Crystallographic studies of the Crimean-Congo haemorrhagic fever virus and tomato spotted wilt virus N proteins

Stephen D. Carter

Submitted in accordance with requirements for the degree of Doctor of Philosophy

The University of Leeds
Astbury Centre for Structural and Molecular Biology
Institute of Molecular and Cellular Biology

February 2015

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

This research has been carried out by a team, which has included (Stephen D. Carter, Dr. Cheryl Walter, Dr. Eric Bergeron, Dr. Antonio Ariza and Rebecca Surtees). My own contributions, fully and explicitly indicated in the thesis, have been to all research within this thesis except the mutagenesis, mini-genome and phylogenetic analysis. Chapters 2 and 3 in this thesis are based on work based on work from a jointly authored publication;

CARTER, S. D., SURTEES, R., WALTER, C. T., ARIZA, A., BERGERON, E., NICHOL, S. T., HISCOX, J. A., EDWARDS, T. A. & BARR, J. N. 2012b. Structure, function, and evolution of the Crimean-Congo hemorrhagic fever virus nucleocapsid protein. *Journal of virology,* 86**,** 10914-23.

The work in this publication which is directly attributable to me includes the expression, purification, crystallization, structure determination of the CCHFV N protein. The caspase-3 protease cleavage anaysis of the CCHFV N protein is also attributable to me. The contributions from the other authors in this paper include mutagenesis by Dr. Cheryl Walter and Rebecca Surtees, and minigenome analysis by Dr. Eric Bergeron, the phylogenetic analysis in chapter 3 is by Dr. Cheryl Walter, and final refinement of the CCHFV N protein is by Dr. Antonio Ariza.

Acknowledgments

I wish to thank most importantly my supervisors John Barr and Thomas Edwards, for their help, support, direction and advice over the past three years. I would like to thank the States of Jersey and Jerry Knapp for financial support.

I would also like to thank to all the members in the virology lab, especially Carsten Zothner for guidance whilst working in the lab. Also thanks to Arwen Pearson for help with trips to the Diamond Synchrotron, and Chi Trinh for crucial advice on X-ray crystallography.

Special thanks to the Carter family, Uncle Chris, Auntie Sandra, Vikki and Daniel, for there support over the past eight years. Special thanks to my Mum who has been hugely supportive. I dedicate this thesis to my Dad.

Abstract

The *Bunyaviridae* family is a group of segmented negative sense RNA viruses (sNSV) that cause severe disease in both plants and humans. This group of viruses is further classified into the *Orthobunyavirus*, *Tospovirus*, *Nairovirus*, *Phlebovirus* and *Hantavirus* genera.

Crimean-Congo haemorrhagic fever (CCHF) is a potentially fatal tick borne viral meta-zoonotic haemorrhagic disease, which is caused by Crimean-Congo haemorrhagic fever virus (CCHFV), a member of the *Nairovirus* genus. Despite the medical importance and severity of disease, the molecular biology of CCHFV remains poorly understood, primarily because outbreaks are sporadic, and its handling requires the highest containment facilities, bio-safety level 4 (BSL-4).

Tomato spotted wilt virus (TSWV) is a member of the *Tospovirus* genus, which are the only plant infecting members of the bunyaviruses. TSWV causes serious disease in over 80 plant families, including peppers, potatoes, lettuce and tomatos, causing approximately 80 million dollars worth of loss to the food industry per year.

The pathogenesis of CCHFV and TSWV depends on the formation of the ribonucleoprotein (RNP) complex, in which their genomes and antigenome are entirely encapsidated by the virus-encoded nucleocapsid (N) protein. Only in the form of the RNP is the genome replicated, transcribed and packaged into new progeny particles. This work has implemented X-ray crystallography to better understand the structural and functional roles of the CCHFV and TSWV N proteins.

This work presents the 2.1 Å crystal structure of the CCHFV N protein, which shows a high degree of structural homology with the N protein from another sNSV member Lassa virus (LASV), and reveals important insight into sNSV phylogenetics, and the interaction of CCHFV with the host cell. This work also includes the expression, purification, and crystallization of the TSWV N protein. Crystals generated from TSWV N protein diffracted to 2.6 Å and structure determination is ongoing.

Table of Contents

Acknowledgments .... **iii**

Abstract .... **iv**

List of figures .... **ix**

Abbreviations .... **xiv**
  Amino acids .... **xvii**
  Nucleotides .... **xvii**

1. Chapter - Introduction to thesis **1**
  1.1 Negative sense RNA viruses **1**
  1.2 Molecular biology of bunyaviruses **1**
    1.2.2 Bunyavirus genome structure, gene products, and RNA synthesis 3
    1.2.3 Non-structural proteins encoded by the S segment 7
    1.2.4 Non-structural proteins encoded by the M segment 8
    1.2.5 Assembly and release 8
    1.2.6 Reverse genetics of bunyaviruses 10
  1.3 Transmission **11**
  1.4 Introduction to CCHFV and the *Nairovirus* genus **12**
    1.4.1 Reservoirs and Vectors 12
    1.4.2 Molecular and cellular biology of CCHFV N protein 13
      1.4.2.1 Subcellular localization of CCHFV N and its interactions with host cell partners 13
        1.4.2.1.1 MxA 14
        1.4.2.1.2 Caspase-3 protease cleavage of CCHFV N protein 14
      1.4.2.2 Genetic diversity and relationships between strains 15
    1.4.3 Epidemiology and Ecology 17
      1.4.4 Threat of CCHFV emergence and outbreaks 18
    1.4.5 Clinical Manifestations 20
    1.4.6 Treatments 21

1.5 Introduction to TSWV and other tospoviruses **21**
  1.5.1 Tospovirus species **22**
  1.5.2 Transmission **23**
  1.5.3 Epidemiology **23**
  1.5.4 Comparison with animal infecting bunyaviruses **24**
  1.5.5 Cell to cell movement **25**
  1.5.6 RNA silencing **26**
  1.6 Arenaviruses **28**
  1.7 Biology of the bunyavirus structural proteins **29**
    1.7.1 Structural proteins encoded by the M segment 29
    1.7.2 Viral proteins encoded by the L segment 32
    1.7.4 Functions of bunyavirus N proteins in addition to RNA binding 37
  1.8 Architecture of the RNPs from segmented and non-segmented negative sense viruses. **38**
    1.8.1 RSV and VSV 38
    1.8.2 IAV 41
    1.8.3 Similarities and differences between IAV, RSV and VSV 42
  1.9 Architecture of bunyavirus RNPs **43**
    1.9.1 Phlebovirus N protein crystal structures 43
    1.9.2 Orthobunyavirus N protein crystal structures 46
    1.9.3 Hantavirus and tospovirus N proteins 50
    1.9.5 Structure of the LASV NP 50

1.2 Summary of thesis aims **53**

2. Chapter 2 - Materials, Methods, and X-ray Crystallography theory **54**
  2.1 Bacterial strains **54**

2.2 Plasmid vectors 55
2.3 Manipulation of recombinant DNA 55
2.3.1 Primers 55
2.3.2 Polymerase Chain Reaction (PCR) 55
2.3.3 PCR and enzyme digestion cleanup 55
2.3.4 Generation of DNA fragments by restriction endonuclease digestion 55
2.3.5 Purification of DNA fragments 56
2.3.6 Analysis of DNA concentration 56
2.3.7 Ligation of DNA fragments 56
2.4 Transformation of chemically competent *E. coli* bacterial cells 57
2.4.1 Star BL21 (DE3), Gold BL21 (DE3), Rosetta BL21 (DE3), Rosetta2 BL21 (DE3), and DH5α in house competent *E. coli* cells 57
2.4.2 Controls of chemically competent *E. coli* cells 57
2.5 Growth of bacterial cultures and preparation of plasmid DNA 57
2.5.1 Purification of plasmid dsDNA from *E. coli* cells 57
2.5.2 dsDNA sequencing 57
2.6 Protein biochemistry 57
2.6.1 Transformation of bacterial expression strains with pET28a-SUMO CCHFV N 57
2.6.2 Small scale expression of CCHFV N protein 58
2.6.3 Large scale expression of CCHFV N protein 58
2.6.4 Nickel (Ni) Superflow Plus resin protein purification 59
2.6.5 Dialysis of imidazole from CCHFV N protein 59
2.6.6 Proteolysis using the SUMO protease 60
2.6.7 Size exclusion chromatography (SEC) 60
2.6.8 Sodium-dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) 60
2.6.9 Coomassie brilliant blue R-250 staining 61
2.6.10 Bio-Rad assay 61
2.6.11 Circular dichroism 61
2.6.12 Mass spectrometry 61
2.6.13 Phenol and chloroform extraction of RNA from CCHFV N protein 62
2.7 *In vitro* caspase-3 protease cleavage and analysis 62
2.8 Phylogenetic analysis CCHFV N amino acid sequences 63
2.10 Crystallization 64
2.10.1 Preparing N protein for crystallization trials 64
2.10.2 Crystallization factorial screens 64
2.10.3 Optimization of crystals 64
2.10.3.1 Optimization of the CCHFV N protein crystals 64
2.10.3.2 Optimization of the TSWV N protein crystals 65
2.11 X-ray diffraction data collection for CCHFV and TSWV N protein crystals 65
2.11.1 CCHFV N protein crystals 65
2.11.2 TSWV N protein crystals 65
2.12 Selenomethionine derivative protein 66
2.13 Phasing 66
2.13.1 Isomorphous replacement 67
2.13.2 Anomalous scattering 68
2.13.3 Dispersive differences 69
2.13.4 Structure determination 69

2.14 Crystallization and X-ray diffraction theory 70
2.14.1 Crystallization 70
2.15 X-ray diffraction theory 73
2.15.1 Symmetry and the unit cell lattice 73
2.15.2 Diffraction basics 77
2.15.2.1 X-ray waves 77
2.15.2.2 Scattering of an atom 79
2.15.2.3 Bragg's Law of diffraction 81
2.15.2.4 Scattering of an entire molecule 82
2.15.2.5 Ewald sphere 83

2.15.2.6 Calculating electron density ... 85
2.15.3 Data collection ... 87
2.15.3.1 Crystal mounting ... 87
3.15.3.2 X-ray sources ... 88
2.15.3.3 Data processing ... 89
2.15.3.4 Completeness ... 90
2.15.3.5 R-merge ... 91
2.15.3.6 Multiplicity ... 91
2.15.4 Non crystallographic symmetry (NCS) ... 91
2.15.5 Molecular Replacement ... 91
2.15.6 Model building and refinement ... 93

3. Chapter 3 - Structural determination and functional analysis of the CCHFV N protein ... 96
3.1 Introduction ... 96
3.2 Cloning ... 97
3.2.1 Constructing the His$_6$-SUMO CCHFV N ORF expression plasmid ... 97
3.2.2 Cloning of pet28a-SUMO plasmid ... 97
3.3 CCHFV N protein expression and purification ... 99
3.3.1 Expression of His$_6$-SUMO CCHFV N fusion protein in *E. coli* BL21 (DE3) cells ... 99
3.3.2 Solubility of His$_6$-SUMO CCHFV N fusion protein ... 101
3.3.3 Purification of CCHFV N protein with Ni-Superflow Plus resin ... 103
3.3.4 Large-scale purification of CCHFV N protein with Ni-Superflow Plus resin ... 104
3.3.5 Purification of CCHFV N protein using SEC ... 105
3.4 RNA Binding ... 106
3.5 Crystallization of CCHFV N protein ... 107
3.5.1 Factorial crystal screens ... 107
3.5.2 Optimization of factorial CCHFV N protein crystals ... 108
3.6 X-ray diffraction of CCHFV N protein crystals ... 108
3.6.1 X-Ray diffraction of factorial CCHFV N protein crystals ... 108
3.6.2 X-Ray diffraction of optimized CCHFV N protein crystals ... 109
3.7 Self-rotation analysis ... 110
3.8 Circular dichroism ... 111
3.9 The Structure of CCHFV N protein ... 113
3.9.1 Molecular replacement ... 113
3.9.2 Selenomethionine CCHFV N protein ... 113
3.9.3 Heavy atom soaks into native crystals grown from CCHFV N protein ... 116
4.9.4 Electrostatic surface potential ... 123
3.10 Similarity to other sNSV N proteins ... 124
3.11 The N-N dimer interface ... 125
3.12 Consequences of cleavage of the DEVD caspase-3 protease recognition site . 126
3.13 Summary of the CCHFV N protein crystal structure results ... 129

3.14 Discussion ... 130
3.14.1 Cloning ... 130
3.14.2 Expression and purification ... 131
3.14.3 Crystallization and X-ray diffraction of the CCHFV N protein ... 131
3.14.4 Structural comparison of the LASV NP ... 132
3.14.5 Structural comparison between CCHFV N proteins from other strains ... 134
3.14.6 Structural comparison of CCHFV N proteins from strains Baghdad-12 and YL04057 ... 134
3.14.7 Structural comparison of CCHFV N proteins from strains Baghdad-12 and IbAr10200 ... 136
3.14.8 Comparison of electrostatic potential analysis of the CCHFV N proteins from Baghdad-12, YL04057 and IbAr10200 strains ... 139
3.14.9 Functional analysis of selected N protein residues by use of a CCHFV minigenome system performed by Dr. Eric Bergeron, Dr. Cheryl Walter and Rebecca Surtees ... 141

3.14.10 Potential for a gated RNA binding mechanism. 145
3.14.11 CCHFV RNP architecture 147
3.14.12 PISA analysis of dimer interface in *C*2 and $P2_12_12_1$ crystal structures 149
3.14.13 Conformational change in Wang et al CCHFV N protein crystal structures 150
3.14.14 Caspase-3 cleavage of CCHFV N protein 150
3.14.15 CCHFV N protein dsDNA specific endonuclease 153
3.14.16 Possible roles of the CCHFV N endonuclease 154

4. Chapter 4 - TSWV N protein expression, purification and crystallization 155
4.1 Introduction 155
4.2 Cloning 156
4.3 Expression and purification of the TSWN N protein 157
4.3.1 Purification of TSWV N protein with Ni-Superflow Plus resin 157
4.3.2 Large-scale purification of TSWV N protein with Ni-Superflow Plus resin 157
4.3.3 Purification of TSWV N protein using SEC 158
4.3.4 RNA binding of purified TSWV N protein 160
4.3.5 RNA binding analysis of RNase treated TSWV N protein purified using SEC 163
4.3.6 Purification of RNA free TSWV N protein using SEC 163
4.4 Crystallization of TSWV N protein 167
4.4.1 Factorial crystal screens 167
4.4.2 Optimization of factorial tetramer TSWV N protein crystals 168
4.5 X-ray diffraction of TSWV N protein crystals 168
4.5.1 Diffraction of factorial TSWV N protein crystals 168
4.5.2 Optimization of factorial trimer TSWV N protein crystals 172
4.6 Self-rotation of TSWN N protein tetramer crystals 173
4.7 Solving the TSWV N protein crystal X-ray diffraction data 175
4.7.1 Molecular replacement 175
4.7.2 Selenomethionine TSWV N protein 177
4.7.3 Heavy atom soaks into native crystals grown from TSWV N protein 178

4.8 Discussion of TSWV N protein results chapter 184
4.8.1 Introduction 184
5.8.2 Expression and purification 184
4.8.3 RNA binding 185
4.8.4 Crystallization, self-rotation and phasing of the TSWV N protein 185

5. Chapter 5 - Overall conclusions and future directions 191
5.1 Future directions for the CCHFV N protein 191
5.2 Future directions for the TSWV N protein 193
5.3 Futurework 194
5.4 *Bunyaviridae* family RNPs conclusion 195

6. Appendix 199
6.1 Primers for all cloning 199
6.1.1 CCHFV N gene cloning 199
6.1.2 TSWV N gene cloning 199
6.2 Ramachandran plot of *C*2 CCHFV N protein crystal structure 200

7. References 201

List of figures

Figure 1. Schematic of a bunyavirus particle 2

Figure 2. Schematic of the three genomic RNA segments from prototypic viruses across the five genera within the *Bunyaviridae* family 3

Figure 3. Schematic of RNA replication and transcription for members of the *Bunyaviridae* family 4

Figure 4. The model for bunyavirus entry, assembly and budding. 10

Figure 5. Life cycle of CCHFV 13

Figure 6. Adult female dorsal view of a *Hyalomma marginatum* tick 18

Figure 7. Transmission of tospoviruses by thrips 23

Figure 8. Phylogenetic relationships among nsRNA viruses 29

Figure 9. Various bunyavirus surface glycoprotein architectures 30

Figure 10. ECT analysis of the UUKV particles 31

Figure 11. Schematic of the polymerase topology for sNSVs 33

Figure 12. Cryo-EM reconstruction of the IAV polymerase complex 35

Figure 13. Molecular architecture of nsNSV RNPs 39

Figure 14. The crystal structures of the N proteins from VSV and RSV of the order *Mononegavirales* 40

Figure 15. Surface representation displaying the VSV and RSV N protein interfaces involved in oligomerization 40

Figure 16. Molecular architecture of the IAV RNP 42

Figure 17. RVFV N protein hexamer represented in cartoon format 44

Figure 18. Phlebovirus N protein/RNA complexes 45

Figure 19. Model for RVFV RNP formation 46

Figure 20. Crystal structure of BUNV N protein 47

Figure 21. Model for orthobunyavirus RNP formation 48

Figure 22. Orthobunyavirus and phlebovirus N protein/RNA tetramer complexes 49

Figure 23. Model for arenavirus RNP organization 52

Figure 24. Schematic of the complex structure factors for protein with and without heavy atoms 67

Figure 25. Phase diagram of crystal growth 68

Figure 26. The hanging drop and sitting drop methods of vapour diffusion for crystallization 71

Figure 27. Phase diagram of crystal growth. 72

Figure 28. Schematic showing a general orthorhombic unit cell, with edges **a**, **b**, **c** and angles α, β, and ϒ. 74

Figure 29. Schematic of an orthorhombic unit cell with the intersection (022) of two planes. 75

Figure 30. Cubic primitive (P), body-centred or internal (I), and face centred (F) unit cells. 76

Figure 31. Schematic of an X-ray wave (electric field) 77

Figure 32. Schematic of X-ray scattering by an atom 78

Figure 33. Schematic showing the relationship between $\mathbf{S_1}$, $\mathbf{S_0}$ and **S** 80

Figure 34. Schematic of Bragg's Law 82

Figure 35. Schematic of Bragg diffraction in reciprocal space 84

Figure 36. Schematic of the pET28a-SUMO CCHFV N expression construct 97

Figure 37. PCR amplification of the CCHFV N cDNA from pET28a and ligation into the multiple cloning site of pET28a-SUMO. 99

Figure 38. Analysis of expression of His_6-SUMO CCHFV N protein 100

Figure 39. Analysis of His_6-SUMO CCHFV N protein expression in Rosetta BL21 (DE3) *E. coli* cells using a negative control. 101

Figure 40. Analysis of soluble/insoluble fractions of Rosetta and Rosetta2 BL21 (DE3) *E. coli* cells expressing the His_6-SUMO CCHFV N protein. 102

Figure 41. Analysis of solubility His_6-SUMO CCHFV N protein expressed using Rosetta and Rosetta2 expression BL21 (DE3) *E. coli* cells. 103

Figure 42. Analysis of nickel-affinity purification of His_6-SUMO-CCHFV N fusion protein using the spin protocol 104

Figure 43. Analysis of nickel-affinity purification of His_6-SUMO-CCHFV N fusion protein using Ni-Superflow Plus resin 105

Figure 44. Analysis of the CCHFV N protein using SEC 106

Figure 45. RNA binding analysis of the CCHFV N protein purified by SEC 107

Figure 46. Crystallization of CCHFV N protein 107

Figure 47. Optimization of CCHFV N protein crystals using the hanging drop method . 108

Figure 48. X-ray diffraction pattern of CCHFV N *C*2 spacegroup crystals 110

Figure 49. Self-rotation function calculated by MOLREP using the CCHFV N protein *C*2 spacegroup crystals. 111

Figure 50. Circular-dichroism spectrum of monomeric CCHFV N protein 112
Figure 51. Mass spectrometry analysis of the CCHFV N selenomethionine protein 114
Figure 52. X-ray diffraction pattern of the selenomethionine CCHFV N protein crystals. 115
Figure 53. Hg atoms react with three separate cysteine within each globular domain depicted using 2mFo-DFc and mFo-DFc difference maps 117
Figure 54. Crystal structure of the CCHFV N protein solved from *C*2 crystals 119
Figure 55. Ribbon representation of the $P2_12_12_1$ CCHFV N protein crystal structure 121
Figure 56. Manual alteration and refinement of the arm domain in molecule C in the AU of the $P2_12_12_1$ CCHFV N protein crystal using 2mFo-DFc and mFo-DFc difference maps..... 122
Figure 57. Superposition of the three molecules of the CCHFV N protein in $P2_12_12_1$ AU 123
Figure 58. Electrostatic potential analysis of the CCHFV N protein from Baghdad-12... 124
Figure 59. Superposition of the Baghdad-12 CCHFV and LASV N/NP proteins. 125
Figure 60. Cartoon representation of two CCHFV N protomers within the *C*2 crystal lattice. 126
Figure 61. CCHFV N protein remains intact upon cleavage by caspase-3 protease 127
Figure 62. Caspase-3 protease cleavage of fluorogenic substrate 128
Figure 63. Location of the caspase-3 protease cleavage site on the CCHFV N protein .. 128
Figure 64. Superpositioning of the Baghdad-12 CCHFV and LASV N/NP proteins and phylogenetic analysis of sNSVs and nsNSVs by N and L amino acid sequences using the MEGA6 program 133
Figure 65. Superpositioning of the Baghdad-12 and the YL04057 N protein globular domains 135
Figure 66. Superpositioning of the crystal structure of the CCHFV N protein from Baghdad-12 and the monomer form of N protein from the IbAr10200 strain.............. 136
Figure 67. Superpositioning of the crystal structure of the CCHFV N protein from Baghdad-12 and the oligomer form of N protein from the IbAr10200 strain. 137
Figure 68. Conformational differences in the arm domain of the CCHFV N protein...... 138
Figure 69. Electrostatic potential analysis of the CCHFV N protein from Baghdad-12, YL04057 and IbAr10200 strains 140

Figure 70. Identifying important RNA binding residues for analysis using the minigenome system 142
Figure 71. *In vivo* analysis of site-directed CCHFV N protein mutants. 144
Figure 72. RNA gated binding mechanism for LASV NP 147
Figure 73. Crystal structure of the CCHFV N protein from strain IbAr10200 showing homo-oligomerization 149
Figure 74. CCHFV N protein putative dsDNA endonuclease active site 153
Figure 75. PCR amplification of the TSWV N cDNA from a pET-15b plasmid construct . 156
Figure 76. Analysis of elution profile of soluble His$_6$-SUMO TSWV N fusion protein purified using the spin protocol 157
Figure 77. Analysis of elution profile of soluble His$_6$-SUMO TSWV N purified using Ni Superflow Plus resin 158
Figure 78. Analysis of the non-RNase treated TSWV N protein using SEC 159
Figure 79. SDS-PAGE analysis of the non-RNase treated TSWV N protein purified using the S75 gel filtration column 160
Figure 80. RNA binding analysis of the TSWV N protein purified by SEC 161
Figure 81. Analysis of the RNase treated TSWV N protein using SEC 162
Figure 82. Analysis of elution profile of RNase treated TSWV N protein following purification using a Superdex S75 gel filtration column 162
Figure 83. RNA binding analysis of the RNase treated TSWV N protein purified by SEC 163
Figure 84. Analysis of the RNA free TSWV N protein using SEC 165
Figure 85. Analysis of elution profile TSWV N protein following purification using Superdex S75 and S200 gel filtration columns 166
Figure 86. Crystallization of trimer TSWV N protein using the sitting drop method of crystallization 167
Figure 87. Crystallization of tetramer of TSWV N protein using the sitting drop and hanging drop method of crystallization 168
Figure 88. X-ray diffraction pattern of tetramer TSWV N crystal grown using 0.2 M Trimethylamine N-oxide dehydrate, 0.1 M Tris pH 8.5, 20% w/v Polyethylene glycol monomethyl ether 2,000 169
Figure 89. X-ray diffraction pattern of tetramer TSWV N protein crystals 171

Figure 90. Self-rotation function calculated by MOLREP using the TSWV N protein crystals...174

Figure 91. HHpred local alignment between the predicted secondary structure of the TSWV N protein and protein structures submitted onto the PDB 176

Figure 92. Crystallization of tetramer TSWV N selenomethionine protein using the sitting drop method...177

Figure 93. X-ray diffraction pattern of a tetramer TSWV N selenomethionine crystal grown using 0.2 M Trimethylamine N-oxide dehydrate, 0.1 M Tris pH 8.5, 20% w/v Polyethylene glycol monomethyl ether 2,000..178

Figure 94. X-ray diffraction pattern of tetramer TSWV N protein crystal grown using 0.01 M zinc sulphate, 25% v/v PEG-550 MME, 0.25 M MES pH 6.5, and soaked with 100 µM of samarium nitrate 182

Figure 95. Anisotropy of TSWV N protein crystals soaked with samarium nitrate 183

Figure 96. Interpretation of TSWV N protein $P2_1$ self-rotation results..........................187

Figure 97. Pairwise Sequence Alignment (PROTEIN) of the TSWV and LEAV N protein amino acid sequences 188

Figure 98. Ramachandran plots for CCHFV N protein crystal structure 200

List of tables

Table 1. Conserved nucleotide sequence of the 3' and 5' terminal nucleotides from the S, M and L segments of prototypic bunyaviruses..5

Table 2. The seven crystal systems 76

Table 3. Data collection and refinement statistics for *C*2 CCHFV N protein native and derivative crystals..118

Table 4. Data collection and refinement statistics for $P2_12_12_1$ CCHFV N protein crystals 120

Table 5. Data collection, integration and scaling statistics for TSWV N protein crystals 172

Table 5. List of heavy atoms used to soak into native TSWV N protein crystals 180

Abbreviations

APS	Ammonium persulphate
ATP	Adenine tri-phosphate
BCCV	Black Canal Creek virus
BSL-2	Bio-Safety Level 2
BLAST	Basic local alignment tool
bp	Base pair
BTP	Bis-Tris Propane
BUNV	Bunyamwera virus
C-terminal	Carboxy terminal
CCHF	Crimean-Congo haemorrhagic fever
CCHFV	Crimean Congo haemorrhagic fever virus
CD	Circular dichroism
cDNA	Complementary DNA
CHAPS	3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate
Da	Dalton
DEV	Doubly enveloped viruses
DNA	Deoxyribonucleic acid
DNase 1	Deoxyribonulcease 1
dNTP	Dinucleotide tri-phosphate
dsDNA	Double stranded DNA
dsRNA	Double stranded RNA
DTT	Dithiothreitol
DUGV	Dugbe virus
E. coli	*Escherichia coli*
EDTA	Ethylyenediaminetraacetic acid
eGFP	Enhanced green fluorescent protein
ECT	Electron Cryo-Tomography
EM	Electron microscopy
ER	Endoplasmic reticulum
ERGIC	ER-Golgi intermediate compartment
eV	Electron volts
FT	Fourier transform
GBNV	Groundnut bud necrosis virus
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

Hg	Mercury
hr	Hour/s
IAV	Influenza A virus
IFN	Interferon
IGR	Intergenic region
INSV	Impatiens necrotic spot virus
IPTG	Isopropyl β-D-1-thiogalactopyranoside
JUNV	Junin virus
K	Kelvin
kDa	Kilo Dalton
L	Polymerase
LACV	La Crosse virus
LASV	Lassa virus
LB	Luria Bertani
LCMV	Lymphocytic choriomeningitis virus
LEAV	Leanyer virus
MAD	Multi-wavelength anomalous dispersion
MES	2-(*N*-morpholino)ethanesulfonic acid
$MgCl_2$	Magnesium chloride
min	Minute/s
MIR	Multiple isomorphous replacement
MIRAS	Multiple isomorphous replacement with anomalous scattering
miRNAs	Micro RNAs
MME	Monomethyl ether
MOPS	3-(N-morpholino)propanesulfonic acid
MR	Molecular replacement
MRCA	Most recent common ancestor
mRNA	Messenger RNA
N	Nucleocapsid
N-terminal	Amino terminal
NaCl	Sodium chloride
NCS	Non crystallographic symmetry
Ni	Nickel
NP	Nucleoprotein
NS	Negative sense

NSDV	Nairobi sheep disease virus
NSm	Non-structural M segment
nsNSV	Non-segmented negative sense RNA viruses
nsRNA	Negative sense RNA
NSs	Non-structural S segment
NTRs	Non-translational region
ORF	Open reading frame
OTU	Ovarian tumor
PAGE	Polyacrylamide gel electrophoresis
PARP	Poly (ADP-ribose) polymerase
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
PICV	Pichinde virus
RdRp	RNA dependent RNA polymerase
RMSD	Root mean standard deviation
RNA	Ribonucleic acid
RNP	Ribonucleoprotein complex
RSS	RNA silencing suppressor
RStV	Rice stripe virus
RSV	Respiratory syncytial virus
SAD	Single-wavelength anomalous dispersion
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
SEV	Singly enveloped viruses
SIR	Single isomorphous replacement
SIRAS	Single isomorphous replacement with anomalous scattering
siRNAs	Small interfering RNAs
sNSV	Segmented negative sense RNA viruses
ssRNA	Single stranded RNA
SUMO	Small ubiquitin-related modifier
TEMED	Tetramethylethylenediamine
THOV	Thogoto virus
tRNA	Transfer RNA
TSWV	Tomato spotted wilt virus

TULV	Tula virus
Ub	Ubiquitin
UUKV	Uukuniemi virus
UV	Ultraviolet
VSV	Vesicular somatitis virus
WHO	World Health Organisation
Z	Matrix protein

Amino acids

Amino acid	*Single letter code*	*Three letter code*
Alanine	A	ALA
Arginine	R	ARG
Asparagine	N	ASN
Aspartic acid	D	ASP
Cysteine	C	CYS
Glutamic acid	E	GLU
Glutamine	Q	GLN
Glycine	G	GLY
Histidine	H	HIS
Isoleucine	I	ISO
Leucine	L	LEU
Lysine	K	LYS
Methionine	M	MET
Phenylalanine	F	PHY
Proline	P	PRO
Serine	S	SER
Threonine	T	THR
Tryptophan	W	TRP
Tyrosine	Y	TYR
Valine	V	VAL

Nucleotides

Adenine	A		
Cytosine	C	Uracil	U
Guanine	G	Thymine	T

1. Chapter - Introduction to thesis

1.1 Negative sense RNA viruses

Negative sense (NS) RNA viruses replicate and transcribe their genomic RNA either in the cytoplasm or the nucleus by means of a mega-dalton RNA-protein complex, which serves to protect the RNA during delivery into cells and is the template for viral polymerase during RNA synthesis. The viral RNA genome can exist as non-segmented or segmented, necessitating the classification of these viruses into two groups. The non-segmented RNA viruses (nsNSV) comprise four families; namely the *Rhabdoviridae* family (including vesicular stomatitis virus (VSV) and rabies virus), *Paramyxoviridae* family (including measles virus and respiratory syncytial viruses (RSV)), *Filoviridae* family (including Ebola and Marburg viruses) and the *Bornaviridae* family (including Borna disease virus). The segmented negative sense RNA viruses (sNSV) comprise three families, the two segmented *Arenaviridae* family (including Lassa virus (LASV) and Machupo virus), the three segmented *Bunyaviridae* family (including CCHFV) and the 8 segmented *Orthomyxoviridae* family (including influenza A virus (IAV)). *Tenuivirus* and *Emeravirus* genera consist of four and five segmented viruses respectively, and are not yet assigned to a family.

The pathogenesis of nsRNA viruses depends on the formation of the ribonucleoprotein (RNP) complex, which encapsidates the genome and antigenome by the virus-encoded nucleocapsid (N) protein. Only in the form of the RNP is the genome replicated, transcribed and packaged into new progeny particles. Together with the viral RNA, the encoded N protein forms the RNP, and is central to replication of both sNSV and nsNSV. A common feature of these RNA viruses is the employment of RNA synthesis machinery comprised of the virally encoded RNA dependent RNA polymerase (RdRP), which travels along the RNP to produce copies of the viral RNA template in two synthetic modes, namely messenger RNA (mRNA) transcription or genome replication.

1.2 Molecular biology of bunyaviruses

The *Bunyaviridae* family is a large group of sNSVs and is divided into the *Orthobunyavirus*, *Tospovirus*, *Nairovirus*, *Phlebovirus* and *Hantavirus* genera, which together contain at least 350 named isolates. Classification of this family is based on possession of a single stranded three-segmented RNA genome of NS or ambisense polarity, a replication strategy that takes place in the cytoplasm, and an enveloped spherical virion of 80-120 nm, that matures in the Golgi complex (Elliott, 2009). The classification of the *Bunyaviridae* family into individual genera is based on serological relationships, and presence of genus-specific biochemical characteristics of individual proteins such as their size and amino acid sequence. Segment coding strategy used to encode these proteins, the size of RNA segments, and presence of conserved sequence elements of the segment termini also can have implications for genera classification (Elliott, 2009).

The lipid-enveloped membrane of the viral particle contains protruding spike proteins, consisting of disulphide linked viral heterodimer glycoproteins Gn and Gc. Both these glycoproteins are produced from a single polypeptide cleaved by host cell proteases (Bergeron et al., 2007b). Particle maturation results in infectious viruses primed for new rounds of replication in uninfected cells, and requires the assembly of viral genomic RNA, N protein and numerous copies of the polymerase (L) protein into the RNP complex (Figure 1) (Elliott, 1990).

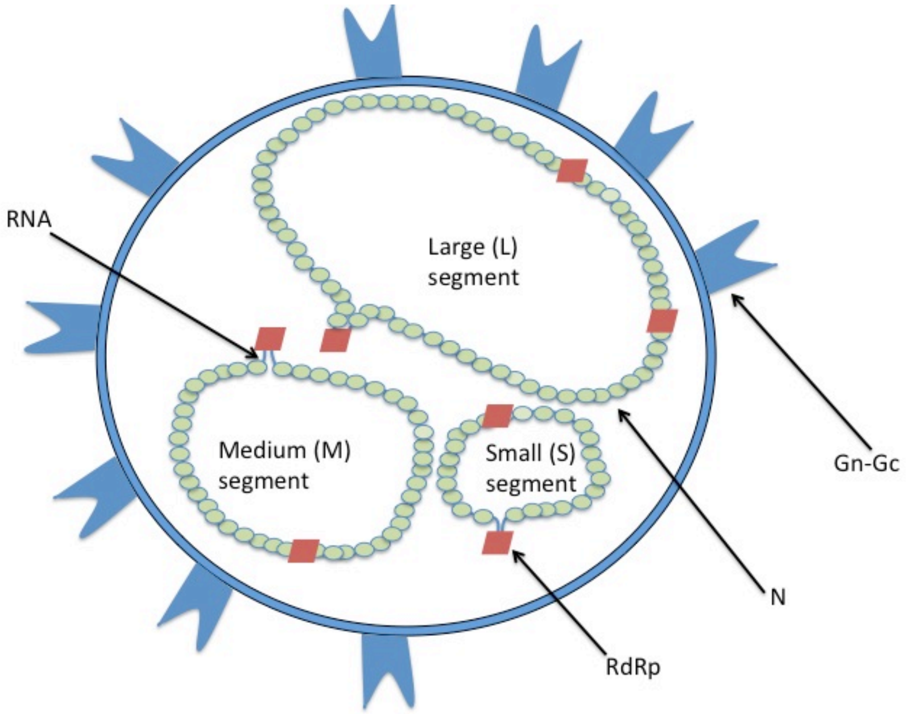


Figure 1. Schematic of a bunyavirus particle. (A) Generally bunyavirus virions comprise a spherical lipid envelope, with embedded spikes comprising Gn–Gc heterodimers positioned in genus-specific arrays on the external membrane surface. Within the interior of the virion the bunyavirus genome comprises three RNA segments (S, M and L), wrapped in the viral N protein.

1.2.2 Bunyavirus genome structure, gene products, and RNA synthesis

The RNA segments of bunyaviruses are named small (S), medium (M), and large (L) reflecting their relative nucleotide length (Figure 2). All bunyaviruses encode an RdRp, or L protein from the L segment, an N protein from the S segment, and a polyprotein precursor from the M segment (Figure 2). The interactions between all these proteins are critical for the formation of the RNP and infectious virus and the section starting from page 30 describes the structural biology of the bunyavirus L protein, N protein and glycoproteins to highlight the parallels and differences between the genera of the *Bunyaviridae* family, and several additional segmented and non segmented negative sense viruses.

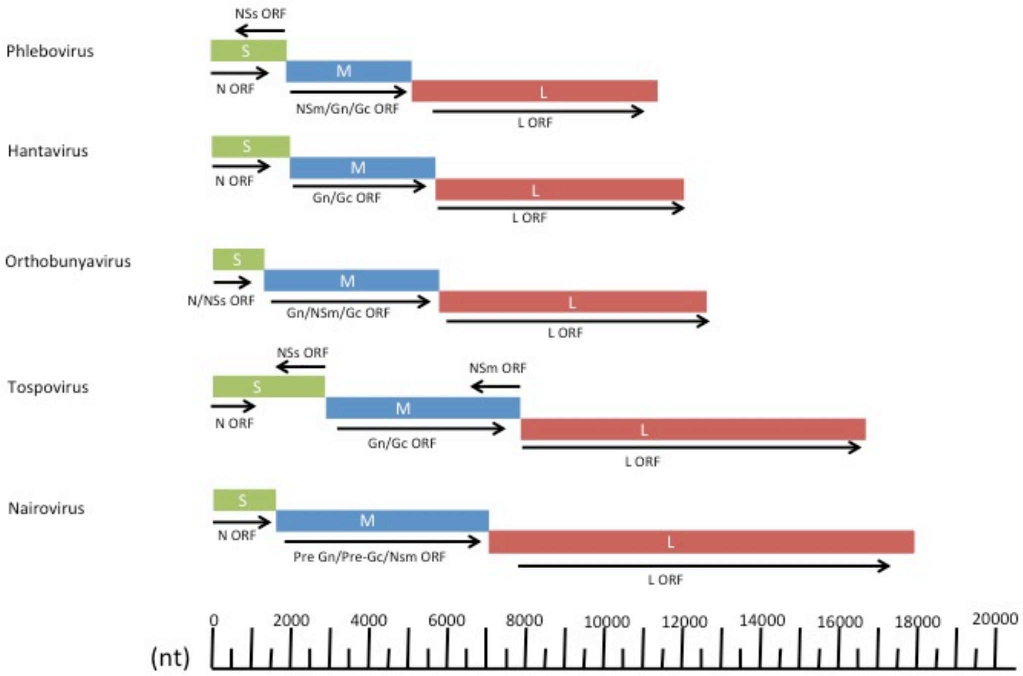


Figure 2. Schematic of the three genomic RNA segments from prototypic viruses across the five genera within the *Bunyaviridae* family. mRNA transcripts encoded by NS templates are indicated by arrows below each segment, whereas ambisense templates synthesized from positive sense coding strategies are denoted by arrows above the segments. The figure was adapted from (Walter and Barr, 2011).

The RNA genome is copied using two different RNA-synthesis activities: mRNA transcription and RNA replication. RNA replication produces full-length complementary copies of both the genomic and antigenomic RNA, whereas transcription products generally are truncated at their 3′ ends relative to the genome template and are extended at their 5′ ends by a host derived capping oligoribonucleotide (Figure 3) (Walter and Barr, 2011).

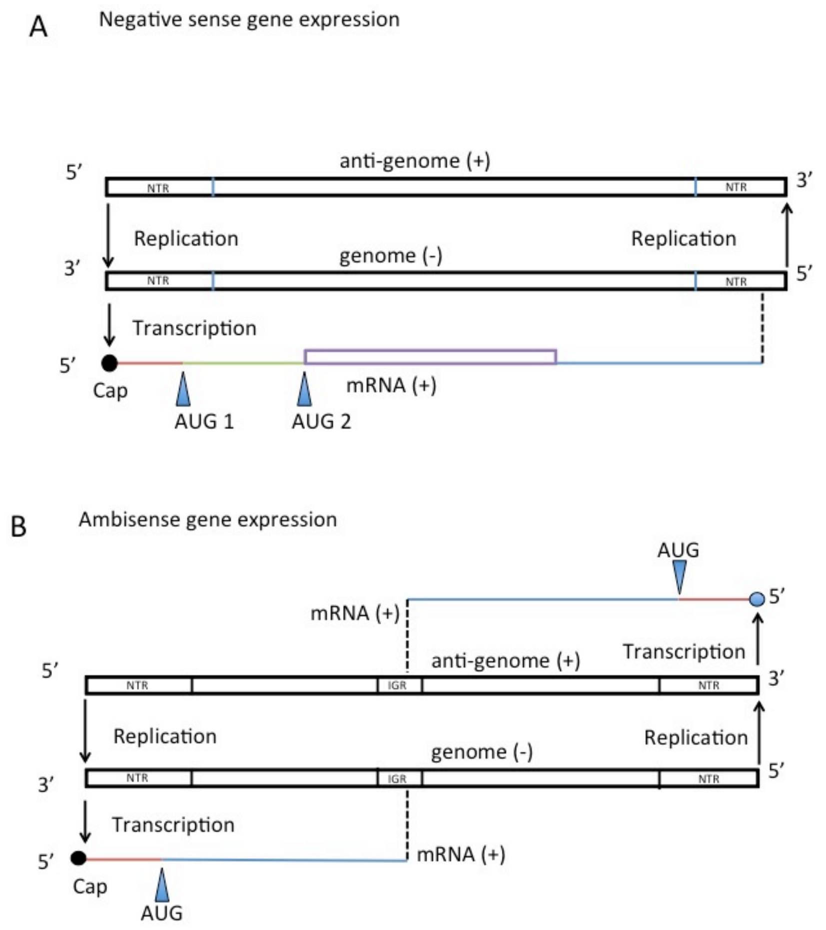


Figure 3. Schematic of RNA replication and transcription for members of the *Bunyaviridae* family. (A) Bunyaviruses which possess genomic RNA segments that contain a single transcriptional unit flanked by non translational regions (NTRs) employ a NS coding strategy for transcription and replication. The transcriptionally silent positive sense antigenome is replicated to generate newly synthesized negative sense genomes, which are packaged into new virus particles. The transcriptional unit of a number of bunyaviruses consists of two overlapping open reading frames (ORF), with two alternating initiation codons. (B) Transcription of mRNAs from both genomic and antigenomic RNA strands through an ambisense coding strategy is exhibited by some bunyaviruses, with transcription-termination signals positioned in a central intergenic region (IGR). The figure was adapted from (Walter et al., 2011).

Contrasting with the nsNSVs - which possess multiple transcriptional units, arranged in a linear array, and sequentially copied - each bunyavirus RNA segment contains one transcriptional unit. Orthobunyaviruses mostly contain a second ORF within the S segment mRNA, encoding a non-structural protein (Figure 2). The significance of the two alternative reading frames is that two proteins can be expressed from one RNA segment, for example the orthobunyaviruses encode from their S segment an N protein and a non-structural protein (NSs) (Weber et al., 2002). Alternative ORFs are also used by tospoviruses (S segment and M segment) and phleboviruses (S segment), by using an ambisense mechanism for transcription to encode for an additional viral protein (Figure 2). Ambisense replication strategies where at least one RNA segment utilizes ambisense nucleotide coding to produce viral protein are employed by four genera (*Arenavirus*, *Phlebovirus*, *Tospovirus* and *Tenuivirus*), distributed between *Arenaviridae* and *Bunyaviridae* families and a small nonclassified genus of plant viruses, the tenuiviruses (Nguyen and Haenni, 2003, Bucher et al., 2003a).

Genus (prototype)	
Orthobunyavirus (Bunyamwera virus)	3' - UCAUACCAUG… 5' - AGUAGUGUGC…
Phlebovirus (Rift Valley fever virus)	3' - UGUGUUUC… 5' - ACACCAAAG…
Nairovirus (Dugbe virus)	3' - AGAGUUUCU… 5'- UCUCAAAGA…
Hantavirus (Hantaan virus)	3'- AUCAUCAUCUG… 5' - UAGUAGUAAGC…
Tospovirus (Tomato spotted wilt virus)	3' – UCUCGUUA… 5' – AGAGCAAU…

Table 1. Complementarity of the 3' and 5' terminal nucleotides from the S, M and L segments of prototypic bunyaviruses.

Located at the 3' and 5' termini of the viral genome within all the genera of the *Bunyaviridae* family each RNA segment strand contains non-translated regions (NTRs) surrounding the single transcriptional unit. Reverse genetics systems have shown that the four animal infecting bunyaviruses contain NTRs which display cis acting signals and involved in the synthesis of RNA and segment packing (Bridgen and Elliott, 1996, Dunn et al., 1995, Flick et al., 2003a, Flick et al., 2004, Flick et al., 2002, Flick et al., 2003b, Kohl et al., 2006, Lopez et al., 1995). The 3' and 5' nucleotides located at the extreme ends of the NTRs show complementarity (Table 1), and are generally highly conserved within each genus, for example the orthobunyaviruses possess RNA segments containing the same 11 nucleotides located at their 3' to 5' termini (Table 1) (Walter and Barr, 2011).

The nucleotide complementarity between the 3′ and 5′ NTRs of each segment allows the possibility that the NTRs may interact through canonical Watson–Crick base pairing forming a "panhandle" structure. Further evidence supporting the NTR interaction is required for segment function and has been reported through both biochemical techniques and direct visualization suggesting the bunyavirus RNA segments exhibit a circular appearance within infected cells and virus particles (Raju and Kolakofsky, 1989, Obijeski et al., 1976, Pettersson and von Bonsdorff, 1975). For example the RNA segment panhandle structure in hantaviruses has been demonstrated to be necessary for efficient and specific binding of the N protein (Mir and Panganiban, 2004, 2005, 2006, Mir et al 2006). Mir et al demonstrated that each panhandle is at least 23 nucleotides long and is bound to the trimeric N protein with high affinity (Mir and Panganiban, 2004, Mir et al., 2006, Mir and Panganiban, 2006, Mir and Panganiban, 2005).

Owing to complementary at the genomic termini, the 3′ and 5′ terminal sequences make up the promoters required for replication and transcription. Findings for Uukuniemi virus (UUKV) (phlebovirus) and Bunyamwera virus (BUNV) (orthobunyaviru*s*) have shown the terminal interaction is strongly implicated as a functional requirement for RNA transcription and replication (Flick et al., 2004). Additionally for BUNV the presence of specific nucleotides located at both ends of the RNA template are needed for transcription, whereas replication of the viral genome depends on inter-terminal complementarity with no apparent sequence specificity (Barr et al., 2005, Barr and Wertz, 2005, Kohl et al., 2004). Furthermore the replication and transcription promoters mapped for BUNV have been shown to incorporate distinct but not overlapping sets of nucleotide sequences (Barr and Wertz, 2004, Barr and Wertz, 2005, Barr et al., 2005, Barr et al., 2003, Kohl et al., 2004). The internal region of the NTR has been shown to be segment specific not only between segments of the same virus, but also between members of the same genus. The roles of these segment specific sequences have been shown to involve regulation of RNA synthesis for both BUNV and Rift Valley fever virus (RVFV) (Barr et al., 2003) (Gauliard et al., 2006). The problem with these *in vivo* studies however is that they have analysed viral replication/transcription and the panhandle structure using mini-genome systems, and involve naked RNA transfected into cells which does not occur during the bunyavirus life cycle; this is because the viral RNA is always coated by N protein.

It is currently unknown how the bunyavirus N protein coated RNA can participate in base-pairing, whist forming a panhandle structure and still be protected in a closed RNA binding groove which can allow the L protein access to the RNA during replication and transcription. One possibility is panhandle formation may occur via extensive protein rearrangements within the single stranded RNA (ssRNA) RNP complexes at the 3’ and 5’ NTRs after access to the RNA by the L protein has occurred, which may accommodate two protected RNAs inside a double stranded RNA (dsRNA) protein complex. Interestingly the binding of hantavirus N to the panhandle *in vitro* has been reported to be an entropy-driven process, which resulted in an initial stable N-RNA interaction followed by a conformational change in the N protein (Mir and Panganiban, 2005).

Since bunyaviruses are not capable of capping their own transcripts the viral mRNAs are not initiated *de novo*, and depend on primers generated from capped host-cell mRNAs by the L protein in a cap snatching mechanism. All members of the sNSVs utilize a cap snatching mechanism for transcription. The notion that cap snatching is highly conserved and is a common feature to all members of the *Bunyaviridae* family, *Orthomyxoviridae* family, and *Arenaviridae* family, has been highlighted by evidence which suggests tomato spotted wilt virus (TSWV) (tospovirus) and IAV share a similar preference for certain capped-RNA primers (van Knippenberg et al., 2005, Geerts-Dimitriadou et al., 2011). Initiation of transcription using these host cell derived capped primers utilizes a ‘prime and realign’ mechanism in which a capped leader RNA primer for transcription is extended for a few nucleotides and then released. Progressive elongation proceeds only after backward re-alignment by virtue of the viral 3′ terminal sequence repeats. *In vivo* studies have shown the TSWV L protein displays preference for cellular mRNA molecules containing residues with base complementary of as little as 1 nucleotide between the mRNA and the 3′ template sequence (Garcin et al., 1995, Duijsings et al., 2001).

1.2.3 Non-structural proteins encoded by the S segment

Most members of the *Orthobunyavirus*, *Tospovirus* and *Phlebovirus* genera encode a NSs protein from the S segment, whose primary role is modulating the host cell anti-viral response through diverse interactions with innate immunity pathways. Because of the absence of a second protein encoding ORF within hantavirus and nairovirus S segments it is interesting to speculate that viruses within these groups make up for the absence of the NSs protein by other means. For example the L proteins of CCHFV and Nairobi sheep disease virus (NSDV) have been shown to contain OTU domains (discussed below on page 36), which have been implicated in down regulating the interferon response (Bakshi et al., 2013, Holzer et al., 2011, van Kasteren et al., 2012). There is also evidence to support the possibility that some sNSV N proteins contain additional features encoded by their considerably larger ORFs. The LASV nucleoprotein (NP) has been shown to have 3′-5′ exoribonuclease and exodeoxyribonuclease activities to down regulate the cellular interferon response, whilst the hantavirus N protein has been shown to function in cap binding and replace the entire eIF4F complex in translation. In addition the CCHFV N protein has been shown to have DNA endonuclease activity, however the functional significance of this remains to be determined (Guo et al., 2012, Hastie et al., 2011a, Panganiban and Mir, 2009).

1.2.4 Non-structural proteins encoded by the M segment

The M segment of the *Orthobunyavirus*, *Phlebovirus* and *Tospovirus* genera also encode a non-structural (NSm) protein. There is evidence suggesting the nairoviruses also generate an NSm protein through proteolytic processing (Bergeron et al., 2012, Bergeron et al., 2007b, Altamura et al., 2007). Whereas the orthobunyavirus and phlebovirus NSm protein is cleaved by cellular proteases from the same polyprotein precursor that yields the Gn and Gc proteins, the tospovirus NSm protein is translated from a separate mRNA encoded by the positive sense antigenome. The NSm protein of orthobunyaviruses is found inside Golgi-associated viral factory tubular structures within infected cells where it is thought to play a role in virus assembly (Shi et al., 2006) (Fontana et al., 2008). The NSm protein of TSWV on the other hand is thought to facilitate the cell to cell transmission through the plant cell wall and the plasmodesmata (Kormelink et al., 1994), whereas as the phlebovirus NSm protein has been shown to be dispensable to the virus (Gerrard et al., 2007), and is thought to play an accessory function in the regulation of apoptosis (Won et al., 2007).

1.2.5 Assembly and release

One of the major distinctions between bunyaviruses and most other nsRNA viruses is that they are generally thought to utilize tube like factories for virus production that are built around the Golgi complex (Kikkert et al., 1999) (Kuismanen et al., 1984) (Murphy et al., 1973) (Matsuoka et al., 1991) (Novoa et al., 2005) (Figure 4). For the animal-infecting bunyaviruses, it is generally thought that particle assembly involves budding of RNPs into the vacuolised lumen of Golgi cisternae stacks, and to a lesser extent, the endoplasmic reticulum (ER)-Golgi intermediate compartment (ERGIC) (Overby et al., 2007a). The *Hantavirus* genus however has been shown to be the exception, with reports that the maturation strategy of Sin Nombre virus and Black Canal Creek virus (BCCV) (New World Hantaviruses) facilitates budding at the plasma membrane of polarised epithelial cells, raising the possibility that hantaviruses utilise alternative assembly pathways (Goldsmith et al., 1995, Ravkov et al., 1997). There has also been an additional report that RVFV particles assemble and bud at the plasma membrane in primary liver cells, opening the question as to whether assembly may be cell-type dependent (Walter and Barr, 2011). In the case of the plant infecting TSWV, particle assembly in plant cells involves enwrapment of RNPs by an entire Golgi stack, leading to the formation of doubly enveloped virus (DEV) particles. Singly enveloped viruses (SEV) are formed in large membrane bound vesicles by the fusion of DEV particles with each other and with ER-derived membranes (Ribeiro et al., 2009).

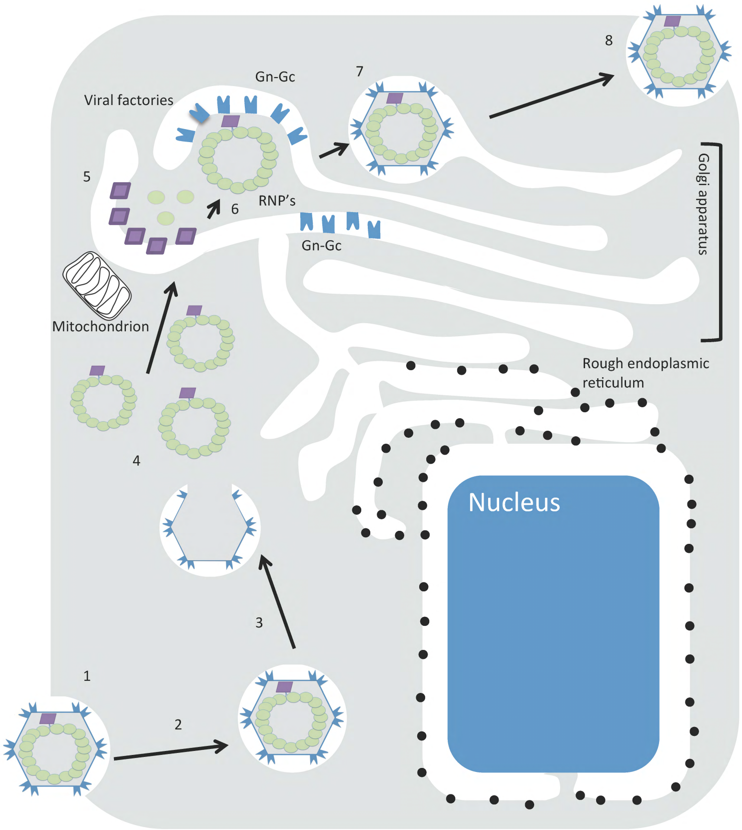


Figure 4. The model for bunyavirus entry, assembly and budding. Clathrin mediated endocytosis is the predominant route of virus entry (1), progression into the cell occurs through early and late endosomes. (2), and infectious RNPs are released following late endosomal acidification (3), and accumulate in the Golgi (4), accumulation of both cellular and viral proteins in close association with Golgi stacks to form viral factories (5) coupled to both mitochondria and the rough endoplasmic reticulum. The Golgi stacks form globular structures serving to protect viral mRNA transcripts and newly replicated S, M, and L complexes, which are generated by the viral polymerase. New bunyavirus particles are generated (6), by association of RNPs with Gn-Gc glycoprotein heterodimers. Gn-Gc heterodimers accumulate in the Golgi and pass through the *trans*-Golgi compartment where modification takes place by the addition of *N*-linked carbohydrates (7). Morphological changes resulting in full infectious virus occur through a period of transition whilst passing through the Golgi apparatus and upon secretion out of the cell (8). The figure was adapted from Walter and Barr (2011).

1.2.6 Reverse genetics of bunyaviruses

Reverse genetic systems capable of generating infectious bunyavirus particles from cloned complementary DNA (cDNA) have become reliable techniques to study this group of medically important viruses, and allows the identification of genetic determinants of transcription, replication, assembly and pathogenesis. For bunyaviruses, the viral genomic RNAs, as well as the minimum requirements for genome transcription, the viral RdRp and N protein must be provided to initiate the viral replication cycle. Furthermore full reverse genetic systems that allow the rescue of infectious virus have been developed for several members of the *Bunyaviridae* family including BUNV (Bridgen and Elliott, 1996), LACV (Blakqori and Weber, 2005), Akabane virus (orthobunyavirus) (Ogawa et al., 2007) and RVFV (Ikegami et al., 2006). Bacteriophage T7 polymerase promoters or mammalian polymerase I (PolI) promoters that drive transcription of the virus genome RNA are commonly used in bunyavirus reverse genetics systems, to generate the viral RdRp, N protein and accessory proteins. These proteins then replicate, transcribe and package the viral genome, allowing rescue of infectious virus.

Reverse genetic systems for rescue of infectious CCHFV and TSWV are not available. An alternative mini-genome version for CCHFV is more widely used and is able to analyse replication, transcription and RNP formation under safer Bio-Safety Level 2 (BSL-2) conditions (Bergeron et al., 2010a). The mini-genome system involves introducing a viral template, in which the viral M segment ORF is replaced with a reporter gene such as enhanced green fluorescent protein (eGFP), luciferase or chloramphenicol acetyltransferase. These reporter genes are flanked by viral NTRs, which contain the necessary signals to allow transcription, replication and subsequent encapsidation of the viral template. The system works by expressing the minimal viral components for virus replication and transcription, namely the N and L proteins, which are under the control of a strong PolI promoter within mammalian cells where they self-assemble into a functional RNP. The RNP both orders and protects the viral RNA, and represents the active template for mRNA synthesis, thus providing a powerful tool to study these processes, without producing infectious virus, and enabling genetic manipulation of viral proteins. There is currently no mini-genome system for tospoviruses.

1.3 Transmission

The lifecycles of the orthobunyaviruse*s*, nairoviruses and phleboviruses rely on the natural replication in haematophagous arthropods (such as mosquitos, midges, sandflies and ticks), and warm-blooded species. The tospoviruses are transmitted by the thrips species of arthropod and are the only plant pathogens within the *Bunyaviridae* family, which offers another basis for classification. Hantaviruses are not vectored by arthropods and instead maintain a persistent infection in rodents (Elliott, 2009, Asher et al., 1988). Bunyaviruses are able to multiply, and spread within all these vectors both transovarially (vertical transmission) (Tesh and Modi, 1987, Miller et al., 1977) and venereally (horizontal transmission) (Thompson and Beaty, 1978).

The aim of this thesis is to elucidate the structure and function the CCHFV and TSWV N proteins and it is therefore important to introduce the Nairovirus and Tospovirus genera to the reader.

1.4 Introduction to CCHFV and the *Nairovirus* genus

CCHFV is grouped within one of seven serogroups of the *Nairovirus* genus, which according to the latest International Committee for the Taxonomy of Viruses report comprise a total of 34 viruses. Of these viruses only CCHFV has serious clinical importance and is the only life threating and human disease causing nairovirus (Burt et al., 1996a). Dugbe virus (DUGV), which belongs to the Nairobi sheep disease group, has high serological and genetic similarity to CCHFV, and has been reported to cause in rare cases a mild febrile illness and thrombocytopenia (Burt et al., 1996b), but is not of clinical importance. Because DUGV causes a similar disease although lacking haemorrhagic fever in immunosuppressed mice to that of CCHFV and both viruses share a close relationship, DUGV has been suggested to be a good potential model for studying CCHFV biology (Coates and Sweet, 1990). The Nairobi sheep disease group also contains NSDV, which is a pathogen of sheep and goats (Bakshi et al., 2013). Hazara virus has also been shown to be closely related to CCHFV (Casals and Tignor, 1974, Foulke et al., 1981), and has not been reported to cause human disease and has also been proposed to be a good model for studying the cellular biology of CCHFV, in safer BSL-2 facilities (Dowall et al., 2012).

1.4.1 Reservoirs and Vectors

CCHFV circulates in an enzootic tick-vertebrate-tick cycle, infecting domestic and wild animals carrying mature ticks, which transmit the infection to large vertebrates such as livestock. Although infection in mammals can last for up to two weeks, they do not display any clinical symptoms (Figure 5) (Gunes et al., 2011). CCHFV is amplified and maintained in nature by numerous domestic and wild animals carrying immature ticks (nymphs) (Hoogstraal, 1979, Appannanavar and Mishra, 2011) (Figure 20), and has been isolated from sheep (Yen et al., 1985), hares (L'Vov D et al., 2002), hedgehogs (Whitehouse, 2004b), a *Mastomys* spp of mouse (Saluzzo et al., 1985), and even domestic birds (Shepherd et al., 1987a, Shepherd et al., 1987b).

Around 31 species of ticks in seven genera of the *Ixodidae* family (hard ticks) have been found to harbor CCHFV and the most efficient and common vectors for CCHFV appear to be members of the *Hyalomma* genus. Transmission of CCHFV also less frequently occurs by other ixodid ticks, including members of the genera *Rhipicephalus*, *Boophilus*, *Dermacentor*, and *Ixodes* (Bell-Sakyi et al., 2012). The factors involved in human transmission are discussed below on page 19.

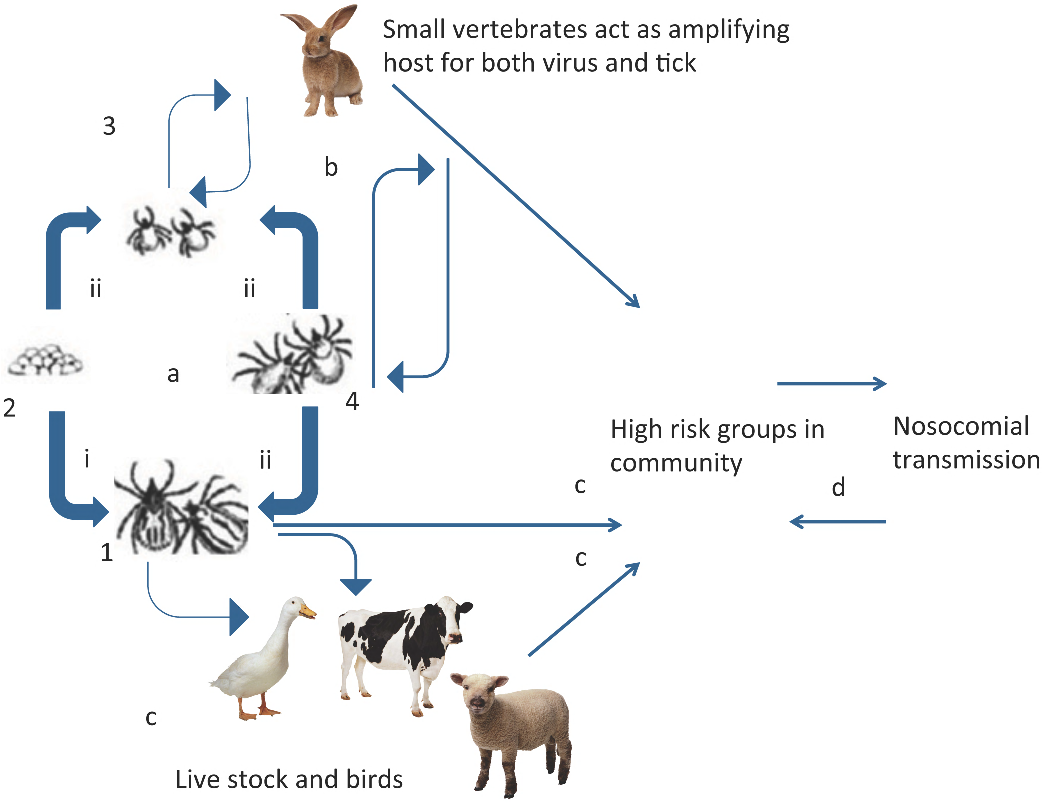


Figure 5. Life cycle of CCHFV. (a) Tick cycle, 1- adult, 2- eggs, 3- larvae, 4- nymph, (i) Trans-ovarian, (ii) Trans-stadial, (b) Tick-small vertebrate cycle, (c) Tick-large vertebrate/bird/human cycle, (d) Human-human cycle (community/nosocomial).

1.4.2 Molecular and cellular biology of CCHFV N protein

1.4.2.1 Subcellular localization of CCHFV N and its interactions with host cell partners

Besides taking part in the functions associated with particle assembly, RNP formation, and viral RNA synthesis the CCHFV N protein interacts with several host cell proteins, including actin and MxA (Andersson et al., 2004b, Andersson et al., 2004a). In the absence of other virally encoded proteins CCHFV N protein localizes to perinuclear regions of the infected cells and following disruption of actin filaments using cytochalasin D, this localization is abolished. Furthermore the disruption of actin filaments by cytochalasin D during virus infection in cells reduced viral titres and implies actin filaments play an important role in the assembly process of CCHFV. Similar results have been seen with another bunyavirus, the BCCV (Ravkov et al., 1998).

1.4.2.1.1 MxA

MxA has also been implicated as a binding partner for CCHFV N protein during infection. MxA belongs to the dynamin family of large GTPases, which are involved in a variety of intracellular transport processes (Haller and Kochs, 2002, Kochs et al., 2002, Sever et al., 2000). MxA has been shown to inhibit the replication of CCHFV, through an interaction with the CCHFV N protein. Andersson et al used confocal microscopy to reveal MxA colocalizes with the N protein of CCHFV in the perinuclear regions of infected cells (Andersson et al., 2004a). Constitutively expressed MxA caused a block in viral genome replication in CCHFV infected cells, while not affecting primary transcription of the incoming genome and translation of viral mRNA. MxA has been proposed to exert its anti-viral effects through sequestration of CCHFV N protein away from sites of replication, however the precise mechanism by which MxA inhibits CCHFV replication remains unclear.

1.4.2.1.2 Caspase-3 protease cleavage of CCHFV N protein

The most striking interaction between CCHFV N protein and a host cell protein is that with the caspase-3 protease and the DEVD motif, which is conserved across all CCHFV strains. *In vivo* analysis by Karlberg et al (Karlberg et al., 2011) demonstrated in multiple cell lines that during viral infection the CCHFV N protein is cleaved by caspase-3 protease into 30 kDa, and 22 kDa fragments corresponding to residues 1-265 and 266-480 respectively (Karlberg et al., 2011). The study also found that CCHFV infection in cells leads to the induction of apoptosis and cleavage of CCHFV N protein by the caspase-3 protease is more pronounced 72 hr after infection (detection of progeny virus was detected 48 hr after infection), which is consistent with the concurrent induction of apoptosis that was observed in infected cells at that time. It is unclear from this analysis whether there is a direct link between apoptosis induction as result of CCHFV replication in cells, or alternatively whether it is an indirect effect mediated by an antiviral response generated by the host cell, such as type I interferons (IFNs) or the dsRNA response, which may act as initiators of apoptosis. Furthermore the cleavage of CCHFV N protein was abrogated by the inhibition of apoptosis and induced progeny viral titers of ~80-90%, which not only highlights the link between apoptosis and caspase-3 protease cleavage of CCHFV N protein but also suggests the caspase-3 protease cleavage of the CCHFV N protein is a host cell defense mechanism. However the proposal that caspase-3 protease cleavage of CCHFV N protein is not beneficial to the virus does not explain the highly conserved nature of the DEVD cleavage site in all CCHFV strains studied to date. Although the purpose of this cleavage by caspase-3 protease is unknown it raises the possibility that nairoviruses may have evolved alternative ways of producing functional non-structural proteins. One possible mechanism for caspase-3 protease cleavage of CCHFV N protein could maintain reduced levels of activated caspase-3 protease cleavage of cellular substrates during the initiation of apoptosis by CCHFV N protein acting as a decoy substrate, and in doing so could delay the downstream-mediated signaling pathways, including apoptosis and the interferon response, with important consequences for viral growth.

1.4.2.2 Genetic diversity and relationships between strains

In recent years a number of studies have demonstrated that the genomic diversity of CCHFV is surprisingly high and has led to the genetic relationship of CCHFV worldwide - based on the S segment - being subdivided into seven genetically distinct groups, Africa 1 (Senegal), Africa 2 (Democratic Republic of the Congo and South Africa), Africa 3 typified by West and South African strains, Europe 4 (Russia, Turkey, Bulgaria, Kosovo and Albania), Europe 2 (Greece), Asia 1 (the Middle East, Pakistan and Iran) and Asia 2 (China, Uzbekistan, Tajikistan and Kazakhstan), and the European 7 subtype (Greece) (Hewson et al., 2004a, Deyde et al., 2006, Burt and Swanepoel, 2005). These studies agree with the generally high rates of mutation and genetic diversity in RNA viruses due to the error-prone nature of their polymerases. However they disagree with the double-filter concept by which the evolution of arboviruses is constrained due to a need to maintain life cycles in both arthropod and amplifying vertebrate hosts (Holland and Domingo, 1998, Weaver, 2006).

Comparative analysis of the whole genomes of 18 CCHFV strains estimated that the evolution rate of the M segment is 3.58 times higher than that of the S segment (Anagnostou and Papa, 2009). Similarly another study found that the genomes of 13 geographically and temporally diverse virus strains were highly variable, with the greatest accumulation of mutations seen again in the surface glycoprotein encoding M segment, with 31% nucleotide and 27% amino acid divergence (Deyde et al., 2006).

The high plasticity within the M segment may have arisen from positive selection operating in a manner reflecting immune surveillance, as well as increasing varied cellular tropism required for efficient attachment to arthropod and vertebrate host cells during different natural cycles. Alternatively lack of constraint may account for the high genetic diversity within the M segment, as much of the amino acid variation resides in the hypervariable mucin-like domain, where side chains can be highly variable and still endure the O-glycan modification (Sanchez et al., 2006, Deyde et al., 2006). These arguments however cannot be applied to the S and L segments (Deyde et al., 2006, Carroll et al., 2010, Anagnostou and Papa, 2009, Mild et al., 2010b, Hewson et al., 2004a, Burt et al., 2009), which have also been shown to exhibit high amounts of genetic diversity between strains.

Whilst mutation is the primary means of increasing RNA genome plasticity, RNA segment re-assortment also contributes to the genetic diversity in sNSVs (Deyde et al., 2006, Burt et al., 2009). The most dramatic example of re-assortment is from IAV, in which segment re-assortment results in antigenic shift and the emergence of pandemic virus strains. Although re-assortment of the CCHFV M segment seems to be the more frequent, examples of S and L segment re-assortment have to a lesser extent also been observed (Deyde et al., 2006).

Recently Chen et al have demonstrated that China strain YL04057 and 79121M18 belong to lineage I based on the M and L segments but belong to lineage IV in the S tree, which apparently represented S segment reassortments (Chen, 2013). One reason for CCHFV reassortments could be co-infection coupled with long-term persistent and vertical transmission in ticks (Burt et al., 2009, Shepherd et al., 1991). The idea that re-assortments maintain fitness during transmission to a different host (tick-vertebrate) is supported by those reassortments SPU415/85, SPU97/85, UG3010, ArD15786, HU8966, Baghdad 12, and Afg09-2990, being isolated from both human and goat (Chen, 2013).

The distinction of group I and III viruses of West Africa demonstrate that multiple virus genetic lineages can also be found in the same geographic areas (Deyde et al., 2006). In some regions such as Xinjiang China, CCHFV strains are derived from multiple origins, including those Xinjiang strains related to those of circulating

Uzbekistan from the Asia 1 group and Fub90009 strain in Asia 2 group. (Sun et al., 2009), and at least four genetic lineages of CCHFV have been detected in Iran (Senegal and Mauritania)(Chinikar et al., 2004, Mild et al., 2010b).

Recombination events have also been observed (Lukashev et al., 2008, Hewson et al., 2004b), and although much less frequent may also contribute, along with the accumulation of mutations and segment re-assortment, to the high genetic diversity of CCHFV.

1.4.3 Epidemiology and Ecology

Phylogeographical analyses between viruses from different geographic regions has not only provided information as to the genetic diversity amongst CCHFV strains, but also provides important information about the direction of migration between endemic geographical areas. Collected phylogeographical analysis by Hewson et al (Hewson et al., 2004a) has revealed closely related viruses can be found in distant geographical locations. For example S RNA sequences of strains from the United Arab Emirates, Madagascar and Pakistan all fall into the Asia 1 group (Hewson et al., 2004b). Moreover it is interesting that Mild et al (Mild et al., 2010a) reported that the same genotypes could be found in countries that are very distant. For example, genotype 1 was found predominately in the Middle East, however was also found west in Madagascar, and genotype 7 was found in west Africa, South Africa and Iran, which suggests CCHFV is capable of migrating over short and long distances. Furthermore Mild et al (Mild et al., 2010a) concluded that because each genotype in the phylogenetic trees showed that CCHFV N did not cluster according to time of isolation (non-temporal evolution), and that several of the sequences were identical despite sampling at different points over a 30 year period, for example genotype 1, Pakistan 1970 and 2002. And therefore through divergent evolution CCHFV has diversified into seven distinct genotypes and most likely has been circulating a very long time, an observation also reported by Anagnotou & Papa (Anagnostou and Papa, 2009).

Although these phylogeographic features can be explained through a number of factors; the major likelihood is that livestock trade has played an important role. For instance the multiple phylogenetic lineages of group I and III viruses in West Africa (Hewson et al., 2004a). Perhaps the multiple genetic lineages, which are found within the Arabian Peninsula, can be explained by the movement of sheep from countries within the Horn of Africa, or Iran and Pakistan. This view is also consistent with the genetic relationships seen between virus strains and detailed molecular epidemiologic analysis of previous outbreaks in the United Arab Emirates and Oman (Deyde et al., 2006, Rodriguez et al., 1997).

The geographic range of CCHFV is exceptionally wide, and coincides with the global distribution of *Hyalomma* ticks, of which *Hyalomma marginatum* (Figure 6) has been implicated as the principal vector of CCHFV in Europe. There are at least 31 tick species that have been incriminated in transmitting the virus in different geographic locations, including at least two species in the family *Argasidae* (soft ticks) and from at least seven genera in the family *Ixodidae*. For instance *H. asiaticum asiaticum* and *H. asiaticum kozlovi* are found to principally harbor CCHFV in Xinjiang, China (Sun et al., 2009), whilst in Turkey the principal vector is *Rhipicephalus bursa,* commonly found on domestic ruminants (Tonbak et al., 2006). Isolates from other tick genera, including *Ornithodoros*, *Boophilus*, *Dermacentor* and *Ixodes* species may also be locally important.

Figure 6. Dorsal view of an adult female *H. marginatum* tick. The figure was adapted from World Health Organization, http://www.who.int/mediacentre/factsheets/fs208/en/.

1.4.4 Threat of CCHFV emergence and outbreaks

The virus can be transmitted to humans through tick-bites or by contact with blood and tissues of viremic hosts, such as livestock or by crushing infected ticks with unprotected hands (Ozkurt et al., 2006). In Turkey between 2004 and 2007 61.7 % (n

= 1031) of CCHF cases reported were farmers, who had close contact with animals, whereas only 9.9% had direct contact with animal blood, animal tissue or other body fluids. Those patients who had contact with animals and those who reported a tick bite were comparable, and therefore it seems likely that tick-infected animals transmit ticks to humans, and that the cause of infections is probably due to bites from ticks. In endemic areas high risk groups for CCHFV infection include farmers, abattoir workers, veterinarians, people living in rural areas and handling livestock, and health personnel have also been described as the risk groups for CCHFV (Vorou et al., 2007). Higher prevalence of ticks within large numbers of livestock maybe the reason for the increased numbers of human CCHF cases in small villages in rural areas, and not surprisingly the seroprevalence for CCHFV is higher in persons living in these areas than in urban areas (Gunes et al., 2009). Although in Iran, for example special markets for animal trading are located on the outskirts of large cities, and CCHFV seroprevalence has found to be higher among persons living in urban areas of this country (Izadi et al., 2006).

Due to the high titres of CCHFV within the blood of infected patients human to human transmission also poses a risk for infection to health workers and relatives who make unprotected contact, especially from those patients with bleeding in the lungs, nose, mouth, gums, and at injection sites (Athar et al., 2005, Altaf et al., 1998, Aradaib et al., 2010, Mertens et al., 2013). Additionally there is evidence reported by Gurbuz et al that transmission from one patient to another by sharing the same room is possible (Gurbuz et al., 2009).

CCHFV is considered an emerging virus and a potential treat to non-endemic areas or those countries where the virus cannot be found. CCHFV is endemic in the Middle East, Africa, Europe (Balkans: Kosovo, Bulgaria, the former Yugoslavia, and Albania) and Asia and continues to emerge and reemerge (Maltezou et al., 2010). There have recently been sporadic outbreaks since 2000 and an increasing number of case reports in Turkey, Kosovo, Albania, Bulgaria, Greece, Iran, Pakistan, Afghanistan, the Russian Federation, Kazakhstan, Tajikistan, Georgia, Mauritania, Kenya, Senegal and South Africa (Leblebicioglu, 2010). Serological evidence has also been reported from Egypt, India, Hungary, Portugal, France and Benin. In Europe, CCHFV is only currently endemic in Bulgaria.

The potential factors associated with emergence of CCHFV in countries where the virus is not currently found are those such as environmental factors including climate change, as well as anthropogenic factors including changes in land use, agricultural practices or hunting activities, or movement of livestock which may influence host-tick-virus dynamics (Maltezou et al., 2010). It is also possible that geographic movement of genetic lineages and emergence of CCHFV in previously non CCHFV endemic countries could occur through long distance carriage of virus via migratory birds, which are either infected or carrying virus infected ticks (Shepherd et al., 1987a, Hewson et al., 2004b, Hewson et al., 2004a, Swanepoel et al., 1998, Zeller et al., 1994). International travel routes have also been highlighted as a potential factor for globally spreading this disease recently, when the first case of CCHFV in the UK was confirmed from an infected 38 year old male who arrived in Glasgow on an international flight from Afghanistan via Dubai (Chamberlain et al., 2013).

1.4.5 Clinical Manifestations

Human infection with CCHFV often results in severe haemorrhagic disease, with a high mortality rate. The severity of disease varies depending on the mode of transmission; for tick–human transmission mortality is 10–50%, but can increase to 80% in the case of nosocomial incidents. The reasons for this phenomenon are not known, but may simply relate to viral dose (Swanepoel et al., 1987, Hewson et al., 2004a). For instance the incubation period may be shorter with blood borne transmissions (Ergonul, 2006).

CCHFV infection typically develops in four distinct phases: incubation, pre-heamorrhagic, heamorrhagic and convalescence periods. Sudden onset of fever, including headache, myalgia, and dizziness is typical of the preheamorrhagic period (Whitehouse, 2004a). In some cases, clinical symptoms such as diarrhea, nausea and vomiting can also be seen (Rodriguez et al., 1997). Haemorrhagic manifestation develops rapidly, typically beginning between the third and fifth day and usually lasts for 2-3 days. The most common bleeding sites are the gastrointestinal tract (hematemesis, melena, and intra-abdominal), uterus (menometrorrhagia), urinary tract (hematuria), respiratory tract (hemoptysis) and nose. The most extensive clinical pathology study of CCHF was that of Swanepoel et al (Swanepoel et al., 1987) whose observations were made on 50 CCHF patients from South Africa, diagnosed between 1981 and 1987. The mortality rate was 30%, with factors contributing to a fatal outcome including cerebral heamorrhage, severe anemia, severe dehydration and shock associated with prolonged diarrhea, myocardial infarction, lung edema, and pleural effusion. For those who do not succumb to death, the convalescence period usually begins 15-20 days after onset of illness.

1.4.6 Treatments

To date, there is no vaccine or specifically targeted treatment available to treat CCHFV infected patients. Ribavirin has shown the most promise and is the only antiviral agent which has been reported to be active *in vitro* (Watts et al., 1989). The World Health Organization (WHO) recommends ribavirin as the antiviral agent for CCHFV infected patients. Several case reports have been published that suggest oral or intravenous ribavirin is effective for treating CCHFV infections (Whitehouse, 2004a). For example in Turkey, twenty-two cases of CCHF were treated with oral ribavirin, immediately after admission. In the ribavirin group, admission times and recovery period were statistically shorter compared to the control group. Of sixty patients, six died (10.0%), two in the ribavirin and four in the control groups, and the fatality rate was 9% in the ribavirin group vs 10.5% in the control group (Ozkurt et al., 2006). This data shows Ribavirin is not effective against CCHFV infection and some contention has arisen suggesting the results are significant enough, with a recent report revealing that the available data in the literature are inadequate to support a claim of efficacy of ribavirin in treating CCHFV infection (Ascioglu et al., 2011).

1.5 Introduction to TSWV and other tospoviruses

Tospoviruses are plant viruses and have notoriety for causing devastating damage to plants, for the largest part due to their extensive host range and pathogenicity. Tospovirus infection encompasses more than 1,000 different plants in approximately 80 different plant families, including dicotyleydens, or those flowering plants whose seed typically has two embryonic leaves or cotyledons, and monocotyledons, which typically have one embryonic leaf (Kormelink et al., 2011).

Their wide host range covers agricultural and ornamental plant species of economic importance, including tomato, potato, pepper, lettuce, ground nut, alstroemeria, cyclamen, begonia, impatiens, as well as numerous weeds (Kormelink et al., 2011). Primarily classification of tospoviruses is divided between two groups, which are based upon by both N protein gene sequence and biological features. Correlation of these two features can also be drawn with geographical distribution, i.e. an American clade and EuroAsian clade (Hassani-Mehraban et al., 2007). The prototypic member and also the best studied of this group is tomato spotted wilt virus (TSWV) (Goldbach et al., 1994, Kormelink, 2011).

Symptoms of tospovirus infection include stunting, necrosis, chlorosis, ring spots and ring/line patterns affecting leaves, stems and fruit, and vary depending on tospovirus species, host plant, time of year and environmental conditions. Factors such as type of tospovirus, host plant, time of year and environmental conditions can affect the severity of the symptoms. Infection rates of 50–90% lead to major losses in commercial vegetable crops, and it is thought that TSWV costs the food industry up to 1 billion dollars per year (Adkins, 2000, Mumford et al., 1996, Prins and Goldbach, 1998)

1.5.1 Tospovirus species

Today the tospovirus genus comprises more than 20 recognised and tentative species (Kormelink, 2011), with new viruses being added to the list frequently. The criteria to establish a new tospovirus species are recognised using the sequence divergence in the N protein (>10% divergence with already classified tospoviruses) and species of thrips vector (Kormelink et al., 2011). Impatiens necrotic spot virus (INSV) was the first tospovirus virus to be recognised as a genetically distinct and separate tospovirus. Sequencing analysis of the INSV N protein revealed it exhibits only 55% sequence identity to that of TSWV N protein. It was shown also to have limited but characteristic host range of which many are cultivated as ornamental plants, e.g *Impatiens* spp (Goldbach et al., 1994).

A basic outline of the organization of the TSWV genome is schematically represented in figure 2. Aside from producing the viral proteins by transcribed mRNA from the 3' of the genome in the same way as the animal infecting bunyaviruses, the plant infecting tospoviruses also transcribe translatable mRNA from the 3' end of antigenome using an ambisense mechanism. The ambisense coding strategy occurs on the S and M segments and resulting mRNA transcripts encode for NSs and NSm proteins respectively.

1.5.2 Transmission

Similarly to animal infecting bunyaviruses, transmission of tospoviruses occurs in a propagative manner by arthropods, predominantly thrips. One major difference however is that tospovirses are not secreted from plant cells and instead accumulate in large vesicles within the cytoplasm. Tospoviruses are mechanically transmitted to neighboring plants by adult thrips, where they replicate and are derived from two wingless larval stages that feed on plant leaves and weeds (Figure 7). TSWV can only be acquired by thrips during the larval stage of their development and after a non-feeding latency period mature adult thrips can transmit the virus for the rest of their life span (Wijkamp et al., 1993).

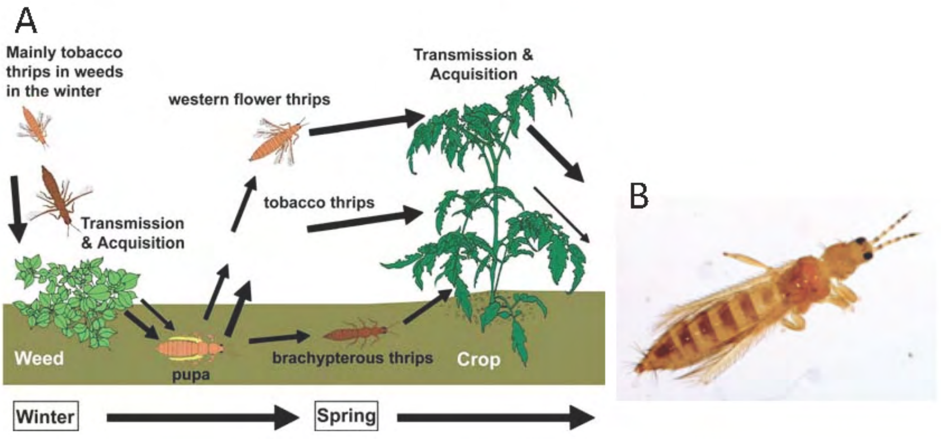


Figure 7. Transmission of tospoviruses by thrips. (A) The immature nymphs acquire the virus from weeds before the fully developed adults transmit the virus to a springtime crop, like tomatos. (B) Adult dorsal view of a western flower thrip (*Frankliniella occidentalis*). The figure was adapted form The University of Agriculture, College of Agricultural & Environmental Sciences http://www.caes.uga.edu/topics/diseases/tswv/vegcrops/tospoviruses/thripsvectors.html#intro.

1.5.3 Epidemiology

Tospoviruses may have evolved recently due the expansion of its vectors, which are linked to natural and/or human influences. The rapid geographical emergence and expansion over the past thirty years of TSWV has been linked with the advanced expansion of the Western flower thrips (*Frankliniella occidentalis*) (Goldbach and Peters, 1994). Events such as these have culminated in the expansion of the host range of TSWV including both horticultural and agronomic crops across temperate, subtropical and tropical regions of the world (Goldbach and Peters, 1994, Kormelink., 2011).

This is best characterised by the emergence of abundant *F. occidentalis* in Texas, and of the *F. fusca* in Georgia, which occurred in the peanut crop in the southern states of the USA. Similarly serious damage in groundnut in India can be attributed to the emergence of the groundnut bud necrosis virus (GBNV), which is also a tospovirus. It is thought that the expansion of irrigated crops and continuous cropping of groundnut and other susceptible hosts has led to rapid expansion of the GBNV (Goldbach and Peters, 1994). In tropic climate zones the main vector of the GBNV virus is considered to be *Thrips palmi* (Goldbach and Peters, 1994).

In 1983 the introduction of *F. occidentalis* in Europe led to the emergence of TSWV in greenhouse cultivations in Western Europe, especially in France, Italy, Portugal and Spain in both field and greenhouse crops. These examples illustrate an active geographical expansion of tospoviruses, correlating mainly to the migration of the thrips vector, which causes an increased incidence of the disease (Goldbach and Peters, 1994).

1.5.4 Comparison with animal infecting bunyaviruses

TSWV and other tospoviruses share similarities with the animal infecting bunyaviruses. Tospovirus particles are enveloped and spherical with a diameter of approximately 80–120 nm. It is worth noting that apart from the plant-infecting rhabdoviruses the presence of an envelope membrane is unique among plant viruses (Adkins, 2000).

Tospoviruses may have emergered recently, not only through global expansion of its vectors, but also through re-evolution from an animal bunyavirus that was able to infect thrips species and subsequently became adapted to plants instead of vertebrates. Firstly the idea that TSWV has descended from an animal infecting bunyavirus is supported by the seemingly limited number of tospoviruses able to infect plants and their low divergence, contrasting to the high number of animal infecting counterparts, which are highly divergent (Goldbach and Peters, 1994). The high level of divergence between phleboviruses, hantaviruses, nairoviruses, and orthobunyaviruses, contrasting the limited number of tospoviruses capable of infecting plants, suggests tospoviruses have evolved from an animal infecting bunyavirus.

It is unsurprising that TSWV and other tospoviruses display a high level of genetic similarity to their animal infecting counterparts. Sequence analysis has revealed TSWV is more closely related based on multiple L and N protein alignments to BUNV than BUNV is to hantaviruses and nairoviruses (Figure 98) (Vieth et al., 2004) (Goldbach and Peters, 1994). Secondly the broad plant host range of tospovirus, coupled with absence of any evolved natural resistance genes indicates TSWV and other tospoviruses have invaded the plant kingdom recently and that co-evolution of bunyaviruses with plants has yet to start (Goldbach and Peters, 1994).

1.5.5 Cell to cell movement

It has been observed that TSWV loses the production of one or both glycoproteins when infection is artificially maintained through omission of the thrips stage of the infection cycle (Resende Rde et al., 1991), suggesting the envelope glycoproteins of bunyaviruses are only involved in receptor recognition during infection of insects and vertebrates. As tospoviruses do not require a receptor for successful infection within plants they have evolved a mechanism to facilitate cell-to-cell movement of the virus across the otherwise impenetrable cellulose barrier of the cell wall. The NSm movement protein has thus developed in tospoviruses to facilitate their spread throughout the plant

It has been demonstrated that the primary roles of the NSm protein is to interact with the N protein, bind ssRNA in a sequence-nonspecific manner and associate with the plasmodesmata (Soellick et al., 2000). *In vivo* studies have demonstrated that the NSm protein can associate with itself and form tubular structures in the absence of other viral proteins, and emanate from the cell surface, altering the size exclusion limits of the plasmodesmata (Storms et al., 1995).

The NSm protein has also been shown to interact with intracellular trafficking plant protein homologues of myosin, kinesin and DnaJ like chaperones (Paape et al., 2006) (Von Bergan). The role of the N protein during this process is not limited to chaperoning the RNA through the re-organized plasmodesmata, but an additional role of contributing to the movement of the RNP complexes has been reported. By using a transiently expressed construct it was demonstrated the N protein, in absence of other TSWV-encoded proteins could facilitate the long distance movement of

Tobacco mosaic virus (TMV) replication complexes. Whether the N protein provides a supporting role to the NSm protein or can solely facilitate transport of infectious virus has yet to be established (Zhang et al., 2012).

1.5.6 RNA silencing

The NSs protein functions to escape the RNA interference (RNAi) antiviral defense mechanism (Baulcombe, 1996, Fire, 2007, Takeda et al., 2002, Bucher et al., 2003a). In plants RNA silencing or post-transcriptional gene silencing (PTGS) is considered as the innate immunity mechanism against viral invaders, and is an evolutionary conserved process whereby dsRNA and imperfect hairpin RNAs (foldback RNAs) that are processed into duplex 21-24 nucleotide short interfering RNA (siRNA) or microRNA (miRNA) by RNA III-type DICER enzymes, to provoke a succession of processes leading to sequence specific breakdown of RNA.

During infection viral dsRNA is produced either from replicative intermediates or alternatively due to the formation of secondary RNA structures. These replicative intermediate RNAs are recognized by specific Dicer-like proteins (Dicer in insects and predominately DCL-4 in plants), and are processed into 21 nucleotide small interfering RNA (siRNA) duplex molecules. The guide strand of the processed duplex small RNAs are then incorporated into a RNP complex termed the RNA-induced silencing complex (RISC), which functions to recognize and degrade complementary (viral) target RNA molecules, through the action of the core Argonaute (AGO) protein. RISC assembly involves guide strand incorporation and passenger strand elimination, by the AGO protein of the siRNA duplex, to subsequently form the single strandard siRNA containing 80S holo-RISC complex. In plants these 21 nucleotide siRNAs have been shown to function as primers for synthesis of secondary siRNAs, by the RdRp which converts RNA target sequences into new long dsRNAs. In this way silencing is not only being amplified but also spread along the entire RNA sequence, known as transitive silencing.

Plant viruses, such as tospoviruses and tenuiviruses overcome production of the replicative intermediates, which lead to the induction of RNAi, by coding for a RNA silencing suppressor (RSS) protein that is able to interfere in the cascade of reactions that lead to viral RNA degradation (Bernstein et al., 2001, Lee et al., 2003, Khvorova et al., 2003, Schwarz et al., 2003, Guo and Ding, 2002). RSS proteins of plant viruses have either shown to restrain RNAi by sequestering siRNAs and thereby prevent their incorporation into RISC. Or alternatively they work to avoid cleavage of long dsRNA into guide siRNAs (Diaz-Pendon et al., 2007, Guo and Ding, 2002).

For tospoviruses, the NSs protein encoded through ambisense coding on the S segment has been identified as the RSS protein using a GFP-silencing assay that involved a co-infiltration of *Agrobacterium tumefaciens* containing a functional 35S-driven GFP and NSs gene construct into *Nicotiana benthamiana* leaves (Takeda et al., 2002, Bucher et al., 2003b). Additionally it was reported by Schnettler et al, that expression of an NSs gene construct in *N. benthamiana* suppresses the accumulation of GFP-specific siRNAs during co-infiltration with an inverted repeat GFP RNA construct, and also reverses the silencing of an eGFP miRNA-sensor construct (Schnettler et al., 2010). Biochemical analysis has shown the NSs from different tospoviruses displays affinity for small RNAs, including si- and mi-RNAs. Furthermore, the TSWV NSs from INSV and Groundnut ringspot virus (GRSV) also display affinity for long dsRNAs, whereas the NSs from Tomato yellow ring virus (TYRV) did not. The TSWV NSs has shown *in vitro* inhibition of DICER mediated cleavage of long dsRNAs (Schnettler et al., 2010). Together the evidence suggests the that NSs from tospoviruses interferers in the plant RNAi innate immune response by sequestering siRNAs and miRNAs to hinder their uploading into their respective RISC complexes. This mechanism has also been supported recently by the report which has shown that the GBNV NSs protein displays NTPase and phosphatase activity (Lokesh et al., 2010), and it has been speculated that this activity could possibly be involved in enzymatically removing the 5′ phosphate from dsRNA and thereby prevent its cleavage by dicer into siRNAs.

1.6 Arenaviruses

The following section will focus on the two-segmented arenavirus NP structure and phylogenetics; it has been included into this section because there are some important aspects of their biology, which will be discussed later in chapters 3 and 4.

The *Arenaviridae* family is distributed worldwide and include viruses that are significant human pathogens. Arenaviruses are subdivided based on serological cross-reactivity, phylogenetic relations and geographical distribution, into the Old World and New World virus complexes. The New World arenaviruses circulate in both North and South America and include Whitewater Arroyo, Tamiami, and Bear Canyon, Tacaribe, Junin, Guanarito, Machupo virus and Sabia viruses. The Old World arenaviruses include those viruses that circulate in Africa, Europe, and Asia including lymphocytic choreomeningitis and LASV (Yun and Walker, 2012). LASV contributes the largest burden to disease out of all the arenaviruses and causes approximately 300,000 to 500,000 infections per year in West Africa (Hastie et al., 2011b).

Arenaviruses process a bisegmented, NS, ssRNA genome with an ambisense coding strategy to produce five known proteins; namely the GP-1 and GP-2 glycoproteins, a NP, a matrix protein (Z), and an L protein. The NP is the most abundant virally encoded protein in an infected cell. The RNP is assembled from the viral RNA and the self interacting NPs and its further association with the L protein, which participate in RNA replication and transcription (Pinschewer et al., 2003). The arenavirus NP also interacts with the Z protein for viral assembly (Eichler et al., 2004, Groseth et al., 2010, Shtanko et al., 2010), and also plays an important role in the suppression of the innate immune system (Martinez-Sobrido et al., 2009, Martinez-Sobrido et al., 2007, Hastie et al., 2011a)

The arenaviruses, bunyaviruses and orthomyxoviruses are phylogenetically related and not surprisingly share features of their intracellular replication cycle, such as cap snatching. The close phylogenetic relationship between arenaviruses and bunyaviruses was highlighted when Vieth et al (Vieth et al., 2004), used secondary structure-assisted alignment of the polymerase region, allowing a reliable reconstruction of the phylogeny of all nsRNA viruses, and which indicated the *Arenaviridae* family are most closely related to the *Nairovirus* genus of the *Bunyaviridae* family (Figure 8).

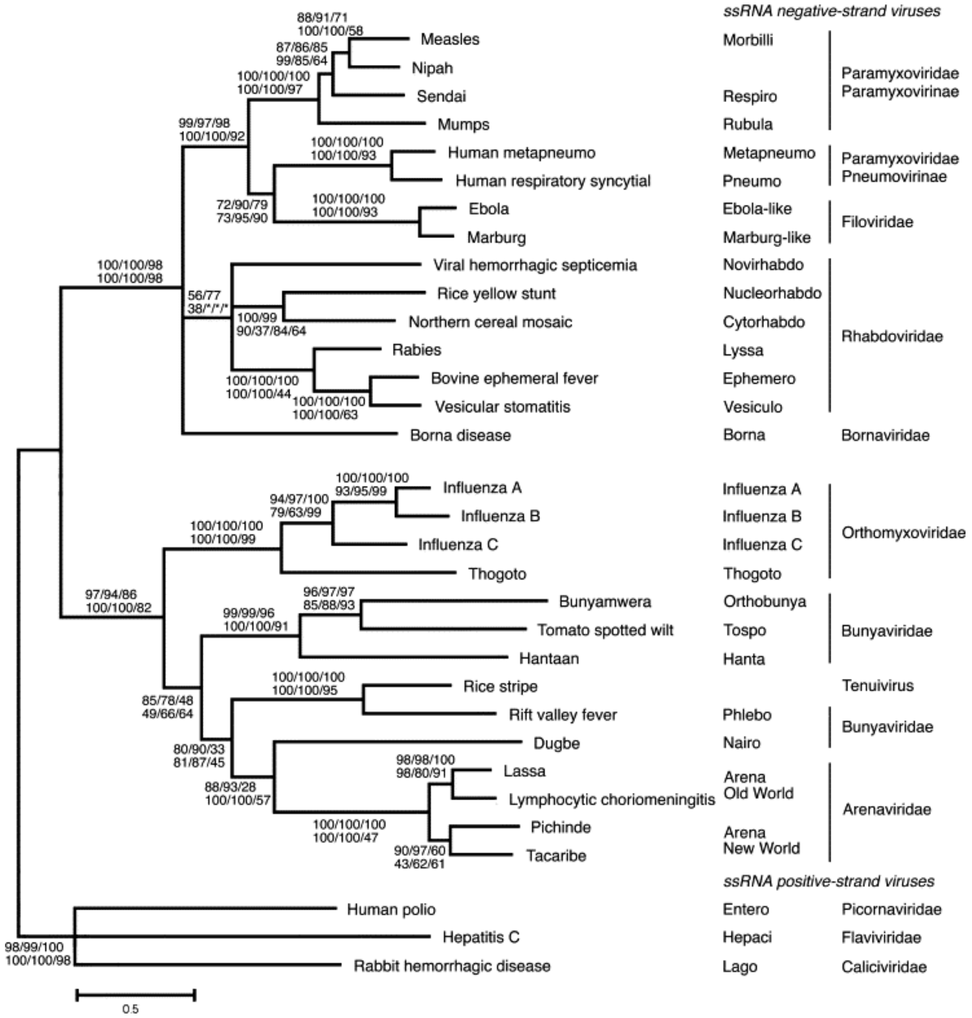


Figure 8. Phylogenetic relationships among nsRNA viruses. Alignment based upon the secondary structure-assisted prediction of polymerase amino acid sequences. The figure was adapted from (Vieth et al., 2004)

1.7 Biology of the bunyavirus structural proteins

1.7.1 Structural proteins encoded by the M segment

For bunyaviruses to form the infectious viral particles the N protein assembled RNP must also crucially interact with the M segment encoded membrane embedded glycoproteins (Figure 9). All members of the *Bunyaviridae* family encode a polyprotein precursor from the M segment, which is inserted co-translationally into the membrane of the endoplasmic reticulum. Cleavage of the precursor by host cell proteases results in Gn and Gc components that form a disulphide linked heterodimer and are transported to, and retained in, the Golgi apparatus. The location of the Gn-Gc heterodimer in the Golgi where it associates with RNPs to mediate assembly and budding of mature virus particles is maintained by a retention signal. Virus assembly, formation of the virus particle and attachment of the virus to new target cells is mediated by the Gn-Gc heterodimer (Schamaljohn & Nichol, 2006).

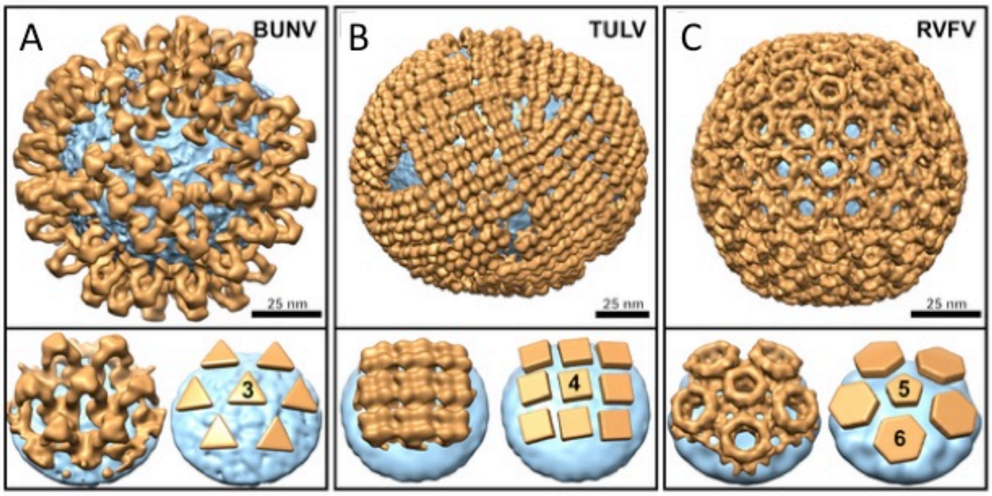


Figure 9. Various bunyavirus surface glycoprotein architectures. (A-C) Structures of Gn–Gc glycoprotein spikes (orange) are shown mapped onto the membrane surface (cyan). (A) Orthobunyavirus (BUNV), (B) Hantavirus (Tula virus; TULV; EMD-1704[11]), and (C) Phlebovirus (RVFV); EMD-1550[7]). Below each panel are the views of a glycoprotein spike cluster (left) and a schematic representation of the spike arrangement (right), with the symmetries of individual spikes annotated. The figure was adapted from (Bowden et al., 2013, Walter and Barr, 2011).

In general the conformational process of membrane envelopment of bunyavirus RNPs is likely triggered by an interaction between the N protein and glycoproteins. In line with this it is very common for enveloped viruses to utilize a matrix protein for contact and maintenance between the RNP and the viral envelope. However members of the *Bunyaviridae* family lack matrix protein and, by analogy to alphaviruses, which also lack a matrix protein, it is quite likely that the envelope glycoproteins of bunyaviruses organise and facilitate the interactions between the envelope and RNP, thus operating as surrogate matrix protein (Hepojoki et al., 2010). This has recently been supported when the crystal structure of the RVFV Gc protein revealed it to be a class II fusion protein, with architecture similar to the fusion protein of positive sense RNA flaviviruses and alphaviruses (Dessau and Modis, 2013). The icosahedral T = 12 lattice of the RVFV viral particle is formed by the assembly of the two glycoproteins, Gn and Gc, which assemble around the lipid envelope where they are anchored creating a rigid outer shell, with an average diameter of 103 nm (Sherman et al., 2009, Huiskonen et al., 2009, Overby et al., 2008). Analysis using electron cryo-tomography

(ECT) has suggested the cytoplasmic tails of the UUKV Gn and Gc directly interact with the RNP, which was visualized as a punctuated moiety, fully occupying the inner core of the virion (Figure 10) (Overby et al., 2008). Furthermore one of the characteristics of *Phlebovirus* genus particle formation, maturation and RNP packing is its dependence on the binding interaction between the Gn protein cytoplasmic tail and N protein (Overby et al., 2007a).

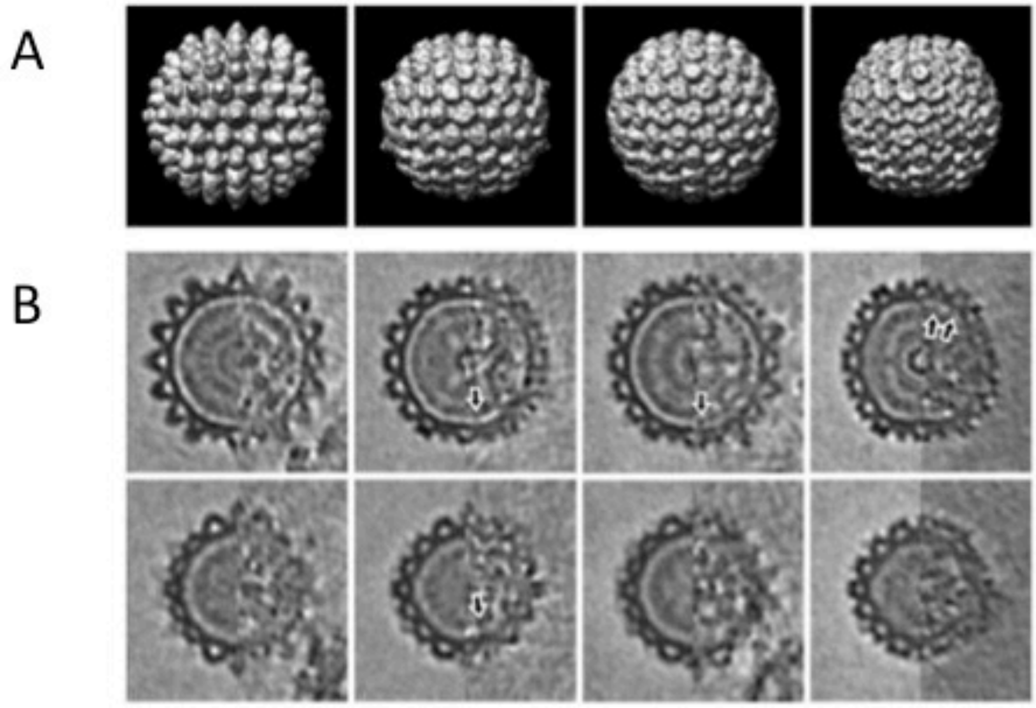


Figure 10. ECT analysis of the UUKV particles. (A) UUKV particles displaying icosahedral 2-fold axis of symmetry. Each column represents data from a single particle, with isosurfaces rendered at 1.5σ above the mean density. (B) Consecutive 8 nm-thick slices through the tomogram, 24 nm apart, are shown for the particles displayed in A. On the left the icosahedrally symmetrised version of the particle is shown within each panel, with the unsymmetrised denoised particle on the right for comparison. Arrows indicate the RNP interactions with the membrane. The figure was adapted from (Overby et al., 2008)

Additional ECT data of Hantaan virus (HTNV) and of Tula virus (TULV) (both hantaviruses) has revealed that the RNP associates with the viral envelope, suggesting that a direct interaction between the RNP and the spike complex exists (Huiskonen et al., 2010, Battisti et al., 2011). The exterior arrangement of the TULV particle revealed ordered patches of tetrameric spike complexes on the lipid membrane, with each spike consisting of Gn-Gc heterodimers (Bowden et al., 2013, Huiskonen et al., 2010)(Figure 9B).

In contrast to the arrangement of the glycoprotein spike complexes for hantaviruses, the orthobunyavirus particle was observed to have no significant areas of the membrane lacking spikes (Huiskonen et al., 2010) (Figure 9C). Similar to phleboviruses and hantaviruses, the BUNV viral membrane surrounds an inner RNP core. ECT of BUNV virions identifies density connecting the RNP core to the membrane, and most likely points to a direct interaction between glycoprotein cytoplasmic tails and the N protein/RNA complex (Bowden et al., 2013).

Observation of a similar interaction between the cytoplasmic tails of the glycoproteins and N protein during TSWV maturation has also been reported (Snippe et al., 2007, Ribeiro et al., 2009). None of these hallmarks have yet to be identified for nairoviruses and the structural biology of particle maturation in this group remains poorly understood. However despite this there have been reports that the Gn protein of CCHFV (nairovirus) localizes to the Golgi complex, a process mediated by retention signals in the cytoplasmic domain and ectodomain of this protein (Haferkamp et al., 2005, Bertolotti-Ciarlet et al., 2005). This data is consistent with the intracellular targeting of most animal bunyavirus glycoproteins to the Golgi complex and strongly suggest this as the assembly site for CCHFV and other nairoviruses. Additionally DUGV particles, have been shown to assemble by budding at the Golgi complex, where the glycoprotein G1 accumulated in vesicles originating from Golgi cisternae (Booth et al., 1991).

1.7.2 Viral proteins encoded by the L segment

The L segments of all members of the *Bunyaviridae* family encode a single L protein (Walter and Barr, 2011). The size of the bunyavirus RdRps vary and are genus specific, with the largest displayed by nairoviruses and tospoviruses (Figure 2) (de Haan et al., 1991) (Honig et al., 2004) (Marriott and Nuttall, 1996) (van Poelwijk et al., 1997). Through alignment and comparison with RdRps from various other groups of nsRNA viruses the well-known and defined polymerase module (region 3), has been shown to exist within bunyavirus RdRps across the entire family (Muller et al., 1994, Kinsella et al., 2004, Poch et al., 1990). The bunyavirus polymerase module encompasses motifs A through D, which has been identified as conserved among all RdRps examined to date (Kinsella et al., 2004).

An additional two motifs, designated motif pre-A and motif E, have also been identified and only found in RdRps for RNA viruses (Figure 11). The pre-A motif can be found in RdRps of positive sense RNA viruses, whilst motif E is thought to be present only in sNSVs (Honig et al., 2004). Furthermore the polymerase module, which can be found across the *Bunyaviridae* family can also be found in the PB1 polymerase subunit of IAV and the two-segmented arenavirus LASV (Vieth et al., 2004) (Biswas and Nayak, 1994).

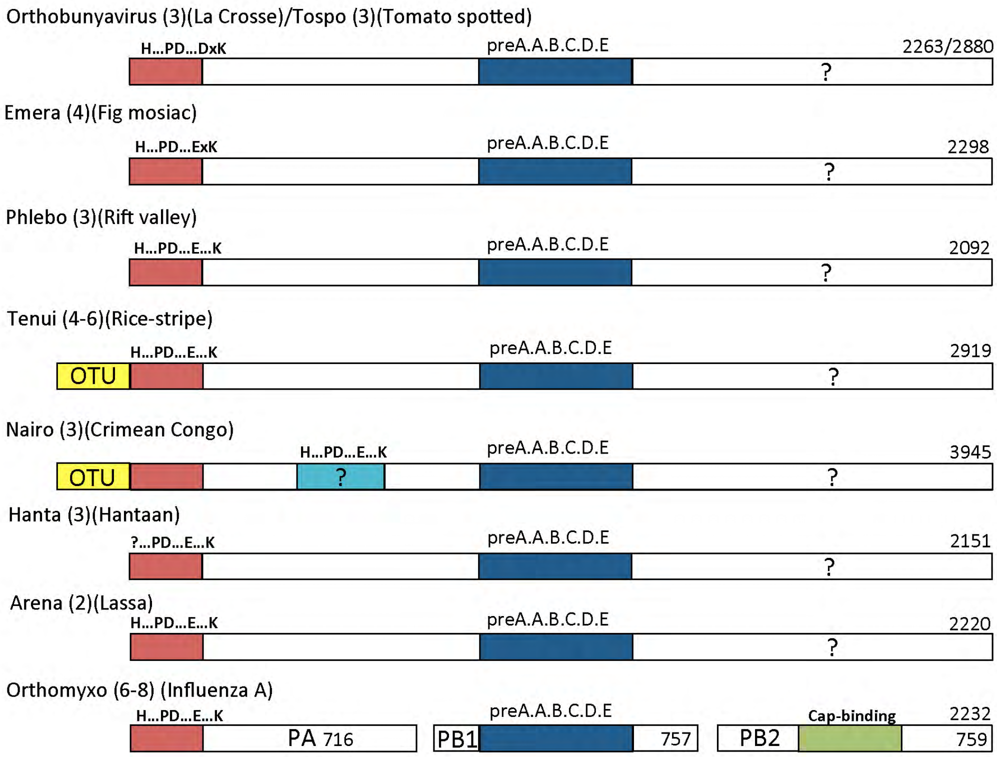


Figure 11. Schematic of the polymerase topology for sNSVs. The arrangement of the polymerase protein from different genera known or suspected to use cap snatching are represented as bars (not drawn to scale). The number of RNA genome segments and a sample of species are indicated in brackets. The amino acid length of each polymerase is labelled at the right hand side. Endonuclease domains are coloured in red and the particular motif of that genus written above. For CCHFV the catalytic motif for the putative endonuclease domain is coloured light blue. Blue bars represent the conserved polymerase domain in the central region. The three separate PA, PB1 and PB2 domains for orthomyxoviruses are represented in co-linear fashion. The green bar represents the cap-binding domain, which possibly indicates the location of this domain in the C-terminal region of the L proteins of bunyaviruses (question mark). The figure was adapted from (Reguera et al., 2010).

The La Crosse virus (LACV) (orthobunyavirus) L protein has been found to utilize an N-terminal endonuclease domain to cleave off small-capped primers from host derived mRNAs (Reguera et al., 2010). The structure of the LACV L protein N-terminal endonuclease domain revealed a catalytic core containing metal binding and catalytic lysine residues. Alignment analysis of this catalytic core motif with amino acid sequences from other sNSVs indicate that homologous L protein N-terminal endonuclease domains also exist not only in other members of the *Bunyaviridae* family, but also in members of arenaviruses, emaraviruses (5-segmented viruses) and tenuiviruses (four-segmented viruses) (Figure 11).

Furthermore the endonuclease domain of LACV L protein has been shown to be structurally related to the IAV endonuclease domain present within the N-terminal of the PA subunit. Knowing that the size of the orthobunyavirus, phlebovirus, and hantavirus L protein amino acid sequences are similar to the length of the three IAV polymerase subunits (Figure 11), an interesting question can be raised to whether the topological arrangement of the L protein domains are in any way homologous to a concatamer arrangement of the IAV three-polymerase subunits PA, PB1, PB2. This arrangement is supported in recent electron microscopy (EM) images of the IAV polymerase, which suggests the location of the polymerase module in the IAV PB1 is sandwiched between the PA and PB2 (Coloma et al., 2009a, Torreira et al., 2007) (Area et al., 2004) (Figure 12). However without any structure or knowing the location of the PB2 cap snatching domain homologue in the bunyavirus L protein, and due to no obvious cross-genera motifs in the C-terminus such a comparison cannot be made. The concept that the location of the cap-binding domain might vary between that of IAV and the bunyaviruses, or even within the *Bunyaviridae* family is also supported by evidence that the distance between the cleavage site and the 5' cap can be variable amid the cap-snatching viruses (Duijsings et al., 2001). Intriguingly evidence has suggested the hantavirus N protein could play this role and directly bind 5' capped RNA, suggesting mechanisms of cap snatching may be genus specific and bringing the cap binding capability of the hantavirus L protein into question (Panganiban and Mir, 2009).

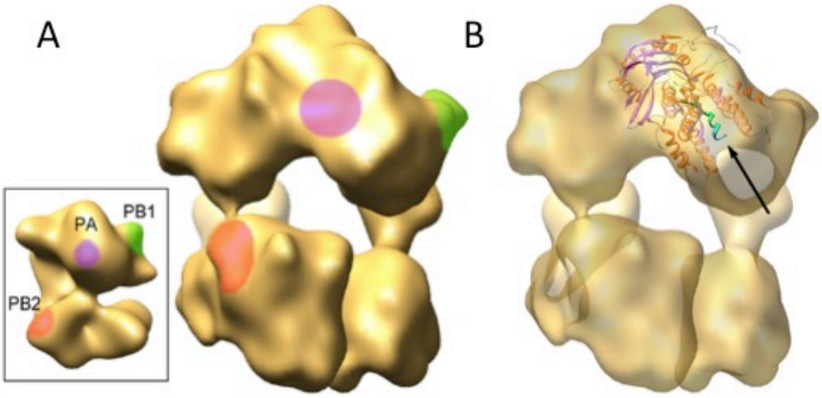


Figure 12. Cryo-EM reconstruction of the IAV polymerase complex. (A) The insert shows the three-dimensional reconstruction for the IAV polymerase complex reported by Area et al (Area et al., 2004). Next to the insert is the IAV polymerase complex reconstruction at higher resolution (Coloma et al., 2009a). The individual domains of the IAV polymerase complex are indicated for the PB1 (green), PB2 (red) and PA (violet) subunits. (B) An identical view of the IAV polymerase shown in A with the docking of the PA (C-terminal)-PB1 (N-terminal) dimer. The N-terminal PB1 peptide is labelled with an arrow and highlighted in green.

RdRps belonging to members of the *Tospovirus* and *Nairovirus* genera are considerably larger than those of other bunyaviruses, with predicted molecular masses of 330 kDa and 450 kDa, respectively (de Haan et al., 1991) (Honig et al., 2004) (Marriott and Nuttall, 1996) (van Poelwijk et al., 1997) (Figure 2). These additional sequences therefore offer additional coding capacity and should comprise multiple functional domains. This is observable when examining the sequence of the N-terminal end of the CCHFV L protein, which displays strong similarity to a cysteine de-ubiquitating protease motif within the ovarian tumor (OTU) super-family. Recently the structure of the OTU protein covalently bound to ubiquitin propylamine, a suicide substrate of the enzyme, was elucidated to 1.7 Å utilizing X-ray crystallographic techniques, and which revealed unique structural elements that define this new subclass of the OTU superfamily (Akutsu et al., 2011, James et al., 2011, Capodagli et al., 2011). This work demonstrated the OTU domain is able to remove both ubiquitin (Ub) and Ub-like ISG-15 modifications from target proteins.

The RIG-I-like receptors (RLR), RIG-I and MDA-5 are responsible for the cellular response for recognizing viral RNA. The RIG-I and MDA-5 protein sensors respond to the presence of viral RNA through the binding of the mitochondrial antiviral-signaling protein (MAVS), and activate two divergent signaling pathways. The first is orchestrated by the activation of TRAF3, which triggers the subsequent activation of the TBK1-IκB kinase (IKKε) complex, leading to the phosphorylation of IRF3, allowing dimerization. Dimerized IRF3 then relocated to the nucleus to initiate transcription of the gene encoding IFN-β (type 1 IFNs). Downstream of MAVS the second pathway involves the TRAF6 adaptor protein and activation NEMO-IKKα/β complex. Once activated the NEMO-IKKα/β complex is responsible for phosphorylation of IκB and subsequent release of NF-κB and subsequent translocation to the nucleus. Once inside the nucleus NF-κB initiates transcription of genes encoding proinflammatory cytokines, such as such as TNF-α.

The production of IFN-β and binding to its receptor activates the Janus kinase-signal transducer and activator of transcription (JAK/STAT), through autocrine and paracrine pathways, which induces the transcription of numerous interferon-stimulated genes (ISGs). ISGs encode RIG-I and MDA5, and the antiviral response molecules OAS, PKR, ISG15, and the Mx proteins, whose activation leads to an antiviral response and limits the spread of viral infection (Bowie and Unterholzner, 2008, Sadler and Williams, 2008).

Apart from phosphorylation the signaling pathways involved in innate immunity also rely on extensive regulation through ubiquitination. ISG15 is a 15 kDa ubiquitin-like protein with two ubiquitin-like domains, and is processed from a 17 kDa precursor, and is one of the most importantly expressed ISGs (Blomstrom et al., 1986, Haas et al., 1987). ISG15 is primarily expressed upon the activation of the JAK/STAT pathway, has been implicated in functioning in a direct or indirect antiviral effect (Harty et al., 2009).

Recent evidence has suggested that one of the major roles of the CCHFV OTU domain involves deubiquitinating the RIG-I protein, and thereby down regulation of RIG-I-mediated innate immune signaling and the antiviral response (van Kasteren et al., 2013). Evidence also exists to suggest the closely related NSDV, and Ganjam viruses

(Genus *nairovirus*) also utilize an OTU domain to block the induction of type 1 and type 2 IFNs (Begum and Wissesman, 1970, Holzer et al., 2011). The OTU domain of the L protein from DUGV has also been shown to possess deubiquitinating and deISGylating activity and inhibits both the TNF-α/NF-κB and JAK/STAT pathways (Bakshi et al., 2013).

It has also been speculated that the relatively small OTU-like domain may be involved in auto-proteolytically cleaving the remainder of the nairovirus L protein N-terminal end (Bergeron et al., 2010a). This cleavage could potentially generate the active L polymerase along with numerous additional protein modules predicted by sequence alignments, to include transcription factors, a gyrase (an enzyme that relieves strain while double stranded DNA (dsDNA) is being unwound by helicase) (Kinsella et al., 2004), helicase (separate two annealed nucleic acid strands) and a topoisomerase (regulates the over winding or under winding of DNA) domains (Honig et al., 2004). Collectively these proteins are dispensable for RNA polymerization functions in the context of a model RNA segment (Bergeron et al., 2010b).

1.7.4 Functions of bunyavirus N proteins in addition to RNA binding

The multifunctional N protein of bunyaviruses not only provides structural uniformity to the RNA genome, but also forms crucial interactions with the membrane glycoproteins Gn and Gc during particle assembly. The N protein also provides a platform by facilitating both the interactions with the viral RdRp and allowing the necessary access to the viral genome (Overby et al., 2007a, Overby et al., 2007b, Ribeiro et al., 2009). One of the long-standing conundrums in nsRNA virology is how the RNPs are able to regulate both transcription and genome replication during infection. Recent analysis of BUNV N protein has suggested the N protein is able to differentially affect replication and transcription activities of assembled RNPs, suggesting it plays a key role in allowing the correct recognition of the bunyavirus RNA template by the L protein (Eifan and Elliott, 2009), although how it achieves this is presently unclear.

1.8 Architecture of the RNPs from segmented and non-segmented negative sense viruses.

1.8.1 RSV and VSV

The mechanim of RNA encapsidation by nsNSVs has been elucidated through high resolution models of the N protein/RNA complexes from the *Rhabdoviridae* family members VSV (Green et al., 2006), and rabies virus (Albertini et al., 2006), and the *Paramyxoviridae* family (Genus *Pneumovirus*) member RSV (Tawar et al., 2009). The N protein structures within each family are highly homologous, and therefore VSV has been chosen to represent rhabdoviruses and RSV, has been chosen to represent paramyxoviruses. Cryo-EM has been used to directly visualize authentic RNPs and allowed superpositioning of the crystal structures to allow for a much more detailed model of RNP structure. Tawar et al (Tawar et al., 2009) reported the right-handed helical EM structure of the RSV N protein/RNA complex (Figure 13A B C, and D). Ge et al (Ge et al., 2010) reported the cryo-EM reconstruction of the entire bullet shaped VSV virion, which included the RNP complex, surrounded by a layer of matrix protein (Figure 13E and F).

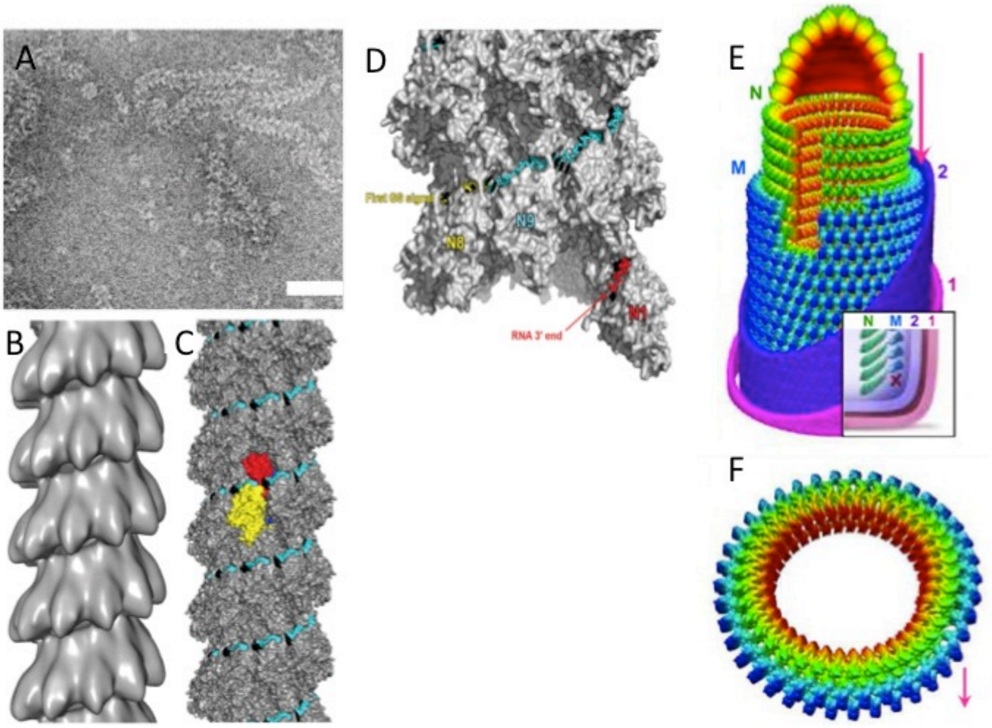


Figure 13. Molecular architecture of nsNSV RNPs. (A) Electron micrograph of negatively stained recombinant RSV N protein helices. Scale bar, 50 nm. (B) Three-dimensional reconstruction at 26 Å resolution of RSV helices from negative stain cryogenic images. The helix has a pitch of 69 Å, with 9.8 N protein subunits per turn. (C) The RSV N protein crystal structure modeled using the contacts into the helix ring, using a pitch of 69 Å which results in 10.35 N subunits per turn (coloured yellow for the N-terminal domain and red for the C-terminal domain, the remaining RSV N proteins are represented by surface representation in gray). (D) Illustration of the location of the RSV 3' RNA within the modeled RSV RNP (nucleotides 1 to 11, red) and initial gene-start (GS) elements (nucleotides 45 to 54, yellow), illustrates how the RNA is accessible for the RSV polymerase complex. (E) Cryo-EM reconstruction in 3D of the VSV virion displaying the M protein, N protein layer, and numbers 1 and 2 label the outer and inner leaflets of the phospholipid bilayer envelope. (F) One complete repeat of the N and M protein helices, involving 75 helical AUs in two turns. The figure was adapted from (Tawar et al., 2009, Ge et al., 2010).

In comparison all these N proteins form linear complexes, which arrange into helical structures. Along the linear N protein complexes the RNA is wrapped within an RNA cavity located between the N- and C-terminal lobes of each subunit, which sequesters the nucleotides through base stacking. The bases of nucleotides 1-4 in rhabdoviruses are stacked and face the solvent side of the RNA cavity, and the bases from nucleotides 5, 7 and 8 are also stacked, but face the interior of the N protein. The RSV N protein uses a similar mechanism, involving stacking of bases from nucleotides 2-4 which face the solvent side of the RNA cavity, whereas bases of nucleotides 5-7 are stacked and face towards the centre of the N subunit (Green et al., 2014b).

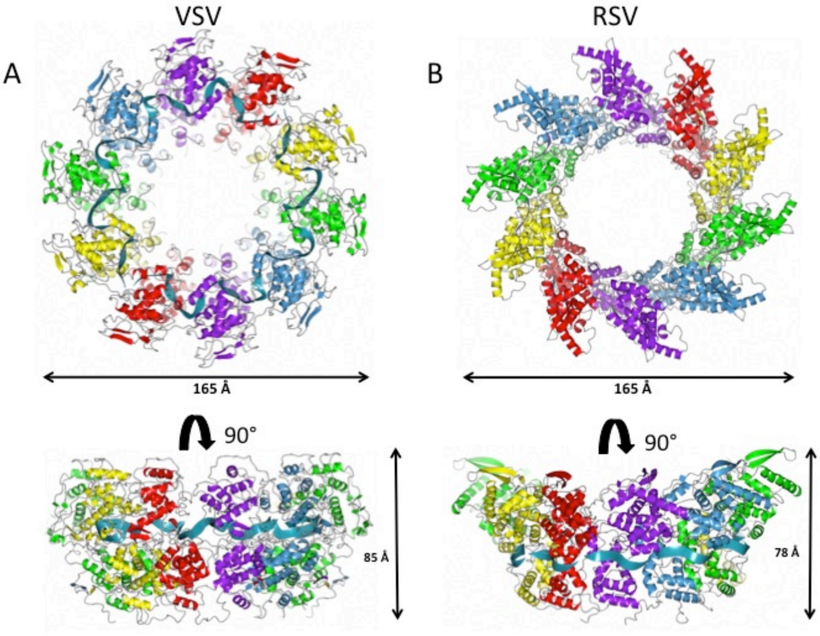


Figure 14. The crystal structures of the N proteins from VSV and RSV of the order *Mononegavirales*. (A) VSV; (PDB code: 2GIC), and (B) RSV (PDB code: 2WJ8). RSV and VSV crystal structure rings are shown in two different orientations, and each subunit is represented by a different colour. The RSV and VSV rings become truncated and form single helical RNPs in which the RNA is wrapped around for protection. The figure was generated using QtMG.

The position of the RNA within the helix however differs between the N protein structures from VSV and RSV. In VSV the N protein forms a cavity within the inside of the helix, whereas for RSV the RNA cavity is positioned facing the exterior of the helix (Figure 14).

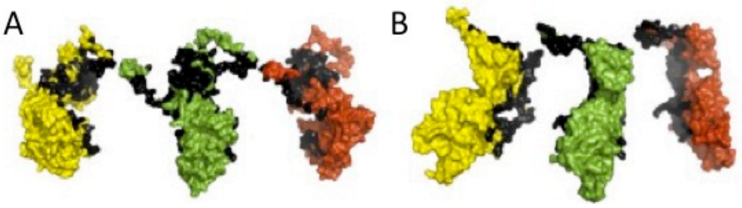


Figure 15. Surface representation displaying the VSV and RSV N protein interfaces involved in oligomerization. (A) VSV, and (B) RSV are shown for three subunits. Each N protein subunit has been radially spaced to display the surfaces that contribute to N-N protein interactions. The three N proteins are coloured red, green and yellow while residues that contact the adjacent protomers are depicted in black. The figure was adapted from (Green et al., 2014a).

1.8.2 IAV

Recently the first complete low resolution model of an RNP from the sNSV IAV virus was reported and revealed striking differences between the RNP structures of segmented and non-segmented negative sense RNA viruses. Elucidation of the atomic structure of authentic RNPs from IAV by two separate groups (Arranz et al., 2012, Moeller et al., 2012) was achieved using ECT, sub volume averaging and previously published crystallographic data. Both groups reported a model in which the RNP consists of two antiparallel strands adopting a double helical structure, leading towards and away from the polymerase complex located at one end of the RNP (Figure 16). The N-N protein interaction involves an extended loop in the C-terminal region that interacts with the C-terminal domain of the adjacent subunit. Because of the lack of high resolution data for the IAV NP N terminus, the atomic models generated for IAV RNPs did not reveal whether the N-terminal region of the NP is involved in the interactions with adjacent subunits (Arranz et al., 2012).

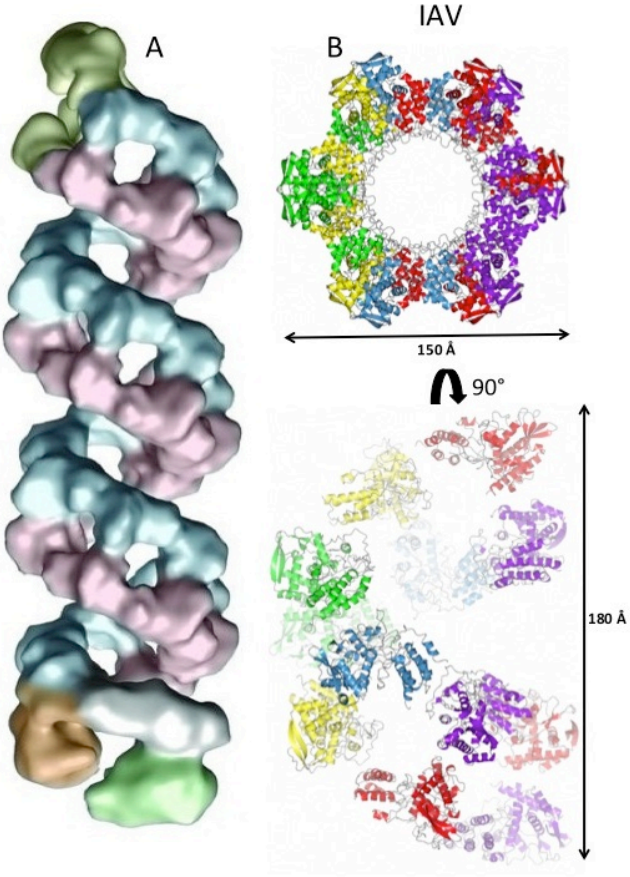


Figure 16. Molecular architecture of the IAV RNP. (A) The three-dimensional model of the IAV RNP complex shows the anti-parallel double helix. The data was derived from an ECT reconstruction and sub volume averaging of infectious RNPs into which (B) the crystal structure of the monomer was modeled (PDB code: 4BBL). The IAV RNPs wind two antiparallel RNAs in opposite directions along its entire length. The figure was generated using QtMG.

1.8.3 Similarities and differences between IAV, RSV and VSV

Each of the RNPs from IAV, RSV and VSV are packaged inside a lipid envelope and appear structurally distinct depending on the virus. The RNP inside the envelope of paramyxoviruses is filamentous, in rhabdoviruses it forms a bullet shape, and in orthomyxoviruses the RNP adopts a double helical structure. The N proteins from RSV and VSV form linear complexes utilising both the N- and C- terminal regions to interact with neighbouring N protein subunits in each direction along the helix respectively (Figure 15). Whereas in IAV the extended loop in the C-terminal region interacts with the C-terminal domain of the next N protein. It has not yet been determined whether the IAV NP N-terminal domain is involved in the interactions with adjacent N protein subunits. The RNP structures from IAV, RSV and VSV reveal N molecules encapsidate the RNA by sequestering nucleotides into a protein cavity. The architecture of the RNA within the RNP in IAV adopts a closed conformation by interaction of both ends with the polymerase. Whereas the RNA within the RNP of VSV and RSV adopts an open conformation, with both the 5’ and 3’ ends located at opposite ends of the helix.

1.9 Architecture of bunyavirus RNPs

The pathogenesis of all bunyaviruses depends on the formation of a RNP complex, in which the genome and antigenome are always associated and entirely encapsidated during the virus life cycle by the S segment encoded N protein. Furthermore, as with all nsRNA viruses, this association is critical and the *Bunyaviridae* family genome is only able to participate in either RNA synthesis or segment packaging when in the form of the RNP.

1.9.1 Phlebovirus N protein crystal structures

The crystal structure of RVFV (phlebovirus) N protein was the first bunyavirus N protein for which high resolution data was available; solved to a resolution of 1.92 Å (Raymond et al., 2010a). The RVFV N protein model displayed N- and C-terminal helical lobes, separated by a short linker helix. Interestingly the folds of both lobes were novel, with neither showing structural homology with any of the other nsRNA virus N protein previously solved. Raymond et al used denaturation and refolding to remove the RNA from the size exclusion chromatography (SEC) purified N protein/RNA multimer complex, which resulted in predominately monomeric RVFV N protein and found a non-helical appearance of the RVFV N protein within the symmetry of the resulting crystal. The observations of no obvious extended terminal loops involved in interacting with adjacent monomers and no positively charged binding groove, suggested at the time that RVFV utilizes a novel strategy of RNP assembly.

The mechanism for oligomerization of the RVFV N protein was uncovered when Ferron et al (Ferron et al., 2011), solved the RVFV N protein to 1.6 Å in hexameric oligomer form with a distinct helical arrangement of the N protein with crystallographic 6-fold and 3-fold symmetry. These ring-shaped hexamers formed a functional RNA binding site, which was assessed using mutagenesis experiments. Additionally, EM analysis demonstrated that the RVFV N hexamer oligomer in complex with RNA also forms rings in solution, and the single particle EM reconstruction of the hexamer N protein/RNA complex is consistent with the crystallographic N protein hexamers. The ring like organization of the RVFV N protein hexamers in the crystal is stabilized by circular interactions of the N-terminal of the adjacent RVFV N protein monomer, which by extending the N-terminal arm, binds to a hydrophobic pocket in the core domain (Ferron et al., 2011) (Figure 17). The two structures also revealed the presence of a hydrophobic pocket at the interface between the two lobes, which was highly conserved among other phleboviruses, and is proposed to represent the site of the previously described interaction with the viral glycoproteins (Overby et al., 2006). Consistent with this observation, the RNPs harvested from infectious virus particles show little evidence of the compact helical RNPs that are characteristic of members of the order *Mononegavirales* (Raymond et al., 2010a). These observations have led to the "beads of a string" model for encapsidation.

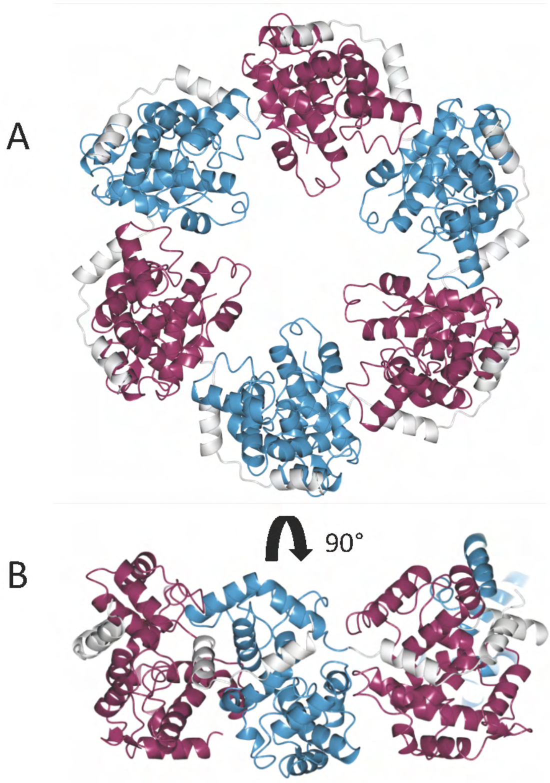


Figure 17. RVFV N protein hexamer represented in cartoon format. (A) Alternating protomers are coloured dark purple and light blue, with the N-terminal arm in white extending from the globular core domain, which fits into the hydrophobic pocket in the surface of the adjacent subunit to form the multimer. (B) RVFV N hexamer rotated 90° to reveal the N- and C-terminal domain interfaces of the individual protomers, where the arm domain binds. The figure was adapted from (Ferron et al., 2011). The figure was generated using QtMG.

The “beads on a string” model for RVFV RNP formation has been strongly supported by three more recent structures, of an RNA bound tetramer, pentamer and hexamer (Figure 18, tetramer PDB code: 4H5P, pentamer PDB code: 4H5O, and hexamer PDB code: 4H5Q). The non–perfect symmetry of the RNA bound multimers demonstrated the flexibility of the arm, which bridge adjacent multimers (Raymond et al., 2012). The core subunit domain displays a narrow deep RNA-binding slot, which becomes a continuous groove and sequesters four nucleotides within the groove and 3 nucleotides (2 for hexamers) at the N/N protein interface by base stacking, so that the bases are nestled into the slot and the phosphate backbone is directed towards the core. This ensures that the RNA is fully encapsidated, including those nucleotides between subunits.

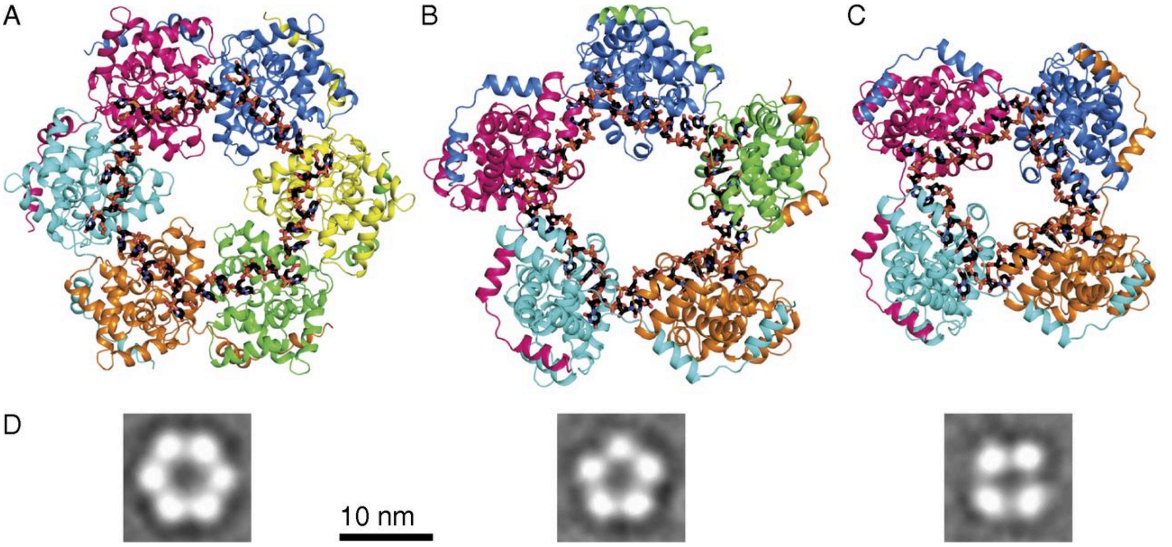


Figure 18. Phlebovirus N protein/RNA complexes. (A) RVFV hexamer bound with 35 nucleotides of ssRNA. (B) RVFV pentamer with 35 nucleotides of ssRNA. (C) RVFV tetramer with 28 nucleotides of ssRNA. N protomers are represented in cartoon format. Oligomerization is facilitated by wrapping of the helical arm of one protomer to the outside of the adjacent protomer. The ssRNA is represented in stick and ball format with black carbon atoms and binds to the RNA binding groove with all bases facing towards the N protein. (D) EM visualization of single particle class averages from hexamer, pentamer, and tetramer multimers, which are representative of multimers released from viral RNPs by RNase digestion. The figure was adapted from (Raymond et al., 2012).

The phlebovirus RNP complex requires very few protein-protein interactions from neighboring N proteins. Presumably the mechanism of oligomerization allows the phlebovirus N protein to become contiguously elongated during RNP formation along the length of the viral RNA to form a long string. In this model the RNA is protected through a continuous N protein pocket, forming the “beads on a string“ like structures seen observations using EM (Raymond et al., 2010b) (Figure 19).

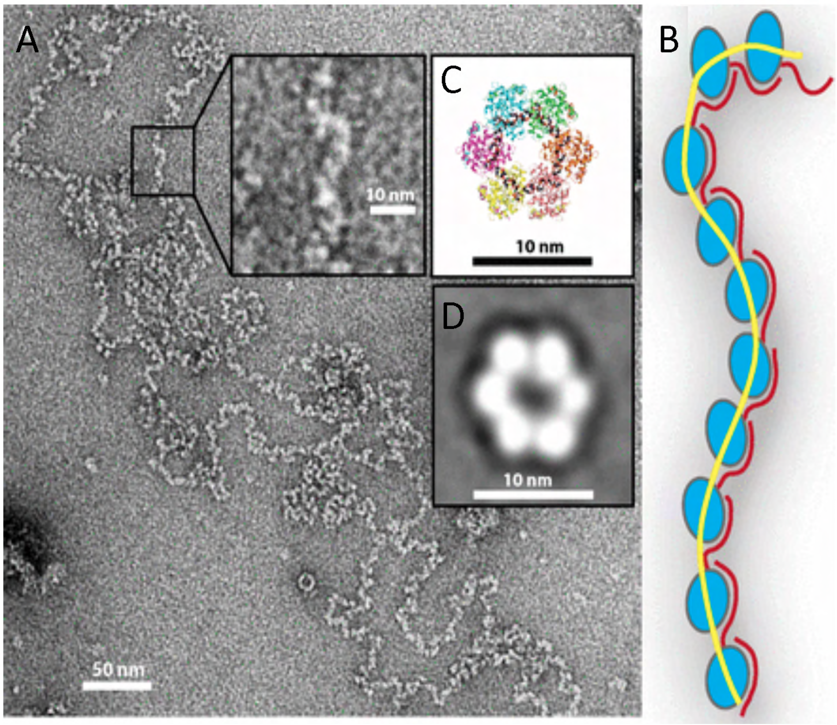


Figure 19. Model for RVFV RNP formation. (A) Authentic RVFV RNPs isolated from virus infected cells visualized by EM with enlarged images to compare with the hexameric structure of RVFV N protein (C) and (D). (B) The “beads on a string” model for RVFV RNP formation is likely achieved through the successive assembly of the monomer-sized N protein/RNA building block. The figure was adapted from (Raymond et al., 2012).

1.9.2 Orthobunyavirus N protein crystal structures

The crystal structures of the orthobunyavirus N protein from BUNV (Figures 20 and 22B, tetramer PDB code: 3ZLA), LACV, Schmallenberg virus (SBV) and Leanyer virus (LEAV) displayed a novel fold for an RNA binding protein (Ariza et al., 2013, Li et al., 2013, Reguera et al., 2013, Niu et al., 2013, Dong et al., 2013b). All structures were crystallized as tetramers as RNA free complexes or in association with RNA derived from the *Escherichia coli* (*E. coli*) expression hosts. For the N protein/RNA complex each monomer binds 11 nucleotides, which sit within an RNA binding pocket, located at the junction of the C- and N-terminal lobes. RNA sequestration occurs primarily from electrostatic and hydrophobic interactions within the large RNA pocket. The structures showed that within the RNA binding pocket there are positively charged residues that twist three bases stacked consecutively on each other (2 to 4) and base stacking of F178 (LEAV; PDB code: 4J1G) dramatically twists the RNA backbone further, giving rise to a pseudo-right handed helical structure (Niu et al., 2013). The RNA is kept inside the ring by a sharp turn in the RNA backbone, which is accommodated by the last base of the eleventh nucleotide stacking against the second base of the next protomer.

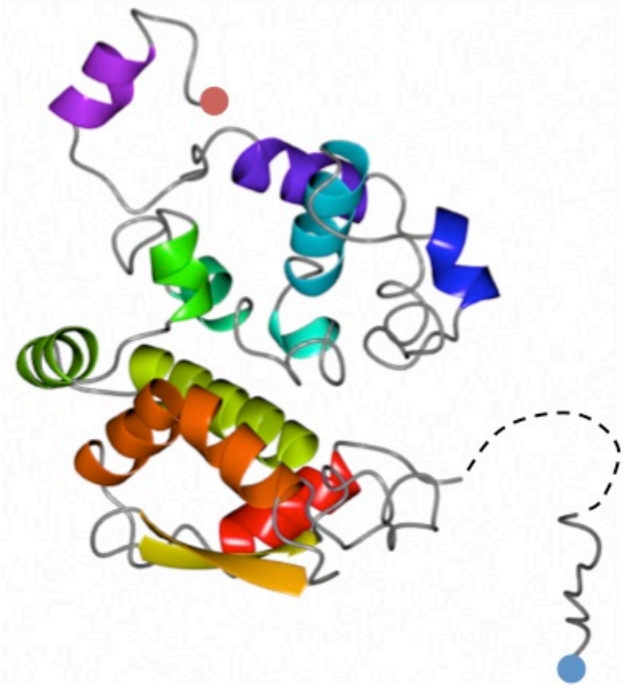


Figure 20. Crystal structure of BUNV N protein. Cartoon representation of the monomer BUNV N protein, with N- and C-termini marked with a blue and red sphere, respectively. Dashed lines represent the disordered region in the crystal structure. The figure was generated using QtMG.

The orthobunyavirus N protein structures also revealed a novel mechanism for oligomerization. The interactions between adjacent monomers on either side involve extensions from both the C- and N-termini in a head to tail arrangement of multimerization. Two models for RNA encapsidation and RNP formation have been proposed based on the comparison of N protein crystal structures and EM images of authentic orthobunyavirus RNPs purified from infected cells (Figure 21A and B).

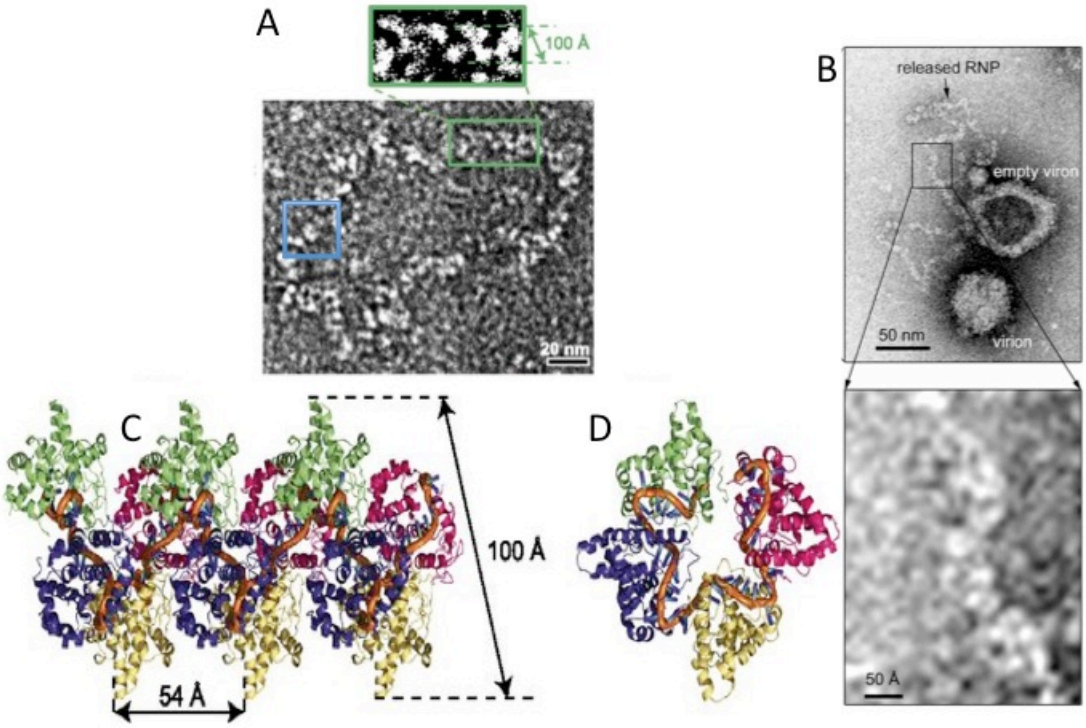


Figure 21. Model for orthobunyavirus RNP formation. (A) EM visualization of LACV RNPs from disrupted virions. The size of the RNP shown is ~160 × 75 nm. Displayed in some areas the LACV RNP display regions of flexibility and in others apparent helical characteristics. Helical areas are highlighted in the green box and shown in close up, whilst a flexible area is shown in the blue box. (B) EM image of negative stained BUNV RNP extracted from BUNV infected cells. The inset frames represents regions that look similar to RVFV RNPs and fit the "beads on a string" model for RNP formation. (C) The crystal packing of LACV N protein in a 4_1 helix as observed in the $P4_1$ crystal form. The width and pitch of the helix is indicated. Modeling of the RNA is represented as tubes in orange tube enwrapped inside the internal surface of the helix. This was achieved with successive 11 nucleotides RNA segments, which are positioned by superposing on the core of each helix protomer the B subunit from the tetramer together with its bound 11 nucleotides of RNA. The continuity of the RNA segments indicates a helical model of RNP assembly, in which a much longer RNA could be enclosed within an extended helix. (D) Upper view of LACV N 4_1 helical/RNA model showing resemblance to the tetramer. The figure was adapted from (Ariza et al., 2013, Li et al., 2013, Reguera et al., 2013)

The first model proposes the "beads on a string" scenario similar to that of the filamentous phlebovirus RNPs. The second model proposes a supercoiling scenario in which stacking of truncated proto-helical tetramers, forms a protective groove for RNA binding within the inside of a continuous symmetrical helix. This model has been supported by the closely matched dimensions of both width and pitch of the helical N protein filaments in EM images for infectious LACV RNPs and those of proto helices of

RNA free tetramers in a crystal structure, presumably truncated during crystal formation (Reguera et al., 2013) (Figure 21C and D). The proto helices in the crystal structure are stacked using $P4_1$ crystal symmetry. In this study successive 11 nucleotide RNA segments were superimposed on to the core of each helix promoter using the RNA bound subunit from the closed tetramer (Figure 21C and D). The modeling revealed that the 11 nucleotide segments join up to form a continuous helical arrangement of RNA, with 44 nucleotides per 54-pitch Å of RNP helix. It is possible that both models are required for different stages of the virus life cycle. For instance, highly condensed helical RNPs may facilitate viral RNA packaging into infectious particles whereas extended RNPs may be required for viral RNA synthesis.

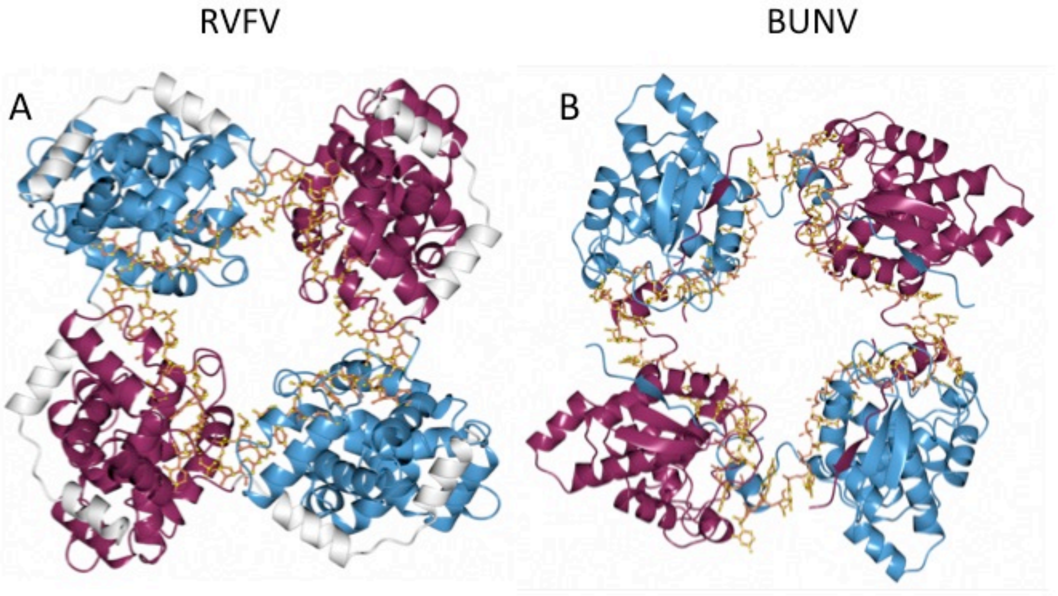


Figure 22. Orthobunyavirus and phlebovirus N protein/RNA tetramer complexes. (A) and (B) display the structures of the N protein/RNA tetramers from (RVFV; PDB code: 4H5P), and BUNV (BUNV; PDB code: 3ZLA) respectively. RVFV and BUNV crystal structure rings are shown, and each subunit is represented by either one of two alternating colours, (dark purple or light blue). For RVFV the N-terminal arm is coloured in white and interact with the adjacent protomer. RVFV and BUNV protein rings become truncated and form single helical RNPs in which the RNA is wrapped around. The figure was generated using QtMG.

Analysis of the BUNV and RVFV N protein crystal structures has allowed a preliminary functional map to be constructed showing regions involved in oligomerization, polymerase binding, and RNP assembly into virus particles (Ariza et al., 2013, Li et al., 2013, Reguera et al., 2013, Niu et al., 2013, Dong et al., 2013b) (Raymond et al., 2012). The BUNV and RVFV N protein crystal structures revealed the stoichiometritry between N protein/RNA complexs within bunyaviruses can vary (Figure 22). The structures of the BUNV and RVFV N protein crystal structures were also distinct from nsRNA virus N protein/RNA complexes solved to date, suggesting the overall architecture of the RNP of bunyaviruses is likely different to those of IAV (Coloma et al., 2009b), RSV (Tawar et al., 2009), VSV (Green et al., 2006) (Figures 13, 14, 15 and 16), and measles virus (Desfosses et al., 2011).

1.9.3 Hantavirus and tospovirus N proteins

In contrast with the phlebovirus and orthobunyaviruses, homotypic oligomerization of the hantavirus N protein has been proposed to involve interactions of adjacent monomers in 'head-to-head' and 'tail-to-tail' arrangements. The hantavirus N protein also forms higher order oligomers, (predominantly trimers) *in vivo* and *in vitro*, and interacting domains have been mapped to both N-and C-terminal regions (Alfadhli et al., 2001, Kaukinen et al., 2001, Kaukinen et al., 2004, Kaukinen et al., 2003, Mir and Panganiban, 2004). The NMR and X-ray crystallography data show the coiled coil domains in the N-terminal region of the hantavirus N protein are insufficient for trimerization, implying important involvement of the C-terminal regions (Boudko et al., 2007, Wang et al., 2008).

Multimerization of the TSWV N protein was proposed to involve a different head-to-tail dimer arrangement, with hydrophobic interacting domains restricted to the extreme terminal regions, however the subsequent identification of additional interacting residues within central regions of the N protein suggests that higher order structures such as tetramers may be the predominant form (Kainz et al., 2004, Uhrig et al., 1999). The differences in the structure and preferred oligomeric states of bunyavirus N proteins may reflect significantly different mechanisms of RNP formation.

1.9.5 Structure of the LASV NP

The crystal structure of LASV NP has recently been elucidated in RNA bound and RNA free forms and has provided insight into the mechanism for arenavirus RNP assembly. The RNA free LASV NP was crystallized as a trimer, and displayed characteristic N- and C-terminal domains. The N-terminal domain folds into a novel structure with a deep cavity, which was initially thought to bind m7GpppN but was later suggested to represent a binding pocket involved in the interaction with a single nucleotide of a bound RNA strand, whereas the C-terminal domain contains 3′-5′ exoribonuclease and exodeoxyribonuclease activities thought to be involved in suppressing innate immune responses (Qi et al., 2010, Brunotte et al., 2011) (Figure 23).

The RNA bound form of the LASV NP was crystallized using the partial N-terminal RNA binding domain in complex with ssRNA and showed RNA binds in a basic crevice located entirely within the N-terminal domain (Figure 23). Hastie et al (Hastie et al., 2011b) proposed that upon RNA binding, the C-terminal domain of NP rotates slightly away from its position in the RNA free trimer, allowing helices α5 and α6 to open away from the RNA-binding crevice to reveal the positive charges residues involved in RNA binding. The gating mechanism proposed may not be unique to nsRNA viruses. The concept of a flexible arm being involved in gating RNA binding has also been proposed for phleboviruses, and orthobunyaviruses, involving the N-terminal arm interacting with an adjacent monomer upon RNA binding, which could be triggered by an as-yet unidentified factor or perhaps the viral genome itself, not only to allow for oligomerization, but also expose the hydrophobic RNA pocket (Ferron et al., 2011).

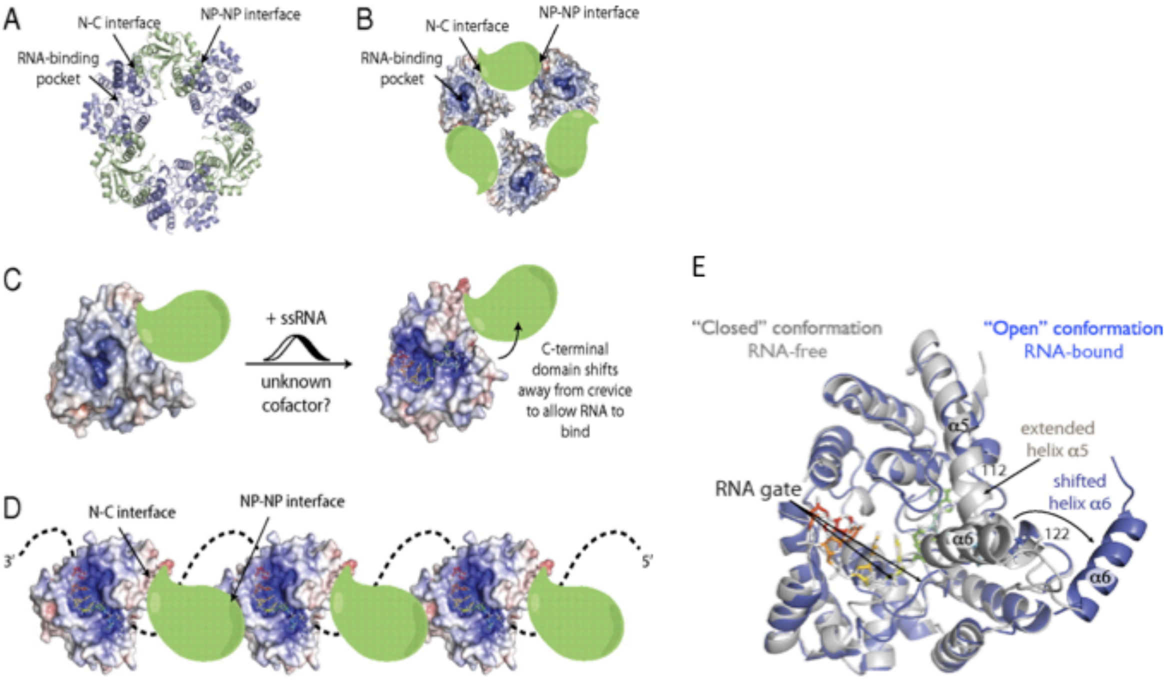


Figure 23. Model for arenavirus RNP organization. (A) Crystal structure of the trimeric, RNA free LASV NP in cartoon representation (PDB code: 3MWP). (B) Electrostatic surface of the N-terminal domain of LASV NP coloured and cartoon representation of the C-terminal domain modeled exhibits the closed form of the LASV NP structure, in which the RNA-binding crevice is not accessible to RNA. (C) The C-terminal domain must change conformation to allow RNA binding within the crevice. An as yet unidentified cofactor or perhaps the viral genome itself could regulate the accessibility of the crevice for RNA binding. (D) The trimer of LASV NP will not form when ssRNA is bound and instead monomers enwrap the RNA genome in a "beads on a string" model from encapsidation. Oligomerization has been proposed to occur when the N-terminal domain interacts with the adjacent C-terminal domain of the adjacent NP. (E) Superimposition of the RNA-free and RNA bound N-terminal domains. RNA binding of the LASV NP is facilitated by conformational changes in the open and closed states by α5 and α6. In RNA-free conformation the RNA gate is closed and α5 extends across the RNA binding pocket, whilst α6 prevents access to the pocket. In the RNA bound state (blue), residues 112-122 are disordered, and α6 opens away from the pocket, accommodating the ssRNA. The figure was adapted from (Hastie et al., 2011b).

1.2 Summary of thesis aims

During this project I have implemented a structural molecular biology approach to investigate the architecture of two bunyavirus N proteins. The N proteins chosen were those encoded by CCHFV and TSWV. The overall aim of the project was to use X-ray crystallography to highlight important structure/function relationships, which could be further analysed using molecular biology techniques.

Before the recent publication of the phlebovirus and orthobunyavirus N protein/RNA crystal structures, there was very limited structural information characterizing the bunyavirus N proteins. This new information has greatly improved the understanding of functions such as RNA binding, oligomerization and assembly of the infectious RNPs. Structural elucidation of the nairovirus and topsovirus N proteins could address important questions about the biology of the *Bunyaviridae* family such as:

- How do the nairovirus and tospovirus N proteins homotypically interact with each other and bind RNA?

- How do the nairovirus and tospovirus N proteins assemble into the RNP?

- Why does the caspase-3 protease cleave the CCHFV N protein?

- How does the caspase-3 protease cleavage of the CCHFV N protein affect its architecture and function?

- What information do the structures of the nairovirus and tospovirus N proteins provide about the evolution of bunyaviruses?

2. Chapter 2 - Materials, Methods, and X-ray Crystallography theory

2.1 Bacterial strains

E. coli DH5α **Genotype:** *F-, Φ80dlacZΔM15, endA1, recA, hsdA1, (rk*$^-$*, mk+), supE44, thi1, deoR, gyrA96, relA1, (ΔlacZYA-argF), U169, λ-,* were made competent in house and used for cloning purposes.

E. coli Gold BL21 (DE3) **Genotype:** *F- ompT, hsdSB, (rB- mB-), gal dcm endA (DE3),* (Stratagene), were made competent in house and used for protein expression.

E. coli Star BL21 (DE3) **Genotype:** *F- ompT, hsdSB, (rB- mB-), gal dcm rne131, (DE3)* (Stratagene), were made competent in house and used for protein expression.

Rosetta™ 2 host strains are BL21 (DE3) derivatives designed to enhance the expression of eukaryotic proteins that contain codons rarely used in *E. coli*. These strains supply tRNAs for 7 rare codons (AGA, AGG, AUA, CUA, GGA, CCC, and CGG) on a compatible chloramphenicol-resistant plasmid. **Genotype:** *F- ompT, hsdSB, (rB- mB-), gal dcm* pRARE2 (Cam^R) (DE3).

Rosetta™ host strains are BL21 (DE3) derivatives designed to enhance the expression of eukaryotic proteins that contain codons rarely used in *E. coli*. These strains supply tRNAs for AGG, AGA, AUA, CUA, CCC, GGA codons on a compatible chloramphenicol-resistant plasmid. The tRNA genes are driven by their native promoter. DE3 indicates that the host is a lysogen of λDE3, and therefore carries a chromosomal copy of the T7 RNA polymerase gene under control of the *lacUV5* promoter. Such strains are suitable for production of protein from target genes cloned in pET vectors by induction with Isopropyl β-D-1-thiogalactopyranoside (IPTG). **Genotype:** *F- ompT, hsdSB, (rB- mB-), gal dcm* pRARE (Cam^R) (DE3).

BL834 (DE3) Methionine deficient *E. coli* cells **Genotype:** F$^-$ *ompT, hsdS*$_B$*, (rB- mB-), gal dcm met* (DE3).

2.2 Plasmid vectors

The plasmid pET-28a-SUMO (made in house in Dr. Thomas Edwards lab) containing the CCHFV N cDNA was used for expression of fusion His$_6$-SUMO CCHFV N protein in BL21 (DE3) *E. coli* cells, and was provided by Roger Hewson (Health Protection Agency, Porton Down, United Kingdom). It carries a kanamycin resistance marker. The pET-15b containing the TSWV N gene was a kind gift form Mark Kainz at Ripon College, Wisconsin, USA. The plasmid pET-28a-SUMO (made in house) containing the TSWV N gene was used for the expression of His$_6$-SUMO TSWV N protein in BL21 (DE3) *E. coli* cells.

2.3 Manipulation of recombinant DNA

2.3.1 Primers

DNA oligonucleotides were purchased from MWG Biotech. A list of primers appears in the appendix, page 198.

2.3.2 Polymerase Chain Reaction (PCR)

PCRs were performed using Phusion DNA polymerase (New England Biolabs) with 50 μl volumes in thin walled 0.2 ml PCR tubes as follows: 5 x Phusion Buffer, 2 μl DMSO, 1.6 μl of a dNTP mix containing 2.5 mM of each dNTP, 2'-deoxyadenosine 5'-triphosphate, 2'-deoxycytidine 5'-triphosphate, 2'-deoxyguanosine 5'-triphosphate, 2'-deoxythymidine 5'-triphosphate, 10-20 ng template of DNA, 1 μmol of the appropriate forward and reverse primers, 1 unit of Phusion DNA polymerase and ddH$_2$0 to make up final volume to 50 μl. A cycling reaction consisted of denaturation at 95 °C for 5 min, 30 x (denaturation at 95 °C for 30 sec, annealing using a gradient 50 °C to 70 °C for 30 sec, and elongation at 72 °C for 1 min).

2.3.3 PCR and enzyme digestion cleanup

The Zymoclean clean and concentrator (Zymo Research), was used to clean up PCR reactions and enzyme digestions. Briefly the mixture was transferred to a Zymo-Spin™ Column² in a collection tube and centrifuged for 30 sec to remove flowthrough. Wash and elution buffers were pre-supplied and protocol conditions were followed according to the manufacturer's instructions (Zymo Research). Typically the elution volume of the DNA sample was 6 μl.

2.3.4 Generation of DNA fragments by restriction endonuclease digestion

PCR products and plasmids were digested with restriction endonuclease enzymes to generate 5' and 3' overhangs. This was performed using 1 unit of enzyme per μg of

DNA (by definition, 1 unit of restriction endonuclease will digest 1 µg of substrate DNA in a 50 µl volume in 60 min). Buffers and reaction conditions were provided and followed according to the manufacturer's instructions (New England Biolabs), which typically contained 10x Buffer 4 (recommended for *BamHI* and *XhoI* double digestion), 10x BSA and ddH$_2$O to make up the final volume. Reactions were carried out for 2 hr at 37 °C. Digestion was carried out for PCR diagnostic purposes and for recombinant DNA manipulation, which typically contained between 1-2 µg and 20-30 µg DNA respectively.

2.3.5 Purification of DNA fragments

Using a UV box DNA bands were visualized using long wavelength UV, and excised from the gels with a sterile scalpel. Gel slices were transferred to a 1.5 ml micro-centrifuge tube and weighed. The QIAquick gel extraction kit protocol (Qiagen) was used to extract and purify the DNA. Briefly the gel slice was mixed with 1:3 w/v of buffer QG and heated at 50 °C, until the gel slice dissolved. Buffers were pre-supplied and protocol conditions were followed according to the manufacturer's instructions (Qiagen).

2.3.6 Analysis of DNA concentration

The Thermo Scientific NanoDrop™ 1000 Spectrophotometer was used to measure the concentration of purified DNA samples. The DNA-50 function was chosen, at 260 nm, where an A_{260} of 1.0 = 50 µg ml^{-1} pure DNA. The machine was blanked/reblanked by adding 1 µl of ddH$_2$O and elution buffer (Qiagen) respectively to the pedestal. 1 µl DNA was analyzed and the concentration was given in ng/µl.

2.3.7 Ligation of DNA fragments

Ligation reactions were performed using T4 DNA ligase (New England Biolabs). Insert versus vector DNA were used for ligation in molar ratios of 15:1 or 3:1. Typically 1 µl of T4 DNA ligase was added to 20 µl of reaction, including 10 x T4 DNA ligase reaction buffer (New England Biolabs) (50 mM Tris-HCl, 10 mM MgCl$_2$, 10 mM DTT, 1 mM ATP, pH 7.5) and ddH$_2$O to make up the 20 µl total volume. Reactions were incubated at 25 °C for 2 hr or overnight at 25 °C. Successful ligations were determined by transformation into competent *E. coli* cells under appropriate antibiotic selection.

2.4 Transformation of chemically competent *E. coli* bacterial cells

2.4.1 Star BL21 (DE3), Gold BL21 (DE3), Rosetta BL21 (DE3), Rosetta2 BL21 (DE3), and DH5α in house competent *E. coli* cells

Frozen competent *E. coli* cells were thawed in ice (50 μl aliquots). 10-100 ng DNA was then added in 1 μl volume for maximum efficiency and incubated for 30 min. Cells were then heat shocked in a water bath at 42 °C for 45 sec, and then immediately put on ice for 2 min. 250 μl of Luria Bertani (LB) medium (Novagen) was then added and the cells and were incubated at 37 °C for one hour. Cultures were plated onto LB agar plates (1% (w/v) tryptone, 1% (w/v) NaCl, 0.5% yeast extract, [pH 7.0]) containing 100 μg/ml of appropriate antibiotic. Typically 200 μl and 50 μl were plated out respectively, using aseptic technique. Plates were incubated at 37 °C for 16 hr.

2.4.2 Controls of chemically competent *E. coli* cells

To control whether these cells were adequately competent a pUC plasmid containing the ampicillin resistance gene was used in transformation experiments. Negative controls included a transformation protocol that did not include pUC plasmid DNA.

2.5 Growth of bacterial cultures and preparation of plasmid DNA

2.5.1 Purification of plasmid dsDNA from *E. coli* cells

A single colony of transformed *E. coli* cells from ampicillin plates was transferred to 5 ml LB medium in a sterile 20 ml universal bottle, containing 100 μg/ml ampicillin and grown overnight at 37 °C. Plasmid extraction and purification was achieved using the plamid miniprep extraction kit (Qiagen). All reaction reagents were provided and the protocol was followed according to the manufacturer's instructions.

2.5.2 dsDNA sequencing

GATC Biotech carried out DNA sequencing commercially. For regular plasmids, 30-100 ng/μl DNA (30 μl total) was required. Primers were provided by the sequencing service and DNA sequences were assessed using the software program 4peaks (Mekentosj.com).

2.6 Protein biochemistry

2.6.1 Transformation of bacterial expression strains with pET28a-SUMO CCHFV N

Bacterial expression cells were transformed with pET28a-SUMO CCHFV N, plated out on kanamycin and chloramphenicol selective media, and incubated overnight at 37 °C. Negative controls included a transformation protocol with a pET28a-SUMO vector which did not include the CCHFV N gene. Colonies were picked and grown overnight in culture, containing 5 ml LB medium.

2.6.2 Small scale expression of CCHFV N protein

4 ml of the overnight culture were subsequently used to inoculate 400 ml of LB and grown at 37 °C until the OD_{600} reached 0.6-0.7. Expression was induced by the addition of 100 mM or 100 μM IPTG, and incubation for a further 12 hr at 18 °C or 37 °C. Cells were harvested by centrifugation at 6000 rpm for 15 min in a Sorvall RC5B Plus centrifuge using a Sorvall SLA-1500 rotor. The resulting pellet was re-suspended in 25 ml of lysis buffer, (500 mM NaCl, 20 mM Tris-HCl, 0.1 % Triton, pH 7), and stored on ice. Cell lysis was performed using one slow freeze thaw cycle at -80 °C and two fast sequential freeze thaw cycles using liquid nitrogen. The soluble fraction was purified typically from individual 1 ml volumes of cell lysates, which were clarified by centrifugation in an Eppendorf 5415 centrifuge at 13000 rpm for 15 min.

2.6.3 Large scale expression of CCHFV N protein

To express optimum concentrations of His_6-SUMO CCHFV N suitable for crystallization Rosetta2 *E. coli* BL21 (DE3) were transformed with pET28a-SUMO CCHFV N, plated out on kanamycin and chloramphenicol selective media, and incubated overnight at 37 °C. Colonies were picked and grown overnight in culture, containing 5 ml LB medium. The next day 4 ml of the overnight culture was subsequently used to inoculate 6 L of LB media and grown at 37 °C until the OD_{600} reached 0.6-0.8. Expression was induced by the addition of 100 μM IPTG (final concentration), and incubated for a further 12 hr at 18 °C. Cells were harvested by centrifugation at 6000 rpm for 15 min in a Sorvall RC5B Plus centrifuge using a Sorvall SLC-6000 rotor. Each pellet was re-suspended in 25 ml of lysis buffer (500 mM NaCl, 20 mM Tris-HCl (pH 7), 0.1% Triton X-100, 5% v/v glycerol, 62.5 μl chicken-egg white lysozyme solution (10 mg ml^{-1}), 1 Unit (U) of deoxyribonuclease 1 (DNAse 1) (1 U/μl) (Invitrogen), 25 U of benzonase (25 U/μl) (Novagen), 20 μl $MgCl^2$ (50 mM)), and stored in a 50 ml falcon tube on ice.

Each 25 ml suspension was subject to cell lysis, which was performed in 20 sequential cycles at 18 microns amplitude sonication; each cycle consisted of 10 sec on and 5 sec off, using a soniprep 150 (MSE). The sample was incubated at room temperature for 10 min and lysates were subjected to an additional cell lysis step using one slow freeze thaw cycle at -80 °C. Finally soluble and insoluble fractions were separated by centrifugation using a Sorvall RC5B Plus centrifuge and an SS-34 rotor at 14000 rpm for 15 min.

2.6.4 Nickel (Ni) Superflow Plus resin protein purification

To assess the ability of the Rosetta2 *E. coli* BL21 (DE3) expressed His_6-SUMO CCHFV N fusion protein to bind to the Ni-Superflow Plus resin (Generon), the desktop spin protocol was implemented using micro-spin columns (Thermo-Scientific). Initially 50 μl of Ni resin was added to the column, and washed to remove any residual ethanol with 200 μl of ddH_2O by centrifugation at 500 rpm for 30 sec in an Eppendorf 5415 D tabletop centrifuge. The resin was subsequently washed as in the previous step in binding buffer (500 mM NaCl, 20 mM Tris-HCl, pH 7), following subsequent addition of 50 μl of the supernatant from 1ml of cell lysate. Ni-bound protein was washed sequentially in wash buffers (500 mM NaCl, 20 mM Tris-HCl, pH 7), and (50 mM imidazole, 500 mM NaCl, 20 mM Tris-HCl, pH 7). The bound protein was eluted using elution buffer (500 mM imidazole, 300 mM NaCl, 20 mM Tris-HCl pH 7).

Large-scale purification of His_6-SUMO CCHFV N protein involved a first step purification method using Ni-Superflow Plus resin (Generon). Typically 2 ml of Ni resin was added to the supernatant of each 25 ml cell lysis suspension and left to bind on a blood mixer for 2 hr at 4 °C. The suspension was centrifuged at 2500 rpm for 3 min, the supernatant fraction was collected and the remaining resin pellet was washed with binding buffer (500 mM NaCl, 20 mM Tris-HCl, pH 7), on a blood mixer as described in the previous step. This was repeated with 3 subsequent wash steps, which included (50 mM imidazole, 500 mM NaCl, 20 mM Tris-HCl, pH 7), (90 mM imidazole, 500 mM NaCl, 20 mM Tris-HCl, pH 7), and (102 mM imidazole, 500 mM NaCl, 20 mM Tris-HCl, pH 7).). The bound protein was eluted using typically 15 ml of elution buffer (300 mM imidazole, 300 mM NaCl, 20 mM Tris-HCl pH 7) and 50 μl from the elution was collected and re-suspended in SDS-PAGE sample buffer for analysis.

2.6.5 Dialysis of imidazole from CCHFV N protein

To remove imidazole from the nickel purified protein the 15 ml elution sample was added to Snakeskin dialysis tubing (Thermo Scientific) and dialyzed overnight at 4 °C against dialysis buffer (500 mM NaCl, 20 mM Tris-HCl, 10 mM EDTA, pH 7).

2.6.6 Proteolysis using the SUMO protease

The SUMO protease was used to separate the His_6-SUMO moiety from the His_6-SUMO CCHFV N fusion protein to produce native CCHFV N protein. Typically the concentration of SUMO protease added was 1:100 to His_6-SUMO CCHFV N fusion protein and incubation was for 12 hr at 4 °C in a dialysis bag.

2.6.7 Size exclusion chromatography (SEC)

SEC was performed using an Akta Purifier (GE Healthcare). S75 and S200 superdex gel filtration preparative columns were used for purification of TSWV and CCHFV N proteins. The S75 analytical column was used for analyzing CCHFV N protein. The unicorn program was used to operate all steps of the purification, which were performed at 4 °C. Briefly the column was equilibrated overnight using filtered/degassed running buffer (10 mM EDTA, 20 mM Tris-HCl, 100 mM NaCl, pH 7), with a flow rate of 0.5 ml min^{-1}, until the absorbance trace at 280 nm reached a steady baseline. The injection loop was washed using running buffer to remove any contaminant protein and injected with 5ml of protein sample, after which the injection valve was selected to inject. The flow rate was set to 1 ml min^{-1} until the sample reached the top of the column, which typically took 7 ml and subsequently the injection valve was set to load and flow rate increased to 2.5 ml min^{-1}. Spectrophotometer readings were taken at both 260 nm and 280 nm and purified protein was collected in 3 ml fractions. Fractions were analyzed using SDS-PAGE to check for purity.

2.6.8 Sodium-dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was used to analyse purified protein fractions using a minigel system (Biorad mini-Protean III). Resolving gel solution (15% acrylamide (v/v), 0.375 M Tris-HCl [pH 8.8], 0.1% SDS (w/v), 0.1% APS (w/v), 0.01% N,N,N',N'-Tetramethylethylenediamine (TEMED) (v/v)) was poured between the glass plates, until four fifths full. A layer of isopropanol (100%) was added to ensure the gel had a flat upper surface after initial polymerization at room temperature. Isopropanol was removed by washing with ddH_2O. Stacking gel (5% acrylamide, 0.126 M Tris-HCl (pH 6.8), 0.1% SDS (w/v), 0.1% APS (w/v), 0.01% TEMED (v/v)) was poured onto the surface of resolving gel, whilst a 10 or 15 well comb was placed between the glass plates before polymerisation was allowed to occur and removed afterwards. 2x Laemmli loading buffer was added to the protein sample (70 mM Tris-HCl (pH 6.8), 4% SDS (w/v) 15% glycerol (v/v) 10 mM DTT, 1% β - mercaptoethanol (w/v), 0.01%

bromophenol blue (w/v)) and heated at 95 °C for 5 min prior to loading. Samples were applied to gels suspended in 1 X SDS-PAGE-running buffer (50 mM Tris-HCl (pH 8.3), 3.84 mM glycine, 0.1% SDS). 1% (v/v). 6 μl of SeeBlue Plus2 pre-stained molecular weight markers (Invitrogen) was used to determine size (kDa) of protein species. Gels were run at 15 V constant until markers were separated to a good resolution.

2.6.9 Coomassie brilliant blue R-250 staining

Acrylamide gels were immersed in Coomassie brilliant blue R-250 stain (50% methanol (v/v), 10% glacial acetic acid (v/v), 0.2% (w/v) Coomassie brilliant blue R-250) for 2 hr under constant rocking at room temperature. Destaining was achieved using a gel destaining solution (50% methanol (v/v), 10% acetic acid (v/v)) until protein bands were visible.

2.6.10 Bio-Rad assay

The protein concentration throughout purification was measured using a spectrophotometer and the colourimetric Bradford assay (Biorad). The spectrophotometer was blanked at wavelength 595 nm using a 1/5 dilution of Bio-Rad dye with ddH_2O in a 1.5 ml cuvette. 1 μl of protein sample was added in a separate dilution, transferred to a cuvette and after colour development the reading at wavelength 595 nm was recorded.

2.6.11 Circular dichroism

The CCHFV N protein was subjected to circular dichroism (CD) to obtain insight into its secondary structure. Recombinant CCHFV N protein was diluted to a concentration of 0.5 mg ml^{-1} in 50 mM sodium phosphate buffer pH 7.5. 200 ml of filtered protein solution was placed in a 1 mm path-length far- and near-UV CD cuvette (Hellma precision cells, synthetic far-UV quartz). CD spectra were measured between 260 nm and 180 nm at a scanning speed of 1 nm s^{-1} using a Chirascan CD spectrometer (Applied Photo-physics). Two spectra were recorded and averaged, followed by baseline correction via subtraction of the buffer spectrum. The mean residue ellipticity was calculated and expressed in units of deg cm^2 $dmol^{-1}$.

2.6.12 Mass spectrometry.

Samples (10 μM in water) were diluted to 50% acetonitrile/0.1 formic acid, and analysed by Z-spray nanoelectrospray ionisation MS using a quadrupole-IMS-orthogonal time-of-flight MS (Synapt HDMS, Waters UK Ltd., Manchester, U.K.) using in-house fabricated gold/palladium coated nanospray capillaries. The MS was operated in positive TOF mode using a capillary voltage of 1.2 kV, cone voltage of 20 V, nanoelectrospray nitrogen gas pressure of 0.1 bar, backing pressure of 2.47 mbar and a trap bias of 4 V. The source and desolvation temperatures were set at 80 °C and 150 °C, respectively. During TOF-MS acquisition, Argon was used as the buffer gas, at a pressure of 4.0×10^{-3} mbar in the trap and transfer regions. Mass calibration was performed by a separate injection of sodium iodide at a concentration of 2 μg μl^{-1}. Data processing was performed using the MassLynx v4.1 suite of software supplied with the mass spectrometer.

2.6.13 Phenol and chloroform extraction of RNA from CCHFV N protein

The RNA binding ability of the CCHFV N protein complex was analysed by RNA extraction using equal volumes of phenol and chloroform, followed by ethanol precipitation. Typically 700 μl of purified protein (1 mg ml^{-1}) was added to 700 μl of phenol and vortexed. The aqueous phase was then taken, and an equal volume of chloroform was added and the mixture was vortexed again. The aqueous phase was again harvested, whilst 10 μl of 4 M NaCl and 1 ml of 100% ethanol were added. RNA was precipitated by incubation for 2 hr at –80 °C. To determine the size of the harvested RNA, samples were electrophoresed on 1.5% TBE non-denaturing agarose gel and compared to the mobility of low range RNA size markers (New England Biolabs) run in adjacent lanes. RNA was stained with ethidium bromide and visualized using long wavelength UV.

2.7 *In vitro* caspase-3 protease cleavage and analysis

Previously published methods were used to express and purify caspase-3 protease (Denault and Salvesen, 2003). The caspase-3 protease (Gene ID: 836) expression plasmid was purchased from Addgene and designed so that, when expressed, the C-terminal end of the protein would possess a His_6 tag (Zhou et al., 1997). SDS-PAGE analysis confirmed caspase-3 protease expression and the concentration estimated by the Bradford assay (Bio-Rad Protein Dye Reagent; Bio-Rad), against a BSA standard concentrated to 10 mg ml^{-1}. The caspase-3 protease purified protein was concentrated to 10 mg ml^{-1}.

Various *in vitro* reactions to measure the caspase-3 protease cleavage of wild-type CCHFV N protein (DEVD), and CCHFV N protein mutants (DEVE and DEVG), were performed using different molar ratios of caspase-3 protease and CCHFV N protein.

Briefly reactions were performed in 1.5-ml microcentrifuge tubes containing caspase-3 protease digestion buffer 200 mM NaCl 20% sucrose, 0.2% (w/v) 3-[3-cholamidopropyl)-dimethylammonio-1-propanesulfonate (CHAPS), 2 mM EDTA, to reach the desired total volume of 40 µl, and were incubated at room temperature overnight. The caspase-3 protease digestion of CCHFV N protein was analyzed using SDS-PAGE and confirmation of caspase-3 protease activity was achieved by monitoring digestion of a flurogenic substrate IX (Merck), derived from poly (ADP-ribose) polymerase (PARP), using a FLUOstar Optima plate reader (BMG Labtech) at excitation and absorption wavelengths of 485 nm and 520 nm, respectively. The CCHFV N protein cleavage products were analyzed using an SEC S200 10/300 GL analytical column (GE Healthcare), which was preequilibrated in buffer (100 mM NaCl, 20 mM Tris-HCl, and 1 mM EDTA, pH 7.5), at a flow rate of 0.5 ml min^{-1}, with UV absorbance monitored at 280 nm.

2.8 Phylogenetic analysis CCHFV N amino acid sequences

MEGA6 (Tamura et al., 2013) was used to analyze the phylogenetic relationships of CCHFV N and L protein sequences with those of other nsRNA viruses with the default settings for multiple sequence alignment analyses using the CLUSTALW (Thompson et al., 1994) algorithm using the BLOSUM (Henikoff and Henikoff, 1992) protein weighting matrix. Phylogenitic trees were generated using the maximum likelihood statistical method and the Dayhoff substitution model (Dayoff et al., 1978). Rates amoung sites were gamma distributed in 5 categories and 95% partial deletion for site coverage cutoff was implemented. The nsRNA viruses used in the comparative N protein sequence phylogeny were DUGV, LCMV, LASV, Pichinde virus (PICV), TACV, BUNV, LACV, RVFV, HTNV, rice stripe virus (RStV), IAV, and Thogoto virus (THOV) and TSWV. In total 500 bootstrap trials were conducted on each submitted alignment file. Virus species abbreviations (GenBank accession numbers) are as follows: DUGV (P15190.1); CCHFV (AAB48503.1); LCMV (AAX49342.1); LASV (ADY11071.1); PICV, (AAC32282.1); TCRV, Tacaribe virus (AAA47903.1); BUNV (P16495.1); LACV (AAM94389.1); RVFV, (ABP88854.1); HTNV (AFA36178.1); RStV (P68559.2); TSWV (AAW48185.1); IAV (VHIVX1); and THOV (YP_145809.1). The amino acid sequences of 14 viral polymerases, including those from six bunyaviruses and four arenaviruses, were also aligned using the same method as for the N protein sequences. GenBank accession numbers for the L protein sequence are as follows: DUGV, Q66431; CCHFV, AAQ98866.1; LCMV, AB196829.1; LASV, CCA30313; PICV, AAL16099.1; TCRV,

AAA47901.1; BUNV, P20470; LACV, Q8JPR2; RVFV, P27316; HTNV, AAG10042.1; RStV, ADE60705.1; VSV, AAA48371.1; IAV, AAA43639.1; and THOV, O41353. The resultant phylogens were generated using M6 Tree Explorer. The percentage each branch was obtained is indicated at the base of each branch.

2.10 Crystallization

2.10.1 Preparing N protein for crystallization trials

The CCHFV N protein pure fractions were combined into one sample and concentrated to 500 μl using centrifugal filter units with a 20,000 Da molecular weight cut off point (Generon). To cryo-protect the CCHFV N protein 10% glycerol (v/v) was added and the concentrated sample separated into aliquots of 50 μl volumes in 1.5 ml Eppendorf tubes, which were snap frozen in liquid nitrogen and stored at -80 °C. The TSWV N protein pure fraction were combined into one sample and concentrated to 750 μl using centrifugal filter units with a 20,000 Da molecular weight cut off point (Generon). Purified TSWV N protein was stored at 4 °C.

2.10.2 Crystallization factorial screens

CCHFV and TSWV N protein factorial crystal screens using the sitting drop method of crystallization were made using five crystal screening kits, which included crystal screen, crystal screen2, index, salt (Hampton research), wizard 1 and wizard 2 (Emerald BioSystems). Initially 1 ml of each condition was added to a 96 well starter block, which was the stock solution used to transfer 60 μl of mother liquor to MRC plates (Molecular Dimensions). The 96 well starter blocks were sealed with foil and kept at 4 °C. The Oryx 6 Douglas crystallization robot (Douglas Instruments) was then used to pipette a 0.2 μl drop consisting of 0.1 μl of mother liquor and 0.1 μl purified CCHFV N (40 mg ml^{-1}), and TSWV N (15 mg ml^{-1}) protein on to each of the two wells next to each individual crystallization screen condition, which would then equilibrate over 59.9 μl well solution. The wasp and front panel programs were used to operate all steps of the preparation and automation of the crystallization robot. Plates were subsequently sealed with transparent microplate sealer (Greiner bio-one) and incubated at 25 °C for two weeks.

2.10.3 Optimization of crystals

2.10.3.1 Optimization of the CCHFV N protein crystals

To improve the quality of initial crystals produced by screening, optimization was implemented using the hanging drop method of crystallization. For CCHFV N protein

1 ml of mother liquor (100 µl of 2 M NaCl, mixed with the chosen amount of 50% PEG-600, and made up to 1 ml with 1 M bis-Tris propane at varied pH) was added to each well of a 24 well plate. Typically 2.5 µl of purified CCHFV N protein and 2.5 µl of mother liquor was placed on to a coverslip. The drop was equilibrated over the corresponding well, which was sealed with the underside of the coverslip, using vacuum grease. Plates were typically incubated at 25 °C for two weeks.

2.10.3.2 Optimization of the TSWV N protein crystals

For TSWV N protein 1 ml of mother liquor (25% PEG-550 mono methyl ether (MME), made up with 0.5 M 2-(*N*-morpholino) ethanesulfonic acid (MES), was added to each well of a 24 well plate. Typically the best crystals would grow from mixing 3 µl purified TSWV N protein at 15 mg ml^{-1}, and 1 µl of mother liquor into a drop and suspending on a cover slip. The drop was equilibrated over the corresponding well, which was sealed with the underside of the coverslip using vacuum grease. Well diffracting TSWV N protein crystals appeared after 9 hr at 25 °C.

2.11 X-ray diffraction data collection for CCHFV and TSWV N protein crystals

2.11.1 CCHFV N protein crystals

Native *C*2 X-ray diffraction data were collected from CCHFV N protein crystals cryo-preserved at 100 K on beamline I04-1 at Diamond Light Source using 12,658 electron volts (eV) X-rays at wavelength 0.9713 Å with a Pilatus 6M detector. Native $P2_12_12_1$ X-ray diffraction data were collected from CCHFV N protein crystals cryo-preserved at 100 K on beamline I02 using 12,658 eV X-rays at wavelength 0.9795 Å with a Pilatus 6M detector. Derivative X-ray diffraction data were collected for CCHFV N crystals soaked with thimerosal cryo-preserved at 100 K on beamline I02 at Diamond Light Source using 12,658 eV X-rays at wavelength 0.9795 Å with a Pilatus 6M detector. Native and derivative data were indexed and integrated using the MOSFLM package (Leslie, 2006) and scaled using SCALA (Evans, 2006b) from the CCP4 suite (Winn et al., 2011).

2.11.2 TSWV N protein crystals

Native X-ray diffraction data were collected from crystals cryo-preserved using 30% PEG-550 MME, and 70% 150 mM NaCl, 20 mM Tris-HCl, at 100 K on beamline I02 at Diamond Light Source using 12,647 eV X-rays at wavelength 0.9803 Å with a Pilatus 6M detector and 0.5° oscillations. Data were indexed with POINTLESS (Evans, 2011) and integrated with XDS and scaled and merged with XSCALE (Evans and Murshudov, 2013) using the automated program XIA2 (Winter et al., 2013, Kabsch, 2010).

2.12 Selenomethionine derivative protein

Expression of TSWV and CCHFV N protein containing selenomethionine was achieved by transforming the pET28a-SUMO CCHFV or TSWV N and pRARE plasmids into the BL834 (DE3) *E. coli* expression cells, which are unable to synthesize their own methionine. Methionine deficient bacterial cells expressing selenomethionine derivative TSWV and CCHFV N proteins were grown in M9 minimal media and nutrients that were pre-made by Molecular Dimensions, with conditions that were followed according to the manufacturer's instructions (Molecular Dimensions).

2.13 Phasing

Phases contain more information than the diffraction intensities, and therefore estimating them is vital for constructing the electron density. Determination of a marker atom substructure is the most common way of calculating phases lost due to the physical process of measuring X-rays by a detector. Traditionally the most common method for calculating phases is by exploiting the large electronic differences provided by heavy atoms in reference to the isomorphous structure. The electronic differences cause intensity changes relative to the reference structure and the differences between the calculated structure factor amplitudes can be used to find the source of the electronic difference, the marker atom. Differences can be introduced by soaking heavy atoms into the crystal, or through protein engineering, such as selenium being introduced into selenomethionine. As a consequence of dispersive differences at different wavelengths and anomalous scattering factor contributions, intensity differences can also arise from native heavy atoms, such as those from metal binding proteins.

For heavy atom derivatives the measurable changes in reflection intensities, are considerable enough to estimate phases. The additive affect from contributions of all atoms which make up the intensity of a reflection, means each complex structure factor of the derivatives F_{PA} is the complex (vector) sum of the partial structure factor contributions from the marker atom F_A and from the protein F_p (Figure 24).

The amplitude of the heavy atom alone is given by

$(F_A = F_{PA} - F_P)$

The phase angle of the heavy atom can then be applied to the find the phase angle of F_p.

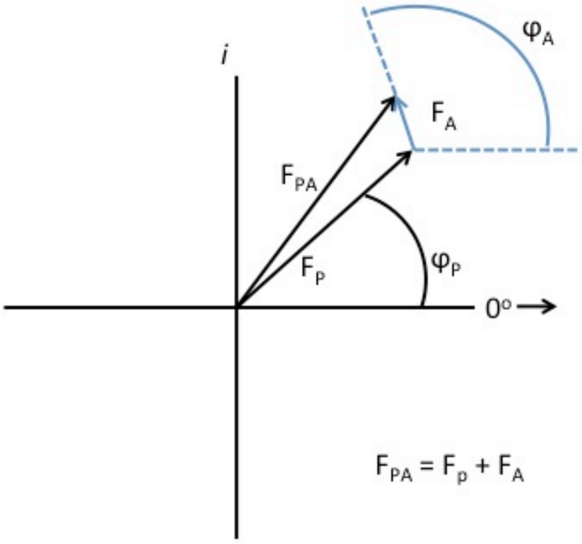


Figure 24. Schematic of the complex structure factors for protein with and without heavy atoms. The complex structure factors for the derivative F_{PA} is the vector sum of the native structure factor for the protein alone F_p and the heavy atom structure factor F_A. The figure was adapted from Biomolecular Crystallography - Bernard Rupp

2.13.1 Isomorphous replacement

The classical sources of electronic differences are heavy atoms, which are commonly soaked into the crystal. The intensity differences between that of the native and derivative crystals are exploited for substructure solution and the method is called either single isomorphous replacement (SIR) or multiple isomorphous replacement (MIR), depending on how many derivatives are used. However when using SIR or single-wavelength anomalous dispersion (SAD) (see below), on their own the phase angle is ambiguous because of the two fold degeneracy and hand uncertainty (Matthew, 1966) (Figure 25). This can be resolved by using additional anomalous signal from anomalous scattering of derivative atoms at certain absorbance wavelengths, and thus the so called single isomorphous replacement with anomalous scattering (SIRAS), or multiple isomorphous replacement with anomalous scattering (MIRAS) are standard methods to break phase ambiguity.

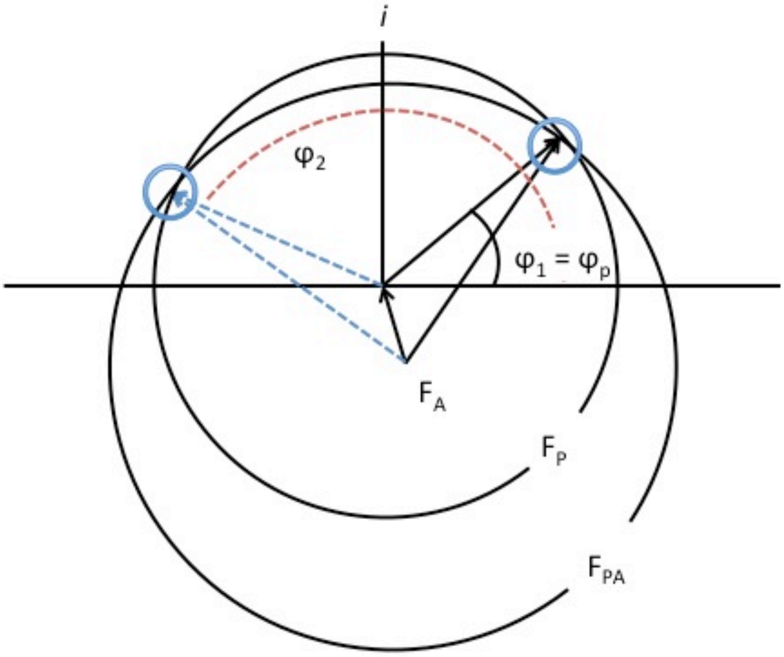


Figure 25. Representation of the phase equation. The complex structure factor for protein, derivative, and heavy atom are related for a generic, centric reflection *hkl*. Determination of the position of the derivative is given by two possible solutions of the protein phase angle for each reflection produced from the entire complex structure factor for the heavy atoms F_A. However one phase angle (φ_2) will be incorrect; the other one (φ_p) will be correct. The figure was adapted from Biomolecular Crystallography - Bernard Rupp

2.13.2 Anomalous scattering

Anomalous scatterers provide anomalous differences within a data set and dispersive differences between data sets of the same crystals recorded at different wavelengths. Besides selenomethionine protein or other heavy atoms introduced by soaking into native crystals, another source of anomalous signal can be natively present atoms in proteins such as zinc, iron, sulphur etc. Anomalous scattering takes advantage of the ability of heavy atoms to absorb X-rays at different wavelengths and promote the breakdown of Friedel's Law.

Friedel's Law states that the intensities associated with two reciprocal points, given by the indices (*h,k,l*) and (*-h,-k,-l*) are equal, unless in the presence of anomalous scattering which will produce a small difference in intensity. Thus the structure factor amplitudes are symmetric about the real axis in the absence of anomalous scattering contributions. However in the presence of anomalous scattering the Friedel pair, instead of having identical amplitudes becomes a Bijvoet pair with different intensities F^+ and F^- for the related reflection. This breaks the centrosymmetry of reciprocal space; with both the magnitude and phase of the reflections in the Bijvoet pair distinctly different from each other and from this disparity the phase information can be extracted.

2.13.3 Dispersive differences

Atomic scattering at different absorption wavelengths can exploit the fact different diffraction intensities can be measured for the same reflection. Giving rise to dispersive differences, which can together with anomalous scattering be used in a multi-wavelength anomalous dispersion (MAD) or (SAD). MAD takes advantage of both orthogonal dispersive and anomalous scattering from the same crystal to resolve the phase ambiguity, where the largest anomalous differences between Bijvoet pairs are measured at the peak wavelengths of the derivative heavy atom. Dispersive differences can also be collected from data sets recorded at the inflection peak, and remote (below the peak). However, the loss of symmetry in the diffraction pattern due to the breakdown in Friedel's Law means that the average redundancy needed will increase, and radiation damage may then come in play for recording three complete datasets at different wavelengths sets from one crystal.

2.13.4 Structure determination

The CCHFV N protein structure was solved by single isomorphous replacement with anomalous scattering (SIRAS) using a mercury derivative obtained by soaking natively grown CCHFV N protein crystals, suspended under the original crystallization conditions, with 5 mM thimerosal for 20 min and back soaking for 10 min to remove any non-specific binding which might cause unwanted absorption of X-rays (Carter et al., 2012a). The derivative data sets were combined with the native data set using CAD (1994), and scaled using SCALEIT (1994), and the Hg substructure positions were determined by generating an isomorphous difference map in Patterson space. Eight prominent Hg peak sites were found and the sites were input into autoSHARP for phase calculation, improvement, and phase extension to 2.1 Å (Vonrhein et al., 2007). Density modification was implemented by PARROT (Cowtan, 2010), after which the automated model building program BUCCANEER (1994, Cowtan, 2006) was used to build 85% of the model. The model was then subjected to iterative cycles of manual building in COOT (Emsley and Cowtan, 2004), and refinement in REFMAC (Murshudov et al., 1997) with TLS parameters applied until 98% of the protein molecules had been modeled. The TLS parameters were determined by the TLSMD server (Murshudov et al., 1997), which defined the globular and arm domains as separate rigid bodies. The ARP/wARP solvent program (Lamzin and Wilson, 1997) was used to initially locate the bulk of water molecules, followed by manual building of the remaining solvent molecules in COOT. The fully refined model has an R_{factor} value of 18.4% and an R_{free} value of 23.4% (Table 3) and was validated using MolProbity (Chen et al., 2010).

2.14 Crystallization and X-ray diffraction theory

X-ray crystallography is one method of analysing macromolecular structure at atomic resolution. Major advances in computing, physics, biology, and chemistry has lead to an exponential growth of structures being solved and deposited into the world's largest repository for macromolecular structures solved using experimental data, the Protein Data Bank (PDB). The PDB holds currently approximately 90,000 models and roughly 79,000 of these are protein and nucleic acid models obtained by X-ray crystallography. The methods and theory involved in macromolecular crystallography have been well documented in a number of textbooks (Crystallography made Crystal Clear – Gale Rhodes and Biomolecular Crystallography - Bernard Rupp). The aim of this section is to provide a basic overview of how researchers use X-ray crystallography to obtain macromolecular structures (Rhodes, 2006, Rupp, 2010).

2.14.1 Crystallization

Owing to the fact that one single molecule is a very weak scatterer of X-rays and a lens is unable to focus X-rays it is not possible to produce a focused image of a single molecule in the way we can see an object with visible light waves. The latter problem can be overcome by using a computer to simulate the lens, whereas to solve the problem of the weak scattering of one molecule and to generate sufficient scattering measurements to feed into the computer, it is necessary to analyse diffraction from crystals, with repeating units, which can be used to amplify the signal, rather than individual molecules.

X-ray crystallographic determination of a protein structure requires diffraction of X-rays from a crystal containing many identical copies in an ordered array of the structure under investigation. The crystals serve to generate an identical repeating unit or lattice in three dimensions, which acts to amplify the strength of the diffracted X-rays. The first main obstacle is to grow crystals that generate sufficient X-ray diffraction for structural determination and this entails growing high-quality crystals from highly pure and mono-disperse protein.

The two common approaches that are widely used for growing reproducible crystals are the hanging drop and sitting drop vapour diffusion methods (McPherson et al., 1982). Both these methods involve mixing soluble pure protein with a precipitant, such as polyethylene glycol or ammonium sulphate, at a concentration of the precipitant that is just below that necessary to precipitate the protein. In principal, precipitation of the protein is achieved by saturating the concentration of the protein with precipitant in a drop by removing water by controlled evaporation. The volume of the drop is made from an amount of soluble protein and an amount of reservoir solution so that the resulting drop has approximately 50% of the precipitant required to cause supersaturation and eventual crystallization. The drop is then suspended underneath a cover slip (hanging drop) or dispensed in a well (sitting drop), which are closed in a container containing a much larger volume of the reservoir solution or 100% of the precipitant to cause supersaturation. Vapour diffusion is induced in this closed system with the net transfer of water from the drop to the reservoir mother liquor, until the precipitant concentration is the same in both solutions (Figure 26).

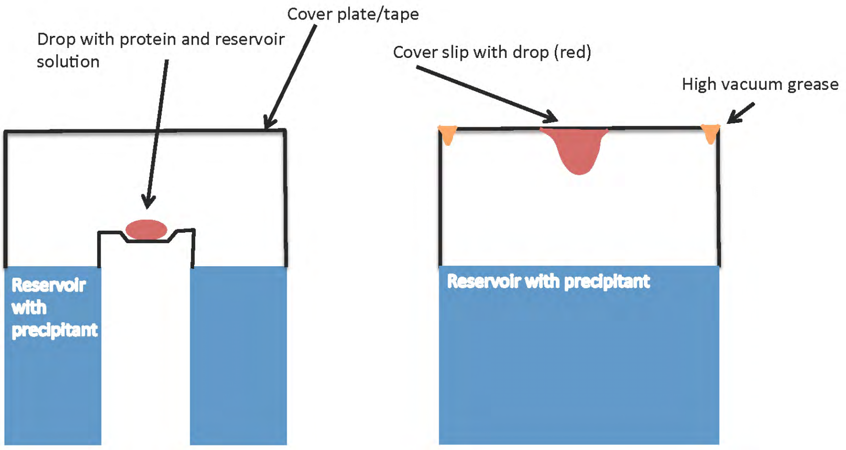


Figure 26. The hanging drop and sitting drop methods of vapour diffusion for crystallization. The figure was adapted from Biomolecular Crystallography - Bernard Rupp

When the combination of protein and precipitation exceeds threshold values the precipitant concentration moves into the nucleation phase, as shown in the phase diagram (Figure 27), and, in theory, precipitation should occur. In some conditions this allows formation of molecular clusters, which act as nucleation sites from which crystals can grow. Slow precipitation is more likely to produce larger crystals, whereas rapid precipitation produces more nucleation sites and many small crystals, or worse (and in most cases) an amorphous solid. As nuclei form the protein concentration in the drop decreases moving the conditions into the metastable growth region (McPherson, 1990).

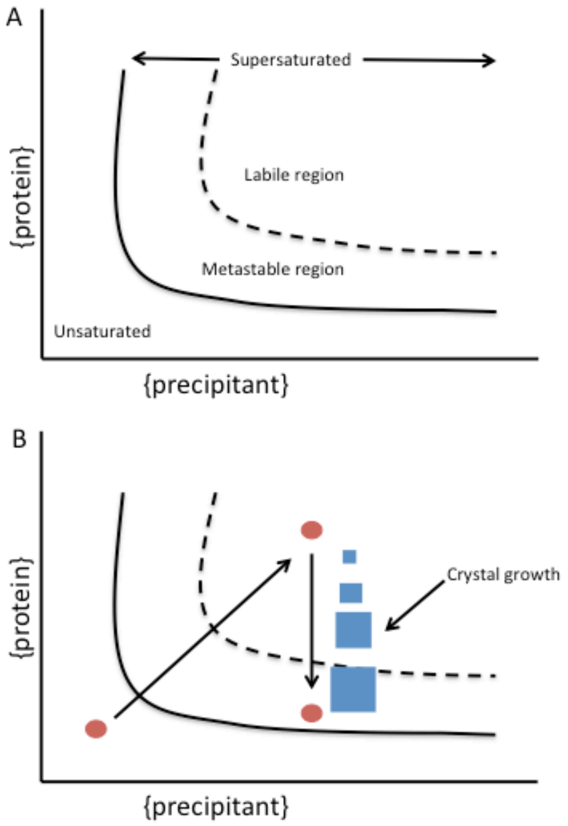


Figure 27. Phase diagram of crystal growth (A) for a hypothetical protein showing the relationship of protein solubility as a function of precipitant concentration. Unsaturated area represents concentrations of protein and precipitant at which the solution is not saturated with protein. Metastable and labile regions are unstable solutions and are supersaturated with protein. Metastable region supports crystal growth only, whilst the labile region supports nucleation and growth. (B) In an ideal crystal growing experiment, once vapour diffusion allows the protein to become supersaturated and reach the region of spontaneous nucleation, nuclei will form, and a crystal may start to grow, depleting the protein solution and allowing the solution into the metastable region for optimal growth. The figure was adapted from Biomolecular Crystallography - Bernard Rupp.

Usually screening for crystallization conditions for structural determination of untested proteins involves a sparse matrix of different solutions using the sitting drop vapour diffusion method (Carter, 1990). This allows hundreds of conditions - different buffers, salts, precipitants and organics - to be analyzed simultaneously for their ability to promote crystal growth in small drops. These types of screens typically produce hits, allowing initial crystallization to be optimized using the hanging drop method, which is amenable to larger drops hanging from a cover slip and typically produce larger crystals with improved diffraction.

2.15 X-ray diffraction theory

2.15.1 Symmetry and the unit cell lattice

In 1912 William Lawrence Bragg discovered that the angle in which diffracted beams emerge from a crystal is related to the spacing of equivalent planes of atoms (Bragg, 1913). This is important because it is the planes within the repeating unit of the crystal that allow the conditions for diffraction to satisfy Bragg's Law (see below). There exists numerous planes within a crystal, and all planes are those defined by the faces of the three-dimensional minimum repeating unit of the crystal, the unit cell. The dimensions of the unit cell are defined by six parameters, the lengths of three unique edges a, b and c; and three unique angles α β and ϒ (Figure 28).

The three-dimensional stacking of the unit cells define the lattice planes in three directions or three axes, in which the **a** vector defines the x axis of the lattice and similarly the y and z axes are formed in the **b** and **c** directions respectively. The lattice planes of the unit cell and all others can be thought of as equally spaced parallel planes, which can cut into the lattice planes and crucially contribute to the conditions of diffraction (discussed below).

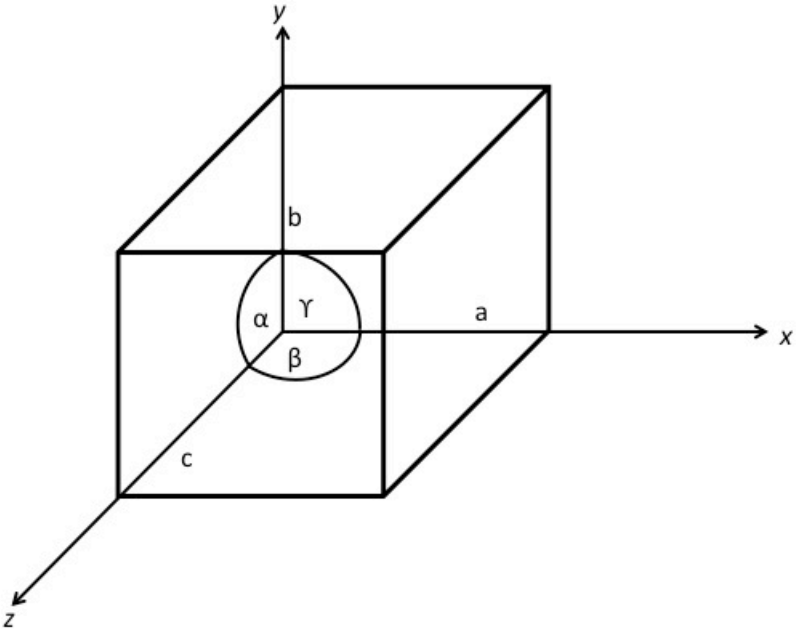


Figure 28. Schematic showing a orthorhombic unit cell, with edges **a**, **b**, **c** and angles α, β, and ϒ. The figure was adapted from Crystallography made Crystal Clear – Gale Rhodes.

These a, b and c axes underlie the three Miller indices (*hkl*), which identify a particular set of equivalent parallel planes within a unit cell and can be thought of as defining how many planes exist and their direction, or alternatively, the number of parts into which the set of planes cut the edge surface of each unit cell. For example a/1, a/2 a/3 or b/1, b/2, b/3 etc. The whole number describes how many times the same equivalent plane cuts at the **a**, **b** or **c** vectors and are therefore characterised as the **a**/*h*, **b**/*k*, **c**/*l* pieces (Figure 29).

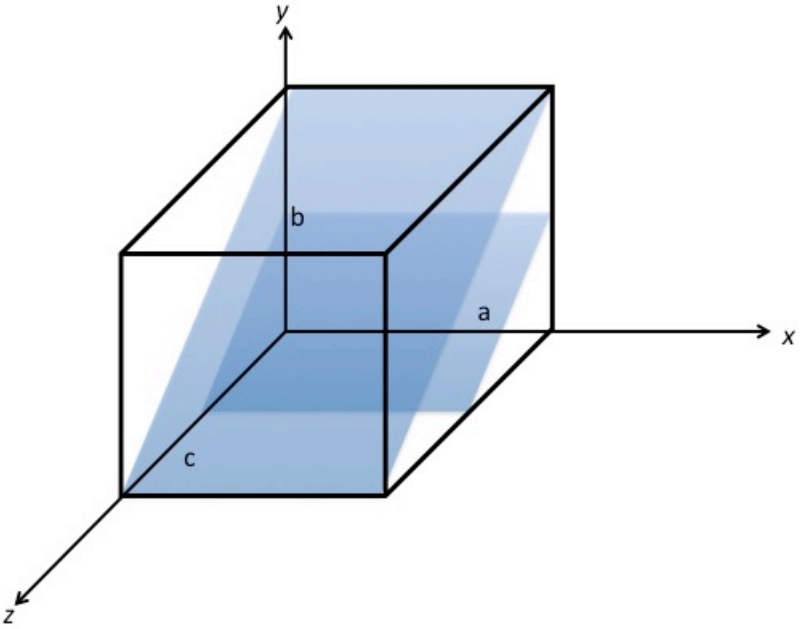


Figure 29. Schematic of an orthorhombic unit cell with the intersection (022) of two planes. a into 0 parts, b into two equivalent parts, and c into two equivalent parts. The figure adapted from Crystallography made Crystal Clear – Gale Rhodes.

The overall symmetry of the unit cell is fundamental to assigning the unit cell dimensions. The proper assignment of the unit cell is that which provides the highest symmetry given by the Bravais lattice, and describes the geometric arrangement of the lattice points, and thereby the translational symmetry of the crystal (Mois et al., 2006). Because there are theoretically several different ways to define the unit cell from one particular lattice, the working unit cell that is most symmetrical for the atomic arrangement of the crystal should be chosen. The simplest cell contains one lattice point and is designated primitive and with the addition of body and face centered cells, there are 13 allowable lattices (Mois et al., 2006) (Figure 30).

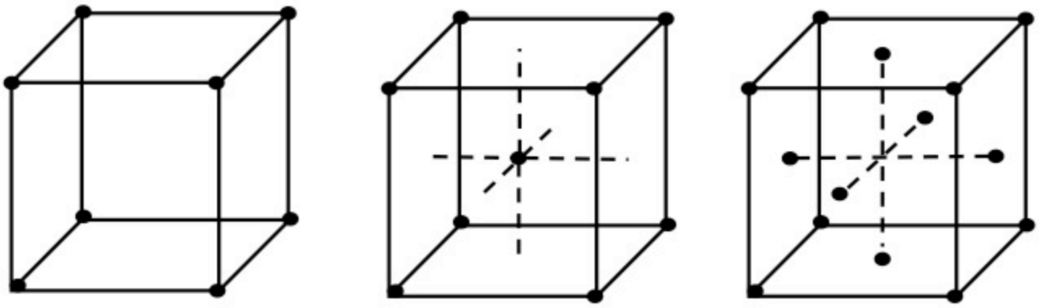
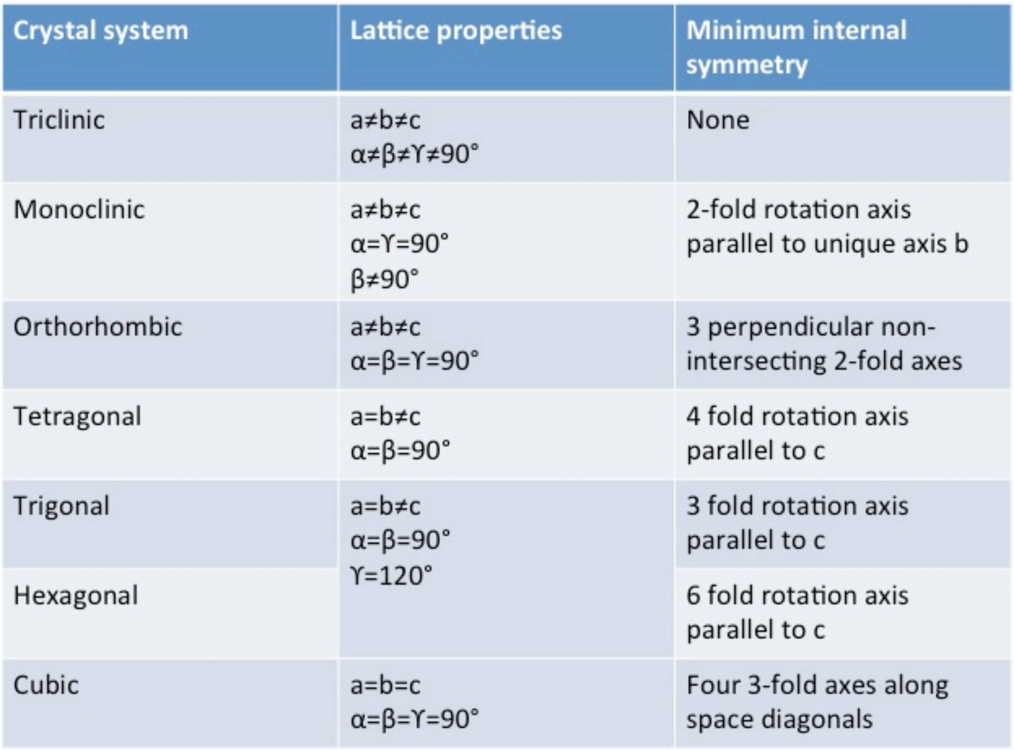


Figure 30. Centered and non-centered cubic unit cells. Cubic primitive (P), body-centered or internal (I), and face centered (F) unit cells. The figure was adapted from Crystallography made Crystal Clear – Gale Rhodes.



Crystal system	Lattice properties	Minimum internal symmetry
Triclinic	a≠b≠c α≠β≠ϒ≠90°	None
Monoclinic	a≠b≠c α=ϒ=90° β≠90°	2-fold rotation axis parallel to unique axis b
Orthorhombic	a≠b≠c α=β=ϒ=90°	3 perpendicular non-intersecting 2-fold axes
Tetragonal	a=b≠c α=β=90°	4 fold rotation axis parallel to c
Trigonal	a=b≠c α=β=90° ϒ=120°	3 fold rotation axis parallel to c
Hexagonal		6 fold rotation axis parallel to c
Cubic	a=b=c α=β=ϒ=90°	Four 3-fold axes along space diagonals

Table 2. The seven crystal systems. The table was adapted from Biomolecular Crystallography - Bernard Rupp

Designation of the space group, which describes both overall and internal symmetry, is achieved by a symbol (for example $P2_1$). This symbol assigns the type of lattice by the capital letter. The first number describes the point group or rotational symmetry operations – without translational contributions - of the crystal system. The subscript number assigns additional screw symmetry operations of the unit cell. The allowable symmetry operations governed by the point group must be compatible with the translational requirements for a specific lattice. The simplest symmetry operations are translation and rotation (only 2, 3, 4 or 6 fold). Combinations of these can produce screw axes and glide planes. Because of the chiral nature of biological molecules symmetry operations such as mirror planes and inversion centres are not allowed. The allowed symmetry operations of a point group within the crystal lattice that can be carried out without changing the appearance of the unit cell dictate which one of the seven systems the crystal relates to (Table 2). The asymmetric unit (AU) is the smallest element within the unit cell that upon application of symmetry operations defined by the space group can generate the entire crystal structure.

2.15.2 Diffraction basics

2.15.2.1 X-ray waves

X-rays are electromagnetic waves that consist of electromagnetic radiation and can be described as a self-propagating transverse oscillating waves of electric and magnetic fields, which are always in phase with each other (Figure 31).

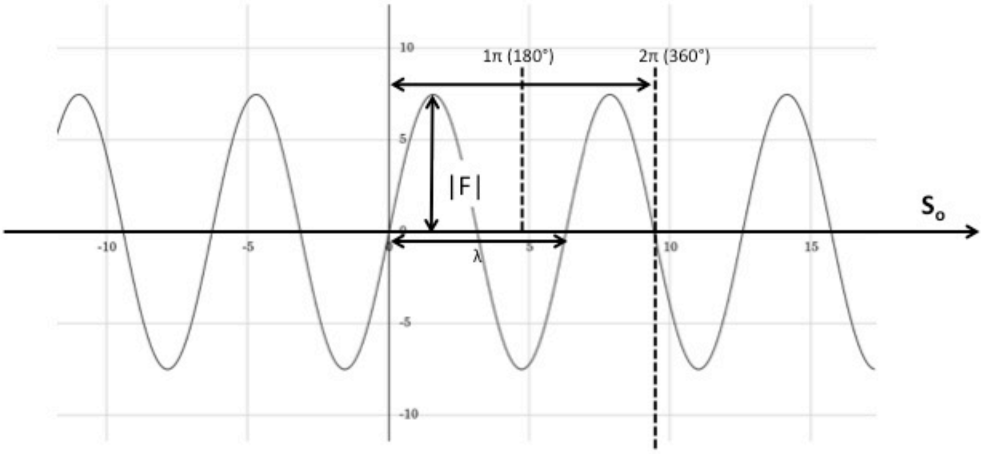


Figure 31. Schematic of an X-ray wave (electric field). The electric field oscillates perpendicular to the propagation vector S_0, and has a magnitude of λ. At any point of the wave the position can be expressed with reference to the periodic origin node in terms of phase shift Δφ ranging from 0 to 2π (in units of radians or rad, 1 rad = $\frac{180}{\pi}$ degrees). The figure was adapted from Biomolecular Crystallography - Bernard Rupp.

The aim of an X-ray diffraction experiment is to produce an X-ray diffraction pattern. The intensity and positions of the reflections (spots) within this pattern are used as the raw data to calculate an electron density map, in which an atomic model can be fitted. The experiment is complete when the model is refined to match the observable data as closely as possible. It is imperative to have an understanding of the way X-ray radiation interacts with the molecules within the crystal at an atomic level in order for this to be carried out.

The way in which the atom scatters the X-rays is important for the resultant X-ray photon to become detectable and produce a reflection (spot). When X-ray radiation collides with an atom the electrons oscillate and emit secondary waves. If the oscillation is at a frequency identical to that of the first electromagnetic wave the result is the emission of radiation of a wavelength and frequency identical to the incident waves. This is called elastic scattering. However if the electron accepts momentum from the incoming photon the result is that the wavelength of the emitted secondary wave will have increased. This is called inelastic scattering. The waves emitted from elastic scattering interfere with each other and give rise to the Bragg reflection, which is detected in the diffraction pattern. Most of the X-rays go straight through the crystal and the incident beam is hardly reduced.

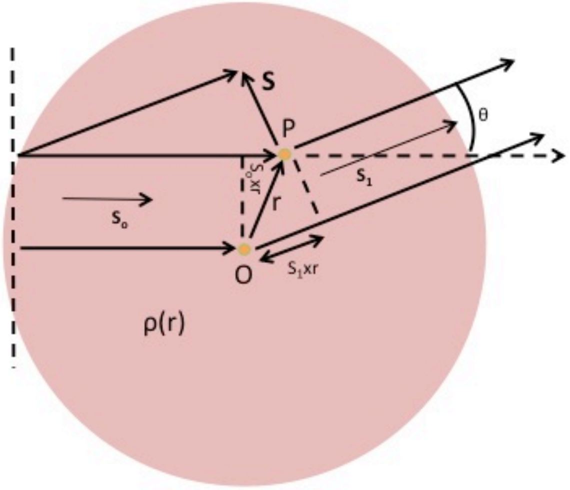


Figure 32. Schematic of X-ray scattering by an atom. The scattering vector of partial waves emanating from a two-electron system $\rho(\mathbf{r})$ of an atom. The red shading shows the spherically symmetric electron density. The two incoming X-ray photons $\mathbf{S_0}$ at wavelength λ, emanating from O and P, will each be scattered by one electron and diffract with the same angle of incidence $\mathbf{S_1}$, but will have different path lengths. The path difference Δp between the lower path through O and the shorter upper path P is given by the difference between scalar vector products $\boldsymbol{r} \cdot \boldsymbol{S_0}$ and $\boldsymbol{r} \cdot \boldsymbol{S_1}$. The figure was adapted from Biomolecular Crystallography - Bernard Rupp

2.15.2.2 Scattering of an atom

When considering a two-electron system with one electron located at the origin O and the other at P, the distance between them is **r** (Figure 32). The scattering of X-rays by an electron is fundamental to the diffraction process. Partial waves, which emanate from the ring of electrons within the atom, have phase differences due to the electric field vector of the photon scattering from electrons in the back of the atom and those in the front. The path length can be calculated by the following equation.

Equation 1.1

$$\Delta p = (\mathbf{S_1} - \mathbf{S_0}) \cdot \mathbf{r}\lambda = \mathbf{S} \cdot \mathbf{r}\lambda$$

The phase difference caused by the path differences of the diffracted waves can be expressed by:

Equation 1.2

$$\Delta\varphi = 2\pi(\mathbf{S_1} - \mathbf{S_0}) \cdot \boldsymbol{r} = 2\pi\mathbf{S} \cdot \mathbf{r}$$

Where

$$\mathbf{S} = (\mathbf{S_1} - \mathbf{S_0})$$

Equation 1.2 introduces the term **S**, which is the scattering vector. The scattering vector is the vector difference between incoming and scattered wave vector (Figure 33 and 34).

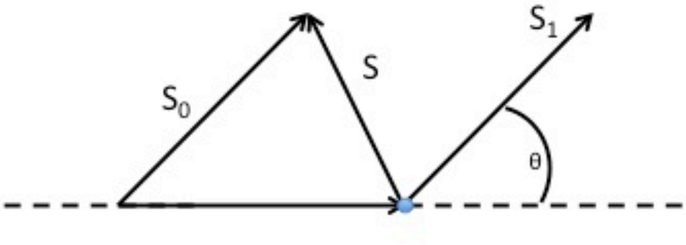


Figure 33. Schematic showing the relationship between $\mathbf{S_1}$, $\mathbf{S_0}$ and $\mathbf{S}$. The figure was adapted from Biomolecular Crystallography - Bernard Rupp.

The atomic scattering factor is a measurement of the scattering angle of the atom and is dependent on the number and position of electrons within the electron cloud. Representation of all partial emanating waves from the atom results in a complete wave integrated over the total volume of the electron cloud. Therefore each partial wave emanating from $\rho(\mathbf{r})$ is an exponential wave vector exp(iφ) with relative phase difference $\varphi=2\pi\mathbf{Sr}$.

Equation 1.3

$$F_S = f_S = \int_{r}^{V\,(atom)} \rho(\,\mathbf{r})\exp(2\pi i\boldsymbol{Sr})\,d\boldsymbol{r}$$

The F_S function is the atomic scattering factor, which describes the scattering of the entire electron cloud of the atom as a function of the scattering direction or angle. As the scattering angle, 2θ, increases there is increasing destructive interference between the X-rays scattered by different parts of the electron cloud and the intensity of the scattered X-rays will decrease with increasing scattering angle. The F_S function is proportional to the square root of the intensity of the scattered X-ray photons recorded within the X-ray diffraction pattern.

From equation 1.3 it can be seen that the Fourier transform (FT) of the electron density $\rho(\mathbf{r})$ of the atom is f_S, so therefore the atomic scattering factor has the form of an integral. This concept is important when considering the superposition of all partial waves from all atoms in a molecule and can be applied to calculate electron density of an entire crystal structure.

2.15.2.3 Bragg's Law of diffraction

It was mentioned earlier that the angle at which diffracted waves emerge from a crystal is related to the spacing of equivalent planes of atoms within the crystal. The Bragg equation expresses X-ray diffraction as reflections of incident X-rays from planes of atoms within the crystals lattice. In figure 34 it can be seen that the incident angle of the approaching wave equals the angle of reflection, and the angle of diffraction from multiple equivalent parallel planes waves is dependent on the distance between these planes. This is known as the Bragg Law of diffraction.

Equation 1.4

$$n\lambda = 2d_{hkl}\sin\theta$$

Where θ=angle incident radiation makes with lattice planes, d = distance between lattice planes, λ = wavelength and n is an integer.

From the equation 1.4 and figure 34 it can be seen that Bragg's Law is only satisfied when n is an integer, in that the additional distance travelled by B between the planes is an integral multiple of the wavelength and therefore wave A and B (Figure 34), will interfere coherently to produce a strong diffracted X-ray beam (Bragg, 1913). For other angles of incidence, which do not satisfy this Law, waves emerging from successive planes are out of phase and will cause destructive interference.

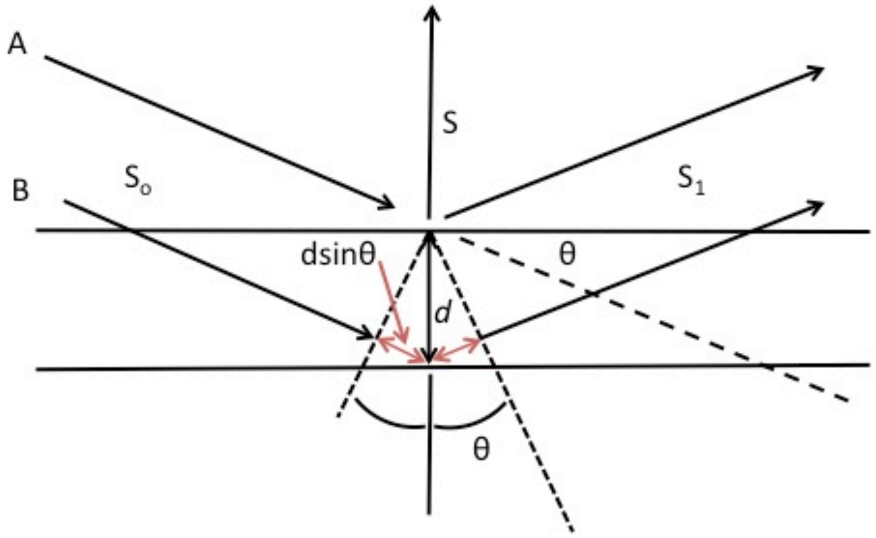


Figure 34. Schematic of Bragg's Law of diffraction. Representation of X-ray diffraction as partial reflections from two separate sets of planes within a crystal allows interpretation of the Bragg equation. The two excited partial waves have total path difference of $2d_{hkl} \cdot \sin\theta$, which must always equal to maximum of n for maximum constructive interference. The figure was adapted from Biomolecular Crystallography - Bernard Rupp.

2.15.2.4 Scattering of an entire molecule

The scattering factor of an entire molecule is the superposition of the partial scattering contribution from each atom *j*. In equation 1.3 $\rho(\mathbf{r})$ represents the electron density at point **r** and the integration is carried out over the entire space **r**. However many atoms at positions **r***j* within a unit cell contribute to the scattering from an entire molecule. The partial wave summation with respect to the origin describes the scattering of the unit cell and is given by the equation:

Equation 1.5

$$\mathbf{F_S} = \sum_{j=1}^{atoms} f_{sj}^{o} \cdot \exp\,(2\pi i \boldsymbol{S}\boldsymbol{r_j})$$

Where $\mathbf{F_s}$ is the scattering function or structure factor of an entire molecule.

The term f_{sj}^{o} is the scattering factor of atom j. The atom j is still treated as a simple sphere of an electron density. The scattering function of a molecule will eventually repeat in certain directions of the crystal. In the direction **a**, the phase difference between unit cells is given by 2π**Sa**. The scattering function in direction **a** is:

Equation 1.6

$$\mathbf{F}_{\mathbf{s}}^{\mathbf{a}} = \sum_{u=0}^{u-1} \mathbf{F}_{\mathrm{s}}^{cell} \cdot \exp\,(2\pi i(u\mathbf{Sa})$$

The scattering function of an entire crystal will repeat in certain directions along the three lattice directions due to translational equivalents, u in **a**, v in **b**, and w in **c**. For instance the 002 planes along the **c** axis will be repeated in the adjacent unit cell. Where these repeats are exact integers along the unit cell vectors and have the same phase, where **S** = integers u, v, w along the respective axis the vectors $\mathbf{F}_{\mathbf{s}}$ will add up in exactly the same direction. These are known as the Laue conditions. The total scattering is therefore amplified proportionally with increasing number of repeating units and is expressed by three sumations:

Equation 1.7

$$\mathbf{F}_{\mathbf{s}}^{cryst} = \mathbf{F}_{\mathbf{s}}^{cell} \cdot \sum_{u=0}^{u-1} \mathbf{F}_{\mathbf{s}}^{cell} \cdot \exp\,(2\pi i(u\mathbf{Sa}) \cdot \sum_{u=0}^{u-1} \mathbf{F}_{\mathbf{s}}^{cell} \cdot \exp\,(2\pi i(v\mathbf{Sb}) \cdot \sum_{u=0}^{u-1} \mathbf{F}_{\mathbf{s}}^{cell} \cdot \exp\,(2\pi i(w\mathbf{Sc})$$

Laue conditions are created when $\mathbf{a} \cdot \boldsymbol{S}$ is equal to an integer h, $\mathbf{b} \cdot \boldsymbol{S}$ an integer k and $\mathbf{c} \cdot \boldsymbol{S}$ an integer l and no scattering is observed from a crystal unless these conditions are obeyed. These diffraction conditions form the basis of the Bragg equation and can be written to derive the following relationship equation:

Equation 1.8

$$\boldsymbol{d}^* = \frac{1}{d_{hkl}}$$

Where **d*** is equal to the reciprocal lattice vector that is perpendicular to the lattice plane in real space, and defined in length as $\frac{1}{d}$ (the reciprocal of the inter planer distance vector d_{hkl}). The length of this vector is designated as **a*** **b*** **c*** for the respective planes. The reciprocal lattice is a second theoretical lattice, which consists of axes, which are collinear with the real space axes of a unit cell and are designated **a* b* and c***, so **a*** = $\frac{1}{a}$.

2.15.2.5 Ewald sphere

Together with the construction of the Ewald sphere the reciprocal lattice is a useful concept and permits understanding of how certain reflections from sets of planes satisfy Bragg's Law and produce a reflection, which can be measured by a detector. The Ewald sphere is created with radius $\frac{1}{\lambda}$ where λ is wavelength of the incident X-ray beam $\mathbf{S_0}$ passing through crystal C at its centre towards O, the origin of the reciprocal lattice at its surface (Figure 35).

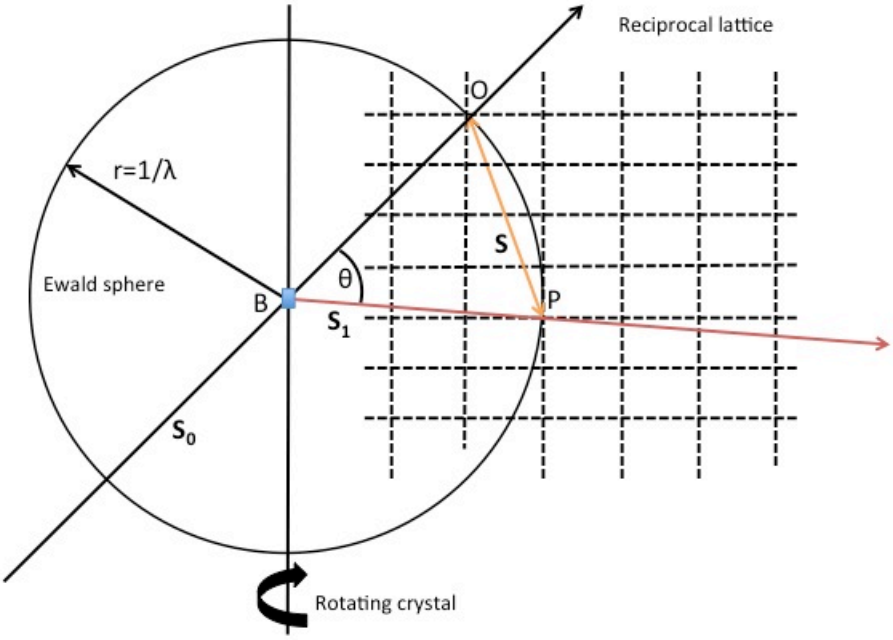


Figure 35. Schematic of Bragg diffraction in reciprocal space. $\mathbf{S_1}$ emerges from the crystal when lattice point P intersects the Ewald sphere. More intersects come into contact with the sphere as the crystal is rotated. The figure was adapted from Biomolecular Crystallography - Bernard Rupp.

By rotating the crystal about B, reciprocal lattice points d_{hkl} with indices *hkl* are brought into contact with the circle of radius $\frac{1}{\lambda}$. When this happens the lines OP and BP can be drawn. The angle of BPO is θ.

The direction of the diffracted X-ray beam is described by the line BP. The line represents a plane in the set that reflect the X-ray beam and lines parallel to it are separated an integral multiple of d_{hkl}. If the wave vector **S** drawn is perpendicular to this plane and its endpoint is touching the surface on the sphere (at point P for example) then BP represents the scattered beam $\mathbf{S_1}$. The direction of the beam BP is given by the diverging beam at angle 2θ and precisely takes it through point P.

With the purpose of collecting a full set of reflections from the diffracting crystal for structural determination, it is necessary to bring all lattice points of the reciprocal lattice into contact with the Ewald sphere. This is achieved by rotating the crystal so that all lattice points satisfy diffracting conditions on the surface of the Ewald sphere.

2.15.2.6 Calculating electron density

From a crystal irradiated with X-rays in a given direction h the scattering function is given by a complex structure factor Fh. The structure factor amplitudes are directly obtainable from measured reflections, as the amplitude of F_{hkl} is proportional to the square root of the reflection intensity I_{hkl}. Fourier integration of the scattering from all electrons within the unit cell is given by the following modification eqauation 1.5:

Equation 1.9

$$\mathbf{F}_{(\mathbf{s})} = \int_{r}^{V\,(cell)} \rho(\mathbf{r}) \cdot \exp(2\pi i \mathbf{S}\mathbf{r})\, dr$$

The fractional coordinates x, y, z of the unit cells and V the volume can be expressed by dv. Where $\rho(\mathbf{r})$ is the electron density at point $\mathbf{r}$ within the unit cell.

The direction of scattering is added to this equation with the reciprocal and direct space positions.

Equation 2.0

$$\mathbf{S} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$$

Equation 2.1

At position

$$\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$$

Equation 2.2

Giving

$$\mathbf{Sr} = \boldsymbol{hx} = \boldsymbol{ky} = \boldsymbol{lz}$$

The complex structure factor $\mathbf{F}(h)$, is the Fourier transform which can be computed from the contents of the crystallographic unit cells by summation of scattered partial waves emanating from all atoms in the unit cell, where ρ(**x, y, z**) is the position of the atoms real space electron density at point **x, y, z**. This is then integrated in real space over the entire unit cell (in fractional coordinates) and allows the calculation of the diffraction pattern using equation 2.4.

Equation 2.3

$$\mathbf{F}(h) = V \int_{x=0}^{1} \int_{y=0}^{1} \int_{z=0}^{1} \rho(xyz) \cdot \exp(2\pi i \mathbf{hx})\, dxdydz$$

With a back transformation reciprocal R' to real space R: We obtain.

Equation 2.4

$$p(x,y,z) = \frac{1}{V} \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} \mathbf{F}(hkl) \cdot \exp(-2\,\pi i(\boldsymbol{hx} + \boldsymbol{ky} + \boldsymbol{lz})$$

Normalization of the volume $\frac{1}{V}$ preceding the discrete Fourier summation provides correct units of $\frac{e^-}{\text{Å}^3}$ for the electron density.

By separating the complex structure factor $\mathbf{F}(h)$ into an explicit amplitude term F_h and phase angle α_h term using the Euler relation.

Equation 2.5

$$\mathbf{F}(h) = F_h \cdot \exp\,(i\alpha_h)$$

And

Equation 2.6

$$\rho(x) = \frac{1}{V} \sum_{h=-\infty}^{+\infty} \boldsymbol{F}_h \cdot\ \exp\,(\,i\alpha_h) \cdot \exp\,(-2\pi i(\boldsymbol{hx}))$$

The Fourier coefficients $\boldsymbol{F}_h$ given in this final form summation for the electron density include both the structure factor amplitude and its associated phase. However, these originate from different sources during X-ray crystallographic work. $\mathbf{F}_h$ is proportional to the intensity of the diffracted X-rays and can thus be quantified from the X-ray diffraction pattern. During an X-ray diffraction experiment only the intensities of the diffracted waves are measured and the phases are lost. In order to compute ρ(*x,y,z*) fully the phase information (α_{hkl}) must be recovered by different means

2.15.3 Data collection

When preparing an X-ray diffraction experiment, important consideration about crystal preparation, exposure of the crystal to X-radiation and data collection, need to be carried out to ensure the amount and quality of data collected is sufficient for structure determination.

2.15.3.1 Crystal mounting

Successful data collection begins with proper harvesting, cryoprotection and mounting of the protein crystals. These steps are implemented to prevent mechanical destruction of the crystal lattice, and minimize radiation damage caused by ionising X-rays.

Capillary mounting and cryo-mounting using a loop are two harvesting techniques employed to reduce these negative affects. Capillary mounting is the most common method for harvesting crystals for collecting diffraction data at room temperature. The capillary is used to suck up a crystal together with a droplet of mother liquor leaving just enough solvent around the surface of the crystal to keep the crystal hydrated and provide some surface tension to adhere the crystal to the inner wall of the capillary. The benefits of capillary mounting are important if problems occur during cryo-cooling, such as damage to the crystal lattice during freezing, which can lead to an unacceptable increase in anisotropy, mosaicity and poor diffraction.

Exposing crystals to X-rays at room temperature can rapidly result in primary and secondary radiation damage, leading to the formation of free radicals from both the macromolecules within the crystal and the solvent. The integrity of the repeating unit of the crystal is compromised by free radicals directly altering the protein, or altering the surrounding solvent, changing the size of the unit cell, which rapidly destroys the quality of the diffraction.

Preserving crystals at cryogenic temperatures throughout the data collection process reduces the diffusion of free radicals through the crystal caused by collecting data at room temperature, particularly at the high intensity synchrotron sources and can dramatically increase the amount and quality of the data obtainable. This means that a full dataset can be collected from a single crystal.

The routine method for harvesting crystals is by flash cooling crystals harvested by a nylon loop in liquid nitrogen at 100K. In this method the crystal is harvested within a nylon loop and is directly quenched in liquid nitrogen if there is sufficient cryo-protectant already in the mother liquor. Alternatively the loop containing the crystal is transferred in the correct percentage of a cryoprotectant and subsequently flash frozen in liquid nitrogen. The antifreeze properties of the cryo-protectant prevent the mother liquor water from crystallizing around the surrounding area of the crystal, which invariably destroys the diffraction quality. The most common cryo-protectants are glycerol, ethylene glycol, MPD, sucrose, and low molecular weight PEGs.

3.15.3.2 X-ray sources

The reliable production of high intensity monochromatic X-rays is a prerequisite for any X-ray diffraction experiment. Generation of X-ray radiation is commonly achieved using two modes: laboratory rotating anode generators and the synchrotron. Rotating anode sources produce X-rays through electron bombardment of a metal target (most commonly copper). When a high-energy electron collides with the target metal it displaces a core electron and a core hole is generated. Subsequently a higher orbital electron drops in to replace and it emits the excess energy as an X-ray photon with a frequency characteristic of the metal target.

Sources using metal targets such as copper or molybdenum will produce two emissions of distinct wavelengths. This is not suitable for X-ray crystallography because this would result in two Ewald spheres at $1/\lambda$ and two amalgamated sets of reflections. Monochromatic X-rays (single-wavelength) are produced by filtering out the secondary Kβ radiation. For copper-rotating anode generators this is achieved by using nickel filters, which absorb strongly at the wavelength for the CuKβ emission, and allow transmission of CuKα.

Synchrotron sources produce substantially higher intensity monochromatic X-rays. The giant ring of the synchrotrons allows electrons or positrons to circle in an orbit with the effect of radiation loss. Electrons are produced in an electron gun and are pre-accelerated in a linear accelerator, by high radio frequency electrical fields. The linear accelerator then injects these electrons into the storage ring in bunches, and they are further accelerated to nearly the speed of light. The electrons travel in a closed path in an ultra high vacuum and are maintained in circular motion by bending magnets and kept stable by multi-pole steering electromagnets. The emission of X-radiation occurs through deflection of accelerated electrons into a curved motion by the bending magnets. Even more powerful devices called undulators or wigglers use high alternating magnetic fields to wiggle electrons resulting in additional bending of the beam, which produces very intense and focused radiation.

Monochromatic X-rays are produced by focusing the polychromatic radiation (by a set of mirrors) onto a silicon crystal monochromator, which disperses the X-rays by diffraction angle according to Braggs Law. Only the desired wavelength is selected through narrow slits into a collimator. The collimator also allows only parallel X-rays with low divergence to exit, directed towards the sample crystal.

The high flux of X-ray photons and collimation of the beam produced from synchrotron radiation allows collection of weak reflections, which would otherwise be too weak to be collected using a laboratory X-ray source, and dramatically increases speed in which the data is collected. In addition the optimal tuning of the beam size allows collection that optimizes the signal to noise ratio for micron or sub-micron sized crystals and can reduce the consequences of radiation damage.

2.15.3.3 Data processing

Following data collection the images are processed to produce the *hkl* Miller indices and intensity of each reflection. The two programs used to process the reflections in the diffraction patterns produced in this project were MOSFLM (Leslie, 2006) and XDS (Kabsch, 2010). Parameters such as radiation wavelength, crystal to detector distance, oscillation angle and the position of the beam must be accurately recorded during data collection and input into the program. Indexing assigns reciprocal *hkl* miller indices to each reflection and also determines the space group, unit cell parameters and estimates the mosaicity of the crystal. Typically MOSFLM (Leslie, 2006) uses two to four images all offset by 90° to detect full reflections which are marked with a red cross, whilst partial reflections are yellow. The Fourier indexing algorithm lists the possible lattice types ranked by a penalty score. Each lattice produces coloured boxes, which mark the spot predictions according to the initial indexing. The correct lattice is chosen as the one with the highest symmetry and with spots falling into predicted spot positions. The mosiacity of the crystal is also adjusted to optimise the fit of the predicted and observed pattern. The mosiacity is a measure of the alignment of the unit cells within the crystal. A larger mosiacity value indicate a large distribution of mis-orientation angles of the unit cells in the crystal and has the effect of broadening diffraction spots between slices as the crystal is rotated. Therefore, diffraction spots from crystals displaying high mosiacity will be observed in the diffraction condition longer than ones with low mosiacity. Over a wide rotation range each diffraction spot will be more likely to overlap with other diffraction spots and in such a case fine slicing may improve processing. Another additional problem is mosaicity of a crystal can also smear spots radially, that is along a line drawn from the spot to the center of the diffraction image.

In the next stage MOSFLM refines the cell constants and experimental parameters, using around 20-25 images. This then leads to the integration of all partial and full indexed intensities from all images with the refined unit cell parameters and correct mosiacity. The integrated intensities are then merged to produce a list of averaged intensities using the program SCALA (Evans, 2006a), which also provides a number of useful statistics and analysis. Because of a variety of factors relating to detector, crystal and diffracted beam, and incident beam the volume intensities from each image in a data set are are not all on the same scale. These factors can be grouped together into a single correction factor dependent scale, or "BATCH SCALE". Scala scales together multiple symmetry related observations of reflections, and merges multiple observations into an average intensity. The appropriate scales are applied when merging symmetry equivalent reflections, so that outliers can be rejected, for example those symmetry related reflections whose intensities are too different from the others. Together with the error associated with all compared measurements, and the agreement between multiple measurements of the same reflection SCALA can correctly measure the partiality of each reflection and convert this to an estimate of the reflections full intensity. All reflection intensities are then merged to produce an averaged intensity I*hkl* and the quality of the data is analysed with various statistics.

2.15.3.4 Completeness

SCALA expresses the number of unique crystallographic reflections measured in a data set as the percentage of the total number of possible reflections present at a given resolution. The completeness is therefore the ratio between the total data theoretically possible in the AU of the reciprocal space and the number of observed unique reflections.

2.15.3.5 R-merge

This is a measure of agreement among multiple measurements of the same reflection, with different measurements being in different frames of data or different data sets. R-merge being the traditional measure of agreement between observed and calculated data is poorly suited for determining the high resolution limit and current standard protocols discard much useful information.

Equation 2.7

$$R_{merge} = \frac{\sum_{hkl} \sum_{i=1}^{n} |I_I(hkl) - \bar{I}(hkl)|}{\sum_{hkl} \sum_{i=1}^{n} I_i\ (hkl)}$$

However the data quality R-values are not comparable to the R values from model refinement and recently a statistic that estimates the correlation of an observed data set with the underlying (not measurable) true signal; CC*, has been introduced to provide a single statistically valid guide for deciding which data are useful (Karplus and Diederichs, 2012).

2.15.3.6 Multiplicity

Overlapping symmetry related and equivalent reflections contribute to redundancy in a crystallographic data set. Multiplicity therefore calculates average number of independent measurements of each reflection in a crystallographic data. The symmetry of the crystal results in the presence of equivalent reflections in different regions of reciprocal space, and also overlapping regions of the three-dimensional diffraction pattern, resulting in multiple measurements of the same reflection.

2.15.4 Non crystallographic symmetry (NCS)

Local symmetry is present when more than one identical object is present in the AU. Identical molecules exhibiting symmetry relationships, other than those connected with crystal symmetry do not display additive diffractive contributions. By knowing and applying the non-crystallographic relationship of these molecules in the unit cell the experimentally phased electron density maps can be used to add the signals together and improve the signal to noise ratio and map quality within the AU. Non-crystallographic symmetry restraints can be applied throughout the refinement.

2.15.5 Molecular Replacement

Molecular replacement (MR) is a way of calculating the phases of the protein structure under study using a model of a known related structure. Typically the known structure needs to have 30% or higher sequence identity. In MR the structure is rotated and translated in the unit cell or AU so that the model molecule is in the same orientation and position in the unit cell as the unknown molecule.

The phases are then calculated from the positioned model and combined with the measured Fobs obtained from the diffraction intensities measured from the crystal. Six-dimensional molecular replacement searches or the so called brute force methods, break the orientation of the model into three angles of rotation about the orthogonal axes x, y and z, with their origins at the centre of the model and search all possible orientations along each translational grid point. However searching with three rotational and three translational degrees of freedom of an entire macromolecule with the present computing power is impractical because the sheer numbers of trials are too great.

A more computationally efficient method takes apart the 6-dimensional searches into a more efficient 3-dimensional rotation search, followed by a 3-dimensional translational search once the correct rotation has been solved. The rotational search is allowed because it is possible to search for the best orientation independently of location using the Patterson function. The Patterson function is essentially the Fourier transform of the intensities rather than the structure factors and describes both a set of intra and intermolecular vectors, which are unique for each protein, however related proteins have similar vector distributions.

By rotating the Paterson map of the known structure around the origin in the unit cell using three-dimensional rotational searches the intramolecular distance vectors can be compared and depend only on the orientation of the molecule, not its translational position. The rotation function can therefore be used to determine the highest correlation between the intramolecular Patterson vectors for various model orientations and compare them with the Patterson map of the desired protein.

Once the orientation of the molecule has been found it can then be translated to search for the best position, by correlating the observed intensities and the Patterson map intermolecular vectors using the symmetry-related molecules of the known structure as it is moved within the cell.

It should be noted that because phases contain more information than intensities and the electron density reconstruction is generated entirely from the phases from the replaced search model, the electron density map is highly susceptible to model bias. There are various Fourier syntheses available to generate maps to minimize phase bias and improve the model; these are discussed below.

2.15.6 Model building and refinement

Once initial phases have been obtained an electron density map can be created, and the process of building a model into the map can be undertaken. This is not too much of a problem with accurate phases and high resolution, however at lower resolution model building becomes more challenging and it is important when building a model into poor maps to avoid introducing into the electron density any model bias.

If the phases are good enough to create an interpretable mapi t is possible to begin to fit amino acid residues manually using model building programmes such as COOT (Emsley and Cowtan, 2004) or by using automated model building programmes such as BUCCANEER (Cowtan, 2006). This entails many cycles of adjusting the atomic coordinates to improve the agreement between amplitudes from the current model and the original measured amplitudes in the native data set.

Global reciprocal space restrained refinement is carried out throughout the model building process of the model and aims to improve the agreement of the observed and calculated structure factor amplitudes (Fobs – Fcalc). After the experimental phases have been used to generate the initial map, subsequent phases are calculated from the refined model itself, using the reverse Fourier transform. Depending on the quality of the initial map these phases should be more accurate and more density should appear, improving the quality of the electron density maps for the next round of building.

During refinement, restraints are applied to the movement of atoms in the molecule based on local stereochemical and conformational parameters, and B factors. Overall parametres are also applied, such as scale factor and overall B factors, bulk solvent corrections, and anisotropy corrections. These are all refined against all experimental data to produce the best agreement between the observed structure factor amplitudes and model structure factor amplitudes.

To overcome incorrect phases dominating the electron density reconstruction 2Fobs-Fcalc maps are created to let the intensity data influence improvement of the model. This helps to overcome any model bias that may be generated from the use of simple Fobs maps. 2Fobs-Fcalc maps highlight regions of the model that are inconsistent with the observed structure factors amplitudes.

The more widely used *Sigma A* figure of merit weighted map (2mFo-DFc) is the maximum likelihood estimate of the full model map (i.e. the best map possible) or its simple modification of 2Fo-Fc map where Fo/Fc are replaced by their maximum likelihood estimates for each model phase (m), and of atomic coordinate errors in the current model (D), which is the most accurate way to reveal the ambiguities which are in conflict with the current model.

Difference maps are calculated by subtracting the calculated structure factor amplitudes (|Fcalc|) from some multiple of the observed amplitudes (|Fobs|) within each Fourier term. The difference between each term is highlighted in the difference electron density maps (Fobs-Fcalc), depending on whether Fobs or Fcalc is larger. Resulting in either negative or positive density. Positive density corresponds to regions where more electron density should be present than implied by the model and negative density implies the opposite.

This allows changes to be made to the model in an iterative manner, aiming each time to improve the agreement between the Fobs and Fcalc. The quality of the model is assessed after each round of refinement. This is based on a pair of statistics known as Rfactor and Rfree, which are calculated by the following equation.

Equation 2.9

$$R = \frac{\sum_{hkl} |F_{obs}(hkl) - F_{calc}(hkl)|}{\sum_{hkl} F_{obs}(hkl)}$$

R-free and Rfactor compare the overall agreement between the amplitudes of the model and native data. The agreement between observed and calculated data improves as the accuracy of the model increases, and therefore the R factor decreases.

Because the Rfactor statistic can be made arbitrarily low for models having sufficient parameters to overfit the data, Brunger introduced Rfree (Brunger, 1992). The Rfree is calculated in the same manner as the Rfactor, but is based only on a small set of randomly chosen intensities; usually 5% are set aside and not used during refinement. The Rfree measures how well the model can predict this subset of the measured reflection intensities and if the model is genuinely correct the Rfree should agree with the Rfactor, which has used the remaining 95% of the data. The final values for the R factors are dependent on the resolution of the data and a desirable target R-factor for a protein model refined with data to 2.5 Å is around 0.2. Once the Rfactor and Rfree cannot be improved any further by alteration and refinement, the model must be validated before being finalised. MOLPROBITY (Davis et al., 2004) is just one server that can carry out validation on a structure. This service highlights bad contacts and unrealistic geometry in both side chains and main chains.

3. Chapter 3 - Structural determination and functional analysis of the CCHFV N protein

3.1 Introduction

- CCHFV is a tick-borne zoonotic virus responsible for serious human disease characterized by haemorrhagic manifestations and multiple organ failure, and is associated with a fatality rate of up to 50% (1, 2).

- CCHFV is a member of the *Bunyaviridae* family, and together with members of the *Arenaviridae* and *Orthomyxoviridae* families these viruses are known as segmented negative sense RNA viruses (sNSV) by virtue of their multistranded genomes.

sNSVs genomes do not exist as naked RNAs, but instead are encapsidated by the viral N protein to form RNP complexes. RNPs associate with their cognate RdRp to form active templates for viral RNA synthesis, resulting in generation of encapsidated replication products and unencapsidated mRNAs. RNP formation is thus essential for virus multiplication, and therefore represents a potential therapeutic target.

The objective of this project was to express and purify homogeneous CCHFV N protein for X-ray crystallography analysis. This was firstly achieved by finding a suitable recombinant expression tag and *E. coli* strain in which to produce abundant and soluble CCHFV N protein. Next Ni-Superflow Plus resin using His_6 trap methodology was used to purify the CCHFV N recombinant protein away from *E. coli* contaminants. SEC was employed to look at the oligomeric state of the native CCHFV N protein, and purify the CCHFV N protein into a homogeneous state suitable for crystallization trials. Crucially for solving the atomic structure of the CCHFV N protein, this project required growing reproducible CCHFV N native protein crystals, which diffracted to high resolution, and a means to obtain phases.

The aim of this project was to implement X-ray crystallography to understand better the structural and functional roles of the CCHFV N protein, such as RNA binding and oligomerization. This will then aid in the design of experiments and functional assays to support these findings and crucially elucidate important targets for rational drug design to inhibit important process crucial to the virus life cycle, offering therapeutic interventions to the treat the disease (Wade, 1997).

3.2 Cloning

3.2.1 Constructing the His$_6$-SUMO CCHFV N ORF expression plasmid

The project's first key objective was to create a plasmid construct, designed to express the CCHFV N protein fused to both small ubiquitin-related modifier (SUMO) (type3) and His$_6$ tag via the N-terminus (Figure 36). The SUMO tag was used to overcome any expression and solubility issues, which are common when expressing recombinant mammalian protein in bacterial cells, whilst the His$_6$ tag was implemented for affinity purification purposes, before eventual cleavage of the His$_6$-SUMO moiety using the SUMO protease to yield the near native CCHFV N protein (Figure 36).

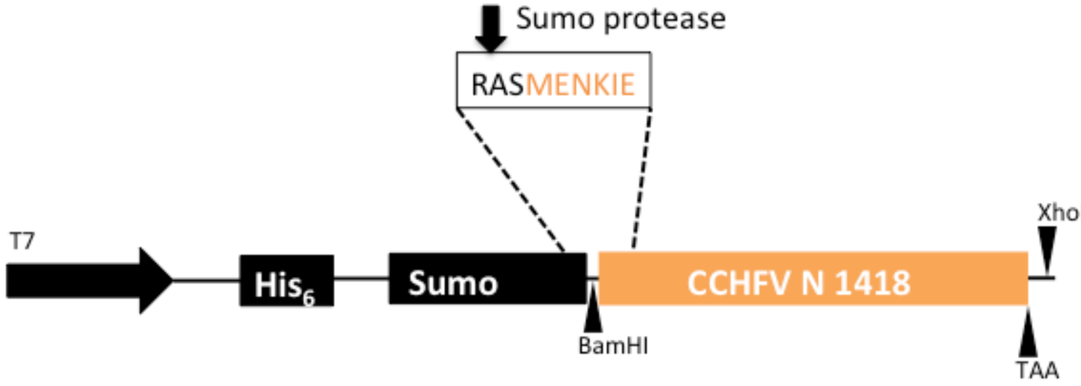


Figure 36. Schematic of the pET28a-SUMO CCHFV N expression construct. Designed to express CCHFV N protein N-terminally fused to His$_6$ and SUMO, and containing a SUMO protease cleavage site. SUMO protease cleavage site is shown boxed.

3.2.2 Cloning of pet28a-SUMO plasmid

Initial cloning aimed to create a construct suitable for expression of His$_6$-SUMO CCHFV N protein in *E. coli*. This involved using PCR to amplify the CCHFV N gene from an existing pET28a clone containing the CCHFV cDNA ORF, whilst flanking the 5' and 3' ends with *BamHI* and *XhoI* sites respectively. These sites would then be utilized to ligate the amplified gene into the multiple cloning site of the expression plasmid pET28a-SUMO, using compatible ends. The position of the 5' end *BamHI* site enables the CCHFV N protein to be N-terminally flanked by the SUMO moiety and His$_6$ tag when expressed and after purification allows subsequent cleavage of this moiety using SUMO protease, whilst leaving only a serine on the N-terminal of the CCHFV N protein.

PCR amplification of the CCHFV N cDNA gene fragment using forward and reverse primers (appendix) was confirmed by analyzing the reaction sample using 1% agarose gel electrophoresis, which revealed a band of approximately 1500 bp, which corresponded to the expected size of the CCHFV N cDNA gene (1497 bp) (Figure 37A). Before *BamHI* and *XhoI* restriction enzymes were added to produce the respective sticky ends flanking the CCHFV N cDNA required for ligation the PCR was concentrated using a clean up column (Zymo Research).

To introduce compatible ends into the pET28-SUMO plasmid single digestion reactions using *BamHI* and *XhoI* enzyme respectively were implemented to confirm each enzyme was working sufficiently (Figure 37B). The single digestion of the pET28a-SUMO plasmid was analyzed using agarose gel electrophoresis and the activity of each enzyme was confirmed by the change in migration from supercoiled to linear species of pET28a-SUMO cDNA (Figure 37B). The pET28a-SUMO plasmid and PCR fragment were subsequently double digested with both *BamHI* and *XhoI* digestion enzymes and the band corresponding to the cleaved plasmid and PCR fragment were respectively gel purified to be finally cleaned up in a gel extraction kit. Ligation of the double digested PCR fragment containing the CCHFV N gene in the pET28-SUMO was successful and DNA sequencing confirmed the un-mutated CCHFV N ORF was cloned into the correct position within the pET28a-SUMO vector.

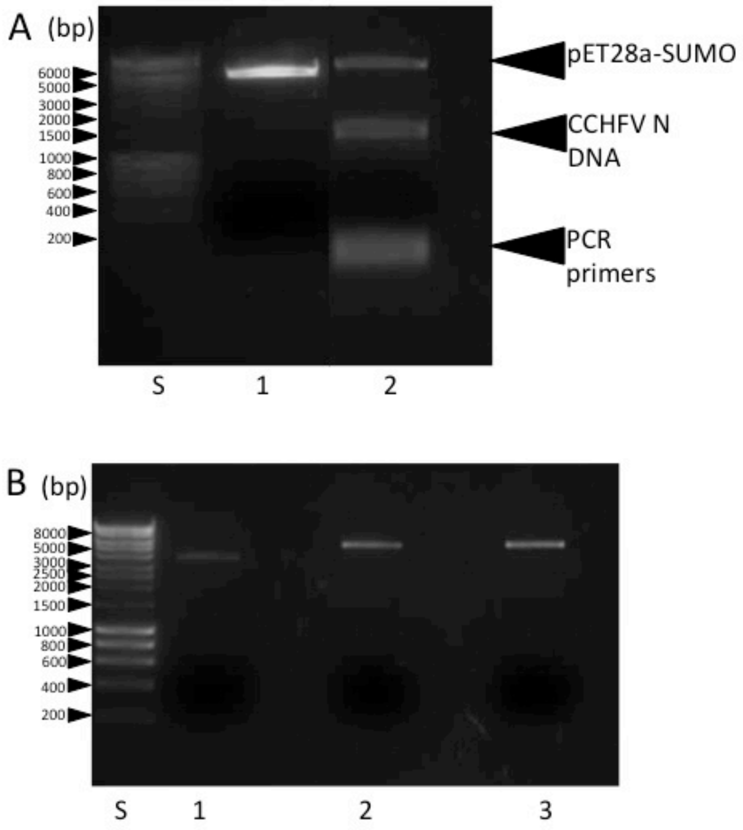


Figure 37. PCR amplification of the CCHFV N cDNA from pET28a and ligation in the multiple cloning site of pET28a-SUMO. 1% agarose gel electrophoresis analysis of single and double restriction digest of the pET28-SUMO plasmid using *XhoI* and *BamHI* enzymes. (A) DNA markers (Bioline) (Lane S), Double digested pET28a-SUMO (Lane 1), PCR amplification of the CCHFV N cDNA (Lane 2). Single digestions were carried out to test the activity of each enzyme. (B) DNA markers (Bioline) (Lane S), Supercoiled pET28a-SUMO plasmid (Lane 1), Single digestion of the pET28a-SUMO using the *BamHI* restriction enzyme (Lane 2), Single digestion of the pET28a-SUMO using the *XhoI* restriction enzyme (Lane 3).

3.3 CCHFV N protein expression and purification

3.3.1 Expression of His_6-SUMO CCHFV N fusion protein in *E. coli* BL21 (DE3) cells

Expression screens were initially made to determine the ability of different *E. coli* cells to express the His_6-SUMO CCHFV N fusion protein. The expression of His_6-SUMO CCHFV N fusion protein was tested by screening 200 ml LB induced with 200 μM IPTG overnight at 30 °C with Rosetta, Rosetta2, Gold and Star *E. coli* BL21 (DE3) cells, which were transformed with the pET28a-SUMO CCHFV N plasmid. The CCHFV N protein was expressed abundantly using Rosetta and Rosetta2 cells. No significant expression of His_6-SUMO CCHFV N fusion protein was seen whilst inducing in the Star or Gold cells (Figure 38).

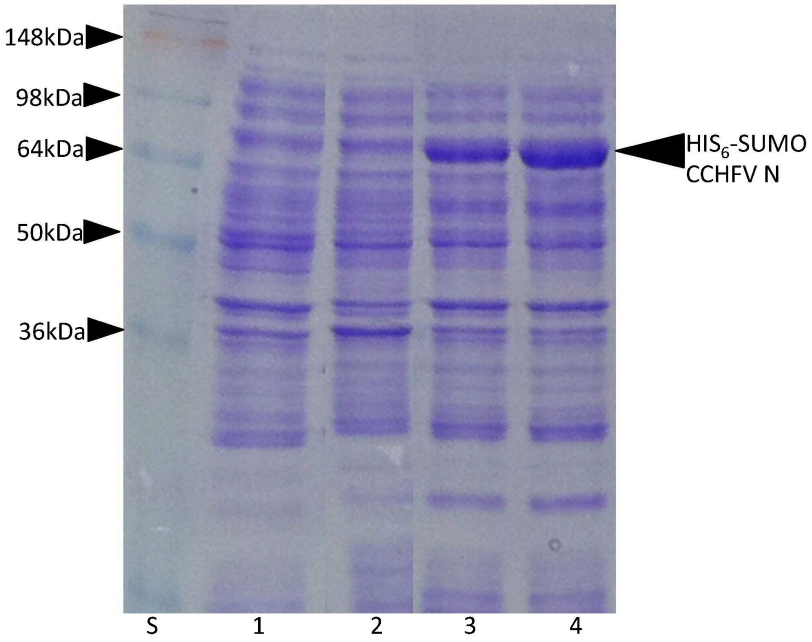


Figure 38. Analysis of expression of His$_6$-SUMO CCHFV N protein. By screening a range of bacterial expression cells transformed with pET28a-SUMO CCHFV N and induced with IPTG. SDS-PAGE analysis followed by Coomassie stain of whole cell lysates of Gold BL21 (DE3) cells (Lane 1), Star BL21 (DE3) cells (Lane 2), Rosetta BL21 (DE3) cells (Lane 3), and Rosetta2 BL21 (DE3) cells (Lane 4). Molecular weight markers (kDa) (Lane S).

Comparison of cultures transformed with pET28a-SUMO CCHFV N and an empty pET28a-SUMO plasmid confirmed the identity of a unique ~70 kDa species, which expressed abundantly in the Rosetta2 cells (Figure 39), and matched closely to the expected molecular weight of the His$_6$-SUMO CCHFV N fusion protein.

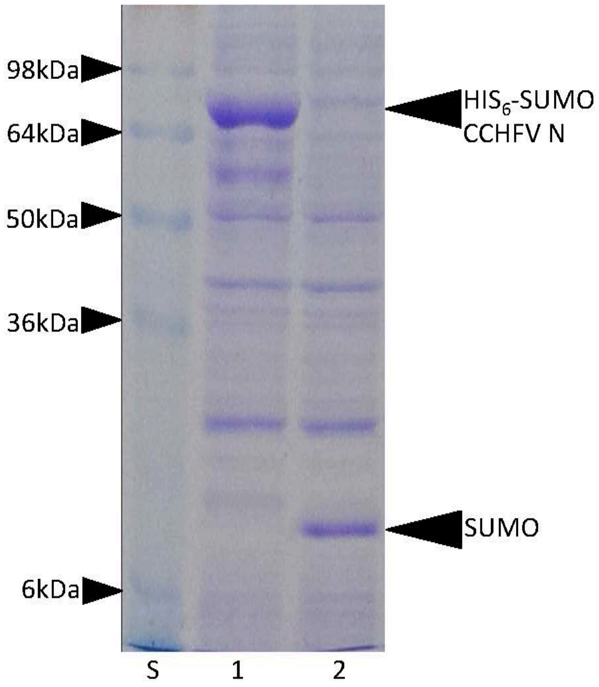


Figure 39. Analysis of His$_6$-SUMO CCHFV N protein expression in Rosetta2 *E. coli* cells using a negative control. (A) SDS-PAGE followed by Coomassie stain of whole cell lysate of Rosetta2 *E. coli* cells transformed with pET28a-SUMO CCHFV N and induced with IPTG (Lane 1) and pET28-SUMO plasmids and induced with IPTG (Lane 2). Molecular weight markers (kDa) (Lane S).

3.3.2 Solubility of His$_6$-SUMO CCHFV N fusion protein

It was found that cultures induced at 30 °C, with 200 mM IPTG, expressed completely insoluble protein for both Rosetta and Rosetta2 expression cells (Figure 40). The solubility of the Rosetta, and Rosetta2 expressed His$_6$-SUMO CCHFV N fusion protein was assessed by analyzing soluble and insoluble fractions using SDS-PAGE. CCHFV N protein solubility was best achieved by inducing cultures with 200 μM IPTG overnight at 18 °C (Figure 41).

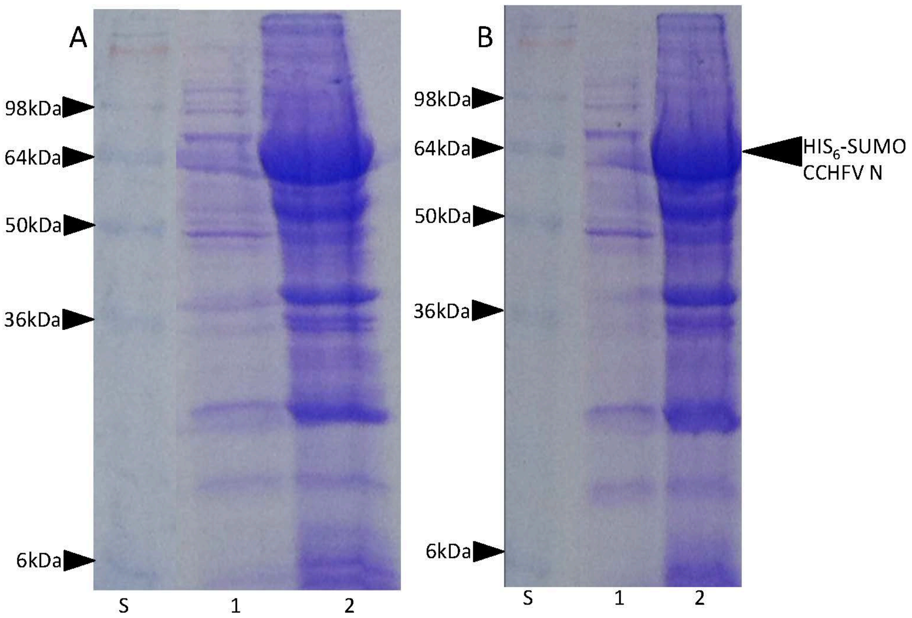


Figure 40. Analysis of soluble/insoluble fractions of Rosetta and Rosetta2 *E. coli* cells expressing the His$_6$-SUMO CCHFV N protein. Using SDS-PAGE followed by Coomassie stain. (A) The soluble fraction (Lane 1), and insoluble fraction (Lane 2) of Rosetta2 cells induced with 200 mM IPTG at 30 °C. Molecular weight markers (kDa) (Lane S). (B) The soluble fraction (Lane 1), and insoluble fraction (Lane 2) of Rosetta cells induced with 200 µM IPTG at 30 °C, Molecular weight markers (kDa) (Lane S).

By comparison the Rosetta2 bacterial cells expressed more soluble His$_6$-SUMO CCHFV N fusion protein than the Rosetta bacterial cells (Figure 41), and therefore the Rosetta2 *E. coli* cell line was implemented for all further expression experiments (Figure 41).

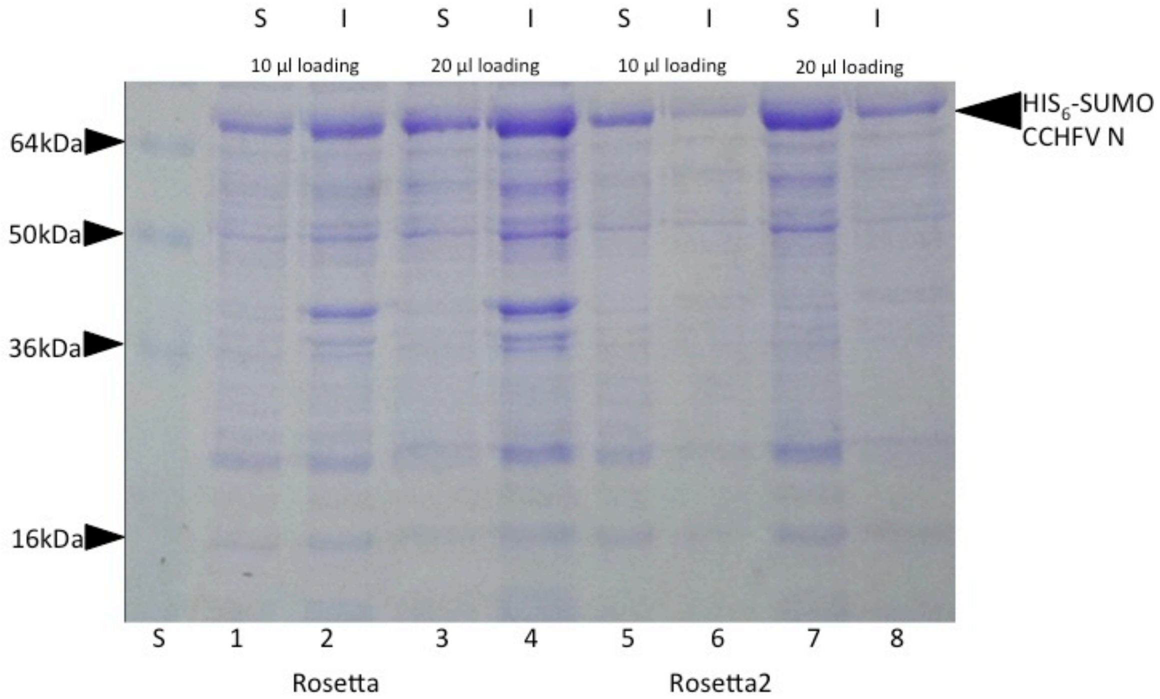


Figure 41. Analysis of solubility of His$_6$-SUMO CCHFV N protein expressed using Rosetta and Rosetta2 expression *E. coli* cells. SDS-PAGE followed by Coomassie stain analysis of soluble (Lanes 1, 3, 5, and 7), and insoluble fractions (Lanes 2, 4, 6, and 8) of 200 ml lysates from pET28-SUMO CCHFV N transformed Rosetta and Rosetta2 cells, induced with 200 µM IPTG at 18 °C. Molecular weight markers (kDa) (Lane S).

3.3.3 Purification of CCHFV N protein with Ni-Superflow Plus resin

The soluble fraction from an induced 200 ml culture of Rosetta2 was analyzed to determine to what degree the expressed His$_6$-SUMO CCHFV N fusion protein bound to nickel resin. The spin protocol was implemented in this case because only a small amount of soluble fraction was being analyzed. The majority of the expressed His$_6$-SUMO CCHFV N fusion protein successfully bound to the resin and was eluted from the spin column using 500 mM imidazole, whilst also being purified away from most of the non-specific bound contaminants with washes using binding buffer and 50 mM imidazole washes (Figure 42).

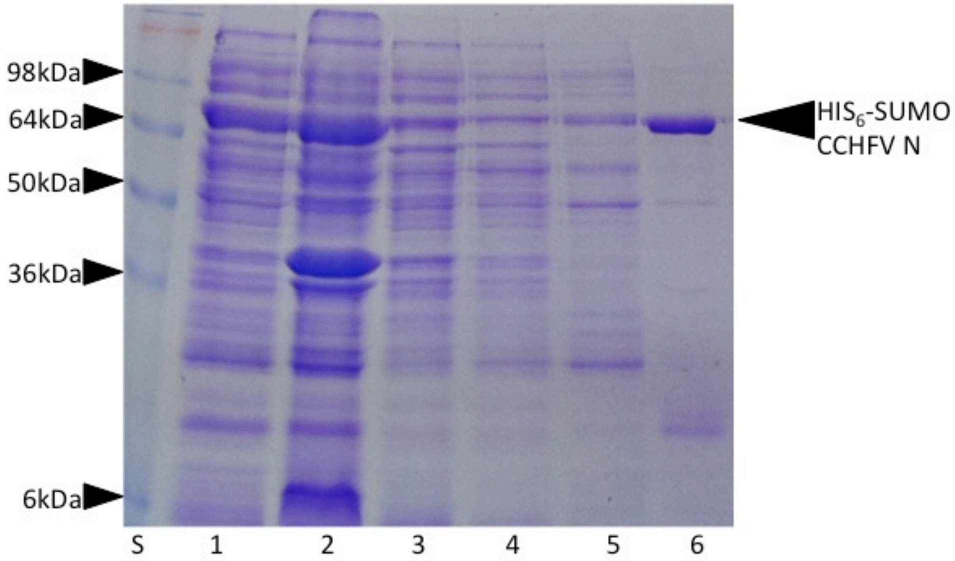


Figure 42. Analysis of nickel-affinity purification of His$_6$-SUMO CCHFV N fusion protein using the spin protocol. Fractions of the elution profile were analysed using SDS-PAGE followed by Coomassie stain. Soluble fraction (Lane 1), Insoluble fraction (Lane 2), Flowthrough (Lane 3), Binding buffer wash (Lane 4), 50 mM imidazole wash (Lane 5), 500 mM imidazole elution (Lane 6), Molecular weight markers (kDa) (Lane S).

3.3.4 Large-scale purification of CCHFV N protein with Ni-Superflow Plus resin

Expression and purification of high quantities of protein suitable for crystallization trials was initially achieved with 2 L culture of Rosetta2 bacterial cells containing the pET28a-SUMO CCHFV N plasmid. For meeting the requirement of even higher quantities of His$_6$-SUMO CCHFV N fusion protein this was later increased to a 6 L culture. The majority of the soluble fraction containing His$_6$-SUMO CCHFV N fusion protein bound to the Ni-Superflow Plus resin. After washes with binding buffer, 50 mM imidazole, 90 mM imidazole, and 120 mM imidazole, the majority of the non-specific contaminants were removed. The His$_6$-SUMO CCHFV N fusion protein bound was eluted from the resin and purified away from most of the contaminants using 300 mM imidazole (Figure 43). Typically the elution volume of the His$_6$-SUMO CCHFV N fusion protein was 15-20 ml, which was dialyzed, cleaved with SUMO protease overnight and filtered to avoid precipitation when concentrating.

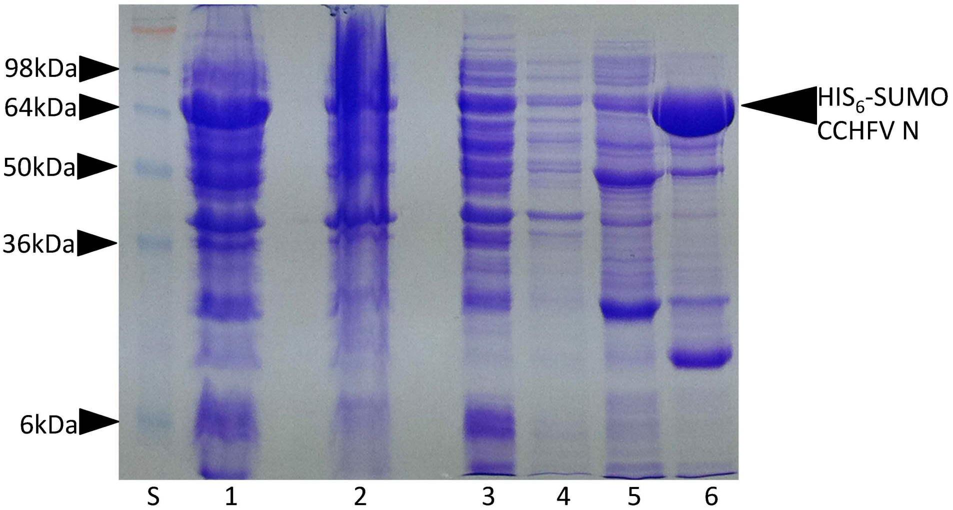


Figure 43. Analysis of nickel-affinity purification of His_6-SUMO-CCHFV N fusion protein using Ni-Superflow Plus resin. Fractions of the elution profile were analysed using SDS-PAGE followed by Coomassie stain. SDS-PAGE followed by Coomassie stain analysis of elution profile of soluble His_6-SUMO CCHFV N purified using the Ni-Superflow Plus resin. Molecular weight markers (kDa) (Lane S). Insoluble fraction (Lane 1), Flowthrough (Lane 2), Binding buffer wash (Lane 3), 50 mM imidazole wash (Lane 4), 90 mM imidazole wash (Lane 5), 300 mM imidazole elution (Lane 6).

3.3.5 Purification of CCHFV N protein using SEC

Typically 60-100 mg of CCHFV N protein from each 2 L culture was purified and concentrated to 5 ml, which was required for SEC. The SEC chromatogram elution profile of the CCHFV N protein produced three distinct peaks (Figure 44A). The main peak 2 was calibrated to be approximately 52 kDa, which closely corresponded to a CCHFV N monomer (54 kDa), whilst the much shallower and faster eluting species (peak 1) was calibrated to be approximately 107 kDa, which corresponded to a dimer (108 kDa). In addition the 260/280 nm ratio of the monomer peak indicated that it does not contain any nucleic acid, whilst the 260/280 nm ratio of the dimer indicated that it likely bound nucleic acid. Fractions corresponding to the center of both main peaks for both CCHFV N protein species were analyzed using SDS-PAGE gel for purity (Figure 44B). The native CCHFV N eluted with no apparent contamination from other non-specific proteins brought through from the Ni-Superflow Plus resin purification. Peak 3 corresponded to cleaved SUMO protease (Figure 44A).

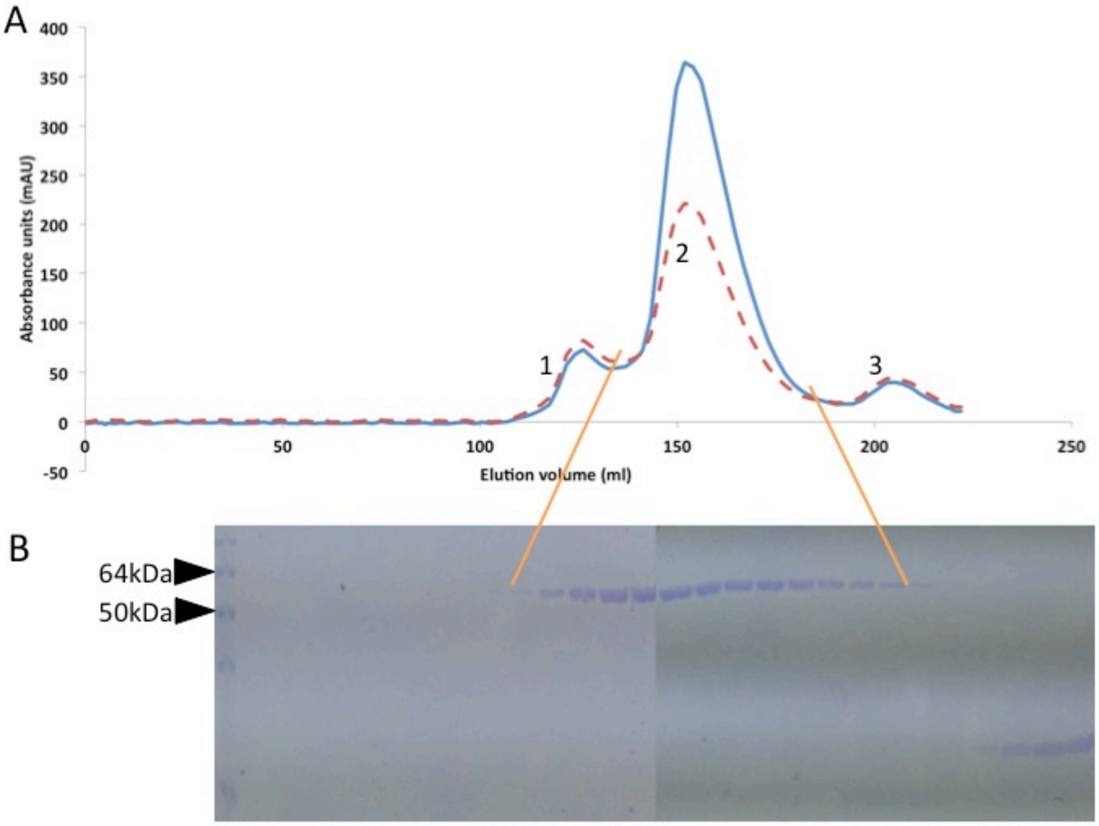


Figure 44. Analysis of the CCHFV N protein using SEC. (A) Bacterially expressed fusion CCHFV N protein was purified by nickel-affinity chromatography, after which the His$_6$-SUMO tag was removed by SUMO protease and applied to a Superdex S75 gel filtration column at a flow rate of 2.5 ml min^{-1}. The elution profile of the CCHFV N protein is shown in blue and dashed red at 280 nm and 260 nm wavelengths respectively. (B) SDS-PAGE followed by Coomassie stain analysis of elution profile CCHFV N protein following purification using a Superdex S75 gel filtration column. Orange lines indicate the peak 2 fractions analysed on SDS PAGE correspond to the S75 elution profile.

3.4 RNA Binding

The RNA binding content of the CCHFV N monomer protein peak was analyzed using phenol/chloroform extraction, and subjected to electrophoresis on a 1% agarose gel alongside ssRNA size markers. Controls for the experiment were also run in adjacent lanes and included running both RNA harvested from non-RNase treated BUNV tetramer protein and RNase treated BUNV N tetramer protein. There was no apparent RNA harvested from the CCHFV N protein monomer peak (Figure 45), whilst an expected sized band appeared from the non-RNase treated BUNV tetramer protein peak (Mohl and Barr, 2009).

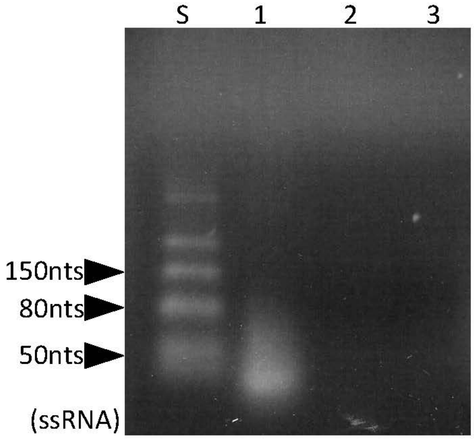


Figure 45. RNA binding analysis of the CCHFV N protein purified by SEC. RNA was harvested from purified CCHFV N protein monomers and subjected to electrophoresis on a 1% agarose gel (Lane 3), alongside ssRNA size markers (Lane S). Positive controls included non-RNase treated BUNV N protein extracted from a SEC tetramer peak (Lane 1), and the negative control included RNase treated BUNV N protein from the same SEC tetramer peak (Lane 2). This figure was generated with the help of Dr. Cheryl Walter, University of Leeds.

3.5 Crystallization of CCHFV N protein

3.5.1 Factorial crystal screens

Fractions collected from the centre of the SEC purified CCHFV N monomer protein peak were pooled together and concentrated until the concentration reached 15 mg ml^{-1} in approximately 500 μl volume. The concentration and purity of the CCHFV N protein was deemed sufficient to attempt crystallization trials using the sitting drop method. Crystals grew in 0.2 M magnesium acetate tetrahydrate, 0.1 M sodium cacodylate trihydrate pH 6.5, 20% w/v polyethylene glycol (PEG) 8,000, (Figure 46).

Figure 46. Crystallization of CCHFV N protein using the sitting drop vapour-diffusion method and carried out at 25 °C (Scale bar = 15 μm). Crystals grown in condition 18 from the crystal screen (Hampton Research) crystallization screening kit.

3.5.2 Optimization of factorial CCHFV N protein crystals

To improve the quality of these crystals, several rounds of screening using the hanging drop method were carried out, using a range of PEGs, different buffers (MOPS, MES, BTP), and salts, and ultimately diffraction quality crystals were grown in 0.2 M NaCl, 25% v/v PEG-600, 1 M Bis-Tris Propane (BTP) pH 6.5 (Figure 47). The best drop ratio was found to be 2.5 µl of CCHFV N protein and 2.5 µl reservoir solution, using a CCHFV protein concentration of 40 mg ml^{-1}.

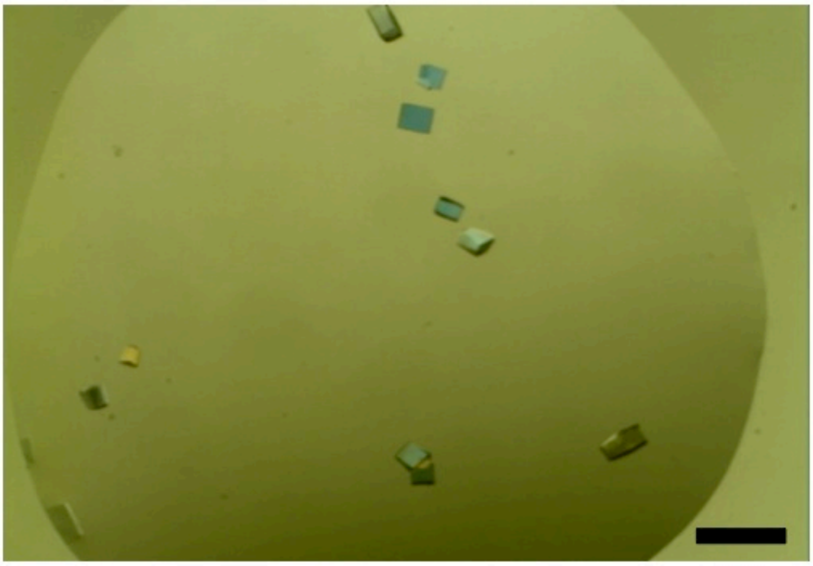


Figure 47. Optimization of CCHFV N protein crystals using the hanging drop vapour-diffusion method. Optimization produced crystals that diffracted to at least 2.1 Å and was carried out at 25 °C. Crystals were produced in drops containing 0.2 M NaCl, 25% v/v PEG-600, 1 M BTP pH 6.5 (Scale bar = 150 µm).

3.6 X-ray diffraction of CCHFV N protein crystals

3.6.1 X-Ray diffraction of factorial CCHFV N protein crystals

Crystals grown using the sitting drop vapour-diffusion method (Figure 46), were analyzed for X-ray diffraction at the Diamond Light source on beamline I02. Crystals diffracted to a maximum Bragg spacing of 4.5 Å and belonged to space group $P2_12_12_1$, with unit-cell parameters a = 73.84, b = 86, c = 264.00 Å, α= 90°, β= 90° γ= 90°. This data set was indexed and integrated using the Xia2 package (Kabsch, 2010) (1994) from the Diamond synchrotron suite of automated programs. These data were scaled with a maximum Bragg spacing of 4.5 Å with an R-merge of 12%. In the $P2_12_12_1$ space group there are three CCHFV N proteins per AU with a solvent content of 56%.

3.6.2 X-Ray diffraction of optimized CCHFV N protein crystals

The native data set used for structure solution was generated from crystals grown using the hanging drop vapour-diffusion method, (Figure 47). Freezing these crystals straight out of the drop was sufficient to give the best quality diffraction with maximum Bragg spacing of 2.1 Å diffraction (Figure 48). The crystals belonged to the space $C2$, with unit-cell parameters a = 149.8, b = 71.9, c = 102.0 Å, α= 90°, β= 111.2°, γ= 90°. It was calculated by the strategy function within the MOSFLM package that rotating this $C2$ space group crystal 120° with 0.5° oscillations was sufficient to collect a full X-ray diffraction data set of the CCHFV N protein, and produced 240 images. This data set was indexed and integrated using MOSFLM package (Leslie, 2006) and scaled using SCALA (Evans, 2006b) from the CCP4 suite (Winn et al., 2011). The data was scaled with a maximum Bragg spacing of 2.1 Å with an R-merge of 57%.

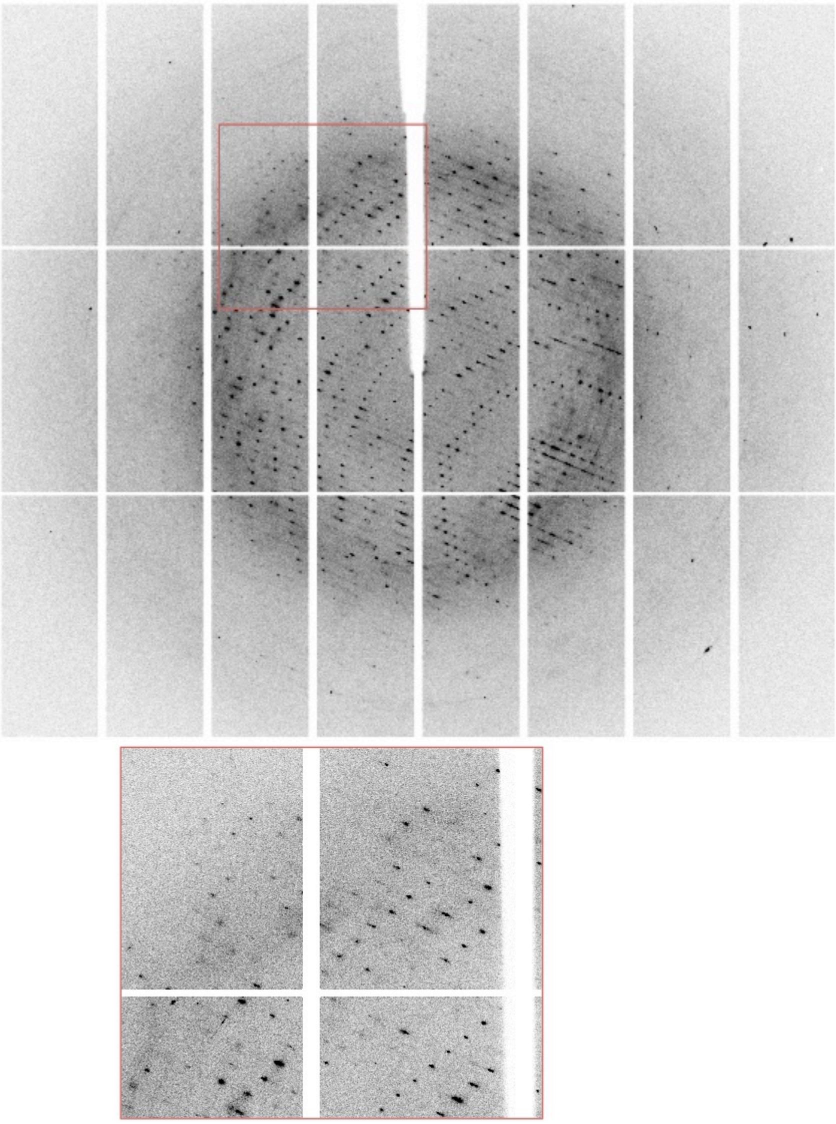


Figure 48. X-ray diffraction pattern of CCHFV N protein crystals. A 0.2° oscillation X-ray diffraction pattern collected from beamline I04-1 using 12,658 eV X-rays at wavelength 0.9795 Å with a Pilatus 6M detector. Reflections are observed to beyond 2.1 Å. The insert boarded in red shows the zoomed in image (x2) of the diffraction pattern boxed in red.

3.7 Self-rotation analysis

Self-rotation functions were calculated with the MOLREP program (Vagin and Teplyakov, 2010). The two strongest peaks κ=180° at (ϕ = 180°) and (ϕ = 0°) are derived from the crystallographic two-fold symmetry axes. The two additional peaks observed on the κ=180° plot correspond to NCS twofold symmetry (Figure 49). Considering the Matthews co-efficient (2.38 Å3 Da^{-1} with 48.35% solvent) (Matthews, 1968), and in conjunction with the twofold NCS in the self-rotation function, two CCHFV N protein molecules are likely to be present in the AU. There was no evidence of four fold or six fold crystallographic symmetry for κ=90° (Figure 49) or κ=60° (data not shown). Tetrameric or hexameric complexes similar to those seen for other bunyaviruses do not appear to be present, unless a combination of AUs results in, for instance, a tetramer (Mohl and Barr, 2009). It is also possible that apo CCHFV N protein does not oligomerize, or that the twofold NCS is unrelated to higher order RNP assemblies.

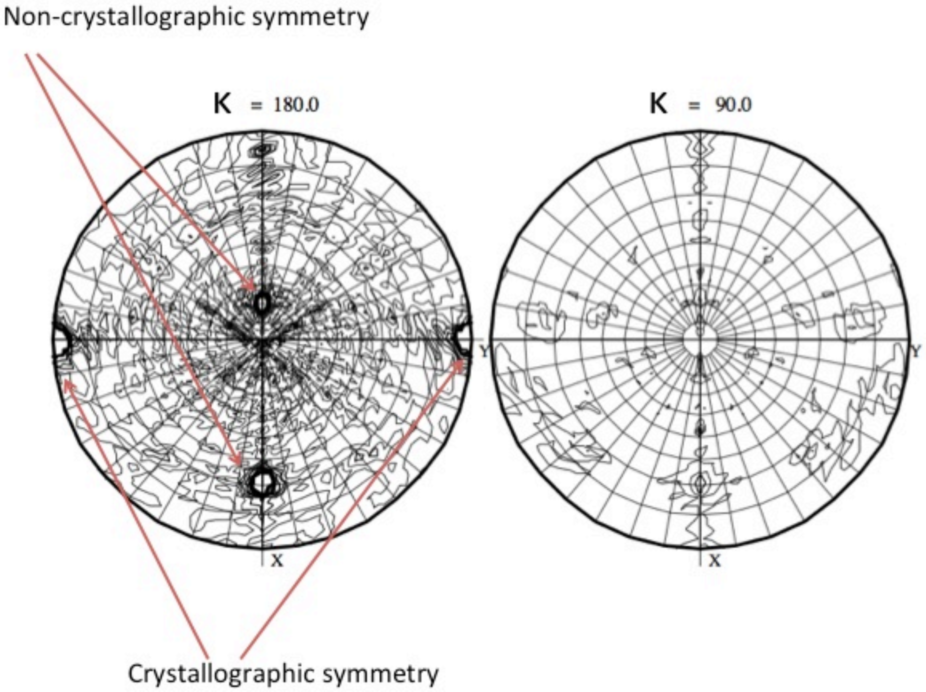


Figure 49. Self-rotation function calculated by MOLREP (Vagin and Teplyakov, 2010) using the CCHFV N protein *C*2 spacegroup crystals. Self-rotation searches with κ= 180° and κ= 90° were used to identify twofold and fourfold rotation angles, respectively.

3.8 Circular dichroism

Analysis of the CD spectrum, displaying a characteristic positive band at ~190 nm and the negative band at 208 nm, suggested a predominately α-helical content (Figure 50). The high proportion of α-helices in the CCHFV N protein is common for many nsRNA virus N proteins and is consistent with the crystal structure of the RVFV and othobunyavirus N proteins (Ferron et al., 2011, Ariza et al., 2013, Dong et al., 2013b, Niu et al., 2013, Reguera et al., 2013, Li et al., 2013).

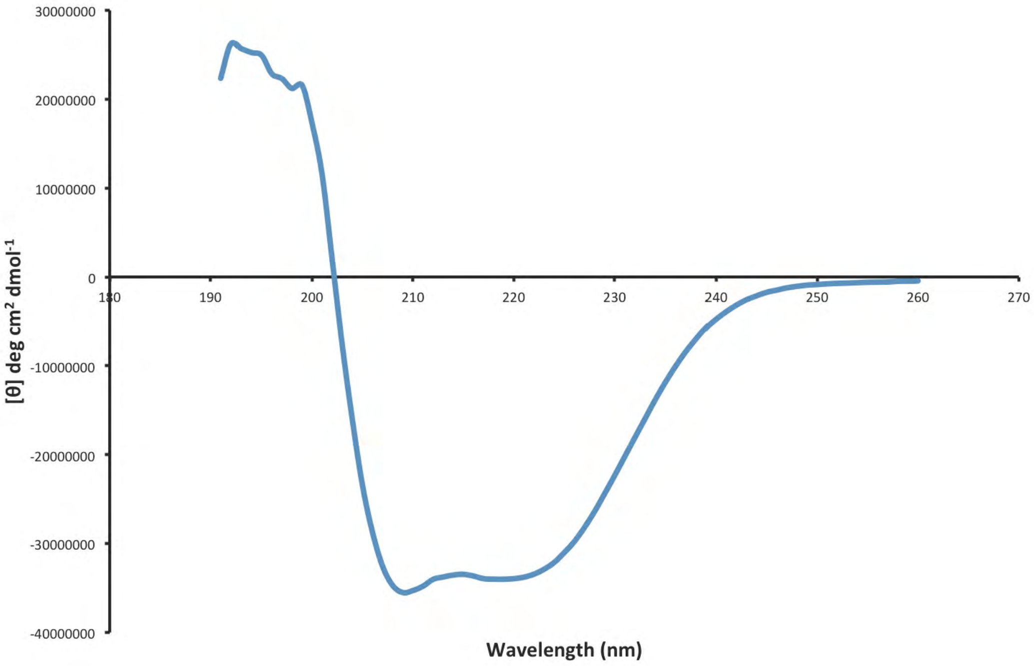


Figure 50. Circular dichroism spectrum of monomeric CCHFV N protein.

3.9 The Structure of CCHFV N protein

3.9.1 Molecular replacement

Initial attempts to determine initial phase information by molecular-replacement approaches using Phaser and MOLREP (McCoy, 2007, Vagin and Teplyakov, 2010), with the RVFV N protein (Raymond et al., 2010; Ferron et al., 2011) as a search model failed (the only bunyavirus N protein at the time available). By searching for CCHFV N sequence similarity using the basic local alignment tool (BLAST) tool, (Altschul et al., 1990), no model in the PDB with sequence identity above 20% was found. All of the remaining N proteins from nsNSV and sNSV for which there are high resolution structures also failed as molecular replacement models with none displaying sequence identity above 20% with the CCHFV N protein.

3.9.2 Selenomethionine CCHFV N protein

By replacing the sulphurs in methionine residues with selenium atoms in the CCHFV N protein the initial phases needed for substructure solution can be generated, without the need for soaking heavy atoms into native crystals. This offers the assurance that the heavy atoms are contained within the protein and in known positions within the amino acid sequence, and is advantageous for phasing because the heavy atoms will be fully occupied in each repeating unit of the crystal.

Bacterial expression of the CCHFV N protein containing selenomethionine was achieved using BL834 (DE3) *E. coli* expression cells, which are methionine auxotrophs, grown in M9 minimal media. Using the same induction conditions for expression as for the native CCHFV N protein sufficient amounts of protein were produced for crystallization. Mass spectrometry analysis of the native and selenomethionine derivative CCHFV N proteins confirmed selenomethionine had been successfully incorporated. The major species for the native CCHFV protein identified by mass spectrometry was 54,044 Da, closely corresponding to the predicted mass of 54,045 Da (Figure 51A). The major species for the selenomethionine protein was 54,643 Da, which corresponded to the predicted mass of 54,644 Da for selenomethionine CCHFV N protein (Figure 51B).

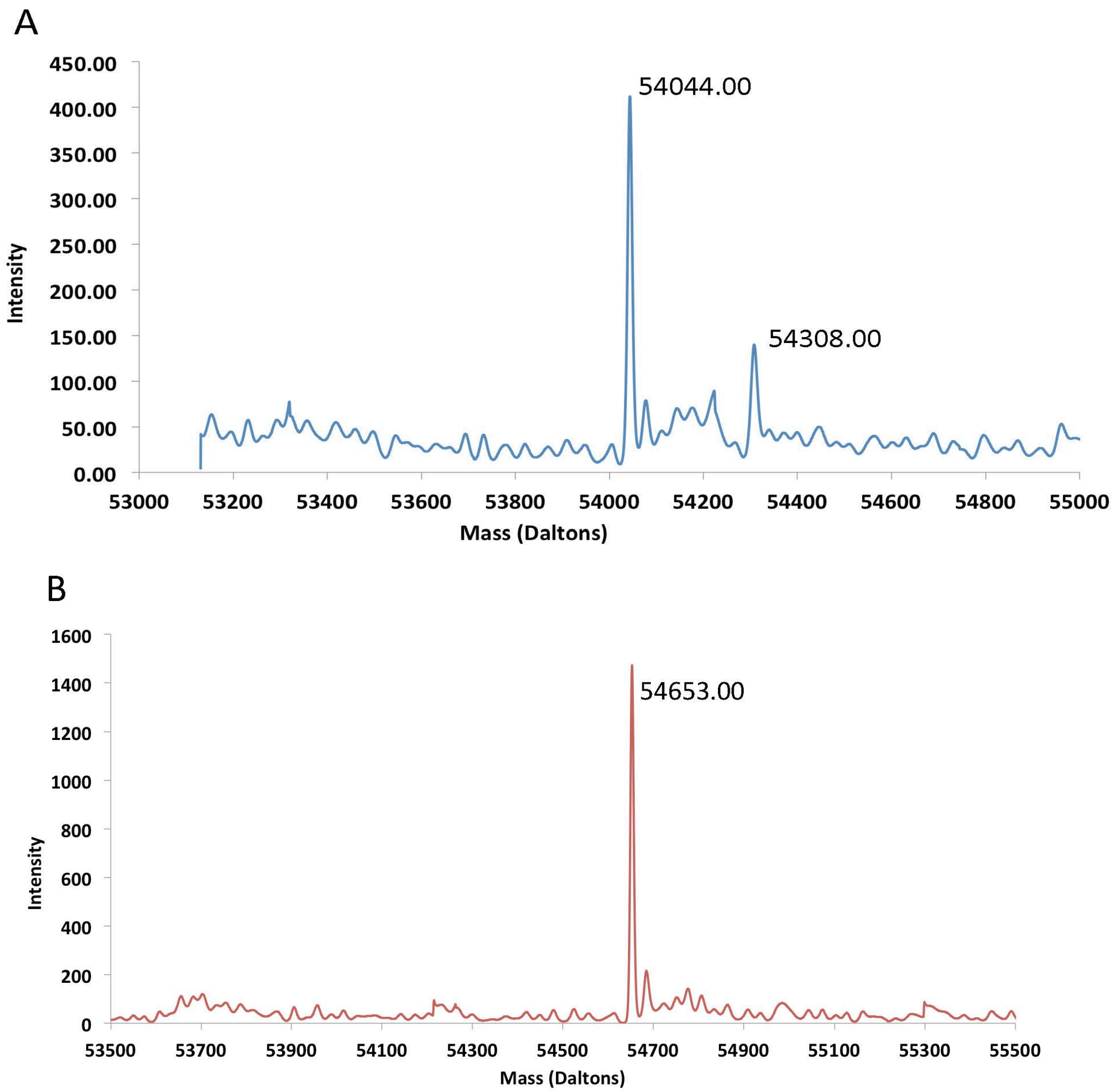


Figure 51. Mass spectrometry analysis of CCHFV N selenomethionine protein of native (A) and selenomethionine derivative (B) CCHFV N proteins, identified peak at 54,044 Da and 54,653 Da, which closely matched the predicted mass of 54,045 Da and 54,654 Da respectively.

Selenomethionine derivative CCHFV N protein crystals were found to grow in the same crystallization conditions for the native CCHFV protein (0.2 M NaCl, 25% v/v PEG-600, 1 M BTP pH 6.5), however the quality of diffraction, completeness of the datasets and their resolution were not good enough to obtain sufficient anomalous signal for substructure solution (Figure 52).

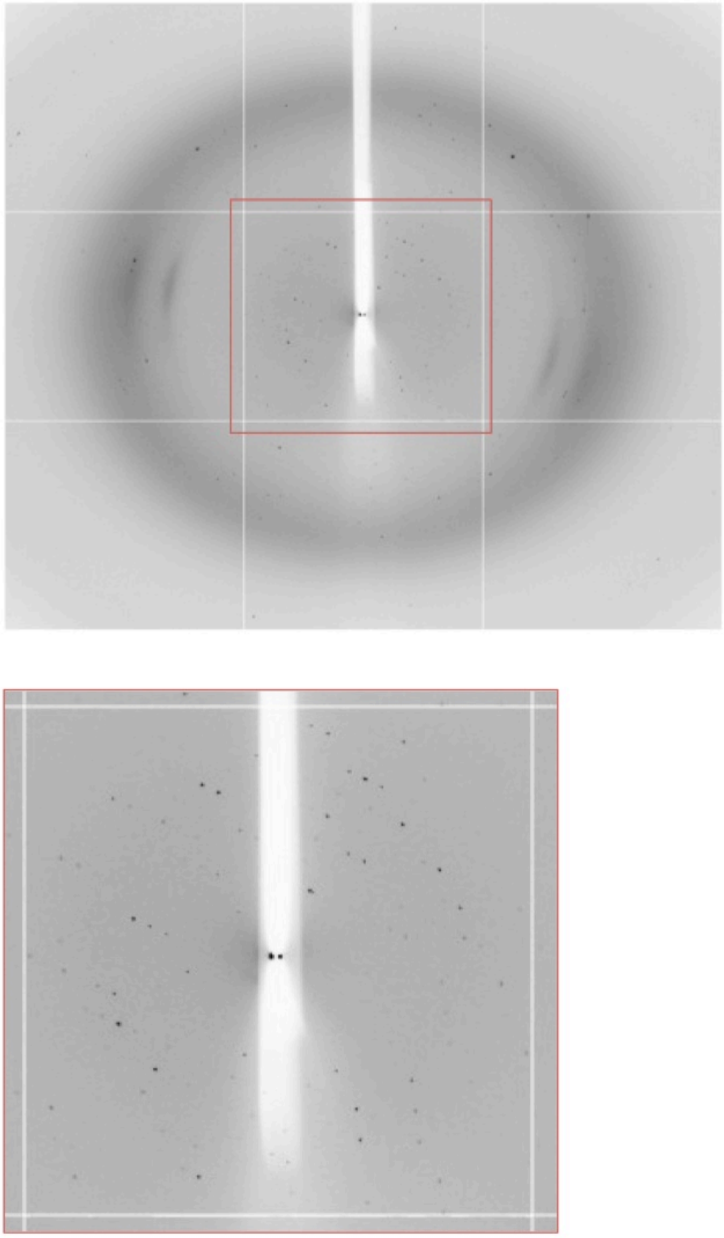


Figure 52. X-ray diffraction pattern of selenomethionine CCHFV N protein crystals. A 0.9° oscillation X-ray diffraction pattern using beamline I02, collected at the peak absorbance of selenium using 12,658 eV X-rays. The insert boarded in red shows the zoomed in image (x2) of the diffraction pattern boxed in red.

3.9.3 Heavy atom soaks into native crystals grown from CCHFV N protein

Two derivative data sets were collected from CCHFV N protein crystals, one which had been soaked in mother liquor containing 2 mM platinum chloride for 20 min, back soaked for 5 min in mother liquor, and another with 2 mM thimerosal for 30 min, back soaked for 10 min in mother liquor. Although the platinum chloride derivative data set was integrated (Leslie, 2006) and scaled (Evans, 2006b) to 2.8 Å, no detectable anomalous signal could be found.

The CCHFV N protein structure was solved with phases generated from a data set collected at 3 Å from crystals soaked with the mercury derivative thimerosal and extended to the native data set at 2.1 Å, using single isomorphous replacement with anomalous scattering (SIRAS). The CCHFV N protein derivative crystals were collected using 12,658 eV X-rays wavelength of the 0.9795 Å with a Pilatus 6M detector. In all 8 mercury sites in each AU were found, with three Hg atoms possessing sites in the globular domain and 1 Hg present within the arm of each CCHFV N protein molecule, with all Hg sites coordinated by cysteine residues (Figure 53). The statistics for the fully refined structure are given in Table 3.

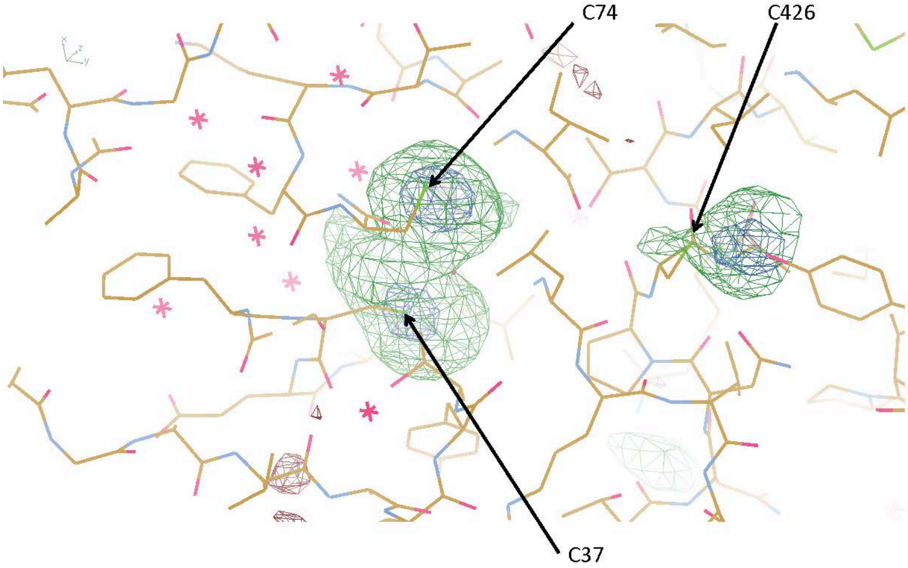


Figure 53. Schematic of 2mFo-DFc and mFo-DFc difference maps shows Hg atoms react with three separate cysteine within each globular domain. The 2mFo-DFc map is coloured blue and contoured at 6 sigma (i.e. 6.0 standard deviations above the mean electron density), and shows where Hg atoms are positioned. The mFo-DFc map is coloured green (3.0 sigma) and red (-3.0 sigma). Cysteine residues coordinating Hg atoms are labelled with arrows.

Parameter	Value	
	Native crystal	**Crystal soaked with thimerosal (8 Hg sites/AU)**
Wavelength (Å)	0.9795	0.9795
Space group	*C*2	C2
Cell parameters		
a (Å)	150.4	149.4
b (Å)	72.1	72.5
c (Å)	101.2	101.8
β (°)	110.7	111.1
Total no. of reflections	185,069 (27,462)	108,583 (16,046)
No. of unique reflections	56,441 (8,347)	20,325 (2,951)
Resolution (Å)	2.10 (2.21–2.10)	3.0 (3.16–3.00)
R_{merge} (%)	4.4 (46.7)	5.6 (13.7)
Completeness (%)	95.5 (97.2)	99.0 (99.4)
Redundancy	3.3 (3.3)	5.3 (5.4)
$I/\sigma(I)$	13.9 (2.5)	19.1 (9.6)
V_M (Å^3/Da)	2.38	2.39
No. of molecules per AU	2	2
R_{crystal}	18.39	
R_{free}	23.35	
No. of reflections in working set	53,584	
[a]Free *R* value set (% [no. of reflections])	5.1 (2,857)	
No. of nonhydrogen atoms	7,672	
No. of water molecules	313	
Mean B value (Å^2)	38.21	
RMSD from ideality		
Bonds (Å)	0.022	
Angles (°)	1.559	
Ramachandran statistics (%)		
Preferred region	97.02	
Allowed region	2.76	
Outliers	0.22	
MolProbity score (95th percentile)	1.65	

[a] Except where noted, numbers in parentheses are for the highest resolution shell.

Table 3. Data collection and refinement statistics for CCHFV N protein native and derivative crystals.

The structure of the CCHFV N protein was determined to give insight into its various functions, including RNA binding, homo-oligomerization, interaction with the viral glycoproteins to form the virus particle, and caspase-3 protease cleavage. The crystal structure of the full-length CCHFV N protein (residues 1 to 482) of CCHFV strain Baghdad-12 was determined in two space groups. Native crystals belonging to the monoclinic space group *C*2 and orthorhombic $P2_12_12_1$ spacegroup with two and three CCHFV N monomers in the AU respectively. For the model from the *C*2 data each monomer was assigned 474 residues, with residues 183 to 191 forming a disordered loop that is missing in the electron density for both protomers (Table 3; Figure 54). The Ramachandran plot for structural refinement can be found in figure 99.

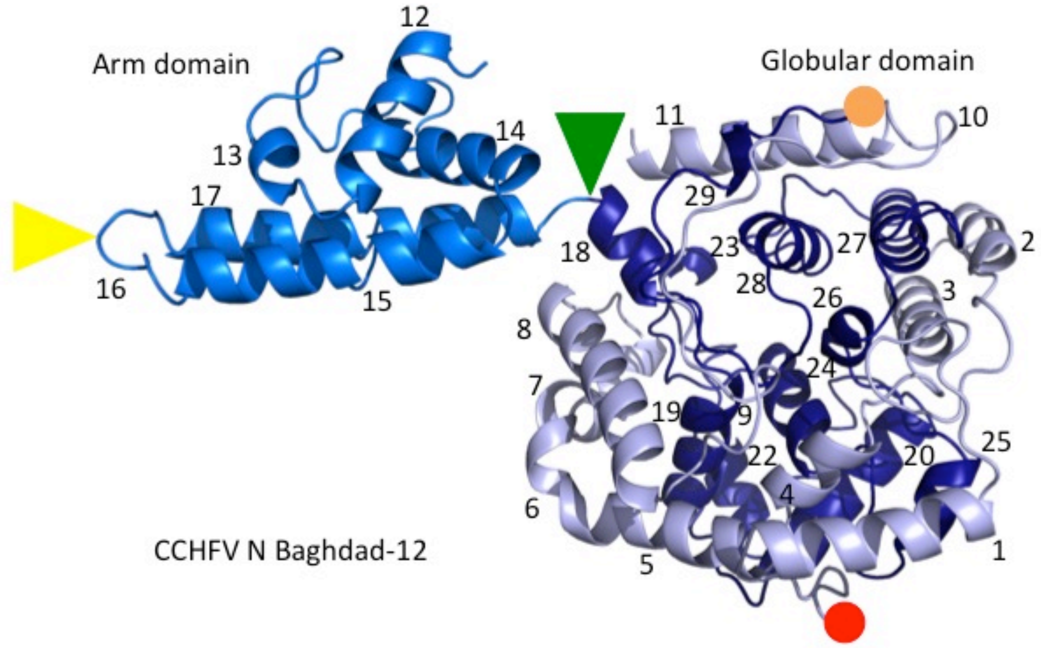


Figure 54. Crystal structure of the CCHFV N protein solved from *C*2 crystals. (A) The crystal structure of CCHFV N monomer is shown by cartoon representation. The globular domain is depicted with alpha helices coloured in light blue or density blue, which represent the N-terminal and C-terminal portions respectively. The arm portion of the CCHFV N protein is shown in marine blue. The C-terminus of the globular domain is shown as an orange sphere, and the N-terminus is shown as a red sphere. Coloured arrowheads point to the caspase-3 protease DEVD cleavage motif (yellow) and the segregation of helices 17 and 18 (green). The figure was generated using PyMol.

N monomers consist of a globular core comprising 23 alpha helices (α1 to α11 and α18 to α29), with a prominent additional structural element termed the "arm," comprising two long alpha helices (α15 and α17), that extends away from the core, with an exposed loop at the apex supported by a small three-helix bundle (α12 to α14) (Figure 54). The C-terminal half of the N protein provides the helices at the core of the globular domain, which in turn are surrounded by helices from the N-terminal half. The monomers are packed in a head-to-tail arrangement, with the interface between monomers comprising helices α6 and α7 and helices α21 and α22 within the globular domain.

Native crystals belonging to the orthorhombic space group $P2_12_12_1$ were also obtained. Despite diffracting to only 4.7 Å resolution, the structure could be unambiguously solved by molecular replacement using the refined model generated from the $C2$ spacegroup and shows three CCHFV N monomers in the AU (Table 4; Figure 55).

Parameter	**Value**
	Native crystal
Wavelength (Å)	0.9795
Space group	$P2_12_12_1$
Cell parameters	
a (Å)	73.3
b (Å)	85.3
c (Å)	262.4
α (°)	90
β (°)	90
γ (°)	90
Total no. of reflections	33923 (2491)
No. of unique reflections	9030 (641)
Resolution (Å)	4.71 (4.83–4.71)
R_{merge} (%)	1.27 (48.2)
Completeness (%)	99.4 (99.3)
Redundancy	3.8 (3.9)
$I/\sigma(I)$	6.9 (2.8)
No. of molecules per AU	3
VM (Å^3/Da)	1.59

R_{factor} 19.70

R_{free} 32.01

Table 4. Data collection and refinement statistics for $P2_12_12_1$ CCHFV N protein crystals

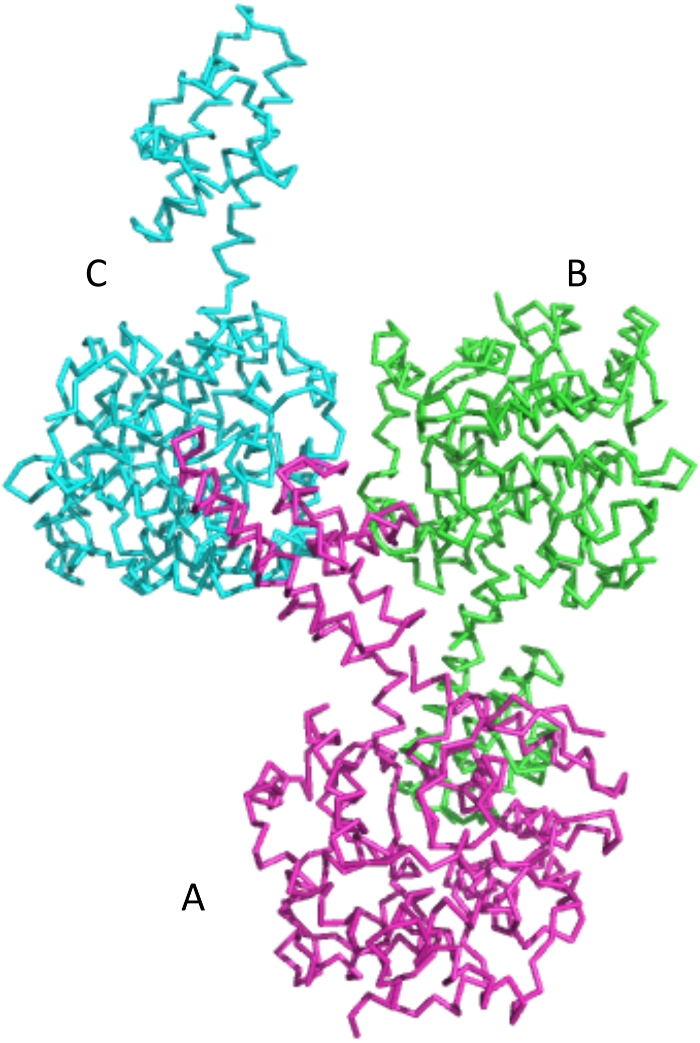


Figure 55. Ribbon representation of the $P2_12_12_1$ CCHFV N protein crystal structure solved by molecular replacement. Each AU is occupied with three CCHFV N monomers named A, B and C, and are coloured magenta, light green and turquoise respectively. The dimer interface in the $C2$ CCHFV N protein crystal structure can also be seen the $P2_12_12_1$ spacegroup (between monomers B and C). The figure was generated using PyMol.

The CCHFV N molecules are in the same conformation in both space groups. Due the poor resolution of the $P2_12_12_1$ data, only changes to alpha helices (α15 and α17) within the stalk region of molecule C, were unambiguously guided by clear positive 2mFo-DFc density (Figure 56).

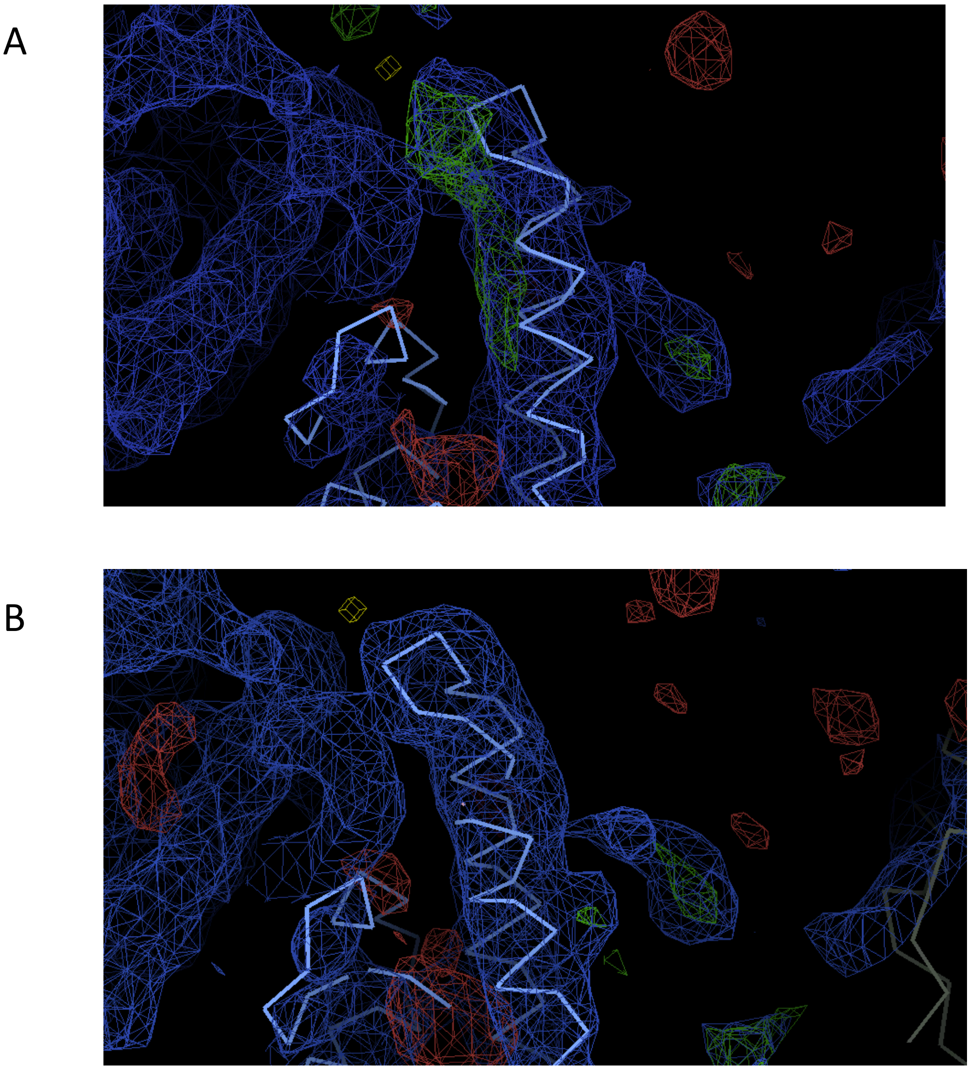


Figure 56. Manual alteration and refinement of the arm domain in molecule C in the AU of the $P2_12_12_1$ CCHFV N protein crystal using 2mFo-DFc and mFo-DFc difference maps. The 2mFo-DFc map is coloured blue and contoured at 1 sigma (i.e. 1.0 standard deviations above the mean electron density), and shows. The mFo-DFc map is coloured green (3.0 sigma) and red (-3.0 sigma). (A) Before alteration and refinement using REFMAC (Murshudov et al., 1997). (B) After alteration and refinement using REFMAC . Maps were displayed using Coot (Emsley and Cowtan, 2004).

In the $P2_12_12_1$ space group, two monomers were stacked in an identical head to tail arrangement (Figure 55, monomers B and C), with a third monomer alongside, which does not share the same “dimer” interface as in the *C*2 AU (Figure 60) - neither with the 2 molecules in the $P2_12_12_1$ AU nor with lattice neighbours.

Superposition of the body domains revealed the position of the stalk domains with respect to the globular domain is shared between all of the three CCHFV N proteins, with an overall root mean square deviation (RMSD) of 0.535 between molecules A an B, 0.576 between molecules A and C, and 0.589 between molecules B and C (Figure 57).

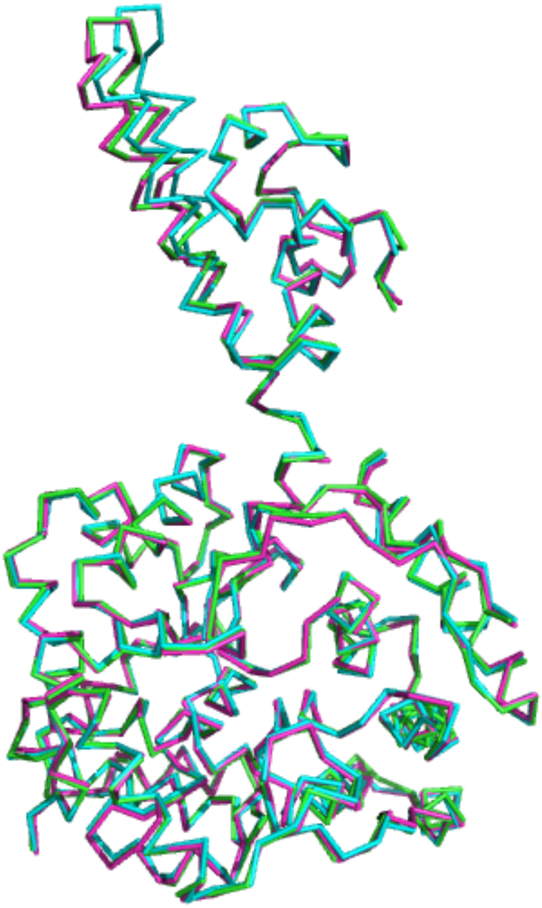


Figure 57. Superpositioning of the three molecules of the CCHFV N protein in $P2_12_12_1$ AU. The three molecules named A, B and C, are coloured in magenta, light green and turquoise respectively. The figure was generated using PyMol.

4.9.4 Electrostatic surface potential

The crystal structure of the CCHFV N protein allowed calculation of the electrostatic surface potential and revealed two positively charged regions. The first region is composed of residues K90, K91, K98, E112, E121, K132, K135, R140, S149, D171, R177, Q300, Y470 K473, which form a continuous positively charged area located on the surface of the globular and under the arm domain, which is termed the "platform" (Figure 58A). The additional region, composed of residues K342, K343, H453, K411, H456, E387, I304, W313 is adjacent to a deep basic crevice, and is referred to as the "pocket" (Figure 58B). These two regions comprise basic residues well conserved among all nairoviruses and therefore suggest a role in RNA binding (Wang et al., 2012)

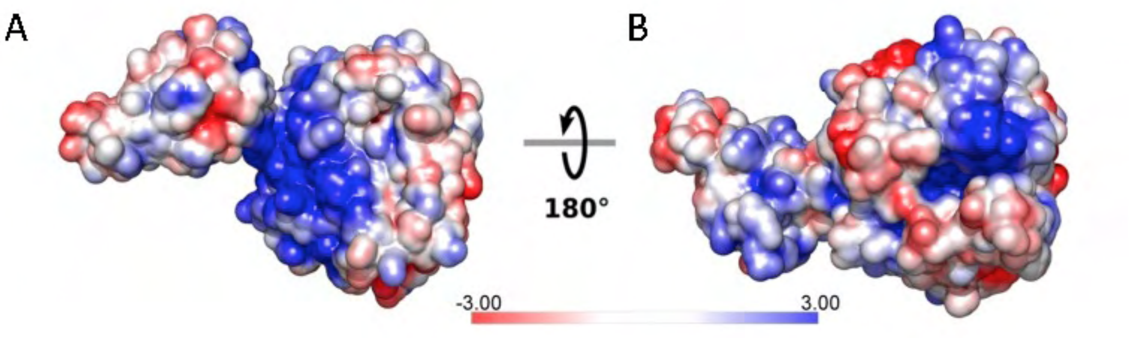


Figure 58. Electrostatic potential analysis of the CCHFV N protein from Baghdad-12 (A) The electrostatic surface potential of N protein from Baghdad-12 (generated with APBS) suggests a positively charged RNA binding "platform", solely located within the surface of the globular domain (Unni et al., 2011). (B) Rotation by 180 degrees suggests a smaller additional positively charged groove, termed the "pocket". The figure was generated using APBS in PyMol.

3.10 Similarity to other sNSV N proteins

The *Bunyaviridae* family is diverse with significant genetic differences between members within each of the five established genera and not surprisingly the CCHFV N protein has no primary sequence homologs within the family or with other nsRNA virus N proteins. Therefore the tertiary fold of the CCHFV N protein crystal structure was analysed using the DALI server to gain insights into bunyavirus phylogenetic relationships. Alignment between the CCHFV and RVFV N globular domains revealed essentially no tertiary structural homology (Z score = 1.7), and zero similarity was detected within the arm domain. Furthermore alignment using the DALI server did not find any structural homology between CCHFV N protein and any of the orthobunyavirus protein crystal structures (Ariza et al., 2013, Reguera et al., 2013, Dong et al., 2013b, Li et al., 2013, Niu et al., 2013).

Interestingly, the DALI server identified the highest scoring structural alignment with the N protein from the two segmented LASV (arenavirus) (Hastie et al., 2011a, Hastie et al., 2011b, Brunotte et al., 2011, Qi et al., 2010). The strongest structural similarity was found within the N-terminal residues 1 to 340, which comprise the RNA binding domain of LASV NP (Figure 59) (Hastie et al., 2011b). Furthermore the DALI server identified structural homology for all forms of the LASV NP deposited in the PDB, with Z-scores between 12.4 and 15.3. Superimposition of the globular domains from the CCHFV N protein and RNA free LASV NP gave a Cα root mean square deviation of 3.38 Å for 237 residues with 8.01% sequence identity. There was however a significant step down to the next highest score (Z-score 4.4; PDB 1w63) to what appeared to be an unrelated structure.

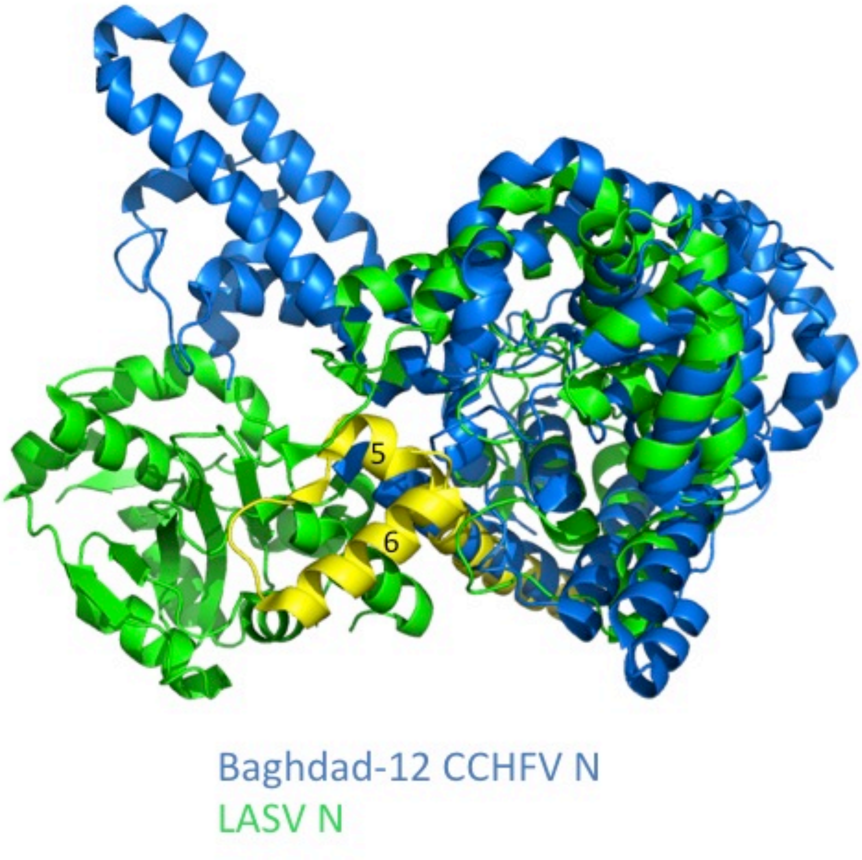


Figure 59. Superposition of the Baghdad-12 CCHFV and LASV N/NP proteins. (A) Structural homology between that of CCHFV and LASV N/NP protein structures. The CCHFV protein is shown in marine blue. The LASV N structure is predominately shown in green, and the α5, loop/α6 are numbered and coloured yellow. Hastie et al (Hastie et al., 2011b) reported this region moves upon RNA binding. Whilst α5 is shared by both N proteins there is no structural equivalent of the α6 RNA gate in the CCHFV N protein. The figure was generated using PyMol.

3.11 The N-N dimer interface

The encapsidation of the genomic RNA by all sNSVs involved the assembly of the N protein into multimeric protein complexes. It is plausible that the interface seen between the 2 NCS related molecules in the *C*2 AU represents a “real” dimer interface, which is present after oligomerization on RNA. The same arrangement is seen between two of the three NCS molecules in the $P2_12_12_1$ spacegroup (Figure 60). Analysis to support whether these two NCS related molecules represent a biologically relevant dimer, was undertaken by PISA (Krissinel and Henrick, 2007).

Analysis of the dimer interface in both $P2_12_12_1$ and $C2$ crystal forms using the PISA server (Krissinel and Henrick, 2007) revealed an average buried surface area of 1,015 Å^2, a ΔG value of −9.5 kcal/mol, and a complexation significance score (CSS) of 0.0 (a score of 1 suggests a stable interface, and a score of 0 suggests a nonstable interface, i.e., crystal packing and therefore not a dimer interface (Figure 60).

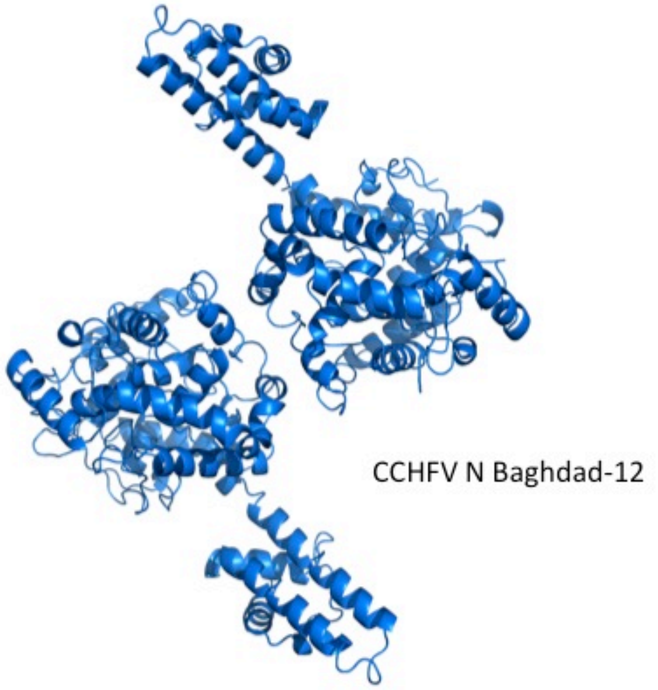


Figure 60. Cartoon representation of two CCHFV N protomers within the $C2$ crystal lattice. The interface between the two CCHFV N protein monomers was investigating using PISA. The figure was generated using PyMol.

3.12 Consequences of cleavage of the DEVD caspase-3 protease recognition site

The crystal structure of CCHFV N revealed the location of the DEVD motif is positioned at the apex of the arm domain (Figure 54, yellow arrowhead). A reaction containing recombinant caspase-3 protease and purified wild-type CCHFV N protein was set up, and cleavage at the DEVD site was monitored by analysis of digestion products separated using SDS-PAGE (Figure 61A). Surprisingly, caspase-3 protease cleavage of wild-type N protein resulted in incomplete digestion even following overnight incubation at an equal molar ratio (Figure 61A, lane 4), indicating that the CCHFV N protein was a poor substrate for cleavage despite the apparent accessibility of the DEVD motif (Figure 54). Caspase-3 protease activity was confirmed using a fluorogenic peptide derived from PARP (Figure 62), and as expected, CCHFV N mutants with altered cleavage motifs (DEVE and DEVG) were uncleaved (Figure 61B, lanes 2 and 3).

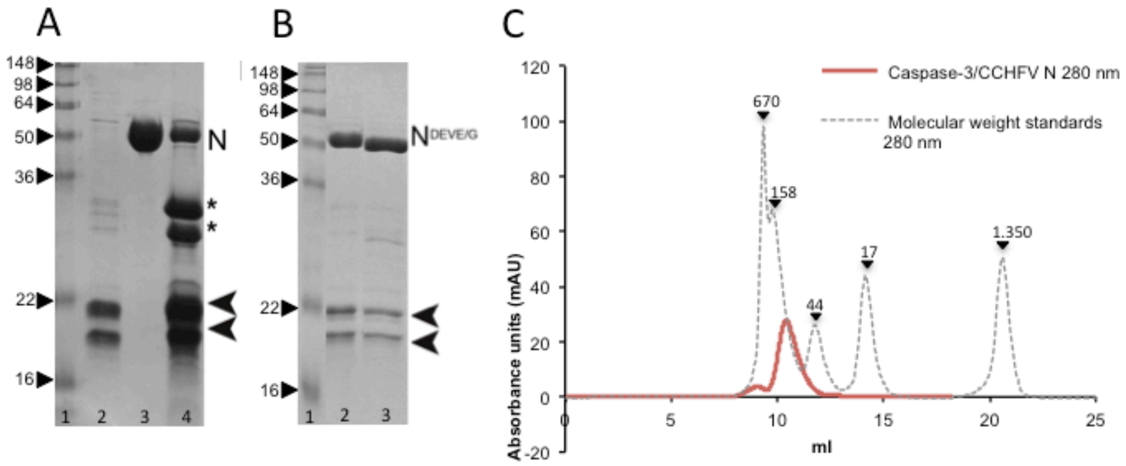


Figure 61. CCHFV N protein remains intact upon cleavage by caspase-3 protease. (A) SDS-PAGE gel stained with Coomassie showing caspase-3 protease cleavage of wild-type CCHFV N protein into two fragments of the expected sizes. Size markers in kDa (Lane 1), purified caspase-3 protease (Lane 2), purified CCHFV N protein (Lane 3) CCHFV N protein and caspase-3 protease (1:1) (Lane 4). N marks the full-length CCHFV N protein, arrowheads mark the two fragments of caspase-3 protease, and asterisks mark the N- and C-terminal cleaved fragments upon caspase-3 cleavage. (B) CCHFV N protein DEVD mutants (DEVE or DEVG) are not cleaved by caspase-3 protease. Size markers in kDa (Lane 1) CCHFV N-DEVE plus caspase-3 protease (Lane 2), CCHFV N-DEVG plus caspase-3 protease (Lane 3). (C) SEC analysis of CCHFV N protein and caspase-3 protease digestion products. The following molecular mass standards (dashed line) are given in kDa: thyroglobulin, 670; β-globulin, 158; ovalbumin, 44; myoglobulin, 17; and vitamin B_{12}, 1.35.

SEC analysis of the reaction products generated by caspase-3 protease digestion of CCHFV N protein resulted in the cleaved N protein being eluted with an apparent molecular mass based on molecular weight standards, corresponding to that of an intact monomer (Figure 61C). Analysis of the chromatogram with lower molecular weight markers revealed no evidence of the detached N- or C-termini cleavage products in solution (approximately 24 to 30 kDa) (Figure 61C).

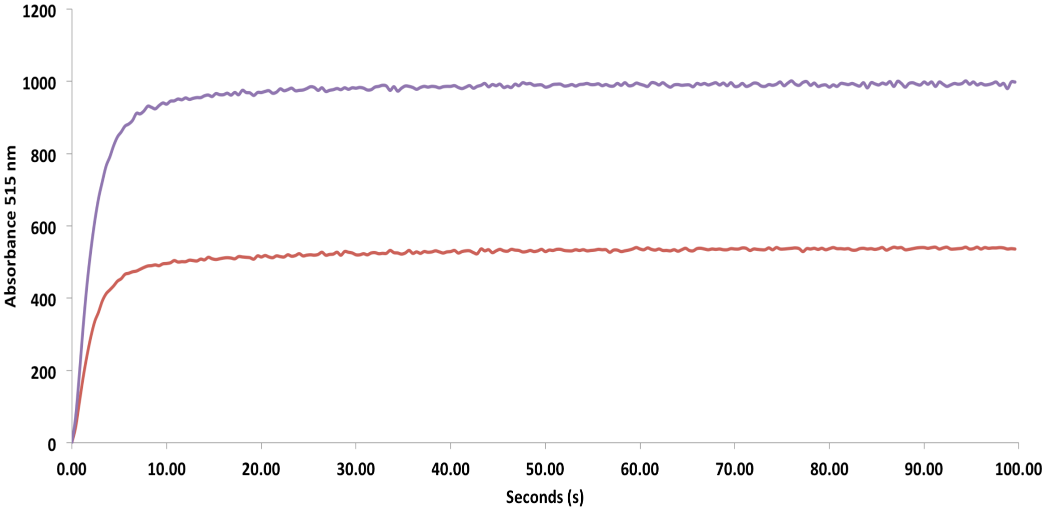


Figure 62. Caspase-3 protease cleavage of fluorogenic substrate. 3 mM of caspase-3 protease was added to 10 mM (red line) and 20 mM (purple line) of fluorogenic peptide derived from PARP, and absorbance was measured at 515 nm. Negative controls including an assay using no substrate in which no absorbance was measured.

The results indicate that upon cleavage by caspase-3 protease the CCHFV N protein remains intact, and this is most likely a consequence of the close contact of between the helical structural elements within the globular domain (Figure 63).

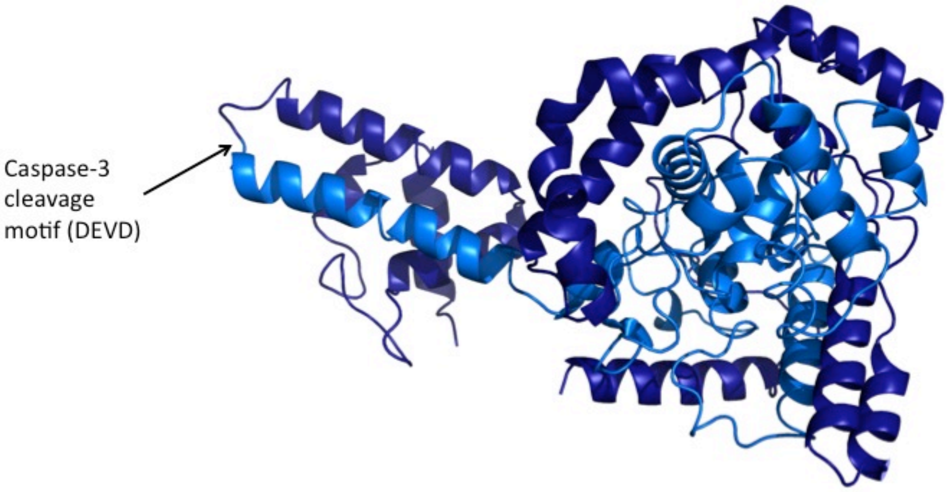


Figure 63. Location of the caspase-3 protease cleavage site on the CCHFV N protein. The two halves of the CCHFV N protein generated by caspase-3 protease cleavage are represented by marine blue (amino acids 1-266) and light blue (amino acids 267-482). The figure was generated using PyMol.

3.13 Summary of the CCHFV N protein crystal structure results

CCHFV N protein is composed of a globular core comprising 15 alpha helices (α1-8; α14-20) and a prominent additional structural element called the 'arm' comprising two long alpha helices (α9-10) extending away from the globular core, presenting an exposed loop at its apex, and supported by a small three helix bundle (α11-13; Figure 51). The C-terminal half of N provides the helices at the core of the globular domain, and in turn is surrounded by helices from the N-terminal half.

In the *C2* space group, the monomers are packed in a head-to-tail arrangement, with the AU dimer interface comprising helices α3, α6-8 within the globular domain. The $P2_12_12_1$ space group also displays the same dimer interface, with an additional monomer alongside, making up three CCHFV N proteins in the AU (Figure 55).

Electrostatic surface potential suggests a possible RNA binding 'platform' adjacent to the arm. Structural comparisons indicate that the CCHFV N globular domain exhibits a higher degree of structural homology with the N-terminal domain of LASV NP, a member of the *Arenaviridae* family, and appears essentially structurally un-related to RVFV N protein, a member within the *Bunyaviridae* family.

Caspase-3 protease cleaves the polypeptide chain at the exposed DEVD motif; however, the cleaved N protein remains an intact unit, likely due to the intimate association of N- and C-terminal fragments in the globular domain.

3.14 Discussion

The initial aim of this project was to express and purify highly concentrated recombinant CCHFV N protein using *E. coli*. To address this aim an approach was implemented, and involved expression of recombinant CCHFV N protein N-terminally linked to a His_6-SUMO tag. The decision to use the SUMO moiety was made to overcome any potential expression or solubility pitfalls, which are inherent in recombinant expression of mammalian proteins within *E coli*. The SUMO tag is cleaved by the SUMO protease, which is highly efficient and recognizes secondary structure, whilst leaving only a serine on the N-terminus of the native protein.

The next strategy of this project endeavored to better understand the functional and structural properties of the CCHFV N protein. Firstly by generating abundant near native CCHFV N protein it was demonstrated by SEC that CCHFV N protein exists predominately as a monomer in solution and lacks the propensity to form higher molecular weight oligomers like other members of the *Bunyaviridae* family. Next phenol/chloroform extraction of RNA revealed the CCHFV N protein monomer peak was not binding to *E. coli* RNA during the expression and purification processes. Finally X-ray crystallography was implemented to characterize the atomic structure of the CCHFV N protein. The crystal structure of the CCHFV N protein revealed important details about important functions such as RNA binding and also revealed previously unrecognized evolutionary relationships with other nsRNA viruses.

3.14.1 Cloning

Initial attempts to clone the CCHFV N cDNA into the pET28a-SUMO plasmid were successful, and the construct was then used in expression trials, using a number of different *E. coli* cell types, induction temperatures and IPTG concentrations. Expression of CCHFV N protein was best achieved using Rosetta2 *E. coli* cells, which supply rare tRNAs for 7 rare codons (AGA, AGG, AUA, CUA, GGA, CCC, and CGG), and offer universal translation, otherwise limited by the codon usage of the Star and Gold strains of *E. coli.*

3.14.2 Expression and purification

Expression of soluble CCHFV N protein was found to be optimum when inducing the culture at 18 ˚C with 100 μM IPTG. The Ni-Superflow Plus resin purification was highly efficient in providing highly concentrated and pure CCHFV N protein, ready for second round purification using SEC. The Superdex S75 gel based column separated CCHFV N protein away from contaminants, and also separated different molecular weight species of CCHFV N protein and the His_6-SUMO tag. The majority of the CCHFV N protein was monomeric, however larger dimer species were seen, and is consistent with its role of interacting with itself and forming large infectious RNP complexes.

It is not uncommon for nsRNA viral N proteins to bind heterogeneous *E. coli* RNA co-purified from the expression host (Ariza et al., 2013, Ferron et al., 2011). Crystallizing and solving the CCHFV N protein bound to *E. coli* RNA can be beneficial for elucidating important mechanisms such as oligomerization, RNA binding and RNP assembly. However one inherent problem arising from attempting to crystallize N protein with *E. coli* RNA is the heterogeneity arising from different bacterial RNAs of various sequences and lengths binding to the N protein. The 260/280 nm ratio of the monomer SEC peak indicated the CCHFV N protein was RNA free, and this was further supported using the phenol/chloroform method for extraction of RNA. As there was no heterogeneous *E. coli* bound to the monomeric CCHFV N protein, it was deemed sufficient for crystallization because the sample was considered homogeneous, and therefore more likely to produce diffracting crystals.

3.14.3 Crystallization and X-ray diffraction of the CCHFV N protein

This process allowed the expression of pure CCHFV N protein that was suitable for structural analysis using X-ray crystallography. Factorial crystal screens yielded crystals from 0.2 M magnesium acetate tetrahydrate, 0.1 M sodium cacodylate trihydrate pH 6.5, 20% w/v PEG-8,000. Crystals diffracted to 4.5 Å and where found to process $P2_12_12_1$ symmetry with the unit cell dimensions a = 73.8, b = 86.0, c = 264.0 Å, α= 90°, β= 90°, γ= 90°. Sequential rounds of optimization were carried out using a range of PEGs, different buffers (MOPS, MES and BTP) and salts; ultimately, diffraction-quality crystals were grown in 0.2 M NaCl, 25%(v/v) PEG-600, 0.1 M BTP pH 6.5 using the hanging-drop vapour diffusion method. Optimization yielded crystals, which diffracted to 2.1 Å with $C2$ symmetry and unit cells dimensions a = 149.8, b = 71.9, c = 102.0 Å, α= 90°, β= 111.2°, γ= 90°.

The diffraction data were deemed sufficient to attempt to solve the phase problem. Solving the phases was undertaken experimentally as there was no sufficiently close homologue in the PDB suitable for successful molecular replacement. Structural determination was achieved by collecting X-ray diffraction data at the mercury absorbance peak, from a CCHFV N protein crystal soaked with thimerosal.

3.14.4 Structural comparison of the LASV NP

The DALI server revealed the CCHFV N globular domain displays a higher degree of similarity with the N-terminal RNA binding domain of the LASV NP, than to N protein within the *Bunyaviridae* family. Surprisingly, this suggests that CCHFV is more closely related to the arenavirus LASV (Figure 64A) and more distantly to other bunyaviruses. However the hypothesis that CCHFV and LASV are more closely related is not supported by a phylogenetic analysis of N sequences of sNSVs performed using the Dayoff model in MEGA6 (Figure 64B), which suggests nairoviruses and arenaviruses are distantly related. Phylogenetic analysis of partial nucleotides from L segment of sNSVs by Vieth et al (Vieth et al., 2007) performed using the PHYLIP 3.6 program package (Felsenstein, 2004), revealed a close relationship between nairoviruses and arenaviruses than to other bunyaviruses. Phylogenetic analysis using full length nucleotide sequences from the L segment performed by Muscle alignment and the Dayoff model in MEGA6 (Figure 63C) and ClustalW and ClustalX (Carter et al., 2012b) gave similar results, with trees showing bunyaviruses and arenaviruses as closely related based on the L protein. The phylogenetic analysis suggests the nairoviruses are more closely related to arenaviruses based on the L segment phylogeny, but are more closely related to members of its own family based on the S segment phylogeny (Figure 64B and C). The phylogentic analysis of the N protien suggests the nairovirus and areanavirus are not closely related but most likely emerged through a once common ancestor and the their structully similar N proteins have evolved through divergent evolution.

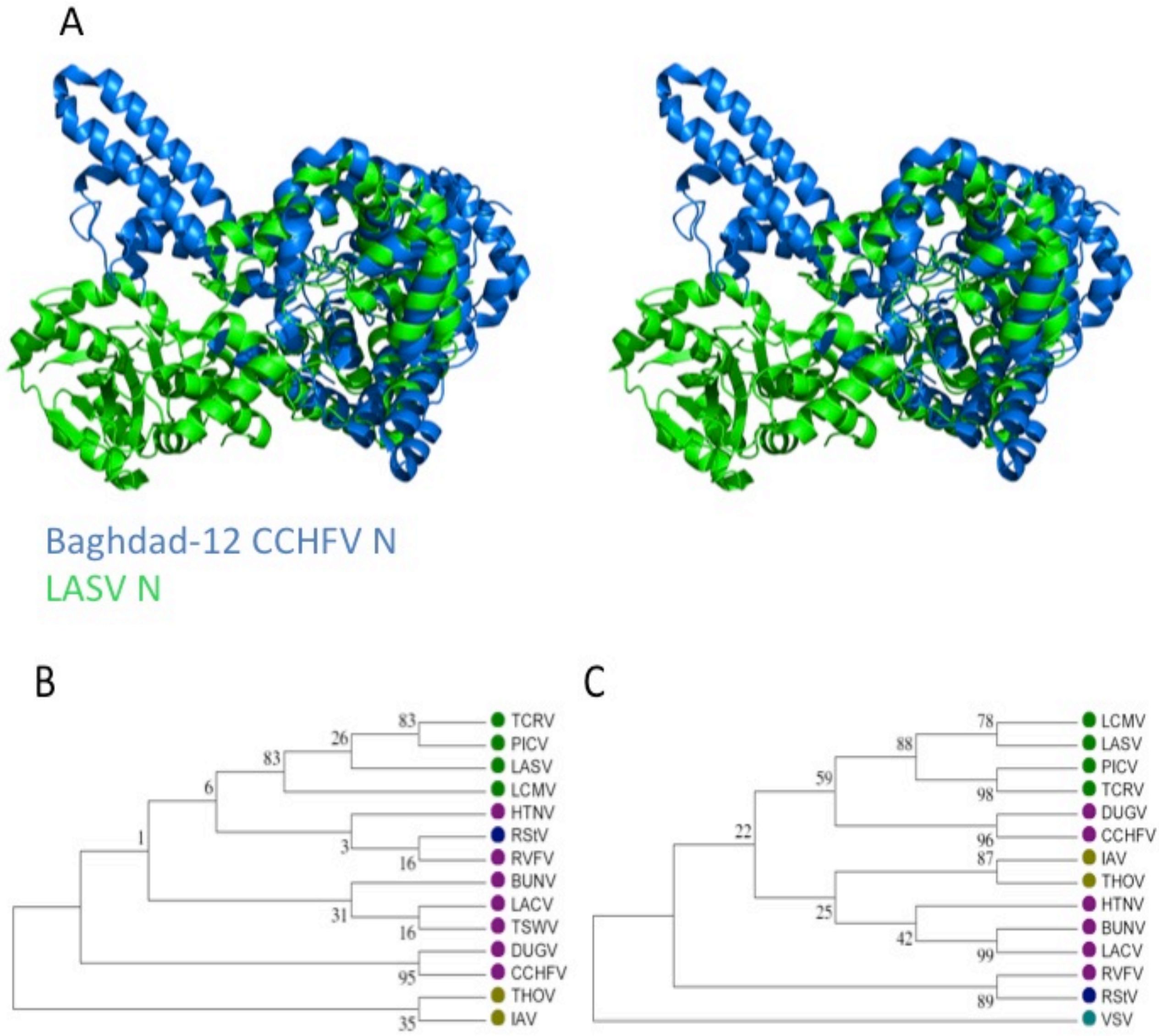


Figure 64. Superpositioning of the Baghdad-12 CCHFV N and LASV N/NP using the stereo cross eye and phylogenetic analysis of nsRNA viruses by N and L amino acid sequences using the MEGA6-Beta2 program. (A) Structural homology between that of CCHFV N and LASV NP crystal structures. A stereo cross eye of CCHFV N displayed in marine blue and the LASV NP shown in green. The figure was generated using PyMol. (B) and (C) The MEGA6-Beta2 program was used to align the amino acid sequences of 14 viral N and L proteins respectively, including those from seven (N protein phylogeny) and six (L protein phylogeny) bunyaviruses (labelled in purple), and four arenaviruses (labelled in green). Orthomyxoviruses are labelled in gold and the tenuivirus is labelled in blue. In the L protein phylogenetic plot (B), the sNSV rhabdovirus is labelled in light blue. The resultant phylogens were generated using M6 Tree Explorer. The percentage each branch was obtained is indicated at the base of each branch.

3.14.5 Structural comparison between CCHFV N proteins from other strains

The structures of the CCHFV N protein monomer of strain YL04057 (Guo et al., 2012) and IbAr10200 *strain (Wang et al., 2012) were published recently. Guo et al (Guo et al., 2012), revealed that in vitro CCHFV N protein has weak RNA binding affinity and has DNA endonuclease activity with distinct metal dependence. Wang et al (Wang et al., 2012), solved the CCHFV N structure in two conformations, which they suggest resembles the oligomerized RNP form and monomeric form respectively.*

3.14.6 Structural comparison of CCHFV N proteins from strains Baghdad-12 and YL04057

Superimposition of the two domains from strains YL04057 (Guo et al., 2012) and Baghdad-12 (Carter et al., 2012b) revealed a high degree of similarity in their folds, with a Cα RMSD of 0.975 Å for 357 residues with 92.16% sequence identity for the globular domain, and a Cα RMSD of 0.642 Å, for 99 residues, with 77.78% identity, for the arm domain (Figure 65A).

Superpositioning of the entire molecule revealed the conformation of the arm significantly changed position, with respect to the globular domain. The conformational change has caused the arm to rotate about 180 degrees and the apex of the loop to shift by a distance of 39.59 Å (at the Cα position of Asp266, at the apex of the arm, in each protein) (Figure 65A). Closer inspection at the base of this region has shown that the helix supporting the arm domain in strain YL04057 has twisted and divided into two shorter helices (α17 and α18) in strain Baghdad-12 (Figure 65A).

Alignments of the primary sequences of strains Baghdad-12 and YL04057 has revealed that there are 29 amino acid differences among 481 positions (Figure 65B). Interestingly the amino acid sequences at the division of helices 17 and 18 are identical, which argues against strain differences being the major determining influence in the change of conformation in the arm position (Figure 65B).

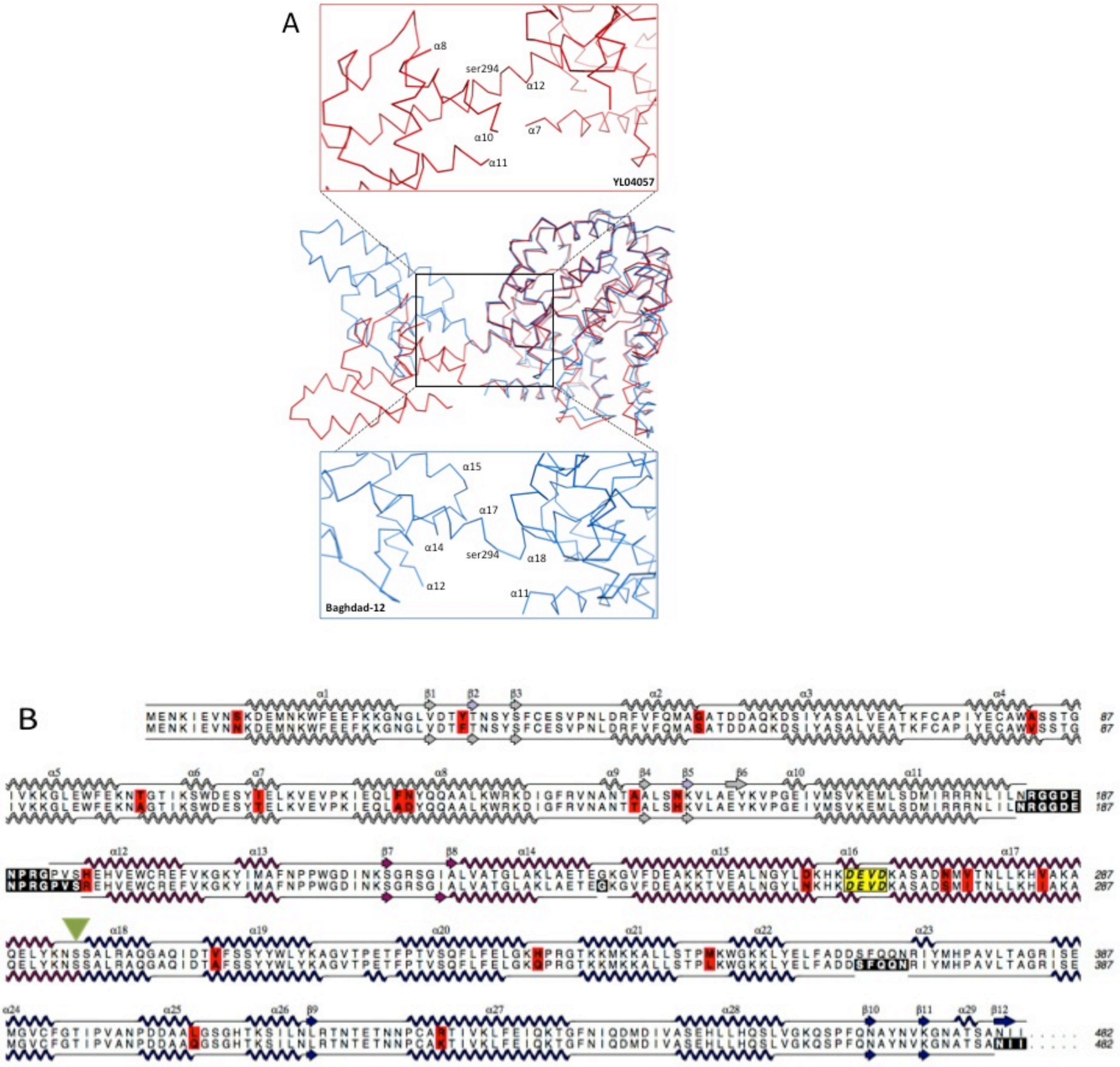


Figure 65. (A) Superpositioning of the Baghdad-12 N and YL04057 N protein globular domains. Superposition (center) of the Baghdad-12 N structure (marine blue) and the YL04057 N structure (red) globular domains reveals the arm has a radically different conformation. Each region between the arm and globular domain for each strain has been expanded, with relevant helices labelled. The helices for Baghdad-12 are numbered as shown in panel B, whereas helices for YL04057 are numbered according to the work of Guo et al (Guo et al., 2012). (B) Secondary structure represented schematically for both strains of CCHFV are matched with their respective primary amino acid sequences of the N protein. ALINE (Bond and Schuttelkopf, 2009), was used to create this figure. Amino acid differences between strains are highlighted in red, residues unassigned electron density are highlighted in black, and twist between helices 17 and 18 (at Ser294) is labelled with a green arrow, and the DEVD motif is highlighted in yellow. Panel A was generated using PyMol.

3.14.7 Structural comparison of CCHFV N proteins from strains Baghdad-12 and IbAr10200

The overall architecture of the globular body and stalk from the CCHFV N Baghdad-12 strain from *C*2 crystals is similar to that of the recently published CCHFV N IbAr10200 (Figure 66). Superposition of the globular domains of the Baghdad-12 and the monomeric CCHFV N protein from the IbAr10200 strain, revealed an average Cα RMSD of 0.559, with an overall similarity of 95.4% in amino acid sequence. Superimposition of the arm domains of the Baghdad-12 and the monomeric CCHFV N protein from the IbAr10200 strain, revealed an average Cα RMSD of 0.732, with an overall similarity of 94.5% in amino acid sequence similarity. Overall the sequence identity between the CCHFV N proteins from Baghdad-12 and IbAr10200 is 95.2% similar.

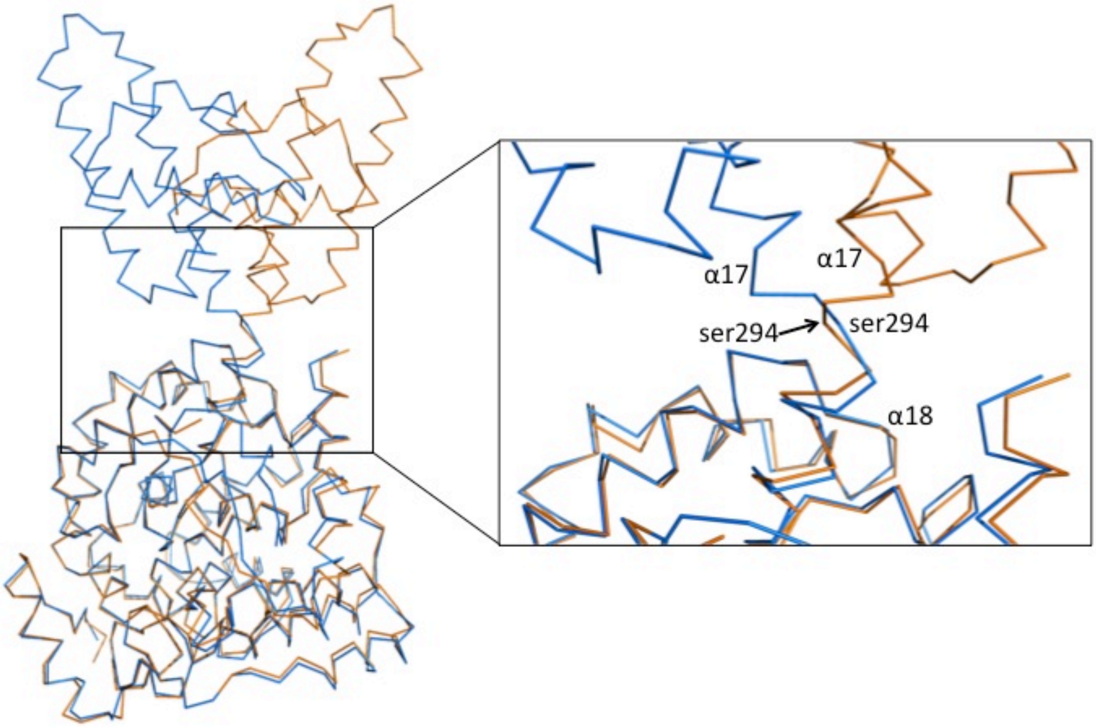


Figure 66. Superpositioning of the crystal structure of the CCHFV N protein from Baghdad-12 (marine blue) and the monomer form of N from the IbAr10200 strain (orange). Representation in ribbon format. An expanded view of the region between the arm and the globular domain from each strain is shown in the black box, with relevant helices labelled. The figure was generated using PyMol.

Superposition of the two entire CCHFV N proteins from the IbAr10200 strain and Baghdad-12 revealed they display different conformations in their stalk regions (Figure 66). Interestingly the conformational change between CCHFV N protein from Baghdad-12 and the monomeric CCHFV N protein from the IbAr10200 strain induces the nearly identical division of alpha helix 17 as described above (Figure 65A). The arm is rotated by about 180 degrees, and the apex of the loop is shifted by a distance of 40 Å (at the Cα position of Asp266, at the apex of the arm, in each protein) (Figure 65A and 66).

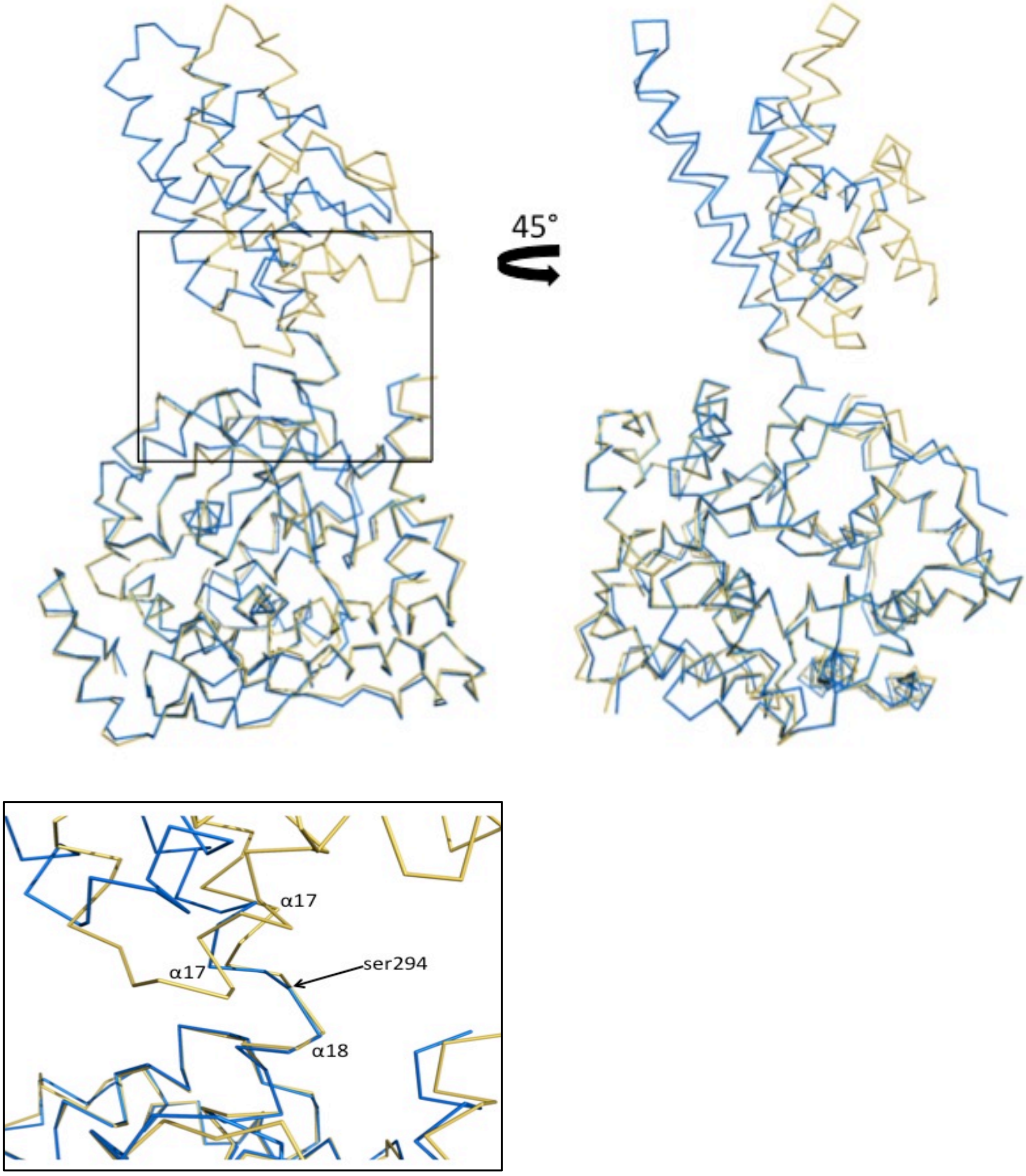


Figure 67. Superpositioning of the crystal structure of the CCHFV N protein from Baghdad-12 and the oligomerized form of N protein from the IbAr10200 strain. Representation in ribbon format Baghdad-12 (marine blue) and IbAr10200 (oligomer) (yellow) CCHFV N proteins. An expanded view of the region between the arm and the globular domain for each strain is shown in the black box, with relevant helices labelled. The figure was generated using PyMol.

Additionally superimposition of the oligomerized form of the IbAr10200 CCHFV N protein and the Baghdad-12 CCHFV N protein (Figure 67), revealed a less drastic conformational change and does not involve rotation and splitting of alpha helix 17 into two smaller helices, but instead the helix shifts its position 30 degrees into a more upright position (Figure 67).

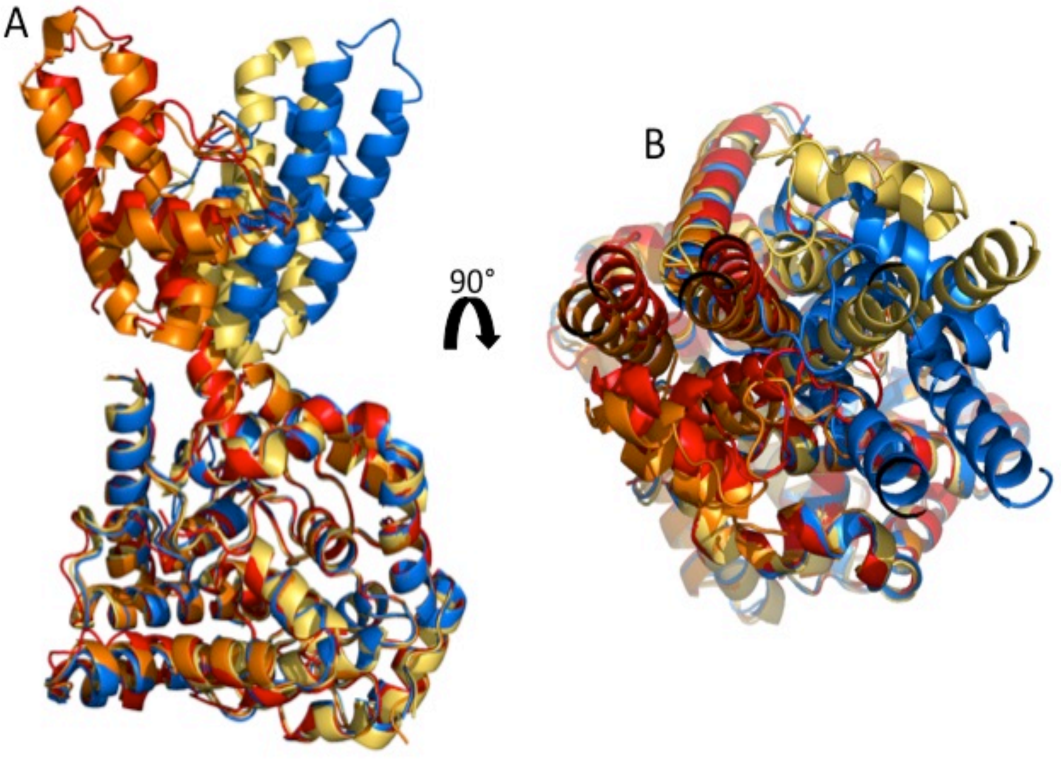


Figure 68. Conformational differences in the arm domain of the CCHFV N protein. (A) Superposition of CCHFV N from IbAr10200 (yellow, (oligomer)) and (orange, (monomer)), Baghdad-12 (marine blue) and YL04057 (red) strains in cartoon representation. (B) Top view of the aligned CCHFV N protein structures. The figure was generated using PyMol.

The flexibility of the arm domain may have important consequences for activities in critical CCHFV N protein functions such as RNA binding or oligomerization. For example the two different arms positions taken up by the CCHFV N protein from strains Baghdad-12 and YL04057 may represent interchangeable forms involved in RNA binding, or oligomerization or perhaps are dictated by the different primary sequences of their respective strains (Figure 65B).

However as there are no differences in sequence within the helix where the conformational change occurs it is unlikely that strain differences account for the range of positions seen between the N proteins. Furthermore the two positions of the two arms and conformation of each helix linking each to the globular domains displayed in strain IbAr10200 correspond very closely to the alternating conformations of the other two strains respectively. Therefore the single α-helix linking the globular and arm domains is unlikely to provide a rigid link and is able to adopt a number of conformations, independent of the strain sequence (Figure 68).

Although there is the possibility that these conformational differences may have been forced by crystal packing, it is likely that they are also reflected in solution, and have attained their position through the bacterial expression process. One consequence of the shifted arm position is an altered electrostatic surface potential, which may affect the RNA binding ability of the respective proteins. Such a shift in conformation may only be stimulated through the presence of RNA or as yet an unknown factor, similar to the proposed gating mechanism of the LASV NP (Figure 23). One possible model for RNA binding could involve the arm representing part of a gating mechanism allowing a switch between RNA-bound and unbound states. Furthermore it is also plausible that the switch in arm position is responsible for conversion of monomeric N into higher-order complexes, which are required for RNP assembly.

3.14.8 Comparison of electrostatic potential analysis of the CCHFV N proteins from Baghdad-12, YL04057 and IbAr10200 strains

Comparison of the electrostatic potential of the CCHFV N proteins from strains Baghdad-12, YL04057 and IbAr10200 strains revealed differences in the surface areas of the main putative RNA binding platform (Figure 69A, D and E). Analysis of the electrostatic surface potential the CCHFV N protein of YL04057 suggests the change in conformation of the arm in this protein has changed the shape of the positively charged RNA binding domain. The conformation of the arm also noticeably alters the electrostatic surface potential of the monomer and oligomerized forms of the CCHFV N protein from IbAr10200 strains, which argues against sequences differences altering the positive charge surface between the YL04057 and Baghdad-12, and suggests the conformational change of the arm is the main contributor behind the change in electrostatic potential.

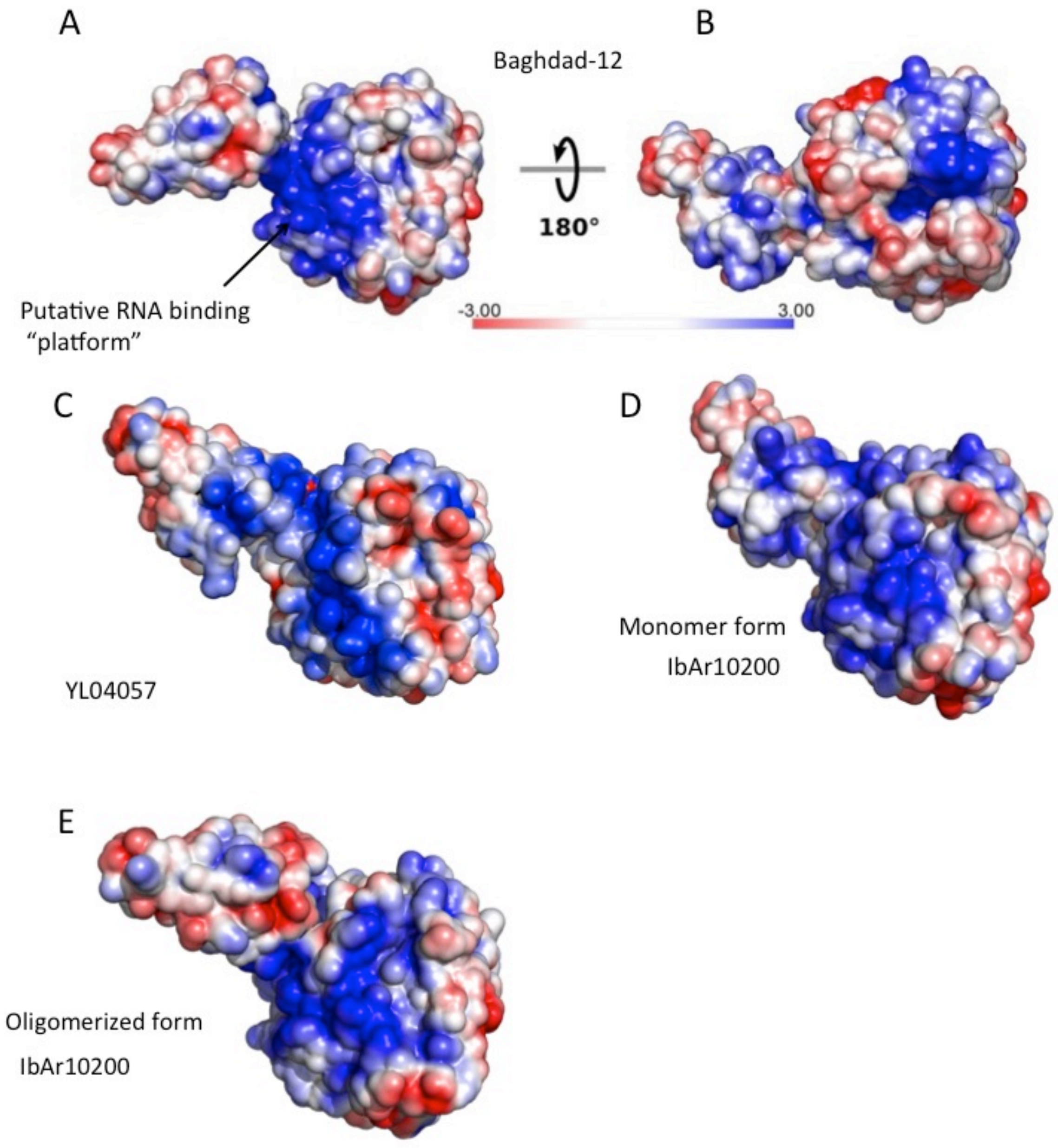


Figure 69. Electrostatic potential analysis of the CCHFV N protein from Baghdad-12, YL04057 and IbAr10200 strains. (A) The electrostatic surface potential of N protein from Baghdad-12 showing the positively charged RNA binding "platform", solely located within the surface of the globular domain (Unni et al., 2011). (B) Rotation by 180 degrees showing the smaller additional positively charged groove, termed the "pocket." (C) The electrostatic surface potential of N protein from Yl04057. (D) The electrostatic surface potential of the monomer form of N protein from IbAr10200. (E) The electrostatic surface potential of the oligomer form of N protein from IbAr10200. The scale bar shows the contour levels for the electrostatic potential at the solvent-exposed surface, in kT/e. The figure was generated using APBS in PyMol (Unni et al., 2011).

3.14.9 Functional analysis of selected N protein residues by use of a CCHFV minigenome system performed by Dr. Eric Bergeron, Dr. Cheryl Walter and Rebecca Surtees

The central region of the protein protrudes from the globular domain and resembles an arm. The positively charged exposed surfaces identified on the exterior of the globular domain guided mutagenesis and revealed the strict requirement of several residues on these surfaces for CCHFV gene expression, suggesting involvement in RNA binding.

To characterize the role of the CCHFV N protein residues in both multimerization and RNA binding, the CCHFV minigenome system was used to examine a panel of CCHFV N proteins with various mutations within the proposed RNA binding surfaces (Figure 70), caspase-3 protease cleavage site and the potential dimer interface, for their ability to promote CCHFV-specific RNA synthesis. The CCHFV N mini-genome system utilizes a luciferase based reporter gene expression, which is promoted by formation of RNP templates from RNA and N protein, which can be measured in mammalian cells. Expression of the luciferase reporter gene system in the minigenome system (Figure 71) depends on the CCHFV N protein's ability to bind RNA, interact with itself to encapsidate and replicate RNA; thus, any drop in genome activity likely results from changes in either RNA binding or multimerization activities (Figure 71C). Expression of altered N proteins was examined by western blotting from equivalent quantities of cells expressing the minigenome system (Figure 71D).

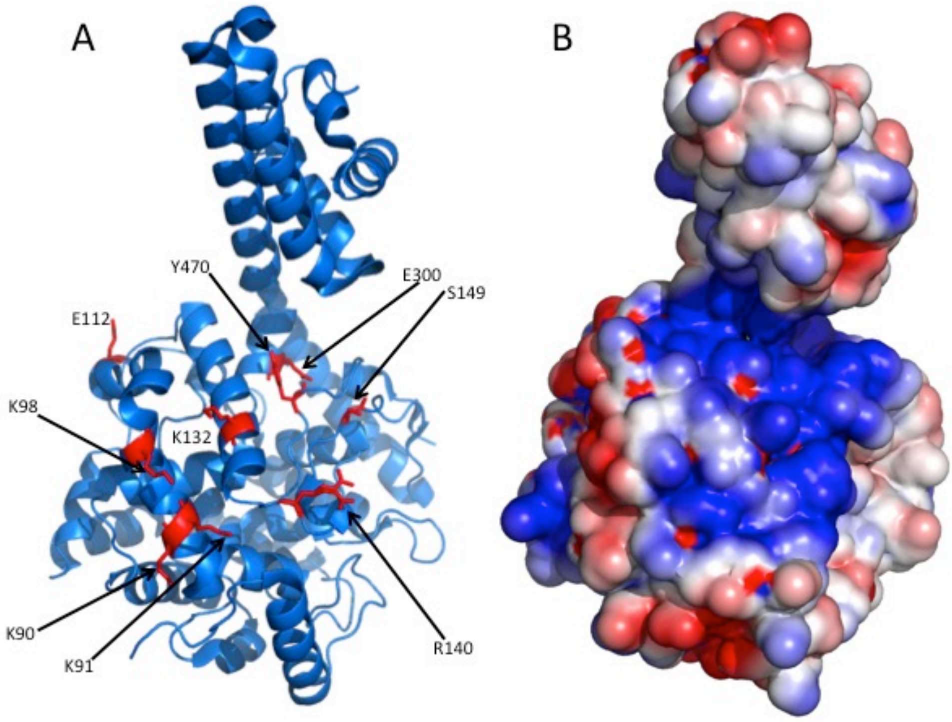


Figure 70. Identifying important RNA binding residues for analysis using the minigenome system. (A) Cartoon representation of Baghdad-12 CCHFV N crystal structure. Residues coloured red and labelled are those chosen based on the electrostatic surface potential. (B) The electrostatic surface potential of CCHFV N from Baghdad-12 (generated with APBS), displaying the RNA binding “platform”. The figure was generated using APBS in PyMol (Unni et al., 2011).

Expansion of the putative RNA binding “platform” using electrostatic surface potential analysis identified nine residues (Figure 70), which were mutated to alanine and analysed using the mini genome system. The genome activity of six of these mutants was unchanged, (K91A, K98A, E112A, R140A, S149A, and Y470A); however, the three remaining mutants showed genome activity that was either considerably reduced (K90A) or effectively abolished (K132A and Q300A) (Figure 71C). Of those mutated positively charged residues within the RNA binding pocket, the H453A, K342A and K343A mutants displayed almost wild-type activity. The H456A and K411A mutants on the other hand resulted in significantly reduced or abolished RNP activity, respectively (Figure 71B). Attempts to analyse residues E387A, I304G, and W313A mutants, located in the pocket, using the mini-genome system failed because they were poorly expressed or undetectable in transfected cells (Figure 71D).

Collectively the results of this analysis recognised three separate residues (K132, Q300, and K411) that are indispensible for mini-genome activity and another two residues (K90 and H456) whose mutation reduced the functionality of the mini genome system. The charge characteristics of these residues and their location on the CCHFV N protein surface, point to their activity being involved in RNA binding.

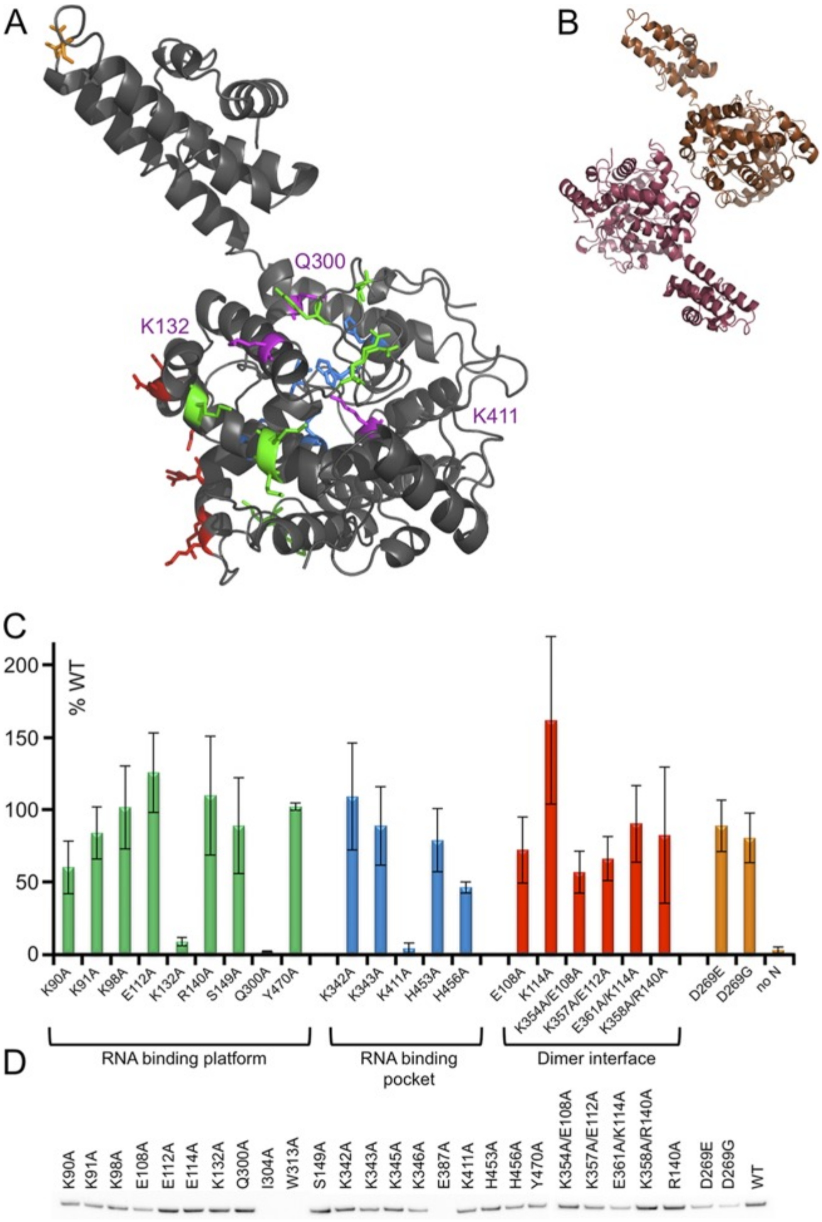


Figure 71. *In vivo* analysis of site-directed CCHFV N protein mutants. (A) CCHFV N protein in cartoon representation, highlighted amino acids selected for mutation and analyse in the minigenome system. Residues highlighted in green represent the putative RNA binding located on the globular domain, residues highlighted in blue comprise the RNA binding pocket, and D269 of the DEVD motif is shown in orange, and residues shown in red represent the dimer interface. Residues K132, Q300, and K411 abolished minigenome activity and are highlighted in magenta. (B) Cartoon representation of the CCHFV N protein dimer present in both crystal forms, investigated using the minigenome system, suggests a putative dimer interface. (C) Bar chart coloured as described above showing reporter gene expression for mutants. The height of each bar is the percentage of that of wild type (WT) N. CCHFV M minigenome *Renilla* luciferase activities were normalized over the control firefly luciferase activity and are reported as percentages of wild-type N activity. Error bars represent the standard deviation from 3 independent experiments. (D) SDS-PAGE followed by Western blot analysis using a polyclonal antibody raised against the CCHFV N protein to assess respective expression levels of N protein mutants. Panels A and B were generated using PyMol

The non-crystallographic dimer interface (Figure 71B), was also investigated using the mini genome system (Figure 71B) and N proteins bearing mutations within this interface were generated using both single-residue mutants (E108A and K114A mutants) and double-residue mutants (K354A/E108A, K357A/E112A, K358A/R140A, and E361K/K114A mutants) (Figure 71A, red residues). Not one of the single mutants or double mutants abrogated the activity of the mini-genome system, which suggests these residues are likely not essential for CCHFV N protein oligomerization (Figure 71B and C).

The double mutants K354A/E108A and K357/E112A did produce a reduction in RNP activity and perhaps these residues correspond to roles in RNP assembly, possibly by promoting oligomerization (Figure 71C). Furthermore the mini-genome system was used to explore the mutants with altered DEVD motifs (DEVG and DEVE), which are not cleavable by caspase-3 protease. None of these mutants abrogated the RNP formation and the mini-genome activity was indistinguishable from that of wild-type N (Figure 71C).

3.14.10 Potential for a gated RNA binding mechanism.

The electrostatic charge distribution on the CCHFV N surface revealed two possible RNA binding regions: the platform and the pocket. Superpositioning of globular domains of CCHFV N from strains Baghdad-12, YL04057, and IbAr10200, suggest that the arm is free to take-up a range of conformations, which could significantly alter surface electrostatics and therefore RNA binding (Figure 69A, B, C, D and E). The mobility of the arm could be allowed by the flexibility of the loop between the arm and the globular domain (residues 183 to 191; not visible in either the YL04057, IbAr10200 and Baghdad-12 N protein crystal structures), and perhaps allows the arm domain to form a "gate" in an RNA binding mechanism, similar to the LASV NP and RVFV N gates (Figure 72).

The proposed RNA gating mechanism for LASV NP involves helices α5 and α6, which are repositioned to allow the binding of RNA in a previously unexposed surface (Hastie et al., 2011b). Interestingly the position of helix α5 in LASV NP corresponds to the beginning of the flexible loop leading to the CCHFV arm domain, suggesting CCHFV N may operate in a similar mechanism for RNA binding, and may involve the arm domain taking part in conformational switching between open and closed states (Figure 72).

Structural alignment has revealed that the RNA binding domain of the LASV NP, initially thought to be a cap-binding pocket, corresponds to the CCHFV N protein positively charged pocket, which has been identified as an active site for endonuclease catalysis by Guo et al (Guo et al., 2012, Qi et al., 2010, Hastie et al., 2011b). The finding that the CCHFV N protein RNA binding platform does not correspond to the RNA binding domain within the LASV NP, but instead aligns with the positively charge pocket, suggest three possibilities. Firstly the CCHFV N protein could bind RNA via two surfaces (platform and pocket) or secondly that conformational rearrangements create a continuous RNA binding surface. Thirdly RNA binding may occur only via only one RNA binding surface, namely the platform, and the positively charged pocket facilitates the binding of DNA for endonuclease catalysis, in an unknown role yet to be uncovered.

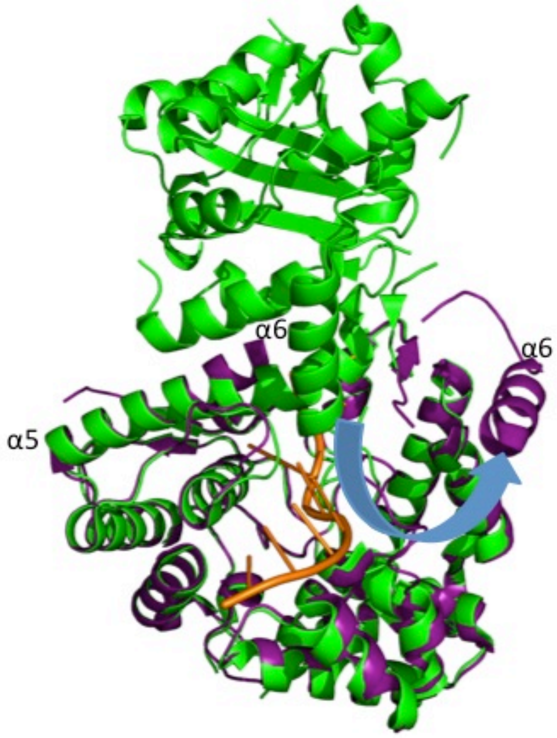


Figure 72. RNA gated binding mechanism for LASV NP. The superimposition of two separate crystal structures of the LASV NP: N-terminal coloured in purple (amino acids 8-336, PDB code: 3T5Q), and N and C-termini coloured in green (amino acids 7-569, PDB code: 3MWT), shows the alpha helix six in two conformations. The change in conformation of alpha helix six (blue arrow), within the crystal structure of the N-terminal of the LASV NP allowed the co-crystallization of a short ssRNA molecule into the proposed RNA binding site and shows the flexibility of this region which could potentially be involved in a potential RNA gating mechanism for RNA binding. The figure was generated using PyMol.

It could be possible that the RNA binding surfaces are not presented in the apo crystal structures and in order to reveal them a combination of arm movement and other structural rearrangements may be involved. The concept of the arm adopting different positions and allowing for structural rearrangement has been supported by comparison with the apo crystal structures of CCHFV N proteins from three different strains of CCHFV (Guo et al., 2012, Wang et al., 2012, Carter et al., 2012b). Superpositioning of the CCHFV N protein from strains IbAr10200, Baghdad-12 and YL04057, revealed that while the globular domains align very similarly, the arm domains assume very different positions (Figure 68).

3.14.11 CCHFV RNP architecture

The crystal structures of the CCHFV N protein from strain IbAr10200 supported the hypothesis that the flexibility of the arm domain is an important feature, which allows the N protein to function in important activities such as RNA binding and oligomerization to assemble the RNP (Wang et al., 2012). The CCHFV N protein crystal structure from Wang et al (Wang et al., 2012) revealed that running through the crystal along the crystallographic *c* axis was an antiparallel double superhelix of nine molecules per turn per helix (Figure 73). One fundamental difference between the double helix arrangement of CCHFV N protein and the architecture of N protein seen in the phleboviruses, orthobunyaviruses, and nsNSVs such RSV and VSV, is the lack of either N- and C-terminal extensions contributing in binding to adjacent N proteins. Instead the arm domain and areas of the central portion of the globular domain drive the N-N interaction in the Wang et al superhelix homo-oligomer structure.

The homo-oligomer arrangement of CCHFV N protein in the proposed RNP has a similar arrangement with the N protein structure of the IAV RNP complex visualized using ECT. Furthermore it is not uncommon for monomeric N protein to crystallize into an oligomerized form that is biologically relevant (Ferron et al., 2011), and it may be that the N protein/RNA complex is required for oligomerization to occur in solution.

Oligomerization of one CCHFV N protein helix in this model is driven by the interaction of the stalk with the base of the globular body of the adjacent protomer. Although the degree of bending within the helix means the contacts between the protomer are slightly different, all the interactions between the stalk and body involve extensive hydrogen bonds and van der Waals contacts between the same residues. In particular, the interactions observed between main chain residues 352 to 354 of the α17-α18 loop of the globular domain and residues 266 to 269 at the α12-α13 loop in the arm domain of the adjacent molecule, were previously identified as the caspase-3 protease cleavage site. None of these interactions are seen in the Baghdad-12 *C*2, or $P2_12_12_1$ CCHFV N protein crystal structures and analysis of the residues involved in the dimer interface seen in both forms argues against this being a functional element within the RNP, however an intermediate of RNP assembly can not be ruled out.

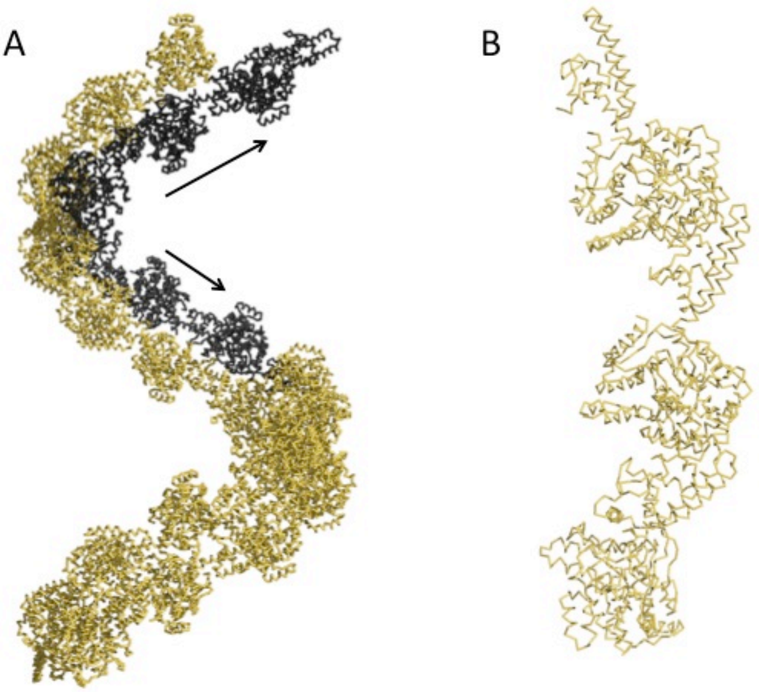


Figure 73. Crystal structure of the CCHFV N protein from strain IbAr10200 showing homo-oligomerization. (A) The antiparallel double superhelix of CCHFV N protein. The double helix is coloured yelloworange, except for three CCHFV N protomers in the antiparallel strand, which are coloured black. The two arrows point towards each direction of each antiparallel CCHFV N protein strand. (B) AU of the $P3_1$ axis. Adapted from (Wang et al., 2012). The figure was generated using PyMol.

3.14.12 PISA analysis of dimer interface in *C*2 and $P2_12_12_1$ crystal structures

The PISA results suggest that even the largest interface between the two CCHFV N protein in the AU of the Baghdad-12 *C*2 and $P2_12_12_1$ crystal structures are not sufficient to generate a stable interface in solution. For comparison, the oligomeric interfaces present in the RVFV N protein structure have an average buried surface area of 1,643 Å^2, a ΔG value of −18.5 kcal/mol, and a CSS of 1.0. However, lattice contacts within the same RVFV N protein crystal form present an average buried surface area of 437 Å^2 and a ΔG value of −3.9 kcal/mol; thus, the CCHFV N interface was scored at values intermediate between the two. The PISA analysis therefore suggest this interface may not represent part of the authentic oligomerization domain, and this is also supported by there being no similar interface in the homo-oligomer arrangement of CCHFV N protein proposed by Wang et al (Wang et al., 2012).

3.14.13 Conformational change in Wang et al CCHFV N protein crystal structures

To further understand the RNA binding sites of the CCHFV N protein Wang et al attempted to co-crystallize the N protein with ssRNA of primer-lengths 12 to 16 nucleotides long, and unexpectedly crystallized the CCHFV N protein with only one CCHFV N protein in the AU (monomer) and without additional electron density resembling bound RNA.

The structure revealed a dramatic shift in the conformation in the arm domain of the CCHFV N protein, upon incubation with primer-length ssRNAs. In particular the rotation of helices of α13 and α14 into a continuous long α-helix at the hinge S294, resulting in the rotation of the stalk domain by ~168°. The same alternative conformation can be seen between CCHFV N protein from strains Baghdad-12 and YL04057 at exactly the same position, S294. This rotation of the stalk domain was proposed to disrupt the head to tail interaction via the α12-α13 loop at the apex of the CCHFV N protein and also cause steric hindrance between the antiparallel dimers in the superhelical oligomer structure (Figure 73).

3.14.14 Caspase-3 cleavage of CCHFV N protein

Wang et al revealed the DEVD site was located directly between the head-to-tail interactions of two CCHFV N proteins, and this location obscures binding and subsequent cleavage by the caspase-3 protease. Wang et al also reported that the monomeric form of the CCHFV N protein is more susceptible to cleavage upon adding a short ssRNA mimicking a primer to the oligomeric form of CCHFV N protein. Interestingly the work in this thesis agrees with Wang et al, and shows that Baghdad-12 CCHFV N protein structure displays a similar conformation to the oligomeric form of IbAr10200 strain N protein, and is also a poor substrate, although susceptible to cleavage by caspase-3 protease over many hours. Unfortunately Wang et al (Wang et al., 2012) did not reveal how long their CCHFV N protein was incubated with caspase-3 protease.

As a result, Wang et al proposed the monomeric CCHFV N will be released from the ends of the double superhelical CCHFV N polymer, allowing the exposure of the caspase-3 protease cleavage site situated at the loop region at the apex of the arm domain to activated caspase-3 protease. The reason for the functional link between the apex of the arm domain becoming accessible to caspase-3 protease upon the presence of primer length RNA is uncertain. It is also uncertain whether the homo-oligomers seen in the CCHFV N crystal structure for strain IbAr10200 are biologically representative of authentic CCHFV RNPs from infected cells or alternatively an artefact of crystal packing. Directly visualizing infectious CCHFV RNPs using EM may provide clues into the assembly of the CCHFV N protein, like that which has already been achieved for the RNPs of orthobunyaviruses and phleboviruses (Ferron et al., 2011, Ariza et al., 2013, Li et al., 2013, Niu et al., 2013, Reguera et al., 2013, Dong et al., 2013b).

It was previously shown that stimulus of apoptosis and depletion of full length CCHFV N by caspase-3 protease cleavage via the DEVD motif, which is conserved across all strains of CCHFV would be detrimental to the replication cycle of the virus in host cells, and therefore represents a host cell mechanism against CCHFV infection (Karlberg et al., 2011). However the striking feature of the CCHFV N protein structure is the accessibility of the DEVD motif situated within the far end of the arm domain (Figure 54). The exposed position of the DEVD motif of the monomer CCHFV N protein, and not being exposed within the RNP, together with its strict conservation in all CCHFV strains, suggests that cleavage of the CCHFV N protein is somehow beneficial to the virus, and not a host cell mechanism to protect the cellular environment from viral replication. Surely if cleavage by caspase-3 protease were beneficial to the host cell, then the DEVD would be quickly selected out of the fast mutating CCHFV genome through viral evolution. However importantly the DEVD motif could be maintained by other more important aspects of the viral life cycle and interestingly the DEVD motif has been shown to be involved in taking part in interaction between monomers within the homo-oligomer helix model, which may have important implications for viral RNA synthesis, such as replication and transcription. Currently there is no evidence to suggest how much CCHFV N protein is free and not RNP associated within a cell.

The structural significance of the caspase-3 protease cleavage site within the CCHFV N protein was investigated to provide clues into the functional importance of this event during CCHFV infection. Therefore the fate of the CCHFV N protein fragments following cleavage was investigated, and performed by observing caspase-3 protease cleavage of the N protein *in vitro*. It was demonstrated that upon cleavage of the CCHFV N by recombinant caspase-3 protease the cleavage products remained closely associated as a single complex, perhaps resembling the native CCHFV N protein. The cleavage of the N protein could however induce altered tertiary or quaternary structure, which could have functional consequences, including the adoption of different oligomeric states or alterations in RNA binding affinity and interactions with viral/host cell protein partners.

One proposed reason for the caspase-3 protease cleavage of CCHFV N protein could involve reducing the overall rate of activated caspase-3 protease cleavage of cellular substrates during the onset of apoptosis by acting as a decoy. By the CCHFV N protein acting as an alternative caspase-3 protease substrate it could prevent the downstream-mediated signaling pathways, including apoptosis and the interferon response, with important consequences for viral growth. Interestingly Wolff et al (Wolff et al., 2013) have recently demonstrated that infection with Junin virus (JUNV) (arenavirus) in cell culture leads to cleavage of the viral NP into discrete products resembling caspase-3 protease cleavage of the CCHFV N protein. Although in contrast JUNV infection does not induce apoptosis in cells, the JUNV NP has been shown importantly to function as a decoy substrate for caspase-3 protease cleavage in order to inhibit the onset apoptosis. The absence of apoptosis in JUNV infected cells could reflect the ability of the NP to be cleaved through two or possibly more identified caspase cleavage motifs, and it is therefore more potent at preventing apoptosis than the CCHFV N protein. Interestingly the purified JUNV particles have been shown to incorporate both the 47 kDa and 40 kDa cleavage products of the NP, whereas no cleaved N protein is present in CCHFV (Bergeron et al., 2007a), and suggests these truncated forms of JUNV NP may play additional roles in viral assembly or preventing apoptosis.

3.14.15 CCHFV N protein dsDNA specific endonuclease

Guo et al (Guo et al., 2012), reported that the CCHFV N protein displayed dsDNA specific endonuclease activity, located within the globular domain, within the pocket of negatively charged amino acids identified using APBS (Unni et al., 2011).

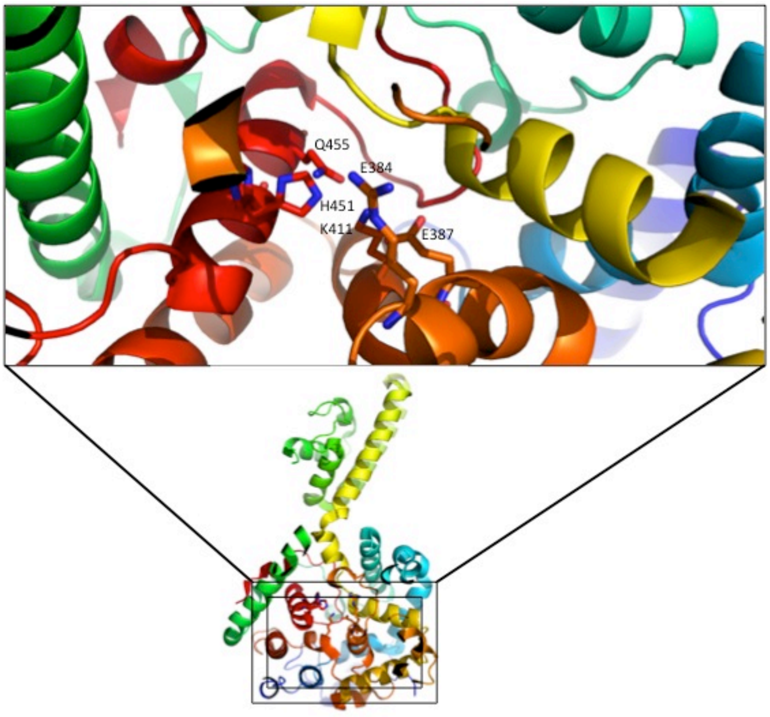


Figure 74. CCHFV N protein putative dsDNA endonuclease active site. Residues located in the globular domain whose hydrophilic side chains are near the proposed small RNA binding "pocket" domain (Figure 69B). The figure was generated using PyMol. The figure was adapted from Guo et al (Guo et al., 2012).

The authors reported that endonuclease activity was specific for dsDNA and not RNA, was inhibited by EDTA, and active in the presence of divalent cations, including Co^{2+}, Mg^{2+}, and Mn^{2+}. Guo et al also identified a putative active site, which consisted five amino acids, Q455, H451, E384, E387, and K411 (Figure 74), which when mutated to alanine abrogated endonuclease activity.

The proposed site for dsDNA endonuclease activity of CCHFV N, argues against the positively charge pocket being involved in RNA binding, and points to an as yet unidentified function (Figure 74). The endonuclease activity could be an artifact of *in vitro* analysis, and analysis *in vivo* may reveal the CCHFV N protein acts as an RNA endonuclease, which would make more sense, as the CCHFV N protein does not locate to the nucleus, but resides in the cytoplasm. In this work residue K411 favors a direct role in CCHFV gene expression and not the DNase activity proposed by Guo et al (Guo et al., 2012).

3.14.16 Possible roles of the CCHFV N endonuclease

If the endonuclease activity of the CCHFV N protein is real and important for the life cycle of the virus, it could play a similar role to the LASV NP. The LASV NP which has C-terminal RNA 3' to 5' exoribonuclease activity, has been shown to suppress interferon induced innate immunity (Hastie et al., 2011a, Qi et al., 2010). Qi et al proposed the LASV NP prevents the virus-induced IFN induction by degrading the PAMP RNAs that otherwise would trigger the viral sensors in the cells (Qi et al., 2010). Qi et al also reported that the LASV NP N-terminal domain could digest both DNA and RNA substrates. It has been shown both LASV and CCHFV antagonize IRF-3 activation pathway leading to the host IFN response to infection (Akutsu et al., 2011, Kinsella et al., 2004, Martinez-Sobrido et al., 2007), and evidence suggests they use different mechanisms. In contrast to immune suppression ability of the LASV NP, the CCHFV N protein has not been shown to have any significant effect on Sendai virus-induced activation of the IFN-β and IFN-stimulated response (Guo et al., 2012). The possibility that CCHFV protein has IFN suppression ability using a different mechanism to the LASV NP and perhaps uses DNA endonuclease activity to perform this cannot be ruled out. An example of DNA exonuclease suppressing production of IFN has been demonstrated for the human TREX1 protein. The human TREX1 protein has been shown to degrade small single stranded DNAs and dsDNAs, which accumulate during cellular apoptosis. Failure to remove DNA fragments by defective TREX1 has been reported to lead to cellular DNA receptor activation, triggering a persistent production of IFNs, which has been linked to some human autoimmune diseases (de Silva et al., 2007, Stetson et al., 2008, Crow and Rehwinkel, 2009).

4. Chapter 4 - TSWV N protein expression, purification and crystallization

4.1 Introduction

- TSWV is a member of the tospovirus genus, which is the only plant-infecting genus of the *Bunyaviridae* family. Over 800 plant species in 80 plant families are susceptible to TSWV and infection rates of 50-90% lead to major losses in global commercial vegetable crops. Major crops susceptible to TSWV infection are tomato, pepper, lettuce, potato, papaya, peanut and tobacco.

- There is currently no high resolution structural information for any TSWV encoded protein to improve our knowledge of important aspects of the virus life cycle, such RNA replication and interaction with the host cell.

The aim of this project was to attain high resolution structural information on the TSWV N protein using X-ray crystallography. The first stage aimed to express and purify highly concentrated TSWV N protein suitable for crystallization trials. This was firstly achieved by finding a suitable recombinant expression tag and *E. coli* strain in which to produce abundant and soluble TSWV N protein. Next Ni-Superflow Plus resin using His_6 trap methodology was used to purify the TSWV N recombinant protein away from *E. coli* contaminants. Whilst SEC was employed to look at the oligomeric state of the native TSWV N protein, it was also used to purify the TSWV N protein into a homogeneous state suitable for crystallization trials. Crucially for solving the atomic structure of the TSWV N protein, the project required growing reproducible TSWV N native protein crystals, which diffracted to high resolution, but also required a means to obtain phases of the diffracted X-rays.

4.2 Cloning

Based on the successful expression and purification of the CCHFV N protein described in chapter 3, a cDNA ORF representing the TSWV N protein was cloned into the pET28a-SUMO plasmid vector with the aim of expressing a His$_6$-SUMO-TSWV N protein.

Briefly, PCR amplification of the TSWV N cDNA within the pET15b plasmid was performed using primers (appendix) designed to introducte *BamHI* and *XhoI* sites at the 5' and 3' ends respectively, and would flank the newly synthesized TSWV cDNA ORF. This new cDNA fragment was then inserted into the corresponding restriction sites within the pET28a-SUMO multiple cloning site, so that when expressed the TSWV N protein would be fused to a His$_6$-SUMO tag via the N-terminus. The tag would subsequently be cleaved to yield the whole TSWV N protein with a non-native serine left on the N-terminus.

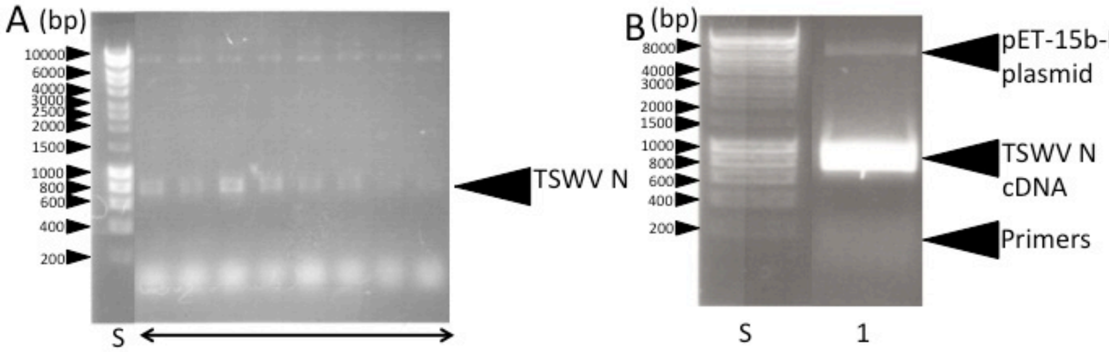


Figure 75. PCR amplification of the TSWV N cDNA from a pET-15b plasmid construct. (A) 1% agarose gel electrophoresis analysis PCR amplification of the TSWV N cDNA. The double pointed arrow underlines lanes analyzing gradient PCR reactions. (B) Amplified TSWV cDNA were pooled together, gel purified and cleaned up, ready for ligation with *BamHI* and *XhoI* double digested pET28a-SUMO. Molecular weight standards (Bioline) (Lane S). Combined PCR amplification of the TSWV N cDNA in A (Lane 1).

Successful amplification of the TSWV N cDNA by PCR was confirmed by analysis using 1% agarose gel electrophoresis, which revealed a band corresponding to the expected size of 774 bp for the TSWV N gene (Figure 75A and B). *BamHI* and *XhoI* restriction enzymes were added to produce the respective sticky ends flanking the TSWV N cDNA required for ligation, the PCR product was cleaned up and concentrated (Zymo Research).

The double digested PCR amplified TSWV N cDNA was subsequently 1% agarose gel purified using a UV box, cleaned and concentrated to 30 ng μl^{-1} using a DNA gel cleaning kit (Qiagen). The TSWV N cDNA fragment was sufficient to be ligated to the double digested pET28a-SUMO plasmid. The PCR fragment containing the TSWV N gene was ligated successfully and DNA sequencing confirmed the un-mutated TSWV N ORF was in the correct position within the pET28a-SUMO vector.

4.3 Expression and purification of the TSWN N protein

4.3.1 Purification of TSWV N protein with Ni-Superflow Plus resin

The expression and purification of the His$_6$-SUMO TSWV N construct was initially tested using Rosetta2 cells and nickel resin using the spin protocol. The majority of the His$_6$-SUMO TSWV N fusion protein was successfully bound to the resin and subject to three washes producing highly pure protein, which eluted using 300 mM imidazole elution (Figure 76).

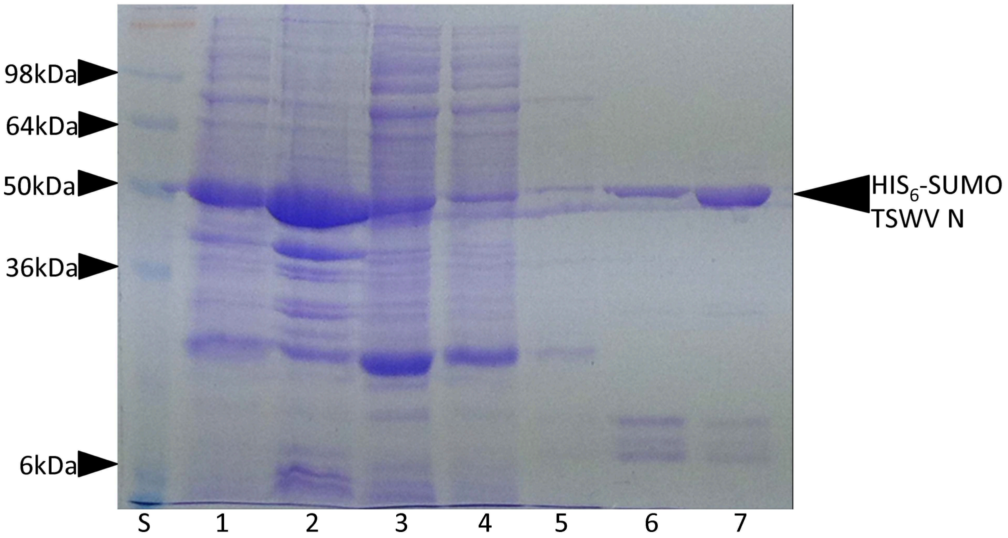


Figure 76. Analysis of elution profile of soluble His$_6$-SUMO TSWV N fusion protein purified using the spin protocol. SDS-PAGE followed by Coomassie stain of soluble fraction (Lane1), Insoluble fraction (Lane2), Flowthrough (Lane 3), Binding buffer wash (Lane 4), 50 mM imidazole wash (Lane 5), 90 mM imidazole elution (Lane 6), 300 mM imidazole elution (Lane 7). Molecular weight markers (kDa) (Lane S).

4.3.2 Large-scale purification of TSWV N protein with Ni-Superflow Plus resin

The results for the expression and purification using the spin protocol were used to scale up the production of pure TSWV N protein suitable for further purification using SEC and crystallization trials. The percentage of total soluble His$_6$-SUMO TSWV N protein capable of successfully binding and eluting from the Ni Superflow Plus resin was small (Figure 77, lanes 4 and 8). However this amount of TSWV N protein generated from the Ni-Superflow Plus resin purification step was deemed sufficient for further rounds of purification using SEC.

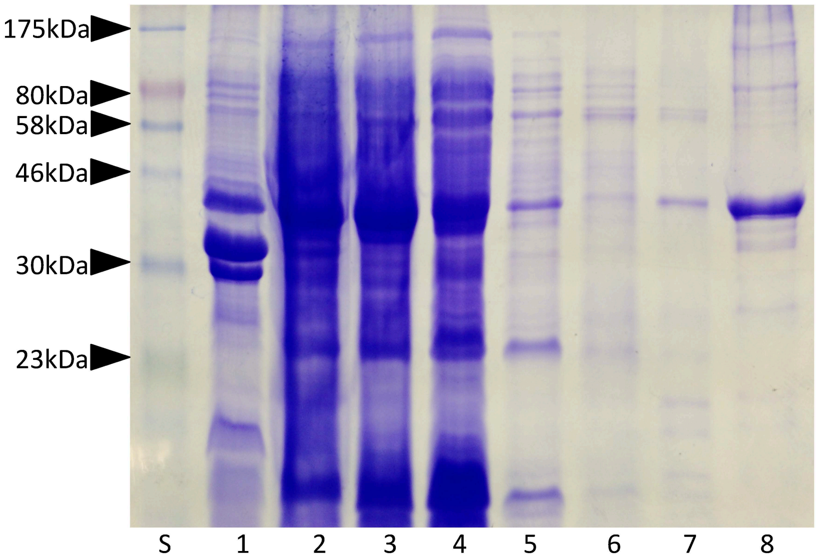


Figure 77. Analysis of elution profile of soluble His_6-SUMO TSWV N purified using Ni-Superflow Plus resin. SDS-PAGE followed by Coomassie stain of molecular weight markers (kDa) (Lane S), Insoluble fraction (Lane 1), Cell lysate (Lane 2), Soluble (Lane 3), Flow through (Lane 4), Binding buffer wash (Lane 5), 50 mM imidazole wash (Lane 6), 90 mM imidazole elution (Lane 7), 2 M NaCl and 300 mM imidazole elution (Lane 8).

4.3.3 Purification of TSWV N protein using SEC

Typically around 60-100 mg of TSWV N protein from each 2 L culture was purified. The TSWV N protein sample purified using nickel-affinity purification was concentrated to 5 ml, and injected onto a S75 SEC column. The SEC chromatogram elution profile of the TSWV N protein produced one distinct peak. The oligomeric state of the TSWV N protein could not be determined because the main peak eluted close to the void volume of the column. The main peak did however display a very high 260/280 nm ratio, and indicated the TSWV N protein was bound to RNA, presumably from the *E. coli* expression host (Figure 78). SDS-PAGE followed by Coomassie stain confirmed the peak contained TSWV N protein, which was pure (Figure 79).

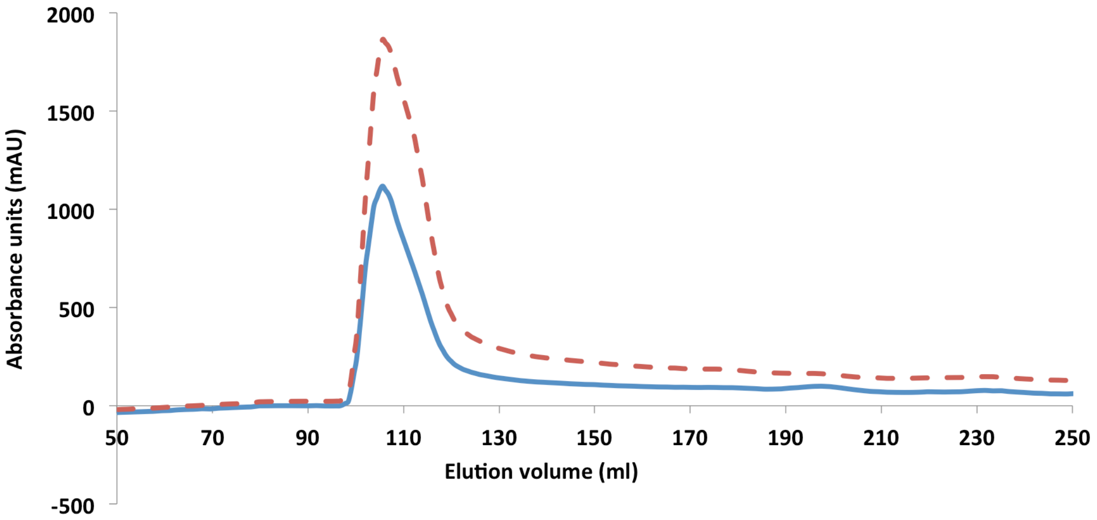


Figure 78. Analysis of the non-RNase treated TSWV N protein using SEC. Bacterially expressed fusion TSWV N protein was purified by nickel-affinity chromatography, after which the His_6-SUMO protein tag was removed by sumo protease and applied to a Superdex S75 gel filtration column at a flow rate of 2.5 ml min^{-1}. The elution profile of the TSWV N protein is shown in blue and dashed red at 280 nm and 260 nm wavelengths respectively.

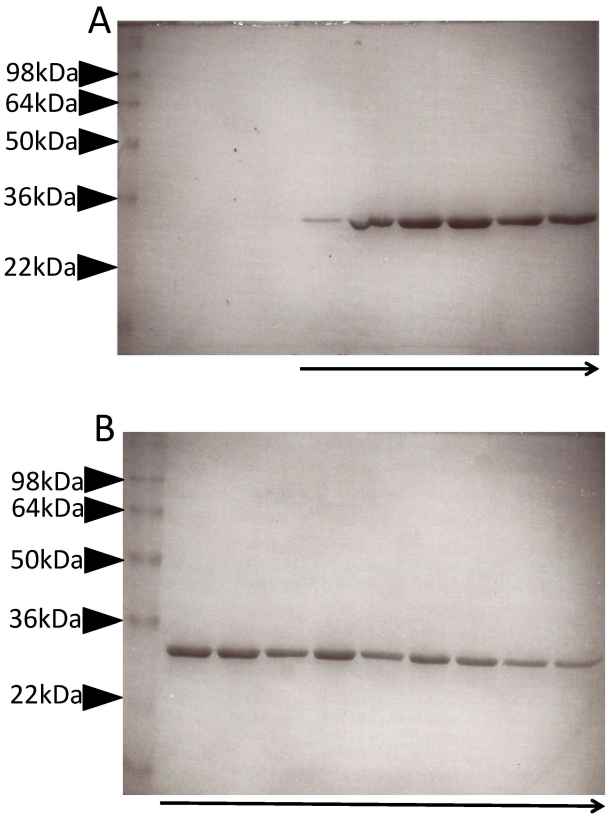


Figure 79. SDS-PAGE analysis of the non-RNase treated TSWV N protein purified using the S75 gel filtration column. SDS-PAGE followed by Coomassie stain. Black arrows in A and B indicate the direction of the elution profile of the main peak seen by the S75 SEC elution profile in figure 78.

4.3.4 RNA binding of purified TSWV N protein

The 260/280 nm ratio of the TSWV N protein eluted from the main SEC peak indicated it was co-purifying with nucleic acid. To establish whether nucleic acids were binding to the TSWV N protein phenol/chloroform extraction was performed. Controls for the experiment were also run in adjacent lanes and included running both RNA harvested and RNase treated from the BUNV N protein. Phenol/chloroform RNA extraction confirmed the TSWV N protein was bound to bacterial RNA, which appeared on the 1% agarose gel as a broad band, containing a range of different sized RNAs. The size range of the TSWV RNA band varied between approximately 45 bp to 150 bp (Figure 80). RNA extraction analysis of the BUNV tetramer N protein peak (Mohl and Barr, 2009) also displayed the expected sized band for RNA, which migrated faster than the RNA extracted from the TSWV N protein.

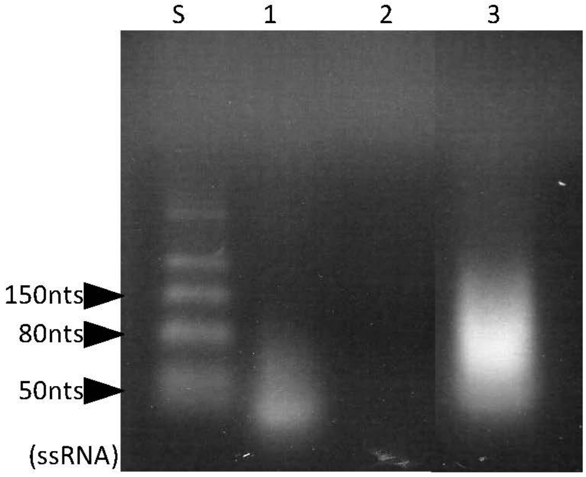


Figure 80. RNA binding analysis of the TSWV N protein purified by SEC. RNA was harvested from N protein purified from the peak of TSWV protein SEC elution profile in figure 78 and subjected to electrophoresis on a 1% agarose gel (Lane 3), alongside ssRNA size markers (Lane S). Positive controls included BUNV N protein extracted from a SEC tetramer peak (Lane 1), and the negative control included RNase treated BUNV N protein from the same SEC tetramer peak (Lane 2). This figure was generated with the help of Dr. Cheryl Walter, University of Leeds.

For every litre of culture the TSWV N protein sample purified using nickel-affinity purification was treated with 10 μg of RNase to remove the bound RNA for 30 min at room temperature, concentrated to 5ml, and finally injected onto a Superdex S75 SEC column. The SEC chromatogram elution profile of the TSWV N protein produced four distinct peaks (Figure 80). The second eluting peak (Figure 81, peak 2), based on molecular weight standards corresponded to a trimer. The oligomer state of the first peak (Figure 81, peak 1), most likely represents tetramers, however this could not be unambiguously established because the S75 SEC column is not suited for separating complexes over 90 kDa. The SEC elution profile also produced a peak corresponding to a TSWV N monomer (Figure 81, peak 3), and a peak corresponding to SUMO protein (Figure 81, peak 4). The initial eluting peaks contained the majority TSWV N protein and the monomer peak contained very little TSWV N protein, as seen by using SDS-PAGE and Coomassie stain analysis (Figure 82). Even though RNase had been added before SEC purification all of the TSWV N protein elution peaks displayed a very high 260/280 nm ratio and indicated bacterial RNA remained bound to the TSWV N protein.

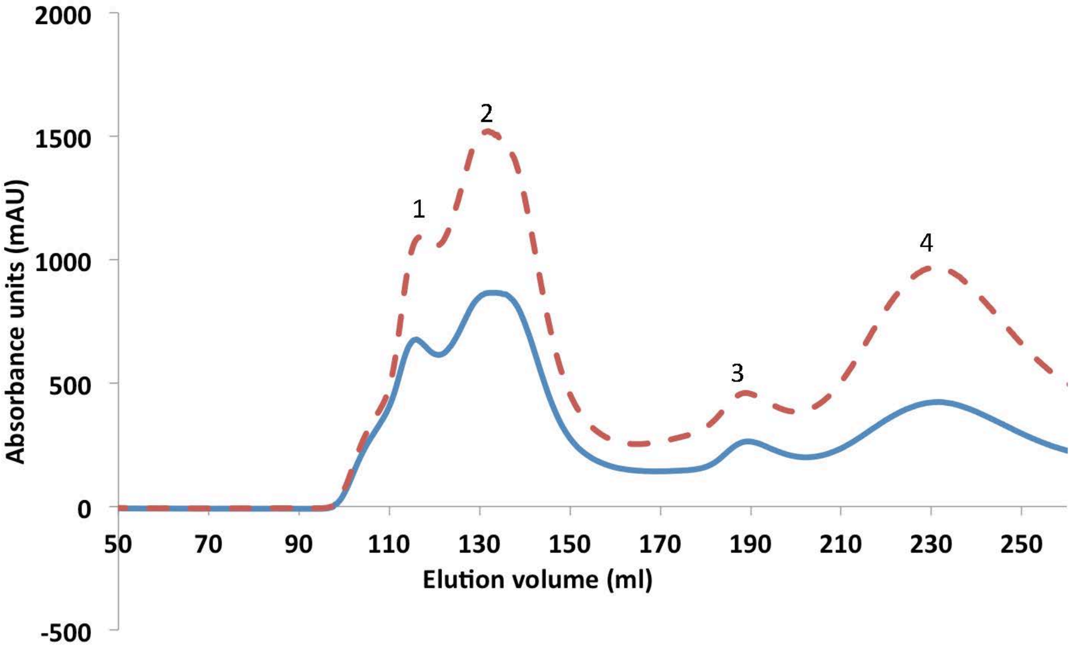


Figure 81. Analysis of the RNase treated TSWV N protein using SEC. TSWV N protein purified by nickel-affinity chromatography was subject to RNase treatment and applied to a Superdex S75 gel filtration column at a flow rate of 2.5 ml min^{-1}. The elution profile of the TSWV N protein is shown in blue and dashed red at 280 nm and 260 nm wavelengths respectively.

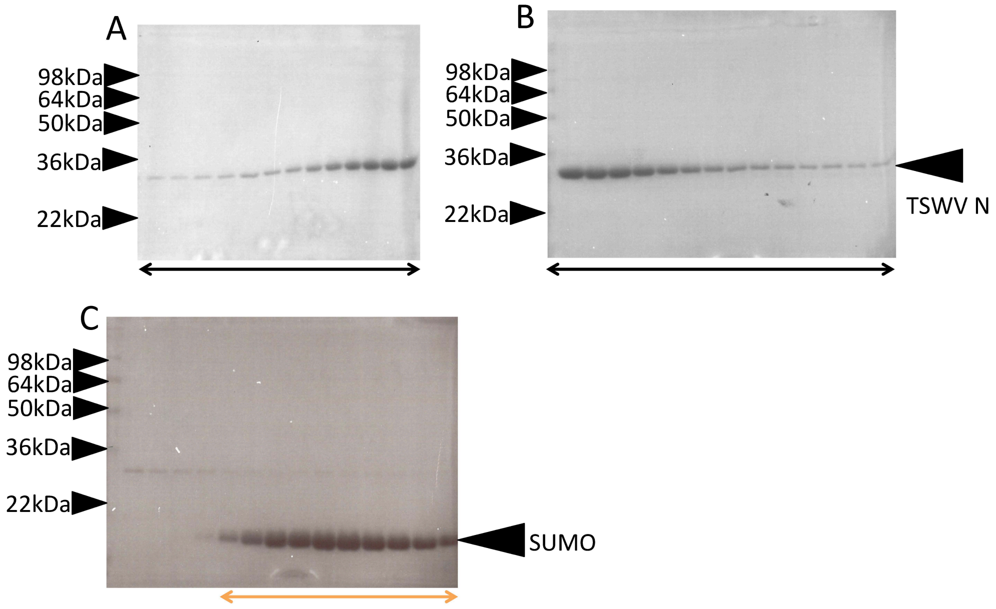


Figure 82. Analysis of elution profile of RNase treated TSWV N protein following purification using a Superdex S75 gel filtration column. SDS-PAGE followed by Coomassie stain of fractions (A) and (B) double black arrows underline the lanes containing fractions from the first three peaks. (C) The double orange arrow underlines the lanes containing fractions with SUMO protein purified using SEC.

4.3.5 RNA binding analysis of RNase treated TSWV N protein purified using SEC

The high 260/280 nm ratio of the TSWV N protein eluted from the first two SEC peaks was further analyzed using phenol/chloroform extraction, and subjected to electrophoresis on a 1% TBE agarose gel alongside ssRNA size markers. Phenol/chloroform RNA extraction confirmed the TSWV N protein remained bound to bacterial RNA, even after RNase treatment. The RNA band from the first peak migrates more slowly than the RNA from the second peak (Figure 83).

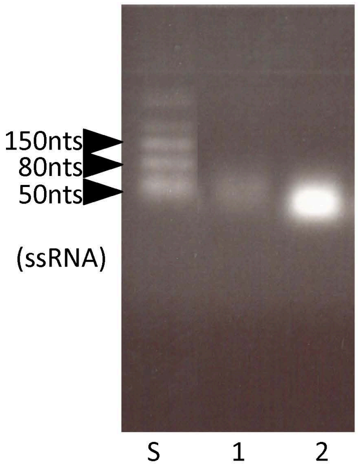


Figure 83. RNA binding analysis of the RNase treated TSWV N protein purified by SEC. RNA was harvested from purified N protein from the first (Lane 1) and second (Lane2) peaks of the SEC elution profile in figure 81 and subjected to electrophoresis on a 1% agarose gel, alongside ssRNA size markers (Lane S).

4.3.6 Purification of RNA free TSWV N protein using SEC

Fractions collected from the center of the first two TSWV N protein SEC elution peaks 1 and 2 were pooled together respectively and concentrated until the concentration reached 15 mg ml^{-1} in approximately 500 μl volume (Figure 81, peak 1 and 2). The concentration and purity of the TSWV N protein was deemed sufficient to attempt crystallization trials using the sitting drop method. Only characteristically diffracting salt crystals were produced and no TSWV N/RNA crystals were grown from any of the factorial conditions, temperatures and drop concentrations screened. In an attempt to produce TSWV protein crystals capable of diffracting suitably for structural determination a strategy to remove the bacterial RNA from the TSWV N protein was developed. This involved altering the final elution step of the nickel-affinity purification to include 2 M NaCl, which would aim to remove any bacterial RNA bound to the TSWV N protein by disrupting the electrostatic N protein/RNA interactions.

Following this step the SEC exclusion running buffer also contained 2 M NaCl. The SEC chromatogram elution profile of the TSWV N protein produced three distinct peaks (Figure 84A). The first large peak (peak 1) displayed a very high 260/280 nm ratio and contains very little protein. This peak most likely represents high molecular weight RNA complexes removed from the TSWV N protein. The chromatogram profile and SDS-PAGE analysis revealed the TSWN N protein was eluted from the size-exclusion column in two oligomeric forms. Comparison of these two species with molecular weight standards indicated that the faster eluting peak 2 had an apparent mass of 110 kDa, and was likely N protein tetramer that has a predicted mass of 112 kDa (Figure 84A). The slower eluting species in peak 3 had an apparent mass of 80 kDa, and was likely N trimer (Figure 80A). The OD_{260nm}/OD_{280nm} ratio of the two peaks was measured to determine the presence of bound nucleic acid. The OD_{260nm}/OD_{280nm} ratios of 0.7 from peak 3 (Figure 84A) and 0.6 from peak 2 (Figure 84A) suggested the TSWV N protein was free of nucleic acid. Analysis of the fractions across the SEC profile of the two main peaks using SDS-PAGE (Figure 85B) revealed the TSWV N protein eluted with little contamination.

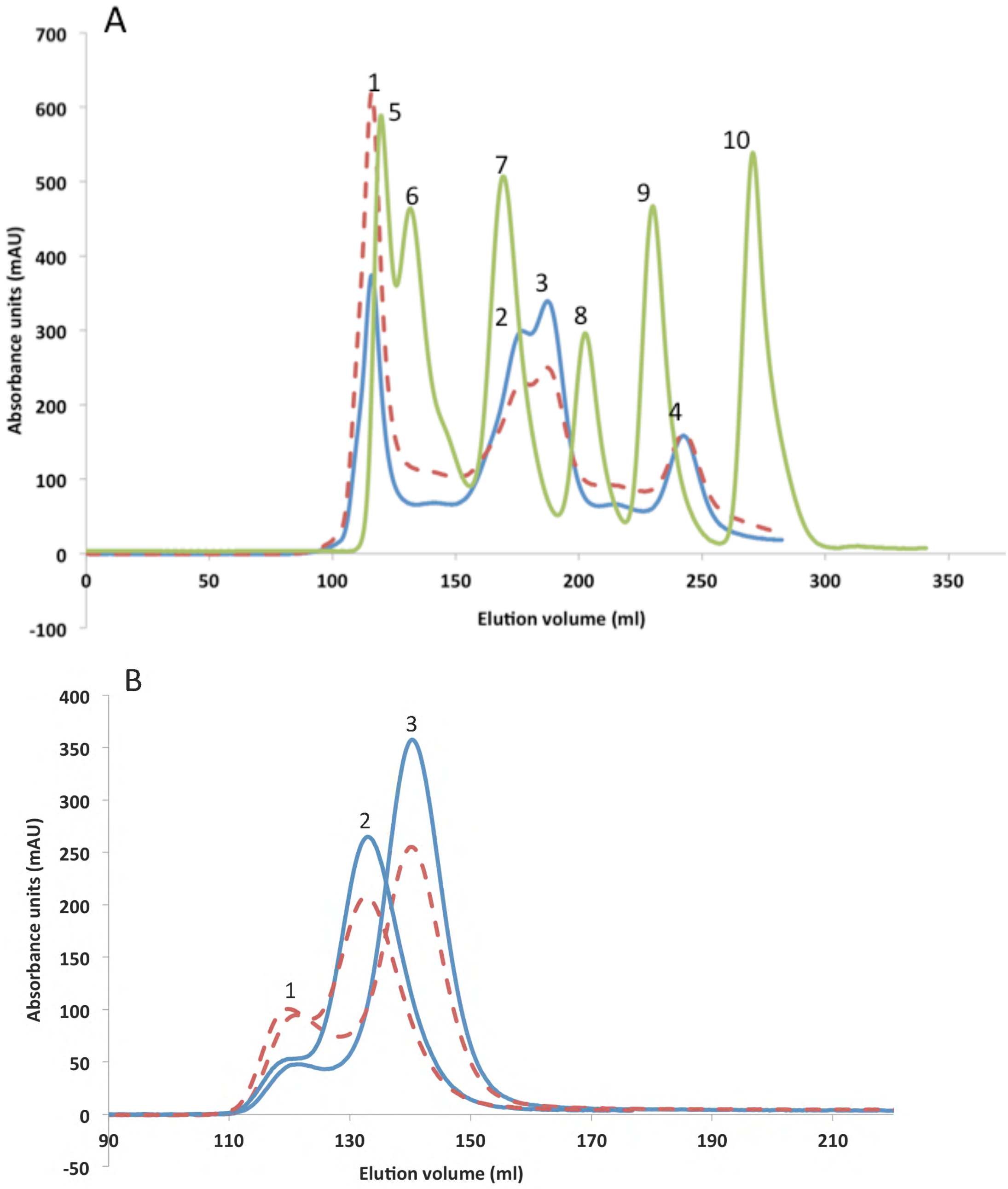


Figure 84. Analysis of the RNA free TSWV N protein using SEC. (A) Bacterially expressed His$_6$-SUMO tagged TSWV N protein was purified by nickel-affinity chromatography, after which the His$_6$-SUMO tag was removed by sumo protease and applied to a Superdex S200 gel filtration column at a flow rate of 1.5 ml min^{-1}. Endogenous *E. coli* RNA bound to the TSWV N protein was separated using 2 M NaCl. The elution profile of the TSWV N protein is shown in blue and dashed red at 280 nm and 260nm wavelengths respectively. Molecular weight standards are shown in green. The following molecular mass standards (green line) are given in kDa: high molecular weight aggregates (Peak 5) thyroglobulin 670, (Peak 6) β-globulin 158, (Peak 7) ovalbumin 44 (Peak 8), myoglobulin 17 (Peak 9), and vitamin B$_{12}$ 1.35 (Peak 10). (B) The trimer and tetramer S200 peak fractions were concentrated to 5 ml respectively and injected onto a Superdex S75 gel filtration column to further purify way any contaminants and buffer exchange the 2 M NaCl into a more suitable buffer for crystallization.

The trimer and tetramer peaks eluted from the S200 column were subject to further purification and injected individually onto a S75 Superdex preparative column (Figure 84B). Peak 2 is TSWV N protein tetramer and peak 3 is TSWV N protein trimer. This step was included not only to exchange the sample into a more suitable buffer for crystallization but also to further purify away any contaminants left from the S200 SEC run.

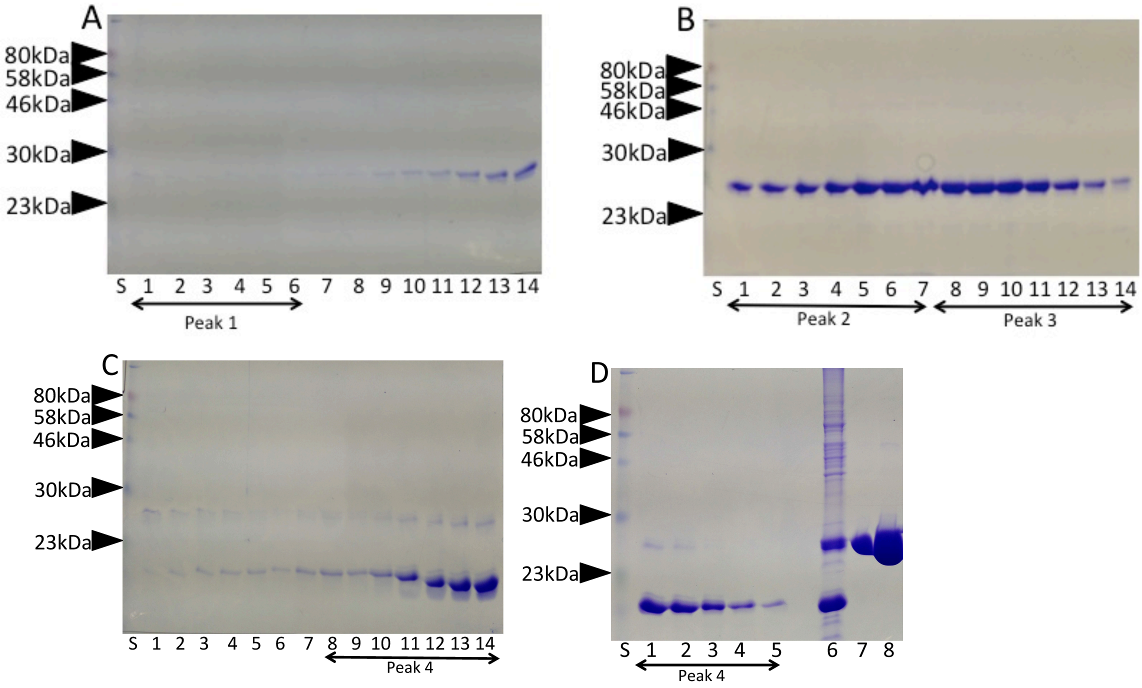


Figure 85. Analysis of elution profile TSWV N protein following purification using Superdex S75 and S200 gel filtration columns. SDS-PAGE followed by Coomassie stain of S200 peak fractions (A) Peak 1 (Lanes 1-6), (B) Peak 2 (Lanes 1-7), and Peak 3 (Lanes 8-14), (C) Peak 4 (Lanes 8-14). (D) Peak 4 (Lanes 8-14), Cleaved TSWV N protein injected onto S200 SEC column (Lane 6), S75 purified tetramer (Lane 7), S75 purified trimer (Lane 8). Double pointed arrows indicate S200 peaks figure 84A .

The elution profile chromatogram of the trimer and tetramer TSWV N protein using the S75 column revealed an unexpected peak (Figure 84B, peak 1). The OD_{260nm}/OD_{280nm} ratio of this peak revealed it contained a high nucleic acid content and most likely contained RNA. Analysis of the OD_{260nm}/OD_{280nm} ratio of the SEC S75 trimer and tetramer TSWV N peaks revealed substantially lower nucleic acid content compared to before they were applied to the S200 Superdex column. The OD_{260nm}/OD_{280nm} ratio of peaks were 0.5 and strongly indicated the trimer and tetramer proteins were free of nucleic acid.

4.4 Crystallization of TSWV N protein

4.4.1 Factorial crystal screens

Fractions collected from the center of the S75 purified TSWV N protein trimer and tetramer SEC elution peaks were pooled together respectively and concentrated until the concentration reached 15 mg ml^{-1} in approximately 500 μl volume (Figure 85D, lanes 7 and 8). The concentration and purity of the TSWV N protein was deemed sufficient to attempt crystallization trials using the sitting drop method. Crystals for the trimer species grew in 0.1 M HEPES pH 7.5, 4.3 M sodium chloride, and 0.01 M zinc sulphate hydrate, 0.1 M MES monohydrate pH 6.5, 25% v/v PEG-550 MME (Figure 86). Crystals for the tetramer TSWV N protein crystallized in 0.01 M zinc sulphate hydrate, 0.1 M MES monohydrate pH 6.5, 25% v/v PEG-550 MME (Figure 87A) and 0.2 M trimethylamine N-oxide dihydrate, 0.1 M Tris-HCl pH 8.5, 20% w/v PEG-2,000 MME (Figure 88A).

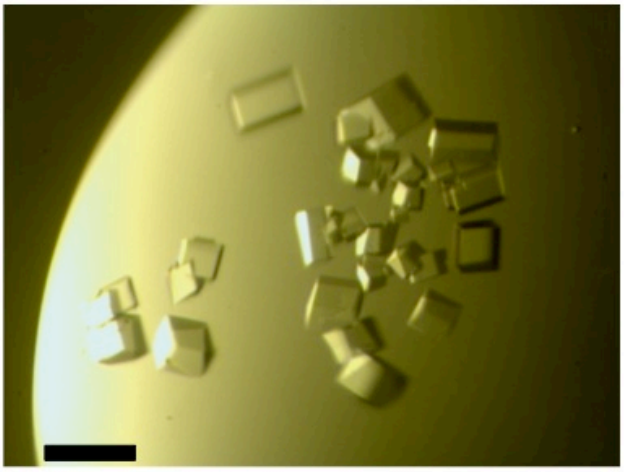


Figure 86. Crystallization of trimer TSWV N protein using the sitting drop method of crystallization. Crystallization trials were carried out at 25 °C using the sitting drop vapour-diffusion method. Crystals grown in condition 27 from the crystal screen 2 (Hampton Research) crystallization screening kit. (Scale bar = 25 μm).

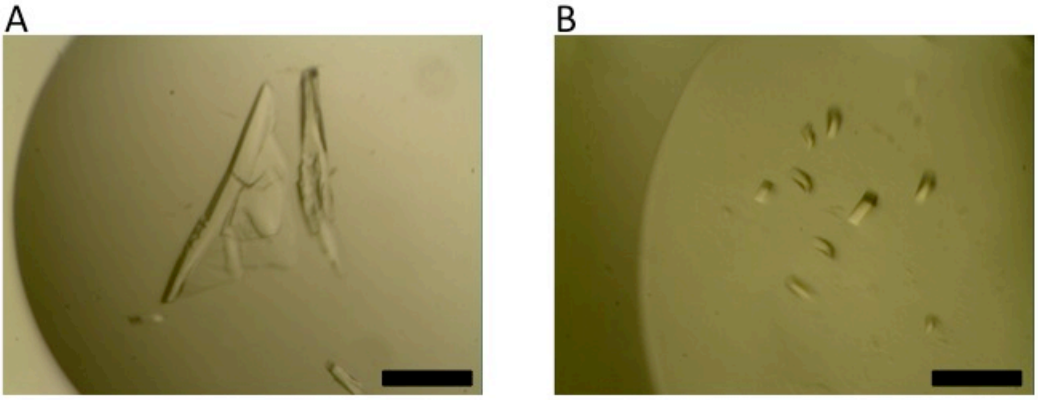


Figure 87. Crystallization of the TSWV N protein tetramer using the sitting drop and hanging drop method of crystallization. Crystallization trials were carried out at 25 °C using the sitting drop vapour-diffusion method. (A) Crystals grown in condition 27 from the crystal screen 2 (Hampton Research) crystallization screening kit (Scale bar 20 µm) (B) Optimization was based on initial sitting drop screening conditions, which produced TSWV protein crystals diffracting to at least 2.6 Å and was carried out at 25 °C using hanging drop vapour-diffusion method. Crystals grown in 25% v/v PEG-550 MME, 0.25 M MES pH 6.5. (Scale bar = 100 µm).

4.4.2 Optimization of factorial tetramer TSWV N protein crystals

To improve the quality of the factorial tetramer TSWV N protein crystals grown using the sitting drop method for crystallization, several rounds of optimization were carried out. This included omitting the zinc sulphate, testing different pHs and changing the percentage of PEG-550 MME. The best crystals were obtained from hanging drops prepared using 25% v/v PEG-550 MME, 0.25 M MES pH 6.5 (Figure 87B).

4.5 X-ray diffraction of TSWV N protein crystals

4.5.1 Diffraction of factorial TSWV N protein crystals

Factorial sitting drop crystallization screening yielded diffracting crystals from two conditions. Firstly sitting drop factorial crystals grown using 0.2 M trimethylamine N-oxide dehydrate, 0.1 M Tris-HCl pH 8.5, 20% w/v PEG-2,000 MME diffracted, but very weakly (Figure 88B). Factorial crystals hits found in 0.2 M trimethylamine N-oxide dihydrate, 0.1 M Tris-HCl pH 8.5, 20% w/v PEG-2,000 MME, were not optimized for native TSWV N protein crystals.

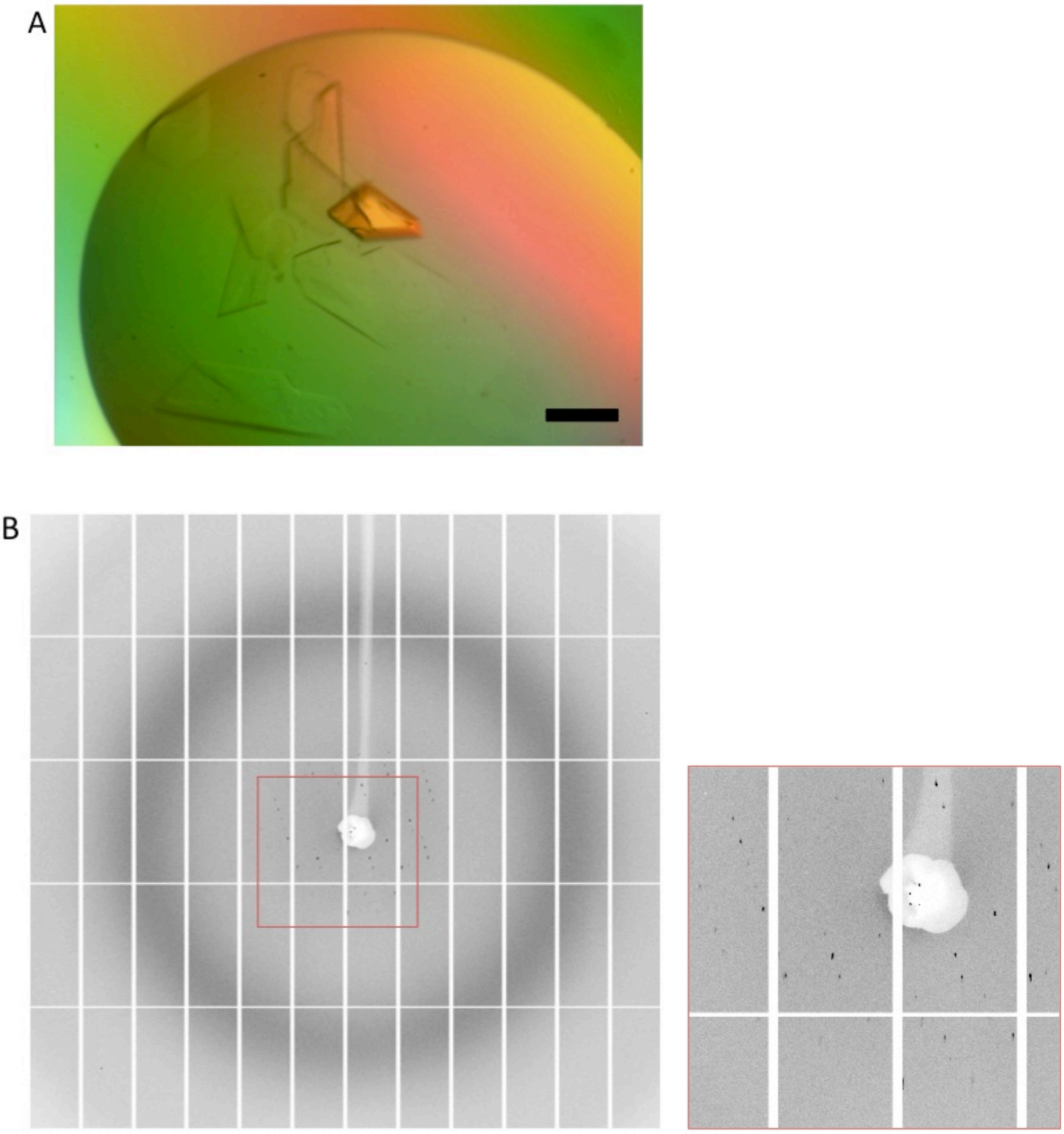


Figure 88. X-ray diffraction pattern of a tetramer TSWV N protein crystal grown using 0.2 M trimethylamine N-oxide dehydrate, 0.1 M Tris-HCl pH 8.5, 20% w/v PEG-2,000 MME. (A) Crystals grown using the sitting drop method using 0.2 M Trimethylamine N-oxide dehydrate, 0.1 M Tris pH 8.5, 20% w/v PEG-2,000 MME (Scale bar = 25 μm). (B) A 0.5° oscillation image collected on beamline I24 using 12.8 KeV X-rays, in which reflections are observed to beyond 9 Å. The insert boarded in red shows the zoomed in image (x3) of the diffraction pattern boxed in red.

Secondly initial factorial crystal hits grown in 0.01 M zinc sulphate hydrate, 0.1 M MES monohydrate pH 6.5, 25% v/v PEG-550 MME using the sitting drop vapour-diffusion method, diffracted to a maximum Bragg spacing of 4 Å and belonged to space group group $P2_1$, with unit-cell parameters a = 72.4, b = 105.9, c = 79.8 Å, α= 90°, β= 105.8°, γ= 90°. Crystals optimised from this condition (see above) using the hanging drop vapour-diffusion method (Figure 87B) diffracted to a maximum Bragg spacing of 2.6 Å (Figure 89). It was calculated by the strategy function within the

MOSFLM package that rotating this $P2_1$ space group crystal 360° with 0.5° oscillations was sufficient to collect a full X-ray diffraction data set of the TSWV N protein, and which produced 720 images. The resulting data set was indexed and integrated using the XDS package (Kabsch, 2010) using Xia2 (Evans, 2006b) from the Diamond synchrotron suite of automated programs. The data were scaled with a maximum Bragg spacing of 2.55 Å with an R-merge of 61.5% in the highest resolution shell. These crystals belonged to the $P2_1$ spacegroup, with similar unit-cell parameters as the factorial crystals (Table 5). In the $P2_1$ space group there are four TSWV N proteins N per AU, with a solvent content of 53.16%.

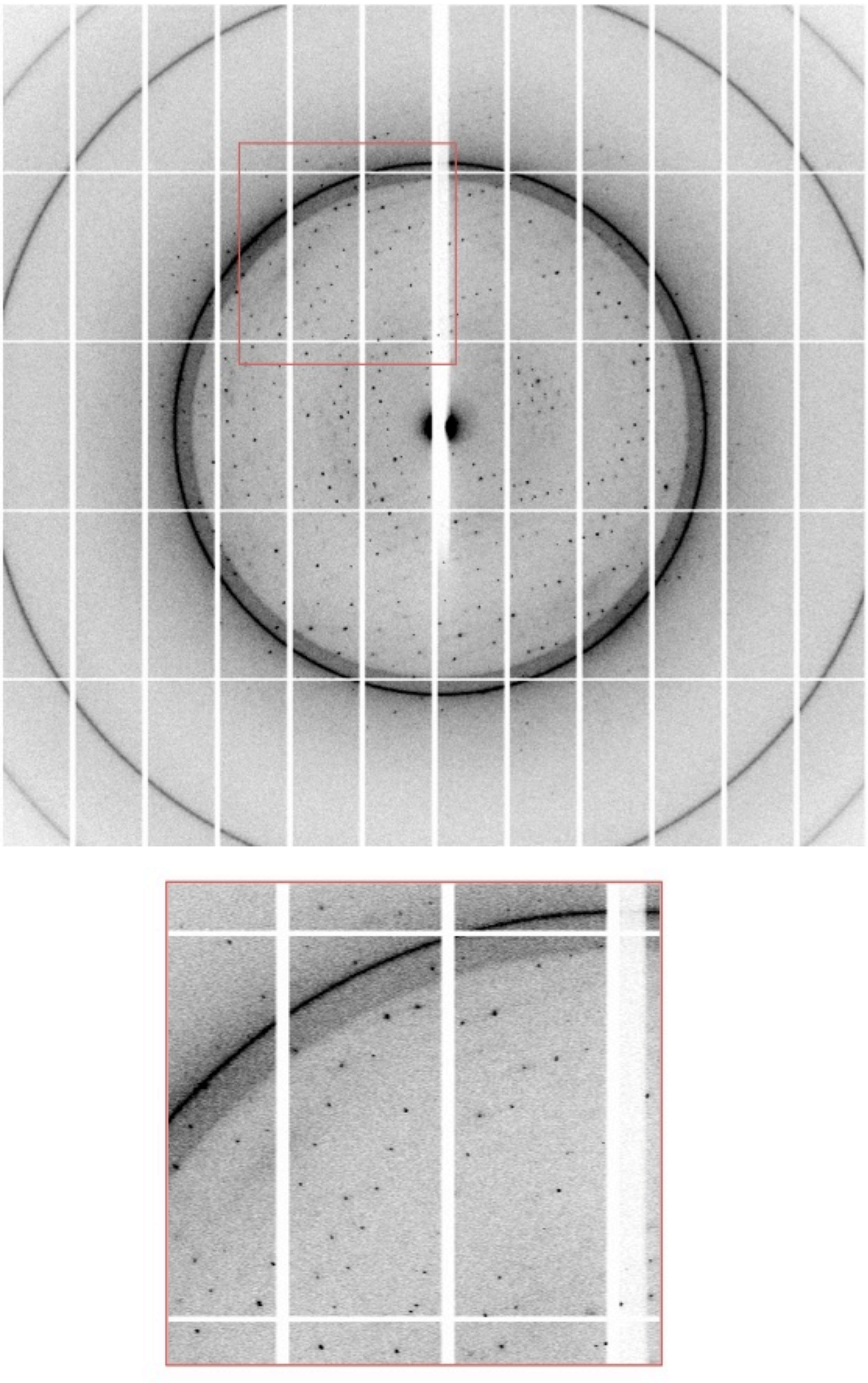


Figure 89. X-ray diffraction pattern of tetramer TSWV N protein crystals. A 0.5° oscillation X-ray diffraction pattern collected at beamline I02 using 12,647 eV energy, and reflections are observed to beyond 2.6 Å. The insert boarded in red shows the zoomed in image (x2) of the diffraction pattern boxed in red.

Parameter	Value
Wavelength (Å)	0.98030
Space group	$P12_11$
Cell parameters	
a (Å)	72.4
b (Å)	105.9
c (Å)	79.8
β (°)	105.8
Total no. of reflections	160,835 (11,932)
No. of unique reflections	37735 (2749)
Resolution (Å)	2.55 (2.62–2.55)
R_{merge} (%)	3 (61.5)
Completeness (%)	99.7 (99.7)
Redundancy	4.3(4.3)
$I/\sigma(I)$	21.2 (2.3)
No. of molecules per AU	4
VM (Å^3/Da)	2.62

Table 5. Data collection and refinement statistics for tetramer TSWV N native crystals.

4.5.2 Optimization of factorial trimer TSWV N protein crystals

Several rounds of optimization were also carried out to improve the quality of the trimer TSWV N protein crystals grown using the sitting drop method for crystallization (Figure 86). Although the trimer TSWV N protein crystals did not diffract from any of the hits found using factorial screening, attempts were made to improve these crystals using the hanging drop vapour diffusion method for crystallization and additive screening using the sitting drop method, however no improvement to diffraction was found.

4.6 Self-rotation of TSWN N protein tetramer crystals

Self-rotation functions were calculated with the MOLREP program using the X-ray diffraction data collected from tetramer TSWV N protein crystals (Vagin and Teplyakov, 2010). The two strongest peaks for κ=180° at ϕ = 180° and ϕ = 0° are derived from the crystallographic twofold-symmetry axes. The two additional peaks observed on the κ= 180° plot correspond two two-fold NCS peaks (Figure 90). There is another NCS peak, slightly off origin represented in the κ=90° slice, indicating a 4 fold non-crystallography symmetry axis. There was no evidence of three fold or six fold crystallographic symmetry for κ=120° (Figure 90), or κ=60° (Figure 90). The Matthews coefficient suggests a tetramer is the most likely component in the AU, with 2.62 Å3 Da^{-1} with 53.16% solvent (Matthews, 1968). Considering that a probable tetramer was observed during SEC, in conjunction with self-rotation analysis suggesting there is a four-fold NCS axis present, a tetramer form is likely to be present in the AU.

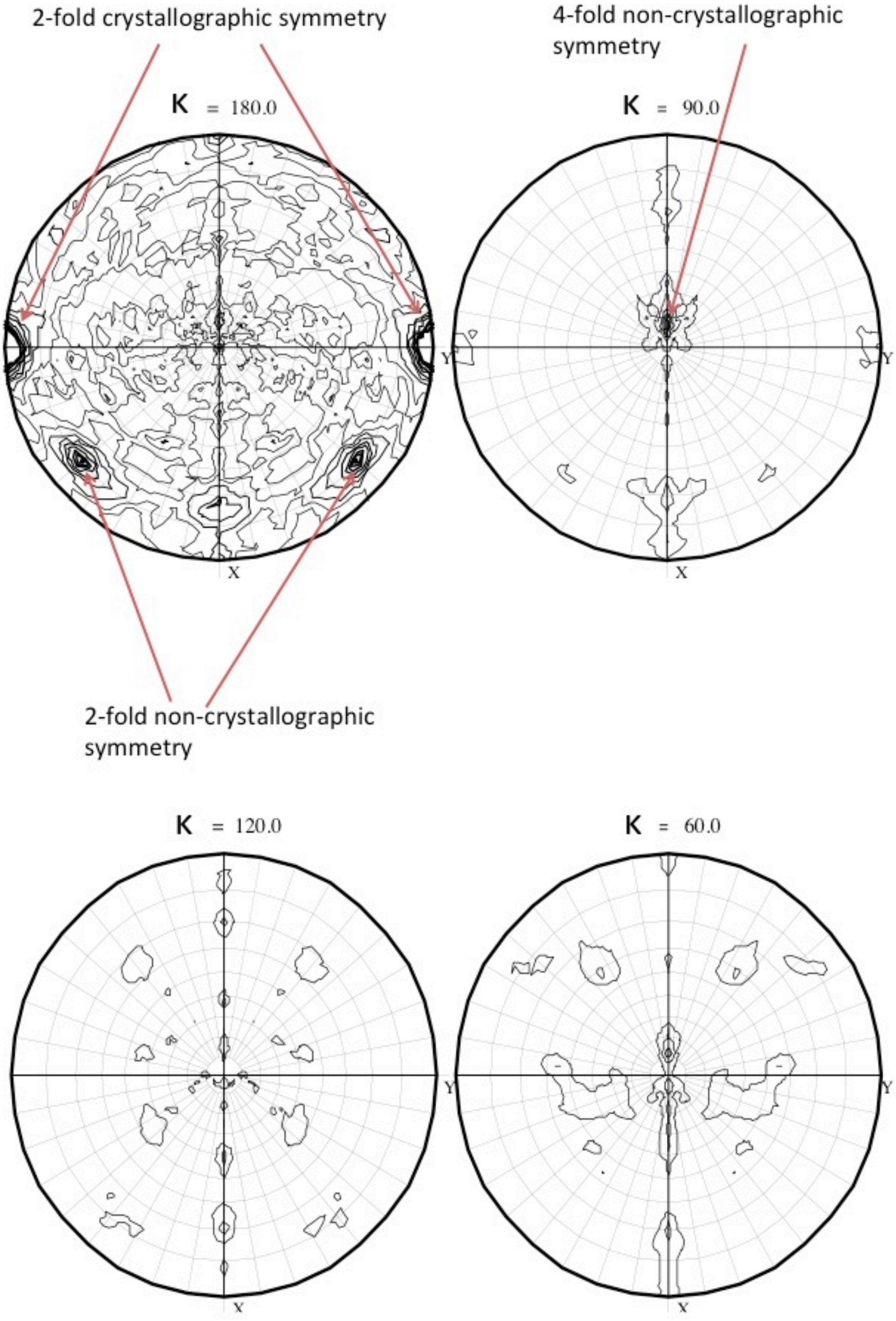
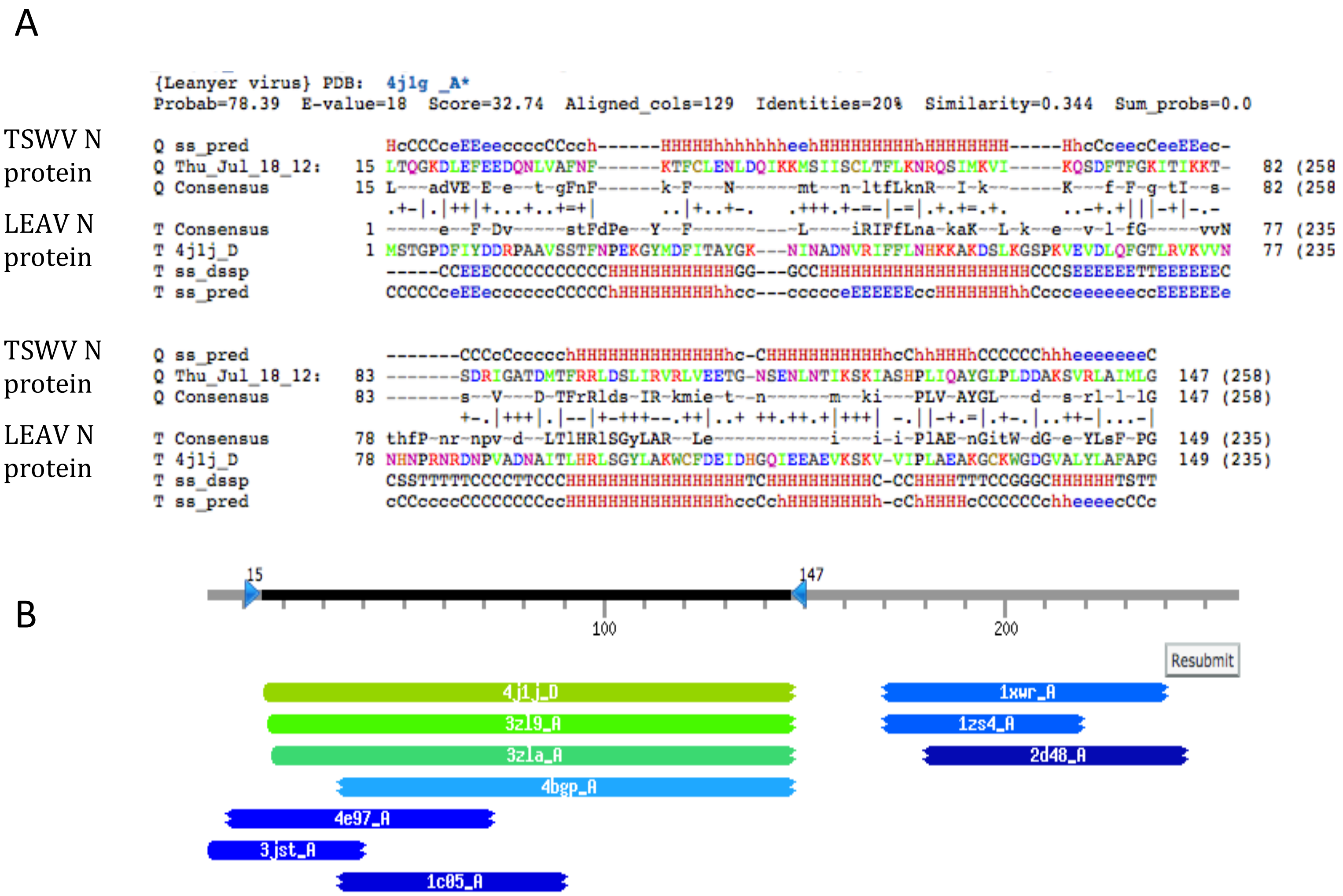


Figure 90. Self-rotation function calculated by MOLREP using the TSWV N protein crystals. Self-rotation searches with κ= 180° and κ= 90°, κ= 120°, and κ= 60° were used to identify two-fold, four-fold, three fold and six fold rotation angles, respectively.

4.7 Solving the TSWV N protein crystal X-ray diffraction data

4.7.1 Molecular replacement

Determination of the initial phases by molecular-replacement approaches using Phaser and MOLREP programs (McCoy et al., 2007, Vagin and Teplyakov, 2010) with the RVFV N protein (Raymond et al., 2010; Ferron et al., 2011) as a search model was not successful. The crystal structures of the recently solved N proteins from orthobunyaviruses (Dong et al., 2013b, Dong et al., 2013a, Ariza et al., 2013, Li et al., 2013, Niu et al., 2013, Reguera et al., 2013), which have approximately 20% sequence homology with the TSWV N protein, also failed as molecular replacement search models. The remainder of N proteins from nsNSV and sNSV for which there are high resolution structures also were not successful for molecular replacement. Using the NCBI BLAST tool, (Altschul et al., 1990) it was found that there was no model in the PDB with sequence similarity above 20% compared to the TSWV N protein sequence. HHpred (Soding et al., 2005) was used to search for domain homology similarities based on the secondary structure prediction of the TSWV N protein and found the orthobunyaviruses N protein crystal structures matched with the highest similarity between residues 15-149 (Figure 91). However this fragment obtained from all the orthobunyavirus N protein crystal structures was unsuccessful in obtaining the suitable phases for structural determination of the TSWV N protein crystal structure.



```
{Leanyer virus} PDB:  4jlg _A*
Probab=78.39  E-value=18  Score=32.74  Aligned_cols=129  Identities=20%  Similarity=0.344  Sum_probs=0.0

Q ss_pred             HcCCCceEEeeccccCCcch------HHHHHhhhhhhheehHHHHHHHHHhHHHHHHHH-----HhcCceecCeeEEec-
Q Thu_Jul_18_12:   15 LTQGKDLEFEEDQNLVAFNF------KTFCLENLDQIKKMSIISCLTFLKNRQSIMKVI-----KQSDFTFGKITIKKT-   82 (258
Q Consensus        15 L---adVE-E-e--t-gFnF------k-F---N------mt--n-ltfLknR--I-k-------K---f-F-g-tI--s-   82 (258
                      .+-|.|++|+...+..+=+|      ..|+..+-.   .+++.+-=-|-=|.+.+=.+.     ..-+.+|||-+|-.-
T Consensus         1 -----e--F-Dv-----stFdPe--Y--F---------L----iRIFfLna-kaK--L-k--e--v-l-fG-----vvN   77 (235
T 4j1j_D            1 MSTGPDFIYDDRPAAVSSTFNPEKGYMDFITAYGK---NINADNVRIFFLNHKKAKDSLKGSPKVEVDLQFGTLRVKVVN   77 (235
T ss_dssp             -----CCEEECCCCCCCCCCCHHHHHHHHHHHHGG---GCCHHHHHHHHHHHHHHHHHHHHCCCSEEEEEETTEEEEEEC
T ss_pred             CCCCCceEEecccccccCCCCChHHHHHHHHHHhhcc---cccccEEEEEEccHHHHHHHhhCccceeeeeeccEEEEEEe


Q ss_pred             -------CCCcCcccccchHHHHHHHHHHHHHHHHhc-CHHHHHHHHHHHHhcChhHHHHhCCCCCChhheeeeeeeC
Q Thu_Jul_18_12:   83 -------SDRIGATDMTFRRLDSLIRVRLVEETG-NSENLNTIKSKIASHPLIQAYGLPLDDAKSVRLAIMLG  147 (258)
Q Consensus        83 -------s--V---D-TFrRlds-IR-kmie-t--n------m--ki---PLV-AYGL---d--s-rl-l-lG  147 (258)
                             +-.|+++|.|--|+-+++--.++|..+ ++.++.+|+++| -.||-+.=|.+-.|..++-|...-|
T Consensus        78 thfP-nr-npv-d--LTlHRlSGyLAR--Le-----------i---i-i-PlAE-nGitW-dG-e-YLsF-PG  149 (235)
T 4j1j_D           78 NHNPRNRDNPVADNAITLHRLSGYLAKWCFDEIDHGQIEEAEVKSKV-VIPLAEAKGCKWGDGVALYLAFAPG  149 (235)
T ss_dssp             CSSTTTTTCCCCTTCCCHHHHHHHHHHHHHHHHHHTCHHHHHHHHHHHC-CCHHHHTTTCCGGGCHHHHHHTSTT
T ss_pred             cCCcccccCCCCCCCCccHHHHHHHHHHHHHHHHhccCchHHHHHHHHHh-cChHHHHcCCCCCCchheeeecCCc
```

Figure 91. HHpred local alignment between the predicted secondary structure of the TSWV N protein and structures submitted onto the PDB. (A) The alignment with the highest identity was between residues 1-149 of the LENV (orthobunyavirus), and residues 15-147 of TSWV N. (B) The highest matched PDBs aligned to the predicated secondary structure of the TSWV N protein, which include orthobunyavirus N proteins and other non-N proteins.

HHpred also matched small sections of the predicted secondary structure of the TSWV N protein with non-N proteins, which include but are not limited to lysozyme hydrolase (PDB code: 4E97A), coliphage lambda transcription activator protein CII (PDB code: 1XWR), and transcriptional coactivator/pterin dehydratase from *Brucella Melitensis* (PDB code: 3JST).

4.7.2 Selenomethionine TSWV N protein

Bacterial expression of the TSWV N containing selenomethionine was achieved using *E. coli* cells, which are methionine auxotrophs, grown in M9 minimal media. Using the same induction conditions for expression as for the native TSWV N protein sufficient amounts of protein were produced for crystallization. However the TSWV N protein, containing selenomethionine did not crystallize in the same condition as the native protein. Optimization of the crystallization was attempted, by changing pH, temperature, drop volumes, drop ratios, but TSWV N protein crystals using the native crystallization condition could not be reproduced using selenomethionine protein.

Next the selenomethionine protein was re-screened for crystallization at a concentration of 15 mg ml^{-1} using the sitting drop vapour diffusion method. Crystals grew in 0.2 M trimethylamine N-oxide dihydrate, 0.1 M Tris-HCl pH 8.5, 20% PEG-2,000 MME (Figure 92). This condition has previously produced crystals for the native TSWV tetramer, however these crystals diffracted very weakly (Figure 93). Optimization of this condition, using a range of different PEG MMEs, salts, different buffers (MOPS, BTP, and MES), temperatures and also omitting the 0.2 M trimethylamine N-oxide dihydrate did not improve the size, quality or diffraction of these crystals. Additionally an additive screen (Hampton research) was implemented by adding 10% v/v of each additive to the condition, and seeding was also attempted using crushed up native protein crystals. Both of these approaches failed to produce better diffracting crystals.

Figure 92. Crystallization of tetramer TSWV N selenomethionine protein using the sitting drop vapour-diffusion method at 25 °C, and crystals were seen in condition 62 from the Index crystallization screening kit (Hampton Research).

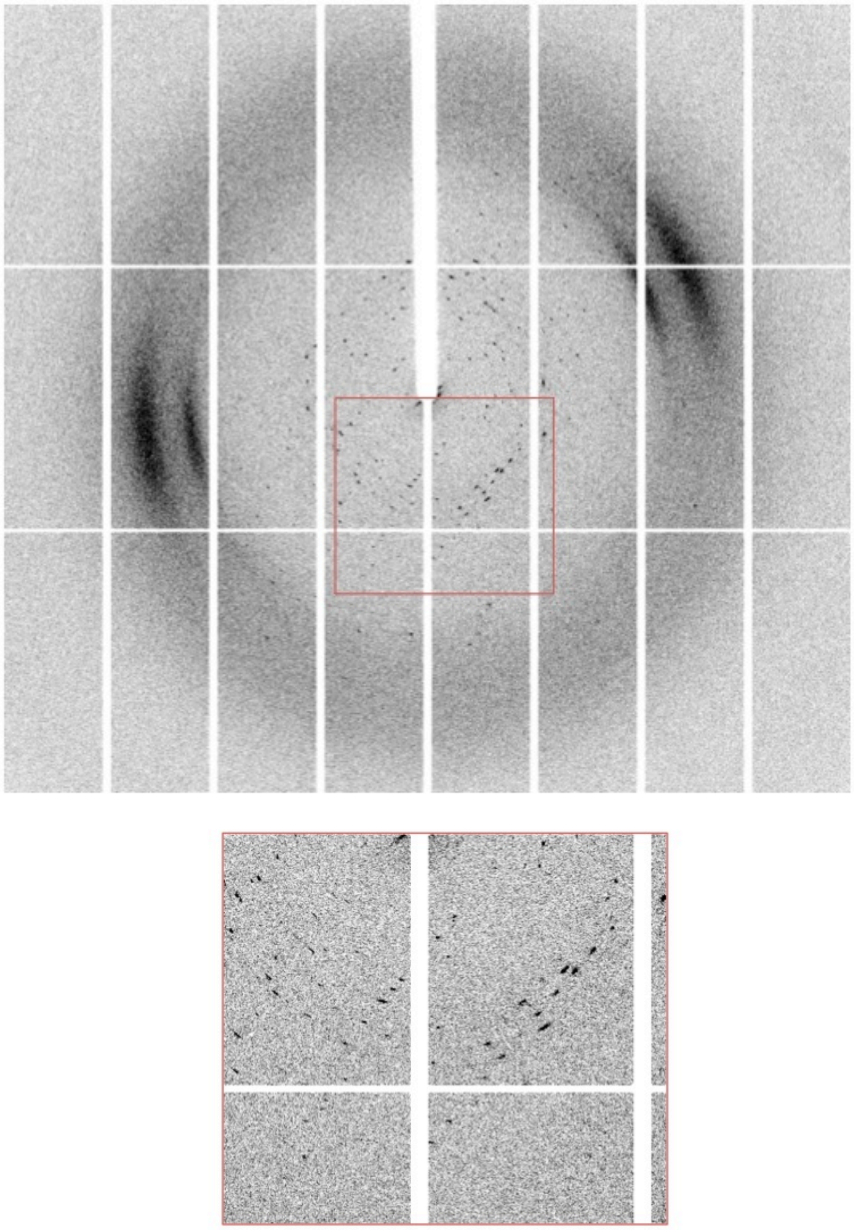


Figure 93. X-ray diffraction pattern of a tetramer TSWV N selenomethionine crystal grown using 0.2 M trimethylamine N-oxide dihydrate, 0.1 M Tris-HCl pH 8.5, 20% w/v PEG-2,000 MME. A 0.5° oscillation image collected on beamline I04-1 using 13,477 eV X-rays, and reflections are observed to beyond 9 Å. The insert boarded in red shows the zoomed in image (x2) of the diffraction pattern boxed in red.

4.7.3 Heavy atom soaks into native crystals grown from TSWV N protein

X-ray diffraction was collected from TSWV N crystals soaked in mother containing 150 mM NaCl for various periods and with various heavy atoms ranging in different concentrations. Heavy atom soaks were followed by a 5 min back wash to remove weak binding atoms (Table 6).

Heavy atom	Concentrations	Times	Result
Thimerosal	100 µM, 1 mM	1 min, 5 min and 10 min	Diffraction poor
Uranyl acetate	100 µM, 1 mM, 10 mM, 20 mM	1 min, 5 min, 10 min, 60 min, 4 hr, 8 hr	Diffracted to 3 Å resolution, bad anisotropy and no detectable anomalous signal
Samarium nitrate	100 µM, 1 mM, 100 mM,	30 sec, 1 min, 5 min	Crystals cracked, bad anisotropy, anomalous signal not sufficient for phasing, poor diffraction 5 Å
Gold(I) potassium cyanide	1 mM, 10 mM, 40 mM	30 sec, 1 min, 5 min	Poor diffraction, anisotropic, no anomalous signal
Gold chloride	1 mM, 10 mM, 20 mM	30 sec, 1 min, 5 min	Poor diffraction, anisotropic, no anomalous signal
Rhubidium chloride	1 mM, 10 mM	2 min, 10 min	Diffracted to 3 Å, no anomalous signal
Mercury potassium iodide	1 mM, 10 mM	2 min, 3 min, 5 min, 10 min	Diffraction poor
Cerium (III) acetate,	100 µM, 1 mM	1 min, 2 min, 5	No diffraction

sesquihydrate		min	
Potassium tetracyanoplatinate(II) trihydrate	100 µM, 1 mM, 10 mM	1 min, 2 min, 10 min	Poor diffraction
p-Chloromercuriphenylsulfonic acid	1 mM, 2 mM	10 sec, 20 sec	Diffracted to 8 Å, anisotropic.
Mercury acetate	2 mM	1 min, 3 min, 10 min	No diffraction
Mercurochrome	1 mM, 2 mM, 10 mM	1 min, 5 min	Good diffraction, no anomalous signal
Potassium chloroplatinate	1 mM, 2 mM	1 min, 5 min, 10 min	Diffracted to 8 Å, no anomalous signal
Trimethyl lead acetate	1 mM, 5 mM	2 min, 5 min, 15 min	Poor diffraction
Lead nitrate	4 mM	2 min, 5 min, 10 min	Bad diffraction
4-chloromercuribenzoic acid	2 mM, 4 mM	2 min, 5 min, 10 min	Bad diffraction, anisotropic
cis-Platinum(II) diammine dichloride	1 mM, 2 mM, 10 mM	2 min, 5 min, 20 min	Diffraction to 4 Å, anisotropic no anomalous signal
5-Amino-246-Triiodo Isophthalic Acid	10 mM	1 min, 10 min	Did not diffract
Hexadecyltrimethylammonium bromide	1 mM, 10 mM	2 min, 3 min, 5 min	Bad diffraction, anisotropic

Table 6. List of heavy atoms used to soak into native TSWV N protein crystals. The table includes the concentrations of the heavy atoms and the times for each concentration, followed by a brief description of the X-ray diffraction observed.

Those crystals soaked with heavy atoms, uranyl acetate, samarium nitrate cis-platinum (II) diammine dichloride were integrated (Leslie, 2006) and scaled (Evans, 2006b). Samarium nitrate soaked crystals produced the best diffracting crystals, (Figure 94) however anomalous signal was only found at 5 Å resolution using AutoSol (Terwilliger et al., 2009), and was not sufficient to produce any discernible electron density maps to build an initial model. Considerable anisotropy was seen in all data sets collected from crystals soaked with heavy atoms. An example of the anisotropy from a samarium nitrate soaked crystal can be seen in figure 95. Isomorphous difference maps were produced using Fourier differences by combining native and samarium nitrate derivative data sets using CAD (1994), and scaling them together using SCALIT (1994). Isomorphous difference maps can be used to find initial positions of heavy atoms in substructure solution, suitable for phase refinement, however no peaks using this method could be seen over four sigma. Additionally all images with high R_{merge} were excluded from the samarium nitrate derivative data set, however this did not improve the quality of the data, and phase information still could not be found.

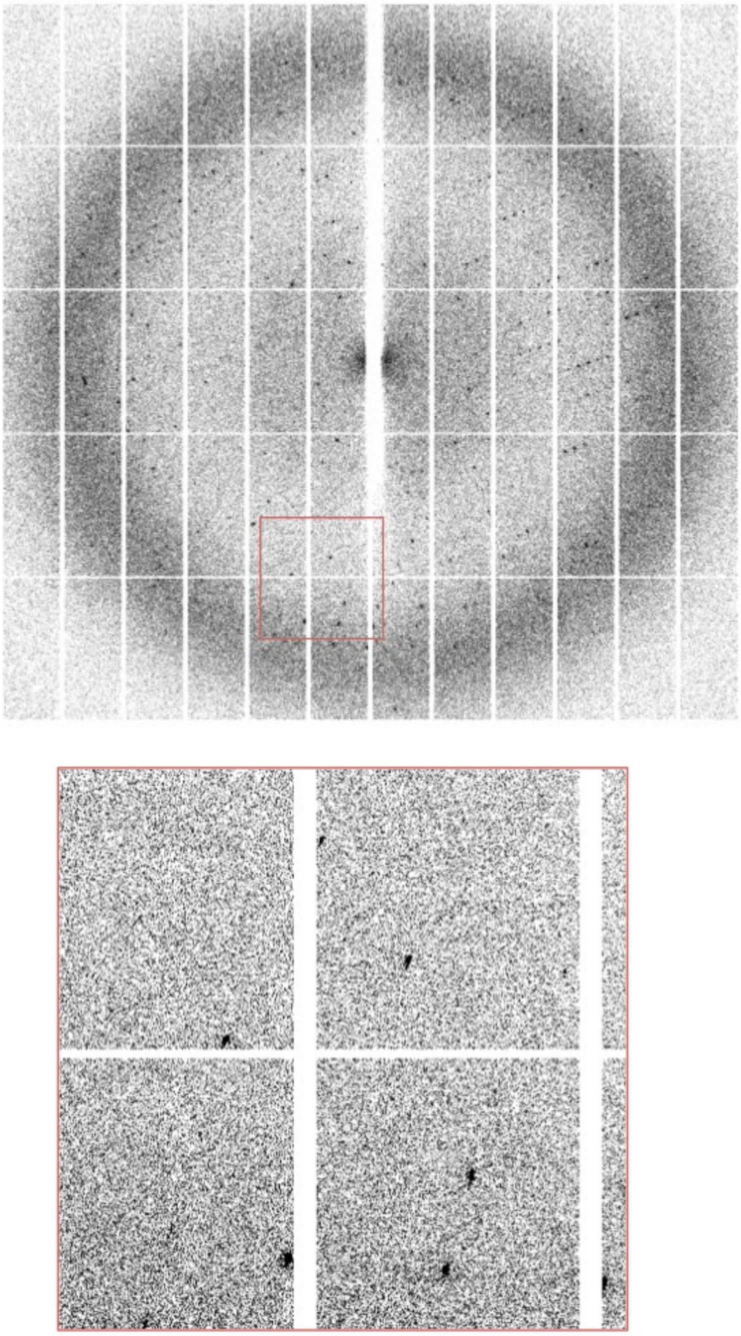


Figure 94. X-ray diffraction pattern of tetramer TSWV N protein crystal grown using 10 mM zinc sulphate, 25% v/v PEG-550 MME, 0.25 M MES pH 6.5, and soaked with 100 µM of samarium nitrate. The insert boarded in red shows the zoomed in image (x3) of the diffraction pattern boxed in red.

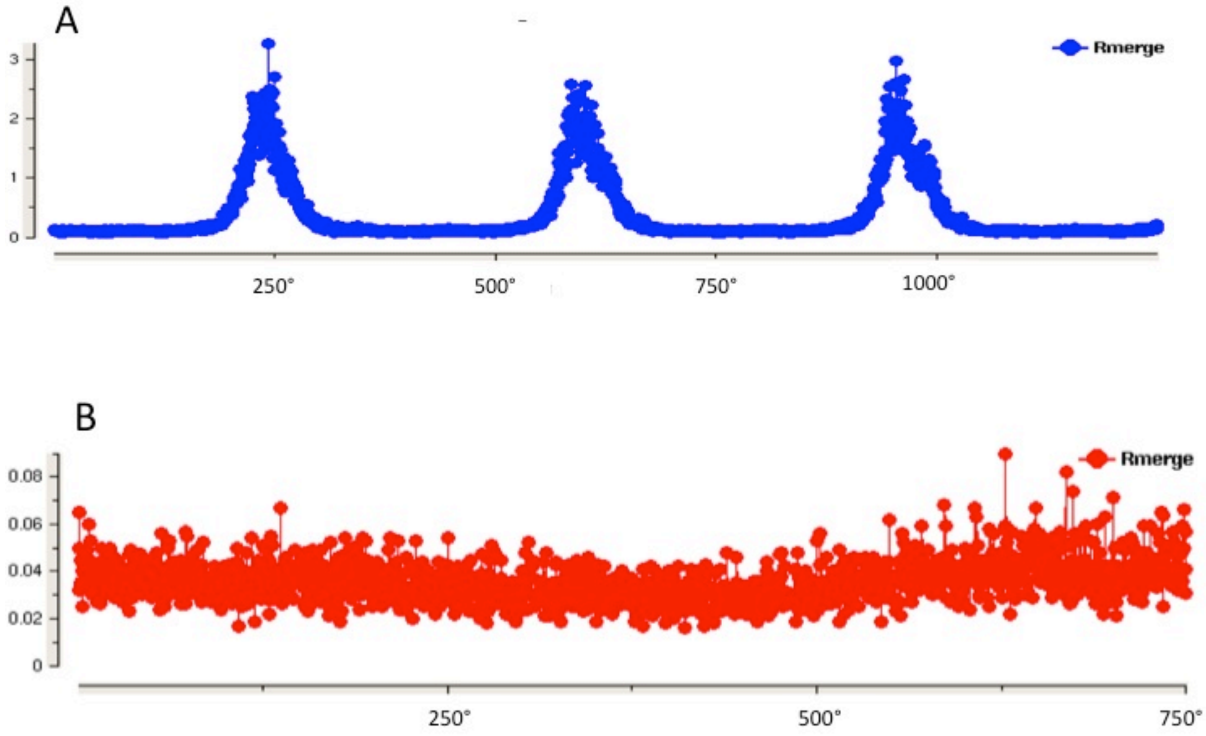


Figure 95. Anisotropy of TSWV N protein crystals soaked with samarium nitrate. (A) R_{merge} vs rotation in degress (°) for TSWV N protein crystals soaked with 1 mM of samarium nitrate for 2 min diffracted to 2.8 Å. (B) R_{merge} vs rotation in degress (°) for native TSWV N protein crystals diffracted to 2.7 Å.

4.8 Discussion of TSWV N protein results chapter

4.8.1 Introduction

The overall aim of this project was to better understand the evolution of the plant infecting tospoviruses by comparing the high resolution TSWV N protein structure with the N proteins from the three animal infecting genera for which high resolution data are already available. A high resolution TSWV N protein would also provide important information about RNA binding, oligomerization and RNP formation.

The initial goal was to express and purify highly concentrated recombinant TSWV N protein suitable for crystallization trials. To address this aim, the same approach used for expression and purification of CCHFV N protein was implemented to express recombinant TSWV N protein N-terminally linked to a His_6-SUMO tag.

By generating abundant near native TSWV N protein it was demonstrated by SEC that TSWV N protein has the propensity to bind *E. coli* RNA, which could be removed by including a high salt wash. The RNA free TSWV N protein exists as high molecular weight complexes, similar to other members of the *Bunyaviridae* family. Finally X-ray crystallography was implemented to characterize the atomic structure of the TSWV N protein. Crystals diffracting to 2.6 Å were produced from the TSWV N protein, and displayed $P2_1$ symmetry, with unit cell dimensions a = 72.4, b = 105.9, c = 79.8 Å, α= 90°, β= 105.8°, γ= 90°.

5.8.2 Expression and purification

The results detailed in chapter 4 provide the overexpression and purification protocol for making soluble TSWV N protein suitable for crystallization. Chapter 4 also details the crystallization and high resolution X-ray diffraction of the TSWV N protein. During writing of this thesis Komoda et al (Komoda et al., 2013), reported the expression of the TSWV N protein using an N-terminal fused tag including a D-box (MNHK), a His_6 tag (HHHHHH), and a TEV cleavage site (ENLYFQG). The near native TSWV N fusion protein was shown to form predominately trimers in solution and was crystallized in 0.1 M Na-HEPES pH 7.5, 25% (w/v) PEG 1,000, 50 mM NaCl by the hanging drop vapour-diffusion method. The authors of this work also reported TSWV N protein crystals diffracted to 3.25 Å, and belonged to spacegroup $P2_1$ with unit cell parameters a = 66.8, b = 97.2, c = 72.0 Å, α= 90°, β= 112.8°, γ= 90°. Although Komoda et al have not yet published the structure they concluded they were able generate phases from selenomethionine TSWV N protein crystals (Komoda et al., 2013). Although this project has not been successful in generating phases and determining the three-dimensional structure of the native TSWV N protein, it has established a protocol for overexpression and purification.

4.8.3 RNA binding

Initially the TSWV N protein was bound to *E. coli* RNA of various lengths, and existed as higher molecular weight oligomers, which eluted near the void volume of the Superdex S75 SEC column. The length of the RNA based on molecular weight markers was between 50 and 150 nucleotides. Following RNase treatment the TSWV N protein still remained bound to RNA, and eluted as two predominant oligomeric profiles using SEC, and most likely corresponded to TSWV N trimer and tetramer proteins. The extracted RNA from both species displayed two distinct bands on the agarose gel, and each migrated below the 50 nucleotide marker at different speeds. The migration of the RNA from the first peak (Figure 81), migrated to a similar distance to the RNA extracted from the similar sized BUNV N tetramer protein, which binds 12 nucleotides of RNA per N protein monomer, suggesting the TSWV N protein may bind RNA using similar stoichiometry. Interestingly the RNA bound to tetramer TSWV N protein, migrated more slowly than the RNA from the trimer TSWV N protein species, which suggests the RNA is protected from RNase, possibly by a hydrophobic cavity similar to that seen for orthobunyaviruses and phleboviruses.

4.8.4 Crystallization, self-rotation and phasing of the TSWV N protein

Factorial screening of both the non-RNase and RNase treated trimer/tetramer TSWV N protein failed and yielded no protein crystals. Co-purification of bacterial expressed nsRNA virus N proteins and RNA is not uncommon (Ariza et al., 2013, Li et al., 2013), and can be beneficial for elucidating important mechanisms such as RNA binding and RNP assembly. However one inherent issue that arises from attempting to crystallize N protein with bacterial RNA is the heterogeneity problem arising from introducing different bacterial RNAs of various sequences and lengths into the crystal, which can reduce the quality of diffraction, or make crystallization less likely.

To improve the probability of yielding diffracting protein crystals the *E. coli* RNA was removed and separated from the TSWV N protein using a high salt wash during the last step in the nickel-affinity and first SEC purification respectively. Once the RNA has been removed the TSWV N protein exists as tetramers and trimers in solution, and both species can be crystallized. However only crystals from the tetramer sized TSWV N protein produced diffraction suitable for structural determination.

Interestingly Komoda et al (Komoda et al., 2013) described the expression and purification of TSWV N protein forming trimers, which produced both native and selenomethionine protein crystals. One possibility is the crystallization of the trimer complex of TSWV N protein used by Komoda et al is more favorable for crystallization due to the extra uncleaved N-terminal amino acids providing new crystal contacts between the monomers. Strikingly both crystals from TSWV N protein tetramer in this work and the trimer protein by Komoda et al (Komoda et al., 2013), have $P2_1$ symmetry and display unit cells which are very similar.

In contrast to the work by Komoda et al this work has been able to produce TSWN N protein free from any fusion or His_6 tags, and therefore is more biologically relevant. Once solved, having TSWV N protein free of any N-terminal tags will provide more valuable information into such important functions such as RNA binding, oligomerization and possibly RNP formation. The native TSWV N protein structure will also provide higher resolution data and will improve the allocation of side chains during the refinement process. The TSWV N protein produced in this work is also suitable for further biochemical analysis such as fluorescence anisotropy to establish an RNA binding assay for the TSWV N protein in order to determine the best RNA for co-crystallization.

For TSWV N protein, with a molecular weight of 28,914 Da, the Matthews coefficient (Matthews, 1968), suggested four TSWV N proteins per AU. The self-rotation analysis suggested the presence of two fold crystallographic symmetry, which relates to the $P2_1$ spacegroup. Additionally the self-rotation revealed two two-fold NCS axes on the $\kappa=180°$ plot and a four-fold NCS axis on the $\kappa = 90°$ plot, which suggests a tetramer is making up the AU (the four-fold non-crystallographic symmetry relates to the axis looking down into the tetramer and the two two-fold NCS corresponds to the symmetry along the same axis (Figure 96). Based on the Matthews co-effiecient self- and rotation analysis it is likely that a TSWV N protein tetramer occupies the AU and is made up of a dimer of dimers, possibly in a ring configuration, resembling the closely relating orthobunyavirus tetramer. Molecular replacement was attempted using the orthobunyavirus N protein tetramer, however this failed.

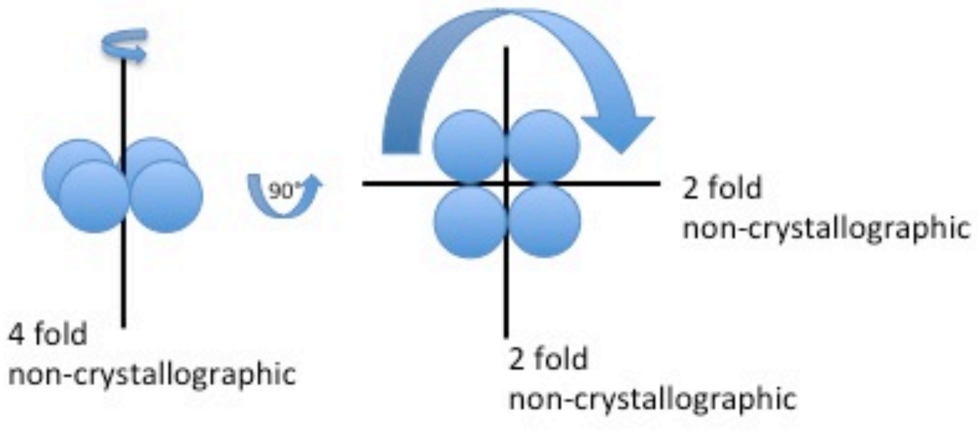
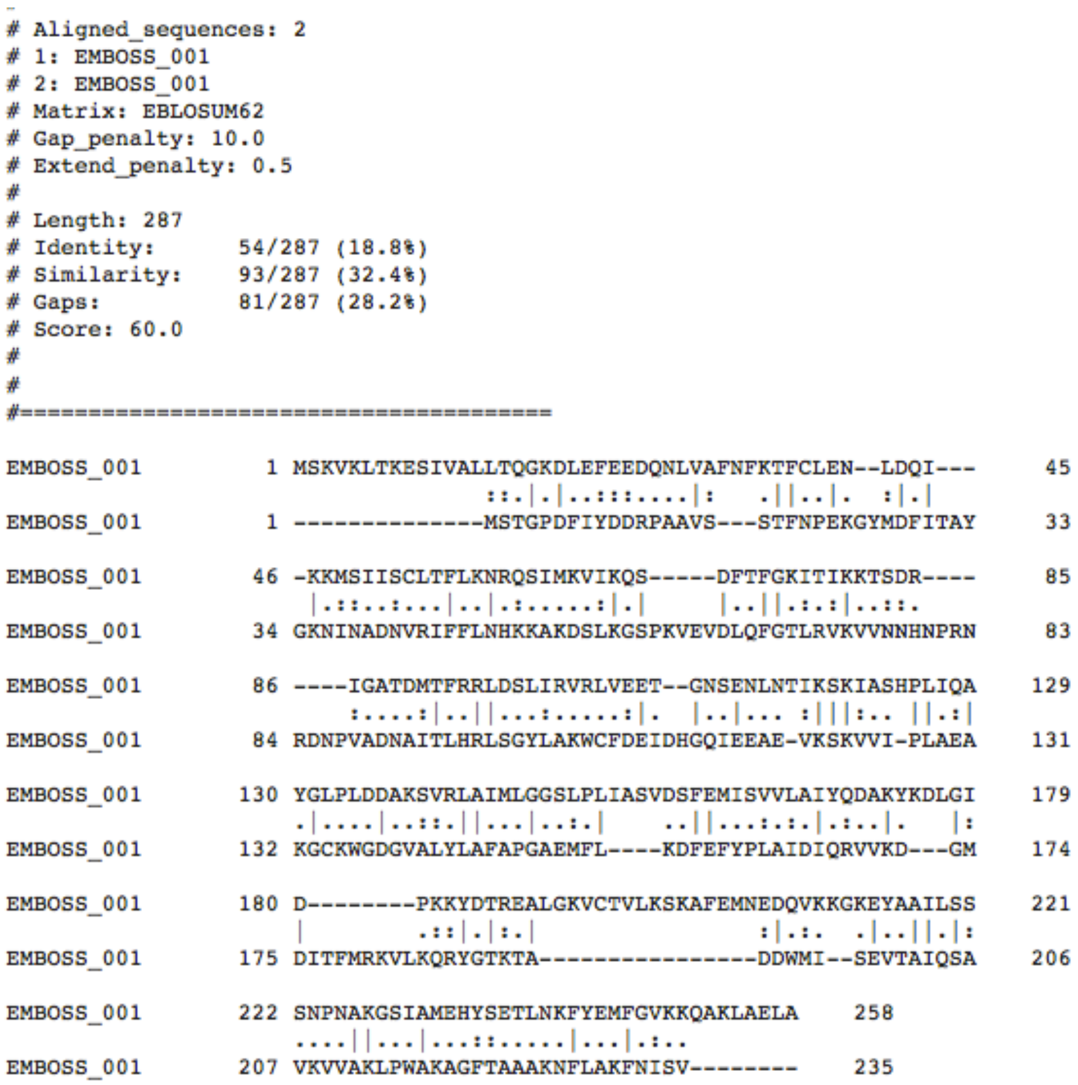


Figure 96. Interpretation of TSWV N protein $P2_1$ self-rotation results. Most probable arrangement of the TSWV N protein tetramer based on the κ=180° and κ=90° plots.

When combining diffraction data for MIRAS to solve the phases needed for substructure solution, derivative crystals soaked in heavy atoms need to be isomorphic, but heavy atom binding can often cause non-isomorphism between derivative crystals. Future work using AutoSol to generate sufficient phase information for substructure solution could involve combining datasets and performing MIRAS using only isomorphous data sets collected from several derivative crystals soaked with several different heavy atoms (Terwilliger et al., 2009).



```
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 287
# Identity:      54/287 (18.8%)
# Similarity:    93/287 (32.4%)
# Gaps:          81/287 (28.2%)
# Score: 60.0
#
#
#=======================================

EMBOSS_001         1 MSKVKLTKESIVALLTQGKDLEFEEDQNLVAFNFKTFCLEN--LDQI---     45
                                   ::.|.|..:::....|:   .||..|.  :|.|
EMBOSS_001         1 --------------MSTGPDFIYDDRPAAVS---STFNPEKGYMDFITAY     33

EMBOSS_001        46 -KKMSIISCLTFLKNRQSIMKVIKQS-----DFTFGKITIKKTSDR----     85
                      |.::..:...|..|.:.....:|.|     |..||.:.:|..::.
EMBOSS_001        34 GKNINADNVRIFFLNHKKAKDSLKGSPKVEVDLQFGTLRVKVVNNHNPRN     83

EMBOSS_001        86 ----IGATDMTFRRLDSLIRVRLVEET--GNSENLNTIKSKIASHPLIQA    129
                         :....:|..||...:.....:|.  |..|... :|||:.. ||.:|
EMBOSS_001        84 RDNPVADNAITLHRLSGYLAKWCFDEIDHGQIEEAE-VKSKVVI-PLAEA    131

EMBOSS_001       130 YGLPLDDAKSVRLAIMLGGSLPLIASVDSFEMISVVLAIYQDAKYKDLGI    179
                     .|....|..::.||...|..:.|    ..||...:.:.|.:..|.   |:
EMBOSS_001       132 KGCKWGDGVALYLAFAPGAEMFL----KDFEFYPLAIDIQRVVKD---GM    174

EMBOSS_001       180 D--------PKKYDTREALGKVCTVLKSKAFEMNEDQVKKGKEYAAILSS    221
                     |        .::|.|:.|                :|.:.  .|..||.|:
EMBOSS_001       175 DITFMRKVLKQRYGTKTA----------------DDWMI--SEVTAIQSA    206

EMBOSS_001       222 SNPNAKGSIAMEHYSETLNKFYEMFGVKKQAKLAELA    258
                     ....||...|...::.....|...|.:..
EMBOSS_001       207 VKVVAKLPWAKAGFTAAAKNFLAKFNISV--------    235
```

Figure 97. Pairwise Sequence Alignment (PROTEIN) of the TSWV and LEAV N protein amino acid sequences. http://www.ebi.ac.uk/Tools/psa/emboss_needle/.

No molecular replacement could be found using the CCHFV, RVFV, LASV, or orthobunyavirus N protein crystal structures. Interestingly an analysis of the secondary structure prediction of TSWV N protein using the HHpred server, identified the highest scoring structural alignment with structures published in PDB, was the orthobunyavirus N protein. This is not surprising considering the TSWV N and LEAV N proteins share similar sizes and belong to the same viral family (Figure 97).

Further analysis of the bunyavirus N protein from all genera has supported the close evolutionary relationship shared by orthobunyaviruses and tospoviruses (Figure 64B). The predicted structural homology between the orthobunyaviruses and tospoviruses suggest it is likely their N proteins do share similar architecture in the overall fold; however this is not supported by sequence alignment, which revealed they share only 18.8% sequence similarity (close to random) (Figure 97). There could be two reasons why the molecular replacement has not worked, either they share some tertiary structure which is not similar enough for the orthobunyavirus N protein to generate the correct phases, or they possess completely different folds. Interestingly although the N-terminal domain of LASV NP displayed structural homology to the globular domain of CCHFV N protein, it was not similar enough as a molecular replacement model to generate correct phases.

The next stage of the project attempted to produce selenomethionine TSWV N protein crystals to generate anomalous signal needed to solve the TSWV N protein structure. Generation of the selenomethionine TSWV N protein entailed using the protocol for the expression and purification of native TSWV N protein, which was successful. However the derivative TSWV N protein could not reproduce crystals using the native condition. One crystallization condition (0.2 M trimethylamine N-oxide dihydrate, 0.1 M Tris-HCl pH 8.5, 20% PEG-2,000 MME), which was originally found by factorial screening using the native protein did yield TSWV N selenomethionine crystals, however these were of a much poorer quality and displayed severe anisotropy. In replacing the sulphur atoms of the methionine side chains in the TSWV N protein with selenium a change to the properties of the TSWV N protein may have resulted, such as solubility or may have affected important crystal contacts.

Heavy atom soaks introduced anisotropy into the native TSWV N protein crystals. Because proteins within a crystal are not equally ordered in all directions of the lattice, a direct consequence of this inherent anisotropic packing is that protein crystals do not usually diffract X-rays equally well in all directions of the lattice. There was very little anisotropy in the diffraction patterns of the native TSWV N protein crystals, however all heavy atom soaks resulted in the crystallographic datasets being anisotropic and the diffraction spots were smeared. The failure to find the differences in intensity needed to find positions of the heavy atoms within the samarium nitrate crystals may be the combined result of low samarium occupancy, anisotropic disorder and poor diffraction spot quality (Figure 94). Additionally a low symmetry spacegroup like $P2_1$, causes more of a challenge when sufficient redundancy is needed, coupled with overexposure of high intensity X-rays and limiting radiation damage.

5. Chapter 5 - Overall conclusions and future directions

5.1 Future directions for the CCHFV N protein

The crystal structure of the CCHFV N protein has guided mutagenesis of the CCHFV N protein and pin pointed important amino acids, which may be involved in RNA binding and multimerization. Following mutation of the S segment, targeted amino acids can be analysed by testing the activity of functional RNPs within mammalian cells using the mini-genome system. Furthermore the high resolution structure of the CCHFV N protein has allowed alternative roles of the CCHFV N protein to be analysed, including caspase-3 protease cleavage. Alternative roles of bunyavirus N proteins have also been demonstrated in the *Hantavirus* genus and include functions relating to viral gene expression, such as cap snatching and manipulation of the innate immune response within host cells.

The two CCHFV proteins found in the AU of the *C2* and $P2_12_12_1$ crystal forms could be authentic RNP interactions or they could be packing against each other simply as lattice contacts within the crystal. Analysis of fully assembled RNPs from the mini-genome or deactivated RNPs from infectious CCHFV virus should be visualized using EM techniques, such as cryo-EM, which could establish any clues to the nature of RNP assembly for CCHFV, and highlight similarities to those proposed for LASV, IAV, RSV, RVFV and BUNV.

The crystal structure of the CCHFV N protein has revealed its globular domain exhibits a high degree of similarity with the N-terminal domain of the NP from the two segmented sNSV LASV, a member of the *Arenaviridae* family. Strikingly the CCHFV N protein displays very little tertiary structural similarity with any N protein within its own family, including orthobunyaviruses and phleboviruses for which high resolution structural data are available. The finding that orthobunyavirus and phlebovirus N proteins appear essentially structurally unrelated to the CCHFV N protein, which itself shows extremely high similarity with the LASV NP, agrees with important phylogenetic relationships between nairoviruses and arenaviruses which had previously been highlighted by Vieth et al (Vieth et al., 2007). However this observation is not further supported by the phylogenetic analysis of N protein amino acid sequences (Figure 64B), and suggests the nairovirus and areanavirus are not closely related but do have structurally similar N proteins that most likely have evoled divergently through a once common ancestor.

Arenaviruses and nairoviruses also share unique aspects of cellular biology that are absent from the remaining genera in the *Bunyaviridae* family, such as glycoprotein precursor processing by SKI-1 proteases (Lenz et al., 2001, Vincent et al., 2003). Taken together this work confirms important phylogenetic evidence and provides the most convincing evidence yet that nairoviruses have an ancestor in common with a past member the of *Arenaviridae* family, and secondly suggests the accepted classification of bunyaviruses as three segmented RNA viruses may be over generalised and that the apparent diversity within the *Bunyaviridae* family highlighted by this work may have implications for taxonomic status.

The crystal structure of the CCHFV N protein also has revealed that caspase-3 protease DEVD motif is located at the apex of the arm domain. Not only is the DEVD motif conserved among CCHFV strains, but also the exposed nature of the DEVD motif, suggested cleavage would benefit the CCHFV life cycle, and does not represent a host cell defence mechanism against the virus.

The homo-oligomeric form of the CCHFV N protein captured in the crystal form published by Wang et al (Wang et al., 2012), revealed the DEVD site was located directly at the point of the head-to-tail interactions of two CCHFV N proteins. Wang et al revealed that the DEVD motif would be protected within this location and obscures binding and subsequent cleavage by the caspase-3 protease. Wang et al also reported that the monomeric is more susceptible to cleavage by caspase-3 protein upon adding a short ssRNA mimicking a primer, and also observed increased RNP reporter activity when inhibiting cleavage by mutating the DEVD motif when measured in the context of the mini-genome system. This work does however does not explain any reasons why they chose the length of the primer ssRNA mimic, or why it would change the conformation of the CCHFV N protein. Additionally they did not provide any link between primer ssRNA, caspase-3 motif and replication of the virus. Furthermore Wang et al also linked this result with their previous work, which reported that the cleavage of CCHFV N protein was abrogated by the inhibition of apoptosis and increased the yield of progeny virus by 80–90% (Karlberg et al., 2011). This work failed to point out that by inhibiting apoptosis there would be more viable cells available for virus infection and therefore more progeny viruses.

Unfortunately, the functional relevance of the DEVD motif cannot be tested within the context of *in vivo* CCHFV infection in mammalian cells, because the CCHFV genome cannot be manipulated and there is no current rescue system, similar to other nsRNA viruses such as IAV.

Interestingly caspase protease cleavage sites are not just limited to CCHFV N protein and LASV NP, but also include the NP of human-infecting strains of IAV. The caspase cleavage site of IAV NP has been shown to possess important roles in the virus life cycle (Zhirnov et al., 1999, Zhirnov and Syrtzev, 2009). Mutations, which abrogate caspase cleavage at an N-terminal recognition site, abolished the ability of the virus to be rescued, whereas infectious virus with mutations located within a C-terminal caspase cleavage site rapidly reverted to the wild type virus to restore cleavability. These findings indicate that such alterations are important to virus fitness and viability, and therefore pathogenesis, suggesting in addition to RNA binding a second functional role exists, which could be targeted by antivirals.

5.2 Future directions for the TSWV N protein

Despite attempts to generate correct phases for the TSWV N protein failing, this work did establish a reliable protocol for producing soluble and stable TSWV N protein suitable for crystallization and *in vitro* biochemical studies. This protocol was used to produce selenomethionine TSWV N protein successfully, however this derivative protein was unable to crystallize in the same condition as the native protein or all the alternative conditions that were tested. The failure to crystallize selenomethionine derivative TSWV N protein in the native condition is not actually an obstacle, because other strategies including re-screening, seeding, and micro seeding (Zhu et al., 2005) can be employed to generate diffraction quality crystals needed to generate phases. However in this case the TSWV N selenomethionine protein was unable to crystallize in the native condition through seeding from native crystals and from those selenomethionine crystals grown from in 0.2 M trimethylamine N-oxide dihydrate, 0.1 M Tris-HCl pH 8.5, 20% PEG-2,000 MME. Additionally no selenomethioine TSWV N protein crystals were generated thorugh re-screening factorial conditions, including using different temperatures and micro seeding using fresh crystals from native crystals and selenomethionine TSWV N protein crystals generated through the condition 0.2 M trimethylamine N-oxide dihydrate, 0.1 M Tris-HCl pH 8.5, 20% PEG-2,000 MME.

In order to pursue this project to generate correct phases to solve the TSWV N protein crystal structure a number of avenues could be employed. Firstly the trimer TSWV N protein reported by Komoda et al (Komoda et al., 2013), was used to generate diffracting crystals from which phases were generated. The trimer TSWV N protein in this work did not generate any well diffracting crystals in factorial screens and instead the tetramer TSWV N protein crystals were pursued. By testing the condition by Komoda et al, selenomethionine trimer TSWV N protein crystals could be reproduced, from which phases could be obtained and the resulting model could be used in MR to solve the tetramer TSWV N protein diffraction data.

Perhaps a more practical approach could involve attempting to screen the native TSWV N tetramer protein again for crystallization and attempt to find an alternative condition, which yields crystals possessing a higher quality of X-ray diffraction. By using TSWV N tetramer protein crystals with a higher symmetry spacegroup, and higher resolution would be more beneficial for phasing: firstly less data will be needed to reach the needed redundancy required to unambiguously locate the positions of the heavy atoms and secondly improved resolution could be useful in attempting to solve for phases using anomalous scattering from the sulphur atoms within the TSWV N protein.

5.3 Futurework

Elucidation of the crystal structure of TSWV and CCHFV N protein/RNA complex is also needed to establish which amino acids take part in binding viral RNA. Fluorescence anisotropy could be used to establish an RNA binding assay for the CCHFV N protein, in order to determine the best RNA for co-crystallization. Co-crystallization of the TSWV and CCHFV N proteins with RNA would not only be more biologically more relevant than the apoprotein form, but also may provide previously unidentified interactions between promoters and clues into the TSWV and CCHFV RNP assembly. Fluorescence anisotropy has been used for the SBV and BUNV proteins and can also be used to analyse mutants designed after X-ray crystallographic analysis.

Further work could also involve analysing recombinant TSWV N protein and digested endogenous viral RNPs by direct visualization using negative stain EM, there by generating single particle reconstructions to compare the oligomer state of each and establish whether either the tetramer or trimer proteins seen solution are in anyway biologically relevant, or just artefacts. This type of analysis could also determine how homogeneous the TSWV N protein is in solution, which could have implications for crystallization, especially for the selenomethionine protein. To establish the precise oligomeric state of the protein within the RNP viewing undigested authentic RNPs from infectious TSWV would be needed. Using negative stain EM, or cryo-EM to directly visualize viral RNPs has been achieved for many nsRNA viruses and greatly improved the understanding in how RNPs assemble and how viral RNA is protected from the exterior environment of the cell.

The same methods used to express and purify the CCHFV N protein, could also be used to express, purify and crystallize other N proteins within the *Bunyaviridae* family, such as those within the *Hantavirus* genus.

5.4 *Bunyaviridae* family RNPs conclusion

The X-ray crystallographic studies in this thesis have shed new light on the N/ protein RNA and N/N protein interactions involved in CCHFV RNP formation. There are now published studies for which high resolution structures are available for the nairovirus, orthobunyavirus and phlebovirus N protein/RNA and N/N protein complexes. Comparison of bunyavirus N protein structures has revealed that they not only possess unique folds but also use different mechanisms for RNP formation. The most important conserved feature among the bunyavirus N proteins is their ability to possess a positively charged RNA binding groove.

Cryo-EM data of the orthobunyavirus RNP suggests a “beads on a string” complex, which takes up a compacted helical form for packing within the virus particle and a flexible form allowing the RNA to be become available for RNA synthesis. The two apparent conformations of the orthobunyavirus RNP are however in contention with the proposed model for phlebovirus RNP, which has only been shown to form flexible single RNP complexes. The larger size of the CCHFV N protein and absence of any apparent C- and N-terminal extensions may even further point towards less commonality between bunyavirus RNP complexes and indicate that RNP formation in the nairovirus genus is noticeably different. One possibility is the CCHFV N protein may employ the arm domain for oligomization as an alternative to the N- and C-terminal arms used by the phleboviruses and orthobunyaviruses. In fact the involvement of the arm domain in oligomerization has been observed in the homo-oligomers shown within the crystal by Wang et al (Wang et al., 2012). The work by Wang et al (Wang et al., 2012), needs to be however validated by direct visualization of the endogenous nairovirus RNPs using cryo-EM, as the observed homo-oligomers could be artifacts of crystal packing.

Both X-ray crystallographic and cryo-EM data of N proteins from nsNSVs and sNSVs (including bunyaviruses) have revealed they possess features with stricking similarities and differences. For example structural comparison of the N proteins from phleboviruses, orthobunyaviruses, RSV (nsNSV) and VSV (nsNSV), have revealed they share commonality in the way they protect the viral RNA, within the centre of the N- and C- terminal portions of the N protein.

Recent reported structures of the phlebovirus N protein have shown they differ from their sNSV (IAV) and nsNSV (RSV and VSV) N protein counterparts, by displaying a non-helical filament RNP structure. There is evidence however to suggest that orthobunyaviruses share similarity with the RSV, VSV and IAV N proteins by facilitating the formation of a helical structures to encapsidate their viral RNA.

An additional major difference is the IAV NP facilitates oligomerization by a C-terminal tail loop, whilst phleboviruses in contrast extend an N-terminal arm towards the neighboring partner N protein. Contrastingly the orthobunyavirus N protein clearly possesses both an N- and C-terminal arm extensions, which is paralleled in RNP formation in nsNSVs (Tawar et al., 2009). Not surprisingly because of the clear structural differences between the inter-N interactions amoung nsNSVs the N- and C-terminal extensions are not conserved between the bunyaviruses.

Recent evidence from (Wang et al., 2012, Guo et al., 2012), including evidence within this thesis suggests the CCHFV N protein does not contain a cavity within the globular domains where the RNA is sequestered. The crystal structure of the CCHFV N proteins also reveals no discernible N- and C-terminal loops which would be suggestive of a role in oligomerization. The arm of the CCHFV N protein is a likely candidate for the main contributor to oligomerization, and the homo-oligomer structure captured in a crystal structure by Wang et al (Wang et al., 2012) supports this. Furthermore the CCHFV and LASV N/NP proteins, share very close structural homology, and both display possible oligomerization domain arms, however the possibility that the CCHFV N protein requires N- and C-terminal extensions for oligomerization cannot be ruled out.

The mechanism for nairovirus RNP formation likely has similarity to LASV, however more data is needed to fully uncover the exact steps involved. LASV has been proposed to undergo a conformational shift upon RNA binding, and it is this shift which has been suggested to facilitate subsequence oligomerization into the functional RNP complex. The nairovirus N protein may undergo a similar transition and reveal the appropriate RNA binding and oligomerization surfaces when the correct viral RNA sequences are present or alternatively by an as yet unknown cellular factor.

Many questions can now be addressed from the advent of the orthobunyavirus RNP structure such as how does the N protein maintain a non-aggregated state competent for RNP encapsidation and does it selectively package its viral genomic RNA over host RNA, and if so what mechanism is used to pack the three RNA segments. Moreover how does the orthobunyavirus adapt its structure to allow readout of nucleotide sequences by the polymerase during RNA synthesis?

Such questions are also likely to be tackled for CCHFV and TSWV RNP formation using similar EM/X-ray crystallography complementation approaches, and are crucial for our understanding of not just nairovirus and tospovirus biology, but also the biology of the *Bunyaviridae* family as a whole, which will likely lead improved prevention and therapeutic intervention.

Although the studies within this thesis answer some key questions about CCHFV N protein RNA binding and oligomerization, there are still many questions that can be posed. For example how does the CCHFV N protein facilitate oligomerization? One possibility is the principal involvement of the arm domain, like that that shown in the oligomer structure by Wang et al (Wang et al., 2012). Or alternatively it could use a similar mechanism to that of the phlebovirus and orthobunyavirus N protein's involving the exposure of N- and C-terminal arms, through major conformational changes and also allow for the exposure of the RNA binding surfaces needed for RNP formation. Another question could be does the CCHFV N protein need RNA before it can oligomerize, and if so are their any nucleotide or amino acid specific sequences required to initiate RNA binding

The greater our knowledge in the way these proteins function in RNA binding and participate in homo-oligomerization, the better our chances of being able to target these important interactions. For now elucidation of the CCHFV N protein crystal structure can be used to design inhibitors using *in silico* based methods, which could potentially prevent RNA binding, the N-N protein interaction to ultimately hinder RNP formation (Wade, 1997). However a much more in-depth understanding of nairovirus RNP formation is needed. Testing of these inhibitors could then be re-laid back through the mini-genome system and further uncover characteristics of the CCHFV N protein.

6. Appendix

6.1 Primers for all cloning

6.1.1 CCHFV N gene cloning

Forward 5’ ACT GAT GGA TCC ATG GAA AAC AAA ATC GAA GTG AAC AGT AAG G 3’
Reverse 5’ GCA ATC TCG AGT TAG ATG ATG TTG GCG CTG GTG GCA TTC C 3’

6.1.2 TSWV N gene cloning

Forward 5’ ACT GGT GGA TCC ATG TCT AAG GTT AAG CTC 3’
Reverse 5’ ATC AGT CTC GAG TCA AGC AAG TTC TGC GAG 3’

6.2 Ramachandran plot of *C*2 CCHFV N protein crystal structure

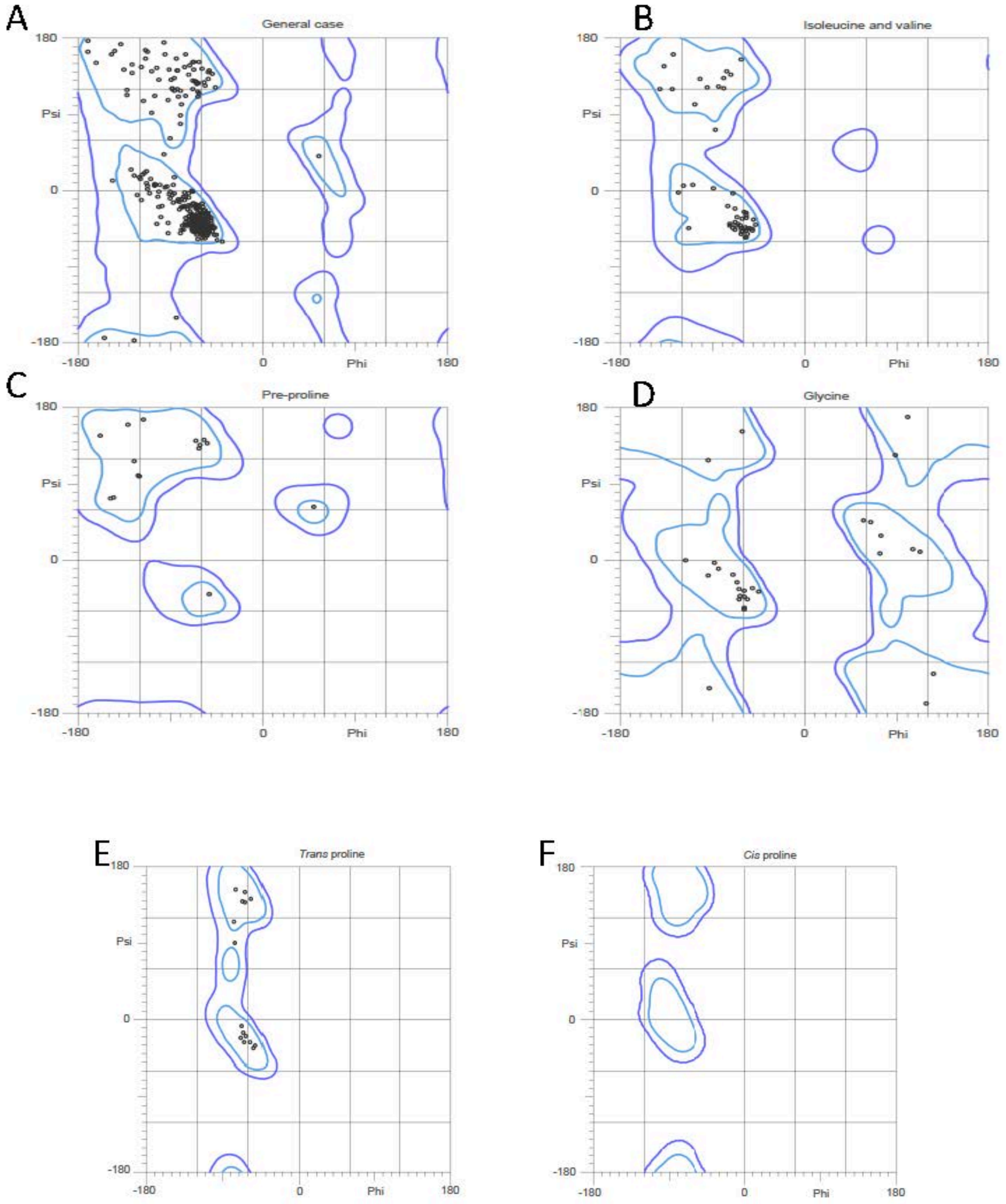


Figure 98. Ramachandran plots for *C*2 CCHFV N protein crystal structure. Plots were generated by molprobity and individual plots are shown for (A) All residues, (B) isoleucine and valine, (C) Pre-proline (D) Glycine, (E) Trans proline, (F) Cis proline. 98.8% (479/485) of all residues were in favoured (98%) regions and 100% (485/485) of all residues were in allowed (>99.8%) regions, with no outliers.

7. References

1994. The CCP4 suite: programs for protein crystallography. *Acta crystallographica. Section D, Biological crystallography,* 50**,** 760-3.

ADKINS, S. 2000. Tomato spotted wilt virus-positive steps towards negative success. *Molecular plant pathology,* 1**,** 151-7.

AKUTSU, M., YE, Y., VIRDEE, S., CHIN, J. W. & KOMANDER, D. 2011. Molecular basis for ubiquitin and ISG15 cross-reactivity in viral ovarian tumor domains. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 2228-33.

ALBERTINI, A. A., WERNIMONT, A. K., MUZIOL, T., RAVELLI, R. B., CLAPIER, C. R., SCHOEHN, G., WEISSENHORN, W. & RUIGROK, R. W. 2006. Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science,* 313**,** 360-3.

ALFADHLI, A., LOVE, Z., ARVIDSON, B., SEEDS, J., WILLEY, J. & BARKLIS, E. 2001. Hantavirus nucleocapsid protein oligomerization. *J Virol,* 75**,** 2019-23.

ALTAF, A., LUBY, S., AHMED, A. J., ZAIDI, N., KHAN, A. J., MIRZA, S., MCCORMICK, J. & FISHER-HOCH, S. 1998. Outbreak of Crimean-Congo haemorrhagic fever in Quetta, Pakistan: contact tracing and risk assessment. *Trop Med Int Health,* 3**,** 878-82.

ALTAMURA, L. A., BERTOLOTTI-CIARLET, A., TEIGLER, J., PARAGAS, J., SCHMALJOHN, C. S. & DOMS, R. W. 2007. Identification of a novel C-terminal cleavage of Crimean-Congo hemorrhagic fever virus PreGN that leads to generation of an NSM protein. *Journal of virology,* 81**,** 6632-42.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *Journal of molecular biology,* 215**,** 403-10.

ANAGNOSTOU, V. & PAPA, A. 2009. Evolution of Crimean-Congo Hemorrhagic Fever virus. *Infect Genet Evol,* 9**,** 948-54.

ANDERSSON, I., BLADH, L., MOUSAVI-JAZI, M., MAGNUSSON, K. E., LUNDKVIST, A., HALLER, O. & MIRAZIMI, A. 2004a. Human MxA protein inhibits the replication of Crimean-Congo hemorrhagic fever virus. *Journal of virology,* 78**,** 4323-9.

ANDERSSON, I., SIMON, M., LUNDKVIST, A., NILSSON, M., HOLMSTROM, A., ELGH, F. & MIRAZIMI, A. 2004b. Role of actin filaments in targeting of Crimean Congo hemorrhagic fever virus nucleocapsid protein to perinuclear regions of mammalian cells. *Journal of medical virology,* 72**,** 83-93.

APPANNANAVAR, S. B. & MISHRA, B. 2011. An update on crimean congo hemorrhagic Fever. *Journal of global infectious diseases,* 3**,** 285-92.

ARADAIB, I. E., ERICKSON, B. R., MUSTAFA, M. E., KHRISTOVA, M. L., SAEED, N. S., ELAGEB, R. M. & NICHOL, S. T. 2010. Nosocomial outbreak of Crimean-Congo hemorrhagic fever, Sudan. *Emerg Infect Dis,* 16**,** 837-9.

AREA, E., MARTIN-BENITO, J., GASTAMINZA, P., TORREIRA, E., VALPUESTA, J. M., CARRASCOSA, J. L. & ORTIN, J. 2004. 3D structure of the influenza virus polymerase complex: localization of subunit domains. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 308-13.

ARIZA, A., TANNER, S. J., WALTER, C. T., DENT, K. C., SHEPHERD, D. A., WU, W., MATTHEWS, S. V., HISCOX, J. A., GREEN, T. J., LUO, M., ELLIOTT, R. M., FOOKS, A. R., ASHCROFT, A. E., STONEHOUSE, N. J., RANSON, N. A., BARR, J. N. & EDWARDS, T. A. 2013. Nucleocapsid protein structures from orthobunyaviruses reveal insight into ribonucleoprotein architecture and RNA polymerization. *Nucleic acids research*.

ARRANZ, R., COLOMA, R., CHICHON, F. J., CONESA, J. J., CARRASCOSA, J. L., VALPUESTA, J. M., ORTIN, J. & MARTIN-BENITO, J. 2012. The structure of native influenza virion ribonucleoproteins. *Science,* 338**,** 1634-7.

ASCIOGLU, S., LEBLEBICIOGLU, H., VAHABOGLU, H. & CHAN, K. A. 2011. Ribavirin for patients with Crimean-Congo haemorrhagic fever: a systematic review and meta-analysis. *The Journal of antimicrobial chemotherapy,* 66**,** 1215-22.

ASHER, D. M., GOUDSMIT, J., POMEROY, K. L., GARRUTO, R. M., BAKKER, M., ONO, S. G., ELLIOT, N., HARRIS, K., ASKINS, H., ELDADAH, Z. & ET AL. 1988. Antibodies to HTLV-I in populations of the southwestern Pacific. *J Med Virol,* 26**,** 339-51.

ATHAR, M. N., KHALID, M. A., AHMAD, A. M., BASHIR, N., BAQAI, H. Z., AHMAD, M., BALOUCH, A. H. & BASHIR, K. 2005. Crimean-Congo hemorrhagic fever outbreak in Rawalpindi, Pakistan, February 2002: contact tracing and risk assessment. *Am J Trop Med Hyg,* 72**,** 471-3.

BAKSHI, S., HOLZER, B., BRIDGEN, A., MCMULLAN, G., QUINN, D. G. & BARON, M. D. 2013. Dugbe virus ovarian tumour domain interferes with ubiquitin/ISG15-regulated innate immune cell signalling. *J Gen Virol,* 94**,** 298-307.

BARR, J. N., ELLIOTT, R. M., DUNN, E. F. & WERTZ, G. W. 2003. Segment-specific terminal sequences of Bunyamwera bunyavirus regulate genome replication. *Virology,* 311**,** 326-38.

BARR, J. N., RODGERS, J. W. & WERTZ, G. W. 2005. The Bunyamwera virus mRNA transcription signal resides within both the 3' and the 5' terminal regions and allows ambisense transcription from a model RNA segment. *Journal of virology,* 79**,** 12602-7.

BARR, J. N. & WERTZ, G. W. 2004. Bunyamwera bunyavirus RNA synthesis requires cooperation of 3'- and 5'-terminal sequences. *Journal of virology,* 78**,** 1129-38.

BARR, J. N. & WERTZ, G. W. 2005. Role of the conserved nucleotide mismatch within 3'- and 5'-terminal regions of Bunyamwera virus in signaling transcription. *Journal of virology,* 79**,** 3586-94.

BATTISTI, A. J., CHU, Y. K., CHIPMAN, P. R., KAUFMANN, B., JONSSON, C. B. & ROSSMANN, M. G. 2011. Structural studies of Hantaan virus. *J Virol,* 85**,** 835-41.

BAULCOMBE, D. C. 1996. RNA as a target and an initiator of post-transcriptional gene silencing in transgenic plants. *Plant molecular biology,* 32**,** 79-88.

BELL-SAKYI, L., KOHL, A., BENTE, D. A. & FAZAKERLEY, J. K. 2012. Tick cell lines for study of Crimean-Congo hemorrhagic fever virus and other arboviruses. *Vector Borne Zoonotic Dis,* 12**,** 769-81.

BERGERON, E., ALBARINO, C. G., KHRISTOVA, M. L. & NICHOL, S. T. 2010a. Crimean-Congo hemorrhagic fever virus-encoded ovarian tumor protease activity is dispensable for virus RNA polymerase function. *J Virol,* 84**,** 216-26.

BERGERON, E., ALBARINO, C. G., KHRISTOVA, M. L. & NICHOL, S. T. 2010b. Crimean-Congo hemorrhagic fever virus-encoded ovarian tumor protease activity is dispensable for virus RNA polymerase function. *Journal of virology,* 84**,** 216-26.

BERGERON, E., CHAKRABARTI, A. K., BIRD, B. H., DODD, K. A., MCMULLAN, L. K., SPIROPOULOU, C. F., NICHOL, S. T. & ALBARINO, C. G. 2012. Reverse genetics recovery of Lujo virus and role of virus RNA secondary structures in efficient virus growth. *Journal of virology,* 86**,** 10759-65.

BERGERON, E., VINCENT, M. J. & NICHOL, S. T. 2007a. Crimean-Congo hemorrhagic fever virus glycoprotein processing by the endoprotease SKI-1/S1P is critical for virus infectivity. *Journal of virology,* 81**,** 13271-6.

BERGERON, E., VINCENT, M. J. & NICHOL, S. T. 2007b. Crimean-Congo hemorrhagic fever virus glycoprotein processing by the endoprotease SKI-1/S1P is critical for virus infectivity. *J Virol,* 81**,** 13271-6.

BERNSTEIN, E., CAUDY, A. A., HAMMOND, S. M. & HANNON, G. J. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature,* 409**,** 363-6.

BERTOLOTTI-CIARLET, A., SMITH, J., STRECKER, K., PARAGAS, J., ALTAMURA, L. A., MCFALLS, J. M., FRIAS-STAHELI, N., GARCIA-SASTRE, A., SCHMALJOHN, C. S. & DOMS, R. W. 2005. Cellular localization and antigenic characterization of crimean-congo hemorrhagic fever virus glycoproteins. *J Virol,* 79**,** 6152-61.

BEGUM, F. & WISSESMAN, C. L. 1970. Hazara (HAZ) strain. JC 280. *The American journal of tropical medicine and hygiene,* 19**,** Suppl:1095-6.

BISWAS, S. K. & NAYAK, D. P. 1994. Mutational analysis of the conserved motifs of influenza A virus polymerase basic protein 1. *Journal of virology,* 68**,** 1819-26.

BLAKQORI, G. & WEBER, F. 2005. Efficient cDNA-based rescue of La Crosse bunyaviruses expressing or lacking the nonstructural protein NSs. *Journal of virology,* 79**,** 10420-8.

BLOMSTROM, D. C., FAHEY, D., KUTNY, R., KORANT, B. D. & KNIGHT, E., JR. 1986. Molecular characterization of the interferon-induced 15-kDa protein. Molecular cloning and nucleotide and amino acid sequence. *J Biol Chem,* 261**,** 8811-6.

BOND, C. S. & SCHUTTELKOPF, A. W. 2009. ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta crystallographica. Section D, Biological crystallography,* 65**,** 510-2.

BOOTH, T. F., GOULD, E. A. & NUTTALL, P. A. 1991. Structure and morphogenesis of Dugbe virus (Bunyaviridae, Nairovirus) studied by immunogold electron microscopy of ultrathin cryosections. *Virus Res,* 21**,** 199-212.

BOUDKO, S. P., KUHN, R. J. & ROSSMANN, M. G. 2007. The coiled-coil domain structure of the Sin Nombre virus nucleocapsid protein. *J Mol Biol,* 366**,** 1538-44.

BOWDEN, T. A., BITTO, D., MCLEES, A., YEROMONAHOS, C., ELLIOTT, R. M. & HUISKONEN, J. T. 2013. Orthobunyavirus ultrastructure and the curious tripodal glycoprotein spike. *PLoS pathogens,* 9**,** e1003374.

BOWIE, A. G. & UNTERHOLZNER, L. 2008. Viral evasion and subversion of pattern-recognition receptor signalling. *Nat Rev Immunol,* 8**,** 911-22.

BRAGG, W. L. 1913. The diffraction for short electromagnetic waves by a crystal. *Proceedings of the Cambridge Philosophical Society*, 17**,** 43-57.

BRIDGEN, A. & ELLIOTT, R. M. 1996. Rescue of a segmented negative-strand RNA virus entirely from cloned complementary DNAs. *Proceedings of the National Academy of Sciences of the United States of America,* 93**,** 15400-4.

BRUNGER, A. T. 1992. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature,* 355**,** 472-5.

BRUNOTTE, L., KERBER, R., SHANG, W., HAUER, F., HASS, M., GABRIEL, M., LELKE, M., BUSCH, C., STARK, H., SVERGUN, D. I., BETZEL, C., PERBANDT, M. & GUNTHER, S. 2011. Structure of the Lassa virus nucleoprotein revealed by X-ray crystallography, small-angle X-ray scattering, and electron microscopy. *The Journal of biological chemistry,* 286**,** 38748-56.

BUCHER, E., SIJEN, T., DE HAAN, P., GOLDBACH, R. & PRINS, M. 2003a. Negative-strand tospoviruses and tenuiviruses carry a gene for a suppressor of gene silencing at analogous genomic positions. *Journal of virology,* 77**,** 1329-36.

BUCHER, E., SIJEN, T., DE HAAN, P., GOLDBACH, R. & PRINS, M. 2003b. Negative-strand tospoviruses and tenuiviruses carry a gene for a suppressor of gene silencing at analogous genomic positions. *J Virol,* 77**,** 1329-36.

BURT, F. J., PAWESKA, J. T., ASHKETTLE, B. & SWANEPOEL, R. 2009. Genetic relationship in southern African Crimean-Congo haemorrhagic fever virus isolates: evidence for occurrence of reassortment. *Epidemiol Infect,* 137**,** 1302-8.

BURT, F. J., SPENCER, D. C., LEMAN, P. A., PATTERSON, B. & SWANEPOEL, R. 1996a. Investigation of tick-borne viruses as pathogens of humans in South Africa and evidence of Dugbe virus infection in a patient with prolonged thrombocytopenia. *Epidemiol Infect,* 116**,** 353-61.

BURT, F. J., SPENCER, D. C., LEMAN, P. A., PATTERSON, B. & SWANEPOEL, R. 1996b. Investigation of tick-borne viruses as pathogens of humans in South Africa and evidence of Dugbe virus infection in a patient with prolonged thrombocytopenia. *Epidemiology and infection,* 116**,** 353-61.

BURT, F. J. & SWANEPOEL, R. 2005. Molecular epidemiology of African and Asian Crimean-Congo haemorrhagic fever isolates. *Epidemiol Infect,* 133**,** 659-66.

CAPODAGLI, G. C., MCKERCHER, M. A., BAKER, E. A., MASTERS, E. M., BRUNZELLE, J. S. & PEGAN, S. D. 2011. Structural analysis of a viral ovarian tumor domain protease from the Crimean-Congo hemorrhagic fever virus in complex with covalently bonded ubiquitin. *Journal of virology,* 85**,** 3621-30.

CARROLL, S. A., BIRD, B. H., ROLLIN, P. E. & NICHOL, S. T. 2010. Ancient common ancestry of Crimean-Congo hemorrhagic fever virus. *Mol Phylogenet Evol,* 55**,** 1103-10.

CARTER, C.W. 1990. Efficient factorial designs and the analysis of macromolecular crystal growth conditions. Methods: *Companion Methods Enzymology*, 1, 12-24.

CARTER, S. D., BARR, J. N. & EDWARDS, T. A. 2012a. Expression, purification and crystallization of the Crimean-Congo haemorrhagic fever virus nucleocapsid protein. *Acta crystallographica. Section F, Structural biology and crystallization communications,* 68**,** 569-73.

CARTER, S. D., SURTEES, R., WALTER, C. T., ARIZA, A., BERGERON, E., NICHOL, S. T., HISCOX, J. A., EDWARDS, T. A. & BARR, J. N. 2012b. Structure, function, and evolution of the Crimean-Congo hemorrhagic fever virus nucleocapsid protein. *Journal of virology,* 86**,** 10914-23.

CASALS, J. & TIGNOR, G. H. 1974. Neutralization and hemagglutination-inhibition tests with Crimean hemorrhagic fever-Congo virus. *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine,* 145**,** 960-6.

CHAMBERLAIN, J., ATKINSON, B., LOGUE, C. H., LATHAM, J., NEWMAN, E. N. & HEWSON, R. 2013. Genome Sequence of Ex-Afghanistan Crimean-Congo Hemorrhagic Fever Virus SCT Strain, from an Imported United Kingdom Case in October 2012. *Genome announcements,* 1.

CHEN, S. 2013. Molecular evolution of Crimean-Congo hemorrhagic fever virus based on complete genomes. *The Journal of general virology,* 94**,** 843-50.

CHEN, V. B., ARENDALL, W. B., 3RD, HEADD, J. J., KEEDY, D. A., IMMORMINO, R. M., KAPRAL, G. J., MURRAY, L. W., RICHARDSON, J. S. & RICHARDSON, D. C. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography,* 66**,** 12-21.

CHINIKAR, S., PERSSON, S. M., JOHANSSON, M., BLADH, L., GOYA, M., HOUSHMAND, B., MIRAZIMI, A., PLYUSNIN, A., LUNDKVIST, A. & NILSSON, M. 2004. Genetic analysis of Crimean-congo hemorrhagic fever virus in Iran. *J Med Virol,* 73, 404-11.

COATES, D. M. & SWEET, C. 1990. Studies on the pathogenicity of a nairovirus, Dugbe virus, in normal and immunosuppressed mice. *J Gen Virol,* 71 (Pt 2), 325-32.

COLOMA, R., VALPUESTA, J. M., ARRANZ, R., CARRASCOSA, J. L., ORTIN, J. & MARTIN-BENITO, J. 2009a. The structure of a biologically active influenza virus ribonucleoprotein complex. *PLoS Pathog,* 5, e1000491.

COLOMA, R., VALPUESTA, J. M., ARRANZ, R., CARRASCOSA, J. L., ORTIN, J. & MARTIN-BENITO, J. 2009b. The structure of a biologically active influenza virus ribonucleoprotein complex. *PLoS pathogens,* 5, e1000491.

COWTAN, K. 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr,* 62, 1002-11.

COWTAN, K. 2010. Recent developments in classical density modification. *Acta crystallographica. Section D, Biological crystallography,* 66, 470-8.

CROW, Y. J. & REHWINKEL, J. 2009. Aicardi-Goutieres syndrome and related phenotypes: linking nucleic acid metabolism with autoimmunity. *Human molecular genetics,* 18, R130-6.

DAVIS, I. W., MURRAY, L. W., RICHARDSON, J. S. & RICHARDSON, D. C. 2004. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic acids research,* 32, W615-9.

DAYHOFF, M. O., R. M. SCHWARTZ, AND B. C. ORCUTT. *1978*. A model of evolutionary change in proteins. Pp. 345–352 in M. O. Dayhoff, ed., Atlas of Protein Sequence and Structure Vol. 5, suppl. 3. National Biomedical Research Foundation, Washington, D.C.

DE HAAN, P., KORMELINK, R., DE OLIVEIRA RESENDE, R., VAN POELWIJK, F., PETERS, D. & GOLDBACH, R. 1991. Tomato spotted wilt virus L RNA encodes a putative RNA polymerase. *The Journal of general virology,* 72 (Pt 9), 2207-16.

DE SILVA, U., CHOUDHURY, S., BAILEY, S. L., HARVEY, S., PERRINO, F. W. & HOLLIS, T. 2007. The crystal structure of TREX1 explains the 3' nucleotide specificity and reveals a polyproline II helix for protein partnering. *The Journal of biological chemistry,* 282, 10537-43.

DENAULT, J. B. & SALVESEN, G. S. 2003. Expression, purification, and characterization of caspases. *Current protocols in protein science / editorial board, John E. Coligan ... [et al.],* Chapter 21, Unit 21 13.

DESFOSSES, A., GORET, G., FARIAS ESTROZI, L., RUIGROK, R. W. & GUTSCHE, I. 2011. Nucleoprotein-RNA Orientation in the Measles Virus Nucleocapsid by Three-Dimensional Electron Microscopy. *J Virol,* 85**,** 1391-5.

DESSAU, M. & MODIS, Y. 2013. Crystal structure of glycoprotein C from Rift Valley fever virus. *Proceedings of the National Academy of Sciences of the United States of America,* 110**,** 1696-701.

DEYDE, V. M., KHRISTOVA, M. L., ROLLIN, P. E., KSIAZEK, T. G. & NICHOL, S. T. 2006. Crimean-Congo hemorrhagic fever virus genomics and global diversity. *J Virol,* 80**,** 8834-42.

DIAZ-PENDON, J. A., LI, F., LI, W. X. & DING, S. W. 2007. Suppression of antiviral silencing by cucumber mosaic virus 2b protein in Arabidopsis is associated with drastically reduced accumulation of three classes of viral small interfering RNAs. *The Plant cell,* 19**,** 2053-63.

DONG, H., LI, P., BOTTCHER, B., ELLIOTT, R. M. & DONG, C. 2013a. Crystal structure of Schmallenberg orthobunyavirus nucleoprotein-RNA complex reveals a novel RNA sequestration mechanism. *RNA,* 19**,** 1129-36.

DONG, H., LI, P., ELLIOTT, R. M. & DONG, C. 2013b. Structure of schmallenberg orthobunyavirus nucleoprotein suggests a novel mechanism of genome encapsidation. *Journal of virology,* 87**,** 5593-601.

DOWALL, S. D., FINDLAY-WILSON, S., RAYNER, E., PEARSON, G., PICKERSGILL, J., RULE, A., MERREDEW, N., SMITH, H., CHAMBERLAIN, J. & HEWSON, R. 2012. Hazara virus infection is lethal for adult type I interferon receptor-knockout mice and may act as a surrogate for infection with the human-pathogenic Crimean-Congo hemorrhagic fever virus. *The Journal of general virology,* 93**,** 560-4.

DUIJSINGS, D., KORMELINK, R. & GOLDBACH, R. 2001. In vivo analysis of the TSWV cap-snatching mechanism: single base complementarity and primer length requirements. *The EMBO journal,* 20**,** 2545-52.

DUNN, E. F., PRITLOVE, D. C., JIN, H. & ELLIOTT, R. M. 1995. Transcription of a recombinant bunyavirus RNA template by transiently expressed bunyavirus proteins. *Virology,* 211**,** 133-43.

EICHLER, R., STRECKER, T., KOLESNIKOVA, L., TER MEULEN, J., WEISSENHORN, W., BECKER, S., KLENK, H. D., GARTEN, W. & LENZ, O. 2004. Characterization of the Lassa virus matrix protein Z: electron microscopic study of virus-like particles and interaction with the nucleoprotein (NP). *Virus research,* 100**,** 249-55.

EIFAN, S. A. & ELLIOTT, R. M. 2009. Mutational analysis of the Bunyamwera orthobunyavirus nucleocapsid protein gene. *J Virol,* 83**,** 11307-17.

ELLIOTT, R. M. 1990. Molecular biology of the Bunyaviridae. *J Gen Virol,* 71 (Pt 3)**,** 501-22.

ELLIOTT, R. M. 2009. Bunyaviruses and climate change. *Clin Microbiol Infect,* 15**,** 510-7.

EMSLEY, P. & COWTAN, K. 2004. Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography,* 60**,** 2126-32.

ERGONUL, O. 2006. Crimean-Congo haemorrhagic fever. *Lancet Infect Dis,* 6**,** 203-14.

EVANS, P. 2006a. Scaling and assessment of data quality. *Acta crystallographica. Section D, Biological crystallography,* 62**,** 72-82.

EVANS, P. 2006b. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr,* 62**,** 72-82.

EVANS, P. R. 2011. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta crystallographica. Section D, Biological crystallography,* 67**,** 282-92.

EVANS, P. R. & MURSHUDOV, G. N. 2013. How good are my data and what is the resolution? *Acta crystallographica. Section D, Biological crystallography,* 69**,** 1204-14.

FELSENSTEIN, J. 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics,* 5**,** 164-166.

FERRON, F., LI, Z., DANEK, E. I., LUO, D., WONG, Y., COUTARD, B., LANTEZ, V., CHARREL, R., CANARD, B., WALZ, T. & LESCAR, J. 2011. The hexamer structure of Rift Valley fever virus nucleoprotein suggests a mechanism for its assembly into ribonucleoprotein complexes. *PLoS pathogens,* 7**,** e1002030.

FIRE, A. Z. 2007. Gene silencing by double-stranded RNA. *Cell death and differentiation,* 14**,** 1998-2012.

FLICK, K., HOOPER, J. W., SCHMALJOHN, C. S., PETTERSSON, R. F., FELDMANN, H. & FLICK, R. 2003a. Rescue of Hantaan virus minigenomes. *Virology,* 306**,** 219-24.

FLICK, K., KATZ, A., OVERBY, A., FELDMANN, H., PETTERSSON, R. F. & FLICK, R. 2004. Functional analysis of the noncoding regions of the Uukuniemi virus (Bunyaviridae) RNA segments. *Journal of virology,* 78**,** 11726-38.

FLICK, R., ELGH, F. & PETTERSSON, R. F. 2002. Mutational analysis of the Uukuniemi virus (Bunyaviridae family) promoter reveals two elements of functional importance. *J Virol,* 76**,** 10849-60.

FLICK, R., FLICK, K., FELDMANN, H. & ELGH, F. 2003b. Reverse genetics for crimean-congo hemorrhagic fever virus. *J Virol,* 77**,** 5997-6006.

FONTANA, J., LOPEZ-MONTERO, N., ELLIOTT, R. M., FERNANDEZ, J. J. & RISCO, C. 2008. The unique architecture of Bunyamwera virus factories around the Golgi complex. *Cellular microbiology,* 10**,** 2012-28.

FOULKE, R. S., ROSATO, R. R. & FRENCH, G. R. 1981. Structural polypeptides of Hazara virus. *The Journal of general virology,* 53**,** 169-72.

GARCIN, D., LEZZI, M., DOBBS, M., ELLIOTT, R. M., SCHMALJOHN, C., KANG, C. Y. & KOLAKOFSKY, D. 1995. The 5' ends of Hantaan virus (Bunyaviridae) RNAs suggest a prime-and-realign mechanism for the initiation of RNA synthesis. *Journal of virology,* 69**,** 5754-62.

GAULIARD, N., BILLECOCQ, A., FLICK, R. & BOULOY, M. 2006. Rift Valley fever virus noncoding regions of L, M and S segments regulate RNA synthesis. *Virology,* 351**,** 170-9.

GE, P., TSAO, J., SCHEIN, S., GREEN, T. J., LUO, M. & ZHOU, Z. H. 2010. Cryo-EM model of the bullet-shaped vesicular stomatitis virus. *Science,* 327**,** 689-93.

GEERTS-DIMITRIADOU, C., ZWART, M. P., GOLDBACH, R. & KORMELINK, R. 2011. Base-pairing promotes leader selection to prime in vitro influenza genome transcription. *Virology,* 409**,** 17-26.

GERRARD, S. R., BIRD, B. H., ALBARINO, C. G. & NICHOL, S. T. 2007. The NSm proteins of Rift Valley fever virus are dispensable for maturation, replication and infection. *Virology,* 359**,** 459-65.

GOLDBACH, R.W., AND PETERS, D. 1994. Possible causes of the emergence of tospovirus diseases. *Semin. Virol*, 5, 113–120.

GOLDSMITH, C. S., ELLIOTT, L. H., PETERS, C. J. & ZAKI, S. R. 1995. Ultrastructural characteristics of Sin Nombre virus, causative agent of hantavirus pulmonary syndrome. *Arch Virol,* 140**,** 2107-22.

GREEN, T. J., COX, R., TSAO, J., ROWSE, M., QIU, S. & LUO, M. 2014a. Common mechanism for RNA encapsidation by negative strand RNA viruses. *J Virol*.

GREEN, T. J., COX, R., TSAO, J., ROWSE, M., QIU, S. & LUO, M. 2014b. Common mechanism for RNA encapsidation by negative-strand RNA viruses. *J Virol,* 88**,** 3766-75.

GREEN, T. J., ZHANG, X., WERTZ, G. W. & LUO, M. 2006. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science,* 313**,** 357-60.

GROSETH, A., WOLFF, S., STRECKER, T., HOENEN, T. & BECKER, S. 2010. Efficient budding of the tacaribe virus matrix protein z requires the nucleoprotein. *Journal of virology,* 84**,** 3603-11.

GUNES, T., ENGIN, A., POYRAZ, O., ELALDI, N., KAYA, S., DOKMETAS, I., BAKIR, M. & CINAR, Z. 2009. Crimean-Congo hemorrhagic fever virus in high-risk population, Turkey. *Emerg Infect Dis,* 15**,** 461-4.

GUNES, T., POYRAZ, O. & VATANSEVER, Z. 2011. Crimean-Congo hemorrhagic fever virus in ticks collected from humans, livestock, and picnic sites in the hyperendemic region of Turkey. *Vector Borne Zoonotic Dis,* 11**,** 1411-6.

GUO, H. S. & DING, S. W. 2002. A viral protein inhibits the long range signaling activity of the gene silencing signal. *The EMBO journal,* 21**,** 398-407.

GUO, Y., WANG, W., JI, W., DENG, M., SUN, Y., ZHOU, H., YANG, C., DENG, F., WANG, H., HU, Z., LOU, Z. & RAO, Z. 2012. Crimean-Congo hemorrhagic fever virus nucleoprotein reveals endonuclease activity in bunyaviruses. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** 5046-51.

GURBUZ, Y., SENCAN, I., OZTURK, B. & TUTUNCU, E. 2009. A case of nosocomial transmission of Crimean-Congo hemorrhagic fever from patient to patient. *Int J Infect Dis,* 13**,** e105-7.

HAAS, A. L., AHRENS, P., BRIGHT, P. M. & ANKEL, H. 1987. Interferon induces a 15-kilodalton protein exhibiting marked homology to ubiquitin. *J Biol Chem,* 262**,** 11315-23.

HAFERKAMP, S., FERNANDO, L., SCHWARZ, T. F., FELDMANN, H. & FLICK, R. 2005. Intracellular localization of Crimean-Congo Hemorrhagic Fever (CCHF) virus glycoproteins. *Virol J,* 2**,** 42.

HALLER, O. & KOCHS, G. 2002. Interferon-induced mx proteins: dynamin-like GTPases with antiviral activity. *Traffic,* 3**,** 710-7.

HENIKOFF, S. & HENIKOFF, J. G. 1992. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A,* 89**,** 10915-9.

HARTY, R. N., PITHA, P. M. & OKUMURA, A. 2009. Antiviral activity of innate immune protein ISG15. *J Innate Immun,* 1**,** 397-404.

HASSANI-MEHRABAN, A., SAAIJER, J., PETERS, D., GOLDBACH, R. & KORMELINK, R. 2007. Molecular and biological comparison of two Tomato yellow ring virus (TYRV) isolates: challenging the Tospovirus species concept. *Archives of virology,* 152**,** 85-96.

HASTIE, K. M., KIMBERLIN, C. R., ZANDONATTI, M. A., MACRAE, I. J. & SAPHIRE, E. O. 2011a. Structure of the Lassa virus nucleoprotein reveals a dsRNA-specific 3' to 5' exonuclease activity essential for immune suppression. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 2396-401.

HASTIE, K. M., LIU, T., LI, S., KING, L. B., NGO, N., ZANDONATTI, M. A., WOODS, V. L., JR., DE LA TORRE, J. C. & SAPHIRE, E. O. 2011b. Crystal structure of the Lassa virus nucleoprotein-RNA complex reveals a gating mechanism for RNA binding. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 19365-70.

HEPOJOKI, J., STRANDIN, T., WANG, H., VAPALAHTI, O., VAHERI, A. & LANKINEN, H. 2010. Cytoplasmic tails of hantavirus glycoproteins interact with the nucleocapsid protein. *J Gen Virol,* 91**,** 2341-50.

HEWSON, R., CHAMBERLAIN, J., MIOULET, V., LLOYD, G., JAMIL, B., HASAN, R., GMYL, A., GMYL, L., SMIRNOVA, S. E., LUKASHEV, A., KARGANOVA, G. & CLEGG, C. 2004a. Crimean-Congo haemorrhagic fever virus: sequence analysis of the small RNA segments from a collection of viruses world wide. *Virus Res,* 102**,** 185-9.

HEWSON, R., GMYL, A., GMYL, L., SMIRNOVA, S. E., KARGANOVA, G., JAMIL, B., HASAN, R., CHAMBERLAIN, J. & CLEGG, C. 2004b. Evidence of segment reassortment in Crimean-Congo haemorrhagic fever virus. *J Gen Virol,* 85**,** 3059-70.

HOLLAND, J. & DOMINGO, E. 1998. Origin and evolution of viruses. *Virus Genes,* 16**,** 13-21.

HOLZER, B., BAKSHI, S., BRIDGEN, A. & BARON, M. D. 2011. Inhibition of interferon induction and action by the nairovirus Nairobi sheep disease virus/Ganjam virus. *PloS one,* 6**,** e28594.

HONIG, J. E., OSBORNE, J. C. & NICHOL, S. T. 2004. Crimean-Congo hemorrhagic fever virus genome L RNA segment and encoded protein. *Virology,* 321**,** 29-35.

HOOGSTRAAL, H. 1979. The epidemiology of tick-borne Crimean-Congo hemorrhagic fever in Asia, Europe, and Africa. *Journal of medical entomology,* 15**,** 307-417.

HUISKONEN, J. T., HEPOJOKI, J., LAURINMAKI, P., VAHERI, A., LANKINEN, H., BUTCHER, S. J. & GRUNEWALD, K. 2010. Electron cryotomography of Tula hantavirus suggests a unique assembly paradigm for enveloped viruses. *J Virol,* 84**,** 4889-97.

HUISKONEN, J. T., OVERBY, A. K., WEBER, F. & GRUNEWALD, K. 2009. Electron cryo-microscopy and single-particle averaging of Rift Valley fever virus: evidence for GN-GC glycoprotein heterodimers. *Journal of virology,* 83**,** 3762-9.

IKEGAMI, T., WON, S., PETERS, C. J. & MAKINO, S. 2006. Rescue of infectious rift valley fever virus entirely from cDNA, analysis of virus lacking the NSs gene, and expression of a foreign gene. *Journal of virology,* 80**,** 2933-40.

IZADI, S., HOLAKOUIE-NAIENI, K., MAJDZADEH, S. R., CHINIKAR, S., NADIM, A., RAKHSHANI, F. & HOOSHMAND, B. 2006. Seroprevalence of Crimean-Congo hemorrhagic fever in Sistan-va-Baluchestan province of Iran. *Jpn J Infect Dis,* 59**,** 326-8.

JAMES, T. W., FRIAS-STAHELI, N., BACIK, J. P., LEVINGSTON MACLEOD, J. M., KHAJEHPOUR, M., GARCIA-SASTRE, A. & MARK, B. L. 2011. Structural basis for the removal of ubiquitin and interferon-stimulated gene 15 by a viral ovarian tumor domain-containing protease. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 2222-7.

KABSCH, W. 2010. Xds. *Acta crystallographica. Section D, Biological crystallography,* 66**,** 125-32.

KAINZ, M., HILSON, P., SWEENEY, L., DEROSE, E. & GERMAN, T. L. 2004. Interaction Between Tomato spotted wilt virus N Protein Monomers Involves Nonelectrostatic Forces Governed by Multiple Distinct Regions in the Primary Structure. *Phytopathology,* 94**,** 759-65.

KARLBERG, H., TAN, Y. J. & MIRAZIMI, A. 2011. Induction of caspase activation and cleavage of the viral nucleocapsid protein in different cell types during Crimean-Congo hemorrhagic fever virus infection. *The Journal of biological chemistry,* 286**,** 3227-34.

KARPLUS, P. A. & DIEDERICHS, K. 2012. Linking crystallographic model and data quality. *Science,* 336**,** 1030-3.

KAUKINEN, P., KOISTINEN, V., VAPALAHTI, O., VAHERI, A. & PLYUSNIN, A. 2001. Interaction between molecules of hantavirus nucleocapsid protein. *J Gen Virol,* 82**,** 1845-53.

KAUKINEN, P., KUMAR, V., TULIMAKI, K., ENGELHARDT, P., VAHERI, A. & PLYUSNIN, A. 2004. Oligomerization of Hantavirus N protein: C-terminal alpha-helices interact to form a shared hydrophobic space. *J Virol,* 78**,** 13669-77.

KAUKINEN, P., VAHERI, A. & PLYUSNIN, A. 2003. Mapping of the regions involved in homotypic interactions of Tula hantavirus N protein. *J Virol,* 77**,** 10910-6.

KHVOROVA, A., REYNOLDS, A. & JAYASENA, S. D. 2003. Functional siRNAs and miRNAs exhibit strand bias. *Cell,* 115**,** 209-16.

KIKKERT, M., VAN LENT, J., STORMS, M., BODEGOM, P., KORMELINK, R. & GOLDBACH, R. 1999. Tomato spotted wilt virus particle morphogenesis in plant cells. *Journal of virology,* 73**,** 2288-97.

KINSELLA, E., MARTIN, S. G., GROLLA, A., CZUB, M., FELDMANN, H. & FLICK, R. 2004. Sequence determination of the Crimean-Congo hemorrhagic fever virus L segment. *Virology,* 321**,** 23-8.

KOCHS, G., JANZEN, C., HOHENBERG, H. & HALLER, O. 2002. Antivirally active MxA protein sequesters La Crosse virus nucleocapsid protein into perinuclear complexes. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 3153-8.

KOHL, A., DUNN, E. F., LOWEN, A. C. & ELLIOTT, R. M. 2004. Complementarity, sequence and structural elements within the 3' and 5' non-coding regions of the Bunyamwera orthobunyavirus S segment determine promoter strength. *The Journal of general virology,* 85**,** 3269-78.

KOHL, A., LOWEN, A. C., LEONARD, V. H. & ELLIOTT, R. M. 2006. Genetic elements regulating packaging of the Bunyamwera orthobunyavirus genome. *The Journal of general virology,* 87**,** 177-87.

KOMODA, K., NARITA, M., TANAKA, I. & YAO, M. 2013. Expression, purification, crystallization and preliminary X-ray crystallographic study of the nucleocapsid protein of Tomato spotted wilt virus. *Acta crystallographica. Section F, Structural biology and crystallization communications,* 69**,** 700-3.

KORMALINK. R. 2011. *The Molecular Biology of Tospoviruses and Resistance. Strategies. In Bunyaviridae: Molecular and Cellular Biology, 163-191,,*Edited by Alexander Plyusnin, Richard M. Elliott.

KORMELINK, R., GARCIA, M. L., GOODIN, M., SASAYA, T. & HAENNI, A. L. 2011. Negative-strand RNA viruses: the plant-infecting counterparts. *Virus research,* 162**,** 184-202.

KORMELINK, R., STORMS, M., VAN LENT, J., PETERS, D. & GOLDBACH, R. 1994. Expression and subcellular location of the NSM protein of tomato spotted wilt virus (TSWV), a putative viral movement protein. *Virology,* 200**,** 56-65.

KRISSINEL, E. & HENRICK, K. 2007. Inference of macromolecular assemblies from crystalline state. *J Mol Biol,* 372**,** 774-97.

KUISMANEN, E., BANG, B., HURME, M. & PETTERSSON, R. F. 1984. Uukuniemi virus maturation: immunofluorescence microscopy with monoclonal glycoprotein-specific antibodies. *Journal of virology,* 51**,** 137-46.

L'VOV D, N., DZHARKENOV, A. F., ARISTOVA, V. A., KOVTUNOV, A. I., GROMASHEVSKII, V. L., VYSHEMIRSKII, O. I., GALKINA, I. V., LARICHEV, V. F., BUTENKO, A. M. & L'VOV D, K. 2002. [The isolation of Dhori viruses (Orthomyxoviridae, Thogotovirus) and Crimean-Congo hemorrhagic fever virus (Bunyaviridae, Nairovirus) from the hare (Lepus europaeus) and its ticks Hyalomma marginatum in the middle zone of the Volga delta, Astrakhan region, 2001]. *Vopr Virusol,* 47**,** 32-6.

LAMZIN, V. S. & WILSON, K. S. 1997. Automated refinement for protein crystallography. *Methods in enzymology,* 277**,** 269-305.

LARKIN, M. A., BLACKSHIELDS, G., BROWN, N. P., CHENNA, R., MCGETTIGAN, P. A., MCWILLIAM, H., VALENTIN, F., WALLACE, I. M., WILM, A., LOPEZ, R., THOMPSON, J. D., GIBSON, T. J. & HIGGINS, D. G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics,* 23**,** 2947-8.

LEBLEBICIOGLU, H. 2010. Crimean-Congo haemorrhagic fever in Eurasia. *Int J Antimicrob Agents,* 36 Suppl 1**,** S43-6.

LEE, Y., AHN, C., HAN, J., CHOI, H., KIM, J., YIM, J., LEE, J., PROVOST, P., RADMARK, O., KIM, S. & KIM, V. N. 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature,* 425**,** 415-9.

LENZ, O., TER MEULEN, J., KLENK, H. D., SEIDAH, N. G. & GARTEN, W. 2001. The Lassa virus glycoprotein precursor GP-C is proteolytically processed by subtilase SKI-1/S1P. *Proceedings of the National Academy of Sciences of the United States of America,* 98**,** 12701-5.

LESLIE, A. G. 2006. The integration of macromolecular diffraction data. *Acta crystallographica. Section D, Biological crystallography,* 62**,** 48-57.

LI, B., WANG, Q., PAN, X., FERNANDEZ DE CASTRO, I., SUN, Y., GUO, Y., TAO, X., RISCO, C., SUI, S. F. & LOU, Z. 2013. Bunyamwera virus possesses a distinct nucleocapsid protein to facilitate genome encapsidation. *Proceedings of the National Academy of Sciences of the United States of America*.

LOKESH, B., RASHMI, P. R., AMRUTA, B. S., SRISATHIYANARAYANAN, D., MURTHY, M. R. & SAVITHRI, H. S. 2010. NSs encoded by groundnut bud necrosis virus is a bifunctional enzyme. *PLoS One,* 5**,** e9757.

LOPEZ, N., MULLER, R., PREHAUD, C. & BOULOY, M. 1995. The L protein of Rift Valley fever virus can rescue viral ribonucleoproteins and transcribe synthetic genome-like RNA molecules. *J Virol,* 69**,** 3972-9.

LUKASHEV, A. N., IVANOVA, O. E., EREMEEVA, T. P. & IGGO, R. D. 2008. Evidence of frequent recombination among human adenoviruses. *J Gen Virol,* 89**,** 380-8.

MALTEZOU, H. C., ANDONOVA, L., ANDRAGHETTI, R., BOULOY, M., ERGONUL, O., JONGEJAN, F., KALVATCHEV, N., NICHOL, S., NIEDRIG, M., PLATONOV, A., THOMSON, G., LEITMEYER, K. & ZELLER, H. 2010. Crimean-Congo hemorrhagic fever in Europe: current situation calls for preparedness. *Euro Surveill,* 15**,** 19504.

MARRIOTT, A. C. & NUTTALL, P. A. 1996. Large RNA segment of Dugbe nairovirus encodes the putative RNA polymerase. *The Journal of general virology,* 77 (Pt 8)**,** 1775-80.

MARTINEZ-SOBRIDO, L., EMONET, S., GIANNAKAS, P., CUBITT, B., GARCIA-SASTRE, A. & DE LA TORRE, J. C. 2009. Identification of amino acid residues critical for the anti-interferon activity of the nucleoprotein of the prototypic arenavirus lymphocytic choriomeningitis virus. *Journal of virology,* 83**,** 11330-40.

MARTINEZ-SOBRIDO, L., GIANNAKAS, P., CUBITT, B., GARCIA-SASTRE, A. & DE LA TORRE, J. C. 2007. Differential inhibition of type I interferon induction by arenavirus nucleoproteins. *Journal of virology,* 81**,** 12696-703.

MATSUOKA, Y., CHEN, S. Y. & COMPANS, R. W. 1991. Bunyavirus protein transport and assembly. *Current topics in microbiology and immunology,* 169**,** 161-79.

MATTHEWS, B. W. 1968. Solvent content of protein crystals. *Journal of molecular biology,* 33**,** 491-7.

MCCOY, A. J. 2007. Solving structures of protein complexes by molecular replacement with Phaser. *Acta crystallographica. Section D, Biological crystallography,* 63**,** 32-41.

MCCOY, A. J., GROSSE-KUNSTLEVE, R. W., ADAMS, P. D., WINN, M. D., STORONI, L. C. & READ, R. J. 2007. Phaser crystallographic software. *Journal of applied crystallography,* 40**,** 658-674.

MERTENS, M., SCHMIDT, K., OZKUL, A. & GROSCHUP, M. H. 2013. The impact of Crimean-Congo hemorrhagic fever virus on public health. *Antiviral Res,* 98**,** 248-60.

MILD, M., SIMON, M., ALBERT, J. & MIRAZIMI, A. 2010a. Towards an understanding of the migration of Crimean-Congo hemorrhagic fever virus. *The Journal of general virology,* 91**,** 199-207.

MILD, M., SIMON, M., ALBERT, J. & MIRAZIMI, A. 2010b. Towards an understanding of the migration of Crimean-Congo hemorrhagic fever virus. *J Gen Virol,* 91**,** 199-207.

MILLER, B. R., DEFOLIART, G. R. & YUILL, T. M. 1977. Vertical transmission of La Crosse virus (California encephalitis group): transovarial and filial infection rates in Aedes triseriatus (Diptera: Culicidae). *Journal of medical entomology,* 14**,** 437-40.

MIR, M. A., BROWN, B., HJELLE, B., DURAN, W. A. & PANGANIBAN, A. T. 2006. Hantavirus N protein exhibits genus-specific recognition of the viral RNA panhandle. *Journal of virology,* 80**,** 11283-92.

MIR, M. A. & PANGANIBAN, A. T. 2004. Trimeric hantavirus nucleocapsid protein binds specifically to the viral RNA panhandle. *J Virol,* 78**,** 8281-8.

MIR, M. A. & PANGANIBAN, A. T. 2005. The hantavirus nucleocapsid protein recognizes specific features of the viral RNA panhandle and is altered in conformation upon RNA binding. *J Virol,* 79**,** 1824-35.

MIR, M. A. & PANGANIBAN, A. T. 2006. The bunyavirus nucleocapsid protein is an RNA chaperone: possible roles in viral RNA panhandle formation and genome replication. *RNA,* 12**,** 272-82.

MOELLER, A., KIRCHDOERFER, R. N., POTTER, C. S., CARRAGHER, B. & WILSON, I. A. 2012. Organization of the influenza virus replication machinery. *Science,* 338**,** 1631-4.

MOHL, B. P. & BARR, J. N. 2009. Investigating the specificity and stoichiometry of RNA binding by the nucleocapsid protein of Bunyamwera virus. *RNA,* 15**,** 391-9.

MOIS I. AROYO, ULRICH MULLER & WONDRATSCHEK, H. *International tables for crystallography.* Vol. A1 (IUCr, 2006).

MULLER, R., POCH, O., DELARUE, M., BISHOP, D. H. & BOULOY, M. 1994. Rift Valley fever virus L segment: correction of the sequence and possible functional role of newly identified regions conserved in RNA-dependent polymerases. *The Journal of general virology,* 75 (Pt 6)**,** 1345-52.

MUMFORD, R. A., BARKER, I. & WOOD, K. R. 1996. An improved method for the detection of Tospoviruses using the polymerase chain reaction. *Journal of virological methods,* 57**,** 109-15.

MURPHY, F. A., HARRISON, A. K. & WHITFIELD, S. G. 1973. Bunyaviridae: morphologic and morphogenetic similarities of Bunyamwera serologic supergroup viruses and several other arthropod-borne viruses. *Intervirology,* 1**,** 297-316.

MURSHUDOV, G. N., VAGIN, A. A. & DODSON, E. J. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta crystallographica. Section D, Biological crystallography,* 53**,** 240-55.

NGUYEN, M. & HAENNI, A. L. 2003. Expression strategies of ambisense viruses. *Virus Res,* 93**,** 141-50.

NIU, F., SHAW, N., WANG, Y. E., JIAO, L., DING, W., LI, X., ZHU, P., UPUR, H., OUYANG, S., CHENG, G. & LIU, Z. J. 2013. Structure of the Leanyer orthobunyavirus nucleoprotein-RNA complex reveals unique architecture for RNA encapsidation. *Proceedings of the National Academy of Sciences of the United States of America*.

NOVOA, R. R., CALDERITA, G., CABEZAS, P., ELLIOTT, R. M. & RISCO, C. 2005. Key Golgi factors for structural and functional maturation of bunyamwera virus. *Journal of virology,* 79**,** 10852-63.

OBIJESKI, J. F., BISHOP, D. H., PALMER, E. L. & MURPHY, F. A. 1976. Segmented genome and nucleocapsid of La Crosse virus. *J Virol,* 20**,** 664-75.

OGAWA, Y., SUGIURA, K., KATO, K., TOHYA, Y. & AKASHI, H. 2007. Rescue of Akabane virus (family Bunyaviridae) entirely from cloned cDNAs by using RNA polymerase I. *The Journal of general virology,* 88**,** 3385-90.

OVERBY, A. K., PETTERSSON, R. F., GRUNEWALD, K. & HUISKONEN, J. T. 2008. Insights into bunyavirus architecture from electron cryotomography of Uukuniemi virus. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 2375-9.

OVERBY, A. K., PETTERSSON, R. F. & NEVE, E. P. 2007a. The glycoprotein cytoplasmic tail of Uukuniemi virus (Bunyaviridae) interacts with ribonucleoproteins and is critical for genome packaging. *J Virol,* 81**,** 3198-205.

OVERBY, A. K., POPOV, V., NEVE, E. P. & PETTERSSON, R. F. 2006. Generation and analysis of infectious virus-like particles of uukuniemi virus (bunyaviridae): a useful system for studying bunyaviral packaging and budding. *J Virol,* 80**,** 10428-35.

OVERBY, A. K., POPOV, V. L., PETTERSSON, R. F. & NEVE, E. P. 2007b. The cytoplasmic tails of Uukuniemi Virus (Bunyaviridae) G(N) and G(C) glycoproteins are important for intracellular targeting and the budding of virus-like particles. *J Virol,* 81**,** 11381-91.

OZKURT, Z., KIKI, I., EROL, S., ERDEM, F., YILMAZ, N., PARLAK, M., GUNDOGDU, M. & TASYARAN, M. A. 2006. Crimean-Congo hemorrhagic fever in Eastern Turkey: clinical features, risk factors and efficacy of ribavirin therapy. *J Infect,* 52**,** 207-15.

PAAPE, M., SOLOVYEV, A. G., EROKHINA, T. N., MININA, E. A., SCHEPETILNIKOV, M. V., LESEMANN, D. E., SCHIEMANN, J., MOROZOV, S. Y. & KELLMANN, J. W. 2006. At-4/1, an interactor of the Tomato spotted wilt virus movement protein, belongs to a new family of plant proteins capable of directed intra- and intercellular trafficking. *Molecular plant-microbe interactions : MPMI,* 19**,** 874-83.

PANGANIBAN, A. T. & MIR, M. A. 2009. Bunyavirus N: eIF4F surrogate and cap-guardian. *Cell cycle,* 8**,** 1332-7.

PERRIERE, G. & GOUY, M. 1996. WWW-query: an on-line retrieval system for biological sequence banks. *Biochimie,* 78**,** 364-9.

PETTERSSON, R. F. & VON BONSDORFF, C. H. 1975. Ribonucleoproteins of Uukuniemi virus are circular. *J Virol,* 15**,** 386-92.

PINSCHEWER, D. D., PEREZ, M. & DE LA TORRE, J. C. 2003. Role of the virus nucleoprotein in the regulation of lymphocytic choriomeningitis virus transcription and RNA replication. *Journal of virology,* 77**,** 3882-7.

POCH, O., BLUMBERG, B. M., BOUGUELERET, L. & TORDO, N. 1990. Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains. *The Journal of general virology,* 71 (Pt 5)**,** 1153-62.

PRINS, M. & GOLDBACH, R. 1998. The emerging problem of tospovirus infection and nonconventional methods of control. *Trends in microbiology,* 6**,** 31-5.

QI, X., LAN, S., WANG, W., SCHELDE, L. M., DONG, H., WALLAT, G. D., LY, H., LIANG, Y. & DONG, C. 2010. Cap binding and immune evasion revealed by Lassa nucleoprotein structure. *Nature,* 468**,** 779-83.

RAJU, R. & KOLAKOFSKY, D. 1989. The ends of La Crosse virus genome and antigenome RNAs within nucleocapsids are base paired. *Journal of virology,* 63**,** 122-8.

RAVKOV, E. V., NICHOL, S. T. & COMPANS, R. W. 1997. Polarized entry and release in epithelial cells of Black Creek Canal virus, a New World hantavirus. *J Virol,* 71**,** 1147-54.

RAVKOV, E. V., NICHOL, S. T., PETERS, C. J. & COMPANS, R. W. 1998. Role of actin microfilaments in Black Creek Canal virus morphogenesis. *Journal of virology,* 72**,** 2865-70.

RAYMOND, D. D., PIPER, M. E., GERRARD, S. R., SKINIOTIS, G. & SMITH, J. L. 2012. Phleboviruses encapsidate their genomes by sequestering RNA bases. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** 19208-13.

RAYMOND, D. D., PIPER, M. E., GERRARD, S. R. & SMITH, J. L. 2010a. Structure of the Rift Valley fever virus nucleocapsid protein reveals another architecture for RNA encapsidation. *Proc Natl Acad Sci U S A*.

RAYMOND, D. D., PIPER, M. E., GERRARD, S. R. & SMITH, J. L. 2010b. Structure of the Rift Valley fever virus nucleocapsid protein reveals another architecture for RNA encapsidation. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 11769-74.

REGUERA, J., MALET, H., WEBER, F. & CUSACK, S. 2013. Structural basis for encapsidation of genomic RNA by La Crosse Orthobunyavirus nucleoprotein.

Proceedings of the National Academy of Sciences of the United States of America, 110, 7246-51.

REGUERA, J., WEBER, F. & CUSACK, S. 2010. Bunyaviridae RNA polymerases (L-protein) have an N-terminal, influenza-like endonuclease domain, essential for viral cap-dependent transcription. *PLoS pathogens,* 6, e1001101.

RESENDE RDE, O., DE HAAN, P., DE AVILA, A. C., KITAJIMA, E. W., KORMELINK, R., GOLDBACH, R. & PETERS, D. 1991. Generation of envelope and defective interfering RNA mutants of tomato spotted wilt virus by mechanical passage. *The Journal of general virology,* 72 (Pt 10), 2375-83.

RHODES, G. *Crystallography made crystal clear*, 3rd edn, (Academic Press, 2006).

RIBEIRO, D., BORST, J. W., GOLDBACH, R. & KORMELINK, R. 2009. Tomato spotted wilt virus nucleocapsid protein interacts with both viral glycoproteins Gn and Gc in planta. *Virology,* 383, 121-30.

RODRIGUEZ, L. L., MAUPIN, G. O., KSIAZEK, T. G., ROLLIN, P. E., KHAN, A. S., SCHWARZ, T. F., LOFTS, R. S., SMITH, J. F., NOOR, A. M., PETERS, C. J. & NICHOL, S. T. 1997. Molecular investigation of a multisource outbreak of Crimean-Congo hemorrhagic fever in the United Arab Emirates. *Am J Trop Med Hyg,* 57, 512-8.

RUPP, B. 2010. *Biomolecular crystallography,* (Garland Science, Taylor and Francis Group, LLC).

SADLER, A. J. & WILLIAMS, B. R. 2008. Interferon-inducible antiviral effectors. *Nat Rev Immunol,* 8, 559-68.

SALUZZO, J. F., DIGOUTTE, J. P., CAMICAS, J. L. & CHAUVANCY, G. 1985. Crimean-Congo haemorrhagic fever and Rift Valley fever in south-eastern Mauritania. *Lancet,* 1, 116.

SANCHEZ, A. J., VINCENT, M. J., ERICKSON, B. R. & NICHOL, S. T. 2006. Crimean-congo hemorrhagic fever virus glycoprotein precursor is cleaved by Furin-like and SKI-1 proteases to generate a novel 38-kilodalton glycoprotein. *J Virol,* 80, 514-25.

SCHMALJOHN C. S., NICHOL S. T. 2006. *Bunyaviridae. In Fields Virology, 5th edn, vol. 2, 1741–1789.* Edited by Knipe D. M., Howley P. M. Philadelphia, PA: Lippincott Williams & Wilkins

SCHNETTLER, E., HEMMES, H., HUISMANN, R., GOLDBACH, R., PRINS, M. & KORMELINK, R. 2010. Diverging affinity of tospovirus RNA silencing suppressor proteins, NSs, for various RNA duplex molecules. *Journal of virology,* 84**,** 11542-54.

SCHWARZ, D. S., HUTVAGNER, G., DU, T., XU, Z., ARONIN, N. & ZAMORE, P. D. 2003. Asymmetry in the assembly of the RNAi enzyme complex. *Cell,* 115**,** 199-208.

SEVER, S., DAMKE, H. & SCHMID, S. L. 2000. Dynamin:GTP controls the formation of constricted coated pits, the rate limiting step in clathrin-mediated endocytosis. *The Journal of cell biology,* 150**,** 1137-48.

SHEPHERD, A. J., SWANEPOEL, R., LEMAN, P. A. & SHEPHERD, S. P. 1987a. Field and laboratory investigation of Crimean-Congo haemorrhagic fever virus (Nairovirus, family Bunyaviridae) infection in birds. *Trans R Soc Trop Med Hyg,* 81**,** 1004-7.

SHEPHERD, A. J., SWANEPOEL, R., SHEPHERD, S. P., LEMAN, P. A. & MATHEE, O. 1991. Viraemic transmission of Crimean-Congo haemorrhagic fever virus to ticks. *Epidemiology and infection,* 106**,** 373-82.

SHEPHERD, A. J., SWANEPOEL, R., SHEPHERD, S. P., MCGILLIVRAY, G. M. & SEARLE, L. A. 1987b. Antibody to Crimean-Congo hemorrhagic fever virus in wild mammals from southern Africa. *Am J Trop Med Hyg,* 36**,** 133-42.

SHERMAN, M. B., FREIBERG, A. N., HOLBROOK, M. R. & WATOWICH, S. J. 2009. Single-particle cryo-electron microscopy of Rift Valley fever virus. *Virology,* 387**,** 11-5.

SHI, X., KOHL, A., LEONARD, V. H., LI, P., MCLEES, A. & ELLIOTT, R. M. 2006. Requirement of the N-terminal region of orthobunyavirus nonstructural protein NSm for virus assembly and morphogenesis. *Journal of virology,* 80**,** 8089-99.

SHTANKO, O., IMAI, M., GOTO, H., LUKASHEVICH, I. S., NEUMANN, G., WATANABE, T. & KAWAOKA, Y. 2010. A role for the C terminus of Mopeia virus nucleoprotein in its incorporation into Z protein-induced virus-like particles. *Journal of virology,* 84**,** 5415-22.

SNIPPE, M., WILLEM BORST, J., GOLDBACH, R. & KORMELINK, R. 2007. Tomato spotted wilt virus Gc and N proteins interact in vivo. *Virology,* 357**,** 115-23.

SODING, J., BIEGERT, A. & LUPAS, A. N. 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic acids research,* 33**,** W244-8.

SOELLICK, T., UHRIG, J. F., BUCHER, G. L., KELLMANN, J. W. & SCHREIER, P. H. 2000. The movement protein NSm of tomato spotted wilt tospovirus (TSWV): RNA binding, interaction with the TSWV N protein, and identification of interacting plant proteins. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 2373-8.

STETSON, D. B., KO, J. S., HEIDMANN, T. & MEDZHITOV, R. 2008. Trex1 prevents cell-intrinsic initiation of autoimmunity. *Cell,* 134**,** 587-98.

STORMS, M. M., KORMELINK, R., PETERS, D., VAN LENT, J. W. & GOLDBACH, R. W. 1995. The nonstructural NSm protein of tomato spotted wilt virus induces tubular structures in plant and insect cells. *Virology,* 214**,** 485-93.

SUN, S., DAI, X., AISHAN, M., WANG, X., MENG, W., FENG, C., ZHANG, F., HANG, C., HU, Z. & ZHANG, Y. 2009. Epidemiology and phylogenetic analysis of crimean-congo hemorrhagic fever viruses in xinjiang, china. *J Clin Microbiol,* 47**,** 2536-43.

SWANEPOEL, R., LEMAN, P. A., BURT, F. J., JARDINE, J., VERWOERD, D. J., CAPUA, I., BRUCKNER, G. K. & BURGER, W. P. 1998. Experimental infection of ostriches with Crimean-Congo haemorrhagic fever virus. *Epidemiol Infect,* 121**,** 427-32.

SWANEPOEL, R., SHEPHERD, A. J., LEMAN, P. A., SHEPHERD, S. P., MCGILLIVRAY, G. M., ERASMUS, M. J., SEARLE, L. A. & GILL, D. E. 1987. Epidemiologic and clinical features of Crimean-Congo hemorrhagic fever in southern Africa. *Am J Trop Med Hyg,* 36**,** 120-32.

TAKEDA, A., SUGIYAMA, K., NAGANO, H., MORI, M., KAIDO, M., MISE, K., TSUDA, S. & OKUNO, T. 2002. Identification of a novel RNA silencing suppressor, NSs protein of Tomato spotted wilt virus. *FEBS Lett,* 532**,** 75-9.

TAMURA, K., STECHER, G., PETERSON, D., FILIPSKI, A. & KUMAR, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol,* 30**,** 2725-9.

TAWAR, R. G., DUQUERROY, S., VONRHEIN, C., VARELA, P. F., DAMIER-PIOLLE, L., CASTAGNE, N., MACLELLAN, K., BEDOUELLE, H., BRICOGNE, G., BHELLA, D., ELEOUET, J. F. & REY, F. A. 2009. Crystal structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial virus. *Science,* 326**,** 1279-83.

TERWILLIGER, T. C., ADAMS, P. D., READ, R. J., MCCOY, A. J., MORIARTY, N. W., GROSSE-KUNSTLEVE, R. W., AFONINE, P. V., ZWART, P. H. & HUNG, L. W. 2009. Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta crystallographica. Section D, Biological crystallography,* 65**,** 582-601.

TESH, R. B. & MODI, G. B. 1987. Maintenance of Toscana virus in Phlebotomus perniciosus by vertical transmission. *The American journal of tropical medicine and hygiene,* 36**,** 189-93.

THOMPSON, J. D., HIGGINS, D. G. & GIBSON, T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research,* 22**,** 4673-80.

THOMPSON, W. H. & BEATY, B. J. 1978. Venereal transmission of La Crosse virus from male to female Aedes triseriatus. *The American journal of tropical medicine and hygiene,* 27**,** 187-96.

TONBAK, S., AKTAS, M., ALTAY, K., AZKUR, A. K., KALKAN, A., BOLAT, Y., DUMANLI, N. & OZDARENDELI, A. 2006. Crimean-Congo hemorrhagic fever virus: genetic analysis and tick survey in Turkey. *J Clin Microbiol,* 44**,** 4120-4.

TORREIRA, E., SCHOEHN, G., FERNANDEZ, Y., JORBA, N., RUIGROK, R. W., CUSACK, S., ORTIN, J. & LLORCA, O. 2007. Three-dimensional model for the isolated recombinant influenza virus polymerase heterotrimer. *Nucleic acids research,* 35**,** 3774-83.

UHRIG, J. F., SOELLICK, T. R., MINKE, C. J., PHILIPP, C., KELLMANN, J. W. & SCHREIER, P. H. 1999. Homotypic interaction and multimerization of nucleocapsid protein of tomato spotted wilt tospovirus: identification and characterization of two interacting domains. *Proc Natl Acad Sci U S A,* 96**,** 55-60.

ULLMAN D.E., GERMAN T.L., SHERWOOD J.L., WESTCOT, D.M. and CANTONE F.A. 1993. Tospovirus replication in insect vector cells: Immuncytochemical evidence that the nonstrcutural protein encoded by the S RNA of tomato spotted wilt tospovirus is present in thrips vector cells. *Phytopathology* 84, 147-153.

UNNI, S., HUANG, Y., HANSON, R. M., TOBIAS, M., KRISHNAN, S., LI, W. W., NIELSEN, J. E. & BAKER, N. A. 2011. Web servers and services for electrostatics calculations with APBS and PDB2PQR. *Journal of computational chemistry,* 32**,** 1488-91.

VAGIN, A. & TEPLYAKOV, A. 2010. Molecular replacement with MOLREP. *Acta crystallographica. Section D, Biological crystallography,* 66**,** 22-5.

VAN KASTEREN, P. B., BAILEY-ELKIN, B. A., JAMES, T. W., NINABER, D. K., BEUGELING, C., KHAJEHPOUR, M., SNIJDER, E. J., MARK, B. L. & KIKKERT, M. 2013. Deubiquitinase function of arterivirus papain-like protease 2 suppresses the innate immune response in infected host cells. *Proceedings of the National Academy of Sciences of the United States of America,* 110**,** E838-47.

VAN KASTEREN, P. B., BEUGELING, C., NINABER, D. K., FRIAS-STAHELI, N., VAN BOHEEMEN, S., GARCIA-SASTRE, A., SNIJDER, E. J. & KIKKERT, M. 2012. Arterivirus and nairovirus ovarian tumor domain-containing Deubiquitinases target activated RIG-I to control innate immune signaling. *Journal of virology,* 86**,** 773-85.

VAN KNIPPENBERG, I., LAMINE, M., GOLDBACH, R. & KORMELINK, R. 2005. Tomato spotted wilt virus transcriptase in vitro displays a preference for cap donors with multiple base complementarity to the viral template. *Virology,* 335**,** 122-30.

VAN POELWIJK, F., PRINS, M. & GOLDBACH, R. 1997. Completion of the impatiens necrotic spot virus genome sequence and genetic comparison of the L proteins within the family Bunyaviridae. *The Journal of general virology,* 78 (Pt 3)**,** 543-6.

VIETH, S., DROSTEN, C., LENZ, O., VINCENT, M., OMILABU, S., HASS, M., BECKER-ZIAJA, B., TER MEULEN, J., NICHOL, S. T., SCHMITZ, H. & GUNTHER, S. 2007. RT-PCR assay for detection of Lassa virus and related Old World arenaviruses targeting the L gene. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 101**,** 1253-64.

VIETH, S., TORDA, A. E., ASPER, M., SCHMITZ, H. & GUNTHER, S. 2004. Sequence analysis of L RNA of Lassa virus. *Virology,* 318**,** 153-68.

VINCENT, M. J., SANCHEZ, A. J., ERICKSON, B. R., BASAK, A., CHRETIEN, M., SEIDAH, N. G. & NICHOL, S. T. 2003. Crimean-Congo hemorrhagic fever virus glycoprotein proteolytic processing by subtilase SKI-1. *Journal of virology,* 77**,** 8640-9.

VONRHEIN, C., BLANC, E., ROVERSI, P. & BRICOGNE, G. 2007. Automated structure solution with autoSHARP. *Methods in molecular biology,* 364**,** 215-30.

VOROU, R., PIERROUTSAKOS, I. N. & MALTEZOU, H. C. 2007. Crimean-Congo hemorrhagic fever. *Curr Opin Infect Dis,* 20**,** 495-500.

WADE, R. C. 1997. 'Flu' and structure-based drug design. *Structure,* 5**,** 1139-45.

WALTER, C. T. & BARR, J. N. 2011. Recent advances in the molecular and cellular biology of bunyaviruses. *The Journal of general virology,* 92**,** 2467-84.

WALTER, C. T., BENTO, D. F., ALONSO, A. G. & BARR, J. N. 2011. Amino acid changes within the Bunyamwera virus nucleocapsid protein differentially affect the mRNA transcription and RNA replication activities of assembled ribonucleoprotein templates. *The Journal of general virology,* 92**,** 80-4.

WANG, Y., BOUDREAUX, D. M., ESTRADA, D. F., EGAN, C. W., ST JEOR, S. C. & DE GUZMAN, R. N. 2008. NMR structure of the N-terminal coiled coil domain of the Andes hantavirus nucleocapsid protein. *J Biol Chem,* 283**,** 28297-304.

WANG, Y., DUTTA, S., KARLBERG, H., DEVIGNOT, S., WEBER, F., HAO, Q., TAN, Y. J., MIRAZIMI, A. & KOTAKA, M. 2012. Structure of Crimean-Congo hemorrhagic fever virus nucleoprotein: superhelical homo-oligomers and the role of caspase-3 cleavage. *Journal of virology,* 86**,** 12294-303.

WATTS, D. M., USSERY, M. A., NASH, D. & PETERS, C. J. 1989. Inhibition of Crimean-Congo hemorrhagic fever viral infectivity yields in vitro by ribavirin. *The American journal of tropical medicine and hygiene,* 41**,** 581-5.

WEAVER, S. C. 2006. Evolutionary influences in arboviral disease. *Curr Top Microbiol Immunol,* 299**,** 285-314.

WEBER, F., BRIDGEN, A., FAZAKERLEY, J. K., STREITENFELD, H., KESSLER, N., RANDALL, R. E. & ELLIOTT, R. M. 2002. Bunyamwera bunyavirus nonstructural protein NSs counteracts the induction of alpha/beta interferon. *J Virol,* 76**,** 7949-55.

WHITEHOUSE, C. A. 2004a. Crimean-Congo hemorrhagic fever. *Antiviral Res,* 64**,** 145-60.

WHITEHOUSE, C. A. 2004b. Crimean-Congo hemorrhagic fever. *Antiviral research,* 64**,** 145-60.

WIJKAMP, I., VAN LENT, J., KORMELINK, R., GOLDBACH, R. & PETERS, D. 1993. Multiplication of tomato spotted wilt virus in its insect vector, Frankliniella occidentalis. *The Journal of general virology,* 74 (Pt 3)**,** 341-9.

WINN, M. D., BALLARD, C. C., COWTAN, K. D., DODSON, E. J., EMSLEY, P., EVANS, P. R., KEEGAN, R. M., KRISSINEL, E. B., LESLIE, A. G., MCCOY, A., MCNICHOLAS, S. J., MURSHUDOV, G. N., PANNU, N. S., POTTERTON, E. A., POWELL, H. R., READ, R. J., VAGIN, A. & WILSON, K. S. 2011. Overview of the CCP4 suite and current developments. *Acta crystallographica. Section D, Biological crystallography,* 67**,** 235-42.

WINTER, G., LOBLEY, C. M. & PRINCE, S. M. 2013. Decision making in xia2. *Acta crystallographica. Section D, Biological crystallography,* 69**,** 1260-73.

WOLFF, S., BECKER, S. & GROSETH, A. 2013. Cleavage of the Junin virus nucleoprotein serves a decoy function to inhibit the induction of apoptosis during infection. *Journal of virology,* 87**,** 224-33.

WON, S., IKEGAMI, T., PETERS, C. J. & MAKINO, S. 2007. NSm protein of Rift Valley fever virus suppresses virus-induced apoptosis. *Journal of virology,* 81**,** 13335-45.

YEN, Y. C., KONG, L. X., LEE, L., ZHANG, Y. Q., LI, F., CAI, B. J. & GAO, S. Y. 1985. Characteristics of Crimean-Congo hemorrhagic fever virus (Xinjiang strain) in China. *Am J Trop Med Hyg,* 34**,** 1179-82.

YUN, N. E. & WALKER, D. H. 2012. Pathogenesis of Lassa fever. *Viruses,* 4**,** 2031-48.

ZELLER, H. G., CORNET, J. P. & CAMICAS, J. L. 1994. Crimean-Congo haemorrhagic fever virus infection in birds: field investigations in Senegal. *Res Virol,* 145**,** 105-9.

ZHANG, Y., ZHANG, C. & LI, W. 2012. The nucleocapsid protein of an enveloped plant virus, Tomato spotted wilt virus, facilitates long-distance movement of Tobacco mosaic virus hybrids. *Virus research,* 163**,** 246-53.

ZHIRNOV, O. P., KONAKOVA, T. E., GARTEN, W. & KLENK, H. 1999. Caspase-dependent N-terminal cleavage of influenza virus nucleocapsid protein in infected cells. *Journal of virology,* 73**,** 10158-63.

ZHIRNOV, O. P. & SYRTZEV, V. V. 2009. Influenza virus pathogenicity is determined by caspase cleavage motifs located in the viral proteins. *Journal of molecular and genetic medicine : an international journal of biomedical research,* 3**,** 124-32.

ZHOU, Q., SNIPAS, S., ORTH, K., MUZIO, M., DIXIT, V. M. & SALVESEN, G. S. 1997. Target protease specificity of the viral serpin CrmA. Analysis of five caspases. *The Journal of biological chemistry,* 272**,** 7797-800.

ZHU, D. Y., ZHU, Y. Q., XIANG, Y. & WANG, D. C. 2005. Optimizing protein crystal growth through dynamic seeding. *Acta Crystallogr D Biol Crystallogr,* 61**,** 772-5.