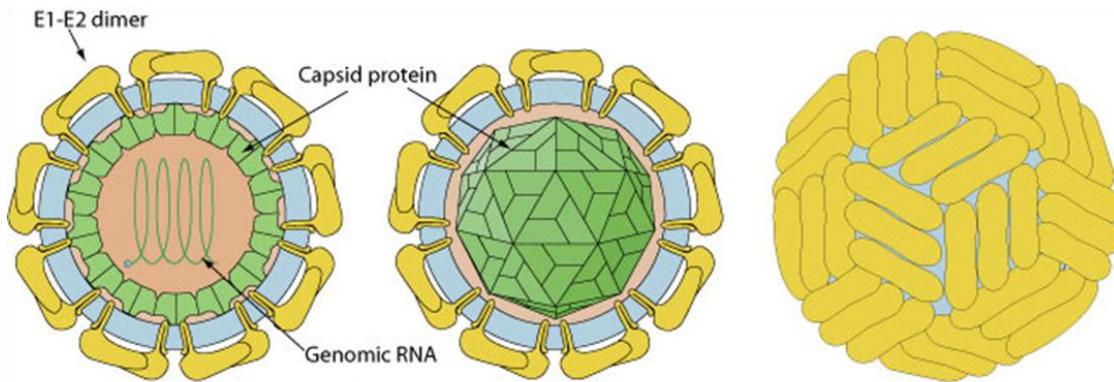


Figure 1.3: Model of flavivirus virion.

Virion model showing the packaging of viral RNA inside a nucleocapsid shell. The capsid protein (green) is surrounded by the envelope proteins E1 and E2 (and  $E^{ms}$  in case of pestiviruses) (yellow). On the surface of the virus the envelope proteins are arranged in an icosahedral symmetry. (Source: ViralZone [www.expasy.org/viralzone](http://www.expasy.org/viralzone), Swiss Institute of Bioinformatics)<sup>1</sup>

Pestiviral genome behaves like a viral messenger RNA molecule and following release into the cytosol it is immediately translated by the host ribosomes as a single polyprotein ( $N^{pro}$ -C- $E^{ms}$ -E1-E2-P7-NS2-NS3-NS4A-NS4B-NS5A-NS5B). The polyprotein is then processed into individual proteins by cellular and viral proteases [29], [32], [35–39] (Figure-1.5). Structural proteins which include the core protein C and envelope glycoproteins  $E^{ms}$ , E1, and E2, are involved in genome packaging and virus assembly [40]. Nonstructural (NS) proteins, in the order they are translated, are  $N^{pro}$ , p7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B. Functions of several NS proteins are known. The HCV protein p7 has been shown to form ion channels in lipid membranes; its role in

<sup>1</sup> Open Source.

pestiviruses remains undetermined although it has been shown to be essential for the generation of infective progeny virus [41,42]. NS2 is a cysteine dependent autoprotease that cleaves at the junction of NS2-3 [43,44]. NS3 is a viral helicase and protease that is involved in the polyprotein processing to generate many of the individual non-structural proteins following translation [40,41]. The protease activity of NS3 requires NS4A as a cofactor. NS5B is the viral RdRp that replicates the pestivirus genome [47]. NS5A also has been shown to be a part of the replication machinery and a binding partner for several cellular proteins [43,44]. Non-structural proteins together with the viral RNA and other cellular proteins form a membrane associated replication complex to carry out viral replication [30,45]. N<sup>pro</sup>, the first translated protein of the pestiviral genome is the only non-structural protein not associated with the viral replication machinery and as such is dispensable for pestiviral replication. It is an auto-protease that cleaves its own C-terminus from the rest of the polyprotein co-translationally and consequently releases the polyprotein for further processing by viral and host proteases [34,46].

Pestiviral structural proteins involved in virion assembly and non-structural proteins involved in genome replication have counterparts with analogous functions in the other two genera of the flaviviridae family (viz., flaviviruses and hepaciviruses). However, a comparison of the genome products shows that pestiviruses encode for two additional proteins; the structural protein E<sup>ms</sup>, found both as membrane associated envelope protein and as secreted form, and the N-terminal protease (N<sup>pro</sup>), a cis-acting leader protease (Figure-1.6)

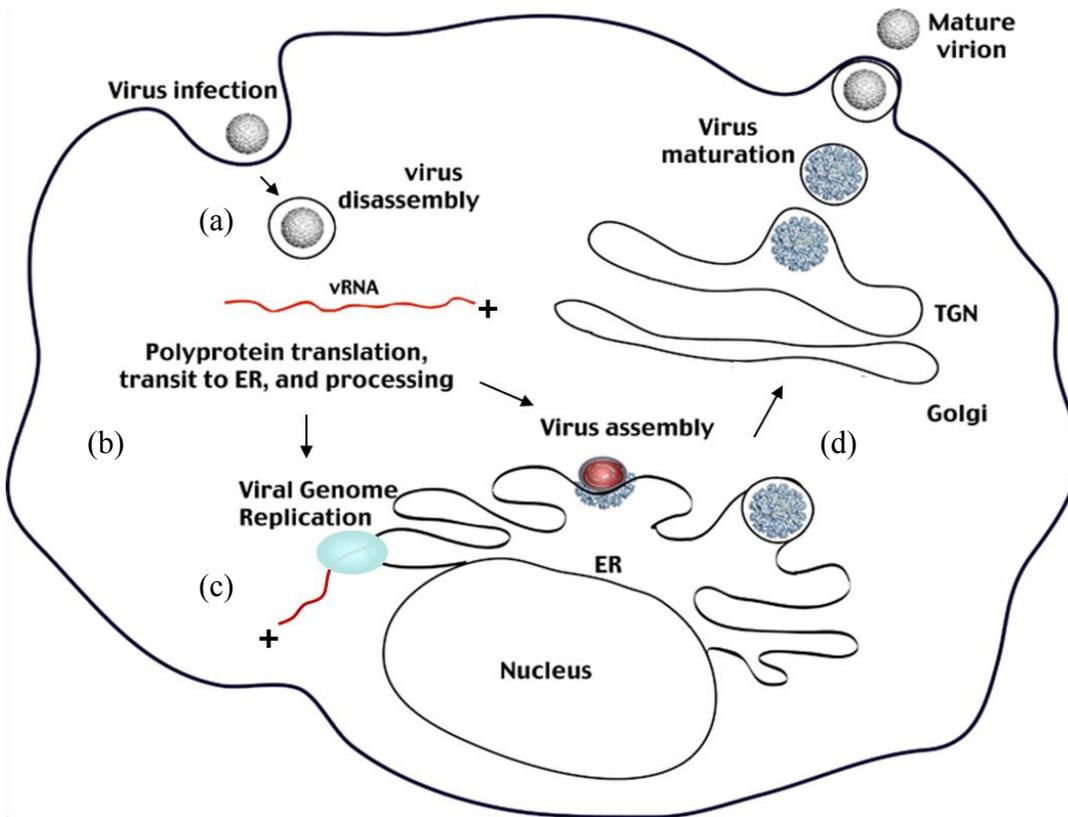


Figure 1.4: Lifecycle of flaviviruses.

(a) Schematic showing viral entry through receptor mediated endocytosis. Genomic RNA (viral +sense ssRNA) is released into the cytoplasm following membrane fusion and viral disassembly in the endosome. (b) Translation of the viral polyprotein takes place on the rough endoplasmic reticulum (rER). Viral mRNA is translated into a single polyprotein, which is then processed into its constituent proteins by viral and host proteases. (c) Following translation and before assembly, viral RNA is replicated via the formation of intermediate dsRNA. (d) Following replication, viral genome is packaged into new viral particles at membrane bound organelles. Mature virions are transported to the plasma membrane through the trans-golgi network following which virus is exocytosed and released outside the cell.

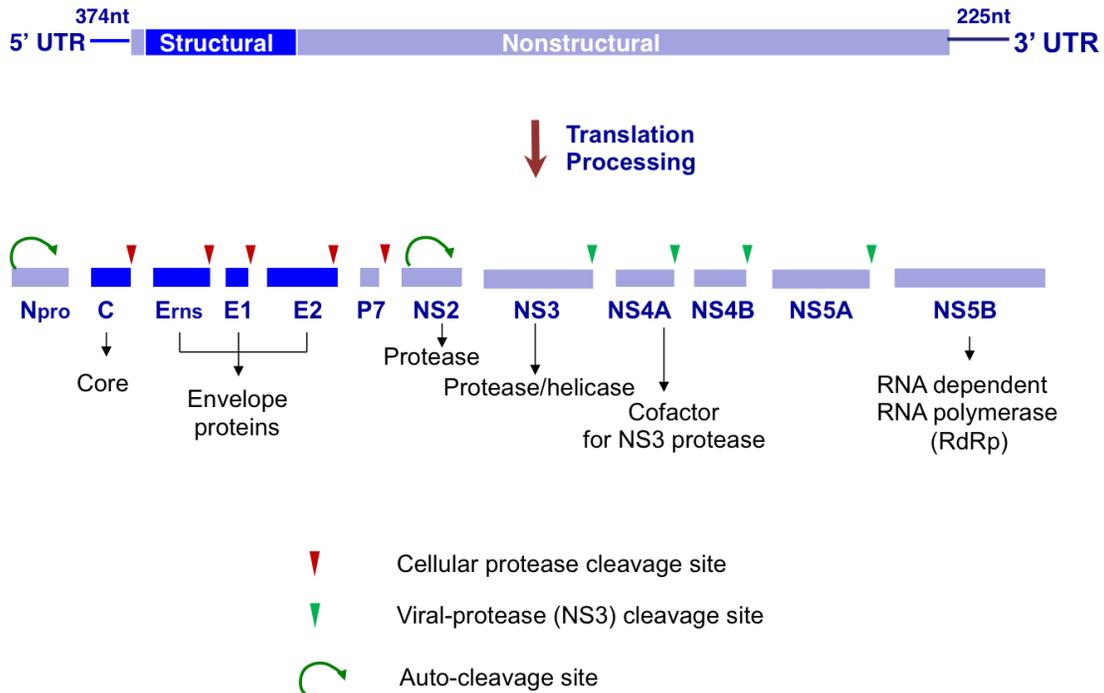


Figure 1.5: Constituents of pestiviral polyprotein.

Genomic +ssRNA (top) is translated into a single polyprotein (bottom). Translation begins at the 5' IRES. Polyprotein is processed into constituent structural and non-structural proteins by viral (green arrows) and host proteases (red arrows). Core C, E<sup>rns</sup>, E1 and E2 constitute structural proteins. N<sup>pro</sup>, NS2, NS3 (viral proteases), NS4A (cofactor for NS3) and the replication complex proteins viz., NS4B (transmembrane protein), NS5A (involved in viral replication), NS5B (RdRp) constitute the viral non-structural proteins.

### 1.5 PESTIVIRUS N-TERMINAL PROTEASE (N<sup>PRO</sup>)

N<sup>pro</sup> is an autoprotease approx. 20 kDa (168aa) in size which when translated cleaves itself off from the remaining polyprotein thus releasing the N-terminus of the subsequent core protein [52]. Figure-1.7 is a sequence alignment of the N-terminal protease of several pestiviruses. Amino acid sequence of N<sup>pro</sup> is highly conserved over all known pestiviruses. The N-terminal half of N<sup>pro</sup> is predicted to be disordered due to the

high density of prolines (1 in every 11 amino acids) in this half of the protein. The C-terminal half of  $N^{\text{pro}}$ , in contrast, is predicted to be mostly structured into  $\beta$ -strands [53].  $N^{\text{pro}}$  is a cysteine protease. The putative catalytic triad was predicted to be formed by Glu22-His49-Cys69 [54]. Mutations in these residues were shown to inhibit the proteolytic activity of  $N^{\text{pro}}$  in a cell-free translation system. However, sequence alignment with other proteases failed to show any significant homology, and a Glu-His-Cys catalytic triad arrangement has not been observed in other known cysteine proteases [54]. Thus, pestivirus  $N^{\text{pro}}$  has been classified into its own cysteine peptidase family, C53 [55].

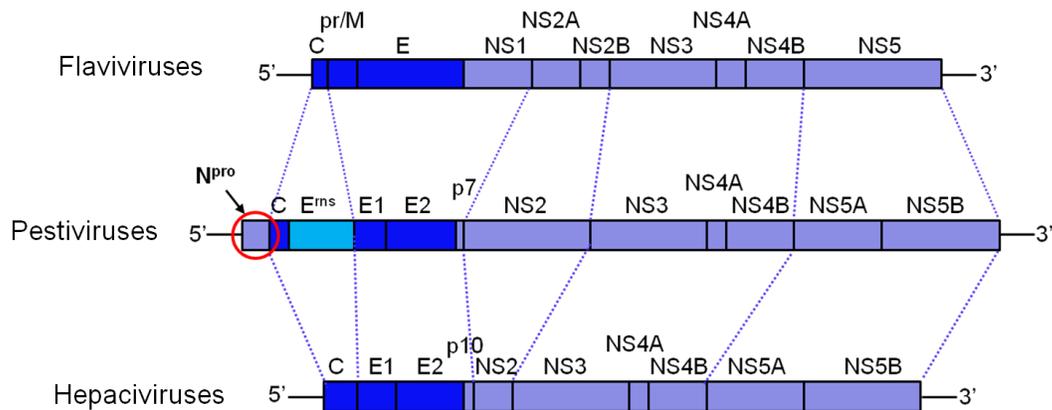


Figure 1.6: Polyprotein composition of *Flaviviridae*.

Comparison of structural and non-structural proteins encoded by viruses of the *Flaviviridae* family. Structural proteins are colored dark and light blue and non-structural proteins are colored gray. Dotted lines demarcate proteins with similar functions in all three viruses. Pestiviruses, as indicated, encode two additional proteins, the envelope protein  $E^{\text{ms}}$  and the N-terminal leader protease,  $N^{\text{pro}}$  (red circle).

The cleavage site of the protease is the C-terminus of  $N^{\text{pro}}$  (Cys168) [52]. The C-terminal half of  $N^{\text{pro}}$  also carries a highly conserved metal-binding motif viz., the TRASH motif (consensus sequence: Cys- $X_{19-22}$ -Cys- $X_3$ -Cys where X is any amino acid) formed by

Cys<sup>112</sup>-X<sub>21</sub>-Cys<sup>134</sup>-X<sub>3</sub>-Cys<sup>136</sup> [49,51]. Interestingly, following the first cleavage reaction, N<sup>pro</sup> blocks its own protease activity and has not been shown to possess any *trans* proteolytic activity upon release from the polyprotein [51]. Thus N<sup>pro</sup> is an auto-cleaving and auto-inhibitory cysteine protease.

N<sup>pro</sup> also plays a key role in the early stages of establishment of pestiviral infection. Viruses trigger the host's innate immune response following infection. One of these responses is mediated by the dsRNA intermediate formed during pestiviral RNA genome replication in the host cell cytoplasm. Cellular RNA helicases (RIG-1, MDA5 etc) recognize this intermediate and activate downstream kinases (TBK-1, IKK-ε) that in turn phosphorylate and activate IRF3 which then dimerizes and translocates into the nucleus. Translocated activated form of IRF3 triggers the transcription of interferon (IFN) α/β genes which coordinate the cells' antiviral and apoptotic responses together with triggering adaptive immunity in the host organism [52,53]; these events are briefly illustrated in figure-1.8.

N<sup>pro</sup> of CSFV and BVDV down-regulate the activation of IFN-α/β following virus infection. This down-regulation was shown to be a result of reduction in protein expression levels of IRF3, the transcription factor necessary for activation of IFN-α/β genes. Deletion of N<sup>pro</sup> gene of CSFV failed to protect SK-6 cells from poly(IC)-induced cell death, indicating that N<sup>pro</sup> is important for subverting host's immune response [54-55]. The down-regulatory effect of N<sup>pro</sup> on IFN-α/β activation was abolished in the presence of epoxomicin, a potent proteasome inhibitor [61–63]. Subsequently it was shown that N<sup>pro</sup> binds IRF3 (the unmodified or unphosphorylated form) in the cytoplasm of the infected cells and targets it for poly-ubiquitination and proteasomal degradation. Mutation of the catalytic residue Cys69 to Ala had a minimal effect on IFN induction in both BVDV and CSFV N<sup>pro</sup>, indicating that the protease activity is not involved in IRF3

binding [64]. Taken together, N<sup>pro</sup> blocks the activation of IFN- $\alpha/\beta$  activation pathway and counters the interferon mediated immediate innate immune response of the host cell independent of its autoprotease activity.

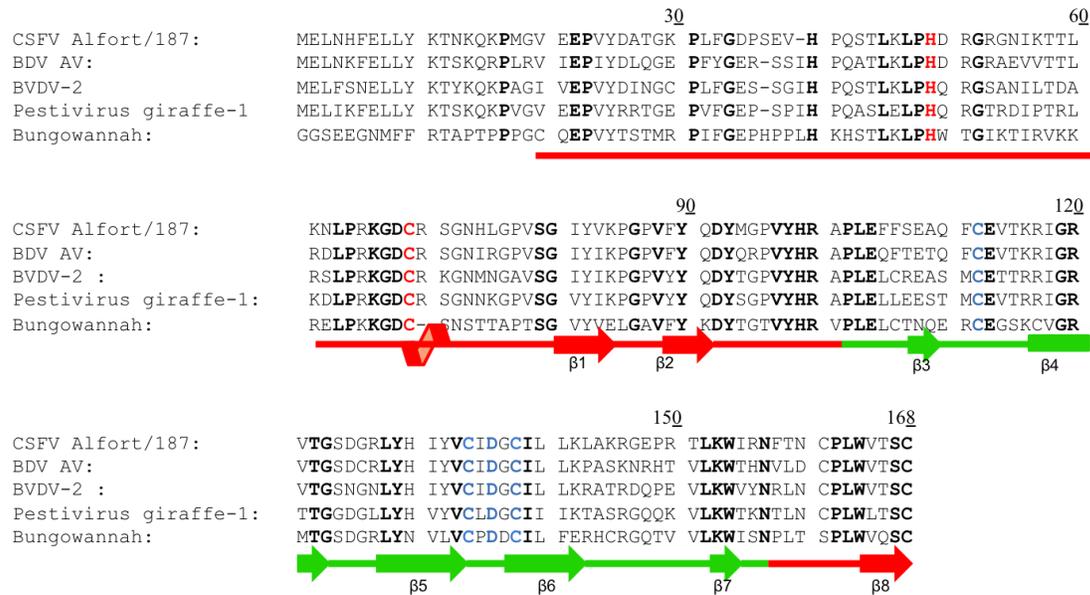


Figure 1.7: Sequence alignment of pestivirus N<sup>pro</sup>.

The alignment includes the CSFV strain Alfort/187, Border Disease virus (BDV) strain AV, BVDV-2, Pestivirus giraffe-1, and the Bungowannah virus (GenBank accession numbers X87939.1, ABV54604.1, AAV69983.1, NP\_777520.1, and DQ901403.1, respectively). Amino acid numbering corresponds to the CSFV Alfort/187 sequence. Residues identical in all pestiviruses are shown in bold. The proposed cysteine protease triad (E22-H49-C69) is colored red. Cleavage site C168 is the C-terminus of N<sup>pro</sup> (green). The TRASH motif sequence (C<sup>112</sup>-X<sub>21</sub>-C<sup>134</sup>-X<sub>3</sub>-C<sup>136</sup>) along with D136 predicted to be part of the Zn coordination site of N<sup>pro</sup> are indicated in blue. The domain organization of N<sup>pro</sup> and the secondary structure elements as determined from its crystal structure (Chapter 2) are overlaid on the sequence. Red and green regions indicate two domains of the protein respectively.

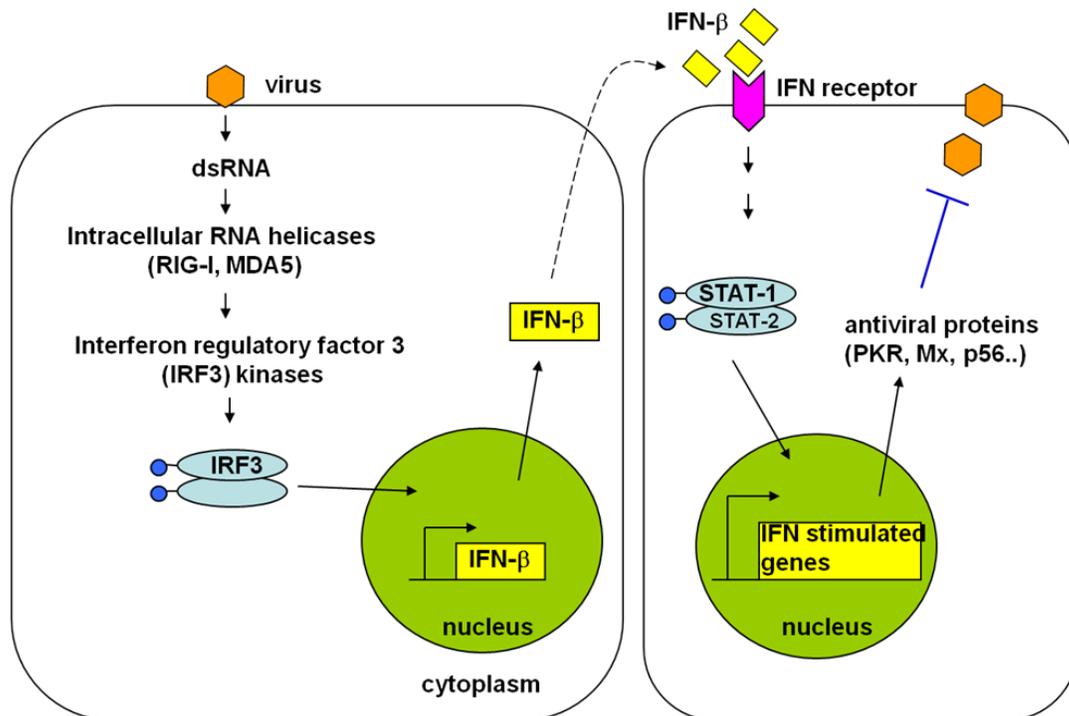


Figure 1.8: IFN- $\alpha/\beta$  mediated innate immune mechanism.

IFN- $\alpha/\beta$  synthesis is triggered by several of the interferon regulatory factors (IRFs) in immune cells. One of the IRFs involved in the IFN activation is IRF3. dsRNA is a pathogen associate molecular pattern (PAMP) of RNA viruses formed during virus replication, and is recognized by host cell cytosolic helicases such as RIG-1 and MDA5 (pathogen recognition receptors or PRRs). PRRs activate downstream kinases such as TBK-1/IKK $\epsilon$  that in turn phosphorylate IRF3. Phosphorylated IRF3 dimerizes, translocates into the nucleus, associates with CBP/P300 (transcription regulatory factor) and activates transcription of IFN- $\alpha/\beta$  genes. Released IFN- $\alpha/\beta$  then coordinates the expression of ISGs (Interferon Stimulated Genes), which are instrumental in carrying out the infected cell's antiviral and apoptotic responses.

## 1.6 SPECIFIC AIMS OF CURRENT STUDY

Despite the importance of N<sup>pro</sup> in viral replication and in counteracting the host immune response, little is known about the protein. The mechanism of autocatalytic cleavage of its C-terminus and the subsequent auto-inhibition of *trans* proteolysis, and

the molecular basis of its interaction with IRF3 and the subsequent down-regulation of IFN activation remain unknown. Although the protease activity itself is not essential for binding IRF3, mutations in either the N- or C-terminal regions of the protein disrupt this interaction (discussed in detail in chapter 5), suggesting that the two activities may share structurally overlapping regions on the protein. Thus overall goal of the project is to biochemically and structurally characterize the pestiviral N-terminal protease and determine the protein's functions in the viral life cycle. Towards this goal, I had three specific aims:

#### **1.6.1 Aim 1: To determine the molecular structure of N-terminal protease of CSFV**

We determined the X-ray crystal structure of wild-type CSFV N<sup>pro</sup> by single wavelength anomalous dispersion (SAD) method. In addition, structure of the cleavage site mutant N<sup>pro</sup>-C168A was also determined using molecular replacement with the wild type N<sup>pro</sup> structure as a model (Chapters 2 and 3). The crystal structure of N<sup>pro</sup> establishes the mechanism of auto-catalysis and subsequent auto-inhibition of protease activity by intramolecular product inhibition, and provides insight into N<sup>pro</sup>'s interaction with IRF-3 and its role in subversion of host immune response.

#### **1.6.2 Aim 2: To determine the molecular mechanism of auto-catalysis and auto-inhibition of CSFV N<sup>pro</sup>**

We hypothesized that the C-terminus of N<sup>pro</sup> remains bound in its own active site after autoproteolysis, and thus inhibits subsequent binding and cleavage of additional substrates (intramolecular inhibition of *trans* activity). We tested this intramolecular product inhibition mechanism by introducing C-terminal deletions to restore *trans* catalytic activity of N<sup>pro</sup>. Additionally, we tested the effect of various mutations on the *in vitro* autocatalytic activity of N<sup>pro</sup> using a N<sup>pro</sup>-GFP fusion construct (Chapter 4).

### **1.6.3 Aim 3: To determine the molecular basis of interaction of CSFV N<sup>pro</sup> with porcine IRF3**

N<sup>pro</sup> targets IRF3 for proteasomal degradation, and thus prevents induction of interferon in infected cells. We mapped the amino acids implicated in interaction of N<sup>pro</sup> with IRF3 in BVDV and CSFV onto two opposite surfaces on the crystal structure of N<sup>pro</sup>. To test interaction between N<sup>pro</sup> and IRF3, porcine IRF3 was expressed and purified. Its interaction with N<sup>pro</sup> will be analyzed using size-exclusion chromatography and analytical ultracentrifugation. To obtain a structure of N<sup>pro</sup>-IRF3 complex, co-expression of CSFV N<sup>pro</sup> and porcine IRF3 in *E. coli* and co-crystallization of the protein complex is under way.

## **1.7 SIGNIFICANCE OF CURRENT STUDY**

N<sup>pro</sup> is a unique cysteine protease that has been characterized into its own family of cysteine proteases (C53). Its protease activity can not be inhibited by known cysteine protease inhibitors. Hence a thorough knowledge of the protease could not only shed light on its function in the virus but also unravel a new class of proteolytic enzymes as yet unidentified. Furthermore, N<sup>pro</sup> is known to play essential roles in the pestivirus life cycle including viral polyprotein processing and evasion of the host immune response via subversion of the IRF3 pathway. Our biochemical, biophysical, and structural studies will elucidate the mechanism of autoproteolysis and inhibition of interferon activation. This information can be used to identify potential targets for therapeutics and suggest strategies for intervention against BVDV and CSFV -induced diseases.

Although live attenuated vaccines are available for both BVDV and CSFV, the use of vaccines in controlling pestiviral epidemics is limited since a stamp-out policy, as opposed to vaccination has been considered economically feasible in the long term.

[8,60]. Added to this, the inability to discriminate antibodies induced by vaccines from those elicited following a natural infection has necessitated generation of emergency marker vaccines and other fast-responding anti-viral strategies. N<sup>pro</sup> deletion marker vaccines for CSFV have been shown to protect the animals from subsequent viral challenge. However, deletion of the whole N<sup>pro</sup> gene showed altered slower growth kinetics compared to wild type N<sup>pro</sup>. If we could selectively retain the auto-cleavage activity of N<sup>pro</sup> and disrupt its antagonistic effect on IFN $\alpha$ / $\beta$  activation, we could enable generation of feasible marker vaccines against CSFV and BVDV. It may be possible to modulate the auto-proteolytic and immune-suppressing functions in a chimaeric vaccine for pestiviruses or another virus infecting ruminants. Similarly, these two functions could be independently targeted through selective anti-viral agents.

## CHAPTER-2

### Molecular Structure of Wild Type N<sup>pro</sup> of CSFV

#### 2.1 INTRODUCTION

Towards the biochemical and structural characterization of pestiviral N<sup>pro</sup> we have successfully expressed and purified full-length recombinant wild type N<sup>pro</sup> of classical swine fever virus from *E. coli*. Using limited proteolysis we have previously shown that N<sup>pro</sup> comprises two approximately equal sized domains, the N-terminal domain carrying the proposed cysteine protease active site (Glu22, His49 and Cys69) and the C-terminal domain that carries the conserved metal binding TRASH motif formed by amino acids Cys112-Cys134-Cys138.

The cysteine protease of N<sup>pro</sup> catalyzes a single *cis*-cleavage reaction at its own C-terminus. Following this, the protease has no *trans* activity. Limited proteolysis experiments demonstrated that the C-terminus of the protein is protected from digestion by exogenous proteinases suggesting that the C-terminus is buried and not solvent accessible. Since N<sup>pro</sup> autoinhibits its own protease activity after the first *cis*-cleavage reaction, we hypothesized that the C-terminus of N<sup>pro</sup> remains buried in the substrate binding pocket and inhibits *trans* activity of the protease by intra-molecular product inhibition.

We have also shown that N<sup>pro</sup>'s TRASH motif coordinates a single zinc atom. Using mammalian two hybrid assays it was determined that zinc binding ability of pestiviral N<sup>pro</sup> and an intact TRASH motif are essential for the interaction between N<sup>pro</sup> and IRF3. This suggested that the TRASH motif site could form a surface mediating binding between N<sup>pro</sup> and IRF3. Individual substitutions of the Cys residues in the

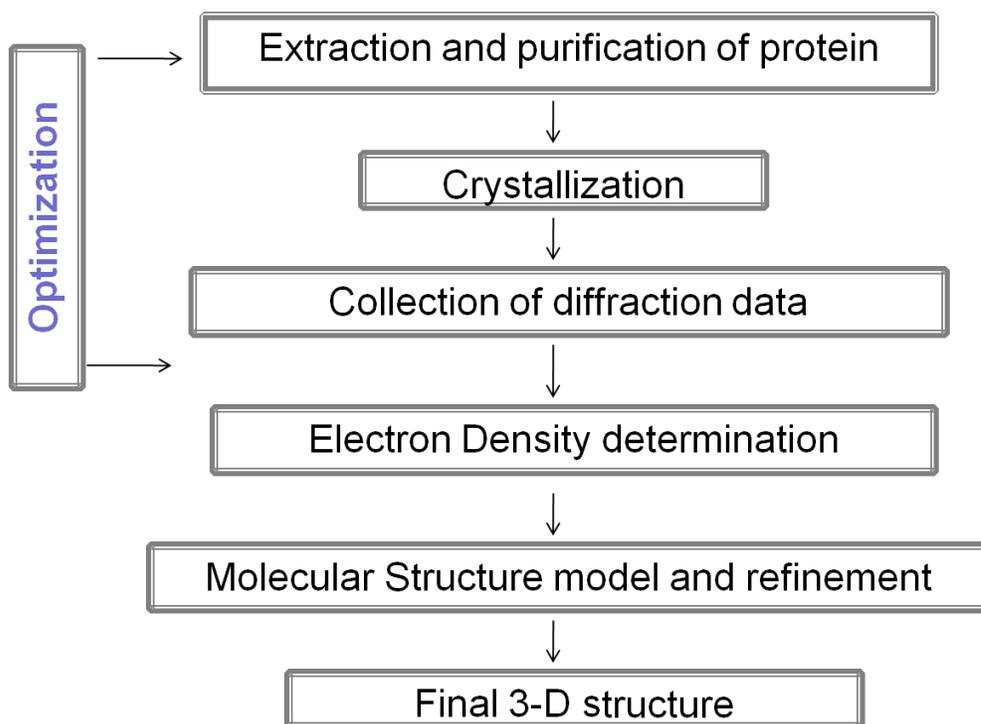
TRASH motif of CSFV N<sup>pro</sup> abolished its interaction with IRF3 and resulted in the loss of virus-mediated IRF3 degradation in infected cells. Accordingly, an increase of IFN $\alpha/\beta$  response was seen in porcine PK-15 cells that were infected by mutant viruses carrying mutations in the Cys residues of the TRASH motif [53]. It was shown that the proteolytic activity of N<sup>pro</sup> is not required for the suppression of IFN $\alpha/\beta$  induction. Cys69Ala mutation that knocked down the N<sup>pro</sup> protease activity did not disrupt the anti-IFN effect of the protein in BVDV and CSFV. However, mutations of Glu22 to Ala or Val and His49 to Leu or Val that resulted in a loss of catalytic activity in N<sup>pro</sup> resulted in a loss of IFN-antagonistic activity, suggesting that structural changes in these mutants that disable catalysis also disrupt IRF3 binding and degradation [58,59,61].

The fold of the two domains of N<sup>pro</sup>, the mechanism of action of its cysteine protease, and the interface of its interaction with IRF3 remain unknown. Further characterization of pestivirus N<sup>pro</sup> would be tremendously aided by a thorough knowledge of the molecular structure of the protein. We thus aimed to determine the 3-dimensional structure of N<sup>pro</sup> of CSFV using X-ray crystallography.

## 2.2 EXPERIMENTAL DESIGN

Below is a flow chart of the procedure for determining the molecular structure of a protein using X-ray crystallography.

### Structure determination using X-ray crystallography



Following is a detailed account of the experimental design along with outcomes of each trial, the results, failures and alternate methods used during the determination of the 3-dimensional molecular structure of wild type N-terminal protease of CSFV.

### 2.2.1 Expression and purification of wild type N<sup>pro</sup>

We obtained clones of wild type CSFV and BVDV N<sup>pro</sup> in pET expression vectors from Dr. Nicolas Ruggli at IVI Switzerland. Briefly, the N<sup>pro</sup> gene of CSFV (strain Alfort/187, GenBank accession number X87939.1) and BVDV (strain NADL, Genbank accession number AJ133738.1) were amplified by PCR and cloned into pCR4-TOPO (Invitrogen). The DNA fragment containing the N<sup>pro</sup> gene was then subcloned into the *NdeI* and *XhoI* restriction sites of pET-15b to obtain pET-6H-throm-N<sup>pro</sup>.

The plasmids containing the cloned constructs of N<sup>pro</sup> were transformed into *E. coli* Rosetta™ (Novagen) cells. Large-scale cultures were started from single colonies containing the target plasmid, in LB medium supplemented with 50 µg/ml Ampicillin and 33 µg/ml Chloramphenicol at 37°C. Cultures were grown to an O.D.<sub>600</sub> of 0.6-0.8 at which point they were induced with 0.5 mM IPTG and growth was continued at 18°C overnight. The harvested cells were suspended in lysis buffer (50 mM sodium phosphate pH 7.0, 300 mM NaCl, 5 mM β-MercaptoEthanol) mixed with EDTA-free protease inhibitor cocktail (Roche). Cells were lysed by sonication with a Misonix XL-2000 sonicator followed by centrifugation to remove cellular debris. Immobilized Metal Affinity Chromatography (IMAC) was used to isolate N<sup>pro</sup> with the N-terminal 6xHis-tag from the soluble fraction of the cell lysate. N<sup>pro</sup> was bound to Talon™ (Clontech) metal affinity resin according to manufacturer's protocol. Bound protein was eluted with a gradient of 5-200 mM Imidazole in 50 mM Sodium phosphate, pH 7.0 and 0.5 M NaCl. Following IMAC purification, fractions containing N<sup>pro</sup> were pooled together, mixed with equal volume of gel filtration buffer (20 mM Tris, pH 7.5) and concentrated using an Ultrafree centrifugal concentrator (Millipore) with a mol. Wt. cutoff of 10 kDa. Concentrated protein was then loaded on to the Sephadex-75 gel-filtration column (GE Healthcare) and eluted with 20 mM Tris-HCl (pH 7.5) containing 5 mM β-

mercaptoethanol, using the AKTA™ purification system (GE Healthcare). Figure-2.1 shows a typical SDS-PAGE from a single purification cycle.

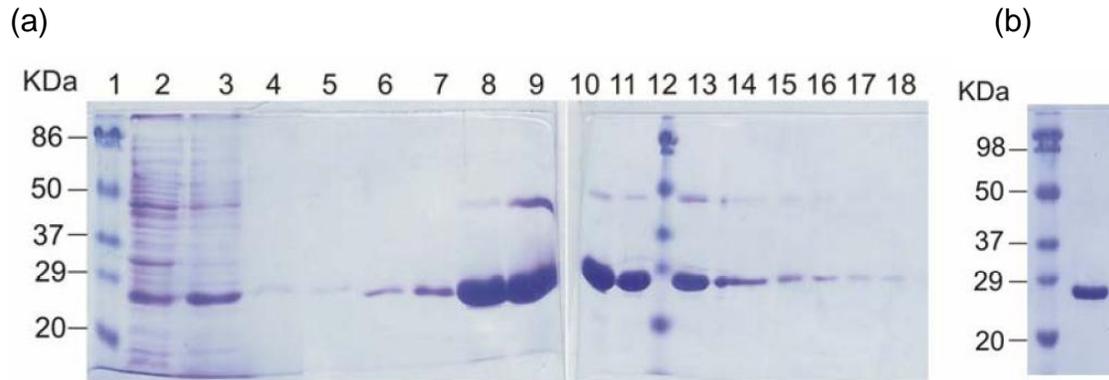


Figure 2.1: Purification of full-length wild-type CSFV N<sup>pro</sup>.

(a) SDS-PAGE analysis of fractions from metal affinity chromatography using Talon™ resin. Lanes 1 and 12: Marker; lane 2: cell lysate; lane 3, flow-through from the column; lane 4, 5 mM Imidazole wash; Lanes 5-11 and 13-18, elution fractions Talon using 5 mM-150 mM Imidazole gradient. (b) SDS-PAGE of the pure full-length N<sup>pro</sup> obtained after size exclusion chromatography.

Concentration of N<sup>pro</sup> was determined by measuring its absorbance at 280 nm ( $A_{280}$ ). Molar extinction coefficient of N<sup>pro</sup> was determined using a modified form of the method described by Pace *et al.* with the ProtParam tool on ExPASy, the bioinformatics resource portal (<http://au.expasy.org/tools/protparam.html>)[67–69]. Molar extinction coefficient of full-length wild type N<sup>pro</sup> of CSFV is  $23295 \text{ M}^{-1}\text{cm}^{-1}$ . The calculated isoelectric point of N-terminal 6xhis-tagged N<sup>pro</sup> is 9.25. We typically saw > 98% purity following the final purification step and yields of around 10mg of recombinant wild type N<sup>pro</sup> per litre of bacterial culture

## 2.2.2 Crystallization trials with full length CSFV N<sup>pro</sup>

### 2.2.2.1 Trial 1

Protein concentration of ~20 mg/ml was determined to be ideal for crystallization trials using the Hampton Pre-Crystallization Test kit. Following concentration, the protein sample was centrifuged at 15,000 rpm for 10 min to remove any precipitate from the mixture; filtered using a 0.2µm centrifugal filter and the resultant sample was used to set up high-throughput crystal trials. Commercially available sparse matrix crystallization screens were used for initial crystallization trials. Trials were conducted by the vapor diffusion method using sitting drops set in 96-well Intelli-plates by Rigaku's Phoenix RE liquid handling robot. However, full-length N<sup>pro</sup> (CSFV and BVDV) proved to be resistant to crystallization.

### 2.2.2.2 Trial 2

Secondary structure prediction algorithms indicated that majority of the sequence of N<sup>pro</sup> would be disordered due to the high proline content of the sequence [53]. We rationalized that by adopting methods that would decrease the intrinsic flexibility in the structure of N<sup>pro</sup> we could induce crystallization. Crystallization trials performed in 2-5 M TMAO (Trimethylamine *N*-oxide), that has been shown to increase secondary structure content of intrinsically disordered proteins [65,66], were unsuccessful. One of the methods described in literature that would potentially achieve a reduction in intrinsic flexibility of the protein is the introduction of bulky methyl groups on lysine side chains using *reductive methylation*. This procedure was shown to aid crystallization of a number of proteins that failed to crystallize or crystallized poorly otherwise [72–75]. The protocol for methylation is similar to the one described by Walter et al 2006 [72]. Briefly, purified recombinant N<sup>pro</sup> was dialysed into buffer containing 50 mM Hepes and 250 mM

NaCl. The protein concentration was kept at 1 mg/ml for methylation reaction since higher protein concentrations led to excessive precipitation of N<sup>pro</sup> during the reaction process. 20 µl of freshly prepared 1 M dimethyl-amine-borane complex (ABC) and 40 µl of 1 M formaldehyde were added per ml of protein solution and the reactions were gently mixed and incubated at 4°C for 2 hr. Further 20 µl of ABC and 40 µl of formaldehyde were added and the incubation continued for 2 hr. Following a final addition of 10 µl ABC, the reaction is incubated overnight at 4°C. Precipitated protein was removed by centrifugation at 15,000 rpm for 10 min. SDS-PAGE analysis showed a visible difference in molecular weight between the unmethylated and methylated CSFV-N<sup>pro</sup> (Figure-2.2). Methylation was further verified using mass spectrometry (MALDI-TOF) of unmethylated and methylated N<sup>pro</sup>. Following completion of methylation, the protein was exchanged in to 20 mM Tris and 200 mM NaCl and concentrated using the Ultrafree centrifugal concentrator (Millipore) with a mol. wt. cutoff of 10 kDa. Due to increase in hydrophobicity from additional methyl groups, the concentration of modified N<sup>pro</sup> used for crystallization was typically lower than the wild type unmodified protein. The concentrated protein was used for high-throughput crystallization trials. Methylation of lysine side chains of N<sup>pro</sup>, however failed to crystallize the protein.

### ***2.2.2.3 Trial 3***

Literature search showed numerous instances of enzymes crystallized in their inactive state, wherein the active site amino acids were mutated to abrogate enzymatic activity, which consequently increased the stability of the active site and in turn stabilized the whole protein in solution. The proposed cysteine protease nucleophile of N<sup>pro</sup> is Cys69. We hence mutated Cys69 to Ala for crystallization purposes. The mutation did not have any effect on the solubility or stability of the protein in solution; however this

mutation also failed to yield any crystals. We tried to crystallize a substrate bound form of the inactive protease ( $N^{\text{pro}}$ -C69A-SDDG) with inclusion of four amino acid residues ( $^{169}\text{SDDG}^{172}$ ) that form the cleaved substrate peptide following the C-terminus cleavage site of  $N^{\text{pro}}$ . This mutant also was resistant to crystallization.

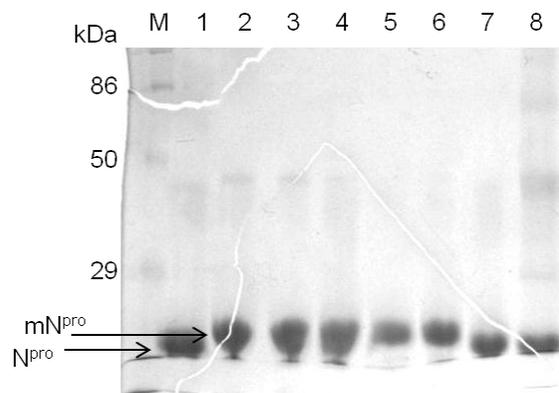


Figure 2.2: Lysine methylation.

SDS-PAGE showing the change in molecular weight of full-length  $N^{\text{pro}}$  after methylation of lysines. Lane M: Mol. Wt. Marker; lane 1: Wild type full length  $N^{\text{pro}}$  before methylation; lane 3-7: Methylation reaction product at 2 hr time points (2, 4, 6, 8 hrs and overnight); lane 7: Unmethylated protein at 4°C over night; lane 8: Precipitate from the reaction. The expected change in molecular weight assuming double methylation ( $-\text{CH}_2-\text{CH}_2$ ) of lysine side chains in wild type  $N^{\text{pro}}$  is  $\sim 370\text{Da}$  ( $N^{\text{pro}}$  has 13 lysines in its sequence).

### 2.2.3 Crystallization trials with individual domains of CSFV $N^{\text{pro}}$

#### 2.2.3.1 Domain organization of $N^{\text{pro}}$

Following the above unsuccessful crystallization attempts, we ventured to isolate individual domains of  $N^{\text{pro}}$  and attempt their independent structure determination. We had previously determined, using limited proteolysis and concomitant mass spectrometry

analysis that N<sup>pro</sup> is divided into two distinct domains with approximately equal molecular weight. Domain arrangement and the cleavage sites of trypsin and chymotrypsin identified using mass spectrometry and peptide N-terminal sequencing are shown in figure-2.3 (reproduced with permission from Szymanski *et al.* 2010<sup>2</sup>). Limited proteolysis with trypsin gave rise to two fragments that were protected from further proteolytic activity up to 2hrs in solution. The N-terminal boundary of the larger fragment with a molecular weight of ~17.3 kDa was identified as Gln15 (Q15). This fragment represented a stable compact core of the protein that was resistant to proteolytic digestion.

#### ***2.2.3.2 Isolation of stable domains of CSFV N<sup>pro</sup>***

We performed limited proteolysis of recombinant full length N<sup>pro</sup> in 20 mM Tris and 200 mM NaCl for 20 min. Following this, size exclusion chromatography with Superdex-75 (GE) prep-grade was used to separate the two proteolytic fragments. However, we were unable to chromatographically separate the two limited proteolysis fragments (17 kDa and 7 kDa) probably due to the small difference in molecular weight of the two fragments (~10 kDa) being beyond the resolution limit of the column.

We then attempted *in situ* proteolysis of full length CSFV N<sup>pro</sup> in crystallization drops with trypsin [76]. Briefly, following the final step of purification of full length N<sup>pro</sup>, the protein was concentrated to ~20 mg/ml. 1% (w/w) of trypsin was added to the concentrated protein sample and incubated at room temperature for 10 min. The protein mixture was then used to set up crystallization trials. This method gave rise to micro crystals, which however could not be scaled up after several attempts. We repeated the above procedure of proteolytically separating domains of N<sup>pro</sup> once again using digestion

---

<sup>2</sup> Refer to Appendix A for the license agreement.

with Endoproteinase Lys-C, a protease that cleaves proteins specifically on the C-terminal side of lysine amino acids [77]. Lys-C cleaves N<sup>pro</sup> at Lys14 and hence produces fragments of similar size as trypsin. As with trypsin, Lys-C cleaved products could not be separated using size exclusion chromatography.

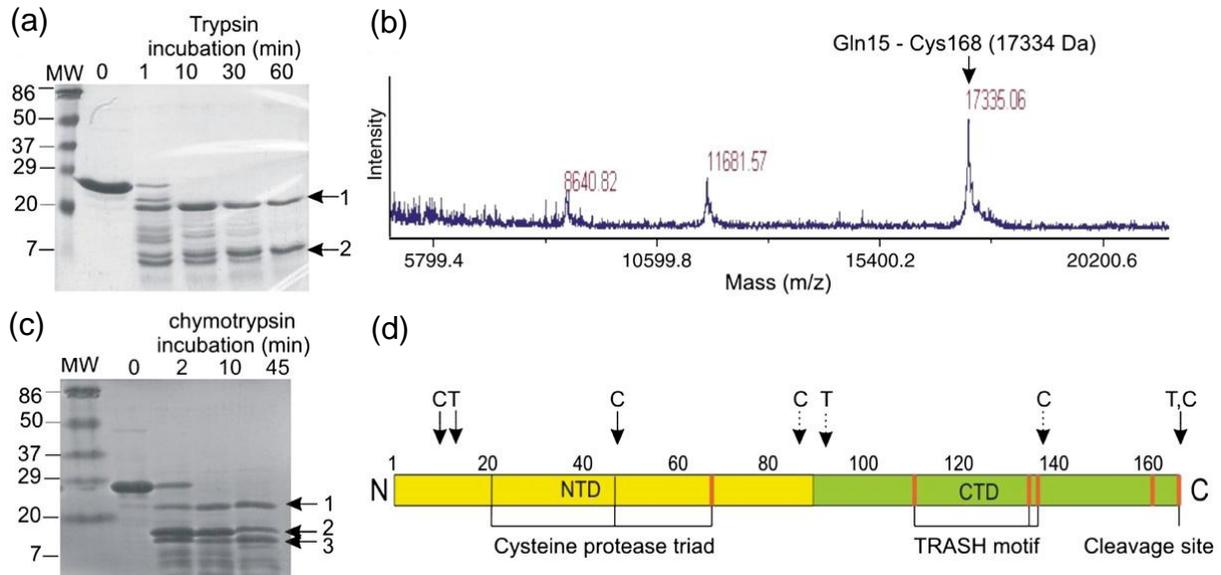


Figure 2.3: Limited proteolysis of CSFV N<sup>pro</sup>.

(a) Proteolysis of N<sup>pro</sup> with trypsin. The two fragments indicated by arrows were sequenced at the N-terminus, and the identified cleavage site was between Lys14 and Gln15 for both fragments. (b) MALDI-MS analysis of the trypsin digest of N<sup>pro</sup>. The mass of the largest peak (17,335 Da) corresponds to the calculated mass for a peptide consisting of residues from Gln15 through Cys168 (17,334 Da). (c) Proteolysis of N<sup>pro</sup> with chymotrypsin. The three fragments indicated by arrows were further analyzed by N-terminal sequencing. (d) Schematic diagram of N<sup>pro</sup>'s domain arrangement. The overall structure of N<sup>pro</sup> can be divided into an N-terminal protease domain (yellow) and a C-terminal Zn-binding domain (green) that are approximately of equal size. Vertical orange lines represent the locations of Cys residues. The locations of the proposed cysteine protease triad and the TRASH motif are indicated. The cleavage sites from the trypsin and chymotrypsin digests are indicated by arrows above the schematic, designated T and C, respectively. The dotted arrows indicate uncertainty in the cleavage sites.

## 2.2.4 Crystallization of N-terminal deletion mutant of N<sup>pro</sup>

Failure to isolate individual domains of N<sup>pro</sup> prompted us to design constructs spanning the domain boundaries determined from limited proteolysis experiments. We obtained the deletion mutant of wild type N<sup>pro</sup> lacking the first 14 amino acids (6xHis-N<sup>pro</sup>-Δ14N) -the cleavage site of trypsin- from the laboratory of Dr. Ruggli. This construct was the N-terminal His-tagged form of the larger cleavage product of N<sup>pro</sup> with trypsin. However, this construct also failed to yield any crystals. We reasoned that the N-terminal His-tag was interfering with the crystallization process since previously we saw micro crystals from an *in situ* proteolysis product of N<sup>pro</sup> with trypsin. We hence sub-cloned the (6xHis-N<sup>pro</sup>-Δ14N) in to pET-15b vector such that it incorporated a thrombin cleavage site (LVPRGS) between the His-tag and N-terminus amino acid Q15 of N<sup>pro</sup>.

### 2.2.4.1 Construction and purification of deletion mutant of N<sup>pro</sup>

The N-terminal deletion mutant of N<sup>pro</sup> lacking the first 14 amino acids (N<sup>pro</sup>-Δ14N) was amplified from full-length wild type N<sup>pro</sup> using the oligonucleotide primers 5' cgcgccagccatatgcaaaaaccaatgggagtggag 3' (forward) and 5' cggatcctcgagtttagcaactgtaaccacaatgg 3' (reverse). The PCR product was sub-cloned into the pET15b expression vector between the *NdeI* and *XhoI* restriction sites, which fused a cleavable 6x-His tag on the N-terminus of the protein. N<sup>pro</sup>-Δ14N was isolated from large-scale bacterial cultures using IMAC in a method similar to the full length N<sup>pro</sup>.

Following elution from Talon™, fractions containing the protein were pooled and dialysed into buffer A (20 mM Tris pH 8.0, 100 mM NaCl and 5 mM β-Mercaptoethanol) overnight. Thrombin protease immobilized on agarose beads (Thermoscientific) was used to cleave the N-terminal 6xHis-tag of the protein construct, which contained the thrombin protease cleavage site (LVPRGS). The cleavage reaction was performed at room temperature in buffer A for 4 hrs using 1 μg of Thrombin for 100 μg of target protein. The

target protein was separated from the cleaved 6xHis-tag by binding the mixture to the Talon™ metal-affinity resin equilibrated in buffer A. Resin was washed with low concentration of imidazole to release bound N<sup>pro</sup>-Δ14N since the protein exhibited non-specific binding to the IMAC resin even in the absence of His-tag.

#### ***2.2.4.2 Crystallization and data collection on N<sup>pro</sup>- Δ14N crystals***

Protein was concentrated to 10 mg/ml using Ultrafree centrifugal concentrator (Millipore) with a mol. wt. cutoff of 10 kDa. Sample was prepared for crystallization in a manner described for the full-length wild type N<sup>pro</sup>. Initial screening for crystallization conditions was done once again using commercially available high-throughput screens (drop volume of 0.2 μl and reservoir volume of 50 μl). Subsequent optimization of crystallization conditions was done using 24-well LINBRO plates by the hanging drop vapor diffusion method with a drop volume of 1 μl and reservoir volume of 500 μl. We obtained diffraction quality crystals of N-terminal deletion mutant of wild type N<sup>pro</sup> (N<sup>pro</sup>-Δ14N) in 23% PEG3350, 0.2 M MgCl<sub>2</sub> and 0.1 M HEPES pH 7.9. Typical size of these crystals was 100 μm × 50 μm × 50 μm.

The crystals were soaked in paratone-oil briefly for cryoprotection before being flash frozen in nitrogen stream. Native data set was collected using the home source at UTMB on the DIP-2030 (MacScience) image plate detector. Crystals of N<sup>pro</sup>-Δ14N diffracted to 2.8 Å. We indexed the unit cell to primitive orthorhombic space group P2<sub>1</sub>2<sub>1</sub>2 with dimensions a=60.08, b=62.6, c=30.8 Å. The asymmetric unit of the crystal unit cell contains a monomer. However, the quality of diffraction data obtained from the crystals of N<sup>pro</sup>-Δ14N was poor with maximum resolution of ~3 Å and mosaicity values as high as 2.5-3.0 in some data sets.

#### **2.2.4.3 Optimization of crystallization of the deletion mutant of N<sup>pro</sup>**

We made several attempts to improve the diffraction quality of the N<sup>pro</sup>-Δ14N crystals. Use of different salts, additives or different buffers did not have any effect on the quality of diffraction data. Since high mosaicity could be a result of abnormal crystal packing we reasoned that this could be a result of high flexibility on the N-terminus of N<sup>pro</sup> and/ or the presence of small amounts of the cleaved His-tag in the protein solution that co-elutes from the metal affinity column with low concentration of imidazole in the wash buffer (Figure-2.4). Subsequently we generated a shorter construct of N<sup>pro</sup> lacking the first 17 amino acids, which got rid of the N-terminal Glutamine (Q15). This construct (N<sup>pro</sup>-Δ17N) no longer displayed any non-specific binding to TALON resin following thrombin cleavage of its N-terminus 6xHis-tag. Figure-2.4 shows a comparison of the elution of N<sup>pro</sup>-Δ14N and N<sup>pro</sup>-Δ17N from TALON following thrombin cleavage of their respective N-terminal 6xHis-tags. The purified N<sup>pro</sup>-Δ17N crystallized in the same conditions as N<sup>pro</sup>-Δ14N. To our surprise, the crystals diffracted to as high as 1.5 Å at the home source and the quality of diffraction was tremendously improved (mosaicity of ~0.6) over the previous crystals. The structure of wild type N<sup>pro</sup> was thereby determined from the N-terminal deletion construct N<sup>pro</sup>-Δ17N.

#### **2.2.5 Data collection and structure determination of Wild type N<sup>pro</sup> using sulfur-SAD**

The N<sup>pro</sup>-Δ17N crystals were cryo-cooled at 100K using paratone as cryo-protectant. High redundancy data was collected using Bruker's Microstar™ microfocus X-ray Source on the Platinum<sup>200™</sup> CCD detector. The data were indexed and merged on Bruker AXS PROTEUM2, software suite for X-ray crystallography. The crystals diffracted to 1.6 Å resolution and belonged to the space group P2<sub>1</sub>2<sub>1</sub>2 with a=60.08, b=62.6, c=30.8 Å. The solvent content was 35% with a monomer in the asymmetric unit.

The structure of N<sup>pro</sup>-Δ17N was solved via the single wavelength anomalous dispersion method (SAD) using the anomalous signal present in sulfur atoms illuminated by a copper K-α home X-ray source. Determination of the positions of sulfur atoms, phasing, and calculation of electron density maps were performed using AutoSol wizard in the Phenix package [78,79]. The initial atomic model was obtained using the Autobuild wizard in Phenix [80]. The final model was achieved using manual model building with the program O [81] followed by iterative cycles of refinement with phenix.refine. All residues from Glu21 to Cys168 were visible in the electron density map except residues 65-71, which encompasses the catalytic Cys69. Ramachandran plot for N<sup>pro</sup>-Δ17N generated using PROCHECK [82] from CCP4 showed 93.8% of the residues in the most favored and 6.3% in additionally allowed regions with no disallowed residues. Data collection, phasing and refinement statistics for the crystal structure N<sup>pro</sup>-Δ17N are given in Table 2.1.

## 2.3 RESULTS

### 2.3.1 Overall structure of wild-type N<sup>pro</sup>

The molecular structure of wild type N<sup>pro</sup> was determined to 1.6 Å resolution. This is the first reported structure of N<sup>pro</sup>. Figure-2.5 shows the cartoon representation of the crystal structure of N<sup>pro</sup> of CSFV. Due to the sequence identity between CSFV N<sup>pro</sup> and N<sup>pro</sup> from other pestiviruses it is conceivable that the structure of CSFV N<sup>pro</sup> is representative of N-terminal proteases from other pestiviruses (such as BVDV and BDV). The structure of N<sup>pro</sup> is predominantly comprised of β-strands that adopt a unique ‘clam shell’-like fold. The protein can be divided into two distinct domains, viz., the catalytic protease domain and the zinc-binding domain.

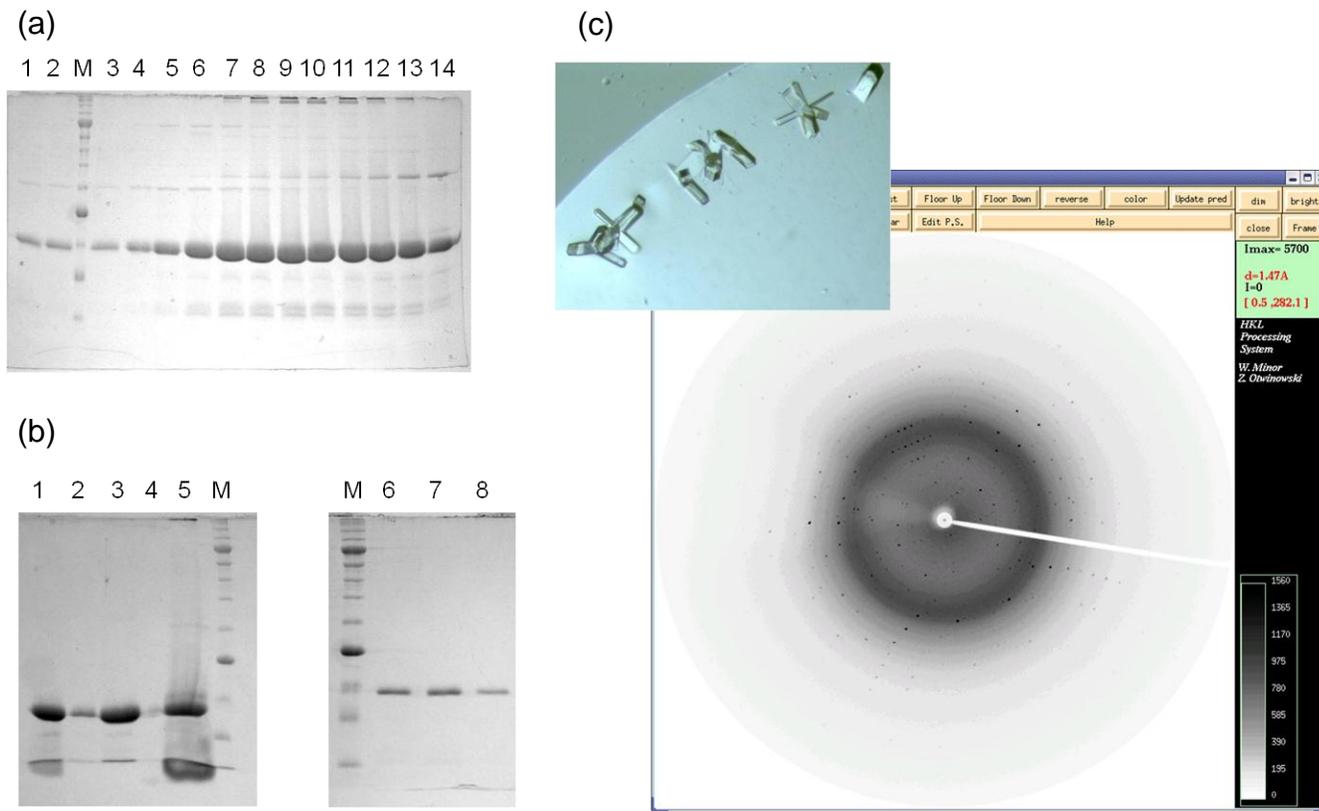


Figure 2.4: Crystallization of N<sup>pro</sup>-Δ17N.

(a) SDS-PAGE analysis of purification of N<sup>pro</sup>-Δ17N using a gradient of 5-150 mM imidazole to elute protein bound to Talon™ (lanes 1-14). (b) Removal of Thrombin cleaved 6xHis-tag from N<sup>pro</sup>-Δ14N (lanes 1-5: lane 1, cleavage product; lane 2, flow through from metal affinity resin; lane 3, 5 mM Imidazole wash; lane 4, 10 mM Imidazole wash; lane 5: 100mM imidazole elute) and from N<sup>pro</sup>-Δ17N (lanes 6-8: flow through from Talon after cleavage of 6xHis-tag). N<sup>pro</sup>-Δ14N binds non-specifically to the resin and low imidazole washes elute small amount of the cleaved 6xHis-tag. This non-specific binding was eliminated in the N<sup>pro</sup>-Δ17N deletion mutant. (c) Crystals of N<sup>pro</sup>-Δ17N and representative diffraction from the crystals.

<b>N<sup>pro</sup> - Δ17N (Sulfur-SAD)</b>	
<b>X-ray data collection</b>	
Wavelength (Å)	1.54178
Resolution limits(Å)	29.7-1.6
Total Reflections	660178
Unique Reflections	21755
Rsym (%)	4.2 (31.8)
I/σ	12.6 (1.98)
Completeness (%)	99.9 (99.8)
Space Group	P2 <sub>1</sub> 2 <sub>1</sub> 2 (No. 18)
Unit cell parameters (Å)	
<i>a</i>	60.133
<i>b</i>	65.398
<i>c</i>	33.315
<b>Phasing Statistics</b>	
Resolution (Å)	29.7-1.6
FOM	0.31
<b>Refinement Statistics</b>	
Resolution limits (Å)	29.7-1.6
Molecules/asymmetric unit	1
R <sub>work</sub> /R <sub>free</sub>	17.2/20
No. of atoms	
Protien	1115
Water	199
Average B-factor (Å <sup>2</sup> )	
Protein	17.53
Water	28.4
R.M.S deviations	
Bond Lengths (Å)	0.006
Bond Angles(°)	1.108
<b>Model Statistics</b>	
Ramachandran Plot	
Most favored (%)	93.8
Additional allowed (%)	6.3
generously allowed	0
Disallowed (%)	0

Table 2.1: Data collection, phasing and refinement statistics of N<sup>pro</sup>-Δ17N.

The protease domain spans the N-terminus through residue 100 and also includes residues 157 through the C-terminus residue 168. The domain harbors the protease active site along with the C-terminal protease cleavage site Cys168. The protease domain comprises mostly coils with a single  $\beta$ -sheet formed by strands  $\beta$ 1,  $\beta$ 2, and  $\beta$ 8. The first two  $\beta$ -strands of the sheet are contributed by the first 100 residues in the sequence. The last 6 residues at the C-terminus (163-168) form the final  $\beta$ -strand, and fold back into the protease active site, positioning the C-terminus Cys168 for cleavage. The N<sup>pro</sup> protease domain was predicted to be disordered, perhaps due to the abundance of proline residues; the domain contains twelve prolines corresponding to an average of one proline for every 7 residues. Most prolines are located in the loops on the surface of the protein, contributing to the unique fold of the protein. A search for similar folds using the DALI server resulted in zero instances, indicating that the catalytic domain of N<sup>pro</sup> has a new fold. The zinc-binding domain of N<sup>pro</sup> spans residues 101 through 156 and forms an anti-parallel  $\beta$ -sheet consisting of five  $\beta$ -strands,  $\beta$ 3,  $\beta$ 4,  $\beta$ 5,  $\beta$ 6, and  $\beta$ 7. This domain carries a conserved metal binding TRASH motif consisting of Cys112-Cys134-Asp136-Cys138, which coordinates a single zinc atom[53]. The TRASH motif is located at one end of the  $\beta$ -sheet. The interface between the protease and zinc-binding domains is mostly hydrophobic, and the C-terminal domain partially covers the final  $\beta$ -strand in the protease domain.

### **2.3.2 Cysteine protease of N<sup>pro</sup> has a catalytic dyad**

N<sup>pro</sup> cleaves the peptide bond between its C-terminus, Cys168 and the first amino acid of the core protein, Ser169 in the polyprotein sequence. Site-directed mutagenesis combined with translation in cell-free lysates showed that amino acid residues Glu22, His49, and Cys69 of N<sup>pro</sup> were essential for its cleavage action [54]. It was proposed from

sequence comparison and mutagenesis studies that Glu22-His49-Cys69 could form the cysteine protease catalytic triad, with Cys69 being the active site nucleophile, His49 the hydrogen bond acceptor which catalyzes the nucleophilic attack on the peptide bond while Glu22 could be the stabilizing group for the tetrahedral intermediate formed between His49 and Cys69. Crystal structure of wild type N<sup>pro</sup> shows that the protease domain carries the proposed cysteine protease triad along with the C-terminal cleavage site. In the structure, only His49 and Cys69 (dashed loop) form the active site pocket enclosing the cleavage site (C-terminus), suggesting the presence of a functional catalytic dyad. Contrary to predictions, Glu22 is not a part of the catalytic site since it is located at a distance of ~25 Å from the C-terminal cleavage site of N<sup>pro</sup>. Interestingly, mutagenesis studies have shown that Glu22 is absolutely essential for the proteolytic activity of N<sup>pro</sup>.

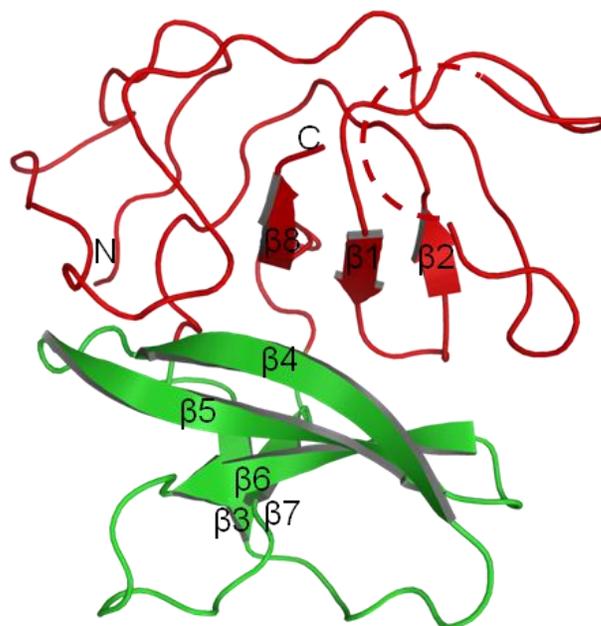


Figure 2.5: Structure of wild-type N<sup>pro</sup> (N<sup>pro</sup>-Δ17N).

Ribbon representation of wild-type N<sup>pro</sup>. Protease domain is shown in red and zinc-binding domain in green. Disordered residues (65 to 71) are indicated by dashed lines. Secondary structural elements, as well as the N- and C-termini, are labeled.

Despite being outside the catalytic pocket of the protease, deletion or alanine-substitution of this residue renders the protease inactive and N<sup>pro</sup> is no longer able to catalyze the cleavage at its C-terminus. Upon closer inspection, we found that the side chain of Glu22 indeed forms a salt bridge with the side chain of Arg100 in the structure (Figure-2.6). Perhaps, the breakdown of the salt bridge between Glu22 and Arg100 due to mutation drastically destabilizes the structure of the protease domain, in effect rendering the protease inactive. Thus Glu22 could enable the catalysis of cleavage at the C-terminus of N<sup>pro</sup> by preserving the structural integrity of the protease domain and could hence be critical to the activity of the cysteine protease.

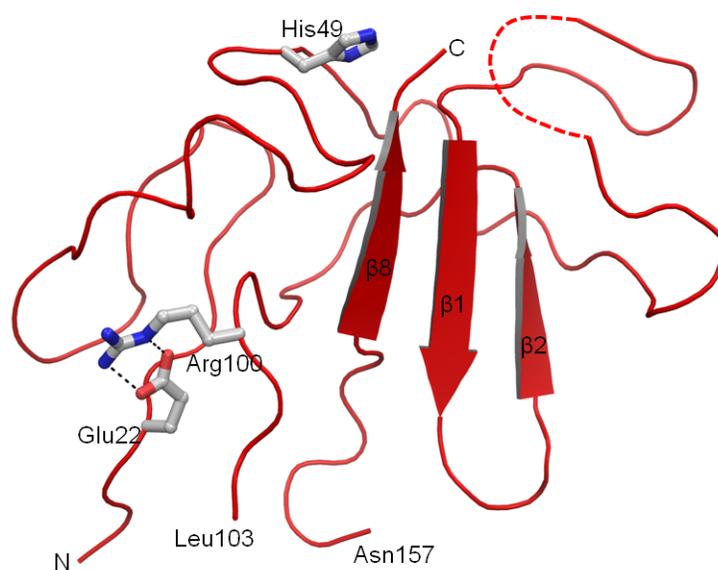


Figure 2.6: The protease domain of N<sup>pro</sup>.

Ribbon representation of the N-terminal protease domain of N<sup>pro</sup>. The catalytic site His49 is shown in ball-and-stick model. Cys69 location could not be determined due to the missing electron density in the structure (red dashes lines). Glu22 originally predicted to be a part of the active site is at a distance of  $\sim 25$  Å from the cleavage site and forms a salt bridge (black dashed lines) with Arg100. The Epsilon oxygens (O $\epsilon$ 1 and O $\epsilon$ 2) of the side chain of Glu22 are at a distance of 2.9 Å from the NH<sub>2</sub> Nitrogen and the Epsilon Nitrogen (N $\epsilon$ ) of the guanidium group of Arg100 respectively.

### 2.2.2 Zinc-binding domain of N<sup>pro</sup>

The C-terminal half of N<sup>pro</sup> carries a highly conserved metal-binding motif viz., the TRASH motif (consensus sequence: Cys-X<sub>19-22</sub>-Cys-X<sub>3</sub>-Cys where X is any amino acid) formed by Cys<sup>112</sup>-X<sub>21</sub>-Cys<sup>134</sup>-X<sub>3</sub>-Cys<sup>136</sup> [49,51]. This is a recently discovered sequence motif shown to be involved in copper homeostasis in cells. Using mass spectrometry coupled with metal content analysis, we previously showed that the TRASH motif of N<sup>pro</sup> together with Asp50 coordinates a single Zn atom in the structure. The C-terminal domain of N<sup>pro</sup> from residues 101 through 156 that carries the Zn coordinating motif has hence been termed the zinc-binding domain of N<sup>pro</sup>. We determined a 1:1 protein to Zn molar ratio in wild-type N<sup>pro</sup> of CSFV and BVDV. However, when TRASH motif cysteines were mutated to Ala N<sup>pro</sup> no longer bound Zn as determined from the above metal analysis [53]. Individual mutations of C112A/R, C134A, D136N, and C138A in the TRASH motif not only resulted in loss of zinc-binding, but also abolished IRF3 binding and subsequent IFN antagonism when introduced into the virus [53].

N<sup>pro</sup>'s zinc-binding domain forms an anti-parallel  $\beta$ -sheet consisting of five  $\beta$ -strands ( $\beta$ 3,  $\beta$ 4,  $\beta$ 5,  $\beta$ 6, and  $\beta$ 7). Though the domain is not directly involved in the proteolytic mechanism, it serves as a structural scaffold for the N-terminal protease domain and shields the C-terminal  $\beta$ -strand. The C-terminal domain likely maintains the structural integrity of the protein until the final  $\beta$ -strand carrying the cleavage site (Cys168) is translated, which then enables the catalytic domain to acquire its active conformation, thus allowing cleavage of the peptide bond at the C-terminus of N<sup>pro</sup>. In the crystal structure, all four residues of the TRASH motif are located at one end of the  $\beta$ -sheet, consistent with previous biochemical data [53]. The zinc-binding site consists of a loop and a  $\beta$ -hairpin that contribute the ligand Cys112 and the other three ligands Cys134, Asp136, and Cys138, respectively. However, a bound zinc atom is missing from

the N<sup>pro</sup>-Δ17N structure. Instead, a disulfide bridge was formed between Cys112 and Cys134. We surmise that the zinc atom escaped the binding site in the absence of a stable reducing agent in the crystallization conditions, which in turn allowed formation of a disulfide bridge.

Several zinc-binding sites in proteins have been described in the literature. The role of zinc in these proteins may be catalytic or structural. The TRASH motif was first described as a novel sequence motif for genes involved in copper homeostasis, and was predicted to have a treble clef fold [56]. The treble clef fold consists of a  $\beta$ -hairpin at the N-terminus and an  $\alpha$ -helix at the C-terminus that contribute two ligands each for zinc-binding [83]. However, the zinc-binding site in N<sup>pro</sup> does not resemble the treble clef fold or any other common zinc-finger motifs. It is close to a zinc ribbon in that the zinc-binding site contains a three-stranded anti-parallel  $\beta$ -sheet. Unlike a typical zinc ribbon that consists of two zinc knuckles (short  $\beta$ -strands connected by a turn) that each contribute two ligands, in N<sup>pro</sup> one ligand comes from the loop connecting  $\beta$ 3 and  $\beta$ 4, and the other three from the strands  $\beta$ 5 and  $\beta$ 6 and the loop connecting them. Since the zinc-binding residues in N<sup>pro</sup> constitute a modified form of the TRASH motif, i.e., C-X<sub>21</sub>-C-X-D-X-C, it is not clear whether the zinc-binding motif in N<sup>pro</sup> forms a subset of the TRASH motif or a new zinc coordinating sequence motif.

An intact zinc-binding site in N<sup>pro</sup> is required for binding IRF3 and targeting it for proteasomal degradation in the host cell [53]. Similar to pestivirus N<sup>pro</sup>, rotavirus NSP1 and herpes virus ICP0 also inhibit IRF3 activation by binding to IRF3 and targeting it for proteasomal degradation [79,80]. Both proteins also contain a conserved zinc-binding RING-finger motif (Cys<sub>3</sub>HisCys<sub>4</sub>) at their N-termini, which has been suggested to act as an E3 ubiquitin ligase. The E3 ligase transfers ubiquitin from the E2 conjugating enzyme to the substrate protein via direct interaction with the substrate protein. Although N<sup>pro</sup>

contains a zinc-binding motif, the structure of the zinc-binding site is rather different from the classical zinc-fingers and does not resemble the RING-finger motif, the typical fold of E3 ubiquitin ligase. Thus, it seems unlikely that N<sup>pro</sup> functions as an E3 ubiquitin ligase and N<sup>pro</sup> may regulate the IRF3 degradation via a mechanism different from that of rotavirus NSP1 and herpes virus ICP0.

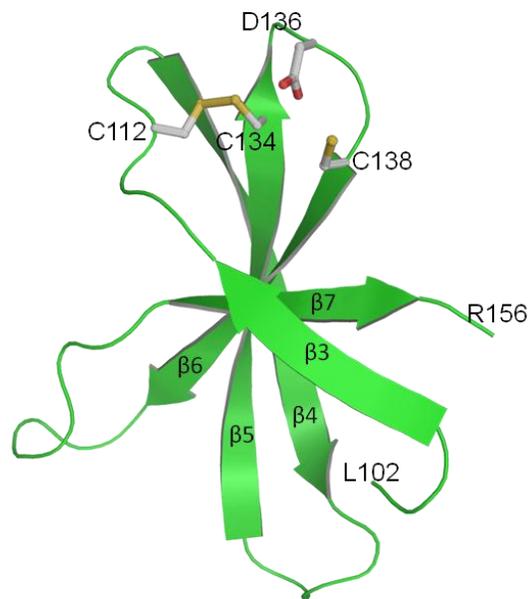


Figure 2.7: Zinc-binding domain of N<sup>pro</sup>.

Ribbon representation of the zinc-binding domain of N<sup>pro</sup>. The TRASH motif (Cys112-Cys134-Asp136-Cys138) is shown in ball-and-stick models on one end of the  $\beta$  sheet. Zinc is missing in the structure, and Cys112 and Cys134 form a disulfide bond, probably due to oxidizing conditions in the crystallization solution.

## CHAPTER-3

### Crystal Structure of N<sup>pro</sup>-Cys168Ala

#### 3.1 INTRODUCTION

Crystal structure of wild-type CSFV N<sup>pro</sup> provided important details about the overall fold of N<sup>pro</sup>, the cysteine protease and the zinc-binding domains, the protease active site and the possible role of Glu22 on the activity of the protease. However the electron density around the enzyme's nucleophile Cys69 could not be determined due to intrinsic flexibility of the encompassing loop [Chapter 2]. One half of the catalytic site of the enzyme could not be built and consequently the interactions between the substrate and the active site of the cysteine protease could not be determined. It was also unclear, from the crystal structure if the side chain sulfur of Cys168 (P1 site) had a role in the catalysis during the cleavage. Furthermore, cysteine protease dyad is located on one end of the domain-I of N<sup>pro</sup> and its surface is solvent exposed. Thus, the enzyme could potentially cleave substrate peptides in *trans*. However, in the absence of the C69-loop density in the structure it is difficult to ascertain how accessible the active site pocket is to external peptides and which orientation of the substrate peptides bind in the active site pocket. We determined the 1.6 Å crystal structure of a cleavage site mutant of N<sup>pro</sup>, described below. The structure provided a high-resolution picture of the N<sup>pro</sup> active site and the interactions that stabilize the substrate binding pocket and aid in the cleavage reaction at its C-terminus.

### 3.2 EXPERIMENTAL DESIGN

Crystal structure of wild type N<sup>pro</sup> represented the product bound form of the cysteine protease. In order to visualize the catalytic site of N<sup>pro</sup> and to study the mechanism of action of the cysteine protease, we tried to crystallize the substrate bound form of the enzyme. The cleavage site for the N<sup>pro</sup> cysteine protease is the peptide bond between its own C-terminus Cys168 and Ser169, the N-terminus of the core protein (2<sup>nd</sup> translated protein of the pestivirus genome). We designed a total of eight catalytic (Cys69, His49) and cleavage site (Cys168) mutants of the protease which were proposed to ablate its catalytic activity and incorporated up to the first four amino acids of the P' site of the substrate peptide, the P1'-P2'-P3'-P4' sequence <sup>169</sup>SDDG<sup>172</sup>, at the C-terminus (table1). *All mutants were designed on the background of the N-terminal deletion construct N<sup>pro</sup>-Δ17N, here on referred to simply as N<sup>pro</sup>.* We hypothesized that the mutations would inactivate the cysteine protease and reduce the conformational flexibility of the active site pocket; further, inclusion of the substrate peptide could trap the enzyme in a catalytically favorable conformation giving us insight into the substrate binding and catalysis by N<sup>pro</sup>.

None of the substrate-bound active site mutants of N<sup>pro</sup> was crystallized even though they were all highly soluble in solution (except N<sup>pro</sup>(H49L)-SDDG). We were finally able to grow crystals of a cleavage site mutant of N<sup>pro</sup>, N<sup>pro</sup>-C168A that was originally designed with the bound substrate (construct 6 in table-3.1). We however, did not see the bound substrate in the crystal structure of the mutant (discussed in detail below). Previous studies have shown that mutation of the cleavage site Cys168 to Glu abrogates the proteolytic activity of N<sup>pro</sup> *in vitro*. This is in agreement with the absolute conservation of the terminal Cysteine-168 over all pestiviral N<sup>pro</sup> sequences. Following the same logic, we mutated Cys168 to Ala expecting the mutant to be catalytically

inactive as well. However, we observed that Cys<sup>168</sup> to Ala mutation did not diminish the auto-proteolytic activity of the protease [discussed in Chapter 4] and the predicted catalytically inactive construct N<sup>pro</sup>-C168A-SDDG resulted in the cleavage of the C-terminal 4 aa peptide following amino acid 168, leading to the truncated construct N<sup>pro</sup>-C168A. We were able to determine the structure of this cleavage site mutant construct to 1.6 Å resolution using molecular replacement using the wild-type crystal structure as a model.

Substrate Bound Mutants of CSFV-N <sup>pro</sup>		
S.No.	Name of mutant	Observed Catalytic Activity
1	N <sup>pro</sup> -SDDG	Active
<i>Catalytic Site Mutants</i>		
2	N <sup>pro</sup> (C69A)	Inactive
3	N <sup>pro</sup> (C69A)-SDDG	Inactive
4	N <sup>pro</sup> (C69A)-S	Inactive
5	N <sup>pro</sup> (H49L)-SDDG	Inactive
<i>Cleavage Site Mutants</i>		
6	N <sup>pro</sup> (C168A)-SDDG*	Active
7	N <sup>pro</sup> (C168A)-S	Active
8	N <sup>pro</sup> (C168A)	Active

\* Crystallized mutant.

Table 3.1: List of designed catalytic and cleavage site mutants of cysteine protease of N<sup>pro</sup>.

### **3.3 MATERIALS AND METHODS**

#### **3.3.1 Plasmid constructs**

N<sup>pro</sup>-Cys168Ala was amplified from the N-terminal deletion mutant, N<sup>pro</sup>-Δ17N, using the oligonucleotide primers 5' ggcagccatatgggagtgagggaaccggtatac 3' (forward) and 5' gtggtgctcgagttagccatcatcagaggcactggtaac 3' (reverse) and the DNA fragment was subcloned into the *NdeI* and *XhoI* restriction sites of pET-28b to obtain pET-6xHis-throm-N<sup>pro</sup>-C168A with a cleavable 6x-His tag on the N-terminus of the protein.

#### **3.3.2 Overexpression and purification of N<sup>pro</sup>-Cys168Ala**

N<sup>pro</sup>-Cys168Ala was purified in the same way as the wild type N<sup>pro</sup>-Δ17N with the only difference being the addition of 10% glycerol to the protein solution during thrombin cleavage of the N-terminal 6xHis-tag. Glycerol was necessary to prevent the protein from precipitating during the cleavage reaction with thrombin at room temperature.

#### **3.3.3 Crystallization of N<sup>pro</sup>-C168A**

Following purification, protein was exchanged into 20 mM Tris-Cl (pH 8.0), 100 mM NaCl and 5 mM β-mercaptoethanol and concentrated to 4.5 mg/ml. Crystallization conditions for N<sup>pro</sup>-C168A were screened in the same way as for the wild type N<sup>pro</sup>-Δ17N. Initial crystallization trials were conducted by the vapor diffusion method with sitting drops in 96-well Inteli-plates using Rigaku's Phoenix-RE™ liquid handling robot. Small crystals appeared within a week. Crystals were further optimized using 24-well LINBRO™ plates by the hanging drop vapor diffusion method with a drop volume of 1 μl and reservoir volume of 500 μl. N<sup>pro</sup>-C168A crystallized in multiple conditions containing PEG3350, PEG8000, ammonium sulfate and lithium sulfate as a precipitate within a pH range of 6.5 to 7.5. Of all the screened conditions we obtained best diffracting crystals in

25% PEG3350, 0.1 M Hepes pH 7.5, and 0.2 M Li<sub>2</sub>SO<sub>4</sub> (condition 1), and 30% w/v PEG8000, 0.1 M Sodium cacodylate trihydrate, pH 6.5, and 0.2 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> (condition 2).

### 3.3.4 Data collection

Crystals of N<sup>pro</sup>-C168A were cryo-cooled at 100K using paratone as cryo-protectant. After the initial round of indexing and strategy determination on HKL2000, a native data set was collected on the home source on Rigaku RAXIS-IV<sup>TM</sup> image plate detector at the Sealy Center for Structural Biology at UTMB. We collected 100 degrees of data with 1° oscillation. Data was indexed and scaled with the HKL3000<sup>TM</sup>, the software package with the triad of Denzo, XDisplayF and Scalepack programs to visualize and interpret X-ray diffraction images [86,87]. The crystals diffracted to 1.6 Å resolution, belonged to the space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> with a=40 Å, b=59 Å, c=74 Å. Matthews coefficient ( $V_M$ ), was calculated to be 2.58 Å<sup>3</sup>/Da and the solvent content of the crystal lattice was determined to be 52.25% [83,84]. The asymmetric unit of the unit cell was a monomer. A representative diffraction image for the crystal is shown in figure-3.1.

### 3.3.5 Structure determination

Molecular Replacement solution with the structure of wild type N<sup>pro</sup> as reference model was calculated using the AutoMR module in Phenix (Python-based Hierarchical Environment for Integrated Xtallography), the macromolecular structure determination suite. AutoMR employs Phaser's Maximum-Likelihood based MR algorithm that generates an ensemble of models, calculates the rotation and translation functions, checks for improper packing of potential solutions and finally does a rigid-body refinement to generate the final molecular replacement solution [90].

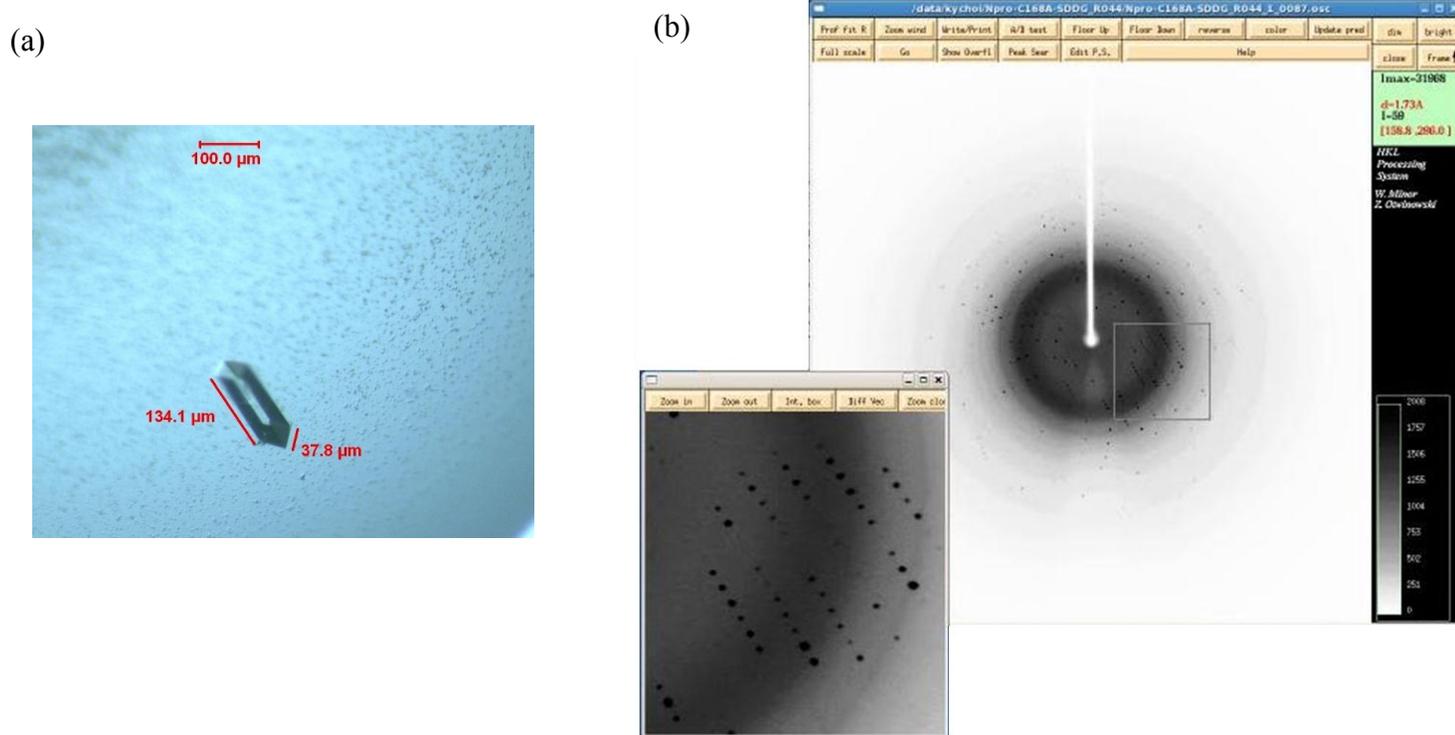


Figure 3.1: Diffraction from N<sup>pro</sup>-C168A crystals.

(a) Typical crystal of N<sup>pro</sup>-C168A. Crystal dimensions and scale are indicated on the image. (b) Single representative frame from the diffraction data collected at the home source at UTMB. The crystals diffracted to 1.6 Å with low mosaicity (~0.6). Inset shows well separated diffraction spots in a region of the frame.

Based on the molecular replacement solution obtained, an initial model was built using Autobuild wizard in the Phenix package [91]. This model was further refined by iterative cycles of manual building, positioning and refinement with O [81] and phenix.refine [92]. During refinement, strong density ( $> 8\sigma$ ) in the Fo-Fc map at the center of the coordination complex indicated the presence of a metal ion at the site. We hypothesized that the metal ion could be zinc (+2) based on the tetrahedral geometry of the coordination complex, the identities of the coordinating groups, viz., His49, His74, Cys69 and carbonyl group from Ala168, and the target distances in the crystal structure between these groups and the metal ion. Single-wavelength anomalous dispersion (SAD) data on Zn was collected on the tunable synchrotron X-ray beamline at the Center for Advanced Microstructures and Devices (CAMD) in LSU, Baton Rouge, Louisiana. Following a fluorescence scan to verify the presence of Zinc, diffraction data was collected at the peak wavelength of 1.2823 Å (9669eV). High redundancy data was collected to a resolution of 2.2 Å. Data was indexed to the orthogonal spacegroup  $P2_12_12_1$ . Unit cell dimensions were identical to those of N<sup>pro</sup>-C168A. Anomalous difference Fourier map was calculated using Phenix map tools to determine the location of zinc in the structure of N<sup>pro</sup>-C168A. The final R and R<sub>free</sub> factors for the N<sup>pro</sup>-C168A structure are 18 and 21.3, respectively. Ramachandran plot shows that 95% of amino acids are in the most favored and additional 5% in the generously allowed region. Data collection, phasing and refinement statistics for the crystal structure of N<sup>pro</sup>-C168A are given in Table-3.2.

## 3.4 RESULTS

### 3.4.1 Overall structure of N<sup>Pro</sup>-C168A

The overall structure of N<sup>Pro</sup>-C168A can be superimposed on the wild-type crystal structure with a R.M.S.D of 0.38 Å for 136 C $\alpha$  atoms. Electron densities of all amino acids from the N-terminal Met18 to Ala144 (including the loop containing the catalytic cysteine, Cys69) and from Arg150 to Ala168 could be clearly determined. The loop between strands  $\beta$ 6 and  $\beta$ 7, containing residues Lys145 through Pro149 in the C-terminal domain was disordered and could not be built (Figure-3.2). Introduction of the Cys168Ala mutation generated an additional screw axis in the crystal lattice (P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> compared to P2<sub>1</sub>2<sub>1</sub>2 for the wild type) and new crystal packing interactions at the active site that contributed to the stabilization of the loop carrying the nucleophilic Cys69. This stabilizing interaction involves the formation of a tetrahedral zinc coordination site by the side chains of Cys69, His49 and the C-terminal carboxylic group of one monomer, and the side-chain of His74 from a second monomer. The typical coordination distances for zinc from main-chain carbonyl group, histidine and cysteine in a metal coordination complex are 2.07 Å, 2.03 Å, and 2.31 Å, respectively [93–95], which are comparable to the distances we observe in our structure. Thus, SAD data at the zinc absorption edge were collected, and anomalous difference Fourier map was calculated to verify the identity of the metal as zinc (Figure-3.3). The difference map showed a clear Zn density at  $> 3.5\sigma$ . Evidence that the coordination of Zn was a result of crystal packing and not a physiologically relevant conformation of the protein came from the fact that His74, which forms the fourth coordination group, is not conserved among pestiviral N<sup>Pro</sup>. In addition, mutation of this residue does not affect the catalytic activity of the protease [54]. It is conceivable that the zinc in the structure was salvaged from the crystallization buffers.

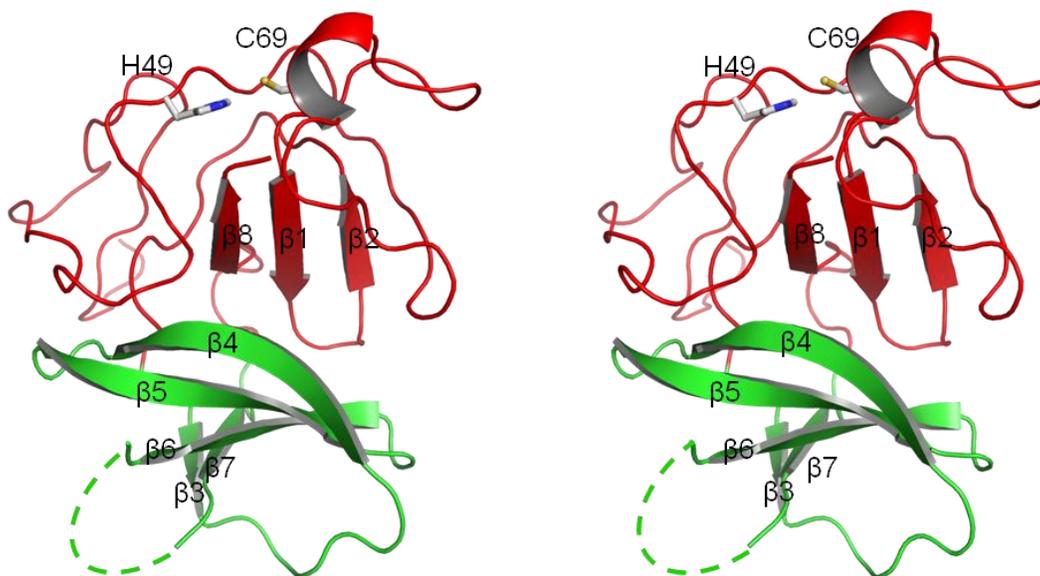


Figure 3.2: Stereo image of structure of  $N^{\text{pro}}$ -C168A.

Catalytic domain-I is in red, zinc-binding domain-II in green. Catalytic dyad His49 and Cys69 are depicted in ball-and-stick models. Dashed lines indicate the missing electron density in domain-II.

We wondered whether the zinc coordination seen at the catalytic site was purely a result of crystal packing in this particular mutant of  $N^{\text{pro}}$ , or it is related to the post-cleavage form of  $N^{\text{pro}}$ . One of the differences between the C168A mutant that has a zinc atom and the wild-type that does not coordinate a zinc atom is that  $N^{\text{pro}}$ -C168A-SDDG construct underwent a single *cis*-cleavage reaction to generate the  $N^{\text{pro}}$ -C168A mutant, while the wild-type  $N^{\text{pro}}$  is terminated following the carboxy-terminal Cys168 and therefore does not require auto-cleavage. We wondered if the process of enzymatic *cis* catalysis could somehow bring about the interactions seen at the active site seen in the C168A crystal structure. Thus we designed a catalytically active construct of  $N^{\text{pro}}$  with four amino acids of the uncleaved substrate peptide (P1'-P4') attached after the cleavage site, viz.  $N^{\text{pro}}$ -

SDDG (construct 1 in table-1) that would allow for the *cis*-cleavage reaction to occur upon translation of N<sup>pro</sup> in vitro.

	N <sup>pro</sup> -Δ17N -C168A (MR)	Zinc Anomalous
<b>X-ray data collection</b>		
Wavelength (Å)	1.5418	1.2823
Resolution limits(Å)	37.7-1.59	45-2.0
Total Reflections	131560	102192
Unique Reflections	24924	21202
Rsym (%)	3.7 (45.2)	5.9 (37.0)
I/σ	40.6 (2.0)	22.88 (2.73)
Completeness (%)	99.2 (87.9)	100 (100)
Space Group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> (No. 19)	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> (No. 19)
Unit cell parameters (Å)		
	<i>a</i>	41.489
	<i>b</i>	58.298
	<i>c</i>	75.464
<b>Phasing Statistics</b>		
Resolution (Å)	37.7-1.6	
FOM		
<b>Refinement Statistics</b>		
Resolution limits (Å)	37.7-1.6	45-2.0
Molecules/asymmetric unit	1	
R <sub>work</sub> /R <sub>free</sub>	18.0/21.3	
No. of atoms		
	Protien	1179
	Ligand/ion	1
	Water	262
Average B-factor (Å <sup>2</sup> )		
	Protein	15.13
	Ligand	3.05
	Water	25.18
R.M.S deviations		
	Bond Lengths (Å)	0.007
	Bond Angles(°)	1.049
<b>Model Statistics</b>		
Ramachandran Plot		
	Most favored (%)	94.8
	Additional allowed (%)	5.2
	generously allowed	0
	Disallowed (%)	0

Table 3.2: Scaling and refinement statistics for N<sup>pro</sup>-C168A native and single wavelength anomalous dispersion (SAD) data on zinc.

The resulting protein would represent the post-cleavage form of N<sup>pro</sup>. It should be noted here that the cleavage site mutant N<sup>pro</sup>-C168A also represents the post-cleavage form of N<sup>pro</sup> since the original construct (N<sup>pro</sup>-C168-SDDG) was designed by including the four amino acids of the substrate peptide following the cleavage site. The only difference between the two post-cleavage constructs of N<sup>pro</sup> (wild type and C168A) is the presence or absence of the amino acid cysteine at the P1 site respectively. We determined the structure of this post-cleavage construct of wild type N<sup>pro</sup> using molecular replacement with the wild type structure. Interestingly, the crystal lattice and crystal packing interactions of the post cleavage form of N<sup>pro</sup> is similar to those of wild-type, and did not resemble those of N<sup>pro</sup>-C168A mutant, even though both post-cleavage form and C168A constructs crystallized under similar conditions. The structures of wild type and post-cleavage N<sup>pro</sup> were essentially identical. Furthermore, similar to the wild type structure, no metal coordination site was seen at the cleavage site for post-cleavage form of wild type; the loop encompassing the nucleophilic Cys69 was once again disordered indicating that the construct was behaving identical to wild type N<sup>pro</sup>. This structure provided conclusive evidence that the zinc at the active site of N<sup>pro</sup>-C168A was in fact an artifact of crystallization and that the region was not a true zinc-binding site in N<sup>pro</sup>.

### **3.4.2 Cysteine protease of N<sup>pro</sup>**

Cysteine proteinases are enzymes that catalyze the cleavage of a peptide bond through the action of a nucleophilic thiol group in their active site. They could either be endo-peptidases or exo-peptidases depending on the location of the peptide bond that is cleaved. Other classified proteinase groups include serine, threonine, aspartate, glutamate and metalloproteinases named after the respective active site nucleophiles and/or the requirement of metal ion for the cleavage reaction to occur.

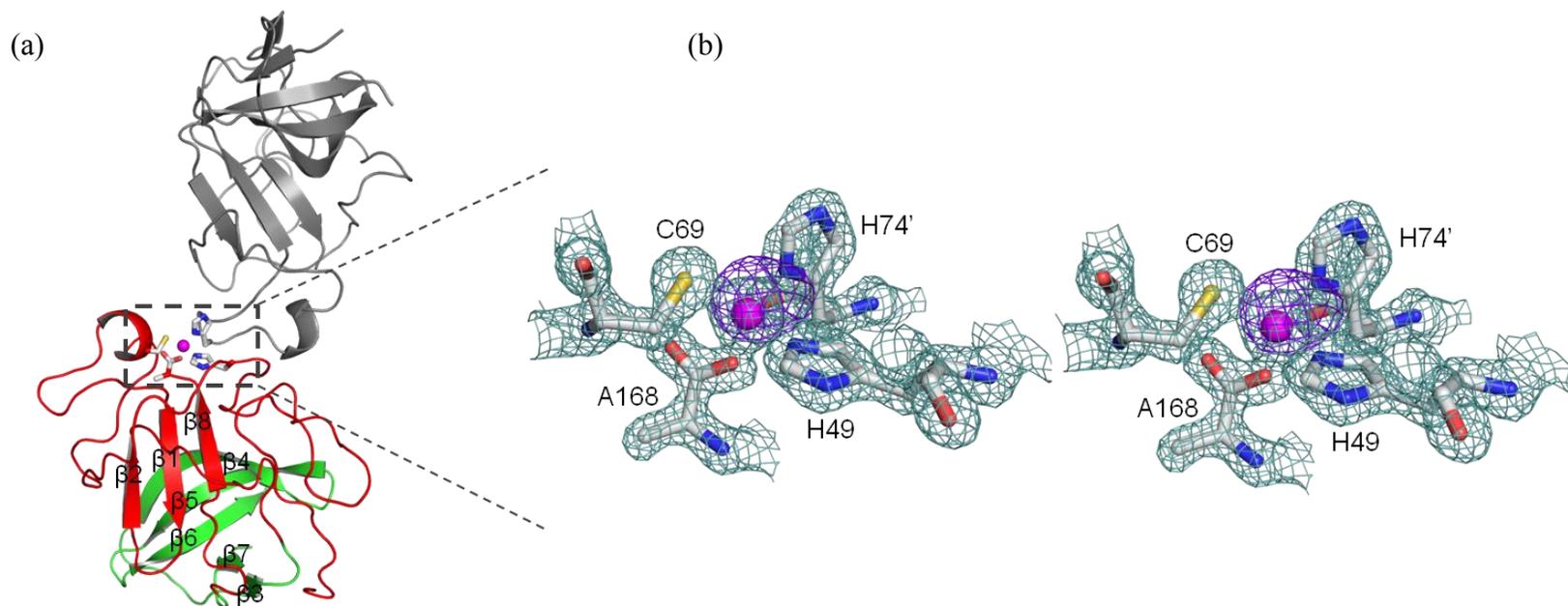


Figure 3.3: Crystal packing and zinc coordination by N<sup>pro</sup>-C168A.

(a) Structure of N<sup>pro</sup>-C168A and its symmetry mate (grey) involved in zinc coordination at the catalytic site. Amino acids involved in the coordination complex (Cys69, His49, COOH of Ala168 and His74') are shown in sticks. Zinc atom is the magenta sphere. (b) Stereo image of 2Fo-Fc electron density around the coordination site at  $2.0\sigma$  (cyan grid) superimposed with the anomalous difference map from SAD data collected at the zinc absorption edge at  $8\sigma$  (purple grid).

The mechanism of action of cysteine proteinases, illustrated below involves four basic steps viz., binding of the substrate peptide, acylation of the enzyme with catalysis of the cleavage reaction, deacylation and dissociation of the cleaved product. These steps are the same as for other enzymes, albeit with mechanistic differences and with the exception that the nucleophile in this case is the cysteine group. Peptide hydrolysis by a cysteine proteinase begins with the non-covalent incorporation of the substrate in the active site of the enzyme containing the thiolate-imidazolium ion pair, to form the enzyme-substrate complex. The ion pair exists between the active site cysteine and an adjacent histidine group in the enzyme. This step is followed by nucleophilic attack of the active site thiolate on the carbonyl carbon of the scissile peptide bond forming a tetrahedral intermediate. This tetrahedral intermediate is unstable and is quickly converted into a more stable thio-ester acyl intermediate involving a covalent bond between the carbonyl carbon and the deprotonated thiolate. Active site histidine donates its proton to the amide nitrogen promoting the conversion of the tetrahedral intermediate into the stable acyl enzyme form with the concomitant release of the amine terminus fragment. This is the acylation part of the catalysis. Deacylation reaction is brought about by the base-catalyzed hydrolysis of a water molecule. Activated water molecule attacks the acyl-enzyme and forms a second tetrahedral intermediate. Imidazole group of histidine catalyzes the breakdown of this intermediate by accepting a proton from the attacking water molecule. This releases the carboxylic acid of the product, regenerating the free enzyme and restoring the thiolate-imidazolium ion pair [96].

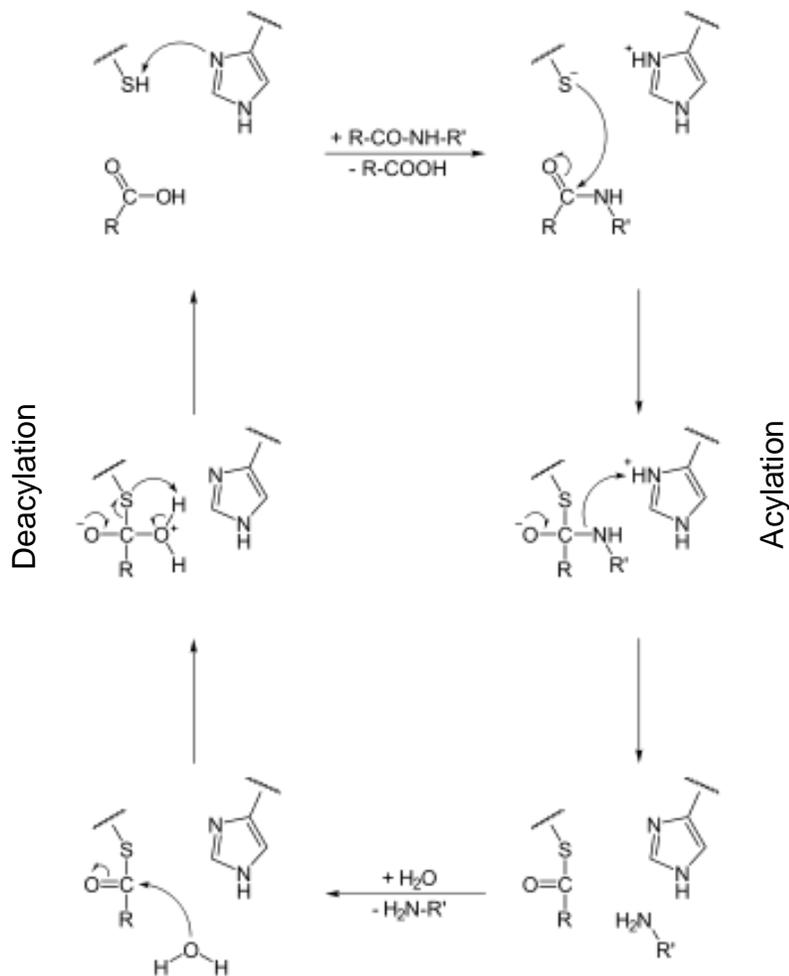


Illustration 3.1: Mechanism of action of cysteine proteinase.

Source: Wikimedia

[http://en.wikipedia.org/wiki/File:Cysteinprotease\\_Reaktionsmechanismus.svg](http://en.wikipedia.org/wiki/File:Cysteinprotease_Reaktionsmechanismus.svg)<sup>3</sup>

Papain is the most studied cysteine protease. In papain, the active site is a triad formed by amino acids Cys-25, His-159 and Asn-175. The amide oxygen of the side chain of Asn-175 is hydrogen bonded to the epsilon nitrogen (N $\epsilon$ ) of His-159 that is not

<sup>3</sup> Open Source

involved in the ion-pair formation with Cys-25 [97]. This hydrogen bond is thought to orient the histidine group for the various reaction steps during enzyme catalysis and also to stabilize the thiolate-imidazolium ion pair [98]. However, unlike in the case of serine proteinases, the presence of a stabilizing/orienting group such as Asp, Glu or Asn is not absolutely essential for catalysis of cleavage in cysteine proteinases [91,92].

The active site of the cysteine protease of N<sup>pro</sup> is a catalytic dyad formed by Cys69 and His49. This was in contradiction to the proposed catalytic triad of Glu22-His49-Cys69. As described in the previous chapter Glu22 could influence the catalysis via long distance structural effects and not directly. Unlike in the wild type structure, electron densities of residues forming the catalytic site and the substrate-binding pocket were clearly determined in the high-resolution crystal structure of N<sup>pro</sup>-C168A (Figure-3.4). In this structure, the nucleophilic Cys69 is part of a single-turn helix formed by aa 68-71. Sulfur of Cys69 is at a distance of 3.8 Å from the N $\epsilon$  group of His49 and 3.5 Å from the terminal carboxylate of Ala168. These distances are comparable to those seen in crystal structure of papain (Figure-3.4). Unlike in Papain, there was no acidic amino acid (Asp, Glu) in the vicinity of His49 that could hydrogen bond to N $\delta$  and stabilize the catalytic ion pair. However, the main chain carbonyl oxygen on Asp50, which is at a distance of 2.8 Å from the side chain N $\delta$  of His49, could function as the hydrogen bond acceptor and histidine-orienting moiety during catalysis in N<sup>pro</sup>.

The nucleophilic attack of the thiolate ion of the active site cysteine is further enabled by the presence of the 'oxyanion hole' in the active site of cysteine and serine proteases. Though there are differences in the relative importance of the oxyanion hole in enzyme catalysis between serine and cysteine proteases, the presence of hydrogen bonds between the main chain or side chain amides and the carbonyl carbon of the scissile bond contribute to the stabilization of the tetrahedral intermediate resulting from the first

nucleophilic attack [96]. In papain the oxyanion hole is defined by the main chain amides of Cys-25 and Gln-19 [99]. In N<sup>pro</sup>, we observed the oxyanion interactions between the main chain amides of Cys69, Gly67, Glu68 and the carbonyl oxygen of Ala168 (scissile carbonyl oxygen). In addition to the catalytic Cys69, Gly67 and Glu68 are highly conserved over all known pestiviral sequences indicating that the oxyanion hole formed by these amino acids could be critical for enzyme activity of N<sup>pro</sup> (Figure-3.4).

### 3.4.3 Substrate binding site of N<sup>pro</sup>

In addition to the interactions at the active site, catalytic activity of a cysteine proteinase requires the interactions between the substrate peptide (P sites) and the enzyme subsites (S sites). *A subsite is defined as the region on the enzyme surface which interacts with one amino acid residue of the substrate* [96]. Papain has 7 defined subsites, S<sub>1</sub>-S<sub>4</sub> and S<sub>1</sub>'-S<sub>3</sub>' on either side of the catalytic cysteine. These subsites interact with the corresponding peptide residues P<sub>1</sub>-P<sub>4</sub> on the amino-terminal side of the scissile bond and P<sub>1</sub>'-P<sub>3</sub>' on the carboxy-terminal side. The composition, size and shape of the enzyme subsites determine to a large extent the efficiency of cleavage of a particular substrate peptide. Thus enzyme subsites determine the specificity of a proteinase for different substrate peptides. In papain, specificity is controlled by the S<sub>2</sub> subsite, a hydrophobic pocket that accommodates primarily bulky hydrophobic or aromatic residues of the P<sub>2</sub> side chain of the substrate [100] (Figure-3.5). Besides the S<sub>2</sub> subsite, there is a lack of clearly defined residue selectivity within the active site of papain although it exhibits a predilection for hydrophobic side chains at the S<sub>1</sub>' subsite, particularly Leu and Trp amino acids [101].

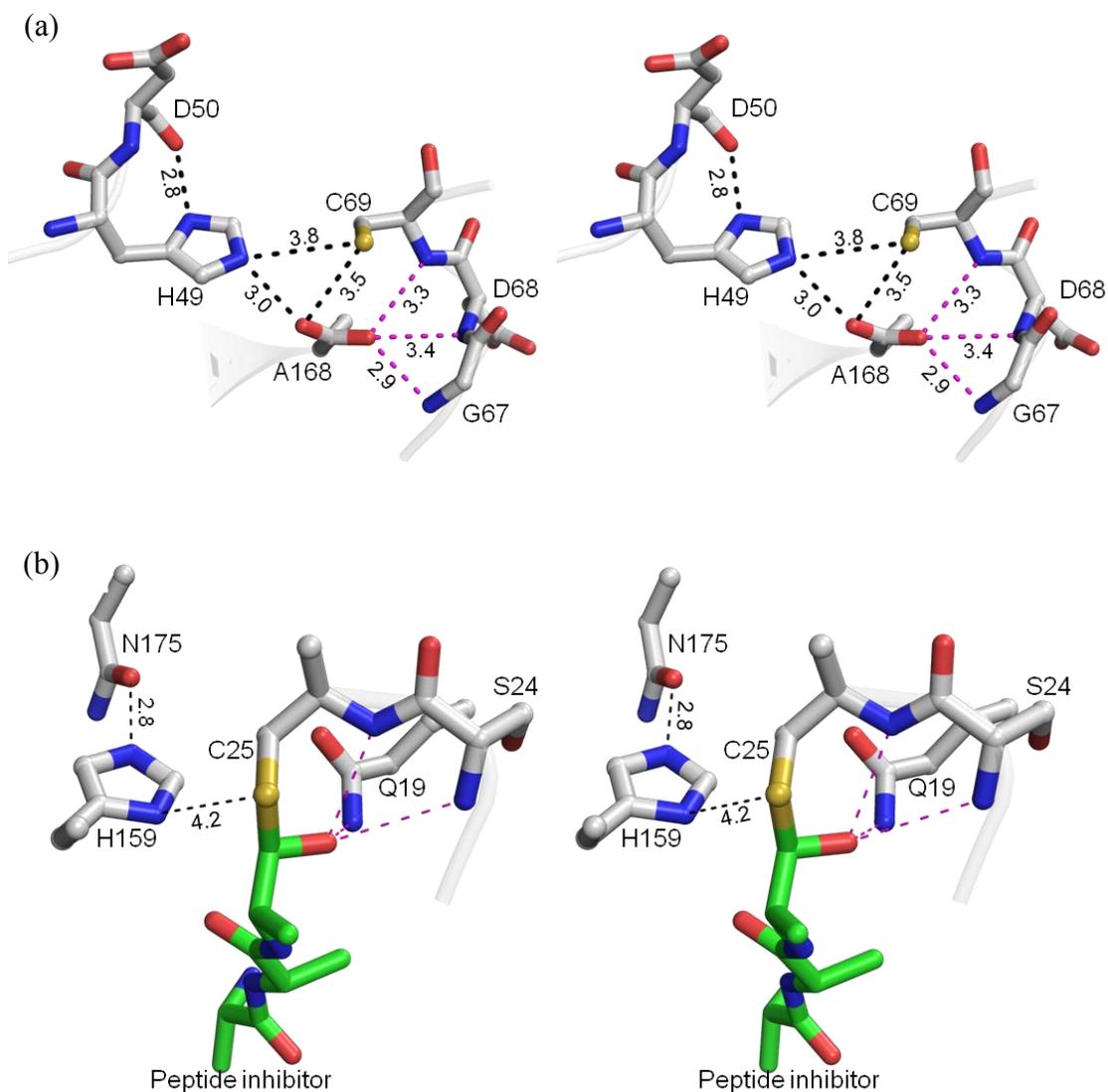


Figure 3.4: Active site of cysteine protease of  $N^{\text{pro}}$  compared to that of papain.

(a) Stereo image of the catalytic dyad of  $N^{\text{pro}}$  (C69-H49). Distances between the thiolate-imidazolium ion pair and the terminal carboxylate are indicated. Carbonyl oxygen of Ala168 forms hydrogen bonds with the backbone amide of G67, D68, C69 (oxyanion hole) stabilizing the substrate in the active site pocket.

(b) Stereo image of the active site of papain. Catalytic triad of papain is formed by C25-H159-N175. Backbone amide hydrogens of C25 and S24 along with side chain of Gln19 form the oxyanion hole in papain. The crystal structure of papain (pdb code: 1BP4) was solved with an aldehyde inhibitor covalently bound to active site cys 25 (green sticks).

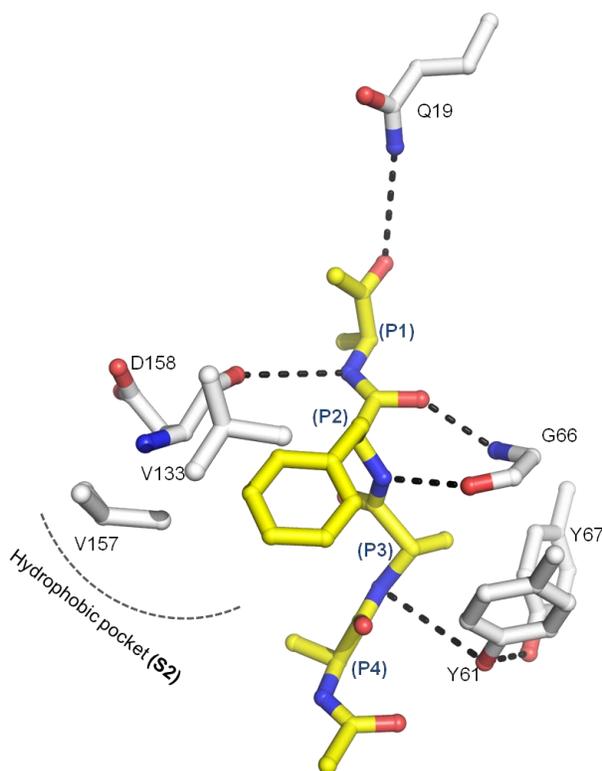


Figure 3.5: Substrate binding site of papain.

Peptide inhibitor of papain (yellow sticks) is stabilized by interactions with the substrate-sites (white sticks). Hydrophobic pocket at the S2 subsite formed by Val133 and Val157 is indicated. Dashed lines indicate stabilizing vanderwaal's and hydrogen bonding interactions between P and S sites. (PDB code: 1PAD)

In  $N^{\text{pro}}$ , the substrate peptide carrying the P sites is the terminal  $\beta$ -strand in the crystal structure (Leu163 to Thr166) on the amino side of the scissile bond, and extends into the N-terminus of the core protein on the carboxy side. The substrate-binding site (S sites) of  $N^{\text{pro}}$  is buried in the catalytic domain. In the crystal structure of  $N^{\text{pro}}$ -C168A we do not observe any defined S' sites in the active site of its cysteine protease. Thus it is likely that the S' sites have shallow pockets and do not contribute much to the cleavage specificity. This is in agreement with mutagenesis studies which showed that variation in sequence of the  $P_1'$  site of the substrate had negligible effect on the activity of the

protease. Other than amino acid proline, cysteine protease of N<sup>pro</sup> has been shown to tolerate all amino acids at the P<sub>1</sub>' position of the substrate peptide [102]. Furthermore, instances of gene duplications of the coding sequence of N<sup>pro</sup>, downstream of the structural proteins have been reported in cytopathic BVDV strains. In these cases, N<sup>pro</sup>'s protease activity was unaffected and upon cleavage N<sup>pro</sup> generated functional N-terminus of the adjoining protein in the sequence [52]. Hence the autoprotease activity of N<sup>pro</sup> is independent of sequences downstream of the scissile peptide bond.

On the other hand, the amino acid residues preceding the cleavage site (Pro162-Cys168) are highly conserved in pestiviruses. Figure-3.6a shows the variation in amino acid residues in this region observed over all known pestivirus N<sup>pro</sup> sequences. In the crystal structure of N<sup>pro</sup>, the final  $\beta$ -strand ( $\beta$ 8) formed by residues Leu<sup>163</sup>-Trp-Val-Thr<sup>166</sup> is extensively hydrogen bonded to the adjoining strand 1 ( $\beta$ 1). In addition to the obvious main chain interactions necessary to maintain the fold of the catalytic domain, the side chains of the substrate also interact with the cysteine protease domain of N<sup>pro</sup> via both hydrophobic and polar interactions and play a role in determining its activity (Fig 6b). Deletion of the terminal five amino acids inactivates the protease *in vivo* and causes the recombinant protein to be expressed in the insoluble inclusion bodies in *E. coli* (Chapter 4). This could be a result of the destabilization of the  $\beta$ -sheet that forms the skeleton of the catalytic domain causing it to misfold and abrogate the catalytic activity of N<sup>pro</sup>.

The enzyme subsites in N<sup>pro</sup> could hence extend from S<sub>1</sub> through S<sub>6</sub> that respectively interact with P<sub>1</sub>-Cys168, P<sub>2</sub>-Ser167, P<sub>3</sub>-Thr166, P<sub>4</sub>-Val165, P<sub>5</sub>-Trp164 and P<sub>6</sub>-Leu163 sites of the substrate peptide (Figure-3.6). No variation in the amino acid identity of the P<sub>1</sub> site, i.e. cysteine-168 has ever been observed over all known pestivirus N<sup>pro</sup> sequences. Achmüller *et al.* reported that mutation of Cys168 to Glu abrogated the activity of the recombinant protease [102]. However, when we tested the same mutation

in our recombinant assays, we noticed partial activity of the cysteine protease (Chapter 4).

(a)

	Sequence Variation in Substrate (P sites)						
	P7	P6	P5	P4	P3	P2	P1
CSFV-Alfort	<b><sup>162</sup>P</b>	<b>L</b>	<b>W</b>	<b>V</b>	<b>T</b>	<b>S</b>	<b>C<sup>168</sup></b>
		<b>I</b>		<b>L</b>	<b>A</b>	<b>T</b>	
					<b>S</b>		
					<b>G</b>		
					<b>Q</b>		

(b)

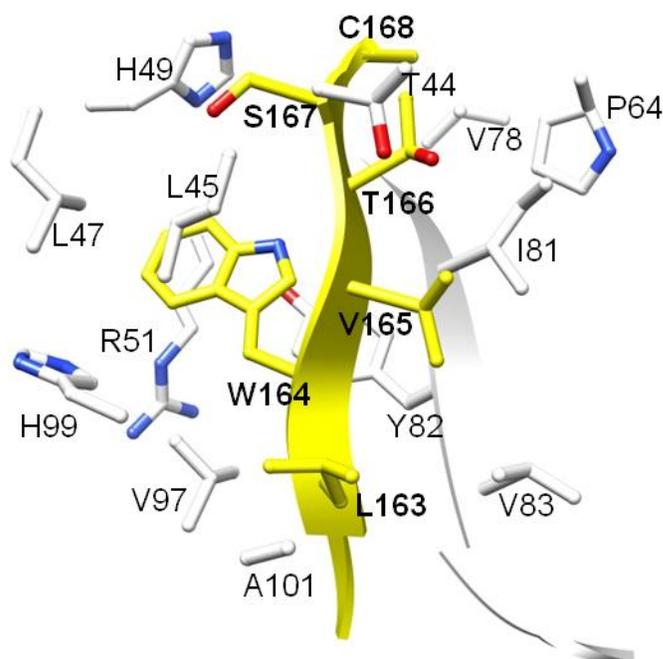


Figure 3.6: Substrate binding site of N<sup>pro</sup>.

(a) Sequence variation in the C-terminal amino acids of N<sup>pro</sup> of known pestiviruses. CSFV-Alfort amino acid sequence used as a reference is in bold. (b) Substrate binding site of N<sup>pro</sup>. The C-terminal  $\beta$ -strand ( $\beta$ 8) is shown in yellow along with the surrounding residues in the protease domain (white). The  $\beta$ -strand is also stabilized by hydrogen bonding interactions with the adjoining strand- $\beta$ 1. Cys168 (P1) is located in a hydrophobic pocket formed by Thr166, Val78, Gly80 (not shown) and Pro64. Trp164 (P4), shown to be critical for protease activity, is embedded in the hydrophobic pocket formed by the amino acids Leu45, Leu47, Tyr82, Val97, and His99.

Furthermore, in our recombinantly expressed N<sup>pro</sup>, mutation of Cys168 to Ala did not have any deleterious effect on the activity of the protease. In fact, in the C168A structure, the side chain of Ala168 is located in a shallow hydrophobic pocket formed by Thr166, Pro64, Val78 and Gly80 (S<sub>1</sub> subsite). This suggested that the geometry of the S<sub>1</sub> subsite of N<sup>pro</sup> could allow amino acid residues with small side chain groups such as glycine, alanine or serine, in addition to cysteine, for optimum activity. A long negatively charged Glu in the C168E protein thus would not fit in the subsite due to steric hindrance. Thus, although N<sup>pro</sup> does not require Cys at the cleavage site, Cys168 is conserved in pestiviruses, suggesting that the residue may have an additional function other than participating in the protease catalysis.

The S<sub>2</sub> subsite (Ser or Thr) of N<sup>pro</sup>, which interacts with P<sub>2</sub> amino acid, could also display similar preferences as the S<sub>1</sub> site i.e. small polar or apolar amino acid side chains. Large sequence variation can be seen in the P<sub>3</sub> position of the substrate peptide of N<sup>pro</sup> over the several pestivirus sequences. However in the absence of experimental data it is difficult to verify if the presence of bulky hydrophobic amino acids at this site could, once again, disrupt the S<sub>3</sub> subsite and result in improper folding of the protease domain itself, in effect rendering the protease inactive.

Specificity determining interactions in N<sup>pro</sup> seem to be accommodated by S<sub>4</sub>, S<sub>5</sub> and S<sub>6</sub> subsites. All three subsites accommodate hydrophobic side chains of the P<sub>4</sub>, P<sub>5</sub> and P<sub>6</sub> amino acid positions in the substrate. In the crystal structure of N<sup>pro</sup>, this region of the protein forms a tight hydrophobic pocket shielded from the solvent by the surrounding hydrophobic amino acids in the loop regions of the catalytic domain (Figure-3.6). Amino acid substitutions that disrupt this hydrophobic pocket would be highly deleterious to the fold of the domain and by extension its catalytic activity. Trp164 at the P<sub>5</sub> position of the substrate peptide has long been shown to be essential for the autocatalytic cleavage at the

C-terminus of BVDV-N<sup>pro</sup>. Trp164 to Ala substitution rendered the protease inactive and prevented the release of N<sup>pro</sup> from the BVDV-encoded polyprotein [46,47]. In the structure Trp164 is packed against the hydrophobic side chains of Tyr82, Leu47, Leu45 and Val97. Perhaps a more conserved substitution to another hydrophobic amino acid such as phenylalanine or isoleucine would retain the hydrophobic packing in this subsite of N<sup>pro</sup>. Further experimental evidence is required to definitively determine the specificity of the enzyme subsites in N<sup>pro</sup>. However, it is evident from the crystal structure that due to the unique fold of the cysteine protease of N<sup>pro</sup> in which the substrate peptide is itself an intrinsic part of the structure of the catalytic domain, the content of these subsites is critical for stabilization of the structure of the domain and by extension the proteolytic activity of N<sup>pro</sup>.

#### **3.4.4 N<sup>pro</sup> is a unique viral cysteine protease**

Cysteine proteases have been isolated from several organisms in prokaryotes and eukaryotes. Most commonly known cysteine protease families include various plant cysteine proteases (Papain from papaya fruit; Bromelain from pineapples etc.), Caspases (mammalian cellular cysteine proteases involved in apoptotic regulation), Calpains (Ca<sup>2+</sup> dependent non-lysosomal cysteine proteases expressed in mammals) and Cathepsins (mammalian lysosomal cysteine proteases). Viral cysteine proteases form a predominant subset of cysteine proteases in nature. They form an integral part of the viral machinery involved in pathogenesis and infection. A major subgroup of these enzymes, derived from different protease families, is involved in viral polyprotein processing. RNA viruses with single-stranded plus sense RNA genome belonging to the *Picornaviridae*, *Coronaviridae* and *Flaviviridae* families translate their genome into polyproteins which are then processed into constituent functional proteins by viral and host cell proteases.

Most of these viral processing enzymes have been found to be cysteine proteases. These processing enzymes could be *cis* and/or *trans* acting enzymes i.e. they can catalyze auto-cleavage at their own C-termini or cleave at other polyprotein sites to release functional viral proteins.

A classification of all known proteinases is available on the web-based database MEROPS (<http://www.merops.co.uk/>). Proteinases are classified into protease *families* based on homology in their structures and/or similarities in the catalytic machinery. There are 89 cysteine protease families, from C1 through C89, classified thus far; some containing just single members (since other proteins homologous to that member are not yet known). Families of enzymes are further grouped together into *clans* representing one or more families that show evidence of their evolutionary relationship by their similar tertiary structures or common sequence motifs around the catalytic residues [103].

Broadly, most viral cysteine proteases identified so far belong to two major clans based on their structural homology and evolutionary relationship with one of the major families of proteases. **Clan PA** comprises all viral cysteine proteases that are evolutionarily related to the Chymotrypsin family. The major viral cysteine protease family C3, formed by viral processing endopeptidases viz., the picornains (proteinase 3C) which are components of the polyproteins of the *Picornaviridae* group of RNA viruses, belong to this clan. Although, the nucleophile is a cysteine, these processing peptidases are evolutionarily related to the chymotrypsin serine protease family (S1) with the active site nucleophile mutated from serine to cysteine in these enzymes. The catalytic nucleophile can be either serine or cysteine arranged similar to His75-Asp120-Ser/Cys213 in the chymotrypsin sequence. All members of this clan share the common double beta-barrel fold characteristic of chymotrypsin with similar arrangement of catalytic residues at the active site. **Clan CA**, the second major clan containing viral

cysteine proteases, comprises all papain-like cysteine proteases from different protease families. The arrangement of the catalytic residues in papain is Cys25-His159-(Asn/Gln175); Asn/Gln helps to orient the imidazolium ring of the catalytic His. Peptidase families containing polyprotein processing enzymes such as murine hepatitis coronavirus papain-like peptidase, leader protease of foot and mouth disease virus (FMDV-L<sup>pro</sup>), porcine reproductive and respiratory syndrome arterivirus-type cysteine peptidase alpha (PRRSV-NSp1a), potato virus Y-type helper component peptidase (HCPro) etc. belong to this clan. These proteinases have a characteristic papain-like fold i.e. a helical bundle, the N-terminus of which carries the catalytic cysteine, and a beta-barrel containing the catalytic site histidine and glutamine groups.

A number of these viral cysteine proteinases are located at the N-termini of their respective virus-encoded polyproteins, and they all mediate a single cleavage event at their own C termini. As the first protein of the pestivirus polyprotein, N<sup>pro</sup> with its elegant auto-cleaving protease domain fits into this group of "accessory leader proteases". However, N<sup>pro</sup> does not share sequence homology with any other known proteases, and thus was assigned to its own family of cysteine proteases, C53. The newly established catalytic His49-Cys69 dyad does not align in either sequence or structure with either types of the cysteine proteases, and the unique fold of the protein reported here supports this classification.

#### **3.4.5 Mechanism of intra-molecular inhibition of cysteine protease of N<sup>pro</sup>:**

N<sup>pro</sup> cleaves the peptide bond between its C-terminal amino acid Cys168 and Ser169, the N-terminus of the subsequent core protein. Following catalysis, the carboxy terminal end of the substrate peptide is released from the active site of N<sup>pro</sup>. The amino terminal end of the substrate peptide however, remains buried within the substrate-

binding site of N<sup>pro</sup> contributing a stabilizing beta-strand to the catalytic domain. In contrast to proteases that exhibit *trans* catalytic activity that have an accessible substrate binding pocket, the catalytic domain of N<sup>pro</sup> entombs the cleaved product peptide in a way that prevents access to exogenous substrates. Hence, the cysteine protease of N<sup>pro</sup>, by virtue of its unique fold, is able to auto-inhibit its own *trans* catalytic activity.

Figure-3.7 is a comparison of the structures of two viral auto-proteinases with that of N<sup>pro</sup>. First is the Sindbis virus capsid protein (Figure-3.7b), a chymotrypsin-like serine proteinase with intramolecular product inhibition of its auto-protease activity. The capsid protein generates its own C-terminus via *cis*-cleavage. It is a protease only for one cleavage, after which it functions as a capsid protein. The crystal structure of the capsid protein shows that the short C-terminal beta-strand occupies the substrate-binding pocket in the active site [99,100]. After autoproteolysis, the C-terminus of the capsid protein remains in the substrate-binding pocket and the enzyme is thus inactive toward additional substrate. In contrast, the leader protease of FMDV, FMDV-L<sup>pro</sup>, (Figure-3.7c) which also catalyzes a single cleavage reaction at its C-terminus, does not exhibit product-inhibition. FMDV-L<sup>pro</sup> is a papain-like protease whose C-terminus is recognized by its own active site groove. However, the interactions between the substrate peptide and its active site are transient and following catalysis, the cleaved C-terminus is released from the binding site [101,102]. These proteases belong to different proteinase families and all of them undergo auto-proteolysis with or without product inhibition. The figure highlights how different viruses have adopted some of the basic proteinase folds that exist in nature for varied purposes. The fold of the cysteine protease of N<sup>pro</sup> is unique. Since the main function of N<sup>pro</sup> is the regulation of IFN $\alpha/\beta$  activation following virus infection, it could have been selected evolutionarily to function independent of the amino

acid sequences downstream of its C-terminus and for its robust mechanism of auto-inhibition.

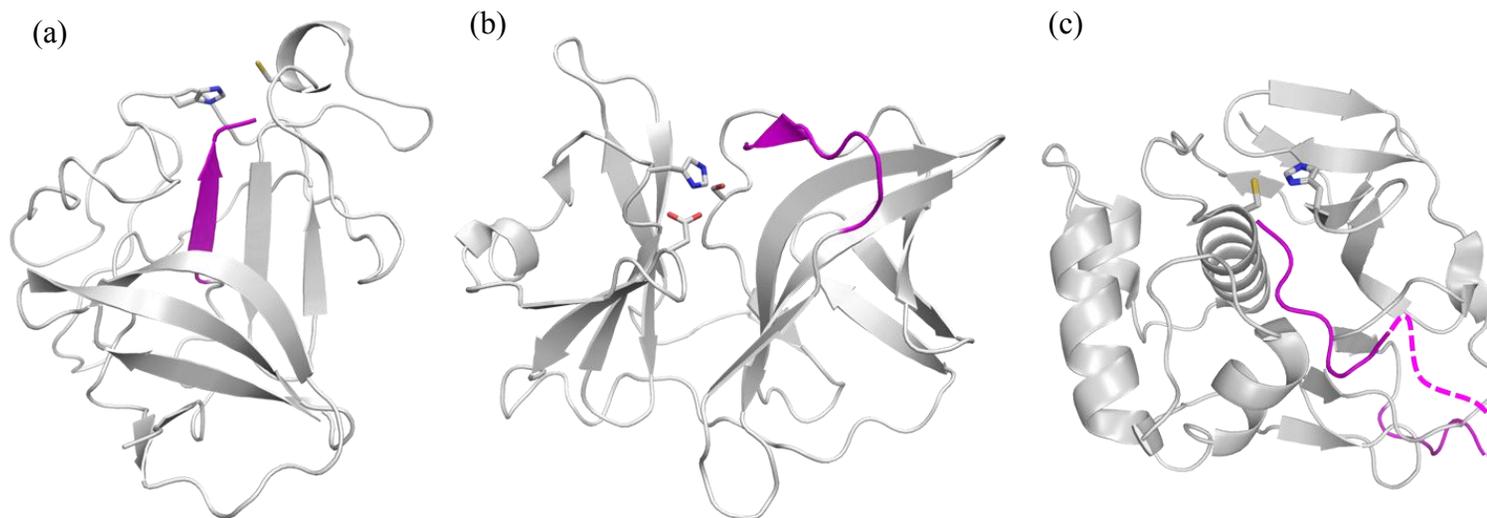


Figure 3.7: Intramolecular product inhibition of *trans* activity in viral auto-proteases.

(a) Cysteine protease of N<sup>pr</sup>, (b) Serine protease of Sindbis virus capsid protein (chymotrypsin-like fold), (c) Foot-and-mouth-disease-virus leader protease (papain-like fold). Catalytic residues are shown in sticks. Auto-inhibitory structural elements are colored magenta. In N<sup>pr</sup> and Sindbis capsin protein, the auto-inhibition of catalytic activity is permanent whereas in FMDV-Lpro, the C-terminal loop only interacts with the substrate site transiently allowing additional *trans* activity of the auto-protease. The substrate bound to Lpro (magenta) in the crystal structure is from an adjoining molecule in the asymmetric unit. Dotted line shows the modeled arrangement of the substrate peptide in the protein during catalysis of cleavage.

## CHAPTER-4

### Catalytic Activity of the Cysteine Protease of N<sup>PRO</sup>

#### 4.1 INTRODUCTION

Pestiviral N<sup>PRO</sup> is an auto-protease that co-translationally catalyzes a single cleavage reaction at its own C-terminus. *Trans* activity of N<sup>PRO</sup> has never been demonstrated and hence the protease was thought to auto-inhibit its own activity. Limited proteolysis experiments on N<sup>PRO</sup> with exogenous proteinases like trypsin and chymotrypsin showed that the C-terminus of the protein is protected from digestion [53]. Since the C-terminus contains the cleavage site of N<sup>PRO</sup>, we hypothesized that following cleavage, the substrate of N<sup>PRO</sup> (C-terminus) is buried in the active site, thus blocking entry to external substrate peptides and auto-inhibiting its *trans* activity. In order to study the catalytic activity of the cysteine protease of N<sup>PRO</sup> we had two aims:

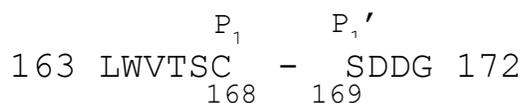
1. to restore the *trans* activity of N<sup>PRO</sup> using C-terminal deletion mutants and test its activity *in vitro*, and
2. to isolate the minimal catalytic domain of N<sup>PRO</sup>.

#### 4.2 TRANS ACTIVITY OF N<sup>PRO</sup>

##### 4.2.1 Experimental design

##### 4.2.1.1 Study of the *trans* activity of CSFV N<sup>PRO</sup> using C-terminal deletion mutants

N<sup>PRO</sup> cleaves the Cys168-Ser169 bond in the viral polyprotein. The substrate site of CSFV N<sup>PRO</sup> is the amino acid sequence:



The cleavage site Cys168, and the P1' site Ser169, are identical in all pestivirus sequences. In addition, residues preceding the cleavage site (P2-P3-P4-P5-P6) are highly conserved. Specifically, it has been shown that Trp164 to Ala mutation in BVDV and deletion of C-terminus from aa161-168 in CSFV abolished the auto-proteolytic activity of N<sup>pro</sup> disrupting its release from the polyprotein [46,47]. Trp164 is the fifth residue preceding the cleavage site; it is not known whether N<sup>pro</sup> has specificity for this position. Alternatively, mutation of Trp164 to Ala in N<sup>pro</sup> may influence folding of the protein, such that the enzyme is no longer able to carry out catalysis. We had shown using limited proteolysis experiments with trypsin, chymotrypsin and carboxypeptidase-C that the C-terminus of N<sup>pro</sup> is inaccessible to proteolytic digestion. We investigated the possibility that this region is buried in the active site pocket of N<sup>pro</sup> and hence acts as an auto-inhibitor while itself is protected from proteolytic digestion. If this is the case, it might be possible to restore the *trans* catalytic activity of N<sup>pro</sup> by the removal of this blocking peptide from the active site pocket. Towards this end and to check the substrate specificity of the cysteine protease of N<sup>pro</sup>, we designed two C-terminal deletion mutants of CSFV N<sup>pro</sup>, viz., N<sup>pro</sup>-Δ4C and N<sup>pro</sup>-Δ5C lacking the last four and five amino acids respectively. We also synthesized substrate peptides carrying the natural cleavage site of N<sup>pro</sup> with the following two sequences:

Peptide 1: WVTSC-SDDG

Peptide 2: VTSC-SDDG

The peptides were synthesized in UTMB's Peptide Synthesis core. N<sup>pro</sup>-Δ4C or N<sup>pro</sup>-Δ5C would be incubated with the substrate peptides, and the aliquots analyzed by MALDI-MS. In the event of a cleavage reaction we expect to find cleaved peptide products of

594Da and 392Da, and 408Da and 392Da from Peptide-1 and Peptide-2 respectively. In the absence of catalytic activity we expect to see MS peaks corresponding to full length uncleaved peptides of 968Da and 783Da, respectively. Mass spectrometry was chosen for activity assays because it is a direct and fast method to determine a cleavage site. Full-length WT CSFV N<sup>pro</sup> would be used as a negative control. If N<sup>pro</sup>-Δ4C and N<sup>pro</sup>-Δ5C had indeed showed *trans* activity, variations in the peptide sequences would be tested to determine the substrate specificity of N<sup>pro</sup>. Further, additional C-terminal truncation mutants would be constructed, and their catalytic activity measured in order to identify the minimum protease domain.

#### 4.2.2 Results

The two C-terminal deletion mutants of CSFV N<sup>pro</sup> were generated from the full-length wild type construct of N<sup>pro</sup> using site directed mutagenesis to insert a stop codon after amino acids Leu163 and Trp164 for N<sup>pro</sup>-Δ4c and N<sup>pro</sup>-Δ5c, respectively. The constructs were expressed in *E. coli* Rosetta cells following similar purification protocol as for the wild type constructs. However, the deletion mutants could not be purified from the soluble fraction of the cell lysate irrespective of the different culture conditions tested. Following X-ray crystal structure determination, an analysis of the C-terminal substrate site of N<sup>pro</sup> showed that indeed the conserved hydrophobic residues in this region of the protease form a densely packed core of the catalytic domain. Trp164 is at the heart of this core and is involved in a number of stabilizing hydrophobic packing interactions. Moreover, the beta-strand formed by aa161-168 is an integral part of the β-sheet stabilizing the catalytic domain. The extensive hydrogen bonding interactions between the backbone and side chains of this β-sheet would be severely disrupted upon deletion of the last four or five residues (Figure-3.6 chapter 3). Hence the C-terminal deletion

mutants are intrinsically unstable in solution, which in turn cause the protein to misfold and accumulate in *E. coli* inclusion bodies.

### 4.3 *IN VITRO* PROTEOLYTIC ACTIVITY OF N<sup>PRO</sup>

From the X-ray crystal structure of wild type N<sup>PRO</sup> and that of its cleavage site mutant N<sup>PRO</sup>-C168A, we arrived at the following hypotheses:

1. Amino acids 18-168 of CSFV N<sup>PRO</sup> (N<sup>PRO</sup>-Δ17N) form the stable crystallizable core of the protein, which is sufficient for the catalysis of the cleavage reaction at the C-terminus of N<sup>PRO</sup>.
2. Cysteine protease of N<sup>PRO</sup> is a catalytic dyad formed by Cys69 and His49 and does not include Glu22 as was originally predicted. Glu22 is at a distance of ~25 Å from the active site and is not directly involved in the catalysis of the auto-cleavage. However, since Glu22 forms a salt bridge with Arg100 in the crystal structure, this could stabilize the fold of the protein and hence affect the activity of the cysteine protease via long-range interactions.
3. The structure of the proposed catalytically inactive mutant N<sup>PRO</sup>-C168A-SDDG did not have the density of the four amino acids following the cleavage site Cys168 viz., 169-SDDG-172. We hypothesized that the Cys168 to Ala mutation preserved the catalytic activity of the protease, contrary to what was expected from published data and the highly conserved nature of Cys168 in pestivirus sequences.
4. The catalytic domain-I and the zinc-binding domain-II are visually distinct in the structure suggesting the presence of independently folding regions in the protein. We thus hypothesized that since the active site is located entirely in domain-I, the

entire Zn-binding domain of N<sup>pro</sup> could be dispensable for protease activity implying that it might be possible to isolate a minimal protease domain of N<sup>pro</sup>.

Existing evidence in literature on the activity of cysteine protease of N<sup>pro</sup> is based on studies done in the context of the whole virus or in cell cultures [46,47,50,59,61]. Various fusion constructs of N<sup>pro</sup> with either the core protein or EGFP (Enhanced Green Fluorescent protein) were used to study its protease activity *in vivo*. The mutations and deletions made to study its activity were not guided by knowledge of the tertiary structure of N<sup>pro</sup> and hence incomplete. In order to test the above hypotheses based on the crystal structure of the protein and to isolate a minimal cysteine protease domain of N<sup>pro</sup>, we needed an *in vitro* protease activity assay for recombinant N<sup>pro</sup>. Since purified N<sup>pro</sup> is inactive in solution and attempts to restore its *trans* activity rendered the protease insoluble, we designed a simple reporter fusion construct of N<sup>pro</sup> and GFP. The fusion construct was expressed in *E. coli* in the soluble fraction of the cell lysate; N<sup>pro</sup> would enzymatically cleave itself from the fusion construct following translation and generate the authentic N-terminus of the target protein (in this case GFP). Using this construct, we tested the catalytic activity of recombinant N<sup>pro</sup> and the effect of various mutations on the protease-catalyzed *cis* cleavage. We used the N-terminal deletion mutant of N<sup>pro</sup> (N<sup>pro</sup>-Δ17N) in all our fusion constructs, as this was the construct that was used in our crystallization trials. The final construct contains an N-terminal His tag and a thrombin cleavage site sequence, the amino acid sequence of N<sup>pro</sup>-Δ17N followed by four residues from the core protein, GFP and finally a C-terminal his tag (shown below).

**6xHis-throm-N<sup>pro</sup> (Δ17N)-SDDG-GFP-6xHis**

Because N<sup>pro</sup> processes the self-cleavage during translation, we added the His-tag at both N- and C-terminus, so that the resulting proteins, whether they are cleaved or not, can be captured by a one-step purification using immobilized metal ion affinity chromatography (IMAC).

### 4.3.1 Materials and methods

#### 4.3.1.1 Construction of N<sup>pro</sup>-GFP fusion constructs

The N-terminal deletion mutant of N<sup>pro</sup> lacking the first 17 amino acids was amplified from the full length construct using the oligonucleotide primers 5' ggcagccatattgggagtgagggaaccgtatac 3' (forward) and 5' cgaattcggatccgcatcatcagagcaactggtaac 3' (reverse). The DNA fragment containing the N<sup>pro</sup> gene was sub-cloned into the *NdeI* and *BamHI* restriction sites of pET-28b vector. Plasmid containing GFPuv, the “cycle 3” variant of green fluorescent protein from *Aequorea victoria* (jellyfish) described by Cramer *et al.*, was obtained from Dr. Jose Barral at UTMB [108,109]. GFPuv was amplified using the oligonucleotide primers: 5' ggatccgaattcatggctagcaaaggagaagaactttcac 3' (forward) and 5' cgagtgcggccgctttgtagagctcatccatgcatg 3' (reverse) that contain *EcoRI* and *NotI* restriction sites, respectively. GFPuv was inserted into the fusion vector pET28b+N<sup>pro</sup> digested with *EcoRI* and *NotI* to generate the final N<sup>pro</sup>-GFP fusion construct: *NdeI*-N<sup>pro</sup>-*BamHI*-*EcoRI*-GFPuv-*NotI*. This construct would attach a cleavable 6xHis-tag on the N-terminus of the fusion product and another 6xHis-tag on the C-terminus of GFPuv in the fusion polypeptide. The mutants N<sup>pro</sup>-C168A-GFP and N<sup>pro</sup>-C168E-GFP were similarly built using the reverse primers 5' cgaattcggatccgcatcatcagaggcactggtaac 3' and 5' cgaattcggatccatcatcagattcactggtaaccacaatggac 3' respectively, for the amplification of

mutant N<sup>pro</sup>. Catalytic site mutant N<sup>pro</sup>-C69A-GFP was amplified from the full length construct supplied by Dr. Ruggli using the same oligonucleotide primers as for the wild type N-terminal deletion mutant ( $\Delta$ 17N) of N<sup>pro</sup>. All other mutants, viz., N<sup>pro</sup>-C69A-GFP, N<sup>pro</sup>-C69S-GFP, N<sup>pro</sup>-E22A-GFP, N<sup>pro</sup>-R100A-GFP and the double mutant N<sup>pro</sup>-E22R-R100A-GFP were generated using Stratagene's Quickchange™ site-directed mutagenesis kit following manufacturer's protocol.

#### **4.3.1.2 Construction of catalytic domain of N<sup>pro</sup>**

Forward primers used to generate both catalytic domain constructs of N<sup>pro</sup>, N<sup>pro</sup> $\Delta$ 103-156 and N<sup>pro</sup> $\Delta$ 105-155 were the same as for wild type N<sup>pro</sup> $\Delta$ 17N-GFP. Two overlapping reverse primers carrying the linker sequence and the sequence of the final beta-strand of N<sup>pro</sup> were designed for both constructs: The reverse primers for N<sup>pro</sup> $\Delta$ 103-156 were 5' caatggacagtcggtgaaattgccgctgctgccaggggctctatgtagacc 3', and 5' gtggtgctcgcgtagcaactggtaaccacaatggacagtcggtgaaattgcc 3'. The sequence was cloned into the pET28b vector digested using *NdeI* and *XhoI*. For the second catalytic domain construct N<sup>pro</sup> $\Delta$ 105-155 we used the reverse primers 5'caatggacagtcggtgaaattctgccgctgctgccctctagaggggctctatggta 3' and 5' cgaattcggatccatcatcagagcaactggtaaccacaatggacagtcggtgaaatttc 3' to generate a GFP fusion protein and inserted it into pET28b vector digested with *NdeI* and *BamHI*. All plasmid constructs were verified by DNA sequencing at UTMB's Molecular Genomics Center.

#### **4.3.1.3 Pull down assay for activity of N<sup>pro</sup>**

Plasmids carrying the cloned constructs of N<sup>pro</sup>-GFP were transformed into *E. coli* Rosetta™ (Novagen) cells. Protein was expressed in the same way as for wild type N<sup>pro</sup>. 1L LB cultures started from single colonies containing the target plasmid were grown at

37°C to an O.D. at 600 nm of 0.6-0.8, at which point they were induced with 0.5 mM IPTG and growth was continued at 18°C overnight. The harvested cells were suspended in lysis buffer (50 mM sodium phosphate (pH 7.0) and 300 mM NaCl) containing 5 mM  $\beta$ -mercaptoethanol and EDTA-free protease inhibitor cocktail tablets (Roche). Cells were lysed by sonication with a Misonix XL-2000 sonicator until the solution was clear and homogeneous. Cell lysate was centrifuged at 15,000 rpm for 15min to remove cellular debris. Soluble fraction of the cell lysate was used for the His-tag pull down assay. Soluble fraction was applied to 1ml of pre-equilibrated Talon™ (Clontech) metal affinity chromatography resin and the proteins were allowed to bind to the resin by incubation for 1hr at 4°C. Following binding, the resin was washed with Talon-wash buffer (50 mM Sodium phosphate buffer, pH 8.0 containing 0.5 M NaCl) and then with Talon-wash buffer containing 5 mM imidazole to remove non-specifically bound proteins. Finally, specifically bound His-tagged proteins were eluted by competing with 150 mM imidazole in Talon™ wash buffer. The fractions were then analyzed using SDS-PAGE to determine the cis-cleavage of N<sup>pro</sup>-GFP fusion proteins.

### 4.3.2 Results

#### 4.3.2.1 Cysteine protease of N<sup>pro</sup> is not specific for cysteine at the C-terminus (168)

The *in vitro* autocatalytic *cis* cleavage reaction of the N<sup>pro</sup> protease was analysed using the N<sup>pro</sup>-GFP fusion construct. The fusion proteins containing wild type N<sup>pro</sup>- $\Delta$ 17N and the catalytic cysteine mutant N<sup>pro</sup>-Cys69Ala were used as positive and negative controls for the activity assay. N<sup>pro</sup>- $\Delta$ 17N-GFP gave rise to two bands corresponding to CSFV-N<sup>pro</sup> at 19 kDa and GFPuv at 29 kDa. No cleavage of the fusion construct was seen for the catalytically inactive mutant N<sup>pro</sup>-C69A-GFP resulting in a 48 kDa band corresponding to 6xHis-N<sup>pro</sup>-GFP-6xHis (Figure-4.1). A qualitative analysis of the

catalytic activity of N<sup>pro</sup> was possible even by visual inspection of the eluate from Talon column. Full length GFP released by catalytically active N<sup>pro</sup> appeared bright fluorescent green whereas the fusion construct N<sup>pro</sup>-GFP had weak color in solution initially, probably owing to slower folding kinetics of GFP in the fusion construct [109–111].

Amino acid cysteine at position 168 (the cleavage site) is evolutionarily conserved over all pestivirus sequences, and was thought to be essential for the proteolytic activity of N<sup>pro</sup>. However, we did not see the density corresponding to amino acids following the cleavage site of N<sup>pro</sup> in the crystal structure of our construct, N<sup>pro</sup>-C168A-SDDG that carried four amino acids after the C-terminus of N<sup>pro</sup>. We hence tested the activity of cleavage site mutants Cys168Ala and Cys168Glu of N<sup>pro</sup>. As predicted from the crystal structure, Cys168Ala mutant of N<sup>pro</sup> was catalytically active and behaved entirely like the wild type construct. Hence we concluded that cysteine at position 168 has no role in the autocatalysis reaction and the protease of N<sup>pro</sup> is not specific for a cysteine at this position.

Achmüller *et al.* had previously shown that Cys168Glu mutant was catalytically inactive in solution. Their observations were based on the activity of refolded N<sup>pro</sup>. However, in our assay we see a partially active cysteine protease of N<sup>pro</sup> when Cys168 is mutated to Glu (Figure-4.1). A possible reduction in the cleavage efficiency in this mutant could be due to the longer charged side chain of glutamic acid compared to that of cysteine or alanine. As described in chapter-3 the geometry of the S1 subsite of N<sup>pro</sup> is such that only amino acid residues with small side chain groups such as glycine, alanine or serine, in addition to cysteine, would be accommodated. A long negatively charged Glu in the C168E protein thus would not fit in the subsite due to steric hindrance and hence in turn result in reduced activity of the protease.

#### 4.3.2.2 *Glu22 is essential for the stability of N<sup>pro</sup> and not for catalysis*

Studies have consistently shown that deletion of the N-terminal 22 amino acids of N<sup>pro</sup> or the mutation of Glu22 to Val/Leu abrogated the catalytic activity of the protease and prevented the release of N<sup>pro</sup> from the pestivirus polyprotein [50,58,61]. Sequence comparisons to other known cysteine proteases lead to the prediction that Glu22 could be a part of the catalytic triad forming the cysteine protease of N<sup>pro</sup>. However, the crystal structure of N<sup>pro</sup> clearly showed that Glu22 was not part of the active site and was in fact at a distance of 25 Å from Cys69. Crystal structure also showed ionic interactions between the side chain oxygen atoms of the  $\gamma$  carboxyl group of Glu22 and the secondary and primary amines of the side chain guanidinium group of Arg100. We predicted that these ionic interactions highly stabilize the tertiary structure of the catalytic domain of N<sup>pro</sup>, since the extensive distribution of prolines in this region of the protein prevents the formation of stable secondary structural elements such as beta-strands and alpha helices. Mutations disrupting these interactions would hence be detrimental to the fold of the catalytic domain and hence render the protease inactive. Based on the structure we also predicted that mutating Arg100 to Ala would have similar effects as Glu22 to Ala on the catalytic activity of N<sup>pro</sup>. Further, a possible reversal of the ionic interactions could be engineered by switching Glu22 to Arg and Arg100 to Glu. If the activity of the cysteine protease in the double mutant reverts to that of wild type N<sup>pro</sup>, this would corroborate our hypothesis that indeed the ionic interactions at this site are critical to the overall fold of the protein and to the activity of the cysteine protease.

We first tested the *in vitro* catalytic activity of N<sup>pro</sup>-Glu22Ala using the N<sup>pro</sup>-GFP fusion construct. Glu22 to Ala substitution, in fact showed partial catalytic activity in solution. This was in agreement with the observations reported by Ruggli *et al.* but in contrast to the complete loss of activity that was reported earlier by R umenapf *et al.*

However, it is important to note that even though a 29 kDa band corresponding to cleaved GFP was observed (the cleavage site was verified as being authentic from N-terminal peptide sequencing) a cleaved N<sup>pro</sup> (19 kDa band) was not visible in the SDS-PAGE. This suggested that N<sup>pro</sup> was quickly degraded following the cleavage reaction at its C-terminus. We believe this could be due to increased flexibility in the catalytic domain of N<sup>pro</sup> brought about by the loss of the stabilizing ionic interactions between Glu22 and Arg100 (Figure-4.1). We are in the process of determining whether a disruption of ionic interactions between Glu22 and Arg100 is indeed the cause for the loss of activity seen in the N<sup>pro</sup>-Glu22Ala mutant by mutating the residue Arg100 to Ala. Furthermore, we designed a double mutant N<sup>pro</sup>-Glu22Arg-Arg100Glu in order to reverse the orientation of charge but retain the nature of the ionic interaction between the two sites in the protein structure. Further analysis of this data is pending.

#### ***4.3.2.3 Cysteine protease of N<sup>pro</sup> requires both domains for activity***

N<sup>pro</sup> has two very different roles in the pestiviral life cycle, one being its autocatalytic release from the polyprotein and the other subversion of IFN $\alpha/\beta$  activation by targeted proteasomal degradation of IRF3. The X-ray crystal structure of N<sup>pro</sup> delineates two structurally distinct domains of the protein, the cysteine protease domain-I and the zinc coordinating domain-II. This observation prompted us to test whether the two distinct functions of N<sup>pro</sup>, i.e., autoproteolysis and subversion of interferon activation, are carried out independently by the two domains of the protein. This knowledge could be applied to the design of N<sup>pro</sup>-deletion marker vaccines for pestiviruses. One of the strategies in the design of these vaccines is substitution of the N<sup>pro</sup> gene with the murine ubiquitin gene. Following polyprotein translation the ubiquitin gene would be cleaved

after its C-terminus by cellular ubiquitin carboxyl-terminal hydrolase (UCH1) generating the authentic N-terminus of the subsequent core protein [48,107].

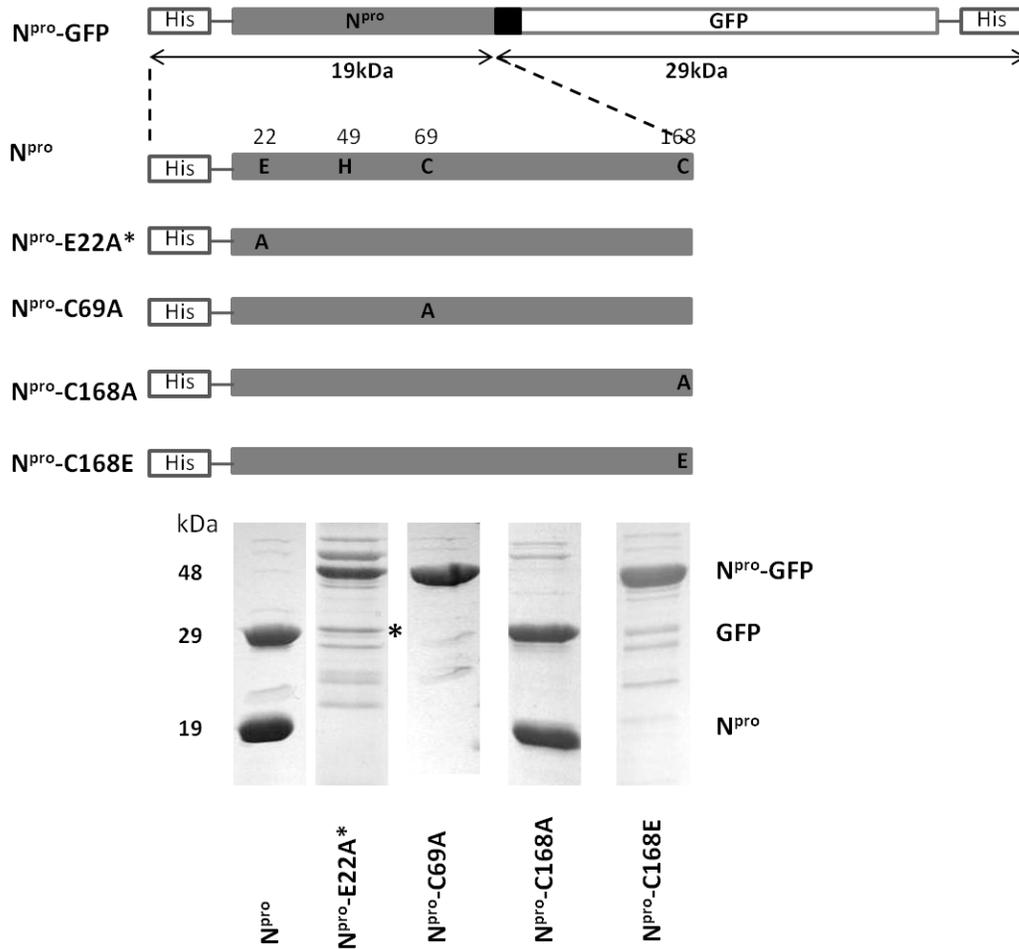


Figure 4.1: *In vitro* protease activity of N<sup>pro</sup>.

The N<sup>pro</sup>-GFP fusion construct expressed in the pET28b vector in *E. coli* consisting of N-terminal 6xHis-tag, N<sup>pro</sup>- $\Delta$ 17N, GFPuv and C-terminal 6xHis-tag is represented schematically. The E22A, C69A, C69S, C168A and C168E mutations and their locations on the sequence of N<sup>pro</sup> are also shown. Cleavage products were analysed on SDS-PAGE (bottom). Upon auto-cleavage of the N<sup>pro</sup>-GFP fusion construct (48 kDa), full length N<sup>pro</sup> (19 kDa) and GFP (29 kDa) are released. Wild type N<sup>pro</sup> and N<sup>pro</sup>-C69A served as positive and negative controls for catalytic activity, respectively. E22A construct was largely unstable in solution and showed only partial activity. C168A was fully active whereas C168E, behaved like E22A, showing partial activity.

Furthermore, viruses with deletion of the N<sup>pro</sup> gene show altered slower growth kinetics compared to wild type N<sup>pro</sup> perhaps because of the slower generation of the authentic core protein following polyprotein translation or due to a non-ideal IRES structure [113]. It is suggested that deletion of the whole N<sup>pro</sup> gene could disrupt the viral IRES, which is thought to extend into the N<sup>pro</sup> coding region [113–115]. Thus, if we could selectively retain the auto-cleavage activity of N<sup>pro</sup> and disrupt its antagonistic effect on IFN $\alpha$ / $\beta$  activation, we would be able to produce more effective vaccines against these viral diseases. Similarly, these two functions could be independently targeted via specifically designed small molecules.

In order to determine if the protease domain alone (without the zinc-binding domain) is able to carry out auto-proteolysis we designed two constructs of the catalytic domain guided by the crystal structure of N<sup>pro</sup>. In the crystal structure the catalytic domain of N<sup>pro</sup> is constituted by amino acids 17 through 101 and once again from aa 161 through 168. The C-terminal beta-strand carrying the cleavage site cysteine forms an integral part of the beta-sheet in the catalytic domain. Two protease domain constructs, N<sup>pro</sup> $\Delta$ 103-156 and N<sup>pro</sup> $\Delta$ 105-155, were designed; a short flexible linker with the sequence GSSG was inserted synching the two loops that run from domain-I to domain-II. The sequences of the two constructs along with their tertiary structure are shown in figure-4.2.

Following expression in *E. coli* (Rosetta), the two constructs were purified using Talon™ metal affinity resin similar to that of the wild type N<sup>pro</sup>. However, following purification from the soluble fraction of the cell lysate, both constructs of the catalytic domain degraded rapidly into smaller fragments (Figure-4.2). Thus, the N<sup>pro</sup>( $\Delta$ 105-155)-GFP fusion construct was additionally engineered to check the catalytic activity of the domain. The fusion protein was also purified using the same IMAC procedure as above. We did not detect a band corresponding to the size of the catalytic domain construct (~10

kDa) on the SDS-PAGE of the His-tag pull down assay, although we observed a band corresponding to the size of GFP (Figure-4.2). N-terminal sequencing of the cleaved GFP product showed that the site of cleavage was after the first glycine of the linker sequence (GSSG) built into the catalytic domain construct, indicating that the N<sup>pro</sup> construct did not have autoproteolytic activity. This also suggested to us that the catalytic domain construct had either misfolded or remained unfolded, which led to its rapid degradation and cleavage at a random site in the sequence of the GFP-fusion construct.

A closer inspection of the crystal structure and amino acid composition of the two domains of N<sup>pro</sup> provided an explanation for this result. The expected fold of the catalytic domain construct as designed with the flexible linker is shown in Figure-4.3a. However, the catalytic domain may not acquire its natural fold in the absence of the zinc-binding domain. Figure-4.3b shows the packing of amino acid side chains at the interface between the two domains in the structure of N<sup>pro</sup>. The interface between the catalytic domain-I and the Zinc-binding domain-II of N<sup>pro</sup> is formed by amino acids that are predominantly hydrophobic in nature. For instance, the domain-I interface is formed by Ile81, Val83, Phe89, Ala101, Phe158, Leu163, Val165, Thr166 and the interface on the side of domain-II is constituted by Phe105, Phe106, Ile118, Val121, Tyr129, Ile131, Val133, Ile139, Leu141. The cavity that divides the two domains of N<sup>pro</sup> in the structure is in fact inaccessible to solvent. The amino acid side chains shield one another from the surrounding solvent thus forming a strong hydrophobic pocket in the structure. In essence, the structure of N<sup>pro</sup> cannot be divided into independently folding domains without disrupting this hydrophobic pocket, which forms the core of the protein. Isolating the catalytic domain by itself would expose these hydrophobic amino acids in the core of the protein to the solvent, causing the protein to misfold, and the catalytic site is no longer formed as in the native structure. We believe that the molecular structure of N<sup>pro</sup> is

such that domain-II forms a scaffold to support the fold of the catalytic domain and together enable the catalysis of the cleavage reaction at the C-terminus. Hence the entire length of the protein sequence (barring the first 19 amino acids) would be required to maintain the catalytic activity of the cysteine protease of pestivirus  $N^{\text{pro}}$ .

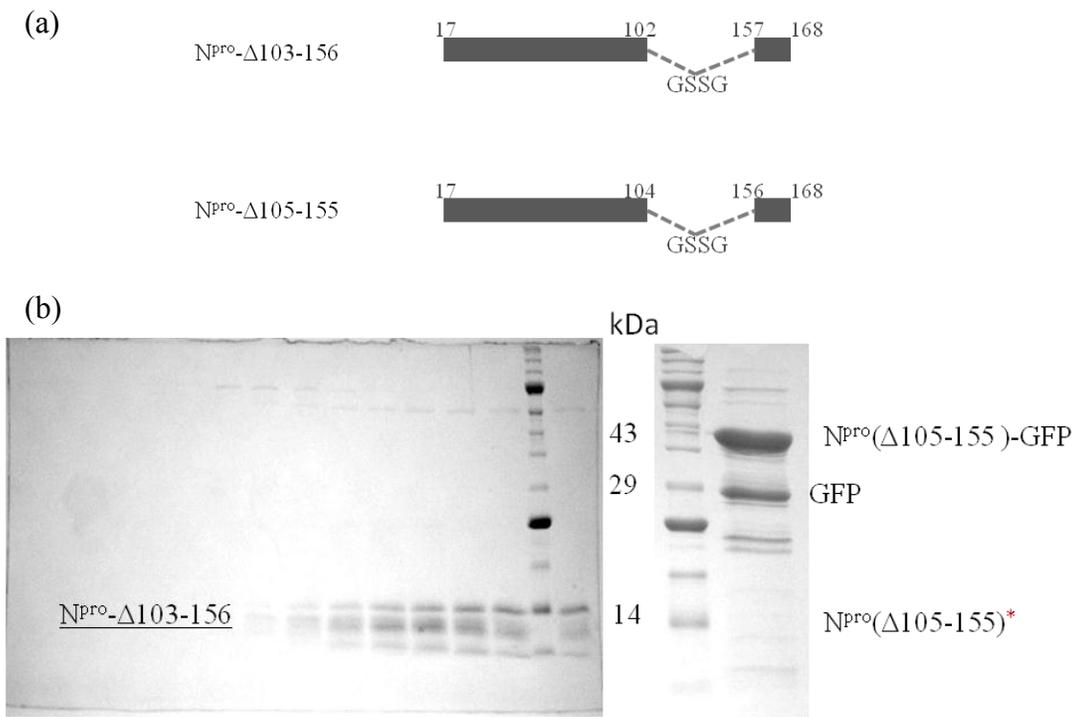


Figure 4.2: Catalytic domain construct of  $N^{\text{pro}}$ .

(a) Schematic representation of the catalytic domain constructs of  $N^{\text{pro}}$  that were designed with the flexible linker GSSG. The  $\Delta$ -numbering represents the amino acids that were deleted in the design of the two constructs. (b) SDS-PAGE of the purification of  $N^{\text{pro}}-\Delta 103-156$  construct. The 14 kDa band corresponding to the size of the catalytic domain quickly degraded into smaller molecular weight fragments visible on the gel. (c) SDS-PAGE of the His-tag pull down assay of  $N^{\text{pro}}(\Delta 105-155)$ -GFP fusion construct. \*Note the 14 kDa band corresponding to the size of the catalytic domain construct is not observed on the gel. Band corresponding to the size of GFP (29 kDa) had an unexpected N-terminus (verified using N-terminal peptide sequencing) indicating degradation of  $N^{\text{pro}}(\Delta 105-155)$

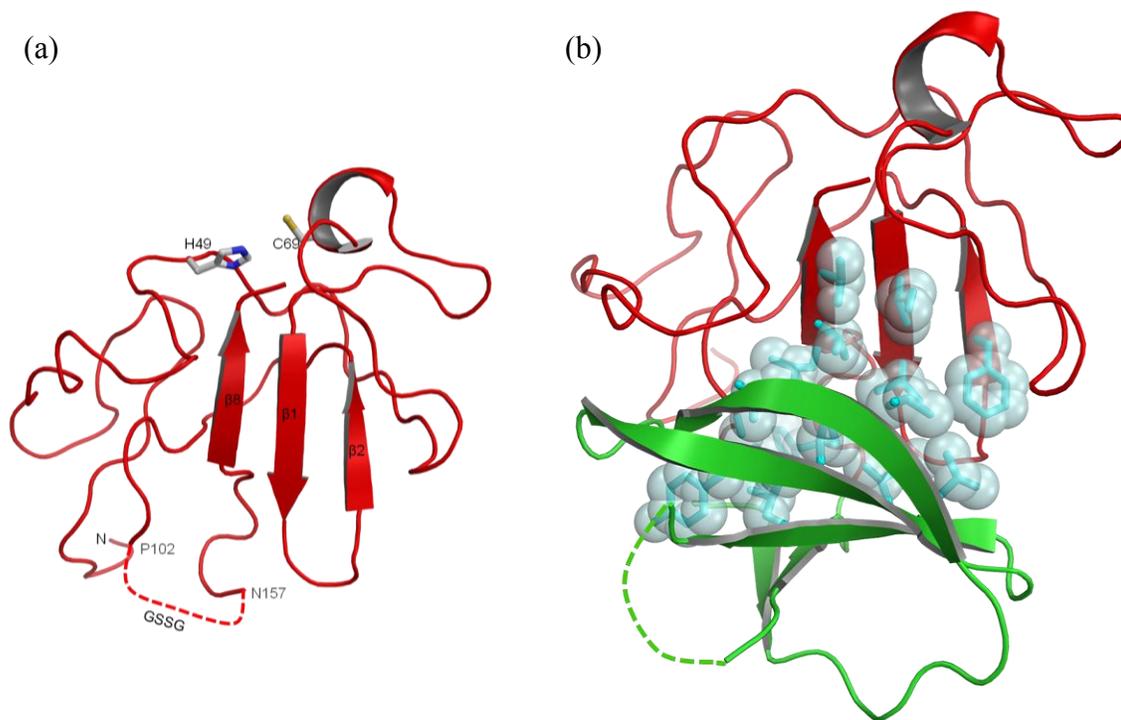


Figure 4.3: Catalytic domain of N<sup>pro</sup>.

(a) Expected fold of the catalytic domain of N<sup>pro</sup> in the construct N<sup>pro</sup>Δ103-156. (b) Hydrophobic core of N<sup>pro</sup>: Structure of N<sup>pro</sup> highlighting the packing of hydrophobic side chains at the domain interface (cyan ball-n-stick models). Deletion of either domain would expose these hydrophobic residues to solvent and disrupt the energetically favorable packing leading to misfolding. Thus the domains act as structural scaffolds for each other and not as independently folding entities.

## CHAPTER-5

### Interaction of N<sup>pro</sup> with Interferon Regulatory Factor-3

#### 5.1 INTRODUCTION

The hallmark of innate immune response against viruses is the activation of Interferon (IFN) type-1 (IFN $\alpha/\beta$ ) signaling in immune cells [116]. IFN $\alpha/\beta$  synthesis is triggered by several of the Interferon Regulatory Factors (IRFs); particularly the activation is mediated by IRF3. IRF3 is expressed constitutively in various cell types. Its activation is triggered by phosphorylation by cellular kinases such as TBK-1/IKK $\epsilon$  which in turn are activated by pattern recognition receptors (PRR) in the immune cells. PRRs recognize anomalous non-self motifs called pathogen-associated molecular patterns or PAMPS such as dsRNA intermediate formed during RNA virus replication. Two groups of PRR have been identified which play a role in the activation of the type-1 IFN signaling following RNA virus infection. The Toll-like Receptors (TLRs) recognize dsRNA or ssRNA that's released by an infected cell during virus infection; intracellular dsRNA intermediate is recognized by cytosolic helicases like RIG-1/MDA5. Both pathways result in the phosphorylation and activation of IRF3 which translocates into the nucleus and activates the transcription of IFN $\alpha/\beta$  [52,111]. It has been proposed that pestiviruses have two mechanisms to counter the innate immune response essential for establishment of infection in dendritic cells and macrophages. First mechanism involves the glycoprotein E<sup>ms</sup> which is unique to pestiviruses; E<sup>ms</sup> is secreted into the extracellular space where it digests released dsRNA and ssRNA molecules before they can bind and trigger the TLR-3 response [112,113]. Inside the infected cell, a second layer of protection against anti-viral response is provided by the pestiviral N-terminal protease.

N<sup>pro</sup> has been shown to bind IRF3, the activator of IFN $\alpha/\beta$ , in the cytosol. Following binding, IRF3 is ubiquitinated and degraded in a proteasome dependent manner. This role of N<sup>pro</sup> in interfering with the activation of IFN response is indeed important for the virus to establish infection, since mutant classical swine fever virus with a deletion of the N<sup>pro</sup> gene is attenuated in pigs (Chapter-1).

### 5.1.1 Map of proposed IRF3 binding interface on N<sup>pro</sup>

Although several studies have demonstrated a physical interaction between N<sup>pro</sup> and IRF3 *in vivo* in both CSFV and BVDV, the exact molecular mechanism of this interaction is unknown. Mutations of the catalytic Cys69 had minimal effect on IFN induction for both BVDV and CSFV N<sup>pro</sup>, indicating that the protease activity is not involved in IRF3 binding. However, mutations of His49 to Val or Leu resulted in a loss of IFN-antagonistic activity, suggesting at least partial structural overlap between the protease and anti-IFN functions of N<sup>pro</sup>.

In CSFV N<sup>pro</sup>, it was shown that the deletion of the first 19 amino acids (but not 22 or more) did not disrupt the suppression of IFN by CSFV N<sup>pro</sup>. Additionally, mutations disrupting the TRASH motif of N<sup>pro</sup>, D136N and C112R abolished IRF3 induction [49,61] and the anti-IFN function of N<sup>pro</sup>, suggesting that an intact Zinc-binding motif is required in order for N<sup>pro</sup> to bind IRF3. Point mutations of Glu22 to Leu/Val and His40 to Leu also led to a loss of IFN-antagonistic activity. In case of BVDV, similar to CSFV, N<sup>pro</sup> deletion mutant that lacks the first 30 amino acids also loses its ability to inhibit IFN- $\alpha/\beta$  induction. However, in BVDV unlike in CSFV, the N-terminus of N<sup>pro</sup> also seems to be involved in the anti-IFN mechanism. For instance, the mutation Leu9Pro in BVDV N<sup>pro</sup> ablated the protein's antagonistic effect towards IFN induction. However, the mutation only prevented ubiquitination of IRF3 but not binding of N<sup>pro</sup> to IRF3

[58,114]. This suggests that the N-terminus of N<sup>pro</sup> could be involved in recruiting a protein that facilitates ubiquitination of IRF3.

To determine if the residues indicated in the anti-interferon response form a localized IRF3-binding surface, we mapped the above-mentioned residues on the 3D structure of N<sup>pro</sup>, along with the conserved residues (Figure-5.1). Large deletions such as the N-terminal 19-22 amino acids or the C-terminal 24 amino acids were not included in the surface mapping, since their loss of function may be caused by the disruption of protein folding. In addition, certain mutations may simply decrease protein stability instead of playing a role in IFN activation. For example, Glu22 was proposed to be important for both proteolytic activity and subversion of IFN functions. In light of the crystal structure of N<sup>pro</sup>, we believe it is likely that the mutation would destabilize the protein folding leading to loss of function.

All other residues implicated in anti-IFN function of N<sup>pro</sup> form two spatial clusters on the opposite sides of the protein surface; one cluster is on a face of the protease domain, and the other on the zinc-binding domain. It's possible that the two patches have distinct roles in the anti-IFN function of N<sup>pro</sup>. Since N<sup>pro</sup> is unlikely to ubiquitinate IRF-3 directly, other cellular proteins probably need to bind to the N<sup>pro</sup>-IRF3 complex for ubiquitination and degradation to occur. Based on the surface distribution of residues involved in IRF3 binding, we speculate that the TRASH motif in the C-terminal Zn binding domain interact with IRF-3, whereas the N-terminus of the protein would bind to a cellular protein involved in the ubiquitination reaction. Identification of cellular proteins that interact with N<sup>pro</sup> is an essential next step towards understanding how binding of N<sup>pro</sup> to IRF3 leads to the degradation of IRF3.

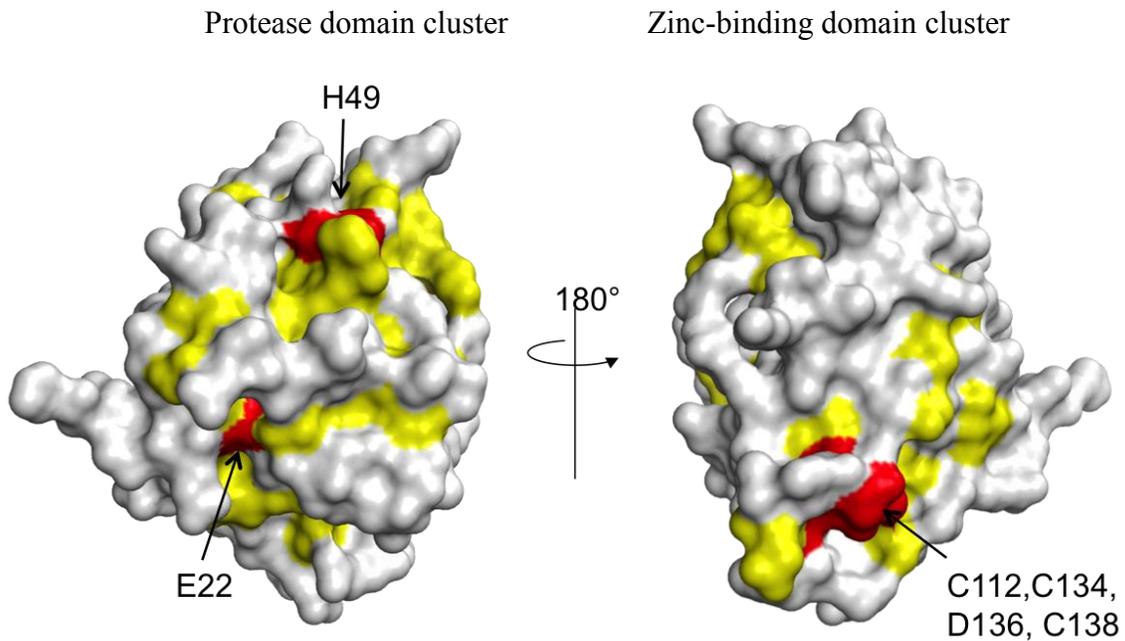


Figure 5.1: Residues of N<sup>pro</sup> involved in the subversion of interferon response.

Surface representation of CSFV N<sup>pro</sup> with residues involved in inactivation of IFN response shown in red. Conserved residues on the surface of N<sup>pro</sup> are in yellow. The protease domain and the zinc-domain surface clusters are on the opposite sides of the protein. The missing residues 1 to 17 at the N-terminal would also likely be present near E22 in the protease cluster.

## 5.2 EXPERIMENTAL DESIGN

Porcine IRF3 is a 48 kDa protein. It has two functional domains, the N-terminal DNA binding domain (DBD) formed by amino acids 1-120 and the C-terminal trans-activating domain (TAD) also called the regulatory domain. The TAD domain constituted by aa190-419 (sequence in pig) comprises three functionally defined regions, the inactivation domain, the auto-inhibition region, and the serine rich region (SRR). The auto-inhibition region prior to activation prevents the dimerization of IRF3 and its interaction with co-factor proteins such as CBP (Creb-Binding Protein) in the nucleus. The SRR carries two distinct phosphorylation sites. Activation of IRF3 in the cell is a tightly regulated process triggered by the activation of upstream protein kinases in

response to viral or bacterial infection. Phosphorylation of the serines in SRR leads to release of auto-inhibition, activation of IRF3, dimerization and subsequent interaction with transcription co-factors in the nucleus. The residues between the DBD and TAD, which carry the nuclear export signal (NES), are believed to be disordered. The amino acid sequence of IRF3 is highly conserved over all higher eukaryotes. Figure-5.2 is a sequence alignment of IRF3 of pig, human and mouse. Greater than 50% of the sequence of pig IRF3 is identical to its human counterpart. Hence it is conceivable that the structure of porcine IRF3 would resemble that of human IRF3. The crystal structures of the regulatory and DNA binding domains of human IRF3 have been solved independently [121–124]. Figure-5.2 also shows an overlay of the determined secondary structure on the sequence of IRF3.

As a first step towards characterizing the nature of interaction between N<sup>pro</sup> and IRF3 *in vitro* we have expressed and purified recombinant porcine IRF3.

### **5.2.1 Expression and purification of soluble recombinant porcine IRF3**

Gene of porcine IRF3 with N-terminal 6xHis-tag was synthesized and cloned into an in-house vector pJexpress414 that carries the T7-polymerase promoter and ampicillin resistance (DNA2.0). Nucleotide sequence of the gene was optimized for expression in *E. coli*. BL21 CodonPlus *E. coli* cells (Stratagene) were transformed with the pJexpress-IRF3 plasmid. Large-scale cultures were grown in 2 L of Luria Broth containing 50 µg/mL of ampicillin at 37 °C to an O.D<sub>600</sub> of 0.6-0.8. Protein expression was induced by the addition of 1 mM IPTG and growth was continued overnight at 18 °C. The cells were resuspended in 50ml of lysis buffer (50 mM sodium phosphate (pH 7.0) and 300 mM NaCl) and lysed by sonication with a Misonix XL-2000 sonicator.

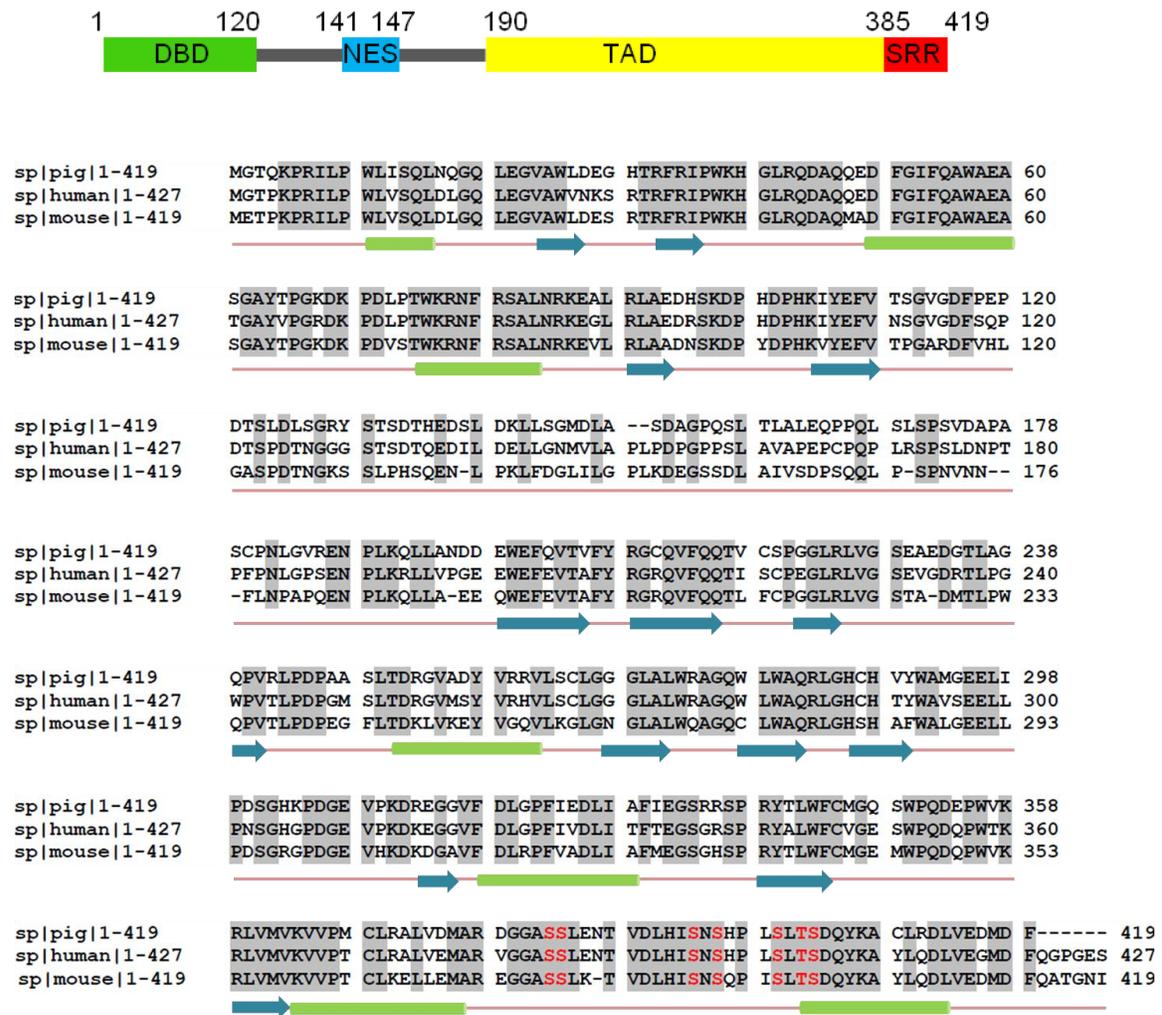


Figure 5.2: Multiple sequence alignment of IRF3.

Alignment of the porcine (uniprot ascn. no. Q764M6), human (uniprot ascn. no. Q14653) and mouse (uniprot ascn. no.: P70671) IRF3 sequences. Identical regions in all three sequences are highlighted in grey. Predicted and experimentally determined secondary structural elements corresponding to human IRF3 have been overlaid on the sequence. Domain organization in IRF3 (top) corresponds to the human sequence. The two phosphorylation sites in the serine rich region are indicated in red in the sequence.

Cell lysate was centrifuged at 15,000 rpm for 15 min to remove cellular debris. Protein in the soluble fraction was loaded onto Talon™ (Clontech) metal-affinity

chromatography resin equilibrated in lysis buffer. Bound IRF3 was eluted using a gradient of 5–150 mM imidazole in wash buffer (50 mM sodium phosphate, pH 8.0, and 0.5 M NaCl). Full length IRF3 was eluted along with shorter fragments of IRF3 that also carried the N-terminal His-tag (Figure-5.2). The fractions containing IRF3 were pooled and concentrated to a final volume of 1 mL using Amicon™ Ultra centrifugal filter device (Millipore). The protein was then loaded onto a Superdex 75 gel-filtration column (GE Healthcare) and eluted with 20 mM Tris–HCl (pH 7.0), 150 mM NaCl containing 5 mM β-mercaptoethanol, using the AKTA purification system (GE Healthcare). Smaller molecular weight fragments of IRF3 could not be fully resolved using this column, probably due to the small difference in molecular weight of the fragments (Figure-5.3).

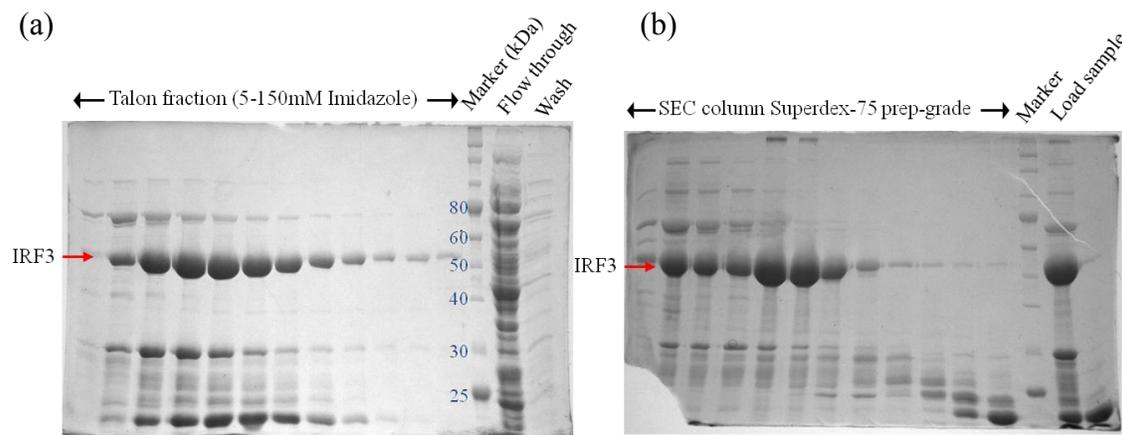


Figure 5.3: Purification of IRF3 (N-terminal 6xHis-tag).

SDS-PAGE of the purification of full length recombinant IRF3 using (a) metal affinity chromatography to isolate N-terminal 6xHis tagged protein from the soluble fraction of the cell lysate, followed by (b) size exclusion chromatography using superdex-75 (16/60) prep-grade column to separate fragments of IRF3 from the full length protein. Molecular weight of IRF3 is 48 kDa but it runs as a higher molecular weight protein (~52 kDa) on the SDS-PAGE.

### **5.2.2 Purification of full length recombinant porcine IRF3 from *E. coli* inclusion bodies**

Purification of soluble recombinant IRF3 resulted in very low purity and yield. We hence resorted to purification under denaturing conditions in order to improve the purity of the final product. We generated a C-terminal 6xHis-tag construct of IRF3. We sub-cloned IRF3 into pET28b vector digested with *NcoI* and *XhoI* in order to generate the C-terminal 6xHis-tagged protein.

Following cell lysis, majority of the over-expressed protein was found in insoluble inclusion bodies. Inclusion body pellet was resuspended in denaturing buffer containing 50 mM Phosphate buffer (pH 8.0), 300 mM NaCl, and 6 M Guanidine Hydrochloride until a homogenous solution was obtained. Denatured IRF3 was isolated from the suspension by binding to Talon metal affinity resin pre-equilibrated in the denaturing buffer. Resin was washed twice with 10ml of denaturing wash buffer (50 mM Phosphate buffer (pH 7.0), 300 mM NaCl, 5.4 M GnHCl) containing 5 mM Imidazole. Finally, denatured full length IRF3 was eluted with a gradient of 5-200 mM Imidazole in the denaturing wash buffer. Figure-5.4a is the SDS-PAGE of isolation of full length IRF3 under denaturing conditions. Following metal affinity chromatography, fractions containing IRF3 were pooled together and dialyzed against 10X volumes of native buffer (20 mM Tris, 200 mM NaCl, 5 mM 2-MercaptoEthanol) 3 times in order to remove all the denaturant from the solution and promote protein folding. Protein solution was concentrated using Amicon™ Ultra centrifugal filter (Millipore) with a molecular weight cut-off of 30 kDa to a final volume of 1ml. Concentrated protein was loaded on to Superdex-75 size exclusion column and column was washed with native buffer using a flow rate of 1ml/min. Figure-5.4b is the SDS-PAGE of fractions from the size exclusion column. The protein is only about 80% pure at this point in the purification procedure.

Further attempts to optimize the procedure by inclusion of ion-exchange chromatography following removal of chaotrope and renaturation are being pursued. Protein refolding was verified using CD-spectrometry analysis. Figure-5.4c is the far UV CD spectrum of the refolded full length IRF3 determined on the CD Spectrometer AVIV model 215 at the Sealy Center for Structural Biology, UTMB. The spectrum is similar to that of wild type full length IRF3 previously reported by Dragan A. I. *et al.* [125]

We are in the process of further optimizing the purification in order to obtain highly pure (>95%) IRF3. We have also co-expressed N<sup>pro</sup> and porcine IRF3 cloned into the pETDuet-1 (Novagen) vector in *E. coli* (clone generated in the laboratory of Dr. Ruggli) in order to co-purify and subsequently co-crystallize the protein complex.

### **5.3 FUTURE DIRECTIONS**

In order to characterize the solution behavior of N<sup>pro</sup> in complex with IRF3, we will study the equilibrium behavior of the complex using size exclusion chromatography, analytical ultra-centrifugation and isothermal titration calorimetry. This will provide information on the stoichiometry of binding, and thermodynamics and affinity of interaction between the two proteins. We will also study the complex using small angle X-ray scattering (SAXS). From this technique, we can determine protein size and a 3D molecular envelope using an ab initio shape determination programs. Since the crystal structures of N<sup>pro</sup> and the individual domains of IRF3 have been determined, the information could be used to generate a model of N<sup>pro</sup>-IRF complex that agrees with the molecular envelope generated using SAXS experiments. Further, we will attempt to determine the co-crystal structure of the complex of N<sup>pro</sup> and IRF3.

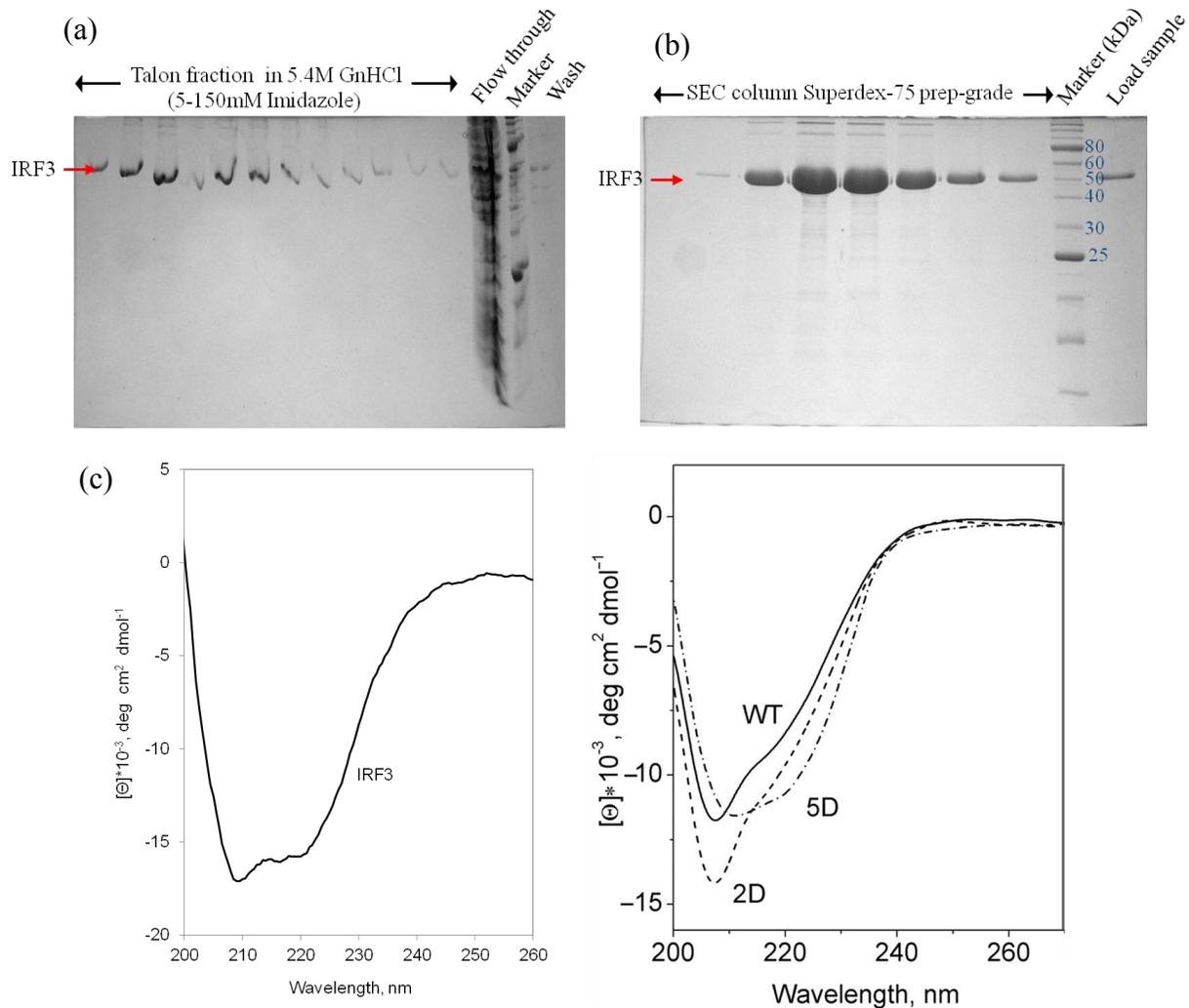


Figure 5.4: Purification of recombinant IRF3 under denaturing conditions.

SDS-PAGE (10%) of the purification of full-length recombinant C-terminal 6xHis-tagged IRF3 under denaturing conditions. (a) Metal affinity chromatography using Talon equilibrated in 6 M GnHCl. Guanidine in the buffer distorts the protein bands in an SDS-PAGE. (b) Size exclusion chromatography using superdex-75 (16/60) prep-grade column following removal of chaotrope from protein solution. (c) Comparison of CD spectrum of refolded full length porcine IRF-3 (left) to the spectra of full length human IRF3 reported by Dragan A. I. *et al.* (right). Both spectra show a dip at 222 nm and another larger dip at 208 nm. 2D and 5D are two phosphomimetic mutants of IRF3. Mutant 2D is a monomer and 5D is a dimer in solution. Dimerization was shown to increase the helical content of the protein.

## CHAPTER-6

### Summary

We reported the crystal structures of CSFV N<sup>pro</sup> and a C168A cleavage site N<sup>pro</sup> mutant to 1.6 Å. N<sup>pro</sup> has a novel fold and is thus structurally distinct from other known cysteine proteases. N<sup>pro</sup> consists of two domains, a protease domain and a zinc-binding domain, consistent with bioinformatics and limited proteolysis results. The cysteine protease triad in N<sup>pro</sup> was suggested to be Glu22, His49, and Cys69 by mutagenesis studies (Rumenapf *et al.* 1998). Our structure shows that His49 and Cys69 form a catalytic dyad in the active site, as predicted. However, Glu22 is 24 Å away from the self-cleavage site Cys168 in the active site, and thus is unlikely to participate in catalysis.

N<sup>pro</sup> cleaves the peptide bond between Cys168 and Ser169 in the viral polyprotein. Our structure shows that the last seven C-terminal residues Pro<sup>162</sup>-Leu-Trp-Val-Thr-Thr-Cys<sup>168</sup> of the protein that constitute the self-cleavage site are not only bound in the protease active site, but also contribute an integral β-strand to the central β-sheet that makes up the active site. This β-strand has specific side chain interactions with the neighboring amino acid residues in the substrate-binding pocket in addition to the typical β-sheet main chain hydrogen bonding. This is consistent with the highly conserved nature of these residues and mutational studies that showed that Trp164 to Ala substitution renders the protease inactive (Wiskerchen *et al.* 1991). Thus, the C-terminus of N<sup>pro</sup> occludes the catalytic site following cleavage, thus inhibiting any *trans* activity of the protease and limiting the activity of the enzyme to a single catalytic turnover. The structure of N<sup>pro</sup>-C168A also showed that the highly conserved cleavage site Cys168 is not required for the catalytic activity of N<sup>pro</sup>. However, the amino acid at the terminal

position could have been selected for the efficiency of cleavage at this site or Cys168 could have other roles in the function of N<sup>pro</sup>.

The C-terminal zinc-binding domain forms an anti-parallel  $\beta$ -sheet comprising five  $\beta$ -strands. We identified the metal-binding TRASH motif, Cys<sup>112</sup>-X<sub>21</sub>-Cys<sup>134</sup>-X<sub>3</sub>-Cys<sup>138</sup> that coordinates a single zinc atom on one end of the  $\beta$ -sheet. Individual C112A/R, C134A, D136N, and C138A mutations in the TRASH motif resulted in loss of zinc binding, and also abolished IRF3 interaction and interferon antagonism. Thus, the intact TRASH motif residues are absolutely required to maintain the anti-interferon activity of N<sup>pro</sup>. The residues involved in subversion of IFN response are localized on two opposite sides of the protein; one cluster is in the N-terminal protease domain, and the other in the C-terminal zinc-binding site. We will test whether the N-terminal protease and C-terminal zinc-binding domain may have different functions in anti-IFN response, i.e., direct interaction with IRF3 vs inducing a down-stream response leading to degradation of IRF3.

Taken together, the N<sup>pro</sup> structures presented here establish the mechanism of autocatalysis and subsequent auto-inhibition of N<sup>pro</sup>, and provide insight into its interaction with IRF3 in subversion of host immune response.

## Appendix A

### ELSEVIER LICENSE TERMS AND CONDITIONS

07/13/2012 09:25 PM

This is a License Agreement between Keerthi Gottipati ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, UK
Registered Company Number	1982084
Customer name	Keerthi Gottipati
Customer address	301 University Blvd Galveston, TX 77555
License number	2947390077354
License date	Jul 13, 2012
Licensed content publisher	Elsevier
Licensed content publication	Journal of Molecular Biology
Licensed content title	Zinc Binding in Pestivirus Npro is Required for Interferon Regulatory Factor 3 Interaction and Degradation
Licensed content author	Michal R. Szymanski, Ana R. Fiebach, Jon-Duri Tratschin, Marco Gut, V. M. Sadagopa Ramanujam, Keerthi Gottipati, Purvi Patel, Mengyi Ye, Nicolas Ruggli, Kyung H. Choi
Licensed content date	14 August 2009
Licensed content volume number	391
Licensed content issue number	2
Number of pages	12
Start Page	438
End Page	449
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Order reference number	
Title of your thesis/dissertation	Structural and Biochemical Characterization of the N-terminal Protease of Classical Swine Fever Virus
Expected completion date	Dec 2012
Estimated size (number of pages)	130
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.0 USD / 0.0 GBP
Total	0.00 USD

## Bibliography

- [1] T. Rüménapf and H. J. Thiel, “Molecular Biology of Pestiviruses,” in *Animal Viruses: Molecular Biology*, T. Mettenleiter C. and F. Sobrino, Eds. Horizon Scientific Press (Norfolk, UK), 2008, pp. 39–96.
- [2] M. S. Collett, D. K. Anderson, and E. Retzel, “Comparisons of the Pestivirus Bovine Viral Diarrhoea Virus with Members of the Flaviviridae,” *Journal of General Virology*, vol. 69, no. 10, pp. 2637–2643, Oct. 1988.
- [3] C. R. Pringle, “The universal system of virus taxonomy of the International Committee on Virus Taxonomy (ICTV), including new proposals ratified since publication of the Sixth ICTV Report in 1995,” *Archives of Virology*, vol. 143, no. 1, pp. 203–210, Jan. 1998.
- [4] C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger, and L. A. Ball, *Virus Taxonomy, VIIIth Report of the International Committee on Taxonomy of Viruses*. Elsevier/Academic Press (London UK), 2005.
- [5] V. Moennig and P. G. W. Plagemann, “The Pestiviruses,” in *Advances in Virus Research Volume 41*, vol. 41, Elsevier, 1992, pp. 53–98.
- [6] V. Moennig, “Introduction to classical swine fever: virus, disease and control policy,” *Veterinary Microbiology*, vol. 73, no. 2–3, pp. 93–102, Apr. 2000.
- [7] P. Nettleton, J. Gilray, P. Russo, and E. Dlissi, “Border disease of sheep and goats.,” *Veterinary Research*, vol. 29, no. 3–4, pp. 327–340, 1998.
- [8] V. Moennig, H. Houe, and A. Lindberg, “BVD control in Europe: current status and perspectives,” *Animal Health Research Reviews*, vol. 6, no. 01, pp. 63–74, Feb. 2005.
- [9] H. Houe, “Economic impact of BVDV infection in dairies,” *Biologicals*, vol. 31, no. 2, pp. 137–143, Jun. 2003.
- [10] H. Houe, “Epidemiology of bovine viral diarrhea virus,” *The Veterinary Clinics of North America. Food Animal Practice*, vol. 11, no. 3, pp. 521–547, Nov. 1995.

- [11] V. Moennig, G. Floegel-Niesmann, and I. Greiser-Wilke, "Clinical Signs and Epidemiology of Classical Swine Fever: A Review of New Knowledge," *The Veterinary Journal*, vol. 165, no. 1, pp. 11–20, Jan. 2003.
- [12] S. Ribbens, J. Dewule, F. Koenen, H. Laevens, and A. D. Kruie, "Transmission of classical swine fever. A review," *Veterinary Quarterly*, vol. 26, no. 4, pp. 146–155, 2004.
- [13] A. L. E. Lindberg, "Bovine viral Diarrhoea virus infections and its control. A review," *Veterinary Quarterly*, vol. 25, no. 1, pp. 1–16, 2003.
- [14] G. Meyers and H. J. Thiel, "Molecular characterization Of Pestiviruses," *Advances in Virus Research*, vol. 847, pp. 53–117, Dec. 1996.
- [15] V. Moennig, "Pestiviruses : a review," *Veterinary Microbiology*, vol. 23, pp. 35–54, 1990.
- [16] J. Brownlie, "Pathogenesis of mucosal disease and molecular aspects of bovine virus diarrhoea virus," *Veterinary Microbiology*, vol. 23, no. 1–4, pp. 371–382, Jun. 1990.
- [17] V. Moennig, H. Frey, E. Liebler, J. Pohlenz, and Liess B, "Reproduction of mucosal disease with cytopathogenic bovine viral diarrhoea virus selected in vitro," *The Veterinary Record*, vol. 127, no. 8, pp. 200–203, 1990.
- [18] B. I. Loehr, H. R. Frey, V. Moennig, and I. Greiser-Wilke, "Experimental induction of mucosal disease: consequences of superinfection of persistently infected cattle with different strains of cytopathogenic bovine viral diarrhea virus," *Archives of Virology*, vol. 143, no. 4, pp. 667–679, Apr. 1998.
- [19] P. F. Nettleton, J. S. Gilmour, J. A. Herring, and J. A. Sinclair, "The production and survival of lambs persistently infected with border disease virus," *Comparative Immunology, Microbiology and Infectious Diseases*, vol. 15, no. 3, pp. 179–188, Jul. 1992.
- [20] S. Edwards, A. Fukusho, P.-C. Lefèvre, A. Lipowski, Z. Pejsak, P. Roehe, and J. Westergaard, "Classical swine fever: the global situation," *Veterinary Microbiology*, vol. 73, no. 2–3, pp. 103–119, Apr. 2000.
- [21] J. Fritzemeier, J. Teuffert, I. Greiser-Wilke, C. Staubach, H. Schlüter, and V. Moennig, "Epidemiology of classical swine fever in Germany in the 1990s," *Veterinary Microbiology*, vol. 77, no. 1–2, pp. 29–41, Nov. 2000.

- [22] I. Greiser-Wilke and V. Moennig, "Vaccination against classical swine fever virus: limitations and new strategies," *Animal Health Research Reviews*, vol. 5, no. 02, pp. 223–226, Dec. 2004.
- [23] M. Beer, I. Reimann, B. Hoffmann, and K. Depner, "Novel marker vaccines against classical swine fever," *Vaccine*, vol. 25, no. 30, pp. 5665–70, Jul. 2007.
- [24] I. Greiser-Wilke, B. Grummer, and V. Moennig, "Bovine viral diarrhoea eradication and control programmes in Europe," *Biologicals*, vol. 31, no. 2, pp. 113–118, Jun. 2003.
- [25] J. F. Ridpath and S. R. Bolin, "Delayed onset postvaccinal mucosal disease as a result of genetic recombination between genotype 1 and genotype 2 BVDV," *Virology*, vol. 212, no. 1, pp. 259–62, Sep. 1995.
- [26] H. B. Ohmann and B. Bloch, "Electron microscopic studies of bovine viral diarrhoea virus in tissues of diseased calves and in cell cultures," *Archives of Virology*, vol. 71, no. 1, pp. 57–74, Mar. 1982.
- [27] A. Wegelt, I. Reimann, H. Granzow, and M. Beer, "Characterization and purification of recombinant bovine viral diarrhoea virus particles with epitope-tagged envelope proteins," *Journal of General Virology*, vol. 92, no. 6, pp. 1352–1357, Feb. 2011.
- [28] M. F. Coria, M. J. F. Schmerr, and A. W. McClurkin, "Characterization of the major structural proteins of purified bovine viral diarrhoea virus," *Archives of Virology*, vol. 76, no. 4, pp. 335–339, Dec. 1983.
- [29] M. S. Collett, R. Larson, S. K. Belzer, and E. Retzel, "Proteins encoded by bovine viral diarrhoea virus: the genomic organization of a pestivirus," *Virology*, vol. 165, no. 1, pp. 200–8, Jul. 1988.
- [30] C. L. Murray, J. Marcotrigiano, and C. M. Rice, "Bovine viral diarrhoea virus core is an intrinsically disordered protein that binds RNA," *Journal of Virology*, vol. 82, no. 3, pp. 1294–304, Feb. 2008.
- [31] C. Riedel, B. Lamp, M. Heimann, M. König, S. Blome, V. Moennig, C. Schüttler, H. J. Thiel, and T. Rümenapf, "The core protein of classical Swine Fever virus is dispensable for virus propagation in vitro," *PLoS Pathogens*, vol. 8, no. 3, p. e1002598, Jan. 2012.

- [32] B. D. Lindenbach, H. J. Thiel, and C. M. Rice, "Flaviviridae: the viruses and their replication," in *Fields Virology*, 5th ed., D. M. Knipe and P. M. Howley, Eds. (Philadelphia, PA) Lippincott-Raven, 2007, pp. 1101–1152.
- [33] M. S. Collett, "Molecular genetics of pestiviruses," *Comparative Immunology, Microbiology and Infectious Diseases*, vol. 15, no. 3, pp. 145–154, Jul. 1992.
- [34] T. Poole, C. Wang, R. Popp, L. Potgieter, A. Siddiqui, and M. S. Collett, "Pestivirus translation initiation occurs by internal ribosome entry," *Virology*, vol. 206, no. 1, pp. 750–754, Jan. 1995.
- [35] M. Collett, M. Wiskerchen, E. Welniak, and S. K. Belzer, "Bovine viral diarrhea virus genomic organization," *Archives of Virology. Supplementum*, vol. 3, pp. 19–27, Jan. 1991.
- [36] H. J. Thiel, R. Stark, E. Weiland, T. Rumenapf, and G. Meyers, "Hog cholera virus: molecular composition of virions from a pestivirus," *Journal of Virology*, vol. 65, no. 9, pp. 4705–4712, Sep. 1991.
- [37] H. Thiel, Meyers G, R. Stark, N. Tautz, T. Rumenapf, G. Unger, and K. Conzelmann, "Molecular characterization of positive-strand viruses: pestiviruses and the porcine reproductive and respiratory syndrome virus (PRRSV)," *Archives of Virology, Supplementum*, vol. 7, pp. 41–52, 1993.
- [38] T. Rumenapf, G. Unger, J. H. Strauss, and H. J. Thiel, "Processing of the envelope glycoproteins of pestiviruses," *Journal of Virology*, vol. 67, no. 6, pp. 3288–3294, Jun. 1993.
- [39] J. Xu, E. Mendez, P. Caron, C. Lin, M. Murcko, M. Collett, and C. Rice, "Bovine viral diarrhea virus NS3 serine proteinase: polyprotein cleavage sites, cofactor requirements, and molecular model of an enzyme essential for pestivirus replication," *Journal of Virology*, vol. 71, no. 7, pp. 5312–5322, Jul. 1997.
- [40] F. Weiland, E. Weiland, G. Unger, A. Saalmüller, and H. J. Thiel, "Localization of pestiviral envelope proteins E(rns) and E2 at the cell surface and on isolated particles," *The Journal of General Virology*, vol. 80, no. 5, pp. 1157–65, May 1999.
- [41] D. Pavlović, D. C. A. Neville, O. Argaud, B. Blumberg, R. A. Dwek, W. B. Fischer, and N. Zitzmann, "The hepatitis C virus p7 protein forms an ion channel that is inhibited by long-alkyl-chain iminosugar derivatives," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 10, pp. 6104–8, May 2003.

- [42] T. Harada, N. Tautz, and H. J. Thiel, "E2-p7 Region of the Bovine Viral Diarrhea Virus Polyprotein: Processing and Functional Studies," *Journal of Virology*, vol. 74, no. 20, pp. 9498–9506, Oct. 2000.
- [43] T. Lackner, A. Müller, A. Pankraz, P. Becher, H. J. Thiel, A. E. Gorbalenya, and N. Tautz, "Temporal modulation of an autoprotease is crucial for replication and pathogenicity of an RNA virus," *Journal of Virology*, vol. 78, no. 19, pp. 10765–75, Oct. 2004.
- [44] T. Lackner, H. J. Thiel, and N. Tautz, "Dissection of a viral autoprotease elucidates a function of a cellular chaperone in proteolysis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 5, pp. 1510–5, Jan. 2006.
- [45] J. K. Tamura, P. Warrener, and M. S. Collett, "RNA-stimulated NTPase activity associated with the p80 protein of the pestivirus bovine viral diarrhea virus," *Virology*, vol. 193, no. 1, pp. 1–10, Mar. 1993.
- [46] P. Warrener and M. Collett, "Pestivirus NS3 (p80) protein possesses RNA helicase activity," *Journal of Virology*, vol. 69, no. 3, pp. 1720–1726, Mar. 1995.
- [47] W. Zhong, L. L. Gutshall, and A. M. Del Vecchio, "Identification and Characterization of an RNA-Dependent RNA Polymerase Activity within the Nonstructural Protein 5B Region of Bovine Viral Diarrhea Virus," *Journal of Virology*, vol. 72, no. 11, pp. 9365–9369, Nov. 1998.
- [48] T. L. Tellinghuisen, K. L. Foss, J. C. Treadaway, and C. M. Rice, "Identification of residues required for RNA replication in domains II and III of the hepatitis C virus NS5A protein," *Journal of Virology*, vol. 82, no. 3, pp. 1073–83, Feb. 2008.
- [49] L. Huang, J. Hwang, S. D. Sharma, M. R. S. Hargittai, Y. Chen, J. J. Arnold, K. D. Raney, and C. E. Cameron, "Hepatitis C virus nonstructural protein 5A (NS5A) is an RNA-binding protein," *The Journal of Biological Chemistry*, vol. 280, no. 43, pp. 36417–28, Oct. 2005.
- [50] C. L. Murray, C. T. Jones, and C. M. Rice, "Architects of assembly: roles of Flaviviridae non-structural proteins in virion morphogenesis," *Nature Reviews. Microbiology*, vol. 6, no. 9, pp. 699–708, Sep. 2008.
- [51] M. Wiskerchen, S. K. Belzer, and M. S. Collett, "Pestivirus gene expression: the first protein product of the bovine viral diarrhea virus large open reading frame, p20, possesses proteolytic activity," *Journal of Virology*, vol. 65, no. 8, pp. 4508–14, Aug. 1991.

- [52] R. Stark, G. Meyers, T. Rumenapf, H. J. Thiel, and T. Rumenapf, "Processing of pestivirus polyprotein: cleavage site between autoprotease and nucleocapsid protein of classical swine fever virus," *Journal of Virology*, vol. 67, no. 12, pp. 7088–95, Dec. 1993.
- [53] M. R. Szymanski, A. R. Fiebach, J. D. Tratschin, M. Gut, V. M. S. Ramanujam, K. Gottipati, P. Patel, M. Ye, N. Ruggli, and K. H. Choi, "Zinc binding in pestivirus N(pro) is required for interferon regulatory factor 3 interaction and degradation," *Journal of Molecular Biology*, vol. 391, no. 2, pp. 438–49, Aug. 2009.
- [54] T. Rumenapf, R. Stark, M. Heimann, and H. J. Thiel, "N-terminal protease of pestiviruses: identification of putative catalytic residues by site-directed mutagenesis," *Journal of Virology*, vol. 72, no. 3, pp. 2544–7, Mar. 1998.
- [55] N. D. Rawlings, A. J. Barrett, and A. Bateman, "MEROPS: the database of proteolytic enzymes, their substrates and inhibitors," *Nucleic Acids Research*, vol. 40, pp. D343–50, Jan. 2012.
- [56] T. J. G. Ettema, M. a Huynen, W. M. de Vos, and J. van der Oost, "TRASH: a novel metal-binding domain predicted to be involved in heavy-metal sensing, trafficking and resistance," *Trends in Biochemical Sciences*, vol. 28, no. 4, pp. 170–3, Apr. 2003.
- [57] J. Hiscott, "Triggering the innate antiviral response through IRF-3 activation," *Journal of Biological Chemistry*, vol. 282, no. 21, pp. 15325–9, May 2007.
- [58] T. Kawai and S. Akira, "Innate immune recognition of viral infection," *Nature Immunology*, vol. 7, no. 2, pp. 131–7, Feb. 2006.
- [59] N. Ruggli, J. Tratschin, M. Schweizer, K. C. McCullough, M. A. Hofmann, and A. Summerfield, "Classical Swine Fever Virus Interferes with Cellular Antiviral Defense : Evidence for a Novel Function of N pro," *Journal of Virology*, vol. 77, no. 13, pp. 7645–7654, 2003.
- [60] O. Bauhofer, A. Summerfield, Y. Sakoda, J. D. Tratschin, M. a Hofmann, and N. Ruggli, "Classical swine fever virus Npro interacts with interferon regulatory factor 3 and induces its proteasomal degradation," *Journal of Virology*, vol. 81, no. 7, pp. 3087–96, Apr. 2007.
- [61] L. Hilton, K. Moganeradj, G. Zhang, Y. H. Chen, R. E. Randall, J. W. McCauley, and S. Goodbourn, "The NPro product of bovine viral diarrhea virus inhibits DNA binding by interferon regulatory factor 3 and targets it for proteasomal degradation," *Journal of Virology*, vol. 80, no. 23, pp. 11723–32, Dec. 2006.

- [62] J. Seago, L. Hilton, E. Reid, V. Doceul, J. Jeyatheesan, K. Moganeradj, J. McCauley, B. Charleston, and S. Goodbourn, "The Npro product of classical swine fever virus and bovine viral diarrhea virus uses a conserved mechanism to target interferon regulatory factor-3," *The Journal of General Virology*, vol. 88, no. 11, pp. 3002–6, Nov. 2007.
- [63] Z. Chen, R. Rijnbrand, R. K. Jangra, S. G. Devaraj, L. Qu, Y. Ma, S. M. Lemon, and K. Li, "Ubiquitination and proteasomal degradation of interferon regulatory factor-3 induced by Npro from a cytopathic bovine viral diarrhea virus," *Virology*, vol. 366, no. 2, pp. 277–92, Sep. 2007.
- [64] L. H. V. G. Gil, I. H. Ansari, V. Vassilev, D. Liang, V. C. H. Lai, W. Zhong, Z. Hong, E. J. Dubovi, and R. O. Donis, "The amino-terminal domain of bovine viral diarrhea virus Npro protein is necessary for alpha/beta interferon antagonism," *Journal of Virology*, vol. 80, no. 2, p. 900, 2006.
- [65] D. J. Paton and I. Greiser-Wilke, "Classical swine fever – an update," *Research in Veterinary Science*, vol. 75, no. 3, pp. 169–178, Dec. 2003.
- [66] N. Ruggli, A. Summerfield, A. R. Fiebach, L. Guzylack-Piriou, O. Bauhofer, C. G. Lamm, S. Waltersperger, K. Matsuno, L. Liu, M. Gerber, K. H. Choi, M. a Hofmann, Y. Sakoda, and J. D. Tratschin, "Classical swine fever virus can remain virulent after specific elimination of the interferon regulatory factor 3-degrading function of Npro," *Journal of Virology*, vol. 83, no. 2, pp. 817–29, Jan. 2009.
- [67] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, and A. Bairoch, "Protein Identification and Analysis Tools on the ExPASy Server," in *The Proteomics Protocols Handbook*, J. M. Walker, Ed. Humana Press (New York), 2005, pp. 571–607.
- [68] H. Edelhoch, "Spectroscopic Determination of Tryptophan and Tyrosine in Proteins," *Biochemistry*, vol. 6, no. 7, pp. 1948–1954, Jul. 1967.
- [69] C. N. Pace, F. Vajdos, L. Fee, G. Grimsley, and T. Gray, "How to measure and predict the molar absorption coefficient of a protein," *Protein Science: a publication of the Protein Society*, vol. 4, no. 11, pp. 2411–23, Nov. 1995.
- [70] I. Baskakov and D. W. Bolen, "Forcing Thermodynamically Unfolded Proteins to Fold," *Journal of Biological Chemistry*, vol. 273, no. 9, pp. 4831–4834, Feb. 1998.
- [71] I. V. Baskakov, R. Kumar, G. Srinivasan, Y. Ji, D. W. Bolen, and B. E. Thompson, "Trimethylamine N-Oxide-induced Cooperative Folding of an Intrinsically

- Unfolded Transcription-activating Fragment of Human Glucocorticoid Receptor,” *Journal of Biological Chemistry*, vol. 274, no. 16, pp. 10693–10696, Apr. 1999.
- [72] T. S. Walter, C. Meier, R. Assenberg, K. F. Au, J. Ren, A. Verma, J. E. Nettleship, R. J. Owens, D. I. Stuart, and J. M. Grimes, “Lysine methylation as a routine rescue strategy for protein crystallization,” *Structure (London, England : 1993)*, vol. 14, no. 11, pp. 1617–22, Nov. 2006.
- [73] W. R. Rypniewski, H. M. Holden, and I. Rayment, “Structural consequences of reductive methylation of lysine residues in hen egg white lysozyme: An x-ray analysis at 1.8-Å resolution,” *Biochemistry*, vol. 32, no. 37, pp. 9851–9858, Sep. 1993.
- [74] M. Kobayashi, M. Kubota, and Y. Matsuura, “Crystallization and improvement of crystal quality for X-ray diffraction of maltooligosyl trehalose synthase by reductive methylation of lysine residues,” *Acta Crystallographica Section D Biological Crystallography*, vol. 55, no. 4, pp. 931–933, Apr. 1999.
- [75] Y. Kim, P. Quartey, H. Li, L. Volkart, C. Hatzos, C. Chang, B. Nocek, M. Cuff, J. Osipiuk, K. Tan, Y. Fan, L. Bigelow, N. Maltseva, R. Wu, M. Borovilos, E. Duggan, M. Zhou, T. A. Binkowski, R. Zhang, and A. Joachimiak, “Large-scale evaluation of protein reductive methylation for improving protein crystallization,” *Nature Methods*, vol. 5, no. 10, pp. 853–4, Oct. 2008.
- [76] A. Dong, X. Xu, A. M. Edwards, C. Chang, M. Chruszcz, M. Cuff, M. Cymborowski, R. Di Leo, O. Egorova, E. Evdokimova, E. Filippova, J. Gu, J. Guthrie, A. Ignatchenko, A. Joachimiak, N. Klostermann, Y. Kim, Y. Korniyenko, W. Minor, Q. Que, A. Savchenko, T. Skarina, K. Tan, A. Yakunin, A. Yee, V. Yim, R. Zhang, H. Zheng, M. Akutsu, C. Arrowsmith, G. V. Avvakumov, A. Bochkarev, L.-G. Dahlgren, S. Dhe-Paganon, S. Dimov, L. Dombrovski, P. Finerty, S. Flodin, A. Flores, S. Gräslund, M. Hammerström, M. D. Herman, B.-S. Hong, R. Hui, I. Johansson, Y. Liu, M. Nilsson, L. Nedyalkova, P. Nordlund, T. Nyman, J. Min, H. Ouyang, H. Park, C. Qi, W. Rabeh, L. Shen, Y. Shen, D. Sukumard, W. Tempel, Y. Tong, L. Tresagues, M. Vedadi, J. R. Walker, J. Weigelt, M. Welin, H. Wu, T. Xiao, H. Zeng, and H. Zhu, “In situ proteolysis for protein crystallization and structure determination,” *Nature Methods*, vol. 4, no. 12, pp. 1019–21, Dec. 2007.
- [77] P. A. Jekel, W. J. Weijer, and J. J. Beintema, “Use of endoproteinase Lys-C from *Lysobacter enzymogenes* in protein sequence analysis,” *Analytical Biochemistry*, vol. 134, no. 2, pp. 347–354, Oct. 1983.

- [78] T. C. Terwilliger, P. D. Adams, R. J. Read, A. J. McCoy, N. W. Moriarty, R. W. Grosse-Kunstleve, P. V. Afonine, P. H. Zwart, and L. W. Hung, "Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard," *Acta Crystallographica. Section D, Biological Crystallography*, vol. 65, no. 6, pp. 582–601, Jun. 2009.
- [79] P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart, "PHENIX: a comprehensive Python-based system for macromolecular structure solution," *Acta Crystallographica. Section D, Biological Crystallography*, vol. 66, no. 2, pp. 213–21, Feb. 2010.
- [80] T. C. Terwilliger, R. W. Grosse-Kunstleve, P. V. Afonine, N. W. Moriarty, P. H. Zwart, L. W. Hung, R. J. Read, and P. D. Adams, "Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard," *Acta Crystallographica. Section D, Biological Crystallography*, vol. 64, no. 1, pp. 61–9, Jan. 2008.
- [81] T. A. Jones, J. Y. Zou, S. W. Cowan, and M. Kjeldgaard, "Improved methods for building protein models in electron density maps and the location of errors in these models," *Acta Crystallographica Section A Foundations of Crystallography*, vol. 47, no. 2, pp. 110–119, Mar. 1991.
- [82] R. A. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton, "PROCHECK: a program to check the stereochemical quality of protein structures," *Journal of Applied Crystallography*, vol. 26, no. 2, pp. 283–291, Apr. 1993.
- [83] S. S. Krishna, "Structural classification of zinc fingers: survey and summary," *Nucleic Acids Research*, vol. 31, no. 2, pp. 532–550, Jan. 2003.
- [84] M. Barro and J. T. Patton, "Rotavirus NSP1 inhibits expression of type I interferon by antagonizing the function of interferon regulatory factors IRF3, IRF5, and IRF7," *Journal of Virology*, vol. 81, no. 9, pp. 4473–81, May 2007.
- [85] K. Saira, Y. Zhou, and C. Jones, "The infected cell protein 0 encoded by bovine herpesvirus 1 (bICP0) induces degradation of interferon response factor 3 and, consequently, inhibits beta interferon promoter activity," *Journal of Virology*, vol. 81, no. 7, pp. 3077–86, Apr. 2007.

- [86] Z. Otwinowski, W. Minor, "Macromolecular Crystallography Part A," *Methods in Enzymology*, vol. 276, C. W. J. Carter and R. M. Sweet, Eds. Academic Press (New York), 1997, pp. 307–326.
- [87] W. Minor, M. Cymborowski, Z. Otwinowski, and M. Chruszcz, "HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes," *Acta crystallographica. Section D, Biological Crystallography*, vol. 62, no. 8, pp. 859–66, Aug. 2006.
- [88] B. W. Matthews, "Solvent content of protein crystals," *Journal of Molecular Biology*, vol. 33, no. 2, pp. 491–497, Apr. 1968.
- [89] B. W. Matthews, "X-ray Crystallographic Studies of Proteins," *Annual Review of Physical Chemistry*, vol. 27, no. 1, pp. 493–493, Oct. 1976.
- [90] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read, "Phaser crystallographic software," *Journal of Applied Crystallography*, vol. 40, no. 4, pp. 658–674, Aug. 2007.
- [91] T. C. Terwilliger, R. W. Grosse-Kunstleve, P. V. Afonine, N. W. Moriarty, P. H. Zwart, L. W. Hung, R. J. Read, and P. D. Adams, "Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard," *Acta Crystallographica. Section D, Biological Crystallography*, vol. 64, no. 1, pp. 61–9, Jan. 2008.
- [92] P. V. Afonine and P. D. Grosse-Kunstleve, R W Adams, "The Phenix Refinement Framework," *CCP4 Newsletter Contribution 8*, 2005.
- [93] M. M. Harding, "Small revisions to predicted distances around metal sites in proteins," *Acta crystallographica. Section D, Biological Crystallography*, vol. 62, no. 6, pp. 678–82, Jun. 2006.
- [94] K. Hsin, Y. Sheng, M. M. Harding, P. Taylor, and M. D. Walkinshaw, "MESPEUS: a database of the geometry of metal sites in proteins," *Journal of Applied Crystallography*, vol. 41, no. 5, pp. 963–968, Aug. 2008.
- [95] I. Dokmanić, M. Sikić, and S. Tomić, "Metals in proteins: correlation between the metal-ion type, coordination number and the amino-acid residues involved in the coordination," *Acta Crystallographica. Section D, Biological Crystallography*, vol. 64, no. 3, pp. 257–63, Mar. 2008.
- [96] A. C. Storer and R. Menard, "Catalytic Mechanism in Papain Family of Cysteine Peptidases," *Methods in Enzymology*, vol. 244, pp. 486–500, 1994.

- [97] D. C. Tessier, "Structural and Functional Roles of Asparagine 175 in the Cysteine Protease Papain," *Journal of Biological Chemistry*, vol. 270, no. 28, pp. 16645–16652, Jul. 1995.
- [98] J. Drenth, H. M. Swen, W. Hoogenstraaten, and A. L. A. Sluyterman, "A mechanism of papain action," *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen - Series C*, vol. 78, no. 2, pp. 104–110, 1975.
- [99] B. Asboth, E. Stokum, I. U. Khan, and L. Polgar, "Mechanism of Action of Cysteine Proteinases: Oxyanion Binding Site Is Not Essential in the Hydrolysis of Specific Substrates," *Biochemistry*, vol. 24, pp. 606–609, 1985.
- [100] E. L. Smith, J. R. Kimmel, and D. M. Brown, "Crystalline Papain," *Journal of Biological Chemistry*, vol. 207, pp. 533–549, 1954.
- [101] M. R. Alecio, M. L. Dann, and G. Lowe, "The specificity of the S1' subsite of papain," *The Biochemical Journal*, vol. 141, no. 2, pp. 495–501, Aug. 1974.
- [102] C. Achmüller, W. Kaar, K. Ahrer, P. Wechner, R. Hahn, F. Werther, H. Schmidinger, M. Cserjan-Puschmann, F. Clementschitsch, G. Striedner, K. Bayer, A. Jungbauer, and B. Auer, "N(pro) fusion technology to produce proteins with authentic N termini in *E. coli*," *Nature Methods*, vol. 4, no. 12, pp. 1037–43, Dec. 2007.
- [103] A. J. Barrett and N. D. Rawlings, "Families and clans of cysteine peptidases," *Perspectives in Drug Discovery and Design*, vol. 6, no. 1, pp. 1–11, Dec. 1996.
- [104] H. K. Choi, L. Tong, W. Minor, P. Dumas, U. Boege, M. G. Rossmann, and G. Wengler, "Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion," *Nature*, vol. 354, no. 6348, pp. 37–43, Nov. 1991.
- [105] L. Tong, G. Wengler, and M. G. Rossmann, "Refined structure of Sindbis virus core protein and comparison with other chymotrypsin-like serine proteinase structures," *Journal of Molecular Biology*, vol. 230, no. 1, pp. 228–47, Mar. 1993.
- [106] A. E. Gorbalenya and E. J. Snijder, "Viral cysteine proteinases," *Perspectives in Drug Discovery and Design*, vol. 6, no. 1, pp. 64–86, Dec. 1996.
- [107] A. Guarné, J. Tormo, R. Kirchweger, D. Pfistermueller, I. Fita, and T. Skern, "Structure of the foot-and-mouth disease virus leader protease: a papain-like fold adapted for self-processing and eIF4G recognition," *The EMBO journal*, vol. 17, no. 24, pp. 7469–79, Dec. 1998.

- [108] A. Cramer, E. A. Whitehorn, E. Tate, and W. P. Stemmer, "Improved green fluorescent protein by molecular evolution using DNA shuffling," *Nature Biotechnology*, vol. 14, no. 3, pp. 315–319, 1996.
- [109] H. Fukuda, M. Arai, and K. Kuwajima, "Folding of Green Fluorescent Protein and the Cycle3 Mutant," *Biochemistry*, vol. 39, no. 39, pp. 12025–12032, Oct. 2000.
- [110] S. E. Jackson, T. D. Craggs, and J. Huang, "Understanding the folding of GFP using biophysical techniques," *Expert Review of Proteomics*, vol. 3, no. 5, pp. 545–59, Oct. 2006.
- [111] J. Huang, T. D. Craggs, J. Christodoulou, and S. E. Jackson, "Stable intermediate states and high energy barriers in the unfolding of GFP," *Journal of Molecular Biology*, vol. 370, no. 2, pp. 356–71, Jul. 2007.
- [112] J. D. Tratschin, C. Moser, N. Ruggli, and M. A. Hofmann, "Classical swine fever virus leader proteinase Npro is not required for viral replication in cell culture," *Journal of Virology*, vol. 72, no. 9, pp. 7681–4, Sep. 1998.
- [113] D. Mayer, M. A. Hofmann, and J. D. Tratschin, "Attenuation of classical swine fever virus by deletion of the viral Npro gene," *Vaccine*, vol. 22, no. 3–4, pp. 317–328, Jan. 2004.
- [114] T. M. Myers, V. G. Kolupaeva, E. Mendez, S. G. Baginski, I. Frolov, C. U. Hellen, and C. M. Rice, "Efficient translation initiation is required for replication of bovine viral diarrhoea virus subgenomic replicons," *Journal of Virology*, vol. 75, no. 9, pp. 4226–38, May 2001.
- [115] S. P. Fletcher, I. K. Ali, A. Kaminski, P. Digard, and R. J. Jackson, "The influence of viral coding sequences on pestivirus IRES activity reveals further parallels with translation initiation in prokaryotes," *RNA*, vol. 8, no. 12, pp. 1558–1571, Dec. 2002.
- [116] O. Haller, G. Kochs, and F. Weber, "The interferon response circuit: induction and suppression by pathogenic viruses," *Virology*, vol. 344, no. 1, pp. 119–30, Jan. 2006.
- [117] K. Honda, A. Takaoka, and T. Taniguchi, "Type I interferon [corrected] gene induction by the interferon regulatory factor family of transcription factors," *Immunity*, vol. 25, no. 3, pp. 349–60, Sep. 2006.
- [118] M. Iqbal, E. Poole, S. Goodbourn, and J. W. McCauley, "Role for Bovine Viral Diarrhoea Virus Erns Glycoprotein in the Control of Activation of Beta Interferon

by Double-Stranded RNA,” *Journal of Virology*, vol. 78, no. 1, pp. 136–145, Dec. 2003.

- [119] I. Magkouras, P. Mätzner, T. Rümenapf, E. Peterhans, and M. Schweizer, “RNase-dependent inhibition of extracellular, but not intracellular, dsRNA-induced interferon synthesis by Erns of pestiviruses,” *The Journal of General Virology*, vol. 89, no. 10, pp. 2501–6, Oct. 2008.
- [120] L. Hilton, K. Moganeradj, G. Zhang, Y. H. Chen, R. E. Randall, J. W. McCauley, and S. Goodbourn, “The NPro product of bovine viral diarrhea virus inhibits DNA binding by interferon regulatory factor 3 and targets it for proteasomal degradation,” *Journal of Virology*, vol. 80, no. 23, pp. 11723–32, Dec. 2006.
- [121] K. Takahasi, N. N. Suzuki, M. Horiuchi, M. Mori, W. Suhara, Y. Okabe, Y. Fukuhara, H. Terasawa, S. Akira, T. Fujita, and F. Inagaki, “X-ray crystal structure of IRF-3 and its functional implications,” *Nature Structural and Molecular Biology* vol. 10, no. 11, pp. 922–927, 2003.
- [122] B. Y. Qin, C. Liu, S. S. Lam, H. Srinath, R. Delston, J. J. Correia, R. Derynck, and K. Lin, “Crystal structure of IRF-3 reveals mechanism of autoinhibition and virus-induced phosphoactivation,” *Nature Structural and Molecular Biology*, vol. 10, no. 11, pp. 913–921, 2003.
- [123] D. Panne, T. Maniatis, and S. C. Harrison, “Crystal structure of ATF-2/c-Jun and IRF-3 bound to the interferon- $\beta$  enhancer,” *The EMBO Journal*, vol. 23, no. 22, pp. 4384–4393, 2004.
- [124] B. Y. Qin, C. Liu, H. Srinath, S. S. Lam, J. J. Correia, R. Derynck, K. Lin, and S. Francisco, “Crystal Structure of IRF-3 in Complex with CBP,” *Structure (Elsevier)*, vol. 13, pp. 1269–1277, 2005.
- [125] A. I. Dragan, V. V. Hargreaves, E. N. Makeyeva, and P. L. Privalov, “Mechanisms of activation of interferon regulator factor 3: the role of C-terminal domain phosphorylation in IRF-3 dimerization and DNA binding,” *Nucleic Acids Research*, vol. 35, no. 11, pp. 3525–34, Jan. 2007.

## **Vita**

Keerthi Gottipati was born in Hyderabad, India on August 22nd, 1983. She is the third of three children born to Drs. Ranga Rao and Raghava Rani Gottipati. Keerthi attended Anna University in Chennai, India, where she received the degree of Bachelor of Technology in Industrial Biotechnology. Following her under-graduate studies, Keerthi enrolled at the University of Texas Medical Branch in August 2005 and received the degree of Master of Science in Biochemistry and Molecular Biology in December 2008. She then joined the laboratory of Dr. Kay Choi in January of 2009 once again at UTMB, to pursue her doctoral studies in protein biochemistry and structure biology.

Permanent Address: Flat 1309, Turquoise Block,  
My Home Jewel,  
Medinaguda, Hyderabad – 500049  
India

This dissertation was typed by Keerthi Gottipati.